

Navigating Uncertainty in Environmental Composite Indicators

Michael J. Burgass^{1, 2 *}, Benjamin S. Halpern^{1, 3, 4}, Emily Nicholson⁵, E.J. Milner-Gulland²

¹ Department of Life Sciences, Imperial College London, Silwood Park Campus, UK.

² Department of Zoology, University of Oxford, UK

³ National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, California, USA

⁴ Bren School of Environmental Science and Management, University of California, Santa Barbara, California, USA.

⁵ Deakin University, Australia. School of Life and Environmental Sciences, Centre for Integrative Ecology, 221 Burwood Highway, Burwood, VIC 3125

*Corresponding author email: m.burgass14@imperial.ac.uk

Abstract

Composite indicators (CIs) are increasingly used to measure and track environmental systems. However, they have faced criticism for not accounting for uncertainties and their often arbitrary nature. This review highlights methodological challenges and uncertainties involved in creating CIs and provides advice on how to improve future CI development in practice. Linguistic and epistemic uncertainties enter CIs at different stages of development and may be amplified or reduced based on subjective decisions during construction. Lack of transparency about why decisions were made can risk impeding proper review and iterative development. Research on uncertainty in CIs currently focuses on how different construction decisions affect the overall results and is explored using sensitivity and uncertainty analysis. Much less attention is given to uncertainties arising from the theoretical framework underpinning the CI, and the sub-indicator selection process. This often lacks systematic rigour, repeatability and clarity. We recommend use of systems modelling as well as systematic elicitation and engagement during CI development in order to address these issues. Composite indicators make trends in complex environmental systems accessible to wider stakeholder groups, including policy makers. Without proper discussion and exposure of uncertainty, however, they risk misleading their users through false certainty or misleading interpretations. This review offers guidance for future environmental CI construction and users of existing CIs, hence supporting their iterative development and effective use in policy-making.

Keywords: systems modelling, ocean health index, environmental performance index, climate change performance index, sustainable society index.

1.1 Introduction

Human activities have large impacts on natural systems (Halpern *et al.* 2008; Buma & Wessman 2011) that are likely to increase in future, given growing human population and demand on natural resources (Kraxner *et al.* 2013; McCauley *et al.* 2015). The resultant changes in natural systems have important consequences for biodiversity (Chapin *et al.* 2000), but also for people through our reliance on provision of ecosystem services for human well-being, health, livelihoods and survival (Costanza *et al.* 1997, 2014; Millennium Ecosystem Assessment 2005). Managing these complex interactions to ensure nature thrives and continues to provide benefits to people requires integrative and interdisciplinary approaches to management that emphasise the complexities of whole social-ecological systems (Folke *et al.* 2005). Effective ecosystem management requires measuring the status and trends of ecosystems to inform which management actions are likely to be effective and if these actions have had their intended effect (Jones *et al.* 2011). Measuring all aspects of complex systems is impossible due to the range of variables and processes present. Variables deemed to be characteristic of the wider system and which are simple enough to be easily measured are often employed as indicators, to act as simplified summaries of system condition and behaviour (Dale & Beyeler 2001).

Good indicator design has been widely discussed (Failing & Gregory 2003; Fulton, Smith & Punt 2005; Parr, Jongman & Kulvik 2010), with general agreement that indicators should: be cost effective; provide reliable information on status and trends; provide information at multiple extents and resolutions; allow frequent reporting; be meaningful to the public; and respond predictably to policy change (Jones *et al.* 2011). In practice, the EU's Streamlining European Biodiversity Indicators project used a stakeholder-based process to apply stringent criteria and reduce over 140 biodiversity indicators to a final 26, while the European Commission assesses indicators based on RACER guidelines; where they should be 'Relevant', 'Accepted', 'Credible', 'Easy to Evaluate' and 'Robust' (Best *et al.* 2008; Eea 2010). Indicators also provide a powerful tool for communicating with stakeholders about the status and trends of ecosystems, as well as helping identify or illuminate

linkages between environmental, human and economic subsystems (Jørgensen, Burkhard & Müller 2013). However, multi-dimensional processes (such as complex ecosystem dynamics) are notably difficult to track with individual indicators due to challenges in linking trends across dimensions (Munda 2005) and capturing interactions between and within sub-systems (Dale & Beyeler 2001). Multiple indicators are recommended to capture different aspects of the relevant systems (Fulton, Smith & Punt 2005), but without techniques to distil or summarise them, can be overwhelming in volume of information (Chatziparadeisis 2007). For example, the marine “Good Environmental Status” goal for EU countries contains 11 descriptors with 29 criteria and 62 individual indicators (European Commission 2010).

“Composite indicators” (CIs) offer a means of aggregating multiple indicators to track and communicate complex systems. CIs are a mathematical combination of a set of indicators that have no common meaningful unit of measurement. They are increasingly used for decision making in a range of sectors such as economics, business statistics, health and academic performance (Munda *et al.* 2009; Paruolo, Saisana & Saltelli 2013). In the environmental sector they are often used for global scale assessments (see Table 1) and to guide policy at local to regional scales (Mendoza & Prabhu 2003; Di Franco *et al.* 2009; Ochoa-Gaona *et al.* 2010). CIs enable direct comparison of disparate social and environmental variables and, due to their clear and unidimensional output, can also gain traction with policy-makers and the general public. Their increasing popularity is unlikely to slow; many have suggested that in order to communicate broad trends effectively and influence conservation policy, meaningful CIs will be required (Balmford *et al.* 2005; Mace & Baillie 2007). CIs are similar to mathematical or computational models in that they are simplified representations of reality, although whereas models are usually based upon scientific theory and detailed biological or physical dynamics, CIs are often simply an aggregation of variables considered relevant to a system or issue (Nardo *et al.* 2008). Modelling studies also typically address, and when possible quantify, inherent uncertainties that arise when simplifying real-world complexities (Kokko 2005). If CIs are to be used more, and more effectively, within conservation, methodological decisions made in their

84 construction, and the consequent uncertainties, should be clearly understood, described and, if
85 possible, represented or treated – just as with any other type of conservation modelling for decision-
86 making (e.g., Regan, *et al.* 2002).

87 Here, we explore the uncertainties that underlie environmental CI construction, with the aim of
88 putting recognition of uncertainty at the heart of CI construction and use. We develop a framework
89 to capture the full range of types and sources of uncertainty in a systematic fashion, using four
90 prominent environmental CIs as primary case studies (but also draw reference to others) and
91 suggest methods to navigate them. We first discuss the methods that are specific to each individual
92 stage and then address those that deal with multiple sources of uncertainty. Finally, we discuss ways
93 forward to improve the development and use of composite indicators in practice.

94 **Table 1** Examples of environmental composite indicators, chosen to display a range of different
95 construction techniques

Composite Indicator	Description	Construction
<p>Ocean Health Index (Halpern <i>et al.</i> 2012) www.oceanhealthindex.org</p>	<p>Evaluates the condition of marine ecosystems according to ten 'goals' of key benefits provided by the ocean. Measures sustainable provision of benefits and gives a score to each country.</p>	<p>Overall score is aggregated from ten equally weighted categories (known as 'goals', each comprised of many individual indicators. Subgroups measure biological, physical, social and economic aspects.</p>
<p>Environmental Performance Index (Hsu <i>et al.</i> 2014) www.epi.yale.edu/</p>	<p>Ranks how well countries perform on high priority environmental issues. Focuses on ranking individual countries.</p>	<p>Nested structure where overall score is aggregated from two equally weighted categories of environmental health and ecosystem vitality. Each category is made up of three and six subgroups respectively, which have between one and four sub-indicators each.</p>
<p>Climate Change Performance Index (Burck & Bals 2011) www.germanwatch.org/en/9472</p>	<p>Evaluates and compares the climate protection performance of countries that are, together, responsible for more than 90% of global energy-related</p>	<p>Nested structure where overall score is aggregated from three categories of emissions trend (50% weighting), emissions level (30% weighting) and climate policy (20% weighting). Each category is made up of between 4 and 9 subgroups which are informed by several</p>

	CO ₂ emissions.	sub-indicators each
<p>Sustainable Society Index</p> <p>(van de Kerk <i>et al.</i> 2014)</p> <p>http://www.ssfindex.com/</p>	<p>Evaluates countries based on their level of sustainability according to human, environmental and economic wellbeing.</p> <p>Focusses on ranking of countries.</p>	<p>Employs a nested structure with three categories; human wellbeing, economic wellbeing and environmental wellbeing.</p> <p>Categories are not aggregated to an overall score due to the correlation between human and environmental wellbeing. Each category is aggregated from 2-3 subgroups which consist of 2-4 sub-indicators in each.</p>

1.2 Characterising Uncertainty and Understanding Trade-Offs

In order to characterise uncertainties within CIs it is first important to understand how CIs are constructed. Although individual CIs differ, Figure 1 shows the construction stages for a typical environmental CI. The typical stages of CI construction are:

- **Theoretical framework** is the overarching conception of the CI and choice of subgroups and categories, which act as the key areas of the system that are of interest to be measured. The theoretical framework can impact technical choices such as weighting and normalization.
- **Data selection** involves construction and normalization of variables or sub-indicators as well as analysis and choice of underlying data.
- **Construction** of the CI includes approaches used for aggregation and weighting of sub-indicators, subgroups and categories.
- **Post-development communication** involves dissemination and communication of results.

