8-2016

Mathematical Approaches to Sustainability Assessment and Protocol Development for the Bioenergy Sustainability Target Assessment Resource (Bio-STAR)

Nathan Louis Pollesch
University of Tennessee, Knoxville, npollesc@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Nathan Louis Pollesch entitled "Mathematical Approaches to Sustainability Assessment and Protocol Development for the Bioenergy Sustainability Target Assessment Resource (Bio-STAR)." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

Louis J. Gross, Major Professor

We have read this dissertation and recommend its acceptance:

Virginia H. Dale, Suzanne M. Lenhart, Vasileios Maroulas

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Mathematical Approaches to Sustainability Assessment and Protocol Development for the Bioenergy Sustainability Target Assessment Resource (Bio-STAR)

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Nathan Louis Pollesch
August 2016

Abstract

Bioenergy is renewable energy made of materials derived from biological, non-fossil sources. In addition to the benefits of utilizing an energy source that is renewable, bioenergy is being researched for its potential positive impact on climate change mitigation, job creation, and regional energy security. It has also been studied to investigate possible challenges related to indirect and direct land-use change and food security. Bioenergy sustainability assessment provides a method to identify, quantify, and interpret indicators, or metrics, of bioenergy sustainability in order to study trade-offs between environmental, social, and economic aspects of bioenergy production and use. Assessment is crucial to inform policymakers, researchers, and stakeholders as they make decisions to support the development of a sustainable bioeconomy in the United States and globally. It is the purpose of this dissertation to identify and derive mathematical techniques that aid in the development of the Bioenergy Sustainability Target Assessment Resource (Bio-STAR). Guiding principles for Bio-STAR include *(i)* adaptability for assessing diverse bioenergy production pathways, *(ii)* flexibility to support a range of analyses that researchers and policymakers may seek to undertake, and *(iii)* mathematical robustness with respect to the operations utilized. Key components of sustainability assessment are defined and presented in the first chapter. Of the key components, *Normalization* and *Aggregation* represent areas in which the mathematical processes utilized are critical to assessment outcomes. As such, mathematical theory is developed for Normalization and Aggregation in sustainability assessment and presented in the second and third chapters, respectively. This theory is applied in the fourth chapter to inform the development of protocols for the Bioenergy Sustainability Target Assessment Resource. Bioenergy is seen as component of a sustainable energy future in the United States. Bioenergy is unique among renewable energy sources in that it can be produced in a variety of ways. Bio-STAR is a tool that will enable policymakers, researchers, and stakeholders to explore these many bioenergy options from a sustainability viewpoint and make decisions that will guide the U.S. and the world towards a sustainable energy future.

Table of Contents

List of Tables

List of Figures

Introduction

Sustainability assessments have found use around the world for applications as diverse as financial investment, energy production, architecture, 3D-printing technology, and city planning (Mori and Yamashita, 2015; Tatari et al., 2015; Gebler et al., 2014; Singh et al., 2009; Mayer, 2008; Böhringer and Jochem, 2007). Accompanying this diversity in application is perhaps an even greater diversity of methodological approaches. Sustainability assessment approaches differ in how, or if, indicators are utilized, data are normalized and weighted, data are aggregated, and spatial and temporal extents for the assessment are defined. Consensus exists that sustainability assessments minimally consider environmental, social, and economic aspects of a system, taking into account potential impacts on future generations (Brundtland et al., 1987). Among other considerations, variation in methodology is due to differing assessment goals, desired application, and philosophical perspectives on acceptable trade-offs between assessment components. Researchers from around the world have, and are continuing to, develop mathematical theory to identify advantages and drawbacks to techniques utilized in sustainability assessment (Dias and Domingues, 2014; Langhans et al., 2014; Pinar et al., 2014; Roberts, 2014a; Ebert and Welsch, 2004). Research presented here builds upon this work and focuses specifically on the normalization and aggregation components that occur in many sustainability assessments.

Researchers at Oak Ridge National Laboratory's Center for Bioenergy Sustainability have identified environmental and socioeconomic indicators for bioenergy sustainability assessment and seek to develop a tool for the visualization and assessment of these indicators (Dale et al., 2015, 2013a; McBride et al., 2011). In addition to contributing to the general body of research related to the mathematics of sustainability assessment, this work is an extension of the research emanating from Oak Ridge National Laboratory's Center for Bioenergy Sustainability to define what sustainability means in a quantitative sense (Efroymson and Dale, 2015; Parish et al., 2016; Dale et al., 2013b; Baskaran et al., 2010). The first chapter presents Key Components for Sustainability Assessment, which includes a review of current methodologies and identifies those techniques that most support the goals highlighted for bioenergy sustainability assessment. The next two chapters utilize mathematical approaches to develop a theoretical basis for the selection of normalization and aggregation methodologies within bioenergy sustainability assessment. The final chapter address the specification of protocols for normalization and aggregation for the Bioenergy Sustainability Target Assessment Resource (Bio-STAR); a visualization and assessment tool that will be made available to support the sustainable operation and development bioenergy production and utilization.

Chapter I

Towards a Bioenergy Sustainability Target Assessment Resource (Bio-STAR): Key Components and Requirement Specification

Publication Statement:

A version of this Chapter is currently being considered for publication by its co-authors, Nathan L. Pollesch and Virginia H Dale. This manuscript was principally composed by N. Pollesch; V.H. Dale contributed Sections 2.3 and 2.4 on indicator selection and categorization as well as numerous comments on the manuscript in previous versions.

Towards a Bioenergy Sustainability Target Assessment Resource (Bio-STAR): Key Components and Requirement Specification

N. Pollesch[a,b,*], V.H. Dale[b]

[a] *Department of Mathematics, The University of Tennessee, 1403 Circle Drive, Knoxville, TN 37996-1320*
[b] *Center for BioEnergy Sustainability, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6035*

Abstract

This paper presents a conceptual framework for the construction of a bioenergy sustainability assessment resource that builds upon previous research. Seven key components for multimetric sustainability assessment are presented and current methodological approaches associated with each component are highlighted. The key components are: (1) Identification of Goals and Extent, (2) Indicator selection and categorization, (3) Data normalization and weighting, (4) Aggregation, (5) Analysis of data quality, (6) Evaluation of assessment results, and (7) Visualization and reporting. We identify specific approaches within each key component that are suitable for the construction of a Bioenergy Sustainability Target Assessment Resource (Bio-STAR). The approaches selected satisfy the following research objectives: Adaptability for assessing diverse bioenergy production pathways, flexibility to support a range of analyses that researchers and policymakers may seek to undertake, and mathematical robustness with respect to normalization, aggregation, and the quantification of data uncertainty.

*Corresponding author

1. Introduction

Bioenergy is a renewable energy source whose utilization can affect processes ranging from climate change to job creation to regional energy security. The 2014 U.S. National Climate Assessment noted that bioenergy derived from forest biomass could improve water quality and emerging bioenergy markets could aid in restoration of forests killed by drought, insects, and fire (Joyce et al., 2014). Depending on context and deployment, it may also affect land use and water use and availability that could move systems away from sustainability goals (Bringezu et al., 2009; Finco and Doppler, 2010; Berndes and Hansson, 2007; Msangi et al., 2007) or closer to sustainability targets (Parish et al., 2012; Dale et al., 2014). The feasibility and effects of the development of a bioeconomy are being investigated around the world. Certification standards have been provided by the Roundtable on Sustainable Biofuels (RSB) and the Global Bioenergy Partnership (GBEP) as examples. The International Organization for Standardization (ISO) has developed sustainability criteria for bioenergy (*ISO 13065:2015*).

Sustainability assessments investigate how various bioenergy production pathways either contribute or detract from sustainability goals at local, regional, and global levels. Given the diversity of production pathways as well as interest in bioenergy as a renewable energy resource globally, sustainability assessment methodologies must be developed that are adaptable for assessing diverse production methods and flexible to support the range of analyses that researchers and policymakers may seek to utilize. These assessment goals must also be reached while maintaining mathematical robustness with respect to aggregation, normalization, and quantifying data uncertainty.

2. Key Components for Bioenergy Sustainability Assessment

Seven key components for the structuring of a bioenergy sustainability assessment methodology are presented and discussed (Table 1.1). Each component is defined and current methodologies are identified. These seven key components serve as the structure upon which the protocols for Bio-STAR are being developed. Accompanying each key component are the associated requirement specification for Bio-STAR that highlights the techniques seen as contributing the most towards the goals of adaptability, flexibility, and mathematical robustness.

Given that sustainability assessments may be seen as a type of mathematical model, one may also look not only to the literature on sustainability assessment construction for inspiration, but also to the literature on modeling. For example, Table 1.2 provides 6 steps highlighted for mathematical model construction from Giordano and Weir (1985). Of these steps, with a little interpretation, we can see that many useful similarities hold. The first step focuses on identifying the problem, which has a direct relationship to setting up the goals of the assessment, i.e. what are you trying to solve or address in your assessment? The second is about making assumptions, this is where one identifies the variables and mechanisms appropriate to solve the problem mathematically. In sustainability assessment, specifically in an indicator-based approach, the variables are the indicators used to represent system function, and one does indeed need to make assumptions as to their interaction in order to identify the correct analyses. In a non-dynamical sustainability assessment there is no model to solve in the third step of solving and interpreting the model. However analysis of data fits naturally into this step and including normalization which can aid in interpretation of the indicators. Step four is a verification step necessary in to test the model for function, applicability, and usefulness with respect to the goals and problem identified in step one. The implementation step, could include, for sustainability assessment, incorporating data from multiple systems not

Table 1.1: Key components for sustainability assessment

Key Component	*Topics for Key Component*
Identification of Goals and Extent of Analysis	• Identification of stakeholders' needs • Clarification of intended assessment use and function • Specification of spatial and temporal extent to be included in assessment • Establish criteria for evaluation of assessment

Indicator Selection and Categorization	• Definition of relevant system dimensions to be quantified and representative indicators
	• Example categories for indicators:
	  – System function represented
	  – Spatial and temporal extent and domain of impact
	  – Measurability scale

Data Normalization and Weighting	• Identification of relevant normalization techniques, for example:
	  – Z-Score Standardization
	  – Unit Equivalence
	  – Ratio Normalization
	  – Distance to target
	• Identification of if, and how, explicit weighting of indicators will take place

Aggregation	• Determination of if, and to what extent, aggregation will take place • Determination of possible/appropriate categories to aggregate • Identification of appropriate aggregation functions to employ

Analysis of Data Quality	• Identification of statistical techniques appropriate to study:
	  – Variation in individual indicators
	  – Variation in aggregates of indicators
	• Determination of data availability and techniques to address any missing data

Evaluation of Assessment Results	• Determination of if, and how, sensitivity analysis will take place. Potential analysis include:
	  – Indicator interactions and correlation
	  – Indicator and/or measurement contributions to aggregate values (if applicable)
	• Specification of if, and how, assessment results may be compared
	• Does assessment meet criteria established during the identification of goals and extent?

Visualization and Reporting	• Identification of relevant indicator properties to convey
	  – Variability and number measurements of indicators
	  – Spatial extent of measurement and/or impact
	  – Normalization bounds and weights (if applicable)
	• Identification of resolution of data to present i.e. aggregate or individual measures for indicators and/or groups of indicators
	• Relevant comparability of results

Table 1.2: Steps in the construction of a mathematical model *adapted from Giordano and Weir (1985)*

Step	*Description*
Step 1	Identify the problem
Step 2	Make assumptions
	• Identify and classify the variables
	• Determine interrelationships between the variables and submodels
Step 3	Solve or interpret the model
Step 4	Verify the model
	• Does it address the problem?
	• Does it make common sense?
	• Test it with real-world data
Step 5	Implement the model
Step 6	Maintain the model

considered in the initial steps. The final step of maintenance of a model or assessment may take place in a variety of forms. Some maintenance items may be considered academic maintenance, such as updating indicators and analyses based on current methodological approaches. Other more practical maintenance exists for an assessment tool, such as maintaining databases, links, and proper physical function, which are especially relevant for tools like Bio-STAR which will be made available online for general use.

2.1. Identification of Goals and Extent of Analysis

Sustainability assessments have been created for applications as diverse as financial investment tools to urban development and environmental vulnerability (Singh et al., 2009; Mayer, 2008; Böhringer and Jochem, 2007). The creation of a sustainability assessment resource is a large undertaking. There are many techniques available to address each component of an assessment. Given the diverse applications that utilize sustainability assessments, the idea of a standard assessment methodology may be impractical. Therefore, beyond adherence to mathematical theory when and where applicable, it can be argued that the most important guide in assessment creation is a well-defined set of application and outcome goals. This statement of goals is critical in that their identification is necessary to select appropriate techniques from the myriad of methodological options available within each component.

In order to aid identification of analysis techniques to be utilized, a clear statement of the goals and uses of the sustainability assessment should be developed. The 2012 United States Environmental Protection Agency (US EPA) publication entitled *A Framework for Sustainability Indicators at EPA* also points to the identification of goals and application of the assessment and specifically gives four lenses, or applications, of an assessment as being (1) public reporting, (2) decision making, (3) research planning, and (4) program evaluation (Fiksel et al., 2012). Differing applications

of the assessment demand differing sets of concerns. For example, if an application for the assessment involves eventual monetary or economic incentives attached to outcomes, the assessment may demand strict guidelines for data processing and use. If an assessment is to be used for a labeling effort in different parts of a region or the world, identifying comparability standards between sites as well as standards for data collection might be emphasized. The assessment techniques chosen, as well as the depth to which certain specifications within the assessment need to be made, vary with differing applications and goals.

The spatio-temporal extent of an assessment may be determined by both indicator selection and the treatment and collection of data related to a given context. Some assessment efforts are isospatial and strive to assess the changes of one particular bioenergy system through time, while others are isochronal and have the goal to compare multiple systems for a fixed point in time or for equal durations of time. Even others attempt to both treat both changing spatial and temporal conditions at a specific site and allow for cross-site comparisons.

Specific choices for techniques in normalization, aggregation, indicator selection, and the other key components can tailor the assessment to the goals of stakeholders. A clear statement of assessment goals helps to guide those constructing the assessment and to ensure utilization of appropriate techniques and methodology.

2.2. Bio-STAR Requirements: Identification of Goals and Extent of Analysis

The interest in assessing different bioenergy production pathways at various extents and contexts has made adaptability of the assessment approach a high priority. This adaptability should be manifest in the ability of the assessment to be used to *(i)* compare different production sites, feedstocks, or parts of the bioenergy supply chain[1], *(ii)* consider changes over time, and *(iii)* assess regions that have differing policy goals and differing ecological and socioeconomic sensitivities.

Differing groups of stakeholders have different desires when it comes to preferred complexity in assessment results. The target audience for bioenergy sustainability assessments, as identified in Dale et al. (2013a), includes policy makers, business people, and other stakeholders involved in aspects of bioenergy ranging from land managers to those involved with logistics and conversion. While some of this target audience will seek to have access to as much of the minutiae as possible, others may only want to access aggregated 'big-picture' information. Many will likely fall somewhere within this spectrum. Consideration of these varying demands has made the ability to adjust the complexity of assessment results a goal for Bio-STAR. In order to achieve this flexibility, novel aggregation, normalization, and uncertainty analysis approaches need to be developed.

In addition to different desires for complexity in assessment results, differing groups of stakeholders may also seek to make different comparisons with the assessment results provided. These comparisons may vary spatially, such as comparing different locations or different regions that include multiple locations. The comparisons may also take place for a location being compared to itself over a period of time. Due to the necessity of Bio-STAR to adjust to varying socioeconomic and environmental sensitivities, in some cases comparison of bioenergy systems may not be appropriate. The development of comparability standards is another goal for Bio-STAR, and one that is discussed in further detail in section 2.12.

Multiple considerations arise in relation to the scope of assessment and the extents over which indicators have influence. Spatial and temporal boundaries for bioenergy sustainability assessments

[1] The bioenergy supply chain includes feedstock production and logistics, conversion to biofuels, biofuel logistics and biofuel end uses (Dale et al., 2013a; McBride et al., 2011).

have an influence on assessment outcomes and may be determined through various methods. The desired spatial boundaries of an assessment may also differ by stakeholder group. Policymakers may wish to consider political boundaries while researchers may want to use geologic boundaries, such as a watershed. A *fuelshed* is the geographic area from which a bioenergy conversion facility receives its feedstocks, and this spatial boundary, which is neither political or geologic in nature, may also be considered desirable given the explicit connection to bioenergy production. These preferences for spatial boundaries have an influence on assessment protocols creation. However, indicator measurements included in the assessment have their own levels of spatial influence that should also be considered in the determination of the assessment boundaries. In order to ensure the assessment can be adapted to multiple bioenergy pathways and utilized for assessment by various researchers, Bio-STAR should have the ability to consider multiple spatial and temporal extent formulations. The reason that a specific spatial protocol may not satisfy the needs of the assessment is due in part to the varying spatial domains particular to the sustainability indicators identified in McBride et al. (2011) and Dale et al. (2013a), detailed discussion of this topic takes place in Chapter IV of this dissertation. Further, given that Bio-STAR should be applicable over the entire bioenergy supply chain, creating adaptability to consider the varying extents of indicators at different phases in the supply chain is important (Efroymson et al., 2013).

2.3. Indicator Selection and Categorization

Indicators can be defined as measures that provide information about potential or realized effects of human activities on phenomena of concern (Heink and Kowarik, 2010). Indicators of progress toward sustainability can serve several purposes. They can be used to assess the condition of the environment, to monitor trends in conditions over time, to provide an early warning signal of changes in the environment, or to diagnose the cause of an environmental problem (Cairns Jr et al., 1993). That purpose is typically determined by the goals of the assessment, which derives from both context of the issue and the stakeholders perspectives (Dale et al., 2016). Indicators are often used to assess social, economic and environmental sustainability of the system under consideration. Effective indicators can help to identify and quantify the sustainability attributes of alternatives.

Sustainability attributes of agricultural practices in general have been discussed and defined by the Millennium Ecosystem Assessment (Millennium Ecosystem Assessment , MEA), the National Sustainable Agriculture Information Service (Sullivan, 2003; Earles and Williams, 2005), and Dale and Polasky (2007). In addition, several national and international efforts have identified sustainability indicators for bioenergy, including the Roundtable on Sustainable Biomaterials (RSB, 2011), U.S. Biomass Research and Development Board, Global Bioenergy Partnership (GBEP, 2011), Council on Sustainable Biomass Production (CSBP, 2010), and the International Organization for Standardization (ISO, 2015). The number of indicators coming from these efforts is large and diverse, though they share two characteristics. First, these lists include numerous, broadly-defined indicators. Second, many of the indicators focus on assessments of management practices and their predicted effects rather than on measurements that relate to realized effects. However, current understanding of the effects of bioenergy management practices is limited, especially for systems not yet in wide use, such as cellulosic bioenergy (McBride et al., 2011).

Indicator selection is one of the first opportunities to introduce bias into the analysis. Without the guidance of a well-established set of goals, one may inadvertently develop an assessment that includes indicators that are impractical or that might actually cloud the assessment results. "The selection and definition of pertinent indicators, to a large extent, defines the whole issue" (Moldan et al., 2012). The set of indicators selected for assessing progress toward sustainability of bioenergy

systems should be applicable to the scales of interest and be useful to diverse stakeholders. Hence input from a breadth of stakeholders is an important part of the selection process. Stakeholders include those directly affected by, and making use of, the bioenergy system as well as policy makers and experts. Additionally, indicator variables need not be thought of as being fixed for inclusion or not. As analysis is undertaken or areas of concern change, it may be necessary to reassess indicator inclusion and adjust the assessment accordingly. The selected metrics should be clearly specified, science-based indicators that can, for example, support decisions about implementation and expansion of more sustainable bioenergy options over time.

The organization of indicators into groups is important because of the effects it can have on providing meaningful assessment results and for helping to identify appropriate normalization, aggregation, and data analysis procedures. Some categories that can provide benefits in indicator analysis are spatial or temporal extent and variability of indicators, level or scale of measurability of indicators, and categorization by the phenomenological system dynamics the indicator represents. Indicators of progress toward sustainability are often organized into social, economic and ecological categories, and further sub-groups within those categories.

An example of temporal variability occurs in measurements of water quality and quantity, where dissolved oxygen content varies diurnally (Griffith, 2012) and a storm event or spring ice-melt has large effects on flow rates and sediment loads in streams. Without considering natural variation in water quality indicators, assessment results may lose their explanatory power. Understanding the level or scale[2] of measurability of an indicator impacts the family of aggregation functions that can be appropriately used for data compression and synthesis in a mathematically consistent fashion (Ebert and Welsch, 2004; Grabisch et al., 2009). Another way in which assessments have been analyzed is with respect to the concepts of weak and strong sustainability (Mori and Christodoulou, 2012; Hacking and Guthrie, 2008; Mayer, 2008). This binary classification of the assessment depends on indicators belonging (exclusively) to the environmental, social, or economic categories. Weak or strong sustainability is determined by the compensatory behavior of the aggregation process across categories, but, in order for this view to be taken, indicators must first be categorized.

Organization or categorization of indicators uses multiple characteristics of the indicators chosen. Although there are numerous ways in which indicators can be organized, a minimum set of suggested indicator attributes by which to categorize includes level of measurability, spatial and temporal extent of influence (site specific, regional, national, global) as well as degree of spatial and temporal variability, and grouping into environmental, social, and economic categories.

2.4. Bio-STAR Requirements: Indicator Selection and Categorization

A team at Oak Ridge National Laboratory selected key indicators of bioenergy sustainability and proposed how they are best used in particular contexts. That effort built from indicators proposed by the Roundtable on Sustainable Biofuels (RSB, 2011), Global Bioenergy Partnership (GBEP, 2011), Council on Sustainable Biomass Production (CSBP, 2010), the International Organization for Standardization (ISO, 2015), and other efforts (see McBride et al., 2011). The selected indicators were (1) chosen from the plethora of indicators proposed by many groups to be those that seem most useful to decision makers, (2) selected to be applicable across the entire bioenergy supply chain, and (3) identified as a minimum set of indicators that are practical, doable and incorporate key areas of interest to science (Dale et al., 2013a). The proposed environmental and

[2]The scale of measurability refers to a characteristic of its measurement units, the four common scales of measurability are nominal, ordinal, interval and ratio (Stevens, 1946).

socioeconomic indicators represent a suite designed to reflect major sustainability considerations for bioenergy. McBride et al. (2011) identified major environmental categories of sustainability to be soil quality, water quality and quantity, greenhouse gases, biodiversity, air quality, and productivity and proposed 19 indicators that fit into those categories. In addition, 16 socioeconomic indicators were identified that fall into the categories of social well-being, energy security, trade, profitability, resource conservation, and social acceptability (Dale et al., 2013a). Together these twelve categories of indicators provide a checklist of key measures that reflect major environmental and socioeconomic effects across the full supply chain for bioenergy. This approach provides a basis to compare changes in sustainability over time for a specific bioenergy pathway or to compare across pathways. For example, it has been used to determine key concerns of using Eucalyptus for bioenergy in the southeastern United States (Dale et al., 2013b) and of algal-based biofuels (Efroymson and Dale, 2015; Efroymson et al., 2016). The proposed indicators provide a means to quantify and evaluate sustainability of bioenergy systems across different regions and production systems.

The indicators of McBride et al. (2011) and Dale et al. (2013a) have also been investigated for measurability type. Ratio scale measurability is the most prevalent scale of measurability. Pollesch and Dale (2015) showed that of the 19 environmental indicators, only a single indicator, namely, the biodiveristy indicator *Presence of Taxa of Special Concern*, does not belong to a ratio scale. Initial investigation into the the 16 socioeconomic indicators has shown a similar prevalence of ratio scale measurable variables. This measurability scale information can be used to identify aggregation functions that will help ensure consistent aggregation within Bio-STAR. Utilizing a sample data set from Vonore switchgrass (*Panicum virgatum*) production (Parish et al., 2016), research into how best to classify indicators by their spatio-temporal extent and variability is underway, with results forthcoming. One product of this preliminary research into spatio-temporal extent of indicator measures has led to the development of a methodology to organize spatio-temporal attributes in order to aid in the deployment of a flexible assessment structure under strict protocols for normalization and aggregation.

A framework for selecting and evaluating indicators from the checklist provides a means to identify those indicators that, in particular contexts, are useful for assessing sustainability of bioenergy systems (Figure 1.1). The steps in the framework include defining the goals, determining tradeoffs, and setting objectives for analysis and criteria for indicator selection. Then the user of the framework identifies and ranks indicators, applies them in an assessment, and evaluates their effectiveness, while identifying gaps that prevent goals from being met, assessing lessons learned, and moving toward improved practices. The framework emphasizes that the selection of appropriate criteria and indicators is driven by the specific purpose of an analysis. Realistic goals and measures of bioenergy sustainability can be developed systematically with the help of the framework presented here.

2.5. Data Normalization and Weighting

Sustainability analyses bring together information representing various aspects of environmental and socioeconomic functioning. Although the inclusion of indicators that collectively capture key aspects of system functioning is nearly prerequisite for sustainability assessment, how and if comparison of those indicators takes place is often contentious. If comparison, analysis, or simplification of the data is to take place, normalization and weighting are two processes that naturally arise to enable it.

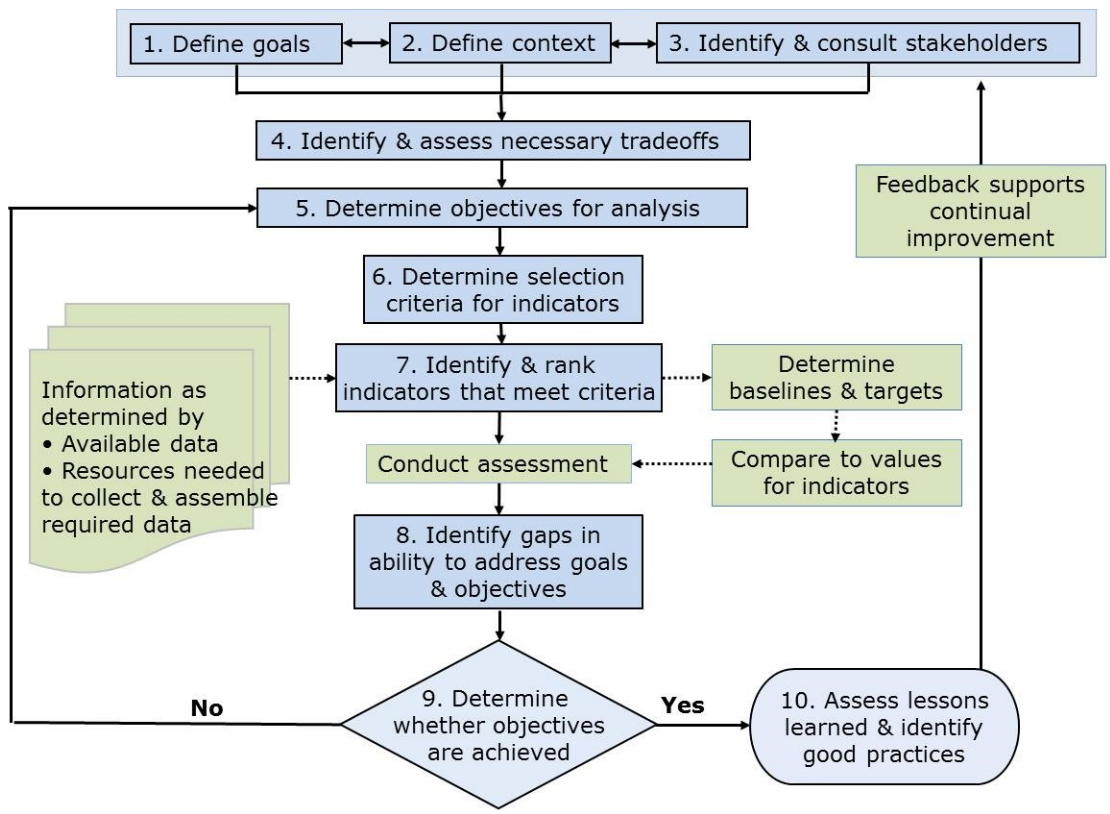


Figure 1.1: A framework for selecting indicators of progress toward sustainability (from Dale et al., 2015). Steps for the framework are shown in blue; supporting components of the assessment process are in green. Note that steps 1, 2 and 3 interact and occur concurrently

2.5.1. Data Normalization

Normalization is the procedure employed to transform differing indicator measures onto similar scales or to unit-free measures. Normalization can take place through a variety of mathematical transformations and is sometimes referred to as standardization or unit scaling. Examples of normalization procedures are z-score standardization, distance-to-target normalization, ratio normalization and unit equivalence normalization, which includes conversion of all measurements to a standard unit such as greenhouse gas equivalents, units of embodied energy (emergy), or the monetization of measurements. Monetization seeks to transform indicator measurements to monetary units, such as dollars (Costanza et al., 1998). Emergy approaches are similiar, but instead of monetary values, emergy units are utilized, such as the emJoule (Scienceman, 1987). Z-score standardization refers to transforming data measurements using the mean and standard deviation values. Distance-to-target normalization refers to the use of baselines and/or target values as references by which all data are transformed. All of these, as well as other normalization procedures can be found within the sustainability assessment literature (see Singh et al., 2009; Böhringer and Jochem, 2007). Pollesch and Dale (*in press*) also includes a sampling of normalization functions found in sustainability assessment. Which procedure is appropriate for a given assessment depends on indicators chosen and the goal of the assessment. However, an understanding of how normalization impacts the identification of appropriate aggregation functions, comparability of assessment results, and implicit weights for indicator data is imperative.

2.5.2. Weighting of Data

Weighting is another topic that arises when considering how data for different indicators are to be compared or when data across diverse systems are compared. Weighting can be explicit, to convey importance or preference for certain indicators, or weighting can be implicit as a result of the rescaling in the normalization procedure and the aggregation procedure chosen. With respect to aggregation using the weighted geometric or arithmetic means, for a set of n indicator variables, weights are represented as values, w_i, for $i = 1, 2, ..., n$ in the interval $[0, 1]$ such that the sum of the weights is 1, $\sum_{i=1}^{n} w_i = 1$. How, or if, to weight indicator data is another topic that should be addressed as goals for the assessment are established. If explicit weighting is to take place, there are a variety of methods to determine explicit weights for indicators. Subjective approaches include weights based on expert opinion and perspectives of stakeholders living within the system (Mayer, 2008). Objective measures, such as those derived from mathematical approaches, include varieties of factor/component analyses (principle component analysis, PCA, for example) and statistical regression techniques may also be utilized (Esty et al., 2005; Singh et al., 2009).

2.6. Bio-STAR Requirements: Data Normalization and Weighting

Of the data normalization techniques currently utilized in sustainability assessment, normalization using distance-to-target methodology contributes most to the goals outlined for Bio-STAR. The concepts that underlay distance-to-target normalization are supported in the sustainability assessment literature. Moldan et al. (2012) point out that, "humans ultimately need an assessment of how far they are from sustainable targets (Stiglitz et al., 2009)" and that "the benefit of specific, quantitative, time-bound targets is then straightforward: the indicators can be linked to them and interpreted clearly on a distance-to-target basis." Mayer (2008) also argues that "indicators are more helpful if they give information on the state of the system with respect to policy targets or biophysical limits." Distance-to-target normalization contributes to the goal of adaptability in that

it allows for direct inclusion of context-relevant baselines and targets tied to the natural geophysical sensitivities of a given system as well as to clearly defined policy and stakeholder goals.

Explicit weighting of indicators is an area that is often debated, as it necessarily walks the line between preference imposition and the introduction of bias. Distance-to-target normalization provides a transparent, implicit weighting procedure and forces weights for indicators to be tied to specific baselines and target values. Booysen (2002) states that "equal weights should be the norm with the burden of proof falling on differential weighting." Although procedures exist for articulating explicit weights for indicators, both empirically derived,or based on expert opinion, explicit weighting of indicator variables for aggregation is not a necessary component of Bio-STAR due to the utilization of distance-to-target normalization. Although setting baselines and targets may be said to also walk the line between preference imposition and the introduction of bias, the transparency and adaptability of the distance-to-target normalization process is a preferred feature. Normalization using distance-to-target contributes most directly to the goals defined for Bio-STAR.

2.7. Aggregation

For analysis of indicator data, complexity can surface simply because of the number of different indicators and the amount of data included in the assessment. Sustainability assessments have used up to 2688 indicator variables (The Living Planet Index, McRae et al., 2012). In addition to the potential inclusion of a large number of indicators in an assessment, each indicator has an associated data set comprised of multiple observations. When all of these data are collected, it leaves the analyst and/or stakeholder with a large amount of information to interpret and synthesize meaningfully.

The extraction of meaningful results from complex, high-dimensional data sets can be difficult. It is for this reason that simplified representative measures are calculated and aggregation of data takes place. The basic statistical processes for aggregating multiple measurements of a single indicator are well known; such as taking a list of data measures and producing summary statistics, such as the mean and standard deviation. However, in summarizing data measurements for several different indicators that span multiple environmental, social, or economic dimensions of a system, the task becomes less clear. Also, spatio-temporal mismatches in extent and resolution of indicator measurements create unique challenges when attempting to synthesize information related to indicators of progress towards sustainability.

Given the benefits of aggregation within a sustainability assessment, researchers have been working to identify appropriate aggregation functions and techniques for environmental and sustainability assessment. Ebert and Welsch (2004) define *meaningful environmental indices* and utilize social choice theory and mathematics to identify *meaningful* aggregation functions for *comparable* or *non-comparable* indicator data based on the data's scale of measurability[3]. Scales of measurability refer to the classifications of data types as nominal, ordinal, interval, and ratio scale formalized by Stevens (1946). Meaningful aggregation functions related to the measurability scales of indicator variables within the assessment provide one approach that can be used to choose aggregation function to utilize. Taking a different approach, the work of Zhou et al. (2006) compares aggregation functions by an objective measure defined as *loss of information*.

[3] The terms meaningful and comparable versus non-comparable have specific mathematical definitions. See Ebert and Welsch (2004) or Grabisch et al. (2009) for formal definitions and Pollesch and Dale (2015) for how those terms apply to sustainability assessment

2.8. *Bio-STAR Requirements: Aggregation*

The aggregation component within a sustainability assessment is, at its essence, a compression of information. The results of this compression within the aggregation component of a sustainability assessment must be performed in a consistent manner in order to make assessment results useful and reliable. Consistency can be ensured by rigid specifications for assessment use (i.e., one must always aggregate specified indicators or groups of indicators in specified ways), however, an ideal assessment uses a framework that is both flexible and consistent. Flexibility is a requirement for Bio-STAR due to the diversity of indicators that are included in the assessment and acknowledges that different stakeholders desire more or less detail depending on their goals and the application. Consistency of aggregate results within a flexible assessment framework can be achieved without overwhelmingly rigid specifications, but in order to do so, a deeper understanding of the fundamental properties of the aggregation process is required. This deeper understanding of the aggregation process is gained by utilizing results from the the mathematical study of aggregation functions.

The work of Grabisch et al. (2009) provides formal definitions and properties associated to aggregation functions as well examples of where inconsistencies can arise in the aggregation process. Although formal definitions are not provided here, mathematical properties of aggregation functions can be interpreted and applied to sustainability assessment (Pollesch and Dale, 2015). For example, compensatory behavior of indicators within aggregation, consistency of aggregation results across multiple sites using multiple indicators, and meaningful selection of aggregation functions based on measurability properties of the indicator can all be derived from basic mathematical properties of aggregation functions. Given that both consistent outputs and flexible usage are desired in the assessment, it is therefore a requirement that, when at all possible, the choice of aggregation functions and processes that underlay the assessment be tied to fundamental mathematical principles.

In addition to ensuring consistent aggregation, quantification of error of aggregate outputs given error of indicator measure is also desirable. In order to complete error analysis in the aggregation component, statistical approaches may be utilized in addition to results from the mathematical study of aggregation functions. Preliminary research into uncertainty quantification of aggregate values of the geometric mean on indicator measures normalized using the distance to target method is discussed in the section that follows. This preliminary work explores the possibility of describing sustainability indicators as random variables in order to utilize statistical techniques related to uncertainty quantification as well as to provide mechanisms for aggregation in the face of missing or poor quality indicator data.

For indicators or groups of indicators described using probability distributions, the possibility of using information theoretic aggregation has been investigated. Specifically, one may create mixture distributions where, initially, each component in the mixture is the probability distribution for a given indicator. This mixture distribution can be sampled, and from those samples, one may use information criteria[4] measures to identify less complex distributions to represent all the indicators that contributed to the data set. This method has thereby aggregated the information contained in the multiple distributions for indicators to a simpler distribution that may be used to describe the behavior of the multiple indicators included in the mixture.

[4]Information criteria quantify a balance between a distribution's ability to describe a data set, through use of a likelihood function, and the complexity of distribution parameters. Akaike Information Criteria, Schwarz Bayesian Information criteria, and Bozdogan's Information Complexity Criteria are a few of the criteria that can be used (Akaike, 1971; Schwarz et al., 1978; Bozdogan, 1990).

Table 1.3: Three non-methodological issues contributing to sustainability index use failure. (*Wilson et al. (2007), via Mayer (2008)*)

(1) A lack of consensus on what sustainability is in a quantitative sense
(2) Insufficient data availability to calculate indices correctly
(3) When indices do give clear advice, an unwillingness of policymakers to follow the advice

2.9. Analysis of Data Quality

Sustainability assessments, from development to utilization, demand a balance between theory and practicality. Detailed information about highly complicated systems through the use of thousands of indicators is often bypassed for a smaller set of practical, measurable, indicators. Even though feasibility is a priority in indicator selection (McBride et al., 2011; Dale and Beyeler, 2001), missing or poor quality data has been identified as a major concern for the successful application of assessment results. Specifically, Mayer (2008), identified "insufficient data availability to calculate indices correctly" as a an issue contributing to sustainability index use failure (Table 1.3), and Myllyviita et al. (2013) discuss the difficulty in obtaining reliable data for specific indicators included in their assessment of wood-based bioenergy sustainability.

