2018

Assessing and supporting the use of fuzzy cognitive maps to simulate complex phenomena

Eric Lavin

ABSTRACT

ASSESSING AND SUPPORTING THE USE OF FUZZY COGNITIVE MAPS TO SIMULATE COMPLEX PHENOMENA

Eric Lavin, M.S.
Department of Computer Science
Northern Illinois University, 2018
Philippe J. Giabbanelli, Director

To study complex phenomena, modelers have used a wide variety of modeling techniques (e.g., agent-based modeling). While these are powerful and useful modeling methods, they are not perfect. To improve our models of complex problems, we need to apply methods that handle uncertainty, such as fuzzy cognitive maps (FCM). This thesis will show that while we use simulation models to evaluate complex phenomena, we do not apply FCMs in these evaluations very often. While they are commonly used in participatory modeling to handle uncertainty in small models, this thesis proposes and evaluates a new method to handle uncertainty in much larger models using FCMs. Specifically, we use parallel design of experiments (DoE) and demonstrate that previously published large models can be simplified. We will present a design of experiments which will extend the ability of Fuzzy Grey Cognitive Maps (FGCM) to better handle uncertainty and identify the main factors that determine the simulation output. We also explore an approximation to allow for execution within a time limit. Beyond handling uncertainty, this thesis will examine whether we need to run several simulations to compare different FCMs or if we can just compare their structures as graphs. This work is limited by the small number of FGCMs currently published. In the future, our work can be used to create a hybrid model for agent-based models with FCMs or to identify the rules necessary to create a heterogeneous population of agents.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

MAY 2018

ASSESSING AND SUPPORTING THE USE OF FUZZY COGNITIVE MAPS TO SIMULATE COMPLEX PHENOMENA

BY

ERIC LAVIN

A THESIS SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

Thesis Director:
Philippe J. Giabbanelli

ACKNOWLEDGEMENTS

This thesis would have been impossible without the support and encouragement of numerous individuals. I would like to recognize all of the people who assisted in the production of this thesis. First, I want to express my profound gratitude to my advisor, Prof. Philippe J. Giabbanelli, who provided endless support, inspiration and encouragement. I thank him for his patience and guidance, without which this thesis would not have been possible. I could not have found a better mentor to assist me in my educational and vocational endeavors.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Hamed Alhoori and Prof. Minmei Hou for their insightful comments and interest in my research which pushed it in the right direction. A special thanks to Dr. Scott Rosen whose distinct point of view expanded the perspective of my thesis.

I would also like to thank my friends and colleagues at Northern Illinois University: Magda Baniukiewicz, Venkata Sai Sriram Pillutla, Robby Ruschke, Marcus Nguyen, Nick Rosso, and Vishrant Gupta for countless enlightening discussions that inspired me to improve and make the myriad of hours in lab always enjoyable.

This thesis was supported by the College of Liberal Arts and Sciences, the Department of Computer Science, Northern Illinois University. Travel for conferences was additionally funded by the Graduate School.

DEDICATION

To my Mom, for her unfailing encouragement

To my Dad, for his unwavering support

To Angelica for her boundless patience

"To kill an error is as good a service as, and sometimes even better than, the establishing of a new truth or fact."

- CHARLES DARWIN, LETTER TO A. STEPHEN WILSON, 1879

CONTENTS

Page

List of Tables ix

List of Figures xi

Chapter

1 INTRODUCTION 1

1.1 Introduction. 1

1.2 Objectives 3

1.3 Methods 3

1.3.1 Overview 3

2 BACKGROUND 14

2.1 Constructing FCMs. 14

2.1.1 Defining FCMs 14

2.1.2 Foundations of Fuzzy Logic. 16

2.1.3 Two Methods for Creating FCMs 22

2.1.4 Software Solutions. 25

2.2 Using FCMs 29

2.2.1 Structural Analysis. 29

2.2.2 Simulating Scenarios 31

2.3 Optimizing FCMs. 34

2.3.1 Improving Convergence 34

2.3.2 Simplifying FCMs 36

Chapter Page

3 A SYSTEMATIC REVIEW OF SIMULATION MODELS OF OBESITY FOR PUBLIC HEALTH: METHODOLOGICAL ISSUES AND DIRECTIONS FOR FUTURE RESEARCH 40

3.1 Introduction. 41

3.2 Methods 44

3.2.1 Overview 44

3.2.2 Selection and Management of Articles 44

3.2.3 Model Characterization. 47

3.2.4 General Quality Assessment 50

3.2.5 Additional Quality Assessment for Agent-Based Models. 52

3.3 Results 53

3.3.1 Assessment across articles 54

3.3.2 Items not Attempted and Temporal Trends 55

3.3.3 Citation Patterns 57

3.4 Discussion 59

3.4.1 Impact of Selected Criteria on Model Quality 60

3.4.2 Question 1: Are Current Models Developed Adequately? 64

3.4.3 Question 2: Are we Improving the Quality of Models? 66

3.4.4 Question 3: Are Best Practices Shared and Re-Used? 67

3.4.5 Limitations 67

3.5 Conclusions. 69

4 ANALYZING AND SIMPLIFYING MODEL UNCERTAINTY IN FUZZY COGNITIVE MAPS 70

4.1 Introduction. 71

Chapter Page

4.2 Background . . . 73

4.2.1 The Simulation Approach: Fuzzy Cognitive Maps . . . 73

4.2.2 Factorial Design of Experiments . . . 76

4.3 Methods . . . 78

4.3.1 New Open-Source Library for Fuzzy Cognitive Maps . . . 78

4.3.2 Proposed Process to Simplify Models . . . 80

4.3.2.1 Overview . . . 80

4.3.2.2 Steps 1–3: Distributing and Running Experiments in Parallel . . . 80

4.3.2.3 Steps 4 and 5: Analyzing Experimental Results to Simplify the Model . . . 82

4.4 Experimental evaluation . . . 83

4.5 DISCUSSION . . . 87

4.6 CONCLUSION . . . 88

5 SIMPLIFYING UNCERTAINTY IN FUZZY COGNITIVE MAPS UNDER TIME LIMITATIONS . . . 90

5.1 Introduction . . . 91

5.2 Background . . . 92

5.2.1 Factorial Design . . . 92

5.2.2 Fractional Factorial Design . . . 94

5.3 Methods . . . 97

5.3.1 Determining p . . . 97

5.3.2 Acquiring Resolution IV . . . 98

5.4 Case Studies . . . 98

5.5 Discussion . . . 99

5.6 Conclusion . . . 106

Chapter Page

6 SHOULD WE SIMULATE MENTAL MODELS TO ASSESS WHETHER THEY AGREE? 107

6.1 Introduction. 108

6.2 Background. 111

6.2.1 Fuzzy Cognitive Maps 111

6.2.1.1 Why Are Fuzzy Cognitive Maps Used As Mental Models? . . . 111

6.2.1.2 How Do Fuzzy Cognitive Maps Work?. 113

6.2.2 Network Centrality. 115

6.3 Simulation setup 116

6.3.1 Dataset 116

6.3.2 Validation. 119

6.3.3 What-if Scenarios 120

6.4 Results 121

6.5 Discussion 124

6.6 Conclusion 125

7 CONCLUSION 126

7.1 Limitations 126

7.2 Future Work 128

7.2.1 Given the First Few Steps From the Run of an FCM, Can We Use Time Series Analysis to Predict the Final Step at Stabilization? 128

7.2.2 FCMs as Meta-Models for Agent-Based Models 129

7.2.3 Using FCMs as Agents for Heterogeneity in Agent-Based Models 130

Bibliography 133

LIST OF TABLES

Table		Page
1.1	Overview of Problem Settings for Three Common Designs of Experiments (DoE)	5
2.1	Examples of participatory modeling in the real world	24
2.2	Current FCM foftware	26
2.3	Schools of centrality metrics	31
3.1	Selection Criteria of the Original Reviews.	47
3.2	Our Selection Criteria to Expand on the Original Reviews.	48
3.3	Aggregate Results.	56
3.4	Aggregate Results Continued	57
4.1	Example of a Factorial Design with Eight Links.	77
4.2	Characteristics of the Three Case Studies.	84
4.3	Number and Percentage of Links that can be Set to a Single Value.	85
4.4	Three Levels of Hardware Used to Perform Experiments	86
5.1	Design for a 2^3 Factorial Design.	94
5.2	Resolution III 2^{3-1} Fractional Factorial Design, with Defining Relation $I = ABC$	95
5.3	Description of the Possible Resolutions in a Fractional Factorial Design	96
5.4	Characteristics of the Three Case Studies	99
5.5	Comparison of Significant Edges Found with a Full Factorial and Fractional Factorial Design Using a 5% Threshold to Determine the Significant Edges	100

Table Page

5.6 Number and percentage of Links that can be Set to a Single Value, Depending on the Threshold for Contributing to Variance 100

5.7 Significance of main effect of case study 1 according to a full factorial design. . . . 102

6.1 The Values of Our 19 Concepts were Set Depending on the What-If Scenario or Validation Test 118

LIST OF FIGURES

Figure Page

1.1 Snowball sampling 7

1.2 Example of an FCM after running three iterations. 9

1.3 FCM membership function displaying overlap of linguistic terms and selection of final value.. 9

1.4 An example of two different FGCMs.. 11

1.5 Ranking simulation output to centrality. 12

2.1 An element x either is (1) or is not (0) part of a crisp set f 17

2.2 Structure of a fuzzy controller. 18

2.3 **(a)** A triangular membership function. **(b)** Gaussian membership function. **(c)** A trapezoidal membership function. 20

2.4 The defuzzified results following both Mamdani and Larsen rules.. 21

2.5 Model built using MetalModeler.. 27

2.6 Scenario built in MentalModeler. Images are the sole intellectual property of Dr S. Gray. Reproduced with S. Gray's permission.. 28

2.7 Graph built from FCM WIZARD. 29

2.8 View of concept outputs from FCM WIZARD. 30

2.9 Processes for optimizing FCMs. 35

2.10 Process to build a cluster from concepts 38

2.11 Process to determine the causal relationship between clusters.. 39

3.1 Types of articles selected and relation with the original three reviews.. 46

Figure		Page
3.2	Flow Chart of Search.	48
3.3	Cumulative distribution of the number of articles (y-axis) and number of items not attempted (x-axis)..	58
3.4	Citation network.	58
4.1	General workflow of our approach	79
4.2	Average and standard deviation of time to apply transfer function either sequentially or in parallel, depending on the number of concepts.	81
4.3	Results on case studies 1 and 2 with a threshold of 5% and a subset of concepts stabilizing	85
5.1	For a model with 18 factors.	104
6.1	Aggregation of three individual FCMs into one, using a weighted average.	119
6.2	Correlation Between Centrality and Simulation Ranking for Different Scenarios and Groups..	123
6.3	Across All Scenarios and Groups of Stakeholders, only Katz Centrality had a Very High Fit.	124
6.4	Fit Across All Stakeholders in Each Group for Scenario 2	124
7.1	A concepts value during a simulation trending towards a stable value but still changing by more than ϵ.	129
7.2	The agents of grass, wolves and sheep have been made into individual concepts along with their attributes.	130
7.3	Different components that make the drinking agent heterogeneous.	131
7.4	Each concept has a range of possible values instead of the edges..	132

CHAPTER 1

INTRODUCTION

1.1 Introduction

There are numerous reasons to run simulation models, ranging from explaining and investigating the dynamics of a problem to forming predictions [68]. A more specific use for simulation is to reduce complexity. This also provides a (virtual) environment where there is no risk to test possible solutions. Testing solutions in the real world can potentially cause detrimental effects or a working selection can be found that is suboptimal. In contrast, virtual testing does not share the risk of physical detriment and can be run to find optimal solutions.

To examine and reduce complex problems we first answer the question of what it means for a problem to be complex. An important point there is to distinguish the difference between a problem that is complicated and one that is complex. A problem is considered complicated when it can be decomposed into a set of elements, that can be individually addressed piece by piece. On the other hand, complex problems result from networks of interacting causes that cannot be fixed separately to form a solution [150]. These interactions generally result in cycles (also known as loops) that cause self-feedback. While we can now differentiate between complex and complicated systems, we must define what complexity means to us. This is important because within the modeling and simulation (M&S) community there is no widely accepted definition of complexity. Such definitions include the difficulty of describing a problem, information entropy, the difficulty of creating a model, or logical depth [32]. For our case, we consider complexity as the systems which have loops and numerous possible combinations of components (i.e., combinatorial complexity [155]).

These characteristics contribute to the difficulty of managing such systems and are found across a variety of applications, as will be discussed in this thesis proposal. Note that even "small" systems with few components can be complex by this definition if they exhibit highly interconnected components with cycles and a large search space.

There are numerous examples of complex problems; one such example would be obesity [40]. Originally obesity was studied as a simple matter of energy imbalance, in which energy intake from food exceeds energy expended over time [69]. However, obesity research has shifted towards a more socio-ecological point of view that accounts for economic, cultural and political determinants for an individual's obesity [44]. Understanding that obesity is a complex problem is only the beginning, though. There are several modeling methods that researchers can apply to capture that complexity. Such methods include system dynamics (SD) [73, 84], cellular automata (CA) [21], agent-based models (ABM) [127], and fuzzy cognitive maps (FCM) [52].

FCM is perhaps the less known method to most modelers among these. FCM was introduced by Kosko [97] with two key advantages: it is easily understandable by experts in the problem domain, and it gives values to causal maps based on qualitative opinions. FCM's also specialize in dealing with the vagueness and uncertainty found in complex social problems [51, 54, 55, 151]. The uncertainty comes from not being able to find an exact value for a variable since it may exist as a distribution or a confidence interval. For example, current studies may not be able to conclude the exact strength of causation between poor body image and depression, but meta-reviews can provide a range of causal strengths. Due to these advantages, FCM's are used in a variety of fields [159] such as medical diagnosis and support [139], even in settings where the slightest inaccuracies can have very harmful consequences (e.g., radiotherapy [138] or tumor characterization [143]). The thesis is organized as follows: we state the objective of the thesis and the expected results. From there we discuss the proposed methods to complete the thesis.

1.2 Objectives

The focus of this thesis is to assess how and whether FCM's are being applied in complex problems and developing novel M&S techniques. We focus on the complex problem of obesity in an observable sense, such as social norms and public messaging. This goal will be accomplished through the completion of three specific aims:

1. Evaluating whether FCM's are being applied in complex problems, particularly obesity. We developed an evaluation framework based on recommendations and best practices in modeling and simulation methods applied to public health (Chapter 3).
2. Simplifying FCMs for problems with large uncertainty. We designed and implemented methods to reduce the search space (Chapters 4 & 5).
3. Evaluating whether structural analyses are sufficient instead of performing extensive simulations for selected tasks. We examine the graph structure of FCMs from stakeholders with differing levels of expertise in a problem (Chapter 6).

1.3 Methods

1.3.1 Overview

This thesis is motivated by the need to improve the use of FCMs in modeling complex problems. In our first aim, we conducted a literature review of simulation models for obesity since existing reviews which follow a public health perspective, such as the use of the models in policy and potential pitfalls. On the other hand, we approached from a simulation perspective to assess model quality. As mentioned before, we are interested in studying problems that have uncertainty

and loops which FCMs excel at dealing with. Obesity was chosen for this study due to being well known for its complexity, especially containing those features [53]. We developed a set of best practices (detailed in Aim 1) for obesity modeling based on guidelines in M&S research as well as characteristics specific to obesity. These practices will include assessing previous models on aspects such as the inclusion of time frame, social interactions, heterogeneity. Starting with previous reviews [105, 125, 166], we evaluated previously developed models to track improvement over time. In our second aim, we combine an extension of FCMs known as fuzzy grey cognitive maps (FGCM) with efficient design of experiments (DoE) techniques [50] to identify the edges of an FGCM that can be simplified. Like an FCM, FGCMs are directed networks where the edges correspond to the strength on the causation between factors. However, the FGCM as proposed by Salmeron [159] gives the edges a range of possible values to better cope with uncertainty. That is, FGCMs can represent scenarios with high uncertainty, but it becomes difficult to simulate them as the search space becomes massive. Using a proper DoE, we identify the importance of each link, and we can thus simplify some of the ranges, in turn creating a much smaller search space in which to perform simulations. By looking at Table 1.1 we can observe the three families of DoE that were considered. For our work, we believe that a factorial design would be the best fit, since we can treat each edge as a factor and know their min and max values regardless of the range distribution.

For the third aim, we take a radical standpoint by questioning whether simulations are needed at all for certain tasks. Specifically, in participatory decision making where participants create mental models in the form of FCMs, the question is whether they agree on what should be done regarding the problem. Theoretically, two FCMs can have very similar structure, but minute differences in weights (of edges) or initial states (of nodes) can lead to very different outcomes. Thus, to know whether two participants have a same view, we are once again faced with a massive search space, in which their two mental models would have to be simulated extensively and the output compared. Instead, we examined on real-world data whether some structural/network metrics are sufficient to conclude that their worldview concurs, thus saving the need for extensive simulations.

Table 1.1: Overview of Problem Settings for Three Common Designs of Experiments (DoE)

Design of Experiment	Application Setting
Factorial Design	Experiment consists of 2 or more factors each with a discrete number of possible values. Test all possible combinations of factors and values [120].
Randomized Complete Block Design	Primarily used in agriculture where experimental units are grouped into replicates. Used to control variation in an experiment by accounting for spatial effects [120].
Latin Squares	Organizes experiment into a grid with equal rows and columns. Treatment to each experimental unit is applied on each row and column [120].

Aim 1: Evaluating whether FCMs are being applied in complex problems, particularly obesity

The aim is to evaluate the current models being used to simulate obesity. First, we compiled a series of best practices for modeling obesity. Then, we applied them on our corpus. The corpus (i.e., set of simulation models for obesity) starts with models from the only three reviews of health models and noncommunicable diseases [105, 125, 166]. The snowball sampling from these three reviews is justified by the wide search criteria, which would be difficult to improve upon, employed by the reviews as shown in Box 1.

Box 1) Selection criteria of original reviews

- [166] Examples from several teams of the National Collaborative Childhood Obesity Research Envision network, including statistical, social network, agent-based, and system dynamics.
- [105] Reviewed models focusing on health and economic consequences of obesity, future rates of obesity by historical trends, models relating dietary and physical activity to obesity, and models of specific interventions and policies.
- [125] Took papers published in the English language between January 2003 and July 2014. Only selected freely available studies of actual applications (not illustrations) of ABM for public health programs. Used following sets of keywords: (1) agent-based model, agent-based modeling or individual-based modeling; (2) noncommunicable diseases, noninfectious diseases, nontransmissible diseases, or chronic diseases; (3) heart disease, diabetes, stroke, cancer, chronic respiratory disease, or neurodegenerative diseases; and (4) unhealthy diet, physical activity, exercise, smoking, alcohol, hyperlipidemia, high cholesterol, high triglyceride, overweight, of obesity.

This initial corpus will be expanded through the steps shown in Box 2.

Box 2) Our selection criteria for models

1. The technique of snowball sampling (Figure 1.1) is applied up to distance 3 to find simulation models.
2. Reduce sample of models to only simulation models using our set of criteria. Thus, we do not include mathematical and statistical based models, data mining studies, frameworks for potential obesity models.

Snowball sampling (Figure 1.1) is done such that we find all papers that cite the initial reviews (distance 1), then all papers that cite those papers (distance 2), and so forth.

Collecting the models is just the first step. We then develop a framework on how to evaluate public health models from a simulation perspective. Through evaluation of the reviews and models, I will focus on finding key criteria in simulation models to assess the improvement of models over time in a simulation perspective. With this framework, we can focus on the improvement of simulation methods in public health studies.

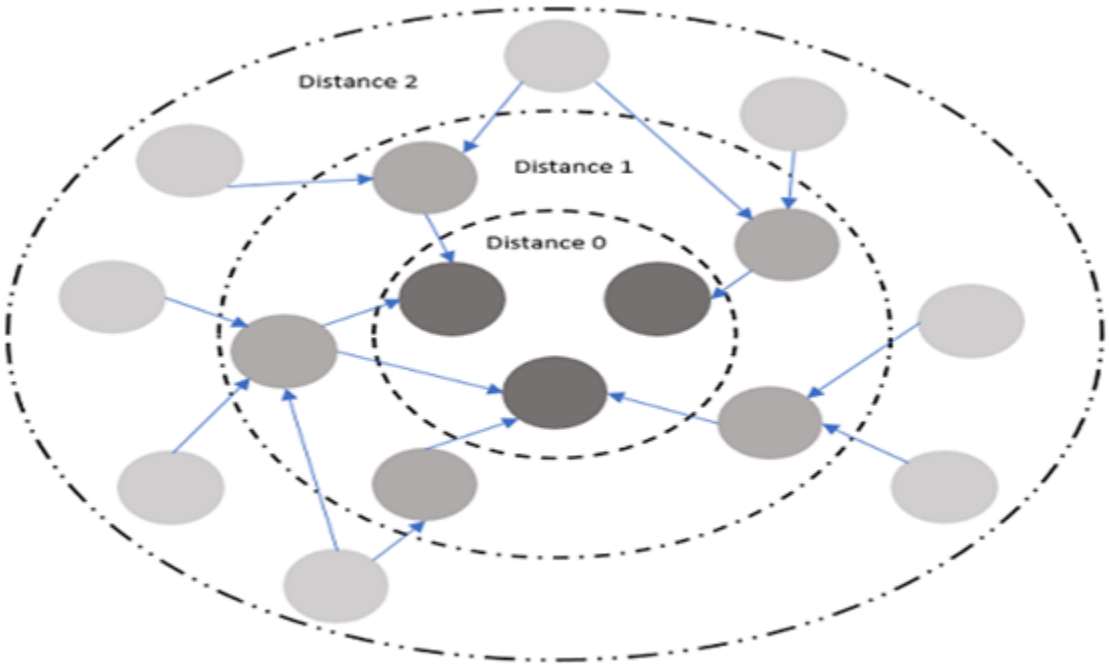

Figure 1.1: Each circle represents a paper that would be grabbed by the sample. The dashed lines represent the layers of the "snowball", that is, the ones in the very center would be the initial review (distance 0). One level up shows all sample taken at distance 1 which means they cite the original article as shown by the directed edges.

Aim 2: Simplifying FCMs for problems with large uncertainty

Our aim here is to better deal with uncertainty in complex problems and effectively utilize parallelization in simulation optimization [156]. To achieve this aim we make use of FCMs be-

cause they are a modeling technique designed to cope with uncertainty. An FCM models behavior through three key constructs:

1. Nodes, representing concepts of the system such as states or entities. Nodes have a weight in the range [0, 1] indicating the extent to which the concept is present at a simulation step.
2. Weighted directed links, representing causal relationships. Their weight is from the range [-1, 1], where negative weights indicate that increases in the source node cause a decrease in the target node. Conversely, positive weights indicate that increases in the source node cause an increase in the target node.
3. An inference function, which updates the value of each node based on the weights of both the links going into it and the nodes that these links connect to.

Formally, the number of nodes is denoted by n. The weights of the directed links can be represented as an $n \times n$ adjacency matrix A, where A_{ij} is the weight of the link from *i* to *j*. The value of each concept at step *t* of the simulation is represented by $V_i(t), i = 1 \dots n$. At each step of the simulation, these values are updated using the following standard equation:

$$V_i(t+1) = f\left(V_i(t) + \sum_{j=1, j\neq i} V_j(t) \times A_{j,i}\right) \tag{1.1}$$

where f is a clipping function (also known as the transfer function) ensuring that the values of nodes remain in the [0, 1] range. For example, in an ecological model, a node could stand for the density of fish in each space, where 0 means no fish and 1 means a maximal density. That value cannot go beyond 1 since it is maximal, and the density cannot be negative either. The clipping function must be monotonic (to preserve the order of nodes' values) and it is recommended to use a sigmoidal function when modeling planning scenarios [177]. We employed the widely used hyperbolic tangent [52, 111] as sigmoidal functions keep saturation levels of nodes within the range of [-1,1] [111]. Figure 1.2 provides examples of updating an FCM by repeatedly applying equation 1.1.

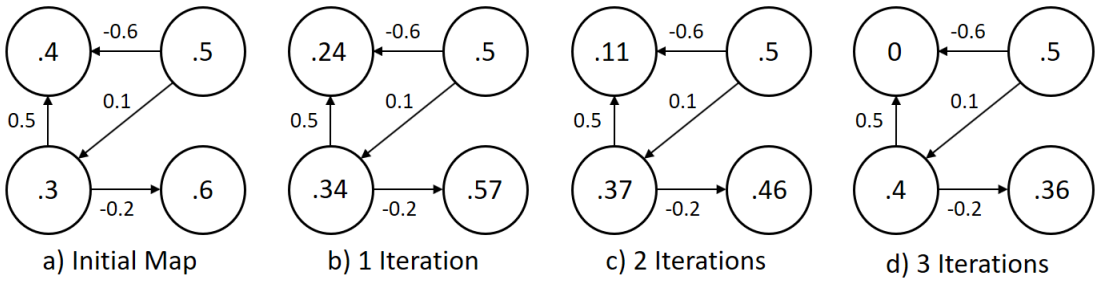

Figure 1.2: Example of an FCM after running three iterations.

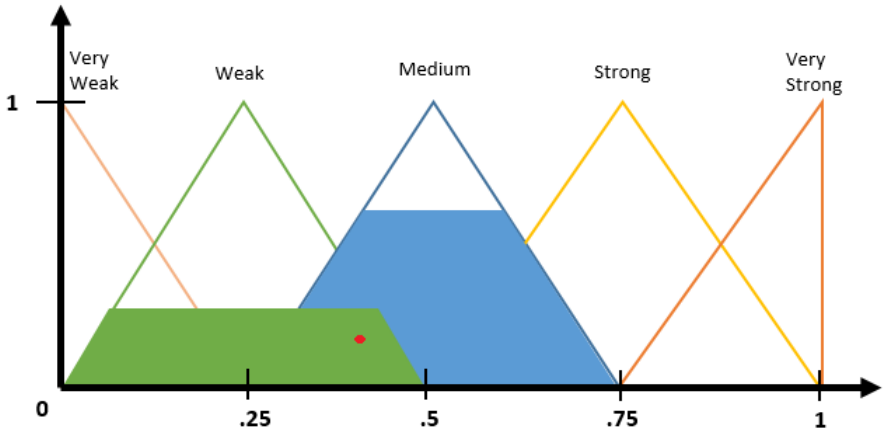

Figure 1.3: FCM membership function displaying overlap of linguistic terms and selection of final value.

As mentioned above, an FCM has weighted causal links. The values on these links are created by fuzzy logic. For this, experts give linguistic terms (e.g., weak, strong, very strong) for the causal relationship between nodes. It must also be considered that what one individual views as medium may be high or low for others. To deal with this problem, membership functions overlap as shown in Figure 1.3 below. The shapes which can be altered are then filled up proportionally to the number of responses given, such as in our example where two-thirds said medium and one-third said strong. The final edge value is the centroid of the filled area represented here by a red dot.

An FCM does not include the concept of time; its time steps do not map to physical time (although there is an extension to remedy this limitation). Consequently, an FCM does not run for a time period. Rather, Equation 1.1 is applied until a chosen subset of nodes reaches a stable value. This subset is determined by the application context. For example, in a previous FCM for obesity, they were interested in the long-term trends for obesity and required a single concept to stabilize to stop iterating [52]. Other concepts such as "food intake" or "weight discrimination" did not have to stabilize to answer the question, would the level of obesity change in reaction to a new intervention? [52]. Formally, consider a subset $S \subseteq V$ needs to stabilize. Then the simulation will end when:

$$|V_i(t+1) - V_i(t)| \leq \epsilon, \forall i \in S \tag{1.2}$$

where ϵ is set to a very small positive value. For simulation packages, an additional condition is set for a maximum number of steps in case Equation 1.2 is never satisfied. While this is rarely necessary in practice it must be accounted for. FCMs design to cope with uncertainty has made them a popular choice in participatory modeling [67, 126]. The causal strength between concepts is determined by participants who evaluate the connection. However, participants may not agree or may view the concepts in different ways. As mentioned before, Salmeron [159] proposed FGCMs to give a range of possible values to the causal links. This creates the problem a large search space of possible edge values. In Figure 1.4 below you can notice that the edges of those FGCMs have a range instead of a singular value as shown in Figure 1.2.

An FGCM may assign uncertainty across too many links. Fixing the less important links would thus make our model more tractable. There are several approaches to assess the importance of different parameters (causal links) in a simulation model [160]. At one extreme, one may perform a simple sensitivity analysis which varies one parameter at a time while others are fixed at typical values. This is statistically inefficient and does not account for interactions. At the other extreme,

one can generate all possible combinations of parameter values, but exhaustively exploring this search space may be infeasible. We articulate the model with no user preference, meaning to optimize the model with no user input [157]. A good Design of Experiment (DoE) thus provides information about the contribution of parameters and their interactions, in a way that is feasible given the resource requirements (e.g. computation time). It is commonly used to determine the relative importance of parameters in a performance study [50].

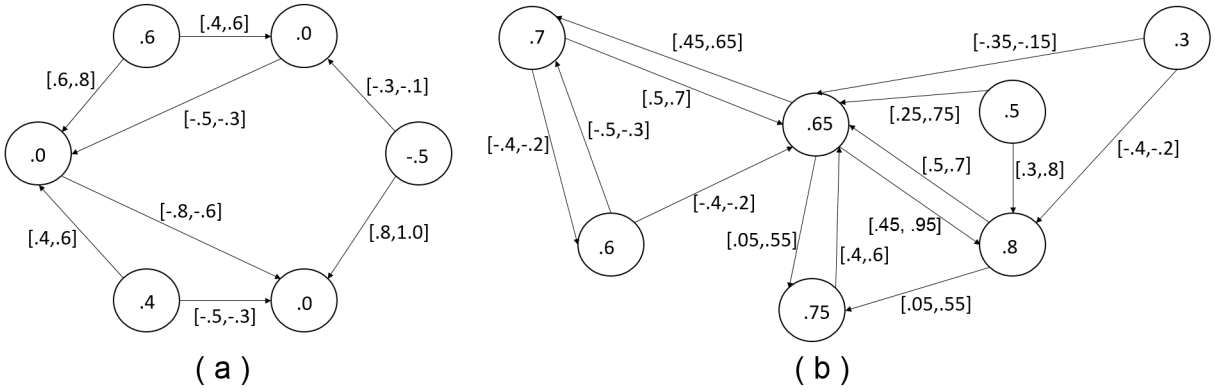

Figure 1.4: An example of two different FGCMs. Each edge has a range of possible values rather than a set value(a) is an FCM for a security system [160], (b) is an FCM for a supervisor in radiotherapy treatment planning [138]

Aim 3: Evaluating whether structural analyses are sufficient instead of performing extensive simulations for selected tasks

In this aim we examine the ability to compare separate individuals FCMs. In this problem, we focus on an ecological fishing case study in which FCMs were used. In this case study, there are four separate groups of people: water managers, fishermen, club managers, and experts. With present knowledge, even if we know that separate groups agree on the concepts of a problem, there is no guarantee that they agree on the links or that their final conclusions will be the same. Thus, we examine if individuals agree on the causal maps structure then they will agree on the outcome. For an FCM, this is no guarantee since the strength of the causal links may vary greatly between individuals. We compare the mental models using centrality, which finds the most important (cen-

tral) concepts in a network [124]. We test that if concepts rank similarly in centrality then they will agree on the final output. We do this by correlating the ranking of a node's centrality with the ranking of its final output as shown by Figure 1.5. To study this correlation, we investigated the FCMs gathered for the previously mentioned case study by our collaborators, Drs Gray and Arlinghaus [64], who agreed to share the data for this thesis.

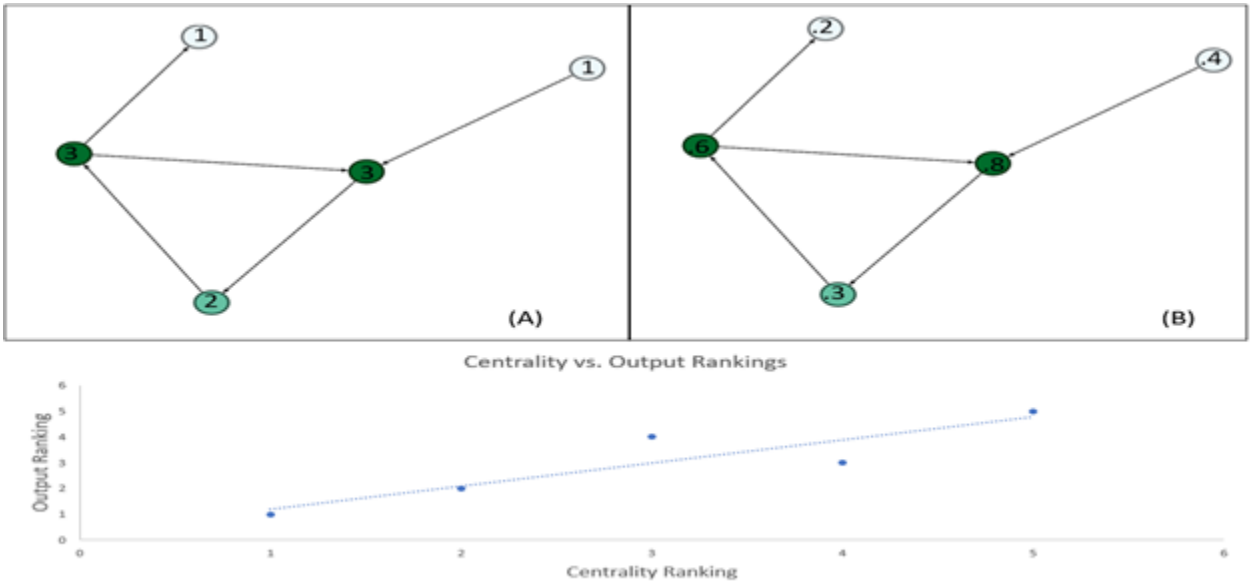

Figure 1.5: (a) Displays the degree centrality for each node of an FCM. (b) Shows the simulation output. (c) Ranking simulation output to centrality the graph shows the correlation between the rank of centrality (more central higher rank) and ranking of output (higher output, higher rank).

