Do all of these meetings matter? Three essays concerning the impacts of collaborative watershed governance

Tyler Scott

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Craig Thomas, Chair

Ann Bostrom

Stanley Asah

Gerald Herting [GSR]

Program Authorized to Offer Degree:
Daniel J. Evans School of Public Affairs

University of Washington

Abstract

Do all of these meetings matter? Three essays concerning the impacts of collaborative watershed governance

Tyler Scott

Chair of the Supervisory Committee:
Professor Craig Thomas
Daniel J. Evans School of Public Affairs

My dissertation presents three related projects. In each project, I make a novel contribution to the understanding of collaborative environmental governance in theory and practice. I focus on a common application of collaborative environmental governance - collaborative watershed planning and management groups. Broadly, I address the question of whether -and how-collaborative governance improves environmental outcomes. Accordingly, each project is intended to address questions that are critical in this regard. Within each chapter I also make a significant methodological contribution to policy and management research. For each of the three chapters I use novel data and demonstrate the use of an analytic method that as-of-yet has not gained traction in policy and management research.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I wish to express my immense gratitude to my committee chair, Craig Thomas, for his support and guidance in all facets of my doctoral training. Ryan Scott and Grant Blume were also instrumental to my work; I owe each of them a debt of gratitude both for their scholarly input and their friendship. Lastly, I want to express my sincere thanks to Richard Zerbe for his many years of support and encouragement, and to the Evans School, particularly current Ph.D. Program Chair Ann Bostrom, former Chair Bill Zumeta, and Dean Sandra Archibald, for their support as well.

DEDICATION

to my incredible wife, Kate, and my family for all of their love and support

Chapter 1

INTRODUCTION

OVERVIEW

Transboundary problems such as nonpoint source water pollution continue to be a vexing environmental policy challenge. By involving relevant stakeholders in planning, implementation, and management, non-regulatory policy instruments such as collaborative environmental governance seek to increase the comprehensiveness and scale of policy efforts so as to better address such problems (Holley et al. 2013; Healy et al. 2014). "Collaborative management" and "collaborative governance" are normatively popular and have been widely employed in environmental policy applications worldwide (Bingham and O'Leary 2008; Donahue and Zeckhauser 2011). However, it is unclear whether government funding for collaborative environmental governance efforts pays off in terms of improved environmental outcomes (Thomas and Koontz 2011; Koontz and Thomas 2006). As a result, there is also a deficit of rigorously tested theory that would enable policy makers to wield collaborative governance as a strategic, context-appropriate policy tool to achieve environmental protection and resource management goals.

My dissertation presents three related projects. In each project, I make a novel contribution to the understanding of collaborative environmental governance in theory and practice. I focus on a common application of collaborative environmental governance–collaborative watershed planning and management groups (Gerlak et al. 2012; Grafton and Hussey 2011; Hoornbeek et al. 2012; Imperial 2005; Lubell 2004-a; Lubell 2004b; Lubell 2004c; Mazmanian and Kraft 2009; Sabatier 2005; Thomas and Koontz 2011). Broadly, I address the question of whether –and how– collaborative governance improves environmental outcomes. Accordingly, each project is intended to address questions that are critical in this regard. Within each chapter I also make a significant methodological contribution to policy and management research. For each of the three chapters I use novel data and demonstrate the use of an analytic method that as-of-yet has not gained traction in policy and management research.

The first chapter addresses two research questions: (1) Does collaborative environmental governance improve environmental outcomes? and (2) How do publicly supported collaborative groups with different levels of responsibility, formalization, and representativeness compare in this regard? Using a representative watershed quality data series, the EPA's National Rivers and Streams Assessment (NRSA) and Wadeable Streams Assessment (WSA), in conjunction with a watershed management regime database coded for this analysis, it tests the relationship between collaborative governance and watershed quality for 357 watersheds nationwide. Since these are observational data, a multilevel propensity score matching method is used to control for selection bias. Using an augmented inverse propensity weighted estimator (AIPW) (Glynn and Quinn 2010), I estimate the average treatment effect (ATE) of collaborative governance for six different water quality and habitat condition metrics. This research presents some of the first evidence about the impacts of collaborative governance that is based upon an objective outcome metric (Carr et al. 2012; Koontz and Thomas 2006). Collaborative watershed groups are found to improve water chemistry and in-stream habitat conditions. I then use hierarchical linear regression modeling (HLM) to examine how group responsibilities, membership diversity, and formalization affect the predicted impact of a collaborative group. Groups that engage in management activities (in comparison to coordination or planning) are found to achieve greater environmental gains. Limited differentiation is found with regards to the presence of a group coordinator, increased goal specificity, or greater stakeholder diversity.

The second chapter examines collaborative management groups from the perspective of policymakers seeking to increase coordination within a policy network. While governments often support collaborative groups as a tool to address perceived network failures such as a lack of coordination, the net impact such groups have is unclear. I use valued exponential random graph models (ERGMs) (Krivitsky 2012; Desmarais and Cranmer 2012a) to model relationships of varying strength amongst a regional network of organizations involved in 57 collaborative groups in the Puget Sound region of Washington state. This provides a unique opportunity to study the interplay between numerous groups and organizations within a large-scale network. Valued ERGMs are a recently developed extension of standard

ERGMs that model valued instead of binary ties (Krivitsky 2012), which allows for a more detailed and nuanced understanding of network structure; this, in this chapter I also make a methodological contribution to the policy literature. Findings suggest that participation in collaborative groups does motivate stronger levels of coordination and cooperation amongst organizations network ties between individual organizations; however, this effect is strongest amongst pair of organizations that: (a) do not have a pre-existing tie; and (b) are not already members of another group. These results support a transaction-cost based perspective of how government-sponsored collaborative groups can influence network coordination; further, they also provide an empirical example of the Ecology of Games (Lubell 2013), in which multiple collaborative institutions have interactive effects on one another within a policy network.

My third and final chapter examines the extent to which different collaborative watershed council actions improve water quality. I couple longitudinal data concerning 2500 state grants given to local watershed councils in the state of Oregon with 20 years of ambient water quality monitoring data sampled at 141 sites around the state. I use state funding as a proxy for watershed council actions, testing whether council actions improve water quality and further comparing the impacts of specific actions such as monitoring, education, administrative support, and restoration. In modeling these effects, I also make a methodological contribution by demonstrating how spatio-temporal ecological and epidemiological modeling techniques can be used to test policy theory and analyze policy impacts using extant data. Specifically, I use integrated nested Laplace approximation (INLA) (Rue et al. 2009) and stochastic partial differential equations (SPDE) (Lindgren et al. 2011) to fit a hierarchical Bayesian model that accounts for spatial and temporal dependency. I find that watershed council actions such as education, outreach, and administrative support functions engender strong improvements in water quality. The impacts of restoration actions are positive on average but of lesser magnitude and greater uncertainty.

BACKGROUND

There is a rich body of public administration literature that considers governance arrangements that cannot be adequately characterized as market-based or state-controlled. These "polycentric governance" arrangements (Ostrom et al. 1961), in which there are many centers of decision making that: (1) are formally independent of one another; but that (2) function within an interdependent system of relations, have long been a significant focus of the theoretical literature and policy makers (Ostrom 2010). Modern literature refers to this third control structure typology as network governance, but by any name a vast literature concerning both policy networks (Adam and Kriesi 2007; Isett et al. 2011) and collaborative governance (Ansell and Gash 2008; Emerson et al. 2012; Margerum 2011) continues to examine how complex networks of actors and institutions address collective action problems. Much of the policy networks literature has focused on how the institutional landscape and structural characteristics affect policy outcomes (Kriesi et al. 2006; Kickert et al. 1997). The strand focusing on collaborative governance largely concerns how collaborative groups build consensus (Schneider et al. 2003; Weible et al. 2009) or influence inter-actor exchange (Bodin and Crona 2009; Crona and Hubacek 2010).

Key Terms

Due to the frequent, but diverse, usage of the terms "governance," "network," and "collaborative" in the literature, it is important to unpack this terminology and define these concepts in order to position my research. "Governance" refers to a process in which public actors make policies, deliver services, or implement policies within a network (or networks) of actors (Frederickson 2005; Torfing and Sorensen 2007; Rhodes 1997). Governance is characterized by a high degree of interdependency amongst actors and a complex decision-making process (Klijn et al. 2010). Bressers (2009, p. 125) poses that "governance" is an enlargement of the concept of public policy (Bressers and Kuks 2003). Thus, in keeping with the findings of Ostrom et al. (1961), governance is not really a new state of affairs, but rather a basis for scientific variables that can be used for empirical studies (Bressers 2009). Defining a "network" can be difficult, however, as the literature lacks a coherence about what a network is and what a network does (Borzel 1998; Wachhaus 2009). In response to this, numerous analytical frameworks (Adam and Kriesi 2007; Agranoff and McGuire 1998; Klijn 1997; Klijn et al. 1995; Rhodes 1997) and literature reviews (Berry et al. 2004; Isett et al. 2011; Provan et al. 2007; Robinson 2007) have sought to clarify and categorize this literature, but there remains no definitive understanding. Broadly "networks" are simply defined as "collections of actors who pursue repeated, enduring exchange relations with one another and, at the same time, lack a legitimate organizational authority to arbitrate and resolve disputes that may arise during the exchange" (Podolny and Page 1998, p. 59). By expanding the concept of exchange, networks can be broadly defined as "sets of individuals bound by communication, relationships, positions, or interest area" (Margerum 2011, p. 33). More strictly, a network is simply "three or more legally autonomous organizations [or individuals] that work together to achieve not only their own goals but also a collective goal" (Provan and Kenis 2008, p. 231). This definition can encompass social networks (interpersonal relationships, e.g., Putnam 2000), inter-organizational networks (structures and processes in which organizations interact, e.g., Alexander 1993), and political networks (power positions and configurations, e.g., Knoke 1990).

Inter-organizational networks

The focus of the research proposed in this project is on interorganizational networks. In keeping with the preceding definitions, "governance networks" thus involve many organizations that are connected because of their dependence on resources or commitments to realize aims/solve problems (Agranoff and McGuire 2001; Koppenjan and Klijn 2004) and are relatively stable and characterized by intensive/regular interactions (Agranoff and McGuire 2003; Kickert et al. 1997; Meier and O'Toole 2001, 2007). The literature concerning the benefits of forming and maintaining connections to other organizations characterizes them broadly, including information access, issue understanding, conflict reduction, and implementation support (Moreland et al. 1993; Gigone and Hastie 1993; Cragan and Wright 1990; Hill and Lynn 2003; Susskind et al. 1999; Sabatier et al. 2005). For instance, Margerum and Holland (2000) find specific instances where a regional coordination group was able to jointly fund several projects that none of the organizations could have funded individually. Not all benefits are experienced collectively; network ties can also benefit the interests and goals of an individual organization (Lubell 2005).

The costs associated with such actions are typically concrete time and staff resources that are required in order to engage in network relations. Interorganizational deliberation and coordination are constrained by the transaction costs associated with such actions. For instance, Thomas (2003) finds that travel time is a key predictor of involvement in coordination activities, and Margerum (2011) describes interviews with high-level managers who are forced to selectively participate in specific meetings of various coordinative groups due to the overall volume of groups and group activities. More broadly, researchers often discuss the transaction costs associated with network relations in terms of norms of reciprocity (Putnam 2000) or shared beliefs and preferences (Schneider et al. 2003; Sabatier and Jenkins-Smith 1993). Thus the structure and function of an organizational network is a result of the costs and benefits associated with inter-organizational relationships.

Collaborative Management as a Policy Tool

In this network governance context, public policy makers often fulfill a role as network manager (Klijn and Koppenjan 2000); in other words, instead of carrying out tasks directly, policy makers attempt to address collective action dilemmas indirectly by changing network rules and influencing network relationships (Klijn and Koppenjan 2006). Inter-organizational (or inter-stakeholder) collaborative groups are one of the most prominent (and well-documented) mechanisms by which environmental policy makers attempt to alter the structure and function of an organizational network (see Ansell and Gash 2008; Emerson et al. 2012; Margerum 2011, for recent discussions). Margerum (2011, p. 6) defines collaboration as "an approach to solving complex problems in which a diverse group of autonomous stakeholders deliberates to build consensus and develop networks for translating consensus into results." Such groups represent "a governing arrangement where one or more public agencies directly engage non-state stakeholders in a collective decision-making process that is formal, consensus-oriented, and deliberative and that aims to make or implement public policy or manage public programs or assets" (Ansell and Gash 2008, p. 544) (see also Emerson et al. 2012; Imperial 2005; Margerum 2011). Thus, a "collaborative group" refers to a planning and management entity that uses a deliberative structure and seeks to encompass relevant stakeholders (Ansell and Gash 2008).

The indirect nature of collaborative approaches might at first seem to be completely at odds with the notion of direct government influence and strategic public policy interventions; however, it is important to bear in mind that the modern paradigm of "governance" implies that government cannot act autonomously, and must instead engage in multi-actor processes in which it is only one of many relevant actors, as its "core purposes can only hope to be realized in such settings" (Bressers 2009, p. 130). In other words, even in a network context government must act as a purposeful, strategic network manager (Klijn and Koppenjan 2000; Klijn and Koppenjan 2006). Thus, the coordination of governance efforts is vested in traditional public authorities that were already previously responsible. In fact, Ansell and Gash (2008).

Given this, I contend that a collaborative group represents a tool that policy makers can use to alter the current way in which organizations in a given network interact. Salamon (2002) defines policy tools as instruments of public action that can be defined as "an identifiable method through which collective action is structured to address a public problem" (Salamon 2002, p. 19). Many scholars, including Ansell and Gash (2008), Imperial (2005), Connick and Innes (2003), and Donahue and Zeckhauser (2011) have considered collaborative groups through such a lens, particularly emphasizing groups that are concerned with public issues. Moreover, Koontz et al. (2004) address this issue even more directly, as they focus on the role government plays initiating and supporting collaborative governance.

Empirically, it appears that public-sector decision-making matches with this theoretical perspective. In the last few decades the US Environmental Protection Agency (EPA) and other Federal agencies such as the US Forest Service have increasingly emphasized collaborative approaches in recognition of this issue (Carr et al. 1998). States also increasingly emphasize collaborative approaches, even in some cases (such as Washington's Puget Sound Partnership and Oregon's Watershed Enhancement Board) creating state agencies officially charged with facilitating and supporting local collaborative groups (Lurie and Hibbard 2008; Chaffin et al. 2012).

Impacts of Collaborative Management

Generally, this push towards expanded stakeholder involvement and consensus-oriented decision-making is regarded as normatively desirable because it is theorized to broaden public engagement (Fung and Wright 2003; Weber 2003; Wondolleck and Yaffee 2000). The prevailing wisdom is that the formal and informal interactions engendered by collaborative groups and their outputs can increase policy effectiveness without requiring formal structural changes or reorganization (Bardach 1998; Provan and Milward 1995). Popular theory holds that increased participation of non-state actors in policy making leads to more ecologically rational decisions than in top-down modes of governance (Dryzek 1997; Smith 2004a). Similarly, Sabatier et al. (2005) and the National Research Council (Dietz and Stern 2008) find that participation of non-state actors leads to improved compliance with decisions and thus better ecological outcomes than does top-down modes of governance.

As mentioned at the outset, there is a vast and growing body of literature concerning the formation, structure, and function of collaborative groups. Much of this research has concerned what Carr et al. (2012) describe as process evaluations and intermediary outcome evaluations (Ulibarri 2015; Lubell 2004b; Leach et al. 2013; Scholz et al. 2008). Process evaluations concern the procedural fairness, legitimacy, and other aspects that related to the participation process itself. Well-functioning collaborative groups serve to increase the amount of information included in the decision-making process and increase representation (Carr et al. 2012). Intermediary outcome evaluations examine either ancillary or intermediate outcomes that are not the ultimate goal of the policy itself. For instance, participatory watershed management has been shown to foster stronger ties between participants (Collins et al. 2007), increase trust amongst stakeholders (Leach and Sabatier 2005), engender more creative policies (Newig and Fritsch 2009b), and enhance learning amongst stakeholders (Leach et al. 2013).

“Third-level” (Innes and Booher 1999), "resource management outcomes" (Carr et al. 2012) are much more difficult to evaluate, for a variety of reasons including but not limited to data challenges, inability to posit causality, and the long-term nature of many environmental outcomes (Koontz and Thomas 2006; Thomas and Koontz 2011). Thus, while we know that collaborative groups–and collaborative governance approach more broadly–have many desirable social and political outcomes, we know very little about whether they succeed at improving environmental conditions (Bingham et al. 2003; Gray 2000; Innes and Booher 1999; Koontz and Thomas 2006; O'Leary and Bingham 2003; Thomson et al. 2008). Environmental outcomes, such as changes in land cover or biological diversity, are difficult to measure. Not only can such assessments require extensive, costly data collection and analysis using remote sensing and ecological studies, but generally there is a long time horizon between governance outputs and environmental outcomes (Koontz and Thomas 2006). Some researchers have attempted to address these difficulties by measuring participant perceptions of environmental change (Leach et al. 2002; Leach and Sabatier 2005; Lubell 2004b; Ulibarri 2015), but such perceptions are obviously subjective, and actor perceptions about environmental change are typically positively correlated with personal effort, feelings about the collaborative group, or feelings about collaborative partners (Coglianese 2003; Leach and Sabatier 2005; Lubell and Lippert 2011); in other words, it is possible that these sentiments are biased by a "halo effect" in which stakeholders perceive environmental improvement due to their positive participatory experience.

Particularly as related to policy decision-making, the current body of research regarding collaborative groups has not sufficiently looked outside of collaborative groups themselves. In particular, there has been little examination as to: (1) how the structure and implementation of collaborative management policies relate to environmental quality; (2) the degree to which collaborative group outputs (such as meetings or group exercises) ultimately affect restoration and recovery network relationships (i.e., whether collaboration achieves benefits in its hypothesized fashion); and (3) the extent to which formal support for collaborative activities (e.g., funding and administrative support provided by state actors) engenders environmental benefits. Because of this, there is little theoretical guidance as to when such interventions might be warranted (Schrank and Whitford 2011)or what effects policy makers might anticipate from such efforts (Carlsson and Berkes 2005; Crona and Hubacek 2010).

In the three dissertations chapters, I address each of these questions. The next chapter examines the relationship between collaborative management and environmental outcomes, specifically the differential effect of specific collaborative management structures and strategies. My chapters addressing the second and third questions then follow.

Chapter 2

DOES COLLABORATION MAKE ANY DIFFERENCE? LINKING COLLABORATIVE GOVERNANCE TO ENVIRONMENTAL OUTCOMES.

INTRODUCTION

“Collaborative governance" and “collaborative management" are normatively popular concepts that have been widely employed in environmental policy applications worldwide (Ansell and Gash 2008; Hall and O'Toole 2000, 2004; Innes and Booher 2004; McGuire 2006; Newig and Fritsch 2009a). Collaboration has been shown to enhance cooperation and foster belief change amongst stakeholders (Lubell 2004c; Leach et al. 2013), generate funds and support for alternative policy measures when problems are too diffuse or difficult to address via regulation (Margerum 2011; Margerum and Holland 2000), and increase the implementation success of policies and programs (Agranoff and McGuire 2003; Meier 2005). However, we still know very little about the relationship between collaboration and environmental outcomes (Carr et al. 2012; Koontz and Thomas 2006) or how the environmental outcomes of collaborative approaches compare to those of other policy alternatives (Margerum 2011; Schneider et al. 2003).

This issue, whether collaborative environmental governance improves environmental outcomes, is the primary research question addressed in this analysis. My second research question builds upon the first, asking: What design and implementation characteristics make groups more or less effective at improving environmental outcomes? Research shows that collaboration alone does not necessarily yield improved outcomes (Newig and Fritsch 2009a), but there is little existing evidence informing how policy makers might best wield collaborative governance as a strategic, context-appropriate policy tool. I address these questions using a common application of collaborative environmental governance–collaborative watershed planning and management groups (Gerlak et al. 2012; Grafton and Hussey 2011; Hoornbeek et al. 2012; Imperial 2005; Lubell 2004b, 2004c; Mazmanian and Kraft 2009; Sabatier 2005; Thomas and Koontz 2011). The next section describes how this analysis fits within–and builds upon–the existing literature. I then specify the empirical approach used in this analysis and explain it is appropriate given the data and research questions. Subsequent sections then detail my data collection process and coding scheme, present model results, and discuss findings.

THEORETICAL RATIONALE

As a wide and growing body of synthesis literature attests, public policy scholars are interested in studying the role and impact of collaborative governance in a variety of policy sectors (Bingham and O'Leary 2008; Carr et al. 2012; Donahue and Zeckhauser 2011; Huxham 2003; Innes and Booher 2010; McGuire 2006; O'Leary and Bingham 2009; Sabatier 2005). However, much of this work concerns the quality of the collaborative process (Ansell and Gash 2008; Coglianese 1997; Frame et al. 2004; Leach 2006; Leach et al. 2002; Langbein 2000; Lubell 2005; Sabatier and Shaw 2009) or addresses changes in intermediate outcomes such as: (1) stakeholder cooperation and consensus (Collins et al. 2007; Fuller 2009; Lubell 2004b, 2004c; Lubell 2005; McGuire and Silvia 2010; Schively 2007; Susskind 1996; Weible et al. 2004); (2) the production of plans and other outputs (Beierle 2002; Innes 1996; Innes and Booher 1999; Margerum 2011; Leach and Sabatier 2005; Lubell 2005; Newig and Fritsch 2009b; Wondolleck and Yaffee 2000); and (3) stakeholder perceptions of outcomes (Leach 2006; Provan and Milward 1995; Leach et al. 2002) that result from collaborative approaches (Carr et al. 2012; Koontz and Thomas 2006). For instance, Lubell (2005) shows how collaborative groups with strong procedures and well-codified practices can enhance stakeholder trust and collective action beliefs, thereby increasing support for collaborative policy efforts. Procedural and intermediate outcomes can be significant in their own right, but it is important to recognize that policymakers use collaborative governance as a tool for improving policy outcomes (Hoornbeek et al. 2012; Koontz et al. 2004). In other words, policymakers purposefully choose to engage in collaborative planning and management (Huxham 2003; Vangen and Huxham 2003) as a means by which to "make or implement public policy or manage public programs or assets" (Ansell and Gash 2008, 544). Relatively few works have focused on the role of government in initiating and supporting collaborative groups in this fashion (Huxham 2003; Koontz et al. 2004; Mandell 1999, 2001; Vangen and Huxham 2003; Schneider et al. 2003).

Initiating and maintaining collaborative governance takes time and effort; accordingly, for policymakers there are “trade-offs associated with participating in... collaborative efforts that divert scarce resources from other activities" (Layzer 2008, 290). These trade-offs naturally raise a question of efficacy: Does collaborative governance improve environmental outcomes? While there are many in-depth case studies that point to specific outcomes (Ansell and Gash 2008; Margerum 2011; Newig and Fritsch 2009b), there is little systematic evidence in this regard (Koontz and Thomas 2006; Carr et al. 2012). Collaborative governance is believed to help facilitate decision-making, better address interrelated problems, carry greater legitimacy and improve implementation (Sabatier et al. 2005). Other benefits attributed to collaboration include access to information, implementation support, and reduced conflict (Gigone and Hastie 1993; Hill and Lynn 2003; Moreland et al. 1993; Sabatier et al. 2005; Susskind et al. 1999). At the same time, collaborative processes can be time consuming and difficult (Margerum 2011), and there are legitimate concerns about whether collaborative institutions are, as asked by Lubell (2004a), “all talk and no action.” My primary hypothesis (H1) is that collaborative watershed governance results in improved environmental outcomes:

H1: Collaborative watershed governance results in improved environmental outcomes.

Policymakers not only face the general choice of whether to support collaborative governance, but also regarding the specific form that their collaborative efforts will take. The current literature contains several typologies and theoretical frameworks that characterize collaborative groups in terms of: (1) conceptual themes such as geographic scale, institutional scale, inclusiveness, or stakeholder incentives (Ansell and Gash 2008; Cheng and Daniels 2005; Emerson et al. 2012; Margerum 2011); or (2) comparisons between agency-led and independent collaborative institutions (Bidwell and Ryan 2006; Moore and Koontz 2003). For instance, (Margerum 2008; Margerum 2011) distinguishes between the institutional scales on which collaboration occurs, while Moore and Koontz (2003) characterize groups in terms of in terms of seating (e.g., agency-based or stakeholder based). However, none of these typologies pertain specifically to the choices policymakers face when designing and implementing a collaborative group within a given institutional context. Similarly, Ansell and Gash (2008) and Emerson et al. (2012) each pose prominent theoretical frameworks that identify key variables, such as participatory inclusiveness and stakeholder incentives, which mediate outcomes. While these frameworks speak broadly to institutional design, they do not distinguish between specific group characteristics. Thus, along with testing the direct "treatment effect" of a collaborative watershed group, I operationalize this literature by comparing collaborative groups in terms of the concrete design and implementation choices public managers must make, such as designating group tasks or inviting group members.

Specifically, I test three collaborative group attributes believed to be key drivers of group impact: (H2) the level of management responsibility accorded to the collective ("Group Responsibility"); (H3) diversity of representation in the group ('Stakeholder Representation"); (H4) group formalization ("Group Formalization"). In the remainder of this section, I provide a brief overview of each sub-hypothesis and orient each within the literature.

Group Responsibility

Group Responsibility (H2) contrasts groups that serve as coordinating bodies or engage in outreach, monitoring, or planning from groups that engage in management activities, such as serving as the lead entity for salmon recovery actions or managing land use in the watershed. Groups conducting management activities presumably engage in more intensive and ongoing collaboration. Incentives to manipulate and act co-optively are checked in situations in which actors expect to engage in ongoing cooperation (2008, 560). Repeated interactions influence the willingness organizations to collaborate (Innes 1998; Moreland et al. 1993), and more intensive collaborative processes are shown to increase information exchange and produce higher-quality decisions (Beierle 2002). However, increasing the intensity of interactions (e.g., from information sharing to planning to joint implementation) requires greater stakeholder engagement and investment (Margerum 2011; Sabatier et al. 2005; Wondolleck and Yaffee 2000). Along with requiring greater time and effort (Hill and Lynn 2003; Sabatier et al. 2005), higher intensity collaborative efforts necessitate increased power sharing amongst participants (Margerum 2011). Lubell et al. (2002) find that as these types of transaction costs increase, it is more likely that actual collaboration will be supplanted by nominal, in-name-only collaboration. Thus, more group responsibility might not result in a larger impact if groups are unable to adequately fulfill such a role.

H2: Increased responsibility for a collaborative group is associated with beneficial environmental outcomes.

Stakeholder Representation

Collaborative endeavors are theorized to be more effective when they incorporate a broader range of information and perspectives (Burby 2003; Innes and Booher 1999; Margerum 2011; Wondolleck and Yaffee 2000) because this increased breadth facilitates better decision-making (Dryzek 1997; Smith 2004b; Gregory et al. 2001), improved compliance (Sabatier et al. 2005), and more effective policy implementation (Burby 2003; Carlson 1999). While Anderson et al. (2013) demonstrate that being more responsive to stakeholders does preclude technically sound management, the literature expresses concern that attempting to incorporate the interests and knowledge of all relevant stakeholders potentially results in diluted–and thus ineffectual–plans and policies (Coglianese 1997; Coglianese 1999; Koontz et al. 2004). Further, an increased number of organizations can make it more difficult to develop key linkages (Alexander 1995; Gray 1989), and incorporating additional jurisdictional levels (horizontally and hierarchically) can make group actions less tractable (Margerum 2011). To examine this, Stakeholder Representation (H3) considers the extent to which a group is comprised solely of local governments (cities, counties, and special districts) or also includes higher-level public organizations (e.g., state and Federal agencies), tribal governments, and external organizations such as businesses, agricultural interests, non-governmental organizations (NGOs), and universities.

H3: Diverse representation in a collaborative group is associated with beneficial environmental outcomes.

Group Formalization

Formalization (H4) distinguishes between collaborative efforts that are more ad hoc and those that have a stronger institutional presence (Alexander 1993; Huxham and Vangen 2005; Imperial 2005; Margerum 2011). While formal group structures and processes are found to enhance collaborative group function and longevity (Ferguson 2004; Margerum and Born 2000) (and increased resource support in general is found to enhance group efficacy (Curtis and Byron 2002; Parker et al. 2010; Yaffee et al. 1996)), it remains unclear how specific resource expenditures, such as hiring a dedicated coordinator or producing more specific plans and agendas, affect group impact. I compare groups on two aspects of formalization: (1) the presence of a dedicated coordinator; and (2) whether a group has itemized goals or objectives. In some cases a coordinator can provide key administrative support and ease group tensions (Imperial 2005; Huxham and Vangen 2000; Margerum 2002; Susskind and Cruikshank 1987; Susskind et al. 1999). Likewise, better-specified goals and objectives can "help motivate groups to resolve conflicts" (Margerum 2011, 121) (see also Mattessich et al. 2001; Susskind and Cruikshank 1987), enable groups to better assess their efficacy and focus their efforts (Anderson 1995; Hoch 2000; Innes and Booher 1999; Levy 2013; Margerum 2011; Wondolleck and Yaffee 2000), and clearly allocate responsibilities (Margerum and Holland 2000). On the other hand, coordinators are not free, and there can be significant opportunity costs associated with efforts to further formalize group processes or better specify plans (Margerum 2011; Wood and Gray 1991). Nonetheless, I hypothesize that more formalized groups will be more strongly associated with improved water quality

H4: Increased formalization of a collaborative group is associated with beneficial environmental outcomes.

METHODOLOGY

Estimating the effect of collaborative watershed groups

A direct comparison between the treatment group (watersheds with an active collaborative group) and control group (watersheds without an active collaborative group) is inappropriate, since self-selection into the treatment group attributable to characteristics that also affect watershed conditions. I address the issue of selection bias using a matching method (Rosenbaum and Rubin 1983) that estimates the average treatment effect (ATE)[1] (Cameron and Trivedi 2005) using an augmented inverse propensity weighted estimator (AIPW) (Glynn and Quinn 2010).

The AIPW estimator ($\widehat{ATE}_{AIPW}$) (see also Robins et al. 1994; Scharfstein et al. 1999) involves two basic elements: (Step 1) fitting a model that estimates the probability of "treatment" (in this case, the presence of an active collaborative group) as a function of relevant observables (i.e., a propensity score, or the estimated probability that a given observation falls in the treatment group (Rosenbaum and Rubin 1983; Cameron and Trivedi 2005)); and (Step 2) fitting two models that estimate the outcome variable[2] of interest under treatment and control conditions, respectively, and weighting each outcome estimate by the propensity scores estimated in Step 1 in order to produce a weighted average of the two regression estimators (Glynn and Quinn 2010). Essentially, the two regression models fit in Step 2 are used to estimate a contrast between what would happen if every observation were put in the control group and what would happen if every observation were put in the treatment group (Freedman and Berk 2008; Robins and Rotnitzky 1995). This adjustment is applied to the standard inverse propensity weight (IPW) estimator (which simply estimates the ATE as the average difference between the treatment and control groups after weighting each obser-

---

[1] The ATE is defined theoretically as $ATE = \mathbb{E}[Y(1) - Y(0)]$

[2] As the presence of a collaborative group can predate both surveys or just the NRSA, I do not model the *change* in outcomes between the WSA and NRSA, since both the WSA and the NRSA present potentially relevant "post-treatment" outcomes.

vation by its corresponding propensity score) to take advantage of the information in the conditioning set (the data used to estimate the propensity scores) and to improve the small sample properties of the IPW estimator (Glynn and Quinn 2010). I specify the $\widehat{ATE}_{AIPW}$ estimator and describe the technical details of this approach, in particular the analytical advantages of the AIPW estimator relative to the IPW estimator, in Appendix A.[3]

The $\widehat{ATE}_{AIPW}$ estimator only removes selection bias if it suitably accounts for the factors that motivate selection into the treatment group (Cameron and Trivedi 2005, 873). For this analysis, this assumption is well-founded, as Lubell et al. (2002) provide a comprehensive analysis of the contextual factors that motivate the formation of collaborative watershed groups. By including variables in the propensity score model that Lubell et al. (2002) identify as key drivers, I am confident that this model removes a great deal of the omitted variable bias. The multilevel logistic regression model used to estimate propensity scores ($Pr(Z = 1)$) is specified:

$$Pr(Z_i = 1|X_i) = logit^{-1}(\gamma_{e[i]} + \theta_{o[i]} + X_i\beta) \tag{2.1}$$

where the probability of being in the treatment group is modeled as a function of covariate vector X, which includes the variables identified by Lubell et al. (2002) as important predictors of group presence. Specifically, for each observation i, X includes developed, forested, and agricultural land cover, population density, active NPDES permits (for a five year period prior to the WSA or NRSA), the ratio of NPDES enforcement actions to permits (within the same five year period), watershed area, and median income.[4] To allow for the possibility that groups occur more frequently in particular geographic regions and become more prevalent

[3] Though it is also possible to estimate the treatment effect by including relevant covariates and the estimated propensity scores directly in a standard regression model, an advantage of the AIPW estimator is that it relaxes the linearity assumption of a regression model, instead differencing the outcomes of collaborative watersheds and the weighted matched non-collaborative watersheds (Black and Smith 2004).

[4] p(Treatment) ∼ Agriculture + % Forest + % Developed + Watershed Area + Pop. Density + Median Income + NPDES Permits + NPDES Enf. Ratio + State + Year (Ecoregion and Year are random effects)

over time, I estimate propensity scores using a multilevel logistic regression model that fits random intercept terms $\gamma_{e[i]}$ for each Omernik Level II Ecoregion e and $\tau_{t[i]}$ for each year t.[5]

Two important empirical considerations for the propensity score estimation model are: (1) that the "conditioning set," i.e., the variables with which propensity scores are estimated, are relatively similar between the treatment and control groups; and (2) that the distributions of estimated propensity scores for the treatment and control groups generally encompass the same range so as to provide common support (Glynn and Quinn 2010; King and Zeng 2006). For instance, if estimated propensity scores for observations in the control group are between 0.05 and 0.88, a treatment observation with a propensity score of 0.95 is not adequately supported by the model, since the model was fit without any observations with a propensity score greater than 0.88. Since the AIPW estimator weights observations in accordance to their observed similarity, the propensity score distributions do not need to be perfectly congruent, but it is at least important that they sufficiently overlap. Appendix B examines the covariate balance between the treatment and control groups in greater detail, demonstrating that the selection model has common support, i.e., that the estimated propensity scores for the observations in the treatment and control groups span a similar range (and that, while not identical, the frequency distribution of scores for each group largely overlap).

For Step 2 of the AIPW estimation procedure, I use a pair of multilevel models to estimate the water quality outcomes under treatment (collaborative governance) and control

[5] I use Ecoregion instead of state as a geographic grouping indicator because state-level random effects result in overfitting. For a few states in the data, there are either no treatment (Massachusetts, New Hampshire, New Jersey, Oklahoma, and Kansas) or control (Georgia) observations. In reality however, the "population" of watersheds in each state includes watersheds with and without an active collaborative group. While the state in which an observation occurs is an important predictor of selection, in this case the state variable is too good of a predictor for the propensity score model (since predictions are based solely on observed data). Ecoregion is an excellent proxy, because the nine different Level II Omernik Ecoregions in these data are able to capture geographic context (political, social, and environmental variables that might influence selection and make the presence of a collaborative group more likely) without being subject to sampling zeros that greatly increase the number of estimated propensity scores at or near 0 or 1.

conditions. Each model includes: (1) observed covariates at the individual observation level to minimize omitted variable bias;[6] and (2) models group-level random effects so as to adjust for lack of independence amongst samples taken in multiple time periods from the same site or from different sites in the same geographic region.[7]

Each model groups observations by state, 4-digit Hydrologic Unit Code (HUC4) sub-basin, and year, as well as by the two points of randomization in the WSA and NRSA sampling design: Level II Omernik Ecoregion and Strahler stream order (both described below in the Data section).[8] At the first level of the model I estimate water quality outcomes for individual stream-year i in sub-basin w, state s, year t, ecoregion e, and stream order o (Equation 2.2)[9]:

---

[6]Note that these covariates do not need to be identical to the conditioning set used to estimate the propensity scores in Step 1 (Glynn and Quinn 2010); thus, in Step 1 I specify only those covariates identified by Lubell et al. (2002) as being key predictors of collaborative governance, and in Step 2 I include some of these same covariates but also additional variables which are related directly to water quality outcomes.