If assessment results are to be utilized confidently, some indication of the uncertainty in data measures needs to be conveyed to stakeholders, researchers, and policymakers. This quantification of data quality would ideally capture uncertainty at the single indicator level and at any level of aggregation the assessment permits. Measures of confidence in single indicator measures benefit from a well established, statistical, methodology. However, methods to provide measures of uncertainty in aggregate values for diverse sets of indicators are less common and would benefit from further development.

2.10. Bio-STAR Requirements: Analysis of Data Quality

An ideal assessment conveys not only aspects of system function but also uncertainty in indicator measures and gives stakeholders quantification of data quality. Preliminary research indicates that some lesser known mathematical properties of aggregation functions as well as statistical approaches that utilize random variables to describe the distribution of indicators aid in the development of this key component of sustainability assessment.

One may look to multivariate statistical approaches as well as to the the study of aggregation functions to capture uncertainty through aggregation and to provide confidence measures for sustainability assessment results. With respect to aggregation theory, some basic properties of the aggregation function(s) employed in the assessment may be utilized for defining bounds on aggregate output. Specifically, Grabisch et al. (2009) discuss Lipschitz continuity of the arithmetic mean, which can be used to derive bounds on aggregate values given relative error of input values. Proofs for bounds using the arithmetic mean for normalized indicators with relative errors are given in Pollesch and Dale (2015). An understanding of error bounds for aggregate values in the assessment is relevant for both uncertainty quantification and validation in assessing the quality of sustainability indicator data.

As opposed to considering each set of indicator measurements as simply data points, one may instead choose to determine underlying statistical distributions for each indicator, as was mentioned in section 2.8 above. For example, after normalization using distance to target, indicator values belong

in the interval between 0 and 1; this leads naturally to description of the normalized indicator data using a Beta random variable (see Figure 1.2). Furthermore, aggregation on the normalized measures that are conducted using internal[5] aggregation functions, such as the arithmetic or geometric mean, are also in the interval between 0 and 1 and may be described using a Beta distribution. Although the process of fitting data to distributions of random variables adds complexity, once indicators are described as random variables, the researcher can then utilize previously developed statistical and probabilistic techniques to assess and convey aspects of data quality within the sustainability assessment. However, parameterizing distributions for indicators and having confidence in resulting distribution fits depends on having multiple replicates of indicator measurements to begin the process.

2.11. Evaluation of Assessment Results

The process of evaluating assessment results may be undertaken through a variety of sensitivity analysis techniques. These techniques can focus on a single indicator or on the entire suite of indicators and can employ statistical or analytical techniques (Saltelli et al., 2008). The quantification of interrelationships between indicators both before and after aggregation (i.e. determining a correlation structure in the assessment), and the analysis of the comparability of assessment results given choices in normalization and weighting of indicators are seen as two major aspects in the evaluation of sustainability assessment results. A proper treatment of these two aspects can aid researchers in the interpretation of assessment outcomes and guide appropriate presentation of assessment results.

2.11.1. Quantifying Interrelationships of Indicators

When sustainability assessments use composite or aggregate values, analyses of particular importance are those that impart an understanding of the interplay between single (or groups) of indicators and the aggregate value used to represent them. In addition to traditional sensitivity analysis techniques, the study of aggregation functions provides techniques for sensitivity analysis of aggregate assessment values. Specifically, Grabisch et al. (2009) provide theoretical frameworks for exploration of the sensitivity properties of the aggregation functions chosen through the use of an *importance index*. The importance index quantifies input components to the aggregation function that have the largest impact on the output value. The mathematical formulation of the importance index is given in Table 1.4.

The interrelationship of indicator variables is a topic that has been discussed in assessment literature for some time. Different researchers have argued different sides for an "ideal" correlation structure between indicator variables (Bollen and Lennox, 1991). Well-known statistical techniques may be utilized to investigate correlation between indicator measures. When it comes to aggregate values and the relationships between indicator variables and aggregate outputs, analysis can take place using the *interaction index* of two indicators with respect to a given aggregation function. Specifically, if the *interaction* of indicators i, j is positive, the contribution to the aggregate is high only if both i, j are high values. If the interaction of i, j is negative, then the contribution to the aggregate is high if one of the values is high. The interaction of indicators i, j is said to be null if the contribution of the indicators is the sum of the individual values (Grabisch et al., 2009, 2003). The

[5] The term *internal* refers to a specific mathematical property that essentially guarantees that the aggregate output will fall somewhere between the minimum and maximum values of all the input values. For a formal definition, see Grabisch et al. (2009)

Probability density function of Beta distribution with parameters α and β, with $\alpha, \beta > 0$:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

for $0 < x < 1$ where $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$

Probability density functions of Beta distribution under varying α and β parameter values:

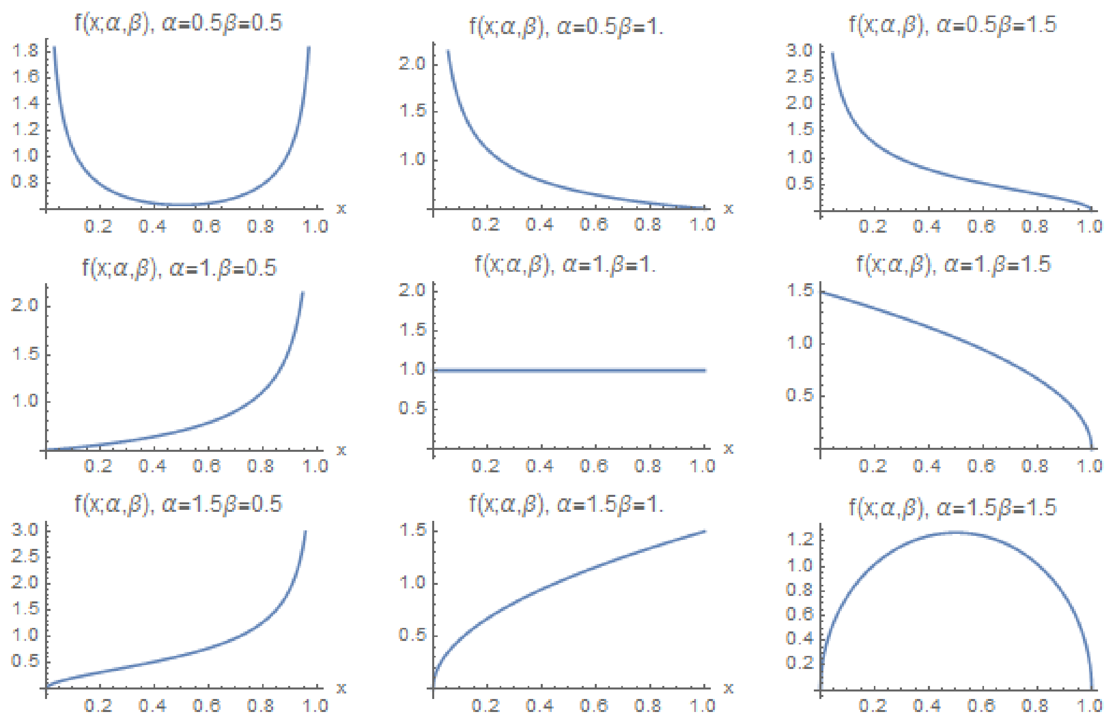


Figure 1.2: Beta distribution probability density function (PDF) and plots of PDF for various parameters to show flexibility of distribution to describe normalized indicator measurement values. Note: Beta(1,1) is the Uniform random variable on (0,1) and could represent a normalized indicator variable that is equally likely to take on any value between (0,1)

exact formulation of the interaction index is omitted from this paper but can be found in Grabisch et al. (2009).

2.11.2. Comparability of Assessment Results

Different systems have different environmental, social, or economic sensitivities that can demand special attention within an analysis. Comparability of assessment results is a major concern when dealing with multiple sites with large spatial or temporal variance leading to differences in baselines and targets for a given indicator in a given system. This variability has long been recognized, and weighting of indicator components is a common technique for quantifying these differences and treating them within the assessment. Weighting, however, as discussed in Section 2.5, is a contentious issue given its ability to introduce bias to cloud the interpretation of those results (Böhringer and Jochem, 2007).

Comparability of assessment results may be seen as directly trading-off with the flexibility of the assessment itself. Consider two sites that differ in their environmental sensitivities; one may picture bioenergy production in regions with dramatically different water resources. Southern California and Northern Wisconsin are examples of two areas with different levels of water stress (Shi et al., 2013; Padowski and Jawitz, 2012). Consider measurements of an indicator such as consumptive water use ($m^3/ha/day$), as identified by McBride et al. (2011), in both systems. Holding all other system indicator values constant, a 10% change in water use efficiency at the Northern Wisconsin bioenergy production site may be much less significant for the surrounding system than a 10% change in the Southern California system. In an assessment that treats all locations equally, these two sites would be indistinguishable. Mathematically, the results are comparable since the same numerical scheme is applied to both circumstances, however, practically speaking, one might question comparison of the results given the different known environmental risk of changing water use in both areas. If different targets and baselines are used for this indicator in these two locations, the assessment results would distinguish between the systems. However, in this case, numerical comparison of the results becomes clouded, since the results in each system are determined using different mathematical schemes. This interplay of comparability and flexibility of the assessment is an aspect that demands consideration in the development of the sustainability assessment technique.

2.12. Bio-STAR Requirements: Evaluation of Assessment Results

Within Bio-STAR, inclusion of an analysis of assessment outcomes is desired. Given that numerically, indicator interaction takes place within the aggregation and normalization steps, some methods developed within a mathematical aggregation theoretic framework can be utilized. The *importance index* for an aggregation function can be calculated to show the relative impact of each indicator on the aggregate output and serve to identify indicators that may have a disproportionate impact (see Table 1.4).

If the aggregation function chosen is *symmetric*[6] (such as the arithmetic or geometric means), then the importance index for all components is equal. In the case where a weighted arithmetic or weighted geometric mean is used, the aggregation function is not symmetric. Thus, when weights for the components are introduced to the aggregation function, the importance index is not be equal for all components. Given that no explicit weighting is used within the Bio-STAR assessment, the

[6]The term *symmetric* refers to aggregate output a function being the same if the order of inputs to the function change. For a formal definition, see Grabisch et al. (2009)

Table 1.4: Formal definition of the *importance index* used for assessing relative impact of indicators on aggregate output for a given aggregation function

The *Importance Index* (Grabisch et al., 2009)
Let $a, b \in \mathbb{R}$ such that $a < b$, let $\boldsymbol{A}$ be an integrable[a] aggregation function in $[a, b]^n$ and consider $i = 1, 2, ..., n$ (where n is the number of components being aggregated). **Definition:** The *(total) variation of* $\boldsymbol{A}$ *with respect to coordinate* i is the function $\Delta_i \boldsymbol{A} : [a, b]^n \to \mathbb{R}$ defined by: $$\Delta_i \boldsymbol{A}(\boldsymbol{x}) := \boldsymbol{A}(b_{\{i\}}\boldsymbol{x}) - \boldsymbol{A}(a_{\{i\}}\boldsymbol{x})$$ Note: $b_{\{i\}}\boldsymbol{x}$ and $a_{\{i\}}\boldsymbol{x}$ are the input vector, $\boldsymbol{x}$, with the i^{th} component replaced by the upper bound on the domain, b, and lower bound, a, respectively. This equation for total variation of a component i in aggregation function $\boldsymbol{A}$ is simply the difference in the aggregate output for a function when component i is set to the maximum input value, b, for that component and when component i is set to the minimum input value, a. **Definition:** The *importance index of coordinate* i *on* $\boldsymbol{A}$ is defined by $$\phi_i(\boldsymbol{A}) := \frac{1}{(b-a)^n} \int_{[a,b]^n} \frac{\Delta_i \boldsymbol{A}(\boldsymbol{x})}{b-a} d\boldsymbol{x}$$ Note: The importance index of component i with respect to the aggregation function $\boldsymbol{A}$ is a way of averaging the total variation over all values of $\boldsymbol{x} \in [a, b]$.
[a] An *integrable* aggregation functions is one for which the integral of the aggregation function (for intervals included in its domain) is a finite value. In this case, the indicators take values from a to b and $\boldsymbol{A}$ is integrable if $\int_{[a,b]^n} \boldsymbol{A}(\boldsymbol{x}) d\boldsymbol{x} < \infty$

importance index needs to be tied to both the range of normalized observations of the indicators as well as to the range of baseline and target measures. Currently there is no explicit formulation on how to use the importance index under this scenario, however, as Bio-STAR is developed, this use will be investigated. Pollesch and Dale (*in press*) provide a method for calculating *scale-modified implicit weights* for indicators within a distance-to-target normalization scheme, which considers both baseline and target values as well as relative impact on an aggregate under changes in non-normalized indicator measures. With similar information, the interaction index[7] will also be investigated as to its application for Bio-STAR analysis of data quality.

Given that assessment results often compare bioenergy production pathways, production systems, and a single system over different periods of time, comparability standards must accompany the results of the Bio-STAR or be built into the aggregation and normalization protocols. The analysis of systems, pathways, and the same systems at different points in time is differentiated in Bio-STAR by the baseline and target values assigned to each indicator. As it was discussed previously, implicit weights are a consequence of normalization through setting baselines and targets. Although normalization takes place to make indicator measurements more comparable, a benefit may be derived from the development of comparability standards as well as the determination of comparability thresholds. In order to attain meaningful comparisons of assessment results, one can make efforts to compare only results derived from *similarly* defined baselines and targets. A quantification of site *similarity* may be achieved through any number of clustering or grouping techniques applied to the targets and baselines provided. The inclusion of comparability standards can aid in the overall usefulness of the assessment resource by fostering meaningful comparisons and discouraging comparisons between bioenergy systems that are seen as fundamentally different. As with many other aspects of assessment development, what comparisons are meaningful and what *fundamentally different* means depends on assessment goals and desired use.

2.13. Visualization and Reporting

To be impactful, even the most meaningful assessment results depend on successful communication. Appropriate data presentation and visualization are major considerations in construction of an assessment. Relevant reporting has been identified as a *requirement for improving the state of sustainable development analysis* 1.5. Although assessment goals and the bioenergy system context often have the largest impact on how and what information is presented, there is a balance to be struck, much as in the indicator selection and aggregation steps, as to the information that is necessary and useful in the visualization and reporting of assessment results.

Just as assessment methodologies range in complexity, so do the techniques employed in visualization. Assessment visualizations range from simple tables of numerical indicator data to complicated graphical user interfaces. The use of tables in various arrangements is a common way in which bioenergy sustainability data are presented. Recently published examples include Hayashi et al. (2014), García et al. (2015), and Maes et al. (2015). Radar charts are commonly utilized where radii are used for different indicators within the assessment and the distance of a point from the center plotted along a given radius represents the measure or value of the indicator (Parish et al., 2016); although Dias and Domingues (2014) point to inconsistencies that can arise when areas of radar charts are used to provide aggregate values for multiple indicators. More elaborate visualization tools have also been created and adopted, such as the *Dashboard of Sustainability*

[7]The *interaction index* (*via* Grabisch et al. (2009)) is more complicated mathematically, so its exact mathematical definition is omitted from this paper.

with many nested levels of information (International Institute for Sustainable Development, 2001). For a collection of hundreds of graphical representations of the concepts behind sustainability, not specifically bioenergy sustainability, see the website of S. Mann and the article *Visualising Sustainability* (Mann, 2009). The clear communication of assessment results affects the usefulness of the assessment and even the assessment construction itself. As such, visualization and communication of results are critical to the development of any sustainability assessment methodology.

Normalization and aggregation are both methods for transforming data into measures that are more easily interpretable. Uniformity of scales and increasing comparability of indicator data is the goal of normalization, and lowering dimensionality and clarifying assessment results is the goal of aggregation. Both aggregation and normalization are steps in the reduction of complex, high-dimensional data sets; as such, they are also both directly related to visualization and communication of results strategies. In addition to dealing with measures from multiple dimensions for multiple indicators in a given system, each indicator may itself possess multiple attributes that are useful to convey. These indicator-specific attributes have relevance defined by assessment context and may lead to additional normalization and aggregation protocol specifications. For example soil quality measures may vary with soil type, slope, and land-use history; these contextual attributes for indicators may have varying importance within the assessment depending on assessment goals. Topics that have been discussed earlier, such as data variability within an indicator or within an aggregate measure, indicator weights, target and/or baseline levels for indicator, scale of measurement, and categorization related to system function are all pieces of information that one might want to report in the assessment result. Creative visualization strategies of assessment information can alleviate the need for data compression and the inherent loss of information that takes place when aggregation occurs.

2.14. Bio-STAR Requirements: Visualization and Reporting

Bio-STAR has overarching goals of adaptability for assessing multiple bioenergy contexts and flexibility in assessment to accommodate multiple analysis that a researcher or stakeholder may seek to undertake. These ideals necessitate an analysis structure that allows users to balance trade-offs in analysis such as compression of data with information loss, as well as the spatio-temporal trade-offs of studying the 'here-and-now' with the 'everywhere-and-forever'. As such, developing the reports and visualization of sustainability indicators and their analyses are challenging.

The visualization and reporting of Bio-STAR analyses seeks to provide access to the following system dimensions and aspects of the data. For individual bioenergy sustainability indicators:

1. Raw and normalized indicator measures
2. Spatio-temporal context for measurements
3. Indication of baselines and targets used for normalization
4. Quantification(s) of measurement uncertainty and measures of central tendency
5. Access to additional relevant indicator attributes and meta-data that accompany indicator measurements

Although the required dimensions of the system and data are mostly analogous for aggregate measures, the analysis of such aspects as aggregate spatial domain, aggregate temporal domain, uncertainty in aggregate value(s) can be significantly more complex. Additionally, as aggregate

measures are considered, further dimensions of the aggregate values may be of interest, such as importance index values, interaction index values, and a reporting of the correlation structure.

Adherence to appropriate reporting standards may also be considered a requirement for the proposed bioenergy sustainability assessment resource. At this point, no reporting standards for bioenergy sustainability data exist within the United States; perhaps an auxiliary product of Bio-STAR could be the beginning of development of such standards along with this assessment resource. Given that this tool will utilize the indicators for bioenergy sustainability provided by Dale et al. (2013a) and McBride et al. (2011), with respect to reporting one should note that Dale et al. (2013a) include *transparency* as a socioeconomic indicator for bioenergy sustainability assessment. In this case, *transparency* is defined as *the percent of indicators for which timely and relevant performance data are reported*. In this way, Bio-STAR itself is contributing towards sustainability goals.

Given the complexity of the problem, and challenges associated with the assessment of bioenergy sustainability data, visualization and reporting rely on development of custom software or the utilization of currently existing tools. Preliminary research has begun for the encoding and deployment of Bio-STAR protocols using the software package *Tableau* (http://www.tableau.com/). As the assessment resource is developed and desired functionality is encoded, the software, developed or otherwise, must be able to incorporate all of the various considerations identified not only in the assessment phase, but also all those specified within this visualization and reporting section. Software development and the actual deployment of the tool are challenging but benefit from this requirement specification and from rigorous protocol development.

3. Conclusion

Utilizing the structure of the seven key components, this paper addresses each key component for the development of Bio-STAR. Although the techniques identified within each key component have general applicability, the approaches outlined within the requirement specification for each key component are chosen to integrate research goals and the environmental and socioeconomic indicators for bioenergy sustainability provided by Dale et al. (2013a) and McBride et al. (2011). A summary of requirement specifications for Bio-STAR by key component is provided in Table 1.6.

This paper also reviews and addresses challenges currently identified in the sustainability assessment literature as well as anticipates those challenges that may arise in the construction of the desired sustainability assessment tool. Not only do the requirements specified for each key component build from the various guidelines and assessments that have been created in the past, they also introduce a set of mathematical concepts and tools that have not yet been utilized within sustainability assessment. Beginning with the Sustainability Assessment Resource for Bioenergy, this paper serves as a road-map for the construction of sustainability assessments that are adaptable for assessing diverse bioenergy production pathways, flexible enough to support a range of analyses, and mathematically robust with respect to normalization, aggregation, and the quantification of uncertainty.

4. Acknowledgments

This research was supported by the U.S. Department of Energy (DOE) under the Bioenergy Technologies Office and the University of Tennessee under subcontract 4000132470. Oak Ridge National Laboratory is managed by the UT-Battelle, LLC, for DOE under contract DE-AC05-00OR22725.

Table 1.5: Importance of reporting and visualization as included in "Requirements for improving the state of sustainable development analyses" (From *Towards a Science of Sustainability, Levin and Clark (2010)*)

(1) Fundamental research into what kinds of evaluation and monitoring systems are most needed
(2) Systems analysis of what is already being adequately measured and what is not at relevant scales
(3) Operational support for collaborative processes to design and put in the missing pieces
(4) **Synthesis efforts to report out the results in forms useful for decision support at relevant scales of management and governance**

Table 1.6: Key components applied to the development of Bio-STAR

Key Component	*Requirements Specified for Bio-STAR*
Identification of Goals and Extent	• Adaptability for assessing diverse bioenergy pathways • Flexibility to support range of analyses • Mathematical robustness with respect to normalization, aggregation, and uncertainty quantification • Assessment that considers entire bioenergy supply chain. Fuelshed to be investigated for spatial extent and per-year temporal extent
Indicator Selection and Categorization	• 19 Environmental indicators identified in McBride et al. (2011) • 16 Socioeconomic indicators identified in Dale et al. (2013a) • Indicator categorization: – Environmental indicators divided into 6 sub-groups; soil quality, water quality and quantity, greenhouse gases, biodiversity, air quality, and productivity – Socioeconomic indicators divided into 6 sub-groups; social-well being, energy security, trade, profitability, resource conservation, and social acceptability – Indicators classified by measurability scale, ratio scale measurability most prevalent
Data Normalization and Weighting	• Normalization using distance-to-target approach – Indicator's target and baseline are site-specific – Indicators normalized to values in interval [0,1] • No explicit weighting of indicators
Aggregation	• Assessment to support aggregation to level of user preference • Aggregation protocols to ensure implicit weights of indicators are equal within the overall aggregate • Measurability scale of data used to specify aggregation function • Information theoretic aggregation of indicator distributions to be researched
Analysis of Data Quality	• Confidence measures to be reported for – Variation in individual indicators – Variation in aggregates of indicators • Beta distributions to be studied for representing normalized indicator data
Evaluation of Assessment Results	• Aggregation theoretic sensitivity to be researched – Importance index – Interaction index • Comparability threshold to be developed using site-specific baseline and target definitions
Visualization and Reporting	• Visualization of data resolution/complexity to be specified by user • Indicator properties to display include: – Measurement variability/confidence – Spatial extent of measurement – Normalization baseline and target values

Chapter II

Normalization in Sustainability Assessment: Methods and Implications

Publication Statement:

A version of Chapter II has been submitted by its co-authors, Nathan L. Pollesch and Virginia H. Dale, and is currently in review for publication in the journal *Ecological Economics*. This manuscript was principally composed by N. Pollesch with writing contributed by V.H. Dale providing context for the research in literature and numerous comments on the manuscript in previous versions.

Normalization in Sustainability Assessment: Methods and Implications

N. Pollesch[a,b,*], V.H. Dale[b]

[a] *Department of Mathematics, The University of Tennessee, 1403 Circle Drive, Knoxville, TN 37996-1320*
[b] *Center for BioEnergy Sustainability, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6035*

Abstract

One approach to assessing progress towards sustainability makes use of multiple indicators spanning the environmental, social, and economic dimensions of the system being studied. Diverse indicators have different units of measurement, and normalization is the procedure employed to transform differing indicator measures onto similar scales or to unit-free measures. Given the inherent complexity entailed in interpreting information related to multiple indicators, normalization and aggregation of sustainability indicators are common steps after indicator measures are quantified. However, it is often difficult for stakeholders to make clear connections between specific indicator measurements and resulting aggregate scores of sustainability. Motivated by challenges and examples in sustainability assessment, this paper explores various normalization schemes including ratio normalization, target normalization, Z-score normalization, and unit equivalence normalization. Methods for analyzing the impacts of normalization choice on aggregate scores are presented. Techniques are derived for general application in studying composite indicators, and advantages and drawbacks associated with different normalization schemes are discussed within the context of sustainability assessment. Theoretical results are clarified through a case study using data from indicators of progress towards bioenergy sustainability.

Keywords: aggregation, bioenergy sustainability, composite indicator, normalization, sustainability assessment, standardization

---

*Corresponding author

1. Introduction

Sustainability is an inherently complex topic with different meanings pertaining to different contexts. Indicators of progress toward sustainability are measures that characterize conditions under which resource uses are more sustainable and are often tracked over time or compared for alternative practices. A variety of indicators of progress toward sustainability have been identified and both the breadth as well the number of indicators used for each assessment varies by application (Dale et al., 2013a; Mori and Christodoulou, 2012; McBride et al., 2011; Singh et al., 2009; Mayer, 2008). Indicators typically involve social, economic and environmental measures in order to capture the three major aspects of sustainability.

Sustainability assessments often rely on a variety of indicators. Different indicators are measured and reported in units pertinent to the particular metric. Having a common unit of measure is useful for comparison and synthesis of indicators. The synthesis of indicators can be done analytically, statistically, or graphically. Combining of measurements of multiple indicators to produce sustainability scores, composite indices, or aggregates is done to reduce dimensionality and can provide a single holistic value. Industry reports and national inventories are typically based on these highly aggregated data (Heijungs et al., 2007; Bare et al., 2006).

Normalization is the process of transforming units of measurement from the original units to common measurement units or to measurements that are unit less. This process is also referred to as unit scaling or standardization, with terminology varying based on the functions utilized in the process and by discipline. For clarity, this paper uses the term *normalization* to refer to all such processes transforming diverse units to common or unit-less quantities. When indicator units vary, normalization is seen as a necessary step prior to aggregation (Nardo et al., 2005). Mathematical research into the structure of sustainability assessments has focused on the aggregation step (Pollesch and Dale, 2015; Langhans et al., 2014; Roberts, 2014a; Zhou et al., 2006; Ebert and Welsch, 2004). Freudenberg (2003) gives a comparison of two different normalization procedures on a composite assessment outcome. Although researchers are often aware of the effect that a given choice of normalization scheme has on assessment outcome, no formal analysis of the implications of the normalization procedure on assessment outcome has emerged in the sustainability assessment literature.

In this paper the consequences of using different normalization functions within an aggregate score of sustainability are explored. The normalization and subsequent aggregation process used to derive composite sustainability scores vary greatly (Mori and Christodoulou, 2012; Singh et al., 2009; Mayer, 2008). Moldan et al. (2012) observe that for sustainability assessment "the selection and definition of pertinent indicators, to a large extent, defines the whole issue." Indicator selection may define the "whole issue," yet how measurements of those indicators are interpreted depends critically on the assessment structure. Even a well-defined set of indicators and accompanying high quality data can lead to completely different assessments of a system depending on the normalization and aggregation procedure employed. A recent example illustrating how something as arbitrary as the order in which indicators appear on a spider diagram can have large effect on sustainability ratings as calculated through surface area is given in Dias and Domingues (2014).

After terminology is established, this paper provides examples of normalization methods in sustainability assessment to show variety and form. Next, we move to an in-depth look at four common normalization procedures in combination with different aggregation functions to elucidate exact dependencies between aggregate sustainability scores and the normalization schemes utilized in their calculation. Through a case study of composite scores of bioenergy sustainability, the question of how changes in non-normalized indicator measures impact aggregate scores of sustainability is

addressed. The paper concludes with the discussion of advantages and drawbacks for normalization schemes, reinforced by results derived in the case study provided.

2. Normalization Methods

The normalization process is used throughout scientific research and is motivated by a variety of circumstances. In sustainability assessment, the major motivation for normalization is to transform measurements of indicators, typically obtained in different units, to a common unit of measurement to compare them or to prepare them for inclusion in an aggregate score of sustainability.

2.1. Terminology and Notation

There are a wide variety of functions that can be applied to data in order to normalize. To aid in the discussion and analysis of these different normalization procedures and functions, the underlying terminology is established below.

- *Indicator Bearing:* Sustainability indicators can differ as to whether smaller or larger values of the indicators are interpreted as being ideal[1] or if there is some ideal value from which the measure should not differ in magnitude too much. In this paper we use the term *indicator bearing* to describe this attribute of indicators. Indicator bearings are referenced through a variety of terminology in the literature, for example "direct correlation with utility" and "inverse correlation with utility" (Maxim, 2014). Krajnc and Glavič (2005) use "positive impact" or "negative impact," and Dias and Domingues (2014) use "criteria is to maximize" or "criteria is to minimize." In this paper, the terms *larger-the-better* (LTB), *smaller-the-better* (STB), and *distance-to-ideal* (DTI) are used given their straight-forward meaning. It should be noted that these are not the only types of indicator bearings for normalization and that some normalization strategies, such as *unit equivalence normalization* (Table 2.1), do not discriminate indicators in these regards.
- *Normalization Schemes:* Since each assessment may include indicators that are of many bearing types, families of normalization functions are used to take these differences into account. For example, if *target normalization* (Table 2.1) is employed, the form of the function applied differs by indicator bearing types. Families, or groups, of normalization functions are referred to here as *normalization schemes.* In the case where the normalization procedure does not discriminate among indicator bearing types, the scheme may consist of just a single function; Z-score normalization (Table 2.1) is an example.
- *Internal Normalization:* Another differentiating factor between normalization functions occurs if they use the entire data set for a given indicator to normalize a single measurement of that indicator. These normalization functions are referred to as *internal*[2]. Examples of this type of normalization function are ratio normalization functions for STB and LTB type indicators (Table 2.1). Normalization functions that are not internal depend on predefined, exogenous values, such as target and baseline levels or unit conversion factors.

[1]The term *ideal* can be interpreted in a variety of ways. Frequently, analysis goals and assessment context guide the interpretation as to what is considered an ideal and, correspondingly, a non-ideal or baseline measurement value.

[2]The term *internal*, with respect to a normalization function, is used to identify those functions that utilize the entire data set for an indicator to normalize any given measurement value from the set. This term is not to be confused with the *internality* of an aggregation function, which describes the aggregation function's compensatory behavior (see Pollesch and Dale (2015) or Grabisch et al. (2009) for a formal definition).

- Notation for normalized and non-normalized indicator measurements also varies. In this paper, non-normalized measures are denoted by a superscript '*'. Subscripts are used to convey a variety of information, such as which indicator is being considered and/or which measurement of that indicator is being referenced. For example, if measurement j of indicator i is normalized, the notation for the normalized measure would be x_{ij} and the non-normalized measure would be x_{ij}^*. For consistency and clarity, examples referenced in this paper are translated into this notation when possible.

2.2. Examples of Normalization in Sustainability Assessment

A plethora of normalization functions are utilized in sustainability assessment. Examples given next provide a glimpse into the variety and form.

Krajnc and Glavič (2005) propose two different normalization schemes, the second of which is employed in the *Sustainable Development Index*. The first scheme normalizes measurements relative to the average of the indicator measures (in this case a total of T measures have been taken over time) so that

$$x_i = \frac{x_i^*}{\frac{1}{T}\sum_{j=1}^{T} x_j^*}$$

The second scheme is given by

$$x_i = \frac{x_i^* - \min_j \{x_j^*\}}{\max_j \{x_j^*\} - \min_j \{x_j^*\}} \text{ and } x_i = 1 - \frac{x_i^* - \min_j \{x_j^*\}}{\max_j \{x_j^*\} - \min_j \{x_j^*\}}$$

for larger-the-better and smaller-the-better type indicators, respectively. All the normalization functions from Krajnc and Glavič (2005) given above are internal.

In the *Holistic Sustainability Assessment Tool for Bioenergy* of Hayashi et al. (2014), the indicator measures are normalized by

$$x_i = \begin{cases} \frac{(x_i^* - T_i)}{(x_{max_i} - T_i)} & x_i^* > T_i \\ \frac{(x_i^* - T_i)}{(T_i)} & x_i^* \leq T_i \end{cases} \text{ and } x_i = \begin{cases} \frac{(x_i^* - T_i)}{(x_{max_i} - T_i)} & x_i^* > T_i \\ \frac{(x_i^* - T_i)}{(T_i)} & x_i^* \leq T_i \end{cases}$$

for larger-the-better and smaller-the-better indicators, respectively. The authors state that x_{max_i} is determined either by historical data or legislation and T_i is a threshold value. These functions score indicator measures from -1 to 1, with -1 being the least sustainable and 1 being the most sustainable. The score is 0 when the indicator is the same measure as the threshold value, T_i. This scheme can be internal or not depending on how x_{max_i} is defined.

Maxim (2014) uses the following functions for the sustainability assessment of electricity generation technology:

$$x_i = \frac{x_i^* - \min_j \{x_j^*\}}{\max_j \{x_j^*\} - \min_j \{x_j^*\}} \text{ and } x_i = \frac{\max_j \{x_j^*\} - x_i^*}{\max_j \{x_j^*\} - \min_j \{x_j^*\}}$$

for larger-the-better and smaller-the-better indicators, respectively. Thus, normalized values fall into the interval $[0, 1]$. This is another example of an internal normalization scheme.

Castoldi and Bechini (2010) use a set of normalization functions in their construction of an integrated sustainability assessment of cropping systems. Normalization is carried out by use of

Table 2.1: Common normalization function definitions and notations: Internal normalization functions, those for which the normalized value of x_j depends on the entire data set $\boldsymbol{x}^*$, and the normalization functions that create dimensionless quantities are identified

Scheme, notation, and definition	**Indicator Bearing**	**Internal**	**Dimensi-onless**
Ratio Normalization			
$R_{L,j}(\boldsymbol{x}^*) = \frac{x_j^*}{\max\{\boldsymbol{x}^*\}}$	LTB	✓	✓
$R_{S,j}(\boldsymbol{x}^*) = \frac{\min\{\boldsymbol{x}^*\}}{x_j^*}$	STB	✓	✓
$R_{D,j}(x_j^*, T) = \frac{\min\{x_j^*, T\}}{\max\{x_j^*, T\}}$	DTI		✓
Z-Score Normalization			
$Z_j(\boldsymbol{x}^*) = \frac{x_j^* - \bar{x}^*}{S_N(\boldsymbol{x}^*)}$	n/a	✓	✓
where $\bar{x}^* = \frac{1}{n}\sum_{j=1}^n x_j^*$, $S_N = \left(\frac{1}{n}\sum_{j=1}^n (x_j^* - \bar{x}^*)^2\right)^{1/2}$			
Unit Equivalence			
$C_j(x_j^*, c_f) = x_j^* c_f$	n/a		
where c_f is a *conversion factor* from x_j^*'s to desired units			
Target Normalization to Interval $[0, 1]$			
$T_{L,j}(x_j^*, T, B) = \begin{cases} 0, & x_j^* \leq B \\ 1 - \frac{T - x_j^*}{T - B}, & B < x_j^* < T \\ 1, & x_j^* \geq T \end{cases}$	LTB		✓
$T_{S,j}(x_j^*, T, B) = \begin{cases} 1, & x_j^* \leq T \\ 1 - \frac{x_j^* - T}{B - T}, & T < x_j^* < B \\ 0, & x_j^* \geq B \end{cases}$	STB		✓
$T_{D,j}(x_j^*, T, B_l, B_u) = \begin{cases} 1 - \frac{T - x_j^*}{T - B_l}, & B_l < x_j^* < T \\ 1, & x_j^* = T \\ 1 - \frac{x_j^* - T}{B_u - T}, & T < x_j^* < B_u \\ 0, & \text{else} \end{cases}$	DTI		✓

Note: LTB: Larger-the-better, STB: Smaller-the-better, DTI: Distance-to-ideal, $\boldsymbol{x}^* = \{x_1^*, x_2^*, ..., x_n^*\}$, T is a target or ideal value for a given indicator, B is a baseline or non-ideal value for a given indicator (B_l and B_u used when an upper and lower baseline are required), $\bar{x}^*$ is the sample mean, S_N is the sample standard deviation, and c_f is a conversion factor to change units of $\boldsymbol{x}^*$ to alternate units (ex. dollars or greenhouse gas equivalents).

Table 2.2: Benchmarking normalization function from *Pinar et al. (2014)*. Indicators are normalized to values between 0 and 1 based on expert judgments of their sustainability level

Normalized value	**Sustainability level**
0	Extremely unsustainable
0.25	Still not sustainable but not as severely as in the previous case
0.50	Discrete level of sustainability, but still far from target
0.75	Satisfactory level of sustainability, yet not on target
1	Fully sustainable

continuous simple functions such that $x_i = 1$ if x_i^* is within some range of sustainability optimality thresholds; x_i takes on values $(0, 1)$ for x_i^* measures between optimal and anti-ideal thresholds and takes on the value of 0 outside of the anti-ideal thresholds. This may be seen as a generalization on the *distance to ideal* normalization function within the target normalization scheme. This normalization scheme is not internal and depends on predefined optimal thresholds for indicators.

Sadamichi et al. (2012) convert all measures to greenhouse gas equivalents in their sustainability assessment of biomass utilization for energy in east Asian countries. Transformation of indicator measures to a different and common unit of measurement for comparison can also take place by transforming to monetary units, such as dollars, or embodied energy units, such as emJoules, see Odum et al. (2000) for example, and falls broadly under the normalization scheme that is referred to in this paper as *unit-equivalence normalization.*

The normalization method utilized in Pinar et al. (2014) for the *FEEM (Fondazione Eni Enrico Mattei) Sustainability Index* is termed *benchmarking.* A benchmarking function is defined that assigns a normalized value to each indicator based on its level of sustainability, determined by "reliable and authoritative literature and international legislation sources." Specifically, Pinar et al. (2014) use the function given in Table 2.2. The benchmarking normalization function is not an internal normalization function, as it depends on indicator values each being mapped to some value based on a qualitative valuation of their level of sustainability.