For each group, we constructed a composite map, and using that composite map conducted a variety of centrality measures, such as Katz centrality, degree centrality, betweenness centrality, load centrality, closeness centrality, and eigenvector centrality. It is important to remember that each map is a graph G, which is composed of a set of vertices V connected by edges E. The degree, $d(v)$, of a vertex $v \in V$ is the number of edges that include that vertex. Since our maps are directed graphs (e.g., has directed edges), we are concerned with the out-degree $d^{+}(v)$ and the in-degree $d^{-}(v)$. The out-degree is the number of edges with v as the origin, and the in-degree is the number of edges with v as a destination. Degree centrality for a directed graph can

measure the number of in-degree, the out-degree, or both. However, degree centrality is one of the simplest centrality measures. There are more advanced ones, such as betweenness centrality [157]. Betweenness centrality is a centrality measure reliant on shortest paths. The shortest path is a path from vertices u to v that traverses the smallest number of vertices between them. The betweenness centrality of a vertex $v \in V$ is found by taking the sum of the number of shortest paths that v is on between two vertexes, normalized by the total number of shortest paths between two vertexes. Formally we present that as

$$g(v) = \sum_{s \neq v, t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1.3}$$

where σ_{st} is the total number of shortest paths from s to node t, and $\sigma_{st}(v)$ is the number of those paths that pass-through v. We correlate the rankings of concepts centrality with their simulation output. This will tell us that if the individuals agree on the structural importance of the concepts, they are likely to agree on what happens to them through interactions. If so, then we can then confirm that the structures agree and run fewer simulations since they would agree on the outcome. This can be taken a step further by taking a composite map created by all stakeholders and comparing the structure and outcome of the composite map compared to the individuals.

CHAPTER 2

BACKGROUND

2.1 Constructing FCMs

2.1.1 Defining FCMs

Fuzzy cognitive maps are a modeling and simulation method which builds on causal maps representing the mental models of individuals or groups. First introduced by Kosko [97] in 1986, FCMs have two key advantages:

1. They are easily understandable by experts in the problem domain.
2. They give values to causal maps based on qualitative opinions.

An FCM is a modeling technique that is designed to cope with uncertainty. This is accomplished through three key constructs:

- Nodes, representing concepts of the system such as states or entities. Nodes have a weight in the range [0, 1] indicating the extent to which the concept is present at a simulation step.
- Weighted directed links, representing causal relationships. Their weight is from the range [-1, 1], where negative weights indicate that increases in the source node cause a decrease in the target node. Conversely, positive weights indicate that increases in the source node cause an increase in the target node.

- An inference function, which updates the value of each node based on the weights of both the links going into it and the nodes that these links connect to.

Formally the number of nodes is denoted by *N*. The weights of the directed links can be represented as an $N \times N$ adjacency matrix *A* where A_{ij} is the weight of the link from *i* to *j*. The value of each concept at step *t* of the simulation is represented by $V_i(t), i = 1 \ldots$ N. At each step of the simulation, these values are updated using Equation (1.1), where *f* is a clipping function (also known as the *transfer* function) ensuring that the values of nodes remain in the [0, 1] range. For example, in an ecological model, a node could stand for the density of fish in each space, where 0 means no fish and 1 means a maximal density. That value cannot go beyond 1 since it is maximal, and the density cannot be negative either. The clipping function must be monotonic (to preserve the order of nodes values) and it is recommended to use a sigmoidal function when modeling planning scenarios [177]. We employed the widely used hyperbolic tangent [52, 111] as sigmoidal functions to keep saturation levels of nodes within the range of [0,1] [177]. Figure 1.2 provides examples of updating an FCM by repeatedly applying equation (1.1).

An FCM does not include the concept of time; time steps do not map to physical time (although there is an extension to remedy this limitation). Consequently, an FCM does not run for a time period. Rather, Equation (1.1) is applied until a chosen subset of nodes reaches a stable value. This subset is determined by the application context. For example, in a previous FCM for obesity, they were interested in the long-term trends for obesity and required a single concept to stabilize to stop iterating [52]. Other concepts such as "food intake" or "weight discrimination" did not have to stabilize to answer the question. How would the level of obesity change in reaction to a new intervention? [52]. Formally, consider a subset $S \subseteq V$ that needs to stabilize. Then the simulation will end when Equation (1.2) is met, where ϵ is set to a very small positive value (e.g., 0.01). For simulation packages, an additional condition is set for a maximum number of steps in case

equation (1.2) is never satisfied. While this is rarely necessary in practice, it must be accounted for.

2.1.2 Foundations of Fuzzy Logic

FCMs are designed to cope with uncertainty, which has made them a popular choice for participatory modeling [65, 126]. In participatory modeling, causal strength between concepts is determined by participants who evaluate the connection. However, participants may not agree or may view the concepts in different ways. Fuzzy logic is used to give quantifiable values to qualitative variables. We accomplish this through the use of fuzzy sets. To understand fuzzy sets we must first understand what a crisp set is. For a crisp set, whether an element is in the set or not is a binary answer. Either it is or it is not (Definition 2.1.1) [99].

Definition 2.1.1. Crisp Set.

In a crisp set, whether the element is in the set is binary: it is (1), or it is not (0). For a concept A, the membership function $\mu_A(x)$ is:

$$\mu_A(x) = 1 \text{ if } x \in A$$

$$\mu_A(x) = 0 \text{ if } x \notin A$$

However, within society we are used to dealing with a large amount of vagueness or uncertainty. An example such as, "I'll turn the heat on when it gets cold," does not clearly state what qualifies as cold or does not. As such we need a way to represent this concept of vagueness. This is done through the application of fuzzy logic and fuzzy sets. Fuzzy sets unlike crisp sets allow us to show how much x is in a set through a range of [0,1] [176] (Figure 2.1.2).

Definition 2.1.2. Fuzzy Set.

A fuzzy set is an extension of the crisp set by allowing partial membership, so $\mu(x)$ is a function that puts values in range [0,1]. This value is the degree to which x belongs to A [176].

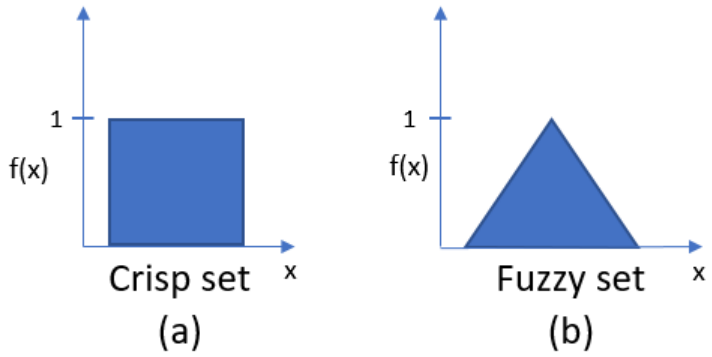

Figure 2.1: An element x either is (1) or is not (0) part of a crisp set f (a). Fuzzy Logic extends this by allowing different levels, as exemplified by a triangle function (b).

From fuzzy sets we focused on fuzzy control, which is a knowledge-based control approach that makes effective use of all information related to a system (e.g., sensors, measurements of key features, and experts) [176]. Fuzzy controllers follow the basic structure of "if-then" rules. Fuzzy control systems are useful under the following situations [176]:

- There are experienced humans (i.e., experts) who can satisfactorily provide qualitative control rules in terms of vague and fuzzy sentences.
- Applications where there is large uncertainty or unknown variation in parameters and structures.

Fuzzy control is primarily focused on model-based methods. Fuzzy control for models allows mathematical ways to ensure performance and robust analysis. However, as these methods are more mathematically involved, they are less transparent while ensuring proper performance. Overall, a fuzzy controller involves three general structures: the database, the rule base, and the processing unit (which is composed of three parts) [176] (Figure 2.2):

1. Fuzzification: Defines the membership function.
2. Inference engine: Applies fuzzy rules to meaning from the qualitative input.
3. Defuzzification: Extracts quantitative value from the fuzzy set.

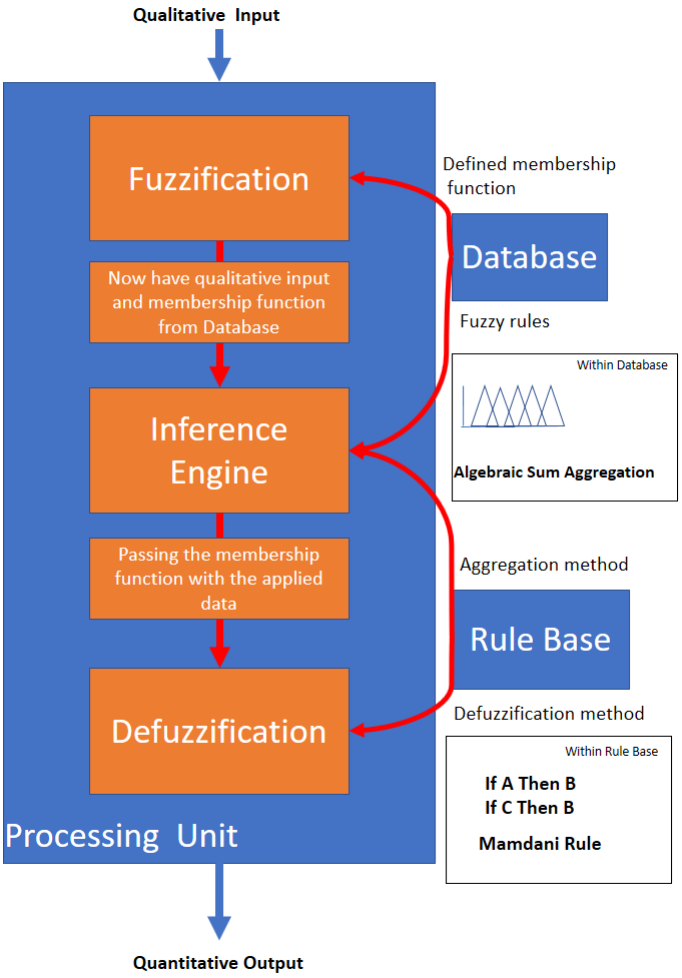

Figure 2.2: Structure of a fuzzy controller. Processing unit is shown as composition of three rules.

The process starts with qualitative input. The database holds information on the membership function such as the number, shape, and distribution of the fuzzy sets along with the meaning of the linguistic terms and if-then rules stating their connections. Information on the membership function is applied in the fuzzification step to create the membership function. The membership function and input are then sent to the inference engine. The inference engine applies the meaning of the qualitative measurements to the membership function. The meaning appropriately assigns the input to the membership functions by the supplied if-then rules. From there we begin the defuzzification process. This applies the rule set that determines in what manner we get the quantitative output. These steps will be further detailed throughout this section in relation to our temperature example mentioned earlier.

It is important not to confuse the database with a relational database. The database in our case contains the configuration of the fuzzy controller such as the number, shape and distribution of the membership function [176]. The rule base contains the set of rules that determine how the controller behaves. Finally, the processing unit exists for the execution of the rules. The execution of the rules is primarily handled by the inference engine.

To clarify this process, we will walk through an example. We want to determine the strength of the relationship between the temperature and the number of people in a store. The first step of the process is the fuzzification through parameters determined in the database. This is done by membership functions that map each point from the input to the range of $[0, 1]$ [184]. A membership function can be defined by a variety of shapes such as triangles, trapezoids, gaussian curves and many others (Figure 2.3).

However, just having the membership function is only part of the solution. The chosen membership function (we used a trianglular membership function for our example) and qualitative input (very low, low, medium, high, very high) are sent to the inference engine. The inference engine will apply rules to the qualitative input to place them into the membership function. That is, if one input translates to medium, then it is also partly high and partly low due to overlap between

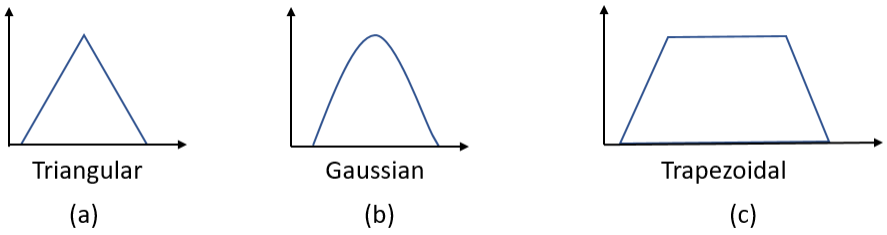

Figure 2.3: **(a)** A triangular membership function. **(b)** Gaussian membership function. **(c)** A trapezoidal membership function.

membership functions. The inference accomplishes this process by combining the rule base with logic. Since fuzzy logic is an extension of binary logic, it contains the logical operators for AND, and OR. Assume we have two fuzzy sets *A* and *B*, the logical operators are defined for fuzzy sets as follows [184].

Definition 2.1.3. AND Operator

The AND operator defines the intersection between two fuzzy sets $A(x)$ and $B(x)$. Its result is the set

$$C(x) = A(x) \wedge B(x) = min(A(x), B(x)), \forall x \in X \tag{2.1}$$

Definition 2.1.4. OR Operator

The OR operator defines the union between two fuzzy sets $A(x)$ and $B(x)$. Its result is the set

$$C(x) = A(x) \vee B(x) = max(A(x), B(x)), \forall x \in X \tag{2.2}$$

Following the inference we now must apply our rule base for defuzzification. Rule bases are a set of if-then rules that define the problem. So if an individual believes that temperature has a medium impact on the number of customers in the store, then we see an increase in the medium member of the function as well as the intersection with the high and low sections. If this were done with three individuals and two said there is a medium relationship and one said a high relationship between temperature and the number of customers, then we would have $\frac{2}{3}$ of the medium member

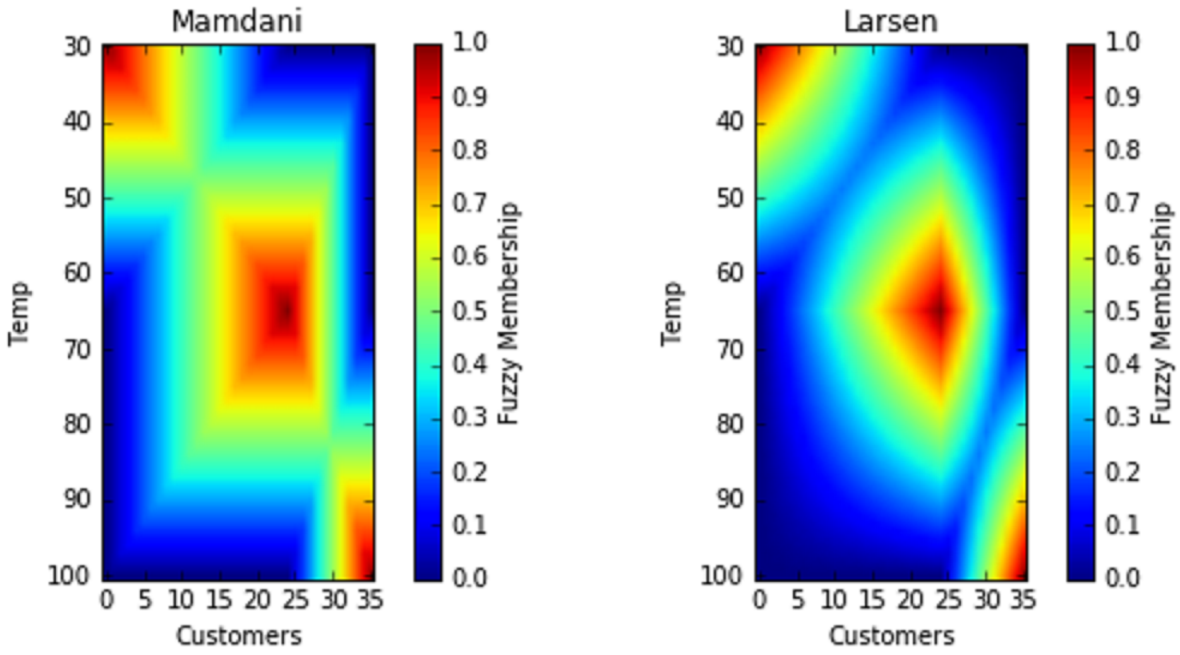

Figure 2.4: The defuzzified results following both Mamdani and Larsen rules. The defuzzified values are similar in all areas if not the same.

and $\frac{1}{3}$ of the weak member (Figure 1.3). While rule sets are dependent on the problem, they all boil down to the base of, "If A, Then B " [184]. Two of the more common inference methods for rule sets are:

1. **Mamdani**: $R_M(x, y) = A(x) \wedge B(x)$ (takes the min.)

2. **Larsen**: $R_L(x, y) = A(x) \times B(y)$ (takes the product.)

It is worth noting that while the inference methods can provide slightly different outcomes with the end result, they are generally quite similar (Figure 2.4). If there is a reason to suspect the inference method greatly affects the final value then the rules can be tested using several inference methods [51, 52].

Finally after applying our inference, we reach the final part of processing, defuzzification. We need to combine the membership function into a single geometric shape. The process is called *aggregation*. In general, aggregation will remove all parts of the membership function not in use

and keep the parts that have been filled. There are several families of aggregations that have been developed [194]. These include the *family Algebraic Sum* $f(x, y) = x + y - x \times y$ and the *family Hamacher Sum* $f(x, y) = (x + y - 2 \times x \times y)/(1 - x \times y)$ [52]. Once the data has been aggregated we want to defuzzify it, meaning to get a singular value by mapping the fuzzy set to a crisp set. One such method is the *center of gravity* (centroid), which is the average of all points in our aggregated shape (or the center of mass). The centroid of the shape is then calculated with the x coordinate being used as the defuzzified value (Figure 1.3). There are several other methods for the defuzzification of a membership function as well [180].

To clarify this process we applied a basic abstract example of a fuzzy controller where we have three experts provide fuzzy responses for the connection between two separate but related entities. First, experts give vague linguistic terms (e.g., weak, strong, very strong) for the causal relationship between entities. Our triangle membership function is then filled up proportionally to the number of responses given, such as in our example where two-thirds said medium and one-third said strong. The final quantifiable output value (found through defuzzification) is the centroid of the filled area represented here by a red dot.

2.1.3 Two Methods for Creating FCMs

We have discussed how FCMs are defined and how we quantify the qualitative input. The next question is, how do we determine the structure of the FCM such as concepts and relationships, and how strong those relationships are? There are two methods:

1. Participant driven
2. Data driven

Participant driven, also known as *participatory modeling*, integrates a range of views from different stakeholders in a problem [90]. For this, researchers gather stakeholder groups to deter-

mine what concepts are relevant in a problem and their relationships [62, 81]. The best practices in collecting data for FCM for participatory modeling involve interviews and/or questionnaires [133].

When gathering data from participants, Özesmi and Özesmi suggest the following practice [133]. Have a facilitator working with individuals or groups of individuals in the map construction. Participants are shown an unrelated FCM example. From there, the participants create a list of important concepts within their problem. Participants are then asked to draw links between concepts they believe are connected and to give quantitative edge values for the strength of that relationship (in range $[-1, 1]$). Maps built through participatory modeling can be analyzed to represent an individual's view of a problem, or aggregated, to represent the collective knowledge of stakeholder groups [133]. Aggregation can be done by:

- Combining all edges and nodes from individual maps into a singular large map, or
- Having the group of stakeholders work together to create a singular map.

Participatory modeling has been used for years to see how stakeholders understand problems and can be done with a varying number of participants (Table 2.1).

The second expert-based method applies questionnaires and can involve extraction from text depending on methodology. The basic practice is described by Axelrod [7]. When using questionnaires, the first goal is to identify the concepts in a problem. There are a few ways to do this:

1. Text extraction
2. Preliminary subject expert questionnaire
3. Expert meeting

Text mining can be used to extract important concepts from peer-reviewed sources and other documents. This is feasible since more information is available online and in a format that can be mined. The original methods are the next two, which require experts in the discipline. The experts complete a preliminary survey that will identify the primary concepts in a problem. If a survey is not possible, or many experts are physically available, researchers can hold a meeting with the

Table 2.1: Examples of participatory modeling in the real world

# of Participants	# of Edges(Average)	# of Nodes(Average)	Year	Reference
51	43	24	2003	[132]
8	9	17	2010	[181]
59	28	18	2012	[116]
20	27	24	2010	[128]
30	43	24	2001	[195]
31	28	19	1999	[129]
13	43	24	2001	[130]
35	43	24	2001	[131]
56	31	24	2002	[23]
44	31	25	2001	[22]
11	28	16	2013	[147]
29	145	26	2012	[136]
12	19	16	2009	[87]
15	63	23	2015	[81]
27	27	117	2012	[62]
7	26	15	2012	[52]
5	44	24	2014	[51]

experts to identify the primary concepts [7].

Once the primary concepts have been identified, a survey to determine the causal relationships between concepts is created. This survey asks for the relationship between ordered pairs of concepts in a questionnaire format. Additionally, the survey can ask participants to compare the relative strength of relationships. For example, there could be a question on whether the relationship $V_1 \rightarrow V_2$ is stronger than $V_3 \rightarrow V_4$. This allows for less bias and more accurate estimations of causal weights [148]. These questions can apply linguistic variables that can later be quantified by fuzzy logic [51]. For this relationship survey, Axelrod [7] suggests having not only experts of the discipline but also "lay" experts. These experts are knowledgeable or concerned about the problem but not defined from a technical discipline.

The final case for creating FCMs is to infer the causal relationships from data. This can be done through the application of machine learning. Learning approaches for FCMs concentrate

on learning the connection matrix (i.e., causal relationships, the edges and their weights) [140]. The data is gathered by expert intervention and/or historical records. The learning algorithms are classified into three types on the basis of their learning paradigm (type of knowledge used). They are [140]:

- *Hebbian based*: Unsupervised learning methods that use historical data and a learning formula to iteratively adjust FCM weights. These include the balanced differential algorithm [85] and the nonlinear Hebbian learning [137] methods.
- *Population based*: Unsupervised methods that train FCMs to mimic the input data. These methods are computationally expensive and include methods such as particle swarm optimization [145] and evolutionary algorithms [98].
- *Hybrid approaches*: An approach that implements both Hebbian-population-based methods. The Hebbian methods are used to approximate the edge matrix while the population-based methods refine these results such as with the nonlinear Hebbian learning differential evolution [135].

The application of these methods can automate the creation of FCMs. It can be noticed that most of these algorithms focus on the weight matrix, as that is the part of the model with the most uncertainty. Furthermore, the population-based methods are primarily optimization methods. As such, they are also applied when optimizing FCM weight matrices, which is covered in more detail in Section 2.3.1

2.1.4 Software Solutions

For the expanded use of FCMs packages in research, it is natural that sets of tools have been developed. Currently there are three main softwares that can be used: (i) MentalModeler, (2) FCM

Table 2.2: Current FCM foftware

Criteria	**Mental Modeler**	**FCM TOOLs**	**FCM WIZARD**	**Our Tool**
Formatted input file	Yes. Accepts .mmp files.	No.	Yes. Accepts .fcm files.	Yes. Formatted text file.
Formatted output file	Yes. Export .mmp files.	Yes. Exports excel file.	Yes. .fcm files. and .png maps	No.
Graphical view of FCM	Yes.	No.	Yes	Yes.
GUI	Yes.	Yes.	Yes	No.
Parallel processing supported	No.	No.	No	Yes. Can run a single FCM in parallel or several FCMs
Observes change in concept values over time	No.	Yes.	Yes	Yes.
Requirements for use	Installed software with creators permission	MATLAB 7.9	Installed software with creator permission	Python 2

TOOL, (3) FCM Wizard. We detail each of these tools and compare them with the library we created for creating and simulating FCMs (Table 2.2, detailed in Chapter 4)

MentalModeler was introduced in 2013 [63]. It was created with a focus on participatory modeling. MentalModeler was created with three purposes in mind: (1) construct a designed qualitative conceptual model (Figure 2.5), (2) develop scenarios and evaluate system change under plausible conditions (Figure 2.6), and (3) revise the model based on output. MentalModeler was developed using Java and has a GUI consisting of a window and tabs for the user. Modelers can import a model using their file format (.mmp), open an already-existing model they used on that machine, or create a new model. Users can add concepts using a "plus sign" on the screen. Relationships can be drawn between concepts by dragging from a tab on the bottom of one concept to the concept it influences. Once a relationship is established, a slider is used to set the quantitative value. Fur-

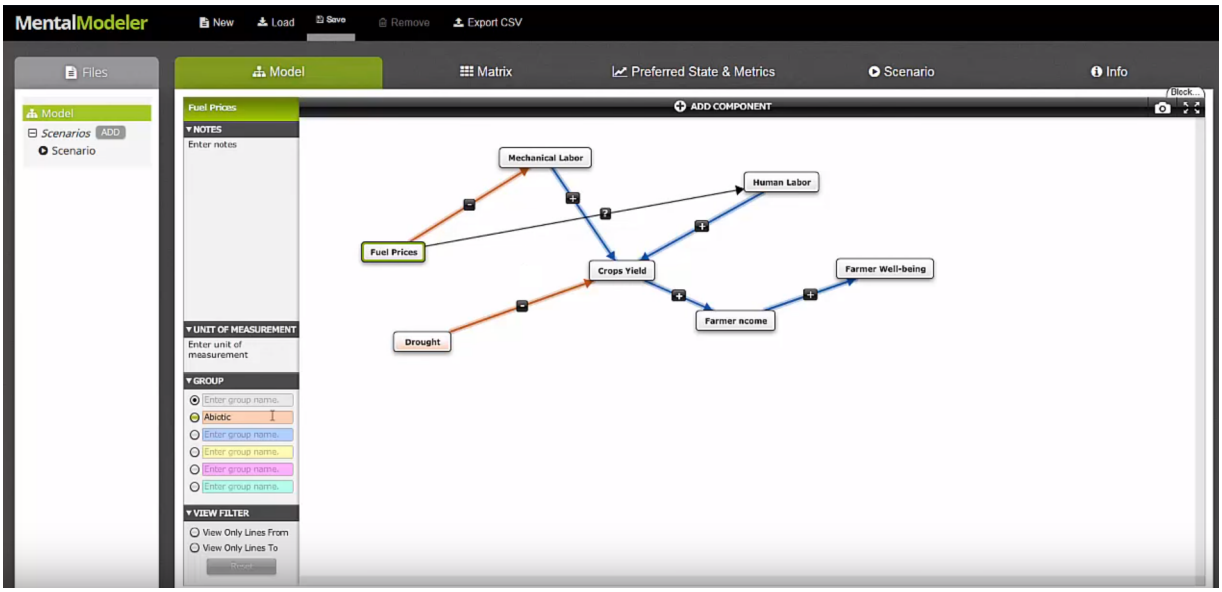

Figure 2.5: Model built using MetalModeler. Images are the sole intellectual property of Dr S. Gray. Reproduced with S. Gray's permission.

thermore, the user can switch tabs to see the matrix form of the map. Users are also able to view structural components of the map such as the centrality, density, and total number of components and edges. A scenario tab allows the user to make a variety of scenarios and have a graphical view of the change in concepts values.

FCM TOOLs was introduced in 2010 [144]. FCM TOOL is designed for building FCMs using MATLAB 7.9. It incorporates Microsoft Excel for the storage and retrieval of data from files. Through this tool users can either create FCMs from scratch or adjust already-existing FCMs by modifying their parameters. FCM TOOL implements a GUI by applying a window and a tabbed menu navigation scheme. Parameters that can be adjusted are the model's concepts, their initial values, and the weights of their relations. After simulation runs, the final concept values are displayed and plotted as a graph. Users can choose to plot only a subset of concepts as well. Furthermore, there is a section to write concluding remarks and comments about a simulation. Within the window, a table is also shown that displays the settings of the simulation. Finally,

Figure 2.6: Scenario built in MentalModeler. Images are the sole intellectual property of Dr S. Gray. Reproduced with S. Gray's permission.

simulation runs are kept so the user can run a variety of simulations in one window. The desired simulations can then be saved in an Excel format file.

FCM WIZARD was introduced in 2016 and is particularly popular among modelers focused on methods, while MentalModeler is focused on participants. Adding nodes and edges is easily accomplished by selecting buttons on the main page. Concepts can be selected and customized by color (highlight different nodes in different colors) or assigned a label and initial value. Similarly, edges can be selected to set their causal value in a pop-up box. FCMs can also be opened and exported in their custom file format (.fcm). Moreover, an image of a graph can be exported as a .png file (Figure 2.7). They also provide an average operator to aggregate maps together. FCM WIZARD provides simulation capability supporting several activation rules and and provides several transformation functions that can be further customized. The value of concepts throughout the simulation can be shown in a table as well as graphically (Figure 2.8). Most importantly though, FCM WIZARD allows for the use of learning algorithms, both supervised and unsupervised, to

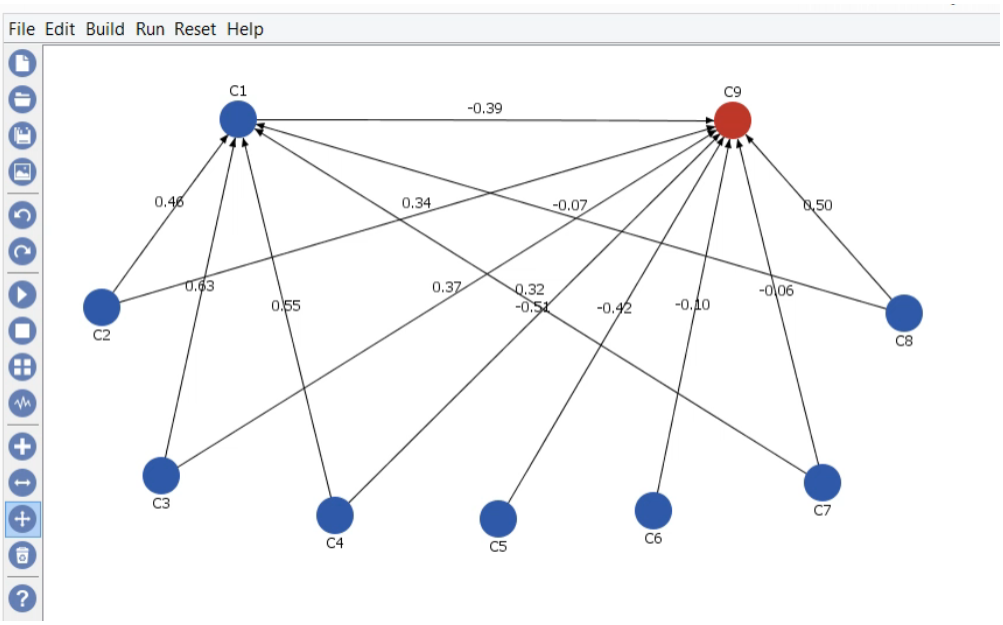

Figure 2.7: Graph built from FCM WIZARD.

create and adjust FCMs (compute causal relations, optimize network topology, and improve convergence). These algorithms take as input a file in the Attribute Relation File Format (.arff).

2.2 Using FCMs

2.2.1 Structural Analysis

FCMs are at heart weighted directed networks that do not allow self loops. This means that we are able to evaluate FCMs from a network perspective as well as a simulation perspective. That is to say, we can measure aspects of the model's structure [64].

By observing the number of concepts and edges we can get a rough idea of how complex a model is. However, that is basic structural information that does not provide a deeper understanding of the model.

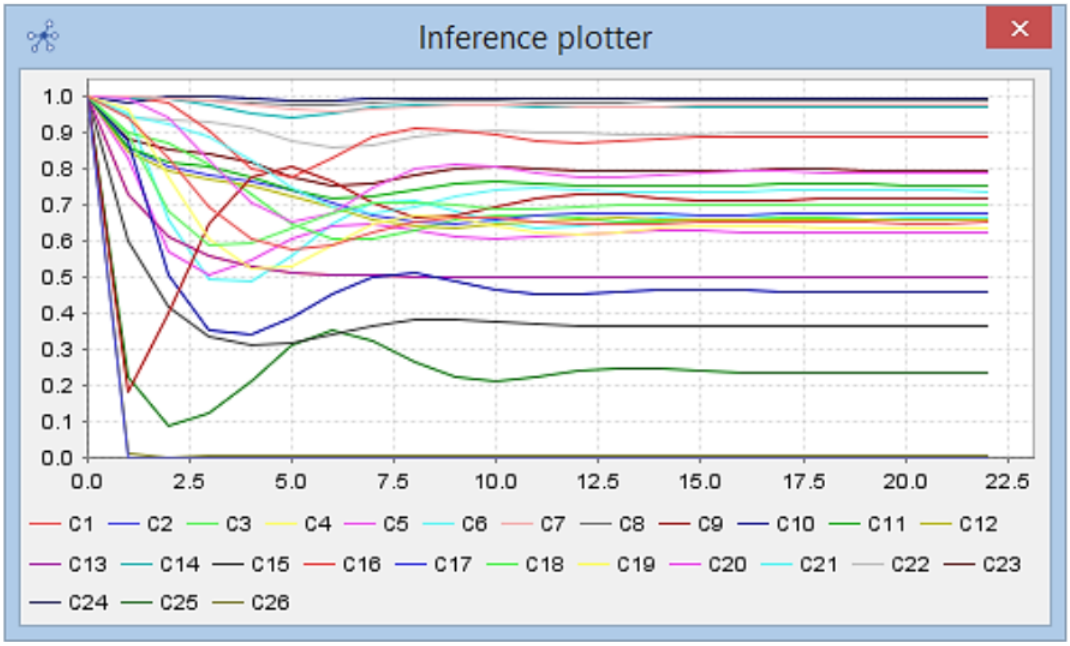

Figure 2.8: View of concept outputs from FCM WIZARD.

Complexity, though, can be measured by the amount of *receivers* (sinks, Definition 2.2.2) and *transmitters* (sources, Definition 2.2.1) [36].

Definition 2.2.1. Transmitter.

Components in a system that affects other system components but are not effected by others. [36]

Definition 2.2.2. Receiver.

Components in a system that are affected by other components; however, these components do not affect other parts of the system [36].

From these two types of concepts, we are able to compute the *complexity* by taking the ratio of receivers to transmitters: $(\frac{\text{number of receivers}}{\text{number of transmitters}})$. This measures the degree to which driving forces such as transmitters are considered. A higher complexity is indicative of more complex systems thinking [36].