[7]The advantage, relative to a more common fixed effects approach, is that a multilevel model accounts for uncertainty associated with each group-level adjustment (Gelman and Hill 2006; Raudenbush and Bryk 2001) by shrinking the adjustment towards the overall sample mean as the size of the group decreases. In other words, as the within-group sample size decreases, the model places more credence upon the whole sample estimate, and vice versa. This "partial pooling" takes advantage of more available information (Greenland et al. 1991; Poole 1991) and avoids overstating differences between groups (Gelman 2006; Gelman and Hill 2006). For data in which individual observations are nested within higher-level groupings, a multilevel model produces more reasonable inferences (Gelman 2006) and more reliable estimates (Gelman et al. 2012).

[8]Since these groups are 'non-nested,' such that two observations can be in the same HUC4 sub-basin but different states, or vice versa, the model is a 'cross-classified' model (Gelman and Hill 2006; Raudenbush and Bryk 2001)

[9]Empirically, each model is specified as: Outcome Metric $\sim$ Site Disturbance + % Agriculture + % Forest + % Developed + Pop. Density + Median Income + Road Density + HUC4 + State + Year + Stream Order + Ecoregion (HUC4, State, Year, Stream Order and Ecoregion are random effects)

$$Y_i = \alpha_{w[i]} + \lambda_{s[i]} + \tau_{t[i]} + \gamma_{e[i]} + \theta_{o[i]} + \sum_l \delta_l Site_i + \varepsilon_{iwsteo} \tag{2.2}$$

where Y_i represents the dependent variable, a given quality metric (e.g., nitrogen level) for sample i. Accordingly, $\alpha_{w[i]}$ represents the conditional intercept estimate for i given that it is in HUC4 basin w; similarly, $\lambda_{s[i]}$ represents the conditional intercept estimate for state s, $\tau_{t[i]}$ the conditional intercept estimate for year t, $\gamma_{e[i]}$ the conditional intercept estimate for ecoregion e, and $\theta o[i]$ the conditional intercept estimate for stream order o. Next, δ_l represents a vector of control parameters for a given site 1 to l ($Site_{lji}$) (listed in Footnote 9. Finally, ε_{iwsteo} represents the random error associated with observation i. Note that each of the random intercepts are themselves modeled; for instance HUC4 groups are modeled as:

$$\alpha_w = \alpha_0 + \mu_w \tag{2.3}$$

in which α_0 represents the average outcome across basins and basin-level random error is denoted as μ_w. State, year, ecoregion, and stream order random effects are modeled in the same way (i.e., the group level outcome as a function of the across group outcome and group-level random error); these equations are omitted for space considerations.[10] Even though the AIPW estimator has many advantages, it remains possible that resultant ATE estimates are biased upwards due to unobserved factors that are positively related both the the presence of a group and water quality outcomes. These data remain observational in nature, and accordingly these results should not be considered to necessarily provide an unbiased causal estimate. Nonetheless, conditioning on observed variables identified in the literature as being key to selection likely absorbs most of the influence of unobserved nonrandom drivers. Even

[10]In discussing multilevel models, it is important to note that the standard heuristics applied to fitting parameters in ordinary least squares regression and similar (e.g., logistic) models, statistical significance, is inappropriate for determining which group-level indicators to leave in and which to leave out (Gelman and Hill 2006, 271). For instance, the model includes a grouping indicator for each group, not just the indicators found to be statistically significant. This is because the focus of the analysis is not on examining inter-group differences, but rather on generating the best possible estimate.

if omitted variable bias remains, it is likely to be small, and in the absence of more rigorous experimental designs these estimates provide better evidence than currently exists for policy makers considering initiation or support of a collaborative watershed group.

Comparing Different Types of Groups

The second part of this analysis aims to compare the predicted effects of different types of collaborative watershed groups. To estimate how group characteristics affect predicted group impact, I fit a single multilevel model that expands upon Equation 2.2 by adding three additional terms:

$$Y_i = \alpha_{w[i]} + \lambda_{s[i]} + \tau_{t[i]} + \gamma_{e[i]} + \theta_{o[i]} + p(C_i) + C_i + \sum_k \beta_k Collab_{ki} C_i + \sum_l \delta_l Site_i + \varepsilon_{iwsteo} \quad (2.4)$$

Equation 2.4 adds three elements to Equation 2.2: the propensity score ($p(C_i)$ for each observation as estimated in Equation 2.1, a main effect for collaborative group presence (C_i), and a summation term (Σ_k) representing a vector of the predicted change in water quality for each group characteristic 1 to k for observation i ($Collab_{ki}$), conditional on the presence of a group C_i. In other words, each observation i is associated with a binary variable (C_i) reflecting whether that observation is within a watershed with an active collaborative group, and then a series of interaction terms ($C_i * Collab_{ki}$) which model the difference for groups with and without a given characteristic (e.g., group coordinator). Having an active group is an obvious prerequisite for having a group coordinator or any other group characteristic. Accordingly, these interaction terms provide a more meaningful–and empirically grounded–interpretation, since the potential impact of any specific management characteristic should rightfully be expressed as altering the predicted impact of a collaborative group and not independently. For observations without an active collaborative group, each interaction term automatically cancels out (since $C_i = 0$). In the next section, I describe the data used to fit these models.

DATA

Dependent Variables

The data used to assess water quality outcomes come from two national surveys, the Wadeable Streams Assessment (WSA) and the National Rivers and Streams Assessment (NRSA). The WSA, conducted in 2004-2005, sampled 1392 stream sites that were randomly selected from all streams of a given size within an ecological region. In other words, the sampling was stratified by ecological region and stream size.[11] The probability-based design used stratification to generate a random, representative sample by ecoregion and EPA region. This presents a unique opportunity for empirical research since most research on collaborative governance selects on either the independent (management characteristics) or dependent (outcome) variables. The NRSA, conducted in 2008 and 2009, re-sampled 357 original WSA sites. These 357 sites for the basis of this analysis. Figure 2.1 shows the location of each site.

The WSA and NRSA assess the ecological condition of each site according to a series of measurements of chemical stressors, metrics of physical condition, and biological indicators. From these data, six variables are selected to provide a holistic representation of stream condition and water quality: total phosphorus content and total nitrogen content (chemical stressors caused by human activities such as mining or agriculture), water turbidity and

---

[11] The WSA surveyed only perennial, wadeable streams. Perennial refers to streams that flow year round under conditions of normal precipitation. The WSA sampling protocol is stratified by Strahler stream order. "Wadeable streams," i.e., those that can be sampled without using a boat, are generally considered to be of orders 1 through 5. However, Strahler ordering does not directly correspond to stream size; rather, the Strahler protocol orders models streams as directed graphs, analogous to a tree. Ordering proceeds in reverse from bottom to top, thus a "leaf" stream, i.e., one that has no tributaries, is of Order 1. The Ohio River is an 8^{th} order stream, the Mississippi River is a 10^{th} order stream, and the Amazon River in South America is a 12^{th} order stream. The sample was also stratified by the 9 (of 15 total) Omernik North American Level II ecoregions that occur in the continental United States, such as the Great Plains and Mediterranean California.

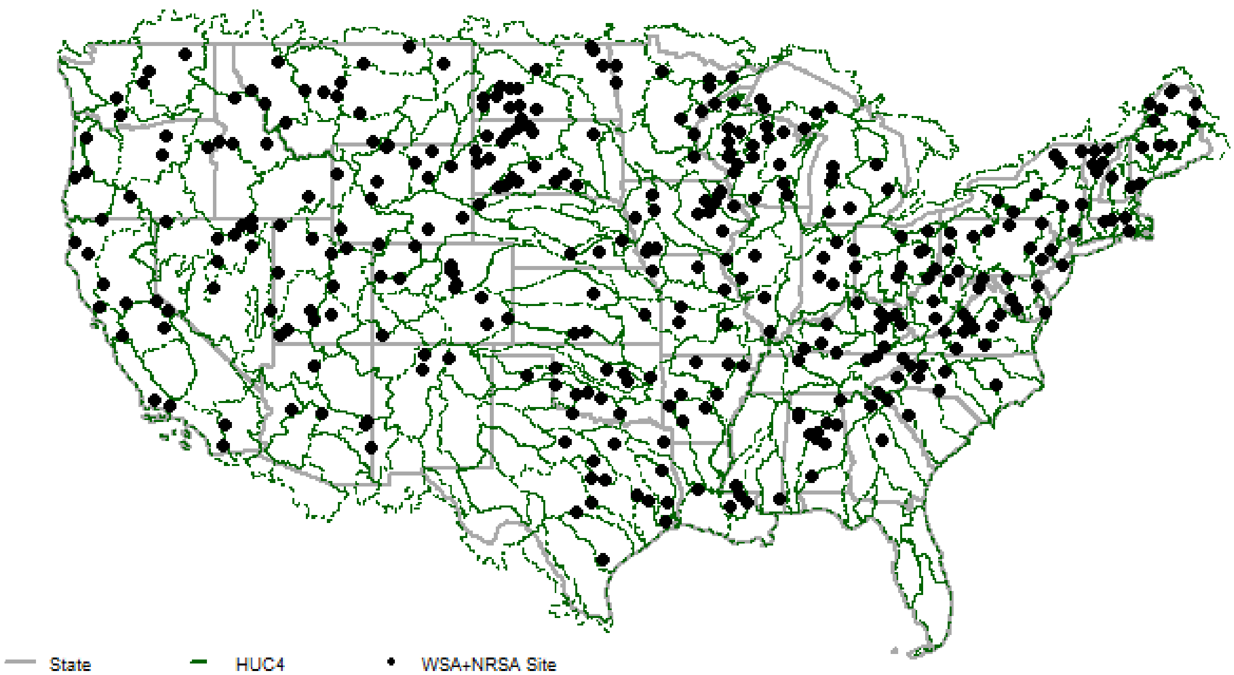


Figure 2.1: Sites Sampled in WSA and NRSA

in-stream natural habitat (physical indicators reflect more proximate habitat destruction), and indices of riparian vegetation and benthic community abundance (biological indicators of condition).[12][13] In order to measure water quality and stream condition holistically, two variables were are specifically chosen from each broader category (chemical, physical, and biological). The particular indicators used selected on the basis of presence in both the WSA and NRSA and completeness of the data. Along with the preceding footnotes, Table

---

[12]Total phosphorus and total nitrogen content are both measured in absolute terms, using micrograms per liter (ug/L) as units. Turbidity is measured in Nephelometric Turbidity Units (NTUs), using a tool called a nephelometer, which gauges the amount of light reflected by the particles in water. in-stream habitat complexity and riparian cover are both calculated using line-transect surveys which calculated the summed areal proportion of each cover type. For instance, to calculate habitat complexity the surveyor assesses coverage at specific points in a 10 meter by 20 meter littoral plot. These data are then used to estimate the areal proportion of the reach that contains natural cover for fish and other aquatic fauna. Because this metric is a summation of the proportion of the reach that is covered by several different kinds of cover, including boulders, large woody debris, and overhanging vegetation, this value can be greater than 1. In the data used for this analysis, sites range in value for the variable from 0 to 2.58. Similarly, because riparian cover is a summation of the proportion of the streamside riparian area that is covered by canopy, mid-layer, and ground-level cover, this value can be greater than 1 as well. Sites range in value for this metric from 0 to 2.18 in the data.

[13]The benthic condition index is more complicated. To assess benthic condition, the WSA and NRSA generate an index for macroinvertebrate assemblage by assessing "least disturbed" sites in each ecoregion, using these sites as the basis of comparison for assessing stream conditions. There are numerous ways to assess the condition of a macroinvertebrate community, including abundance, composition, diversity, and various submetrics related to particular taxa. Further, the appropriateness and significance of these various metrics can differ by region. Thus, for each of the 9 ecoregions within which sampling was stratified, a particular subset of 6 benthic community metrics were chosen upon which to generate a macroinvertebrate multimetric index (MMI) for each ecoregion (each individual metric is scored on a 1 to 10 scale, after which all six metrics are summed and then normalized to a 0 to 100 scale). Metrics were chosen on the basis of sensitivity to human disturbance, commonness amongst sites, independence of candidate metrics, and applicability across ecoregions (*National Rivers and Streams Assessment: 2008-2009* 2013). For instance, in the Xeric Ecoregion (comprised of the Great Basin, much of Southern California, and the Intermountain West), the MMI incorporates metrics for Non-insect % Distinct Taxa, % Individuals in Top 5 Taxa, Scraper Richness, Clinger % Distinct Taxa, EPT Richness Distinct Taxa, and Tolerant % Distinct Individuals. In total, there are 21 different metrics that are part of the MMI for at least one ecoregion.

2.1 provides more detail about the dependent variables.

Table 2.1: Outcome Metrics (Unstandardized)

	Mean	SD	Units	Details
Phosphorus	131.68	402.49	ug/L	Total nitrogen content
Nitrogen	1216.69	2206.71	ug/L	Total phosphorus content
Turbidity	289.08	166.67	NTU	Turbidity level
Benthic Health	351.53	206.20	Index Score (0-100)	Benthic multi-metric index
Riparian Cover	298.19	183.53	Sum Areal Prop.	Ground + mid + canopy cover
In-stream Habitat	197.37	133.28	Sum Areal Prop.	All natural cover types

To facilitate comparison comparison across outcomes, each outcome metric is log-transformed (to achieve a more normal distribution), and then mean centered and divided by two standard deviations (Gelman 2008). This is particularly important for phosphorus, nitrogen, and turbidity, which as shown in Table 2.1 each have an extremely positive skew (i.e., a few observations have very high values). While this method of standardization makes direct interpretation more difficult than does using untransformed or log-transformed inputs (which can be interpreted as simple elasticities), it offers three advantages for this analysis. For the ATE estimates, standardized effects can be compared across metrics which are originally on different scales. Further, with regards to the regression models used in the second part of the analysis, this method of standardization renders continuous variables on a similar scale to untransformed binary variables (Gelman 2008). This allows for comparison between binary or categorical variables of interest (related to the presence of a collaborative group) and other model inputs. Finally, the parameter estimates for transformed continuous inputs compare predicted change associated with said variable moving from a low or high value (or vice versa), as the coefficient reflects the change in outcome predicted by a 2 standard deviation change in the input.

Covariates

Publicly available external data are also incorporated into this analysis, primarily for modeling propensity scores as described above. Watershed land cover data (the percentages of HUC8 land cover that are impervious, used for agricultural purposes, or covered by wetlands and forests) are obtained from the National Land Cover Database (NLCD). Income and population data are obtained from the American Community Survey (ACS). Government spending data are procured from the U.S. Census Bureau (stemming from the Census of Governments). National Pollutant Discharge Elimination System (NPDES) permitting and enforcement data are obtained from the U.S. EPA. These data, and the scripts used to produce these data, are available by request.

Independent Variables

In order to develop a watershed management database, data were collected from (1) legislative documents that allocate management responsibilities and funds to groups; (2) group reports, mission statements, membership lists, and constitutional documents; and (3) watershed management plans (specifically the portion of each plan that discusses the use and role of public involvement) for each of the 357 watersheds that were sampled for both the NRSA and WSA. In very few cases are the majority of these sources available for a given watershed, so a primary challenge is to apply a uniform coding scheme to diverse sources.

The coding process for each watershed begins at the EPA's "Surf Your Watershed" site for the HUC8 designation associated with the observation.[14] This page provides background information including the state(s) and county(ies) with land area in the watershed, the primary watershed name, and links to various monitoring websites and in some cases local watershed organizations. I then proceeded to search for the documentation described above, starting with links provided on the EPA page, proceeding to state and local government documentation and databases, and finally an conducting an extensive Google search using

[14] http://cfpub.epa.gov/surf/locate/index.cfm

keywords (e.g., “watershed council”, “river management group”, etc.) and local geographic names to find groups and data sources without a presence in official channels. All sources used to develop these data are available from the author. This multi-source approach, taking advantage of the various resources available on state and Federal agency websites and databases, is quite similar (though expanded) to the approach used by (Moore and Koontz 2003) to identify and survey watershed groups in the state of Ohio.

In determining whether a watershed is considered for the purpose of this analysis to be managed collaboratively, only groups in which at least one governmental entity participates are included. Since the focus of this research is on the use of collaborative governance for public purposes, cases of interest are those in which public agencies act as “initiators and instigators of collaborative governance" (Ansell and Gash 2008, 545) by devoting time and resources to the group. The ultimate question then is whether such public expenditure improves policy outcomes, in this case water quality. This coding strategy proves inclusive, encompassing a wide variety of inter-organizational collaborative institutions with the exception of local citizen groups. These advocacy-oriented groups are not of interest in this particular study given my specific focus on collaborative efforts that are initiated or supported by public managers (i.e., instances in which a public entity has chosen to devote resources towards collaborative governance).

Variables of interest are coded as follows:

Dedicated Coordinator: Groups were coded ”1” if the group does have a designated coordinator or director and ”0” otherwise. This variable does not reflect the coordinator’s FTE.

Objective Formalization: Anderson (1995) and Margerum (2011) distinguish between three ways in which groups codify their aims and purposes: (1) “mission statements": a broadly conceived sentence (or paragraph) that provides a general statement about the impetus and aims of the group; (2) “goals": itemized, but unspecific, tenets that “describe a desired future condition" (Margerum 2011, 126) (e.g., “I. Improve water quality in river; II. Increase awareness about environmental behavior in community."); and (3) “objectives": itemized statements that outline specific actions intended by the group and/or specific metrics by which the group is able to measure its output or outcomes (Anderson 1995). In practice, the distinctions between goals and objectives are somewhat blurry; perhaps most problematically, the list of aims published by a group often contains a mix of both goals and objectives (i.e., some items are specific and measurable and some are not). Thus, since this analysis does not delve deeply into the content of group goals and objectives, it is most appropriate for this analysis to code a binary variable comparing groups that only publish a mission statement (‘0’) and groups which develop an itemized list of goals and objectives (‘1’). This facilitates a comparison between groups which more clearly codify their purposes by developing an itemized list of motives and tasks and those that do not.

Diversity of Representation: As specified above, since this study concerns publicly supported collaborative governance efforts, the baseline requirement for a watershed group to be coded as such is that the group includes a public institution as a member. Thus, the “null value" for a group’s diversity is a group that is comprised solely of local governmental representatives. Groups are scored for the presence of tribal governments, businesses, local stakeholders (e.g., advocacy organizations), non-governmental organizations (e.g., Nature Conservancy), research or educational organizations such as universities or colleges, agricultural interests, Federal agencies, and state agencies. A group receives either a “1” (present) or “0” (absent) reflecting membership by each other type of organization. These values are then summed. Thus, if a group is constituted solely from representatives of local government, tribes, and the business community, then said group’s score for the number of stakeholder types included is a 2 (since membership by local government is requisite for inclusion in the analysis).

Group Responsibility: In order to develop a comprehensive coding scheme for the types of responsibility policy makers accord to a collaborative group, seven general categories of tasks that emerge inductively from the data are employed: Planning, management, outreach, monitoring, coordination, projects, and education. Collaborative group activities such as joint policy implementation are more intensive than activities such as information sharing because they entail greater transaction costs (Margerum 2007; Wondolleck and Yaffee 2000). Practical distinctions between many of these activities are not always concrete (for instance, a group that uses "restoration projects" for "education" and for "outreach"). Even without more detailed data concerning group activities and responsibilities, a general contrast emerges between groups that engage in management activities and those that do not. Many groups serve as information sharing forums or conduct restoration, education, or outreach projects; others engage in management activities such as overseeing endangered species recovery efforts or land use planning and management. For this variable, a group is coded as "1" if it has management responsibilities (e.g., the group itself is the lead entity on an environmental restoration plan or for Endangered Species Act recovery actions, or a group manages land use in the watershed) and "0" if it does not have such responsibilities.

Table 2.2 summarizes the distribution of these variables across groups. In the data collection process it became apparent that groups vary considerably in terms of their "presence" in grey literature (e.g., agency reports) and on the internet. Some group websites contain an archival section from which past yearly reports and older documents are accessible, or a specific page which references staff or organizational members. For other groups, a more deductive approach is necessary. For instance, a group resolution might be co-signed at the bottom by group members. These data would then be used to record membership of different stakeholder types. This heterogeneity increases the potential for Type II error, either the conclusion that a group does not exist (or more likely) overlooking a specific group characteristic simply because a given document or textual reference is not found or is not available. While I am unable to eliminate this potential source of bias, I am confident that any bias is likely to be quite small for several reasons. First, I employ a consistent data

Table 2.2: Group Variables

Variable	**Levels**	**n**	%
Has Coordinator	None	54	23.2
	Dedicated Coordinator	179	76.8
	all	233	100.0
Goal Formalization	0	138	59.2
	1	95	40.8
	all	233	100.0
Group Responsibility	0	129	55.4
	1	104	44.6
	all	233	100.0
Total Stakeholder Types	0 (Local gov. only)	12	5.2
	1	13	5.6
	2	15	6.4
	3	25	10.7
	4	50	21.5
	5	64	27.5
	6	35	15.0
	7	14	6.0
	8	5	2.1
	all	233	100.0

discovery and coding protocol to limit bias due to collection methods. Second, Leach et al. (2002) show that concerted efforts in small geographic areas are successful in identifying additional groups and group characteristics; investigator time and resource limitations, in that I am only able to devote a few hours of time to any one observation and am unable to visit any sites, are thus the main cause of Type II error. For this reason, I expect that underidentification is random, meaning that it increases standard errors but does not bias the results (Lubell et al. 2002). As with Lubell et al. (2002), this analysis sacrifices the level of detail that would be affordable with a regional approach in favor of national generalizability. Third, it is possible that a group's choice not to maintain an active public presence and provide up-to-date records is not randomly distributed. While this could also bias treatment estimates, the implications that such a choice holds for a group's environmental impact is unclear, and there is not compelling rationale that would indicate this significantly biases the results. Finally, the data collection process I employ is similar to methods that have been used–and published–in the past (e.g., Lubell et al. 2002; Leach et al. 2002; Moore and Koontz 2003).

The coding process itself is similar to that of qualitative document analysis (QDA) (Altheide et al. 2008; Altheide and Schneider 2012), often used in political science. Since QDA involves the qualitative coding of textual sources for meaning, precision and impartiality are primary methodological concerns (Guba and Lincoln 1994). This analysis is concerned with manifest structures, rather than latent concepts. Thus, coding in this case is primarily a question of identification, rather than one of subjective interpretation. For instance, if a group document lists an individual as being a "coordinator" or "executive director," than a group is coded as having a coordinator. Thus, I do not believe that partiality is a significant concern in this analysis. To address precision, I provide an "audit trail" (Platt 2006) in Appendix C that presents the coding protocol applied to each textual resource. This provides an overview of the analytical process applied to each data source. Likewise, in adherence to the recommendation of Guba and Lincoln (1994) to provide full access to data so that findings can be replicated and verified, the author intends to make available the data sources employed (including group websites, plans, reports, etc.) for each assessed watershed. These are available by email using the contact information above.

RESULTS

H1: Collaborative Group Presence

In evaluating these results one should be concerned not only with the statistical significance of the parameters of interest, but also with how the estimated effect of a variable behaves across all six outcome metrics. Colloquially, one might interpret increased levels of nitrogen, phosphorus, and suspended solids and decreased vegetation, in-stream habitat, and benthic abundance as "bad for the environment" and the converse as "good for the environment." In interpreting ATE estimates and model coefficients then, it is important to note that the dependent variables are not uniform in directionality. So that each estimated parameter reflects the direction of predicted change in the outcome variable, the directionality of each variable is kept "as-is." Generally, if collaborative watershed management is associated with environmental improvement, one might expect to observe a negative ATE for the phosphorus, nitrogen, and turbidity level models, and a positive ATE for benthic community health, riparian cover, and habitat complexity models. The same holds true for subsequent regression models. However, not all policies and programs will affect all of these variables simultaneously. For instance, a program that targets sources of nonpoint pollution such as fertilizer use might significantly affect water chemical content but have no bearing on riparian habitat. Thus, while using six metrics in concert provides a holistic conception of water quality, one should not necessarily expect any effect to perform in a wholly consistent way across each outcome metric. I discuss this issue in greater detail in the context of the model results below.

Of the 357 sites sampled under both the WSA and NRSA, 124 are found to have a collaborative watershed management group at the time of the WSA sample, and 167 are found to have a collaborative watershed management group at the time of the NRSA sample. However, one issue regarding the assignment of watersheds into the "treatment" group is that the various outputs of a collaborative group (such as plans or joint projects) do not likely have an immediate effect on on-the-ground conditions; instead, it is likely that such any effect would take time to be realized. For this reason, it makes little pragmatic sense to model a sample taken in the same year in which a group was formed as being in the treatment group. While there is limited evidence about how long it takes for group actions to manifest, Leach et al. (2002) and Leach and Sabatier (2005) find that perceived success (on the part of participants) increases after groups have been active for approximately four years (of course, as discussed previously it is unclear how perceived success relates to actual outcomes). Based on these results of Leach et al. (2002) and Leach and Sabatier (2005), I model all watersheds in which a collaborative group has been active for at least four years prior to the sample date as being in the "treatment" group. This results in a treatment group size of 233 (87 WSA samples and 146 NRSA samples), with 481 observations in the control group. The treatment estimates obtained using the AIPW estimator are shown in Table 2.3. Standard errors for each ATE are estimated via bootstrapping (Glynn and Quinn 2010; Funk et al. 2010).

Table 2.3: Average Treatment Effect (ATE) (Y>=4)

	Phoshorus	Nitrogen	Turbidity	Benthic	Riparian	In-stream
ATE	-0.08^*	-0.10^*	-0.07^*	0.03	0.05	0.11^*
	[−0.16; −0.02]	[−0.17; −0.02]	[−0.16; −0.01]	[−0.06; 0.11]	[−0.02; 0.11]	[0.02; 0.18]

* 0 outside the confidence interval

Table 2.3 presents bootstrapped confidence intervals for each ATE estimate (the average effect of a collaborative group that has been active for at least 4 years prior to the observation). These bounds represent 95% confidence intervals for each ATE as estimated by 500 bootstrap samples. I label as significant any ATE estimate for which the bootstrapped 95% confidence interval does not contain zero. Four of the six ATE estimates (phosphorus, nitrogen, turbidity, and in-stream habitat complexity) are thus found to be significant with 95% confidence. All four of these ATE estimates also have a sign suggesting that collaborative groups engender environmental improvement. For interpretation, it is helpful to think of the ATE estimates as if they are each a regression coefficient associated with a binary treatment variable, in this case a collaborative watershed management group that has been active for at least four years. Again, each outcome metric is log-transformed and then standardized by mean centering and then dividing by two standard deviations (see Gelman 2008). Thus, the expected phosphorus level for a watershed in the treatment group is (i.e., treatment = 1 vs. treatment = 0) is 21.5% less than a watershed in the control group. Since the standardized unit is 2 standard deviations of the log-transformed phosphorus variable (the standard deviation of which equals 1.52), we can multiply the coefficient by twice the standard deviation, and then exponentiate the result to produce a multiplicative effect estimate of 0.78 ($exp(-0.08 * 1.52 * 2) = 0.78$). This predicts that a watershed with a collaborative group will have a phosphorus level 22% below that of an untreated watershed. Similarly, the suggested effect on nitrogen and turbidity are a reduction of 23% ($sd = 1.29$) and 21% ($sd = 1.65$), respectively. In-stream habitat complexity is is predicted to increase by 15% ($sd = 0.62$). The suggested effects on benthic community health and riparian cover are both negligible and insignificant.

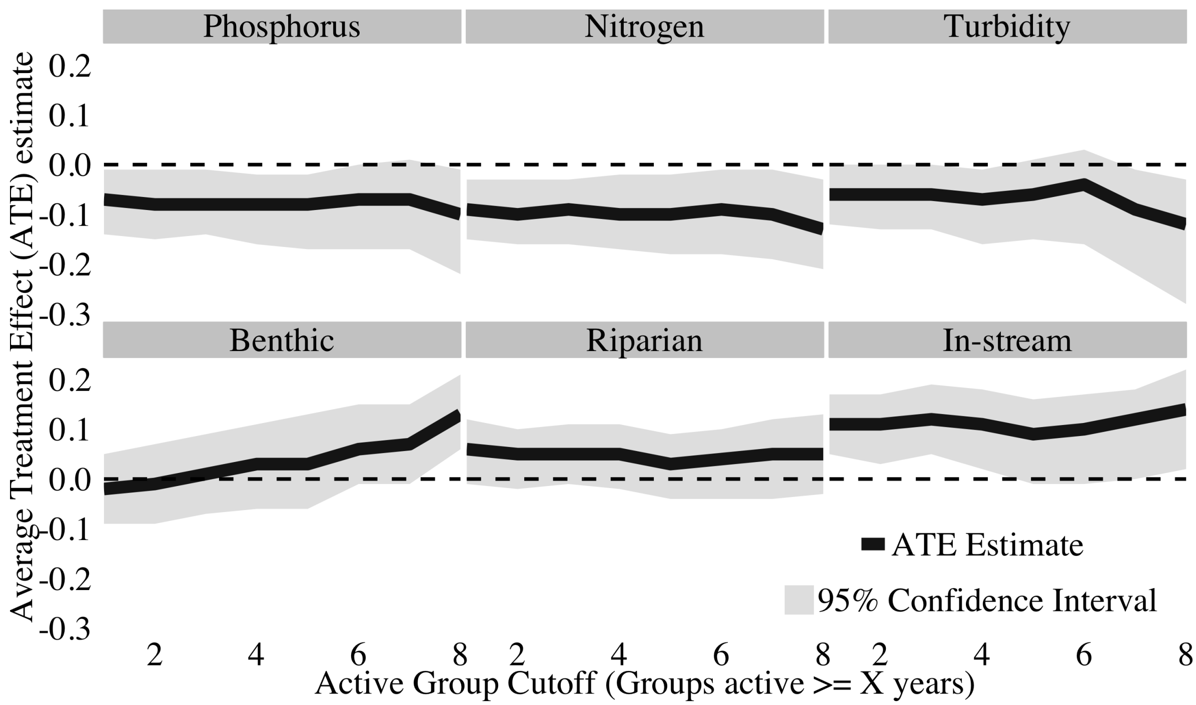


Figure 2.2: ATE estimates with varying cutoff

These results can perhaps be explained by considering the extent to which a collaborative watershed group might have influence over each of these metrics. Of these six metrics, riparian cover is most subject to the influence of the land owner directly proximate to the sample site; it is not likely that actions elsewhere in the watershed meaningfully influence riparian cover at the site. Thus, finding a significant increase in riparian cover is perhaps a "hard case," in that it would require the group to exert some form of influence directly on that plot of land. Conversely, land use and management actions taken elsewhere in the watershed that reduce net erosion or chemical pollution are likely to indirectly affect stream conditions at the sample site. Simply put, one might say that riparian cover more closely depends on actions taken *at* the sample site, whereas in-stream vegetation, turbidity, or phosphorus content to a greater extent depend on actions taken *somewhere* in the watershed. The negligible predicted difference in benthic health is perhaps explained by the link between riparian cover and benthic health, as benthic heath is shown to be sensitive to proximate conditions such as riparian cover (Sweeney et al. 2004). Further, the impacts of upstream logging and other disturbances on benthic community health are shown to resonate up to 40 years after such behavior has ceased (Zhang et al. 2009); thus, it is possible that benthic conditions change on a much longer time scale and thus most groups have simply not been active long enough for there to be a detectable effect.

Figure 2.2 presents the results of a sensitivity analysis which would seem to support this interpretation. In Figure 2.2, the cutoff for an active group (e.g., all groups active at least four years or six years) varies along the x-axis for each outcome metric, while the ATE estimate associated varies on the y-axis. Generally, the parameter estimates remain fairly consistent, supporting the use of the four-year cutoff that I explore most fully in this analysis. As expected, one result that occurs as the cutoff is raised (i.e., requiring a group to have been active for more years to be considered part of the treatment group) is that the confidence interval surrounding the ATE estimate becomes slightly wider due to the decreasing sample size of the treatment group as the cutoff becomes more stringent.[15]. The most interesting finding emerging from the sensitivity analysis is that the ATE for benthic health increases steadily as the cutoff for an active group increases. When the ATE is estimated using only groups that have been active at least 8 years or more, the presence of a collaborative group is predicted to have a statistically significant positive effect on benthic health (shown in the panel as the confidence interval does not span the dashed line representing an estimate of zero effect). Since benthic community health is perhaps slowest to respond to new management practices, this lends further support for the contention that

[15]In the data, there are 291 observations associated with an active group. As the number of years required to be considered part of the treatment group increases, the treatment group sample size declines to 270 (2 or more years), 260 (3), 233 (4), 219 (5), 193 (6), 178 (7), and 159 (8)

collaborative governance does have a beneficial effect on water quality overall. I further address these results in the discussion section below.

Table 2.4 presents the multilevel regression models used to test group characteristics. Each outcome metric is shown in a separate column. Control variables that are not substantively interesting, specifically the propensity scores used to control for selection bias, are not included in Table 2.4. All continuous numeric inputs to each model are standardized via the method described above. Table 2.4 also does not present the random intercept adjustments modeled for HUC4, state, and year groups (fit to account for spatial and temporal dependencies), and for stream order and ecoregion (fit to account for the points of randomization in the WSA and NRSA design). Table 2.4 presents bootstrapped confidence intervals for each parameter; the level of significance specified in the table refers to the maximum bootstrapped interval at which a parameter is "significant," i.e. does not contain 0. This is the optimal way to test hypotheses related to linear mixed model effects, since residual degrees of freedom are uncertain for a multilevel model[16] (Bates et al. 2014; Bolker et al. 2009).

Before assessing the variables of interest, it is important to consider the consistency of parameter estimates for known sources of environmental degradation included as control variables in each model. These models are able to identify established causes of water quality changes, such as road density and agricultural land use. For instance, the results in 2.4 show that an increase agricultural land usage within a watershed has a significant positive effect on phosphorus and nitrogen levels (i.e., increased pollution levels); this speaks to the face validity of this modeling approach.[17] Note that the estimated effect of an active group in

[16] The multilevel model is a compromise between a complete pooling (no fixed effects) and no-pooling (fixed effects) model where the precise amount of pooling differs for each group. Thus, it is unclear what the correct df used to calculate the t or F statistic and thus test a given parameter should be, since the appropriate degrees of freedom presumably differ across each group of observations.

[17] Phosphorus, nitrogen are strongly linked to agricultural land use (Tong and Chen 2002). Similarly, in 2.4 road density is positively related to stream turbidity, phosphorus level, and nitrogen level, and negatively related to benthic condition. This fits with established ecological findings; for instance, roads increase erosion and sediment yield, thereby increasing stream turbidity (Forman 1998; Montgomery 1994), and

Table 2.4 is not directly comparable to the ATE estimates in Table 2.3, because each model in Table 2.4 has an additional interaction terms that acts on the treatment variable. The regression-based ATE estimates will likely differ in any case given that the AIPW estimator uses a non-parametric differencing approach.[18]

H2: Group Responsibility

The interaction term “WG * Management” in Table 2.4 represents the predicted difference in each outcome metric between a groups that have actual management responsibilities and those that do not (e.g., groups that serve as coordinative bodies or that engage in stakeholder outreach and education). As described in the specification of the model above, group characteristics are interacted with group presence because a characteristic only has meaning in the context of an active group (for instance, a group must be active to have any type of responsibility, management or otherwise). Using interaction terms ensures that the group characteristic coefficients can be interpreted as representing the predicted difference between groups with and without said characteristic. Table 2.4 suggests that a group with management responsibilities has a significant negative impact on phosphorus levels and a significant positive impact on benthic community health. Using the same method of interpretation applied to the ATE estimates above (multiplying the parameter estimate by two times the standard deviation of the variable and then exponentiating the result to get a multiplicative effect), a group with management responsibilities is associated with a 37% ($sd = 1.52$) lower

water runoff from roads carries heavy metal pollutants that can harm benthic communities (Forman 1998; Horner and Mar 1983)

[18] A potential complicating factor in testing group characteristics is that correlation between characteristics might hinder simultaneous estimation (i.e., multicollinearity). I tested for this possibility by fitting a distinct model for each group characteristic, and comparing these isolated estimates to the parameter estimates from the unrestricted model including all group characteristics; parameters from the restricted models (one characteristic each) were almost identical to those in the unrestricted model. Thus, I present only the unrestricted model results.

