Environmetrics
DOI: 10.1002/env.1055 (on Early View since Nov 18, 2010)
This version is the authors' manuscript accepted for publication.

Latent Health Factor Index: A Statistical Modeling Approach for Ecological Health Assessment

Grace S. Chiu[1, 5] (corresponding author),

Peter Guttorp,[2] Anton H. Westveld,[3] Shahedul A. Khan[1] and Jun Liang[4]

[1]Department of Statistics and Actuarial Science, University of Waterloo, Ontario N2L 3G1, Canada.

[2]Department of Statistics, University of Washington, Seattle, Washington 98195, USA.

[3]Department of Mathematical Sciences, University of Nevada Las Vegas, Nevada 89154, USA.

[4]Canadian Institute for Health Information, 90 Eglinton Avenue East, Suite 300, Toronto, Ontario M4P 2Y3, Canada.

[5]**Correspondence**:

CSIRO Mathematics, Informatics and Statistics, GPO Box 664, ACT 2601, Australia.

E-mail: grace.chiu@csiro.au Telephone: +61 2 6216 7114 Facsimile: +61 2 6216 7111

SUMMARY

Multimetric indices (MMIs) are appealing scalar-valued tools for policy makers when rating ecosystems with respect to biological conditions that are not directly measurable. For conventional assessment of ecological health using MMIs, the quantitative calibration of health qualities can be specific to the investigator, and to the geographical region and time frame of interest. We propose a statistical-model-based approach that provides a systematic mechanism to construct MMIs; our approach aims to address some common issues of conventional practices, including the loss of information from data, spatio-temporal restrictions, and concerns over arbitrariness and costs. Our latent health factor index (LHFI) is obtained via statistical inference for an unobservable *health factor* term in a mixed-effects analysis-of-covariance regression that directly models the relationship among metrics, a very general notion of health, and factors that can influence health. We illustrate the approach by constructing an LHFI for a freshwater system using benthic taxonomic data in various Bayesian hierarchical formulations of generalized linear mixed models, implemented by Markov chain Monte Carlo techniques. The concept of the LHFI is also applicable to medical and other contexts.

KEY WORDS: Bayesian inference; ecosystem health; hierarchical models; mixed-effects models; multimetric index

1 INTRODUCTION

Many scientific disciplines involve assessing underlying conditions with a single number computed based on various measurable characteristics. We generically refer to these conditions as *health* throughout this article. A familiar example is an economic index. Another example is the body mass index (BMI), which combines a person's height and weight measurements to yield a scalar-valued quantification of obesity, a form of poor health. Scalar-valued assessments are naturally appealing for their structural simplicity and supposed ease of interpretation, particularly in decision making contexts such as disease diagnosis. In certain applications, the definition of the scalar index may incorporate scientific theory in the subject matter. Other applications of scalar indices may lack a unified, systematic approach for the construction or interpretation of the index. For example, how one should interpret BMI values in different situations has long been a contentious issue (e.g. López-Alvarenga *et al.*, 2003). To assess aquatic ecosystem health, conventional indices such as the benthic index of biotic integrity (B-IBI) (Kerans & Karr, 1994) and its variants (e.g. McCormick *et al.*, 2001) are also constructed by studying and combining multiple indicator variables, or *metrics*, to reflect a very general notion of underlying health of field sites. IBI variants and *observed-to-expected* (O/E) indices (e.g. Hawkins *et al.*, 2000) are types of reference-based health indices: sites whose health is under scrutiny are gauged against sites identified as comparable in every aspect except for the unstressed *reference* conditions. These days, unstressed sites can be difficult to locate due to widespread environmental degradation across the globe, or inaccessible to scientists due to their remoteness. For this and other reasons, reference criteria are often admittedly arbitrary (e.g. CEH Web; Hawkins *et al.*, 2000; Kennard *et al.*, 2006), leading to calibration schemes that are specific to geographical regions and time frames (e.g. Moss *et al.*, 2001). More broadly, existing index construction approaches are based largely on human intuition. Given that such indices ought to reflect unobservable conditions of interest, the present

lack of a unified, systematic approach to quantify qualitative features is of concern. More in-depth discussions of this and other issues appear elsewhere (e.g. Chiu & Guttorp, 2004, 2006; Chiu *et al.*, 2008; Steedman & Regier, 1990; Suter, 1993).

Despite all, conventionally constructed multimetric indices (MMIs) remain popular in policy-making contexts (e.g. SEQ; Stoddard *et al.*, 2005) for good reasons: a scalar MMI value is a readily communicable "report card," and is appealing in its high biological content (from subject-matter expertise involved in metric selection and index construction) and structural simplicity (being easily computable from the chosen metrics). However, to enhance scientific integrity of an MMI, desirable statistical inferential properties should not be overlooked. As Dr. Neil McKenzie (Chief, Land and Water) of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) advocated in the 2009 CSIRO Workshop on Nationally Relevant Environmental Monitoring, "environmental information increases in utility when it reduces uncertainty for a decision maker."

For proper reduction of uncertainty, first the amount of uncertainty — propagated through the various stages of data collection / index construction / inference — should be properly assessed. Often, this is a challenging task for conventional indices, due to the various (semi-)qualitative schemes involved in collapsing multimetric data into scalar. Thus, we propose a new approach for constructing MMIs that combines the statistical advantages of model-based techniques and the communicability of conventional indices (e.g. BMI, B-IBI). Regarding the former, (1) propagation of uncertainty is built into our approach, and (2) the method allows one to assess health based (a) solely on species composition, or (b) additionally on environmental factors that influence health. A consequence of this flexibility is the ability of the investigator to predict health using easily measurable environmental variables without the need for costly species sampling and subsequent laboratory assays to identify species composition.

We illustrate our method in the context of assessing freshwater ecosystem health. For this, we take metrics that have been expertly identified as informative, then model their interdependence by regressing them on health as a latent covariate. This unobservable factor can be estimated statistically, thereby yielding a scalar assessment of underlying health conditions. If desired, observable environmental factors — e.g. stream order and other physical traits, spatial and/or temporal locations of sites, and human demographic variables that could directly influence site health — may be included as extra covariates. The resulting *latent health factor index* (LHFI) can then be compared to existing indices for the same data. If both are deemed to contain similar information about health, then the LHFI may be preferred for its unambiguous quantitative nature. This is because (1) its construction is based almost entirely on standard modeling principles while accounting for propagation of uncertainty, (2) its performance is tractable statistically, (3) it involves no intermediate dimension reduction procedures that are qualitative in nature and distort valuable information from available data, and (4) when covariates are included to explain the latent health, then (a) prediction of site health and its proper inference are straightforward, and (b) the fitted model can help resource managers to identify external factors that influence health. Specifically, the significance of their impact on health can be statistically assessed and classified, thus readily providing policy makers with unambiguous guidelines for prioritizing conservation measures. Indeed, both ranking sites on health and classifying factors according to their impact on health, as well as assessing their uncertainty, can be achieved in a single step of fitting the model.

Our methodology and its rationale and principles are given in Section 2. Statistical inference

for health, including its prediction, is discussed in Section 3. In Section 4, we construct an LHFI for the 1997 Puget Sound Lowland (PSL) freshwater ecosystem. Although not necessarily of present scientific interest, the health of this ecosystem is highly relevant to ecosystem health assessment in general when various approaches are compared. Thus, we discuss in Section 5 how our PSL LHFI compares to existing indices in numerical value, and in statistical and biological performance. In Section 6, we suggest how one may make biological interpretations from a model-based LHFI without relying on external or prescribed reference conditions. Overall findings and advice are summarized in Section 7.

2 A FULLY QUANTITATIVE MODELING FRAMEWORK FOR THE LHFI

In some biomonitoring studies, it is preferred to assess ecological health by basing entirely on faunal metrics that have been selected according to their apparent relationships with non-faunal environmental variables (e.g. Stoddard *et al.*, 2005). To maximize the utilization of relevant data when constructing a report-card index, other studies regard any health-related variable, faunal or otherwise, as a potential metric (e.g. SEQ). In either case, mainly study-specific, semi-qualitative schemes are used to calibrate multiple metrics into a scalar index. Here, we present our quantitatively oriented LHFI scheme that is intended as an alternative: it relies on systematic statistical model building procedures, and allows the investigator to include or exclude non-faunal environmental variables for index construction.

Many ecosystems subject to conservation measures involve bodies of water, so that much focus is on the ecosystem defined at the level of the aquatic environment. To assess aquatic ecosystem health, benthic macroinvertebrate faunal data are typically considered. In general, benthic taxonomic data are collected by obtaining a mud sample according to a standardized protocol, separating the animals from this mud, then sorting each animal into one of many taxa (e.g. Cuffney *et al.*, 1993; Ferraro *et al.*, 2006; Marshall *et al.*, 2006). This collection of animals forms the field sample. Replicate field samples per site are common. Ecologists identify various numerical aspects of the field sample composition to reflect ecosystem health. For instance, an abundance of animals from predatory taxa reflects a healthy ecosystem that can sustain a large number of predators. Similarly, a field sample rich in stress-sensitive taxa indicates an ecosystem that has been subject to minimal stress. Corresponding numerical measures, known as *metrics*, are used to construct health indices.