Table 2.3: Schools of centrality metrics

Type of Centrality measurement	**Example**	**Definition**
Flow indice	Betweenness centrality	Fractions of shortest paths find the amount of flow through a concept.
Reachability indice	Closeness centrality	Shortest paths measure concepts ability to reach other concepts.
Feedback indice	Katz centrality	Concepts accumulate influence by concepts feeding into them.
Local indice	Degree centrality	Measures centrality based on only the concepts neighbors.

Definition 2.2.3. Density.

The fraction of edges present compared to the total number of possible, or how connected a graph is. It can be calculated by

$$\frac{|E|}{|V|(|V|-1)}$$

where $|E|$ is the number of edges and $|V|$ is the number of concepts [124].

A higher *density* means structural areas can change due to the connectivity of the model [133]. We can further measure the *centrality* of concepts. The centrality of a concept is the relative importance of the concept in relation to the centrality measurement. There are several measurement types of centrality that observe different aspects of a graph. Therefore, when measuring centrality it is important to know what aspect of centrality is important for the graph (e.g., does it influence several nodes, is it a connection between two clusters). There are a few different types of centrality measurements and common metrics for those types (Table 2.3) [124].

2.2.2 Simulating Scenarios

Scenario techniques have been found to be useful when uncertainty and complexity are present [164]. As we have discussed already, FCMs have a large amount of uncertainty as the causal relationships

are created by experts, and we use them on complex problems. This makes scenario planning an excellent choice for simulating FCMs. Simulating scenarios allow us to simulate strategic thinking by simulating multiple possible futures [3]. Scenarios are built by "telling a story." This more directly means scenarios are "a set of hypothetical events set in the future constructed to clarify a possible chain of causal events as well as their decision points" [91]. Thus, we can say that scenarios are a simulation of possible future events that we can view as a chain of causal events. Moreover, scenario planning techniques are frequently used to better articulate mental models about the future [3].

There is no singular approach to scenario planning [3]. There are currently three schools of thought on scenario development [3]:

1. *Intuitive Logics* school.
2. *Probabilistic Modified Trends* school.
3. *French* school (La Prospective).

The Intuitive Logics school takes a non-mathematical approach. This method relies on knowledge and credibility to create the scenario. They follow a basic methodology that ranges from 5 to 15 steps. However, the proposed methodology by the Stanford Research Institute International is the most popular [3]. While this school of thought creates flexible and internally consistent scenarios, it is dependent on the planners and is easily biased [3].

The Probabilistic Modified Trends school incorporates two separate matrix-based methodologies: trend impact analysis and cross-impact analysis [14]. These techniques extrapolate trends and perform a probabilistic modification on them. The trend impact analysis relies on historic data extrapolation without considering the effects of unprecedented future events [14], whereas cross-impact analysis captures the interrelationships between key influencing factors to avoid fore-

casting events in isolation [14]. By applying this methodology, individuals are able to account for historical data in creating scenarios.

La Prospective methodology follows the belief that the future is not a predetermined temporal continuity. Instead, it can be deliberately created and modeled [3]. This method of scenario construction calls upon four essential concepts: the base, the external context, the progression, and the images [35]. The base is created by taking a thorough analysis of the present. The environment surrounding the model, such as the social, economic, and political context, creates the external context. The progression is the historical simulation derived from the base and the external context. The reality at the time for the future in the scenario is called the image [35].

For example, consider an FCM concerning fishing in a lake. Following the Intuitive Logics school, we would have an expert determine the scenario. As such they would define the density of fish expected in the water, the amount of fishermen and other conditions involving the lake. The concepts would be initialized to those values and we would simulate from there. For this scenario, it is the environment envisioned by the expert, thus any misconceptions they have about the problem would be reflected in the scenario. For the same example following the Probabilistic Modified Trends school, we would search for historical data on the amount of fishermen, the population of fish and any other concepts considered in the model. Using statistical methods we would develop likely distributions for each concept and build the scenarios off those distributions. This way we can simulate based on historical data if it is available. Finally, using the La Prospective school of thought we would examine the context and present situation of the lake we are simulating, such as, Are there laws restricting the number of fishermen? Is it a season when we are likely to have a lot of fish? Are they still reproducing? Using these we define the environment of the lake we wish to simulate and match it to the historical progression of the lake over time. This can give a reliable simulation to many scenarios due to the flexibility in defining the scenario.

What-if scenarios provide not only the simulations we run to see how modifying a concept can alter the whole system but also provide validation for the model. It has been found that a

reliable way to validate FCMs is to create a what-if scenario in which there is no doubt in the final outcome [49]. These scenarios can be seen as reliable by following the La Prospective thinking and once we have confirmed the validity of the model. The other scenario-building methodologies can be applied depending on the existence of historical data and reliable experts to create viable scenarios.

2.3 Optimizing FCMs

2.3.1 Improving Convergence

A weakness of FCMs is the large variance on the causal relationship between concepts given by experts. Machine learning methods can be used to optimize the weight matrix for getting outputs within a specified range. Population-based methods learn from historical data to optimize the weight matrix. This learning is done through the use of an objective function [140]. There are numerous population-based methods used for FCMs which include but are not limited to particle swarm optimization (PSO) and evolutionary algorithms [140].

Particle swarm optimization is a stochastic optimization algorithm that belongs to the swarm intelligence family of algorithms. These algorithms use a population of individuals to examine promising regions of the search space. The entire population is called a *swarm*, and the individuals are called *particles*. Each particle will move at an adaptable velocity in its search space and keep a memory of its optimal position within that search space [145].

For PSO we have the concepts of an FCM $V_1, \ldots, V_N$ and $V_{out_1}, \ldots, V_{out_m}$ be the output concepts with $1 \leq m \leq N$. All other concepts are considered input or interior concepts [145]. The goal is to limit the output into a range:

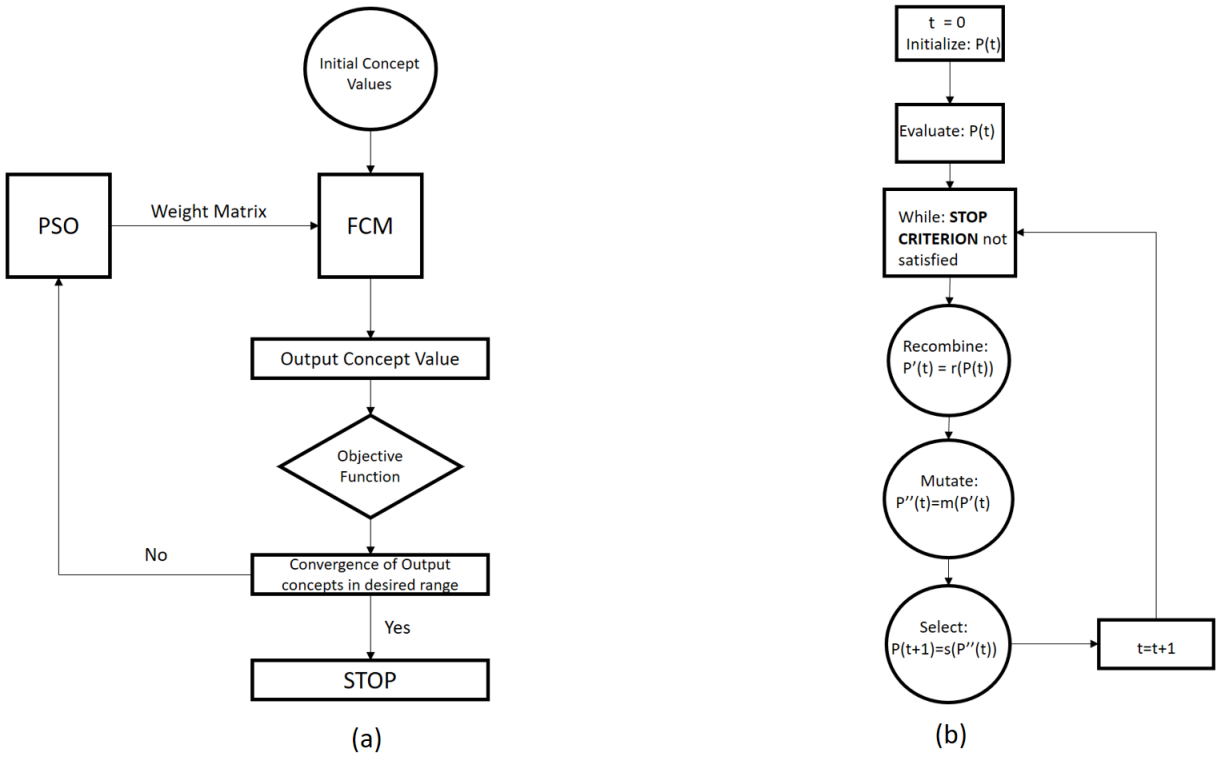

Figure 2.9: (a) Processes for optimizing FCMs. Flow for optimizing weight matrix using PSO. (b)Flow for optimization using evolutionary algorithms.

$$V_{out_i}^{min} \leq V_{out_i} \leq V_{out_i}^{max}, i = 1, \ldots, m$$

where the range of the output concept is determined by experts. The weight matrix is adjusted to keep the output concepts in range. The weights are adjusted using the objective function until the selected concepts end in the desired range while still maintaining the meaning of the edges [145] (Figure 2.9 (a)). This method produces a plethora of weight matrices that lead to the FCMs' convergence in the desired concept ranges. These solutions can be statistically analyzed and examined by experts to make sure the chosen output(s) is within the constraints of a problem.

Beyond the particle swarm, evolutionary algorithms (EA) have been applied to optimize the weight matrix of FCMs. An evolutionary algorithm will imitate the process of natural evolution to find a solution to complicated optimization problems [98]. The general structure of an EA is to make slight alterations to the present weight matrix to get new weight matrices (the next generation) and compare the results of the new matrices. The most accurate solution is then used to produce a new generation. This is done until a matrix is created that is sufficiently accurate according to some stop criteria. In Figure 2.9 (b), functions r and m transform the matrix to give us more suitable offspring to choose from. We then apply the function s to select the better model between the two. The process is then repeated using the newly selected model.

2.3.2 Simplifying FCMs

FCMs are commonly used in complex problems. With complex problems it is difficult to reduce the complexity while maintaining the accuracy. There are, however, recent efforts to simplify FCMs to reduce their complexity. Complexity is dependent on the number of concepts and the relationships in a model. Thus, to reduce the complexity of a model we can either reduce the relationships or reduce the number of concepts. Presently, methods exist to reduce the number of

concepts in a model [79, 80, 141], and one of our contributions explored in Chapter 4 is to simplify relationships.

These studies reduce the number of concepts by clustering similar concepts together [141]. Concepts have two properties when considering similarity [141]:

1. *Reflexive*: Concepts are similar to themselves, so concept V_i is similar to V_i.
2. *Symmetric*: If concept V_i is similar to concept V_j, then V_j is similar to V_i.

It is important to note, however, concepts do not follow the transitive property. That is, if concept V_i is similar to concept V_j, and V_j is similar to concept V_k, then it does not necessarily mean that V_i is similar to V_k [141].

This method starts by constructing the clusters of concepts. A cluster is built for each concept in the FCM. Figure 2.10 depicts the process for building a cluster from a single concept. When a new concept is considered for the cluster, it must be considered similar to every element in the cluster. This similarity is determined by a function that will compute a similarity value if that value is less than ϵ, with $0 \leq \epsilon \leq 1$. One such example of an *isNear* function is given in Equation (2.3) [141].

$$isNear(i, j, \epsilon, K) = \frac{\sum (w_{ik} - w_{jk})^2{}_{k \neq i, k \neq j, k \notin K} + \sum (w_{ki} - w_{kj})^2{}_{k \neq i, k \neq j, k \notin K}}{8m} \tag{2.3}$$

If the new concept is found to be similar to all concepts in the cluster, then the new concept is added to the cluster. It is worth noting that, if $\epsilon = 0$, then no elements will merge into clusters, and if $\epsilon = 1$, you will end with two distinct clusters [141]. A good epsilon is determined by trial and error and expert analysis.

Once we have reduced the number of concepts by clustering the similar ones, we still need to account for the relationships and their weights between concepts in the clusters. This is done through aggregating weights between clusters. For example, if we have two clusters a & b, we take the average of the weighted relationship for their concepts and set that to be the new relationship

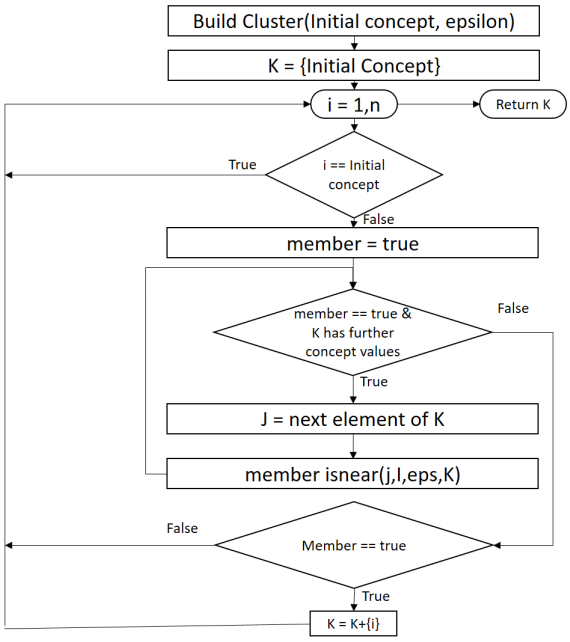

Figure 2.10: Process to build a cluster from concepts. *Initial concept* is the original concept in its own cluster, and *epsilon* is the threshold for how similar concepts need to be to be clustered. The *isNear* function determines if the two concepts should be clustered.

weight between the clusters [141] (Figure 2.11). This can cause problems such as self loops which are not permitted in FCMs if one were to follow Kosko's model.

The goal of this concept reduction is to lessen the complexity of the model to allow more transparency and comprehensibility by non-subject experts. It has been found that even by reducing the number of concepts through clustering, the reduced FCM performs to give a similar output to the original model [141]. While the merging has given successful results, there is still work to be done to select the reduction method in the future.

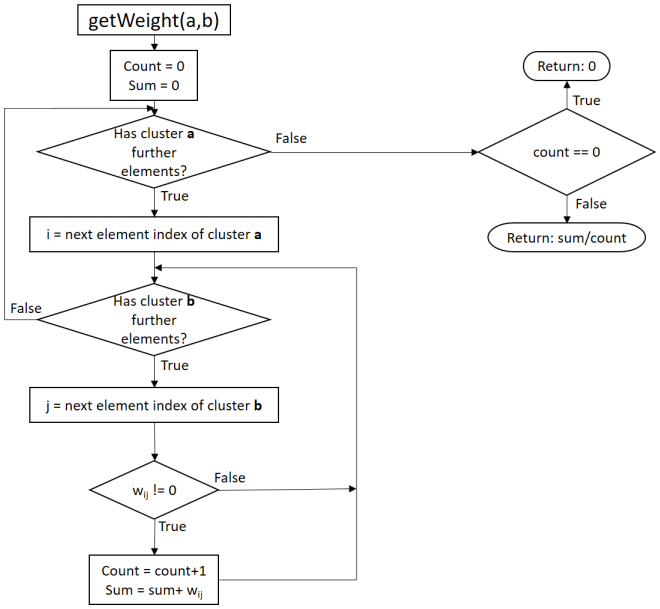

Figure 2.11: Process to determine the causal relationship between clusters. *a* and *b* are clusters of concepts to be merged.

CHAPTER 3

A SYSTEMATIC REVIEW OF SIMULATION MODELS OF OBESITY FOR PUBLIC HEALTH: METHODOLOGICAL ISSUES AND DIRECTIONS FOR FUTURE RESEARCH

In the previous chapter we discussed the creation and applications of FCMs. It was shown that FCMs are an effective modeling technique for complex problems, in particular with participatory modeling. However, does this mean that FCMs are currently being applied in complex problems such as public health? In this chapter, we present a literature review of public health simulation models to detect whether or not FCMs are being used to model this complex problem. Beyond that, the literature review examines the current modeling techniques used in public health modeling and criteria that can be used as guidelines to complete simulation models with more methodological rigor.

My contributions consisted of:

(i) Performing the search for simulation models.

(ii) Developing the guidelines and evaluation of the simulation models based on those guidelines.

(iii) My co-authors assisted in the search for simulation models and development of the guidelines as well as analysis of the citation network.

3.1 Introduction

The ongoing obesity crisis continues to be a huge health concern and an economic burden. If current rising trends continue, 2.7 billion adults will be overweight by 2025 [192]. The increase from 2.0 billion overweight adults in 2014 highlights the difficulty of tackling obesity [192]. This difficulty stems in part from the complexity of obesity, as it is part of a system of heterogeneous elements interacting with one another and adapting to changing circumstances [31, 40]. Since obesity is a complex problem without a definite solution for sustainable weight loss, simulation models can provide a much-needed decision support tool. *Simulation models* are broadly defined as the imitation of a system [155], which would include a concept map as well as an agent-based model or a system dynamics model. Our focus is on *dynamic* simulation models, in which the imitation of a system takes place through time [155]. For instance, the values of virtual entities are updated as the simulation progresses. By modeling the complex system of which obesity is a part of, it becomes possible to systematically search for interventions with high potential. This helps apply systems thinking to the management of obesity in a safe manner, since interventions are only tested in the virtual world of the model rather than by directly impacting the population [38]. Obtaining a comprehensive view of obesity through a model also helps with evaluating an intervention, since we then know what variables are related to the intervention and should be monitored [20]. It is thus clear that, from a public health standpoint, the possibilities offered by models are not merely to represent data but to guide future policy on obesity by highlighting where to focus the next actions and data collection efforts [169].

Different modeling approaches provide insights on different facets of a problem's complexity. Badham proposed three categories for models [8]: qualitative aggregate models, quantitative aggregate models (also called *macro-simulation*), and quantitative individual models (also called *micro-simulation*). Qualitative modeling techniques are "modeling techniques used for under-

standing an issue, problem structuring, integrating perspectives and communicating the system structure" [8]. The Foresight Obesity Map is a qualitative model articulating how weight-related factors are connected, and it prompted conversations to understand who was responsible for which factors and how to achieve better coordination [182]. These types of maps continue to be developed, with a recent example being the diagram of obesity causes from Allender *et al.* [2]. System dynamics (SD) models, which belong to the category of quantitative aggregate models, have also been developed to help policymakers. They are a typical example of quantitative aggregate model since "quantitative aggregate models use equations to describe relationships between the averages of pairs (or larger groups) of system components" [8]. For example, these models can be used as a virtual platform to test population health approaches [183] or can be a focal point to develop systems thinking capacity with regards to policies [118]. Fuzzy cognitive maps (FCM), which are also quantitative aggregate models, were created to help practitioners navigate the complexity of obesity in their patients [52]. In the last category, "individual-oriented modelling techniques track individuals and calculate results by counting the relevant individuals" [8]. Among individual oriented models, a large number of network-based and agent based models (ABM) have been developed and used to suggest policy interventions; in particular, we refer the reader to the work of Giabbanelli [49, 51], Shoham [166, 167] and Yang [196, 197]. We note that cellular automata (CA) also provide individual oriented models but have been used much less commonly for obesity research.

Our focus is on dynamic simulation models, which are either quantitative aggregate models or quantitative individual models. A simulation model can compute how a system can *change*, for example, under hypothetical scenarios of interest to policymakers [38]. To illustrate this, policymakers may be interested in a "what-if" question such as: What if we were to implement zoning regulations? A simulation model for this situation would be able to compute key metrics (e.g., physical activity, prevalence of obesity) in reaction to selected zoning regulations. In contrast, a concept map is still a model of the problem but cannot compute values in reaction to hypothet-

ical scenarios. The 2010 review by Levy *et al.* introduced the use of modeling to the obesity community and focused on its potential contribution to policy [105]. Two additional reviews were produced in 2015. The review by Shoham *et al.* focused on modeling social norms [166], since many models have been devoted to capturing peer effects on weight dynamics following the work of Christakis and Fowler [19]. While the review by Nianogo and Onyebuchi was interested in non-communicable diseases more generally, it devoted a significant portion to models for obesity [125]. All three reviews catered to a health audience. The rationale for our review differs from the past three reviews due to our focus on the technical quality of the models. Our review seeks to address the following three specific questions:

1. Are current models developed adequately from a simulation viewpoint and given the specific needs of obesity modeling?

2. Are we improving the quality of models, with recent ones satisfying more needs than older ones?

3. Is there sufficient cross-pollination of ideas in the field to ensure that best practices are shared and re-used?

While other reviews continue to be published, we note that they still focus on creating policy-relevant models [107], in contrast with our unique positioning in examining models with respect to technical soundness. Our main contribution is thus to provide the first technical assessment of simulation models in obesity. This assessment can be used to build the foundations for the next generation of models, thus improving the quality of the evidence base for policies regarding obesity.

The chapter is organized as follows. We start by explaining how articles were selected and examined in section 3.2. In particular, we define and state how we assessed whether each model dealt with our selected standard simulation questions (e.g., calibration, validation) and obesity-specific

needs (e.g., heterogeneity). Then in Section 3.3, we provide the results of our assessment across articles as well as over time (to check whether there is a change in model quality). Section 3.4 is discussion, which starts by examining the importance of each selected feature for the quality of a simulation model. This contextualizes our findings, as knowing the importance of a feature tells us about the consequences for not including it in a model. Finally, in Section 3.5, we offer a brief conclusion on the importance of developing methodologically sound models and supporting cross-pollination of modeling practices going forward.

3.2 Methods

3.2.1 Overview

In this section, we start by explaining how our dataset of articles was obtained and detail the application of our exclusion criteria. Then, we summarize how three types of data were extracted from the articles. We first categorized models based on aspects such as the stated purpose of the model and the intervention level at which their policies operate. We note that categories are neither right or wrong, but simply state the type of model. Then we define and apply criteria for simulation models, both in general as well as for agent-based models specifically.

3.2.2 Selection and Management of Articles

Our corpus consists of articles reporting the development of simulation models for obesity. We assembled the corpus by first extracting articles about the models discussed in the three previous reviews [105, 125, 166]. These three reviews on health models and noncommunicable diseases were selected due to being the only three of their kind at the time of the study. We used these three

reviews as a start and expanded using snowball sampling to update the evidence base used in these reviews. Intuitively, we found all papers that cite the initial reviews and then papers that cite those papers to a certain distance. Figure 3.1 illustrates the connections between models in snowball sampling.

Snowball sampling is not a probabilistic method; that is, it does not recruit a random sample) [158]. We use it to expand on the three previous reviews, which used a variety of search strategies as shown in Table 3.1. That is, we build on the rigor and exhaustive approach of these reviews and update them using snowball sampling. After performing the snowball sampling up to distance 3, we applied our exclusion criteria (Table 3.2) which resulted in narrowing the sample of articles from n=60 to n=33 (Figure 3.2).

We found 20 articles that either focused on other chronic diseases, did not provide models for public health (e.g.,, [72]), or did not qualify as simulation models per the criteria used here. For example, statistical methods composed of regressions were removed [42, 189]. We note that the models preserved may have *involved* non-simulation techniques such as regressions (e.g., to provide initial data to an agent-based model or analyze its output [199]), but the model itself had to be a simulation model. There were three articles removed which did not provide models themselves but rather guidelines about modeling [92, 95, 172]. For example, the article by Karanfil *et al.* proposed a paradigm to build policy-relevant models of obesity by integrating aggregate and individual techniques [92]. Finally, we removed four articles that applied a quantitative individual technique (network modeling) but performed data mining rather than building a simulation model [26, 47, 48, 103]. For example, Leahey *et al.* collected the network of participants undergoing obesity treatment and searched for associations [103] between network features (i.e., weight status of peers, level of obesogenic influence among peers) and individual data (e.g., weight loss, baseline weight). In line with reporting guidelines for PRISMA-P 2015 [119], we also note that only articles written in English were considered, although none was excluded for language reasons.

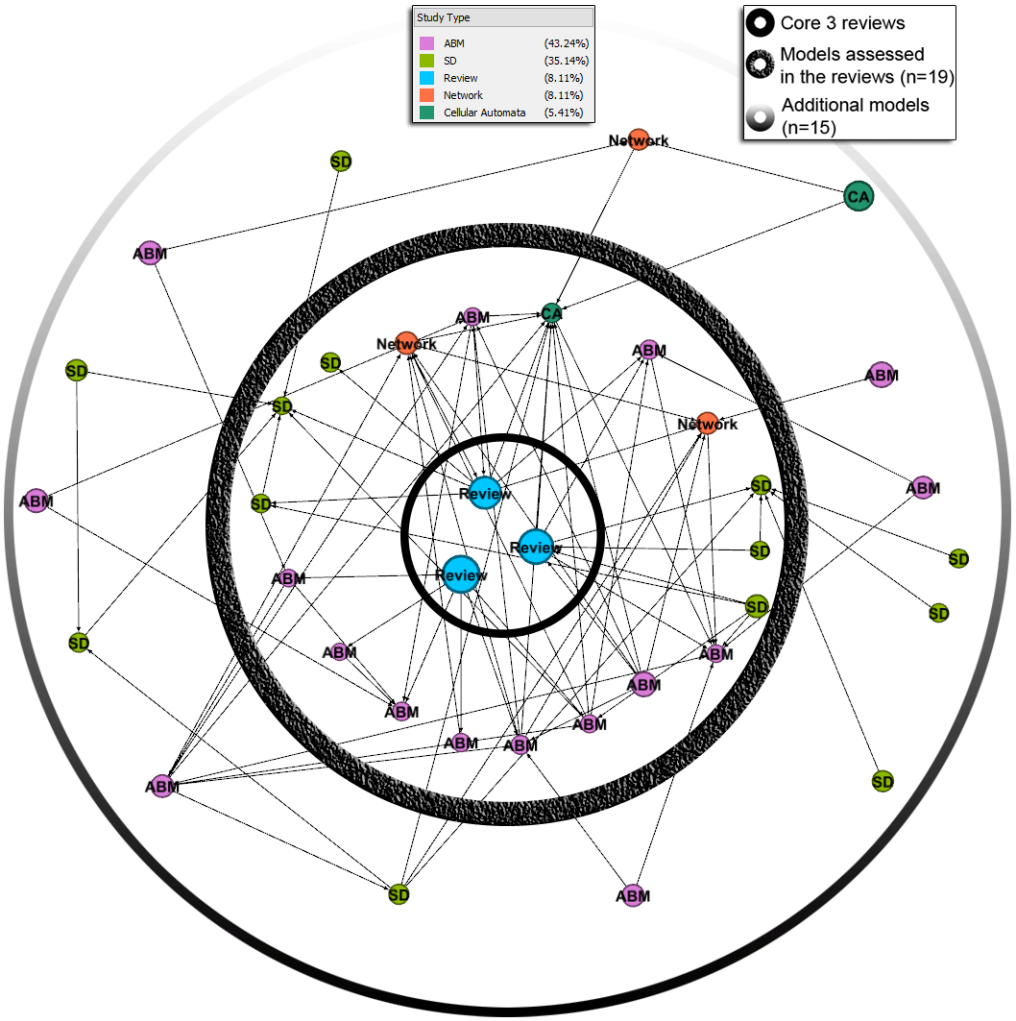

Figure 3.1: Types of articles selected and relation with the original 3 reviews. Each node within the black ring are the original reviews (distance 0). The nodes obtained a level up show samples taken at distance 1, which means they cite the original reviews.

Table 3.1: Selection Criteria of the Original Reviews.

Ref.	Criteria
[166]	Models from several teams of the National Collaborative Childhood Obesity Research Envision network, including statistical, social network, agent-based, and system dynamics.
[105]	Models focusing on health and economic consequences of obesity, future rates of obesity by historical trends, models relating dietary and physical activity to obesity, and models of specific interventions and policies.
[125]	Models published in the English language between January 2003 and July 2014. Only selected freely available studies of actual applications (not illustrations) of Agent Based Model for public health programs. Used the following sets of keywords: • Agent-based model, agent-based modeling or individual-based modeling. • Noncommunicable diseases, noninfectious diseases, non-transmissible diseases, or chronic diseases. • Heart disease, diabetes, stroke, cancer, chronic respiratory disease, or neurodegenerative diseases. • Unhealthy diet, physical activity, exercise, smoking, alcohol, hyperlipidemia, high cholesterol, high triglyceride, overweight, of obesity.

The exclusion criteria were applied by three independent reviewers after completion of graduate training in simulation methods (EAL, ATS, NAH). When disagreement between reviewers occurred, the final decision was made by a supervisory expert in simulation models of obesity (PJG). All extracted data was stored as Excel spreadsheets, available as supplementary material.

3.2.3 Model Characterization

We broadly characterized each article using up to five items:

Table 3.2: Our Selection Criteria to Expand on the Original Reviews.

Selection step	**Substep**	**Actions**
1		The technique of snowball sampling (Figure 1.1), is applied up to distance three from originally cited papers.
2		The sample is reduced by applying exclusion criteria in three consecutive steps. We removed:
	i	Articles which either focused on other chronic diseases, did not provide a model usable for public health (e.g.,, physiology only), or did not qualify as dynamic simulation model.
	ii	Articles that did not provide a model but a framework.
	iii	Articles that used a quantitative individual technique to perform data mining rather than run a simulation.

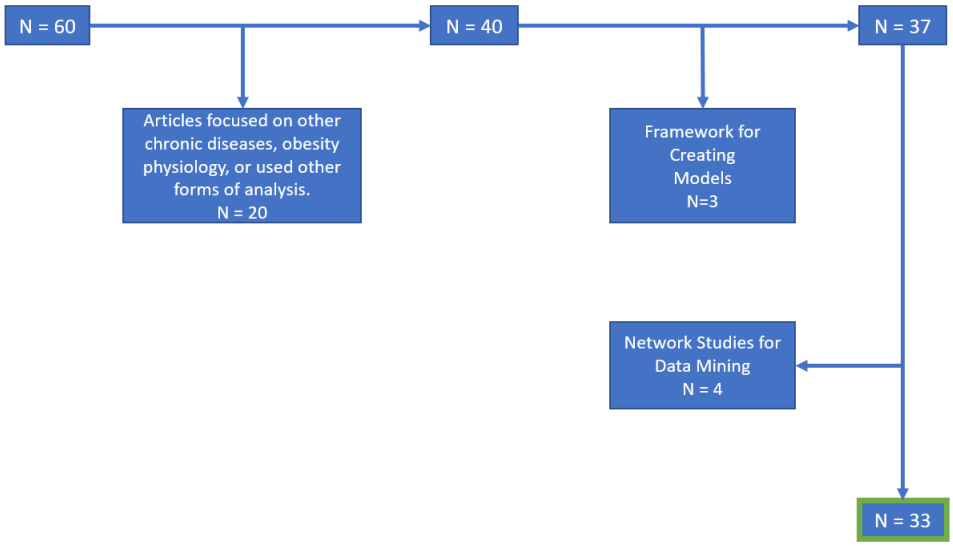

Figure 3.2: Flow Chart of Search.

- The modeling technique used
- Whether the quantitative model was individual or aggregate
- Whether the objective of the model was to be descriptive and/or predictive

- For predictive models, whether they were designed for intervention and/or prevention
- For preventive models, whether they were designed for primary and/or secondary prevention

A same article may be viewed as using different techniques. For example, one can build an agent-based model in which there is no physical space but only interactions between agents, which closely resembles a network model. Alternatively, one could claim to have developed a network model, but if the only topology studied is a grid, then this is *de facto* equivalent to a cellular automaton. We thus categorized articles not only based on the reported modeling technique but also on the clearer distinction between aggregate techniques versus individual level techniques. The article was categorized based on the purpose of its model as either predictive or descriptive (which can also be called prospective and retrospective, respectively [75]) based on the primary goal of the simulation. A descriptive model's goal is to assess the importance of different factors. For instance, a descriptive model built on an already-implemented policy can help with its evaluation, such as finding what may have caused a failure or success [75]. This is illustrated by the model used in [27], which observes the impact of weight on making new connections among youths.

In contrast, predictive models run simulations to observe changes in the long run in reaction to "what-if" questions (e.g., What would happen to obesity if certain taxes on soda were implemented? or zoning regulations on fast-food outlets?). When the model has predictive abilities, we further assess its target level of intervention and/or prevention. Intervention occurs at either the population or individual level. Intervention at the individual level is observing changes based off of individual modification, such as an individual altering one's friendship network. Population-level intervention methods would be changes in policy or similar cases where the entire population is equally affected. Prevention levels are similarly categorized into primary and secondary.[1] As summarized by Hoelscher *et al.*, "Primary prevention is defined as a public health effort targeting

[1]The Centers for Disease Control and Prevention (CDC) provides comprehensive examples of primary prevention strategies, such as increasing the consumption of fruits and vegetables in the population. Secondary prevention can correct departures from a state of normal weight via various interventions, of which many examples can be found in the literature on childhood obesity.

the *entire population* to prevent the development (incidence) of, or to decrease, the prevalence of obesity. In contrast, secondary prevention focuses on weight reduction among *overweight and obese* [individuals] to prevent long-term disease progression and development of comorbidities" (emphasis added) [83].

3.2.4 General Quality Assessment

The field of obesity is united by the problem and approaches it with a large variety of methods. There is currently no technical guidance for simulation models supporting public policies regarding obesity. Technical improvements of models for tobacco regulations are sometimes used as a proxy to inform models for obesity since the two fields share numerous features (e.g., policy resistance, heterogeneity of individuals) [75]. Consequently, we created a list of nine items drawn from classical items from simulation methods (e.g., calibration, validation, sensitivity analysis, replicability) and items specific to modeling obesity (e.g., inclusion of heterogeneity and social interactions). Note that being "specific to obesity" does not mean that these items *uniquely* apply to obesity, as they may be of use for other complex problems such as smoking. Rather, it is to say that when modeling obesity (and not just any complex problem), these aspects are known to be relevant.

Each aspect is briefly explained in turn while its significance is detailed in the discussion. Note that for brevity's sake, we explain our items in seven categories, but two of them were assessed independently (calibration and validation, sensitivity and uncertainty analysis) thus making nine items. Also note that since each modeling technique uses different terminology for the representation of a single element in the model, we refer to them all as *entities* (e.g., nodes in a network model, agents in an agent-based model, cells in a cellular automaton).