A variety of normalization procedures are employed in sustainability assessment, each of which has its own properties and unique impact on any aggregate measure of sustainability derived from the normalized measures. Although it is not within the scope of this paper to fully analyze the normalization functions provided above, the analysis and case study included in this paper sheds light on some of the behavior of common normalization procedures encountered, namely ratio normalization and target normalization to the interval [0,1]. The examples in the case study provide background for describing the ways in which normalization functions can be analyzed to understand important properties of their behavior.

2.2.1. Ratio and Target Normalization Schemes

Given that ratio normalization and target normalization to the interval [0,1] are used for analysis in this paper, a brief discussion of these two schemes is useful. Ratio normalization is named as such because measures are transformed by taking the ratio of individual measurements to extremal measurements (minimum or maximum) of the data set. When indicators are of smaller-the-better type, the minimum value from the data set is used to transform all other measurements, and

hence the minimum value normalizes to a value of 1. For larger-the-better type indicators, it is the maximum value that is used to transform all other measurements, and the maximum value normalizes to a value of 1. The smaller-the-better and larger-the-better normalization functions are internal, so normalized values do not have meaning relative to exogenous system-defined targets or baselines. Also, for indicators with these bearings, the normalized value's significance comes from their relation only to the extremal elements from the data set in which they belong.

Target normalization compares individual measurements to predefined baseline and target values. These values can be system specific and tied to the environmental or socioeconomic sensitivities of the system being studied. They can also be uniform values, such as those provided by government regulations in the case of baselines, that may apply to multiple systems included in the study. Arguments for linking sustainability assessment outcomes to target or ideal levels, which is what target normalization accomplishes, can be found in the work of Moldan et al. (2012), Stiglitz et al. (2009) and Mayer (2008). Moldan et al. (2012) argues that "The benefit of specific, quantitative, time bound targets is then straightforward: The indicators can be linked to them and interpreted clearly on a distance-to-target basis." Mayer (2008) states that " indicators are more helpful if they give information on the state of the system with respect to policy targets or biophysical limits." Unlike ratio normalization, extremal elements in a data set do not influence normalized values when target normalization is used.

Beyond advocating for the inclusion of targets and baselines in sustainability, there is some discussion about how to determine and assign specific target and baseline values. Moldan et al. (2012) discusses ways in which target levels can be defined for sustainability indicators and provides example resources for their definition. Specifically, that study cites EEAiS: Star Portal Smeets et al. (1999), Millennium Development Goals, Eurostat, and Organisation for Economic and Co-Operation and Development (OECD) as potential resources for target references (Smeets et al., 1999; Eurostat, 2009; Nations, 2010; OECD, 2003). Further discussion and comparison of these normalization methods are provided throughout this paper.

3. Analyzing Normalization Functions

Studying the mathematical structure of the normalization functions provides insights into the implications that a given choice of normalization scheme may have on sustainability assessment outcomes. The four normalization schemes that are considered in this paper are *ratio normalization, Z-score normalization, unit equivalence normalization, and target normalization to the interval [0,1]* (Table 2.1). This paper analyzes how changes in the original, non-normalized data for a given indicator can cascade to alter composite sustainability scores. This investigation has implications for, not only the sensitivity of aggregate outcomes based on the normalization scheme chosen, but also the implicit weight or impact that a given normalization scheme has on particular measures of indicators. The comparability of assessment results based on the normalization procedure employed is also discussed.

3.1. Internal Normalization

Whether a normalization function is internal or not can have a large impact on how changes in non-normalized values can affect the total aggregate outcome. This effect occurs because each normalized indicator measurement depends on the full data set for that indicator. Internal normalization functions from literature were identified in the previous section. Of the four normalization schemes defined in Table 2.1, Z-score normalization is an internal normalization function; and unit

equivalence normalization and target normalization to interval [0,1] are not. Within the ratio normalization scheme, the larger-the-better and smaller-the-better normalization functions are internal, while the distance-to-ideal function is not internal; hence internality is not necessarily a property of the normalization scheme but rather of individual normalization functions. In the case of ratio normalization functions, the use of $\min\{\boldsymbol{x}^*\}$ and $\max\{\boldsymbol{x}^*\}$ cause them to be internal, while Z-score normalization is internal from both the explicit use of the mean value, $\bar{x}^*$, and the calculation of the standard deviation, S_N. The case study provided in Section 5 motivates the importance of knowing if a normalization function is internal or not.

3.2. Derivatives of Normalization Functions

The goal of investigating changes in the output of some normalization function naturally leads to the calculation and investigation of the derivatives of the normalization functions. Table 2.3 presents derivatives of the functions included in the four normalization schemes from Table 2.1 with respect to an arbitrary j^{th} non-normalized measurement, x_j^* . Two important properties considered here are piecewise differentiability of the normalization functions and how the derivative function depends on the variable being differentiated.

Many of the normalization functions are piecewise-defined and thus are differentiated piecewise. For ratio normalization functions, the presence of the $\min\{\boldsymbol{x}\}$ and $\max\{\boldsymbol{x}\}$ have a particular influence on the calculation of the derivative. Specifically, for $R_{L,j}(\boldsymbol{x}^*)$, as long as the non-normalized value being changed is not the maximum of the data set, $\max\{\boldsymbol{x}\}$, and does not become the maximum of the data set, the derivative is a constant value $\frac{1}{\max\{\boldsymbol{x}\}}$. However, in the case where the value changing is the maximum (or becomes the maximum), the behavior is quite different. Thus a complete characterization of the derivative must take these different possibilities into account (see Section 6.1). Similarly, target normalization behavior changes as measurements near the target and baselines values are varied and surpass these thresholds. How normalized values change near these threshold values and the impact of this behavior on aggregate sustainability scores are shown in further detail in the example included in Section 6.

In both internal and non-internal cases, how the variable of differentiation x_j^* appears in the derivative function is important (see Section 5.4). For example, the derivative of $R_{S,j}(\boldsymbol{x}^*)$ has fundamentally different behavior from nearly all other normalization functions considered due to the appearance of $(x_j^*)^2$ in the derivative (see Table 2.3). This difference leads to an impact on an aggregate score of sustainability that varies depending both on the value that is changing and the magnitude of the change, whereas the impact of normalization functions whose derivatives do not contain an x_j^* term is proportional to the change alone. An example of this effect is shown in the case study and discussed further in Section 6 below.

3.3. Comparability and Normalization

Assessments are often created with the goal of comparing alternative scenarios, different systems, or the same system at different points in time. The normalization scheme chosen has an affect on the comparability of results. Internal normalization schemes transform indicator measures based on the values present only in a particular data set. For a very simple example, consider two systems that are to be assessed and compared through measurements of a single, smaller-the-better type indicator. Let the first system have values (2, 5, 6, 2, 10) and the second system have values (20, 50, 60, 20, 100) for the indicator measured. If ratio normalization is used, these two very different data sets would normalize to equivalent the measures (1/5, 1/3, 2/5, 1, 1). Z-score standardization in this case behaves identically, since both data sets normalize to measures of $(0, -\frac{3}{\sqrt{11}}, -\frac{3}{\sqrt{11}}, \frac{1}{\sqrt{11}}, \frac{5}{\sqrt{11}})$.

Table 2.3: Normalization function derivatives: Using functions defined in Table 2.1, change in normalized value with respect to a change in the data point, x_j^*, is presented

Change in normalized value with respect to change in x_j^*

Ratio Normalization

$$\frac{\partial}{\partial x_j^*}(R_{L,j}(\boldsymbol{x}^*)) = \begin{cases} \frac{1}{\max\{\boldsymbol{x}^*\}}, & x_j^* < x_k^* \forall k \neq j \\ 0, & \text{else} \end{cases}$$

$$\frac{\partial}{\partial x_j^*}(R_{S,j}(\boldsymbol{x}^*)) = \begin{cases} \frac{-\min\{\boldsymbol{x}^*\}}{(x_j^*)^2}, & x_j^* < x_k^* \forall k \neq j \\ 0, & \text{else} \end{cases}$$

$$\frac{\partial}{\partial x_j^*}(R_{D,j}(x_j^*, T)) = \begin{cases} \frac{1}{T}, & x_j^* < T \\ 0, & x_j^* = T \\ \frac{-T}{(x_j^*)^2}, & x_j^* > T \end{cases}$$

Z-Score Normalization

$$\frac{\partial}{\partial x_j^*}(Z_j(\boldsymbol{x}^*)) = \frac{\left(\left(\frac{1}{n}\sum_{j=1}^n (x_j^* - \bar{x}^*)^2\right)^{\frac{1}{2}}(1 - \frac{1}{n})\right) - \left(\frac{n^{-1/2}(x_j^* - \bar{x}^*)^2}{(\sum_{j=1}^n (x_j^* - \bar{x}^*)^2)^{1/2}}\right)\left(\frac{\sum_{k \neq j}(x_k^* - \bar{x}^*)}{-n(x_j^* - \bar{x}^*)} + (1 - \frac{1}{n})\right)}{\frac{1}{n}\sum_{j=1}^n (x_j^* - \bar{x}^*)^2}$$

Unit Equivalence Normalization

$$\frac{\partial}{\partial x_j^*}(C(x_j^*, c_f)) = c_f$$

Target Normalization to Interval [0,1]

$$\frac{\partial}{\partial x_j^*}(T_{L,j}(x_j^*, T, B)) = \begin{cases} \frac{1}{T-B}, & B < x_j^* < T \\ 0, & \text{else} \end{cases}$$

$$\frac{\partial}{\partial x_j^*}(T_{S,j}(x_j^*, T, B)) = \begin{cases} \frac{-1}{B-T}, & T < x_j^* < B \\ 0, & \text{else} \end{cases}$$

$$\frac{\partial}{\partial x_j^*}(T_{D,j}(x_j^*, T, B_l, B_u)) = \begin{cases} \frac{1}{T-B_l}, & B_l < x_j^* < T \\ \frac{-1}{B_u-T}, & T < x_j^* < B_u \\ 0, & \text{else} \end{cases}$$

Note: $\boldsymbol{x^*} = \{x_1^*, x_2^*, ..., x_n^*\}$, T is a target or ideal value for a given indicator, B is a baseline or non-ideal value for a given indicator (B_l and B_u used when an upper and lower baseline are required), $S_N = [\frac{1}{n}\sum_{j=1}^n (x_j - \bar{x})^2]^{1/2}$ is the sample standard deviation, $\bar{x}^* = \frac{1}{n}\sum_{j=1}^n x_j^*$ is the sample mean, and c_f is a conversion factor to change units of $\boldsymbol{x}^*$ to alternate units, such as dollars or greenhouse gas equivalents.

However, if these measures were normalized using unit equivalence normalization, the order of magnitude difference would be maintained. Depending on context and the indicator being measured, the order of magnitude difference showing up in normalized values may or may not be a necessary or desirable trait. For target normalization, the transformed values for an indicator depend not only on the individual measurement but also on targets and/or baseline(s) defined. If the same targets and baselines were used for both systems, the results would be distinguishable. Dependence, in both cases, of normalized values on targets and baselines leads to questions of comparability.

4. Normalization and Aggregate Measures of Sustainability

Thus far we have defined terminology, presented examples of common normalization functions, and shown a sample of the variety of these functions that can be found in the sustainability assessment literature. We have also provided derivatives of the functions included in four normalization schemes and defined relevant properties that can be used to classify types and behaviors of these normalization functions. Consideration now moves to how a change in a non-normalized value impacts an aggregate score of sustainability given a choice of normalization scheme and aggregation function(s). To carry out this analysis, its helps to place the normalization process into a context relevant to sustainability assessment.

Interpretation of composite sustainability scores is predicated by an understanding of the differential impacts of indicators. Langhans et al. (2014) present trade-off diagrams to show, for given aggregation functions and two indicators (each indicator represented by a single measurement), how much an increase in one indicator needs to be accompanied by an increase in the other indicator to have the same impact on the composite score. Pinar et al. (2014) also provide examples of how relative importance of indicators and interaction among indicators can be computed within their FEEM Sustainability Index.

In this paper, differential impacts, or sensitivities, of the aggregate sustainability score to changes in non-normalized indicator measurements are investigated by the use of derivative functions. Let $\mathcal{S}$ denote an aggregate sustainability score, where $\mathcal{S} = A(x_1, x_2, ..., x_n)$ for some aggregation function A. The change of the aggregate output, $\mathcal{S}$, with respect to a change in an indicator, x_i, is investigated by computing the partial derivative

$$\frac{\partial}{\partial x_i}\left(A(x_1, x_2, ..., x_n)\right) \tag{1}$$

However, the partial derivative in (1) assumes that there is one representative value for each indicator, x_i. In practice, the x_i values are often the aggregate of multiple measurements, j, for a given indicator, i. Adding in this detail, let $x_i = \alpha_i(x_{ij})$, for some aggregation function α_i that combines the various measurements for indicator i. Now we can ask how $\mathcal{S}$ is impacted by changes of individual indicator measurements, x_{ij}. Calculating the change in $\mathcal{S}$ as an indicator measurement, x_{ij}, changes leads to computing the partial derivative by use of the chain rule

$$\frac{\partial}{\partial x_{ij}}\left(A(\alpha_1(x_{1j}), \alpha_2(x_{2j}), ..., \alpha_n(x_{nj})\right) = A'(\alpha_1(x_{1j}), \alpha_2(x_{2j}), ..., \alpha_n(x_{nj}))\alpha_i'(x_{ij}) \tag{2}$$

where the prime notation '$\,'\,$' represents the partial derivative with respect to x_{ij}. Again, this is equation is often not representing the full picture because normalization is performed on individual indicator measurements before aggregation. Let $x_{ij} = f_i(x_{ij}^*)$ be the output of some normalization

Table 2.4: Arithmetic and geometric mean definitions and derivatives: For the arithmetic mean (AM), geometric mean (GM), weighted arithmetic mean (WAM), and weighted geometric mean (WGM) a change in aggregate value with respect to a change in an input component, x_i, is presented

$A(\boldsymbol{x})$: **Aggregation Function**	$\frac{\partial}{\partial x_i}(A(\boldsymbol{x}))$
Arithmetic Means	
$AM(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\frac{\partial}{\partial x_i}(AM(\boldsymbol{x})) = \frac{1}{n}$
$WAM(\boldsymbol{x}, \boldsymbol{w}) = \sum_{i=1}^{n} w_i x_i$	$\frac{\partial}{\partial x_i}(WAM(\boldsymbol{x}, \boldsymbol{w})) = w_i$
Geometric Means	
$GM(\boldsymbol{x}) = \prod_{i=1}^{n} (x_i)^{1/n}$	$\frac{\partial}{\partial x_i}(GM(\boldsymbol{x})) = \frac{1}{n}(\prod_{j \neq i} x_j)(\prod_{i=1}^{n} x_i)^{1/n-1}$
$WGM(\boldsymbol{x}, \boldsymbol{w}) = \prod_{i=1}^{n} (x_i^{w_i})$	$\frac{\partial}{\partial x_i}(WGM(\boldsymbol{x}, \boldsymbol{w})) = w_i x_i^{w_i - 1} \prod_{j \neq i} (x_j^{w_j})$

Note: $\boldsymbol{x} = \{x_1, x_2, ..., x_n\}$ and $\boldsymbol{w} = \{w_1, w_2, ..., w_n\}$ where $0 \leq w_i < 1$ and $\sum_{i=1}^{n} w_i = 1$

function, f_i, for indicator i that is acting on the raw indicator data, x_{ij}^*, where all raw data for a given indicator are normalized using the same normalization function, f_i. Thus, for a full treatment of how $\mathcal{S}$ is impacted by changes in non-normalized indicator data, we need to add this final detail to our derivatives. This leads to the following:

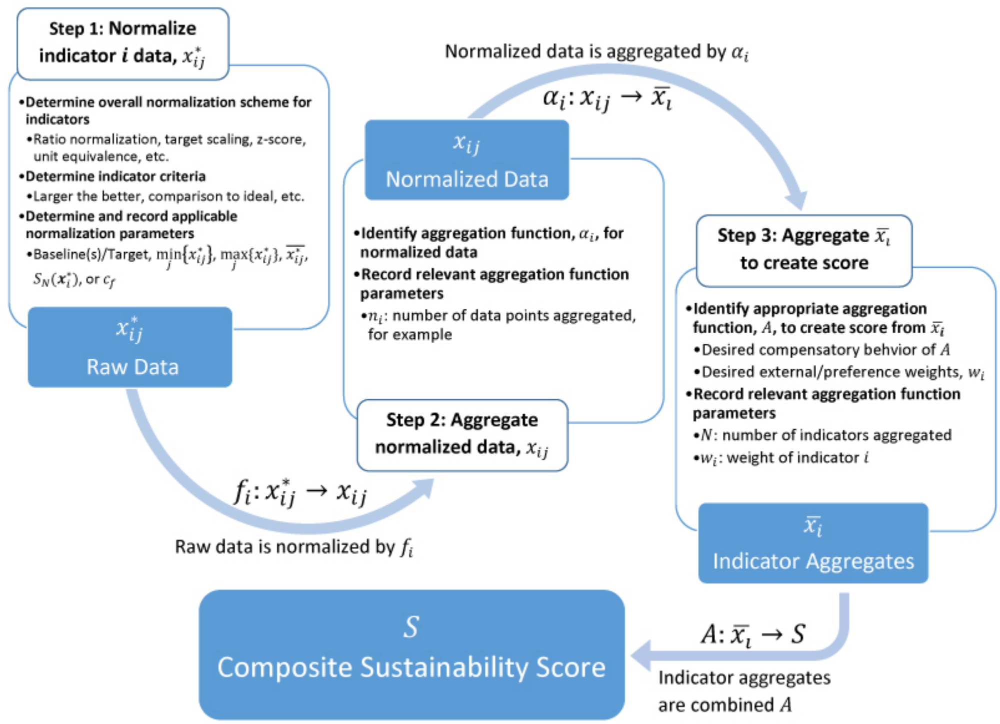
$$\frac{\partial}{\partial x_{ij}^*}\left(A(\alpha_1(f_1(x_{1j}^*)), \alpha_2(f_2(x_{2j}^*)), ..., \alpha_n(f_n(x_{nj}^*)))\right)$$
$$= A'(\alpha_1(f_1(x_{1j}^*)), \alpha_2(f_2(x_{2j}^*)), ..., \alpha_n(f_n(x_{nj}^*)))\alpha_i'(f_i(x_{ij}^*))f_i'(x_{ij}^*) \tag{3}$$

In this case, the prime notation ‘ $'$ ’ represents the partial derivative with respect to x_{ij}^*. Equation (3) is calculated to determine the impact of a variable, in this case a raw data measurement x_{ij}^*, that has been acted on by a normalization and aggregation function, f_i and α_i, respectively, before being acted on by A to determine the final output for $\mathcal{S}$. Although Equation (3) is beginning to look like a bit of a monster, if the aggregation functions A and α_i are those of common employ, such as the weighted or unweighted arithmetic or geometric mean, the derivatives are quite straightforward (see Table 2.4). The derivative in Equation (3) serves as the road map for the analysis that takes place in the case study that follows.

A summary of how raw data are transformed is as follows: First raw data, x_{ij}^*, for indicators are normalized using f_i, then normalized data, x_{ij}, are aggregated for each indicator using α_i, and finally multiple indicator aggregates, x_i, are combined using some aggregation function A to derive a sustainability score $\mathcal{S}$. Figure 2.1 summarizes this procedure and presents a flowchart that describes the process. Figure 2.1 also identifies examples of relevant properties and parameters for consideration at each step as measurements move from raw data, x_{ij}^*, to a sustainability score, $\mathcal{S}$.

5. Case Study

The following case study is presented to link the varying properties of normalization functions discussed in Section 3 to behaviors in example aggregate scores of sustainability. The sustainability



Notation: x_{ij}^* non-normalized measurement j of indicator i, f_i normalization function for indicator i, x_{ij} normalized measurement j for indicator i, α_i aggregation function for normalized measures of i, $\bar{x_i}$ aggregate of normalized measures of indicator i, and A is the aggregation function for indicator aggregates $\bar{x_1}, \bar{x_2}, ...$

Figure 2.1: Flowchart of a normalization and aggregation procedure utilized in multi-criteria sustainability assessment. Beginning with raw data for an indicator, data is transformed and aggregated multiple times before its eventual inclusion in a sustainability index or score

assessment structure outlined in Figure 2.1 is followed. This application uses Equation (3) to understand how changes in raw data impact a sustainability score, $\mathcal{S}$, under eight different scenarios. In this case each scenario is a choice of a normalization scheme, which determine functions f_i, and the choice of an aggregation function A, that is used to compute $\mathcal{S}$ from the combined normalized indicator measures. The scenarios included in the case study are the combinations of two different normalization schemes, ratio and target normalization to [0,1], with four different aggregation functions: the weighted and non-weighted arithmetic and geometric means.

5.1. Background Information on Assessing Progress Towards Bioenergy Sustainability

This case study builds from work to identify a limited set of indicators of progress toward sustainability for bioenergy systems and data collected for those indicators. Researchers at the Center for BioEnergy Sustainability at Oak Ridge National Laboratory have identified 35 indicators covering environmental, social, and economic aspects of sustainability of bioenergy systems (Dale et al., 2013a; McBride et al., 2011). Under research of the Southeastern Partnership for Integrated Biomass Supply Systems (IBSS), data were collected for a number of these indicators for switchgrass (*Panicum virgatum*) to examine actual yields and production costs under a wide range of physical settings and realistic farm management conditions in east Tennessee (Parish et al., 2016). Switchgrass is native to the southeastern United States and was planted within an eleven-county area to support a demonstration-scale ethanol production biorefinery in Vonore, Tennessee, that was operated by DuPont Cellulosic Ethanol (Tiller, 2011). To illustrate how normalization can affect aggregate scores of sustainability, this case study focuses on aggregation of three indicators: phosphorus levels in an adjacent water catchment (mg/L), yield of switchgrass (tons/acre), and percent organic matter (%OM) of soil in fields within the study using two normalization scenarios. The combination of these indicators has been chosen due to availability of high quality data sets. A comprehensive assessment of progress towards bioenergy sustainability would utilize many more of the 35 indicators identified in McBride et al. (2011) and Dale et al. (2013a). Given that the focus of this paper is on the analysis methodology presented, using this small set of indicators is done to aid in clarity and tractability. The approach developed in this paper is general and can be applied to a variety of assessments scenarios when aggregate scores are derived through normalized indicator measurements where the number of indicators, normalization scheme, and aggregation functions can vary.

5.2. Ratio Normalization Scenarios

In the ratio normalization scenarios, the derivation of the composite sustainability score $\mathcal{S}$ follows the procedure outlined below. Results are provided in Table 2.5.

1. Individual measures, x^*_{ij}, of each indicator are normalized to values, x_{ij}, under the function $x_{ij} = f_i(x^*_{ij})$. In this example, $f_1 = R_{S,j}$ (smaller-the-better, ratio normalization), $f_2 = R_{L,j}$ (larger-the-better, ratio normalization), and $f_3 = R_L, j$ (larger-the-better, ratio normalization). See Table 2.1 for definitions.

2. The arithmetic mean is employed to aggregate the normalized measurements, such that $\alpha_i(f_i(x^*_{ij})) = \frac{1}{n_i}\sum_{j=1}^{n_i} f_i(x^*_{ij})$, where n_i is the number of measurements for indicator i, to give an aggregate value for the normalized measures, which is denoted as $\bar{x}_i = \alpha_i(f_i(x^*_{ij}))$.

3. The aggregate normalized measures for each indicator, $\bar{x}_i$, are used to to calculate $\mathcal{S}$ under the aggregation function $A(\bar{x_1}, \bar{x_2}, \bar{x_3})$. In this case A is taken to be each of the arithmetic mean

Table 2.5: Ratio normalization of bioenergy sustainability indicators: Indicators, notation, and appropriate parameters for the indicator data sets are presented

x_i : Indicator (*units*)	n_i	$\min_j \{x^*_{ij}\}$	$\max_j \{x^*_{ij}\}$	$\bar{x_i}$
x_1 : Phosphorus (mg/L)	113	0.003	0.491	0.036
x_2 : Yield ($tons/acre$)	10	0	6.58	0.422
x_3 : % Organic Matter (%)	120	0	7.24	0.410
$\mathcal{S}$: Composite score derived from ratio normalized indicators				
$\mathcal{S} = AM(\bar{x_1}, \bar{x_2}, \bar{x_3}) = 0.290$				
$\mathcal{S} = GM(\bar{x_1}, \bar{x_2}, \bar{x_3}) = 0.185$				

Note: n_i is the number of measurements for indicator i, $\bar{x_i}$ is the arithmetic mean of the ratio normalized measurements of indicator i

(AM), geometric mean (GM), weighted arithmetic mean (WAM), and weighted geometric mean (WGM).

5.3. Target Normalization Scenarios

For the target normalization scenarios, the derivation of the composite sustainability score $\mathcal{S}$ follows the procedure outlined below. Notation has been changed from x to y in the target normalization scheme to distinguish between the two different normalization procedures; the underlying data sets for the indicators are the same in both cases. Results, along with baseline and target values used for the target normalization scenarios, are provided in Table 2.6.

1. Individual measures, y^*_{ij}, of each indicator are normalized to values, y_{ij}, under the function $y_{ij} = f_i(y^*_{ij})$. In this example, $f_1 = T_{S,j}$ (smaller-the-better, target normalization), $f_2 = T_{L,j}$ (larger-the-better, target normalization), and $f_3 = T_{L,j}$ (larger-the-better, target normalization). See Table 2.1 for definitions.

2. The arithmetic mean is employed to aggregate the normalized measurements, such that $\alpha_i(f_i(y^*_{ij})) = \frac{1}{n_i}\sum_{j=1}^{n_i} f_i(y^*_{ij})$, where n_i is the number of measurements for indicator i, to give an aggregate value for the normalized measures, which will be denoted as $\bar{y_i} = \alpha_i(f_i(y^*_{ij}))$.

3. The aggregate normalized measures for each indicator, $\bar{y_i}$, are used to to calculate $\mathcal{S}$ under the aggregation function $A(\bar{y_1}, \bar{y_2}, \bar{y_3})$. In this case A is taken to be each of the arithmetic mean (AM), geometric mean (GM), weighted arithmetic mean (WAM), and weighted geometric mean (WGM).

Baseline and target levels for yield (*tons/acre*) are derived from expert opinion based on extensive data for the case study as described in Parish et al. (2016). Phosphorus concentration (mg/L) target and baseline levels have been set to 0 (mg/L) and 0.1 (mg/L) based on what is considered a critical concentration (Walker, 2000). The values for % organic matter baseline and target levels are based on Brady et al. (1996), who provide 1.5%-4.0% as a range of values for %OM in Ultisol, the dominant soil type in the bioenergy cropping region for this case study.

Table 2.6: Target normalization of bioenergy sustainability indicators: Indicators, notation, and appropriate parameters for the indicator data sets are presented

y_i : Indicator (*units*)	n_i	Baseline	Target	$\bar{y_i}$
y_1 : Phosphorus (mg/L)	113	$B = 0.1$	$T = 0$	0.123
y_2 : Yield ($tons/acre$)	10	$B = 0$	$T = 8$	0.348
y_3 : % Organic Matter (%)	120	$B = 1.4$	$T = 4$	0.545
$\mathcal{S}$: **Composite score derived from target normalized indicators**				
$\mathcal{S} = AM(\bar{y_1}, \bar{y_2}, \bar{y_3}) = 0.339$				
$\mathcal{S} = GM(\bar{y_1}, \bar{y_2}, \bar{y_3}) = 0.286$				

Note: The notation y_i is used to distinguish between the two different normalization procedures, the underlying data set for each indicator is the same in both cases. n_i is the number of measurements for indicator i, $\bar{y_i}$ is the arithmetic mean of the target normalized measurements of indicator i.

5.4. Quantifying Impacts of Indicator Measurements

Aggregate values for each indicator and for the composite sustainability score $\mathcal{S}$ are different when internally normalized by ratio normalization and when tied to external target and baseline values in the target normalization process. These differences are to be expected. Comparing and applying meaning to the different aggregate results derived from ratio and target normalization is not recommended, given how different the two normalization approaches are. However, what can be contrasted is how the composite score of sustainability, $\mathcal{S}$, is impacted, as non-normalized data measures change in each normalization scenario explored.

The impact, or weight, that individual indicator measurements carry into the score $\mathcal{S}$ can become unclear in composite scores of sustainability. For example, in this case study, one may ask if the measurements of phosphorus are having more influence on $\mathcal{S}$ than the measurements of yield? One might also ask, what role the normalization and aggregation functions chosen have on determining any differential impacts on $\mathcal{S}$ for specific indicators? Tables 2.5 and 2.6 give values for $\mathcal{S}$ as calculated through the arithmetic and geometric mean for the ratio and target normalization schemes. Tables 2.7 and 2.8 give the partial derivatives of those scores with respect to changes in non-normalized indicator measures for each of the three indicators as calculated using Equation (3); these derivatives serve as the starting point in elucidating differing impact on changes in $\mathcal{S}$ from different normalization schemes. The derivatives in Tables 2.7 and 2.8 give exact formulas for analysis; the plots given in Figures 2.2 and 2.3 provide another way to visualize differences in impact of changes in the non-normalized indicator measures given the different normalization functions applied. All derivative functions can be calculated with respect to an arbitrary non-normalized measurement x^*_{ij}; however, in order to create the visualizations in Figures 2.2 and 2.3, a specific indicator measurement in the data set must be chosen to vary. In this case the median value was chosen and varied, and the corresponding value of $\mathcal{S}$ was calculated and plotted. Together, the derivatives and the visualizations provide two tools that can be used to study how normalization affects this composite score of sustainability.

With the results presented in Tables 2.5 through 2.8, one can see not only how the different normalization functions affect $\mathcal{S}$ but also how the aggregation functions for individual indicators, α_i, and the aggregation function A affect the value of $\mathcal{S}$ as non-normalized measures are changed. As discussed in Section 3.2, given the piecewise definition of many of the normalization functions, a similar piecewise definition of the derivatives is needed. For clarity, the derivatives presented in Tables 2.7 and 2.8 represent the behavior of $\mathcal{S}$ for x^*_{ij} values changing away from the minimum and maximum values, for the ratio normalization scheme, and in between the target and baseline values, for the target normalization scheme. Further discussion of how $\mathcal{S}$ changes as non-normalized values take on the minimum and maximum values is given in detail in Section 6.1.

6. Discussion

In order to understand the different influences that a normalization function can have on a composite sustainability score, properties of normalization functions have been discussed, and an analysis using partial derivatives of the aggregation and normalization functions has been presented. Even though calculation of the derivatives shown in Table 2.7 and Table 2.8 adds a step to the assessment process, the application and interpretation of the quantities derived improves overall understanding of the composite score for the sustainability indicators.

The first useful information added by calculating the derivatives is nearly by definition; derivatives indicate the per unit change in the composite score, $\mathcal{S}$, due to a per unit change in a non-

Table 2.7: Change in composite scores of ratio normalized bioenergy sustainability indicators as a function of non-normalized indicator measurements. The weighted arithmetic and geometric mean derivatives have been left in a general form to show influence of the weights, w_i, without need of a particular specification

$\frac{\partial}{\partial x^*_{ij}}(\mathcal{S})$: Change in composite score with respect to change in x^*_{ij}

Arithmetic Mean : $\mathcal{S} = AM$

$$\frac{\partial}{\partial x^*_{1j}}(AM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (\tfrac{1}{3})(\tfrac{1}{n_1})\left(\frac{-\min_j\{x^*_{1j}\}}{(x^*_{1j})^2}\right) = (-8.85 \times 10^{-6})(x^*_{1j})^{-2}$$
$$\frac{\partial}{\partial x^*_{2j}}(AM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (\tfrac{1}{3})(\tfrac{1}{n_2})\left(\frac{1}{\max_j\{x^*_{2j}\}}\right) = 0.0051$$
$$\frac{\partial}{\partial x^*_{3j}}(AM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (\tfrac{1}{3})(\tfrac{1}{n_3})\left(\frac{1}{\max_j\{x^*_{3j}\}}\right) = 0.0004$$

Geometric Mean : $\mathcal{S} = GM$

$$\frac{\partial}{\partial x^*_{1j}}(GM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (\tfrac{1}{3})(\bar{x_2}\bar{x_3})^{1/3}(\bar{x_1})^{-2/3}(\tfrac{1}{n_2})\left(\frac{-\min_j\{x^*_{1j}\}}{x^*_{1j}}\right) = -(4.33 \times 10^{-6})(x^*_{1j})^{-2}(\bar{x_1})^{-2/3}$$
$$\frac{\partial}{\partial x^*_{2j}}(GM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (\tfrac{1}{3})(\bar{x_1}\bar{x_3})^{1/3}(\bar{x_2})^{-2/3}(\tfrac{1}{n_2})\left(\frac{1}{\max_j\{x^*_{2j}\}}\right) = 0.0012(\bar{x_2})^{-2/3}$$
$$\frac{\partial}{\partial x^*_{3j}}(GM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (\tfrac{1}{3})(\bar{x_1}\bar{x_2})^{1/3}(\bar{x_3})^{-2/3}(\tfrac{1}{n_3})\left(\frac{1}{\max_j\{x^*_{3j}\}}\right) = 0.0001(\bar{x_3})^{-2/3}$$

Weighted Arithmetic Mean : $\mathcal{S} = WAM$

$$\frac{\partial}{\partial x^*_{1j}}(WAM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (w_1)(\tfrac{1}{n_1})\left(\frac{-\min_j\{x^*_{1j}\}}{x^*_{1j}}\right)$$
$$\frac{\partial}{\partial x^*_{2j}}(WAM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (w_2)(\tfrac{1}{n_2})\left(\frac{1}{\max_j\{x^*_{2j}\}}\right)$$
$$\frac{\partial}{\partial x^*_{3j}}(WAM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (w_3)(\tfrac{1}{n_3})\left(\frac{1}{\max_j\{x^*_{3j}\}}\right)$$

Weighted Geometric Mean : $\mathcal{S} = WGM$

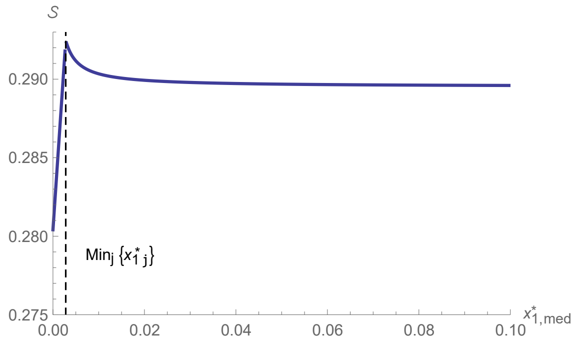
$$\frac{\partial}{\partial x^*_{1j}}(GM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (w_1)(\bar{x_2}^{w_2}\bar{x_3}^{w_3})(\bar{x_1})^{1-w_1}(\tfrac{1}{n_2})\left(\frac{-\min_j\{x^*_{1j}\}}{x^*_{1j}}\right)$$
$$\frac{\partial}{\partial x^*_{2j}}(GM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (w_2)(\bar{x_1}^{w_2}\bar{x_3}^{w_3})(\bar{x_2})^{1-w_2}(\tfrac{1}{n_2})\left(\frac{1}{\max_j\{x^*_{2j}\}}\right)$$
$$\frac{\partial}{\partial x^*_{3j}}(GM(\bar{x_1}, \bar{x_2}, \bar{x_3})) = (w_3)(\bar{x_1}^{w_1}\bar{x_2}^{w_2})(\bar{x_3})^{1-w_3}(\tfrac{1}{n_3})\left(\frac{1}{\max_j\{x^*_{3j}\}}\right)$$

Note: The derivatives above hold for the case when, for indicator 1, $x^*_{1j} > x^*_{1k} \forall k \neq j$, otherwise $x^*_{1j} = \min_j x^*_{1j}$ and thus takes on the constant normalized value of 1, and thus the derivative is 0. For indicators 2 and 3, the case is similar, but the derivatives hold when $x^*_{2j} < x^*_{2k}$ and $x^*_{3j} < x^*_{3k} \forall k \neq j$. The weights must satisfy $0 \leq w_i < 1$ and $\sum_{i=1}^{n} w_i = 1$.