By combining metrics to form a health index, conventional schemes essentially regard health as the response and metrics as covariates or driving factors of health measurements. In reality, some metrics are indicators of health. Thus, in a statistical model, such metrics would appear more naturally as responses, being driven by health as a latent covariate. If desired, we allow health to appear hierarchically in a latent regression on non-faunal variables that have a potential impact on the field site's overall health. Schematically, the two approaches can be contrasted as follows.

$$\begin{aligned} &\textbf{CONVENTIONAL}: && \textbf{metrics} \longrightarrow \textbf{health} \\ &\textbf{LHFI MODELING}: && \textbf{covariates} \longrightarrow \textbf{health} \longrightarrow \textbf{metrics} \end{aligned} \tag{1}$$

In the absence of explicit scientific theory for index definition, the above role reversal between

metrics and health is fundamental to the scientific integrity of our index construction approach. This is because their relationship is directly modeled in an intrinsically quantitative framework, in a causal hierarchy, without ambiguous metric manipulation that is common to existing ecological health indices. While subject-matter expertise determines what constitutes biologically meaningful metrics, statistical principles enable effective preservation of information contained by these metrics during index construction, resulting in a biologically meaningful health index.

2.1 *Building the latent health factor model*

For the benefit of decision makers in environmental policy and monitoring, we exploit the practicality of latent variable modeling, while retaining the scientific and practical appeal of MMIs. Researchers in other disciplines have employed explicit statistical inference for latent quantities to assess unobservable traits (e.g. Hays *et al.*, 2000; Martin & Quinn, 2002; Pietrobon *et al.*, 2004; Rabe-Hesketh & Skrondal, 2008; Rosas, 2009; Skrondal & Rabe-Hesketh, 2008; Stock & Watson, 1989; Ward & Hoff, 2007). These methods may be grouped as follows: variants of factor analysis, variants of item response modeling, and context-specific latent variable modeling. The context of biomonitoring puts our method in the third group.

To consider the relationship in the schematic of (1), multivariate-response models could be considered, although they often require complex parameterizations for non-Gaussian data. Instead, we rely on the simple principles of analysis of (co)variance (ANO(CO)VA), expressed as a generalized linear mixed model (GLMM). For benthic data, let Y_{ijk} denote the (possibly transformed) value of the jth metric for the ith site's kth independent replicate, i=1,...,n, j=1,...,J, and k=1,2,...,K. In light of (1), Y_{ijk} can thus be explained, in an ANOVA model, by site effect H_i (regarded as the site's underlying health) and metric effect β_j. Assume randomly chosen sites. Then, if desired, health can be regressed on $\boldsymbol{x}_i$, a vector of observable covariates. Let $f_{\boldsymbol{\theta}}$ be the function of latent regression with coefficients $\boldsymbol{\theta}$. Altogether, we have a hierarchical GLMM (ANOVA if $f_{\boldsymbol{\theta}}$ is taken to be constant, ANOCOVA otherwise)

$$\nu_{ij} \equiv g\big(E[Y_{ijk}|H_i,\beta_j]\big) = H_i + \beta_j\,, \qquad H_i = f_{\boldsymbol{\theta}}(\boldsymbol{x}_i) + \varepsilon_i \tag{2}$$

where g is an appropriate link function and ε_i's are independent and identically distributed (iid) errors. Our main interest is in H_i. Although health itself is latent, its estimate $\widehat{H}_i$ from the model fit is an explicit quantification of site health. Note that no site-metric interaction is modeled since homogeneity induced by random selection of sites can be assumed. Finally, we model β_j's as random effects with mean 0 (to avoid confounding with the intercept from H_i) and an appropriate covariance structure. Modeling β_j's as random has the following advantages: (i) a non-diagonal covariance allows for any dependence of Y_{ijk}'s over j due to informational overlap of metrics (see Section 4); (ii) this dependence may be readily investigated by comparing models that are nested with respect to $\boldsymbol{\Sigma}$ (e.g. identity nested in diagonal nested in unstructured); and (iii) having unequal variances for β_j's allows for the notion of an unknown, unequal weighting of metrics.

When non-faunal environmental covariates are utilized for index construction, deciding on the functional form of $f_{\boldsymbol{\theta}}$ in the ANOCOVA model may require some effort. An ANOVA model without covariates can be fitted initially, then an exploratory analysis conducted on the resulting

$(\widehat{H}_i, \boldsymbol{x}_i)$'s to identify an appropriate form of $f_{\boldsymbol{\theta}}$ to be used for (2). Alternatively, a simple form such as linear may be blindly fitted, then diagnosed subsequent to the fit.

2.2 *Model for combining spatial and other types of domains*

When developing an ecological health index, neighboring geographical domains may be similar enough to share the same metrics, yet different enough that traditional metric calibration devised for one region may not effectively reflect the health conditions of another.

Suppose our J metrics are deemed adequate for multiple spatial domains. The goal is to assess in one combined study the ecological health of sites from all domains. However, formal spatial models are typically impractical: ecologists often warn of the sparsity / large variability of ecological data of this sort (mainly due to the substantial costs of observing and processing each field sample) that prevent statistical detection of underlying spatial correlation patterns. The traditional approach for, say, IBI variants would then require recalibration of all J metrics to account for the different spatial scales. Painstaking effort aside, personal preferences could play a heavy role in this recalibration, adding ambiguity to the health assessment. Chiu & Guttorp (2006) advocate the *gold standard* SHIPSL scoring / calibration scheme, where metrics are standardized against predetermined values of region-specific "gold-standard" mean and standard deviation. However, these authors warn that implementation could be challenging in practice. To preserve biological and statistical integrity while formal spatial modeling is impractical, our LHFI approach handles the issue as follows. Let $Y_{i(\ell)jk}$ denote the value of the jth metric from replicate sample k in the ith site nested within spatial domain ℓ. In the presence of environmental covariates, a spatial effect term λ_ℓ can be added to yield $\nu_{i(\ell)j} = H_{i(\ell)} + \beta_j$ and $H_{i(\ell)} = \lambda_\ell + f_{\boldsymbol{\theta}}(\boldsymbol{x}_{i\ell}) + \varepsilon_{i(\ell)}$, where $f_{\boldsymbol{\theta}}(\boldsymbol{x}_{i\ell})$ may be removed if covariates are absent, and λ_ℓ modeled as random or fixed depending on the context. For the intercept λ_ℓ to be logically regarded as the overall health of the ℓth domain, all domains should share the same set of metrics and covariates. Then, the simple addition of this spatial effect term in the LHFI model allows us to study the health conditions by estimating $H_{i(\ell)}$ over all sites simultaneously and without ambiguity. The same principles may be applied to multiple temporal domains, through a temporal effect term (e.g. *year*), possibly ordinal, to the model in a similar fashion. To account for both types of domains, a spatio-temporal interaction term may be included. Similarly, *stream order* (size category) could be considered a type of domain, and incorporated as such.

Currently, we are unaware of quantitative biomonitoring indices that account for spatial and temporal differences in a statistically sound manner; ours attempts to do so when field sites are sparse. In very rare cases, field sites are sampled densely. Higgs & Hoeting (in press) use data from one such study in Maryland (USA) to illustrate a spatial model for rating streams based on the stream's IBI score. In principle, an LHFI can be constructed for this system of streams by modifying (2) to incorporate spatial dependence among ε_i's. LHFI construction for general inter-regional and/or -temporal studies is ongoing (Chiu *et al.*, in progress; Wu, 2009). In the remainder of this article, we return our focus to (2) with iid ε_i's.

3 COMPUTING THE LHFI: BAYESIAN MODEL INFERENCE

According to Gelman & Hill (2007), the hierarchical Bayesian framework is the most direct way to handle models with latent structures, as each level of latent regression in the *model hierarchy* has a direct correspondence to a specific level in the *parameter hierarchy*. As a bonus, unlike some classical techniques, this framework does not rely on asymptotics that may be inappropriate due to small sample sizes and/or unbalanced designs that are common in ecological and other contexts.