Time Frame. The time frame is defined as the length of simulated time by a (dynamic simulation) model. While some models such as SDs give easy-to-recognize time frames (i.e., in weeks or years), other models such as ABMs measure in ticks (i.e., discrete time steps) that can be converted into physical time. This feature was assessed by examining for how long a model was run by authors and whether authors provided a reason for that length of time.

Social Interactions. We describe *social interactions* as a connection between entities in a simulation model and their ability to influence each other. This is exemplified by Bahr *et al.* [9] in which a network model connects the individual entities and their influence on their neighbors. We evaluated this feature by looking at each model to see whether the obesity of an entity is impacted by its neighbors. The importance of this factor is well known in obesity [11, 82]. Further, the highly cited work of Christakis and Fowler has drawn attention to the links between a person's social network and body mass [19].

Heterogeneity. The complexity of obesity has been well acknowledged [31, 40]. A hallmark of complexity is the presence of *heterogeneity*, that is, the tremendous variations found among people. The effect of heterogeneity has been measured as high for weight-related factors [29], and such differences among individuals can lead to very different levels and trends of obesity [127] even when differences are very small [153]. Heterogeneity was observed by checking model parameters and whether they accounted for variations across individuals rather than using a single value.

Multiple Datasets. To have *multiple datasets* is to have data gathered for the simulation model from different sources. This was satisfied when authors combined multiple sources in creating a model. Multiple sources do not include using one data source but with multiple time points (i.e.,, one longitudinal dataset was not counted as consisting of several datasets).

Calibration and Validation. A model in general should perform both data calibration and validation. *Calibration* is the process by which adjustments are made to the model parameters within the margins of uncertainty to obtain a representation of the phenomena in question [46, 73]. *Validation* is defined as "the process of ensuring that the model is sufficiently accurate for the

purpose at hand" [155]. To examine this in models, we first searched for the presence of common calibration and validation techniques (i.e., trends or partial time series). If neither of those were found, then we searched for any mention of calibration or validation.

Sensitivity and Uncertainty Analysis. *Sensitivity and uncertainty analysis* of a model is defined by how a change in parameters will affect the outcome of a model, which affects how much trust is given to the simulated outcomes [16, 162]. Specifically *sensitivity analysis* refers to methods used to assess how sensitive (e.g., in terms of variance) the outputs are to changes in the inputs [154, 162]. This was assessed as we did for calibration and validation. We first looked to see if the more common methods of analysis were used (e.g., single parameter, regression, factorial). If none of those were used, we searched for the use of any method to perform the analysis.

Replicability (or reproducibility). A study is not replicable when researchers either (i) cannot (re)create the simulation model or (ii) are not able to reach the same conclusion (e.g., because they do not have access to the data used to run the model) [154]. Conversely, a study is replicable when an independent researcher can create the same model and obtain the same results as the authors of the study. We examined (i) whether or not a model was fully defined so it could be re-created (e.g.,, using formal definitions, pseudo-code, equations, or publicly accessible source code) and (ii) whether the data set could be publicly accessed.

3.2.5 Additional Quality Assessment for Agent-Based Models

In addition to the nine general items detailed in the previous section, we include two items that are only applicable to agent-based models (which represent about half of the total studies). The nine general items were drawn from both simulation methods and obesity research. Similarly, the two items used here draw from both. The first item checks whether an ABM is built using individual-level data, which is important for model quality from a simulation perspective. The

second item examines whether the ABM explicitly captured the physical space in which individuals exist.

Individual-Level Data. ABMs explicitly represent each individual in the target population. These virtual individuals (i.e.,, agents) have attributes that can distinguish them from each other [125]. If their attributes are independently drawn from some distributions, then it ignores the important correlations between the distributions (e.g.,, when assigning height independently of weight). To capture these correlations, models can draw from individual-level datasets, in which all parameters of the ABM take values together for a given individual. Note that this is different from assessing *heterogeneity*, which asks whether differences are modeled; here we assess the quality of the data used to model these differences.

Spatial Awareness. For a model to have *spatial awareness* it needs to account for the geography (i.e., physical area) around an agent. We evaluated this in models by searching for mention of the environment in which the agent lived and its impact on the agent. Geography plays an important role in obesity, particularly when it comes to inequalities regarding drivers of weight. For a comprehensive treatment of the topic, we refer the reader to the edited volume, *Geographies of Obesity: Environmental Understandings of the Obesity Epidemic* [146].

3.3 Results

The previous section detailed the items that would be asked of all articles, as well as the specific two that could be asked for ABMs specifically. A detailed application of these items to three selected publications is provided on a public online repository at `https://osf.io/n6pja/`. In this section, we start by reporting the results across the dataset. For example, this shows the percentage of articles that include heterogeneity or not. Then, we examine the number of articles and the number of items that they did not include (e.g.,, no time frame justification, no heterogene-

ity, no validation). In particular, we explore whether the numbers of items not included changes over time, which would indicate a difference in quality between more recent models and earlier versions. Finally, we investigate citation patterns to understand whether sufficient mixing exists in the field to promote the use of better techniques. The data displayed in all three subsections is provided as online supplementary materials.

3.3.1 Assessment across articles

Our results are presented in Table 3.4, with the underlying data accessible as online supplementary material (`https://osf.io/n6pja/`). In terms of model categorization, models were almost evenly split as descriptive or predictive. The interventions suggested were generally at least the population level, though many targeted both individuals and populations. A few models used only individual-level interventions [73, 185, 200]. The prevention level used was always at least primary. Many models experimented with policies that promoted healthy eating norms throughout the population, which affects both obese and non-obese individuals (i.e., primary and secondary interventions). There were no models that used only secondary-level interventions, though the model by El-Sayed *et al.* [37] experimented with using primary and secondary interventions independently of each other. The time range was similarly split between short and long term, with slightly fewer models in the intermediate range between five and ten years [76, 165].

In terms of quality assessment, a 'No' in Table 3.4 means that nothing was done in this item, while the remaining part either states that it was done ('Yes') or details the manner in which it was performed. For example, we see two-thirds of articles did not justify their time frame, and those that did tended to use simulation methods. The items least frequently attempted are the justification of a time frame, making a simulation replicable, and performing sensitivity analysis. These items were not attempted in half or more of the dataset. On the positive side, the items most frequently

attempted include an informal description of the model, calibration, and validation. We note that the quality with which calibration and validation was done is still limited, as models mostly rely on comparing trends rather than fitting with real-world time series data. Similarly, we note that even if only half of the models performed sensitivity analysis, this was most often done by varying a single parameter at a time, which is statistically inefficient and does not account for interactions, so conclusions may be incorrect [88]. Results on the quality assessments specific to ABMs are mixed. Most ABMs had access to individual data, but most ignored spatial components and thus limited agents to social interactions similarly to a network model.

3.3.2 Items not Attempted and Temporal Trends

We now focus on the number of items that were not attempted, which expresses the worst case. The data presented here is accessible as online supplementary material (`https://osf.io/n6pja/`). In Figure 3.3, the x-axis shows the number of items not attempted. We note that 0 is missing (as not a single article fulfilled all of the items).

From the cumulative distribution in Figure 3.3-a, we observe over half of the dataset did not address four or more items. We see a linear increase in the number of articles as the number of items rises from 1 to 6, which reflects that the number of articles was almost uniform across these items (Figure 3.3-b). We notice a sharp drop from 7 onward, with only four more articles out of which three were earlier models. This raised the question as to whether there was a statistically significant improvement of models between early and more recent ones.

Table 3.3: Aggregate Results

Main Category	Sub-category	Value	Prevalence (%)
General quality assessment	Time Frame Justification	Simulation Method	24.24
		Policy Agenda	9.09
		None	66.66
	Social Interactions	Yes	69.70
		No	30.30
	Heterogeneity	Yes	66.67
		No	33.33
	Multiple Datasets	Yes	63.64
		No	36.36
	Model Calibration	Trend	51.52
		Partial Time Series	30.30
		None	18.18
	Model Validation	Trend	57.58
		Partial Time Series	18.18
		None	24.25
	Sensitivity Analysis	Regression	6.06
		Fractional Factorial Design	6.06
		Full Factorial Design	6.06
		Single Parameter	30.30
		No	51.52
	Uncertainty Analysis	Yes	57.58
		No	42.42
	Replicability: description	Yes	81.82
		No	18.18
	Replicability: Source code	Yes	9.09
		Formal Description	33.33
		No	57.58
	Replicability: Dataset	Yes	69.70
		No	24.24
		N/A	6.06
Model-specific quality assessment	Agent Based Model: Individual Level Data	Yes	36.36
		No	12.12
		N/A	51.52
	Agent Based Model: Spatial Awareness	Yes	15.15
		No	48.48
		N/A	36.36

Table 3.4: Aggregate Results Continued

Main Category	Sub-category	Value	Prevalence (%)
Model Characteristics	Type	SD	36.36
		ABM	48.48
		Network	6.06
		Cellular Automata	3.03
		SABM	6.06
	Purpose	Predictive	54.54
		Descriptive	42.42
		Predictive and Descriptive	3.03
	Predictive Model:Intervention Level	Population	42.42
		Individual	9.09
		Population and Individual	9.09
		N/A	39.39
	Predictive Model:Prevention Level	Primary	27.27
		Secondary	0.00
		Primary and Secondary	24.24
		N/A	48.48

Figure 3.3-b further subdivides the articles between those published before the median date (February 2014) or after. We conducted a two-way ANOVA using R version 3.3.1 to test if there was a significance between the year a study was conducted and the number of items not attempted. The data was split into two subsets, corresponding to being published prior to February 2014 (Figure 3.3-b, orange) or after (Figure 3.3-b, blue). We found there was no significant connection between the year of the study and the amount of items not attempted, $P > .638$ and $P > .87$.

3.3.3 Citation Patterns

Having established that there was no technical improvement between earlier and more recent models, we explored whether this could be due to the field being insufficiently aware of the tech-

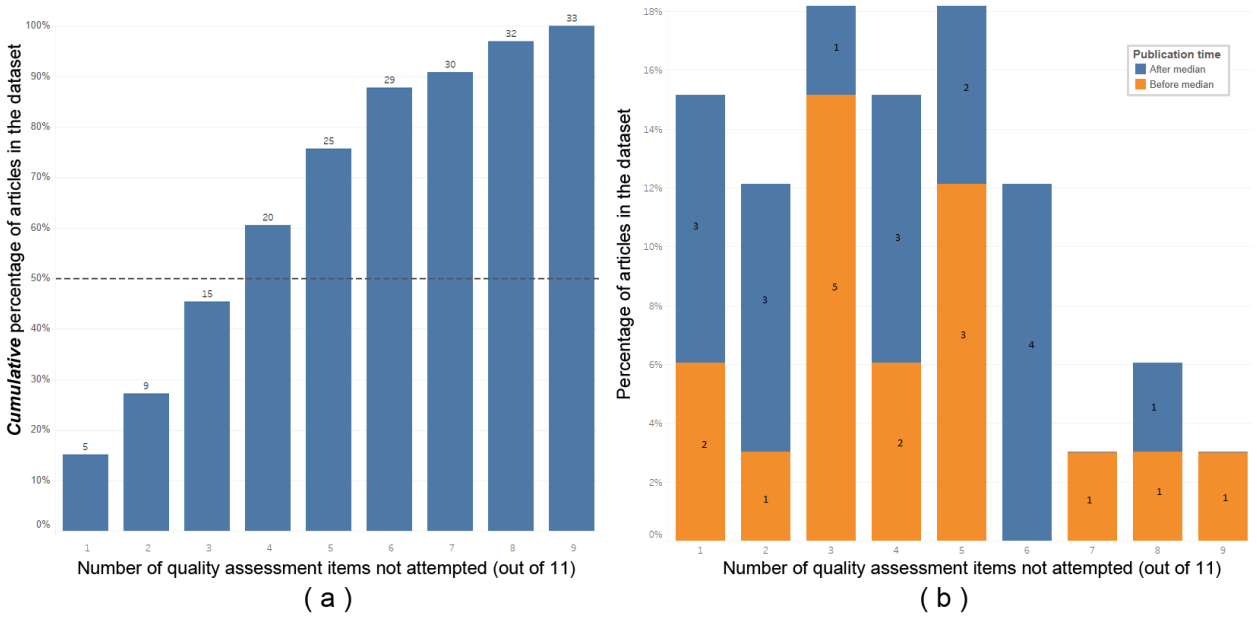

Figure 3.3: Cumulative distribution of the number of articles (y-axis) and number of items not attempted (x-axis). A histogram of the same dataset is further divided to show articles published either before or after the median of February 2014 (a).

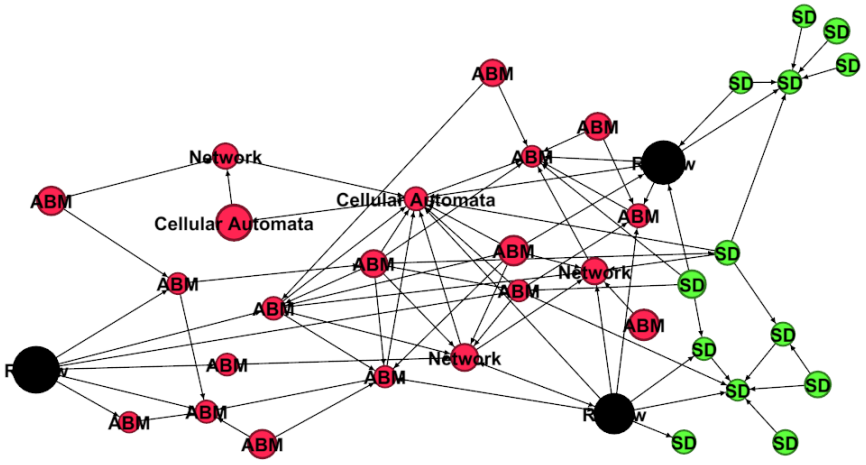

Figure 3.4: Citation network. Individual-level models in rose, aggregate models in green, and the original three reviews in black.

nical rigor displayed in some models. We thus construct a citation network in which each node corresponds to an article (including the original three reviews) and a directed edge from node a to b shows that a cited b. The network is shown in Figure 3.4 and is accessible as online supplementary material (`https://osf.io/n6pja/`).

The main two types of models were agent-based models (ABM) and system dynamics (SD). Network analysis showed that articles of either type tended to cite articles of the same type. About 75% of ABM papers cited at least one other ABM paper but only 56.2% cited at least one non-ABM paper. The difference was particularly pronounced for SD papers: 76.9% of SD papers cited at least one other SD paper but only 23.1% of SD papers cited at least one non-SD paper. When simplifying the type of model to aggregate (e.g.,, SD) or individual (e.g.,, ABM), we found that, on average, 95.83% of individual models shared the same technique as a neighboring node, while 88.69% of aggregate models shared the same technique as neighboring models. Alternatively, this says that individual models only made 4.13% of citations to aggregate models, while aggregate models made 11.23% of their citations to individual models.

Figure 3.4 suggests two groups of SD models (top right and bottom right). We further explored this possibility using community detection algorithms, as implemented in Gephi version 0.9.1. The modularity class model confirmed that there are two groups of SD papers, though both groups predominantly cite other SD papers.

3.4 Discussion

The growing use of simulation models in public health as it relates to obesity has resulted in dozens of models built on different techniques without guidelines or systematic methodological evaluation. This paper seeks to review past models of obesity based on their technical merit and, from this, to suggest priorities for the design of models going forward such that they can con-

tribute deeper insight into effective public health policies. We examined a total of n=33 articles obtained from the original three reviews as well as a snowball sample to capture additional models. One-third of the articles were aggregate models (using system dynamics) while the remaining two-thirds were individual-level models (using mostly agent-based modeling but also cellular automata or network models). For each model, we recorded its type and purpose, and examined nine items based on best practices in simulation models as well as requirements specific to obesity. Two additional items were included for agent-based models only. To contextualize the impact of meeting some items or not, the next subsection details why these items are needed in light of research on either simulation or obesity. Then we examine the answer to each of our three research questions and conclude on the limitations of the study.

3.4.1 Impact of Selected Criteria on Model Quality

In the methods section we defined the important features, how we evaluated them within each model and the importance of the feature in terms of obesity. Here we discuss the effects of these features on simulation.

Time Frame. In simulation models the time frame must be considered for either the real-world effectiveness in case of policy, for instance, or by statistical methods to best make the model most accurate. First, if a model is focusing on policies, then the policies studied in the model have to be designed, implemented, and evaluated in the real world. The expected duration within which to evaluate policy performance [34] should be reflected by a model. For example, in British Columbia, general elections are held every four years and shape the political agenda; consequently, a model aiming at informing policymakers may need to provide results within four years rather than within decades [183]. Second, finding the appropriate time frame is a well-studied problem in the area of simulation [102], and statistical methods have been devised to select the run length,

such as the marginal standard error rule (MSER) or the MSER-5 [190, 191]. In other words, there are systematic methods to assess for how long a model needs to run. By not following these methods, results may report a (misleading) transient output, and erroneous inferences may be made regarding the potential long-term benefits of the virtual intervention. This is particularly important in the context of obesity, as the search for effective long-term weight management interventions is ongoing [78].

Social Interactions. Links between a person's social network and body mass have been the subject of debates on both the methodology and causation; for example, the correlation between one's weight and the weight of peers can also be due to selecting friends of similar body types [27]. Nonetheless, a review has confirmed that social influence plays an important role in weight dynamics, for example, by influencing social norms [74]. In a policy context, it is thus recommended to take into account social interactions. They can be included in individual-level models, by taking into account the explicit interactions between specific pairs of individuals, and can also be accounted for in aggregate models [166].

Heterogeneity. The danger of ignoring heterogeneity by focusing on "one size fits it all" interventions has long been met with criticism from practitioners who observed that success would then only occur when a person happens to meet the average profile assumed for the intervention [39, 77]. Consequently, any obesity model must take care to appropriately model heterogeneity in the population as it pertains to the objective of the model [186].

Multiple Datasets. Reliable data is key while modeling; while one (sufficiently comprehensive) dataset may meet the needs of a model, there are several reasons to have *multiple datasets*. First, it is important for *consistency*. The Bradford Hill criteria defines consistency as the ability to observe consistent findings with different samples. For example, one would be able to know whether the conclusion of a simulation holds for different population samples or are highly specific to (possibly unknown) features of the population used in a specific sample. Second, and related to consistency, multiple datasets contribute to creating accurate models. In data mining, training

using data from a multitude of compatible sources causes a significant improvement of classification accuracy [163]. Furthermore, taking data from multiple yet similar data sources can aid in mitigating the bias from any of the individual data sources [86]. Thus making use of multiple similar data sources will increase the generalizability and accuracy of a model as well as lessen the bias that occurs within individual data sources.

Calibration and Validation. If the model is predictive, the data used to initialize the model should continue on from a short time before the end of the observed data. This allows for a *warmup* time of the model against real-world observations of what is being modeled [174]. That is, to start the model within the time of your observed data, and if the outcomes of the model in this warmup period match the observed data, then the model can be allowed to predict further into the future based off these initial conditions [178]. This process is known as data assimilation, and while much more highly used in geosciences, its application is still highly relevant to that of any predictive model application. Whether the model is descriptive or predictive, data should be used so that the model is functioning in the correct and meaningful way.

Sensitivity and Uncertainty Analysis. The reality of modeling and simulation is that models are often used as authoritative evidence, and overconfidence in their results is a very real and common problem [155]. It is thus important to convey to possible users (such as policymakers) how parameters may affect a policy's outcome and the extent to which we can trust the simulated outcomes. The routine should thus start by assessing the possible sources of uncertainty for a model [16, 162]. This assessment can take various forms, including structural, parameter, stochastic, and data uncertainty [16, 17]. When any of these uncertainty sources appear in a model, they should be assessed as it pertains to the modeling objective, including implications of that uncertainty on the results and any means by which the uncertainty was addressed in the model. While the primary purpose of uncertain analysis is to help determine the overall level of confidence one can have in the results, it can also be used to determine the importance of collecting additional information when a decision is being made [16]. Some modeling work has made a contribution to

obesity research in this regard by highlighting what data should be obtained before sufficiently accurate simulation models can be made [45].Performing sensitivity analysis has numerous benefits, including [134] ensuring the robustness of the model (or identifying inputs that are not robust), increased understanding of the model structure, and finding potential errors in the relationship between input and output variables. In sum, a model that lacks uncertainty and/or sensitivity analysis necessitates low confidence in the conclusions due to a lack of confidence evaluation on the part of the study's author(s).

Replicability is especially important when surprising or contradictory results occur and for important studies. The issue of replicability has been neglected for some time now by the scientific community at large, and researchers have warned of a looming replication crisis in many different fields, with many scientists being unable to reproduce the results of other research and often even their own [10]. The importance of replicability is illustrated in the National Institutes of Health's (NIH) FY2016-2020 Strategic Plan. In the context of modeling and simulation, efforts should be made to achieve as much technical transparency as possible while still respecting intellectual property rights [188]. We note that replicability is a continuum rather than a binary. A model can most easily be re-created when its source code in a commonly used language is hosted on a public repository. There are risks to reproducibility when the source code is instead hosted on a person's website, as it may not be available in the future. When the source code is not available, researchers may be able to write it if there is a sufficient description of the model's structure. A 'sufficient' description is also a continuum, ranging from using standard protocols (such as the ODD protocol [70]) to providing ad-hoc/incomplete descriptions.

Individual-Level Data. Several problems occur when using individual-level models without individual-level data. First, a distribution would have to be assumed for each attribute. This can be unrealistic, for example when assuming normally distributed values when they may not be. Second, even when the right distributions are found for each attribute in isolation, we note that attributes can be highly correlated [51]. For example, an agent's food intake level is closely related to the agent's

level of exercise. A study by Hall *et al.* suggested that a small average daily difference between the two would be enough to explain the observed average weight gain in the population [72]. By ignoring this dependency, a model could create a population whose dynamics at baseline are already erroneous.[2] In sum, an ABM should be able to initialize its agents from sufficient data to represent the diversity found in the real world. Not being able to do so can create unrealistic populations, thus making an ABM invalid for use.

Spatial Awareness. There is a spatial element to both food and physical activity. For example, the prices of both healthy and unhealthy foods were found to differ between areas with high poverty rates and more affluent areas [12]. In particular, the ratio between the price of healthy and unhealthy foods can create a significant obstacle to making the healthy choice the easy choice [31]. The large variations in the food environment observed in some regions have even prompted the specific ABMs [6]. Similarly, there is spatial variation in physical activity and its enablers (i.e.,, aspects of the built environment required to perform physical activity), which was investigated in over five articles using ABM by Yong Yang and colleagues (for the most recent reference see [196]). In conclusion, the space plays an important role in obesity and numerous models have demonstrated that it can (and should) be captured in an ABM.

3.4.2 Question 1: Are Current Models Developed Adequately?

Three main technical issues were found across all types of models. First, there were insufficient statistical analyses; nearly half of the models had no sensitivity analysis or uncertainty analysis. These analyses are normally routine modeling practices. They are also particularly important from

[2]We acknowledge that policy models *could* use multivariate distributions to capture correlated distributions [100]. Since each distribution can have a very different form (e.g., socio-economic status can be a power law while weight follows a normal), few multivariate distributions are flexible enough for this purpose (e.g., the Johnson system [28]) and they can be difficult to fit to the data [41]. This difficulty may explain why, in the absence of individual-level data, simulation models of obesity resort to including very few (if any) correlations.

a policy standpoint, as policymakers may ask about the extent to which results can be trusted and may vary the parameters used by a model (e.g., for adaptation to a local context) without being aware of the outcome's sensitivity to these parameters. We also note that, when sensitivity analysis was performed, it was most often done by varying a single parameter at a time. This is well established as statistically inefficient (i.e., not enough information is gathered about the model's response proportionally to the number of variations to make), and the fact that it ignores interactions may result in erroneous conclusions [88]. Thiele and colleagues summarized previous research by stating that modelers in some fields

> are amateurs with regard to computer science and the concepts and techniques of experimental design. They often lack training in [...] sensitivity analysis and for implementing and actually using these methods. Certainly, comprehensive monographs on these methods exist, but they tend to be dense and therefore not easily accessible. [173]

Our findings suggest that this is also the situation for modeling public policies in obesity. We thus agree with previous research that popular software tools should facilitate sensitivity analysis, for example by providing access through extensive statistical packages [1, 173]. However, adding packages will only make a difference if modelers use them. The former Institute on Systems Science and Health (ISSH) played an important role in introducing the main tools (system dynamics, agent-based modeling and network models) to those interested in using them for public health. The fact that many of the characteristics found in the models are at the level of introductory modeling courses [58] suggests a need for advanced institutes or workshops, building on the curriculum introduced by the ISSH that has become more mainstream and emphasizing statistical rigor.

Second, there frequently was insufficient data. This can be viewed in different items. Nearly 40% of the reviewed models did not use multiple datasets, thus limiting the ability to make consistent and accurate findings. When models did engage in calibration or validation, they mostly did so by looking at trends or comparing with partial time series.

In other words, modelers considered their model "correct" (either when calibrated or validated) based on the trajectory of the output or matching a few data points (e.g., when the change in obesity in the model from one year to the next was in line with datasets [49]). There were no reviewed models that performed either calibration or validation against a complete time series. We also observed that not all of the individual-level models actually used individual-level data. In sum, this points to the importance of obtaining more fine-grained data. Given the current data-heavy state of the world, this can be addressed. The main barrier may no longer be the nonexistence of data but rather our ability to find it, which should not be neglected given the multiplicity of repositories or the tendency of modeling papers to cite little outside of their group. Indeed, we see the same datasets often being used by the same groups of authors, for example through the the Health Survey for England (HSE) data [25, 198], or the National Longitudinal Study of Adolescent to Adult Health (Add Health) data [175, 201]. Only the National Health and Nutrition Examination Survey (NHANES) data has so far been used by a larger variety of groups [43, 84, 152]. This suggests that there needs to be greater guidance about which datasets are the most robust for typical weight-related factors. Finally, while most of the reviewed models were specified through informal descriptions or equations, only a handful provided source code [9, 175]. Improved transparency is important to support replication efforts.

3.4.3 Question 2: Are we Improving the Quality of Models?

We did not find a statistically significant difference between the year a study was conducted and the number of items that were not met. Therefore, we conclude that newer models are not better than earlier ones in addressing all necessary aspects. However, some aspects may still have been better addressed (e.g., as many models are validated, but they may be validated using better techniques than before). As we performed a worst-case analysis, our conclusion is only at the level

of items (e.g., do newer models justify their time frame more than before?) rather than within items.

3.4.4 Question 3: Are Best Practices Shared and Re-Used?

Overall, we found very little re-use of existing models, outside of groups adapting their model to different contexts. There was very limited cross-pollination between modeling techniques of different types, even when our nine general assessment items could be applied across techniques. As modelers often specialize in individual or aggregate techniques, we expected that models of one type also frequently cite models of the same type. However, we found that it was not the case for articles in the aggregate techniques (all of which used system dynamics in our sample). Even though the articles focused on policy models for obesity and used the same technique, there appears to be two groups at work with limited interactions as seen through the citation networks.

The limited re-use may be due by a combination of factors involving a lack of awareness of previous models and not being able to access their code. We also find it preoccupying that the source code was never provided on repositories (e.g., the Open Science Framework or GitHub) and that authors rely on hosting code on webpages that may no longer exist in the future. This points to the need for more systematic efforts in using established repositories and disseminating models to promote re-use.

3.4.5 Limitations

There were several limitations to this study. As the field of modeling for public health in obesity is highly dynamic, several new articles are routinely released. Our data collection includes articles published up to the first half of 2016. Other articles have been published since. For

example, in September 2016, Li *et al.* released two new agent-based models supporting public health interventions and policies for improving dietary behaviors and preventing obesity [106]. The forthcoming 2017 Routledge *Handbook of Applied System Science* will also include an agent-based model, but with a focus on transforming the space to support physical activity [108]. As part of our review, we examined whether the quality of models was improving as time goes on. We found no statistically significant difference in earlier models compared to their more recent counterparts. Consequently, we do not expect the latest 2016 models to differ noticeably. However, we think that the field would benefit from conducting a regular technical assessment of models, particularly after our recommendations aforementioned may have had an effect.

Another limitation was in the criteria that we chose. While many more criteria could have been proposed, we focused on what is commonly agreed to be important for the complexity of obesity or a good practice from a simulation standpoint. As the field matures, more ambitious criteria may be added. In particular, we saw few attempts in the models to capture the intergenerational effects of obesity, whether physiologically or by household environment. In fact, several models do not include birth and death and instead have "endless" individuals. Similarly, when birth and death are captured, it may be a random arrival process in a model that does not keep track of a lineage. A notable exception is the study by Thomas *et al.*, who looked at how population BMI distributions changed over the course of multiple decades. By including birth, they were able to model individuals growing up with obese parents and thus more susceptible to obesity [174]. Additionally, the study found that the model needed to incorporate natural deaths so that persons did not pool into higher BMI indices but that instead the model more naturally showed the real progression of the population from different BMI states [174]. Future assessments may thus ask not only whether the model captures social interactions but also which social structures are taken into account.

Similarly to the inclusion of intergenerational effects being scarce, current models do very little in conveying their runtime. That is, even if models can be highly complex, they generally

do not state their algorithmic complexity or the time it takes to run them on specific computer architectures. This can create unexpected challenges for replicability, as one may have access to even the source code of a model but not to the computational power necessary to run it. While this risk is limited given that modelers often rely on software such as AnyLogic, the need for more computational power should be assessed as the field matures.

3.5 Conclusions

Obesity is an important and complex societal problem. A growing number of models are developed to support the identification and evaluation of potential solutions. However, we found that these models suffer from several technical shortcomings, including insufficient statistical analyses, a lack of data, and an insufficient use of repositories to disseminate and re-use models. We suggested several ways to address these barriers, and we recommend that technical assessments be regularly conducted with extended criteria as the field matures.

CHAPTER 4

ANALYZING AND SIMPLIFYING MODEL UNCERTAINTY IN FUZZY COGNITIVE MAPS

In the previous chapter, we discussed guidelines for creating simulation models in public health. We saw that modelers very rarely considered FCMs when creating the complex models for public health. FCMs, however, are a prime method for modeling complex problems due to their ability to cope with uncertainty. In this chapter, we focus on a method to better account for uncertainty in FCMs and how to reduce that same uncertainty. Furthermore, we introduce our new open-source Python library for the creation and simulation of FCMs.

This chapter was published in the following peer-reviewed conference article [101]:

- **Eric A. Lavin** and Philippe J. Giabbanelli. Analyzing and Simplifying Model Uncertainty in Fuzzy Cognitive Maps. Proceedings of the 2017 Winter Simulation Conference.

My contributions consisted of (i) creating and implementing the FCM library for Python; (ii) completing the factorial design, and (iii) performing all tests on a HPC and singular machines.

4.1 Introduction

Fuzzy cognitive mapping (FCM) is a modeling method that can represent the "mental model" of individuals by articulating different factors and the dynamics of their interactions. Intuitively, an FCM can be seen as a causal network equipped with an inference engine. While FCM dates back to the late 1980s [97] and has been used in a variety of fields [142], it is increasingly popular in participatory modeling specifically. Indeed, modern platforms (e.g.,, `MentalModeler.com`) now allow individuals to asynchronously and collaboratively develop simulation models by sharing their knowledge in an accessible and standardized format [55, 63]. This is particularly used in ecological modeling, where FCMs allow us to synthesize the different perspectives of the many participating stakeholders [30, 65, 126], and in health where complex problems may require a large number of experts for each subdomain [49]. While FCMs share broadly similar constructs with the modeling technique of system dynamics (capturing factors and dynamic interactions, use in participatory settings), FCMs are specifically designed to handle situations with uncertainty and vagueness. By building on fuzzy logic, FCMs can be seen as a formal tool to manage the imprecision found in real-world problems. For example, participants would assess the strength of a causal link in the FCM using linguistic terms (e.g., very high, medium) that correspond to fuzzy membership functions. Fuzzy logic would transform their answers into one number for this causal link, which the inference engine uses when updating factors.

However, in real-world problems, we may not be able to assign only one number to a causal link. This may be due to conflicting evidence, a lack of agreement between participants, or the vagueness of what a link may represent. For example, in our FCM of obesity, we found that experts' assessments on one-third of the FCM's links had significant variations [49]. Salmeron extended the formalism of FCMs into fuzzy grey cognitive maps (FGCMs) by giving a range to all links rather than one number [159] (Figure 1.4). While this provides a vehicle to represent

vagueness, it also creates a much larger search space in which to run the model: Which value should be used for each link in one simulation run? This problem may be avoided altogether by systematically taking the mid-point of the range [159], but this solution has two issues. First, it brings us back to a normal FCM, thus defeating the purpose of capturing vagueness as a range to start with. Second, this simplification may ignore a lot of the search space. That is, the conclusions made by running the model this way may not be representative of what would have been obtained by running it with the other values included in each link's range. While this simplification can thus lead to erroneous conclusions, the other extreme of running the model for all possible combinations of values would just be infeasible. In this chapter, we analyze how to simplify an FGCM while having a minimal impact on limiting the model's outputs.

Design of experiments (DoE) has long been used in the field of simulation to address questions of that type [88]. For example, a 2^k factorial design would inform modelers about how much of their model's variance is due to single parameters or combinations of these parameters. Parameters that contribute little to the variance (either directly or in combination with others) can then be simplified by being set to one value. In our case, each link is a parameter, and the goal is to identify the links whose range can be replaced by a single number. This simplifies the model after careful analysis, rather than using the same approach for all links regardless of how sensitive they are for the model. Using the DoE approach requires running a model many times. This is unusual for FCMs, which are deterministic models and thus only run once. Even when FCMs are used as part of stochastic simulations (e.g., to represent the mental models of populations), the number of runs remains small [51]. Consequently, current software packages run FCMs sequentially, which means that running enough of them for a DoE can be prohibitive timewise. In addition, it is unknown whether FCMs would lend themselves well to simplifications. In theory, one could design an FCM where all links contribute equally to variance, thus our simplification after a DoE would be no better than taking the mid-point of all ranges to start with. This chapter addresses these limitations as follows:

- We develop a new, open-source library that runs fuzzy cognitive maps in parallel.
- Using this library, we examine whether DoE techniques can help to simplify three previously developed fuzzy grey cognitive maps (which extend FCMs with uncertainty on each link).
- The time required to analyze and simplify the models is examined on different configurations ranging from personal computers to high-performance clusters.