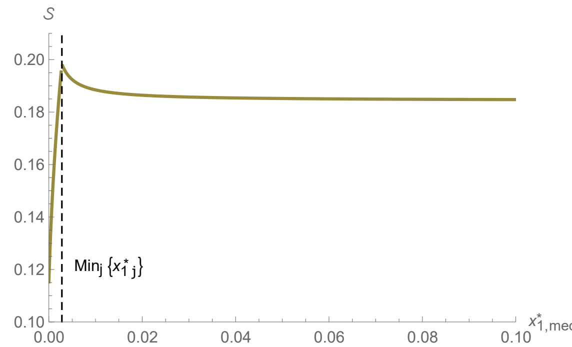
Table 2.8: Change in composite scores of target normalized bioenergy sustainability indicators as a function of non-normalized indicator measurements. The weighted arithmetic and geometric mean derivatives have been left in a general form to show influence of the weights, w_i, without need of a particular specification

$\frac{\partial}{\partial y^*_{ij}}(\mathcal{S})$: **Change in composite scores with respect to change in y^*_{ij}**
Arithmetic Mean : $\mathcal{S} = AM$
$\frac{\partial}{\partial y^*_{1j}}(AM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (\frac{1}{3})(\frac{1}{n_1})(\frac{-1}{B_1-T_1}) = -0.0295$
$\frac{\partial}{\partial y^*_{2j}}(AM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (\frac{1}{3})(\frac{1}{n_2})(\frac{1}{T_2-B_2}) = 0.0042$
$\frac{\partial}{\partial y^*_{3j}}(AM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (\frac{1}{3})(\frac{1}{n_3})(\frac{1}{T_3-B_3}) = 0.0011$
Geometric Mean : $\mathcal{S} = GM$
$\frac{\partial}{\partial y^*_{1j}}(GM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (\frac{1}{3})(\bar{y_2}\bar{y_3})^{1/3}(\bar{y_1})^{-2/3}(\frac{1}{n_1})\left(\frac{-1}{B_1-T_1}\right) = -0.0169(\bar{y_1})^{-2/3}$
$\frac{\partial}{\partial y^*_{2j}}(GM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (\frac{1}{3})(\bar{y_1}\bar{y_3})^{1/3}(\bar{y_2})^{-2/3}(\frac{1}{n_2})\left(\frac{1}{T_2-B_2}\right) = 0.0017(\bar{y_2})^{-2/3}$
$\frac{\partial}{\partial y^*_{3j}}(GM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (\frac{1}{3})(\bar{y_1}\bar{y_2})^{1/3}(\bar{y_3})^{-2/3}(\frac{1}{n_3})\left(\frac{1}{T_3-B_3}\right) = 0.0004(\bar{y_3})^{-2/3}$
Weighted Arithmetic Mean : $\mathcal{S} = WAM$
$\frac{\partial}{\partial y^*_{1j}}(AM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (w_1)(\frac{1}{n_1})(\frac{-1}{B_1-T_1})$
$\frac{\partial}{\partial y^*_{2j}}(AM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (w_2)(\frac{1}{n_2})(\frac{1}{T_2-B_2})$
$\frac{\partial}{\partial y^*_{3j}}(AM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (w_3)(\frac{1}{n_3})(\frac{1}{T_3-B_3})$
Weighted Geometric Mean : $\mathcal{S} = WGM$
$\frac{\partial}{\partial y^*_{1j}}(GM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (w_1)(\bar{y_2}^{w_2}\bar{y_3}^{w_3})(\bar{y_1})^{1-w_1}(\frac{1}{n_1})\left(\frac{-1}{B_1-T_1}\right)$
$\frac{\partial}{\partial y^*_{2j}}(GM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (w_2)(\bar{y_1}^{w_1}\bar{y_3}^{w_3})(\bar{y_2})^{1-w_2}(\frac{1}{n_2})\left(\frac{1}{T_2-B_2}\right)$
$\frac{\partial}{\partial y^*_{3j}}(GM(\bar{y_1}, \bar{y_2}, \bar{y_3})) = (w_3)(\bar{y_1}^{w_1}\bar{y_2}^{w_2})(\bar{y_3})^{1-w_3}(\frac{1}{n_3})\left(\frac{1}{T_3-B_3}\right)$

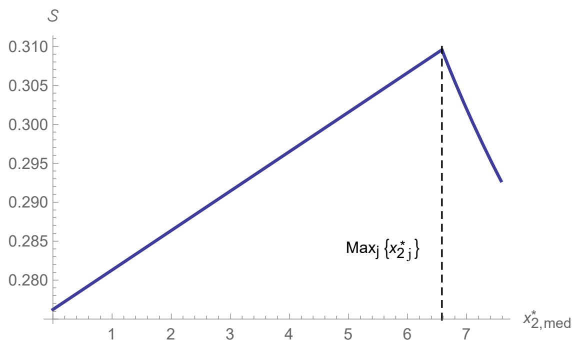
Note: These derivatives hold for the case when y^*_{ij} falls within the interval created by the targets (T_i) and baselines (B_i) for the respective indicators. The weights must satisfy $0 \leq w_i < 1$ and $\sum_{i=1}^{n} w_i = 1$.



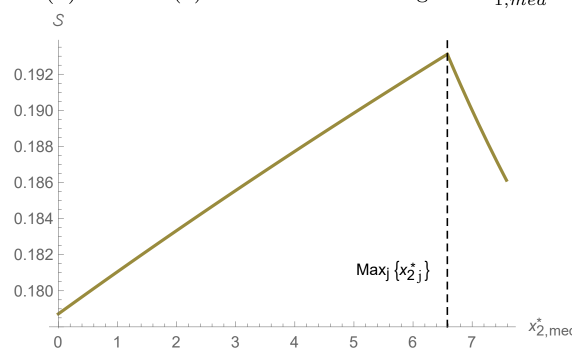
(a) $\mathcal{S} = AM(\bar{x})$ as a function of changes in $x^*_{1,med}$



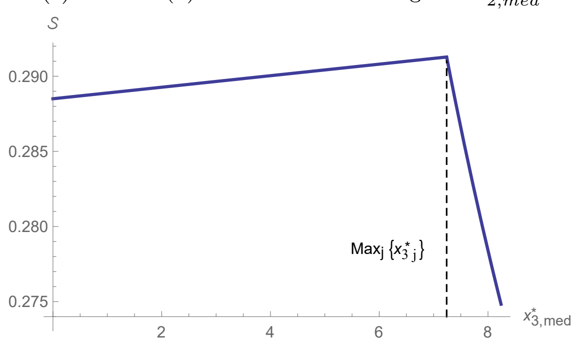
(b) $\mathcal{S} = GM(\bar{x})$ as a function of changes in $x^*_{1,med}$



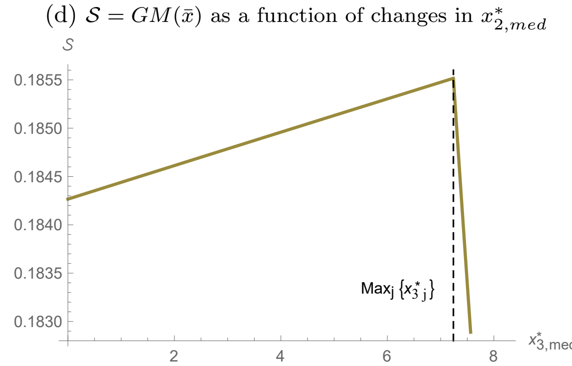
(c) $\mathcal{S} = AM(\bar{x})$ as a function of changes in $x^*_{2,med}$



(d) $\mathcal{S} = GM(\bar{x})$ as a function of changes in $x^*_{2,med}$

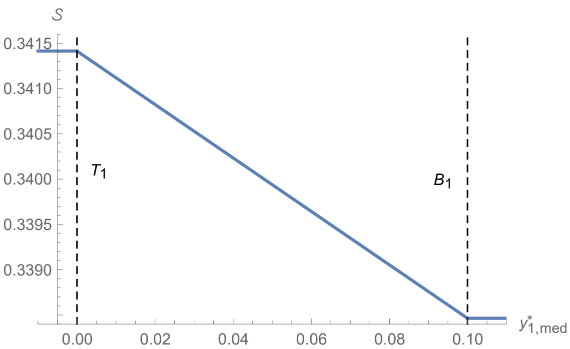


(e) $\mathcal{S} = AM(\bar{x})$ as a function of changes in $x^*_{3,med}$

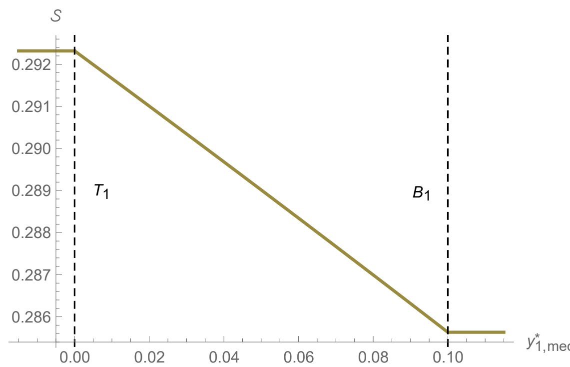


(f) $\mathcal{S} = GM(\bar{x})$ as a function of changes in $x^*_{3,med}$

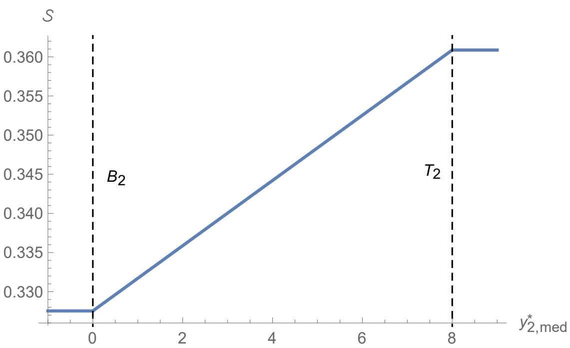
Figure 2.2: Changes in $\mathcal{S}$ in response to changes in median values of from the data sets for phosphorus ($x^*_{1,med}$), yield ($x^*_{2,med}$), and %OM ($x^*_{3,med}$) under ratio normalization scheme. Dashed lines show $\min_j x^*_{ij}$ and $\max_j x^*_{ij}$ values from Table 2.5. In all cases, the median value from each data set was varied in order to show the effect that changing non-normalized indicator measures has on $\mathcal{S}$. Notice the behavior of $\mathcal{S}$ when the median value becomes the $\min_j x^*_{ij}$ in (a,b) and the $\max_j x^*_{ij}$ in (c,d,e,f). This dramatic change is due to the fact that ratio normalization is an internal normalization process, and the dependence of all normalized values in the data set on the minimum or maximum value of that data set. All functions depicted correspond to functions presented in Table 2.7



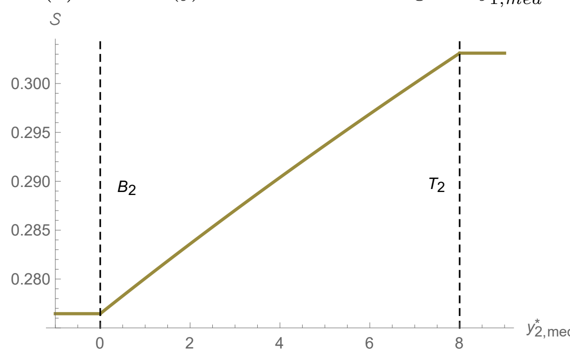
(a) $\mathcal{S} = AM(\bar{y})$ as a function of changes in $y^*_{1,med}$



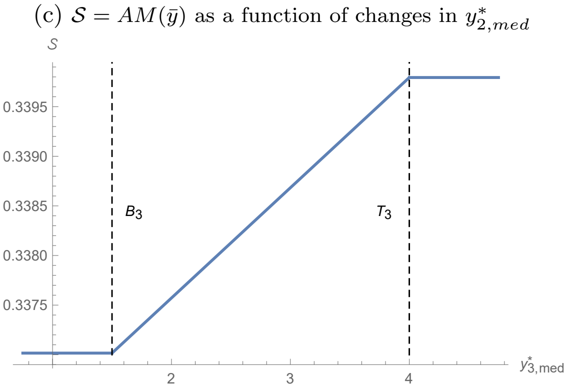
(b) $\mathcal{S} = GM(\bar{y})$ as a function of changes in $y^*_{1,med}$



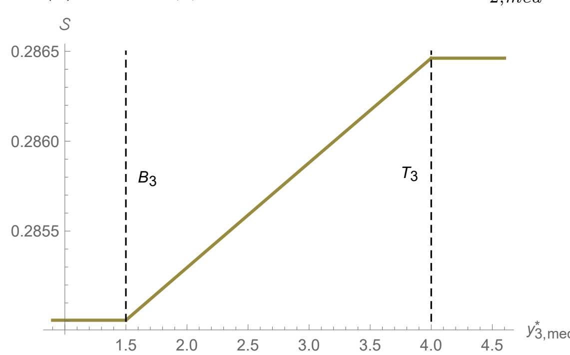
(c) $\mathcal{S} = AM(\bar{y})$ as a function of changes in $y^*_{2,med}$



(d) $\mathcal{S} = GM(\bar{y})$ as a function of changes in $y^*_{2,med}$



(e) $\mathcal{S} = AM(\bar{y})$ as a function of changes in $y^*_{3,med}$



(f) $\mathcal{S} = GM(\bar{y})$ as a function of changes in $y^*_{3,med}$

Figure 2.3: Changes in $\mathcal{S}$ in response to changes in median values of from the data sets for Phosphorus ($y^*_{1,med}$), Yield ($y^*_{2,med}$), and %OM ($y^*_{3,med}$) under target normalization scheme. In all cases, the median value from each data set was varied in order to show the effect that changing non-normalized indicator measures has on $\mathcal{S}$. Dashed lines show normalization parameters from Table 2.6. When the median value is changed to values outside the baseline and target intervals, there is no response in $\mathcal{S}$ to further changes due to the normalized value becoming a constant 0 or 1. All functions depicted correspond to functions presented in Table 2.8

normalized measurement, x_{ij}^*. As such, differences in the derivatives quantify differential sensitivities in the sustainability score by indicator. For example, Table 2.7 shows that the composite score $\mathcal{S}$ is nearly 12 times as sensitive to a per unit change in a yield measurement as it is to a unit change in a %OM measure when ratio normalization and the arithmetic mean are used.

Beyond just the different sensitivities of the aggregate score, $\mathcal{S}$, with respect to changes in indicator measurements, Figures 2.2 and 2.3 show that different normalization procedures lead to fundamentally different ways in which indicator measurements impact the composite score. This difference is apparent not only across the normalization schemes presented, but fundamental differences can emerge within the same scheme. For example, setting aside behavior changes near the extremal values, within the ratio normalization scheme if an indicator is of the type smaller-the-better, then a change in a measurement of that indicator has an impact on the composite score that depends on the value of the measurement changing. This is due to the presence of x_{1j}^* term in the derivative (see Figures 2.2a and 2.2b). However, if the indicator is of the type larger-the-better, then the impact of a change in a measurement of that indicator on $\mathcal{S}$ is independent of the value that is changing (see Figures 2.2c, 2.2e, 2.2d, and 2.2f). With respect to the extremal values, it can also be seen that as the minimum (maximum) value changes in smaller-the-better (larger-the-better) type indicators, there is a dramatic change in the score of $\mathcal{S}$ (2.2). This change is due to the internal normalization property of the ratio normalization functions, and, when the minimum or maximum changes for a data set, all of the other measurements in the data set also change causing $\mathcal{S}$ to be very sensitive to changes in the extremal values of the data set.

In the target normalization scheme, Figure 2.3 shows that in both smaller-the-better and larger-the-better indicator bearing, behavior does not display the fundamental differences that can be be seen between these two bearings in the ratio normalization scheme. In the case of the final aggregate score calculated through the arithmetic mean, changes in $\mathcal{S}$ are constant when non-normalized indicator measures change between the baseline and target values. For the geometric mean, although the change between baseline and target values appears to be constant, it is not, as the derivatives in Table 2.8 show. In both cases changes in $\mathcal{S}$ are 0 when non-normalized measures change beyond the baseline and target measures defined. Target normalization is free from the dramatic changes in $\mathcal{S}$ that appear in ratio normalization as extremal values change in the data set. In target normalization, if a non-normalized measure changes to a value beyond the baseline or target, it ends up having no impact on $\mathcal{S}$ because the normalized value for that measurement is constant at either 0 or 1 beyond those threshold values.

6.1. On the Piecewise Nature of Derivatives Encountered

All functions of the ratio normalization and target normalization schemes are piecewise differentiable; this fact leads to complicated behavior of the derivatives. Up to this point of the paper, there has been limited discussion on how to analyze the behavior of $\mathcal{S}$ when non-normalized measures change to become the minimum or maximum values in the ratio normalization scheme and when non-normalized measures move beyond the baseline and target values in target normalization.

Figure 2.2 addresses the behavior of $\mathcal{S}$ in the ratio normalization scenarios. Notice that in all six plots that the derivatives given in Table 2.7 hold until the indicator value becomes the maximum or minimum value for the data set, at which point one needs to consider how a change not just in x_{ij}^* impacts $\mathcal{S}$, given by $\frac{\partial \mathcal{S}}{\partial x_{ij}^*}$ but also how changing $\max_j \{x_{ij}^*\}$ impacts $\mathcal{S}$, given by $\frac{\partial \mathcal{S}}{\partial \max_j \{x_{ij}^*\}}$. In practice for ratio normalization, there are four different cases one needs to consider to capture all the behaviors that may occur as indicator values change:

1. x_{ij}^* *is not the maximum* (or minimum, respectively), and changes *do not cause* it to become so. In this case, one can analyze the impact of changes in x_{ij}^* without need to consider $\frac{\partial \mathcal{S}}{\partial \max_j \{x_{ij}^*\}}$ (these are the derivatives shown in Table 2.7).

2. x_{ij}^* *is not the maximum* (or minimum, respectively), and changes *cause* it to become so. In this case, one must consider both $\frac{\partial \mathcal{S}}{\partial x_{ij}^*}$ and $\frac{\partial \mathcal{S}}{\partial \max_j \{x_{ij}^*\}}$.

3. x_{ij}^* *is the maximum* (or minimum, respectively), and changes *do not cause* it to become otherwise. In this case one need only consider $\frac{\partial \mathcal{S}}{\partial \max_j \{x_{ij}^*\}}$.

4. x_{ij}^* *is the maximum* (or minimum, respectively), and changes *cause* it to become otherwise. In this case one again needs to consider both $\frac{\partial \mathcal{S}}{\partial x_{ij}^*}$ and $\frac{\partial \mathcal{S}}{\partial \max_j \{x_{ij}^*\}}$.

In the target normalization scheme, the changes in $\mathcal{S}$ as a function of the changes in indicator values are more easily captured, even though they are also defined piecewise. This behavior is due to the fact that target normalization is not an internal normalization process. The cases for target normalization have to do with the measurement value changing between the baseline and target values defined and the changing beyond those values. As it is shown in Figure 2.3, there is no dramatic change as indicator values move outside the interval defined by the baselines and targets. In fact, once an indicator measurement moves beyond the baseline or target values, the change in $\mathcal{S}$ becomes exactly 0.

7. Opportunities for Further Research

Further research may seek to investigate additional normalization functions not included in this paper. The techniques developed in this paper are general in their application, with regard to the type of normalization functions and aggregation functions that can be analyzed and used. In addition to expanding analysis to other normalization schemes, determining the sensitivity in the calculation of a composite sustainability score, $\mathcal{S}$, to other normalization scheme parameters, such as the targets and baseline(s) defined in target normalization, would also be valuable. Additional topics of interest include the quantification of implicit weights and studying normalization in the context of meaningful aggregation. These two topics are discussed next and examples are included.

7.1. Quantification of Implicit Weights

In some specific cases, the differential impacts on $\mathcal{S}$ can be written as a set of weights associated with each indicator given the normalization function, f_i, indicator aggregation function, α_i, and final aggregation function, A that are chosen. Further development of methods to quantify implicit weights could prove useful and would provide stakeholders a quick way to determine the relative importance placed on each indicator resulting from the mathematical structure of the sustainability score. The simplest case of when these implicit weights can be calculated occurs when the the partial derivatives, found using Equation (3), are constant. For example, consider the derivatives in Table 2.8 when $\mathcal{S} = AM$, the arithmetic mean and values are not changing beyond baseline or target measures. We have the derivatives for $\frac{\partial \mathcal{S}}{\partial y_{1,j}^*} = -0.0295$, $\frac{\partial \mathcal{S}}{\partial y_{2,j}^*} = 0.0042$, and $\frac{\partial \mathcal{S}}{\partial y_{3,j}^*} = 0.0011$.

The implicit weights, call them w_i, for each indicator measurement are

$$w_i = \frac{\left|\frac{\partial \mathcal{S}}{\partial y^*_{i,j}}\right|}{\sum_{k=1}^{3}\left|\frac{\partial \mathcal{S}}{\partial y^*_{k,j}}\right|},$$

specifically, this scenario produces weights of $w_1 = 0.85$, $w_2 = 0.12$, $w_3 = 0.03$ for phosphorus, yield, and %OM matter indicators, respectively. However, it should be pointed out that these are changes in $\mathcal{S}$ per a unit change in the indicators within the target and baseline range. For indicators, such as the water quality indicator of Phosphorus, a unit change (say from 0 to 1 mg/L) is very large and would in fact move any measurement within the target and baseline values to a value outside of that range. Once again, the challenge of working with multiple indicators on various scales shows up. In an instance such as this, which is likely to be very common in sustainability assessment, the question becomes, how can one use the information contained in the derivatives to quantify implicit weights adjusted to the scales of the indicators?

Using the derivatives, $\frac{\partial \mathcal{S}}{\partial y^*_{1,j}} = -0.0295$, $\frac{\partial \mathcal{S}}{\partial y^*_{2,j}} = 0.0042$, and $\frac{\partial \mathcal{S}}{\partial y^*_{3,j}} = 0.0011$, we can capture a more relevant quantity related to changing an indicator measure by, instead of considering a single unit change, using the baseline and target values to provide a range for changes that are relevant to the indicator. Specifically, one can multiply the derivative value by the difference in the baseline and targets, $(-0.0295)|T_1 - B_1| = (-0.0295)|0 - 0.1| = 0.00295$, $(0.0042)|T_2 - B_2| = (0.0042)|8 - 0| = 0.0336$, and $(0.0011)|T_3 - B_3| = (0.0011)|4 - 1.5| = 0.00275$ for Phosphorus, Yield, and %OM, respectively. This use of the derivatives, baselines, and targets has immediately provided something useful; this calculation is the analytical analog to derive the numerical quantities that one can gather by taking the difference of the maximum and minimum values of the plots provided in Figures 2.3a, 2.3c, 2.3e, respectively. With these scale adjusted responses of $\mathcal{S}$ to changes in indicator measurements, we can now revisit our quantification of implicit weights and calculate *scale-adjusted implicit weights*, $\hat{w}_i$, in the following way,

$$\hat{w}_i = \frac{|T_i - B_i|\left|\frac{\partial \mathcal{S}}{\partial y^*_{i,j}}\right|}{\sum_{k=1}^{3}\left|\frac{\partial \mathcal{S}}{\partial y^*_{k,j}}\right|}. \tag{4}$$

Using this formulation, the scale adjusted weights are $\hat{w}_1 = 0.075$, $\hat{w}_2 = 0.855$, $\hat{w}_3 = 0.070$ for phosphorus, yield, and %OM matter indicators, respectively. These scale-adjusted weights now represent the relative impact each indicator measurement has on the aggregate output $\mathcal{S}$ as the measurement varies between the baseline and target values.

7.2. Normalization and Meaningful Aggregation

How normalization functions transform *measurability scales* of data can also be investigated in order to utilize results from previous research into *meaningful* statements made with aggregate values, sometimes just referenced as *meaningful aggregation.* The topic of meaningful aggregation arises in sustainability and environmental assessment (Pollesch and Dale, 2015; Roberts, 2014a; Böhringer and Jochem, 2007; Zhou et al., 2006; Ebert and Welsch, 2004) and uses the measurability scale of indicator data to provide a method for selecting aggregation functions. Examples of *measurability scales* include ratio, interval, ordinal, and nominal (Stevens, 1946). Knowledge of these scales, along with an application of Luce's principle (Luce, 1959), is used to ensure that, when

data are transformed, they are transformed in such a way that the information contained in the data is maintained; such transformations are referred to as *meaningful transformations.* If one determines how or if the normalization function changes the scale of measurement of the data being considered, it is then possible to utilize results of previous research in meaningful aggregation and to create an aggregate score of sustainability that adheres to the principles therein. For example, ratio-scale measurable indicators occur frequently in sustainability assessment. These indicators are identified by differences between data points having meaning, ratios of data points having meaning, and the existence of a non-arbitrary zero point for the data being measured. Pollesch and Dale (2015) showed that of the 19 environmental indicators for bioenergy sustainability identified in McBride et al. (2011), all but one indicator is ratio-scale measurable. For an example of how normalization functions affect scales of measurement, consider an indicator that is ratio-scale measurable.

- Applying any of the functions in the ratio normalization scheme to a ratio-scale measurable indicator results in a unit less ratio-scale measurable indicator. The non-arbitrary zero value stays the same, and the normalized value now defines a new ratio scale.
- Z-score standardization of ratio-scale measurable data assigns a value of zero to the mean value of the data set, and the unit less quantity represented by a Z-score is also ratio-scale measurable. The normalized value represents the number of standard deviations the original value is away from the mean, thus Z-score standardization transforms ratio-scale measurable data to a new ratio scale.
- Unit equivalence normalization is scalar multiplication, and thus for any non-zero conversion factor c_f, the normalized measurability scale of ratio-scale measurable data is a new ratio scale.

As an opportunity for future research, further investigation as to how the normalization process changes measurability scale can be carried out for different combinations of measurability scale types and normalization functions. This analysis would allow identification of meaningful aggregation functions for indicators included in a sustainability assessment.

8. Conclusions

This paper investigates properties of normalization functions and explores the implications that different choices of normalization schemes can have when normalized values are included in aggregate measures of sustainability. Ratio normalization, Z-score normalization, unit equivalence normalization, and target normalization schemes are analyzed for their behavior in terms of internal normalization, the structure of their derivatives, and comparability of normalized values. We introduce the term *bearing* to unify the variety of terminology present in literature that is used to discuss this property of indicators. The case study motivates the theoretical analysis of normalization schemes by demonstrating how the properties of normalization functions manifest in the simple three-indicator bioenergy sustainability assessment provided.

Quantification of sustainability is approached using a variety of metrics, many of which utilize indicators as stand-alone measures or within aggregate values. Indicator approaches for assessing progress towards sustainability include, at a minimum, information about the economic, social, and environmental aspects of the system being studied. Given the large number of indicators that can be used within an assessment, there is often stakeholder interest and a benefit in combining sustainability indicators. Although clarity is seen as a benefit when combining indicators, as indicator

measurements are combined, this benefit comes at the cost of lost information and data resolution. This is inherent in any aggregation procedure. Gasparatos and Scolobig (2012) provide a good discussion on tradeoffs arising in sustainability assessment. Normalization of indicators, although almost always prerequisite for aggregation of indicators, elicits tradeoffs within the analysis as well.

The case study shows differences of behavior between ratio and target normalization schemes. The consequence of ratio normalization functions being internal is especially evident in Figure 2.2 where the aggregate score of sustainability $\mathcal{S}$ is impacted greatly as measures change the extremal values of the data set. The ratio normalization scheme has fundamental differences in the behavior of smaller-the-better and larger-the-better normalization functions. Specifically, for smaller-the-better bearing indicators, the impact of changes on $\mathcal{S}$ in non-normalized measures differs depending on where those measures are in relation to the minimum value for that data set (see Figures 2.2a and 2.2b), whereas larger-the-better type indicators do not have this dependence. This discrepancy between indicator bearing type does not occur when target normalization is used. Both normalization schemes have behaviors that change as non-normalized measures are varied near threshold values. These are the minimum or maximum values, in the case of ratio normalization, and baseline and target values in target normalization. This set of behaviors differs near threshold values in complexity and influence on predictability of how aggregate outcomes are impacted by changes in non-normalized values across these two normalization schemes.

This research highlights some of the advantages and disadvantages associated with normalization schemes used in sustainability assessment and the calculation of a composite score of sustainability. The internal normalization procedures, Z-score and ratio normalization are easier to implement on a data set given that they do not require externally defined targets and baselines encountered in target normalization or the multitude of conversion factors required for unit equivalence normalization. However, the internal normalization procedures have disadvantages when it comes to the dependence exhibited in the normalized values on extremal values of the data set and how that dependence manifests in aggregate sustainability scores derived from those normalized values. With respect to ratio normalization, the difference between how changes in smaller-the-better and larger-the-better type indicators can impact normalized values and aggregate values derived thereof is of concern. The different cases that one might encounter due to the piecewise differentiability of ratio and target normalization functions are not present in Z-score and unit equivalence normalization. The change in normalized value with respect to changes in non-normalized measures presented in Table 2.3 show that unit-equivalence normalization has the simplest partial derivative expression of the four schemes, while Z-score normalization produces quite a complicated expression for the partial derivative, even without needing to consider the different scenarios of the piecewise defined derivatives for ratio and target normalization.

Of the four normalization schemes explored in-depth in this paper, target normalization stands out as a candidate for use within sustainability assessment. In sustainability assessment, context is extremely important, and a strength of target normalization is that it allows for the inclusion of contextually relevant normalization parameters in the forms of baseline and target values. This context specificity also aids in the interpretation of normalized values. Additionally, as discussed previously, functional forms across bearing type within target normalization are more consistent than those used within the ratio normalization scheme. Although target normalization is a standout when it comes to sustainability assessment for the reasons just provided, it is recommended that advantages and disadvantages of normalization schemes be considered before inclusion in any assessment application; this paper will aid researchers in this regard.

This paper will also help researchers and stakeholders by providing methods to clarify connec-

tions between normalization scheme and the accompanying impact that normalization functions choice can have on the aggregation of indicators measuring progress towards sustainability. The derivatives based approach shown in this paper was chosen to elucidate how general properties of normalization functions, such as internality, manifest to create specific dependencies in aggregate assessment outcomes. The derivatives based approach also provides a foundation upon which other analysis can be developed. Specifically, the scale adjusted implicit weights formulation (Equation (4)) shows promise, with further development, to become a standard method for reporting indicator specific sensitivities that can accompany aggregate scores of sustainability.

9. Acknowledgments

This research was supported by the U.S. Department of Energy (DOE) under the Bioenergy Technologies Office. Oak Ridge National Laboratory is managed by the UT-Battelle, LLC, for DOE under contract DE-AC05-00OR22725. Discussions with Fred Roberts, Esther Parish, Harlan Stech, and Bruce Peckham have been helpful. Comments by Lou Gross on an earlier draft are also appreciated.

The case study in the paper builds from research of the Southeastern Partnership for Integrated Biomass Supply Systems (IBSS), which is funded through Agriculture and Food Research Initiative Competitive Grant no. 2011-68005-30410 from the USDA National Institute of Food and Agriculture.

Chapter III

Applications of Aggregation Theory to Sustainability Assessment

Publication Statement:

A version of Chapter III was originally published by Nathan L. Pollesch and Virginia H. Dale: Pollesch, N, and VH Dale. 2015. Applications of Aggregation Theory to Sustainability Assessment. *Ecological Economics*, 114:117-127. This manuscript was principally composed by N. Pollesch with writing by V.H. Dale included in the introduction. V.H. Dale provided motivation for the research as well as numerous comments on previous versions of the manuscript.

Applications of Aggregation Theory to Sustainability Assessment

N. Pollesch[a,b,*], V.H. Dale[b]

[a] *Department of Mathematics, The University of Tennessee, 1403 Circle Drive, Knoxville, TN 37996-1320*
[b] *Center for BioEnergy Sustainability, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6035*

Abstract

In order to aid operations that promote sustainability goals, researchers and stakeholders use sustainability assessments. Although assessments take various forms, many utilize diverse sets of indicators numbering anywhere from two to over 2000. Indices, composite indicators, or aggregate values are used to simplify high dimensional and complex data sets and to clarify assessment results. Although the choice of aggregation function is a key component in the development of the assessment, there are few literature examples to guide appropriate aggregation function selection. This paper applies the mathematical study of aggregation functions to sustainability assessment in order to aid in providing criteria for aggregation function selection. Relevant mathematical properties of aggregation functions are presented and interpreted. Cases of these properties and their relation to previous sustainability assessment research are provided. Examples show that mathematical aggregation properties can be used to address the topics of compensatory behavior and weak versus strong sustainability, aggregation of data under varying units of measurements, multiple site multiple indicator aggregation, and the determination of error bounds in aggregate output for normalized and non-normalized indicator measures.

Keywords: aggregation functions, bioenergy sustainability, compensatory functions, distance to target, indicators, mathematical aggregation theory, sustainability assessment, uncertainty, weak versus strong sustainability

---

*Corresponding author

1. Introduction

A challenge for assessing sustainability is that it is not a single entity that can be readily measured. Instead sustainability is a combination of several aspects of the physical and biotic environment, social welfare and economic wellbeing. Furthermore, it is an aspiration rather than a state. Its meaning is largely determined by contextual circumstances (Efroymson et al., 2013). Yet it is important to be able to measure, quantify, and discuss progress toward that goal.

Current sustainability assessment approaches often represent sustainability using multiple indicators, multiple variables, or multiple data points. At a minimum, the consensus is that sustainability needs to incorporate environmental, social, and economic conditions, which are referred to as the three pillars of sustainability (Mori and Christodoulou, 2012; Hacking and Guthrie, 2008; Mayer, 2008; Brundtland et al., 1987). In practice, sustainability indices can incorporate data from over 2600 indicator variables (The Living Planet Index, McRae et al., 2012). To add further complexity, each input variable often has an associated data set containing multiple observations. This large amount of data about diverse components of sustainability is difficult to manage and nearly impossible to visualize without some sort of compression or reduction of dimensionality.

Aggregation functions are one method employed to accomplish this task of clarifying and simplifying data. Aggregation theory is the area of mathematics that explores the form and properties of such aggregation functions. In ecological economics the topic of aggregation comes up in regard to spatial aggregation (Su and Ang, 2010), valuation of ecosystem benefits (Tait et al., 2012; Lele and Srinivasan, 2013), calculation of conservation benefits (Winands et al., 2013), and combining information across sectors (Lenzen, 2007; Marin et al., 2012).

This study introduces basic properties, definitions, and theory related to the process of aggregation in order to aid in providing a rigorous mathematical baseline for further development of sustainability assessment techniques and methodologies. This paper deals with the conditions that must be met in order for information to be combined in an accurate, consistent, and overall robust manner. Five examples highlight some of the many relationships that can be derived between mathematical aggregation theory and sustainability assessment. These examples include mathematical interpretation of weak and strong sustainability, a proof that provides a simple bound for aggregate outputs under varying levels of relative error using the arithmetic mean, and two examples of how grouping and aggregation can lead to inconsistent results depending on how aggregation takes place. The final section discusses multiple invariance properties with respect to the scale of measurability of the indicators to be aggregated and includes an example of how a simple change in measurement units can create inconsistent aggregate outputs. The 2004 paper by Ebert and Welsch, which provides a guideline for choosing aggregation functions, is interpreted and placed into the larger mathematical aggregation theoretic context.

2. Basic Properties of Aggregation Functions

The process of aggregation is ubiquitous in the sciences. However, the word aggregation can take on different meanings within different disciplines. The book *Aggregation Functions* (Grabisch et al., 2009), presents a comprehensive mathematical treatment of aggregation functions and their properties and is a unique resource within the mathematics literature. The definitions provided in Grabisch et al. (2009) are adopted here. Beginning with the formal definition for an aggregation function, the following section establishes the basic terms and properties used. For each set of properties presented, a mathematical definition is provided along with interpretations related to sustainability assessment to provide context.

2.1. Definition of an Aggregation Function

In general, an aggregate value is a single representative value for an arbitrarily long set of related values. An aggregation function is the mathematical operation that maps the input values to the representative output value or 'aggregate'. Formally, for some nonempty real interval $\mathbb{I} \subseteq \mathbb{R}$ containing the values to be aggregated, an *aggregation function* in $\mathbb{I}^n$ is a function:

$A^{(n)} : \mathbb{I}^n \to \mathbb{I}$ that

(i) is nondecreasing (in each variable)

(ii) fulfills the following boundary conditions

$$\inf_{\boldsymbol{x} \in \mathbb{I}^n} A^{(n)}(\boldsymbol{x}) = \inf \mathbb{I} \text{ and } \sup_{\boldsymbol{x} \in \mathbb{I}^n} A^{(n)}(\boldsymbol{x}) = \sup \mathbb{I} \tag{5}$$

where n represents the number of variables in the argument of the function, that is, the number of values to be aggregated or the dimension of the input vector, $\boldsymbol{x}$. In general, an aggregation function $A^{(n)}(\boldsymbol{x})$ will be written as $A(\boldsymbol{x})$ with the number of variables in its argument suppressed. Also note that the domain associated with a given aggregation function often changes with assessment context.

As an interpretation, condition (i) states that if any input value increases, the aggregate output value cannot decrease. Condition (ii) dictates what must happen at the boundary values. For example, if a set of indicators are normalized to values between 0 and 1, then the nonempty interval is given by $\mathbb{I} = [0, 1]$ and an aggregation function $A^{(n)}(\boldsymbol{x})$ must satisfy, $A^{(n)}((0, ..., 0)) = 0$ and $A^{(n)}((1, ..., 1)) = 1$.

Table 3.1 gives some common aggregation functions and their definitions. The aggregation functions most frequently used in practice for sustainability assessment are the arithmetic and weighted arithmetic means (Singh et al., 2009; Böhringer and Jochem, 2007). Although the mathematical properties used to describe function behavior are numerous, certain properties of functions have particular importance to aggregation and are included here. The properties presented may help determine appropriate choices of aggregation functions given the sustainability indicator variables selected and the intended use within the assessment. Some of the mathematical definitions and properties presented, such as continuity, are familiar to mathematicians, while others, such as internality, conjunctivity, and disjunctivity as well as some of the grouping-based properties are less familiar. However, within the context of sustainability assessment and aggregation theory, even familiar properties of functions can take on new meanings. The function property definitions in this paper follow the format of Grabisch et al. (2009), and interpretations relevant to sustainability assessment are provided when possible. Examples relating selected properties to sustainability assessment follow each set of properties provided.

2.2. Continuity Properties

Continuity relates closeness in the input variable(s) to closeness in the output variable(s) where closeness is defined using a specified norm. As such, continuity is important for understanding how the aggregation function performs with variable data or noise. Stronger and weaker forms of continuity exist. A strong form, Lipschitz continuity, allows for computing exact bounds in the output error of the aggregation function by knowing the error present in the input. An example of how the property of Lipschitz continuity of an aggregation function may be put to practical use in sustainability assessment is given next. Table 3.2 includes definitions for standard continuity and Lipschitz continuity, for comparison and reference.