Many different types and sources of uncertainty exist, emerging from one or more of these stages and requiring different approaches (Fig. 2). Epistemic uncertainty arises from a lack of knowledge of the dynamics and state of a system and includes uncertainty from limitations of measurement devices, insufficient data, extrapolations and interpolations, and variability over time or space. Linguistic uncertainty is a result of scientific vocabulary being under-specific, ambiguous, vague, context dependent, or exhibiting theoretical indeterminacies (Regan *et al.* (2002); Table S1). These uncertainties can be reduced or amplified based on decisions taken during construction (Table S2). Explorations of uncertainty in CIs have typically focussed on mathematical rules of construction, primarily related to statistical coherence and precision of the CI, and explored using mathematical techniques such as sensitivity and uncertainty analysis. However, these techniques are still not universally applied. Different construction methods are discussed and summarised in Nardo *et al.*'s (2008) guidance handbook for CI construction. They note the importance of construction decisions, especially an appropriate theoretical framework. Yet they offer little advice to constructors, stating

121 the soundness of the framework and fitness for purpose of the CI is best assessed by the peer
122 community. As such, the theoretical framework for CIs has received less attention in the literature
123 than other sources of uncertainty.

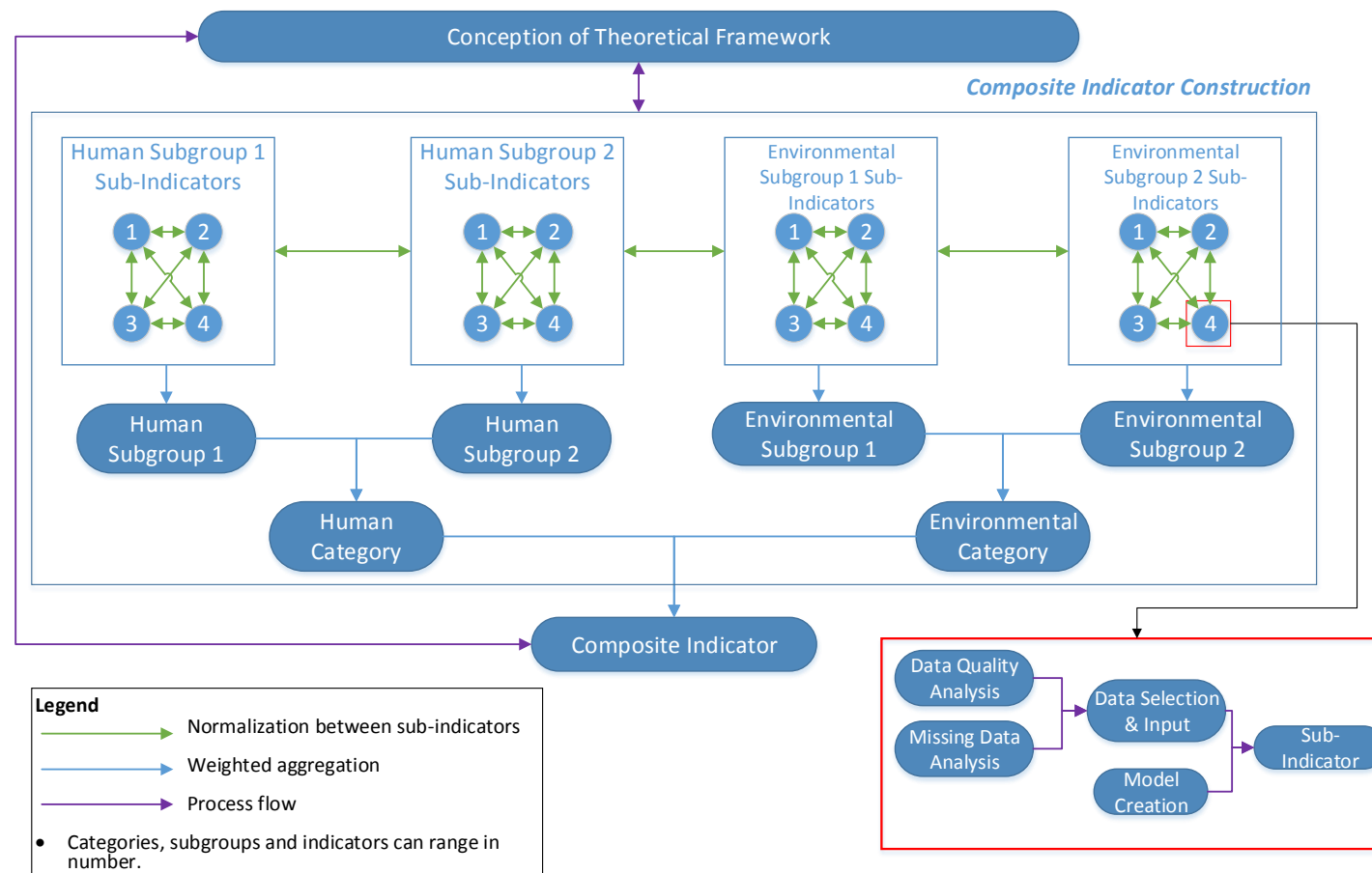


Figure 1. Typical construction of a composite indicator. The theoretical framework drives the mathematical construction with some to and fro likely as the index is pieced together. The red box indicates how a typical sub-indicator might be constructed – ideally the desired model would drive the sub-indicator creation, but in reality data availability is often the driving factor. The subgroup boxes show normalization and aggregation of sub-indicators, which in essence create sub-composite indicators.

1.2.1 Theoretical Framework

The theoretical framework is the starting point for all CIs, and comprises understanding and defining the system to be measured and the contributing categories and subgroups (Nardo *et al.* 2008; Mendola & Volo 2017). For example, the Environmental Performance Index (EPI) has two distinct categories of Ecosystem Vitality and Environmental Health, which contain various relevant subgroups such as 'Biodiversity and Habitat' and 'Air Pollution' respectively. These subgroups are in this case in essence also CIs, as they are aggregated from several sub-indicators. The final use of the CI is considered here as it affects later decisions on data and construction; for example Buckland *et al.* (2005) suggest criteria for how a biodiversity CI should perform in order to inform choices around construction. Despite numerous CIs existing, little guidance is found in the literature on how to successfully develop a theoretical framework for a CI. General indicator framework advice exists (e.g., using Driving Force-Pressure-State-Impact Response approaches (OECD 2003)), but the distinctive nature of each CI often requires creation of a unique framework that represents the conceptual thinking underlying the indicator. For example, the Ocean Health Index's (OHI) 'goals' (categories) were selected based on subject experts reviewing the literature on what the public expects from a healthy ocean. Likewise EPI scanned the literature and policy documents to split their index into two 'objectives' and smaller core categories (based on merging the Pressure-State-Response and Driving Force-State-Response frameworks (Hsu *et al.* 2013)). The Sustainable Society Index (SSI) used the Brundtland+ definition of sustainability to pick indicators, which were subsequently grouped into five categories (Van de Kerk & Manuel 2008).

A clear relationship between what a CI measures and its structure helps provide clarity to the user. However, many environmental concepts (e.g., sustainability) are not well defined, potentially introducing linguistic uncertainty into a CI from the start. This can make it difficult to identify appropriate categories and subgroups. Even widely used terms such as "biodiversity" have associated linguistic uncertainties, which means capturing and monitoring a vague concept can be

difficult and highly uncertain (Morar, Toadvine & Bohannon 2015). Lack of clarity causes model uncertainty, as the CI may not actually measure the construct to which it relates.