For us, let $\boldsymbol{H} = (H_1, \ldots, H_n)^T$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^T$. Let $\boldsymbol{\nu}$ denote the vector of ν_{ij}'s; $\boldsymbol{Y}$, the vector of Y_{ijk}'s; $\mathbb{X}$, the design matrix whose rows are $\boldsymbol{x}_i$'s; and $\boldsymbol{\Omega}$, the vector of remaining model parameters, including $\boldsymbol{\theta}$ and those from the distributions of $\boldsymbol{\beta}$ and ε_i's. In the Bayesian context, all but $\mathbb{X}$ are considered random quantities. Next, let P be the generic label for a distribution. Then, $P(\boldsymbol{\Omega})$ is the prior for $\boldsymbol{\Omega}$, $P(\boldsymbol{Y}|\boldsymbol{\nu})$ or $P(\boldsymbol{Y}|\boldsymbol{H}, \boldsymbol{\beta})$ is the likelihood, $P(\boldsymbol{H}|\boldsymbol{\Omega}, \mathbb{X})$ is the distribution of $\boldsymbol{H}$, and $P(\boldsymbol{\beta}|\boldsymbol{\Omega})$ is that of $\boldsymbol{\beta}$. In the absence of concrete preconceptions of $\boldsymbol{\Omega}$, a diffuse prior $P(\boldsymbol{\Omega})$ is commonly applied. Bayesian inference for $\boldsymbol{H}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Omega}$ is made based on the joint posterior $P(\boldsymbol{H}, \boldsymbol{\beta}, \boldsymbol{\Omega}|\boldsymbol{Y}, \mathbb{X})$. We assume independence of $\boldsymbol{H}$ and $\boldsymbol{\beta}$, so that

$$P(\boldsymbol{H}, \boldsymbol{\beta}, \boldsymbol{\Omega}|\boldsymbol{Y}, \mathbb{X}) \propto P(\boldsymbol{Y}|\boldsymbol{H}, \boldsymbol{\beta})\, P(\boldsymbol{H}|\boldsymbol{\Omega}, \mathbb{X})\, P(\boldsymbol{\beta}|\boldsymbol{\Omega})\, P(\boldsymbol{\Omega})\,. \tag{3}$$

Then, one can take the marginal posterior mean (or median / mode) of H_i to be our LHFI:

$$\widehat{H}_i \equiv \widehat{H}_i(\boldsymbol{Y}, \mathbb{X}) = E(H_i|\boldsymbol{Y}, \mathbb{X}) = \int H_i \int\int\int P(\boldsymbol{H}, \boldsymbol{\beta}, \boldsymbol{\Omega}|\boldsymbol{Y}, \mathbb{X})\, d\boldsymbol{\beta}\, d\boldsymbol{\Omega}\; d\boldsymbol{H}_{-i} \tag{4}$$

where $\boldsymbol{H}_{-i}$ is obtained by removing H_i from $\boldsymbol{H}$. Estimation uncertainty can be assessed by highest posterior density (HPD) intervals. Once the right-hand-side of (3) is determined, obtaining HPD intervals is routine and unambiguous. In contrast, confidence intervals for existing indices such as IBI and SHIPSL variants rely on the non-parametric bootstrap, and are negatively biased in location and width in general (Chiu & Guttorp, 2006). Although closed forms may not exist for (3) or (4), samples can be simulated from (3) by numerical methods such as Markov chain Monte Carlo (MCMC). Approximating (4) based on posterior draws is then trivial. The remaining nuisance parameters can be estimated similarly.

3.1 *Predicting site health*

The ability to reliably predict site health has significant practical implications, especially for ecosystems that traditionally involve the painstaking and costly collection and laboratory analysis of taxonomic data. For example, each benthic sample may contain thousands of minute animals to be sorted and identified. Instead of handling benthic animals directly per field sample per site, it becomes invaluable to have reliable and easily observable surrogate data that can reflect ecosystem health equally well for some sites. In a statistical modeling framework, covariates can be surrogates for the response. Thus, given a set of covariates, their quality as health indicators in their own right can be gauged via prediction inference.

An issue with monitoring ecosystem health by common indices is the inability to make sound statistical inference on the predictions of site health. For IBI / SHIPSL variants and O/E indices, one may first compute the index values, then regress them on observable covariates and make

predictions of future site health via the fitted regression. For this two-step approach, uncertainty in the raw taxonomic data is unaccounted for. In contrast, prediction / interpolation of health at a "future site" (e.g. one with no taxonomic data) and its inference is much more straightforward given Model (2) that has been fitted to a set of sites with taxonomic and covariate data. This inference is done via the posterior predictive distribution $P(H^*|\boldsymbol{Y}, \mathbb{X}, \boldsymbol{x}^*)$, where a "$*$" denotes a future value. Specifically, first consider a single Monte Carlo sample from the joint posterior (3). Extract from this sample those components of $\boldsymbol{\Omega}$ that are relevant to (2). Now, substitute these components together with $\boldsymbol{x}^*$ into (2) to simulate a Monte Carlo draw from $P(H^*|\boldsymbol{Y}, \mathbb{X}, \boldsymbol{x}^*)$. Repeat this process until a collection of simulated draws are obtained from $P(H^*|\boldsymbol{Y}, \mathbb{X}, \boldsymbol{x}^*)$. Then, $\widehat{H}^*$ is approximated by, e.g. the mean, of these draws. Predictive HPD intervals are also easily approximated using appropriate quantiles of the simulated collection of draws.

The advantage of our predictive inference approach is that it accounts for the modeled relationship among metrics, health, and environmental variables simultaneously, "rolling up" uncertainty from different levels of the data hierarchy in an unambiguous fashion. Predictive inference can also be used in cross-validation for model evaluation, as demonstrated by Chiu *et al.* (2008). These authors report that despite having many unknown quantities but relatively few sites and replicates, various formulations of the ANO(CO)VA GLMM for the PSL data all show reasonable predictive power and no apparent overfitting.

4 LHFIs FOR THE PSL

4.1 *Data*

We apply our methods to the 1997 PSL benthic data (obtained from Chiu & Guttorp, 2006) to construct and compute LHFIs. These data were collected from n=18 sites scattered over 9 streams (Table 1), with K=3 replicate field samples per site. For the PSL, an observed animal could belong to one of 80 taxa, and the animal count per taxon could range from 0 to more than 1,000, but is equal or close to 0 for many taxa. Biologists have previously identified 10 useful metrics for the PSL (Table 1), whose values are computed based on the 80 counts. Here, all 10 metrics are highly correlated due to their definitions: 7 describe taxa richness (count), and 3 describe relative abundance (%). For #Intol and %Tol, non-tolerant taxa are not necessarily intolerant, as some taxa are classified as neither tolerant nor intolerant. Also note that %Tol and %Dom3 are negatively associated with health (Morley, 2000). Chiu & Guttorp (2004) take the obvious transformations %NonTol=100%$-$%Tol and %NonDom3=100%$-$%Dom3, so that higher values of the index correspond to higher metric values. They also show that it is beneficial, at least statistically, to convert the taxa richness counts to percentages, before combining them with relative abundance metrics to form a health index. They suggest making #Tx the denominator in converting the other 6 count metrics into *relative richness* (%), in the same way that sample cardinality N_{ik} (total number of animals in the field sample) is used to define relative abundance. This way, all J=9 variables now share the same scale over all sites and replicates. (To avoid handling metrics on different scales, Chiu *et al.*, 2008, also consider just the 7 count metrics, including #Tx, in a Poisson regression.)

In addition to metrics data, associated with each stream are data for environmental covariates

(taken from Morley, 2000) that include *urbanization* and *Global Positioning System (GPS) co-ordinates*. Urbanization perceivably has a strong influence on stream health; its values are in the form of percentages of total impervious area in the sub-basin to which the stream belongs. Thus, some sites share the same urbanization value. GPS co-ordinates are latitudes and longitudes recorded with sensitive instruments, and are unique to each site. Since urbanization and longitude are highly correlated, we omit the latter from our consideration.

4.2 *Model A*

To formulate the LHFI model, all 9 metrics may be considered observed success probabilities, where "success" is an occurrence of the taxon (towards richness) or animal (towards abundance) that indicates a healthy stream. Therefore, we construct a logistic model. We first consider urbanization as the sole covariate due to its preconceived impact on health. When we diagnose the fit of this model, we will also decide if latitude should be included in the latent regression.