Table 3.1: Example aggregation functions

Function Name	Formula	Assumptions \ Notes
Arithmetic Mean	$A(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^{n} x_i$	$A : \mathbb{I}^n \to \mathbb{I}, \boldsymbol{x} :\in \mathbb{I}$
Weighted Arithmetic Mean	$A(\boldsymbol{x}) := \sum_{i=1}^{n} w_i x_i$	$A : \mathbb{I}^n \to \mathbb{I}, \boldsymbol{x} :\in \mathbb{I}$ $(w_1, ..., w_n) \in [0,1]^n$ $\sum_{i=1}^{n} w_i = 1$
Ordered Weighted Average	[a]$A(\boldsymbol{x}) := \sum_{i=1}^{n} w_i x_{(i)}$	$A : \mathbb{I}^n \to \mathbb{I}, \boldsymbol{x} :\in \mathbb{I}$ $(w_1, ..., w_n) \in [0,1]^n$ $\sum_{i=1}^{n} w_i = 1$
Geometric Mean	$A(\boldsymbol{x}) := (\prod_{i=1}^{n} x_i)^{1/n}$	$A : \mathbb{I}^n \to \mathbb{I}, \boldsymbol{x} :\in \mathbb{I}$ [b]If $n > 1$ then $\mathbb{I} \subseteq (0, \infty)$
Weighted Geometric Mean	$A(\boldsymbol{x}) := \prod_{i=1}^{n} x_i^{w_i}$	$A : \mathbb{I}^n \to \mathbb{I}, \boldsymbol{x} :\in \mathbb{I}$ $(w_1, ..., w_n) \in [0,1]^n$ $\sum_{i=1}^{n} w_i = 1$ If $n > 1$ then $\mathbb{I} \subseteq (0, \infty)$
Minimum	$A(\boldsymbol{x}) := \min\{x_1, ..., x_n\}$ (or $\mathrm{OS}_1(\boldsymbol{x}) := x_{(1)}$)	Also written $\mathrm{Min}(\boldsymbol{x}) = \bigwedge_{i=1}^{n} x_i$ and OS_1 is the 1st order statistic
Maximum	$A(\boldsymbol{x}) := \max\{x_1, ..., x_n\}$ (or $\mathrm{OS}_n(\boldsymbol{x}) := x_{(n)}$)	Also written $\mathrm{Max}(\boldsymbol{x}) = \bigvee_{i=1}^{n} x_i$ and OS_n is the nth order statistic

[a] $x_{(i)}$ represents the ith lowest coordinate of $\boldsymbol{x}$, s.t. $x_{(1)} \leq \cdots \leq x_{(k)} \leq \cdots \leq x_{(n)}$.
[b] The Geometric Means are not aggregation functions on every domain, specifically, for $n > 1$ then $\mathbb{I}$ must satisfy $\mathbb{I} \subseteq (0, \infty)$.

Table 3.2: Continuity properties

Property	Definition	Interpretation\ Notes
Standard Continuity	$F : \mathbb{I} \to \mathbb{R}$, F is *continuous* at $x' \in \mathbb{I}$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in \mathbb{I}$ $\|x - x'\| < \delta$ $\Rightarrow \|F(x) - F(x')\| < \epsilon$	In essence, continuous functions have the property that *small* changes in input values ($\|x-x'\| < \delta$) will result in *small* changes in the output values ($\|F(x) - F(x')\| < \epsilon$).
Lipschitz Continuity	$F : \mathbb{I}^n \to \mathbb{R}$, F is *Lipschitz continuous* (with respect to the norm, $\|\| \cdot \|\|$)[a] with *Lipschitz constant* c if $\|F(\boldsymbol{x}) - F(\boldsymbol{x'})\| \leq c\|\|\boldsymbol{x} - \boldsymbol{x'}\|\|$ for all $\boldsymbol{x}, \boldsymbol{x'} \in \mathbb{I}^n$	With Lipschitz continuity, knowledge about variation in input values ($\|\|\boldsymbol{x} - \boldsymbol{x'}\|\|$), can be used to give an exact bound for the variation in output values. This differs from the definition of standard continuity and is a stronger property. All Lipschitz continuous functions obey (standard) continuity, but not all continuous functions are Lipschitz continuous.

[a] A *norm* is used to convey a sense of distance between variables. With regards to indicator variable input, if $\boldsymbol{x}$ is the vector of *true* input values (without measurement error), and if $\boldsymbol{x'}$ represents the actual indicator variable measurements (with measurement error), then the norm of two inputs, $||\boldsymbol{x} - \boldsymbol{x'}||$, represents how different the true and measured values are (or how much error is present in the actual measurements).

2.3. Example: Lipschitz Continuity and Error Estimation in the Arithmetic Mean

Error estimation and uncertainty quantification through the aggregation process may be approached by utilizing a variety of techniques. Certain aggregation functions have properties that allow one to provide exact bounds in output error depending on the input error (e.g. the arithmetic mean and its Lipschitz continuity).

Consider the following example: Let $\boldsymbol{x} = (x_1, x_2, ..., x_n), x_i \in [0,1]\ \forall\, i$ be a vector whose components are a set of n indicators to be aggregated using the arithmetic mean, $A(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^{n} x_i$. Assume that there is variability in the measures of each of the components, x_i, and that each indicator has a maximum relative error equal to ϵ for some $\epsilon > 0$. Let $\bar{x}_i$ be the best estimate of indicator x_i. Let $\hat{x}_i = \bar{x}_i + \bar{x}_i * \epsilon$, be the upper bound of measures of indicator i. Let $\check{x}_i = \bar{x}_i - \bar{x}_i * \epsilon$ be the lower bound in the measurement of indicator i. Let $\hat{\boldsymbol{x}} = (\hat{x}_1, ..., \hat{x}_n)$ and $\check{\boldsymbol{x}} = (\check{x}_1, ..., \check{x}_n)$. It follows that $||\hat{\boldsymbol{x}} - \check{\boldsymbol{x}}||_1$ is the largest distance (using the L_1 norm[1]) between any two vectors of indicator measures. Further, this difference simplifies as follows:

$$
\begin{aligned}
||\hat{\boldsymbol{x}} - \check{\boldsymbol{x}}||_1 = & \\
&= \sum_{i=1}^{n} |\hat{x}_i - \check{x}_i| \\
&= \sum_{i=1}^{n} |(\bar{x}_i + \bar{x}_i * \epsilon) - (\bar{x}_i - \bar{x}_i * \epsilon)| \\
&= \sum_{i=1}^{n} 2|\epsilon| * |\bar{x}_i| \\
&= 2|\epsilon| * ||\bar{\boldsymbol{x}}||_1
\end{aligned}
$$

Using the Lipschitz continuity of the arithmetic mean (Grabisch et al., 2009), with Lipschitz constant $\frac{1}{n}$, the following bound must hold for the largest error in our aggregate value:

$$
\begin{aligned}
|A(\hat{\boldsymbol{x}}) - A(\check{\boldsymbol{x}})| &\leq \frac{1}{n}||\hat{\boldsymbol{x}} - \check{\boldsymbol{x}}||_1 \\
&= \frac{2|\epsilon|}{n}||\bar{\boldsymbol{x}}||_1
\end{aligned} \tag{6}
$$

Inequality (6) gives a simple bound to the output in the aggregate value of a set of indicator variables with the same relative error, ϵ, using the arithmetic mean as the aggregation function. A relevant simplification comes when one considers measures that have been normalized to fall between the values of 0 and 1. In this case, $||\bar{\boldsymbol{x}}||_1 \leq n$ and thus $|A(\hat{\boldsymbol{x}}) - A(\check{\boldsymbol{x}})| \leq 2|\epsilon|$.

Since it is not generally expected that each of the indicators will have the same relative error in measurement, a more realistic example may include relative errors, ϵ_i, for each indicator variable x_i. To treat this case in which each indicator has its own maximum relative error, let $\hat{x}_i = \bar{x}_i + \bar{x}_i * \epsilon_i$ be the upper bound of measures of indicator i, let $\check{x}_i = \bar{x}_i - \bar{x}_i * \epsilon_i$ be the lower bound of measures of indicator i. Also, let $\hat{\boldsymbol{x}} = (\hat{x}_1, ..., \hat{x}_n)$ and $\check{\boldsymbol{x}} = (\check{x}_1, ..., \check{x}_n)$. With different relative errors for each indicator, the largest relative error among the set of relative errors determines the bound. Let

[1] Let $\boldsymbol{x} = (x_1, ..., x_n)$ and $\boldsymbol{y} = (y_1, ..., y_n)$ be vectors in $\mathbb{R}^n$. The L_1 norm of the difference between $\boldsymbol{x}$ and $\boldsymbol{y}$ denoted $||\boldsymbol{x} - \boldsymbol{y}||_1$, is given by $||\boldsymbol{x} - \boldsymbol{y}||_1 = \sum_{i=1}^{n} |x_i - y_i|$ (also see Table 3.2 for a brief explanation of a *norm*).

$\epsilon_{max} = \max\{\epsilon_1, ..., \epsilon_n\}$, let $\hat{x}_i^{max} = \bar{x}_i + \bar{x}_i * \epsilon_{max}$, let $\check{x}_i^{max} = \bar{x}_i - \bar{x}_i * \epsilon_{max}$, and let $\hat{\boldsymbol{x}}_{\boldsymbol{max}} = (\hat{x}_1^{max}, ..., \hat{x}_n^{max})$ and $\check{\boldsymbol{x}}_{\boldsymbol{max}} = (\check{x}_1^{max}, ..., \check{x}_n^{max})$. It follows that $||\hat{\boldsymbol{x}} - \check{\boldsymbol{x}}||_1 \leq ||\hat{\boldsymbol{x}}_{\boldsymbol{max}} - \check{\boldsymbol{x}}_{\boldsymbol{max}}||_1$ and one can obtain a bound in the aggregate value, similar to the bound above, given by:

$$\begin{aligned} |A(\hat{\boldsymbol{x}}) - A(\check{\boldsymbol{x}})| \leq \frac{1}{n}||\hat{\boldsymbol{x}} - \check{\boldsymbol{x}}||_1 &\leq \frac{1}{n}||\hat{\boldsymbol{x}}_{\boldsymbol{max}} - \check{\boldsymbol{x}}_{\boldsymbol{max}}||_1 \\ &= \frac{2|\epsilon_{max}|}{n}||\bar{\boldsymbol{x}}||_1 \end{aligned} \tag{7}$$

In this case if indicator measures are normalized using distance-to-target to fall in the interval between 0 and 1, then $||\bar{\boldsymbol{x}}||_1 \leq n$, leading to $|A(\hat{\boldsymbol{x}}) - A(\check{\boldsymbol{x}})| \leq 2|\epsilon_{max}|$.

Inequalities (6) and (7) show, for a set of indicator variables and some known relative error in their measurement, there is a precise bound for the arithmetic mean maximum range of aggregate output. Although this formulation is not relevant for unbounded error terms such as when ϵ is assumed to follow some probability distribution, the result is useful for sensitivity analysis as well as the quantification of uncertainty in aggregate output given the uncertainty in indicator input values, both of which are key components in sustainability assessment.

2.4. Internality, Conjunctivity, and Disjunctivity Properties

The degree to which, and if, compensation should occur between indicator variables to be aggregated is an often contentious topic (Mori and Christodoulou, 2012; Hacking and Guthrie, 2008; Mayer, 2008). These disagreements are based on the fact that compensation between indicator variables implies that the quantities represented by those variables are in some sense substitutable. The use of the term *compensatory* or *compensation* in this paper describes the ability, in the aggregate output, of 'high' input component values to offset 'low' input component values, and vice-versa. Hacking and Guthrie (2008) point out that the Living Planet Index (McRae et al., 2012) utilizes a mathematically appropriate aggregation function but is flawed in that it assumes the substitutability of different species. The properties of internality, conjuctivity, and disjunctivity (defined in Table 3.3) are all related to how functions deal with extreme elements of input vectors and if the aggregate output falls strictly within the range of component input values or not. Thus, this set of properties is directly related to the compensatory behavior of the aggregation function. Given the relationship between the Min($\boldsymbol{x}$) and Max($\boldsymbol{x}$) functions and the order statistic functions (see Table 3.1), the notion of conjunctivity and disjunctivity can be generalized to *k-conjunctive* (or *k-disjunctive*) functions. For example, a k-conjunctive function remains unchanged when any of the ordered components, $x_{(k+1)}, ..., x_{(n)}$ are replaced with values greater than or equal to the k^{th} smallest ordered element, $x_{(k)}$. These types of functions can be used to 'ignore' the upper $n - k$ elements of an input vector.

An intuitive definition of an aggregation function includes the mathematical concept of internality[2], given the prevalence of means and averages in assessment aggregation. If non-compensatory aggregation is sought in sustainability assessment, then a further investigation of the types of aggregation functions that are not internal is necessary (see Grabisch et al., 2009, chapter 3 for a treatment of conjunctive and disjunctive aggregation functions).

[2]The term *internality* used in this paper is purely mathematical, describing the behavior of the aggregation function in mapping an input vector to an output value that falls within the range of input component values. It is not related to the economic concept of imposition of costs as being internal or external that utilize the same terminology.

Table 3.3: Internality, conjuntivity, and disjunctivity properties

Property	Definition	Interpretation\ Notes
Conjunctive	[a] $F : \mathbb{I}^n \to \bar{\mathbb{R}}, \boldsymbol{x} \in \mathbb{I}^n$ F is *conjuntive* if $\inf \mathbb{I} \leq F(\boldsymbol{x}) \leq \mathrm{Min}(\boldsymbol{x})$	The output of the function F must be bounded (above) by the Min($\boldsymbol{x}$) function. This condition means that, in a conjunctive function, no low input component can be compensated for by a high input component.
Disjunctive	$F : \mathbb{I}^n \to \bar{\mathbb{R}}, \boldsymbol{x} \in \mathbb{I}^n$ F is *disjunctive* if $\mathrm{Max}(\boldsymbol{x}) \leq F(\boldsymbol{x}) \leq \sup \mathbb{I}$	Similar to conjuntivity, but the output of the function F must be bounded (below) by the Max($\boldsymbol{x}$) function. Meaning that no low input component values may compensate for a high input component value.
Internal	$F : \mathbb{I}^n \to \bar{\mathbb{R}}, \boldsymbol{x} \in \mathbb{I}^n$ F is *internal* if $\mathrm{Min}(\boldsymbol{x}) \leq F(\boldsymbol{x}) \leq \mathrm{Max}(\boldsymbol{x})$	Internal aggregation functions allow for *compensatory effects* between input component values. Here compensatory effects are taken to mean those that allow, for example, high input components to offset low input components in the aggregate output. Averages or mean aggregation functions are internal functions.
[a] $\bar{\mathbb{R}} = [-\infty, \infty]$ represents the *extended real line* and $\mathbb{I} \subseteq \bar{\mathbb{R}}$		

2.5. Example: Weak Versus Strong Sustainability and Internality

The concepts of *weak* and *strong sustainability* were introduced to capture the idea that certain natural capital stocks are unique, essential, and their loss could have irreversible effects on human well-being (Pearce et al., 1994). An 'overall capital stock' approach that allows for compensation between these unique, non-substitutable natural capital stocks and capital stocks not categorized as such, was deemed to fall within a *weak sustainability* framework. A framework that respected the non-substitutability of these natural stocks was deemed as adhering to *strong sustainability.*

The concepts presented in Pearce et al. (1994) have since been used to categorize sustainability assessments (Mori and Christodoulou, 2012; Hacking and Guthrie, 2008; Mayer, 2008). A weak sustainability assessment approach permits compensation of indicators across the three pillars into a representative aggregate value. A strong sustainability assessment, on the other hand, specifies that no aggregation of indicators across the three pillars should be allowed during assessment (Mori and Christodoulou, 2012).

An example of how aggregation theory may be applied to give a mathematical definition to weak versus strong sustainability is useful. Assume that a 'low' value represents a nadir (or anti-ideal) state for a given indicator, and a 'high' value represents an ideal state. Within a strong sustainability assessment framework, for example, no low indicator should offset a high economic or social indicator level and no low social indicator should offset a high environmental or economic indicator level, and so forth. Such compensatory effects, from an aggregation theoretic perspective, fall under the set of properties related to interality, conjunctivity, and disjunctivity. With this in mind, the following definition is suggested:

Let $\mathbb{I}^i_{env}, \mathbb{I}^j_{soc}, \mathbb{I}^k_{econ}$ represent disjoint subsets of indicators containing environmental, social, and economic indicators, respectively. Let $\mathbb{I}^n_{sust} = \mathbb{I}^i_{env} \cup \mathbb{I}^j_{soc} \cup \mathbb{I}^k_{econ}$ represent the entire set of sustainability indicators and Let $F(\boldsymbol{x})$ be an aggregation function s.t. $F : \mathbb{I}^n_{sust} \to \mathbb{I}$, then:

Definition An aggregation function F satisfies *strong sustainability* for any $\boldsymbol{x} \in \mathbb{I}^n_{sust}$, if F is *conjunctive* or *disjunctive.*

This definition states that only conjunctive or disjunctive aggregation functions will satisfy strong sustainability when acting on input vectors that include components from more than one of the three pillars of sustainability.

With respect to the concept of weak and strong sustainability, this example can easily be extended to use different indicator categorizations. For instance, as opposed to assuming that social, environmental, and economic capital stocks are non-substitutable across the three pillars as categories (and are completely substitutable within), one can utilize indicator categorizations that identify the non-substitutable indicator components specifically. This approach would enable moving beyond the necessity of categorizing all indicators as being environmental, social, or economic, while still following the conceptual guidelines for strong sustainability as presented in Pearce et al. (1994).

A further interesting example of a function in the sustainability assessment literature that can be either internal or conjunctive is provided by Díaz-Balteiro and Romero (2004). The authors consider n systems being assessed by m indicators. Letting W_j be a weight, or relative importance factor, for indicator j, and $\bar{R}_{ij}$ represent the normalized value for system i and indicator j, they propose the following index function, IS_i:

$$IS_i = (1 - \lambda)[\min_j (W_j \bar{R}_{ij})] + \lambda \sum_{j=1}^{m} W_j \bar{R}_{ij}$$

Where $\lambda \in [0,1]$ represents a *compensation parameter*. Notice that the internality/conjunctivity of this function is controlled by the value of the parameter λ. This function is a convex combination of the internal function $\sum_{j=1}^{m} W_j \bar{R}_{ij}$, the weighted arithmetic mean, and the conjunctive $\min_j(W_j \bar{R}_{ij})$ function. In this case, any $\lambda \in (0,1]$ will make the function IS_i internal; when $\lambda = 0$ the function IS_i is conjunctive. Although omitted here, it can be shown that IS_i satisfies the formal conditions (see section 2.1) of an aggregation function.

2.6. Grouping Based Properties

An ideal assessment method allows investigators or policy makers to both compress information when simplicity is demanded and also control the aggregation to maintain acceptable levels of information retained verses lost. Such a method may utilize aggregated values of groups or categories of indicators to focus on particular dimensions of the assessment. Additionally, this flexible assessment approach may have aggregation taking place multiple times and at multiple levels of the data structure, which can lead to inconsistent results depending on the aggregation function chosen. Properties related to the behavior of an aggregation function with respect to grouping and aggregation at multiple levels are found in Table 3.4. The behavior of the aggregation function with respect to the ordering of inputs (or groups of inputs) is captured by symmetry related properties presented in Table 3.5. These properties and behaviors are discussed below.

2.6.1. Properties Related to Aggregation at Multiple Levels

Repeated aggregation, or aggregation at multiple levels, is common in sustainability assessment. It arises when aggregate values are used to calculate other aggregate values. One example is when indices are used within indices. The inclusion of any biodiversity index as an indicator in a further aggregate value is an exemplar. The *Environmental Sustainability Index* calculation uses the *National Biodiversity Index* as an indicator. More recently Dobbs et al. (2011), propose that Shannon diversity and evenness index be used as an indicator for urban forest ecosystem services and goods assessment (Esty et al., 2005; Dobbs et al., 2011). Another example can be found in Gómez-Limón and Sanchez-Fernandez (2010), where the *risk of abandonment of agricultural activity* is an index included in their composite indicator of agricultural sustainability. The mathematical properties of *associativity* and *decomposability* (Table 3.4) are related to the behavior of function output with respect to aggregation at multiple levels.

The definitions given in Table 3.4 use notation that may be unfamiliar. To further decode the notation and what the definitions mean, consider the following example. Let $\{x_1, ..., x_{30}\}$ be a set of 30 sustainability indicator variables to be aggregated. Where the first 5 indicator values, $\{x_1, ..., x_5\}$ are all related to air quality, the next 10 are related to water quality, $\{x_6, ..., x_{15}\}$ and the final 15 indicators, $\{x_{16}, ..., x_{30}\}$ are related to soil quality. Let F be our aggregation function and let $\boldsymbol{x} = (x_1, ..., x_5), \boldsymbol{x'} = (x_6, ..., x_{15}), \boldsymbol{x''} = (x_{16}, ..., x_{30})$, so $\boldsymbol{x} \in \mathbb{I}^5, \boldsymbol{x'} \in \mathbb{I}^{10},$ and $\boldsymbol{x''} \in \mathbb{I}^{15}$. The aggregate value of the air, water, and soil quality indicators individually is given by $F(\boldsymbol{x}), F(\boldsymbol{x'}),$ and $F(\boldsymbol{x''})$, respectively. $F(\boldsymbol{x}, \boldsymbol{x'}, \boldsymbol{x''})$ gives the total aggregate value for all indicators, where F is now taking as input, these three vectors of different dimensions, where the dimensions are related to number of indicators in the different groupings of the total set of indicators. The value, $F(F(\boldsymbol{x}), F(\boldsymbol{x'}), F(\boldsymbol{x''}))$, may also be used to represent the total aggregate value of all indicators. In this case, F is only taking 3 input values, namely, the three values found to represent each group individually. The notation $F : \cup_{n \in \mathbb{N}} \mathbb{I}^n \to \bar{\mathbb{I}}$ indicates that our aggregation function F needs to have the flexibility to take as input arguments of varying numbers of indicators. And to reiterate, if it is found that

Table 3.4: Associativity and decomposability properties

Property	Definition	Interpretation\ Notes
Associativity	$F : \cup_{n \in \mathbb{N}} \mathbb{I}^n \to \mathbb{I}$ F is *Associative* if $F(x) = x$ for all $x \in \mathbb{I}$ and if $F(\boldsymbol{x}, \boldsymbol{x'}) = F(F(\boldsymbol{x}), F(\boldsymbol{x'}))$ for all $\boldsymbol{x}, \boldsymbol{x'} \in \cup_{n \in \mathbb{N}} \mathbb{I}^n$	Associativity preserves the output of an aggregation of n indicators under the situation where first a subset of k components ($k < n$) are aggregated, then that output value is aggregated with the rest of the components.
Decomposability	[a] $F : \cup_{n \in \mathbb{N}} \mathbb{I}^n \to \mathbb{I}$ F is *Decomposable* if $F(x) = x$ for all $x \in \mathbb{I}$ and if [b] $F(\boldsymbol{x}, \boldsymbol{x'}) =$ $F(k \cdot F(\boldsymbol{x}), k' \cdot F(\boldsymbol{x'}))$ for all k, k' non-negative integers, and $\boldsymbol{x} \in \mathbb{I}^k, \boldsymbol{x'} \in \mathbb{I}^{k'}$	Decomposability is a similar property to associativity but requires knowledge of how many input values, k and k', the aggregate values to be aggregated again contain.

[a] The notation, $\cup_{n \in \mathbb{N}} \mathbb{I}^n$, is used to represent that the function F can map input vectors of varying number of arguments, n for any $n \in \mathbb{N}$, to the output interval $\mathbb{I}$.
[b] The notation $F(k \cdot F(\boldsymbol{x}), k' \cdot F(\boldsymbol{x'})) = F(\underbrace{F(\boldsymbol{x}), ..., F(\boldsymbol{x})}_{k \text{ of these}}, \underbrace{F(\boldsymbol{x'}), ..., F(\boldsymbol{x'})}_{k' \text{ of these}})$

$F(\boldsymbol{x}, \boldsymbol{x'}, \boldsymbol{x''}) = F(F(\boldsymbol{x}), F(\boldsymbol{x'}), F(\boldsymbol{x''}))$, then the aggregation function F is associative. If it is the case that $F(\boldsymbol{x}, \boldsymbol{x'}, \boldsymbol{x''}) = F(5 \cdot F(\boldsymbol{x}), 10 \cdot F(\boldsymbol{x'}), 15 \cdot F(\boldsymbol{x''}))$, then the aggregation function F is decomposable (see Table 3.4)

To give specific examples of aggregation functions with associativity or decomposability as properties, $\prod_i x_i$ and $\sum_i x_i$ are associative, while aggregation functions such as the arithmetic and geometric mean, $\frac{1}{n}\sum_{i=1}^{n} x_i$ and $(\prod_{i=1}^{n} x_i)^{1\backslash n}$ are decomposable but not associative. The following example shows how inconsistency in aggregate value output can arise if associativity and decomposability of the aggregation function are ignored.

2.6.2. Example: Aggregation of Subsets of Indicators and Associativity

The larger the scope and the more data included in the assessment can lead to aggregation taking place multiple times to produce assessment results. In addition to the examples of indices within indices, often, there is statistical analysis of replicates of data measurements for a given indicator to produce a single representative measurement for that indicator. Specifically, mean or median values of repeated indicator measurements are often used to calculate a composite indicator index rather than the raw data directly. The following example highlights how using a common aggregation function, the arithmetic mean, may cause inconsistencies under two different approaches to find a representative value for a data set.

Consider the following example using a simplified assessment scenario of two indicators. Let x_1 and x_2 represent the indicators and assume each indicator has multiple measurement observations. Indicator x_1 has been measured 5 times to give the following observations, (0.4,0.6,0.5,0.5,0.5). Indicator x_2 has been measured 4 times, with observations of its measure as (0.1,0.1,0.3,0.3). If one is interested in an aggregate value for either indicator x_1 or x_2 individually, then the arithmetic mean of the measures of x_1 is 0.5, and for x_2 the arithmetic mean is of its measures is 0.2.

If a composite (or aggregate) value is sought for the two indicators x_1 and x_2 together, two approaches are:

Approach A Aggregating the mean values of the measurements for x_1 and x_2, 0.5 and 0.2 ,respectively. Using the arithmetic mean results in an overall representative value of 0.35

Approach B Aggregating all the measurement data to find the arithmetic mean of the entire list of measures for both indicators, (0.5,0.4,0.6,0.5,0.5,0.1,0.1,0.3,0.3). Following this approach gives an overall representative value of $0.36\bar{6}$

The inconsistency in results between approaches grows as the disparity in the number of observed measures for each indicator increases. For example, if indicator x_1 is measured five additional times at a value of 0.5, to give ten total observations, the arithmetic mean of the observations of x_1 is still 0.5. Approach A yields the same overall result of 0.35 for indicators x_1 and x_2. However, Approach B, which reports the arithmetic of all measurements for both indicators, yields an overall result of 0.41.

The discrepancy arising in the above example is because the arithmetic mean is not *associative*. The arithmetic mean is, however, *decomposable* and therefore can be used consistently in the two different approaches given in the example as long as the number of observed measures used to arrive at the mean values of 0.5 and 0.2 are known (see Table 4). The decomposability property applied to this circumstance gives that since five measures contribute to the mean value of 0.5 and four

measures contribute to the mean value of 0.2, the aggregate value of [3]

$$(5{\cdot}0.5,4{\cdot}0.2) = (0.5,0.5,0.5,0.5,0.5,0.2,0.2,0.2,0.2)$$

will be same as the aggregate value of (0.5,0.4,0.6,0.5,0.5,0.1,0.1,0.3,0.3) using the arithmetic mean.

The assessment scenario just presented is abridged for clarity. However, variation in the number of measurements of different indicators is a common occurrence, given the diversity of indicators included in sustainability assessments and the differing measurement techniques that accompany each indicator. This example highlights that if one wishes to carry out further aggregations using aggregate values, then the grouping properties of aggregation functions becomes important in order to arrive at consistent values.

2.6.3. Symmetry Properties

The order in which the indicators appear in the input vector may or may not influence the aggregate value. Practically speaking, in the case when equal weights are assumed among indicators, the ordering of indicator appearance in the input vector should have no impact on the output value. The dependence of aggregate output value on input value ordering is captured by symmetry-related properties.

Symmetry is also extended to two-dimensional arrays, or matrices, of indicator values. In this setting, the property of *bisymmetry* is used describe aggregate-value behavior under both varied grouping and varied ordering to arrive at aggregate values. Specifically, one can consider a set of indicator values organized in a $n \times n$ or $n \times p$ dimensional matrix, and ask if the total aggregate value is consistent with the aggregate value found by first aggregating the rows and then the columns as well as the value found by first aggregating the columns and then the rows. This type of question might arise when considering a set of indicators to be aggregated for multiple sites to arrive at a total aggregate value for all sites.

To understand what symmetry of an aggregation function guarantees, consider the following example for a three dimensional input vector. Let $\boldsymbol{x} = (x_1, x_2, x_3) \in \mathbb{I}^3$ and $A(\boldsymbol{x})$ be our aggregation function, if $A(\boldsymbol{x})$ is symmetric then $A(x_1, x_2, x_3) = A(x_1, x_3, x_2) = A(x_2, x_1, x_3) = A(x_2, x_3, x_1) = A(x_3, x_1, x_2) = A(x_3, x_2, x_1)$.

2.6.4. Example: Strong Bisymmetry and Multiple Site, Multiple Indicator Aggregation

The final example provided in this section concerns the aggregation of multiple indicators across multiple sites.

For an example, consider sustainability of a bioenergy production site as measured by n indicators and p different production sites that are being assessed using the same set of indicators. Let $x_{ij} \in \mathbb{I}$ be the measure of indicator i at some site j. For an aggregation function, $F : \cup_{n \in \mathbb{N}} \mathbb{I}^n \to \mathbb{I}$, , let $F_{*j} = F(x_{1j}, x_{2j}, ..., x_{nj})$ be the aggregate value of all n indicators at site j, and let $F_{i*} = F(x_{i1}, x_{i2}, ..., x_{ip})$ be the aggregate value of a single indicator, i, at all p sites.

If a researcher is interested in finding one representative value for the sustainability of the all the p production sites within the region, they can follow different approaches:

Approach A First aggregate a single indicator across all sites, and then aggregate that value across indicators.

[3] The notation (5·0.5,4·0.2) is defined in Table 3.5

Table 3.5: Symmetry properties

Property	Definition	Interpretation\ Notes
Symmetry	$F : \mathbb{I}^n \to \bar{\mathbb{R}}$ F is *symmetric* if [a] $F(x) = F([\boldsymbol{x}]_\sigma)$ for all $\boldsymbol{x} \in \mathbb{I}^n$ and $\sigma \in \Sigma_{[n]}$	Symmetry preserves the output of an aggregation of n indicators under any permutation of the input components.
Bisymmetry	$F : \mathbb{I}^n \to \bar{\mathbb{R}}$ F is *bisymmetric* if $F(F(x_{11}, ...x_{1n}), ..., F(x_{n1}, ..., x_{nn})) = F(F(x_{11}, ...x_{n1}), ..., F(x_{1n}, ..., x_{nn}))$ for all $n \times n$ square matrices $\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix} \in \mathbb{I}^{n \times n}$	Within the matrix framework, bisymmetry ensures that the output of the function being applied to the function values determined by the row entries is equivalent to the output value determined by the function being applied to the function values determined by the column entries.
Strong Bisymmetry	$F : \cup_{n \in \mathbb{N}} \mathbb{I}^n \to \mathbb{I}$ F is *strongly bisymmetric* if $F(x) = x$ for all $x \in \mathbb{I}$ and if for any $n, p \in \mathbb{N}$ we have $F(F(x_{11}, ...x_{1n}), ..., F(x_{p1}, ..., x_{pn})) = F(F(x_{11}, ...x_{p1}), ..., F(x_{1n}, ..., x_{pn}))$ for all $p \times n$ matrices $\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pn} \end{pmatrix} \in \mathbb{I}^{p \times n}$	Strong Bisymmetry extends the property of bisymmetry to rectangular $n \times p$ matrices for $n \neq p$.

[a] The notation, $[\boldsymbol{x}]_\sigma$ represents a permutation, σ, of the components of the input variable, $\boldsymbol{x}$. $\Sigma_{[n]}$ is used to denote the set of permutations of the n components.

Approach B First aggregate all indicators for a particular site, and then aggregate those values for all sites.

The pertinent question is whether both approaches yield the same overall output from the aggregation. Or, symbolically:

$$F(F(x_{11},...,x_{n1}),....,F(x_{1p},...,x_{np})) \stackrel{?}{=} F(F(x_{11},...,x_{1p}),...,F(x_{n1},...,x_{np}))$$

If the function used to carry out the aggregation is *strongly bisymmetric*, then consistency is guaranteed. Otherwise, these two approaches may result in different overall assessments of the all the sites, using all the indicators, depending on the order in which aggregation takes places.

3. Invariant and Meaningful Aggregation Functions by Level of Measurability

Given the variety of sustainability assessment approaches, developing procedures for the construction of aggregation methodologies adds uniformity and rigor to the field of research. However, in order for aggregation procedures to be derived, commonalities amongst the diverse approaches must be identified.

Ebert and Welsch (2004) provide an excellent example. Their work to define a *meaningful environmental index* is derived from concepts of *invariance* and *meaningfulness* and has since been used as a standard in the comprehensive evaluation of sustainability indices in Böhringer and Jochem (2007) and Singh et al. (2009). More recently, the work of Roberts (2014b), discusses meaningfulness related to landscape ecology and biodiversity measures. Ebert and Welsch (2004) note that the topics of meaningfulness and invariance are well-known in social choice theory.

Ebert and Welsch (2004) has its basis in the theory of measurement. The *level of measurability* or *scale of measurability* is a classification of a given indicator variable based on the way in which the indicator can be quantified[4]. The classic examples of *levels of measurability* for variables are nominal, ordinal, interval, and ratio (Stevens, 1946). This fundamental classification of indicator variables and the properties associated with each level of measurability give rise to the concepts of invariance and meaningfulness for a given scale of measurability. After a brief introduction to the theory underlying the work of Ebert and Welsch (2004), the rest of this section connects their work to the larger context of aggregation functions, ending with examples in which invariance and meaningfulness can be applied to the development of sustainability assessment methodologies.

3.1. Admissible Transformations

Transforming variables between different units of measure is something that scientists often do. Whether it be from U.S. customary units to metric units for mass or length or from Celsius to Fahrenheit, these different units are seen as equivalent. However, when aggregation occurs among these variables, inconsistencies can arise if the scales on which they are measured are not taken into account.

[4]One should be clear that *scale of measurability* of an indicator variable has no connection to the spatial or temporal extent to which a variable belongs

3.1.1. Example: Inconsistent aggregate output under measurement unit transformations

In the creation of a sustainability assessment tool, allowing for transformation of data measurements between equivalent units, such as inches to centimeters or parts per million to parts per billion, is certainly desirable.

In this example, consider a researcher who is comparing two different bioenergy production sites, site A and site B. The researcher wishes to track the total impact of nitrogen (N) and phosphorus (P) concentration in streams adjacent to the production site and then focus mitigation efforts on the site with the larger loading of nitrogen and phosphorus. Nitrogen concentrations are taken in units of milligram per liter (mg/L) and phosphorus in centigram per liter (cg/L). Site A has a nitrogen concentration of 0.970 mg/L and phosphorus concentration of 0.0051 cg/L. Site B has a nitrogen concentration of 0.950 mg/L and phosphorus concentration of 0.0082 cg/L.

The researcher calculates the arithmetic mean of nitrogen and phosphorus indicators and finds that site A has a value of $(0.970 + 0.0051)/2 = 0.488$ and that site B has a value of $(0.950 + 0.0082)/2 = 0.479$. The researcher concludes that *site A has a higher aggregate value of nitrogen and phosphorus loading than site B* and concludes to focus their mitigation effort on site A.

Consider now that the researcher decides to record both phosphorus and nitrogen measures using the same units, milligrams per liter. After a quick change of units, phosphorus is now recorded as 0.0510 mg/L at site A and 0.0820 mg/L at site B. The arithmetic means are taken again. Site A has an aggregate value of $(0.970 + 0.0510)/2 = 0.511$, while site B has an aggregate value of $(0.950 + 0.0820)/2 = 0.516$. The researcher now concludes that *site B that has a higher aggregate value of nitrogen and phosphorus loading than site A*. The researcher is left with a contradiction.

Both mg/L and cg/L are ratio scale measurable units. The inconsistent result shown in this example is due to changing between different ratio scale measurable units while using the arithmetic mean to aggregate. In order to ensure consistency under unit transformations that are often taken for granted, measurability scale invariant transformations need to be defined and aggregation functions identified that respect those transformations. In this example, had an aggregation function been chosen that is *meaningful on independent ratio scales* (the geometric mean for example), this inconsistent site ranking would not have occurred.

3.2. Admissible Transformation Formulations by Scale of Measurability

Ordinal, interval, and ratio scale measurable data all appear in sustainability assessments. Ordinal scale data represent a ranking or an order, but differences between numbers do not have meaning. Many surveys utilize ordinal scale measurable data. In recent research, Kopmann and Rehdanz (2013) include *Life Satisfaction* measured on a scale from 1 to 10 (where 1= very dissatisfied and 10= very satisfied) in their human well-being approach to assess value of natural land areas. Interval scale data are similar to an ordinal scale, except that differences between data points are meaningful. Interval scale measurable data also have arbitrary zero values that do not indicate the absence of the measured variable. Temperature as measured in Celsius is a classic example of interval scale measurement, since the difference between 20 and 21 degrees is the same as between 6 and 7, but 0 degrees does not represent the absence of temperature. Variables measured on a ratio scale are similar to interval scale measurable variables, except that the zero value is unique and non-arbitrary. Many indicators used for sustainability assessment are ratio scale measurable variables (see Table 3.8), such as bulk density measurements in soils, to measurements of CO_2 emissions from a power plant over a given period of time. Besides nominal, ordinal, interval, and ratio, additional scales of measurability have been defined. Scales are also not fixed, as some data may be

transformed between scales, temperature in Celsius (interval scale) to temperature in Kelvin (ratio scale) is an example.