Once the system has been defined, effort usually focusses on next defining the categories and subgroups which form the structure of the CI. These are outward facing aspects that gain significant attention from constructors as they form the overall communication of the CI, and once selected often act as guidance in determining the indicators that fill them. This approach means that the paths taken to arrive at the final output are often unclear and linkages between subgroups are not fully understood. Relying heavily on a pre-determined theoretical framework rather than developing a conceptual model of the specific processes underlying a given CI runs the risk of arbitrary selection of sub-indicators, which may not be properly representative of the system (Nicholson *et al.* 2012).

1.3 Data

As in any quantitative analysis, it is important to understand uncertainties in the data that form the basis of a CI (Figs 1 & 2). Challenges here include choosing which data best represent components of the theoretical framework, what methods to use to understand uncertain data, and how to deal with the uncertainties.

1.3.1.1 Data quality

Composite indicators almost exclusively use existing data collected by various sources other than the CI creator, over different temporal and spatial scales (Kaufmann, Kraay & Mastruzzi 2011). Data are subject to varying levels of uncertainty depending on the credibility of the source, data collection methods, timing of sampling, measurement error, natural variation, and data interpretation. Data uncertainties are rarely adequately quantified (Munda *et al.* 2009), and most uncertainty in CIs is irreducible. At a minimum, a general audit of data quality should therefore be undertaken, with data assessed for relevance, accuracy, timeliness, accessibility, interpretability and coherence before being selected for inclusion (Nardo *et al.* 2008). The EPI and OHI do this by setting rough quality standards for data inclusion, however it is not revealed which data were discarded nor how robust

the included data are. Such a process is inherently subjective and users of CIs are not always able to discern where strong or weak data lie. The SSI acknowledges that “*the reliability of data remains a serious concern*”, but similarly does not indicate where its strongest or weakest data are found. No discussion of data quality was found for the Climate Change Performance Index (CCPI). Lack of clarity around data that are entered into, or excluded from, a CI might dissuade users or suggest that the CI is a risky basis for policy-making.

Pedigree matrices can be an effective way of assessing unquantifiable uncertainties in data (Van Der Sluijs *et al.* 2005). This technique involves using qualitative expert judgement to assess parameters through pedigree criteria, which are chosen as the most relevant and applicable criteria to assess parameter strength. Responses are then coded in the pedigree matrix (e.g. between 0 (weak) and 4 (strong)) to reduce arbitrariness and subjectivity. Experts are consulted individually so that consistency across scores indicates a common view of the underpinnings of the parameters, whereas disagreement reflects an ignorance of these underpinnings (Van der Sluijs *et al.* 2002). This approach thus helps to move consideration of uncertainties beyond those that are quantifiable to the large range of qualitative uncertainties. Assigning data quality scores, as done by the Living Planet Index, is one means of being transparent about data quality issues (Collen *et al.* 2009). Scores are assigned to data based on their source, methodology and whether a measure of variation was included. Scores can be used either to represent uncertainty, test how overall results differ using different quality data, or adjust weightings for lower quality data. The EPI used a similar method when it was known as the Environmental Sustainability Index, however it was terminated as the method was deemed too subjective and problematic when experts disagreed on assessment criteria for grading (Hsu *et al.* 2013). If data quality is not deemed sufficient then reporting units can be excluded. The EPI did this in their 2012 assessment by removing North Korea, as several anomalous results raised serious questions over data quality.

1.3.1.2 *Treating missing data*

Data that underpin CIs inevitably contain gaps, requiring decisions about methods used to address these gaps (See Table S3). Although modern imputation techniques (such as multiple imputation and maximum likelihood estimation) exist, the use of such techniques may be constrained by time, budget or expertise of the team; these trade-offs need to be reported and justifications given on why certain methods were or were not used. Understanding and displaying where missing data exist is important, as some reporting units (e.g., countries) may be composed of significant amounts of imputed data, which may slip through unnoticed if not transparently logged.

The EPI attempted to collect data for 232 countries but calculations were only performed for 178 due to missing or incomplete data. Conversely, the OHI provides a score for all 221 Exclusive Economic Zones and 15 high seas regions. The OHI fills gaps using a hierarchical decision tree with four different methods: temporal, using data from previous years; alternate datasets used as proxies; spatial, using averages from nearby regions; special rules applicable to particular instances (Halpern *et al.* 2015). The SSI has seemingly high data coverage, with less than 10% gaps (Saisana & Philippas 2012), whereas the CCPI offers no discussion of missing data. However, none of the four case study CIs offers an easy insight into which particular data have been imputed, although the OHI has implemented a gap-filling tracking methodology that will be incorporated and presented in global 2016 scores (Frazier *et al.*, in review). CI documentation should be open about which data have been filled, and which reporting regions have been deleted, so the subsequent uncertainty can be properly recognised (Frazier, Longo & Halpern 2016). Regions that contain significant amounts of missing data should be highlighted or removed from the assessment. Analysis of how data gaps affect the overall outcome of the CI is important to guide targeted data collection and inform approaches to gap filling.

1.3.1.3 *Data selection and sub-indicator construction*

Data selection and sub-indicator construction are intrinsically linked. Ideally, sub-indicators would be selected systematically based on their relevance to what is being measured (Nardo *et al.* 2008;

Riedler *et al.* 2015). In reality, the data required to construct an ideal sub-indicator might not be available, be of questionable quality, or have substantial gaps, meaning a trade-off is required. This may involve discarding a preferred indicator in favour of one which is supported by better data, risking introducing severe model error into the CI, or including weaker data in a preferred sub-indicator, meaning it is likely to be less robust.

Typically, individual indicators are chosen using pre-determined selection criteria. Dale and Beyeler (2001) suggest that indicators should be: easily measured; sensitive to stress; respond to stress predictably; anticipatory; predict changes that can be averted; integrative; and have low variability in response to extraneous influences. However, the criteria stated as being important vary widely between indicators. Our case study CIs show some consistencies in selection criteria, but also differences (Table 2). Lists of selection criteria can enable a more consistent set of indicators, but they do not give much insight into the actual selection process because they do not give information pertaining to why a particular individual indicator or indicator group was chosen and others were discarded, or any relationships between the selected and discarded indicators (Niemeijer & de Groot 2008a). The criteria in Table 2 focus mainly on data quality and availability rather than understanding the relevance of indicators to the dynamics of the system, or how indicators might behave as per Dale & Beyeler's (2001) list above. Without knowledge of the system's dynamics, it is unclear how the sub-indicators are linked and if they accurately represent the system. Correctly designing sub-indicators is important as they are the basis of a CI; uncertainties here will propagate through the CI, with a 'garbage in-garbage out' logic (Nardo *et al.* 2008).

Table 2 Selection criteria for data and sub-indicator construction used in the case study indices

Environmental Performance Index (Hsu <i>et al.</i> 2013)	Ocean Health Index (Ocean Health Index 2015)	Sustainable Society Index (Van de Kerk & Manuel 2008)	Climate Change Performance Index
--	--	---	---

Data must be relevant to what is to be measured	Data must be relevant to what is to be measured	Data must be relevant to what is to be measured	Not located
Indicator provides empirical data on ambient conditions or on-the-ground results for the issue of concern.	Must be able to be scaled to a meaningful reference point	Data must be measurable	
Data must have established scientific methodology and based on peer review or institutions charged with data collection	Data must be freely accessible	Data must be from public sources, scientific or institutional	
Must have adequate global and temporal coverage	Must have adequate global and temporal coverage	Must be available for all countries (or all but smallest countries)	
Data represent the best measure available.	Data quality should be considered	Data must be reliable	
Data have been consistently measured across time	Data must be recent and ideally regularly updated	Data must be recent and regularly updated	

		Independent from other indicators with no overlap	
--	--	---	--

1.3.1.4 Data Normalization

Data that are used in sub-indicators are often in many different formats and therefore must be normalized to the same scale for aggregation (Jacobs *et al.* 2004). This allows comparison of disparate indicators within a single framework. Decisions also have to be made regarding outliers, which can cause problems by becoming unintended benchmarks, skewing data and biasing statistical approaches to weighting. Popular normalization techniques include (Saisana & Saltelli 2011):

- **Ranking** – Simply ranks units in order and therefore does not preserve specific information. Final output is rank only.
- **Standardization (or z-scores)** – converts indicators to a continuous variable with a mean of zero and standard deviation of one. Assumes normality, meaning outliers can have a large effect, which might not be desirable.
- **Min-Max** – normalizes indicators within a given range (e.g. 0-1) by subtracting the minimum value and dividing by the range. Outliers can distort the CI for similar reasons.
- **Distance to target** – Normalizes indicators by dividing the unit's value by a reference target. Can be sensitive to outliers when the best performing unit is used as a target.