The 9 metrics can be broken down into J_1=6 pertaining to richness, and J_2=3 to abundance. Let s=1,2 denote the respective groups. Next, for replicate k from site i, let $Y_{ij(s)k}$ denote the total number of successes for metric j nested in group s, each success occurring with probability $p_{ij(s)\bullet}$. Also, let $\nu_{ij(s)\bullet}$ denote the logit-transformed $p_{ij(s)\bullet}$. Initially we blindly consider linearity in the latent regression, with urbanization x_i centered to remove dependence between the regression intercept and slope. Finally, we take metric effects β_j's to be independent but heterogeneous. Linearity and the simple structure for Cov($\boldsymbol{\beta}$) will be diagnosed after fitting the model. Then, we rewrite (2) as

$$\left(Y_{ij(s)k}|T_{i\bullet sk}, p_{ij(s)\bullet}\right) \overset{ind}{\sim} \text{Binomial}(T_{i\bullet sk}, p_{ij(s)\bullet}), \quad T_{i\bullet sk} = \begin{cases} (\#\text{Tx})_{ik} & \text{if} \quad s=1 \\ N_{ik} & \text{if} \quad s=2 \end{cases}, \tag{5}$$

$$\ln \frac{p_{ij(s)\bullet}}{1-p_{ij(s)\bullet}} \equiv \nu_{ij(s)\bullet} = H_i + \beta_{j(s)}, \qquad \sigma_{1(1)} = \sigma_{2(1)} = \sigma_{3(1)}, \tag{6}$$

$$H_i = \theta_0 + \theta_1(x_i - \overline{x}) + \varepsilon_i, \qquad (\varepsilon_i|\sigma_H) \overset{iid}{\sim} N(0, \sigma_H^2), \quad (\beta_{j(s)}|\sigma_{j(s)}) \overset{ind}{\sim} N(0, \sigma_{j(s)}^2), \tag{7}$$

for j=1,...,J_s and k=1,2,3. The model stipulates that success probability is affected by site health and metric, but not by metric type (richness / abundance). It further assumes a common variance for the metric effects of *Eph., Ple.*, and *Tri.* taxa (s=1, j=1,2,3), known collectively as EPT taxa. (Analyses by Chiu *et al.*, 2007, suggest that metric type is insignificant, and that the three $\sigma_{j(s)}$'s under consideration are very similar although the rest are not.)

For priors, we take θ_1 and θ_2 to be iid normal with mean 0 and variance 100, and σ_H^2 and $\sigma_{j(s)}^2$'s to be iid inverse-Gamma with shape and scale both equal to 1. Our choice of hyperparameters leads to relatively diffuse priors to reflect our ignorance of the modeled parameters. MCMC sampling from the resulting posterior was implemented with OpenBUGS (Thomas *et al.*, 2006); hierarchical centering was required for the implementation to reduce runtime (see Appendix I).

Based on two Markov chains, all of $\boldsymbol{H}, \boldsymbol{\beta}$, and $\boldsymbol{\Omega}$ were unambiguously estimated, except for non-EPT $\sigma_{j(s)}$'s, which exhibited minor convergence problems (see Appendix II). As no such problem was encountered for H_i's, we define LHFI-A as the mean of the H_i draws from both chains combined. Index values and corresponding 95% HPD intervals (Smith, 2007) appear in

gray in Figure 1. Posterior summary statistics for $\boldsymbol{\Omega}$ are given in Table 2. From Figure 2(a), we see no obvious violations against the linearity assumption of Model A, in light of sites MA1 and LB4 (shown as "M" and "L" in the figure) being potential outliers. Nor do we see the need to regress latent health on latitude due to the large scatter in Figure 2(b), regardless of those two sites.

4.3 *Model B*

The formulation of Model A is based entirely on binomial distributions associated with the metrics. However, one could fine-tune the dependence among $Y_{ij(s)k}$'s based on the disjoint nature of #Eph, #Ple, and #Tri that define a quadrinomial variate. To incorporate this into the LHFI model, we further break down the richness metrics into two subgroups by letting s=0 represent EPT richness metrics, and s=1 for the remaining three. The group of abundance metrics remains as s=2. Thus, each group consists of 3 metrics. As before, $Y_{ij(s)k}$'s are binomial for s=1, 2. However, for s=0,

$$\begin{aligned}\Big(Y_{i1(0)k}\,,\; Y_{i2(0)k}\,,\; Y_{i3(0)k}\,,\; T_{i\bullet 1k}-\textstyle\sum_{j=1}^{3}Y_{ij(0)k} \;\Big|\; T_{i\bullet 1k}\,,\; p_{i1(0)\bullet}\,,\; p_{i2(0)\bullet}\,,\; p_{i3(0)\bullet}\Big)\\ \sim \text{Multinomial}\big(\,T_{i\bullet 1k}\,;\; p_{i1(0)\bullet}\,,\; p_{i2(0)\bullet}\,,\; p_{i3(0)\bullet}\,,\; 1-\textstyle\sum_{j=1}^{3}p_{ij(0)\bullet}\,\big)\end{aligned} \tag{8}$$

where $p_{ij(0)\bullet}$ is the probability of an observed taxon from site i falling in the jth EPT category. Note that all 6 richness metrics share the same margin $T_{i\bullet 1k}$, irrespective of s=0 or 1. As large values of $p_{ij(0)\bullet}$'s are indicative of good health, we consider the multinomial-logit

$$\ln\frac{p_{ij(0)\bullet}}{1-\sum_{j=1}^{3}p_{ij(0)\bullet}} \equiv \nu_{ij(0)\bullet} = H_i + \beta_{j(0)}\,. \tag{9}$$

Altogether, Model B is binomial-multinomial mixture logit, comprising (8)–(9) for s=0, with $\sigma_{1(0)}$=$\sigma_{2(0)}$=$\sigma_{3(0)}$; (5)–(6) for s=1,2; and (7) for all s=0,1,2.

The prior for $\boldsymbol{\Omega}$, as well as the characteristics of the resulting two Markov chains of posterior draws (see Appendix II), are all as for Model A above. Again, we combine both chains to form LHFI-B. Index values and corresponding 95% HPD intervals are given in solid black in Figure 1. Posterior summary statistics for $\boldsymbol{\Omega}$ are in Table 2.

4.4 *Model C*

Finally, one might wish to consider as part of the model the dependency of the richness counts over sites and metrics, for the following reason. The nature of the dependence between pairs of richness counts is expected to vary by site and metric. Although random site selection is assumed, any given stream may yield multiple selected sites, so that Y_{ijk} and $Y_{i'jk}$ may be dependent. More obvious may be the dependence between Y_{ijk} and $Y_{ij'k}$ for, say, the jth metric being #Eph and the j'th, #Cl, as many *Eph.* taxa fall in the clinger category; the covariance structure between Y_{ijk} and $Y_{ij'k}$ differs among different pairs of metrics.

In fact, the dependence of pairwise covariance on (i, i') is already reflected by the latent regression in (2), and that having correlated β_j's can further account for the dependence on (j, j'); see Appendix III. In particular, we modify Model B by having $\boldsymbol{\beta} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ has jth

diagonal element σ_j^2 and off-diagonal (j, j')th element $\sigma_{jj'}$. In principle, one may wish to impose a covariance structure that is based on the conceptual relationship among metrics. However, except for some special structures that may be unrealistic, it is often challenging to efficiently sample from the posteriors of the covariance parameters (Westveld & Hoff, conditionally accepted). Thus, we assume the popular inverse-Wishart prior for an unstructured $\mathbf{\Sigma}$, with q degrees of freedom and scale matrix $\mathbb{S}$, so that $E(\mathbf{\Sigma}) \propto \mathbb{S}$. Hyperparameters are chosen to yield reasonably diffuse proper priors; we take q=9 (smallest possible) and $\mathbb{S}$ to have diagonal values 1 and positive off-diagonal values, arbitrarily set as 0.5. The latter reflects the prior notion that all 7 metrics are positively associated with latent health, and hence, perhaps with each other. Other hyperparameters for $\mathbf{\Omega}$ are as for Model B.

From two chains of posterior draws, we observe convergence problems for many $\mathbf{\Sigma}$ matrix entries (perhaps unsurprisingly; see Appendix II). Nonetheless, H_i chains mixed well marginally; hence, we define LHFI-C using the combined chain. Index values and 95% HPD intervals appear as dashed lines in Figure 1. Posterior summary statistics for selected $\mathbf{\Omega}$ elements are in Table 2.

5 DISCUSSION OF RESULTS

Pairwise correlations among the three LHFIs are all equal to 1.00. This suggests that Models A to C (variants of the same model) contain essentially the same information about the posterior mean for each H_i, which we define as the LHFI. This does not imply that the models yield the same inference, as we must also consider the reliability of the resulting health assessment. Figure 1 indicates that LHFI-C has substantially more uncertainty (longer HPD intervals) than A and B, which show almost identical properties but for a minor location shift. Thus, although Model C incorporates the natural correlation among metric values over sites and metrics into the LHFI model, the extra complexity of the model did little in practice to improve our inference for health given these data. Some remarks on this complexity appear in Appendix IV.

Note the extremes, "R" (RO1) and "T" (TH1), in Figure 2(a). Indeed, health conditions of Thornton Creek have been found to be so poor that even the untrained from surrounding communities have recognized and discussed the problems through local news and residents' forums. The opposite is true for Rock Creek. Our LHFI successfully reflects these extremes, and agrees with the B-IBI and SHIPSL in this regard. Figure 3 shows pairwise relationships among B-IBI, SHIPSL, and LHFI-B for all sites. (Recall that all three LHFIs are virtually perfectly correlated.) There is a strong positive correlation ($r \approx 0.9$) between our index and either existing one, but the relationship is slightly curvilinear. The curvature can be explained by the non-linearity of Models A to C, whereas both B-IBI and SHIPSL are linear combinations of metric scores (calibrated values). The strong correlation demonstrates that our LHFIs are no less informative about the sites' health conditions. And because of its highly quantitative nature, we therefore advocate the model-based LHFI as a comprehensive assessment of overall health.