Each scale of measurability has its own set of transformations that maintain the information contained in the data. Since the ordering or ranking is the information stored by ordinal scale data, functions that transform data on the ordinal-scale need to maintain that order. An *admissible transformation on the ordinal scale* is of the form, $x \mapsto \varphi(x)$, where φ is any strictly increasing function[5] (8). Interval scale measurable data has the same restriction as ordinal scale data, but they must also maintain the distance between measurements. An *admissible transformation on the interval scale* is of the form, $x \mapsto rx + s,$ where $r > 0$ and $s \in \mathbb{R}$ (9). For ratio scale measurable data, order, distance, and the unique zero point must be maintained through transformations of the data. An *admissible transformation on the ratio scale* is of the form, $x \mapsto rx,$ where $r > 0$ (10).

As an example to motivate consideration of admissible transformations, recall the *Life Satisfaction* ordinal response scale from 1 to 10, where 1=very dissatisfied and 10=very satisfied, from Kopmann and Rehdanz (2013). For simplicity, assume that there are 3 respondents to this survey question and their responses are {1,9,3}. Here respondent 1 is the least satisfied, respondent 2 is the most satisfied, and respondent 3 is somewhere in between. Formulation (8) states that in order to maintain the information in the response data, that any later transformation must be a strictly increasing function. Choosing $\varphi(x) = 2x$ as a simple example of a strictly increasing function, under the use of φ, the responses are now transformed from $\{1,9,3\} \xrightarrow{\varphi} \{2,18,6\}$. Respondent 1 is still the least satisfied, respondent 2 the most satisfied, and the transformed data set has the same rank order in responses. If a function that is not strictly increasing was chosen, such as $\hat{\varphi}(x) = (x-5)^2$ then $\{1,9,3\} \xrightarrow{\hat{\varphi}} \{16,16,4\}$ the result is that respondent 1 and 2 are equally satisfied, and respondent 3 is the least satisfied; the use of a non-strictly increasing function as a transformation has fundamentally changed the information contained within the data set.

3.2.1. Meaningfulness of Functions and Indices by Scale Type

Functions defined as *meaningful* obey the principle that an admissible transformation of the input variable(s) should lead to an admissible transformation of the output variable. This is known as Luce's principle (Grabisch et al., 2009; Luce, 1959). Consider a function F, a set of input variables, $\boldsymbol{x} = (x_1, ..., x_n)$, a set of admissible transformations for input variables , $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_n)$, and an admissible transformation for the output variable with respect to the set of transformations for the input variables, $\Psi_{\boldsymbol{\varphi}}$.

A function F that satisfies the following equation:

$$F(\varphi_1(x_1), ..., \varphi_n(x_n)) = \Psi_{\boldsymbol{\varphi}}(F(x_1, ..., x_n)) \tag{11}$$

is defined to be *meaningful* (Grabisch et al., 2009).

Formulation (11) is the most general condition for a meaningful function given some set of admissible transformations. Each input variable, x_i, can be measured on a different scale of measurability and the output variable, $F(\boldsymbol{x})$, can be measured on a scale different from all the x_i. Each x_i can have its own class of admissible transformations, φ_i. The output variable, $F(\boldsymbol{x})$ can also have its own admissible transformation, $\Psi_{\boldsymbol{\varphi}}$, different from all the x_i. Although abstract, equation (11) is used to derive the definitions that are given in the next section. Additionally, the general case for which all

[5]A function F is strictly increasing if $a < b \Rightarrow F(a) < F(b)$

input variables and all output variables can be on any scale of measurability has simplifications that are utilized in practice. From here out, focus is placed on interval and ratio scales, and although similar results exist for ordinal scale meaningful functions, those results are omitted[6].

Simplifications exist to equation (11) that arise when input variables and the output variable are measured on the same scales, or, in the most restrictive case, when variables are on the same scale and use the same measurement units. Beginning with the assumption that all x_i and $F(\boldsymbol{x})$ are on the same scale of measurability (all ratio scale measurable for example), Grabisch et al. (2009) focus on three simplifications. When all input variables and output variables are measured on the same scale, but none share identical units of measurement, this is termed *meaningful on independent scales*. For example, the scenario of aggregating 3 indicators measured in parts per million (ppm), g/cm^3, and temperature in Kelvin, respectively, to arrive at an output variable that is measured on a ratio scale different from all these three would be captured by the term *meaningful on independent ratio scales*. A further simplification comes when input variables and the output variable are measured using the same scale and all input variables share the exact same unit of measurement. The term *meaningful on a single scale* is used to describe this case. If one sought to aggregate input variables all measured in Fahrenheit to an output variable measured in Celsius, a meaningful function in this case would be termed *meaningful on a single interval scale*. The final, and least general case, is when all the input and the output variables are measured on the same scale and in the same units. The term *invariant* is once again utilized to describe functions that are meaningful under this circumstance. If one wished to meaningfully aggregate input variables all measured in hectares to an output variable also measured in hectares, then, since hectares are ratio scale measurable units, this would be necessitate the use of a *ratio scale invariant* function. Although the final circumstance is the least general derived using equation (11), it frequently arises. When multiple samples are taken for a single indicator and aggregated to find a representative value for that indicator, this circumstance applies. The mathematical formulations that accompany the simplifications discussed in this paragraph for both ratio and interval scale measurable variables are presented next.

Admissible transformations for ratio scale measurable variables are classified in (10) above. Using this form of an admissible transformation, and the general form of meaningful functions presented in (11), the following definitions arise when all input and output variables are measured on ratio scales:

A function $F : \mathbb{I}^n \to \mathbb{R}$ is

Ratio scale invariant if for any $r > 0$,

$$F(r\boldsymbol{x}) = rF(\boldsymbol{x}) \tag{12}$$

for all $\boldsymbol{x} \in \mathbb{I}^n$ such that $r\boldsymbol{x} \in \mathbb{I}^n$

Meaningful on a single ratio scale if for any $r > 0$, there exists $R(r) > 0$ such that,

$$F(r\boldsymbol{x}) = R(r)F(\boldsymbol{x}) \tag{13}$$

for all $\boldsymbol{x} \in \mathbb{I}^n$ such that $\boldsymbol{rx} \in \mathbb{I}^n$

[6]For more about aggregation on ordinal scales see Grabisch et al. (2009), Chapter 8

Meaningful on independent ratio scales if for any $\mathbf{r} \in (0, \infty)^n$, there exists $R(\mathbf{r})$ such that,

$$F(\boldsymbol{rx}) = R(\boldsymbol{r})F(\boldsymbol{x}) \tag{14}$$

for all $\boldsymbol{x} \in \mathbb{I}^n$ such that $\boldsymbol{rx} \in \mathbb{I}^n$

Admissible transformations for interval scale measurable variables are classified in (9) above. Using this form for admissible transformations, and the general forms of meaningful functions from (11), the following definitions arise for functions applied to interval scale measurable input and output variables:

A function $F : \mathbb{I}^n \to \mathbb{R}$ is

Interval scale invariant if for any $r > 0$ and $s \in \mathbb{R}$,

$$F(r\boldsymbol{x} + s\mathbf{1}) = rF(\boldsymbol{x}) + s \tag{15}$$

for all $\boldsymbol{x} \in \mathbb{I}^n$ such that $r\boldsymbol{x} + s\mathbf{1} \in \mathbb{I}^n$

Meaningful on a single interval scale if for any $r > 0$ and any $s \in \mathbb{R}$, there exists $R(r, s) > 0$ and $S(r, s) \in \mathbb{R}$ such that,

$$F(r\boldsymbol{x} + s\mathbf{1}) = R(r, s)F(\boldsymbol{x}) + S(r, s) \tag{16}$$

for all $\boldsymbol{x} \in \mathbb{I}^n$ such that $r\boldsymbol{x} + s\mathbf{1} \in \mathbb{I}^n$

Meaningful on independent interval scales if for any $\mathbf{r} \in (0, \infty)^n$ and any $\boldsymbol{s} \in \mathbb{R}^n$, there exists $R(\boldsymbol{r}, \boldsymbol{s})$ and $S(\boldsymbol{r}, \boldsymbol{s}) \in \mathbb{R}$ such that,

$$(\boldsymbol{rx} + \boldsymbol{s}) = R(\boldsymbol{r}, \boldsymbol{s})F(\boldsymbol{x}) + S(\boldsymbol{r}, \boldsymbol{s}) \tag{17}$$

for all $\boldsymbol{x} \in \mathbb{I}^n$, such that $\boldsymbol{rx} + \boldsymbol{s} \in \mathbb{I}^n$

Although the definitions given above progress from least to most general in terms of the types of input and output variables that are considered, it is not the case that a function being meaningful on independent scales implies that the function will be meaningful on a single scale or that a function being meaningful on a single scale implies it will also be invariant on the scale. The relationship that holds between the definitions is that if a function is scale invariant, it will also be meaningful on a single scale, and, if a function is meaningful on independent scales, it will also be meaningful on a single scale (Grabisch et al., 2009). With the definitions and conditions formalized for meaningful and invariant functions on given scales, the next section provides example functions that satisfy the six meaningful definitions given above.

3.2.2. Meaningful Aggregation Functions

General results for functions that satisfy the defined meaningfulness and invariance equations have been studied for nearly three decades. Results are available for all six of the different invariance and meaningfulness scenarios presented above, as well as others (Grabisch et al., 2009; Aczél and Roberts, 1989; Aczél et al., 1986). Using slightly different terminology, Ebert and Welsch (2004), present some of these results in the context of defining their meaningful environmental index as well.

Table 3.6: Aggregation rules for variables by Ebert and Welsch *via Böhringer and Jochem (2007)*

	Noncomparable	Full Comparable
Interval scale	Dictatorial ordering	Arithmetic mean
Ratio scale	Geometric mean	Any homothetic function

Table 3.7: Meaningfulness of common aggregation functions *adapted from Grabisch et al. (2009)*

Aggregation Function	R.S.I.	S.R.S.	I.R.S.	I.S.I.	S.I.S.	I.I.S.
Arithmetic Mean	✓	✓		✓	✓	
Geometric Mean	✓	✓	✓			
[a] $P_k(\boldsymbol{x}) := x_k$	✓	✓	✓	✓	✓	✓
[b] $OS_k(\boldsymbol{x}) := x_{(k)}$	✓	✓		✓	✓	
Weighted Arithmetic Mean	✓	✓		✓	✓	
Weighted Geometric Mean	✓	✓	✓			
Ordered Weighted Average	✓	✓		✓	✓	
$\sum_{i=1}^{n} x_i$	✓	✓			✓	
$\prod_{i=1}^{n} x_i$	✓	✓				

Ratio scale invariant (R.S.I.), meaningful on a single ratio scale (S.R.S.) meaningful on independent ratio scales (I.R.S.), interval scale invariant (I.S.I), meaningful on a single interval scale (S.I.S.), and meaningful on an independent interval scales (I.I.S)
[a] *$P_k(\boldsymbol{x})$ is the projection onto the k^{th} element, x_k of the input vector $\boldsymbol{x}$*
[b] *$OS_k(\boldsymbol{x})$ is the projection on the k^{th} ordered element, $x_{(k)}$ of the input vector $\boldsymbol{x}$ (All other function definitions may be found in Table 3.1)*

Ebert and Welsch (2004) provided a derivation and examples of ratio noncomparable, ratio full comparable, interval noncomparable, and interval full comparable orderings that satisfy various forms of continuity and monotonicity. In their work, the term 'noncomparable' is similar to the term *independent* in classifying scales of input and output variables. The meaning of the term 'full comparable' is similar to the use of the word *single* above. The results of their paper have since been simplified and presented in the form of a compact table which is given in Table 3.6 (Böhringer and Jochem, 2007).

If focusing on *aggregation functions*, then one only needs to consider functional forms that satisfy the different meaningful and invariance properties above and that are nondecreasing and fulfill the boundary conditions provided in equation (5). In doing so, nearly identical results arise as to those presented in Ebert and Welsch (2004). Table 3.7 is adapted from Grabisch et al. (2009) and presents aggregation functions that satisfy the different meaningfulness properties discussed thus far.

Grabisch et al. (2009) also provide deeper results for meaningfulness with respect to ratio and interval scale measurable variables than just examples of functions that satisfy different meaningfulness properties. Their results give complete descriptions for the types of aggregation functions

that satisfy meaningfulness on independent scales for ratio (14) and interval (17) scale measurable variables.

Meaningful Aggregation Functions on Independent Ratio Scales

Proposition 7.8 (Grabisch et al., 2009)

[7]Let $\mathbb{I} = |0, b|$ with $b \in (0, \infty]$. A function $F : \mathbb{I}^n \mapsto \mathbb{I}$ is a *meaningful aggregation function on independent ratio scales* if and only if

$$F(\boldsymbol{x}) = a \prod_{i=1}^{n} x_i^{a_i} \tag{18}$$

where $a_1, ..., a_n \in [0, \infty), \sum_{i=1}^{n} a_i > 0$ and $a > 0$ if $b = \infty$, while $a = b \prod_{i=1}^{n} b^{-a_i}$ if $b < \infty$

Meaningful Aggregation Functions on Independent Interval Scales

Proposition 7.34 (Grabisch et al., 2009)

An aggregation function $F : \mathbb{I}^n \mapsto \mathbb{I}$ is *meaningful on independent interval scales* if and only if

$$F(\boldsymbol{x}) = cx_i + d \tag{19}$$

for some i, where $c > 0$ and $d \in \mathbb{R}$ satisfy $ca + d = a, cb + d = b$, where $a = \inf \mathbb{I}, b = \sup \mathbb{I}$

The major results of Ebert and Welsch (2004) are echoed in Table 3.7 and propositions 7.34 and 7.8 from Grabisch et al. (2009). Specifically, Grabisch et al. (2009) give that, for equation (19), these solutions for meaningful aggregation functions on independent interval scales contained within bounded intervals are projections onto a single coordinate, $P_k(\boldsymbol{x})$ (all indicator variables will be contained in bounded intervals). Further, $P_k(\boldsymbol{x})$ may be seen as a *dictatorial ordering*, which is the terminology used by Ebert and Welsch (2004), because one chooses a single input element x_k to represent the rest of the input values by. The geometric mean and weighted geometric means are the only example aggregation functions that satisfy all three of the ratio scale meaningfulness properties; these clearly fit under the categorization of equation (18). $P_k(\boldsymbol{x})$ is the only example of an aggregation function that satisfies all six of the invariance and meaningfulness properties for interval and ratio scale measurable variables.

3.3. Application of Meaningfulness Properties to Sustainability Assessment

In order to apply successfully the invariance properties to develop an aggregation strategy, a proper classification of what measurability scenario exists for the chosen indicators is crucial. This necessitates categorization of the indicator variables included in the sustainability assessment into their scales of measurability. It further demands an understanding of how measurability scales change through any normalization or rescaling process.

Within sustainability assessment, one might encounter any of the meaningfulness or invariance scenarios presented above. Ratio scale measurable variables are common amongst indicator variables, since many scientific quantities fall on this scale naturally. As an example, Table 3.8 contains the list of recommended environmental sustainability indicator variables for assessing bioenergy sustainability as provided by McBride et al. (2011), and 18 out of the 19 indicators are ratio scale

[7]The notation $|0, b|$ represents any real interval with endpoints 0 and b.

measurable. The indicators are from diverse categories such as air, soil, and water quality, water quantity, greenhouse gas emissions as well as productivity and biodiversity.

Within some categories, such as soil and water quality, many of the variables are measured on the same scale in the same units, for example indicators 1-3 are all measured in Mg/ha, and indicators 5-8 are all measured in mg/L and kg/ha/year. This consistency allows one to utilize ratio scale invariant functions for aggregation within those particular groups. However, it may be simplest to choose a function that is meaningful on independent ratio scales, for then any aggregation on 18 of 19 indicator variables may be carried out using the same function. In the McBride et al. (2011) example, ratio scales dominate. Also, normalization procedures can produce measures on relative scales that are also ratio scale measurable.

Many sustainability assessment researchers argue that normalization by using a distance to target method is an appropriate way to deal with variables that reside on different scales of measurability (Mayer, 2008; Moldan et al., 2012). When it comes to normalized or relative values, Grabisch et al. (2009) point to the work of Roberts (1994) as identifying functions of the form given in (18) as appropriate aggregation functions on these scales. However, with respect to normalization, Ebert and Welsch (2004) contend that normalization introduces further ambiguities to the system when there are arbitrary normalization rules. They go on further to say that if one simply used unnormalized data measures that are on independent ratio scales, a meaningful environmental index would result by using a geometric mean as the aggregation function. Whether or how to normalize indicators is indeed something that varies by project, and in either case, as long as one understands the scale of measurability that the raw or normalized indicators fall on, meaningful aggregation functions can be identified and used.

Flexibility for assessing sustainability in different contexts is enhanced by defining site specific baselines and targets for distance to target normalization. If the same baselines and targets are set for all sites, then the normalized variables can be aggregated in a meaningful fashion using a ratio scale invariant function. However, in a scenario where normalization using distance to target is used and each site has its own baselines and targets, the normalized variables no longer fall on identical scales and so ratio scale invariant functions are no longer appropriate. This result occurs because when different baselines and targets are used, a per unit change in the indicator at one site corresponds to a different change in the normalized variable than a per unit change in the same indicator at a different site. This observation highlights that even slightly different normalization by site has the ability to change the scales for identical indicators and an aggregation function that is meaningful on independent ratios scales is needed for the normalized values from different sites.

4. Conclusion

Sustainability assessments are often complex, utilizing high-dimensional data sets and multifaceted analyses of the diverse indicator data. Aggregation is a key component in many sustainability assessments and a step that has large impact on the outcome of assessment results. To build upon the existing guidance for the construction of sustainability assessments, this paper introduces mathematical concepts that can be used to introduce further rigor and consistency within the aggregation component of the assessment. The concepts presented draw mostly from the mathematical study of aggregation functions, for which Grabisch et al. (2009) provide an excellent resource. Beyond the presentation and justification of relevant mathematical properties of aggregation functions, examples provide context and motivation for further investigation into this branch of mathematics

Table 3.8: Recommended environmental indicators for bioenergy sustainability with measurability scales *adapted from McBride et al. (2011)*

Category	Indicator	Units	Measurability Scale
Soil quality	1. Total organic carbon (TOC)	Mg/ha	Ratio scale
	2. Total nitrogen (N)	Mg/ha	Ratio scale
	3. Extractable phosphorus (P)	Mg/ha	Ratio scale
	4. Bulk Density	g/cm^3	Ratio scale
Water quality and quantity	5. Nitrate concentration in streams (and export)	Concentration: mg/L; export: kg/ha/year	Ratio scale; ratio scale
	6. Total phosphorus (P) concentration in streams (and export)	Concentration: mg/L; export kg/ha/year	Ratio scale; ratio scale
	7. Suspended sediment concentration in streams (and export)	Concentration: mg/L; export kg/ha/year	Ratio scale; ratio scale
	8. Herbicide concentration in streams (and export)	Concentration: mg/L; export kg/ha/year	Ratio scale; ratio scale
	9. Peak storm flow	L/s	Ratio scale
	10. Minimum base flow	L/s	Ratio scale
	11. Consumptive water use (incorporates base flow)	Feedstock production: m^3/ha/day; biorenery: m^3/day	Ratio scale; ratio scale
Greenhouse gases	12. CO_2 equivalent emissions (CO_2 and N_2O)	kg C_{eq}/GJ	Ratio scale
Biodiversity	12. Presence of taxa of special concern	Presence	**
	14. Habitat area of taxa of special concern	ha	Ratio scale
Air Quality	15. Tropospheric ozone	ppb	Ratio scale
	16. Carbon monoxide	ppm	Ratio scale
	17. Total particulate matter less than 2.5μm diameter ($PM_{2.5}$)	μg/m^3	Ratio scale
	18. Total particulate matter less than 10μm diameter (PM_{10})	μg/m^3	Ratio scale
Productivity	19. Aboveground net primary productivity (ANPP)/yield	g C/m^2/year	Ratio scale

***Due to variation of habitat for species of special concern in different contexts, this indicator does not have specified units of measurement (McBride et al., 2011).*

that can be used in sustainability assessment. The paper concludes with a discussion of the work of Ebert and Welsch (2004) and meaningful indices and aggregation functions. It is shown that meaningful aggregation can take place using a variety of aggregation functions, depending on the scale in which the indicator variables are measured. Whether indicator data are normalized or not, meaningful aggregation functions can be defined and utilized for the synthesis and compression of high-dimensional assessment data.

As new sustainability assessments are constructed and existing assessment utilized, this paper should provide deeper understanding of how inconsistencies can arise in sustainability assessment in relation to the aggregation function(s) utilized. The properties of aggregation functions presented are by no means exhaustive and were chosen due to their particular relevance and to raise awareness of opportunities to introduce mathematical rigor within sustainability assessment.

4.1. Acknowledgements

This research was supported by the U.S. Department of Energy (DOE) under the Bioenergy Technologies Office. Oak Ridge National Laboratory is managed by the UT-Battelle, LLC, for DOE under contract DE-AC05-00OR22725. Comments by Lou Gross and Radko Mesiar on an earlier draft are also appreciated. Discussions with Keith Kline, Esther Parish, Jean-Luc Marichal, Bruce Peckham, and Suzanne Lenhart have been helpful.

Chapter IV

Bioenergy Sustainability Target Assessment Resource (Bio-STAR): Protocols and Relevant Indicator Attributes for Normalization and Aggregation

Publication Statement: A version of Chapter IV was submitted as part of a report to the Department of Energy in May, 2016. Its co-authors Nathan L. Pollesch, Virginia H. Dale, and Mike R. Hilliard are currently not pursuing further publication. This chapter is the beginnings of a guide or manual to accompany the Sustainability Target Resource for Bioenergy (Bio-STAR), and may be modified for inclusion once Bio-STAR development is complete. This manuscript was principally composed by N. Pollesch. Motivation for the research, including guidance on goals for specified protocols were provided by V.H. Dale. M.R. Hilliard aided, and continues to aid, in the development of the visualization for Bio-STAR analyses and the implementation of the protocols.

Bioenergy Sustainability Target Assessment Resource (Bio-STAR): Protocols and Relevant Indicator Attributes for Normalization and Aggregation

N. Pollesch[a,b,*], V.H. Dale[b], M.R. Hilliard[c]

[a] *Department of Mathematics, The University of Tennessee, 1403 Circle Drive, Knoxville, TN 37996-1320*
[b] *Center for BioEnergy Sustainability, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6035*
[c] *National Transportation Research Center, Oak Ridge National Laboratory, Knoxville, TN 37932*

Abstract

One approach to assess progress towards sustainability utilizes indicators which are identified to convey various aspects of system behavior and function. A comprehensive sustainability assessment for a given system requires diverse indicators of progress towards sustainability, representing aspects of environmental, social, and economic functioning. Aggregation and normalization of indicator data are commonly employed techniques for analysis and interpretation. These analyses are complicated, in part, by the variety of indicators encountered and variability in the datasets that represent them. Variability in sustainability assessment data stems from multiple sources. In order to ensure consistent results, protocols for normalization and aggregation are established within sustainability assessments. Such protocols utilize spatial, temporal, or other relevant attributes of indicators and their associated datasets. Building from previous research into normalization and aggregation in sustainability assessment, this paper establishes protocols for these analyses within the Bioenergy Sustainability Target Assessment Resource (Bio-STAR). In addition to presenting a concise set of protocols, this paper also identifies attributes of bioenergy sustainability indicators and associated data that are relevant to enact the protocols established.

*Corresponding author

1. Introduction

With a focus on bioenergy, the papers by McBride et al. (2011) and Dale et al. (2013a) identify 35 indicators that can be used to assess progress towards sustainability. These indicators span the *three pillars of sustainability*: environmental, social, and economic aspects of system function. The 35 indicators are organized into 12 sub-categories that represent areas as diverse as soil quality and external trade. After indicators are identified, data are gathered on those indicators and analyzed in order to gain insight into progress that a given bioenergy production operation is making towards sustainability goals; analysis also frequently compares bioenergy production routes.

Analysis of bioenergy sustainability indicator data can take place in a variety of forms. Normalization and aggregation are two common ways in which sustainability assessments transform, synthesize, and process data for interpretation and analysis. Normalization can transform data from various indicators so that they can be more easily compared and/or interpreted. Aggregation plays multiple roles within an assessment. These roles include synthesizing multi-dimensional data to provide summary values of a single indicator and groups of indicators. Aggregation is also used to provide summary values of indicator measures over a specified span of time or region in space.

Although the process of normalization may be used in the absence of any aggregation, normalization is frequently seen as a necessary step before aggregation can take place (Nardo et al., 2005). The interplay of these two processes demands that consideration be given not only to how each individual process is carried out, but also to how they interact within a sustainability assessment. Examples of inconsistencies that can arise within sustainability assessment due to the normalization and aggregation choices are not difficult to find within the literature (Pollesch and Dale, *in press*; Pollesch and Dale 2015; Dias and Domingues, 2014; Roberts, 2014a; Ebert and Welsch, 2004) . Protocols for normalization and aggregation provide the structure for these processes and are necessary to bolster consistency of assessment results.

Protocols for aggregation and normalization utilize various attributes of sustainability indicators and the data gathered to represent them. In this paper the word *attribute* is used as a non-technical term to discuss various properties, qualities, or features associated to an indicator or measurement. However the carryover of the term to computing parlance is acknowledged, and in most instances, informative as one moves to consider methods for organizing sustainability indicators and data. Attributes for sustainability indicators and their accompanying data range from spatio-temporal information about measurements to a wide variety of technical and contextually relevant information. Accompanying the protocols for normalization and aggregation is a discussion and identification of the relevant attributes of indicators and their associated data that one must utilize in order to enact the protocols specified.

2. Background on the use and development of Bio-STAR

Sustainability is a goal toward which progress is desired. Its definition depends on context and hence relates to the specific social, political, cultural, economic, and environmental setting as well as the temporal and spatial boundaries selected. It is useful to define context-specific targets for the components of sustainability that are of interest. Those components fall into twelve categories. Economic aspects are energy security, trade, and profitability, and social components are social well-being, resource conservation, and social acceptability (Dale et al., 2013a). Environmental concerns deal with soil quality, water quality and quantity, greenhouse gases, biodiversity, air quality, and productivity (McBride et al., 2011). Context-specific indicators can be determined for each of these

components (Efroymson et al., 2013). Then for each indicator category, baseline and targets can be determined.

Indicators are typically interpreted in view of baseline conditions and the particular context of a proposed bioenergy system. Baseline conditions are a set of observations or data that are used for comparison to new activities or for a reference case. Ideally, the comparison between measured indicator values and baseline conditions reveals marginal effects of a bioenergy system (McBride et al., 2011). Sometimes baseline conditions are represented by measurements taken before bioenergy operations are initiated.

Targets are built from information about sustainability of particular bioenergy systems given possible values of indicators and inform management responses to those values (McBride et al., 2011). Some targets are thresholds or ranges, for which measurements below, above, or between certain values are acceptable. Other targets are desired trends such as a continued increase in water quality over 5 years.

The audience for Bio-STAR includes stakeholders interested in or affected by the assessment and implementation of sustainable bioenergy systems. Initial stakeholders may include researchers, land-use planners and other individuals or institutions interested in understanding bioenergy sustainability and the trade-offs among social, economic, and ecological indicators. Stakeholders also include the individuals, groups, businesses or organizations that can affect or be affected by a specific process or project under consideration (ISO, 2009). Stakeholder values, perspectives, and information needs influence the goals, time frame, underlying assumptions and other aspects of the decision-making process (Johnson et al., 2013). Hence, the identification of stakeholders interacts with the process of defining overarching goals and prioritizing issues and indicators in a given situation (Dale et al., 2015).

A tool to visualize sustainability progress must be structured to facilitate user adaptation and frequent updates over time. The development of research goals and questions, baseline, and targets for sustainability involves an iterative process, which has tremendous influence on subsequent data collection and analysis. Because sustainability requires mechanisms for continual improvement and contextual conditions change with time and spatial extent of analysis, Bio-STAR is designed to allow users to set and adjust key contextual data over time. Such a flexible approach is both necessary and risky in terms of potential for misrepresentation or misinterpretation of results. The process of defining questions, baseline, and targets requires attention to assure transparency and internal consistency. Protocols to accompany all included analyses in Bio-STAR are being defined to boltser consistency and mathematical robustness.

3. Normalization in Bio-STAR

Normalization plays an important role when one seeks to synthesize data from diverse indicators. Some approaches to normalization also aid in interpretation of data (Pollesch and Dale, *in press*). The term *normalization* in this paper refers to the process of transforming the units of diverse datasets to common units or unit-less quantities. Prevalent normalization procedures are target normalization, Z-score normalization, unit equivalence normalization, and ratio normalization; a large number of variations on these normalization schemes and other unique schemes exist in addition to those of common employ (Pollesch and Dale, *in press*; Singh et al., 2009; Böhringer and Jochem, 2007). .

The first protocol in Bio-STAR is that target normalization is utilized as the normalization function. With this choice made, three other protocols follow in order to fully utilize the benefits

Table 4.1: Target normalization function definitions and notations adapted from (Pollesch and Dale, *in press*). Larger-the-better (LTB), Smaller-the-better (STB), Distance-to-ideal (DTI)

Target normalization to interval $[0, 1]$ function forms	*Indicator Bearing*
$T_{L,j}(x_j^*, T, B) = \begin{cases} 0, & x_j^* \leq B \\ 1 - \frac{T - x_j^*}{T - B}, & B < x_j^* < T \\ 1, & x_j^* \geq T \end{cases}$	*LTB*
$T_{S,j}(x_j^*, T, B) = \begin{cases} 1, & x_j^* \leq T \\ 1 - \frac{x_j^* - T}{B - T}, & T < x_j^* < B \\ 0, & x_j^* \geq B \end{cases}$	*STB*
$T_{D,j}(x_j^*, T, B_l, B_u) = \begin{cases} 1 - \frac{T - x_j^*}{T - B_l}, & B_l < x_j^* < T \\ 1, & x_j^* = T \\ 1 - \frac{x_j^* - T}{B_u - T}, & T < x_j^* < B_u \\ 0, & \text{else} \end{cases}$	*DTI*
x_j^* is the j^{th} non-normalized value for $j = 1, 2, ..., n$ indicator measurements. $\boldsymbol{x}^* = \{x_1^*, x_2^*, ..., x_n^*\}$, T is a target value for a given indicator, B is a baseline value for a given indicator. B_l and B_u are used for lower and upper baseline values, respectively, in the distance-to-ideal bearing function	

of target normalization. A summary of the normalization protocols in Bio-STAR is as follows:

1. Target normalization will be used in Bio-STAR. Three different normalization functions (Table 4.1) can be used for different bearing types, these include:
   (a) Larger the better
   (b) Smaller the better
   (c) Distance to ideal
2. Normalization baseline and target parameter values are provided in the units specified for the indicator measures being normalized
3. Accompanying each set of normalization parameters are attributes for their appropriate spatial and temporal application
4. In the case when more than a single set of normalization parameters can be utilized, by default the parameters with the highest *resolution*[1] will be utilized

3.1. Motivation for Normalization Protocols in Bio-STAR

Of the numerous normalization schemes, target normalization stands out as an excellent candidate for use within sustainability assessments. Pollesch and Dale, (*in press*, show target normalization's suitability for use in sustainability assessment is due to:

- Mathematically consistent behavior among *indicator bearing*[2] type
- Ability to input contextually relevant parameters, such as baseline and target levels for indicators
- Normalized values being clearly interpretable. Specifically, if an indicator is at or above a target level, or at or below a baseline level
- If any intentional differential weighting of indicators is desired, this can be achieved transparently through the setting of target and baseline normalization parameters

The second protocol states that normalization parameters must be provided in specified units and is straight-forward; correct units are a necessity for the proper mathematical conversion to a unitless normalized quantity. Specifying spatial and temporal applicability of normalization parameters under the third protocol has two main purposes, the first being that in order to convey the proper spatio-temporal context, spatially and temporally explicit information about the normalization parameters is necessary. This spatio-temporal information is necessary due to the fact that within Bio-STAR the flexibility to define multiple sets of target and baseline values for a given indicator for various regions in space and spans of time is desired. The second purpose of the third protocol relates to aggregation; in order to ensure that only spatio-temporally comparable measures are being

[1] *Resolution* refers to the level of detail available for measurements of an indicator, usually in either space or time. Resolution is commonly qualified as *low* or *high*. A set of low resolution data in time would be measured yearly, while higher resolution data would be measured monthly. Low and high qualifications of resolution depend on context and expected frequency of measurement

[2] *Indicator bearing* is an attribute of indicator that specifies if, for a given measure, more is better, less is better, or there is some ideal value from which measures should not differ too much

aggregated, one must have a way to keep track of the spatio-temporal context of the normalization parameters. The final normalization protocol provides a default operation when more than one set of normalization parameters can be applied to a given dataset. A scenario with more than one set of normalization parameters could happen if, for example, an indicator is measured at the county-level and a county-level target has been set as well as a state-level target for that indicator. In this case, Bio-STAR will default to use the county-level normalization parameters defined. The last protocol may also be adjusted by the user should they wish to normalize all candidate data using the same normalization parameters.

3.2. Relevant Indicator and Data Attributes to Enact Bio-STAR's Normalization Protocols

In order to normalize indicator measurements following the protocol specification for Bio-STAR a variety of information must be present in the assessment tool about the indicator as well as the data being normalized. With respect to target normalization, one must determine and record the indicator bearing for each indicator that is to be normalized. It should be noted that given the assessment context, the bearing type for a given indicator may change. Once a bearing type has been determined for an analysis as a function of the indicator and the assessment context, one must then identify baseline and target parameter values. Determining appropriate target and baseline values for normalization of bioenergy sustainability data is often challenging. Mayer (2008) points out that "Setting sustainability policy targets can be a difficult process, although targets can be set through political choices and theoretically- and model-derived limits." In this paper discussion is limited to the implementation of target and baseline values, methodology for identification of those values is outside of the scope. It is worth noting that in an assessment utilizing target normalization, assessment results depend critically on baseline and target values, and in practice one should thoroughly justify each parameter choice. Normalization parameters must be given in the units specified for the indicator; therefore measurement unit is another attribute that is relevant for the normalization process.

The spatial and temporal extent of indicator measurements and normalization parameters are necessary to conduct normalization based on the protocols provided. Specifically, one must know when and where the indicator measures represent as well as what span of time and region in space the normalization parameters can be utilized. The information and attributes necessary to carry out the fourth normalization protocol, defaulting to the highest resolution normalization parameter set, is slightly more subtle. The term *highest resolution* presupposes that a hierarchy exists among spatial and temporal extents for the normalization parameters. Researchers familiar with treating environmental data within political boundaries know that simple spatial hierarchies do not always exist. For example, what if normalization parameters have been defined at a watershed level and at a county level where the measurement to be normalized falls into both? How does one decide which is the higher resolution in the absence of a true hierarchy? In this case, instead of a hierarchy, a relationship between spatial extents exists. Rules must be determined as to how to resolve conflicts of this type that may arise in normalization of sustainability indicator measurements.

With respect to Bio-STAR, research is currently underway for developing methods to store and organize measurements of bioenergy sustainability indicators, including how to treat the various spatial and temporal attributes of datasets as well formalizing methods to handle the various hierarchies and relationships that arise among them. A well-designed structure to organize and access the relevant attributes of bioenergy sustainability indicator data is important for being able to quickly complete the normalization and aggregation operations present within Bio-STAR. With

normalization protocols specified and relevant attributes identified, we now move into a discussion of aggregation.

4. Aggregation Protocols for Bio-STAR

Aggregation of indicator measurements takes place to provide summary statistics for a single indicator and also to synthesize data from multiple indicators. In practice, the number of indicator measurements, the spatial resolution, and the temporal resolution can vary greatly across indicators; this variation can also take place for a single indicator. Sources of variation add complication to the aggregation process when one seeks to ensure that consistent results are produced under varying analysis scenarios. The process of aggregation always involves a trade-off between simplicity and information; as multiple values become represented by a single value, information is lost.

Aggregation and disaggregation can take place to move up or down spatial and temporal hierarchies. Aggregation of this form is routinely conducted to simply summarize data, but within the context of sustainability assessment, spatial or temporal mismatches are often made commensurate through the process of aggregation. For example, consider a farm consisting of multiple fields being assessed with two indicators, call them indicators A and B. Indicator A is measured multiple times within each field while indicator B is measured just once for the entire farm. One can aggregate all sub-field-level measures of indicator A for a given field; next one can aggregate all the derived field-level measures of A on the farm to arrive at a final value, with the spatial resolution of farm, that matches the native spatial resolution of indicator B. Disaggregation can occur to move from lower to higher resolution, rectifying measures in the opposite direction. For example, the farm-level measure of indicator B can be disaggregated into higher spatial resolution measures by making assumptions of how that indicator is associated to the individual fields in the farm, or areas within the fields. In both cases data processing is necessitated, and assumptions of relationships of the data and the spatial extents must be made. Assessment results depend critically on the assumptions made and how this data manipulation is conducted.

In an effort to support anticipated aggregation scenarios of users and to adhere to the overall goals of Bio-STAR, the following principles and protocols are defined:

1. Aggregation will take place using the arithmetic mean. Alternatively, the *Minimum* function is being explored for use as a non-compensatory aggregation function within Bio-STAR (see Table 4.2).

2. Only numerically comparable measurements can be aggregated. This is achieved by specifying that all indicator measurements are recorded in the units specified and by specifying that only normalized measurements can be aggregated in Bio-STAR. This specification applies to measurements to be aggregated for a single indicator and to measurements to be aggregated that represent multiple indicators.