The meaning of the results of a CI could be affected by which technique is chosen and should therefore be considered during the theoretical framework stage. Distance to target is a popular method as it allows for the inclusion of political goals, for example the EPI uses the Convention of Biological Diversity's 17% target of terrestrial and inland areas under protection as its critical habitat protection indicator. Normalizing by such a political goal provides a clear benchmark that is relevant and can be easily communicated and frameworks exist for robust selection of quantitative

management targets (Samhuri *et al.* 2012). If using the other more arbitrary methods without testing different techniques, subjective judgement error will mean the outcome of the CI is affected an unknown amount by the choice of normalization.

1.3.2 Construction

This phase determines how normalized sub-indicator scores are aggregated and weighted. Although weighting should be considered during the theoretical framework stage, it is discussed separately here, as it can be an iterative process.

1.3.2.1 Weighting

Weights are often used as measures of perceived importance of the subgroup to the system. For example, the CCPI rates its categories of climate policy, emission trends and emission levels at 20%, 50% and 30% respectively (Burck, J. & Bals 2011). However, lack of knowledge about subgroup importance, or unwillingness to prioritise one area above another, frequently results in equal weight being allocated, which although seen as neutral, is still a weighting decision. The SSI uses equal weights due to a lack of scientific basis for the attribution of weights (Van de Kerk & Manuel 2008). Likewise the OHI uses ten equally weighted 'goals', as the literature does not distinguish which factors are most important for a healthy ocean (Halpern *et al.* 2012), although the OHI explicitly includes a goal weighting term and encourages development and use of weights for localised assessments. The EPI generally sets weights based on the quality of data and relevance of the indicator to the issue it is measuring. Less robust or relevant data are therefore given a lower weighting (Hsu *et al.* 2014). However, despite poor data, the indicator concerned may be key to describing system dynamics; giving it a low weight may therefore reduce the meaningfulness of the CI.

Despite weights often being assigned to different subgroups and/or categories as importance coefficients, variation and correlation in data mean assigned or desired weights might not act as intended. Weighting may have to be an iterative process in order to achieve a desired weighting structure. For example, the EPI performed a sensitivity analysis in 2012 and found that although

environmental health and ecosystem vitality had been given equal weights, the greater variation in environmental health scores meant that countries that performed better in environmental health were more likely to perform better in the overall EPI. The weighting was therefore not actually equal and was subsequently adjusted to account for this phenomenon (Hsu *et al.* 2013). However, adjusting weighting to account for variations can be problematic as the assigned weighting is therefore not reflective of the importance, as ecosystem vitality has a larger weighting on face value. Weightings and importance could therefore be misinterpreted by users and therefore should be properly recorded and communicated. Likewise, without such analysis and understanding, weights that are assigned as importance measures may not actually perform as desired in the CI.

1.3.2.2 Aggregation

Two widely used options for aggregation have gained attention in the CI literature; linear and geometric aggregation. Linear aggregation involves a summation of (weighted) sub-indicator scores (usually averaged around a mean), while geometric aggregation involves aggregation by the geometric mean (i.e., using the product of values).

The choice of aggregation method can be a source of model error and subjective judgement uncertainty as it can fundamentally alter how the CI performs. The SSI aggregates through a geometric mean (van de Kerk *et al.* 2014) and the Human Development Index switched from linear to geometric aggregation in 2010 (Klugman 2010), while the OHI, EPI and CCPI all use linear aggregation. Key considerations are:

- **Compensability** –Linear aggregation allows complete compensability and geometric partial compensability, which means good performance in one indicator can offset poor performance in another. For example, consider a hypothetical subgroup of a biodiversity index, made up of sub-indicators for protected areas, endangered species and critical habitats (equally weighted). Two countries, A and B, score 9.0, 7.6, 0.4 and 6.8, 6.0, 4.2 respectively. If linear aggregation were used then both countries perform similarly, with a

resulting score of roughly 5.6. However, when using the geometric mean, A's overall score is reduced from 5.6 to 3.0, while B's score remains at 5.6. Should no amount of compensability be acceptable and for weights to be truly interpreted as importance coefficients, non-compensatory aggregation methods should be used (Munda 2008; Munda & Nardo 2009; Munda *et al.* 2009; Cinelli, Coles & Kirwan 2014). However, this method has seen limited use and therefore there has been encouragement towards multi-criteria approaches to assess robustness (Munda *et al.* 2009). Given that many CIs now provide transparency to indicator level, there is less importance placed on the overall score, which can be sensitive to aggregation.

- **Improvement of scores** –An improvement in A's critical habitat score from 0.4 to 1.4 in our hypothetical index sees an overall improvement under the geometric mean from 3.0 to 4.6, and 5.6 to 6.0 under the arithmetic mean. The larger jump in the geometric mean could encourage focus on lower performing metrics, which may be beneficial from a policy perspective, but could also dissuade action on higher-performing metrics even if those actions would be beneficial.
- **Communication** – linear aggregation is more straightforward for communication and engagement as users can clearly trace scores from the bottom level to the top. It also rewards proportionally to weights, whereas geometric aggregation rewards units with higher scores.

Aggregation to an overall single value is appealing for media traction and communication but may not always be appropriate. For example, the SSI chooses not to aggregate to a single figure based on the strong negative correlation of human wellbeing to environmental wellbeing and thus gives results for three separate composite indicators (the third being economic wellbeing; Saisana & Philippas 2012).

1.3.3 Uncertainty in Post-Development

Communicating a composite indicator can be a complicated undertaking, which can decrease linguistic uncertainty if well done, or increase it by not being transparent and using confusing or vague language. Poor communication may mean a CI is seen as ineffectual or, worse, used improperly. Targeting a technical audience may mean the CI can be critiqued and iteratively improved, but it may not gain the desired public or political traction that a more populist CI would. Likewise, uncertainties can be openly presented or not discussed, but reaffirming complexity within a highly simplified measure is challenging for communication.

A key challenge is whether to focus on the final single numeric output or delve deeper into the CI. By aggregating to a single number, composite indicators can potentially send over-simplistic messages. The same overall score can be achieved in many different ways; one way to overcome this is to give attention to the categories, subgroups and potentially even sub-indicators. All the case study CIs presented here give more information than just ranking countries/overall scores. Detail is given on how scores are achieved, which sub-indicators make up the subgroups and how these have changed over time. The SSI offers downloads of the normalized scores, while the EPI and OHI offer normalized raw data and scores. CCPI offers a qualitative performance review for each country by category but does not discuss sub-indicators or provide data. In all the case studies it is also unclear if or how sub-indicators and subgroups interact, e.g. where an increase in one might cause a decrease in another, although the OHI's inclusion of pressures derived from each category (or 'goal') helps users understand potential trade-offs. Understanding linkages and interactions is potentially critical from a policy point of view, as it is unclear how attempts to alter the status of one particular sub-indicator or subgroup will affect the others. A systems modelling approach could give a starting point for decision makers trying to understand how interactions occur and what their consequences might be.

1.4 Navigating Uncertainty

The range of uncertainties present in CI construction means there is not a single way to treat or represent them all (Table 3). Single issue solutions have been addressed in the sections above, but here we lay out potential approaches to improving CI construction that address multiple uncertainties simultaneously.

