Fisheries Research 182 (2016) 172–176

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Statistical tools to assess the breadth and depth of shrimp aquaculture certification schemes

Michael F. Tlusty[a,b,*], Matthew Thompson[a], Heather Tausig[a]

[a] New England Aquarium, Boston, MA 02110, USA
[b] University of Massachusetts at Boston, Boston, MA 02125, USA

ARTICLE INFO

Article history:
Received 30 April 2015
Received in revised form 6 October 2015
Accepted 9 October 2015
Available online 29 October 2015

Keywords:
Aquaculture
Certification
Cluster analysis
Continuous improvement
Multiple analysis of variance
Seafood
Shrimp
Standard
Statistics
Sustainability

ABSTRACT

Nearly two decades ago, significant concern about the environmental impacts of aquaculture production gave rise to environmental certification schemes as a means to ensure production adhered to less impactful environmental standards. Currently, some governments with more robust regulations are also engaging by creating national voluntary sustainability certification schemes. While its likely that the majority of aquaculture standards originated independently of one another, they generally overlap on the main impact issues but with differing requirements for compliance, making it difficult to differentiate more unique or robust schemes. Differentiation and encouragement of the adoption of the more rigorous standards is one means to encourage the industry to lessen environmental and social impacts. One confounding factor in benchmarking studies is the lack of a consistent methodology making wider comparisons difficult. Here we created a tool for standards comparison that began with a broad reaching set of factors based on the FAO Technical Guidelines on Aquaculture Certification and the International Principles for Responsible Shrimp Farming. Our analysis first compared if the factors were addressed (the breadth of the scheme), as well the mechanism by which compliance was required (the scheme depth), a proxy for how rigorously each factor was addressed. This analysis compared 112 factors divided into five impact areas (community, environment, food safety, feed and marine resource use, and supply risk), for three national shrimp aquaculture certification schemes including Indonesia, Thailand, and Vietnam, and three global schemes including Aquaculture Stewardship Council, Global Aquaculture Alliance and GLOBALG.A.P. The global schemes were found to be of greater breadth and depth than the national schemes. As an analysis tool, the breadth–depth (B–D) graphical analysis was compared to more rigorous statistical methods including multiple analysis of variance and cluster analysis. Overall, the B–D analysis provided a relatively simple means to assess rigor (breadth and depth) of multiple certification standards when compared using a broad baseline set of factors. It was further observed that the global certification programs overlapped in breadth and depth largely because of an uneven application across the five impact areas. The use of this analysis can be implemented to better understand similarities and differences between standards, and can be foundational in developing and adjusting schemes to ensure they are unique and operating at different levels of rigor, which can be a roadmap towards increasing the sustainability of aquaculture production.

1. Introduction

In the face of burgeoning global population growth, creating a food secure future is one of the world's foremost challenges (FAO et al., 2014). Given that food production contributes almost 30% of global greenhouse gas emissions (Vermeulen et al., 2012), one of the more immediate means to ensure food security is to lessen the impacts of food production on ecosystem health (Dobermann and Nelson, 2013; Searchinger et al., 2013). Multiple approaches are necessary to improve the food system including but not limited to those addressing policy (Rosegrant and Cline, 2003), livelihoods, capabilities, and entitlements (Pritchard et al., 2013), as well as changes to access, utilization and production (Dobermann and Nelson, 2013; FAO, 2013; Searchinger et al., 2013). Many of these approaches require longer-term solutions, such as technological, behavioral, and policy innovations and modifications. However, shorter-term solutions exist with one of the more popular being environmental certification, a strategy that incentivizes more environmentally-friendly modes of food production (Ward and Phillips, 2008). Well-designed and effectively-implemented

* Corresponding author.New England Aquarium, Boston, MA 02110, USA.
E-mail address: mtlusty@neaq.org (M.F. Tlusty).

http://dx.doi.org/10.1016/j.fishres.2015.10.008
0165-7836/

certification schemes can fuel improvements by providing market recognition to those working to address and minimize the common adverse impacts of food production (Bush and Oosterveer, 2015).

One upshot of the popularity of this approach is that there are now in excess of 30 certification schemes for aquaculture production (Lee, 2008). From a theoretical standpoint, both vertical (Bush and Oosterveer, 2015) and horizontal (Tlusty, 2012) differentiation of schemes is a means to create a journey toward a more sustainable state (Tlusty, 2012). Over time, markets, and major buyers can source products from schemes of increasingly greater rigor, which can create the steps toward a more sustainable food production system. In the face of these growing challenges, there is a call by industry to understand "equivalency" amongst the schemes. However, to date there has been no consistent methodology to assess the comparative rigor (Tlusty et al., 2012) of seafood certification schemes. While a number of benchmarking and comparative studies have been conducted (including but not limited to Boyd and McNevin, 2012; Sullivan et al., 2012; Thrane et al., 2009; Trade Map, 2015; Volpe et al., 2011; WWF Switzerland and Norway, 2007) they vary in their methodology and lack statistical approaches, often resulting in insufficient power to differentiate certification schemes (e.g., Sustainable Fisheries Partnership, 2011).

