Open Access Articles

Combining and aggregating environmental data for status and trend assessments: challenges and approaches

Combining and aggregating environmental data for status and trend assessments: challenges and approaches

Kathleen G. Maas-Hebner · Michael J. Harte ·
Nancy Molina · Robert M. Hughes · Carl Schreck ·
J. Alan Yeakley

Abstract Increasingly, natural resource management agencies and nongovernmental organizations are sharing monitoring data across geographic and jurisdictional boundaries. Doing so improves their abilities to assess local-, regional-, and landscape-level environmental conditions, particularly status and trends, and to improve their ability to make short- and long-term management decisions. Status monitoring assesses the current condition of a population or environmental condition across an area. Monitoring for trends aims at monitoring changes in populations or environmental condition through time. We wrote this paper to inform agency and nongovernmental organization managers, analysts, and consultants regarding the kinds of environmental data that can be combined with suitable techniques and statistically aggregated for new assessments. By doing so, they can increase the (1) use of available data and (2) the validity and reliability of the assessments. Increased awareness of the difficulties inherent in combining and aggregating data for local- and regional-level analyses can increase the likelihood that future monitoring efforts will be modified and/or planned to accommodate data from multiple sources.

Keywords Data aggregation · Lurking variable ·
Simpson's paradox · Modifiable areal unit problem ·
Change of support problem · Environmental monitoring

K. G. Maas-Hebner (✉)
Department of Fisheries and Wildlife, Oregon State
University, Corvallis, OR 97331, USA
e-mail: kathleen.maashebner@oregonstate.edu

M. J. Harte
College of Earth, Ocean and Atmospheric Sciences, Oregon
State University, Corvallis, OR 97331, USA

N. Molina
Cascadia Ecosystems, 620 SE 14th Court, Gresham, OR
97080, USA

R. M. Hughes
Amnis Opes Institute and Department of Fisheries and
Wildlife, Oregon State University, 2895 SE Glenn, Corvallis,
OR 97333, USA
e-mail: hughes.bob@amnisopes.com

C. Schreck
Oregon Cooperative Fish & Wildlife Research Unit, US
Geological Survey, Oregon State University, Corvallis, OR
97331, USA

J. A. Yeakley
Department of Environmental Science & Management,
Portland State University, PO Box 751, Portland, OR 97207,
USA

Introduction

Status and trend monitoring is often conducted by natural resource and land management agencies and organizations. Status monitoring characterizes the current condition of a species or an environmental condition, whereas trend monitoring aims to assess changes in a species or condition over time (Roni 2005; Olsen and Peck 2008). Natural resource managers, policy makers, and scientists may also pool monitoring data across geographic and jurisdictional boundaries to increase

the cost-effectiveness of environmental assessments and to help inform management decisions at ecologically relevant scales. However, environmental data are frequently collected in localized or spatially discontinuous patterns or gathered in surveys targeted at a limited set of objectives and cover only a portion of the region or population of interest. Inevitably, new questions arise and it becomes expedient to combine datasets that contain different variables or to assemble data from spatially disconnected studies to address more regionalized questions. In the Pacific Northwest of the USA, several organizations are creating standardized field protocols and centralized databases to increase the capacity of management agencies and nongovernmental organizations to share and integrate data into regional status and trend assessments (e.g., Northwest Environmental Data Network 2005; Mulvey et al. 2009; Pacific Northwest Aquatic Monitoring Partnership 2014; StreamNet 2014). Combining shared data creates challenges that need to be carefully considered and addressed to produce meaningful and valid assessments.

In 1997, the State of Oregon established the Independent Multidisciplinary Science Team (IMST) to provide independent, rigorous scientific review of the Oregon Plan for Salmon and Watersheds (Oregon Plan 1997) and other issues related to the management of Oregon's native fish and watersheds (Oregon Revised Statute 541.914). Over the following 18 years, the IMST reviewed multiple salmonid conservation and recovery plans, revised water quality standards, habitat conservation plans and environmental impact statements, and monitoring programs (http://www.fsl.orst.edu/imst/index.html). In the course of those reviews, the IMST identified several areas where state and federal agencies could better integrate monitoring programs, share data, and increase their capacity to track environmental status and trends at state and regional levels (e.g., IMST 2010, 2011a, b, 2013). Our goal for this paper is to inform natural resource managers, analysts, and consultants about the kinds of data that can be combined and statistically aggregated for new assessments to increase the use of available data and to increase the validity and reliability of assessments. We do so in four sections. First, we briefly review how objectives should be defined and key statistical elements addressed before any data are combined or aggregated. Second, we describe potential issues that analysts may encounter when combining datasets and basic techniques to combine data from disparate sources. Third, we