Table 3: Range of uncertainties in composite indicators and how to treat them

Source of Uncertainty	Issue	Reason for Issue	Potential solution
Theoretical Framework	Is theoretical framework representative of the system?	<ul style="list-style-type: none"> No systematic process Subjective Lack of transparency and repeatability 	<ul style="list-style-type: none"> Systems Modelling Systematic expert judgement/stakeholder engagement Transparency and iterative improvement
Data	Accuracy of data	<ul style="list-style-type: none"> Data quality rarely assessed and therefore not really considered 	<ul style="list-style-type: none"> Data scoring/pedigree matrices Systematic expert judgement/stakeholder engagement Uncertainty analysis
	Amount of missing data	<ul style="list-style-type: none"> Unclear where data gaps are and number of them. Gap filling methods are subjective 	<ul style="list-style-type: none"> Transparency and iterative improvement Uncertainty/sensitivity analysis Advanced monte-carlo gap-filling methods
	Is indicator an accurate and desired representation of the system	<ul style="list-style-type: none"> Led by data availability, stakeholder or constructor values therefore subjective. Unclear how indicators relate to system 	<ul style="list-style-type: none"> Systems modelling
	Representation vs quality	<ul style="list-style-type: none"> Trade-off between data accuracy and missing data v how well it represents the system 	<ul style="list-style-type: none"> Transparency and iterative improvement Systematic expert judgement/stakeholder engagement

Source of Uncertainty	Issue	Reason for Issue	Potential solution
		<ul style="list-style-type: none"> • Subjective 	
Data Normalization	Different methods	<ul style="list-style-type: none"> • Subjective 	<ul style="list-style-type: none"> • Transparency and iterative improvement • Uncertainty/sensitivity analysis
Weighting	Arbitrary weighting	<ul style="list-style-type: none"> • Unclear how weights were assigned. • "Neutral" weighting still a weighting decision • Subjective 	<ul style="list-style-type: none"> • Systems modelling • Systematic expert judgement/stakeholder engagement
	Implicit weights may be different to assigned weights	<ul style="list-style-type: none"> • Statistical properties mean assigned weights don't always work as intended 	<ul style="list-style-type: none"> • Correlation analysis • Uncertainty/sensitivity analysis
Aggregation	Different methods	<ul style="list-style-type: none"> • Subjective 	<ul style="list-style-type: none"> • Transparency and iterative improvement • Uncertainty/sensitivity analysis
Communication	Different interested parties	<ul style="list-style-type: none"> • How to communicate to public/policy makers/scientists 	<ul style="list-style-type: none"> • Transparency and iterative improvement • Multi-layered approach of engagement/analysis

377

378 1.4.1 Systems Modelling

379 Without a proper understanding of the system and how individual indicators represent its dynamics,
380 CIs risk severe structural uncertainty and improperly informing management decisions. There is a
381 lack of guidance in the literature on how to construct CI theoretical frameworks and as such they
382 tend to follow general approaches such as Driving Force-Pressure-State-Response framework (OECD
383 2003) or be purpose-built, often informed by literature, experts or stakeholders, focussing on
384 perceived key areas of the system which form the sub-categories (Nardo *et al.* 2008). However, the
385 processes involved in this construction often go undocumented, meaning the resultant structure

could be seen as arbitrary and techniques are impossible to replicate. Sub-categories are then populated by relevant sub-indicators. A key feature of any environmental indicator, including CIs, is that it is able to reflect changes in a system. It is therefore crucial that the theoretical framework and selected sub-indicators accurately represent the system. Without first understanding system dynamics, and testing the behaviour of the sub-indicators as the system changes, it is impossible to know if the chosen CI sub-indicators do accurately reflect the system. This means it is not clear whether changes in the sub-indicators are reflecting real system change, or whether critical data gaps exist which could impact on the ability of the CI to track system change.

Systems modelling is an effective approach to representing understanding and thus provides a systematic, transparent and repeatable way to aid sub-indicator selection and theoretical framework development. This approach defines variables or processes which are most important to a system's dynamics, and their interactions, thereby mapping the system and the linkages within it (Niemeijer & de Groot 2008b). Using quantitative modelling approaches to select indicators is well explored in environmental science and is seen as one of the most effective methods of understanding how indicators respond to change. It has been suggested or used as an approach for selecting indicator sets in forest management (Brang *et al.* 2002; Mäkelä *et al.* 2012) and commonly used for testing and refining indicators in fisheries (Fulton, Smith & Punt 2005; Branch *et al.* 2010). Complex social-ecological models such as 'Atlantis' (Fulton *et al.* 2011a) can provide a basis for understanding systems and picking out key indicators to be included in a CI. Once sub-indicators have been selected, a systems model could be altered based on policy options to test how CIs react to underlying changes in the data; CIs are then able to be validated and act as a decision making aid (Nicholson *et al.* 2012). However, given CIs often aim to represent highly complex concepts or systems, quantitative models might not be available, or qualitative conceptual and/or expert-based models may be more appropriate. Qualitative modelling approaches, which map out a system diagram, have proved useful frameworks, for example in indicator selection for sustainable tourism (Margoluis *et al.* 2009;), fisheries (Vugteveen *et al.* 2015), land use change (Benini *et al.* 2010; Van

Oudenhoven *et al.* 2012), land degradation (Gisladdottir & Stocking 2005; Agyemang, McDonald & Carver 2007), urbanization (Jago-on *et al.* 2009) and water management (Chung & Lee 2009). Conceptual modelling like this allows stakeholders to come together and help develop the model, which enables sharing and inclusion of perspectives and values. A conceptual systems model can be used to select indicators by first defining clearly a concrete question to be answered. In the case of CIs, this is likely to be broad (i.e. measuring sustainability of countries) but could also be very specific (i.e. the effect of increasing protected areas on biodiversity). Broader questions will naturally require indicators that span multiple issue areas and will not be as fine-tuned as those that are selected to help track more specific questions; indeed, specific questions may require more detailed construction of the systems model. In order to select indicators, nodes must first be identified: Root nodes which have many outgoing arcs typically provide information on sources of issues; central nodes with many incoming and outgoing arcs are usually important for the most general of indicators as they provide information on many issues; end of root nodes with many incoming arcs allow the gauging of multiple issues at once. These nodes can then provide guidance on the types of indicators that are required to answer the question. Niemeijer & de Groot (2008a) provide a detailed example using this approach to select indicators for ecological impact of nitrogen fertilization on surface waters. Such an approach could then allow CIs as a means to track specific issues, by using the systems model to pull out sub-indicators relevant to a specific issue. A systems model means indicators are not selected arbitrarily, have a wider and understandable function and can be interpreted effectively, minimising the risk of incorrect analysis.

A system model could also inform weightings by pinpointing those system areas that are more central or peripheral, meaning final weights could be selected by an informed systems approach, thereby constraining uncertainty. Selected sub-indicators found to be suboptimal due to data issues could be highlighted, aiding in representing uncertainty as well as showing key data gaps that should be filled in order to have a more robust understanding of the system. Starting with a systems

approach could help balance the current top-down framework creation and reduce arbitrary indicator selection.

Of course in a CI, subgroups are often based on stakeholder values or perceptions, and this is recognised as an important facet of CIs. While stakeholder engagement is always important, it is particularly critical where highly complex models include large amounts of uncertainty or system dynamics are unclear. We therefore encourage an iterative back-and-forth approach between the modellers who can point to expected important variables based on their system models and stakeholders who can likewise do the same based on societal values.

1.4.2 Systematic expert judgement and stakeholder engagement

Expert-led or stakeholder-participation approaches are often used for CI construction, but with little information supplied on how or why decisions were made. This lack of transparency means the results of engagement are often unknown and methods unrepeatably. Structured elicitation of knowledge from experts is well explored and can be a powerful tool if used correctly. Key lessons include; elicit knowledge from groups rather than individuals, carefully choose members, strive for group heterogeneity, calibrate and weight experts, train experts, and give feedback (Burgman *et al.* 2011a; b; Sutherland & Burgman 2015). Using these techniques opens up many options throughout constructing a CI; such as providing judgement on theoretical framework and sub-indicator selection, estimating data accuracy by providing bounds or data scoring, providing guidance on weighting and help with communication and analysis of results. McBride *et al.* (2012) demonstrate how such techniques were applied to the IUCN Red Listing of Australian birds in order to reduce bias and error amongst experts. Furthermore, their elicitation was carried out online, showing that lengthy and costly workshops do not always have to be undertaken.