The negative trend in Figure 2(a) also supports the common practice of habitat conservation by controlling urbanization. Indeed, all three models yield statistical evidence that urbanization impacts stream health negatively: 95% HPD intervals for θ_1 are below zero (approximately -3.4 to -0.4; see Table 2). While this negative effect might have been a foregone conclusion based on bio-

logical knowledge, our LHFI models provide direct quantitative evidence to support this biological notion. Such results have profound implications in practice. A policy maker may be presented with several factors that have potential impact on ecosystem health. Meanwhile, due to limited resources, s/he may be forced to devise conservation policies in response to selected factors only. To illustrate this, we can additionally include latitude in Model B; results are not discussed except for the inference on the latitude effect. The extra covariate has virtually no impact on the posterior distributions of H_i's (or of other unknown quantities). A typical 95% HPD interval for the corresponding coefficient includes 0, suggesting statistically insignificant effect on health due to latitude. (Generally, it is advisable to keep in the model any covariate that subject-matter experts have previously identified as influential to health. Such a model incorporates expert knowledge in an unambiguous fashion, and it undoubtedly provides a more comprehensive picture of the relationship among metrics, factors, and latent health, than if covariates were kept in the model merely for statistical significance.) Thus, our latent factor hierarchical modeling approach provides the policy maker with a scientific mechanism to classify factors according to their observed impact on health: an HPD interval that is negative indicates detrimental effects, one that covers 0 indicates undetectable impact, and a positive one implies positive impact. (The credible level may require adjusting in the context of multiple testing; e.g. see Westfall *et al.*, 1997) Since this mechanism accounts for uncertainty propagation, resulting conclusions are more realistic than from the traditional procedure of regressing a faunal-only MMI on covariates after index construction.

A quantitative comparison among models may be of interest also. A common basis of comparison is the deviance information criterion (DIC), which assesses how well the model can predict $Y_{ij(s)k}$'s. The DIC can be used to compare performance among models for identical data (Spiegelhalter *et al.*, 2002); smaller values are preferred. Our DIC values are taken from the OpenBUGS output, and shown in Table 2. (Theory behind the DIC is beyond the scope of our article.) Here, the DIC is 4651.0 for Model A and 4606.0 for Models B and C. Thus, posterior predictive power is apparently gained by accounting for the multinomial dependence among EPT metrics.

6 INTERPRETING THE LHFI IN THE ABSENCE OF PRESCRIBED REFERENCE CONDITIONS

As discussed in Section 1, a reference-based scheme typically suffers from non-transferability between geographical / temporal domains, and relies heavily on the availability of unstressed ecosystems, despite their rarity and/or the cost in locating and observing them. Hence, one may wish to consider no reference-based calibrations at all, but rely on a scheme of relative rating among several sites included in a single study. As a compromise for the lack of a full spectrum of health conditions, it may be useful to gauge health against a heavily degraded ecosystem which, sadly, is likely easier to locate than very healthy ones. When included in the study, a badly degraded site then serves as the baseline for "internal referencing," a concept originally proposed by Chiu & Guttorp (2006). Much in the same way that a one-way ANOVA assesses the *effectiveness* among several treatments relative to the least effective treatment, an inference-based comparison can rate sites according to their *health* (i.e. LHFI) relative to an unhealthy site; externally defined baseline or reference conditions are perhaps less crucial.

With our approach, ratings can be defined relative to the posterior distribution of health H_{worst} for the site identified (before or after fitting the LHFI model) as the worst degraded in the region. This method, as we now illustrate, is inspired by the comments from a reader of an article related to this paper. Let us take TH1 to be the worst degraded site in the PSL region. Thus, H_{worst}=H_{TH1}, and it acts as a baseline value for other sites. To assess Site BB1 situated along Big Bear Creek, a simple approach then is to compute a z-score for its LHFI value (posterior mean of H_{BB1}) relative to the posterior $P(H_{\text{TH1}}|\boldsymbol{Y}, \mathbb{X})$. Assuming Model B, we have

$$z_{\text{BB1}} = \frac{E(H_{\text{BB1}}|\boldsymbol{Y}, \mathbb{X}) - E(H_{\text{TH1}}|\boldsymbol{Y}, \mathbb{X})}{\sqrt{Var(H_{\text{TH1}}|\boldsymbol{Y}, \mathbb{X})}} = \frac{-0.788 - (-2.272)}{0.461} = 3.22\,.$$

A subject-matter expert may now translate z=3.22 back to practical terms, and decide on the overall degree of degradation for BB1. Note that the z-score is appropriate here, as $(H_{\text{TH1}}|\boldsymbol{Y}, \mathbb{X})$ is approximately normally distributed (not shown).

Ideally, the study would include healthier sites that are recognized as "nearly pristine." In this case, the above gauge could be replaced or used alongside its "mirror image," i.e. the same procedure but applied to the best site in the study. One could extend this principle further by using posterior quantiles for the best site in the study to define future unstressed sites. For instance, a new site may be added to the current study, and the LHFI model re-fitted. The earlier best site will now have an updated posterior distribution $P(\tilde{H}_{\text{best}}|\tilde{\boldsymbol{Y}}, \tilde{\mathbb{X}})$ due to the inclusion of the new site, where '$\sim$' indicates the update; but qualitatively, the site remains "nearly pristine." Now, one may declare that the new site is unstressed if its LHFI value falls above, say, the 90th percentile of $P(\tilde{H}_{\text{best}}|\tilde{\boldsymbol{Y}}, \tilde{\mathbb{X}})$. Similarly, the site could be labeled as "exceedingly degraded" if its index value falls below, say, the 10th percentile of $P(\tilde{H}_{\text{worst}}|\tilde{\boldsymbol{Y}}, \tilde{\mathbb{X}})$. Note that this approach is not restricted to new sites taken from the same spatial or temporal domain as the others, so long as the model from Section 2.2 is deemed sensible.

Gauging ecosystem health with this "internal referencing" scheme may reduce the disadvantages associated with externally defining reference conditions. Of course, a minor level of ambiguity is still inevitable despite our quantitatively oriented approach, such as in the percentile cut-offs used to define categories of health, which should be left to subject-matter experts to decide.

7 CONCLUSION

In this article, we described the methodology for constructing LHFIs and contrasted it with that for some existing health indices. We used the LHFI approach to assess the health of the 1997 PSL freshwater ecosystem, and compared it to existing B-IBI and SHIPSL measures. Major advantages stem from a modeling framework that allows proper inference for all crucial quantities, except for the exceedingly diffuse metric (co)variances, which are difficult to estimate with a finite number of MCMC draws. Of course, the same technique could lead to very different results and conclusions when applied to another dataset.

Irrespective of the data, LHFI modeling is systematic and unambiguous for ecosystem health studies. It attempts to retain the user-friendliness of conventional scalar health indices while over-

coming several hurdles not clearly addressed by the approaches taken to construct the latter. Relying on standard modeling principles, our approach exhibits the following desirable properties. (1) *Honest statistical inference of health and impact from environmental factors.* Building an LHFI for ecosystems involves virtually no qualitative procedures and deals directly with the raw metrics and associated covariates; hence, it can incorporate auxiliary information as part of the index in a systematic fashion, and allows proper assessment of uncertainty that trickles down the hierarchy of the multi-level relationship among variables under consideration. Thus, compared to common indices, ours provides more "honest" statistical inference for current and predicted site health. Alongside index construction is the ability to assess the significance of impact on health from observable factors. (2) *Versatility and adaptability to studies involving datasets that arise from multiple strata (geographical or otherwise and/or observed on different macroscopic scales).* We use a domain effect term in the LHFI model, a standard practice for scientific comparisons among strata, to address the age-old difficulty encountered in inter-regional and -temporal studies. This also avoids the complexity and impracticality of formal spatio-temporal models in biomonitoring studies. The same principle has been applied by Wu (2009) on non-spatio-temporal stratification. (3) *Less need for external reference conditions.* When truly unstressed conditions are unavailable, we propose "internal referencing" against those sites on either extreme of health that can be easily included in a study. Scientific comparisons via statistical modeling is universal, and constructing a health index as such is intended to achieve the same purpose as reference-based techniques, while avoiding some of the associated disadvantages.

The statistical principles used to construct the LHFI would not diminish the biological worthiness of the resulting index, as subject-matter expertise remains vital in variable selection and results interpretation before and after model fitting. In the terminology of Fjelland (2002) page 168, here statisticians play the role of non-experts in the "extended peer communities" of ecologists, and because they are naturally "closer to the problem" of developing quantitative methods, their contribution can only enhance the overall value of the methodology in scientific applications. Our technique has been embraced by some ecologists and quantitative scientists who work alongside them; studies that involve inter-regional and -temporal data to be converted to LHFIs are ongoing. Other directions of LHFI research appear in Appendix IV.