The remainder of this chapter is organized as follows. In Section 4.2, we provide a technical background on the computations involved in running an FCM, and on the design of experiments. Then, we explain the design of our solution in Section 4.3. This includes our new library for FCMs, using it efficiently to perform simulation runs in parallel over various hardware configurations and analyzing results to simplify the model. Section 4.4 provides an experimental evaluation of our approach to simplify three published models, ranging from 8 to 25 links. Section 4.5 highlights some of the limitations of our approach, particularly when it comes to simplifying very large models. We conclude by summarizing our current achievements in systematically simplifying small to medium models.

4.2 Background

4.2.1 The Simulation Approach: Fuzzy Cognitive Maps

A fuzzy cognitive map (FCM) models the behavior of a system through three key constructs:

(i) Nodes, representing concepts of the system such as states or entities. Nodes have a weight in the range [0, 1], indicating the extent to which the concept is present at a simulation step.

(ii) Weighted directed links, representing causal relationships. Their weight is from the range [-1, 1] where negative weights indicate that increases in the source node cause a *decrease* in

the target node. Conversely, positive weights indicate that increases in the source node cause an *increase* in the target node.

(iii) An inference function, which updates the value of each node based on the weights of both the links going into it and the nodes that these links connect to.

Formally, the number of nodes is denoted by n. The weights of the directed links can be represented as an $n \times n$ adjacency matrix A, where $A_{i,j}$ is the weight of the link from i to j. The value of each concept at step t of the simulation is represented by $V_i(t), i = 1 \dots n$. At each step of the simulation, these values are updated using the standard equation (Equation 1.1), where f is a clipping function (also known as *transfer* function) ensuring that the values of nodes remain in the [0, 1] range. For example, in an ecological model, a node could stand for the density of fish in a given space, where 0 means no fish and 1 means a maximal density. That value cannot go beyond 1 since it is maximal, and the density cannot be negative either. The clipping function has to be monotonic (to preserve the order of nodes' values) and it is recommended to use a sigmoidal function when modeling planning scenarios [177]. In this chapter, the function we employ is the widely used hyperbolic tangent $tanh$ [49, 71, 111].

An FCM does not include the concept of time; its steps do not map to physical time (although extensions exist to remedy this limitation). Consequently, an FCM does not run for a time period. Rather, Equation (1.1) is applied until a subset of nodes reaches a stable value. The subset is determined based on the application context. For example, in our previous FCM for obesity, we were interested in the long-term trends for obesity and required this one concept to stabilize in order to stop iterating. Other concepts such as food intake or weight discrimination did not have to stabilize in order to answer the question: How would the level of obesity change in reaction to a new intervention? [49]. Formally, consider that a subset $S \subseteq V$ needs to stabilize. Then the simulation will end when Equation (**??**) is satisfied, where ϵ is set to a very small positive value. Simulation packages generally include an additional condition whereby the simulation will

stop after a set maximum number of steps, in case the condition stated by equation (1.2) is never met. This additional condition is rarely necessary in practice, as will be illustrated in Section 4.4. Consequently, an FCM is an asymmetrical network of continuous concepts, of which some are required to converge to an equilibrium point or limit cycles. For a broader discussion on methods to improve convergence velocity (e.g.,, particle swarm optimization), we refer the reader to [123, 142].

Since there is no randomness in equations (1.1) and (1.2), an FCM does not need repeated simulation runs. The main simulation packages for FCMs, such as `FCM TOOLS` or `MentalModeler.com` thus do not even mention parallelism [63, 122]. While there has been research on parallelism as it relates to FCM, it has mostly been on parallel implementations of techniques used to design FCMs from data (e.g.,, genetic algorithms [170]) rather than on running the FCM itself. Similarly, there have been extensions of the FCM framework which *may* then involve parallelism; these typically propose to design models by combining multiple FCMs either through arbitrary network topologies [51] or via the coordination of a central entity [171]. However, a parallel implementation was not presented, as the algorithms appear to run sequentially.

FCMs are designed to cope with uncertainty, which contributes to their popularity as a participatory modelling technique [30, 65, 126]. Fuzzy logic is indeed used to compute the weight of each link based on linguistic variables (e.g., very strong) picked by participants who evaluate its causal strength. However, participants may not agree or may interpret what the link represents in different ways. Thus, there can be significant variations between participants' assessment of causal strength. In one of our FCMs, significant variations were observed on one-third of the links [49]. To represent these variations and deal with the uncertainty of real-world problems, Salmeron extended the FCMs presented in this section. In particular, he equipped each link with a range as seen in Figure 1.4. We refer to this extension as fuzzy grey cognitive maps, or FGCMs.

4.2.2 Factorial Design of Experiments

As explained in the introduction, an FGCM may have uncertainty on too many links. Fixing the less important ones would thus make the model more tractable. There are several approaches to assess the importance of different parameters in a simulation model [88]. At one extreme, one may perform a simple sensitivity analysis that varies one parameter at a time while others are fixed at typical values. This is statistically inefficient and does not account for interactions. At the other extreme, one can generate all possible combinations of parameter values, but exhaustively exploring this search space may be infeasible. A good design of experiment (DoE) thus provides information about the contribution of parameters and their interactions, in a way that is feasible given the resource requirements (e.g., computation time). In particular, a 2^k factorial design of experiments reduces the number of levels of each parameter to 2. It is commonly used to determine the relative importance of parameters in a performance study, and readers can refer to [199, 50] for examples.

To prepare a 2^k factorial design, one creates a table with all possible combinations of parameter values. Each row defines the setting of a simulation run. The runs are performed, and results are stored in additional columns. For example, in Table 4.1 we have eight parameters (i.e., links of an FGCM) so we start with eight columns and 256 rows for all possible combinations. The first eight rows are shown in Table 4.1. Note that this table shows the first eight combinations out of $2^8 = 256$. Many entries are identical and represented by ... to avoid displaying redundant information. Simulation results in this example are the values of nodes C_1 and C_2 upon stabilization (equation 1.2).

Table 4.1: Example of a Factorial Design with Eight Links.

$C_2 \to C_1$	$C_5 \to C_1$	$C_6 \to C_4$	$C_3 \to C_1$	$C_3 \to C_2$	$C_5 \to C_4$	$C_4 \to C_2$	$C_6 \to C_2$	C_1	C_2
-.8	.8	.4	-.5	.4	-.3	-.5	.6	-.9678	.6107
...	...	...	...	...	...	-.5	.8	-.9744	.7472
						-.3	.6	-.9758	.7822
					-.3	-.3	.8	-.9781	.8430
					-.1	-.5	.6	-.9698	.6478
					...	-.5	.8	-.9752	.7662
						-.3	.6	-.9762	.7915
-.8	.8	.4	-.5	.4	-.1	-.3	.8	-.9783	.8488

Once results are generated, we can calculate the effects, i.e., how much of the variance in the results is due to the parameters and their *interactions*, i.e., if there are two parameters A and B, we would calculate the effects from A, B, and AB (2nd order interaction). While using a 2^k design allows us to compute effects up to the k-th order interaction, it is common to stop after the 3rd order if effects become very small [199, 50]. Several textbooks detail how to calculate effects, and for an updated coverage we refer to Chapters 6 and 7 from [120]. While we are not aware of previous work computing effects from a 2^k design in parallel, it is important to understand why it can be done for this chapter, as k can be large. In short, calculating effects involves (i) representing all parameter values as -1 and 1 (thus using a truth table), (ii) generating additional columns for interacting effects by multiplying the respective columns (e.g.,, the signs for AB are obtained by multiplying the columns for A and B), and (iii) multiplying each binary column by results from the simulation runs and adding them. This third step can intuitively be understood as a weighted sum, which qualifies as "embarrassingly parallel" since computing a weighted sum can be divided into independently computing parts of that sum.

4.3 Methods

4.3.1 New Open-Source Library for Fuzzy Cognitive Maps

The code for our library is publicly available on a third-party research repository at `https://osf.io/qyujt/`. The library is written for Python 2 and builds on NetworkX (as data structure for the underlying network of an FCM) and NumPy (to compute the matrix operations of equation 1.1). Note that the matrix operation from equation (1.1) is already performed in parallel by NumPy (using the Basic Linear Algebra Subroutines) [179]. Additional parallelism may also be done at the level of the transfer function (f in equation (1.1)), which consists of ensuring that

the new values for the concepts are within the [0, 1] interval. We tested whether such parallelism was useful in our situation by computing the average time to apply the transfer function in parallel or sequentially, depending on the number of concepts. Results of our benchmarking (Figure 4.2) suggest that parallel processing is *slower* than sequential processing (due to associated overheads) when there are less than 50 concepts, similar up to 100 concepts, and *faster* after 100 concepts. As the case studies in this chapter all have less than 50 nodes, we opted for the faster sequential implementation in our library.

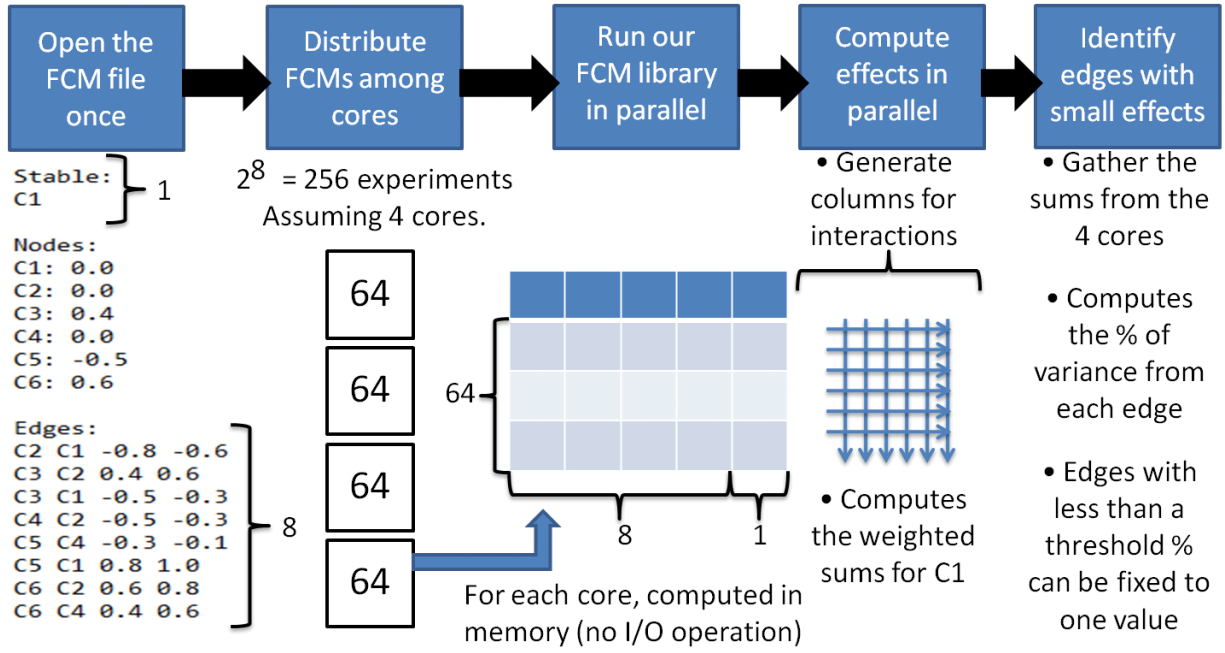

Figure 4.1: General workflow for our approach (top). Each component of the workflow is exemplified using our first case study and assuming a hardware with four cores. When several successive steps are necessary for a component, they are listed as successive bullet points.

4.3.2 Proposed Process to Simplify Models

4.3.2.1 Overview

Our proposed process involves five main steps, depicted in Figure 4.1. In short, we open the file only once, load the FCM, determine how many concepts to stabilize on, and how many links we need to assess for possible simplification. All combinations of binary link values are then generated and the corresponding FCMs are run in parallel, with each computing core running an approximately equal number of FCMs. Rather than aggregating the raw results from each FCM to calculate the effects, each core calculates the effects based on the FCMs that it ran. That is, each core generates the truth table with all factors (i.e., individual links and their interactions) and computes the weighted sums between each factor and each FCM concept to stabilize. The sums are then gathered across the cores so that we know how much of the overall variance in each FCM concept is due to each factor, across all simulation runs. Links whose contribution to variance (either directly or through interactions) is less than a given threshold can be simplified. The next subsection details how experiments are performed in parallel, and the last subsection explains how results are computed.

4.3.2.2 Steps 1–3: Distributing and Running Experiments in Parallel

While one FCM may be considered a relatively small model in terms of the number of links k, using a 2^k factorial design means that the search space grows exponentially with k. It is thus important to efficiently run FCMs in parallel. Our solution defines an FCM in a file listing the concepts, their initial values, the links (with their two possible values), and the concepts that need to stabilize. Reading from this file for each of the 2^k runs would create a significant I/O bottleneck.

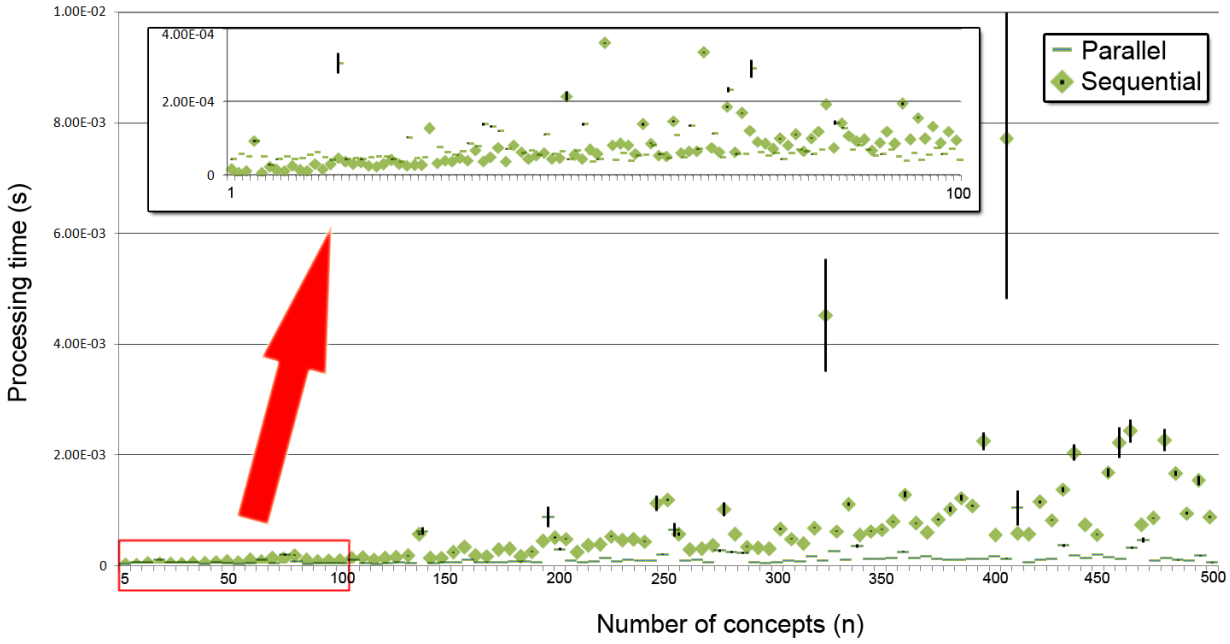

Figure 4.2: Average and standard deviation of the time to apply a transfer function, either sequentially or in parallel, depending on the number of concepts. The inset shows all number of concepts from 1 to 100 included. The main figure shows concepts from size 5 to 500 by steps of 5. Each data point was computed over 500 repeats. The benchmarking script is available at https://osf.io/qyujt/.

This is avoided by reading the file once and then distributing the computations among the c cores available on the current machine. The FCMs have the same causal structure and initial value for the nodes: they only differ in the values of their links. Thus, we treat all FCMs as equal[1] by dividing the 2^k experiments into c sets and allocating each set to a core.[2]

To further minimize I/O operations, we do not *explicitly* send to each core a list of all experiments in its set. Instead, we only send the ID of the starting experiment and how many experiments are in the set. A core uses these two numbers to iteratively find the combination of parameter (link) values corresponding to an experiment and generate the next one, until all of them have been performed. To do this, each experiment has an ID (from 0 to $2^k - 1$ included) that encodes the

[1]Some causal values may lead an FCM to converge faster, thus ending a simulation run sooner. However, additional research would be needed to reliably identify such initial settings and to use that information when distributing the computations. Benefits may be limited as a single FCM runs within milliseconds as discussed in Section 4.4.

[2]We rely on the `multiprocessing` library for Python, which uses *logical cores* to take advantage of hyperthreading. For example, a workstation with 20 physical cores running two threads each will be seen as having 40 logical cores.

combination of parameter values. To obtain these values, the ID is decoded in binary (over k bits), where the n-th bit specifies whether to use the low (0) or high (1) value for the n-th link. For example, if we have an FCM with 3 links, we perform $2^3 = 8$ experiments numbered from 0 to 7. Experiment 3 would be coded as 011 in binary, stating that the first link must be assigned its low value (0), the second its high value (1), and the third its high value (1).

4.3.2.3 Steps 4 and 5: Analyzing Experimental Results to Simplify the Model

The results of the simulations performed by one core produce a table similar to Table 4.1. The simplest approach would be to aggregate all these tables and analyze them (i.e., calculate the effects). However, this would have two significant drawbacks. First, given that we have an exponential time complexity to simulate the combinations of an FCM with k factors, we would also have an exponential space complexity when attempting to store and aggregate all raw results. Examples are provided at `https://osf.io/qyujt/` with the intermediate output files showing the aggregate tables for k=14 (20Mb file) and k=25 (2.5Gb file). This significant burden on I/O operations would, in turn, slow down the process. Second, aggregating results in a single table with 2^k rows would mean that one core would then analyze the table, which is inefficient and non-scalable. Our solution avoids both drawbacks: we do *not* store and combine raw results. Instead, we use the fact that the analysis can be performed in parallel. That is, the three steps to calculate effects (Section 2.2) are performed on each core based only on its simulation outputs. Each core thus produces a partial weighted sum. These sums are then added on one core, which computes the final results; how much of the variance in each stabilizing FCM concept is produced by each of the links and their interactions.

Once the variance has been analyzed, we have precisely quantified the importance of each link. If a link has an almost negligible contribution to the variance, it means that using a high or low

value for that link has a negligible impact on simulation outputs. Such links can thus be set to any value within the range (e.g.,, the mid-point), which allows for a careful simplification of the model. However, defining the boundary as being negligible depends on the simulation setting. The output of one model may be used to continuously control a sensitive device (e.g.,, medication dosage) while the output of another informs a binary decision (e.g.,, whether to embark on a given policy or not). Thus, we employ a user-defined threshold T_{var}. A link is simplified if *the sum* of its contributions to variance exceeds T_{var} for at least one stabilizing concept. For example assume $T_{var} = 5\%$ and three links A, B, C with the following contributions: $A : 91.8\%$, $B : 3.2\%$, $C : 2\%$, $AB : 1\%$, $AC : 1\%$, $BC : 0.5\%$, $ABC : 0.5\%$. The total contributions involving each link are: $A + AB + AC + ABC = 94.3\% > T_{var}$ for A, $B + AB + BC + ABC = 5.2\% > T_{var}$ for B, and $C + AC + BC + ABC = 4 < T_{var}$ for C. In this situation, the link C would be simplified by setting its value rather than using a range. We note that our approach produces a very conservative estimate (allowing us to confidently set a link's value), as it counts contribution through interactions with the *same* impact as contributions from the link alone.

4.4 Experimental evaluation

The contributions of this chapter (as stated in Section 4.1) are threefold: proposing a new method and implementation to simplify models and evaluating it with respect to (i) how much of a model can be simplified and (ii) how long it takes to perform the computations. The previous section detailed the method and its implementation, thus this section is devoted to the evaluation. The case studies for the evaluation are three models (Table 4.2) published from 2012 to 2015. They were designed for widely different contexts and have from 8 to 25 links. The files specifying each case study are available at `https://osf.io/qyujt/`, using the format shown in Figure 4.1 (first step). The two smaller models are also shown in Figure 1.4.

Table 4.2: Characteristics of the Three Case Studies.

#links	#nodes	#stabilizing nodes	Description	Reference
8	6	1	Controller for a security system. Concepts include intruder localization accuracy or distance to potential intruder.	[160]
14	7	2	Supervises radiation therapy. Concepts include the final dose delivered to the patient or the dose prescribed.	[161]
25	15	1	Evaluates the trajectory of individuals in terms of obesity. Involves food intake, depression, or stress.	[49]

To evaluate how much of each model could be simplified, we used two thresholds: $T_{var} = 5\%$ and $T_{var} = 10\%$. We also examined how to stop a simulation in two ways. First, using the default scenario designed by each model's authors, in which only a designated subset of nodes must stabilize, and second, an extreme setting in which we required *all* nodes to stabilize. This setting makes it much harder to simplify a link's value because a link can now affect many more outputs other than the designated subset of nodes. This extreme was chosen to assess whether, even in the most draconian situation, we would still be able to simplify a model. Results are shown in Table 5.6. In the most common setting (column 1: $T_{var} = 5\%$, stabilize the author-designated subset) we observe that about half of each model (42% to 50%) can be simplified. That is, we do not need to use a range for about half of the links and can set their value without significantly changing simulation outcomes. A more lenient setting (column 3 with increasing $T_{var} = 10\%$) allows us to simplify 52% to 62% of a model. An extremely strict alternative (column 2 where *all* concepts are potential outcomes) has varied effects: no link can be set in the smaller model (versus half using a subset of nodes as outcomes), while almost half of the links can still be set in the larger model.

Detailed results for the most common setting are provided in Figure 4.3, where each link is labeled with its total contribution to variance (including interactions). While we may intuitively

Table 4.3: Number and Percentage of Links that can be Set to a Single Value, Depending on the Threshold for Contributing to Variance and Whether All or a Subset of Factors had to Stabilize.

Case Studies	Threshold 5% Subset stabilizing	Threshold 5% All stabilizing	Threshold 10% Subset stabilizing	Threshold 10% All stabilizing
1, 8 links	4 (50%)	0 (0%)	5 (62%)	0 (0%)
2, 14 links	6 (42%)	2 (14%)	8 (57%)	2 (14%)
3, 25 links	12 (48%)	11 (44%)	13 (52%)	11 (44%)

expect links directly impacting the stabilizing concept(s) to be more important, Figure 4.3a exemplifies that it isn't necessarily the case: two of the links directly impacting the one stabilizing concept have a contribution of 0.52% and 0.75% so they can be simplified (colored green). Similarly, we see in Figures 4.3b-c that links directly impacting one of the stabilizing concepts (dashed) are not necessarily important. This shows that simulating the system as a whole is important to adequately identify the contribution of each link.

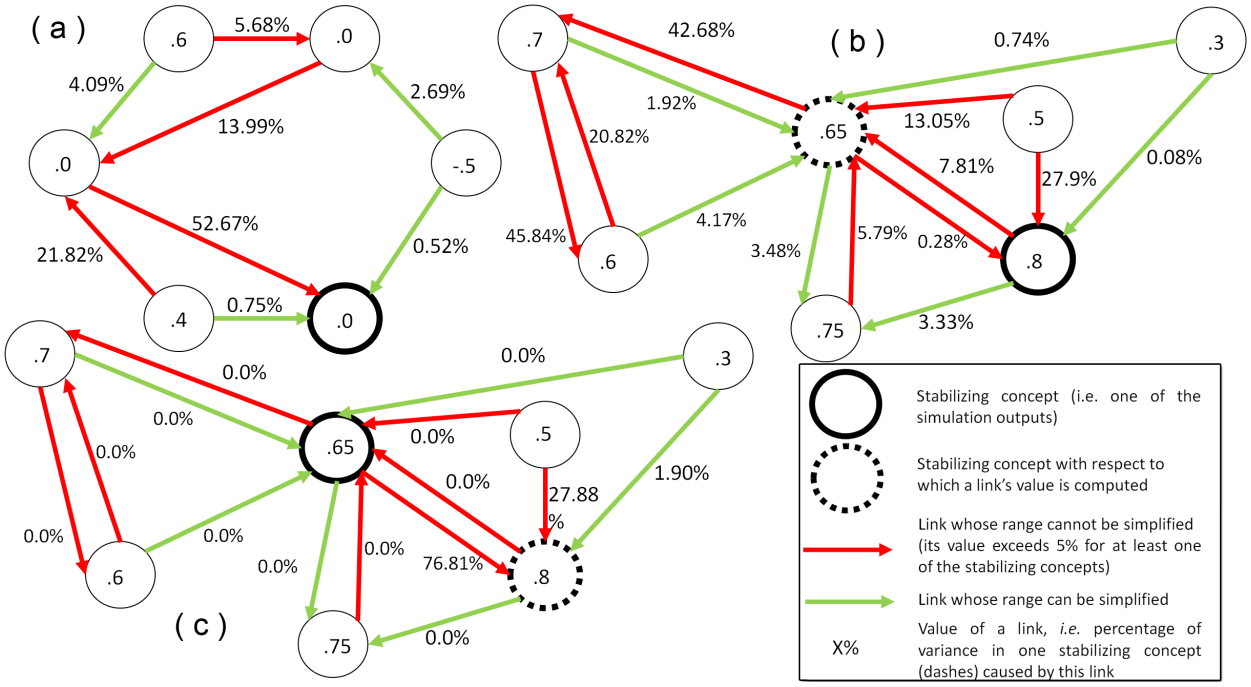

Figure 4.3: Results on case studies 1 and 2 with a threshold of 5% and a subset of concepts stabilizing (first result column in Table 5.4). Values are rounded to the nearest two decimals.

Table 4.4: Three Levels of Hardware Used to Perform Experiments

CPU	Memory	Description (year of purchase)
Intel Core i5-3210M 2.50 GHz 2-core	6Gb 800Mhz (DDR3)	Personal laptop (2012)
2 Intel Xeon E5-2650 v3 2.3 GHz 10-core	32Gb 2133MHz (DDR4)	Professional workstation (2015)
20 nodes, each with 2 Intel X5650 2.66 GHz 6-core	20 nodes, each with 72 GB RAM	High-performance computer cluster (2012)

Although results suggest that our approach can simplify half of a model under typical requirements, our approach has significant computational costs. To evaluate in which settings these costs may constitute an obstacle to the use of our approach, we computed the time it took to perform the computations on different architectures ranging from an older personal laptop to a modern workstation or a high-performance computer cluster (Table 4.4). The time for the first two case studies (8 links and 14 links) was computed over 100 repeats, with the criterion that all nodes should be stabilized. On average for the smaller case study, it took $0.0238s \pm 0.0243$ on the laptop, $0.0058s \pm 0.0057$ on the workstation, and $0.0029s \pm 0.0141$ on the cluster. For the medium-sized model, it took on average $5.5911s \pm 2.0671$ on the laptop, $0.4615s \pm 0.1962$ on the workstation, and $0.0928s \pm 0.0949$ on the cluster. Each of the individual 100 timings for both cases are available online. Models with up to 14 links can thus be simplified almost instantaneously, even when using entry-level laptops. The largest model with 25 links was run once on the cluster, where it took about 28 hours using only the author-defined subset of nodes to stabilize and over 260 hours when all nodes had to stabilize. It is thus increasingly infeasible to provide immediate model simplification as the number of links grows beyond 14. Beyond this, a model can still be simplified (at least up to 25 links), but not immediately and only using specific hardware.

4.5 DISCUSSION

We used different hardware configurations to evaluate the time necessary to perform all steps leading to the simplification of models 1 (8 links) and 2 (14 links). The first model could be simplified in 0.02 seconds (on average) using an entry-level laptop, while the second one took 5.59 seconds (on average) with the same laptop. This has important practical implications. It means that facilitators can run a workshop, gather ranges for each link from the participants, and immediately identify the important edges. This, in turn, can be used to guide the conversation with participants on important edges, for example by devoting more time to discussing their possible ranges. Model 3 (25 links) could only be handled using a computing cluster, where it took about 28 hours when considering only one node as output or over 260 hours when all nodes could be output. This suggests that models between 15 and 25 links can become intractable on entry-level hardware. We thus conclude that our approach is feasible on *typical* model sizes and standard hardware but becomes intractable as we start handling large FCMs.

While many FCMs tend to be relatively small, there exists a few with a large number of links. For example, a radiotherapy model was proposed with 66 links [161]. This would lead to 2^{66} experiments, which is over 70 quintillion experiments. Similarly, our expanded version of the obesity model had 269 links and would count among the largest FCMs created to date [31]. Given their already massive number of links, such FCMs could benefit from a simplification. Efficiently simplifying large FCMs should thus be an object of future research. We note that, while small FCMs are typically generated within a facilitated workshop setting, large FCMs can be created over weeks through an asynchronous collaborative process. Such a setting does not require the ability to simplify a model using entry-level hardware within seconds. Instead, it would tolerate longer processing times (on the order of days to weeks) and may utilize higher-end hardware such as a computing cluster.

While our approach focused on simplifying the ranges associated to links, there can also be ranges associated with nodes. For example, stakeholders may consider that a variety of settings exist for a given concept, or policymakers may use population distributions rather than an "average person" when initializing a concept. Our approach can also be employed for this setting; instead of generating 2^k experiments for the k links, we would generate 2^{k+n} experiments by also taking into account the n nodes. Results would then show which nodes and links can be simplified. The main consequence is on computational requirements, which in turn impacts the type of hardware that one needs. In model 1, there would be $2^{k+n} = 2^{8+6} = 2^{14}$ experiments, which is feasible on an entry-level laptop within seconds. However, starting with model 2, we would already have $2^{14+7} = 2^{21}$ experiments, which may require a computing cluster and days. As mentioned above, this would benefit from research on simplifying larger models.

The endpoints are typically the only information provided about a range [160, 161]. Our 2^k factorial design of experiments thus assumes that these endpoints are representative. However, additional data may be available. For example, when each participant has to assign a weight to a link through a questionnaire, the set of questionnaires provides a distribution about the link. If such distributions are heavy tailed, then the two endpoints of the range may not be representatives and could instead be outliers. As simplifying a model should make use of all available information, future research may explore alternative design of experiments using distributions rather than just endpoints.

4.6 CONCLUSION

Fuzzy cognitive maps (FCMs) can represent the mental model of stakeholders as a causal network equipped with an inference engine. Stakeholders may have widely different views, feel unsure about specific aspects of a complex problem, or wish to capture when the evidence is

inconclusive. This can be represented by using a range for each causal link, rather than assigning it a specific weight. The issue then becomes, if all links have a range of values, which ones should we use when running a simulation? Similarly, stakeholders often need to identify the parameters that matter when they design interventions on complex problems. This requires knowing which ranges are unimportant and which ones strongly impact the model output. Ranges have so far been dealt with by simplifying them to their mid-point [159], but this systematic simplification does not take into account which ranges matter and which ones do not. In this chapter, we presented, implemented, and evaluated a new approach to identify which ranges are important and to simplify models accordingly. Our approach uses a 2^k factorial design, where k is the number of links, to evaluate the contribution of each link to variance in the output (i.e.,, final value of selected nodes in the model). Given the exponential cost of a 2^k factorial design, our implementation uses parallelism not only to run the simulations but also to analyze the variance. Our evaluation assessed (i) whether our approach can identify unimportant links in previously published models and (ii) whether our approach is feasible on typical model sizes and standard hardware.

Results from three previously published models show that, under a commonly used setting, almost half of the models can be simplified (42% to 50% of the links contribute to less than 5% of the variance). A more lenient setting allows us to simplify an additional 10% of the model (52% to 62% of the links contribute to less than 10% of the variance). Even in the extreme case where the lowest tolerance threshold is used *and* all concepts of the model are considered as possible outputs, some models may be simplified. This setting, however, exhibits more variability (0% to 44% of the links). We thus conclude that, on previous models, our approach can successfully identify unimportant links. A closer investigation as to which links could be simplified further demonstrated that they could not be straightforwardly identified, for example by assuming that links directly impacting an output would be important to that output's variance.

CHAPTER 5

SIMPLIFYING UNCERTAINTY IN FUZZY COGNITIVE MAPS UNDER TIME LIMITATIONS

In the previous chapter, we presented a design of experiments that allows us to identify the significant edges in an FCM. As the number of edges increased, the number of experiments that needed to be run increased exponentially. In this chapter, we propose a design of experiments to approximate the significance of each edge to the simulation outcome while running fewer experiments.

My contributions consisted of:

(i) Implementing a method to determine the number of simulation runs necessary for the output to be within a given confidence interval.

(ii) Automatically generating a fractional factorial design in which the effects of the main factors are not confounded.

(iii) Applying the method on three previously published case studies, with one serving as validation.

5.1 Introduction

When facilitating a session with stakeholders (e.g., for participatory modeling), a modeler has a limited amount of time to go through several complex tasks. In the case of fuzzy cognitive maps (FCM), the meeting is used to determine the concepts and causal relationships in the FCM. However, not all causal relationships in an FCM are equally important. It is vital to be able to identify the important relationships while all participants are still available to provide input. To identify these important edges, a full factorial analysis can be performed as in [101]. This approach faces several limitations. The number of experiments needed grows exponentially with the number of edges, meaning this only works on small FCMs. Once the number of edges reaches around 15 to 25, the number of experiments becomes unmanageable. This cannot be solved by distributing computations on a high-performance cluster as was done in [101], as this only slightly increased the amount of manageable edges, and it is not a resource available to all meetings with stakeholders (e.g.,, insufficient funds to purchase cloud computing capacity or lack of knowledge in using it). Since it is not possible to use a factorial design of experiments (DoE) to find the important edges in a reasonable amount of time, we need to approximate. When a full factorial design is not possible, we instead explore performing a fractional factorial design that can estimate the important factors while running a smaller number of the total experiments. The main contributions in this chapter are:

1. Examining if a fractional factorial design can be applied to approximate a solution.
2. Creating the best approximation based on the number of computations that can be performed.

The remainder of this chapter is organized as follows. In Section 5.2 we provide a technical background on factorial design and fractional factorial design of experiments. Then we describe our methods for implementing a fractional factorial design in Section 5.3. Next, we describe our case studies to demonstrate our design as well as validation in Section 5.4. In Section 5.5 we

explain the results and discuss limitations in the study. Finally, we have concluding remarks in Section 5.6.