discuss issues related to statistical aggregation including potential consequences of improper aggregation. Fourth, we recommend measures for improving our capabilities to survey environmental status and trends and give some current examples of such survey programs for aquatic ecosystems. This paper is not a comprehensive guide for combining, analyzing, or interpreting data; rather, it should serve as a starting point for thoughtful discussions and considerations regarding survey planning and procedures.

Objectives, target populations, and sampling frames

Any environmental assessment that is undertaken, whether or not it includes combining data from disparate sources, needs to have well articulated and achievable objectives. This is a critical step, just as it is when a monitoring program is created or revised. Similarly, survey statisticians should be included in all phases of assessment planning and analysis, including objective development (Reynolds 2012). Objectives should specify the target population (and subpopulations if applicable), spatial domain, time frame, population elements, and sampling frame. The target population may be a stream network, species, or forest type. Population elements make up the population (e.g., stream segment, age class or cohort, watershed of a specific size). The sampling frame specifies from where samples are to be drawn from with respect to the target population, which may or may not be the same as the target population. The population elements and sampling frame determine if a sample is part of the target population. The survey design(s) used to sample the target population will determine the basis from which conclusions or inferences can be drawn and which techniques are appropriate for combing datasets. Objectives should also specify the attributes of concern, sampling protocol, and units of measure for each variable (Hughes and Peck 2008) and, for trends analysis, the type of change being investigated.

Once the statistical population, sampling frame, and objectives have been clearly articulated, there are several other issues related to using or combining data that need to be considered. These issues include data credibility and reliability (Evans et al. 2001; Canfield et al. 2002; Hanson 2006) and data inconsistencies over time, and, among observers (Darwall and Dulvy 1996;

Rieman et al. 1999), noncomparability of data (Boyce et al. 2006; Roper et al. 2010), insufficient sample sizes (Gouveia et al. 2004), differences in sampling effort (Cao et al. 2002; Fayram et al. 2005; Smith and Jones 2008), data completeness (e.g., low sampling frequency and short time-frames; Rieman et al. 1999; Gouveia et al. 2004), and incomplete spatial and/or temporal coverage of data (Goffredo et al. 2004; Smith and Michels 2006). Identifying and resolving these issues will be possible if detailed metadata records exist for all datasets.

Potential issues encountered when combining datasets

Four statistical issues may arise that the analyst should be aware of when working with any environmental dataset: pseudoreplication, spatial autocorrelation, cross-scale correlation, and lurking variables. We describe these briefly below.

Generally, the more replicates used, the greater the statistical precision of the resulting data analysis. However, the lack of sample independence or the lack of true replicates can lead to *pseudoreplication*, a common error associated with ecological studies (Hurlbert 1984; Heffner et al. 1996; Millar and Anderson 2004). When samples are pseudoreplicated, such as those from sites with naturally different ecological potentials or those from sites whose locations may affect observations at other sites, the natural random variation exhibited by a variable is not properly quantified (Millar and Anderson 2004). Pseudoreplicated sample sizes appear higher than they truly are, giving the illusion of greater statistical power than what actually exists. Consequently, inferential statistics must be used with great care because most tests are designed for samples of independent observations. Inaccuracies are typically manifested in biased standard errors that misrepresent (typically by underestimating) the variation in the data and artificially inflate the significance of statistical comparisons. This greatly increases the chance of reaching conclusions of significance for phenomena that only happened by random chance. Where psuedoreplication exists, it may be possible to use a linear effects model to analyze the data by separating the different sources of variability, which will generate correct inferences from the data (Chaves 2010).