Our approach is rooted in a simple statistical concept of ANOCOVA model building, and may be easily adapted to any context of health assessment, be it ecological, medical, or otherwise. Once a list of relevant observable variables has been identified, constructing an LHFI reduces to forming a statistical model that efficiently describes the relationship among these variables and latent health. Some variables may be explanatory to health, and vice versa for others. Although latent variable modeling techniques have become widely popular in many sciences, its use to yield a direct quantitative "report card" composite measure of health in a general sense is apparently uncommon. Our proposed methodology is a simple but universal and versatile approach that is potentially valuable to many scientific disciplines in which a scalar assessment of health is desirable.

ACKNOWLEDGMENTS

This research is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Discovery and Strategic Project Grants to G. Chiu, the latter grant sub-contracted from Prof. Jon Grant, Department of

Oceanography, Dalhousie University. We thank the Editor, Associate Editor, and referee for their invaluable comments; Dr. John van Sickle, Environmental Protection Agency Western Ecology Division, for sharing his ideas and providing some of his publications and data as reference; and various readers of our papers on this research for their helpful suggestions. The corresponding author thanks various research scientists and statisticians of CSIRO Mathematics, Informatics and Statistics, for sharing their experience with ecological health indices and embracing the method proposed in this paper; Professor Michael Dowd, Department of Mathematics and Statistics, Dalhousie University, for stimulating discussions on MCMC techniques and post-normal science; and Professor Jennifer Hoeting, Department of Statistics, Colorado State University, for her suggestions and moral support.

References

Brooks S, Gelman, A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**: 434–455.

The CEH Web. RIVPACS – reference condition. http://www.ceh.ac.uk/products/software/RIVPACS-ReferenceCondition_000.html [4 November 2009]

Chao A. Species estimation and applications. In *Encyclopedia of Statistical Sciences*, Kotz S, Read CB, Balakrishnan N, Vidakovic B (eds). Wiley InterScience: Hoboken, 2006; Online service. DOI: 10.1002/0471667196.ess5051

Chiu GS. On identifiability of covariance components in hierarchical generalized analysis of covariance models. Working Paper 2008-09, Department of Statistics and Actuarial Science, University of Waterloo, 2008.

Chiu G, Guttorp P. New developments involving the stream health index for the Puget Sound Lowland. Technical Report 079, Northwest Research Center for Statistics and the Environment, University of Washington, 2004.

Chiu, G, Guttorp P. Stream health index for the Puget Sound Lowland. *Environmetrics* 2006; **17**: 285–307.

Chiu GS, Guttorp P, Khan SA, Liang J, Westveld AH. An ecological latent health factor index via a random-effects model for taxa richness and composition. Working Paper 2006-02, Department of Statistics and Actuarial Science, University of Waterloo, 2007.

Chiu GS, Guttorp P, Westveld AH, Khan SA, Liang J. A latent health factor index modelling approach via generalized linear mixed models, with application to ecological health assessment. Working Paper 2008-08, Department of Statistics and Actuarial Science, University of Waterloo, 2008.

Chiu GS, Wu MA, Lu L, Grant J. A latent health factor index for the Richibucto estuarine ecosystem. In progress.

Cuffney TF, Gurtz ME, Meador MR. Methods for collecting benthic invertebrate samples as part of the national water-quality assessment program. Open file report 93-406, United States Geologic Survey, Washington, D.C., 1993.

Ferraro SP, Cole FA, Olsen AR. A more cost-effect EMAP benthic macrofaunal sampling protocol. *Environmental Monitoring and Assessment* 2006; **116**: 275–290.

Fjelland R. Facing the problem of uncertainty. *Journal of Agricultural and Environmental Ethics* 2002; **15**: 155–169.

Gelfand AE, Sahu SK, Carlin BP. Efficient parametrisation for normal linear mixed models. *Biometrika* 1995; **82**: 479–488.

Gelman A, Hill J. *Data Analysis Using Regression and Multilevel / Hierarchical Models*. Cambridge University Press: New York, 2007.

Hawkins PH, Norris RH, Hogue JN, Feminella JW. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 2000; **10**: 1456–1477.

Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st Century. *Medical Care* 2000; **38**: II-28–II-42.

Higgs MD, Hoeting JA. A clipped latent variable model for spatially correlated ordered categorical data. *Computational Statistics and Data Analysis*.

Kennard MJ, Harch BD, Pusey BJ, Arthington AH. Accurately defining the reference condition for summary biotic metrics: a comparison of four approaches. *Hydrobiologia* 2006; **572**: 151–170.

Kerans BL, Karr JR. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications* 1994; **4**: 768–785.

López-Alvarenga JC, Montesinos-Cabrera RA, Velázquez-Alva C, González-Barranco J. Short stature is related to high body fat composition despite body mass index in a Mexican population. *Archives of Medical Research* 2003; **34**: 137–140.

MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. *Occupancy Estimation and Modeling*. Elsevier: Amsterdam, 2006.

Marshall JC, Steward AL, Harch, BD. Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an Australian dryland river: the benefits and costs of using species abundance data. *Hydrobiologia* 2006; **572**: 171–194.

Martin AD, Quinn KM. Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis* 2002; **10**: 134–153.

McCormick FH, Hughes RM, Kaufmann PR, Peck DV, Stoddard JL. Development of an index of biotic integrity for the Mid-Atlantic Highlands region. *Transactions of the American Fisheries Society* 2001; **130**: 857–877.

Morley SA. *Effects of urbanization on the biological integrity of Puget Sound Lowland streams: restoration with a biological focus*. M.S. Thesis, University of Washington, 2001.

Moss AJ, Smith MJ, Harch BD, Bunn SE, Storey AW, Kennard MJ, Marshall CJ. Ecosystem health guidelines for rivers and streams in South East Queensland. In *Design and Implementation of Baseline Monitoring (DIBM3)*. Brisbane: South East Queensland Healthy Waterways Partnership Ecosystem Health Monitoring Program, 2001.

Pietrobon R, Taylor M, Guller U, Higgins LD, Jacobs DO, Carey T. Predicting gender differences as latent variables: summed scores, and individual item responses: a methods case study. *Health and Quality of Life Outcomes* 2004; **2**: 59.

Rabe-Hesketh S, Skrondal A. Classical latent variable models for medical research. *Statistical Methods in Medical Research* 2008; **17**: 5–32.

Rosas G. Dynamic latent trait models: an application to Latin American banking crises. *Electoral Studies* 2009; **28**: 375–387.

Skrondal A, Rabe-Hesketh S. Latent variable modelling. *Statistical Methods in Medical Research* 2008; **17**: 3–4.

Smith BJ. An R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software* 2007; **21**: Online service.

South East Queensland Healthy Waterways Partnership. http://www.healthywaterways.org/EcosystemHealthMonitoringProgram/ProductsandPublications/AnnualReportCards.aspx [4 November 2009]

Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. *Statistical Methodology* 2002; **64**: 583–639.

Steedman RJ, Regier HA. Ecological bases for an understanding of ecosystem integrity in the Great Lakes Basin. In *An Ecosystem Approach to the Integrity of the Great Lakes in Turbulent Times* Special Publication No. 90-4, Edwards CJ, Regier HA (eds). Great Lakes Fishery Commission: Ann Arbor, 1990; 257–270.

Stock JH, Watson MW. New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual* 1989; **4**: 351–394.

Stoddard JL, Peck DV, Olsen AR, Larsen DP, van Sickle J, Hawkins CP, Hughes RM, Whittier TR, Lomnicky G, Herlihy AT, Kaufmann PR, Peterson SA, Ringold PL, Paulsen SG, Blair R. Western Streams and Rivers Statistical Summary. EPA Publication 620/R-05/006, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C., 2005.

Suter GW. A critique of ecosystem health concepts and indexes. *Environmental Toxicology and Chemistry* 1993; **12**: 1533–1539.

Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS open. In *R News* Vol. 6/1, Plummer M, Murrell P (eds). R Foundation for Statistical Computing: Vienna, 2006; 12–17.

Ward MD, Hoff PD. Persistent patterns of international commerce. *Journal of Peace Research* 2007; **44**: 157–175.

Westfall PH, Johnson WO, Utts JM. A Bayesian perspective on the Bonferroni adjustment. *Biometrika* 1997; **84**: 419–427.

Westveld AH, Hoff PD. A Bayesian mixed effects model for longitudinal social network data. *Annals of Applied Statistics*.

Wu MAC. *A latent health factor model for estimating estuarine ecosystem health.* M.Math. Thesis, University of Waterloo, 2009.