Here, we developed a two-dimensional analytical method for comparing the standards set by seafood certification schemes. The analysis assessed both the number of factors each standard considers (breadth) as well as the aspiration of each factor (depth). While we previously proposed (Tlusty et al., 2012) that increasing both breadth and depth would result in a more sustainable scheme, nothing implicitly suggested that breadth and depth would scale linearly. It could be argued that standards could vary according to both breadth and depth and be undifferentiated (few factors, low aspiration), single issue (a few factors, addressed aspirationally), broad but general (many factors but of low aspiration), or rigorous (many factors of high aspiration, see Graphical abstract). We then compared the results of this breadth–depth (B–D) analysis to analyses using more conventional statistical tools including multiple analysis of variance and cluster analysis. These three analytical methods were used to assess the relative positioning of six shrimp aquaculture certification schemes including 3 national (including Indonesia, Thailand, and Vietnam) and three global sustainability standards (including Aquaculture Stewardship Council, Global Aquaculture Alliance Best Aquaculture Practices, and Global G.A.P.). These results of these analyses were discussed in the light of using multiple certification schemes to create positive movement toward greater sustainability of aquaculture production.

2. Methods

The shrimp-specific aquaculture standards for certification schemes assessed here included three national schemes; Indonesia (CBIB, Decree of the Minister of Maritime Affairs and Fisheries No. KEP.02/Men/2007), Thailand (Thai Agricultural Standard (TAS) 7401-2009 Good Aquaculture Practices for Marine Shrimp Farm), and Vietnam (National Standard on Good Aquaculture Practices in Vietnam (VietGAP) No. 1503/QĐ-BNN-TCTS), as well as three global schemes; Aquaculture Stewardship Council (Draft Standards for Responsible Shrimp Aquaculture. Version 3.0 for Guidance Development and Field Testing), Global Aquaculture Alliance (Draft Aquaculture Facility Certification. Finfish and Crustacean Farms. Rev. 4/13), and GLOBALG.A.P. (Version 4.0. Edition 4.0-1_FEB2012, generic or shrimp specific and published or draft). A full description of schemes is provided in Supplemental information 1. These schemes were compared using a newly derived B–D analysis, a two dimensional analysis to assess the breadth and depth of a standard. In essence, this analysis determined if a factor was covered by the standard (breadth), and if so, the degree to which compliance was required, a proxy for how rigorously the factor was addressed (depth). The breadth dimension of this analysis was based on the FAO (2011) Technical Guidelines on Aquaculture Certification and the International Principles for Responsible Shrimp Farming (FAO et al., 2006). The 112 factors discussed in these guidelines were divided into five impact issues (community, environment, food safety, feed and marine resource use, and supply risk, nine to 31 factors per impact issue, see Supplemental information 2). Each factor was evaluated if it was covered by the standard, and the breadth value was calculated as the average percentage of factors addressed per the 5 impact issues. Each component factor addressed by the standard was then evaluated using a seven-point scale to estimate relative rigor (Table 1). The scale began at zero (factor was not addressed), where a score of one indicated legal compliance and the maximum of seven indicated the factor was aspirationally addressed by the scheme (see Table 1). Interim scores between two and six highlighted a progression from non-audited recommendations, through management plans, to performance-based metrics. The breadth and depth scores for each factor were presented graphically as averages and 95% confidence intervals of the five impact issues for each standard. Note that zero values were truncated from the calculation of the depth value, and were only used when the breadth value was zero.

Differences between standards were also analyzed as a multiple analysis of variance (MANOVA) in R (R Core Team, 2014) with orthogonal post-hoc comparisons examined for statistical similarity ($p > 0.05$) with the Pillai–Bartlett Trace multivariate test statistic (Fox et al., 2013). Contrasts included national vs. global scheme standards, along with paired comparisons of adjacent standards based on ranks of Eigen values (determined by the dist( ) function in R). This B–D analysis was compared to the results obtained through cluster analyses (JMP 8.0.2.2, SAS, Cary NC). Two cluster analyses were conducted, the first on the average breadth and depth scores for each impact issue, and the second on all 112 factors. Differences in these two cluster analyses assessed how the aggregation of factors into impact areas may influence interpretation of results. For each cluster analysis, a minimum (0) and maximum (100 for breadth, and seven for depth) were also forced into the model as a means to bound most and least rigorous schemes. Data were ordered by the first principle component, clusters were hierarchically determined with the Ward method, and resultant clusters were plotted with distance scales. For the cluster analysis, depth scores defaulted to 0 only if the breadth was also 0.

3. Results

For all certification schemes, there was a positive correlation between breadth and depth. As the number of factors addressed within a standard increased, the assessment became more rigorous (Fig. 1). The B–D analysis found differences between the national and global schemes, and that the CBIB standard was of lower breadth and depth than the other national schemes. There was a great deal of overlap in the B–D analysis of the global schemes standards.

The MANOVA analysis of these data supported the B–D analysis, but with a greater ability to distinguish differences between adjacent standards. There was no statistically significant difference between the impact factors (MANOVA, Pillai test statistic = 0.16, $F_{4,24} = 0.56$, $p > 0.8$), but the schemes were statistically significantly different (MANOVA, Pillai test statistic = 0.89, $F_{5,24} = 3.85$, $p > 0.001$). The national and the global schemes differed (orthogonal contrast, ASC + GAA + GG-CBIB-TAS-VG, MANOVA, Pillai test statistic = 0.53, $F_{1,24} = 13.22$, $p < 0.001$). Within this larger structure, Eigen values ranking of standards found the order to be CBIB TAS VG GAA GG