Spatial autocorrelation occurs when measurements taken at sites in close proximity exhibit values more similar to one another than would be expected if variation was distributed randomly across space or through time. In other words, the value of a measurement depends on, or can be predicted from, values measured at nearby sites, which often may be the case in ecological studies (e.g., Van Sickle and Hughes 2000; Herlihy et al. 2006; Pinto et al. 2009). Therefore, one cannot assume that samples are independent (Bataineh et al. 2006). Spatial autocorrelation can result from characteristics inherent in a species' growth or ecology (e.g., clonal growth, conspecific attraction), its distribution (e.g., Ficetola et al. 2012), or other external factors (e.g., the tendency for some environmental disturbances to be correlated with vegetation patterns; Lichstein et al. 2002). Fortin et al. (1989) and Lichstein et al. (2002) described methods for identifying and overcoming spatial autocorrelation in ecological analyses, and Dormann et al. (2007) provided methods to address spatial autocorrelation in species distribution.

Cross-scale correlation (i.e., correlations between habitat variables measured at different spatial scales) has been documented by researchers pursuing multiscale habitat relationship studies (Battin and Lawler 2006; Kautza and Sullivan 2012; Marzin et al. 2013). Where cross-scale correlations exist, erroneous conclusions may be drawn about the strength of relationships among predictor and response variables measured at a particular spatial scale (Battin and Lawler 2006; Lawler and Edwards 2006). For example, Marzin et al. (2013) reported that fish and macroinvertebrate assemblages were related to poor water quality and impoundment at the stream reach scale, but at the catchment scale, assemblages were related to a gradient from forest to agricultural covers. Battin and Lawler (2006) reviewed statistical techniques for detecting cross-scale correlations among variables measured at different spatial scales. Lawler and Edwards (2006) demonstrated how variance decomposition (Whittaker 1984) can be used as a diagnostic tool for revealing the amount of variation in a variable of interest explained by habitat variables measured at different spatial scales. Several studies have shown how site- and catchment-scale predictor variables, as well as their shared variance, explain differing amounts of biological response variables (e.g., Sály et al. 2011; Marzin et al. 2013; Macedo et al. 2014).

Lastly, the association between two or more variables can be induced, masked, or modified by the presence of an unknown or *lurking variable* (Sandel and Smith 2009). Spatial pattern, spatial scale, historic human interventions, and abiotic conditions can be common lurking variables that account for variation in environmental and ecological data. In aquatic assemblage data, the size of the water body and geographic location from which samples are drawn can have enormous implications for results (Hughes and Peck 2008), but calibration techniques are available (e.g., McGarvey and Hughes 2008; Pont et al. 2009; Terra et al. 2013b). Historical human interventions (e.g., removal of native vegetation, imposition of road networks, channelization, mines, dams) can create long-lasting legacies that affect the present conditions of stream geomorphology, water quality, and aquatic habitat (Frissell and Bayles 1996; Harding et al. 1998; Walter and Merritts 2008; Brown et al. 2009). Year-to-year and seasonal variability may also confound aggregation results. If not adequately accounted for in sampling designs and analyses, a lurking variable can be problematic when assessing the effects of management actions on environmental conditions.

Methods available to combine data from multiple sources

Because the statistical methods needed to combine environmental information from different sources will require case-specific formulations (Cox and Piegorsch 1994), this section should not be viewed as an all-inclusive guide to techniques but rather an illustration of approaches that can be used. Olsen et al. (1999) cautioned that if studies were designed without the anticipation of combining additional data, some of these techniques might not be feasible. To successfully merge datasets and identify possible data incompatibilities, it essential that all datasets include comprehensive and up-to-date metadata records for sampling locations and protocols (Boyce et al. 2006). Rigorous metadata documentation includes descriptions of the data, sampling design and data collection protocols, quality control procedures, preliminary data processing used (e.g., derivatives or extrapolations, estimation procedures), professional judgment used, and any known anomalies or oddities of the data (National Research Council 1995; Pont et al. 2006; Hughes and Peck 2008).

Combining data from different probability-based sampling designs

Combining data from different studies is most straight forward if the sampling designs are probability-based (Olsen et al. 1999). Most long-term natural resource and environmental surveys use *probability-based* or *probabilistic* survey sampling designs such as simple random, systematic, stratified, and cluster designs. In recent years, the spatially balanced generalized random tessellation stratified (GRTS) design (Stevens and Olsen 2004; Olsen et al. 2012) has been developed and implemented by several resource agencies (e.g., US Environmental Protection Agency (USEPA), Oregon Department of Fish and Wildlife, Bonneville Power Administration). McDonald (2012) recommends using the GRTS design for large-scale and long-term ecological programs because of the design's flexibility and broad spatial coverage. Probabilistic survey designs have the characteristic that every element in the population has a known and positive (i.e., >0) probability of being chosen; consequently, unbiased estimates of population parameters that are linear functions of the observations (e.g., population means) can be constructed from the data.