APPENDIX I: HIERARCHICAL CENTERING

To minimize MCMC mixing problems, we employ partial hierarchical centering (Gelfand *et al.*, 1995) to reformulate parts of each model before implementation; see Appendix in Chiu *et al.* (2007) for the full rationale. In general, we must explore several formulations of partial hierarchical centering to identify one under which to obtain posterior samples efficiently. For example, the following formulation of the relevant parts of Model A performs satisfactorily in our study:

$$(\tilde{H}_i|\theta_1,\sigma_H,\boldsymbol{x}) \sim N(\theta_1(x_i-\overline{x}),\sigma_H^2)\,, \quad \left(b_{j(s)}|\theta_0,\sigma_{j(s)}\right) \overset{ind}{\sim} N(\theta_0,\sigma_{j(s)}^2)\,,$$
$$\nu_{isj\bullet} = \tilde{H}_i + b_{j(s)}\,, \quad H_i = \theta_0 + \tilde{H}_i\,, \quad \beta_{j(s)} = b_{j(s)} - \theta_0\,.$$

APPENDIX II: MCMC DETAILS

For each model, two independently generated Markov chains of posterior samples form the basis

of our posterior inference. The chains started at different random initial values. Each chain consisted of 10,000 draws, reduced from removing a burn-in of b draws then thinning by a lag of ℓ. For each of Models A and B, b=5,000 and ℓ=5; what appeared to be mixing problems were encountered only for the non-EPT $\sigma_{j(s)}$'s, in the form of noticeably different chain dispersions with tail values on the order of 10^3; however, marginal posterior medians were comparable between chains. Nor did the Brooks-Gelman-Rubin convergence diagnostic plots (Brooks & Gelman, 1998) exhibit patterns that would cause much concern. Indeed, it is understandable that extensive coverage of the support of an exceedingly diffuse posterior may require an impractically large number of simulated draws. Chiu (2008) also observes that for the PSL data, diffuseness of the posterior is an indication of weak identifiability, likely a direct result of the diffuseness of the prior. In light of how readily the marginal chains for the other parameters converged, and how similar the non-EPT $\sigma_{j(s)}$ medians were between chains, our minor mixing problems do not appear particularly problematic.

Similar characteristics were observed for Model C, for which b=250,000 and ℓ=50. After the substantial burn-in, chains for all parameters but $\boldsymbol{\Sigma}$ converged exceptionally quickly. Posterior draws for matrix entries of $\boldsymbol{\Sigma}$ were on the order of 10^6, again with substantially different dispersion between chains, although chain medians were highly comparable. For the purpose of estimating the problematic (co)variance parameters for these models, we restrict our attention in each case to the chain that yielded a larger posterior variability for the parameters in question, as the smaller variability for the rejected chain could have resulted from an initial value that confined the Markov chain to a smaller subset of the parameter space.

APPENDIX III: RESPONSE CORRELATION OVER SITES AND OVER METRICS

We show that having (a) health regressed on covariates and (b) correlated metric effects in (2) can address the natural correlation among metric values over sites and over metrics.

First, consider the potential shortcomings of Models A and B which assume independence among metric effects. We do so through a Gaussian analog of the models (with dependence on s dropped from the notation to reduce clutter):

$$W_{ijk} \equiv \ln Y_{ijk} = H_i + \beta_j + \varepsilon_{ijk}\,, \qquad (\varepsilon_{ijk}|\sigma_\varepsilon) \overset{iid}{\sim} N(0,\sigma_\varepsilon^2)\,, \tag{10}$$

$$(H_i|\theta_0,\theta_1,\boldsymbol{x},\sigma_H) \overset{ind}{\sim} N\left(\theta_0+\theta_1(x_i-\overline{x}),\sigma_H^2\right)\,, \qquad (\beta_j|\sigma_j) \overset{ind}{\sim} N(0,\sigma_j^2)\,. \tag{11}$$

Conditioned on $\boldsymbol{\Omega}$=$(\boldsymbol{\theta},\sigma_H,\sigma_1,\ldots,\sigma_7)$, the mean and covariance structures of the data coming from sites (i,i') and from metrics (j,j') are

$$E(W_{i'jk}|\boldsymbol{\Omega}) = E(H_{i'}+\beta_j+\varepsilon_{i'jk}|\boldsymbol{\Omega}) = \theta_0+\theta_1(x_{i'}-\overline{x})\,, \tag{12}$$

$$E(W_{ijk}|\boldsymbol{\Omega}) = E(H_i+\beta_j+\varepsilon_{ijk}|\boldsymbol{\Omega}) = \theta_0+\theta_1(x_i-\overline{x}) = E(W_{ij'k}|\boldsymbol{\Omega})\,, \tag{13}$$

$$\text{Cov}(W_{ijk},W_{i'jk}|\boldsymbol{\Omega}) = \text{Cov}(H_i+\beta_j+\varepsilon_{ijk},\ H_{i'}+\beta_j+\varepsilon_{i'jk}|\boldsymbol{\Omega}) = \text{Var}(\beta_j|\sigma_j) = \sigma_j^2\,,$$

$$\text{Cov}(W_{ijk},W_{ij'k}|\boldsymbol{\Omega}) = \text{Cov}(H_i+\beta_j+\varepsilon_{ijk},\ H_i+\beta_{j'}+\varepsilon_{ij'k}|\boldsymbol{\Omega}) = \text{Var}(H_i|\sigma_H) = \sigma_H^2\,.$$

Now, take the priors from the models. Then, by the law of total covariance, one can easily show

that the marginal covariances become

$$\begin{aligned} \mathrm{Cov}(W_{ijk}, W_{ij'k}) &= \psi + c[1 + (x_i - \overline{x})^2]\,, \quad (14) \\ \mathrm{Cov}(W_{ijk}, W_{i'jk}) &= \psi + c[1 + (x_i - \overline{x})(x_{i'} - \overline{x})] \quad (15) \end{aligned}$$

where c is the hypervariance of the normal prior, and ψ depends on the inverse-Gamma hyperparameters only. Thus, given site i, (14) implies that the correlation between (the log-values of) any pair of metrics is constant over metrics (i.e. independent of (j, j')). However, as discussed in Section 4.4, metric values could be naturally correlated over metrics and over sites. Conveniently, dependency over sites is addressed by regressing latent health on site-specific covariates according to (15): given metric j, the correlation of metric values between any pair of sites depends on (i, i'). However, this dependence would have been lost should the dependence on $\boldsymbol{x}$ be removed from (11), leaving (12), (13), (14), and (15) simply as

$$\begin{aligned} E(W_{ijk}|\boldsymbol{\Omega}) = E(W_{i'jk}|\boldsymbol{\Omega}) = E(W_{ij'k}|\boldsymbol{\Omega}) = \theta_0\,, \\ \mathrm{Cov}(W_{ijk}, W_{i'jk}) = \mathrm{Cov}(W_{ijk}, W_{ij'k}) = \psi + c\,. \end{aligned}$$

Just as the latent regression introduces correlation over sites, dependence among metric effects β_j's conveniently incorporates correlation over metrics into the model, by replacing independent β_j's in (11) with $\boldsymbol{\beta} \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$. Adding this to the latent regression turns (14) and (15) into

$$\begin{aligned} \mathrm{Cov}(W_{ijk}, W_{i'jk}) &= c\,[1 + (x_i - \overline{x})(x_{i'} - \overline{x})] + E(\sigma_j^2)\,, \\ \mathrm{Cov}(W_{ijk}, W_{ij'k}) &= c\left[1 + (x_i - \overline{x})^2\right] + \psi + E(\sigma_{jj'})\,. \end{aligned}$$

The hyperparameter $\mathbb{S}$ in the inverse-Wishart prior for Model C can be specified such that $E(\sigma_j^2)$ and $E(\sigma_{jj'})$ — and hence, the covariances — depend on j and (j, j'), respectively. For the PSL data, various such priors were employed, but they all led to virtually identical estimates. As it turns out, we find little evidence from these data that β_j's are correlated.

APPENDIX IV: REMARKS

One may wonder if proper estimation of $\boldsymbol{\Sigma}$ could be an issue for Model C, since no replication for any given j appears at the level of $P(\boldsymbol{\beta}|\boldsymbol{\Omega})$ alone; to the contrary, the pooling of information from other random quantities in the model aids inference for $\boldsymbol{\Sigma}$. To see this, again consider the Gaussian analog from (10). Letting δ_{ijk}=β_j+ε_{ijk}, we have $([\delta_{i1k}, \delta_{i2k}, \ldots, \delta_{iJk}]^T|\boldsymbol{\Sigma}, \sigma_\varepsilon^2) \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma}$+$\sigma_\varepsilon^2\mathbb{I})$, where $\mathbb{I}$ is the identity matrix. Thus, replication of each j exists over (i, k) for the estimation of $\sigma_{jj'}$'s. Indeed, Chiu (2008) shows that for the same PSL data analyzed by Poisson counterparts of our three models (involving the 7 count metrics only), insightful inference is in no way hindered even by diffuseness of (proper) priors, as substantial *Bayesian learning* is achieved for $\boldsymbol{\Sigma}$. Those conclusions are expected to extend to our logistic models which are structurally identical to the Poisson counterparts.