5.2 Background

To understand a model, more work needs to be done than just observing a single simulation run of the model. The parameter values (i.e., inputs) need to be drawn from all ranges for which the model is applicable in order to best understand the causal effects between factors. For this, you need to change the inputs in methodical ways for a series of experiments, where each experiment is a run of the simulation [120]. In general, this means we modify the set of inputs *x* and observe how it alters the response variable *y* (i.e., the output). There can be multiple response variables that we are interested in monitoring. The objective of the experiments can be a solution to determine [120]:

- The most influential factors on our response variable
- Where to set the influential x's so that y is almost always near the desired nominal value
- Where to set the influential x's so that variability in y is small
- Where to set the influential x's so that the effects of the uncontrollable variables $z_1, z_2, \ldots, z_q$ are minimized.

In our case we want to find the most influential factors on our response variable. For this, a commonly applied method is a factorial design of experiments.

5.2.1 Factorial Design

When conducting a study on the effects of two or more factors, a factorial design is the most efficient type of experiment [120]. A factorial design means that we run an experiment for all possible combinations of factor values. That means if we have two factors *A* and *B*, with x and y

levels respectively, we would conduct $x \times y$ experiments. The effect of a factor is determined by the change in the response (output) variable by modifying the level of a factor. This is called the main effect because it refers to the primary factors in the experiment [88]. For example, if changing *A* from its low level to its high level causes a change in the response variable, that is the main effect of *A*. However, we may find that a difference in the response between levels of one factor is not the same at all levels of other factors. That is, a factor may not change the model's output to the same extent depending on how other factors are set. This indicates *interaction* between factors [120]. For example, consider that we run our experiment with our factors A and B and test a low level and high level for B. If changing the level of B alters the main effect of A, that means there is some interaction effect between A and B.

In a 2^k factorial design, where k is the number of factors, each factor will only have two levels: high and low (Table 5.1). The -1 indicates to use the low level, and the +1 indicates to use the high level. The final column shown is the value of the response variable. The bottom two rows show the influence of each factor of the response. To determine the significance of a factor, we multiplied the response variable by the factor level for that experiment. We then sum that result across all experiments. For example, in Table 5.1 we get the effect of factor A to be 80 by

$$(-1 * 14) + (1 * 22) + (-1 * 10) + (1 * 34) + (-1 * 46) + (1 * 58) + (-1 * 50) + (1 * 86)$$

The factorial design is very computationally expensive though. The number of experiments to run grows exponentially with the number of factors, quickly becoming unmanageable even if a single experiment takes very little time. When the set of experiments is manageable, all effects are normalized so the conclusion of a factorial experiment tells us the contribution (%) of change in the output due to all factors and interactions considered.

Table 5.1: Design for a 2^3 Factorial Design.

I	A	B	C	AB	AC	BC	ABC	y
1	-1	-1	-1	1	1	1	-1	14
1	1	-1	-1	-1	-1	1	1	22
1	-1	1	-1	-1	1	-1	1	10
1	1	1	-1	1	-1	-1	-1	34
1	-1	-1	1	1	-1	-1	1	46
1	1	-1	1	-1	1	-1	-1	58
1	-1	1	1	-1	-1	1	-1	50
1	1	1	1	1	1	1	1	86
320	80	40	160	40	16	24	9	Total
40	10	5	20	5	2	3	1	Total/8

5.2.2 Fractional Factorial Design

When we do not have the time or resources to run a full factorial design, an alternative is to run a fractional factorial design. A fractional factorial design will reduce the number of runs needed to approximate the effects of factors on our response variable [120]. This means we ran 2^{k-p} experiments, where p is set by the user to reflect how many experiments are feasible given the resource constraints. The quality of the approximation is best when higher order effects are seen as negligible [88]. Fractional designs are based on three key ideas [120]:

1. *Sparsity of effect principle*. This is driven by a few main effects and low-order interactions.
2. The *projection principle*. Fractional designs can be projected into larger designs in the subset of significant factors. For example, a design of resolution R can be projected to a full factorial design of 2^{R-1} for every subset of factors.
3. *Sequential experimentation*. It is possible to combine the runs of two fractional factorial designs to create a larger design to better estimate factors and interactions.

For example, consider that we have 2^3 experiments to run as we did in Table 5.1, meaning we have three factors A, B, C. However, we can only run four experiments or 2^{3-1} experiments (k =

Table 5.2: Resolution III 2^{3-1} Fractional Factorial Design, with Defining Relation $I = ABC$

A	B	C=AB
-1	-1	1
1	-1	-1
-1	1	-1
1	1	1

3, p = 1), also known as one-half fraction design [120]. In this case we call *ABC* our generator (also called a 'word') for this design. By accounting for the identity *I*, we have I = ABC as the defining relation for the design [120]. Table 5.2 estimates the main effects with the following equations [120]:

- $\mathbb{A}$ = $\frac{1}{2}$(a - b - c + abc)
- $\mathbb{B}$ = $\frac{1}{2}$(-a + b - c + abc)
- $\mathbb{C}$ = $\frac{1}{2}$(-a - b + c + abc)

We denote by $\mathbb{F}$ the effects of the set of factor F on a selected output. These effects, as shown in the previous equations, are computed through linear combinations. The examples show how to compute $\mathbb{A}, \mathbb{B}$ and $\mathbb{C}$. We observe that $\mathbb{A} = \mathbb{BC}$ since $A = BC$. This means that main effect A is *aliased* (also known as *confounding*) by interaction effect *BC* [120, 88]. The $\frac{1}{2}$ is called the *principle fraction* [120].

These aliases determine the *resolution* of the design. The resolution identifies which interactions and effects are aliased with each other. As exemplified in Table 5.2, the resolution is depicted by roman numerals (Table 5.3). The resolution of a design is the length of the shortest defining word. Thus for our earlier example, we would have had a resolution III design because $C = AB$, thus our defining relation is $I = ABC$. The length of the defining relation is 3; therefore, the design has resolution III. The higher the resolution, the more interaction effects that can be accounted for.

Table 5.3: Description of the Possible Resolutions in a Fractional Factorial Design

Resolution	**Ability**	**Example**
III	Estimates the main effects. However, main effects can be confounded with two-factor interactions.	2^{3-1} with defining relation I = ABC
IV	Estimates the main effects unconfounded. Estimates the two-factor interaction effects, but these can be confounded with each other.	2^{4-1} with defining relation I = ABCD
V	Estimates main effects unconfounded by three-factor (or less) interactions. Estimate two-factor interaction effects unconfounded by two-factor interactions. Estimate three-factor interaction effects, but these may be confounded with two-factor interactions.	2^{5-1} with defining relation I = ABCDE
VI	Estimates main effects unconfounded by four-factor (or less) interactions. Estimate two-factor interaction effects unconfounded by three-factor (or less) interactions. Estimate three-factor interaction effects, but these may be confounded with other three-factor interactions.	2^{6-1} with defining relation I = ABCDEF

5.3 Methods

5.3.1 Determining *p*

If we had enough time on a given hardware, we would perform all 2^k experiments to deal with k factors. However, when there is not enough time, we need to consider 2^{k-p}, where p is the number of factors we are going to alias, such as factor C in Table 5.2. The amount of time we can afford thus drives the value of p. To determine *p* we want to determine the maximum number of experiments that can be run in a predetermined amount of time. For this we follow these steps:

1. Initialize the FCM and simulate it 100 times to find the average simulation time.
2. Use the confidence interval method in[155, p. 154] [155] to determine the average simulation time within a 95% confidence interval.
3. Determine how many experiments can be run given (i) the average time a single experiment takes, (ii) the number of factors k, and (iii) the user's defined time limit on a given hardware.

Step 3 will give us some value j where 2^j is how many experiments can actually be run in our time limit. For example, in case study 1 we have $k = 25$. That means we have to run $2^{25} = 33,554,432$ experiments for a full factorial design. After calculating our average simulation time of 0.003744 in steps 1 and 2, we determine we can only run between 2^{19} and 2^{20} experiments in our given time constraint of 20 minutes. Thus, $j = 19$ as we 2^{19} is the most complete design we can run. We know p is their difference, so $p = k - j$ or $p = 25 - 19$, so $p = 6$.

5.3.2 Acquiring Resolution IV

Our goal is to estimate the main effect of each edge. To avoid confounding the main effect with the second-level interaction, we need to have at least a resolution IV design. To have a resolution IV design means we need a defining relation such that I = ABCD as shown in Table 5.3. This means that none of our main effects are confounded with the second-degree interaction, but second-degree interactions are confounded with each other. To guarantee resolution IV, we need the shortest defining relation to be four characters long ($I = ABCD$). Thus for any factor we must approximate for, we need a three-character-long alias. For example, with I = ABCD, if we need to alias D, we get the defining relation of D = ABC. This will guarantee a resolution IV design since the identifying word is length 4. For larger examples we generated $k - p$ factors. So if $k - p = 7$ with $k = 9$ and $p = 2$, we generated the string $'abcdefg'$ and created p permutations of that string of length 3. We used those permutations to alias our needed p factors. For example, from the previous string we need to alias factors i and j.They can be aliased with the length 3 permutations of our original string $i = abc$ and $j = def$.

5.4 Case Studies

To evaluate our method, we have found three case studies, described in Table 5.4. These case studies were published between 2010 and 2012. These FGCMs were too large to perform a full factorial analysis on a regular laptop, as may be found in a session during participatory modeling. We note that case study 1 was used in Chapter 4 with a full factorial analysis, via a high-performance computing cluster. We thus use it as validation for our method. That is, for a reasonable value of p (i.e.,, not an excessive approximation), we checked whether the significant edges found in the full factorial case are the same as in the fractional factorial case studied here.

Table 5.4: Characteristics of the Three Case Studies

#Links	p	#Nodes	#Stabilizing nodes	Description	Reference
25	6	15	1	Evaluates the trajectory of individuals in terms of obesity. Involves food intake,depression, or stress.	[49]
44	26	17	3	IT implementation risks in both open source and commercial IT projects.	[159]
66	47	16	3	A radiotherapy treatment planning procedure	[161]

Table 5.5 compares the significant edges found by both methods. It can be seen that the fractional only identified one of the significant edges identified by the full factorial method and two incorrect edges. Thus we were unsuccessful in validating the fractional factorial in approximating the full factorial design. Table 5.6 shows the number of edges that can be simplified in an FGCM. However, since the design was not fully validated, this may cause us to classify significant edges as non-significant.

5.5 Discussion

In this chapter we explored using a fractional factorial design rather than a full factorial design to find the important causal relationships in our FGCMs. This would allow modelers to identify important relationships while working with stakeholders under constraints of time and equipment. Thus, the modeler could have the participants discuss the important relationships in more detail to reduce uncertainty. We attempted to create a resolution IV fractional factorial design, which means that the main effects would not be confounded with other main effects or the second-level interactions. Our validation case (case study 3 in [101]) showed that the fractional factorial could not identify the important relationships that were identified by the full factorial design. By observing the results of the full factorial design in Table 5.7, we can see the main effects are not the primary

Table 5.5: Comparison of Significant Edges Found with a Full Factorial and Fractional Factorial Design Using a 5% Threshold to Determine the Significant Edges

	Full Factorial	Fractional Factorial
Significant Edges	Socio economic status→ Psychosocial Barrier Age→ Psychosocial Barrier **Fitness→ Exercise** Exercise→ Physical health Female Gender→ Depression Exercise→ Fitness Exercise→ Fitness Exercise→ Obesity Age→ Fitness Exercise→ Depression Female Gender→ Psychosocial Barrier Obesity→ Fitness Belief in personal responsiblity→ Weight discrimination	**Fitness→ Exercise** Stress→ Food Intake Obesity→ Weight discrimination

Table 5.6: Number and percentage of Links that can be Set to a Single Value, Depending on the Threshold for Contributing to Variance and Whether All or a Subset of Factors had to Stabilize

Case Studies	Threshold 5% Subset stabilizing	Threshold 5% All stabilizing	Threshold 10% Subset stabilizing	Threshold 10% All stabilizing
1, 25 links	22 (88%)	9 (36%)	22 (88%)	14 (56%)
2, 44 links	35 (79%)	25 (56%)	35 (79%)	30 (68%)
3, 66 links	43 (65%)	5 (7%)	50 (75%)	15 (22%)

means of determining the important factors, but their interactions are. This means that in larger FCMs the interactions may be as important as the main effects. To account for interactions accurately we would need a design of at least resolution V. To increase the resolution we would need to increase the length of the defining relation. This means we have a wider variety of possible aliases for factors. There are further limitations that if handled may also allow us to identify the important relationships in a short amount of time. First, if the simulations were run on GPUs instead of CPUs we would be able to run significantly more experiments, and GPUs can now be found in the laptop that a modeler may bring to a meeting. This means we could reduce the *p* for our 2^{k-p} and get a closer approximation and perhaps more reliable work. Furthermore, the fractional factorial is the natural extension of the factorial design if we want to run it in less time. However, it is not the only possible solution.

We could instead attempt the latin hypercube sampling-partial rank correlation coefficient and the sensitivity heat map method [193], factor screening methods, or hybrid designs. These methods can be further expanded for better approximations of the full factorial design. These methods are defined as factor screening methods, which means they are meant to identify the key factors in a model. One popular method is *sequential bifurcation* (SB) [94]. This method relies on two key assumptions [94]:

1. The model can be approximated by a first-order polynomial with noise.

$$y = \beta_0 + \beta_1 \times x_1 + \cdots + \beta_k \times x_k + \epsilon$$

   With our factors $x_j (j = 1, \ldots, k)$ and β is the effect of a factor.

2. All main effects have known signs and are non-negative.

$$\beta_j \geq 0, (j = 1, \ldots, k)$$

Table 5.7: Significance of main effect of case study 1 according to a full factorial design.

Edges	Significance
Female Gender→Psychosocial Barrier	0.2708
Socio economic status→Psychosocial Barrier	0.2708
Age→Psychosocial Barrier	0.2708
Age→Fitness	0.2708
Fitness→Exercise	0.2708
Exercise→Fitness	0.2708
Exercise→Physical health	0.2708
Exercise→Depression	0.2708
Exercise→Obesity	0.2708
Obesity→Fitness	0.2708
Obesity→Physical health	0.2708
Female Gender→Depression	0.2708
Depression→Antidepressants	0.0356
Antidepressants→Obesity	0.0275
Antidepressants→Food Intake	0.0481
Food Intake→Obesity	0.0
Stress→Physical health	0.0
Stress→Food Intake	0.0
Female Gender→Fatness percieved as a negative trait	0.0
Socio economic status→Fatness percieved as a negative trait	0.0
Fatness percieved as a negative trait→Weight discrimination	0.0
Belief in personal responsiblity→Weight discrimination	0.0806
Weight discrimination→Depression	0.0
Obesity→Weight discrimination	0.0
Stress→Depression	0.0176

The first assumption means that all of the factors can be ranked in the order of most important to least important. The second assumption ensures that main effects do not cancel each other out and that we can assign high/low values for experimentation [94]. The efficiency of the method is measured by the required number of simulation runs used to determine the important factors in the model. The method can be visualized in Figure 5.1 and runs as follows [94]:

1. All factors are aggregated into a single group. Then we run two extreme scenarios with all values set low and then all values set high. If the difference in the output is greater than some δ, determined by the experimenter, then we proceed to step 2.

2. The group is split into two subgroups.

3. The two subgroups are tested for significance with the method described in step 1. If a group is not found to be significant, then it is discarded. Significant subgroups return to step 1.

4. All individual factors that are not in subgroups identified as unimportant are estimated and tested.

$$\beta_j = \frac{w_j - w_{j-1}}{2}$$

with w_j being the observed simulation output with the factors 1 through j set to their high levels and the remaining factors set to their low levels.

We can further increase the efficiency of SB (decrease the number of simulated factors) by [94]:

- Labeling individual factors and placing them in increasing order of importance. By utilizing prior knowledge we can begin to cluster known important factors to allow for the creation of fewer subgroups.

- Placing similar factors into the same subgroups.

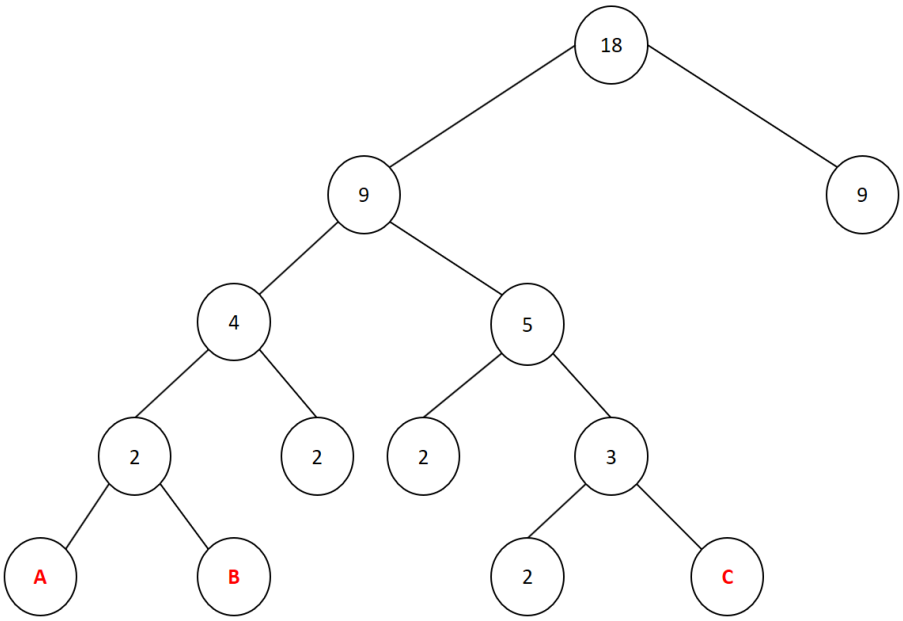

Figure 5.1: For a model with 18 factors. We bifurate (split) it into two subgroups each with nine factors. One group is found insignificant and discarded while the process continues on the other group. We identify three important factors.

- Splitting subgroups such that the number of factors in new subgroups is a power of two, e.g., group of 12 factors is split into subgroups of 8 (2^3) and 4 (2^2) factors.

Figure 5.1 shows that there are splits, and the execution leads to a tree. If one of the two groups is never significant, then your tree simplifies to a path, in which case there are no gains. But the more frequently two groups are significant (= both are retained), the more you split. These splits can be run in their own threads.

In addition to fractional factorial, latin hypercube, and sequential bifurcation, hybrid methods have been introduced to approximate a factorial design. Recent work presents a hybrid method called the *sliced full factorial-based latin hypercube design* (sFFLHD) [187]. This method starts with an orthogonal array, thus all rows are independent of each other. The array is sorted in ascending order by the value in the first column. We then partition the array into equally sized smaller grids (slices) by the content of the first column so all similar values are in a slice (Equation (5.2)).

The remaining columns in a slice contain a different *latin hypercube*. For a latin hypercube, the input distribution of a factor is divided into N equal parts that are sampled once [113] (Equation (5.1)). In this case X_{kj} is the sample from a distribution for the components X_k. That is, each X_k is an input for the factor which is labeled as X_i.

$$
\begin{aligned}
&X_{kj}, (j = 1, \ldots, N) \\
&X_k, (k = 1, \ldots, k) \in X_i, (i = 1, \ldots, N)
\end{aligned}
\tag{5.1}
$$

$$
Z = \left|\begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & 1 & 1 & 2 \\
0 & 2 & 2 & 1 \\
\hline
1 & 0 & 2 & 2 \\
1 & 1 & 0 & 1 \\
1 & 2 & 1 & 0 \\
\hline
2 & 0 & 1 & 1 \\
2 & 1 & 2 & 0 \\
2 & 2 & 0 & 2
\end{array}\right| . \tag{5.2}
$$

When enough batches of experiments have been run of the latin hypercubes, we will have the design of a full factorial design. As such, more experiments are run until a stopping criterion is reached, which means that a full factorial latin hypercube design has been created. The experimenter can choose to continue the process indefinitely beyond the stopping criterion and stop at a later time [187]. sFFLHD in high-dimensional examples was seen to have a lower root mean squared error than other sequential methods [187]. sFFLHD can be run in parallel by putting each

batch of runs into a separate core and then aggregating the results afterward since the run of one group of experiment does not effect the others.

5.6 Conclusion

In this chapter we explored using a fractional factorial design to approximate the factorial design implemented in Chapter 4. We found that the fractional design, while able to execute all experiments in the desired time frame, was not able to accurately approximate the important causal relationships in our models. We then suggest other possible methods to determine the important factors in a limited amount of time.

CHAPTER 6

SHOULD WE SIMULATE MENTAL MODELS TO ASSESS WHETHER THEY AGREE?

This chapter of the thesis focuses on the simulation output of several different groups of individuals. We have discussed that FCMs are an excellent method for participatory modeling. A single FCM can represent the causal map of either an individual or a group. What we have not yet examined is whether the structure of the FCM will assist us in determining the output of the simulation between different groups. In this chapter we examine that as separate groups create different FCMs that have the same central concepts, will they also agree on the ranking of the final outputs?

This chapter will be published in the Proceeding of the 2018 Spring Simulation Conference:

- **Eric A. Lavin**, Philippe J. Giabbanelli, Andrew T. Stefanik, Steven A. Gray, and Robert Arlinghaus. Should We Simulate Mental Models to Assess Whether They Agree?.

My contributions consisted of (i) designing and implementing the case studies and validation scenarios, (ii) computing the correlations between centralities and simulation outputs.

6.1 Introduction

Interventions on complex systems may require coordination among individuals to identify a common purpose, establish suitable and efficient scenarios, and finally implement them. This is grounded in the theoretical concept of *policy coherence*, emphasizing the need to share outcomes across sectors and ensure consistency [33]. A coherent policy is articulated around a specific issue and requires that individuals work together through forms ranging from cooperation (sharing a mission but keeping resources separate) to collaboration (sharing a mission and resources). Modeling approaches have been proposed to capture how individuals can work together, given their different capacities and associated time scales (e.g., organizational and political cycles) [56]. This assumes that a shared mission has already been identified and focuses on the logistics of accomplishing it. At the other end of the spectrum, various modeling approaches are routinely used to elicit individual goals and preferences [65]. For instance, integrated assessment (IA) is a popular approach for socio-environmental problems, which seeks stakeholder participation and involves modeling approaches such as systems dynamics, bayesian networks, agent-based models, and fuzzy cognitive maps [57, 104]. Between eliciting individual preferences and managing the logistics when pursuing a common mission, there is one essential step: identifying shared preferences such that we can agree on the mission.

Qualitative methods are a popular approach to identify shared preferences (e.g., using depth interviews or focus groups), but they can be resource intensive (e.g., training staff to run and analyze sessions, compensating participants for their time) and involve non-trivial logistics given the need for some synchronicity (e.g., staff and one or multiple participants interact at the same time). In contrast, we are interested in fully automatic, scalable quantitative methods. The objective is to identify agreements using few resources (hence making it scalable) and allowing for asynchronous participation. Achieving scalability and supporting asynchronous participation are essential to sup-

port the participation of many stakeholders, residing in any place. That is, it supports the uptake of citizen science in a more inclusive form of the decision-making process, which has known benefits in areas such as conservation planning [66].

To automatize the full process leading to the identification of shared preferences, we need the first part of the process to already be automatic. That is, among all possible processes to elicit individual preferences, we are limited to those which are scalable and asynchronous. Our contribution consists of finding similarities among stakeholders by analyzing individual preferences. Finding similarities among individuals is related to the concept of a *cultural model*. Cultures are about the social behavior and norms of societies; they tell us what is shared across individuals. In Axelrod's landmark model for "The Dissemination of Culture" individuals have a list of features which can take different values (e.g., what hats they like to wear). To elicit such models automatically, participants can simply email their features and associated values, then commonalities are straightforwardly defined as having the same value on the same feature. However, this approach only considers isolated facts; it does not tell us *why* individuals have a given value or the consequences of having it. For instance, one individual may have a black hat because it is a religious requirement, and another may have a black hat because it is a fashion trend. The same observation is thus rooted in different causal antecedents, and the "commonality" is superficial. We can go one level higher by (i) eliciting a network from each individual and (ii) comparing whether the same nodes also share antecedents and consequences. This is known as the problem of comparing causal maps [59]. However, research on systems thinking shows that this is still not a complete comparison. Indeed, Donella Meadows and others [112, 115] have posited that isolated nodes are around the lowest level, while links (i.e., antecedents and consequences) are slightly higher, but the highest level is the *paradigm*, which mobilizes the entire system. Assessing links is thus insufficient to really conclude that individuals will agree on a course of action. For instance, given a cycle with four links, individuals could agree on three links but disagree on the fourth one, and that is enough to create complete disagreement (e.g., by turning a reinforcing loop into a balancing one).

To mobilize a whole system, we can extract mental models from individuals and then simulate a variety of what-if questions on the models. If they generally provide the same simulation output, then we can conclude that the individuals themselves are in broad agreement.

Several studies have shown that mental models suitable for simulations could be extracted in the form of fuzzy cognitive maps (FCMs) [4, 89], which offers some scalability [54] and fully allows for asynchronous participation using platforms such as `MentalModeler` [67]. However, running several what-if scenarios over hundreds or more FCMs is computationally intensive [101] and requires to identify the "right" scenarios. In this chapter, we ask the fundamental question: Can we tell that simulation models (in the form of FCMs) are going to agree only based on their structure, without actually running simulations? This would provide a more accurate assessment of similarity between individuals than current structural approaches (only comparing edges instead of using the whole model) and would present performance advantages over conducting many simulations. We propose to use centrality metrics, as they mobilize the whole network structure. To identify which centrality metric(s) are suitable, we examine whether agreeing on centrality is correlated with agreeing on which simulation outcomes are important across different scenarios. Our experiments are performed using real-world data, collected for the socio-ecological problem of fishery management. The dataset provides a large sample of 264 mental models from different types of stakeholders.

In short, our main contributions are as follows:

1. We propose a new approach to investigate agreements between mental models by using network centrality instead of individual edges (known to be potentially uninformative) and without resorting to a large number of simulations (causing a computation burden).
2. We demonstrate our approach on a large real-world dataset with different types of stakeholders.

3. We show that one centrality metric correlates with simulation outcomes, thus a structural analysis may suffice to comprehensively assess agreements between models instead of running simulations.

The remainder of this chapter is organized as follows. In Section 6.2, we provide a technical background on fuzzy cognitive maps and network centrality. In Section 6.3, we explain our simulation setup by summarizing the context of our dataset, testing its quality using extreme settings, and identifying four what-if scenarios. Our simulation scripts are hosted on the Open Science Framework at `https://osf.io/qyujt/`, within Comparing FCMs. In Section 6.4, we report our results on the correlation between simulation outcomes and selected centrality metrics. The implication of these results for participatory modeling and policy coherence are discussed in Section 6.6. We conclude by summarizing our achievements in identifying agreements across participants more comprehensively than previous structural approaches without extensive simulations.

6.2 Background

6.2.1 Fuzzy Cognitive Maps

6.2.1.1 Why Are Fuzzy Cognitive Maps Used As Mental Models?

A mental model broadly captures an individual's "perspective" on a specific topic. We thus need to refine what goes into a 'perspective'. On the one hand, individuals could provide statements backed by abundant evidence [24], such as using meta-reviews to justify why they think that a system works in a certain way. On the other hand, a perspective may involve information of any type or quality, which would include sensorial imprints (e.g., "Do you like the smell of this fish?"). The perspective that we seek to capture in a mental model lies in between; we are interested in

meanings and understandings rather than sensorial imprints, but we do not require that they are supported by evidence [55]. For instance, we can ask a fisherman how he thinks that fishing is going to impact a fish population, and the answer neither requires evidence nor should it focus on senses.

Research in cognitive sciences has long been concerned with how individuals form and organize knowledge in their memory and how it can be elicited. To capture an individual's mental model, we thus need to tap into the specific part of his or her memory where the knowledge of interest is held. The intention is often to tap into the *semantic* memory, which holds the individual's conceptualization of the world [13], rather than in the episodic memory, which is about specific events. This results in a mental model where individuals generalize and abstract from their experiences, rather than focusing on a specific instance. The practical question is thus, How do we elicit a perspective through an individual's semantic memory? And, as a corollary, To what extent is the resulting mental model an artifact of the elicitation process? McNeese and Ayoub suggested that, if the elicitation method closely aligns with how human knowledge is internally represented, then (i) its elicitation would be easier, (ii) the representations would tie into and open up spontaneous access to associated knowledge within memory, and (iii) the knowledge of one person can be compared with another to search for invariance [114]. Since semantic memory provides functional *relationships* between objects, some elicitation methods (e.g., mind mapping and semantic networks) focus on capturing interrelatedness. They thus produce *networks* or *maps*. In other words, if mental models are published and shared in the form of maps, it owes to the fact that we seek to capture semantic memory whose structure is network based.

A relationship from A to B implies that changes in A *may* have effects on B. This may be characterized with parameters such as intensity of the change, time, and previous history. For instance, merely suggesting to someone not to fish a juvenile pike may have limited to no effect. In contrast, strongly insisting on the illegal or damaging effects of fishing non-harvestable fish may have a larger effect. This effect may not be immediate (i.e.,, a lag may occur), as one may

finish fishing but change strategy the next time. Such dynamic relationships can be *represented* using system dynamics (SD). However, eliciting SD models can be challenging, as individuals may not be readily able to provide a clear number or to precisely estimate the duration of a time lag. This may require in-depth interviews or focus groups to create graphical functions [183]. As we are interested in a more scalable process (see Introduction), we may thus focus on only acquiring some parameters characterizing a relationship. A fuzzy cognitive map (FCM) represents neither time nor history, as only the current value of a concept can influence the next one. The focus is on capturing the intensity of the change in a way that is intuitive to participants and can easily be done through online questionnaires. Rather than being forced to provide numbers that they may not have thought of, participants evaluate the strength of a relationship using linguistic variables such as "Low", "Medium", "Strong". Then fuzzy logic is used to associate a membership function to these variables, and the defuzzification process eventually results in a number used by the model [4, 54].

6.2.1.2 How Do Fuzzy Cognitive Maps Work?

Definition 6.2.1. At a given iteration t, a fuzzy cognitive map $F^t = (V^t, E, f)$ is formed of:

1. A set V^t of n nodes, representing concepts of the system such as states or entities. The values can change through the simulation, and V_i^t indicates the value of node i at iteration t. Values are bounded in the range [0, 1], where 0 indicates the absence of a concept and 1 indicates that it is maximal.
2. A set of weighted, directed edges E representing causal relationships. Weights and directionality are typically represented using an adjacency matrix A, where $A_{i,j}$ is the weight of the link from i to j. If $A_{i,j}$ is negative, then an increase in i causes a decrease in j. Conversely, if $A_{i,j}$ is positive, then an increase in i causes a decrease in j.

3. A clipping function f, also known as transfer function. As nodes' values are updated, the clipping function ensures that the result remains in the range [0, 1].

Rather than representing units, an FCM models the extent to which concepts are present. For example, in an ecological model, a node could stand for the density of fish in a given space, where 0 means no fish and 1 means a maximal density. That value cannot go beyond 1 since it is maximal, and the density cannot be negative either. Concept values are updated over discrete iterations, which do not intend to correspond to real-world time steps. The update uses an inference function, where the next value of each concept V_i^{t+1} is computed in Equation (1.1) based on (i) its current value V_i^t, (ii) the value of connected concepts V_j^t, (iii) the strength of these connections $A_{j,i}$, and (iv) the clipping function f.

The update is repeated (Figure 1.2) until a subset $S \subseteq V$ stabilizes. For instance, if the intention of the mental model is to summarize how various factors impact the long-term presence of fish in a lake, then we update the model until the "fish" concept changes by less than a very small ϵ. Formally, the simulation ends when equation (1.2) is satisfied. Since there is no randomness in equations (1.1) and (1.2), an FCM does not need repeated simulation runs. The choice for f and ϵ is normally left to the modelers. The function f has to be monotonic (to preserve the order of nodes' values) and a sigmoidal function is recommended for planning scenarios [177]. The hyperbolic tangent $tanh$ is widely used [71, 111] and will be employed here as well. The value of ϵ is rarely reported, as it is assumed to be a small constant. We use $\epsilon = .001$. The objective of the elicitation process is to obtain the set of nodes and edges from participants. Initial values for the nodes are set when applying the model to a specific scenario/what-if question.

FCMs are implemented by several libraries and software such as `MentalModeler` [67]. For intensive simulations, we use the Python library introduced in [101].

6.2.2 Network Centrality

From a network perspective, an FCM is composed of nodes and edges. Since we do not look at changes in nodes' values, we use V rather than V^t to refer to the set of nodes. Intuitively, the centrality of a node represents its "importance". The convention is that the higher the centrality, the more important the node.

Definition 6.2.2. Node centrality c induces at least a semi-order on the set of nodes, allowing us to conclude that $x \in V$ is at least as central as $y \in V$ with respect to centrality c if $c(x) \geq c(y)$. Generally, the difference or ratio of two centrality values cannot be interpreted as a quantification of how much more central one node is than the other [96].

Different centrality indices can produce very different ranges of values; they can be between 0 and 1 when counting what *fraction* of some network dynamics involves a node or significantly larger than 1 when counting the *total* contribution of a node. In addition, Definition 6.2.2 emphasizes that the relative difference between the values may not be meaningful; values show whether a node is more or less central than another one, but not *how much* more or less. To address these two potential issues when dealing with multiple centrality indices, we normalize all of them into rankings. That is, the node with the highest centrality value now takes the value of 1, the next highest takes the value of 2, and so on.

Centrality indices can evaluate different aspects to decide that a node is "important". Reachability indices focus on the cost that it takes for a node to reach others, while flow indices assess how much traffic may pass through a node, and feedback indices compute a node's importance based on its neighbor's importance [96]. As we cannot presume that one type of centrality will best correlate with simulation outcomes, we use different types, all implemented using the `NetworkX` library in Python:

1. For reachability indices, we use *closeness centrality*, which is the inverse of the sum of the shortest distances between a node and all others.
2. For flow indices, we use *betweenness centrality* and its variant *load centrality* [61]. Betweenness is the fraction of shortest paths between all pairs that go through a given node. Load is different from betweenness as shown in [15]; it considers that traffic splits equally between neighbors, thus leading to a different approximation of how traffic accumulates.
3. For feedback indices, we use *Katz centrality*. A node accumulates influence through incoming neighbors and to a lesser extent through farther away nodes. A damping factor α adjusts the influence of nodes based on the distance. We use the default value $\alpha = 0.1$ in `NetworkX`.
4. For a local index that depends only on a node's neighbor, we use the *degree*, defined as the total number of edges incidents to a node.