To be combined, datasets must have variables in common (or variables that can be transformed to achieve commonality) and must be capable of being restructured as a single probabilistic sample (Larsen et al. 2007). Cox and Piegorsch (1994, 1996) described three methods for combining data from two or more probabilistic surveys. The first combines weighted estimates from separate probability samples. The estimates for the parameter of interest and its variance are computed for each sample; then, each estimate is weighted inversely proportional to its estimated variance, and then, the weighted estimates are added resulting in a design-based unbiased minimum variance combined estimate (Cox and Piegorsch 1994, 1996). A second method is based on post-stratification (Cox and Piegorsch 1996; Olsen et al. 1999). Strata are defined by using shared frame attributes or subsamples that partition the two probabilistic samples. Both samples are post-stratified by revising sample unit weights proportional to the new stratum size. Revised estimates are then computed for the parameter(s) of interest. Cox and Piegorsch (1996) indicated that dual-frame estimation could be used to combine the estimates or to estimate a nonframe variable or an index based on frame variables.

In the third method, two probabilistic samples are directly combined into one sample. The probabilities of each sampling unit's inclusion in the combined sample are computed from their first- and second-order inclusion probabilities in the original samples (Cox and Piegorsch 1996).

Using some elements from methods described by Cox and Piegorsch (1996), Larsen et al. (2007) combined stream monitoring data from two probability surveys implemented in Oregon to demonstrate how survey design principles can facilitate data aggregation. The data were from the Oregon Department of Fish and Wildlife's integrated aquatic monitoring program (i.e., salmonid populations, stream habitats, water quality and aquatic biotic assemblages; Nicholas 1997) and the US Forest Service's Aquatic and Riparian Effectiveness Monitoring Program, which was focused on indices of watershed health (Reeves et al. 2004). Even though these two efforts targeted questions at different spatial scales and used different indicators, the survey data could be combined into a single probabilistic sample because sound survey design principles were used and because the details of the survey frames and sample selection methods were well documented.

Similarly, the Oregon Department of Environmental Quality conducted a comprehensive assessment of water quality and aquatic habitat in the Willamette River basin by combining 450 randomly selected sites from nine probability-based monitoring programs into a single probabilistic assessment (Mulvey et al. 2009). Although the sampling frames were different for each monitoring program, Mulvey et al. (2009) randomly chose monitoring sites within the basin and assigned differential site weighting factors to the data to address potential sources of bias in randomness. Combining the datasets was possible because the various programs all used the USEPA's Environmental Monitoring and Assessment Program's field sampling methods (Stoddard et al. 2005).

It is also possible to build future data integration into monitoring designs. For example, Larsen et al. (2008) used a GRTS-based "master sample" approach for stream networks in Oregon and Washington. This scheme establishes a framework of potential sampling sites (points, linear networks, or polygons) that can be sampled at a variety of spatial scales, in such a way that spatial balance relative to the resource or feature under consideration is maintained, and the advantages of a probability-based design are retained as successive samples are drawn (Stevens and Olsen 2004). The Oregon Master Sample is being used in select watersheds and consists of almost 180,000 stream sites. The Washington State Department of Ecology (2006) also adopted the master sample concept for use in stream sampling by several different state agencies as part of its status and trend monitoring. More widespread use of the master sample concept could, when applied and managed correctly, significantly strengthen the statistical rigor of regional assessments using data from multiple monitoring programs.

Combining data from probability-based and nonprobability-based sampling designs

Environmental monitoring programs may acquire data derived from both probabilistic and *nonprobability-based* (sometimes called ad hoc, convenient, opportunistic, or targeted) sampling methods; for example, a water quality program may include stations randomly placed to monitor ambient conditions and nonrandom sites to monitor point source pollutants (e.g., Stein and Bernstein 2008). Nonprobability-based sampling selects population elements subjectively, and unlike probabilistic sampling, not every element of the population has a known and positive probability of being chosen. Nonprobability-based data may not be representative of the population of interest, and there is no ability to quantify that uncertainty. Inferences to the population are possible by using statistical models if the available data support the model used, so that unbiased estimates of the model can be obtained.

