Chapter 4
Quantifying Sustainability:
Methodology for and Determinants of an Environmental Sustainability Index

Kobi Abayomi
Georgia Institute of Technology, USA

Victor de la Pena
Columbia University, USA

Upmanu Lall
Columbia University, USA

Marc Levy
CIESIN at Columbia University, USA

ABSTRACT

This chapter consider new methods of component extraction and identification for the Environmental Sustainability Index (ESI) – an aggregation of environmental variables created as a measure of overall progress towards environmental sustainability. Principally, the authors propose and illustrate a parametric version of Independent Component Analysis via Copulas (CICA). The CICA procedure yields a more coherent picture of the determinants of environmental sustainability.

INTRODUCTION

Shrinkage methods - statistical dimension reductions – are important and popular alternatives to numerical models in fields as diverse as climatology, psychology and econometrics. The objective in these methods is to identify a subset of coordinates that sufficiently describe the evolution of specific state variables. From an applied perspective, the goal is to identify (possibly lower dimension) versions of multivariate data via the extraction of salient characteristics. The data may then be recast, modulo

these characteristics, as input to further modeling. From a theoretical perspective, the proposition of a method for dimension reduction depends upon the declaration of characteristics that can offer a sound basis for extraction.

DOI: 10.4018/978-1-60960-531-5.ch004

Breiman states – *Statistics starts with data*; improved methods can illustrate latent phenomena and uncover alternative metrics in extant data [Breiman 2001]. This statistical duality, the hysteretic iteration of statistical theory and data application, is especially instrumental in emerging fields where functional and causal representations are sparse.

Social indexes, in particular environmental indexes, seek to describe as well as predict phenomena that are often poorly measured and ill-defined. An *index* is a metric, often at administrative levels, used to characterize a latent quality.

Gross Domestic Product (GDP) and of the Dow Jones indexes are common economic indices; Pacific Decadal Oscillation (PDO) and El Nino ([Francis 1998], [Gershunov1998]), climatological indices; the National Threat Level could also be called an index. Example environmental indices are the Natural Disaster Hotspots report [CHRR-World Bank 2005]; the Human and Ecosystems Wellbeing Indexes - (HWI) and (EWI) [Prescott-Allen 2001]; and the United Nations Human Development Index - (HDI) [UNDP 2006].

A goal for these environmental indices is the extraction of salient, perhaps latent, characteristics that describe or predict the elusive and undefined sustainability concept. *A fortiori*, the identification of as yet unmeasured information can illustrate the appropriate experimental design and thus guide future measurement (See Fuentes et al. [2007] for a creative example using Bernardo's [1979] fundamental comment on information maximization as a criteria).

Independent Component Analysis (ICA) - and the special case Principal Component Analysis (PCA) - extract uncorrelated and statistically independent components - or bases - of multivariate data. In ICA the model is explicit - the observed data are mixed independent sources; in PCA, implicitly, the data are mixed multivariate Gaussian. These *component analysis procedures* are used to reduce dimension – by yielding a lower order basis – and to parse or elucidate latent factors.

Environmental data are often non-Gaussian, and frequently – characteristically – extreme value [Meyers and Ganipati 2006]. Researchers apply an array of approaches: from spatial-temporal processes [Stein 2007], to stochastic optimization [Tsai and Chen 2004], and hierarchical models [Lin, Gelman, Price,and Krantz 1999]. Environmental statisticians rely upon a suite of statistical methodologies as the underlying processes are complex (as in transport phenomena), multiple (as in wastewater treatment), or latent (as in ecology). Environmental statisticians face particular challenges in modeling environmental processes; these are typically 'out-of-control' and require more sophisticated assumptions.

While the concept of sustainability has been widely embraced, it has been defined only vaguely and has proven difficult to measure with any consensus. There is a critical need for sustainability indicators; environmental statisticians have a stake in making the broad concept of sustainability operational. Researchers can justify an increased focus by providing specific measures – which decision makers can use and the public can judge – of progress or failure.

In this chapter we illustrate the 2002 Environmental Sustainability Index and exploit its dependency structure using a new version of ICA – Copula Based Component Analysis (CICA) – to extract a reduced component set as the determinants of environmental sustainability. This approach is designed to highlight important information, suggest some focal metrics, and discredit others.

A unifying definition for an *index*, in the context of this paper: a function that maps disparate multivariate data onto a scalar at administrative units. An index should be:

1. **Transparent**: The methodology use to construct the index should be clear and unambiguous. Assumptions and decisions that affect index values (`scoring') should be well stated.

Related Content