Table 2.4: Multilevel Model Results

	Phoshorus	Nitrogen	Turbidity	Benthic	Riparian	In-stream
Site Disturbance	**0.10*****	**0.10*****	0.001	−0.03	**−0.14*****	−0.02
	(0.04, 0.17)	*(0.05, 0.16)*	*(−0.07, 0.07)*	*(−0.10, 0.05)*	*(−0.21, −0.07)*	*(−0.09, 0.05)*
% Agricultural	**0.25*****	**0.24*****	0.02	**−0.12****	−0.04	−0.02
	(0.13, 0.37)	*(0.15, 0.36)*	*(−0.09, 0.17)*	*(−0.25, −0.01)*	*(−0.16, 0.08)*	*(−0.15, 0.11)*
% Forest	**−0.14****	**−0.23*****	**−0.18****	**0.21*****	**0.19*****	0.05
	(−0.28, 0.004)	*(−0.30, −0.06)*	*(−0.34, −0.05)*	*(0.06, 0.34)*	*(0.06, 0.32)*	*(−0.09, 0.20)*
% Developed	0.01	0.01	−0.05	**0.16****	0.05	0.02
	(−0.12, 0.13)	*(−0.09, 0.10)*	*(−0.26, 0.01)*	*(0.03, 0.29)*	*(−0.06, 0.17)*	*(−0.11, 0.14)*
Pop. Density	**0.11***	**0.18*****	**0.13***	**−0.20*****	−0.01	−0.05
	(−0.02, 0.24)	*(0.07, 0.27)*	*(−0.01, 0.27)*	*(−0.35, −0.06)*	*(−0.14, 0.12)*	*(−0.18, 0.09)*
Med. Income	**−0.14*****	−0.07	**−0.13*****	0.07	0.01	0.06
	(−0.25, −0.05)	*(−0.13, 0.03)*	*(−0.23, −0.02)*	*(−0.04, 0.17)*	*(−0.09, 0.11)*	*(−0.04, 0.15)*
Road Density	**0.12*****	**0.12*****	**0.14*****	**−0.09****	−0.04	−0.07
	(0.05, 0.20)	*(0.06, 0.19)*	*(0.05, 0.21)*	*(−0.18, 0.000)*	*(−0.12, 0.04)*	*(−0.15, 0.01)*
Watershed Group *(WG)*	−0.001	−0.09	0.01	−0.01	0.11	0.06
	(−0.13, 0.13)	*(−0.20, 0.03)*	*(−0.12, 0.16)*	*(−0.16, 0.14)*	*(−0.03, 0.24)*	*(−0.09, 0.21)*
WG * Goals/Objectives	−0.01	**0.13*****	−0.05	**−0.15****	−0.01	0.01
	(−0.12, 0.10)	*(0.02, 0.21)*	*(−0.16, 0.07)*	*(−0.28, −0.03)*	*(−0.15, 0.09)*	*(−0.12, 0.13)*
WG * Coordinator	−0.04	−0.03	**−0.12***	0.04	−0.10	0.02
	(−0.17, 0.09)	*(−0.14, 0.09)*	*(−0.27, 0.02)*	*(−0.11, 0.19)*	*(−0.23, 0.04)*	*(−0.11, 0.16)*
WG * Management	**−0.15*****	−0.05	0.09	**0.16****	0.04	0.05
	(−0.26, −0.03)	*(−0.15, 0.05)*	*(−0.04, 0.19)*	*(0.02, 0.29)*	*(−0.09, 0.15)*	*(−0.07, 0.18)*
WG * Stakeholders	0.08	0.06	0.05	0.06	−0.05	−0.03
	(−0.03, 0.19)	*(−0.03, 0.17)*	*(−0.06, 0.18)*	*(−0.07, 0.19)*	*(−0.17, 0.07)*	*(−0.15, 0.08)*
BIC	837.28	609.63	919.89	1,157.54	901.90	1,036.49

Note: ***p < .01; **p < .05; *p < .1
p-values refer to bootstrapped confidence intervals that do not contain 0
Models also include propensity scores and random effects for Year, HUC4, State, Ecoregion, and Stream Order.

phosphorus level and a 27% higher benthic index score ($sd = 0.74$). While the sign of the coefficient for the estimated effect on nitrogen level, riparian cover, and in-stream habitat complexity is in the hypothesized direction (reduced pollution, improved habitat condition), these effects are all insignificant; the estimated effect on turbidity is insignificant and not in the hypothesized direction. These results provide limited support for the hypothesis that collaborative groups with management responsibilities have a relatively greater impact on water quality. Table 2.4 shows that the difference in phosphorus level between the two group types is very similar to that of the predicted difference (in terms of both sign and significance) associated with a two standard deviation increase in county median income. Similarly, the difference associated with management groups with regards to benthic community health is similar in magnitude to the change associated with a two standard deviation increase in agricultural land usage.

H3: Stakeholder Representation

The number of stakeholder types in a group is considered as a continuous variable (standardized in the same way as the continuous covariates) in the "WG * Stakeholders" interaction term. While the parameter estimates predict a small increase in pollution (phosphorus, nitrogen, and turbidity) as the number of stakeholder types in a watershed group increase, none of these parameters are statistically significant. As the number of stakeholder types is mean-centered and standardized, this means that there is not a great deal of difference between an ''average group" involving local governmental representatives and four or five additional stakeholder types (the mean number of additional types is 4.3) and either a limited group involving only local governmental representatives (e.g., local city and county officials) or a diverse group involving all coded stakeholder types. This does not corroborate the theory that collaborative institutions are more made more effective by incorporating a broader range of perspectives (Burby 2003; Innes and Booher 1999; Margerum 2011; Wondolleck and Yaffee 2000), but it also does not evidence that broader involvement dilutes policy actions (Coglianese 1997; Coglianese 1999; Koontz et al. 2004).

H4: Group Formalization

Table 2.4 also tests two aspects of group formalization: (1) whether or not a group has a dedicated coordinator; and (2) the level of goal specification a group codifies. The "WG * Coordinator" interaction term represents the predicted difference between a group that has a coordinator and a group that does not. Only one coefficient is significant (turbidity, which is predicted to decrease by 33% [sd = 1.65]), but five of six are of a sign suggesting that groups with a coordinator achieve greater environmental gains. Given that it is fairly accepted that coordinators serve a valuable purpose, it is very interesting that these results do evidence a stronger, more substantive difference between groups that have a coordinator and those that do not. One potential source of variation not captured available in these data is the work level of the coordinator. In some cases, a group coordinator works on a part-time–or even largely volunteer–basis, or serves as coordinator as part of her broader job description at a government agency. Other groups have a coordinator who works full time in support of the group. Presumably, better data that are able to codify coordinator effort level would more carefully test the benefit of having a full time, dedicated coordinator.

The "WG * Goals/Objectives" term in Table 2.4 compare groups that have either itemized goals or objectives to the reference category, groups that publish only a mission statement. While several of the results are insignificant and of almost zero magnitude (phosphorus level and in-stream habitat condition), it is interesting to note that groups with itemized goals and objectives are associated with significant greater (40%, sd = 1.29) nitrogen levels and significantly reduced benthic community health (20%, sd = 0.74). While the lack of significance or a consistent pattern amongst the remaining outcome metrics makes it difficult to draw an overarching conclusion, these results suggest that itemizing purposes and goals does not necessarily make a group any more impactful.

DISCUSSION

The results of this analysis suggest that collaborative watershed groups achieve water quality and instream habitat gains. Watersheds with a collaborative group that has been active at least four years are estimated to have significantly lower levels of phosphorus content, nitrogen content, and turbidity. These watersheds are also estimated to have significantly greater in-stream habitat complexity (e.g., woody debris and aquatic plants). There is no significant estimated difference associated with benthic community health or with riparian cover (but the sign of each parameter is positive, meaning that all six ATE estimates are of a sign suggesting environmental improvement). As discussed above, considering the conceptual linkage between the actions of a collaborative watershed group and site-specific measurables lends explanatory context to these results. Water content metrics are obviously subject to proximate inputs, but they also more broadly reflect land use and environmental behavior throughout the watershed. For instance, if group actions help several farms mitigate fertilizer runoff, such activities would be reflected in a water quality sample taken downstream. Similarly, while in-stream habitat conditions are of course subject to onsite actions such as channelization, upstream land usage such as logging and development (or conversely restoration actions) can result in flow changes and floods that alter downstream habitat (Crispin et al. 1993; Wang et al. 1997).

Given that benthic community health is also shown to be related to upstream activities such as logging (Harding et al. 1998), it is interesting that no significant effect is found on this metric. This result is likely attributable to the fact that benthic communities continue to demonstrate the effects of land use actions decades after the activity has ceased (Zhang et al. 2009) (Harding et al. (1998) actually refer to stream biodiversity as "the ghost of past land use"). Accordingly, linking collaborative group presence to changes in benthic community health likely requires a longer time horizon. Lastly, inability to identify a link between collaborative groups and riparian cover is likely attributable to the fact that this metric least reflects aggregate watershed management and restoration actions and most reflects proximate actions by whichever entity owns that piece of property.

Comparing different types of groups, this analysis identifies a distinction between collaborative groups given management responsibilities (e.g., lead management entity for an Endangered Species Act recovery plan) and those tasked solely with coordination or planning. Two of the six parameters are significant and five of six are of the hypothesized sign, providing limited evidence that groups given management responsibilities stand apart as more effective. These results suggest that the additional costs of collaborative management (as opposed to coordination or planning), such as increased power sharing (Margerum 2011), time and resource commitment (Hill and Lynn 2003; Sabatier et al. 2005), and investment in the process (Wondolleck and Yaffee 2000; Margerum 2011; Sabatier et al. 2005), do result in increased environmental benefits as well.

Little differentiation is found with regards to stakeholder representation. A possible reason for the lack of significance associated with stakeholder representation is that this analysis focuses specifically on government-sponsored collaborative watershed groups. For instance, in the face of existing theory and evidence (Gregory et al. 2001; Burby 2003; Innes and Booher 1999; Margerum 2011; Wondolleck and Yaffee 2000; Dryzek 1997; Smith 2004b) it would seem unlikely that incorporating additional perspectives and interests into the policy process does not matter at all. However, it is very plausible that the role of government sponsored watershed groups and scope of their activities are fairly constrained, such that there is ultimately little variation in outcomes regardless of inputs. For instance, even watershed groups that engage in management actions do not have broad rulemaking and enforcement authority, but rather manage land use or similar issues (that are otherwise typically the responsibility of local governments [Koontz et al. 2004]). Watershed groups cannot pass new laws or implement a market-based water quality trading system. Thus, groups are necessarily limited by their legislative and regulatory environment.

Group impacts are shown to differ somewhat by group formalization, but not necessarily in the hypothesized direction. The results suggest that groups with itemized goals and objectives actually perform worse with regards to phosphorus level and benthic community health. It is not readily clear why this is the case, but given the prevailing wisdom that increased specification helps groups to resolve conflict and better assess efficacy (Margerum 2011; Mattessich et al. 2001; Susskind and Cruikshank 1987), the lack of association between goal specificity and improved environmental outcomes is noteworthy. It is plausible that regardless of goal specificity, the goals or objectives that end up being prioritized are those that closely dovetail with existing regulatory mandates, and thus the nominal goals of the group do not track closely with empirical actions. While this does not explain the significant results opposite of the hypothesized direction, it does perhaps explain why goals and objectives are not linked to environmental improvements. Finally, while the presence of a group coordinator or facilitator is linked only to a significant decrease in turbidity, as discussed above the lack of conclusive results in this regard is likely due to the fact that these data combine coordinators of various types and capacities. Resource limitations and a lack of data availability prevent ascertaining the effort level of a group coordinator.

A prominent theme that emerges from these results is the contrast between the significant ATE estimates associated directly with collaborative group presence and the inability of the group characteristics tested to "account" for the predicted difference between collaborative and non-collaborative watersheds. One potential reason is that the variables tested might not be the variables that drive group effectiveness (as measured by environmental impact). A notable omission, of course, are group funding levels. While one would presume that differential effectiveness associated with funding discrepancies is somewhat of a given, this relationship is worth testing to posit whether public agencies devoting funds to collaborative endeavors are getting any "bang" for their "buck." Of course, resource munificence alone cannot be the sole driver of group effectiveness. For instance, the findings above conclusively identify a significant benefit associated with having a group coordinator.

These results also highlight the essential role of qualitative research in understanding the role and function of collaborative governance. It is unlikely that large-N statistical analysis alone can definitively answer these questions. For instance, the extensive case studies conducted by Margerum (2011) speak to contextual variables and localized drivers of group efficacy that do not necessarily emerge in a larger-N cross-sectional analysis. Given that the uncertainty and complexity of environmental systems makes it difficult to parse effects of collaboration amidst other influences (Koontz and Thomas 2006; Rapp 2008), process tracing and the use of program logic models (Bickman 1987; Margerum 2011; McLaughlin and Jordan 1999) might serve as an evaluatory complement to the systematic analysis of outcomes conducted in this project. Most importantly, these works highlight the idiosyncratic nature of local environmental management; what works well in one context might not work well in another, and this is very difficult to tease out in a large-N statistical analysis.

CONCLUSION

It is easy to lose sight of the fact that collaborative governance requires the expenditure of time and effort by public actors, and that these resources could be applied elsewhere. In other words, collaboration is not just a "concept... [but potentially] a way of solving problems... and achieving results" (Margerum 2011, 306). However, much of what we currently know about the environmental impacts of government-supported collaborative institutions is based upon evidence from small-N case studies or studies that use subjective measures (e.g., stakeholder perceptions or quality of policy outputs) as proxies for environmental outcomes. Previous research (e.g., Koontz 2003; Lubell 2004a; Leach and Sabatier 2005; Hoornbeek et al. 2012; Ulibarri 2015) has shown that collaborative governance has a positive effect both on intermediate outputs and perceived policy or program effectiveness. This analysis uses a unique dataset and a rigorous analytical approach to build upon these works by conducting one of the first large-N statistical analyses that systematically tests the relationship between collaborative governance and environmental outcomes. Most importantly, the use of objective outcome data (water quality and habitat metrics) across a large geographic scale represents a major advancement.

Simply put, these results evidence that collaborative governance institutions (in this case, collaborative watershed groups) do improve ecological outcomes. It is also important to note that I find no indication across any of the six outcomes metrics that collaborative governance engenders worse environmental outcomes. This demonstrates that the lowest common denominator effect (Coglianese 1999) is less of a concern than might be thought, and that fears of collaboration leading to more talk and less action might be somewhat unfounded. Despite the rigorous matching approach employed, it remains possible that the ATE estimates are biased upwards due to unobserved factors that are positively related both to the presence of a group and to water quality outcomes. However, since matching is based upon research evidence regarding the factors that drive collaborative group formation (Lubell et al. 2002) it is likely that this approach successfully reduces omitted variable bias. Further, as discussed, these estimates represent a significant step forward even if some omitted variable bias remains (and given the infeasibility of randomized assignment in this context, it is unlikely that experimental data will be available in the future).

Due to the inconclusive findings with regards to group characteristics, this work does not shed a great deal of light on the question of how collaborative watershed management improves environmental outcomes. The lack of definitive results associated with group characteristics generally accepted as beneficial (e.g., presence of a coordinator, goal specificity) is particularly interesting. One suggestion emerging from the literature is that collaborative watershed governance in general represents a shift in focus towards non-point water quality problems that are not suitably addressed by state and Federal regulatory authorities (Hardy and Koontz 2008; Hoornbeek et al. 2012; Koontz et al. 2004). This shift in focus, however operationalized, might matter more than the specific details of the group itself at a macro level. Group characteristics are not likely irrelevant, but perhaps are more of a contextual issue rather than the basis for collaborative group impact.

Future research directions include the addition of a third wave of data from the 2013-2014 NRSA (to be released in 2016 following laboratory analysis of samples), as well as additional data concerning group budgets, procedures, and activities. Also, while I control for land and resource use in this analysis, future research will take advantage of larger samples to allow for a better understanding of how collaborative groups can be more or less effective (in terms of producing desired outcomes) in specific contexts. This ongoing work is important, as perhaps the central takeaway of this analysis is that we as policy scholars and practitioners need to think more deeply about why we believe that collaborative groups are an effective tool for achieving public policy goals.

Chapter 3

ANALYZING POLICY NETWORKS USING VALUED EXPONENTIAL RANDOM GRAPH MODELS (ERGMS): TO WHAT EXTENT DO GOVERNMENT-SPONSORED COLLABORATIVE GROUPS ENHANCE ORGANIZATIONAL NETWORKS?

INTRODUCTION

This paper examines how government-sponsored collaborative environmental management groups influence the structure of inter-organizational networks. Collaborative groups continue to grow in popularity as a tool for increasing coordination amongst network actors (Margerum 2011). In theory, government-sponsored groups lower the transaction costs organizations face with regards to forming and maintaining inter-organizational ties by subsidizing a degree of these costs. Thus, participants in groups which facilitate 'principled engagement' and increase the 'capacity for joint action' (Emerson et al. 2012) are much more likely to engage directly in consultation, planning, or policy implementation with one another as well (Scott and Thomas 2015).

In practice, when viewed at a larger scale it is unclear whether collaborative groups foster a net increase in cooperation and coordination amongst organizations (i.e., more and stronger inter-organizational ties) or whether they simply engender a different pattern of ties with no net change in overall collaborative behavior (Lubell et al. 2010; Lubell and Lippert 2011). Recent literature building upon the "Ecology of Games" framework (Berardo and Scholz 2010; Lubell et al. 2010; Lubell and Lippert 2011; Lubell et al. 2011; McAllister et al. 2014; Smaldino and Lubell 2014; Niles and Lubell 2012) shows that "any particular collaborative process may have positive or negative feedbacks on the performance of other policy decisions" (Gerlak et al. 2012, 424). Accordingly, in light of the complex array of existing institutions (collaborative groups and other venues) organizations already operate within (see Lubell 2013; Lubell et al. 2011), the standard rationale for government initiation and support of new, additional collaborative groups deserves greater scrutiny.

This analysis is a companion piece to (Scott and Thomas 2015), which also addresses government-sponsored collaborative management groups employed as network interventions. Both analyses leverage a unique dataset concerning a large scale regional network of organizations involved 57 different collaborative management groups involved in ecosystem restoration and recovery. However, (Scott and Thomas 2015) seek to operationalize and test the framework for collaborative governance developed by Emerson et al. (2012), specifically examining the mechanisms of principled engagement and increased capacity for joint action that are theorized to be drivers of collaborative behavior. (Scott and Thomas 2015) test these mechanisms using data concerning participant beliefs about group actions and outcomes, showing a positive and significant relationship between the extent to which a group is reported to facilitated principled engagement with other organizations and the prevalence of coordination and cooperation amongst group members. In contrast, this paper uses data concerning participation in group activities to test whether participation in a collaborative group is associated with a corresponding increase in information sharing, planning, and policy or program implementation with other organizations. Building upon this basic question, I use participation data from multiple groups to test whether participation in several groups diminishes the predicted marginal impact of participation, and further use data concerning whether a network tie existed prior to group membership to compare the extent to which group participation serves to strengthen existing ties versus foster new ones. The overarching goal of this paper is to examine whether there is a prima facie case supporting the use of collaborative management groups as a tool for increasing network coordination, particularly for complex policy networks in which organizations already have many network ties or are members of multiple collaborative groups. I also make a methodological contribution to the policy literature by providing an empirical example of statistical network analysis using values to reflect network ties of differing strengths instead of modeling a binary metric reflecting tie presence or absence.

In the section to follow, I provide the background and theoretical rationale for this research. Particularly, I define and distinguish between key terms such as networks and collaborative groups and embed my hypotheses within the extant literature. I then describe the research design and the recently developed method of valued-tie exponential random graph models (ERGMs) (Krivitsky 2012; Krivitsky and Butts 2013; Cranmer and Desmarais 2011; Desmarais and Cranmer 2012b; Wyatt et al. 2009, 2010) I use to test my hypotheses. After describing the survey instrument and data collection process, I present model results.

Finally, I conclude with a discussion of findings and their broader implications, both for empirical management and for the ongoing collaborative governance literature.

BACKGROUND

The diffuse nature of networks[1] might at first seem to be completely at odds with the notion of direct government influence and strategic public policy interventions; however, the concept of governance[2] implies that government cannot act autonomously. Public agents must instead engage in multi-actor processes in which it is only one of many relevant actors, as its "core purposes can only hope to be realized in such settings" (Bressers 2009, 130). In other words, public agencies still face the same problems of service delivery, but must employ different means to achieve their desired ends.

Within a network governance context, public policy makers often fulfill a role as network manager (Klijn and Koppenjan 2000); instead of carrying out tasks directly, policy makers attempt to address collective action dilemmas indirectly by changing network rules and influencing network relationships (Klijn and Koppenjan 2006). Interorganizational (or

---

[1] Broadly, networks are simply defined as "collections of actors who pursue repeated, enduring exchange relations with one another and, at the same time, lack a legitimate organizational authority to arbitrate and resolve disputes that may arise during the exchange" (Podolny and Page 1998, 59). By expanding the concept of exchange, networks can be broadly defined as "sets of individuals [or organizations] bound by communication, relationships, positions, or interest area" (Margerum 2011, 33). This definition can encompass social networks (interpersonal relationships, Putnam 2000), inter-organizational networks (structures and processes in which organizations interact, Alexander 1993), and political networks (power positions and configurations, Knoke 1990).

[2] "Governance" refers to a process in which public actors make policies, deliver services, or implement policies within a network (or networks) of actors (Frederickson 2005; Rhodes 1997; Torfing and Sorensen 2007). Governance is characterized by a high degree of interdependency amongst actors and a complex decision-making process (Klijn et al. 2010). Bressers (2009, 125) poses that "governance" is an enlargement of the concept of public policy (also Bressers and Kuks 2003). Thus, in keeping with the findings of Ostrom et al. (1961), governance is not really a new state of affairs, but rather a basis for scientific variables that can be used for empirical studies.

inter-stakeholder) collaborative groups are one of the most prominent (and well-documented) mechanisms by which environmental policy makers attempt to alter the structure and function of an organizational network (see Ansell and Gash 2008; Emerson et al. 2012; Margerum 2011, for recent discussions). Such groups represent "a governing arrangement where one or more public agencies directly engage non-state stakeholders in a collective decision-making process that is formal, consensus-oriented, and deliberative and that aims to make or implement public policy or manage public programs or assets" (Ansell and Gash 2008, 544) (see also Emerson et al. 2012; Imperial 2005; Margerum 2011). Even though collaborative effort on the part of a public agency might be required by legislative mandate, a collaborative group represents a voluntary association of legally autonomous public, private, or non-profit organizations (Ansell and Gash 2008; Schneider et al. 2003; Lubell et al. 2002); this serves to distinguish a collaborative group from an organization, defined in this analysis as a legally autonomous entity with a formal hierarchical structure. Thus, a "collaborative management group" refers to a management entity comprised of several independent organizations that uses a deliberative structure and seeks to encompass relevant stakeholders (Ansell and Gash 2008).[3] To avoid confusion, going forward this paper uses “coordination” and “cooperation” in reference to network ties between two organizations and reserves the term “collaboration” as a reference specifically to collaborative management groups.

RATIONALE

If coordination and cooperation with other organizations is potentially advantageous, why do organizations not always elect to do so? Assuming that network organizations act rationally in pursuit of their interests, one might surmise then that the existing structure and function of an organizational network reflects the landscape of costs and benefits associated with

[3]Collaboration of course can also occur outside the auspices of a formal "collaborative group." Margerum (2011, p. 6) defines collaboration as “an approach to solving complex problems in which a diverse group of autonomous stakeholders deliberates to build consensus and develop networks for translating consensus into results.”

network relationships. The benefits of inter-organizational coordination and cooperation include heightened information access, issue understanding, conflict reduction, and implementation support (Moreland et al. 1993; Gigone and Hastie 1993; Cragan and Wright 1990; Hill and Lynn 2003; Susskind et al. 1999; Sabatier et al. 2005). Of course, collaboration is not always beneficial for organizations. Further, even in cases where collaboration might be advantageous, the time and resources required to initiate and maintain network relations can outweigh any potential benefits. Highly practical constraints such as travel time (Thomas 2003) or there being too many different meetings and activities for managers to attend each one (Margerum 2011) constrain inter-organizational collaboration. Researchers also often discuss more intangible transaction costs associated with network relations, such as norms of reciprocity (Putnam 2000) or shared beliefs and preferences (Schneider et al. 2003; Sabatier and Jenkins-Smith 1993), which also incentivize (or disincentivize) collaborative efforts.

Though the behavior of individual organizations ultimately depends on the particular incentives faced and motivations held by each organization, from a policy and management perspective existing policy network structures can prove suboptimal (i.e., ‘’network failure” (Schrank and Whitford 2011) or ‘’system failure” (Carlsson and Jacobsson 1997)) in the same way that existing market structures can lead to negative outcomes (for instance, classic market failures such as overfishing) (Weimer and Vining 2010). In particular, Schneider et al. (2003) find that often network relations are under supplied because ’the costs of creating and maintaining networks [network ties in the context of this analysis] are high and the benefits gained by the policy community... are not reflected in the incentives of individual stakeholders" (Schneider et al. 2003, 144). By producing outputs such as meetings and providing administrative support for joint activities, a collaborative group can potentially address this under supply problem by altering the transaction cost landscape that each organization faces. The sponsoring agency is essentially hoping that subsidizing a degree of the transaction costs related to inter-organizational networking will engender increased collaboration. This type of intervention can create new network content (Koppenjan and Klijn 2004), guide network interactions (Kickert et al. 1997; Mandell 1990), and further inter-actor trust through facilitated interactions (Klijn et al. 2010). For instance, Schneider et al. (2003) find evidence that federal programs such as the National Estuary Program (which supports collaborative work with stakeholders) can help overcome "second-level" collective action problems (i.e., those related to coordination and cooperation) by providing funding, encouraging broader participation, establishing a focal policy arena, and increasing legitimacy amongst network organizations.

Nonetheless, altering the incentive structures organizations face does not necessarily alter the motivations and goals of said organizations (but might in the long run, as I discuss later). Thus, the extent to which government initiation and support of collaborative management groups increases actual inter-organizational cooperation and coordination remains an open question. The basic hypothesis (H1) for this analysis is that participation in a collaborative group is associated with an increase in direct coordination and cooperation with other organizations.

H1: Participation in a collaborative group is associated with an increase in direct coordination and cooperation with other organizations

While a positive association is expected, the relative strength–or tenuousness–of the relationship between collaborative group formation and increases in actual coordination and cooperation amongst individual organizations, which speaks to the impact that a government-sponsored network intervention can achieve, has not been systematically evaluated. Further, I extend this basic hypothesis to examine conditions that are more interesting in theory and practice. First, collaborative processes can have both positive and negative feedbacks within a policy network (Gerlak et al. 2012), The reality that network interventions can have both positive and negative impacts within a complex policy subsystem is one of the overarching messages of the growing Ecology of Games literature (Berardo and Scholz 2010; Lubell et al. 2010; Lubell et al. 2011; McAllister et al. 2014; Smaldino and Lubell 2014; Niles and Lubell 2012). New collaborative groups do not operate in a vacuum. Especially in highly institutionalized settings such as the Puget Sound case I examine in this paper, organizations not only already possess various types of ties to other network organizations, but moreover are typically involved in a complex array of existing collaborative institutions as well (see Lubell 2013; Lubell et al. 2011). The Ecology of Games framework (developed by Long (1958) and reintroduced by Lubell et al. (2010)) emphasizes organizational constraints, namely that organizations have a finite capacity for networking and interaction with other organizations. While this capacity might change in the long run as a function of changing organizational goals, motivations, and beliefs (e.g., Bingham and O'Leary 2008; Innes and Booher 2010; Leach and Sabatier 2005; Lubell 2005), in the short run organizational "demand" for collaboration is essentially fixed. Thus, a new collaborative group might change the existing pattern of network ties without engendering a net increase in overall coordination within a complex policy network (i.e., instead of keeping old ties and forming new ones, organizations drop old ties and form new ones, resulting in a zero-sum change). This analysis does not specifically examine organizational capacity, since my focus is on the general impacts of a government-sponsored network intervention. Nonetheless the implications of the Ecology of Games framework and related literature directly inform Hypothesis 2 (H2):

H2: Participation in multiple collaborative groups diminishes the association between group participation and direct coordination and cooperation with other organizations

Further, network ties are often conceptually dichotomized as either "bridging" or "bonding" ties (Berardo and Scholz 2010; Berardo et al. 2014); bridging ties are "weak ties" (Granovetter 1985) that maximize information flow and bonding ties are "strong ties" characterized by high levels of reciprocity and trust. The "information sharing" ties measured in my data are an example of a bridging tie, while a "joint implementation" tie more closely equates to a bonding tie (I assume that coordinated planning ties occupy a middle ground). Since respondents are asked to report whether a given tie existed prior to their involvement in a PSP-sponsored group (these data are discussed in greater detail below), this indicator can be used to compare the extent to which participation in a collaborative group serves to strengthen existing relationships versus foster new ones. Scholz et al. (2008) find that lack of information about potential partners significantly constrains inter-organizational coordination and cooperation in policy networks. Government-supported collaborative groups can reduce these search costs. For organizations that have an existing tie, however, search costs are presumably not a major barrier to having a stronger relationship; instead, the extent to which organizational goals and motivations do not overlap–or at least are not compatible–is likely a more prominent factor in explaining why two organizations do not have a stronger network tie. My analysis considers short-term network changes; while participation in a collaborative group can alter organizational goals and beliefs in the long run (e.g., Bingham and O'Leary 2008; Leach and Sabatier 2005; Lubell 2005), in the short run group participation likely does more to reduce search costs (i.e.., increase awareness of other organizations). Since lack of awareness better explains why two organizations have no network tie whatsoever than it explains why two organizations do not have a stronger network tie, my third hypothesis (H3) is that the predicted change in tie strength for any two collaborative group participants will be diminished if two organizations have a pre-existing tie:

H3: A pre-existing network tie diminishes the association between group participation and predicted tie strength

The following section describes the modeling approach I used to test these hypotheses.

MODEL

While individual actor attributes affect network tie formation (Handcock et al. 2014a), the very presence or absence of other ties also affects whether other network ties are initiated, maintained, or destroyed (Lubell et al. 2012). This interdependence means that these data violate the standard statistical assumption that observations are independent of one another (Robins et al. 2012). Failing to account for this biases estimates (Kolaczyk 2009; Krackhardt 1988). Thus, I use exponential-family random graph models (ERGMs), which explicitly model tie interdependence (Lubell et al. 2012), to address these hypotheses. The application of ERGMs to policy research is well established in the literature (Feiock et al. 2010; Henry

2011; Lubell et al. 2012). The basic premise of ERGMs is that the observed network (the survey results) constitute one sample from a distribution of network graphs; I can simulate a distribution of similar graphs (i.e., ones that on average have the same number of organizations, inter-organizational ties, and other network structures as does the observed network) and then compare the observed network to this distribution. If the observed number of a specific network structure, for instance a triangle in which $y_{ij} = 1$, $y_{jk} = 1$, and $y_{ki} = 1$, is very high relative to the typical number of said structure present in the distribution of randomly simulated networks, then this would indicate that there is significant "transitivity" (i.e., “a friend of my friend is my friend”) in the observed networks.

The primary consequence of assuming any type of dependence[4] amongst observations is that each and every tie variable must be modeled conditionally based upon all other ties observed in the network (Lusher et al. 2013). Because there is an extremely large number of possible network configurations, it is not feasible to analyze all possible graphs. Instead, the *statnet* R package (Goodreau et al. 2008) implements a Markov chain Monte Carlo (MCMC) procedure that estimates model parameters using maximum likelihood estimation (Handcock 2003). The MCMC routine proposes a single change to the network (i.e., altering one potential tie between two nodes) and compares the probabilities of the previous and proposed networks.[5] When the probability of the proposed network is greater than the current network, the MCMC routine chooses the proposed network and then repeats the same process; when the probability of the proposed network is less than the current network, the MCMC routine only chooses the proposed network a given percentage of the time (Lusher

[4] Transitivity is just one type of dependence. Others include “reciprocity,” the tendency of ties to be reciprocated, and “popularity,” the tendency for popular network members to gain more network ties by virtue of this popularity (see Lusher et al. 2013).

[5] Lusher et al. (2013, p. 9) define the probability of network G as $P_\theta(G) = c * \exp^{\theta_1 z_1(G) + \theta_1 z_1(G) + \ldots + \theta_p z_p(G)}$, which means that the probability of a graph is given by the exponentiated sum of network statistics (z_p) each weighted by the corresponding coefficient values θ_p. Network statistics refer to the count of specific network structures in G, such as triangles.

et al. 2013). By repeating this process over and over again, the MCMC sampling procedure converges on a stationary distribution of network graphs from which sample graphs can be drawn to generate a basis against which to compare the observed data.

Whereas (Scott and Thomas 2015) and most other statistical network analyses model binary network ties (either present or absent), this paper makes a methodological contribution by providing an empirical example of the valued ERGM method. “Valued networks,” or networks with non-binary ties, are still very much an exploratory area within statistical network analysis. Applications of valued ERGMs to non-trivial empirical datasets are rare in the policy literature (Desmarais and Cranmer 2012b; Krivitsky 2012). The binary ERGM approach (as used in Scott and Thomas 2015 requires modeling each type of network tie as constituting a distinct network (e.g., a “joint implementation” network and a “coordinated planning” network). Modeling valued edges enables a more nuanced approach, specifically the ability to model one overall network consisting of inter-organizational relationships of differing intensity (e.g., consultation vs. coordinated activity). This makes a great deal of empirical sense, since the different inter-organizational ties modeled in this paper (consultation, planning, and implementation) reflect relationships of differing intensity rather any sort of firm categorical boundaries. Valued ties are also helpful from an analytical perspective, as it allows for distinction between network structures consisting of more and less intensive ties. For instance, if Organization A and Organization C both engage in policy implementation with Organization B, we might expect the fact that A and C each have a fairly intensive relationship with B to increase the likelihood of A and C themselves sharing a direct tie relative to if A and C were to each have a less intensive relationship with B.

I code inter-organizational ties ordinally by the level of interaction reported using a truncated, discrete geometric-reference ERGM. A consultative tie is coded as 1, a planning tie as 2, and a implementation tie as 3 (all potential ties that are not observed are coded as 0). From Krivitsky (2012), the valued ERGM (using discrete variables) is specified as follows:[6]

[6]Note that the *ergm.count* extension to the *statnet* package is designed primarily for modeling count network

$$Pr_{h,g}(Y = y_i; \theta) = h(y)\frac{\exp(\theta^T g(y))}{c_{h,g}(\theta)}, y \in Y \tag{3.1}$$

with a normalizing constant represented by c and the reference distribution represented by $h(y)$ (Krivitsky and Butts 2013). The normalizing constant simply ensures that all probabilities sum to one. The reference distribution specifies the the model prior to the addition of any model terms; thus it represents the observed baseline distribution of network ties. For the discrete uniform-reference model, $h(y) = 1$ (Krivitsky 2012). Additional terms can then be added to the model to account for structural characteristics of the network, or for exogenous attributes of the nodes (organizations) or edges (collaborative ties). For instance, adding a basic summation term for the value of all observed network relations produces the following specification (Krivitsky and Butts 2013):

$$Pr_{h,g}(Y = y; \theta) \propto h(y) \exp(\theta \sum_{(i,j) \in Y} y_{ij}) \tag{3.2}$$

which is similar to an intercept-only linear regression model. The summation term acts as an intercept because it makes the predicted value of y_{ij} equal to the average observed density of the network (the total value of all ties divided by the total number of possible ties).