6.3 Simulation setup

6.3.1 Dataset

The northern pike is an important recreational fish in Germany [5]. Recreational fishermen using a rod and a line are referred to as *anglers*. The dynamics of the northern pike population depend on several socio-ecological factors, and anglers are one of the agents of ecological change. Appreciating this role through an ecological understanding helps to identify preferred management policies. To examine the anglers' ecological understanding, we recently collected their FCMs [64]. The process started by mailing a solicitation to all 461 angling clubs in the German state of Lower Saxony, of which 41 agreed to participate, and 17 were retained after criteria such as interest to improve their knowledge of pike fisheries management. Up to 25 anglers were accepted as

volunteers from each club. Within a club, we report separately on club managers, water managers, and "regular" anglers (i.e., without a management role for the water or the club), as their role may come with a different level of familiarity regarding socio-ecological dynamics. After excluding participants who provided missing data, we had a total of 136 anglers, 79 club managers, and 32 water managers. To compare their maps with those produced by biologically trained academic experts, we also collected the maps of 17 "volunteers comprised of researchers, post-docs and PhD students employed [at] a research institute specializing in fish ecology and biology, and an inland fisheries institute" [64]. We thus had a total of 264 FCMs coming from four categories of participants. Note that the previous study reported a higher sample size as it also counted participants who were excluded from the analysis. While the *cognitive* map (i.e., the network without its fuzzy quantification) of our problem was not included here due to space limitations, it can be found in [64].

If participants created FCMs without constraints, they may use different terms to refer to the same concept (e.g., number of pikes, pike population, amount of pikes) [59]. We set a model boundary and limited linguistic variability using a list of standardized terms, which allowed to assess whether a concept in one map is the same as in another map. Our standardization involved "independent focus groups with anglers and fishery experts about the biology and fisheries ecology of pike" [64], resulting in a list of 19 concepts (Table 6.1). We created *aggregate* FCMs representing the mental model held by a group of stakeholders in addition to the 264 *individual* FCMs. A node or edge appears in the aggregate if it was expressed by any stakeholder. An edge's value is obtained using a weighted average for the value of the edge in all maps that contained it (Figure 6.1). The FCM of a group thus uses the FCMs of all of its stakeholders. This leads to five aggregate maps: all anglers, all water managers, all club managers, all experts, and all non-experts (i.e., anglers, water managers, and club managers) [64].

The focus of these FCMs is the relationship between the ecosystem (including anglers) and the population of pike over the legal size limit (meaning anglers can legally keep the fish if caught).

Table 6.1: The Values of Our 19 Concepts were Set Depending on the What-If Scenario or Validation Test

	Scenarios				**Validation**	
Concepts	1	2	3	4	1	2
Spawning Grounds	.6	.35			.6	.1
Angling Pressure	.75			.5	.2	.9
Refuges	.25	.5	.25		.6	.1
Pike Population (adults over legal size limit)	.31				.9	
Stocked Pike (adults over legal size limit)	.21					.4
Stocked Pike (under legal size limit)	.25		.5	.25		
Baitfish	.15					
Other Predatory Fish	.15					.8
Algae	.18				.5	.1
Depth of Water	.3					
Wild Pike (under legal size limit)	.5					
Emergent Riparian Plants	.45				.5	.1
Benthic Invertebrates	.25					.1
Zooplankton	.27				.5	.1
Submerged Aquatic Plants	.45				.5	.1
Cormorant	.1					.8
Plant Nutrients	.18					
Turbidity of Water	.2				.1	.8
Surface Area of a Body of Water	.75					

We briefly detail the main concepts of the system. Pike maintain their population by reproducing in a spawning habitat, which requires vegetative mats to keeps eggs in a better oxygenated area [18]. Thus, by keeping highly vegetative areas, we expect an increase in juvenile pike (that cannot be fished for legally), which ultimately increases the number of harvestable pike. This vegetation further increases the amount of zooplankton and feed for other (small) fish, which become part of the food supply for pike. However, pikes are not the top of the food chain in their ecosystem. Other fish threaten pike and their offspring, and these predators also benefit from having more small fish. Furthermore, pikes of all sizes are also threatened by cormorants, a group of aquatic birds. On the human side, strong fishing habits can greatly reduce the amount of fish in a lake. To compensate,

the pike population is boosted through a practice called *stocking*, which consists of placing pike of legal size and below into the environment.

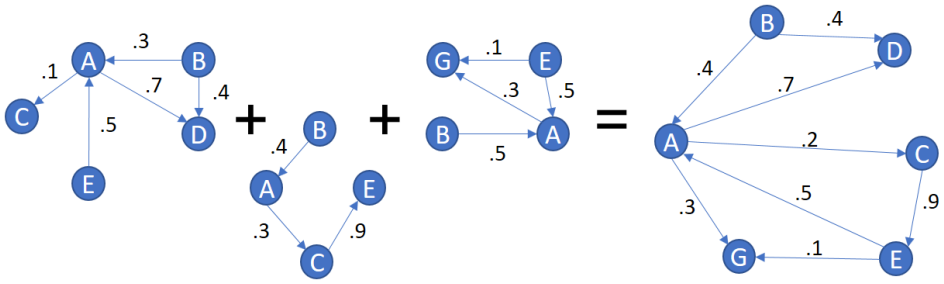

Figure 6.1: Aggregation of three individual FCMs into one, using a weighted average.

6.3.2 Validation

Our dataset only serves to test our proposed methods to correlate simulation outcomes and centrality. Nonetheless, it is important to test its validity, as an improper data collection process may introduce bias in our results. A broad test used for complex models [183], and particularly for FCMs [52], consists of establishing extreme scenarios where there is no doubt in what the correct outcome should be. Starting with a very large population for pike (set to 0.9) and an abundance of other species in the system (other predatory fish = cormorant = 0.8), our two selected scenarios should clearly increase (validation 1) or decrease (validation 2) the total pike population of legal size.

In the validation expected to *increase* the pike population, we use conditions optimal for the breeding of pike as studied by Casselman and Lewis [18]. We want high amounts of refuge and spawning areas (spawning grounds = refuge = .6). We want the water to be clean and not deep, thus we want lower values for the turbidity (turbidity = .1). Along with that, we need a sufficient number of plants for food and cover (algae = emergent plants = submergent plants = zooplankton =

.5). We keep plant nutrients low (plant nutrients = .2) and avoid overfishing through a low angling pressure (angling pressure = .2).

In the validation expected to *decrease* the pike population, we set poor conditions for pike to thrive by following the opposite logic of the previous validation. We make it difficult for pike to reproduce (spawning grounds = refuge = .1). The water will be dirty and deep with a large number of predators (cormorant = predatory fish = turbidity =.8). Food and cover are scarce (algae = emergent plants = submergent plants = zooplankton = invertebrates =.1). Finally, the lake will be significantly used for fishing, thus it has a large number of anglers fishing (angling pressure = .9) and involves stocking (stocked adult pike = .6).

As both validation scenarios use an expert understanding of the system, we apply them to the expert maps. That is, we expect that the expert FCMs resulting from our data collection process will make the same conclusions as derived from the scientific literature. Two-thirds of the expert maps (64.70%) produced the expected increase in scenario 1 and the expected decrease in scenario 2. This validates our data collection process, which was applied to all stakeholders. We note that, while the aggregate FCM of all experts is obtained through a standard process, it predicts an increase in pike in both scenarios (although less so in scenario 2). Thus we report our results on the aggregate FCMs as well as across all individual FCMs.

6.3.3 What-if Scenarios

We implemented four what-if scenarios, representing possible actions to manage the pike population: increase spawning habitat (1), increase refuge (2), increase juvenile stocking (3), and decrease angling pressure (4). Scenarios 1 and 2 have been proposed in different studies as possible actions to manage a stocked lake for pike fishing [18, 149] while scenario 3 was discussed in

our previous work [64]. A study found that a high angling pressure could reduce the fish population by half, thus scenario 4 considers alleviating this pressure [121].

All scenarios are applied in the context of an "average" initial situation. In short, we consider a lake with a medium amount of fish per hectare (which does not attract many predators), a supportive environment for growth of the pike population, and a heavy fishing presence. Specifically, we first examined the characteristics of lakes that are stocked with pike for fishing. The average density of pike in a stocked lake is 31 per hectare, thus we set our value to .31 [149]. Research further shows the average stocking proportionality to be 21 per hectare, so we set it to .21 as well. We assume that there are more pike below the legal fishing size (value of .5), since larger pike are exploited at 2 to 9 times greater rates than smaller pike [149]. To create a supportive environment, we use clear water (low turbidity value of .2), a medium depth of .4 and a non-dense amount of submergent and emergent vegetation (both at .45) [18]. Studies mention the spawning areas for pike to be in shallower water [18], which competes with human desire for shallow areas near the shore. These areas provide refuge from predators and a stable breeding ground. This competition leads us to assume the pike only have a moderate amount of space for spawning and refuge. Thus, we set these values to .25 and .35 respectively. For surface area, we place the value as .75. Since the dynamics of food supply lead to having fewer predators than prey in an ecosystem, we set other predatory fish to .15 and cormorant to .1. Angling pressure was set to .75 to represent the large amount of fishing in the areas.

6.4 Results

Our goal is to assess whether there is a correlation between some centrality metric(s) and simulation outcomes across all four scenarios and stakeholder groups. That is, nodes are ranked by both centrality and simulation outcomes, and we compute the correlation between these two

rankings. To avoid being sensitive to outliers, we use the robust Theil–Sen estimator to fit a line (also known as Kendall robust line-fit method). Instead of the typical [-1, 1] range for correlations, the estimator returns 1 when there is a perfect fit *regardless* of whether it is a positive or negative correlation. Negative or null values indicate a poor fit. Figure 6.2 exemplifies the Theil–Sen estimator. All results are available on `https://osf.io/qyujt/`.

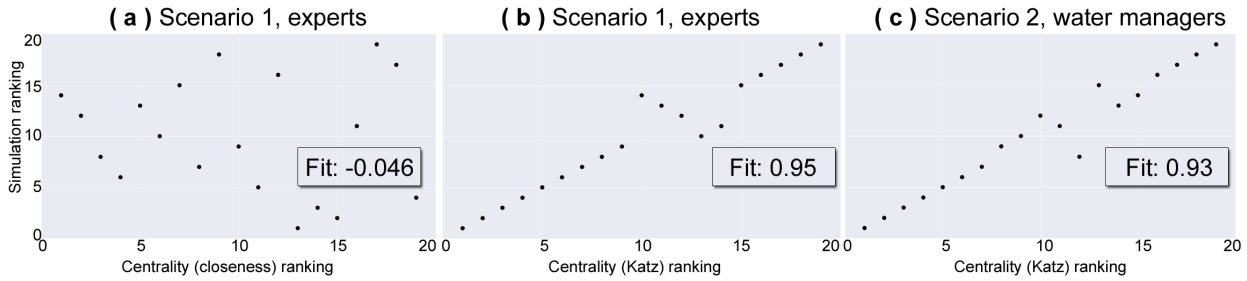

Figure 6.2: Correlation Between Centrality and Simulation Ranking for Different Scenarios and Groups.

Sample results on the aggregate maps are shown in Figure 6.3. The results show that Katz centrality has a very high fit across all scenarios and aggregate maps (from .93 for water managers to 1 for anglers), whereas all other centrality metrics have a very poor fit. As we emphasized in Section 6.2 that the aggregate map can behave differently from most of its underlying FCMs, we also assessed the fit across individual maps for each scenario. Results for scenario 2 are shown in Figure 6.4, and results for other scenarios (available online) support the same two observations. First, the four centrality metrics that show no correlation at the *aggregate* level also show no correlation at the *individual* level. Second, in contrast to the aggregate level, Katz centrality only exhibits a (small) correlation when applied to the expert group and behaves the same as other metrics for other groups. The implication is that Katz centrality is suitable to compare aggregate maps but may only be used to compare individual maps in very specific cases. A possible explanation is as follows. Aggregate maps are highly connected, since an edge exists if a single participant used it. These many connections provide abundant feedback, which the Katz centrality seeks to measure. In contrast, individual maps (and particularly maps of non-experts) have much less feedback.

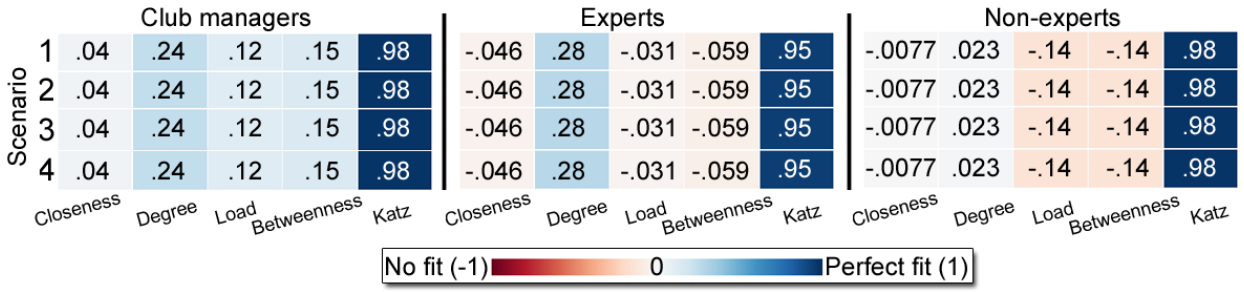

Figure 6.3: Across All Scenarios and Groups of Stakeholders, only Katz Centrality had a Very High Fit.

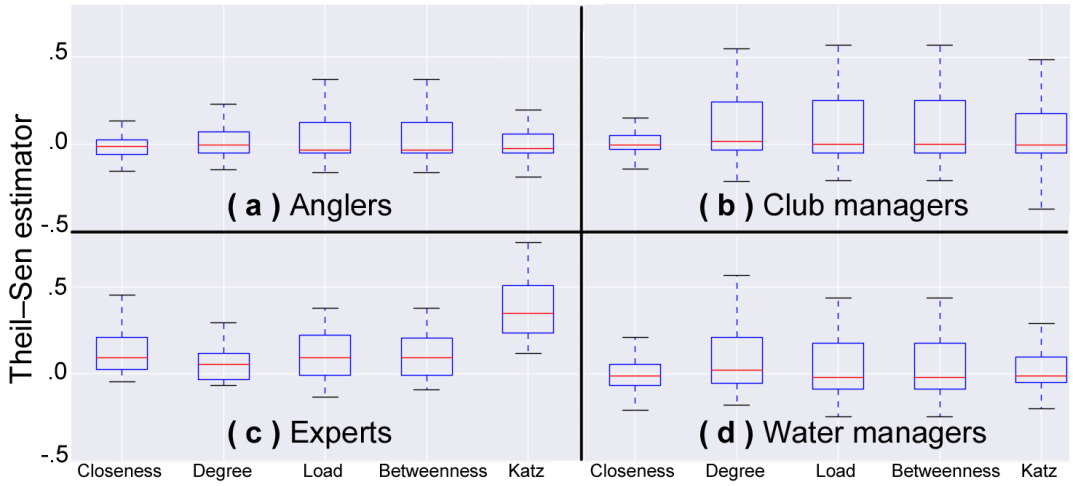

Figure 6.4: Fit Across All Stakeholders in Each Group for Scenario 2

6.5 Discussion

To know whether two simulation models agree, we can check whether they yield similar outputs across an extensive and computationally demanding set of simulations. We examined whether analyzing the *structure* of the models was sufficient, instead of simulating them. Our application context is participatory modeling, where the mental models of stakeholders are represented as fuzzy cognitive maps (FCMs) and we need to identify whether stakeholders agree before pursuing a course of action. As systems science emphasizes that contrasting systems in terms of their independent nodes or edges does not fully capture their dynamics, we instead searched for an

analysis that involves the whole model structure. Since FCMs are built on networks, we used different measures of network centrality to take into account reachability, flow, or feedback. We then assessed whether the importance of nodes as judged by the centrality tended to agree with their importance in terms of simulation outputs. Using the mental models of 264 stakeholders from four groups, we found an almost perfect agreement when applying Katz centrality at the level of a group of stakeholders and a moderate agreement when applying it to expert stakeholders only. Only 19 set terms could be used in individual models, which may create more agreement than if we allowed for complete linguistic variability.

The implication is that if two groups of stakeholders view nodes as being central (in terms of Katz), then they would make the same conclusion as to what happens to these nodes across a broad range of scenarios. In practice, we can thus consider that agreeing on centrality is enough to conclude that two groups share a paradigm. The implication to compare individual models is less clear, as centrality was only moderately useful for a single category of participants. Future research would thus need to assess whether other feedback-based measures of centrality provide better results at the individual level. Alternatively, one may allow us to perform a few simulations (instead of none) and use the time series of node values to supplement the purely topological information used by a centrality measure. This is becoming possible as dynamic centrality measures on temporal network data are increasingly available.

6.6 Conclusion

We demonstrated that we do not need to simulate the mental models of groups to know whether they agree, as network centrality can sufficiently characterize these models. However, further research is needed to adequately characterize the mental models of individuals, which still require simulations to be compared.

CHAPTER 7

CONCLUSION

Simulation models are run to perform a multitude of functions. They can be used to better understand a problem or predict outcomes of policies before they are implemented. An exceedingly powerful use for simulation is the reduction of complexity in a problem. Throughout this thesis we have demonstrated that fuzzy cognitive maps (FCM) are a powerful modeling and simulation method in terms of uncertainty. We have shown that presently FCMs are not heavily used in solving complex problems (Chapter 3). We presented a design of experiments to decrease the uncertainty in FCMs and identify the important causal links in the model (Chapter 4). We further explored approximating our factorial design for execution in a limited amount of time for participatory modeling (Chapter 5). Finally, in Chapter 6, we demonstrated that it is possible to examine the graphical structure of separate FCMs to see if they will agree in their simulation outcomes instead of running simulations. We have seen that while FCMs are not commonly used in complex problems, they can deal with the large amounts of uncertainty that is present. In this chapter we discuss some limitations of these works in Section 7.1 and then present possible future work in Section 7.2.

7.1 Limitations

New work is constantly being conducted and published. That means that accurate literature reviews are only relevant for a limited amount of time. While our literature review targeted a specific subset of obesity models, discrete simulations are important modeling and simulation techniques.

These new works may include more FCMs or follow more of the guidelines that we specified. Therefore, while the work was comprehensive at the time it was done, it may be less comprehensive now.

These new works are of course not limited only to obesity models. While fuzzy grey cognitive maps (FGCM) are still relatively new, they are growing as a method for FCMs to better identify the uncertainty on causal relationships. At the time of the work in Chapter 5, there were very few FGCMs in the published literature. This meant we only had a few viable case studies, ranging greatly in size. For example, after the case study with 25 edges, the next smallest FCM was 44 edges, which was intractable with our present hardware and methods. This means we can estimate a limit with the high-performance cluster (HPC) but not test up to demonstrate it.

Another limitation we faced was hardware. We were able to successfully run our code on a HPC; however, all code was executed on CPUs. Simulating FCMs requires repeatedly executing the same matrix multiplication. This is a task which GPUs excel at, which may have extended the number of factors that could be run on a personal machine or a HPC. We were further limited by how we structured the FGCMs. We only account for the two end points since the factorial analysis assumes the end points to be representative. This does not account for the possibility of a heavy-tailed distribution of possible values where the end points are not representative.

With the amount of time needed to run a full factorial design on FGCMs we can only evaluate small FCMs up to around 25 to 30 edges. It is normal to expect larger FCMs than that. For that we presented the fractional factorial design to identify the important causal relationships in larger FGCMs. In exchange for running fewer experiments, the fractional design can only approximate the main effects and not the effects of interactions between edges.

Finally, there are many different centrality measures. In our examination of the graphical structure of FCMs, we attempted to use measures that represent the different families of centrality measurement.

- For reachability indices, we use *closeness centrality*.

- For flow indices, we use *betweenness centrality* and its variant *load centrality* [61].
- For feedback indices, we use *Katz centrality*.
- For a local index that depends only on a node's neighbor, we use the *degree centrality*.

Most of these types of centrality did not find a correlation between the FCMs graphical structure and simulation outcomes (i.e. reachability, flow, and local indices.) for either aggregate or individual maps. This does not mean other centralities in those schools will not perform well. Furthermore, for feedback indices, we only tested Katz centrality. This indicates that we should further examine feedback indices; however, it may be that only Katz provides any correlation or that other feedback indices may perform better.

7.2 Future Work

7.2.1 Given the First Few Steps From the Run of an FCM, Can We Use Time Series Analysis to Predict the Final Step at Stabilization?

When we run FCMs, we run them until they stabilize or for a max number of steps. This is done with the expectation that if the FCM does not converge, we can stop the simulation when we reach a predetermined number of steps. However, the concept not stabilizing in time does not mean the value is not trending towards a stable value (Figure 7.1). It may just not have reached the stable value within range of ϵ yet.

With times series analysis we can predict the final stable value based on the previous observations. This is because time series, unlike many statistical methods, assume dependence between values [168]. This means that the prediction assumes that the previous values will assist in determining the final value. With that we could run FCMs for shorter periods of time and still get the final stable value of our concept of interest.

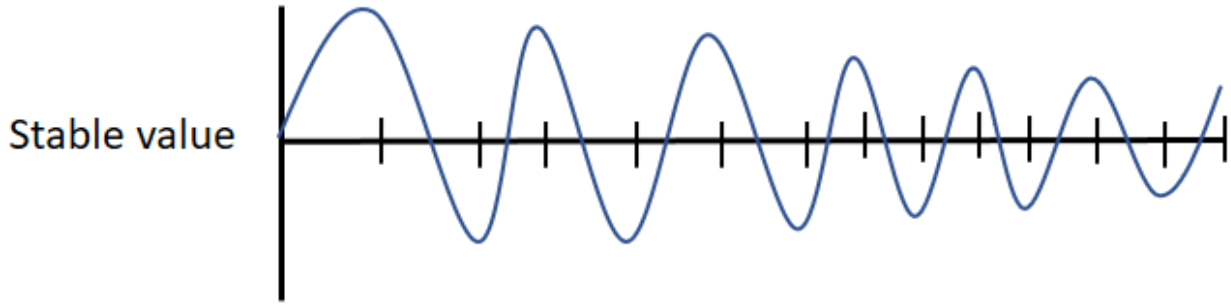

Figure 7.1: A concepts value during a simulation trending towards a stable value but still changing by more than ϵ.

7.2.2 FCMs as Meta-Models for Agent-Based Models

Agent-based models (ABM) are an individual type of simulation model that accounts for heterogeneous populations of individuals [110]. ABMs are comprised of heterogeneous agents and the rules that dictate their actions in the model [117]. However, with the added complexity for accuracy comes the cost of increasing the amount of time needed to execute the simulation. The amount of time needed to run ABMs can be enormous [109], reaching easily into days for a single run when the model is scaled up in size. To deal with the increase in time, we create simpler models of models called *meta-models* [93]. These models are simpler versions of the original model (ABMs in this case) that perform with similar accuracy, within a range of error we determine, but in significantly less time. This way, we can get reliable results in far less time than executing the full model. By using FCMs which are extremely light weight compared to ABMs, we can model similar problems while maintaining the interdependent relationships between the ABMs. By using the agents and the rules that dictate them as concepts, we can recreate the model while maintaining the interdependent relationships between the agents as causal edges. While this can cause a loss in accuracy, we can try to keep the cost within a user-defined tolerance threshold. We can exemplify this through a predator-prey model with grass, sheep, and wolves as agents. We can

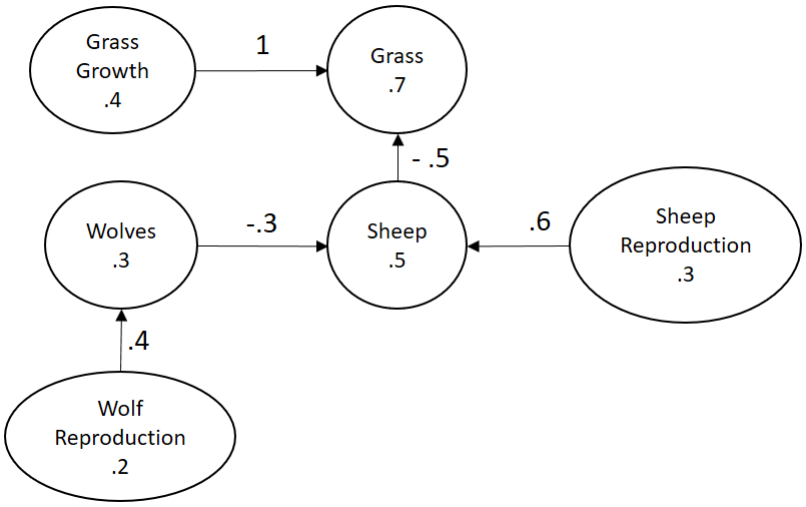

Figure 7.2: The agents of grass, wolves and sheep have been made into individual concepts along with their attributes.

model the agents and their characteristics as concepts in our model. Thus, we can have our grass as a concept but also its regrowth rate as a concept that causes the amount of grass to increase. Furthermore, we are aware that the sheep eat the grass giving us a negative causal relationship between sheep and grass. However, we need to consider the ABM's rules about how much grass the sheep eat. Whether the rules have the sheep all eating just enough to satisfy themselves or eating until they are sick would alter the strength of that causal relationship. Thus instead of a network of interconnected agents, we can build an FCM (Figure 7.2) from the agents and their factors.

7.2.3 Using FCMs as Agents for Heterogeneity in Agent-Based Models

We apply ABMs when we want to represent heterogeneity in a model [60]. That means, individuals do not behave the same. Agents can look very similar componentwise, but they may have different rules governing the interaction between those components. For agents in an ABM to be considered heterogeneous, beyond their target behavior (e.g., eating or drinking), agents usually

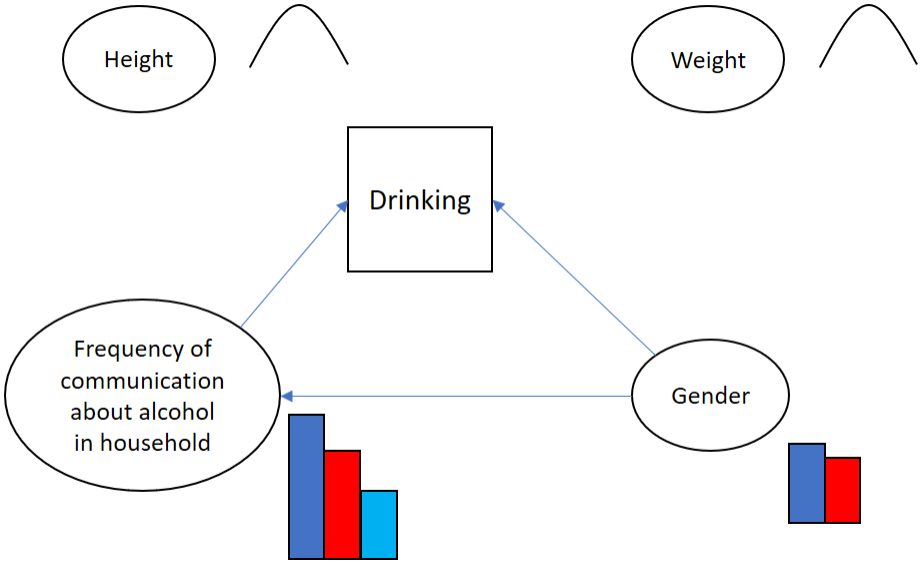

Figure 7.3: Different components that make the drinking agent heterogeneous. You can observe that gender and frequency of communication have an effect on drinking while height and weight do not.

have a set of features or traits (e.g., age, gender), thus there are generally multiple components going into an agent (Figure 7.3).

These components can have a wide possible distribution such as the age or income of an individual. Thus, when we have the multiple components that describe an agent, they are deemed to be heterogeneous. However, not all components may contribute to the usefulness of the heterogeneity. For example, if we were to look at a model of drinking, we may find that using gender to make the agents heterogeneous improves the model, whereas including their age contributes very little or nothing. Thus, we want a method to identify the important components in agent and remove the irrelevant components. We could run a factorial design on the ABM; however, as mentioned previously, factorial analysis is computationally expensive and requires large amounts of time; this is especially true since ABMs are also computationally expensive on their own. Our proposed solution is to model the rules that govern agent actions as FCMs [57] and perform a factorial analysis on the nodes using the same method in Chapter 4. When each type of agent is run by different

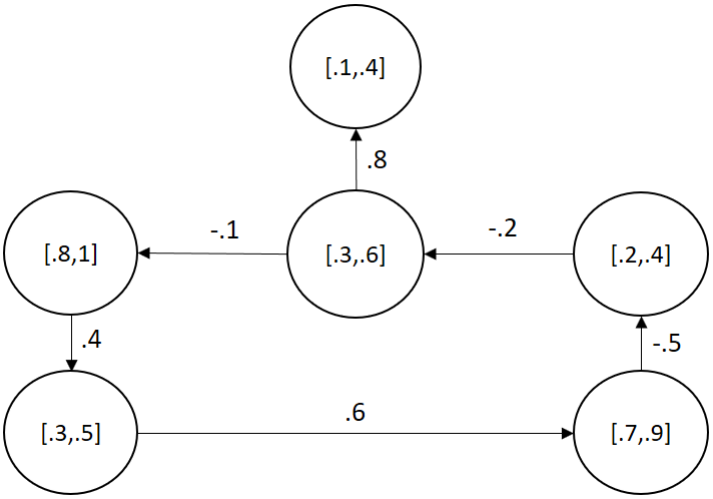

Figure 7.4: Each concept has a range of possible values instead of the edges. Thus we can determine important nodes instead of edges.

rules, we have to create a separate FCM for each type of agent. Each concept in these FCMs will be one of the components considered for heterogeneity in the agents, and the causal relationships will be the rules that determine how the agent acts. We can then perform a factorial analysis on these FCMs and determine which concepts have the greatest impact rather than which edges are important, as done in Chapter 4 (Figure 7.4). This can allow us to determine which components are important to an agent's decisions rather than just take them all and expect the best result.

BIBLIOGRAPHY

[1] K. Alden, J. Timmis, and M. Coles. Easing parameter sensitivity analysis of netlogo simulations using spartan. In *Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems (ALIFE 14)*, 2014.

[2] S. Allender, B. Owen, J. Kuhlberg, J. Lowe, P. Nagorcka-Smith, J. Whelan, and C. Bell. A community based systems diagram of obesity causes. *PLoS One*, 10(7):e0129683, 2015.

[3] M. Amer, T. U. Daim, and A. Jetter. A review of scenario planning. *Futures*, 46:23–40, 2013.

[4] A. Amirkhani et al. A review of fuzzy cognitive maps in medicine: Taxonomy, methods, and applications. *Computer Methods and Programs in Biomedicine*, 142(Sup. C):129 – 145, 2017.

[5] R. Arlinghaus et al. Understanding the heterogeneity of recreational anglers across an urban–rural gradient in a metropolitan area (berlin, germany), with implications for fisheries management. *Fisheries Research*, 92(1):53 – 62, 2008.

[6] A. H. Auchincloss, R. L. Riolo, D. G. Brown, J. Cook, and A. V. D. Roux. An agent-based model of income inequalities in diet in the context of residential segregation. *American Journal of Preventive Medicine*, 40(3):303–311, 2011.

[7] R. Axelrod. *Structure of decision: The cognitive maps of political elites*. Princeton University Press, 2015.

[8] J. Badham. A compendium of modelling techniques. http://i2s.anu.edu.au/sites/default/files/integration-insights/integration-insight_12.pdf, 2010.

[9] D. B. Bahr, R. C. Browning, H. R. Wyatt, and J. O. Hill. Exploiting social networks to mitigate the obesity epidemic. *Obesity*, 17(4):723–728, 2009.

[10] M. Baker. Is there a reproducibility crisis? nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help. *Nature*, 533:452–454, 2016.

[11] K. Ball and D. Crawford. The role of socio-cultural factors in the obesity epidemic. In D. Crawford, R. Jeffery, K. Ball, and J. Brug, editors, *Obesity epidemiology: from aetiology to public health*, pages 105–118. Oxford University Press, 2010.

[12] M. A. Beydoun, L. M. Powell, and Y. Wang. The association of fast food, fruit and vegetable prices with dietary intakes among us adults: is there modification by family income? *Social Science and Medicine*, 66(11):2218–2229, 2008.

[13] J. R. Binder and R. H. Desai. The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15(11):527–536, 2011.

[14] R. Bradfield, G. Wright, G. Burt, G. Cairns, and K. Van Der Heijden. The origins and evolution of scenario techniques in long range business planning. *Futures*, 37(8):795–812, 2005.

[15] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136 – 145, 2008.

[16] A. H. Briggs, M. C. Weinstein, E. A. Fenwick, J. Karnon, M. J. Sculpher, A. D. Paltiel, et al. Model parameter estimation and uncertainty: a report of the ispor-smdm modeling good research practices task force-6. *Value in Health*, 15(6):835–842, 2012.

[17] J. J. Caro, A. H. Briggs, U. Siebert, and K. M. Kuntz. Modeling good research practices – overview a report of the ispor-smdm modeling good research practices task force-1. *Medical Decision Making*, 32(5):667–677, 2012.

[18] J. M. Casselman and C. A. Lewis. Habitat requirements of northern pike (essoxlucius). *Canadian Journal of Fisheries and Aquatic Sciences*, 53(S1):161–174, 1996.

[19] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, July 2007.

[20] P. Craig, P. Dieppe, S. Macintyre, S. Michie, I. Nazareth, and M. Petticrew. Developing and evaluating complex interventions: the new medical research council guidance. *BMJ*, 337:a1655, 2008.

[21] V. Dabbaghian, V. K. Mago, T. Wu, C. Fritz, and A. Alimadad. Social interactions of eating behaviour among high school students: a cellular automata approach. *BMC Medical Research Methodology*, 12(1):155, 2012.

[22] F. Dadaser and U. Özesmi. Participatory management plan for tuzla lake ecosystem: a fuzzy cognitive mapping approach. In *Proceedings of the IV National Environmental Engineering Congress, Mersin, Turkey*, 2001.