Also, note that we formulate our logistic LHFI models with $T_{i\bullet sk}$'s, which are not population

quantities but of the samples. The level of richness and abundance in the species community (i.e. population) that can be supported by the habitat is highly relevant to the notion of ecological health. Thus, a remaining statistical issue is that of *species (occupancy) estimation* (see Chao, 2006; MacKenzie *et al.*, 2006). However, standard sampling protocols for benthic fauna utilize the *species accumulation curve* (see Chao, 2006), constructed based on data often accrued over many previous studies, to design how to obtain benthic samples that best reflect conditions of the entire community. Of course, standard protocols are not foolproof (Ferraro *et al.*, 2006). A new direction of LHFI research could investigate the circumstances under which occupancy estimation should play a role in index construction.

Finally, although not demonstrated for these data, using any of Models A to C to make prediction inference for site health (Section 3.1) would be straightforward for a PSL site that is not accompanied by benthic faunal data but by an urbanization value.

Table 1: Sites sampled from the PSL in 1997, and metrics identified in ecological studies to be effective indicators of PSL stream health

site		*metric*		
name	*location*	*label*	*characteristic*	*type*
BB1	Big Bear Creek	#Tx	all taxa	richness* (count)
BB2		#Eph	*Ephemeroptera* taxa	richness
BB3		#Ple	*Plecoptera* taxa	richness
BB4		#Tri	*Trichoptera* taxa	richness
BB5		#LL	long-lived taxa	richness
BS1	Big Soos Creek	#Intol	intolerant taxa	richness
JE1	Jenkins Creek	#Cl	clinger taxa	richness
LB1	Little Bear Creek	%Tol	tolerant taxa	abundance† (%)
LB2		%Pred	predatory taxa	abundance
LB3		%Dom3	3 most dominant taxa	abundance
LB4				
MA1	May Creek			
MI1	Miller Creek	*# distinct taxa of given characteristic appearing in field sample		
RO1	Rock Creek			
SW1	Swamp Creek	$^{\dagger}\frac{100 \times \#\text{ animals of given characteristic in field sample}}{\text{total \# animals in field sample}}$		
SW2				
SW3				
TH1	Thornton Creek			

Table 2: Selected summary statistics of posterior draws. Values for parameters with a '*' are based on one Markov chain only

	mean	*median*	*2.5th HPD %-ile*	*97.5th HPD %-ile*	*MC error*	*# draws*
			Model A: DIC=4651.0			
θ_0	−1.62	−1.63	−2.51	−0.68	0.01	20 000
θ_1	−2.03	−2.03	−3.59	−0.43	0.01	
$\sigma_{1(1)}, \sigma_{2(1)}, \sigma_{3(1)}$	0.87	0.58	0.10	2.36	0.01	
${\sigma_{4(1)}}^*$	2.31	1.03	0.12	7.21	0.08	10 000
${\sigma_{5(1)}}^*$	12.92	5.50	0.64	37.76	0.56	
${\sigma_{6(1)}}^*$	4.35	1.92	0.16	13.91	0.18	
${\sigma_{1(2)}}^*$	10.11	4.30	0.37	28.62	0.77	
${\sigma_{2(2)}}^*$	4.30	1.85	0.15	12.99	0.17	
${\sigma_{3(2)}}^*$	3.12	1.31	0.12	9.16	0.20	
σ_H	0.58	0.56	0.40	0.79	0.00	20 000
			Model B: DIC=4606.0			
θ_0	−1.07	−1.06	−1.97	−0.20	0.01	20 000
θ_1	−2.08	−2.09	−3.64	−0.49	0.00	
$\sigma_{1(0)}, \sigma_{2(0)}, \sigma_{3(0)}$	0.86	0.57	0.10	2.37	0.01	
${\sigma_{1(1)}}^*$	3.22	1.32	0.13	9.36	0.15	10 000
${\sigma_{2(1)}}^*$	17.41	7.02	0.78	47.77	0.85	
${\sigma_{3(1)}}^*$	3.28	1.38	0.14	9.42	0.15	
${\sigma_{1(2)}}^*$	7.55	3.22	0.25	22.43	0.39	
${\sigma_{2(2)}}^*$	5.85	2.58	0.20	18.33	0.16	
${\sigma_{3(2)}}^*$	2.71	1.00	0.12	7.08	0.22	
σ_H	0.58	0.57	0.40	0.79	0.00	20 000
			Model C: DIC=4606.0			
θ_0	−1.40	−1.40	−5.42	2.57	0.02	20 000
θ_1	−2.09	−2.09	−3.67	−0.52	0.01	
σ_1^{2*}	47.23	2.50	0.07	53.10	0.25	10 000
σ_2^{2*}	196.5	9.52	0.11	168.0	0.01	
σ_3^{2*}	251.7	3.12	0.07	64.60	0.02	
σ_4^{2*}	576.4	6.07	0.08	110.5	0.05	
σ_5^{2*}	53.84	3.63	0.07	79.68	0.24	
σ_6^{2*}	179.8	2.53	0.08	53.04	0.02	
σ_7^{2*}	178.7	2.34	0.07	49.60	0.02	
σ_H	0.58	0.56	0.39	0.78	0.00	20 000

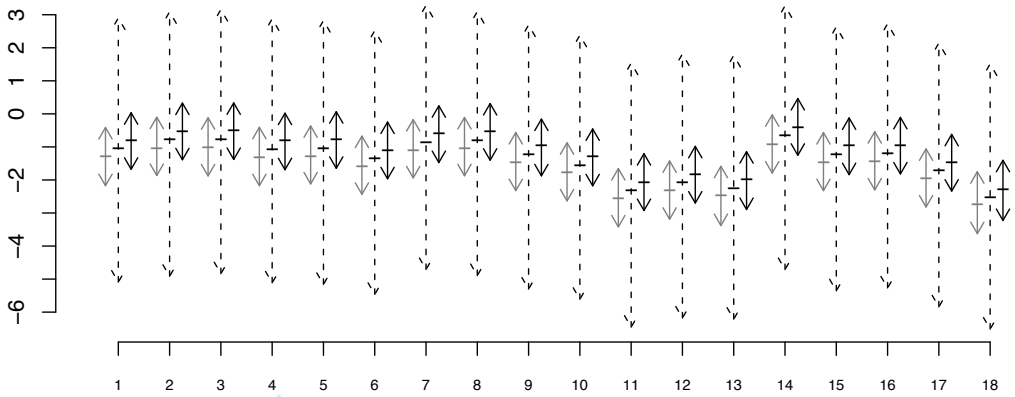

Figure 1: Ninety-five percent HPD intervals for LHFIs A (gray), B (solid black), and C (dashed); a '−' denotes the index value

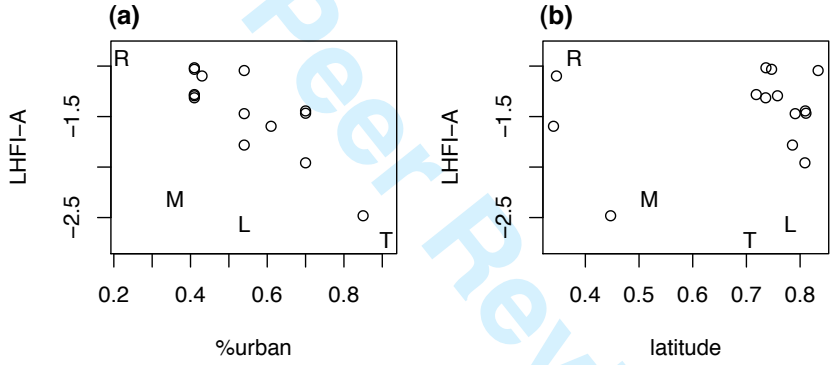

Figure 2: The relationship between covariates and the estimated latent health from Model A (latitudes shown are shifted by −47); points labeled 'L', 'M', 'R', and 'T' correspond to sites LB4, MA1, RO1, and TH1, respectively

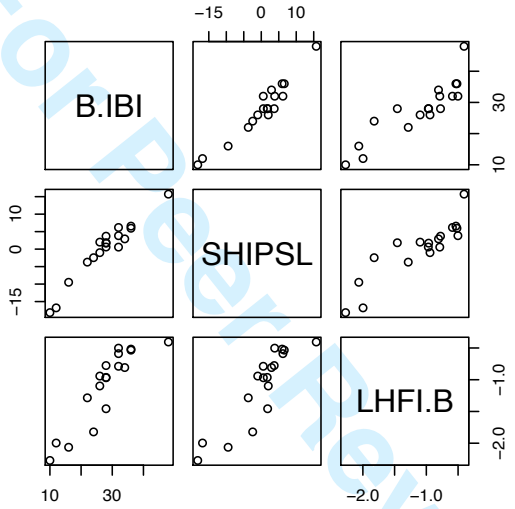

Figure 3: Scatterplots among various health indices for the 1997 PSL data