While the premise of using collaborative groups to increase inter-organizational coordination and cooperation seems intuitive, empirically there are several issues that make supporting this causal claim with rather difficult. Ideally, one would use either extensive longitudinal

data (e.g., the number of communications sent from one node to another. However, since the terms included in the *ergm.count* package used in my model (described in the specification of the baseline model below) do not place meaning on the sums or differences of network ties, but rather simply make comparisons between tie values, it is feasible to apply these tools to rank-order categorical ties (personal conversation with Dr. Pavel Krivitsky). For instance, the Mutual parameter (explained in detail below) models reciprocity using the minimum observed tie value between two nodes i and I. This metric behaves in precisely the same way whether for count data or ordinal categorical data, since $Min(y_{i,j} = 2, y_{j,i} = 1) = 1$ whether the tie values 1 and 2 refer to counts or an ordinal relationship category.

data or a randomized experiment to estimate the causal effect. However, time series data are difficult and costly to maintain. At this point there are not sufficient resources to facilitate long-term longitudinal data collection, and no comparable data currently exists. Even with sufficient resources to implement repeated surveys to such a large organizational network, feedback received from participants in the current survey indicates that many participants would be unwilling or unable to participate in a more extensive study. An experimental evaluation would also likely be unfeasible. Conducting a randomized controlled trial is simply not realistic for such a large scale policy network; it is difficult to envision successfully implementing a randomized experiment in which some organizations in a region are invited to participate in a management or decision-making process while others are not. Further, the dependent variable(s) in a network analysis are ties between units (organizations in this case), not characteristics of individual units. This quite obviously violates the Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1986; Pearl 2000), as the network behavior of each organization (whether in the treatment or the control group) affects that of every other organization.

The model specified above is applied to observational data (described in detail below). Thus, any estimates of the association between collaborative group participation and increased consultation, planning, and implementation with other organizations are potentially biased upwards due to confounding. For instance, it is possible that unobserved characteristics, namely the motivation to form ties with other organizations, are positively correlated with both collaborative group participation (the "treatment" variable) and the formation of network ties (the outcome variable) (i.e., confounding). This would bias estimates upwards because organizations who participate in groups would also be more likely to report inter-organizational ties in general regardless of their participation in a group. The inappropriateness of regression models for network data (Kolaczyk 2009; Krackhardt 1988) precludes the use of traditional regression-based methods, such as propensity score matching, used to control for omitted variable bias and estimate causal effects with observational data. Thus, I employ a deductive, stepwise modeling approach to address potential confounding, as well as the possibility of reverse causality (that organizations tied to one another join the same groups, rather than vice-versa). Further, I also account for the practical consideration of bandwidth limitations (that ties switch according to group membership without a net increase overall) in light of the large number of institutional venues that most organizations have access to (Berardo and Scholz 2010; Lubell et al. 2010; Lubell et al. 2011; McAllister et al. 2014; Smaldino and Lubell 2014; Niles and Lubell 2012; Gerlak et al. 2012). As these data concern a large number of organizations and collaborative groups active within the region, I am able to control for existing group membership patterns and thus avoid overstating the gains attributable to the focal network intervention.

Fitting a series of ERGMs with different specifications for group participation and past network ties allows me to triangulate a potential effect. While this method cannot eliminate potential bias, the estimates provided herein present prima facie evidence for the use of collaborative groups to enhance inter-organizational networks. Given the lack of more rigorous experimental designs, these data are the best available evidence for policymakers to date. I further discuss my empirical approach within the Analysis and Results sections. In the interim, the following section describes the case selection and data gathering process.

CASE SELECTION AND DATA

The analyzed case is that of the Puget Sound Partnership (PSP), an agency in Washington state created in 2007 that is confusingly named because it is a public agency, not an inter-organizational partnership. The PSP is tasked with improving coordination and enhancing collaboration in Puget Sound ecosystem restoration and recovery efforts. As part of its efforts the PSP has initiated or significantly funded 34 different local and regional collaborative groups. What makes this particularly interesting is that the Puget Sound region already had numerous active collaborative groups prior to 2007 in which many organizations were involved (see Scott and Thomas 2015). Thus, PSP-initiated groups exist within a broader context of organizational ties and other collaborative groups as well. This is likely emblematic of the type of environment encountered throughout the US and other highly institutionalized settings, where organizations have many opportunities for networking and must make trade-offs (e.g., which meeting to attend or which organizations to work with) due to time and resource constraints (Lubell et al. 2010). Thus, I also use data concerning membership in 23 local and regional collaborative groups that predate the PSP's network intervention in order to account for existing group membership and participation. In particular, I am interested in whether this mitigates the predicted impact of membership and participation in a PSP group. In total, the sample frame encompasses the membership of 57 local and regional collaborative groups, which provides a comprehensive picture of the overall network inside and outside the sphere of PSP sponsorship.

The survey instrument (presented in Appendix D) uses a "hybrid name generator" technique (Henry et al. 2012; Lubell et al. 2011) in which each respondent is asked to list up to five organizations with which they regularly engage in: (1) joint projects or program implementation (given that many organizations are consulting, funding, or administrative support bodies, this includes activities such as permitting assistance); (2) coordinated planning or strategy development; and (3) informal consultation (e.g., information sharing) (Scott and Thomas 2015. Thus, each respondent could potentially list up to 15 organizations in total, 5 for each type of collaborative activity. The total number of responses for each category are limited to emphasize regular, substantive organizational ties (and make survey completion tractable). The survey instrument also produced data on whether or not each reported tie existed prior to the PSP's intervention and each respondents' level of participation in collaborative groups.

As some group coordinators are unwilling to share their group email lists, the survey response rate must be estimated using published membership rosters. I estimate that the 57 collaborative groups sampled encompass 1600 total members (see also Scott and Thomas 2015). However, many organizations maintain membership in multiple collaborative groups; thus, 902 unique individuals are identified. Further, there are 100 group positions without any published name attached (e.g., groups rosters list only positions, such as "municipal representative," and no name or organization); to be conservative, I assume that each of these unnamed positions correspond to a unique individual, thus producing a total population estimate of 1002. Of these 1002 individuals, 498 accessed the survey instrument. Because 63 respondents did not identify the organization they represent, and 35 did not complete the portion of the survey in which they were asked to identify network ties, the sample used for this analysis comprises 400 individuals (a response rate of around 40%). These 400 individuals represent 221 unique organizations.[7]

ANALYSIS AND RESULTS

Model Fit

The basic premise of the ERGM approach that networks are not static and that an observed network configuration is a realization of an ongoing, dynamic process (Lusher et al. 2013). Accordingly, the observed structure is just one realization from amongst all of the possible structures that could be observed (e.g., imagine the observed network but with one more inter-organizational tie). An ERGM facilitates statistical inference on the processes that drive network structure by considering the set of all possible network configurations, and comparing the observed configuration (the network as revealed by the data) to this theoretical set (placing extra weight on possible configurations that bear increased similarity to the

---

[7]These 221 organizations are largely representative of the population of organizations that participate in environmental restoration and recovery efforts in the Puget Sound region; as expected, many organizations in the sample are themselves public entities. Out of the 221 unique organizations represented, 93 are local government (municipalities, county governments, special districts, and local commissions), 23 are Federal or state agencies or regional governance commissions, and 12 are tribal governments. Local or regional advocacy and outreach groups constitute another 48 unique organizations in the sample. The remainder are universities and research organizations (16), parks and reserves (7), natural resource extraction firms (4), large non-governmental organizations (12), and consulting firms (6). What is important to remember is that representativeness in this case does not refer to role organizations necessarily play in environmental outcomes (or else organizations such as natural resource extraction firms would seem to be highly underrepresented), but rather the degree to which different types of organizations are involved in ongoing environmental management and restoration efforts in the region.

observed configuration). Whereas a standard regression model works–in theory–even in the absence of control variables, hypothesis testing using an exogenous (i.e., non-structural) variable in an ERGM model first requires inductively fitting a baseline model that adequately describes the observed network; the variable of interest can then be added to the baseline model to see if it also accounts for a significant portion of the observed variance (Kolaczyk 2009; Lusher et al. 2013). In other words, generating a baseline ERGM is a "trial-and-error process" (Lusher et al. 2013, 184) that requires a series of stepwise modeling fittings to develop an optimal baseline model to which terms testing the independent variable(s) of interest can then be added.

Snijders et al. (2006) and Lusher et al. (2013) provide a set of "structural" parameters that are recommended as a starting place for fitting ERGMs Structural parameters represent endogenous effects that result due to self-organization within the network independent of actor attributes or characteristics (Lusher et al. 2013). One might think of these structural parameters as capturing the "basic processes" of networking, such as the tendency for popular organizations to attract more ties simply by virtue of being popular, or for the propensity of group-like structures such as triangles to form amongst related organizations (Lusher et al. 2013). While specific types of drivers are of differing importance in different networks, one can model a network by starting with the recommended set of general structural terms (as well as any exogenous variables that are theorized to be relevant) and then removing terms that are not significant or that do not contribute to model fit (Snijders et al. 2006). Accordingly, the model presented in Table 3.1 shows the "best-fit" baseline model produced via this process (without the variables of interest).

ERGMs are fit using a Markov-chain Monte Carlo (MCMC) procedure in which a possible network configuration is generated and then successively updated (e.g., change the value of one tie); the model then samples from the MCMC "chain" (the ongoing sequence of perturbed configurations) to generate a distribution of configurations against which the observed network is compared. The primary challenge in developing a baseline model is ensuring that an ERGM does not become degenerate, a condition which occurs when the

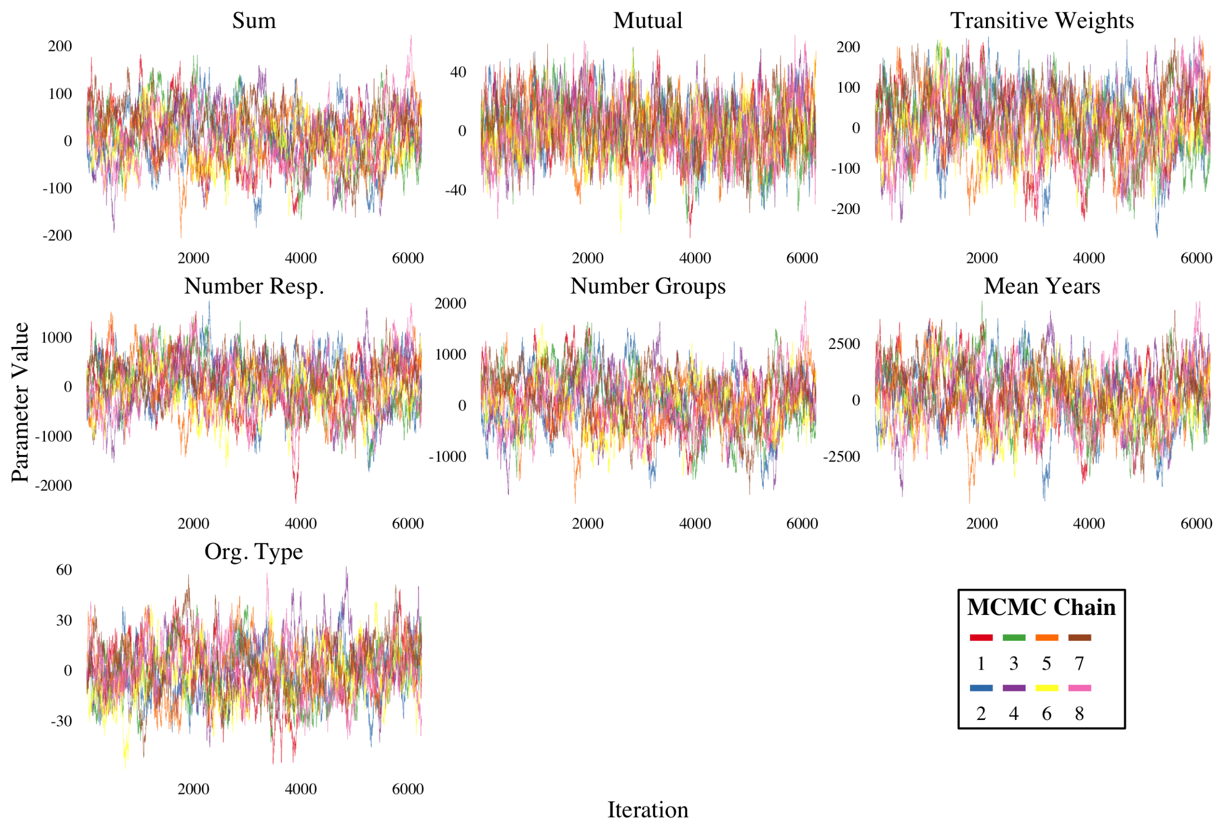
MCMC fitting processes cascades towards a completely empty (no ties) or a completely full (every possible tie) configuration and thus places most or all of the weighting on these unrealistic possibilities (Handcock 2003; Kolaczyk 2009). This basic explanation as to why this happens is that as the iterative fitting process changes one tie, it can simultaneously affect the likelihood associated with multiple network structures (such as triangles, multi-tie paths, etc.), and if this continually happens the model will head towards the empty or full network rather than fully exploring the parameter space. Of particular issue in non-binary (i.e., valued-edge) ERGMs is that the parameter sample space is unbounded depending on the configuration of parameters used, and thus the simulated parameter values can simply keep decreasing or increasing across iterations (Krivitsky and Butts 2013). By viewing trace plots of the MCMC process, one can verify whether the models "mix" adequately or whether they stray outside of the parameter space (Krivitsky 2012).

Figure 3.1 shows traceplots for each parameter across iterations of the MCMC chains.[8] I discuss the interpretation of each parameter below. Each line in Figure 3.1 shows how the maximum likelihood estimate for each parameter changed as each chain progressed. The seemingly random, haphazard path of the chains for each parameter evidence that the model is searching throughout the parameter space and not converging to the full or empty network. This shows that the MCMC process is sampling from a stable distribution and is not degenerate.

Further, Figure 3.2 shows smoothed parameter density plots from each MCMC run. As shown, each chain produces a unimodal distribution and generally speaking, the 8 chains produce similar distributions for each parameter. The x-axis values in Figure 3.2 represent

[8]The models were fit via parallel processing, thus 8 chains were used for each model. Each chain first proceeds through a "burn-in" period (1875 samples are taken and thrown out), with the presumption being that following the burn-in period subsequent draws are from a stable distribution. Each chain then takes 6250 samples, with a thinning interval of 1500 (i.e., the chain proceeds through 1500 perturbations before recording a new sample–this ensures that samples are not all drawn from the same area of the distribution, as would be the case if the chain simply recorded subsequent draws).

Figure 3.1: *Traceplots of MCMC Chains*



the difference between the observed network statistics and the network statistics for each iteration of the MCMC routine. One can verify model convergence by seeing that the density plot for each parameter (normalized by the observed statistic) has a mean near zero and is normally distributed; in other words, the model has converged to the observed network. In this case, one observes that most distributions are centered closely to zero and appear to be roughly normally distributed.[9] Most importantly, the overall mean in each case appears to

[9]Ideally, the distributions for each of the 8 chains would be completely congruent; in practice, however, this is difficult to achieve for a large, non-trivial network. Each model shown in this paper takes approximately 30 hours to fit on a high-performance Linux server cluster. It is expected that were each chain run for a

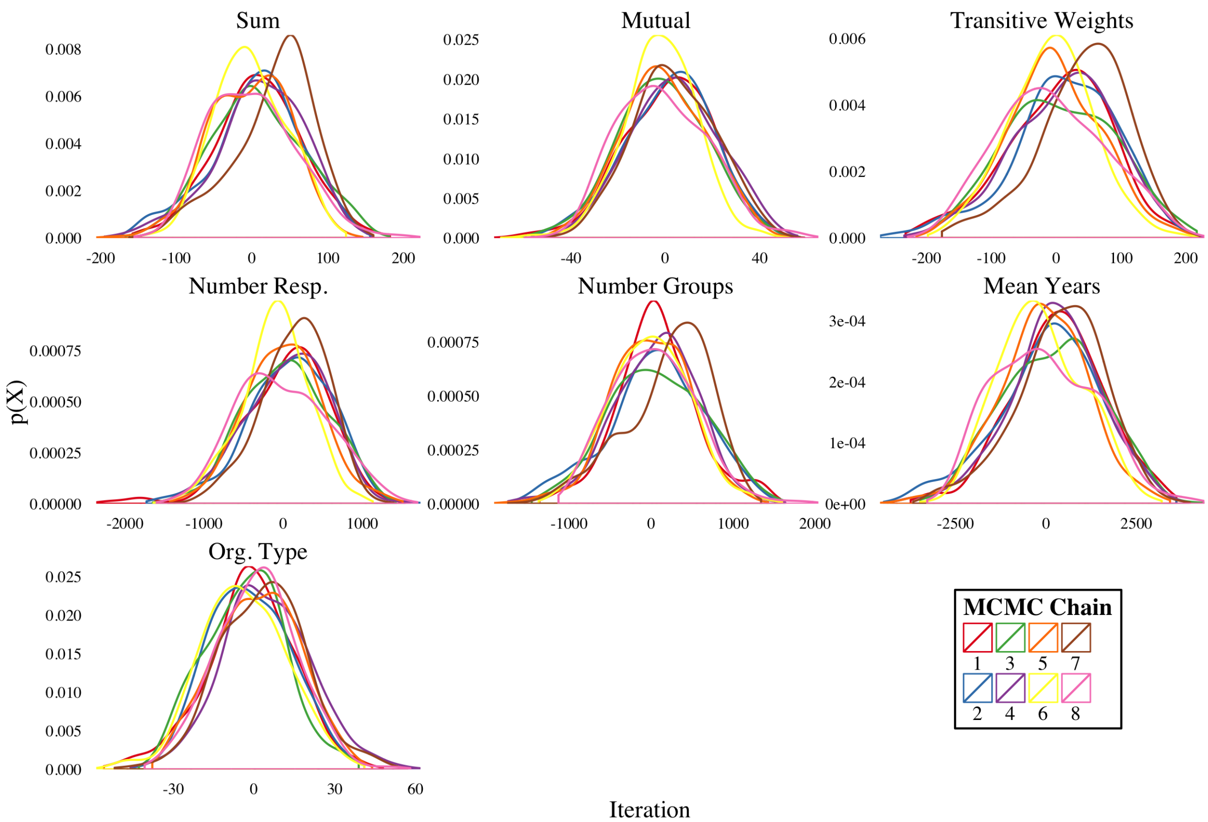
be centered near zero as well, which demonstrates that the fitted model nicely approximates the observed data.

Control Variables

Table 3.1 presents the restricted model with only structural and actor-related control variables. It is helpful to interpret these parameters because it helps reveal how the model actually works. First, the model includes three "structural" parameters, "Sum," "Mutual," and "Transitive Weights." Each coefficient is interpreted as an additive change on the natural log of the expected tie value (which can thus be exponentiated to produce a more easily interpretable multiplicative coefficient). The Sum parameter is a measure of network density, as it reflects the total sum of tie values observed in the network. For instance, a network with more joint implementation ties (each having a value of 3) will have a higher density than a network with more informal consultation ties (each having a value of 1). The strongly negative Sum parameters in Table 3.1 show the network is very sparse. Specifically, the parameter can be exponentiated to show that the expected tie value between two randomly selected organizations is 0.015 ($exp^{-4.18} = 0.015$) (Krivitsky 2012). Whereas in a binary ERGM this value would refer to an odds ratio ($p(y_{ij} = 1)/p(y_{ij} = 0)$), for a valued ERGM this statistic refers to the expected value of the tie ($E(y_{ij}) = 0.015$). It is expected that the expected value is quite low in this case, simply because most organizations in the network do not have any type of tie to most other organizations (with 221 organizations there are

sufficient amount of samples, the distributions would converge. This assumption was tested by comparing the model fitting results presented in this paper to those of models with a smaller MCMC sample size; as the number of samples per chain increases from 2250 to 6250 (i.e., from 10,000 total samples to 50,000 total samples), the parameter distributions for each chain become more congruent. Further, it is important to note that many published ERGMs do not fit MCMC chains in parallel, and instead simply examine whether the single distribution for each parameter is normally distributed and centered on zero. Even though the distributions in Figure 3.2 do not completely overlap, the fact that they show relatively consistent shapes and centerings across all 8 chains still serves to demonstrate more consistent, robust model behavior than does a single distribution for each parameter as produced by a standard single-chain model.

Figure 3.2: *Density Plots of MCMC Parameters*



48,620 potential ties).

Each of the remaining parameters can similarly be exponentiated and then interpreted as acting multiplicatively on the sum term. The Mutual parameter[10] represents the change in probability of observing a tie between two organizations, for instance a tie from A to B, given an observed tie from B to A. The observed values for "mutuality" in the data are taken to be the minimum value or count of the two reciprocal dyads ($min(y_{i,j}, y_{j,i})$. As expected, this shows a strong and significant correlation between both the tie counts associated with a network dyad. In other words, I would expect A to be much more likely to report a tie to B

[10] Which takes the general form $g_{\leftrightarrow} = \sum_{(i,j)\in Y} \min\left(y_{(i,j)} y_{(j,i)}\right)$

Table 3.1: Baseline Models

	Baseline Model	Group Participation	Group Participation^2
Sum	-4.19^{***}	-4.19^{***}	-4.44^{***}
	(0.06)	(0.06)	(0.07)
Mutual	1.70^{***}	1.65^{***}	1.60^{***}
	(0.09)	(0.09)	(0.08)
Transitivity	0.11^{***}	0.10^{***}	0.08^{**}
	(0.03)	(0.03)	(0.04)
Num. Resp.	0.16^{***}	0.12^{***}	0.07^{***}
	(0.005)	(0.01)	(0.01)
Num.Groups	-0.01^{**}	-0.01^{*}	-0.01^{***}
	(0.01)	(0.005)	(0.01)
Years	0.01^{***}	0.01^{***}	0.01^{***}
	(0.002)	(0.002)	(0.002)
Org. Type	0.26^{***}	0.28^{***}	0.25^{***}
	(0.07)	(0.07)	(0.07)
Group Partic.		0.03^{***}	0.11^{***}
		(0.002)	(0.01)
Group Partic.^{2}			-0.002^{***}
			(0.0002)
AIC	−115,800.10	−115,928.30	−116,144.70
BIC	−115,738.60	−115,858.00	−116,065.60

Note: ***p < .01; **p < .05; *p < .1

given that B has already reported a tie to A (consider the informational jump in this case from the basic probability associated with two randomly selected organizations). The expected value of a tie from one organization to another increases by 417% ($exp^{1.70} = 5.47$) for each tie observed in the opposite direction. Given that there are many potential network partners in this network, it seems reasonable to presume that defection risk is relatively high; this result then appears to be in keeping the Berardo and Scholz (2010) Risk Hypothesis, which predicts that reciprocity (i.e., mutual ties) will be favored over bridging ties in networks where defection risk is high. The avoidance of asymmetrical ties (i.e., high levels of reciprocity) has already been shown specifically in estuary networks very similar to those analyzed in this paper by Berardo and Scholz (2010) and Desmarais and Cranmer (2012a) (and Puget Sound generally is one of the sub-cases considered in both analyses).

The Transitive Weights parameter[11] is interpreted in a similar way as that of the Mutual parameter, but in this case refers to the degree to which presence and strength of a two-path[12] between two organizations affects the expected value of a direct tie between the same two organizations. Table 3.1 shows that for each unit by which the highest value two-path between two organizations increases in value (for instance, if $y_{i,j} = y_{j,k} = 2$ instead of $y_{i,j} = y_{j,k} = 1$), the expected tie value between the organizations increases by 12% ($exp^{0.11} = 1.12$). This result is also in keeping the Berardo and Scholz (2010) hypothesis, since transitivity (wherein closure tends to occur such that two-paths become triangles in which all three nodes are linked) is also assumed to be more prominent when defection risk is greater.

The baseline model also fits four node-specific variables that are predicted to affect the probability of observing ties associated with a given node. The Number of Respondents

[11] Which takes the general form $g(y) = \sum_{(i,j)\in Y} \min(y_{i,j}, \max\limits_{k\in N}(\min(y_{i,k}, y_{k,j})))$ in which the strength of a two-path from i to j is defined by the minimum value along said path, and the statistic then reflects the sum of minimums over all dyads (i, j) and the value of the strongest two-path in between each i and j combination.

[12] If there is an edge from A to C and from C to B, A and B are said to be connected via a two-path

(Num. Resp.) parameter reflects the predicted impact on the expected value of a tie associated with an organization that has an additional respondent in the survey. This parameter should be positive, since having an additional respondent increases the number of ties that can possibly be reported by an organization. As expected, this parameter is significant and positive, predicting a 17% ($exp^{0.16} = 1.17$) increase in the expected value of a tie between the focal organization and each other organization. Interestingly, the parameter reflecting the predicted change for participation in an additional group is statistically significant and negative; however, the magnitude of this suggested relationship is minimal (only a 1% decrease in the probability of observing a tie between the focal organization and another organization [$exp^{-0.01} = 0.99$]). The 'Years' parameter models the predicted association between the number of years a respondent has been at their position and the value of a network tie. This parameter is small, but positive, with each year predicting a 1% increase in the value of a tie ($exp^{0.01} = 1.01$). Finally, the 'Org. Type' variable compares the expected tie value to similar organizations. Specifically, organizations are coded as 15 different types of organizations, including state agencies, consulting firms, and tribes.[13] The predicted value of a tie (i.e., the strength of a collaborative tie) to a similar organization is 30% greater than that of a dissimilar organization ($exp^{0.26} = 1.30$); this is consistent with the commonly observed network phenomenon of homophily, in which actors with common characteristics are more likely to be linked in a network (Prell 2012; Kolaczyk 2009).

Group Participation

The second column (Group Participation) of Table 3.1 adds a generic collaborative group participation metric to the restricted baseline model shown in Column 1 (Baseline Model). Of course, simply being a member of a group does not necessarily speak to the degree to which

[13]The full list includes: environmental NGOs, city governments, local advocacy groups, local commissions, special districts, county governments, local outreach and education organizations, consulting firms, regional advocacy groups, tribes, federal agencies, parks and reserves, natural resource extraction firms, universities and research organizations, regional commissions, and state agencies.

an organization actually participates. Instead one would expect that inter-organizational transaction costs are reduced in proportion to the degree of participation, and further that the strength of the association between group membership and tie formation should be commensurate with the degree to which organizations actually participate in collaborative groups. In order to test this, group participation is measured in terms reported engagement in seven different group activities for each group in which a respondent reports membership: (1) send or respond to group emails; (2) attend group meetings; (3) attend other group events; (4) participate in group projects or programs; (5) read or review group reports and documents; (6) produce group reports or documents; and (7) other types of participation. This value is then divided by 7 to produce a fractional measure of participation in a given group. This per-group participation measure is then summed to produce an overall measure of group participation for a given organization.[14] The Group Participation model (Column 2) in Table 3.1 supports H1, suggesting that participation in a collaborative group does increase direct coordination and cooperation with other group members. However, while this parameter is shown to be positive and significantly different from zero, it appears to be of limited practical significance: a one-unit increase in group participation increases the expected value of a tie involving the focal organization by 3% ($exp^{0.03} = 1.03$).

---

[14]This measure is obviously a coarse metric of group participation. Ideally, group participation would be measured using a scalable metric such as hours per month so as to convert different forms of participation into a common metric while accounting for the different time commitments associated with different activities. While these data cannot be obtained for this analysis, a time-based effort measure would be a valuable contribution of future studies. Because there is no way to rule out individual heterogeneity for specific group activities in these data (e.g., group projects could represent something very different–and potentially a very different expenditure of time and effort–even for different individuals in the very same group), I use the basic summation approach to avoid layering additional–possibly unfounded–assumptions on this variable.

Diminishing Returns

One likely reason that the predicted impact of group participation is small in magnitude is that a linear participation metric does not account for potential that, as predicted by H2, there are likely to be diminishing returns associated with a marginal increase in group membership and participation (Lubell et al. 2010). Recent policy literature discussing how organizations operate within complex institutional environments (e.g., Lubell 2013; Berardo and Scholz 2010), along with empirical findings regarding the practical limitations most managers and officials face (Thomas 2003; Margerum 2011), both demonstrate that organizations have a limited capacity for collaborative activities (Lubell et al. 2010). Thus, a government-sponsored collaborative group intended to further enhance network coordination might actually motivate organizations to reallocate their existing efforts across these multiple arenas without a net increase. As a basic test, the third column of Table 3.1 (Group Participation^2) shows a model with an additional quadratic term for group participation. In this model, the predicted linear association between group participation and tie occurrence becomes much stronger, predicted a 12% increase in expected tie value ($exp^{0.11} = 1.12$), and the quadratic term is very small in magnitude but negative. In other words, participation in a collaborative group does strongly predict an increase in tie values associated with the participating organization, but this predicted increase diminishes as the number of groups the organization participates in increases.

Whereas the quadratic term in Table 3.1 tests this phenomenon generally (by examining overall participation in Puget Sound area collaborative groups regardless of origin), I examine the PSP's intervention specifically by dividing the Group Participation variable shown in Table 3.1 into two categories: participation in non-PSP affiliated groups (i.e., those in existence prior to the network intervention) and participation in PSP sponsored groups. Because these data are observational, I am unable to eliminate the possibility that group participation and inter-organizational collaboration are both products of unobserved organizational characteristics such as motivation to engage with other organizations. There are certainly organizations for whom both participation in collaborative groups and direct coordination/cooperation with other organizations is more beneficial, and organizations for whom both of these activities are less beneficial. What I seek to examine in this analysis, however, is whether there is an observable impact of government-sponsored collaborative groups net of these inherent characteristics. In other words, regardless of organizational "demand" for collaboration, does the the predicted strength of an inter-organizational relationship change in association with a marginal change in collaborative group participation? Using a series of additional model specifications, this relationship is triangulated using a series of progressively stricter coding schemes and comparing the coefficients.

Table 3.2 shows the results of three models. Column 1 (Generic Participation) in Table 3.2 presents a model fitting participation (coded using the same process as described above) in non-PSP groups, participation in PSP groups, and a term interacting participation in each group type (the base parameters are omitted in these and subsequent tables, but are included in all models). These results support H2, showing that participation in either type of group (measured simply in terms of the level of reported engagement in group activities) is positively associated with the expected value of an inter-organizational tie, but that there are diminishing returns to participation in both types of groups. In other words, the association between group participation and the expected value of a network tie diminishes to the extent that the organizations involved in PSP-sponsored groups already participated in non-PSP groups. This indicates that the PSP's network intervention had the greatest impact on organizations that were not already heavily involved in collaborative groups. Pragmatically, this makes a great deal of sense; organizations more heavily involved in collaborative groups prior to the network intervention presumably already had greater opportunity to engage with and initiate ties to other organizations.

The next two models in Table 3.2 present increasingly stringent ways of coding for group participation. If participation in collaborative groups is a driver of inter-organizational networking, then the impact of a collaborative group should primarily be experienced with regards to organizations that are actually members of the same group. In other words, while

Table 3.2: Triangulating Participation

	Group Participation	Direct Participation	Co-Participation
Group Partic. non-PSP	0.15***	0.54***	1.13***
	(0.01)	(0.02)	(0.04)
Group Partic. PSP	0.12***	0.86***	1.59***
	(0.01)	(0.03)	(0.06)
Group Partic. non-PSP*PSP	−0.03***	−0.03***	−0.11***
	(0.002)	(0.002)	(0.01)
AIC	−116,174.50	−117,194.20	−117,562.00
BIC	−116,086.60	−117,106.30	−117,474.10

Note: ***p < .01; **p < .05; *p < .1
Note: Each model also includes all terms shown in the Column 1 of Table 3.1.

collaborative groups might certainly increase awareness of and access even to other organizations that are not in the same group, it stands to reason that the association between participation and network ties should be concentrated amongst organizations who actually are members of the same group. Thus, I create a ''Direct Participation" measure, by conditioning total participation on group membership. Each organization is given a distinct Direct Participation score to every other organization. This score is zero if the organizations do not share a group. If two organizations do share at least one group, then the Direct Participation score from Organization A to Organization B reflects group participation reported by Organization A within groups of which Organization B is also a member. If group participation and inter-organizational networking are confounded by a common cause, then the predicted association between the general metric ''Group Participation" and the occurrence of network ties (column one) should not be meaningfully different than the predicted association between "Direct Participation" and the occurrence of network ties (column two). However, what I observe in Table 3.2 is that the coefficients become much stronger. A one unit increase in general Group Participation in a PSP-affiliated group increases the predicted value of a network tie by 13% ($exp^{0.12} = 0.13$), whereas as one unit increase in Direct Participation in a PSP-affiliated group increase the predicted value of a network tie by 136% ($exp^{0.86} = 2.36$).

While Direct Participation focuses on organizations that actually share group membership, the theoretical mechanisms by which collaborative groups are hypothesized to reduce transaction costs (and thus engender increased inter-organizational collaboration) holds that collaboration is motivated by ''principled engagement" and ''increased capacity for joint action" (Emerson et al. 2012). In other words, by engaging directly with one another, organizations build trust and learn more about one another's needs, interests, and capabilities, thereby decreasing the costs associated with searching for, initiating, and maintaining network ties. Practically speaking, this can only occur if both organizations "show up" and actually encounter one another in group activities. Thus, I code a third metric, ''Co-Participation," in which the covariate associated with each possible pair of organizations is the minimum combined Direct Participation measure shared by the two organizations (column three). For instance, if Organization A and Organization B are both members of one group, in which A participates in five of seven activities and B participates in three of seven, then the Co-Participation score for dyad AB (and for dyad BA) is three.[15] The coefficients in column three of Table 3.2 are even stronger than those in column two. A one unit increase in Co-Participation increases the predicted value of a network tie by 390% ($exp^{1.59} = 4.90$).

As with the generic participation metric tested in column one of Table 3.2, the stricter participation metrics fit in columns two and three also speak to the diminishing returns associated with greater levels of participation. Specifically, the negative interaction terms shows a lessened participation effect when an organization participates in both kinds of groups. It seems reasonable to expect, for instance, that if two organizations both participate

[15]Note that I purposefully do not match on specific activities, as doing so would inject a great deal of false specificity into the model. For instance, when two organizations both report that they participate in group events or help write group documents, there is no way to verify whether these organizations participated in the same events or have helped write the same documents. Accordingly, I take the conservative approach of not matching on specific activities; this approach is conservative because if collaborative ties are in fact strongly associated with participation in the same group activities, then the strength of this association will be underestimated in my model.

in all possible activities for the same two collaborative groups, then shared participation in a new additional group is unlikely to make as much of a difference in terms of increasing familiarity between the two organizations. One note of caution, however, is that the negative interaction parameters imply that at very high levels, participation can have the perverse effect of decreasing predicted tie strength; conceptually, while it seems reasonable to assume to assume that participation in too many groups can diminish the predicted impact on tie strength to the point that the predicted impact of a marginal increase in participation is zero, it is difficult to come up with a rationale for high levels of shared participation predicting lower tie strength than exists between two randomly selected organizations in the network (excepting some sort of interpersonal conflict engendered by too much interaction, which is not addressed by these data). Thus, it is likely that these results reflect a somewhat localized marginal estimate and not a wholly linear relationship in which too much participation has a perverse effect.