[23] F. Dadaser and U. Özesmi. Stakeholder analysis for sultan marshes ecosystem: a fuzzy cognitive approach for conservation of ecosystems. *EPMR2002, Environmental Problems of the Mediterranean Region, Nicosia, North Cyprus*, pages 12–15, 2002.

[24] N. C. Dalkey. The delphi method: An experimental study of group opinion. Technical report, RAND CORP SANTA MONICA CALIF, 1969.

[25] B. Dangerfield and A. Zainal. Towards a model-based tool for evaluating population-level interventions against childhood obesity. In *Procs of the International System Dynamics Conference*, 2010.

[26] K. De la Haye, G. Robins, P. Mohr, and C. Wilson. Obesity-related behaviors in adolescent friendship networks. *Social Networks*, 32(3):161–167, 2010.

[27] K. De La Haye, G. Robins, P. Mohr, and C. Wilson. Homophily and contagion as explanations for weight similarities among adolescent friends. *Journal of Adolescent Health*, 49(4):421–427, 2011.

[28] D. J. DeBrota, S. D. Roberts, J. J. Swain, R. S. Dittus, J. R. Wilson, and S. Venkatraman. Input modeling with the johnson system of distributions. In *Proceedings of the 20th Conference on Winter Simulation*, WSC '88, pages 165–179, 1988.

[29] P. Deck, P. J. Giabbanelli, and D. T. Finegood. Exploring the heterogeneity of factors associated with weight management in young adults. *Canadian Journal of Diabetes*, 37:S269–S270, 2013.

[30] E. M. Douglas et al. Using mental-modelling to explore how irrigators in the murray–darling basin make water-use decisions. *Journal of Hydrology: Regional Studies*, 6:1 – 12, 2016.

[31] L. Drasic and P. Giabbanelli. Exploring the interactions between physical well-being and obesity. *Canadian Journal of Diabetes*, 39:S12–S13, 2015.

[32] J. Du. The "weight" of models and complexity. *Complexity*, 21(3):21–35, 2016.

[33] L. Dube et al. From policy coherence to 21st century convergence: a whole-of-society paradigm of human and economic development. *Annals of the New York Academy of Sciences.*, 1331(1):201–215, 2014.

[34] W. N. Dunn. *Public policy analysis*. Routledge, 2015.

[35] J. Durand. A new method for constructing scenarios. *Futures*, 4(4):325–330, 1972.

[36] C. Eden, F. Ackermann, and S. Cropper. The analysis of cause maps. *Journal of management Studies*, 29(3):309–324, 1992.

[37] A. M. El-Sayed, L. Seemann, P. Scarborough, and S. Galea. Are network-based interventions a useful antiobesity strategy? an application of simulation models for causal inference in epidemiology. *American journal of epidemiology*, page kws455, 2013.

[38] J. Epstein. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4), 2008.

[39] L. Epstein, M. Myers, H. Raynor, and B. Saelens. Treatment of pediatric obesity. *Pediatrics*, 101(3):554–570, 1998.

[40] D. Finegood. The complex systems science of obesity. In J. Cawley, editor, *The oxford handbook of the social science of obesity*, pages 208–236. Oxford University Press, 2011.

[41] M. Flynn. Fitting human exposure data with the johnson s(b) distribution. *Journal of Exposure Science and Environmental Epidemiology*, 16(1):56–62, 2006.

[42] K. Fontaine, D. Redden, C. Wang, A. Westfall, and D. Allison. Years of life lost due to obesity. *Journal of the American Medical Association*, 289(2):187–193, 2003.

[43] L. M. Frerichs, O. M. Araz, and T. T. K. Huang. Modeling social transmission dynamics of unhealthy behaviors for evaluating prevention and treatment interventions on childhood obesity. *PLoS ONE*, 8:1–14, 12 2013.

[44] S. Frood, L. M. Johnston, C. L. Matteson, and D. T. Finegood. Obesity, complexity, and the role of the health system. *Current Obesity Reports*, 2(4):320–326, 2013.

[45] D. Gasevic, C. Matteson, M. Vajihollahi, M. Acheson, S. Lear, and D. Finegood. Data gaps in the development of agent-based models of physical activity in the built environment. *Obesity Reviews*, 11(S1):459, 2010.

[46] B. Gerristen. Reduction of uncertainties through data model integration (dmi), 2011. Accessed: 2016-08-29.

[47] S. B. Gesell, K. D. Bess, and S. L. Barkin. Understanding the social networks that form within the context of an obesity prevention intervention. *Journal of Obesity*, 2012, 2012.

[48] S. B. Gesell, E. Tesdahl, and E. Ruchman. The distribution of physical activity in an after-school friendship network. *Pediatrics*, 129(6):1064–1071, 2012.

[49] P. Giabbanelli, A. Alimadad, et al. Modeling the influence of social networks and environment on energy balance and obesity. *Journal of Computational Science*, 3:17–27, 2012.

[50] P. Giabbanelli and R. Crutzen. An agent-based social network model of binge drinking among dutch adults. *Journal of Artificial Societies and Social Simulation*, 16(2):10, 2013.

[51] P. Giabbanelli et al. *Modelling the Joint Effect of Social Determinants and Peers on Obesity Among Canadian Adults*, pages 145–160. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[52] P. Giabbanelli, T. Torsney-Weir, and V. Mago. A fuzzy cognitive map of the psychosocial determinants of obesity. *Applied Soft Computing*, 12(12):3711–3724, 2012.

[53] P. J. Giabbanelli. *Computational models of chronic diseases: understanding and leveraging complexity*. PhD thesis, Science: Department of Biomedical Physiology and Kinesiology, 2014.

[54] P. J. Giabbanelli. Modelling the spatial and social dynamics of insurgency. *Security Informatics*, 3(1):2, May 2014.

[55] P. J. Giabbanelli and R. Crutzen. Creating groups with similar expected behavioural response in randomized controlled trials: a fuzzy cognitive map approach. *BMC Medical Research Methodology*, 14(1):130, 2014.

[56] P. J. Giabbanelli and R. Crutzen. Using agent-based models to develop public policy about food behaviours. *Computational and Mathematical Methods in Medicine*, 2017:5742629, 2017.

[57] P. J. Giabbanelli, S. A. Gray, and P. Aminpour. Combining fuzzy cognitive maps with agent-based modeling: Frameworks and pitfalls of a powerful hybrid modeling approach to understand human-environment interactions. *Environmental Modelling and Software*, 95(Sup. C):320 – 325, 2017.

[58] P. J. Giabbanelli and V. K. Mago. Teaching computational modeling in the data science era. *Procedia Computer Science*, 80:1968 – 1977, 2016. International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA.

[59] P. J. Giabbanelli and A. A. Tawfik. Overcoming the pbl assessment challenge: Design and development of the incremental thesaurus for assessing causal maps (itacm). *Technology, Knowledge and Learning*, Sep 2017.

[60] N. Gilbert. Agent-based models (quantitative applications in the social sciences) sage publications. 2008.

[61] K.-I. Goh, B. Kahng, and D. Kim. Universal behavior of load distribution in scale-free networks. *Physical Review Letters*, 87:278701, 2001.

[62] S. Gray et al. Modeling the integration of stakeholder knowledge in social–ecological decision-making: Benefits and limitations to knowledge diversity. *Ecological Modelling*, 229(Sup. C):88 – 96, 2012.

[63] S. Gray et al. Mental modeler: A fuzzy-logic cognitive mapping modeling tool for adaptive environmental management. In *Proceedings of the 46th International Conference on Complex Systems*, pages 963–974, 2013.

[64] S. Gray et al. The structure and function of angler mental models about fish population ecology. *Journal of Outdoor Recreation and Tourism*, 12(Sup. C):1 – 13, 2015.

[65] S. Gray et al. Using fuzzy cognitive mapping as a participatory approach to analyze change, preferred states, and perceived resilience of social-ecological systems. *Ecology and Society*, 20(2):11, 2015.

[66] S. Gray et al. Combining participatory modelling and citizen science to support volunteer conservation action. *Biological Conservation*, 208(Sup. C):76 – 86, 2017.

[67] S. Gray, D. Mellor, R. Jordan, A. Crall, and G. Newman. Modeling with citizen scientists: Using community-based modeling tools to develop citizen science projects. 2014.

[68] S. A. Gray, E. Zanre, and S. Gray. Fuzzy cognitive maps as representations of mental models and group beliefs. In *Fuzzy cognitive maps for applied sciences and engineering*, pages 29–48. Springer, 2014.

[69] J. Greener, F. Douglas, and E. van Teijlingen. More of the same? conflicting perspectives of obesity causation and intervention amongst overweight people, health professionals and policy makers. *Social science and medicine*, 70(7):1042–1049, 2010.

[70] V. Grimm, U. Berger, D. DeAngelis, J. Polhill, J. Giske, and S. Railsback. The odd protocol: a review and first update. *Ecological Modelling*, 221:2760–2768, 2010.

[71] P. P. Groumpos and C. D. Stylios. Modelling supervisory control systems using fuzzy cognitive maps. *Chaos, Solitons and Fractals*, 11(1–3):329 – 336, 2000.

[72] K. Hall, G. Sacks, D. Chandramohan, C. Chow, Y. Wang, S. Gortmaker, and B. Swinburn. Quantification of the effect of energy imbalance on bodyweight. *Lancet*, 378(9793), 2011.

[73] K. D. Hall. Predicting metabolic adaptation, body weight change, and energy intake in humans. *American Journal of Physiology-Endocrinology and Metabolism*, 298(3):E449–E466, 2010.

[74] R. Hammond. Social influence and obesity. *Current Opinion in Endocrinology, Diabetes and Obesity*, 17(5):467–471, 2010.

[75] R. Hammond. Appendix a: Considerations and best practices in agent-based modeling to inform policy, 2015.

[76] R. A. Hammond and J. T. Ornstein. A model of social influence on body mass index. *Annals of the New York Academy of Sciences*, 1331:34–42, 2014.

[77] J. Harkaway. Obesity and systems research: the complexity of studying complexities. *Familities, Systems and Health*, 18(1):55–59, 2000.

[78] J. Hartmann-Boyce, D. Johns, P. Aveyard, et al. Managing overweight and obese adults: The clinical effectiveness of long-term weight management schemes for adults. *University of Oxford: Oxford*, 2013.

[79] M. F. Hatwágner, A. Buruzs, P. Földesi, and L. T. Kóczy. A new state reduction approach for fuzzy cognitive map with case studies for waste management systems. In *Computational Intelligence in Information Systems*, pages 119–127. Springer, 2015.

[80] M. F. Hatwágner and L. T. Kóczy. Parameterization and concept optimization of fcm models. In *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*, pages 1–8. IEEE, 2015.

[81] S. Henly-Shepard, S. A. Gray, and L. J. Cox. The use of participatory modeling to promote social learning and facilitate community disaster planning. *Environmental Science and Policy*, 45:109–122, 2015.

[82] A. Hill. Social and psychological factors in obesity. In G. Williams and G. Fruhbeck, editors, *Obesity: Science to Practice*, pages 347–366. Wiley-Blackwell, 2009.

[83] D. M. Hoelscher et al. Incorporating primary and secondary prevention approaches to address childhood obesity prevention and treatment in a low-income, ethnically diverse population: Study design and demographic data from the texas childhood obesity research demonstration (tx cord) study. *Childhood Obesity*, 11(1):71–91, 2015.

[84] J. Homer, B. Milstein, W. Dietz, D. Buchner, and E. Majestic. Obesity population dynamics: exploring historical growth and plausible futures in the us. In *24th International System Dynamics Conference*, 2006.

[85] A. V. Huerga. A balanced differential learning algorithm in fuzzy cognitive maps. In *Proceedings of the 16th International Workshop on Qualitative Reasoning*, volume 2002, 2002.

[86] J. P. Ioannidis. Anticipating consequences of sharing raw data and code and of awarding badges for sharing. *Journal of Clinical Epidemiology*, 70:258–260, 2016.

[87] M. E. Isaac, E. Dawoe, and K. Sieciechowicz. Assessing local knowledge use in agroforestry management with cognitive maps. *Environmental Management*, 43(6):1321–1329, 2009.

[88] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. Wiley-Interscience: New York, 1991.

[89] A. J. Jetter and K. Kok. Fuzzy cognitive maps for futures studies—a methodological assessment of concepts and methods. *Futures*, 61(Sup. C):45 – 57, 2014.

[90] N. A. Jones, P. Perez, T. G. Measham, G. J. Kelly, P. d'Aquino, K. A. Daniell, A. Dray, and N. Ferrand. Evaluating participatory modeling: developing a framework for cross-case analysis. *Environmental Management*, 44(6):1180, 2009.

[91] H. Kahn, A. J. Wiener, et al. year 2000; a framework for speculation on the next thirty-three years. 1967.

[92] O. Karanfil, T. Moore, P. Finley, T. Brown, A. Zagonel, and R. Glass. A multi-scale paradigm to design policy options for obesity prevention: Exploring the integration of individual-based modeling and system dynamics. *Sandia National Laboratories. Available online: http://www. sandia. gov/casosengineering/docs/SD*, 202011, 2011.

[93] J. P. Kleijnen. *Statistical tools for simulation practitioners*. Marcel Dekker, Inc., 1986.

[94] J. P. Kleijnen. Factor screening in simulation experiments: review of sequential bifurcation. In *Advancing the frontiers of simulation*, pages 153–167. Springer, 2009.

[95] L. M. Koehly, A. Loscalzo, et al. Adolescent obesity and social networks. *Preventing Chronic Disease*, 6(3):A99, 2009.

[96] D. Koschützki et al. *Centrality Indices*, pages 16–61. Springer Berlin Heidelberg, 2005.

[97] B. Kosko. Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, 24(1):65–75, 1986.

[98] D. Koulouriotis, I. Diakoulakis, and D. Emiris. Learning fuzzy cognitive maps using evolution strategies: a novel schema for modeling and simulating high-level behavior. In *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, volume 1, pages 364–371. IEEE, 2001.

[99] R. Kruse, C. Borgelt, C. Braune, S. Mostaghim, and M. Steinbrecher. *Computational intelligence: a methodological introduction*. Springer, 2016.

[100] M. Kuhl, J. Ivy, E. Lada, N. Steiger, M. Wagner, and J. Wilson. Multivariate input models for stochastic simulation, 2010.

[101] E. A. Lavin and P. J. Giabbanelli. Analyzing and simplifying model uncertainty in fuzzy cognitive maps. *Proceedings of the 2017 Winter Simulation Conference*, 2017.

[102] A. M. Law. *Simulation Modeling and Analysis*. McGraw-Hill Education, 2015.

[103] T. M. Leahey, C. Y. Doyle, X. Xu, J. Bihuniak, and R. R. Wing. Social networks and social norms are associated with obesity treatment outcomes. *Obesity*, 23(8):1550–1554, 2015.

[104] R. A. K. (Letcher) et al. Selecting among five common modelling approaches for integrated environmental assessment and management. *Environtal Modelling and Software*, 47(Sup. C):159 – 181, 2013.

[105] D. Levy, P. Mabry, Y. Wang, S. Gortmaker, T. Huang, T. Marsh, M. Moodie, and B. Swinburn. Simulation models of obesity: a review of the literature and implications for research and policy. *Obesity Reviews*, 12:378–394, 2010.

[106] Y. Li, J. Berenson, A. Gutiérrez, and J. A. Pagán. Leveraging the food environment in obesity prevention: the promise of systems science and agent-based modeling. *Current Nutrition Reports*, pages 1–10, 2016.

[107] Y. Li, M. Lawley, D. Siscovick, D. Zhang, and J. Pagan. Agent-based modeling of chronic diseases: A narrative review and future research directions. *Preventing Chronic Disease*, 13:150561, 2016.

[108] A. Ligmann-Zielinska, S. C. Grady, and J. McWhorter. The impact of urban form on weight loss -combining spatial agent-based model with transtheoretical model of health behavior change. In Z. P. Neal, editor, *The Routledge Handbook of Applied System Science*. Routledge, 2017.

[109] S. Lukens, J. DePasse, R. Rosenfeld, E. Ghedin, E. Mochan, S. T. Brown, J. Grefenstette, D. S. Burke, D. Swigon, and G. Clermont. A large-scale immuno-epidemiological simulation of influenza a epidemics. *BMC Public Health*, 14(1):1019, 2014.

[110] C. Macal and M. North. Introductory tutorial: Agent-based modeling and simulation. In *Proceedings of the 2014 Winter Simulation Conference*, pages 6–20. IEEE Press, 2014.

[111] V. K. Mago et al. Analyzing the impact of social factors on homelessness: a fuzzy cognitive map approach. *BMC Medical Informatics and Decision Making*, 13(1):94, 2013.

[112] L. Malhi et al. Places to intervene to make complex food systems more healthy, green, fair, and affordable. *Journal of Hunger and Environmental Nutrition*, 4(3-4):466–476, 2009.

[113] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, 2000.

[114] M. D. McNeese and P. J. Ayoub. Concept mapping in the analysis and design of cognitive systems: a historical review. *Applied Concept Mapping: capturing, analyzing and organizing knowledge*, 2011.

[115] D. H. Meadows. *Thinking in Systems: A Primer*. Chealsea Green Publishing, 2008.

[116] A. Meliadou, F. Santoro, M. R. Nader, M. A. Dagher, S. Al Indary, and B. A. Salloum. Prioritising coastal zone management issues through fuzzy cognitive mapping approach. *Journal of Environmental Management*, 97:56–68, 2012.

[117] K. Mertens et al. Using structural equation-based metamodeling for agent-based models. In *2017 Winter Simulation Conference*, 2017.

[118] K. Minyard, R. Ferencik, M. A. Phillips, and C. Soderquist. Using systems thinking in state health policymaking: an educational initiative. *Health Systems*, 3(2):117–123, 2014.

[119] L. Moher, Davidand Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement. *Systematic Reviews*, 4(1), 2015.

[120] D. C. Montgomery. *Design and Analysis of Experiments*. John Wiley and Sons, 8 edition, 2013.

[121] T. E. Mosindy et al. Impact of angling on the production and yield of mature walleyes and northern pike in a small boreal lake in ontario. *North American Journal of Fisheries Management*, 7(4):493–501, 1987.

[122] G. Napoles et al. A computational tool for simulation and learning of fuzzy cognitive maps. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8, 2015.

[123] G. Napoles et al. On the convergence of sigmoid fuzzy cognitive maps. *Information Sciences*, 349–350:154–171, 2016.

[124] M. Newman. *Networks: an introduction*. Oxford University Press, 2010.

[125] R. Nianogo and A. Onyebuchi. Agent-based modeling of noncommunicable diseases: a systematic review. *American Journal of Public Health*, 105(3):e20–e31, 2015.

[126] A. Nyaki et al. Local–scale dynamics and local drivers of bushmeat trade. *Conservation Biology*, 28(5):1403–1414, 2014.

[127] M. G. Orr, S. Galea, M. Riddle, and G. A. Kaplan. Reducing racial disparities in obesity: simulating the effects of improved education and social network influence on diet behavior. *Annals of epidemiology*, 24(8):563–569, 2014.

[128] L. Ortolani, N. McRoberts, N. Dendoncker, and M. Rounsevell. Analysis of farmers' concepts of environmental management measures: an application of cognitive maps and cluster

analysis in pursuit of modelling agents' behaviour. *Fuzzy Cognitive Maps*, pages 363–381, 2010.

[129] U. Özesmi. Modeling ecosystems from local perspectives: fuzzy cognitive maps of the kizilirmak delta wetlands in turkey. In *1999 World Conference on Natural Resource Modelling*, pages 23–25, 1999.

[130] U. Özesmi. Bilissel (kognitif) haritalamaya gore halkin talepleri (the wants and desires of the local population based on cognitive mapping). *Yusufeli Baraji Yeniden Yerlesim Plani (Yusufeli Damlake Resettlement Plan), Devlet Su Isleri (DSI)(State Hydraulic Works). Sahara Muhendislik, Ankara*, pages 154–169, 2001.

[131] U. Özesmi and S. Özesmi. A participatory approach to ecosystem conservation: Uluabat lake environmental management plan using fuzzy cognitive maps and stakeholder analysis. In *Proceedings of the IV National Environmental Engineering Congress, Mersin, Turkey*, pages 7–10, 2001.

[132] U. Özesmi and S. Özesmi. A participatory approach to ecosystem conservation: fuzzy cognitive maps and stakeholder group analysis in uluabat lake, turkey. *Environmental Management*, 31(4):0518–0531, 2003.

[133] U. Özesmi and S. L. Özesmi. Ecological models based on people's knowledge: a multi-step fuzzy cognitive mapping approach. *Ecological Modelling*, 176(1):43–64, 2004.

[134] D. J. Pannell. Sensitivity analysis of normative economic models: theoretical framework and practical strategies. *Agricultural Economics*, 16(2):139–152, 1997.

[135] E. Papageorgiou and P. Groumpos. A new hybrid learning algorithm for fuzzy cognitive maps learning. *Applied Soft Computing*, 5(4):409–431, 2005.

[136] E. Papageorgiou and A. Kontogianni. Using fuzzy cognitive mapping in environmental decision making and management: a methodological primer and an application. In *International Perspectives on Global Environmental Change*. InTech, 2012.

[137] E. Papageorgiou, C. Stylios, and P. Groumpos. Fuzzy cognitive map learning based on nonlinear hebbian rule. In *Australasian Joint Conference on Artificial Intelligence*, pages 256–268. Springer, 2003.

[138] E. Papageorgiou, C. Stylios, and P. Groumpos. The soft computing technique of fuzzy cognitive maps for decision making in radiotherapy. *Intelligent and adaptive systems in medicine*, pages 106–127, 2008.

[139] E. I. Papageorgiou. A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. *Applied Soft Computing*, 11(1):500–513, 2011.

[140] E. I. Papageorgiou. Learning algorithms for fuzzy cognitive maps—a study. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(2):150–163, 2012.

[141] E. I. Papageorgiou, M. F. Hatwágner, A. Buruzs, and L. T. Kóczy. A concept reduction approach for fuzzy cognitive map models in decision making and management. *Neurocomputing*, 232:16–33, 2017.

[142] E. I. Papageorgiou and J. L. Salmeron. A review of fuzzy cognitive maps research during the last decade. *IEEE Transactions on Fuzzy Systems*, 21(1):66–79, 2013.

[143] E. I. Papageorgiou, P. Spyridonos, C. D. Stylios, P. Ravazoula, P. P. Groumpos, and G. Nikiforidis. Advanced soft computing diagnosis method for tumour grading. *Artificial Intelligence in Medicine*, 36(1):59–70, 2006.

[144] M. Papaioannou, C. Neocleous, A. Sofokleous, N. Mateou, A. Andreou, and C. N. Schizas. A generic tool for building fuzzy cognitive map systems. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 45–52. Springer, 2010.

[145] K. E. Parsopoulos, E. I. Papageorgiou, P. Groumpos, and M. N. Vrahatis. A first study of fuzzy cognitive maps learning using particle swarm optimization. In *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on*, volume 2, pages 1440–1447. IEEE, 2003.

[146] J. Pearce and K. Witten, editors. *Geographies of obesity: Environmental Understandings of the Obesity Epidemic*. Routledge, 2016.

[147] A. S. Penn, C. J. Knight, D. J. Lloyd, D. Avitabile, K. Kok, F. Schiller, A. Woodward, A. Druckman, and L. Basson. Participatory development and analysis of a fuzzy cognitive map of the establishment of a bio-based economy in the humber region. *PloS one*, 8(11):e78319, 2013.

[148] M. S. Pfaff, J. L. Drury, and G. L. Klein. Modeling knowledge using a crowd of experts. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 60, pages 183–187. SAGE Publications Sage CA: Los Angeles, CA, 2016.

[149] R. B. Pierce et al. Exploitation of northern pike in seven small north-central minnesota lakes. *North American Journal of Fisheries Management*, 15(3):601–609, 1995.

[150] R. Poli. A note on the difference between complicated and complex social systems. *Cadmus*, 2(1):142, 2013.

[151] S. F. Pratt, P. J. Giabbanelli, P. Jackson, and V. K. Mago. Rebel with many causes: A computational model of insurgency. In *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*, pages 90–95. IEEE, 2012.

[152] H. Rahmandad and N. S. Sabounchi. Modeling and estimating individual and population obesity dynamics. In S. J. Yang, A. M. Greenberg, and M. Endsley, editors, *Social Computing, Behavioral - Cultural Modeling and Prediction: 5th International Conference, SBP 2012, College Park, MD, USA, April 3-5, 2012. Proceedings*, pages 306–313. Springer, 2012.

[153] K. Resnicow and R. Vaughan. A chaotic view of behavior change: a quantum leap for health promotion. *International Journal of Behavioral Nutrition and Physical Activity*, 3(25), 2006.

[154] M. G. Richiardi, R. Leombruni, N. J. Saam, and M. Sonnessa. A common protocol for agent-based social simulation. *Journal of Artificial Societies and Social Simulation*, 9, 2006.

[155] S. Robinson. *Simulation: the practice of model development and use*. Palgrave Macmillan, 2014.

[156] S. Rosen, P. Salemi, B. Wickham, A. Williams, C. Harvey, E. Catlett, S. Taghiyeh, and J. Xu. Parallel empirical stochastic branch and bound for large-scale discrete optimization via simulation. In *Proceedings of the 2016 Winter Simulation Conference*, pages 626–637. IEEE Press, 2016.

[157] S. L. Rosen, C. M. Harmonosky, and M. T. Traband. Optimization of systems with multiple performance measures via simulation: Survey and recommendations. *Computers and Industrial Engineering*, 54(2):327–339, 2008.

[158] G. R. Sadler, H.-C. Lee, R. S.-H. Lim, and J. Fullerton. Recruiting hard-to-reach united states population sub-groups via adaptations of snowball sampling strategy. *Nursing and Health Sciences*, 12(3):369–374, 2010.

[159] J. L. Salmeron. Modelling grey uncertainty with fuzzy grey cognitive maps. *Expert Systems with Applications*, 37(12):7581 – 7588, 2010.

[160] J. L. Salmeron. A fuzzy grey cognitive maps-based intelligent security system. In *2015 IEEE International Conference on Grey Systems and Intelligent Services (GSIS)*, pages 29–32, 2015.

[161] J. L. Salmeron and E. I. Papageorgiou. A fuzzy grey cognitive maps-based decision support system for radiotherapy treatment planning. *Knowledge-Based Systems*, 30:151 – 160, 2012.

[162] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley and Sons, 2008.

[163] A. Sarker and G. Gonzalez. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *Journal of Biomedical Informatics*, 53:196–207, 2015.

[164] P. J. Schoemaker. When and how to use scenario planning: a heuristic approach with illustration. *Journal of Forecasting*, 10(6):549–564, 1991.

[165] R. Shahid and S. Bertazzon. Local spatial analysis and dynamic simulation of childhood obesity and neighbourhood walkability in a major canadian city. *Public Health*, 2(4):616–637, 2015.

[166] D. Shoham, R. Hammond, H. Rahmandad, Y. Wang, and P. Hovmand. Modeling social norms and social influence in obesity. *Current Epidemiological reports*, 2:71–79, 2015.

[167] D. A. Shoham. Advancing mechanistic understanding of social influence of obesity through personal networks. *Obesity*, 23(12):2324–2325, 2015.

[168] R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications: with R examples*. Springer Science and Business Media, 2006.

[169] B. Silverman, N. Hanrahan, G. Bharathy, K. Gordon, and D. Johnson. A systems approach to healthcare: agent-based modeling community mental health, and population well-being. *Artificial Intelligence in Medicine*, 63:61–71, 2015.

[170] W. Stach et al. Parallel learning of large fuzzy cognitive maps. In *2007 International Joint Conference on Neural Networks*, pages 1584–1589, 2007.

[171] W. Stach and L. Kurgan. Parallel fuzzy cognitive maps as a tool for modeling software development projects. In *Fuzzy Information, 2004. Processing NAFIPS '04. IEEE Annual Meeting of the*, volume 1, pages 28–33, 2004.

[172] K. Takayanagi and S. Kurahashi. Analysis of the network effects on obesity epidemic. In *Agent and Multi-Agent Systems: Technologies and Applications*, pages 393–403. Springer, 2015.

[173] J. C. Thiele, W. Kurth, and V. Grimm. Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using netlogo and r. *Journal of Artificial Societies and Social Simulation*, 17(3):11, 2014.

[174] D. M. Thomas, M. Weedermann, B. F. Fuemmeler, C. K. Martin, N. V. Dhurandhar, C. Bredlau, S. B. Heymsfield, E. Ravussin, and C. Bouchard. Dynamic model predicting overweight, obesity, and extreme obesity prevalence trends. *Obesity*, 22(2):590–597, 2014.

[175] L. Tong, D. Shoham, and R. Cooper. A co-evolution model for dynamic social network and behavior. *Open Journal of Statistics*, 4(9):765, 2014.

[176] E. Trillas and L. Eciolaza. *Fuzzy logic: an introductory course for engineering students*, volume 320. Springer, 2015.

[177] A. K. Tsadiras. Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps. *Information Sciences*, 178(20):3880–3894, 2008.

[178] UCAR/COMET. Understanding assimilation systems: How models create their initial conditions - version 2, 2009.

[179] S. van der Walt, S. C. Colbert, and G. Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering*, 13(2):22–30, 2011.

[180] W. Van Leekwijck and E. E. Kerre. Defuzzification: criteria and classification. *Fuzzy sets and systems*, 108(2):159–178, 1999.

[181] M. van Vliet, K. Kok, and T. Veldkamp. Linking stakeholders and modellers in scenario studies: The use of fuzzy cognitive maps as a communication and learning tool. *Futures*, 42(1):1–14, 2010.

[182] I. Vanderbroeck, J. Goossens, and M. Clemens. Foresight tackling obesities: a future choices – building the obesity system map. *UK Governments Foresight Programme*, 2007.

[183] T. Verigin, P. J. Giabbanelli, and P. I. Davidsen. Supporting a systems approach to healthy weight interventions in british columbia by modeling weight and well-being. In *Proceedings of the 49th Annual Simulation Symposium*, ANSS '16, pages 9:1–9:10. Society for Computer Simulation International, 2016.

[184] C. Wang. *A study of membership functions on mamdani-type fuzzy inference system for industrial decision-making*. Lehigh University, 2015.

[185] Y. Wang, H. Xue, H.-j. Chen, and T. Igusa. Examining social norm impacts on obesity and eating behaviors among us school children based on agent-based model. *BMC Public Health*, 14(1):1, 2014.

[186] M. C. Weinstein, B. O'Brien, J. Hornberger, J. Jackson, M. Johannesson, C. McCabe, and B. R. Luce. Principles of good practice for decision analytic modeling in health-care eval-

uation: report of the ispor task force on good research practices modeling studies. *Value in Health*, 6(1):9–17, 2003.

[187] D. Weitao, B. Ankenman, P. Sanchez, W. Duana, B. E. Ankenmana, P. J. Sanchezb, and S. M. Sanchezb. Sliced full factorial-based latin hypercube designs as a framework for a batch sequential design algorithm. 2014.

[188] R. West. Data and statistical commands should be routinely disclosed in order to promote greater transparency and accountability in clinical and behavioral research. *Journal of Clinical Epidemiology*, 70:254, 2016.

[189] R. Whitaker, J. Wright, M. Pepe, K. Seidel, and W. Dietz. Predicting obesity in young adulthood from childhood and parental obesity. *New England Journal of Medicine*, 337(13):869–873, 1997.

[190] K. P. White Jr, M. J. Cobb, and S. C. Spratt. A comparison of five steady-state truncation heuristics for simulation. In *Proceedings of the 32nd conference on Winter simulation*, pages 755–760. Society for Computer Simulation International, 2000.

[191] K. P. White Jr and S. Robinson. The problem of the initial transient (again), or why mser works. *Journal of Simulation*, 4(4):268–272, 2010.

[192] World Obesity Federation. New figures indicate 2.7 billion adults worldwide will be overweight by 2025. `http://www.worldobesity.org/site_media/uploads/World_Obesity_Day_Press_Release.pdf`, 2015.

[193] J. Wu, R. Dhingra, M. Gambhir, and J. V. Remais. Sensitivity analysis of infectious disease models: methods, advances and their application. *Journal of The Royal Society Interface*, 10(86):20121018, 2013.

[194] Z. Xu and Q.-L. Da. An overview of operators for aggregating information. *International Journal of Intelligent Systems*, 18(9):953–969, 2003.

[195] N. Yalçın and G. Seçme. Fuzzy cognitive mapping technique to examine the problems and development opportunities for kayseri industry. *Graduation thesis, Erciyes University Industrial Engineering Department*, 2001.

[196] Y. Yang, A. H. Auchincloss, D. A. Rodriguez, D. G. Brown, R. Riolo, and A. V. Diez-Roux. Modeling spatial segregation and travel cost influences on utilitarian walking: Towards policy intervention. *Computers, Environment and Urban Systems*, 51:59–69, 2015.

[197] Y. Yang, A. V. D. Roux, A. H. Auchincloss, D. A. Rodrigue, and D. G. Brown. A spatial agent-based model for the simulation of adults' daily walking within a city. *American Journal of Preventive Medicine*, 40(3):353–361, 2011.

[198] N. Zainal Abidin, M. Mamat, B. Dangerfield, J. H. Zulkepli, M. A. Baten, and A. Wibowo. Combating obesity through healthy eating behavior: A call for system dynamics optimization. *PLoS ONE*, 9:1–17, 12 2014.

[199] D. Zhang, P. Giabbanelli, O. Arah, and F. Zimmerman. Impact of different policies on unhealthy dietary behaviors in an urban adult population: an agent-based simulation model. *American Journal of Public Health*, 104(7):1217–1222, 2014.

[200] J. Zhang, D. A. Shoham, E. Tesdahl, and S. B. Gesell. Network interventions on physical activity in an afterschool program: An agent-based social network study. *American Journal of Public Health*, 105(S2):S236–S243, 2015.

[201] J. Zhang, L. Tong, P. Lamberson, R. Durazo-Arvizu, A. Luke, and D. Shoham. Leveraging social influence to address overweight and obesity using agent-based models: the role of adolescent social networks. *Social Science and Medicine*, 125:203–213, 2015.