Stakeholder engagement is important for ensuring that the CI is useful for the intended audiences, and should be done at an early stage and throughout CI development and implementation. The more diverse the perspectives involved in developing the conceptual framework and exploring sub-

indicator selection, the more likely it is to represent meaningful reality for the endusers (Burgman *et al.* 2011b; Fulton *et al.* 2011b). This engagement can be more difficult when large numbers of different types of stakeholder are involved. Even at the global scale, representatives of particular groups can be consulted. Stakeholder input can be particularly vital, however, when CIs are used at more regional or local scales. Systematic, recordable techniques are useful in order to document engagement outcomes. Halpern *et al.* (2013) used such methods (based on random utility theory and analytical deliberation) to elicit stakeholder preferences for indicator weighting in the OHI in a regional assessment of the California Current. However, such a task was considered by Halpern *et al.* (2013) to be unworkable on a global level as the range of preferences would be so vast. Indicators tend to be less successfully utilised when they are purely scientific; involving stakeholders and leaving room for negotiation in CI construction can be highly beneficial in the messy situations CIs tend to be needed for (Turnhout, Hisschemöller & Eijsackers 2007).

1.4.3 Statistical coherence and robustness

The combined use of uncertainty and sensitivity analysis for CIs is well explored within the literature but is still not universally applied (Saisana *et al.* 2005; Munda *et al.* 2009; Paruolo *et al.* 2013). Uncertainty analysis focuses on how uncertainty in inputs, such as poor data or subjective construction choices about aspects like the weighting scheme, propagates through the CI to affect outputs. Results are usually represented by uncertainty bounds around output values. Sensitivity analysis looks at how each individual source of uncertainty contributes to this variation, and has been used to investigate the robustness of several CIs, including the SSI and EPI. The latest report on the EPI (Athanasoglou, Weziak-Bialowolska & Saisana 2014) found that three of the nine issue areas did not contribute significantly to EPI ranking, suggesting that changes to these indicators should be made. Aggregation function choice was found to account for 94% of sample variance, whilst choice of weighting for the objectives only accounted for 4%. This suggests that further discussions surrounding methodological choice should focus on aggregation method rather than weighting. Ninety percent of SSI countries shifted less than ± 1 position with respect to the simulated median,

suggesting that the 2012 SSI is not unduly driven by methodological assumptions (Saisana & Philippas 2012). The OHI and CCPI have yet to undertake uncertainty/sensitivity analyses, although they are planned for the OHI.

The combined use of uncertainty and sensitivity analysis provides an evaluation of confidence in the mathematical properties of the CI, assessing uncertainties associated with the construction process. This can help with many of the methodological decisions we highlight. However, uncertainty and sensitivity analyses only deal with a limited part of the uncertainty surrounding CIs. For example, they can show how rankings change based on the methodological choices made, or if any bias is present. They cannot, however, detect whether indicators are measuring what they intend to, or whether the CI represents overall system dynamics.

1.4.4 Communication and transparency

A key benefit of CIs is their ability to communicate issues clearly, to a wide audience, by aggregating sub-indicators to a single figure. However, many CIs act as more than a communication tool and are used for tracking trends and decision making. This simplistic output of a single number or score then becomes problematic, as some parties will not believe in complete aggregation as an approach to summarising complex and interacting systems. The deeper a CI can be explored, the more useful it will be for technical audiences. Therefore, complete transparency to sub-indicator level, including documentation of the issues covered here is desirable, which is not usually the case (Freudenberg 2003; Munda 2005; Böhringer & Jochem 2007; Singh *et al.* 2009). Users such as scientists and management authorities may require different or more detailed information, particularly related to how sub-indicators were selected, how they interact, where data gaps are found and why methodological decisions were made. This information is necessary for management, can ensure CIs are quality checked appropriately through peer review, and supports their iterative development. Transparency should involve acknowledging the methodological decisions made during construction and the rationale for employing certain methods over others.

All CIs in Table 1 provide detailed documentation on their methods and results in an accessible format through their website. Currently, however, none of them discuss uncertainty beyond the uncertainty and sensitivity analysis performed for the EPI and SSI. This limits constructive criticism and improvement. Given the variation in sources of uncertainty and methods of treating them, communicating these effectively becomes problematic. Those who solely use CIs for their most simplistic numeric output are unlikely to be interested in the technical uncertainties. Attempting to communicate these uncertainties might dilute the effectiveness of CIs themselves. Therefore, although uncertainty considerations are critical, we believe they should be reserved for more technical audiences.

1.5 Future Considerations

Environmental CIs are increasingly produced, but have often been criticised for their lack of acknowledgement and treatment of uncertainty (Böhringer & Jochem 2007; Jørgensen, Burkhard & Müller 2013; Giampietro & Saltelli 2014). We have provided a comprehensive assessment of the sources of uncertainties and methods to treat, represent or reduce them. Articulation and treatment of uncertainty within CIs is underdeveloped; how it is accounted for will depend on individual CIs' aims and audiences. Transparency and acceptance of uncertainty may be sufficient in some CIs, but others may require reworking of sub-indicators and construction methods. We hope this framework can act as a basis for considering uncertainty at each stage and recording not only why certain decisions were taken, but why others were not. Choices are unavoidable during CI construction but should always be acknowledged.

The focus to date on mathematical techniques for dealing with uncertainties in CIs has meant the role of the theoretical framework and sub-indicator selection has received less attention. Approaches such as systems modelling are fundamental for proper selection and grouping of indicators and their interactions, if a CI is properly to represent reality. This would also increase their usefulness in a policy setting, by testing policy scenarios and selecting specific sub-indicators to help answer more explicit questions. Importantly, development of a CI should be an iterative process.

Expressing the location and importance of different types of uncertainty can then be a catalyst for new data collection or conceptual development that is targeted at reducing the most influential uncertainties. Communicating these uncertainties may need to be done separately from the public-facing communication of the main CI, so as to not dilute its impact. However, there is much scope for novel, simple and clear techniques for communicating uncertainties effectively to technical audiences who require such information. This will build trust in CIs and thereby enhance their ability to support decision-making.

1.6 Acknowledgements

Funding: This work was supported by a Kristian Gerhard Jebsen Scholarship awarded through the Grantham Institute for Climate Change and the Environment, Imperial College London.

The authors would like to thank the reviewers for their comments which greatly improved this manuscript.

1.7 References

- Agyemang, I., McDonald, A. & Carver, S. (2007) Application of the DPSIR framework to environmental degradation assessment in northern Ghana. *Natural Resources Forum*, **31**, 212–225.
- Athanasoglou, S., Weziak-Bialowolska, D. & Saisana, M. (2014) *Environmental Performance Index 2014 JRC Analysis and Recommendations - JRC Science and Policy Reports*.
- Balmford, A., Bennun, L. & Brink, B. Ten. (2005) The Convention on Biological Diversity's 2010 target. *Science*, **3**, 212–213.
- Benini, L., Bandini, V., Marazza, D. & Contin, A. (2010) Assessment of land use changes through an indicator-based approach: A case study from the Lamone river basin in Northern Italy. *Ecological Indicators*, **10**, 4–14.
- Best, A., Blobel, D., Cavalieri, S., Giljum, S., Hammer, M., Lutter, S., Simmons, C. & Lewis, K. (2008) Potential of the Ecological Footprint for monitoring environmental impacts from natural resource use. *Sustainable Development*, 312.
- Böhringer, C. & Jochem, P.E.P. (2007) Measuring the immeasurable - A survey of sustainability indices. *Ecological Economics*, **63**, 1–8.
- Branch, T. a, Watson, R., Fulton, E. a, Jennings, S., McGilliard, C.R., Pablico, G.T., Ricard, D. & Tracey, S.R. (2010) The trophic fingerprint of marine fisheries. *Nature*, **468**, 431–435.
- Brang, P., Courbaud, B., Fischer, A., Kissling-Naf, I., Pettenella, D., Schonenberger, W., Spork, J. & Grimm, V. (2002) Developing indicators for the sustainable management of mountain forests using a modelling approach. *Forest Policy and Economics*, **4**, 113–123.
- Buckland, S.T., Magurran, a E., Green, R.E. & Fewster, R.M. (2005) Monitoring change in biodiversity through composite indices. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **360**, 243–254.
- Buma, B. & Wessman, C. a. (2011) Disturbance interactions can impact resilience mechanisms of forests. *Ecosphere*, **2**, art64.

577 Burck, J. & Bals, C. (2011) *The Climate Change Performance Index Background and Methodology*.

578 Burgman, M., Carr, A., Godden, L., Gregory, R., McBride, M., Flander, L. & Maguire, L. (2011a)

579 Redefining expertise and improving ecological judgment. *Conservation Letters*, **4**, 81–87.

580 Burgman, M.A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff,

581 L. & Twardy, C. (2011b) Expert status and performance. *PLoS ONE*, **6**, 1–7.

582 Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U.,

583 Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C. & Díaz, S. (2000) Consequences of changing

584 biodiversity. *Nature*, **405**, 234–42.