While these observational data cannot eliminate the role of selection bias (namely, that organizations that are more willing to participate in collaborative groups tend to also be more willing to engage in direct collaboration with other organizations), by modeling increasingly strict conceptions of group participation I am able to establish that the predicted relationship between group participation and the level of inter-organizational collaboration is strongest among organizations that both participate to a similar extent in the same group(s). This suggests that the collaborative group itself has some role in driving network structure, net of intrinsic organizational characteristics. While this is by no means conclusive evidence about the causal effect of sponsoring collaborative management groups, these results are what one would expect to observe if providing increased opportunities for organizations to engage with one another does in fact engender increased inter-organizational collaboration. Realistically, it is likely both that this occurs and that some confounding is present.

Pre-existing Network Ties

One potential contributor to the overall magnitude of the Direct Participation and Co-Participation parameters is the combined sparsity of the overall network and the informational value these parameters provide about the network proximity of any two organizations. Specifically, it is plausible that instead of forming relationships within collaborative groups, organizations invite other organizations with whom they already engage to participate in groups. In one sense, this would represents reverse causality, in which network relationships drive collaborative group participation, rather than the other way around. However, the possibility that existing network relationships drive group participation does not rule out the possibility that group participation alters the strength of these existing relationships. In other words, an organization might join a group because it already coordinates or cooperates with other group members, but group participation might still serve to deepen these relationships. Strengthening existing patterns of coordination and cooperation might even be the goal of a network intervention. For this reason, H3 contrasts the predicted impact of group participation on tie strength between two members who share a pre-existing tie with the predicted impact on two members who do not share a pre-existing tie.

Table 3.3: Pre-Existing Ties

	Past Tie
Past Tie (PT)	4.73***
	(0.07)
All Group Co-Part.	1.65***
	(0.06)
All Group Co-Part.^{2}	−0.21***
	(0.01)
All Group Co-Part. * PT	−0.89***
	(0.06)
AIC	−123,046.20
BIC	−122,949.50

Note: ***p < .01; **p < .05; *p < .1
Note: Each model also includes all terms shown in the Column 1 of Table 3.1.

In order to test this, the survey instrument asked each respondent to report whether each reported inter-organizational tie had existed prior to the PSP's network intervention (i.e., more than five years ago). This provides a rough approximation as to whether a network tie predates the intervention (the exception being in cases where the tie was initiated within the last five years, but prior to the organization becoming a group member). While this does not provide as robust a picture as would a true longitudinal analysis, it does provide a general gauge of the extent to which reverse causality might be at play by comparing how the association between group participation and inter-organizational ties differs between ties that existed prior to the PSP's intervention and those that did not.

Table 3.3 presents results from a model that includes all of the covariates specified in the baseline model and adds four additional parameters. First, the model fits a binary term for each potential dyad (pair of organizations) that reflects whether the tie reported from Organization A to Organization B was reported to exist prior to the PSP's network intervention. For a binary network, it would not be feasible to use this metric, since the metric would be a perfect predictor of the affirmative presence of a tie (but not tie absence, since more newly formed ties are reported as well)[16]. However, since this analysis models valued network ties, it can examine how the fact that a tie was also reported to exist before the network intervention impacts the expected value of the tie after the intervention. Given the self-reinforcing nature of network relations (Desmarais and Cranmer 2012b), one might also expect that on average, ties that existed five years ago should be stronger than ties that did not. As expected, this parameter is positive and very strong in magnitude; this result is not substantively interesting, however, since past ties reported in the survey instrument are restricted to currently existing ties. Table 3.3 then fits both a linear and a quadratic term for Co-Participation in both PSP and non-PSP groups. This metric is calculated in the exact same way as the metrics in column three of Table 3.2, but in this case combines participation in both types of group. As in Table 3.2, the linear parameter is positive

[16]Of the 1045 ties observed in the data, 581 are reported to be ties that predate the network intervention.

and strong in magnitude, and the quadratic parameter is negative, again demonstrating a diminishing marginal predicted association.

The primary parameter of interest in Table 3.3 is the fourth parameter, an interaction term between co-participation in collaborative groups and the fact that a tie was reported to have existed prior to the PSP's network intervention. This parameter is negative and large in magnitude; that a tie of any value existed prior to the network intervention decreases the association between co-participation in a collaborative group and the value of a network tie by 69% ($exp^{-0.89} = 0.41$). In other words, shared participation in a collaborative group has much less impact on the strength of a relationship between two organizations who already engaged in some level of coordination or cooperation prior to joining the group. Note however that the strong positive relationship between co-participation and tie strength amongst organizations that did not have a tie prior to the network intervention remains. This supports the conceptual expectation of H3: in the short term collaborative groups are more likely to foster new bridging ties (or "weak ties") amongst participants than to increase the level of coordination and cooperation amongst existing network partners (e.g., transition from a bridging tie to a bonding tie). In a practical sense, these results also indicate that the relationship between collaborative group participation and network ties works in both direction; organizations that share a network tie are more likely to join and participate in the same collaborative group, and organizations that join and participate in the same collaborative group are likely to increase the strength of their network relationship.

DISCUSSION AND CONCLUSION

While it is clear that governments are beginning to intervene to address perceived network failures just as they seek to remedy market failures and other institutional shortcomings, there is not yet a large body of knowledge built up surrounding the use of various policy tools for network interventions. This analysis builds on the work of Schneider et al. (2003), who demonstrate the overall change in network conditions associated with a network intervention by government actors (e.g., denser networks), and Henry (2011), who model the role of social capital as a driver of policy network structure, by examining the relationship between government network interventions and the network behavior of individual organizations. These results advance the growing body of network-related policy theory by developing a transaction cost-based perspective of government network interventions. Specifically, I find a strong positive association between participation in collaborative groups and the number–and intensity–of inter-organizational ties an organization has. This evidences how government intervention in a policy network is associated with changes in network structure and function and complement the findings of (Scott and Thomas 2015), which examines the specific mechanisms by which collaborative groups are hypothesized to reduce the transaction costs organizations face with regards to networking (and thus motivate the formation and maintenance of network ties). By modeling the comparative intensity of an inter-organizational relationship (consultation, planning, or implementation), this analysis is able to model this relationship in a more nuanced fashion than does the policy network literature using binary ERGMs (e.g., (Scott and Thomas 2015), Lubell et al. 2012; Gerber et al. 2013; Schneider et al. 2003). Thus, I make a methodological contribution to the policy literature by providing one of the first empirical demonstrations of valued ERGM analysis in policy research.

Second, discussions of policy networks often overlook the fact that–particularly in highly institutionalized contexts–organizations are faced with numerous networking opportunities and are limited in their ability both to engage in collaborative groups and to maintain ties with individual organizations Lubell et al. (2010) and Lubell et al. (2011). that In spite of the rapacious growth of collaborative institutions in practice, organizations do in fact have a finite capacity for networking. The Ecology of Games literature (e.g., Berardo and Scholz 2010; Lubell et al. 2010; Lubell and Lippert 2011; Lubell et al. 2011; McAllister et al. 2014; Smaldino and Lubell 2014; Niles and Lubell 2012) thus predicts that in some contexts, a network intervention might simply induce organizations to switch the decision-making venue(s) in which they participate or change network partners without engendering an net increase in network coordination. This analysis provides a real-world example supporting this theoretical contention, showing that the association between group participation and tie strength diminishes with each additional group in which an organization participates. For policymakers, this speaks to a need to fashion context-appropriate network interventions in order to have the greatest possible impact and avoid just shuffling the deck. It is important to recognize instances where there are already numerous–if not too many–collaborative venues, and thus a new network intervention is likely to have diminished returns.

Third, I find that in the short run, collaborative group participation is a stronger predictor of new tie formation than it is of strengthened existing ties. I argue that this is because the short term impacts of collaborative group participation do more to alleviate search costs than to align organizational goals and beliefs. Lack of information about potential network partners is shown to greatly constrain network coordination and cooperation (Scholz et al. 2008); participation in a collaborative group can increase awareness of other organizations (actions, goals, capabilities, etc.) and directly reduce search costs. If two organizations are highly compatible but previously unaware of this compatibility, shared group participation might lead to the formation of a new, strong network tie (e.g., joint policy or program implementation in this model). For organizations who share an existing tie prior to participating in the same collaborative group, however, search costs would not seem to be a major barrier to a stronger relationship. One might expect then that the reason two organizations with an existing tie engage in informal consultation rather than coordinated planning is due to other factors, such as a lack of commonality. Collaborative group participation likely does little to address something such as lack of commonality, at least in the short term. Accordingly, these results suggest that in the short run, government sponsored collaborative groups do more to influence the formation of new network ties than they do to strengthen existing patterns of coordination and cooperation.

Going forward, longitudinal network data that track the formation and maintenance collaborative ties over time are likely critical for fully understanding these issues. Though it is expensive and difficult to collect repeated measures network data (even in collecting these cross-sectional data, the author encountered many practitioners who were too busy to participate and had too many competing demands on their time), a longitudinal perspective is required to better understand network dynamics over time and trace the long-term impact of public policy interventions. As stated, this analysis examines only the short run predict impact of a network intervention; that is, I assume that a publicly sponsored collaborative group subsidizes transaction costs–making network ties "cheaper" in a general sense–and then test whether this theoretical cost reduction motivates increased inter-organizational networking. However, there is evidence that interaction and dialogue with other organizations also influences goals and beliefs (Bingham and O'Leary 2008; Innes and Booher 2010; Leach and Sabatier 2005; Lubell 2005); in other words, a network intervention might alter the demand for network ties in the long run by changing organizational goals and motivations. While this analysis finds that collaborative groups do more to foster new ties rather than strengthen old ones in the short term (H3), it is possible that the long term impacts are different. Accordingly, longitudinal data tracking network ties over a longer time period by sampling at multiple points might speak more fully to the extent to which government can motivate inter-organizational coordination and cooperation.

In a similar fashion, more detailed network metrics, such as time spent working in concert or resource transfers between organizations, would also contribute further to our understanding of the role that publicly sponsored collaborative groups play in policy networks–and whether such network interventions achieve their intended purposes. One element of particular relevance in this regard are the interpersonal ties that exist alongside–and potentially interact with–inter-organizational ties. Within a given policy network, we might expect that individuals have shared interests, similar training, and other commonalities; especially in instances where individuals move from one network organization to another (e.g., a public manager joining a consulting firm), individual-level factors might play a large role in inter-organizational networking.

Chapter 4

IS COLLABORATION A GOOD INVESTMENT? MODELING THE IMPACT OF COLLABORATIVE GOVERNANCE ON WATER QUALITY

INTRODUCTION

Governments increasingly rely on collaborative relationships with non-profit organizations to implement policies or provide services (Salamon 2002). Collaborative management with local nonprofit groups gives governments a community-based vehicle through which to implement policies and programs, and provide nonprofits with access to funding and other resources (Nikolic and Koontz 2008). Management arrangements of this form are very common in environmental applications, particularly watershed management and water quality (e.g., Leach et al. 2013; Leach et al. 2002; Margerum 2011; Hardy and Koontz 2008). In this paper I build on the considerable body of research discussing the role that governments play in–and resultant impacts of–supporting collaborative management (e.g., Nikolic and Koontz 2008; Lubell and Fulton 2008; Ansell and Gash 2008; Emerson et al. 2012) by asking a relatively simple question that proves highly elusive in practice: How does government support for collaborative management affect environmental outcomes? To examine this question, I use publicly available water quality monitoring data to explore the impact of 2500 grants given by a state agency, the Oregon Watershed Enhancement Board (OWEB), to local non-profit stakeholder councils engaged in ongoing watershed planning and management activities in watersheds across Oregon over the course of almost 20 years.

To model these data I use Bayesian hierarchical modeling, specifically Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009) for estimating complex hierarchical models and Stochastic Partial Differential Equations (SPDE) (Lindgren et al. 2011) for modeling spatial and temporal dependency, in order to account for the complex spatio-temporal nature of these data.[1] In this paper I also make a methodological contribution to the policy literature by helping to establish the use of these methods for testing public policy and management theory. The INLA approach facilitates large-scale hierarchical models and complex

[1] There are of course numerous methodological approaches for modeling spatially and/or temporally correlated observations; in the methodological discussion below, I discuss why I use the SPDE approach and INLA estimation method in particular.

specifications that account for irregular data and spatial and temporal relationships. This enables the use of publicly available, observational environmental data and helps address some of the analytical challenges that have prevented researchers from linking collaborative management efforts to environmental outcomes in the past (Koontz and Thomas 2006; Thomas and Koontz 2011).

Using this analytical approach, I address two primary research questions: (1) Are collaborative watershed council actions (using funding as a proxy for action) associated with a measurable change in water quality? and (2) How does this predicted impact compare across different types of funded actions? Specifically, how does the predicted change in water quality of funding collaborative group activities that have a direct environmental output (e.g., riparian revegetation) compare to supporting production of indirect outputs (e.g., supporting council administrative activities, which in turn shape future environmental outputs)? In what follows, I first describe the theoretical rationale for this research, and then provide background concerning the case analyzed. I then specify my analytical approach and introduce the INLA and SPDE methods. The remaining sections present the data and model results, and discuss the implications of these findings.

RATIONALE

The idea that policy implementation does not solely involve autonomous actions by public agencies is long-established (e.g., Ostrom et al. 1961) and ubiquitous in modern policymaking (Salamon 2002, p. 8). In fact, it is unclear in practice what an alternative to the general concept of collaboration (Donahue and Zeckhauser 2011; Agranoff and McGuire 2003), in which government agencies communicate, consult, coordinate, or cooperate with other public, private, and nonprofit entities, would even be. A stricter definition of collaborative management, which involves government initiation and/or funding (Ansell and Gash 2008) for efforts "in which a... group of autonomous stakeholders deliberates to build consensus and develop networks" (Margerum 2011, p. 6) in order to "make or implement public policy or manage public programs or assets" (Ansell and Gash 2008, p. 544), raises a more interesting issue: collaboration by this definition does not just happen, but rather policymakers choose collaborative management as a means through which to design and implement policies (Layzer 2008; Hoornbeek et al. 2012; Koontz et al. 2004). Does using funding nonprofit collaborative stakeholder groups result in improved environmental outcomes? For instance, is it more beneficial for the state of Oregon to implement a restoration project directly or to provide grant funds to a local nonprofit watershed council to implement the restoration project instead?

While collaborative management is heavily documented in the policy and management literatures (Emerson et al. 2012; Margerum 2011; Sabatier et al. 2005; Ansell and Gash 2008; Lubell 2004a), the complexity of social-ecological systems makes it difficult to trace how government support for non-profit collaborative groups ultimately impacts environmental outcomes (Koontz and Thomas 2006; Thomas and Koontz 2011). Clearly, the context of a particular locale or project greatly determines the answer to this question. More broadly, however, the literature concerning collaborative management provides a theoretical basis to help understand the rationale for governments to partner with nonprofit collaborative groups. The general rationale for supporting collaborative management, such as by funding a nonprofit collaborative management group, is that collaborative efforts yield more holistic and comprehensive management. For instance, collaborative management is shown to enhance cooperation amongst stakeholders (Lubell 2004a), alter existing stakeholder beliefs (Leach et al. 2013), increase information exchange and learning amongst actors (Beierle 2002; Weible et al. 2009), foster trust and collective action (Lubell 2005), and incorporate a broader range of information (Innes and Booher 1999; Wondolleck and Yaffee 2000). Local nonprofit collaborative groups can be flexible and responsiveness, pivoting to meet local needs and concerns (Nikolic and Koontz 2008).

On the other hand, since collaborative management is deliberative and consensus-oriented (Ansell and Gash 2008), it can involve a great deal of time and effort (Margerum 2011). Further, in contrast with direct government actions where the implementing agency has stricter control over policy efforts (Salamon 2002), policy implementation via collaborative management has a heightened degree of uncertainty from the perspective of the funding agency since the involvement of more actors creates issues of accountability and control (Weber 2003). There is also concern that government support for local collaborative management efforts can detract from the ability of these local organizations to operate with responsiveness and flexibility, reducing the very qualities that are presumed to make such groups effective (Smith 2004b; Nikolic and Koontz 2008).

Nonetheless, the use of collaborative management continues to proliferate (Ansell and Gash 2008; Emerson et al. 2012) as a response to complex environmental problems (Margerum 2011). Among the practical rationales for public managers to choose collaborative management are the expectations that collaborative management will facilitate a more comprehensive understanding of policy problems (Leach et al. 2013), alleviate conflict between stakeholders (Berardo et al. 2014), reduce interorganizational transaction costs (Emerson et al. 2012), and foster greater "buy-in" from stakeholders (Ansell and Gash 2008). In short, policymakers believe that collaborative management will improve the design and implementation of policies and programs and thereby improve policy outcomes.

BACKGROUND

Nonprofit watershed councils in the state of Oregon have proved to be a fruitful object of study for those interested in collaborative management (Griffin 1999; Dakins et al. 2005; Margerum 2002, 2007, 2008; Margerum 2011; Habron 2003; Margerum and Whitall 2004; Hibbard and Lurie 2006; Lurie and Hibbard 2008). My study differs significantly from the cited works in that none of these analyses evaluate the water quality impacts of watershed councils on a long-term, statewide basis (most involve intensive case studies that focus on specific watersheds). Created by the state legislature in 1995, OWEB provides guidance, administrative support, and resources to 87 local watershed councils. This support includes grants to local watershed councils for myriad purposes ranging from environmental assessment to hiring a full-time coordinator. Because grants are awarded by OWEB on a competitive basis, these data are not suitable for providing a generalizable estimate of how funding collaborative management in a randomly selected watershed improves water quality. This is because the grant application and approval process is intended to identify and select motivated parties and favorable circumstances, which results in selection bias at the watershed level of analysis. However, what is really of interest in this case is the efficacy of Oregon's statewide support of a collaborative watershed management system. The case that I consider is the state of Oregon's ongoing financial support for collaborative watershed management. In other words, the purpose of this analysis is to evaluate the statewide OWEB watershed management strategy (of providing public funds to nonprofit management councils) by modeling whether providing funds to watershed councils corresponds to improved water quality. It is difficult to envision a state or regional program not administered on similar grounds (with funds strategically allocated), and thus this analysis thus provides a good conception of the effectiveness of a grant-funded collaborative watershed management system implemented at a regional governance level.

HYPOTHESES

The OWEB statewide, grant-based system also provides a unique way to address one of the challenges typically faced in estimating the impact of collaborative groups: even if budget data are available, it is difficult to determine the activities and relative effort level of each group. OWEB grant funds provide a consistent approximation of collaborative management effort taking place in given place and time in the state. Rather than simply comparing "collaborative" and "non-collaborative" watersheds, grant funding provides a continuous metric of collaborative management that can be used to estimate how increased support for collaborative management relates to environmental outcomes. If collaborative watershed management has an impact on water quality, greater support (as measured by grant funding in this case) should be associated with a larger predicted impact:

- H1: Grant funds given to nonprofit watershed councils are associated with improved water quality in the target watershed.

Building upon this basic hypothesis, I then distinguish between grants that more directly target water quality through environmental restoration projects and grants that indirectly target water quality by funding: (1) assessment, monitoring, and other technical actions; (2) educational and outreach programs; and (3) council activities and personnel. "Program logic" (Bickman 1987), the narrative for how a given program will work to address an identified problem (Margerum 2011; McLaughlin and Jordan 1999), differs greatly for these different types of grants, particularly between restoration projects and the three non-restoration project types. In a simplistic sense, restoration projects can be viewed as a direct purchase of environmental outputs; Typical restoration projects implemented by watershed groups include abatement or prevention of nonpoint or point pollution sources, in-stream flow modifications or water allocation, stream channel restoration, and changes in land-use designations (Leach et al. 2002). These outputs directly contribute to water quality by altering physical conditions. While the extent to which these outputs will alter environmental outcomes might be uncertain, the program logic is fairly straightforward.

Conversely, grant types, including for assessment and monitoring, education and outreach, and council administrative actions, might best be characterized as an investment: Instead of directly purchasing environmental outputs, the program logic of supporting non-restorative actions is essentially that investing in council efforts that produce non-environmental outputs will ultimately engender a return in environmental outputs–and thus outcomes–over time. Much of the collaborative management literature implicitly revolves around this idea. Monitoring and assessment projects, for instance, provide data and information that managers can use for adaptive policymaking (Huntington 2000; Leach et al. 2002; Yaffee et al. 1996). The program logic for investing in monitoring and assessment is that investing in better information will facilitate improved management practices and decision-making, which will thereby result in improved outcomes.

Similarly, outreach projects seek to build community support for environmental efforts in a watershed (Huntington 2000) and educational projects seek to foster agreement on how to improve a watershed (when internal to the council) (Leach et al. 2002) or to promote environmentally responsible behaviors (when external to the council) (Nikolic and Koontz 2008). Education and outreach program logic is that such programs foster learning that alters stakeholder beliefs and practices (e.g., Leach et al. 2013; Leach et al. 2002) or build community support and motivation (Huntington 2000; Emerson et al. 2012), serving to change environmental behaviors or improve implementation efforts and in turn improving environmental outcomes. For instance, Lubell and Fulton (2008) demonstrate that increased exposure to policy networks, such as might be fostered through an outreach project, increases the probability that landowners will adopt environmental practices.[2]

Finally, council support grants fund group coordinators and group administrative functions. In other words, these types of grants support the existence and operation of the nonprofit watershed council itself. Collaborative management is theorized to engender principled engagement amongst participants that fosters shared motivation, which in turn enables joint actions which could not be accomplished separately (Emerson et al. 2012). Adequate support is shown to be critical to for collaborative efforts to be successful in this regard (Lubell et al. 2009). In particular, there are significant transaction costs associated with initiating and maintaining inter-organizational endeavors, and government provision of staffing, infrastructure, and other resources are shown facilitate such efforts (Schneider et al. 2003). Emerson et al. (2012) theorize that collaborative actions are likely to be implemented only to the extent to which the three collaborative dynamics (principled engagement, shared motivation, and capacity for joint action) function. Thus, to put a fine point on a broad

---

[2]There is a large body of literature in the field of environmental learning and education that challenges the overly sanguine perspective on the impacts of education and outreach, what Heimlich (2010, p. 183) refers to as the "false causality of knowledge or attitude leading to behavior," that increasing stakeholder understanding or concern will lead to behavioral change (e.g., Hungerford and Volk 1990). However, Hines et al. (1987) do find in a large-scale meta-analysis of education and outreach program evaluations that issue and action strategy knowledge, sense of responsibility, locus of control, and verbal commitment are all associated with responsible environmental behaviors. Heimlich (2010)'s contention is thus not that education and outreach have no impact in practice, but rather that this process is highly complex and demands a more nuanced theoretical model.

and nuanced literature, the program logic for investing in council support is that a well-functioning collaborative group will result in policy and program outputs that draw on more relevant perspectives (Ansell and Gash 2008; O'Leary et al. 2006; Leach 2006), have increased commitment from participants (Bryson et al. 2006; Ansell and Gash 2008), and could not be produced separately (Emerson et al. 2012), and that these improved outputs will result in improved environmental outcomes.

As with any investment, the program logic for each type of grant speaks to a tradeoff of risk and reward. For instance, education projects are affordable and easy to carry out (Leach et al. 2002), but have a much more uncertain causal link to environmental outcomes than do projects that produce direct environmental outputs. If a project does not alter stakeholder behavior, or if the impacts of stakeholder behaviors are relatively minor in the grand scheme of things, then it will not change environmental outcomes at all. However, if a project is successful in changing consequential stakeholder behaviors or helps the community agree on necessary improvement steps, then it might result in large changes in environmental outcomes at the cost of producing relatively inexpensive outputs. Similarly, a relatively minor investment in improving processes and policies via collaborative approaches might result in considerable impact. With the exception of certain economies of scale, such as that achieved by improving habitat connectivity, OWEB essentially gets only us much restoration as it pays for; other nonprofit watershed council activities can potentially produce more environmental outputs in excess of the original expenditure.

However, as mentioned previously the literature also emphasizes the time-consuming nature of collaborative management (e.g., Margerum 2011; Ansell and Gash 2008; Imperial 2005) and expresses skepticism about possible achievements of deliberative, consensus-based policymaking (Coglianese 2003; Gunton 2003). In any case, since we do not know a great deal about the relationship between collaborative management outputs and environmental outcomes generally (Carr et al. 2012; Koontz and Thomas 2006; Newig and Fritsch 2009a), I pose a conservative hypothesis that assumes that restoration grants, which have the most direct theoretical causal link between outputs and policy outcomes (e.g., revegetation that directly alters streamside riparian areas, which has an effect on water quality), will produce measurable water quality changes and grants for other purposes (with non-environmental outputs) such as education, outreach, administrative support, and technical support will not:

H2: Grants for restoration will be associated with a measurable improvement in water quality, while grants for projects that do not have direct environmental outputs will not.

MODEL AND METHODS

In accordance with these hypotheses, the goal of this paper is to estimate the association between grants given to watershed councils and subsequent water quality. Water quality index observations (Y_{itj}) occur at location i in time period t within stratum (HUC8 watershed in this case) j. Standard regression models are inappropriate in this application, since it is expected that two observations near to each other in spatial location or time, or within the same administrative boundary (e.g., under the purview of a given watershed council) are more similar than two randomly selected observations (and thus exhibit residual dependency, i.e., correlated residuals). These data are thus highly similar to many epidemiological data contexts, as there is an outcome (water quality) and a “risk factor” or “confounder” (grant funded projects) and where the spatial (sample sites in the state of Oregon) and temporal (monthly observations) structure of the data must be accounted for in order to make valid inferences. Accordingly, I use a suite of Bayesian hierarchical modeling methods found primarily in the epidemiological (Cameletti et al. 2013; Blangiardo et al. 2010, 2013), spatial econometrics (Bivand et al. 2008, 2013; Gomez-Rubio et al. 2014), and ecological modeling (Cosandey-Godin et al. 2014; Clark 2005; Clark and Gelfand 2006; Wikle 2003; Cressie et al. 2009; Wikle and Hooten 2010; Xu and Wikle 2007; Cressie and Wikle 2011) literatures.

A Bayesian model derives a statistical result (the “posterior distribution”) via an inferential process that combines the “prior distribution” (what was assumed prior to observing additional data) with the current data model (Bernardo and Smith 2009). For this application, there are two primary advantages of the Bayesian approach. First, since a posterior parameter distribution is estimated by the model, it is easy to obtain the posterior probability that the parameter does or does not exceed a given value (Blangiardo et al. 2013); this is easier to interpret than p-values used in frequentist statistics. Second, Bayesian methods greatly ease the use of hierarchical model structures, which use random effects to model variance at multiple levels of a model. The Bayesian model I use includes random effects that model spatial and temporal dependence amongst observations. Bayesian methods are shown to be highly effective for analyzing data with this type of spatio-temporal structure (Dunson 2001).

Hierarchical Bayesian models are typically fit using Markov chain Monte Carlo (MCMC) algorithms which use a simulation-based approach to model the posterior parameter distributions (Brooks et al. 2011; Robert and Casella 2004; LeSage and Pace 2010). Since these methods rely on a very large number of simulations applied to complex model structures, MCMC methods are greatly time- and computationally-intensive (Blangiardo et al. 2013). In lieu of MCMC, I use a more computationally efficient method, Integrated Nested Laplace Approximation (INLA), developed by Rue et al. (2009) and widely employed for Bayesian hierarchical modeling (Beguin et al. 2012; Martino et al. 2011; Martino and Rue 2010; Lindgren et al. 2011; Cosandey-Godin et al. 2014).

Given that INLA is relatively new and has as-of-yet limited penetration into the policy literature, along with describing the model specification used in this paper I also provide background on the INLA estimation method more generally. A full description of the INLA methodology is beyond the scope of this paper; Rue et al. (2014) and Lindgren and Rue (2013) provide excellent technical background. Based on the specifications of Cameletti et al. (2013)and Cosandey-Godin et al. (2014) for a spatio-temporal point-reference model, water quality observations in a given watershed at a specific time and location are linked to a structured additive predictor η that is defined linearly as:

$$\eta = \alpha_{h[i]} + \sum_{m=1}^{M} \beta_m Site_i + f(.) + \zeta_{t[i]} + \tau_{t[i]} \tag{4.1}$$

where β_m is the coefficient associated with site covariate m for observation i (including elevation and distance from coast); $f(.)$ represents the semi-parametric function used to model the spatio-temporal random effect (described in more detail below); $\tau_{t[i]} = (t_1, ..., t_T)$ represents a smoothed linear trend (a generalized additive model [GAM] term) accounting for long term water quality trajectory; $\zeta_{t[i]} = (t_1, ..., t_T)$ is a seasonal component with periodicity $(p = 12)$ to account for expected seasonal variation in water quality in Oregon, particularly between low-flow (June to September) and high-flow (October to May) months (Cude 2001); and $\alpha_{h[i]}$ is the random intercept estimated for observation i in a given HUC8 watershed h. The advantage of fitting a random group effect as opposed to a fixed effect is twofold. First, the random effect accounts for differing within-group sample sizes by placing more emphasis on the group mean when there are many observations in the group and drawing more broadly from the population mean when there are very few observations in the group (Gelman 2006; Gelman et al. 2013). This helps ensure that predicted differences between watersheds are not simply a product of small within-watershed sample sizes by attenuating the group-level estimates for watersheds with fewer water quality samples towards the overall mean (producing more conservative group-level estimates than would a basic fixed-effect approach). Second, and of particular importance for this analysis, is that the random group effect can itself be modeled as a function of group-level covariates. This includes important water quality control variables, such as the percentage of land in the watershed that is developed (e.g., paved or contains buildings) and that is used for agricultural purposes. It also includes grant funding, the variable(s) of interest. Since grants are given to watershed councils, it makes the most sense to aggregate funding at the watershed level. The HUC8-specific adjustment is thus itself modeled as:

$$\alpha_h = \alpha_0 + \sum_{w=1}^{W} \gamma_w Watershed_{wj[i]} \tag{4.2}$$

where α_0 is the population average and γ_w represents a vector of coefficients corresponding to watershed-level variables 1 to W for an observation (i) in a given watershed (h).

The way INLA facilitates Bayesian inference on model parameters is by assuming that these model parameters collectively constitute a latent field, $\theta = \{\alpha_j, \beta_m, f, \zeta_t, \tau_t\}$. This latent field is in turn assumed to be a defined by a Gaussian multivariate distribution of mean 0 and precision matrix $Q(\psi)$, such that $\theta \sim N(0, Q^{-1}(\psi)$ (Blangiardo et al. 2013; Rue and Held 2005; Rue et al. 2009; Cosandey-Godin et al. 2014). Using this model, an individual observation $y(s_i, t)$ (at location s and time t) is modeled by a subset of θ according to its spatio-temporal characteristics:

$$y(s_i, t)|\theta, \psi \sim p(y(s_i, t)| \sum_j, A_{ij}\theta_j, \psi) \tag{4.3}$$

In Equation 4.3, the observation matrix A_{ij} is where the SPDE method enters in. In order to account for spatial dependence amongst water quality observations, this model includes a random spatial effect known as a Gaussian random field (GRF) (Cosandey-Godin et al. 2014). Modeling a continuous spatial process obviously poses a significant, and largely intractable, computational challenge. The SPDE method posed by Lindgren et al. (2011) indexes the continuous GRF as a discrete random process, or a Gaussian Markov random field (GMRF) (Lindgren and Rue 2013), by dividing the spatial domain (the state of Oregon in this case) into a "mesh" of triangles (Blangiardo et al. 2013). Essentially, this triangular grid is used to approximate the continuous field. The observation matrix A contains the values of the spatio-temporal random field at the specific times and locations contained in the dataset and uses these values for parameter estimation (Cosandey-Godin et al. 2014). Since values are only stored for specific points (triangle vertices in the mesh) (points within a triangle can be estimated by extrapolating between values at neighboring vertices), this provides considerable computational savings. The latent field is thus linked to model likelihood via A, such that $\eta^* = A\eta$ (Cosandey-Godin et al. 2014):

$$p(y(s_i,t)|\theta,\psi) = \prod_{i=1}^{n} p(y(s_i,t)|\eta^*,\psi) \tag{4.4}$$

Further details regarding INLA and SPDE methods are provided in the context of the analysis presented below. First, I describe the data used in this model.

DATA

Dependent Variable

The dependent variable of interest, a water quality index score, is obtained from the Oregon Department of Environmental Quality (ODEQ). The Oregon Water Quality Index (OWQI) is a multimetric index that integrates eight water quality variables (temperature, dissolved oxygen, biochemical oxygen demand, pH, total nitrogen, total phosphorus, total solids, and fecal coliform) into one comprehensive metric (Cude 2001). For each of these eight subindices, the analytical measurement is converted into a quality rating between 10 (worst case) and 100 (ideal); ODEQ then uses a harmonic square mean formula:

$$OWQI = \sqrt{\frac{S}{\sum_{s=1}^{S} \frac{1}{SI_s^2}}} \tag{4.5}$$

where SI_s refers to subindex s in subindices 1 to S (e.g., pH level), to compute the OWQI score. In this method, the most impaired variable imparts the greatest influence on overall index score (Cude 2001), which thus provides a holistic measure of general water quality (since a site cannot have a relatively high overall score if it performs poorly on any metric). It is important to note that the OWQI is best suited as a comparative metric; not only is the index calibrated specifically for streams in the state of Oregon, but it does not reflect the suitability of water for specific uses (e.g., fishing or swimming). It does, however, provide an excellent synopsis of overall water quality that allows for comparison of observations from across the state over time. In particular, the OWQI is designed to facilitate comparisons between watersheds, and thus sub-indices such as pH and total solids are adjusted to account for geologic variability (Cude 2001).

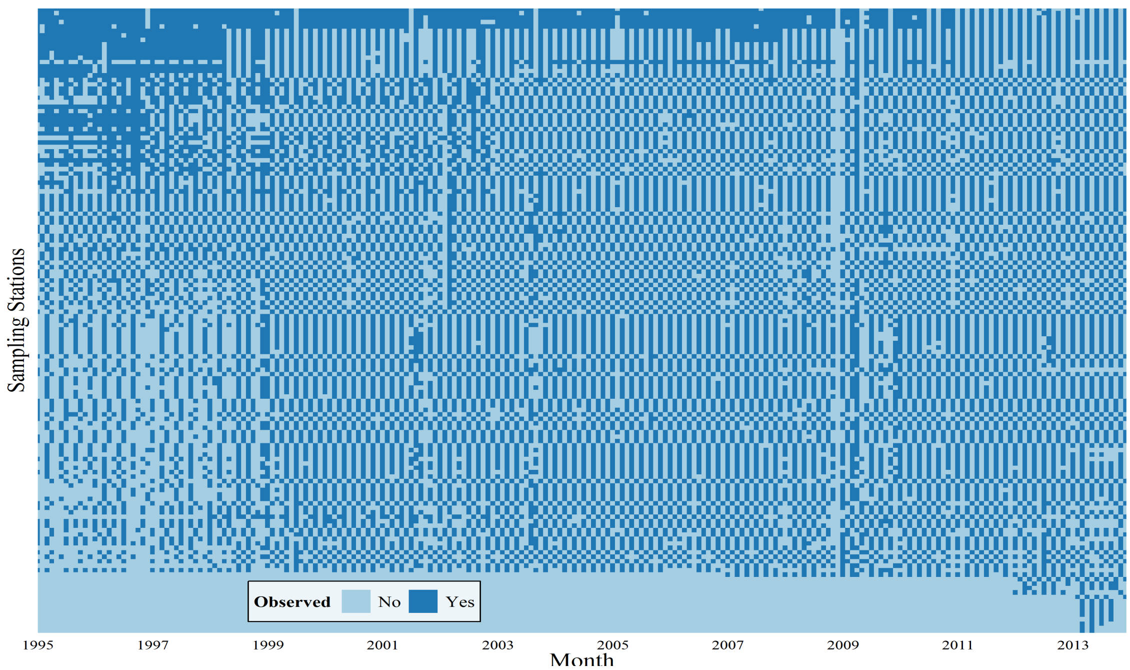


Figure 4.1: *Observations by sampling station: January 1995 to December 2013*

OWQI scores are observed on an intermittent monthly basis from 1990 to 2013 at 141 monitoring stations. Generally, in the data obtained from ODEQ index scores are tabulated every other month. For instance, some stations have observations in June and August while others have observations in July and September. In a few cases, however, there are only three or four total monthly observations in total at a given site. One major empirical advantage of the INLA hierarchical model and the random effects approach I use is that it can readily handle this type of irregular data. Thus, there is no need to drop any data or to impute data for under-sampled sites or unsampled months. In total, there are 16,676 unique observations. Figure 4.1 shows how many observations are included for each of the 141 sample sites; each dot represents a given month between January 1995 and December 2013. Sites are ordered along the Y-axis from the site with the fewest observations to the site with the most observations.