3. Implicit differential weighting of indicators based on number of replicates is to be avoided. Comparability is achieved by specifying that only identical numbers of values, which have been normalized using different normalization parameters, can be aggregated. This specification is motivated by the scenario in which measurements for different indicators are aggregated, however it can arise when multiple normalization parameters have been used for measures of a single indicator.

Table 4.2: Example aggregation functions *adapted from Pollesch and Dale, 2015*

Function Name	Formula	Assumptions \ Notes
Arithmetic Mean	$A(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^{n} x_i$	$A : \mathbb{I}^n \to \mathbb{I}, \boldsymbol{x} :\in \mathbb{I}$
Geometric Mean	$A(\boldsymbol{x}) := (\prod_{i=1}^{n} x_i)^{1/n}$	$A : \mathbb{I}^n \to \mathbb{I}, \boldsymbol{x} :\in \mathbb{I}$ [a]If $n > 1$ then $\mathbb{I} \subseteq (0, \infty)$
Minimum	$A(\boldsymbol{x}) := \min\{x_1, ..., x_n\}$ (or $\mathrm{OS}_1(\boldsymbol{x}) := x_{(1)}^{[b]}$)	Also written $\mathrm{Min}(\boldsymbol{x}) = \bigwedge_{i=1}^{n} x_i$ and OS_1 is the 1st order statistic
Notation: $\boldsymbol{x} = \{x_1, x_2, ..., x_n\}$, n is the number of components of $\boldsymbol{x}$, and $\mathbb{I} \subseteq \mathbb{R}$ is some non-empty real interval [a] The Geometric Mean is not an aggregation function on every domain, specifically, for $n > 1$ then $\mathbb{I}$ must satisfy $\mathbb{I} \subseteq (0, \infty)$ [b] $x_{(i)}$ represents the ith lowest coordinate of $\boldsymbol{x}$, s.t. $x_{(1)} \leq \cdots \leq x_{(k)} \leq \cdots \leq x_{(n)}$		

4. All within-indicator aggregation for a single indicator must be completed before any between-indicator aggregation of multiple indicators can occur. This specification is also to avoid differential weighting of indicators.

5. Spatial and temporal extents of data to be aggregated must match before aggregation can occur.

4.1. Motivation for Aggregation Protocols in Bio-STAR

With respect to the first aggregation protocol, choosing the arithmetic mean as the aggregation function, there are multiple benefits, most of which are related to the choice of target normalization within Bio-STAR. When utilized with target normalization, the arithmetic mean enables straightforward quantification of *scale-adjusted implicit weights* to convey the impact that each indicator has on the aggregate output (Pollesch and Dale, *in press*). Pollesch and Dale (2015) showed that nearly all of the indicator measurements included in McBride et al. (2011) and Dale et al. (2013a) are ratio-scale measurable. As such, studies into *meaningful aggregation* (see Grabisch et al. (2009); Ebert and Welsch (2004)) would lead one to consider the use of the geometric mean. However, in conjunction with the target normalization process in Bio-STAR, the arithmetic mean has the opportunity to convey information about the status of the system more consistently. Increased consistency in conveying information is due to the fact that indicator data beyond the baseline specified will be normalized to a value of 0; any aggregation using geometric means on a data set including a value of 0 will be 0. The Minimum function provides a *non-compensatory* alternative to the *compensatory* arithmetic mean. Compensatory in this sense means that high and low values in the aggregate can offset each other through the aggregation process, while in a non-compensatory aggregation function, such as the Minimum for example, no high value can offset any low value. For stakeholders interested in non-compensatory aggregation, the Minimum function may be made available to users of Bio-STAR depending on their assessment goals and desires.

Restricting aggregation in Bio-STAR to normalized measures is done to ensure that only comparable measures are aggregated. In the case when only measures from a single indicator are being

aggregated, one may question the necessity of requiring normalization given that all measures of a single indicator are in identical units. However, it is because data from a single indicator can be normalized using different normalization parameters depending on spatio-temporal context, that this specification applies to all data being aggregated and not just data that includes multiple indicators. The ability to use multiple normalization parameters for a single indicator also necessitates the specification of the fourth protocol.

The number of replicates included in data sets can vary across indicators and for a single indicator over certain regions in space and spans of time, leading to additional challenges in the synthesis and visualization of data. Variation in number of replicates included in an indicator's data set is not only an inherent attribute of indicators but also due to external factors, such as availability of resources to collect data. Consideration must be given to the number of measurements included in an aggregate to keep track of differential weighting that can occur with most aggregation functions (Pollesch and Dale, *in press*). For example, some indicators may only be measured once annually, while others may be measured hundreds of times in a year. Keeping with this example, assume the indicator measured once a year is considered to be at or above the target level, hence normalized to a value of 1, is aggregated using the arithmetic mean with the measures of the other indicator that are below the baseline value, hence normalized to a value of 0. In this case the value of the first indicator will have little effect on the aggregate due to the small implicit weight it is given compared to the indicator that is measured more frequently. Varying number of replicates is expected to arise frequently in Bio-STAR analyses, and therefore the third aggregation protocol has been specified to avoid implicit differential weighting stemming from differing numbers of measures.

Another important scenario related to differential weighting in the aggregation of multiple indicators is most related to the third protocol. As an example, consider that measures for 4 different indicators, A, B, C and D, have been aggregated to a value of X and that a measure from an additional indicator, E is to be aggregated with those. In this scenario, one must be sure to not just use the arithmetic mean on the 2 values of the aggregate and the additional indicator, X and E, but instead one should aggregate all 5 values, one from each indicator, A, B, C, D, and E. Otherwise, indicator measure E will receive undue weight over the other 4 indicator measures, A, B, C, and D. This is due to the fact that the arithmetic mean is not an *associative* aggregation function (see Grabisch et al. (2009) for the formal definition of an associative aggregation function and Pollesch and Dale (2015) for a numeric example emphasizing the role of associativity in aggregation).

The fifth protocol specifies that spatio-temporal attributes of measures must match for aggregation to occur. This protocol ensures, for example, that a farm-level measurement is not aggregated with a county-level measurement, or that a weekly measure is not aggregated with annual measurement. At its root, the fifth aggregation protocol is a check to make sure that only measures representing similar spatial and temporal extents are combined.

In practice, data for bioenergy sustainability indicators have diverse spatial and temporal extents and are varied in number of replicates. Expected diversity in spatial and temporal extents does not mean that the majority of data cannot be aggregated in Bio-STAR. Instead this leads to the necessity of defining a process for aggregating data with different spatial and temporal extents, varying numbers of replicates, and various normalization targets while still adhering to the protocols identified above. Just as research is underway for constructing an organizational scheme for indicator data, development of actual implementation methods within the organizational scheme for Bio-STAR is also currently underway.

4.2. Relevant Indicator and Data Attributes to Enact Bio-STAR's Aggregation Protocols

The relevant attributes for carrying out the aggregation protocols within Bio-STAR are similar to those for normalization with a few exceptions. Specifically, access to measurements units for indicators, spatio-temporal extent of measures, spatio-temporal hierarchy and relationship definitions are necessary. Within the attributes relevant to aggregation, indicator bearing is not necessary information, however the normalization parameter set used to normalized each measure is important. The normalization parameter set for each measure is necessary to avoid differential weighting as specified in the third protocol. For example, normalized data from different indicators will necessarily have different normalization parameters, therefore, the third protocol necessitates that only identical numbers of those differently-normalized measures can be aggregated. In practice, this means that there will likely be an aggregation of normalized values for each indicator followed by an aggregation of the resulting values. This scenario also arises when a single indicator has data that has been normalized using more than one set of normalization parameters. Therefore, each normalized measure must be accompanied by reference to its normalization parameters. Along these lines, and although it nearly goes without saying, knowledge of the number of replicates for a set of indicator measures is also needed for computing aggregates.

5. Discussion

The normalization and aggregation protocols specified in this document are seen as a minimal working set to begin the development of Bio-STAR. All protocols are motivated by anticipated needs of users of Bio-STAR being matched with theoretical investigations into the mathematical properties of aggregation and normalization functions. Furthermore, since the processes of normalization and aggregation interact within the assessment tool, protocols must be specified that consider this interaction.

The relevant attributes of indicators and datasets that are necessary in Bio-STAR are, in part, determined by the protocols specified. At this point in development, beyond those attributes such as measurement units and number of replicates that are needed to determine normalization parameters and for actual computation of normalized and aggregate values, other relevant attributes are centered around conveying the spatio-temporal extent of measurements. Currently, no aggregation or normalization protocols have been put in place that utilize other contextual attributes of indicators (see Section 6.1). However this functionality may be considered desirable as an extension of Bio-STAR for future applications. The topic of non-spatio-temporal contextual attributes of indicators, explicit weighting of indicators, and methods for conveying uncertainty in indicator measures are addressed in the following section as next steps in Bio-STAR development.

6. Next Steps in Development of Protocols for Visualization and Analysis in Bio-STAR

Three aspects that have not been the focus of development thus far but are considered important for future research are defining and handling additional contextual attributes of indicators, weighting of indicators for analysis, and calculating and conveying uncertainty of indicators measures.

6.1. Contextual Attributes of Indicators

This document has focused on developing protocols for analyses that can handle complexities presented by data for indicators of bioenergy sustainability and cater to various anticipated uses

by stakeholders. Protocols have primarily centered on dealing with spatio-temporal differences in measurements and differences in number of replicates across indicators. Spatio-temporal attributes exist for all indicator measurements, in that all measurements correspond to some region in space and some span of time. However some indicators have relevant attributes that are not common to all indicators that are non-spatio-temporal in nature. Preparing for how these more complicated sets of attributes can be stored, visualized, and used for analysis is important as development continues.

Examples of non-spatio-temporal contextual attributes that may be relevant to bioenergy sustainability are *Previous Land Use* and *Soil Type*. These attributes may be directly relevant to aid interpretation of particular indicators, such as *Above Ground Net Primary Productivity* or to entire groups of indicators such as the Soil Quality sub-category. Although relevant to some indicators, these type of contextual attributes may not be directly relevant to other indicators. For example, previous land use may not be relevant *Total Phosphorus Concentration in Streams* but not *Fuel Price Volatility*. This difference in relevance among indicators for contextual attributes can present a variety of challenges if analyses are to take place utilizing these attributes to specify protocols.

The primary questions related to dealing with these types of contextual attributes are: If, and how, to convey information related to non-spatio-temporal contextual attributes? Which of these attributes must be collected along with the data and how should they be stored and organized? Another set of questions is if, and how, these attributes should influence operations such as normalization and aggregation? Deciding upon a uniform way to gather, organize, and visualize contextual attributes as well as the influence on normalization and aggregation protocols related to contextual attributes is an important area for future research.

6.2. Weighting of Indicators for Analysis

A common way in which data are made contextually relevant is by imposing explicit weights on indicator measurements before they are aggregated or synthesized. These weights are chosen to convey differential values for specific indicators. For example, if the analysis included an aggregate value of all 35 indicators identified by McBride et al. (2011) and Dale et al. (2013a), but the stakeholder felt that in a particular bioenergy context the indicator *Public Opinion* was less important than the indicator *Employment* they may assign a higher weight to a measure of Employment, and lower weight to Public Opinion measurements. Aggregation of weighted indicators may take place using the weighted arithmetic or geometric means. In this section we speak briefly about weighting and how the target normalization process, by virtue of the baselines and target structure, implicitly weights indicators and therefore ties weights to transparent baseline and target values.

When measurements of different indicators are aggregated Bio-STAR, protocol dictates that those measurements be normalized. For normalization to occur, target and baseline values must be established and the normalized values can be interpreted as the percent of target attained. The mathematical form of the normalization functions transforms the native scale of measurements to a new scale, which is determined directly by the baseline and target values. For instance, if a target of 30 full-time equivalent (FTE) jobs and a baseline of 0 FTE jobs are set as normalization parameters for Employment, a measure of 15 FTE jobs is normalized to exactly 50% of the goal. However, in a different context, one may set 20 FTE jobs as the target, and therefore a measure of 15 FTE jobs is normalized to 75% percent of the goal. If this normalized measure of Employment is included in an aggregate of economic indicators, only the values of .50 and .75 are passed to the next step of analysis, and thus the impact of the employment measure on the economic aggregate is weighted by the targets and baselines set for normalizing. The same type of differential impact on an aggregate that includes this normalized employment measure could be achieved by assigning

weights and using a weighted aggregation function. In either of these cases, the measure of 15 FTE jobs is not changing; what is changing is how this measure impacts the synthesized or aggregate value that contains it, and how the value of 15 FTE jobs is interpreted within the bioenergy context.

The benefit in using baselines and targets to impose weights is clear, if an indicator such as Employment is to receive a differential weight over another, such as Public Opinion, it will necessarily be tied to the sustainability context through the clear determination of baselines and target values for each indicator. If it is found that baseline and target setting is insufficient to impose the proper context on normalized and synthesized measures of sustainability indicators, then this finding may dictate further research into the area of explicit weight determination and integration into Bio-STAR.

6.3. Conveying Uncertainty of Indicator Measures

Having an idea of the level of uncertainty for measures of sustainability indicators is important not only for interpretation but also to provide insight into variability inherent in the indicators being measured. Even a simple measure of variation in the data set can influence the confidence a stakeholder may have in a measured indicator and the analyses derived thereof. As such, determining the ways in which uncertainty can be quantified and conveyed becomes an important aspect of visualization and analysis.

Uncertainty exists to a certain degree in all measurements. When those measurements are used in further calculations and analyses, that uncertainty gets transmitted and propagated in a variety of ways depending on the mathematical form of the analyses. Determining simple measures of variability, such as variance for a dataset, is computationally straight-forward, as is storing and visualizing the variability measure. However, as indicator measures are normalized and aggregated, conveying uncertainty becomes more complicated. In the case of aggregate values that include multiple indicators, there is a combined level of uncertainty that arises as a function of all measurements included in the aggregate, the normalization function and parameters used, and the aggregation function employed. Calculating this propagated and aggregated uncertainty presents a challenge, especially in a tool that is structured to contain a large amount of flexibility in terms of the analyses that can be undertaken, such as Bio-STAR. Additionally, challenges related to uncertainty emerge in how best to convey and visualize measures of uncertainty once they are calculated. Regardless of the complexity involved, determining how to calculate and convey uncertainty alongside the various measures and analyses for indicators of sustainability is an important next step in development.

7. Conclusion

Many challenges emerge when handling, analyzing, and visualizing the diverse data sets associated to the environmental and socioeconomic indicators identified by McBride et al. (2011) and Dale et al. (2013a). These are related to spatio-temporal variability, determination of normalization parameters and their implementation, uncertainty in measurements, and the variety of non-spatio-temporal contextual attributes encountered. Aggregation of bioenergy sustainability indicators, even in seemingly simple scenarios, can become complicated quite quickly when trade-offs exist between contending goals.

This paper addresses two key components related to developing Bio-STAR. The protocols presented are an application of the previous theoretical findings related to normalization (Pollesch and Dale, *in press*) and aggregation (Pollesch and Dale (2015) in sustainability assessment. Additionally, attributes of indicators and data that are relevant for enacting those protocols are identified

and discussed. Although there is much to be done before the deployment of Bio-STAR, this paper is seen as a critical first-step in Bio-STAR's development. Meaningful assessment and visualization of multi-metric data related to bioenergy sustainability are critical to aid policymakers and stakeholders in decisions related to bioenergy production and utilization and to make progress towards the overall goal of contributing to a sustainability bioeconomy.

8. Acknowledgments

This research was supported by the U.S. Department of Energy (DOE) under the Bioenergy Technologies Office and the University of Tennessee under subcontract 4000132470. Oak Ridge National Laboratory is managed by the UT-Battelle, LLC, for DOE under contract DE-AC05-00OR22725. Esther Parish and Keith Kline have provided useful comments as the ideas developed, and Lou Gross provided helpful feedback on an earlier draft.

Conclusion

An indicator approach to bioenergy sustainability assessment utilizes environmental and socioeconomic indicators to understand and quantify the impacts of decisions for bioenergy utilization both currently and for future generations. Challenges emerge for visualization, analysis, and synthesis of diverse indicators due to variation in spatio-temporal resolution and extent, data quality and availability, and any number of contextual attributes that individual indicators possess. As it was shown through this collection of manuscripts, challenges in sustainability assessment can be addressed by the development and application of mathematical theory.

Utilizing the environmental and socioeconomic indicators identified by McBride et al. (2011) and Dale et al. (2013a), this collection of manuscripts builds upon previous research from Oak Ridge National Laboratory's Center for Bioenergy Sustainability. Research begins with the identification of Key Components for Sustainability Assessment, in which current methodologies in sustainability assessment are provided and techniques that support the goals for the Bioenergy Sustainability Target Assessment Resource are highlighted. Research then moves to the Normalization process for indicators of sustainability, in which multiple normalization schemes are highlighted and investigated for application in sustainability assessment. Normalization functions are classified and studied; target and baseline normalization is recommended for use in Bio-STAR due to desirable mathematical properties. These properties are related to sensitivity in aggregate output due to changes in non-normalized measures and the ability to tie normalized values to sustainability context. Mathematical aggregation theory is used to study the behavior of common aggregation functions and identify those which may be recommended for use within sustainability assessment. Examples included in the third manuscript of this dissertation show that mathematical theory can be tied to sustainability assessment methodology development. Protocols for normalization and aggregation for Bio-STAR are presented in the final manuscript. These protocols apply results from the research into normalization and aggregation of sustainability indicators in order to support the goals and desired function of Bio-STAR.

The implementation of the results from these manuscripts and the construction of Bio-STAR is currently underway. The research presented in this collection of manuscripts contributes to the construction of an assessment tool that adheres to the goals of *(i)* adaptability for assessing diverse bioenergy production pathways, *(ii)* flexibility to support a range of analyses that researchers and policymakers may seek to undertake, and *(iii)* mathematical robustness with respect to normalization, aggregation, and the quantification of data uncertainty. From a broader perspective, this work is designed to support the development of a sustainable bioeconomy in the United States and globally. Assessment is critical to aid policymakers, researchers, and stakeholders as they investigate impacts of various bioenergy production options both now and for future generations.

References

Aczél, J. and Roberts, F. S. (1989). On the possible merging functions. *Math. Soc. Sci.*, 17(3):205–243.

Aczél, J., Roberts, F. S., and Rosenbaum, Z. (1986). On scientific laws without dimensional constants. *J. Math. Anal. Appl.*, 119(1-2):389–416.

Akaike, H. (1971). Information theory and an extension of the maximum likelihood principle; 1973. *Tsahkadsor, Armenian SSR*, pages 267–281.

Bare, J., Gloria, T., and Norris, G. (2006). Development of the method and us normalization database for life cycle impact assessment and sustainability metrics. *Environmental science & technology*, 40(16):5108–5115.

Baskaran, L., Jager, H., Schweizer, P. E., and Srinivasan, R. (2010). Progress toward evaluating the sustainability of switchgrass as a bioenergy crop using the swat model. *Transactions of the ASABE*, 53(5):1547–1556.

Berndes, G. and Hansson, J. (2007). Bioenergy expansion in the eu: cost-effective climate change mitigation, employment creation and reduced dependency on imported fuels. *Energy Policy*, 35(12):5965–5979.

Böhringer, C. and Jochem, P. E. (2007). Measuring the immeasurable A survey of sustainability indices. *Ecol. Econ.*, 63(1):1–8.

Bollen, K. and Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological bulletin*, 110(2):305.

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social indicators research*, 59(2):115–151.

Bozdogan, H. (1990). On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics-Theory and Methods*, 19(1):221–278.

Brady, N. C., Weil, R. R., et al. (1996). *The nature and properties of soils.* Prentice-Hall Inc., 11 edition.

Bringezu, S., Schutz, H., OBrien, M., Kaupi, L., Howarth, R., and McNeely, J. (2009). *Assessing biofuels: towards sustainable production and use of resources.* United Nations Environment Programme.

Brundtland, G. H., on Environment, W. C., and Development (1987). *Our common future*, volume 383. Oxford University Press Oxford.

Cairns Jr, J., McCormick, P. V., and Niederlehner, B. (1993). A proposed framework for developing indicators of ecosystem health. *Hydrobiologia*, 263(1):1–44.

Castoldi, N. and Bechini, L. (2010). Integrated sustainability assessment of cropping systems with agro-ecological and economic indicators in northern italy. *European journal of agronomy*, 32(1):59–72.

Costanza, R., d'Arge, R., Groot, R. d., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., et al. (1998). The value of the world's ecosystem services and natural capital.

Council on Sustainable Biomass Production (CSBP) (2010). Draft provisional standard for sustainable production of agricultural biomass. `http://www.arb.ca.gov/fuels/lcfs/workgroups/lcfssustain/csbp.pdf`. [Online; accessed 12-May-2016].

Dale, B. E., Anderson, J. E., Brown, R. C., Csonka, S., Dale, V. H., Herwick, G., Jackson, R. D., Jordan, N., Kaffka, S., Kline, K. L., et al. (2014). Take a closer look: biofuels can support environmental, economic and social goals. *Environmental science & technology*, 48(13):7200–7203.

Dale, V. H. and Beyeler, S. C. (2001). Challenges in the development and use of ecological indicators. *Ecol. Indic.*, 1(1):3–10.

Dale, V. H., Efroymson, R. A., Kline, K. L., and Davitt, M. S. (2015). A framework for selecting indicators of bioenergy sustainability. *Biofuels, Bioproducts and Biorefining*, 9(4):435–446.

Dale, V. H., Efroymson, R. A., Kline, K. L., Langholtz, M. H., Leiby, P. N., Oladosu, G. A., Davis, M. R., Downing, M. E., and Hilliard, M. R. (2013a). Indicators for assessing socioeconomic sustainability of bioenergy systems: a short list of practical measures. *Ecological Indicators*, 26:87–102.

Dale, V. H., Kline, K. L., Buford, M. A., Volk, T. A., Smith, C. T., and Stupak, I. (2016). Incorporating bioenergy into sustainable landscape designs. *Renewable and Sustainable Energy Reviews*, 56:1158–1171.

Dale, V. H., Langholtz, M. H., Wesh, B. M., and Eaton, L. M. (2013b). Environmental and socioeconomic indicators for bioenergy sustainability as applied to eucalyptus. *International Journal of Forestry Research*.

Dale, V. H. and Polasky, S. (2007). Measures of the effects of agricultural practices on ecosystem services. *Ecological economics*, 64(2):286–296.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.

Dias, L. C. and Domingues, A. R. (2014). On multi-criteria sustainability assessment: Spider-gram surface and dependence biases. *Applied Energy*, 113:159–163.

Díaz-Balteiro, L. and Romero, C. (2004). In search of a natural systems sustainability index. *Ecological Economics*, 49(3):401–405.

Dobbs, C., Escobedo, F. J., and Zipperer, W. C. (2011). A framework for developing urban forest ecosystem services and goods indicators. *Landscape and urban planning*, 99(3):196–206.

Earles, R. and Williams, N. (2005). Sustainable agriculture: An introduction. *A Publication of ATTRA, the National Sustainable.*

Ebert, U. and Welsch, H. (2004). Meaningful environmental indices: a social choice approach. *J. Environ. Econ. Manage.*, 47(2):270–283.

Efroymson, R. A. and Dale, V. H. (2015). Environmental indicators for sustainable production of algal biofuels. *Ecological Indicators*, 49:1–13.

Efroymson, R. A., Dale, V. H., Kline, K. L., McBride, A. C., Bielicki, J. M., Smith, R. L., Parish, E. S., Schweizer, P. E., and Shaw, D. M. (2013). Environmental indicators of biofuel sustainability: what about context? *Environmental management*, 51(2):291–306.

Efroymson, R. A., Dale, V. H., and Langholtz, M. H. (2016). Socioeconomic indicators for sustainable design and commercial development of algal biofuel systems. *GCB Bioenergy*.

Esty, D. C., Levy, M., Srebotnjak, T., and De Sherbinin, A. (2005). Environmental sustainability index: benchmarking national environmental stewardship. *New Haven Yale Cent. Environ. Law Policy*, pages 47–60.

Eurostat (2009). Measuring progress towards a more sustainable europe. In *2009 monitoring report of the EU sustainable development strategy*. Office for Official Publications of the European Communities, Luxembourg.

Fiksel, J., Eason, T., and Frederickson, H. (2012). A Framework for Sustainability Indicators at EPA. *Natl. Risk Manag. Res. Lab. Cincinnati, OH, USA*.

Finco, M. V. A. and Doppler, W. (2010). Bioenergy and sustainable development: the dilemma of food security and climate change in the brazilian savannah. *Energy for sustainable development*, 14(3):194–199.

Freudenberg, M. (2003). Composite indicators of country performance. *OECD Science, Technology and Industry Working Papers*.

García, C. A., Riegelhaupt, E., Ghilardi, A., Skutsch, M., Islas, J., Manzini, F., and Masera, O. (2015). Sustainable bioenergy options for mexico: Ghg mitigation and costs. *Renewable and Sustainable Energy Reviews*, 43:545–552.

Gasparatos, A. and Scolobig, A. (2012). Choosing the most appropriate sustainability assessment tool. *Ecological Economics*, 80:1–7.

Gebler, M., Uiterkamp, A. J. S., and Visser, C. (2014). A global sustainability perspective on 3d printing technologies. *Energy Policy*, 74:158–167.

Giordano, F. R. and Weir, M. D. (1985). *A First Course in Mathematical Modeling*. Brooks/Cole, Monterey, 1 edition.

Gómez-Limón, J. A. and Sanchez-Fernandez, G. (2010). Empirical evaluation of agricultural sustainability using composite indicators. *Ecological Economics*, 69(5):1062–1075.

Grabisch, M., Labreuche, C., and Vansnick, J.-C. (2003). On the extension of pseudo-boolean functions for the aggregation of interacting criteria. *European Journal of Operational Research*, 148(1):28–47.

Grabisch, M., Marichal, J.-L., Mesiar, R., and Pap, E. (2009). *Aggregation Functions*. Cambridge University Press, New York, 1 edition.

Griffith, M. (2012). Natural Variability — CADDIS: Data Analysis — US EPAA. `http://www.epa.gov/caddis/da_advanced_3.html`.

Hacking, T. and Guthrie, P. (2008). A framework for clarifying the meaning of Triple Bottom-Line, Integrated, and Sustainability Assessment. *Environ. Impact Assess. Rev.*, 28(2-3):73–89.

Hayashi, T., van Ierland, E. C., and Zhu, X. (2014). A holistic sustainability assessment tool for bioenergy using the global bioenergy partnership (gbep) sustainability indicators. *Biomass and Bioenergy*, 66:70–80.

Heijungs, R., Guinée, J., Kleijn, R., and Rovers, V. (2007). Bias in normalization: causes, consequences, detection and remedies. *The International Journal of Life Cycle Assessment*, 12(4):211–216.

Heink, U. and Kowarik, I. (2010). What are indicators? on the definition of indicators in ecology and environmental planning. *Ecological Indicators*, 10(3):584–593.

International Institute for Sustainable Development (2001). The dashboard of sustainability. `http://www.iisd.org`. [Online; accessed 3-August-2014].

ISO (2009). 13824: General principles on risk assessment of systems involving structures. `http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44140`.

Johnson, T. L., Bielicki, J. M., Dodder, R. S., Hilliard, M. R., Kaplan, P. O., and Miller, C. A. (2013). Advancing sustainable bioenergy: evolving stakeholder interests and the relevance of research. *Environmental management*, 51(2):339–353.

Joyce, L., Running, S., Breshears, D., Dale, V., Malmsheimer, R., Sampson, R., Sohngen, B., and Woodall, C. (2014). Ch.7: Forests. In Melillo, J., Tichmond, T. T., and Yohe, G., editors, *Climate Impacts in the United States: The Third National Climate Assessment*, pages 175–194. U.S. Global Change Research Program. doi:10.7920/J0Z60KZC.

Kopmann, A. and Rehdanz, K. (2013). A human well-being approach for assessing the value of natural land areas. *Ecological Economics*, 93:20–33.

Krajnc, D. and Glavič, P. (2005). A model for integrated assessment of sustainable development. *Resources, Conservation and Recycling*, 43(2):189–208.

Langhans, S. D., Reichert, P., and Schuwirth, N. (2014). The method matters: A guide for indicator aggregation in ecological assessments. *Ecological Indicators*, 45:494–507.

Lele, S. and Srinivasan, V. (2013). Disaggregated economic impact analysis incorporating ecological and social trade-offs and techno-institutional context: A case from the western ghats of india. *Ecological Economics*, 91:98–112.

Lenzen, M. (2007). Aggregation (in-) variance of shared responsibility: A case study of australia. *Ecological Economics*, 64(1):19–24.

Levin, S. and Clark, W. (2010). Toward a science of sustainability: report from toward a science of sustainability conference. *Princeton Environmental Institute, Princeton.*

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological review*, 66(2):81.

Maes, D., Van Dael, M., Vanheusden, B., Goovaerts, L., Reumerman, P., Luzardo, N. M., and Van Passel, S. (2015). Assessment of the sustainability guidelines of eu renewable energy directive: the case of biorefineries. *Journal of Cleaner Production*, 88:61–70.

Mann, S. (2009). Computing for sustainability: Visualising sustainability. `http://computingforsustainability.com/2009/03/15/visualising-sustainability/`. [Online; accessed 04-Feb-2015].

Marin, G., Mazzanti, M., and Montini, A. (2012). Linking namea and input output for consumption vs. production perspectiveanalyses: Evidence on emission efficiency and aggregation biases using the italian and spanish environmental accounts. *Ecological Economics*, 74:71–84.

Maxim, A. (2014). Sustainability assessment of electricity generation technologies using weighted multi-criteria decision analysis. *Energy Policy*, 65:284–297.

Mayer, A. L. (2008). Strengths and weaknesses of common sustainability indices for multidimensional systems. *Environment international*, 34(2):277–291.

McBride, A. C., Dale, V. H., Baskaran, L. M., Downing, M. E., Eaton, L. M., Efroymson, R. A., Garten, C. T., Kline, K. L., Jager, H. I., Mulholland, P. J., Parish, E. S., Schweizer, P. E., and Storey, J. M. (2011). Indicators to support environmental sustainability of bioenergy systems. *Ecol. Indic.*, 11(5):1277–1289.

McLachlan, G. and Peel, D. (2004). *Finite mixture models.* John Wiley & Sons.

McRae, L., Collen, B., Deinet, S., Hill, P., Loh, J., Baillie, J., and Price, V. (2012). The Living Planet Index. In Almond, R., editor, *The Living Planet Report 2012*, Gland, Switzerland. WWF International.

Millennium Ecosystem Assessment (MEA) (2005). Ecosystems and human well-being: Synthesis. `http://www.millenniumassessment.org/en/index.html`. [Online; accessed 19-May-2016].

Moldan, B., Janoušková, S., and Hák, T. (2012). How to understand and measure environmental sustainability: Indicators and targets. *Ecol. Indic.*, 17:4–13.

Mori, K. and Christodoulou, A. (2012). Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI). *Environ. Impact Assess. Rev.*, 32(1):94–106.

Mori, K. and Yamashita, T. (2015). Methodological framework of sustainability assessment in city sustainability index (csi): A concept of constraint and maximisation indicators. *Habitat International*, 45:10–14.

Msangi, S., Sulser, T., Rosegrant, M., and Valmonte-Santos, R. (2007). Global scenarios for biofuels: Impacts and implications for food security and water use. In *10th Annual Conference on Global Economic Analysis, Purdue University, Indiana.*

Myllyviita, T., Leskinen, P., Lähtinen, K., Pasanen, K., Sironen, S., Kähkönen, T., and Sikanen, L. (2013). Sustainability assessment of wood-based bioenergy A methodological framework and a case-study. *Biomass and Bioenergy*, 59:293–299.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., and Giovannini, E. (2005). *Handbook on constructing composite indicators*. OECD publishing.

Nations, U. (2010). *The Millennium Development Goals Report 2010*. United Nations Publications, New York.

Odum, H. T., Doherty, S. J., Scatena, F. N., and Kharecha, P. A. (2000). Emergy evaluation of reforestation alternatives in puerto rico. *Forest Science*, 46(4):521–530.

OECD, O. (2003). Environmental indicators–development, measurement and use. *OCDE, Paris*.

Padowski, J. C. and Jawitz, J. W. (2012). Water availability and vulnerability of 225 large cities in the united states. *Water Resources Research*, 48(12).

Parish, E. S., Dale, V. H., English, B. C., Jackson, S. W., and Tyler, D. D. (2016). Assessing multimetric aspects of sustainability: Application to a bioenergy crop production system in east tennessee. *Ecosphere*, 7(2).

Parish, E. S., Hilliard, M. R., Baskaran, L. M., Dale, V. H., Griffiths, N. A., Mulholland, P. J., Sorokine, A., Thomas, N. A., Downing, M. E., and Middleton, R. S. (2012). Multimetric spatial optimization of switchgrass plantings across a watershed. *Biofuels, Bioproducts and Biorefining*, 6(1):58–72.

Pearce, D. W., Atkinson, G. D., and Dubourg, W. R. (1994). The economics of sustainable development. *Annual Review of Energy and the Environment*, 19(1):457–474.

Pinar, M., Cruciani, C., Giove, S., and Sostero, M. (2014). Constructing the feem sustainability index: A choquet integral application. *Ecological Indicators*, 39:189–202.

Pollesch, N. and Dale, V. H. (2015). Applications of aggregation theory to sustainability assessment. *Ecological Economics*, 114:117–127.

Pollesch, N. and Dale, V. H. (*In Press*). Normalization in sustainability assessment: Methods and implications. *Ecological Economics*.

Roberts, F. S. (1994). *Operations Research and The Public Sector*, volume 6 of *Handbooks in Operations Research and Management Science*. Elsevier.

Roberts, F. S. (2014a). Meaningful and meaningless statements in landscape ecology and environmental sustainability. In *Clusters, Orders, and Trees: Methods and Applications*, pages 297–312. Springer.

Roberts, F. S. (2014b). Meaningful and meaningless statements in landscape ecology and environmental sustainability. In Pardalos, P., Goldengorin, B., and Aleskerov, F., editors, *Clusters, Orders, Trees: Methods and Applications*, pages 297–312. Springer.

Roundtable on Sustainable Biomaterials (RSB) (2011). Roundtable on sustainable biofuels - indicators of compliance for the rsb principles & criteria. http://www.rsb.org/pdfs/standards/11-03-08\%20RSB\%20Indicators\%202-1.pdf. [Online; accessed 11-May-2016].

Sadamichi, Y., Kudoh, Y., Sagisaka, M., Chen, S. S., Elauria, J. C., Gheewala, S. H., Hasanudin, U., Romero, J., Shi, X., and Sharma, V. K. (2012). Sustainability assessment methodology of biomass utilization for energy in east asian countries. *J Jpn Inst Energy*, 91(10):960–968.

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

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Scienceman, D. M. (1987). Energy and emergy. *by G. Pillet and T. Murota. Roland Leimgruber, Geneva*, pages 257–276.

Shi, D., Devineni, N., Lall, U., and Piñero, E. (2013). America's water risk: Water stress and climate variability. http://water.columbia.edu/files/2013/09/GB_CWC_whitepaper_climate-water-stress_final.pdf. [Online; accessed 28-July-2014].

Singh, R. K., Murty, H., Gupta, S., and Dikshit, A. (2009). An overview of sustainability assessment methodologies. *Ecol. Indic.*, 9(2):189–212.

Smeets, E., Weterings, R., and voor Toegepast-Natuurwetenschappelijk, N. C. O. (1999). *Environmental indicators: Typology and overview*. European Environment Agency Copenhagen.

Stevens, S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684):677 – 680.

Stiglitz, J., Sen, A., and Fitoussi, J.-P. (2009). The measurement of economic performance and social progress revisited. *Reflections and overview. Commission on the Measurement of Economic Performance and Social Progress, Paris*.

Su, B. and Ang, B. (2010). Input–output analysis of CO_2 emissions embodied in trade: The effects of spatial aggregation. *Ecological Economics*, 70(1):10–18.

Sullivan, P. (2003). Applying the principles of sustainable farming. *The National Sustainable*.

Tait, P., Baskaran, R., Cullen, R., and Bicknell, K. (2012). Nonmarket valuation of water quality: Addressing spatially heterogeneous preferences using gis and a random parameter logit model. *Ecological Economics*, 75:15–21.

Tatari, O., Kucukvar, M., and Onat, N. C. (2015). Towards a triple bottom line life cycle sustainability assessment of buildings. In *Science for Sustainable Construction and Manufacturing Workshop Volume I. Position Papers and Findings*, page 226.

The Global Bioenergy Partnership (GBEP) (2011). The global bioenergy partnership sustainability indicators for bioenergy, first edition (final version, december 15, 2011). http://www.globalbioenergy.org/fileadmin/user_upload/gbep/docs/Indicators/The_GBEP_Sustainability_Indicators_for_Bioenergy_FINAL.pdf. [Online; accessed 11-May-2016].

Tiller, K. (2011). Tennessee case study in full-scale development and integration of switchgrass feedstocks with biorefinery operations: Genera energy and university of tennessee biofuels initiative. *Industrial Biotechnology*, 7(5):357–364.

Walker, F. (2000). Best management practices for phosphorus in the environment. *Agricultural Extension Service, University of Tennessee, PB*, 1645.

Wilson, J., Tyedmers, P., and Pelot, R. (2007). Contrasting and comparing sustainable development indicator metrics. *Ecol. Indic.*, 7(2):299–314.

Winands, S., Holm-Müller, K., and Weikard, H.-P. (2013). The biodiversity conservation game with heterogeneous countries. *Ecological Economics*, 89:14–23.

Zhou, P., Ang, B., and Poh, K. (2006). Comparing aggregating methods for constructing the composite environmental index: An objective measure. *Ecol. Econ.*, 59(3):305–311.