585 Chatziparadeisis, A. (2007) Policy Impact of Composite Indicators. *Statistisches Bundesamt*, 1–5.

586 Chung, E.S. & Lee, K.S. (2009) Prioritization of water management for sustainability using hydrologic

587 simulation model and multicriteria decision making techniques. *Journal of Environmental*

588 *Management*, **90**, 1502–1511.

589 Cinelli, M., Coles, S.R. & Kirwan, K. (2014) Analysis of the potentials of multi criteria decision analysis

590 methods to conduct sustainability assessment. *Ecological Indicators*, **46**, 138–148.

591 Collen, B., Loh, J., Whitmee, S., Mcrae, L., Amin, R. & Baillie, J.E.M. (2009) Monitoring Change in

592 Vertebrate Abundance: the Living Planet Index. *Conservation Biology*, **23**, 317–327.

593 Costanza, R., d’Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S.,

594 O’Neill, R. V., Paruelo, J., Raskin, R.G., Sutton, P. & van den Belt, M. (1997) The value of the

595 world’s ecosystem services and natural capital. *Nature*, **387**, 253–260.

596 Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S.J., Kubiszewski, I., Farber, S. &

597 Turner, R.K. (2014) Changes in the global value of ecosystem services. *Global Environmental*

598 *Change*, **26**, 152–158.

599 Dale, V.H. & Beyeler, S.C. (2001) Challenges in the development and use of ecological indicators.

600 *Ecological Indicators*, **1**, 3–10.

601 EEA (2010) *Assessing Biodiversity in Europe — the 2010 Report*.

602 European Comission. (2010) 2010/477/EU: Commission decision of 1 September 2010 on criteria and

603 methodological standards on good environmental status of marine waters. (notified under
604 document C(2010) 5956). *Official Journal of the European Commission*.

605 Failing, L. & Gregory, R. (2003) Ten common mistakes in designing biodiversity indicators for forest
606 policy. *Journal of environmental management*, **68**, 121–132.

607 Folke, C., Hahn, T., Olsson, P. & Norberg, J. (2005) Adaptive Governance of Social-Ecological Systems.
608 *Annual Review of Environment and Resources*, **30**, 441–473.

609 Di Franco, A., Marchini, A., Baiata, P., Milazzo, M. & Chemello, R. (2009) Developing a scuba trail
610 vulnerability index (STVI): A case study from a Mediterranean MPA. *Biodiversity and
611 Conservation*, **18**, 1201–1217.

612 Frazier, M., Longo, C. & Halpern, B.S. (2016) Mapping uncertainty due to missing data in the global
613 ocean health index. *PLoS ONE*, **11**.

614 Freudenberg, M. (2003) Composite indicators of country performance: a critical assessment. *OECD
615 Science, Technology and Industry Working Papers*, **16**, 35.

616 Fulton, E.A., Link, J.S., Kaplan, I.C., Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton,
617 R., Gamble, R.J., Smith, A.D.M. & Smith, D.C. (2011a) Lessons in modelling and management of
618 marine ecosystems: the Atlantis experience. *Fish and Fisheries*, **12**, 171–188.

619 Fulton, E. a., Smith, A.D.M. & Punt, A.E. (2005) Which ecological indicators can robustly detect
620 effects of fishing? *ICES Journal of Marine Science*, **62**, 540–551.

621 Fulton, E.A., Smith, A.D.M., Smith, D.C. & Van Putten, I.E. (2011b) Human behaviour: The key source
622 of uncertainty in fisheries management. *Fish and Fisheries*, **12**, 2–17.

623 Giampietro, M. & Saltelli, A. (2014) Footprints to nowhere. *Ecological Indicators*, **46**, 610–621.

624 Gisladdottir, G. & Stocking, M. (2005) Land degradation control and its global environmental benefits.
625 *Land Degradation & Development*, **16**, 99–112.

626 Halpern, B.S., Longo, C., Hardy, D., McLeod, K.L., Samhouri, J.F., Katona, S.K., Kleisner, K., Lester, S.E.,
627 O’Leary, J., Ranelletti, M., Rosenberg, A. a, Scarborough, C., Selig, E.R., Best, B.D., Brumbaugh,
628 D.R., Chapin, F.S., Crowder, L.B., Daly, K.L., Doney, S.C., Elfes, C., Fogarty, M.J., Gaines, S.D.,

629 Jacobsen, K.I., Karrer, L.B., Leslie, H.M., Neeley, E., Pauly, D., Polasky, S., Ris, B., St Martin, K.,
630 Stone, G.S., Sumaila, U.R. & Zeller, D. (2012) An index to assess the health and benefits of the
631 global ocean. *Nature*, **488**, 615–20.

632 Halpern, B.S., Longo, C., Lowndes, J.S.S., Best, B.D., Frazier, M., Katona, S.K., Kleisner, K.M.,
633 Rosenberg, A. a., Scarborough, C. & Selig, E.R. (2015) Patterns and Emerging Trends in Global
634 Ocean Health. *Plos One*, **10**, e0117863.

635 Halpern, B.S., Longo, C., McLeod, K.L., Cooke, R., Fischhoff, B., Samhour, J.F. & Scarborough, C.
636 (2013) Elicited preferences for components of ocean health in the California Current. *Marine*
637 *Policy*, **42**, 68–73.

638 Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C. V., Micheli, F., D'Agrosa, C., Fruno, J.F., Casey,
639 K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T.,
640 Selig, E.R., Spalding, M., Steneck, R. & Watson, R. (2008) A Global Map of Human Impact on
641 Marine Ecosystems. , **319**, 948–953.

642 Hsu, A., Emerson, J., Johnson, L., Malik, O., Schwartz, J. & Jaiteh, M. (2014) *2014 Environmental*
643 *Performance Index Full Report and Analysis*. New Haven, CT.

644 Hsu, A., Johnson, L. & Lloyd, A. (2013) *Measuring Progress: A Practical Guide from the Developers of*
645 *the Environmental Performance Index*. New Haven.

646 Jacobs, R., Smith, P. & Goddard, M. (2004) Measuring performance: An examination of composite
647 performance indicators. , 1–112.

648 Jago-on, K.A.B., Kaneko, S., Fujikura, R., Fujiwara, A., Imai, T., Matsumoto, T., Zhang, J., Tanikawa, H.,
649 Tanaka, K., Lee, B. & Taniguchi, M. (2009) Urbanization and subsurface environmental issues:
650 An attempt at DPSIR model application in Asian cities. *Science of the Total Environment*, **407**,
651 3089–3104.

652 Jones, J.P.G., Collen, B., Atkinson, G., Baxter, P.W.J., Bubb, P., Illian, J.B., Katzner, T.E., Keane, A., Loh,
653 J., McDonald-Madden, E., Nicholson, E., Pereira, H.M., Possingham, H.P., Pullin, A.S., Rodrigues,
654 A.S.L., Ruiz-Gutierrez, V., Sommerville, M. & Milner-Gulland, E.J. (2011) The why, what, and

655 how of global biodiversity indicators beyond the 2010 target. *Conservation biology : the journal*
656 *of the Society for Conservation Biology*, **25**, 450–7.

657 Jørgensen, S.E., Burkhard, B. & Müller, F. (2013) Twenty volumes of ecological indicators – An
658 accounting short review. *Ecological Indicators*, **28**, 4–9.

659 Kaufmann, D., Kraay, A. & Mastruzzi, M. (2011) The Worldwide Governance Indicators: Methodology
660 and Analytical Issues. *Hague Journal on the Rule of Law*, **3**, 220–246.

661 Van de Kerk, G. & Manuel, A.R. (2008) A comprehensive index for a sustainable society: The SSI - the
662 Sustainable Society Index. *Ecological Economics*, **66**, 228–242.

663 van de Kerk, G., Manuel, A. & Kleinjans, R. (2014) *Sustainable Society Index 2014*.

664 Klugman, J. (2010) *Human Development Report 2010 The Real Wealth of Nations : Pathways to*
665 *Human Development*. UNDP, New York.