Independent Variables

The independent variables of interest are obtained from the Oregon Watershed Enhancement Board (OWEB). The grant database received from OWEB contains record of all grant projects funded by OWEB, including location (watershed), project start and end dates, project type, and funding amount; this analysis specifically considers 2509 grants given to watershed councils throughout the state between 1997 and 2013. Table 4.1 summarizes the nature of these grants:

Table 4.1: Grant summary statistics

	N	Avg. Amount ($)	Avg. Length (months)
Monitoring/Assessment/Tech. Assistance	579	41972.26	24.79
Council Support	372	104154.25	26.37
Education/Outreach	159	23515.76	17.24
Restoration	1384	69498.97	24.91
All Grants	2494	65345.99	24.61

While in some cases, a grant is targeted at a specific site (as is often the case for restoration grants), many OWEB grants have a more disparate spatial focus. For instance, an outreach project might be targeted at an entire watershed or basin. To model the spatial focus of grants, grants are aggregated by HUC8 for each month (228 months in total from 1995 to 2013). The total value of a grant is divided by the length of the project in months to calculate a monthly value of each grant; for instance, a $60,000 grant project started in March, 2001 and ending in May, 2001 results in a value of $20,000 for each of March, April, and May. While it is unlikely that grant funds are expended uniformly in practice, the "true" model of fund distribution is unclear from these data. The temporal distribution of funds also likely varies greatly between projects in any case. Assuming uniform monthly expenditure across the life of the project provides a consistent approach that is simple to interpret. This method also provides a consistent treatment of projects ongoing as of December, 2013, allocating funds in proportion to the portion of the project timeline that has already passed.

Measuring the impact of grants given to watershed councils poses several significant issues. The causal impact of a grant might not necessarily correspond directly to the actual project period. For instance, a restoration grant used to restore streamside riparian areas should have an ongoing impact on stream turbidity by reducing erosion for years after the project is finished. An OWEB education or outreach grant likewise is expected to have an ongoing impact on stakeholder behavior in the watershed. Of course, in the absence of ongoing maintenance efforts, the effect of restoration or outreach actions likely dissipate or diminish over time. In other words, recording the cumulative total of all grant funding is also inappropriate. To accommodate both the potential for effects that last beyond the actual project duration and for effects that diminish over time, I specify grant funding (for each category of grant) as a rolling sum of active funds for three different time periods: one year, three years, and five years.[3] For instance, for the five-year metric, active grant funds in a given watershed in January 2005 are the sum of all monthly grant funds in that watershed between January 2000 and January 2005.

Covariates

One of the advantages of the spatio-temporal model is that accounting for temporal and spatial relationships between observations serves to address many of the factors that affect water quality. For instance, more or less rainfall than is typical in a given year might affect water quality scores by increasing or reducing water levels; grouping observations

[3]Note that there is a considerable literature speaking to how long it takes for collaborative management efforts to begin to have an impact following inception. For instance, Leach et al. (2002) find that watershed partnerships formed for purposes such as restoration, education, or monitoring often take about 48 months to gain traction. However, what Leach et al. (2002) are in effect referring to is the time it takes for groups to complete projects (particularly how long it takes for a newly formed watershed council to begin to gain traction). In the case of projects funded by OWEB, each grant-funded project has an inception date and a completion date; thus, there is a clear point at which the project, for better, or worse, is finished, and there is no need to apply a temporal lag as would be the case if one were testing the link between watershed council formation and water quality.

by time accounts for this sort of variation. Likewise, the HUC8 random effects describe above account for local characteristics that differ across watersheds, such as management institutions or population. Several additional covariates are included in the model, however, in order to address factors that are not accounted for by controlling for the relative location of an observation in space and time.

First, the western portion of Oregon in between the Pacific Ocean and the Cascade Mountains receives much more rainfall than does the eastern portion and has a much different climate. The distance between a given sample site and the coast makes a considerable difference for observations. Thus, the model includes a variable measuring the Euclidean distance between the sample site and the Oregon coast (mapped using geodata obtained from the Oregon Geospatial Data Library).

Further, land use and land cover are well established as key drivers of water quality (Tong and Chen 2002; Meador and Goldstein 2003). Agricultural land usage is linked to increased water chemical content (Skaggs et al. 1994; Johnes and Heathwaite 1997). Waterways near developed land also demonstrate higher pollution levels (Wang 2001) (e.g., cars that leak oil onto pavement, which then washes into streams). To calculate the proportion of each HUC8 watershed that this used for agricultural purpose or that is developed, I use 30-meter by 30-meter raster (pixel image) data from the National Land Cover Database (NLCD). The NLCD includes national land cover data for 1992, 2001, 2006, and 2011. Using ArcGIS, I first produce a binary "True/False" raster for each land cover type (agricultural land [cropland or pasture], wetlands,[4] forest, and developed land), where each 30-meter by 30-meter pixel is coded as a "1" if that pixel in the comprehensive NLCD raster corresponds to the designated land cover type, and a "0" otherwise. I then use the *sp* (Pebesma and Bivand 2014) and *raster* (Hijmans 2014) packages in R to calculate the mean pixel value of all pixels within a 1000-meter radius surrounding each sample site for each land cover type. The buffer zone land cover values calculated using the 1992 NLCD are then matched to water quality

[4] Wetlands are not included in the model specifications below, as wetlands were not found to be a significant model predictor for any fitted model.

observations from 1995-2001, the 2001 NLCD data to water quality observations from 2002 to 2005, the 2006 NLCD data to water quality observations from 2006 to 2010, and the 2011 NLCD data to water quality observations from 2010 to 2013. While it would be ideal to have more fine-grained land cover data, perhaps on a yearly or quarterly basis, the NLCD data in this case satisfy the purpose at hand; the absolute values are less important than having a consistent way to account for relative differences in land usage amongst watersheds. I also use elevation data obtained from the CGIAR Consortium for Spatial Information using the R *raster* package (Hijmans 2014) to control for the elevation of the sample site.

MODEL FITTING AND SELECTION

Before presenting results specifically pertaining to my hypotheses, I describe how I fit these data into the Bayesian hierarchical model and identify the best-fitting specification. All models are estimated using R (R Core Team, 2013) and the INLA package developed by Rue et al. (2009). As described above, Bayesian models estimation posterior parameter distributions; this implies that each parameter also has a prior distribution. Since there are no "prior data" in this case, prior estimates are specified vaguely using the default recommended INLA settings (Rue et al. 2009) and are said to be "non-informative priors" (Gelman et al. 2013). This means that posterior estimates are almost wholly generated in light of the data (i.e., the priors have little effect on the posterior estimates). The spatial mesh used for the SPDE approach is shown in Figure 4.2:

Water quality sampling stations can be used, but do not need to be, as triangle vertices in the mesh (Lindgren et al. 2011). The mesh is more finely grained in areas where there are more water quality sampling stations; larger triangles represent areas with little or no information (Cosandey-Godin et al. 2014). Essentially, this serves so that the model estimates the field with increased accuracy where there are sufficient data, and conversely does not attempt to model the spatial random effect with great detail where there are no data. This model feature is important, since as Figure 4.2 demonstrates, there fewer water quality sampling stations in much of eastern Oregon. This is due to the fact that there is much less

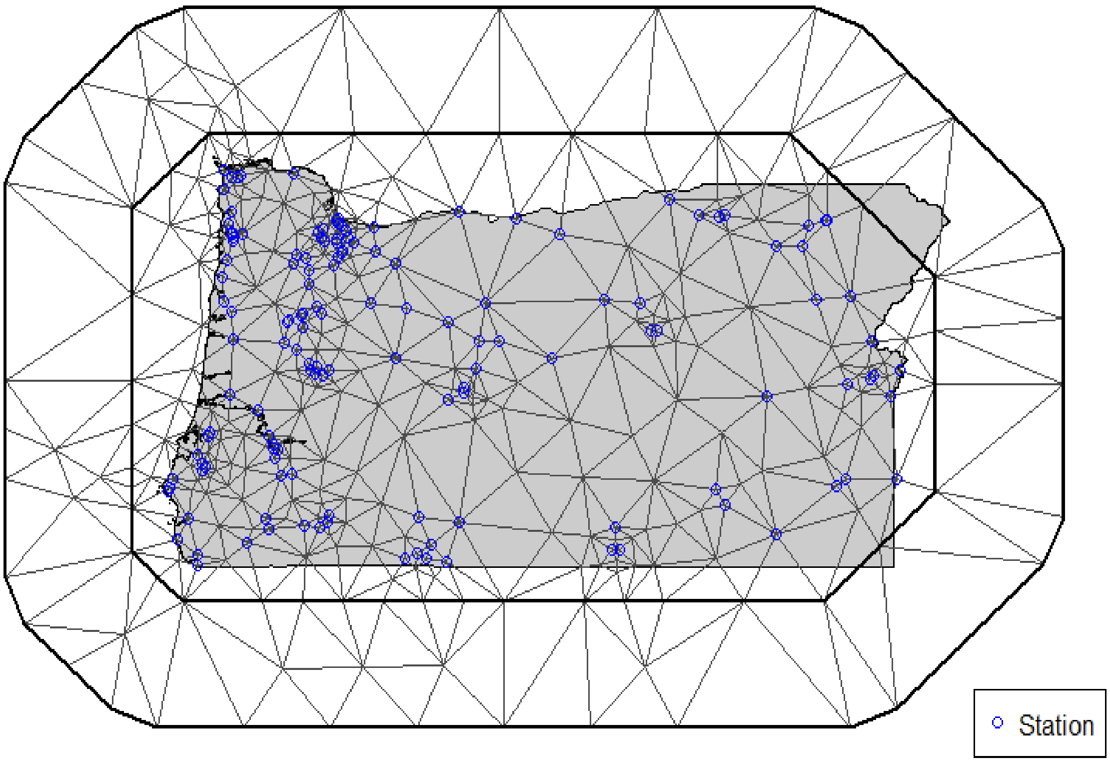


Figure 4.2: *Triangular mesh used for SPDE model*

precipitation and fewer streams in the eastern portion of the state. Conversely, the western part of the state is highly concentrated with sample sites, reflecting the much higher density of streams and rivers in this region. Figure 4.2 also shows that the mesh is extended beyond the boundaries of the sample area; this is done to avoid boundary effects wherein there is increased variance at the borders of a spatial field (since locations near the field border have fewer surrounding values) (Lindgren and Rue 2013).

There are several possible methods by which to model spatio-temporal correlation posed by Cosandey-Godin et al. (2014) that I compare: (Model 0) no spatial correlation; (Model 1) spatial correlation that is constant over time; (Model 2) spatial correlation that has a different realization for each year; (Model 3) spatial correlation that is itself correlated in consecutive years; (Model 4) spatial correlation that is repeated amongst years, so that the correlation between 2000 and 2001 is the same as that between 2000 and 2007; (Model 5) spatial correlation that has a different realization for each month; (Model 6) spatial correlation in which consecutive months in the same year are correlated; and (Model 7) spatial correlation that is repeated amongst months in a given year, so that the correlation between March 2007 and April 2007 is the same as that between March 2007 and September 2007). Table 4.2 presents the results associated with each of these models, as well as a model without an SPDE component, fit without the grant funding data. I select the optimal spatio-temporal model structure based upon Deviance Information Criterion (DIC) scores, which are a version of the traditional Akaike Information Criterion (AIC) score adapted to better suite hierarchical Bayesian models (Ward 2008; Spiegelhalter et al. 2002). As with AIC (and Bayesian Information Criterion [BIC]) scores, lower DIC scores indicate a better-fitting model:

Table 4.2: Compare DIC scores for different SPDE models

	M0	M1	M2	M3	M4	M5	M6	M7
DIC score:	$2,765$	$2,048$	$2,227$	-23	50	$2,068$	$1,637$	$1,686$

As Table 4.2 demonstrates, Model 3 has the smallest DIC value by far, indicating that it provides the optimal method for modeling spatial correlation over time. Model three fits a separate spatial correlation for each year (1995 through 2013), but assumes that the spatial correlation between consecutive years is itself correlated. Relative to the some of the other specifications, this makes a great deal of intuitive sense. First, "spatial correlation" refers to the underlying spatial process assumed to present; while conditions at different sites might vary considerably, the role that spatial distance plays in these conditions likely remains relatively constant. Thus, it makes sense to model all months in a given year as having the same spatial correlation function. Depending on broader ecological, economic, or political changes, however, one might also expect to see the role of spatial relationships change at least somewhat over 20 year period, as is included in this analysis. Assuming that consecutive years are correlated (i.e., spatial correlation in one year is very much like that of the following year), but allowing spatial correlation to change over time accounts for such long-term, incremental changes. As discussed above, the model also includes terms to capture temporal trends and seasonal fluctuation. Figure 3 shows the smoothed time trend and seasonal trend fitted as part of Model 3. It is interesting to note that overall, water quality appears to fluctuate between 1995 and 2013 but there is no discernible upward or downward trend.

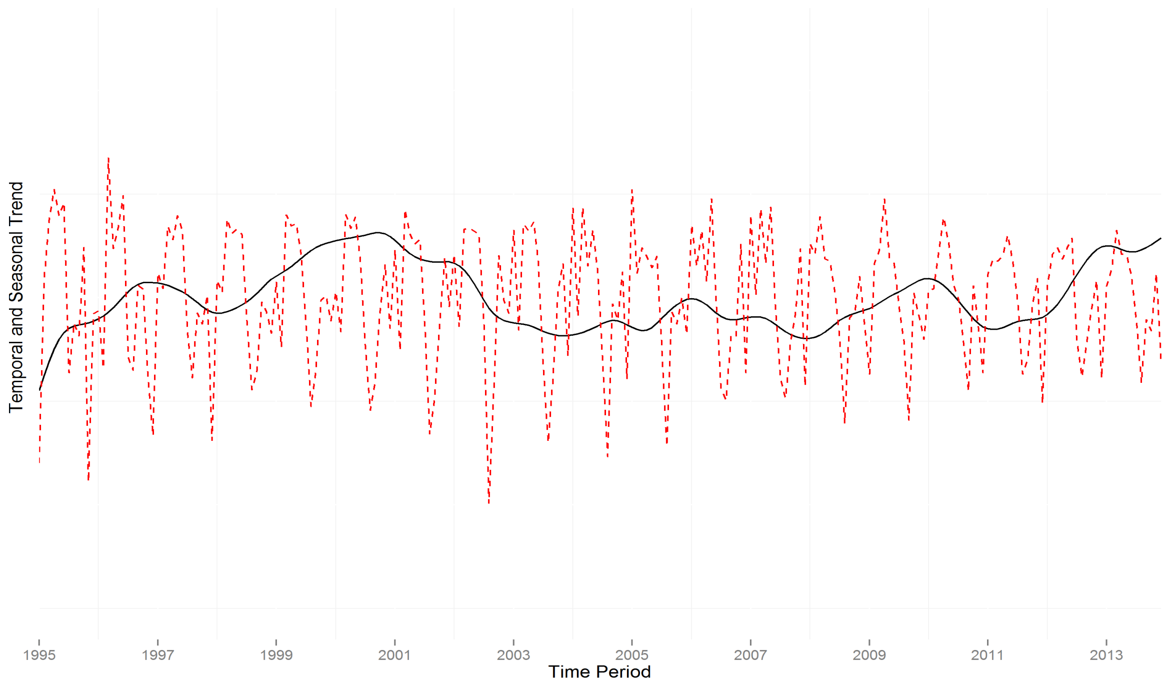


Figure 4.3: *Smoothed temporal trend and seasonality terms*

Having identified the spatio-temporal correlation structure that best fits these data, I now proceed to fit grant funding into the model. First, I model the predicted impact of all grant funds, regardless of type, computed using rolling 1-year, 3-year, and 5-year sums (e.g., where the 3-year sum represents the total grant funds provided in a given watershed for the 36 months prior to the water quality observation). I use the 1-year interval because water quality is typically considered in terms of a 12-month "water year," and the 3-year and 5-year intervals to potentially capture project impacts realized on a more long term basis. While as mentioned above there are few data that explicitly concern the effects of collaborative management on water quality, Lubell et al. (2009, p. 281) note that the perceived effectiveness of collaborative efforts increases with time (see also Leach et al. 2002; Leach 2006)

Table 4.3 shows the results of each of these three models. Since the dependent variable, the Oregon WQI score, is log-transformed, each coefficient from the model is interpreted as $log(OWQI) = \alpha + \beta x$, so that a one unit increase in x changes $log(OWQI)$ by β. An easier way to interpret these coefficients then is to exponentiate each coefficient so that it can be interpreted directly in terms of OWQI score where OWQI increases by a factor of $\exp^{\beta}$. This multiplicative coefficient is how terms in Table 4.3 are presented; a coefficient greater than one indicates an increase in OWQI score, and a coefficient less than one indicates a decrease in OWQI score. As described in previously, Bayesian models produce estimates of the posterior distribution for each parameter; thus, what Table 4.3 presents is the quantile values that encompass 95% of the posterior distribution for each parameter. This is somewhat analogous to the confidence interval derived from a Frequentist approach; since the exponentiated coefficients are interpreted multiplicatively, the primary concern is whether these quantile bounds span 1 (because when $\beta = 1$, $x * \beta = x$, indicating no predicted impact). Bounds that do not span one indicate a "significant" parameter, i.e., one that is statistically unlikely to be equal to zero. The DIC scores for each model in Table 4.3 are very similar, as is expected given the similarity between the 1-year, 3-year, and 5-year metrics; each model has a lower DIC score than Model 3 above, evidencing that the addition of grant funding does improve model fit. Note that each model in Table 4.3 also includes an intercept and random effect terms for HUC8 watershed, time, season, and space.

One important issue to note with regards to Table 4.3 is that, as the reader will recognize, the predicted effects of even well-known influencers of water quality such as agricultural land usage are minimal. Similarly, while elevation and distance from the Pacific Ocean are expected to be highly predictive of water quality, these coefficients are shown to be unim-

Table 4.3: Posterior parameter estimates, all grant types

	1 year sum	3 year sum	5 year sum
% Agric. (100m buffer)	0.999 (0.998, 0.999)	0.999 (0.998, 0.999)	0.999 (0.998, 0.999)
% Forest (100m buffer)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
% Devel. (100m buffer)	1.000 (0.999, 1.000)	1.000 (0.999, 1.000)	1.000 (0.999, 1.000)
Elevation (10m)	1.015 (0.996, 1.034)	1.015 (0.996, 1.034)	1.015 (0.996, 1.034)
Dist. from coast (10km)	0.991 (0.986, 0.996)	0.991 (0.986, 0.996)	0.991 (0.986, 0.996)
12 month total ($100k)	1.004 (1.001, 1.006)		
36 month total ($100k)		1.001 (1.001, 1.002)	
60 month total ($100k)			1.001 (1.000, 1.002)

portant in this model. The reason for this is that when one controls for spatial correlation, as in the SPDE model, spatial correlation is not a phenomenon in and of itself, but rather is a proxy for covariates that vary geographically such as elevation or land usage. Thus, these fixed effects are blurred out due to“spatial confounding” (Hodges 2014). Likewise, fitting a smoothed temporal term accounts not only long term climatic and environmental changes, but also for social, political, and economic changes that might otherwise be represented by time-varying covariates such as land usage. Simple hierarchical models, particularly fit without the SPDE method, show each of these covariates to be highly influential for water quality and of an expected sign (i.e., development and agriculture are negatively linked to water quality, forest and elevation are positively linked to water quality). It is interesting to note that developed land is negatively linked to water quality in the non-spatial model, but positively linked to water quality once spatial correlation is factored in, though why this is the case is unclear. Regardless of covariate behavior, the models that account for spatial correlation explicitly are the best predictive models (as shown by comparing Models 1 through 7 with Model 0 in Table 4.2).

The implication of this issue is that controlling for spatial correlation makes it very difficult to tease out the impact of any variable that is itself spatially distributed. This includes not only the covariates discussed above, but also grant funds that are distributed to particular watersheds and not others. In light of this, it is noteworthy that Table 4.3 does show that increased grant funding is predictive of improved water quality. This association is strongest for grant funds expended in the 12 months prior to the sample, where the middle 95% of the posterior distribution is between 1.001 and 1.008. In other words, net of all other terms in the model, a $100,000 in grant funds received in the past year predicts a 0.1% to 0.8% increase in water quality index score. This supports Hypothesis 1, that providing increased grant funding to collaborative watershed councils will be associated with improvements in water quality. While magnitude of this change might seem very small at first glance, given: (a) the complexity of factors that influence water quality; and (b) the indirect linkages between grant-funded projects and conditions at sample observation sites, it is noteworthy that the model identifies this relationship.

The terms associated with 3 years and 5 years of prior grant receipts are also positive, but the posterior distribution spans 1.000. This speaks to the question raised in the data and modeling section concerning the temporal impact of grant projects. Table 4.3 indicates that the predicted impact of grant funds (considering all grants in total) are most pronounced within 12 months of expenditure, and dissipate somewhat over time. It is important to note that this does not necessarily mean that projects have only a short term impact, but rather likely speaks the complexity of measuring environmental policy impacts more generally: the further in time an observation is from a given policy action, the more difficult it is to disentangle policy impacts from ecological trends and other drivers. I continue to explore the temporal nature of grant funds in the context of specific grant types below. I break down grant funds into four categories: restoration, scientific and technical (grants labeled by OWEB as assessment, monitoring, or technical assistance), and education and outreach (grants labeled by OWEB as either and outreach or and educational project). Given that OWEB funds are allocated competitively, it is unsurprising that different grant types are somewhat correlated. Table 4.4 presents the correlation for 12-month, 36-month, and 60-month rolling grant sums. A more substantive concern, however, is that OWEB uses different grant types as substitutes or else allocates specific grants in conjunction or succession; to the extent that activities sponsored by different grant types are not truly targeting different aspects of the problem but rather are aimed at the same things, comparing different council actions becomes more problematic. To examine this correlation between grant types, I tested numerous interaction specifications, including all two-way interaction terms and a full four-way interaction term (interacting the current funding levels for all four grant types). All interaction terms were negligible, and the non-interaction coefficients varied minimally across the different specifications. This indicates that interaction amongst grant types is not a significant concern in this case, and that the predicted impacts of different council actions (using grant type as a proxy for action) can be viably compared.

Table 4.4: Correlation amongst grant types (12/36/60 month values)

	Science/Tech.	Ed./Outreach	Rest.	Council
Science/Tech.				
Ed./Outreach	0.26/0.28/0.34			
Rest.	0.64/0.77/0.82	0.19/0.24/0.30		
Council	0.60/0.70/0.75	0.40/0.46/0.51	0.64/0.67/0.69	

Table 4.5 presents the results for the model with the full four-way interaction terms. Table 4.5 shows that the mean parameter value for each type of grant funding is positive, indicating (as expected) that grant funds of all kinds improve the impacts of watershed councils. For the 12-month rolling sum, all grants but restoration grants have a posterior parameter distribution for which the middle 95% of the distribution does not encompass 1.000 (again bearing in mind that since water quality is log-transformed, each coefficient is exponentiated and interpreted as a multiplicative factor). A \$100k increase in education and outreach grant funding within 12 months prior to the water quality observation predicts a 8.4% increase in water quality at the mean parameter value; scientific and technical support is associated with a 2.2% increase in water quality, while council administrative support predicts a 2.9% increase.

Using the 36-month rolling sum as a measure of grant funds, all three non-restoration

Table 4.5: Comparing Grant Effects by Type

Grant ($100k)	$\hat{\beta_{12m}}(0.025, 0.975)$	$\hat{\beta_{36m}}(0.025, 0.975)$	$\hat{\beta_{60m}}(0.025, 0.975)$
Restoration	1.003 (1.000, 1.006)	1.001 (1.000, 1.002)	1.001 (1.000, 1.002)
Ed./Outreach	1.084 (1.043, 1.126)	1.038 (1.023, 1.054)	1.028 (1.017, 1.038)
Scientific/Tech.	1.022 (1.008, 1.035)	1.011 (1.005, 1.017)	1.008 (1.004, 1.012)
Council Support	1.029 (1.013, 1.046)	1.011 (1.005, 1.017)	1.007 (1.003, 1.011)
Interaction	0.995 (0.991, 1.000)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)

grant types again have posterior distributions where the middle 95% does not include 1.000. A $100k increase in education and outreach funding predicts on average a 3.8% improvement, a $100k increase in scientific and technical support or in council support both predict a 1.1% increase. The results for 60-month grant totals are similar, with a $100k increase in education/outreach, scientific/technical, or council support associated with water quality improvements of 2.8%, 0.8%, and 0.7%, respectively. In the section below, I discuss the implications of these results. Note that for the 12-month grant totals, the interaction term suggests that having multiple types of grants active at the same time slightly reduces the overall predicted impact of grant funding on water quality (by about 0.5% per $100k). However, the interaction terms at the 36-month and 60-month totals have both mean estimates and posterior bounds that round to 1.000, indicating that there is little-to-no differential impact associated with having multiple types of grants active over longer periods of time.

DISCUSSION AND CONCLUSION

Table 4.5 reveals an interesting pattern: the coefficient means decrease in magnitude (or in two cases remain constant) as the time window for which grants are summed increases, but the posterior distributions become more leptokurtic (i.e., narrower). In other words, as the time window increases, the environmental impact of all grants are estimated to be positive with greater certainty, but the average impact is estimated to lessen. One likely reason for the reduced uncertainty regarding estimated impacts is that the increased time window provides a more consistent measure of ongoing council efforts. These data assume a uniform monthly expenditure profile for the life of every grant type. In practice, different projects (even within a single grant type) certainly have different implementation and output profiles, both during and after the project period. Thus, 36-month and 60-month post expenditure time windows likely best capture all grant expenditures at "full strength" (i.e., we can be more certain that each project has run its course), explaining the narrower posterior distributions. The reason why the mean estimates decrease as the time window increases in Table 4.5 (and also in Table 4.3 for all grants combined) is likely due to the data and measurement approach I employ. Water quality is a product of myriad complex factors, and so it follows that the magnitude of any predicted grant impact should decrease over time simply because there is greater opportunity for intervening factors. While it is possible that the impacts of certain projects do not diminish (or perhaps even increase) over time, from a simple cause-and-effect standpoint the 36-month and 60-month time windows place greater emphasis on projects much more distant to the actual water quality observation.

Measurement challenges also likely account for the fact that I do not identify a significant measurable impact from spending on restoration projects, in spite of the preponderance of ecological and environmental science demonstrating that restoration grants have an environmental impact to one extent or another (assuming the project is conducted). These data were not originally intended to serve the purpose of program evaluation. To best evaluate the efficacy of a grant, one would need to follow up directly with an evaluation protocol that explicitly monitors the outputs of grant-funded projects. However, such data collection efforts would obviously be very costly and time-intensive; OWEB is not able to track impacts in such detail or using comprehensive metrics that allow for comparison across the state. In lieu of direct evaluation data, I use existing Oregon water quality sampling stations, which are selected independently of project sites. This best explains why I find the reverse of what I hypothesize in Hypothesis 2 and do not find a measurable impact from restoration grants. Identifying a measurable impact from restoration projects is particularly challenging in this case because I am essentially selecting a stream point within an HUC8 watershed at random and testing whether spending on restoration projects within that same watershed is associated with water quality at that site. If the sample site is near to the actual restoration project site, then the impact might be clear; if the sample site is nowhere near the project site, then measuring the impact is highly unlikely.

Conversely, these data are likely better suited for evaluating non-restoration projects, since the nonpoint nature of many of these projects (in that restoration occurs at a specific site, whereas an outreach project for instance attempts to influence stakeholder behavior at many sites in the watershed) means the project impact has a broader spatial distribution in practice. For instance, an outreach project might influence various individuals throughout the watershed, and funds supporting council operations might be used to provide tools and services that serve people and organizations throughout the watershed. A randomly selected water quality sample site might be more likely to be nearby a "project site" in this case because there are presumably more "sites" involved. Generally speaking, the sampling approach of the OWQI lends itself towards evaluating broader management efforts, not wholesale changes in specific areas. In fact, the OWQI itself is billed explicitly as a tool for evaluating water quality management niveness (Cude 2001). While the impact of a restoration project might be very pronounced at or near the project site, these impacts are the most difficult to assess using this model. Thus, while this analysis does support the idea that watershed council actions have a meaningful impact on water quality, it is important to emphasize that these results should not discredit the impact of restoration activities.

The primary implication of these findings are that government investment in collaborative management efforts, specifically providing funds to nonprofit watershed councils in this case, is associated with improvement can produce changes in environmental outcomes. Returning to the investment perspective advanced above, I find that OWEB's investment in the production of non-environmental outputs (e.g., educational programs, stakeholder meetings, monitoring and data collection) does appears to engender an environmental return. The extant literature described in this paper provides a great deal of theoretical support as to why governments might invest in local nonprofit collaborative management groups. As evidenced by the considerable funding governments provide to collaborative management bodies (e.g., Sabatier et al. 2005), many practitioners already assume the viability of this type of governance approach. These data provide some of the first available evidence quantifying such returns.

More broadly, this research demonstrates the use of hierarchical Bayesian spatio-temporal modeling as a means by which publicly available observational data produced by states and the Federal government for environmental monitoring purposes can be repurposed to test environmental policy theory and evaluate public environmental programs. While the results of this analysis speak to the tradeoffs amongst different types of environmental projects, the exploratory nature of this work is insufficient to provide direct guidance to policy makers about how they should best allocate limited resources. Nonetheless, this provides a basis for further inquiry. In particular, this analysis advances beyond the limiting "collaborative" versus "non-collaborative" dichotomy to examine a more interesting and relevant question: how should policymakers distribute resources between policies and programs that produce environmental outputs at a relatively fixed input-output ratio (e.g., restoration projects) and those that produce non-environmental outputs (e.g., meetings, educational programs, administrative support) with the potential for returns that exceed inputs (e.g., motivating landowners and resource users to take restorative actions or modify their environmental behaviors in ways that far outstrip the input level of the original outreach program)? Much additional work is needed to understand how these different types of policies ultimately relate to environmental outcomes.

Appendix A

AIPW SPECIFICATION AND PROPENSITY SCORE BALANCE

The Average Treatment Effect (ATE) is estimated using the AIPW via:

$$\widehat{ATE}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{X_iY_i}{\hat{\pi}(Z_i)} - \frac{(1-X_i)Y_i}{1-\hat{\pi}(Z_i)}\right) - \frac{(X_i - \hat{\pi}(Z_i))}{\hat{\pi}(Z_i)(1-\hat{\pi}(Z_i))} * \right.$$
$$\left.[(1-\hat{\pi}(Z_i)\hat{\mathbb{E}}(Y_i|X_i = 1, Z_i) + \hat{\pi}(Z_i\hat{\mathbb{E}}(Y_i|X_i = 0, Z_i)]\right\} \quad \text{(A1)}$$

where $\hat{\pi}(Z_i$ is the estimated propensity score given the set of control variables Z for site i, Y_i is the observed outcome, and X_i is the treatment variable for site i. Equation A1 builds upon the basic inverse propensity weight (IPW) estimator by adjusting for a weighted average of the two regression estimators.[1] Glynn and Quinn (2010, 41) show that $\widehat{ATE}_{AIPW}$ is a consistent estimator for ATE when either: (1) the propensity score model is correctly specified; or (2) the two outcome regression models are correctly specified (see also Scharfstein et al. 1999). This means that the estimate is "doubly robust" (Glynn and Quinn 2010; Bang and Robins 2005) to uncertainty about both the selection process and the outcome model. Since the empirical processes that drive the existence of collaborative watershed groups and water quality conditions are both complex, this is a significant advantage. Using the $\widehat{ATE}_{AIPW}$, I estimate the effect of an active collaborative group for each of the outcome metrics used in this analysis[2].

[1] The adjustment term is $[(1-\hat{\pi}(Z_i)\hat{\mathbb{E}}(Y_i|X_i = 1, Z_i) + \hat{\pi}(Z_i\hat{\mathbb{E}}(Y_i|X_i = 0, Z_i)]$, such that $\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{X_iY_i}{\hat{\pi}(Z_i)} - \frac{(1-X_i)Y_i}{1-\hat{\pi}(Z_i)}\right) - \frac{(X_i-\hat{\pi}(Z_i))}{\hat{\pi}(Z_i)(1-\hat{\pi}(Z_i))}\right\}$ is the basic $\widehat{ATE}_{IPW}$ estimator.

[2] Discussed below in the Data section, six outcome metrics are used in order to comprehensively assess environmental condition: nitrogen content, phosphorus content, turbidity, benthic community health index, in-stream habitat complexity, and riparian cover. Thus, 6 ATE estimates are generated.

While the AIPW estimator has many advantages, it also can exhibit an extremely high variance when weights applied to observations are very small or very large. This occurs when estimated propensity scores are close to 0 or 1 (since weights are derived from the inverse propensity scores). Pohlmeier et al. (2013) thus recommend a shrinkage method which stabilizes the treatment estimators by shrinking the propensity scores ($\hat{p}_i = Pr(Z_i = 1|X_i)$) towards the unconditional mean treatment value (see also Busso et al. 2014; Frolich 2004):

$$\hat{p}_i^2 = (1 - \lambda_i(n))\hat{p}_i + \lambda_i(n)\hat{D} \tag{A2}$$

where $\hat{D}$ is the mean treatment value and $\lambda_i(n)$, the "tuning" parameter used to shrink the scores, is specified as $\lambda_i(n) = 1/\sqrt{n}$ (Pohlmeier et al. 2013). This method reduces weight variance, thereby reducing variance in the ATE estimators as well (Pohlmeier et al. 2013).[3] This procedure results in ATE estimates nearly identical to those generated without tuning, but serves to greatly reduce the estimate standard errors.

[3]Other common techniques applied to very low or very high weights are: (1) to discard all weights above and below specified cutoffs; and (2) to truncate all weights above and below specified cutoffs. The former method discards potentially valuable information and sacrifices efficiency, and given the relatively small sample size for this analysis is not the best option. The latter method essentially shrinks only extreme parameters; the advantage of the shrinking method used in this analysis is that it applies a consistent procedure to all observations.