Appendix

1. Information Theoretic Aggregation for Sustainability Assessment: Model Selection of Beta Mixtures

1.1. Motivation for Research into Mixtures of Beta Distributions

This paper is written as a proof-of-concept example of how mixtures of Beta distributed random variables can be identified, using model selection criteria, to find a mixture distribution that is optimal in an information theoretic context. Beta random variables in this case represent indicators of sustainability. As opposed to aggregating distributions of these indicators based on some known relationship between them, this work is presented as a method for aggregation based on objective information criteria. This work may be useful as the Bioenergy Sustainability Target Assessment Resource (Bio-STAR) is developed and expanded to move from analysis of datasets associated with indicators of bioenergy to a distributional approach for describing indicators. For general background on finite mixture models one can see McLachlan and Peel (2004), although mixtures of Beta distributions are not treated explicitly in the text.

1.2. Introduction

Sustainability assessment describes the process of using measurements of indicator variables [1] that represent various aspects of environmental, economic, and societal aspects of the system being studied. Sustainability assessments have utilized numbers of indicator variables in excess of 2000. Indicator variables are also chosen to provide multiple dimensions of system performance.

Recent identification of indicator variables has been performed by McBride et al. (2011) and Dale et al. (2013a) for bioenergy production sustainability. Given the variety of indicator variables that are utilized in assessment, normalization of measurements of indicator variable data is often used to generate data that can be compared, simplified, or aggregated. Although different normalization methods exist, such as monetization or z-score standardization, target normalization has been identified as the procedure that may be the most appropriate (Mayer, 2008). Suitability of target normalization is argued, in part, since normalized indicator variable measurements are then tied to specific baseline and/or target levels that are relevant to each indicator variable and the associated sustainability goals for that particular indicator variable. Pollesch and Dale, *In Press* also show beneficial mathematical properties of target normalization compared to other common normalization techniques.

In addition to understanding multiple indicator measures for a specific system, sustainability assessments are also used to compare the performance of multiple systems. For example, within the bioenergy production context, one may wish to compare production routes such as switchgrass bioenergy production in Tennessee to miscanthus bioenergy production in Illinois, or corn ethanol in Iowa to sugarcane ethanol in Brazil. The indicators identified by McBride et al. (2011) and Dale et al. (2013a) were chosen to help facilitate such comparisons (see Table A.1).

Comparisons of bioenergy production routes may be done with a goal of identifying production sites that are near to their target levels for some set of indicators or those that are far away. Such a cross-site analysis may benefit from a clustering or group identification technique to distinguish, statistically, those sites that have indicator measurements values that separate them from the other sites. Given that distance-to-target normalized indicator measurements may be complete so

[1] The term *indicator variable* is used in sustainability assessment to describe a quantity being assessed that indicates something about the state of the system. It is not to be confused with the probabilistic use of indicator variable, $\mathbb{1}_{\{\Theta\}}(\theta)$, a function that is 1 when $\theta \in \Theta$ and 0 else.

Table A.1: Recommended environmental indicators for bioenergy sustainability *adapted from McBride et al. (2011)*

Category	Indicator	Units
Soil quality	1. Total organic carbon (TOC)	Mg/ha
	2. Total nitrogen (N)	Mg/ha
	3. Extractable phosphorus (P)	Mg/ha
	4. Bulk Density	g/cm^3
Water quality and quantity	5. Nitrate concentration in streams (and export)	Concentration: mg/L; export: kg/ha/year
	6. Total phosphorus (P) concentration in streams (and export)	Concentration: mg/L; export kg/ha/year
	7. Suspended sediment concentration in streams (and export)	Concentration: mg/L; export kg/ha/year
	8. Herbicide concentration in streams (and export)	Concentration: mg/L; export kg/ha/year
	9. Peak storm flow	L/s
	10. Minimum base flow	L/s
	11. Consumptive water use (incorporates base flow)	Feedstock production: m^3/ha/day; biorenery: m^3/day
Greenhouse gases	12. CO_2 equivalent emissions (CO_2 and N_2O)	kg C_{eq}/GJ
Biodiversity	12. Presence of taxa of special concern	Presence
	14. Habitat area of taxa of special concern	ha
Air Quality	15. Tropospheric ozone	ppb
	16. Carbon monoxide	ppm
	17. Total particulate matter less than 2.5μm diameter ($PM_{2.5}$)	μg/m^3
	18. Total particulate matter less than 10μm diameter (PM_{10})	μg/m^3
Productivity	19. Aboveground net primary productivity (ANPP)/yield	g C/m^2/year

that each normalized measure falls in the interval $(0, 1)$, the Beta distribution is a natural choice for distribution of the random variable used to represent the normalized indicator measures for a given site. This paper develops a technique that uses mixtures of Beta random variables to identify clusters of bioenergy production sites that have similar performance with respect to their sustainability baselines and goals for a given indicator variable.

1.3. Problem Description

Each $X_i \sim Beta(\alpha, \beta)$ included in the mixture may be used to represent an indicator variable i for some site s, where there are S to be considered in the analysis. Alternatively, one might also view each X_i as representing an indicator for the same site, and one can consider the mixture an aggregate distribution for some number of indicators at the same site. In this paper, we will take the viewpoint of a mixture of the same indicator from multiple sites.

A Beta random variable, $X \sim Beta(\alpha, \beta)$, has pdf

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

A beta mixture with S components is defined as:

$$f(X|\Theta) = \sum_{i=1}^{S} \pi_i f(X|\theta_i)$$

where $0 < \pi_i \leq 1$ with $\sum_{i=1}^{S} \pi_i = 1$ are the mixing proportions and $f(x|\theta_i)$ represent the Beta distributions. The parameter set is written as $\Theta = \{\theta, \pi\}$ with $\theta = (\theta_1, ..., \theta_S), \theta_i = (\alpha_i, \beta_i)$ and $\pi = (\pi_1, ..., \pi_S)$.

This paper considers two different problems associated with Beta mixture distributions:

1. Determination of the mixing parameters, π_i, when the number of components and parameters for the Beta distributions in the mixture are assumed to be known.

2. Model selection on the number of mixtures in the model using information theoretic approaches and information criteria for scoring when the number of components and parameters for the Beta distributions in the mixtures are assumed to be known.

For both problems listed above, it is assumed that there is a single normalized indicator being measured at multiple bioenergy production sites. These measurements are then used to parameterize a Beta random variable for each site, which are included in the mixture model.

1.4. Method

Let $X_i \sim Beta(\alpha, \beta)$ represent the distribution of normalized indicator measures for site i. Although this method can be generalized to larger or small numbers of production sites, assume that there are 5 sites included in the assessment. Parameters $\theta_i = (\alpha_i, \beta_i)$ for each X_i are randomly generated by sampling from $Uniform[0, 10]$. Then mixing parameters, π_i are randomly assigned to each X_i by assigning $\pi_i, i = 1, 2, 3, 4$ a sample from $Uniform[0, 1/4]$ and then letting $\pi_5 = 1 - \sum_{i=1}^{4} \pi_i$. The generated mixture model is then created in *Mathematica* by using the function *MixtureDistribution[]* (Figure A.1). In the case where mixing parameters are sought to be estimated,

the mixing distribution is sampled 500 times to create the simulated data set (Figure A.2). For the model selection component, 100 data points are simulated for each X_i individually.

The expectation-maximization (EM) algorithm is used to find mixing parameters,π_i in the first problem, and it is also used to find the entire mixing parameter set, Θ, in the model selection problem (Dempster et al., 1977). Model selection may be completed using a number of information criterion, such as Akaike Information Criteria, Bayesian Information Criteria, or Bozdogan's Information Complexity Criteria, ICOMP. In this problem the Akaike Information Criteria (AIC) as well as the Schwarz Bayesian Information Criteria (SBIC) are used to score models (Schwarz et al., 1978; Akaike, 1971).

The process used used to conduct the EM algorithm is as follows:

1. Calculation of the probability that data point x_j belongs to mixture component θ_i. Using Bayes formula the probability that x_j belongs to component θ_i:

$$f(\theta_i|x_j,\Theta) = \frac{\pi_i f(x_j|\theta_i)}{\sum_{s=1}^{S} \pi_s f(x_j|\theta_s)}$$

2. Update mixture proportions by computing the following:

$$\pi_i^{new} = \frac{1}{S}\sum_{j=1}^{N} f(\theta_i|x_j,\Theta)$$

3. In the case when distribution parameters $\{\alpha_i, \beta_i\}$ are to be estimated:

   (a) Mean values are calculated as follows:

$$\mu_i^{new} = \frac{\sum_{j=1}^{N} x_j f(\theta_i|x_j,\Theta)}{\sum_{j=1}^{N} f(\theta_i|x_j,\Theta)}$$

   (b) Variance values are calculated as follows:

$$\sigma_i^{new} = \frac{\sum_{j=1}^{N} f(\theta_i|x_j,\Theta)(x_j-\mu_k^{new})^2}{\sum_{i=1}^{N} f(\theta_i|x_j,\Theta)}$$

   (c) The mean and covariance values, μ_i^{new} and σ_i^{new} are used to calculate the parameters as follows:

$$\alpha_i^{new} = \frac{\mu_i^{new}\beta_i^{new}}{1-\mu_i^{new}} \text{ and } \beta_i^{new} = \frac{\sigma_i^{new}\mu_i^{new} - \sigma_i^{new} + (\mu_i^{new})^3 - 2(\mu_i^{new})^2 + \mu_i^{new}}{\sigma_i^{new}}$$

Following the iteration scheme laid out above, the stopping criteria in the *While* loop is defined using the Kullback-Leibler divergence defined as:

$$d_{KL}(f_1||f_2) = \int log\left[\frac{f_1(x)}{f_2(x)}\right] f_1(x)dx$$

Table A.2: Components and mixing parameters for Beta distributions used to simulate data in experiment

$X_1 \sim Beta(0.45783, 7.64709)$	$\pi_1 = 0.123537$
$X_2 \sim Beta(2.63789, 1.60009)$	$\pi_2 = 0.17774$
$X_3 \sim Beta(3.69424, 2.97784)$	$\pi_3 = 0.195568$
$X_4 \sim Beta(8.05187, 6.11292)$	$\pi_4 = 0.0438734$
$X_5 \sim Beta(7.8943, 7.01838)$	$\pi_5 = 0.459282$

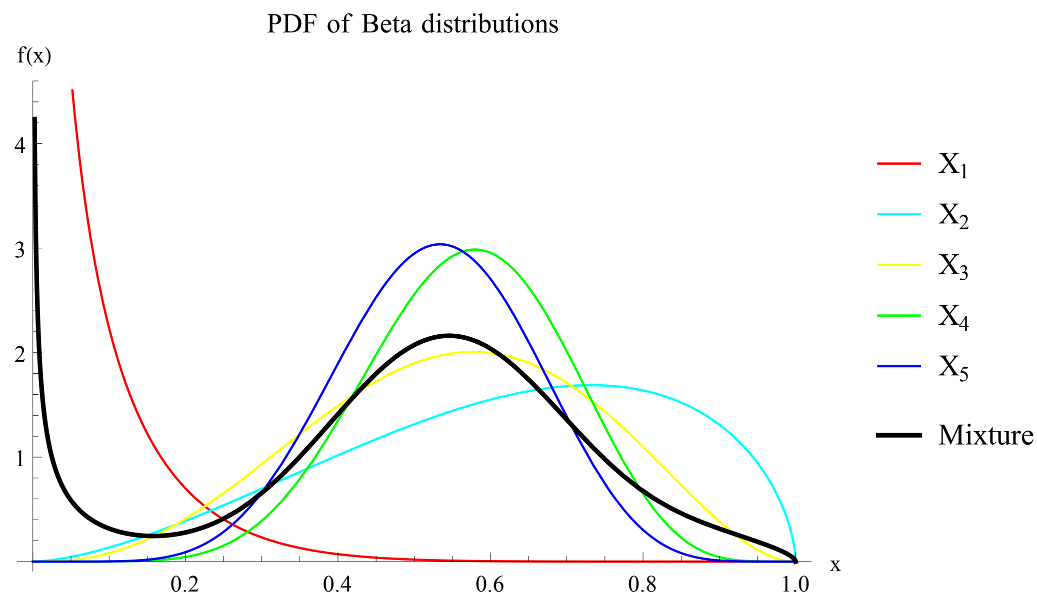


Figure A.1: Probability distribution functions of individual mixture components, $f(X|\theta_i)$, and complete mixture distribution $f(X|\Theta) = \sum_{i=1}^{S} \pi_i f(X|\theta_i)$

Although multiple values for d_{KL} can be used to stop the iteration scheme, a value of $d_{KL} = 1{*}10^{-6}$ is used here.

1.5. Results

1.5.1. Estimation of Mixing Parameters

The five components and true mixing parameters used to simulate the data in the mixture distribution, $f(X|\Theta) = \sum_{i=1}^{S} \pi_i f(X|\theta_i)$, were generated according to the methods previously described and are contained in Table A.2.

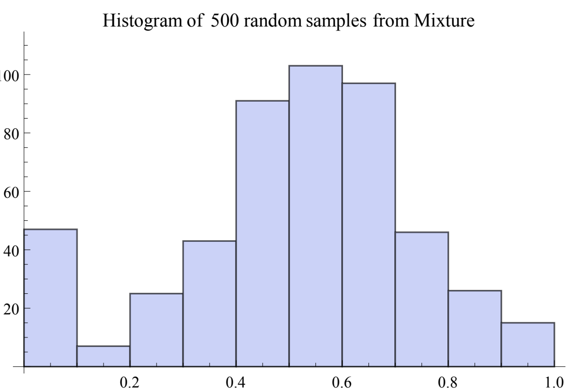


Figure A.2: 500 random samples from complete mixture distribution $f(X|\Theta) = \sum_{i=1}^{S} \pi_i f(X|\theta_i)$

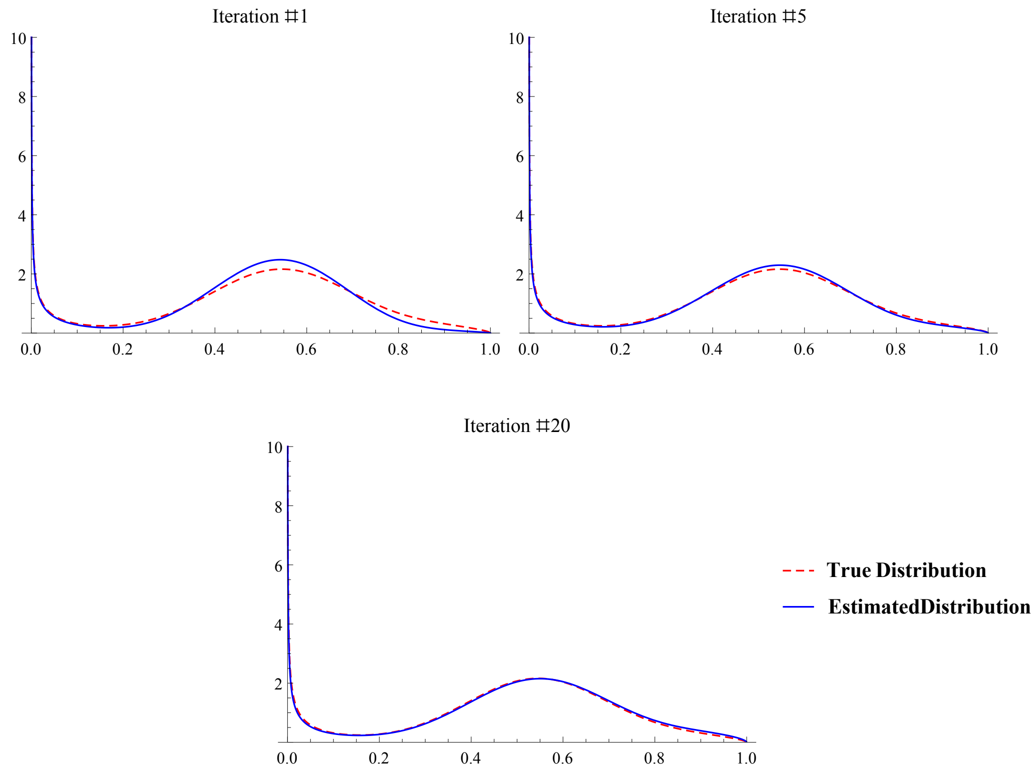


Figure A.3: Convergence of estimated distribution, $f(X|\Theta)^{(t)} = \sum_{i=1}^{S} \pi_i^{(t)} f(X|\theta_i)$. Although not labeled within figure, the ordinate axes are the true and estimated Beta probability distribution functions and abscissa are x.

1.5.2. Results from EM Algorithm for Estimation of Mixing Parameters

Beginning the EM algorithm using randomly generated starting mixing proportions given by

$$\pi^{(0)} = (0.0606277, 0.00793612, 0.057018, 0.122814, 0.751604)$$

produced the results included in Figures A.3, A.4, and A.5 below.

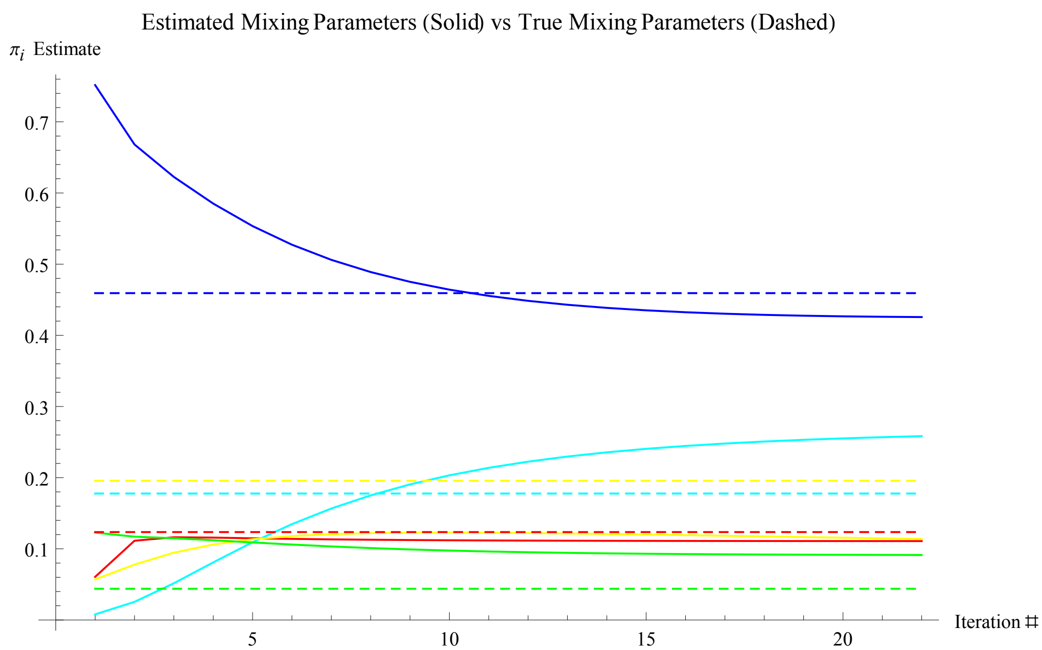


Figure A.4: Convergence of estimated mixing parameters, π_i

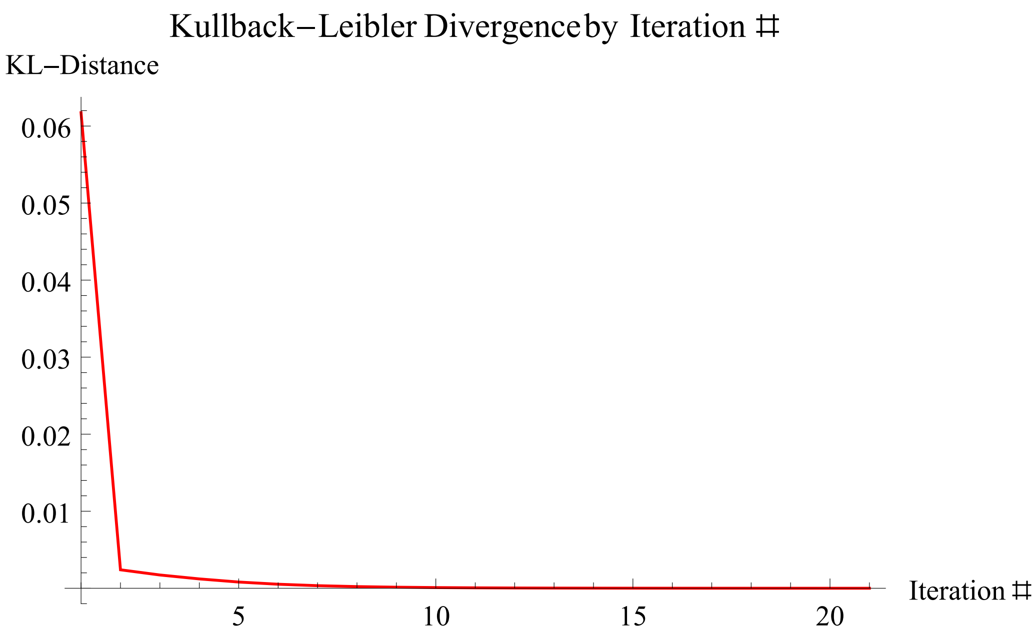


Figure A.5: Graph of KL divergence, d_{KL}, as a function of iteration number

Table A.3: Results of model selection using EM algorithm and AIC and SBIC scoring

Model	Resulting Best Fit Distribution	AIC	SBIC
True	$0.1235 * Beta(0.4578, 7.6471) + 0.1777 * Beta(2.6379, 1.6001)$ $+ 0.1956 * Beta(3.6942, 2.9778) + 0.0439 * Beta(8.0519, 6.1129)$ $+ 0.4593 * Beta(7.8943, 7.0184)$	-265.587	-202.367
5	$0.1190 * Beta(0.2936, 4.1697) + 0.1862 * Beta(12.4762, 10.0026)$ $+ 0.0290 * Beta(22.2633, 13.0948) + 0.5067 * Beta(5.3001, 4.9414)$ $+ 0.1593 * Beta(4.1007, 1.4631)$	-270.52	-207.301
4	$0.3583 * Beta(10.4648, 8.323) + 0.2571 * Beta(5.4156, 5.4778)$ $+ 0.2071 * Beta(3.5420, 1.7887) + 0.1774 * Beta(0.2400, 0.7294)$	-275.782	-225.207
3	$0.1950 * Beta(0.2757, 0.7540) + 0.1040 * Beta(6.8493, 2.0416)$ $+ 0.7010 * Beta(7.1216, 6.1327)$	-281.716	-243.784
2	$0.7588 * Beta(6.2147, 4.9760) + 0.2412 * Beta(0.3163, 0.5550)$	-283.587	-258.299

1.5.3. Discussion of Estimation of Mixing Parameters

The EM algorithm used to find mixing parameters, for the first problem addressed, gave results that may be seen as adequate. Although the EM algorithm converged to a distribution that was in some sense *close* to the true distribution (see Figure A.3), the mixing parameter estimates did not necessarily converge as nicely (see Figure A.4).

1.6. Model Selection for Beta Mixtures Models

Utilizing the same *true model parameters* presented in section 4.1, model selection using information criteria was conducted to determine if 5, 4, 3, or 2 components could best describe the data generated from the 5 component mixture. The EM algorithm technique described in the Methods section was utilized along with the AIC and SBIC information criteria for scoring models after Kullback-Leibler convergence had been achieved, where $AIC = 2 * p - 2\log[f(\hat{\theta}|data)]$ and $SBIC = p * \log[N] - 2\log[f(\hat{\theta}|data]$.

1.6.1. Results from EM Algorithm for Model Selection for Beta Mixture Models

Table A.3 contains the results from the model selection analysis. It was found that the 2 component model was the best to fit the data generated under both information criteria. Figure A.6 shows the 2-component, minimum information mixture model, selected compared to the true model used to generate the data.

1.6.2. Discussion of Model Selection

The model selected in this case was the model that only included a mixture of two Beta random variables. In this case, it was shown that the 2-component model found actually had a lower AIC and SBIC score than the true model. This result is interesting in the fact that two Beta distributions found can be mixed in such a way that fits the 5-component mixture well. This result may be due in part to the fact that the Beta distribution has a rich number of forms it can take on over its support. However, in order to fully make conclusions on optimal fits, one should conduct multiple

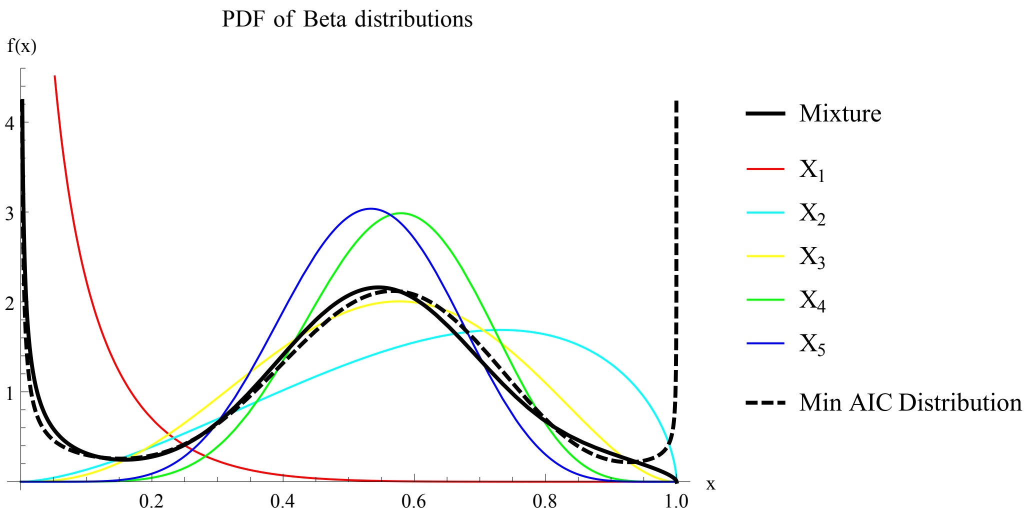


Figure A.6: Comparison of PDF for *true* Beta components and Beta mixture distribution with 2-component, minimum information mixture model, selected through EM algorithm

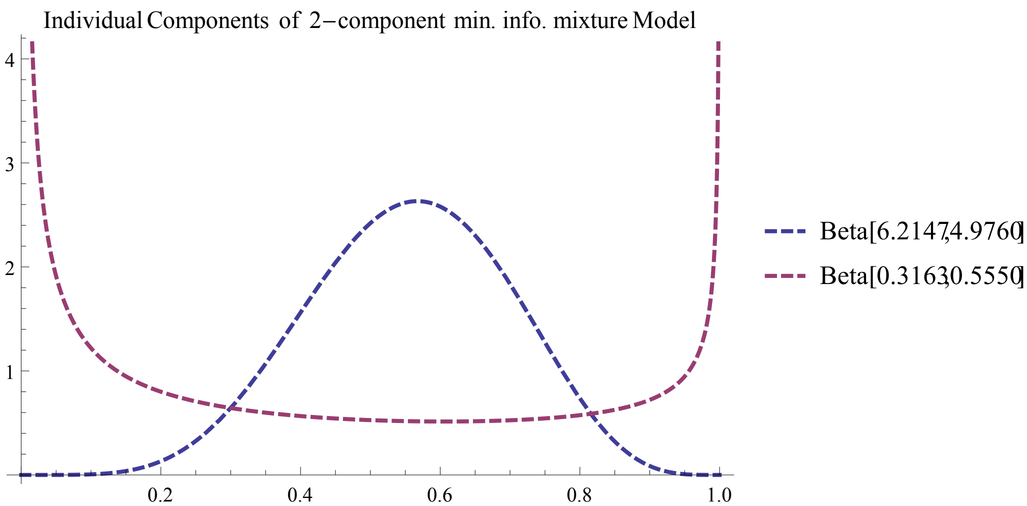


Figure A.7: The PDF of individual components of the 2-component, minimum information mixture model. Although not labeled within figure, the ordinate axes are the true and estimated Beta probability distribution functions and abscissa are x.

simulations using various sets of *true model parameters*. Figure A.7 shows the two components of the minimum information criteria.

1.7. Future Research and Extensions

In this paper, simulated data were used as a proof of concept and to develop the tools necessary to treat the data from bioenergy sustainability assessment sites once it becomes available. Future research will seek to use actual data collected in order to conduct real analysis of bioenergy sustainability indicator measurements.

Given that sustainability is not perceived as the performance of a single indicator variable, but the performance of multiple indicators across multiple categories, in future research, clusters for groups of indicator variables may be approached. This approach would allow one to identify sites that are grouped based on categories of indicators, such as soil quality, or economic performance. Furthermore, aggregation approaches are also considered a key component in sustainability assessment and an understanding of how aggregate values of indicators lead to specific groupings may also be desirable.

On the side of theoretical development, one could explore the Beta distribution (α, β) parameter space systematically and vary not only parameters but also the number of components included in the original mixture. This extended approach may lead to building a clear intuition as to how sets of parameters, and their corresponding distributions interact.

1.8. Acknowledgments

The *Expectation Maximization algorithm* as presented by Alexander Churbanov and available at *http://www.wyomingbioinformatics.org/ achurban/docs/presentationEM.pdf* was very helpful in developing the *Mathematica* notebook in which the analyses were carried out. Further, the Matlab code written by Brani Vidakovic available at *http://www2.isye.gatech.edu/ brani/isyebayes/bank/handout12.pdf* gave further insight into how EM could be implemented.

2. Glossary of Terms

Major sources for this glossary are Dale et al. (2016) and IPCC Glossary of Terms: *https://www.ipcc.ch/pdf/special-reports/srex/SREX-Annex/Glossary.pdf*

Aggregate A collection of items that are gathered together to form a total quantity

Baseline/reference The baseline (or reference) is the state against which change is measured. It might be a 'current baseline,' in which case it represents observable, present-day conditions. It might also be a 'future baseline,' which is a projected future set of conditions excluding the driving factor of interest. Alternative interpretations of the reference conditions can give rise to multiple baselines.

Bearing (Normalization Bearing) This term is used to classify what type of normalization should take place for a given indicator. Common normalization bearings are "larger the better", "smaller the better", and "not too much, not too little". See the manuscript *Normalization in Sustainability Assessment* for a further description (Pollesch and Dale, *in press*).

Bioenergy Renewable energy made from materials derived from non-fossil, biological sources.

Bioenergy supply chain Feedstock production, logistics of accumulating and transporting the feedstock, creation of the energy or fuel, transport and final use of the energy, and decommissioning as appropriate at end of life for equipment or facilities.

Biomass Any living or recently living organic material that has stored sunlight in the form of chemical energy, including plants, residues from agriculture or forestry, or the organic component of municipal and industrial wastes.

Catchment A topographically defined area that collects and drains precipitation, often a subset of a watershed.

Confidence Confidence in the validity of a finding, based on the type, amount, quality, and consistency of evidence and on the degree of agreement. Confidence is expressed qualitatively.

Context The context of a sustainability assessment includes the purpose, the particular biofuel production and distribution system, policy conditions, stakeholder values, location, temporal influences, spatial scale, baselines, and reference scenarios (Efroymson et al. 2013).

Extent The space, time, or degree to which a thing extends; length, area, volume, time, or scope. This term is used to specify the spatial area/volume and time span/period(s) associated with a set of indicator measurements or estimates derived from those measurements. It is used to differentiate from the associated term 'resolution'. Extent is used to specify extent of analysis and extent of measurement.

Extent of Analysis The set of areas/volumes in space and span(s) of time that are considered in an analysis.

Extent of Measurement This is the set of area/volumes in space and span(s) of time that are represented by a single measurement. I.e. how representative of space or time is a single measurement. Example: There is single survey response by a farmer, however this response is meant to represent his entire farm (spatial) and for the year (temporal).

Greenhouse gas Those gaseous constituents of the atmosphere, both natural and anthropogenic, which absorb and emit radiation at specific wavelengths within the spectrum of thermal infrared radiation emitted by the Earth's surface, by the atmosphere itself, and by clouds. This property causes the greenhouse effect. Water vapor (H2O), carbon dioxide (CO2), nitrous oxide (N2O), methane (CH4), and ozone (O3) are the primary greenhouse gases in the Earth's atmosphere. Moreover, there are a number of entirely human-made greenhouse gases in the atmosphere, such as the halocarbons and other chlorine- and bromine containing substances, dealt with under the Montreal Protocol. Besides CO2, N2O, and CH4, the Kyoto Protocol deals with the greenhouse gases sulfur hexafluoride (SF6), hydrofluorocarbons (HFCs), and perfluorocarbons (PFCs). Governance: The way government is understood has changed in response to social, economic, and technological changes over recent decades. There is a corresponding shift from government defined strictly by the nation-state to a more inclusive concept of governance, recognizing the contributions of various levels of government (global, international, regional, local) and the roles of the private sector, of nongovernmental actors, and of civil society.

Land use and land-use change Land use refers to the total of arrangements, activities, and inputs undertaken in a certain land cover type (a set of human actions). The term land use is sometimes used in the sense of the social and economic purposes for which land is managed (e.g., grazing, timber extraction, and conservation). Land-use change refers to a change in the management of land by humans, which may or may not lead to a change in land cover.

Landscape A perspective of how phenomena occur in a region that includes both pattern and process.

Landscape design A spatially explicit, collaborative plan for integrated management of landscape resources and supply chains, developed by an informal or formal group of stakeholders around a set of specified goals.

Predictability The extent to which future states of a system may be predicted based on knowledge of current and past states of the system.

Projection A projection is a potential future development of a quantity or set of quantities, often computed with the aid of a model. Projections are distinguished from predictions in order to emphasize that projections involve assumptions concerning, for example, future socioeconomic and technological developments that may or may not be realized, and are therefore subject to substantial uncertainty.

Range Range is used in this case to represent a summary of the data included. For range of measurements, these are the summary statistics for the minimum and maximum values reported for a given data set. When it comes to temporal and spatial range of measurements, this is keeping track of the timespan or physical area from which measurements have been recorded for a given indicator.

Spatial Range of Measurements The span of area covered in the data set i.e. multiple counties, multiple states, could be given in terms of spatial coordinates as well. Example: Switchgrass yields had a temporal range of measurements from 2011-2012, a spatial range of measurements of multiple counties, and range of values reported from 0 to 6.95 tons/acre.

Temporal Range of Measurements The span of time covered in the data set i.e. the oldest and most recent times a measurement was taken for a given indicator.

Range of reported values The minimum and maximum values of measurements included in the data set for a given indicator.

Renewable energy Energy that comes from resources that nature replenishes on a human timescale such as sunlight, wind, rain, tides, waves, and geothermal heat.

Resilience The ability of a system and its component parts to anticipate, absorb, accommodate, or recover from the effects of an event in a timely and efficient manner, including through ensuring the preservation, restoration, or improvement of its essential basic structures and functions.

Resolution The level of detail available for measurements of an indicator usually in either space or time. May also represent frequency of measurement. Resolution is commonly qualified as 'low' or 'high'.

Stakeholders Individuals and groups who are any part of a production supply chain, as well as those affected positively or negatively by the development and use of the product. Within a bioenergy context stakeholders in a production supply chain includes producers, transporters, and users of the product, its precursors, and it coproducts.

Sustainable development Development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

Sustainability A concept that considers development options in terms of meeting current needs while conserving opportunities for future generations to meet their needs. The term is commonly applied to consider the relative sustainability of two or more trajectories or pathways for development where one is compared to other(s) based on sustainability criteria and indicators such as those associated with land, air, water, ecosystems, the biological and human environment, nonrenewable resources, species diversity, and other clearly defined providers of ecosystem services. Sustainability is not a state but rather reflects aspirational goals (Hansen JW. 1996) and is a dynamic of human values, choices and technology. Developing and using effective and cost-efficient measures of sustainability requires (1) a limited set of indicators; (2) collection of data over appropriate spatial and temporal scales; (3) storage and analysis of those data; (4) stakeholder engagement; and (5) communicating and acting upon results (Dale et al. 2013b).

Trade-offs A situation involving diminishment or loss of one quality in exchange for enhancement of another quality.

Uncertainty An expression of the degree to which a value or relationship is unknown. Uncertainty can result from lack of information or from disagreement about what is known or even knowable. Uncertainty may originate from many sources, such as quantifiable errors in the data, ambiguously defined concepts or terminology, or uncertain projections of human behavior. Uncertainty can therefore be represented by quantitative measures, for example, a range of values calculated by various models, or by qualitative statements, for example, reflecting the judgment of a team of experts.

Vulnerability The propensity or predisposition to be adversely affected.

Vita

Nathan Pollesch was born in Milwaukee, Wisconsin. He graduated from the University of Wisconsin - Stevens Point with a Bachelor's of Science in 2009 where he studied the natural sciences and mathematics. He then attended the University of Minnesota-Duluth, graduating in 2012 with a Master's of Science in Applied and Computational Mathematics. Under the advisement of Dr. Harlan Stech, his Master's thesis utilized dynamical systems to study the effects of nutrient enrichment in the Gulf of Mexico in the wake of the 2010 Deepwater Horizon Oil Spill. He then attended the University of Tennessee - Knoxville where he earned his Ph.D. in Mathematics, with a concentration on Mathematical Ecology and a concurrent M.S. in Statistics in 2016. He was co-advised by Dr. Louis Gross and Dr. Virginia Dale. In addition to mathematical approaches to sustainability assessment, his research efforts at the University of Tennessee included dynamical systems modeling of protein translation with Dr. Mike Gilchrist and stochastic modeling approaches to study fractal porous media models with Dr. Edmund Perfect and fellow graduate student in Environmental Engineering, Micah Wyssmann. He has accepted a position of a post-doctoral researcher at the United States Environmental Protection Agency in Duluth, Minnesota, where he will research and develop methods to assess impacts of pesticides on wildlife populations with P.I. Dr. Matt Etterson.