666 Kokko, H. (2005) Useful ways of being wrong. *Journal of Evolutionary Biology*, **18**, 1155–1157.

667 Kraxner, F., Nordström, E.-M., Havlík, P., Gusti, M., Mosnier, A., Frank, S., Valin, H., Fritz, S., Fuss, S.,
668 Kindermann, G., McCallum, I., Khabarov, N., Böttcher, H., See, L., Aoki, K., Schmid, E., Máthé, L.
669 & Obersteiner, M. (2013) Global bioenergy scenarios – Future forest development, land-use
670 implications, and trade-offs. *Biomass and Bioenergy*, **57**, 86–96.

671 Mace, G.M. & Baillie, J.E.M. (2007) The 2010 biodiversity indicators: Challenges for science and
672 policy. *Conservation Biology*, **21**, 1406–1413.

673 Mäkelä, A., Río, M. del, Hynynen, J., Hawkins, M.J., Reyer, C., Soares, P., van Oijen, M. & Tomé, M.
674 (2012) Using stand-scale forest models for estimating indicators of sustainable forest
675 management. *Forest Ecology and Management*, **285**, 164–178.

676 Margoluis, R., Stem, C., Salafsky, N. & Brown, M. (2009) Using conceptual models as a planning and
677 evaluation tool in conservation. *Evaluation and Program Planning*, **32**, 138–147.

678 McBride, M.F., Garnett, S.T., Szabo, J.K., Burbidge, A.H., Butchart, S.H.M., Christidis, L., Dutson, G.,
679 Ford, H.A., Loyn, R.H., Watson, D.M. & Burgman, M.A. (2012) Structured elicitation of expert
680 judgments for threatened species assessment: A case study on a continental scale using email.

681 *Methods in Ecology and Evolution*, **3**, 906–920.

682 McCauley, D.J., Pinsky, M.L., Palumbi, S.R., Estes, J. a., Joyce, F.H. & Warner, R.R. (2015) Marine
683 defaunation: Animal loss in the global ocean. *Science*, **347**, 247–254.

684 Mendola, D. & Volo, S. (2017) Building composite indicators in tourism studies: Measurements and
685 applications in tourism destination competitiveness. *Tourism Management*, **59**, 541–553.

686 Mendoza, G. a. & Prabhu, R. (2003) Fuzzy methods for assessing criteria and indicators of sustainable
687 forest management. *Ecological Indicators*, **3**, 227–236.

688 Millennium Ecosystem Assessment. (2005) *Ecosystems and Human Well-Being: Synthesis*.

689 Morar, N., Toadvine, T. & Bohannon, B.J.M. (2015) Biodiversity at Twenty-Five Years: Revolution Or
690 Red Herring? *Ethics, Policy & Environment*, **18**, 16–29.

691 Munda, G. (2005) ‘Measuring sustainability’: A multi-criterion framework. *Environment*,
692 *Development and Sustainability*, **7**, 117–134.

693 Munda, G. (2008) *Social Multi-Criteria Evaluation for a Sustainable Economy*. Springer-Verlag, Berlin
694 Heidelberg.

695 Munda, G. & Nardo, M. (2009) Noncompensatory/nonlinear composite indicators for ranking
696 countries: a defensible setting. *Applied Economics*, **41**, 1513–1523.

697 Munda, G., Nardo, M., Saisana, M. & Srebotnjak, T. (2009) Measuring uncertainties in composite
698 indicators of sustainability. *International Journal of Environmental Technology and*
699 *Management*, **11**, 7.

700 Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2008) *Handbook on*
701 *Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing.

702 Nicholson, E., Collen, B., Barausse, A., Blanchard, J.L., Costelloe, B.T., Sullivan, K.M.E., Underwood,
703 F.M., Burn, R.W., Fritz, S., Jones, J.P.G., McRae, L., Possingham, H.P. & Milner-Gulland, E.J.
704 (2012) Making robust policy decisions using global biodiversity indicators. *PloS one*, **7**, e41128.

705 Niemeijer, D. & de Groot, R.S. (2008a) A conceptual framework for selecting environmental indicator
706 sets. *Ecological Indicators*, **8**, 14–25.

707 Niemeijer, D. & de Groot, R.S. (2008b) Framing environmental indicators: moving from causal chains
 708 to causal networks. *Environment, Development and Sustainability*, **10**, 89–106.

709 Ocean Health Index. (2015) The Ocean Health Index Assessment Manual, [http://ohi-](http://ohi-science.org/manual/index.html)
 710 [science.org/manual/index.html](http://ohi-science.org/manual/index.html)

711 Ochoa-Gaona, S., Kampichler, C., de Jong, B.H.J., Hernández, S., Geissen, V. & Huerta, E. (2010) A
 712 multi-criterion index for the evaluation of local tropical forest conditions in Mexico. *Forest*
 713 *Ecology and Management*, **260**, 618–627.

714 OECD. (2003) *OECD Environmental Indicators: Development, Measurement and Use*.

715 Ojima, D.S., Galvin, K. a & Turner II, B.L. (1994) The global impact of land-use change. to understand
 716 global change, natural scientists must consider the social context influencing human impact on
 717 environment. *Bioscience*, **44**, 300–304.

718 Van Oudenhoven, A.P.E., Petz, K., Alkemade, R., Hein, L. & De Groot, R.S. (2012) Framework for
 719 systematic indicator selection to assess effects of land management on ecosystem services.
 720 *Ecological Indicators*, **21**, 110–122.

721 Parr, T.W., Jongman, R.H.G. & Kulvik, M. (2010) *The Selection of Biodiversity Indicators for EBONE*.

722 Paruolo, P., Saisana, M. & Saltelli, A. (2013) Ratings and rankings: Voodoo or science? *Journal of the*
 723 *Royal Statistical Society. Series A: Statistics in Society*, **176**, 609–634.

724 Regan, H., Colyvan, M. & Burgman, M. (2002) A taxonomy and treatment of uncertainty for ecology
 725 and conservation biology. *Ecological applications*, **12**, 618–628.

726 Riedler, B., Pernkopf, L., Strasser, T., Lang, S. & Smith, G. (2015) International Journal of Applied
 727 Earth Observation and Geoinformation A composite indicator for assessing habitat quality of
 728 riparian forests derived from Earth observation data. *International Journal of Applied Earth*
 729 *Observations and Geoinformation*, **37**, 114–123.

730 Saisana, M. & Philippas, D. (2012) *Sustainable Society Index (SSI): Taking Societies' Pulse along Social,*
 731 *Environmental and Economic Issues - The Joint Research Centre Audit on the SSI*.

732 Saisana, M. & Saltelli, A. (2011) Rankings and Ratings: Instructions for Use. *Hague Journal on the Rule*

of Law, **3**, 247–268.

Saisana, M., Saltelli, a. & Tarantola, S. (2005) Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **168**, 307–323.

Samhuri, J.F., Lester, S.E., Selig, E.R., Halpern, B.S., Fogarty, M.J., Longo, C. & McLeod, K.L. (2012) Sea sick? Setting targets to assess ocean health and ecosystem services. *Ecosphere*, **3**, 1–18.

Singh, R.K., Murty, H.R., Gupta, S.K. & Dikshit, a. K. (2009) An overview of sustainability assessment methodologies. *Ecological Indicators*, **9**, 189–212.

Van Der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. & Risbey, J. (2005) Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. *Risk Analysis*, **25**, 481–492.

Van der Sluijs, J.P., Potting, J., Risbey, J., van Vuuren, D., de Vries, B., Beusen, A., Heuberger, P., Corral Quintana, S., Funtowicz, S. & Kloprogge, P. (2002) Uncertainty assessment of the IMAGE/TIMER B1 CO₂ emissions scenario, using the NUSAP method. *Report commissioned by the Netherlands National Research Program on global Air Pollution and Climate Change, RIVM, Bilthoven, The Netherlands*, 225.

Sutherland, W.J. & Burgman, M.A. (2015) Use Experts Wisely. *Nature*, **526**, 7–8.

Turnhout, E., Hisschemöller, M. & Eijsackers, H. (2007) Ecological indicators: Between the two fires of science and policy. *Ecological Indicators*, **7**, 215–228.

Vugteveen, P., Rouwette, E., Stouten, H., van Katwijk, M.M. & Hanssen, L. (2015) Developing social-ecological system indicators using group model building. *Ocean & Coastal Management*, **109**, 29–39.