Appendix B

EXAMINING PROPENSITY SCORE BALANCE

Table B1: Comparison of covariates between treatment and control groups

	Group Active >= 4 years	No Group/Active < 4 years	p-value
% Developed	0.020	0.030	0.007**
% Agricultural	0.250	0.200	0.010*
Watershed Area	4,772.560	4,488.880	0.102
Pop. Density	21.660	44.070	0.178
Median Income	45,264.620	47,677.090	0.032*
NPDES Permits	6.810	10.820	0.009**
NPDES Enforce Ratio	0.690	0.760	0.651
Note:	***p < .001; **p < .01; *p < .05		

To examine covariate balance, Table B1 presents average values for the treatment and control groups for each variable included in the propensity score model. The third column presents the p-value resulting from a standard two-sample t-test comparing the mean values. Three substantive difference that do emerge are that: (1) control observations have, on average, a county population density twice as high as treatment observations; (2) control observations have, on average, about 4 more active NPDES permits within their HUC8 watershed than do treatment observations (since the variance of population densities amongst watersheds is quite high, the the t-test fails to identify a statistical significant difference between the treatment and control groups even though the average difference is quite high in substantive terms); and (3) watersheds with an active group have, on average, about 25% agricultural land, compared to 20% in the control group. This is consistent with predicted role of collaborative groups as primarily targeting non-point source pollution (Hardy and Koontz 2008; Hoornbeek et al. 2012; Koontz et al. 2004) (addressed in the Discussion and

Conclusion sections); watersheds with a higher level of NPDES permits are presumably those for which point source pollution is more significant issue, whereas watersheds with fewer NPDES permits (as well as lower population density and a higher percentage of agricultural land) are likely those for which non-point source pollution is a more significant driver of water quality. It is important to remember, however, that the reason we examine the degree to which the treatment and control groups are balanced on observables is due to concern that the two groups might differ in ways that are unobserved as well (differences in observables can obviously be controlled for in the model). In this case, it is highly plausible that the observables do sufficiently account for variation between the treatment and control groups, since land cover, development, and point source pollution permits provide a comprehensive reflection of local watershed characteristics. This is likely why coarser metrics such as voting records (tested but left out of the final models) are insignificant and do not improve model fit, since the aforementioned covariates do a better job of capturing local heterogeneity.

Since the AIPW estimator relies on observed covariates to estimate selection probability, it is also important that the distributions of each covariate are relatively balanced between the treatment and control groups; otherwise, the selection model lacks common support. Figure B shows that while there are more control observations overall, covariate frequency distributions for the treatment and control groups are highly similar.

While Table B1 shows that the mean covariate values differ somewhat between the treatment and control groups, Figure B demonstrates that the overall distributions for each covariate overlap nicely. While there are more control observations, the range and relative frequency of observations are highly similar between the two groups. Figure B shows a similar distribution for the actual propensity scores. As might be expected, the distribution of propensity scores for the control group are skewed slightly lower than the distribution of propensity scores for the treatment group; nonetheless, the overall distribution for each spans the same range. Also, note that there are very few scores close to 0 or 1; this is in part due to the application of the shrinkage method, and helps ensure stable estimation within the weighting process.

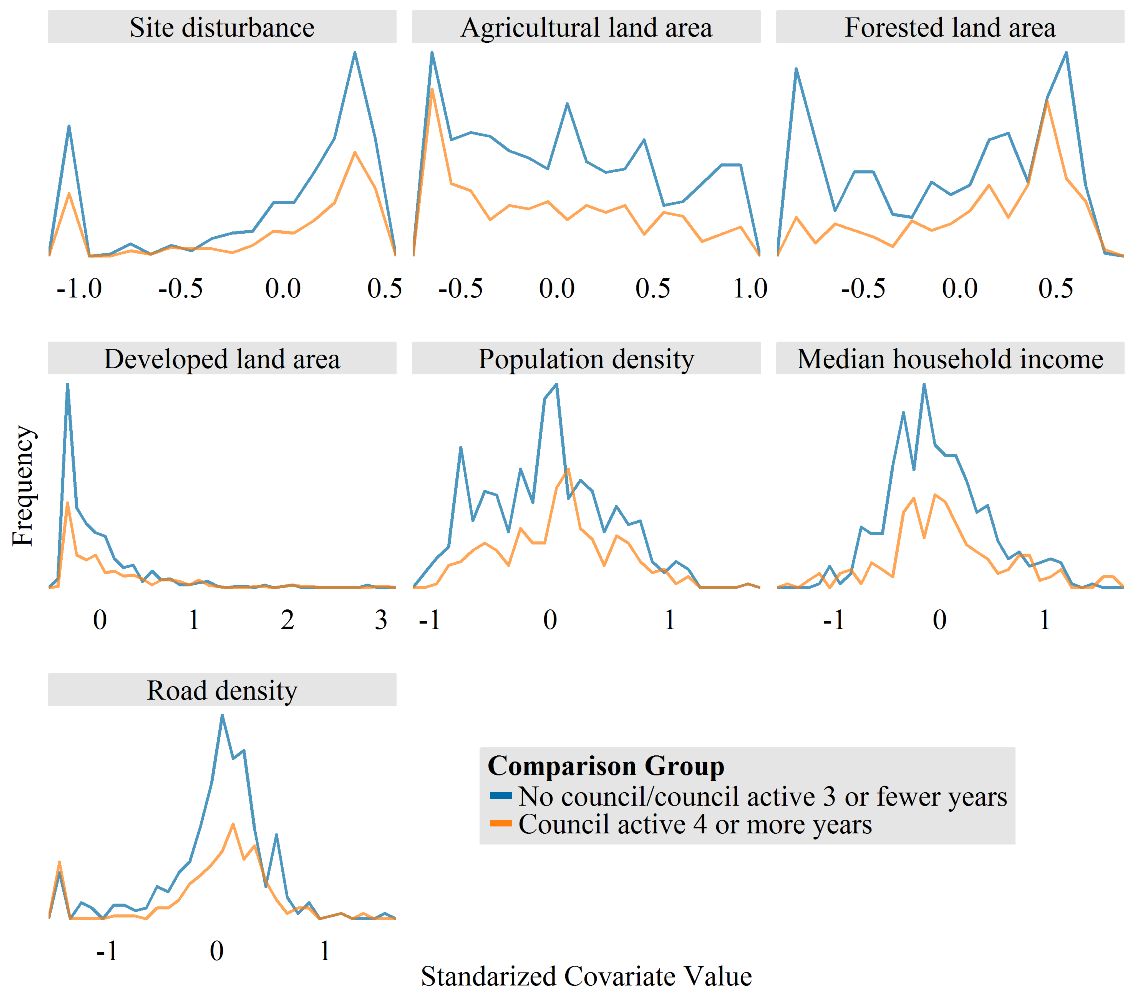


Figure B1: Covariate distributions, treatment and control groups

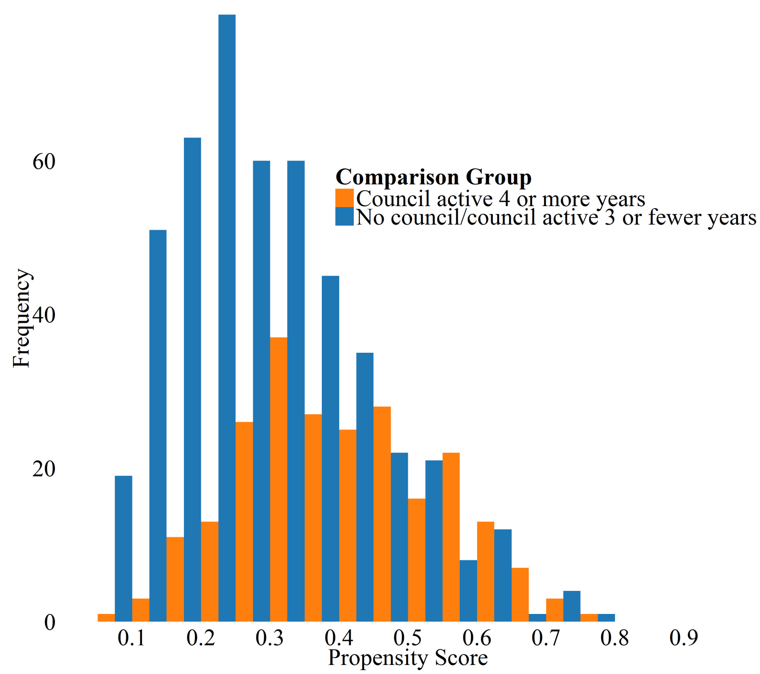


Figure B2: P-score distributions, treatment and control groups

Appendix C

EXAMPLE CODING PROTOCOL

Note: This process is applied iteratively across available documents. Since many groups have distinct documents that reference bylaws, membership, and funding, for instance, a group might initially receive a “0“ for each category of stakeholder representation when coding the bylaws document; these variables will then be updated to reflect new data subsequently produced by analyzing the membership roster.

Q1 Is this textual source an: (1) official group website; (2) annual group report; (3) group bylaw or charter document; (4) piece of authorizing legislation

If no → disregard If yes → proceed to Question 2

Q2 Does the textual source contain language that addresses a group’s purpose?

If no → proceed to Question 5 If yes → proceed to Question 3

Q3 Does the text speaking to a group’s purpose present an itemized set of purposes?

If no → code Objective Formalization as “MISSION STATEMENT" If yes → proceed to Question 4

Q4 Does the itemized set of purposes contain specific, measurable points of reference (e.g., “reduce total nitrogen level“ instead of “improve water quality")

If no → code Objective Formalization as “GOALS" If yes → code Objective Formalization as “OBJECTIVES"

Q5 Does the textual source contain language describing or listing group membership?

If no → proceed to Question 14 If yes → proceed to Question 6

Q6-Q13 Does text describing group membership list a tribe (or business, Federal agency, etc.) as a member of the group?

If no → code Tribal Representation as 0, proceed to next question If yes → code “Tribal Representation" as 1, proceed to next question —for Q6-12, proceed to next stakeholder type; for Q13, proceed to Q14 —

Q14 Does textual source contain specific reference to a group coordinator or facilitator?

If no → code COORDINATOR as 0, proceed to Q21 If yes → code COORDINATOR as 1, proceed to Q21

Q15: Does textual source identify year in which group was formed?

If no → proceed to Q22 If yes → code FORMATION YEAR as specified year

Q16-Q23: Does textual source contain language reference to group actions or responsibilities related to EDUCATION (e.g., group "runs environmental education programs in local schools")?

If no → proceed to next question If yes → code GROUP ACTIVITY as "education"

Q17 Outreach (e.g., group "reaches out to local farmers")

Q18 Coordination (e.g., group "provides forum where agencies can share information")

Q19 Monitoring (e.g., group "conducts ongoing monitoring of stream pollutants")

Q20 Projects (e.g., group is "conducting restoration on local creek")

Q21 Planning (e.g., group is "charged with developing comprehensive action plan")

Q22 Management (e.g., group is "lead local entity for water improvement program")

Q23 Permitting (e.g., group is "administers land use permits within the watershed")

If GROUP ACTIVITY is "Management" or "Permitting" code GROUP RESPONSIBILITY as "1" otherwise code GROUP RESPONSIBILITY as "0"

Appendix D
SURVEY INSTRUMENT DRAFT

Note: This survey instrument was implemented online using Novisurvey, an online survey design and hosting software. This appendix is a textual reproduction of the survey instrument; the survey as taken by respondents differed in appearance.

The Puget Sound Partnership (PSP) would appreciate your completion of this survey of the organizational networks you participate in for Puget Sound recovery. This survey is being conducted in partnership with the Evans School of Public Affairs at the University of Washington. The survey serves two purposes. One is to help the PSP generate network maps that will allow individuals involved in Puget Sound recovery efforts to see which networks are involved in specific recovery activities. The other purpose is to support applied research on the links between network activities and environmental outcomes.

For the purposes of the PSP, your names will be attached to your responses, since it is crucial for the PSP to understand existing networks. For the purposes of research at the University of Washington, your names will be separated from your responses, as they are not needed for academic research. Hence, confidentiality will be maintained in academic publications, but not within the PSP. None of the questions ask you to state opinions or disclose personal information other than your name, the organization you represent, and the organizational networks in which you participate.

If you have any questions about this survey, please contact Professor Craig Thomas at the University of Washington, 206-221-3669, thomasc@washington.edu, or Dave Ward at the Puget Sound Partnership, dave.ward@psp.wa.gov, 206-462-2275. If you have already completed the survey, you do not need to do so again and may exit the survey now. If you have not yet completed the survey, please click "next" to continue.

1. **What is your name?**

We would first like to gather some background information about the collaborative groups working on environmental recovery in the Puget Sound region that you participate in.

2. **Have you regularly participated in any of the following collaborative groups working on Puget Sound recovery in the past five years? Please check all that apply.**

☐ One or more local ECO Nets

☐ One or more Local Integrating Organization(s)

☐ The PSP Leadership Council

☐ The PSP Ecosystem Coordination Board

☐ The PSP Science Panel

☐ The PSP Social Science/Social Strategies Committee

☐ The Puget Sound Salmon Recovery Council

☐ The Puget Sound Recovery Implementation Technical Team

☐ The Puget Sound Watershed Leads Group

☐ One or more local salmon recovery or watershed group(s)

☐ The Puget Sound Federal Caucus

☐ The Puget Sound Environmental Caucus

☐ The Puget Sound Ecosystem Monitoring Program (PSEMP) Steering Committee

☐ One or more Puget Sound Ecosystem Monitoring Program (PSEMP) work group(s)

☐ The Puget Sound Institute

☐ The Puget Sound Nearshore Restoration Project (PSNERP)

☐ One or more county Marine Resource Committee(s)

☐ None of the above

3. **Which local ECO Net group(s) have you participated in within the last**

five years ? Please check all that apply

(question appears if corresponding choice is made for question 2)

☐ Hood Canal ECO Net

☐ King ECO Net

☐ Kitsap ECO Net

☐ Mason ECO net

☐ Pierce ECO Net

☐ San Juan ECO Net

☐ Skagit ECO Net

☐ Snohomish/Camano ECO Net

☐ Straits ECO Net

☐ Thurston ECO Net

☐ Whatcom ECO Net

☐ Whidbey ECO Net

☐ None of the above

4. **Which Puget Sound Ecosystem Monitoring Program (PSEMP) work group(s) have you participated in within the last five years? Please check all that apply.**

   (question appears if corresponding choice is made for question 2)

   ☐ ECM Birds and Mammals

   ☐ ECM Freshwater

   ☐ ECM Forage Fish/Food Webs

   ☐ ECM Marine Waters

   ☐ ECM Modeling

   ☐ ECM Nearshore

   ☐ ECM Salmon

☐ ECM Stormwater

☐ ECM Toxics

☐ None of the above

5. **Which Marine Resource Committee(s) have you participated in within the last five years? Please check all that apply.**

   (question appears if corresponding choice is made for question 2)

   ☐ Clallam County Marine Resource Committee

   ☐ Island County Marine Resource Committee

   ☐ Jefferson County Marine Resource Committee

   ☐ San Juan County Marine Resource Committee

   ☐ Skagit County Marine Resource Committee

   ☐ Snohomish County Marine Resource Committee

   ☐ Whatcom County Marine Resource Committee

   ☐ None of the above

6. **Which local salmon recovery or watershed group(s) have you participated in within the last five years? Please check all that apply.**

   (question appears if corresponding choice is made for question 2)

   ☐ Green/Duwamish Watershed

   ☐ Hood Canal Watershed

   ☐ Island Watershed

   ☐ Lake Washington/Cedar/Sammamish Watershed

   ☐ Nisqually River Watershed

   ☐ Nooksack Watershed

   ☐ North Olympic Peninsula Watershed

   ☐ Puyallup/White & Chambers/Clover Watersheds

☐ San Juan Watershed

☐ Skagit Watershed

☐ Snohomish Watershed

☐ South Sound Watersheds

☐ Stillaguamish Watershed

☐ West Sound Watersheds

☐ None of the above

7. **Which Local Integrating Organization(s) (LIO) have you participated in within the last five years? Please check all that apply.**

   (question appears if corresponding choice is made for question 2)

   ☐ San Juan County/Watershed LIO

   ☐ Whatcom County/Nooksack Watershed LIO

   ☐ Island County/Watershed LIO

   ☐ Snohomish and Stillaguamish Watersheds LIO

   ☐ South Central Action Area (WRIA 8, 9, 10) LIO

   ☐ South Sound Action Area (WRIA 11, 12, 13, 14) LIO

   ☐ Hood Canal Action Area LIO

   ☐ Strait Action Area LIO

   ☐ North Central Action Area (WRIA 15) LIO

   ☐ None of the above

8. **Have you regularly participated in any other collaborative group(s) focused on environmental recovery in the Puget Sound region within the last five years other than those that you previously identified?**

   ☐ Yes ☐ No

9. **Please specify the collaborative groups focused on environmental recovery in the Puget Sound region in which you have regularly participated within the last five years, other than those that you previously identified. You may list up to ten.**

The following section poses questions about the extent of your involvement in collaborative group(s) working on Puget Sound recovery. In some of the following questions, the name(s) of the collaborative group(s) in which you previously identified yourself as a participant are imported into rows that correspond to the response options.

10. **What organization do you primarily represent (work for, volunteer for, or are otherwise affiliated with) on the collaborative group(s) that you previously identified?**

11. **How many years have you worked at or volunteered for the organization that you identified above?**

12. **Do you represent any other organization(s) on the collaborative group(s) that you previously identified?**
    ☐ Yes ☐ No

13. **Please specify which organization you have primarily represented on each group that you previously identified.**
    (choice appears to list new organization for each listed collaborative group)

14. **Please describe the reason(s) that you have been involved in the collaborative group(s) that you previously identified. Please select one answer for**

each group you have participated in.

(choice appears for each listed collaborative group)

☐ Part of your job description

☐ Professional service

☐ Independent stakeholder

☐ None of these

15. **Please describe your regular level of involvement in the following group activities for the collaborative groups in which you participate. For each activity and group you participate in, please check the box if you regularly engage in the specified action.**

(choice appears for each listed collaborative group)

☐ Send or respond to group emails

☐ Attend group meetings

☐ Attend group events beyond group meetings (e.g., ECO Net summit)

☐ Participate in group projects or programs

☐ Read or review group reports or documents

☐ Produce group reports or documents

☐ Other types of group participation

16. **My participation in the following collaborative group(s) has increased the amount of face-to-face communication I engage in with other organizations.**

☐ Strongly disagree

☐ Disagree

☐ Neither agree nor disagree

☐ Agree

☐ Strongly agree

☐ Don't know

☐ NA

17. **My participation in the following collaborative group(s) has increased my awareness of the interests and values of other organizations and stakeholders.**

    ☐ Strongly disagree

    ☐ Disagree

    ☐ Neither agree nor disagree

    ☐ Agree

    ☐ Strongly agree

    ☐ Don't know

    ☐ NA

18. **My participation in the following collaborative group(s) has increased my understanding of commonly used language in the field (e.g., a common definition of "stormwater").**

    ☐ Strongly disagree

    ☐ Disagree

    ☐ Neither agree nor disagree

    ☐ Agree

    ☐ Strongly agree

    ☐ Don't know

    ☐ NA

19. **My participation in the following collaborative group(s) has increased**

my awareness of and/or access to scientific, technical, or policy-specific information.

☐ Strongly disagree

☐ Disagree

☐ Neither agree nor disagree

☐ Agree

☐ Strongly agree

☐ Don't know

☐ NA

20. **My participation in the following collaborative group(s) has increased my access to human resources (e.g., administrative support, IT services).**

☐ Strongly disagree

☐ Disagree

☐ Neither agree nor disagree

☐ Agree

☐ Strongly agree

☐ Don't know

☐ NA

21. **My participation in the following collaborative group(s) has increased my access to financial resources (e.g., grant opportunities).**

☐ Strongly disagree

☐ Disagree

☐ Neither agree nor disagree

☐ Agree

☐ Strongly agree

☐ Don't know

☐ NA

We would now like to ask several questions about the specific organizations (not the larger collaborative groups mentioned above) that you work with on a regular basis for the purposes of environmental protection, restoration, or education in the Puget Sound region. Each question below addresses a particular type of interaction with other organizations. These organizations may or may not be part of the collaborative group(s) you participate in.

22. **With which organizations do you routinely implement or formally support joint projects or programs (e.g., habitat restoration projects or outreach programs)? We would like to know the primary organizations that you routinely work with, not all of them; thus, you may list up to 5 organizations.**

23. **Did your work on joint projects or programs with these organizations begin within the past five years?**
    (choice appears for each listed organization)
    ☐ Yes ☐ No

24. **With which organizations do you routinely coordinate plans or strategies (e.g., setting environmental performance goals, project selection)? We would like to know the primary organizations that you routinely coordinate with, not all of them; thus, you may list up to 5 organizations.**

25. **Did your coordinated planning or strategizing with these organizations be-**

gin within the last five years?

(choice appears for each listed organization)

☐ Yes ☐ No

26. **Aside from the organizations with which you coordinate plans or strategies or implement joint projects, with which organizations do you informally consult on a routine basis (e.g., seek technical guidance, get feedback from local stakeholders)? Please list up to 5 primary organizations.**

27. **Did your informal consultation with these organizations begin within the last five years (e.g., seek technical guidance, get feedback from local stakeholders)?**

(choice appears for each listed organization)

☐ Yes ☐ No

The following section poses questions regarding whether the amount of informal consultation, coordinated planning, and joint project implementation you engage in with other organizations has increased or decreased in the last five years, as well as how useful these specific joint actions are. For each statement, please select the choice that you think is most accurate.

28. **In the last five years, have you implemented or formally supported less or more of the following types of projects or programs jointly with other organizations than you did previously? Please select one answer for each type of activity.**

Habitat restoration:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Research or development:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Land or resource management:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Education or outreach:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Monitoring or enforcement:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

29. **How useful is it for you to do the following types of projects or programs jointly with other organizations? Please select one answer for each type of activity.**

Habitat restoration:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Research or development:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Land or resource management:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Education or outreach:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Monitoring or enforcement:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

30. **In the last five years, have you engaged in less or more of the following types of planning or strategizing jointly with other organizations than you did previously? Please select one answer for each type of planning activity.**

Goal or target setting:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Project or program design:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Land or resource use planning:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Strategy development:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

31. **How useful is it for you to engage in the following types of planning or strategizing jointly with other organizations? Please select one answer for each type of planning activity.**

Goal or target setting:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Project or program design:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Land or resource use planning:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Strategy development:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

32. **In the last five years, have you engaged in less or more of the following types of informal consultation with other organizations than you did previously? Please select one answer for each type of consultation.**

Technical feedback:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Stakeholder input:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

Project/program implementation guidance:

☐ Much less

☐ Less

☐ Neither more nor less

☐ More

☐ Much more

☐ Don't know

☐ NA

33. **How useful is it for you to engage in the following types of informal consultation with other organizations? Please select one answer for each type of resource.**

Technical feedback:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Stakeholder input:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

Project/program implementation guidance:

☐ Not useful

☐ Somewhat useful

☐ Moderately useful

☐ Very useful

☐ Don't know

☐ NA

34. **Do you have any suggestions as to how the Puget Sound Partnership can help improve coordination or cooperation among organizations involved in Puget Sound recovery?**

BIBLIOGRAPHY

Adam, S., and H. Kriesi. 2007. "The network approach." *Theories of the policy process* 2:129–54.

Agranoff, R., and M. McGuire. 1998. "Multinetwork management: Collaboration and the hollow state in local economic policy." *Journal of Public Administration Research and Theory* 8 (1): 67–91.

— . 2001. "Big questions in public network management research." *Journal of public administration research and theory* 11 (3): 295–326.

— . 2003. "Inside the matrix: Integrating the paradigms of intergovernmental and network management." *International Journal of Public Administration* 26 (12): 1401–1422.

Alexander, Ernest R. 1993. "Interorganizational coordination: Theory and practice." *Journal of Planning Literature* 7 (4): 328–343.

— . 1995. *How Organizations Act Together: Interorganizational coordination in theory and practice*. New York, NY: Routledge.

Altheide, David L., and Christopher J. Schneider. 2012. *Qualitative Media Analysis*. 2nd. Qualitative Research Methods 38. Thousand Oaks, CA: Sage.

Altheide, David, Michael Coyle, Katie DeVriese, and Christopher Schneider. 2008. "Emergent qualitative document analysis." In *Handbook of emergent methods*, 127–151. New York, NY: Guilford Press.

Analytics, Revolution, and Steve Weston. 2014. *doParallel: Foreach parallel adaptor for the parallel package*. R package version 1.0.7.

Anderson, Larz T. 1995. *Guidelines For Preparing Urban Plans*. Chicago, IL: Planners Press, American Planning Association.

Anderson, Sarah E., Heather E. Hodges, and Terry L. Anderson. 2013. "Technical management in an age of openness: the political, public, and environmental forest ranger." *Journal of Policy Analysis and Management* 32 (3): 554–573.

Ansell, C., and A. Gash. 2008. "Collaborative governance in theory and practice." *Journal of Public Administration Research and Theory* 18 (4): 543–571.

Bang, Heejung, and James M. Robins. 2005. "Doubly robust estimation in missing data and causal inference models." *Biometrics* 61 (4): 962–973.

Bardach, E. 1998. *Getting agencies to work together: The practice and theory of managerial craftsmanship.* Washington, D.C.: Brookings Inst Press.

Bates, Douglas, Martin Maechler, Benjamin M. Bolker, and Steven Walker. 2014. "lme4: Linear mixed-effects models using Eigen and S4." ARXIV e-print; submitted to Journal of Statistical Software, *submitted to Journal of Statistical Software.*

Beguin, Julien, Sara Martino, Havard Rue, and Steven G. Cumming. 2012. "Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation." *Methods in Ecology and Evolution* 3 (5): 921–929.

Beierle, Thomas C. 2002. "The Quality of Stakeholder-Based Decisions." *Risk Analysis* 22 (4): 739–749.

Berardo, Ramiro, Tanya Heikkila, and Andrea K. Gerlak. 2014. "Interorganizational Engagement in Collaborative Environmental Management: Evidence from the South Florida Ecosystem Restoration Task Force." *Journal of Public Administration Research and Theory*:muu003.

Berardo, Ramiro, and John T. Scholz. 2010. "Self-Organizing Policy Networks: Risk, Partner Selection, and Cooperation in Estuaries." *American Journal of Political Science* 54 (3): 632–649.

Bernardo, José M., and Adrian FM Smith. 2009. *Bayesian theory.* Vol. 405. John Wiley & Sons.

Berry, F. S, R. S Brower, S. O Choi, W. X Goa, H. S Jang, M. Kwon, and J. Word. 2004. "Three traditions of network research: What the public management research agenda can learn from other research communities." *Public Administration Review* 64 (5): 539–552.

Bickman, Leonard. 1987. "The functions of program theory." *New Directions for Program Evaluation*, no. 33:5–18.

Bidwell, R. D, and C. M Ryan. 2006. "Collaborative partnership design: the implications of organizational affiliation for watershed partnerships." *Society and Natural Resources* 19 (9): 827–843.

Bingham, Lisa Blomgren, and Rosemary O'Leary. 2008. *Big Ideas in Collaborative Public Management.* Armonk, NY: M.E. Sharpe.

Bingham, Lisa B., David Fairman, Daniel J. Fiorino, and R. O'Leary. 2003. "Fulfilling the promise of environmental conflict resolution." In *The promise and performance of environmental conflict resolution*, ed. by Rosemary O'Leary and Lisa B. Bingham, 329–351. Washington, D.C.: RFF Press.

Bivand, Roger S., Edzer J. Pebesma, and Virgilio Gomez-Rubio. 2008. *Applied spatial data analysis with R.* Vol. 747248717. Springer.

Bivand, Roger S., Edzer Pebesma, and Virgilio Gomez-Rubio. 2013. "Disease Mapping." In *Applied Spatial Data Analysis with R*, 319–361. Springer.

Black, Dan A., and Jeffrey A. Smith. 2004. "How robust is the evidence on the effects of college quality? Evidence from matching." *Journal of Econometrics* 121 (1): 99–124.

Blangiardo, Marta, Sylvia Richardson, John Gulliver, and Anna Hansell. 2010. "A Bayesian analysis of the impact of air pollution episodes on cardio-respiratory hospital admissions in the Greater London area." *Statistical methods in medical research.*

Blangiardo, Marta, Michela Cameletti, Gianluca Baio, and Håvard Rue. 2013. "Spatial and spatio-temporal models with R-INLA." *Spatial and spatio-temporal epidemiology* 7:39–55.

Bodin, O, and B. I Crona. 2009. "The role of social networks in natural resource governance: What relational patterns make a difference?" *Global Environmental Change* 19 (3): 366–374.

Bolker, Benjamin M., Mollie E. Brooks, Connie J. Clark, Shane W. Geange, John R. Poulsen, M. Henry H. Stevens, and Jada-Simone S. White. 2009. "Generalized linear mixed models: a practical guide for ecology and evolution." *Trends in ecology & evolution* 24 (3): 127–135.

Borzel, Tanja A. 1998. "Organizing Babylon-On the Different Conceptions of Policy Networks." *Public administration* 76 (2): 253–273.

Bressers, H. 2009. "From public administration to policy networks: Contextual interaction analysis." *Rediscovering public law and public administration in comparative policy analysis: tribute to Peter Knoepfel*:123–142.

Bressers, J. T. A., and S. M. M. Kuks. 2003. "Managing water resources in Florida and the Netherlands."

Brooks, Steve, Andrew Gelman, Galin Jones, and Xiao-Li Meng. 2011. *Handbook of Markov Chain Monte Carlo*. CRC Press.

Bryson, John M., Barbara C. Crosby, and Melissa Middleton Stone. 2006. "The design and implementation of Cross-Sector collaborations: Propositions from the literature." *Public administration review* 66 (s1): 44–55.

Burby, R. J. 2003. "Making plans that matter: Citizen involvement and government action." *Journal of the American Planning Association* 69 (1): 33–49.

Busso, Matias, John DiNardo, and Justin McCrary. 2014. "New evidence on the finite sample properties of propensity score reweighting and matching estimators." *Review of Economics and Statistics*, no. In Press.

Butts, Carter T., Mark S. Handcock, and David R. Hunter. 2014. *network: Classes for Relational Data.* R package version 1.9.0. Irvine, CA.

Cameletti, Michela, Finn Lindgren, Daniel Simpson, and Håvard Rue. 2013. "Spatio-temporal modeling of particulate matter concentration through the SPDE approach." *AStA Advances in Statistical Analysis* 97 (2): 109–131.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: methods and applications.* Cambridge university press.

Carlson, C. 1999. "Convening." In *The Consensus Building Handbook: A comprehensive guide to reaching agreement.* Ed. by L Susskind, S. McKearnen, and J. Thomas-Lamar, 169–198. Thousand Oaks, CA: Sage.

Carlsson, Bo, and Staffan Jacobsson. 1997. "In Search of Useful Public Policies — Key Lessons and Issues for Policy Makers." In *Technological Systems and Industrial Dynamics*, ed. by Bo Carlsson, 10:299–315. Economics of Science, Technology and Innovation. Springer US.

Carlsson, L., and F. Berkes. 2005. "Co-management: concepts and methodological implications." *Journal of environmental management* 75 (1): 65–76.

Carr, Deborah S., Steven W. Selin, and Michael A. Schuett. 1998. "Managing public forests: Understanding the role of collaborative planning." *Environmental management* 22 (5): 767–776.

Carr, G., G. Bloschl, and D. P. Loucks. 2012. "Evaluating Participation in Water Resource Management: A review." *Water Resources Research* 48 (11).

Chaffin, B. C., R. L. Mahler, J. D. Wulfhorst, and B. Shafii. 2012. *Collaborative Watershed Groups in Three Pacific Northwest States: A Regional Evaluation of Group Metrics and Perceived Success1.* Wiley Online Library.

Cheng, A. S, and S. E Daniels. 2005. "Getting to "we": examining the relationship between geographic scale and ingroup emergence in collaborative watershed planning." *Human Ecology Review* 12 (1): 30–43.

Clark, James S. 2005. "Why environmental scientists are becoming Bayesians." *Ecology letters* 8 (1): 2–14.

Clark, James S., and Alan E. Gelfand. 2006. *Hierarchical Modelling for the Environmental Sciences: Statistical methods and applications: Statistical methods and applications*. Oxford University Press.

Coglianese, C. 1999. "The limits of consensus: the environmental protection system in transition: Toward a more desirable future." *Environment: Science and Policy for Sustainable Development* 41 (3): 28–33.

Coglianese, Cary. 1997. "Assessing Consensus: The Promise and Performance of Negotiated Rulemaking." *Duke Law Journal* 46 (6): 1255–1349.

— . 2003. "Is satisfaction success? Evaluating public participation in regulatory policymaking." In *The Promise and Performance of Environmental Conflict Resolution*, ed. by R. O'Leary and L. B. Bingham, 69–89. Washington, D.C.: RFF Press.

Collins, Kevin, Chris Blackmore, Dick Morris, and Drennan Watson. 2007. "A systemic approach to managing multiple perspectives and stakeholding in water catchments: some findings from three UK case studies." *Environmental Science & Policy* 10 (6): 564–574.

Connick, S., and J. E Innes. 2003. "Outcomes of collaborative water policy making: Applying complexity thinking to evaluation." *Journal of Environmental Planning and Management* 46 (2): 177–197.

Cosandey-Godin, Aurelie, Elias Teixeira Krainski, Boris Worm, and Joanna Mills Flemming. 2014. "Applying Bayesian spatio-temporal models to fisheries bycatch in the Canadian Arctic." *Canadian Journal of Fisheries and Aquatic Sciences*, no. ja.

Cragan, J. F, and D. W Wright. 1990. "Small group communication research of the 1980s: A synthesis and critique." *Communication Studies* 41 (3): 212–236.

Cranmer, Skyler J., and Bruce A. Desmarais. 2011. "Inferential network analysis with exponential random graph models." *Political Analysis* 19 (1): 66–86.

Cressie, Noel, and Christopher K. Wikle. 2011. *Statistics for spatio-temporal data.* John Wiley & Sons.

Cressie, Noel, Catherine A. Calder, James S. Clark, Jay M. Ver Hoef, and Christopher K. Wikle. 2009. "Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling." *Ecological Applications* 19 (3): 553–570.

Crispin, Val, Robert House, and Dave Roberts. 1993. "Changes in Instream Habitat, Large Woody Debris, and Salmon Habitat after the Restructuring of a Coastal Oregon Stream." *North American Journal of Fisheries Management* 13, no. 1 (): 96–102. doi:10.1577/1548-8675(1993)013<0096:CIIHLW>2.3.CO;2.

Crona, B., and K. Hubacek. 2010. "The right connections: how do social networks lubricate the machinery of natural resource governance?" *Ecology and Society* 15 (4): 18–22.

Cude, Curtis G. 2001. "Oregon Water Quality Index: A tool for evaluating water quality management effectiveness." *Journal of the American Water Resources Association* 37 (1): 125–137.

Curtis, A., and I. Byron. 2002. *Understanding the social drivers of catchment management in the Wimmera region.* Johnstone Centre for Research in Natural Resources / Society.

Dakins, Maxine E., Jeffery D. Long, and Michael Hart. 2005. "Collaborative Environmental Decision Making in Oregon Watershed Groups: Perceptions of Effectiveness." *JAWRA Journal of the American Water Resources Association* 41 (1): 171–180.

Desmarais, Bruce A., and Skyler J. Cranmer. 2012a. "Micro-Level Interpretation of Exponential Random Graph Models with Application to Estuary Networks." *Policy Studies Journal* 40 (3): 402–434.

— . 2012b. "Statistical inference for valued-edge networks: the generalized exponential random graph model." *PloS one* 7 (1): e30136.

Dietz, Thomas, and Paul C. Stern. 2008. *Public participation in environmental assessment and decision making.* Washington, D.C.: National Academies Press.

Donahue, J. D, and R. Zeckhauser. 2011. *Collaborative Governance: Private Roles for Public Goals in Turbulent Times.* Princeton, N.J.: Princeton University Press.

Dryzek, John S. 1997. *The Politics of the Earth: Environmental discourses.* New York, NY: Oxford University Press.

Dunson, David B. 2001. "Commentary: practical advantages of Bayesian analysis of epidemiologic data." *American Journal of Epidemiology* 153 (12): 1222–1226.

Emerson, K., T. Nabatchi, and S. Balogh. 2012. "An integrative framework for collaborative governance." *Journal of Public Administration Research and Theory* 22 (1): 1–29.

Feiock, Richard, In Won Lee, Hyung Jun Park, and Keon-Hyung Lee. 2010. "Collaboration Networks Among Local Elected Officials: Information, Commitment, and Risk Aversion." *Urban Affairs Review* 46 (2): 241–262.

Ferguson, Charlie. 2004. "Governance of collaborations: A case study." *Administration in Social Work* 28 (2): 7–28.

Forman, Richard T. T. 1998. "Roads and Their Major Ecological Effects." *Annual Review of Ecology and Systematics* 29 (1): 207–231.

Frame, Tanis M., Thomas Gunton, and J.C. Day. 2004. "The role of collaboration in environmental management: an evaluation of land and resource planning in British Columbia." *Journal of Environmental Planning and Management* 47 (1): 59–82.

Frederickson, H. G. 2005. "Governance, Governance Everywhere." In *The Oxford Handbook Of Public Management*, ed. by E. Ferlie, L. E. Lynn, and C. Pollitt, 282–304. Oxford Handbooks Online. New York, NY: Oxford University Press.

Freedman, David A., and Richard A. Berk. 2008. "Weighting regressions by propensity scores." *Evaluation Review* 32 (4): 392–409.

Frolich, Markus. 2004. "Finite-sample properties of propensity-score matching and weighting estimators." *Review of Economics and Statistics* 86 (1): 77–90.

Fuller, B.W. 2009. “Surprising cooperation despite apparently irreconcilable differences: Agricultural water use efficiency and CALFED.” *Environmental Science and Policy* 12 (6): 663–673.

Fung, Archon, and Erik Olin Wright. 2003. *Deepening democracy: Institutional innovations in empowered participatory governance.* Vol. 4. Verso Books.

Funk, Michele Jonsson, Daniel Westreich, Chris Weisen, and Marie Davidian. 2010. “Doubly robust estimation of treatment effects.” In *Analysis of Observational Health Care Data Using SAS*, ed. by Douglas Faries, Andrew Leon, Josep Haro, and Robert Obenchain, 85–103. Cary, NC: SAS Institute.

Gelman, A. 2006. “Multilevel (Hierarchical) Modeling: What It Can and Cannot Do.” *Technometrics* 48, no. 3 (): 432–435. doi:10.1198/004017005000000661.

Gelman, A., and J. Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* New York, NY: Cambridge University Press.

Gelman, Andrew. 2008. “Scaling regression inputs by dividing by two standard deviations.” *Statistics in medicine* 27 (15): 2865–2873.

Gelman, Andrew, Jennifer Hill, and Masanao Yajima. 2012. “Why we (usually) don’t have to worry about multiple comparisons.” *Journal of Research on Educational Effectiveness* 5 (2): 189–211.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian data analysis.* CRC press.

Gerber, Elisabeth R., Adam Douglas Henry, and Mark Lubell. 2013. “Political Homophily and Collaboration in Regional Planning Networks.” *American Journal of Political Science* 57 (3): 598–610.

Gerlak, Andrea K., Mark Lubell, and Tanya Heikkila. 2012. "The Promise and Performance of Collaborative Governance." In *Oxford Handbook of US Environmental Policy*, ed. by Sheldon Kamieniecki and Michael E. Kraft, 413–434. Oxford, UK: Oxford University Press.

Gigone, D., and R. Hastie. 1993. "The common knowledge effect: Information sharing and group judgment." *Journal of Personality and Social Psychology* 65 (5): 959–974.

Glynn, Adam N., and Kevin M. Quinn. 2010. "An introduction to the augmented inverse propensity weighted estimator." *Political Analysis* 18 (1): 36–56.

Gomez-Rubio, Virgilio, Roger S. Bivand, and Havard Rue. 2014. "Spatial Models Using Laplace Approximation Methods." In *Handbook of Regional Science*, 1401–1417. Springer.

Goodreau, Steven M., Mark S. Handcock, David R. Hunter, Carter T. Butts, and Martina Morris. 2008. "A statnet Tutorial." *Journal of statistical software* 24 (9): 1–27.

Grafton, R. Quentin, and Karen Hussey. 2011. *Water Resources Planning and Management.* New York, NY: Cambridge University Press.

Granovetter, M. 1985. "Economic action and social structure: the problem of embeddedness." *American Journal of sociology*:481–510.

Gray, B. 1989. *Collaborating: Finding common ground for multiparty problems*. San Francisco, CA: Jossey-Bass Publishers.

Gray, Barbara. 2000. "Assessing inter-organizational collaboration: Multiple conceptions and multiple methods." In *Cooperative strategy: Economic, business, and organizational issues*, ed. by D. Faulkner and M. De Rond, 7–32. New York, NY: Oxford University Press.

Greenland, Sander, Malcolm Maclure, James J. Schlesselman, Charles Poole, and Hal Morgenstern. 1991. "Standardized regression coefficients: a further critique and review of some alternatives." *Epidemiology* 2 (5): 387–392.

Gregory, Robin, Tim McDaniels, and Daryl Fields. 2001. "Decision Aiding, Not Dispute Resolution: Creating Insights through Structured Environmental Decisions." *Journal of Policy Analysis and Management* 20 (3): 415–432.

Griffin, Carol B. 1999. "Watershed Councils: An Emerging Form of Public Participation in Natural Resource Management." *JAWRA Journal of the American Water Resources Association* 35 (3): 505–518.

Guba, Egon G., and Yvonna S. Lincoln. 1994. "Competing paradigms in qualitative research." In *Handbook of Qualitative Research*, ed. by N.K. Denzin and Y.S. Lincoln, 105–117. Thousand Oaks, CA: Sage.

Gunton, Thomas I. 2003. "The theory and practice of collaborative planning in resource and environmental management." *Environments: a journal of interdisciplinary studies* 31 (2).

Habron, Geoffrey. 2003. "Role of adaptive management for watershed councils." *Environmental Management* 31 (1): 0029–0041.

Hall, Thad E., and Laurence J. O'Toole. 2000. "Structures for Policy Implementation An Analysis of National Legislation, 1965-1966 and 1993-1994." *Administration & Society* 31 (6): 667–686.

— . 2004. "Shaping Formal Networks through the Regulatory Process." *Administration & Society* 36 (2): 186–207.

Handcock, Mark. 2003. "Assessing Degeneracy in Statistical Models of Social Networks."

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, Pavel N. Krivitsky, and Martina Morris. 2014a. *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. R package version 3.1.2. The Statnet Project (http://www.statnet.org).

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, Pavel N. Krivitsky, Skye Bender-deMoll, and Martina Morris. 2014b. *statnet: Software tools for the Statistical Analysis of Network Data.* R package version 2014.2.0. The Statnet Project (http://www.statnet.org).

Harding, J. S., E. F. Benfield, P. V. Bolstad, G. S. Helfman, and E. B. D. Jones. 1998. "Stream biodiversity: The ghost of land use past." *Proceedings of the National Academy of Sciences* 95, no. 25 (): 14843–14847.

Hardy, ScottD., and TomasM. Koontz. 2008. "Reducing Nonpoint Source Pollution Through Collaboration: Policies and Programs Across the U.S. States." *Environmental Management* 41, no. 3 (): 301–310. doi:10.1007/s00267-007-9038-6.

Healy, Robert G., Debora VanNijnatten, and Marcela López-Vallejo. 2014. *Environmental policy in North America: Approaches, capacity, and the management of transboundary issues.* University of Toronto Press.

Heimlich, Joe E. 2010. "Environmental education evaluation: Reinterpreting education as a strategy for meeting mission." *Evaluation and Program planning* 33 (2): 180–185.

Henry, A. D. 2011. "Ideology, Power, and the Structure of Policy Networks." *Policy Studies Journal* 39 (3): 361–383.

Henry, A. D, M. Lubell, and M. McCoy. 2012. "Survey-Based Measurement of Public Management and Policy Networks." *Journal of Policy Analysis and Management* 31 (2): 432–452.

Hibbard, Michael, and Susan Lurie. 2006. "Some community socio-economic benefits of watershed councils: A case study from Oregon." *Journal of environmental planning and management* 49 (6): 891–908.

Hijmans, Robert J. 2014. *raster: raster: Geographic data analysis and modeling.* R package version 2.3-0.

Hill, C., and L. Lynn. 2003. "Producing human services: Why do agencies collaborate?" *Public Management Review* 5 (1): 63–81.

Hines, Jody M., Harold R. Hungerford, and Audrey N. Tomera. 1987. "Analysis and synthesis of research on responsible environmental behavior: A meta-analysis." *The Journal of environmental education* 18 (2): 1–8.

Hoch, Charles. 2000. *The Practice of Local Government Planning.* Washington, DC: Published for the ICMA Training Institute by the International City/County Management Association.

Hodges, James S. 2014. *Richly Parameterized Linear Models Additive, Time Series, and Spatial Models Using Random Effects.* Taylor / Francis.

Holley, Cameron, Neil Gunningham, and Clifford Shearing. 2013. *The new environmental governance.* Routledge.

Hoornbeek, John, Evan Hansen, Evan Ringquist, and Robert Carlson. 2012. "Implementing Water Pollution Policy in the United States: Total Maximum Daily Loads and Collaborative Watershed Management." *Society & Natural Resources* 26 (4): 420–436.

Horner, Richard Ray, and Brian Wayne Mar. 1983. "Guide for assessing water-quality impacts of highway operations and maintenance." *Transportation Research Record*, no. 948.

Hungerford, Harold R., and Trudi L. Volk. 1990. "Changing learner behavior through environmental education." *The journal of environmental education* 21 (3): 8–21.

Huntington, H. P. 2000. "Using traditional ecological knowledge in science: methods and applications." *Ecological applications* 10 (5): 1270–1274.

Huxham, Chris. 2003. "Theorizing collaboration practice." *Public Management Review* 5 (3): 401–423.

Huxham, Chris, and Siv Vangen. 2000. "Leadership in the shaping and implementation of collaboration agendas: How things happen in a (not quite) joined-up world." *Academy of Management journal* 43 (6): 1159–1175.

Huxham, Chris., and Siv Vangen. 2005. *Managing to collaborate : the theory and practice of collaborative advantage*. New York: Routledge.

Imperial, M. T. 2005. "Using collaboration as a governance strategy lessons from six watershed management programs." *Administration & Society* 37 (3): 281–320.

Innes, J. E, and D. E Booher. 1999. "Consensus building and complex adaptive systems." *Journal of the American Planning Association* 65 (4): 412–423.

Innes, J. E., and D. E. Booher. 2004. "Reframing Public Participation: Strategies for the 21st Century." *Planning Theory and Practice* 5 (4): 419–436.

Innes, Judith E. 1996. "Planning through consensus building: A new view of the comprehensive planning ideal." *Journal of the American Planning Association* 62 (4): 460–472.

— . 1998. "Information in communicative planning." *Journal of the American Planning Association* 64 (1): 52–63.

Innes, Judith E., and David E. Booher. 2010. *Planning with Complexity: An introduction to collaborative rationality for public policy.* New York, NY: Routledge.

Isett, K. R, I. A Mergel, K. LeRoux, P. A Mischen, and R. K Rethemeyer. 2011. "Networks in public administration scholarship: Understanding where we are and where we need to go." *Journal of Public Administration Research and Theory* 21 (suppl 1): i157–i173.

Johnes, P. J., and A. L. Heathwaite. 1997. "Modelling the impact of land use change on water quality in agricultural catchments." *Hydrological Processes* 11 (3): 269–286.

Kickert, W. J.M, E. H Klijn, and J. F.M Koppenjan. 1997. "Managing networks in the public sector: Findings and reflections." *Managing Complex Networks. Strategies for the Public Sector*:166–191.

King, Gary, and Langche Zeng. 2006. "The dangers of extreme counterfactuals." *Political Analysis* 14 (2): 131–159.

Klijn, E. H. 1997. "Policy networks: an overview." *Managing complex networks. Strategies for the public sector*:14–34.

Klijn, E. H, J. Edelenbos, and B. Steijn. 2010. "Trust in Governance Networks Its Impacts on Outcomes." *Administration & Society* 42 (2): 193–221.

Klijn, E. H, and J. F.M Koppenjan. 2000. "Public management and policy networks." *Public Management an International Journal of Research and Theory* 2 (2): 135–158.

Klijn, Erik-Hans, and Joop Koppenjan. 2006. "Institutional design." *Public Management Review* 8 (1): 141–160.

Klijn, Erik-Hans, Joop Koppenjan, and Katrien Termeer. 1995. "Managing networks in the public sector: A theoretical study of management strategies in policy networks." *Public Administration* 73 (3): 437–454.

Knoke, D. 1990. "Networks of political action: Toward theory construction." *Social forces* 68 (4): 1041–1063.

Kolaczyk, Eric D. 2009. *Statistical analysis of network data: methods and models.* New York, NY: Springer.

Koontz, T. M. 2003. "The farmer, the planner, and the local citizen in the dell: How collaborative groups plan for farmland preservation." *Landscape and Urban Planning* 66 (1): 19–34.

Koontz, T. M, and C. W Thomas. 2006. "What do we know and need to know about the environmental outcomes of collaborative management?" *Public Administration Review* 66:111–121.

Koontz, T. M, J. A Carmin, T. A Steelman, and C. W Thomas. 2004. *Collaborative environmental management: What roles for government?* Washington, D.C.: RFF Press.

Koppenjan, J. F.M, and E. H Klijn. 2004. *Managing uncertainty in networks; a network approach to problem solving and decision making.* London: Routledge.

Krackhardt, David. 1988. "Predicting with networks: Nonparametric multiple regression analysis of dyadic data." *Social networks* 10 (4): 359–381.

Kriesi, H., S. Adam, and M. Jochum. 2006. "Comparative analysis of policy networks in Western Europe." *Journal of European public policy* 13 (3): 341–361.

Krivitsky, Pavel N. 2012. "Exponential-family random graph models for valued networks." *Electronic Journal of Statistics* 6:1100–1128.

Krivitsky, Pavel N., and Carter T. Butts. 2013. *Modeling valued networks with statnet.* Tech. rep. The Statnet Development Team.

Langbein, L. I. 2000. "Regulatory Negotiation versus Conventional Rule Making: Claims, Counterclaims, and Empirical Evidence." *Journal of Public Administration Research and Theory* 10 (3): 599–632.

Layzer, Judith A. 2008. *Natural Experiments: Ecosystem-based management and the environment.* Cambridge, MA: MIT Press.

Leach, W. D, N. W Pelkey, and P. A Sabatier. 2002. "Stakeholder partnerships as collaborative policymaking: Evaluation criteria applied to watershed management in California and Washington." *Journal of Policy Analysis and Management* 21 (4): 645–670.

Leach, W. D., C. M. Weible, S. R. Vince, S. N. Siddiki, and J. C. Calanni. 2013. "Fostering Learning through Collaboration: Knowledge Acquisition and Belief Change in Marine Aquaculture Partnerships." *Journal of Public Administration Research and Theory* 24 (3): 591–622.

Leach, William D. 2006. "Collaborative Public Management and Democracy: Evidence from Western Watershed Partnerships." *Public Administration Review* 66 (Supplement): 100–110.

Leach, William D., and Paul A. Sabatier. 2005. "Are trust and social capital the keys to success? Watershed partnerships in California and Washington." In *Swimming Upstream: Collaborative approaches to watershed management*, ed. by Paul. A. Sabatier, Will Focht, Mark Lubell, Zev Trachtenberg, Arnold Vedlitz, and Marty Matlock, 233–258. Cambridge, MA: MIT Press.

LeSage, James, and Robert Kelley Pace. 2010. *Introduction to spatial econometrics.* CRC press.

Levy, John M. 2013. *Contemporary Urban Planning.* Upper Saddle River, NJ: Pearson Education.

Lindgren, Finn, and Havard Rue. 2013. "Bayesian spatial and spatiotemporal modelling with r-inla." *Journal of Statistical Software.*

Lindgren, Finn, Håvard Rue, and Johan Lindström. 2011. "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (4): 423–498.

Long, N. E. 1958. "The local community as an ecology of games." *American Journal of Sociology*:251–261.

Lubell, M. 2004a. "Collaborative environmental institutions: All talk and no action?" *Journal of Policy Analysis and Management* 23 (3): 549–573.

— . 2005. "Political Institutions and Conservation by Local Governments." *Urban Affairs Review* 40 (6): 706–729.

Lubell, M., A. D Henry, and M. McCoy. 2010. "Collaborative institutions in an ecology of games." *American Journal of Political Science* 54 (2): 287–300.

Lubell, M., and L. Lippert. 2011. "Integrated regional water management: a study of collaboration or water politics-as-usual in California, USA." *International Review of Administrative Sciences* 77 (1): 76–100.

Lubell, Mark. 2004b. "Collaborative Watershed Management: A View from the Grassroots." *Policy Studies Journal* 32 (3): 341–361.

— . 2004c. "Resolving conflict and building cooperation in the National Estuary Program." *Environmental Management* 33 (5): 677–691.

— . 2013. “Governing institutional complexity: The ecology of games framework.” *Policy Studies Journal* 41 (3): 537–559.

Lubell, Mark N., Garry Robins, and Peng Wang. 2011. “Policy coordination in an ecology of water management games.”

Lubell, Mark, and Allan Fulton. 2008. “Local policy networks and agricultural watershed management.” *Journal of Public Administration Research and Theory* 18 (4): 673–696.

Lubell, Mark, William D. Leach, and Paul A. Sabatier. 2009. “Collaborative watershed partnerships in the epoch of sustainability.” In *Toward sustainable communities: transition and transformations in environmental policy*, ed. by Daniel A. Mazmanian and Michael E. Kraft, 255–88. Cambridge, MA: MIT Press.

Lubell, M., M. Schneider, J. T. Scholz, and M. Mete. 2002. “Watershed partnerships and the emergence of collective action institutions.” *American Journal of Political Science* 46 (1): 148–163.

Lubell, M., J. Scholz, R. Berardo, and G. Robins. 2012. “Testing Policy Theory with Statistical Models of Networks.” *Policy Studies Journal* 40 (3): 351–374.

Lurie, Susan, and Michael Hibbard. 2008. “Community-based natural resource management: Ideals and realities for oregon watershed councils.” *Society and natural resources* 21 (5): 430–440.

Lusher, Dean, Johan Koskinen, and Garry Robins. 2013. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. New York, NY: Cambridge University Press.

Mandell, M. P. 1990. “Network management: Strategic behavior in the public sector.” In *Strategies for Managing Intergovernmental Policies and Network.* Ed. by R. W Gage and M. P. Mandell. New York, NY: Praeger.

Mandell, Myrna P. 1999. “The impact of collaborative efforts.” *Review of Policy Research* 16 (1): 4–17.

— . 2001. "Collaboration through network structures for community building efforts." *National Civic Review* 90 (3): 279–288.

Margerum, R. D. 2011. *Beyond Consensus: Improving collaborative planning and management*. Cambridge, MA: MIT Press.

Margerum, R. D, and S. M Born. 2000. "A co-ordination diagnostic for improving integrated environmental management." *Journal of Environmental Planning and Management* 43 (1): 5–21.

Margerum, Richard D. 2002. "Collaborative Planning Building Consensus and Building a Distinct Model for Practice." *Journal of Planning Education and Research* 21 (3): 237–253.

— . 2007. "Overcoming locally based collaboration constraints." *Society & Natural Resources* 20 (2): 135–152.

— . 2008. "A typology of collaboration efforts in environmental management." *Environmental management* 41 (4): 487–500.

Margerum, Richard D., and Rachael Holland. 2000. "Environmental Planning in South East Queensland: An Evaluation of Local Government Activities and the Influence of Collaborative Growth Management." PhD thesis, School of Planning, Landscape Architecture and Surveying, Queensland University of Technology.

Margerum, Richard D., and Debra Whitall. 2004. "The challenges and implications of collaborative management on a river basin scale." *Journal of environmental planning and management* 47 (3): 409–429.

Martino, Sara, Rupali Akerkar, and Havard Rue. 2011. "Approximate Bayesian inference for survival models." *Scandinavian Journal of Statistics* 38 (3): 514–528.

Martino, Sara, and Havard Rue. 2010. "Case studies in Bayesian computation using INLA." In *Complex data modeling and computationally intensive statistical methods*, 99–114. Springer.

Mattessich, Paul W., Marta Murray-Close, and Barbara R. Monsey. 2001. *Collaboration: What Makes it Work.* St. Paul, MN: Amherst H. Wilder Foundation.

Mazmanian, Daniel A., and Michael E. Kraft. 2009. *Toward Sustainable Communities: Transition and Transformations in Environmental Policy.* Cambridge, MA: MIT Press.

McAllister, Ryan RJ, Rod McCrea, and Mark N. Lubell. 2014. "Policy networks, stakeholder interactions and climate adaptation in the region of South East Queensland, Australia." *Regional Environmental Change* 14 (2): 527–539.

McGuire, Michael. 2006. "Collaborative Public Management: Assessing What We Know and How We Know It." *Public Administration Review* 66 (Supplement): 33–43.

McGuire, Michael, and Chris Silvia. 2010. "The Effect of Problem Severity, Managerial and Organizational Capacity, and Agency Structure on Intergovernmental Collaboration: Evidence from Local Emergency Management." *Public Administration Review* 70 (2): 279–288.

McLaughlin, John A., and Gretchen B. Jordan. 1999. "Logic models: a tool for telling your programs performance story." *Evaluation and program planning* 22 (1): 65–72.

Meador, Michael R., and Robert M. Goldstein. 2003. "Assessing Water Quality at Large Geographic Scales: Relations Among Land Use, Water Physicochemistry, Riparian Condition, and Fish Community Structure." *Environmental Management* 31, no. 4 (): 0504–0517. doi:10.1007/s00267-002-2805-5.

Meier, K. J. 2005. "Managerial Networking: Issues of Measurement and Research Design." *Administration & Society* 37 (5): 523–541.

Meier, K. J, and L. J O'Toole. 2001. "Managerial strategies and behavior in networks: A model with evidence from US public education." *Journal of Public Administration Research and Theory* 11 (3): 271–294.

— . 2007. "Modeling public management." *Public Management Review* 9 (4): 503–527.

Montgomery, David R. 1994. “Road surface drainage, channel initiation, and slope instability.” *Water Resources Research* 30 (6): 1925–1932.

Moore, Elizabeth, and Thomas Koontz. 2003. “Research note a typology of collaborative watershed groups: Citizen-based, agency-based, and mixed partnerships.” *Society & Natural Resources* 16 (5): 451–460.

Moreland, R. L, J. M Levine, and M. A Cini. 1993. “Group socialization: The role of commitment.” In *Group Motivation: Social Psychological Perspectives*, ed. by M. A. Hogg and D. Abrams, 105–129. London: Harvester Wheatsheaf.

National Rivers and Streams Assessment: 2008-2009. 2013. Tech. rep. Washington, D.C.: U.S. Environmental Protection Agency.

Newig, Jens, and Oliver Fritsch. 2009a. “Environmental governance: participatory, multilevel, and effective?” *Environmental Policy and Governance* 19 (3): 197–214.

— . 2009b. “More Input-Better Output: Does citizen involvement improve environmental governance.” In *In Search of Legitimacy. Policy Making in Europe and the Challenge of Complexity*, ed. by I Bluhdorn, 205–224. Opladen, Germany: Farmington Hills.

Nikolic, Sara JS, and Tomas M. Koontz. 2008. “Nonprofit organizations in environmental management: A comparative analysis of government impacts.” *Journal of Public Administration Research and Theory* 18 (3): 441–463.

Niles, M. T, and M. Lubell. 2012. “Integrative Frontiers in Environmental Policy Theory and Research.” *Policy Studies Journal* 40:41–64.

O‘Leary, R., C. Gerard, and L. B Bingham. 2006. “Introduction to the symposium on collaborative public management.” *Public Administration Review* 66 (s1): 6–9.

O’Leary, Rosemary, and Lisa B. Bingham. 2003. *The promise and performance of environmental conflict resolution*. Washington, D.C.: RFF Press.

O’Leary, Rosemary, and Lisa Blomgren Bingham. 2009. *The Collaborative Public Manager: New ideas for the twenty-first century*. Washington, D.C.: Georgetown University Press.

Ostrom, E. 2010. "Beyond markets and states: polycentric governance of complex economic systems." *American Economic Review* 100 (3): 641–672.

Ostrom, V., C. M Tiebout, and R. Warren. 1961. "The organization of government in metropolitan areas: a theoretical inquiry." *The American Political Science Review* 55 (4): 831–842.

Parker, Kate Bodane, Richard D. Margerum, Dana C. Dedrick, and Jason P. Dedrick. 2010. "Sustaining watershed collaboratives: the issue of coordinator-board relationships." *Society & Natural Resources* 23 (5): 469–484.

Pearl, Judea. 2000. *Causality: models, reasoning and inference.* Vol. 29. Cambridge Univ Press.

Pebesma, Edzer J., and Roger S. Bivand. 2014. *sp: Classes and methods for spatial data in R.*

Platt, Jennifer. 2006. "Some Specific Problems of Documentary Research Evidence and Proof in Documentary Research, Part Two." In *Theory and Methods in Documentary Research*, ed. by John Scott. Documentary Research 1. Thousand Oaks, CA: Sage Publications.

Podolny, J. M, and K. L Page. 1998. "Network forms of organization." *Annual review of sociology*:57–76.

Pohlmeier, Winfried, Rubin Seiberlich, and Derya Uysal. 2013. "A Simple and Successful Method to Shrink the Weight." *Available at SSRN:*

Poole, Charles. 1991. "Multiple comparisons? No problem!" *Epidemiology*:241–243.

Prell, C. 2012. *Social Network Analysis: History, Theory and Methodology*. Sage.

Provan, K. G, A. Fish, and J. Sydow. 2007. "Interorganizational networks at the network level: A review of the empirical literature on whole networks." *Journal of management* 33 (3): 479–516.

Provan, K. G, and P. Kenis. 2008. "Modes of network governance: Structure, management, and effectiveness." *Journal of public administration research and theory* 18 (2): 229–252.

Provan, K. G, and H. B Milward. 1995. “A preliminary theory of interorganizational network effectiveness: A comparative study of four community mental health systems.” *Administrative science quarterly* 40 (1): 1–33.

Putnam, R. D. 2000. *Bowling Alone.* New York, NY: Simon & Schuster.

Rapp, Valerie. 2008. *Northwest Forest Plan, the First 10 Years (1994-2003): First-decade Results of the Northwest Forest Plan.* Tech. rep. US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Raudenbush, S. W, and A. S Bryk. 2001. *Hierarchical Linear Models: Applications and data analysis methods.* 2nd ed. Thousand Oaks, CA: Sage.

Rhodes, R. A.W. 1997. *Understanding governance: Policy networks, governance, reflexivity and accountability.* Open University Press.

Robert, Christian P., and George Casella. 2004. *Monte Carlo statistical methods.* Vol. 319. Citeseer.

Robins, Garry, Jenny M. Lewis, and Peng Wang. 2012. “Statistical Network Analysis for Analyzing Policy Networks.” *Policy Studies Journal* 40 (3): 375–401.

Robins, James M., and Andrea Rotnitzky. 1995. “Semiparametric efficiency in multivariate regression models with missing data.” *Journal of the American Statistical Association* 90 (429): 122–129.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. “Estimation of regression coefficients when some regressors are not always observed.” *Journal of the American Statistical Association* 89 (427): 846–866.

Robinson, S. E. 2007. “A decade of treating networks seriously.” *Policy Studies Journal* 34 (4): 589–598.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. “The central role of the propensity score in observational studies for causal effects.” *Biometrika* 70 (1): 41–55.

Rubin, Donald B. 1986. "Comment: Which ifs have causal answers." *Journal of the American Statistical Association* 81 (396): 961–962.

Rue, Havard, and Leonhard Held. 2005. *Gaussian Markov random fields: theory and applications.* CRC Press.

Rue, Havard, Sara Martino, and Nicolas Chopin. 2009. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the royal statistical society: Series b (statistical methodology)* 71 (2): 319–392.

Rue, Havard, Sara Martino, Finn Lindgren, Daniel Simpson, Andrea Riebler, and Elias Teixeira Krainski. 2014. *INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation.* R package version 0.0-1404466487.

Sabatier, P. A. 2005. *Swimming Upstream: Collaborative approaches to watershed management.* Cambridge, MA: MIT Press.

Sabatier, P. A, and H. C Jenkins-Smith. 1993. *Policy Change and Learning: An advocacy coalition framework.* Boulder, CO: Westview.

Sabatier, P. A, W. D Leach, M. Lubell, and N. W Pelkey. 2005. "Theoretical frameworks explaining partnership success." In *Swimming Upstream: Collaborative approaches to watershed management*, ed. by P. Sabatier, W. Focht, M. Lubell, Z. Trachtenberg, A. Vedlitz, and M. Matlock, 173–200. Cambridge, MA: MIT Press.

Sabatier, Paul A., and Lauren K. Shaw. 2009. "Are collaborative watershed management groups democratic? An analysis of California and Washington partnerships." *Journal of Soil and Water Conservation* 64 (2): 61A–64A.

Salamon, L. M. 2002. *The Tools of Government: A guide to the new governance.* New York, NY: Oxford University Press.

Scharfstein, Daniel O., Andrea Rotnitzky, and James M. Robins. 1999. “Adjusting for nonignorable drop-out using semiparametric nonresponse models.” *Journal of the American Statistical Association* 94 (448): 1096–1120.

Schively, Carissa. 2007. “A Quantitative Analysis of Consensus Building in Local Environmental Review.” *Journal of Planning Education and Research* 27 (1): 82–98.

Schneider, M., J. Scholz, M. Lubell, D. Mindruta, and M. Edwardsen. 2003. “Building consensual institutions: Networks and the National Estuary Program.” *American Journal of Political Science* 47 (1): 143–158.

Scholz, J. T, R. Berardo, and B. Kile. 2008. “Do networks solve collective action problems? Credibility, search, and collaboration.” *Journal of Politics* 70 (2): 393–406.

Schrank, A., and J. Whitford. 2011. “The Anatomy of Network Failure*.” *Sociological Theory* 29 (3): 151–177.

Scott, Tyler, and Craig Thomas. 2015. “The Effect of Collaborative Partnerships on Interorganizational Networks.” *Public Performance & Management Review* In Publication.

Skaggs, R. W., M. A. Breve, and J. W. Gilliam. 1994. “Hydrologic and water quality impacts of agricultural drainage.” *Critical reviews in environmental science and technology* 24 (1): 1–32.

Smaldino, Paul E., and Mark Lubell. 2014. “Institutions and Cooperation in an Ecology of Games.” *Artificial Life* 20, no. 2 (): 207–221. doi:`10.1162/ARTL_a_00126`.

Smith, Graham. 2004a. *Deliberative Democracy and the Environment*. New York, NY: Routledge.

Smith, Steven Rathgeb. 2004b. “Government and nonprofits in the modern age: Is independence possible?” In *In Search of the Nonprofit Sector*, ed. by Peter Frumkin and Jonathan B. Imber, 3–18. New Brunswick, NJ: Transaction Publishers.

Snijders, Tom AB, Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. "New specifications for exponential random graph models." *Sociological Methodology* 36 (1): 99–153.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. 2002. "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (4): 583–639.

Susskind, L. E, J. Thomas-Lamar, and S. McKearnen. 1999. *The Consensus Building Handbook: A comprehensive guide to reaching agreement*. Thousand Oaks, CA: Sage.

Susskind, Lawrence. 1996. *Dealing With an Angry Public: The mutual gains approach to resolving disputes*. New York, NY: Free Press.

Susskind, Lawrence., and Jeffrey Cruikshank. 1987. *Breaking the Impasse: Consensual approaches to resolving public disputes*. New York, NY: Basic Books.

Sweeney, Bernard W., Thomas L. Bott, John K. Jackson, Louis A. Kaplan, J. Denis Newbold, Laurel J. Standley, W. Cully Hession, and Richard J. Horwitz. 2004. "Riparian deforestation, stream narrowing, and loss of stream ecosystem services." *Proceedings of the National Academy of Sciences of the United States of America* 101 (39): 14132–14137.

Thomas, C. W, and T. M Koontz. 2011. "Research designs for evaluating the impact of community-based management on natural resource conservation." *Journal of Natural Resources Policy Research* 3 (2): 97–111.

Thomas, Craig W. 2003. *Bureaucratic landscapes : interagency cooperation and the preservation of biodiversity*. Cambridge, MA: MIT Press.

Thomson, Ann Marie, James L. Perry, and Theodore K. Miller. 2008. "Linking collaboration processes and outcomes." *Collaborative Public Management*:97–120.

Tong, Susanna TY, and Wenli Chen. 2002. "Modeling the relationship between land use and surface water quality." *Journal of Environmental Management* 66 (4): 377–393.

Torfing, J., and E. Sorensen. 2007. *Theories of democratic network governance.* Basingstoke: Palgrave Macmillian.

Ulibarri, Nicola. 2015. "Tracing Process to Performance of Collaborative Governance: A Comparative Case Study of Federal Hydropower Licensing." *Policy Studies Journal.*

Vangen, Siv, and Chris Huxham. 2003. "Nurturing collaborative relations Building trust in interorganizational collaboration." *Journal of Applied Behavioral Science* 39 (1): 5–31.

Wachhaus, A. 2009. "Networks in contemporary public administration: A discourse analysis." *Administrative Theory & Praxis* 31 (1): 59–77.

Wang, Lizhu, John Lyons, Paul Kanehl, and Ronald Gatti. 1997. "Influences of Watershed Land Use on Habitat Quality and Biotic Integrity in Wisconsin Streams." *Fisheries* 22, no. 6 (): 6–12. doi:10.1577/1548-8446(1997)022<0006:IOWLUO>2.0.CO;2.

Wang, Xinhao. 2001. "Integrating water-quality management and land-use planning in a watershed context." *Journal of Environmental Management* 61 (1): 25–36.

Ward, Trevor J. 2008. "Barriers to biodiversity conservation in marine fishery certification." 2, *Fish and Fisheries* 9:169–177.

Weber, E. P. 2003. *Bringing society back in: Grassroots ecosystem management, accountability, and sustainable communities.* Cambridge, MA: MIT Press.

Weible, C. M, P. A Sabatier, and K. McQueen. 2009. "Themes and variations: Taking stock of the advocacy coalition framework." *Policy Studies Journal* 37 (1): 121–140.

Weible, C., P. A Sabatier, and M. Lubell. 2004. "A Comparison of a Collaborative and Top-Down Approach to the Use of Science in Policy: Establishing Marine Protected Areas in California." *Policy Studies Journal* 32 (2): 187–207.

Weimer, D. L, and A. R Vining. 2010. *Policy Analysis: Concepts and Practice, 5th edition.* Upper Saddle River, NJ: Prentice Hall.

Wikle, Christopher K. 2003. "Hierarchical models in environmental science." *International Statistical Review* 71 (2): 181–199.

Wikle, Christopher K., and Mevin B. Hooten. 2010. "A general science-based framework for dynamical spatio-temporal models." *Test* 19 (3): 417–451.

Wondolleck, J. M, and S. L Yaffee. 2000. *Making collaboration work: Lessons from innovation in natural resource management*. Washington, D.C.: Island Press.

Wood, Donna J., and Barbara Gray. 1991. "Toward a Comprehensive Theory of Collaboration." *Journal of Applied Behavioral Science* 27 (2): 139.

Wyatt, Danny, Tanzeem Choudhury, and Jeff Bilmes. 2009. "Dynamic multi-valued network models for predicting face-to-face conversations." In *NIPS workshop on Analyzing Networks and Learning with Graphs.*

— . 2010. "Discovering long range properties of social networks with multi-valued time-inhomogeneous models." In *AAAI.*

Xu, Ke, and Christopher K. Wikle. 2007. "Estimation of parameterized spatio-temporal dynamic models." *Journal of Statistical Planning and Inference* 137 (2): 567–588.

Yaffee, S. L, A. Phillips, I. C Frentz, P. W Hardy, S. Maleki, and B. Thorpe. 1996. *Ecosystem Management in the United States: An assessment of current experience*. Washington, D.C.: Island Press.

Zhang, Yixin, John S. Richardson, and Xavier Pinto. 2009. "Catchment-scale effects of forestry practices on benthic invertebrate communities in Pacific coastal streams." *Journal of Applied Ecology* 46 (6): 1292–1303.