Combining probabilistic data with nonprobability-based data has significant limitations that must be factored into the analysis. In these situations, spatial and temporal variation cannot be assumed to have been factored into sampling in equivalent ways. The primary problem is that quantitative estimates of variation and uncertainty cannot be calculated from nonprobability-based data, so the validity of the results cannot be quantified. The nature and objectives of the final analysis in which the data will be used will determine how severe a problem this may be.

There are important caveats to address before any of the methods presented in this section are used. The samples used must be synoptic (i.e., sampled from the entire population) or, at the very least, must lack any evidence suggesting that data points were preferentially selected. In addition, the interpretation of any variance

estimate is open to question, and in some methods, quantifying uncertainty is problematic.

Nonprobability data may be placed into a probability-based sample context by simply treating the nonprobability data as a simple random sample (i.e., each sample has the same probability of being chosen). While this solution is suboptimal in a rigorous statistical sense, it allows the analyst to create a "pseudoprobability" structure for the sample. To do so, some information about the entire population must be available, in addition to information from the sample. If only the locations of the sample sites are known, there is still some recourse. All environmental populations have spatial structure because locations near one another are subject to the same natural and anthropogenic stressors and influences (i.e., they are statistically autocorrelated). One approach to imputing a pseudoprobability is simply to say that a sample site represents all of those population elements closer to that site than to any other sample site. The size (number, length, area, or volume) of the total of those elements is then used as a weight for that sample point. The pseudoprobability approach was used in the following methods proposed by Overton et al. (1993) and Brus and de Gruijter (2003).

In this set of approaches, the first step is to choose valid probabilistic samples and "found" datasets. Found sites are chosen from the overall nonprobability sample that conforms to the probability sample characteristics (Overton et al. 1993). One of two methods can be used to determine similarity between probability and nonprobability samples and to produce population estimates: *pseudorandom* and *stratified calibration* (Fig. 1). The pseudorandom approach is used when the variable of interest from the found dataset was also measured in the probability-based survey. If the variable of interest is only known for the found data, then stratified calibration is used.

The concept behind the pseudorandom approach is closely related to the post-stratification technique (Fuller 2009) that is sometimes used to improve a poorly randomized sample after the fact. To combine the samples, the sampling frame attributes are used to classify the probabilistic sample into homogeneous groups or subpopulations (Overton et al. 1993). Found sites are then assigned to the subpopulations. Pseudorandom samples are defined by treating the nonrandom sample as if it were a stratified random design with simple random sampling within the strata. Population estimates can then be calculated from the combined data.

The stratified calibration technique is used when the desired population attribute was not measured in the probability sample. The initial steps are the same as for the pseudorandom approach described in the preceding paragraph; similarity between the datasets is established, the probability sample is stratified, subpopulations are identified, found sites are assigned to the subpopulations, and predictor equations for desired attributes are developed for each subpopulation (Overton et al. 1993). If two subpopulations have similar predictor relationships, they are combined; if not, they are kept separate. Some populations may not have corresponding found data, so no predictor equation can be developed. The desired attribute is then predicted for the probability sample, and population estimates can be calculated.

Astin (2006) combined nonprobability-based data with probabilistic data and census data to select and calibrate water quality indicators in the Potomac River basin. Data originated from Maryland, Virginia, and Pennsylvania. Because each monitoring group used variations of the USEPA's *rapid bioassessment protocols for streams and rivers* (Plafkin et al. 1989), an additive or multimetric framework based on the protocols was used to combine the data. Astin (2006) assumed that repeat observations taken at fixed sites were independent and that the sites were representative of the range of conditions found in the basin.

With respect to these two approaches, Overton et al. (1993) cautioned that there is an unprovable assumption that the sites are representative because found sites were not chosen randomly. Brus and de Gruijter (2003) asserted that if one is not confident in the representativeness of the found sites or one does not want to make this assumption, then the methods proposed by Overton et al. (1993) should not be used. The approach of Brus and de Gruijter maintains the assumption of representativeness, and the validity of results from estimating means of the nonprobability data is ensured by collecting and combining additional data through probabilistic sampling. The approach of Brus and de Gruijter involves overlaying a grid onto the nonprobability and randomly sampled data and calculating the difference in the means by interpolation through point kriging. The error in estimating the mean for each nonprobability sample is calculated by the difference of the true mean and the average of
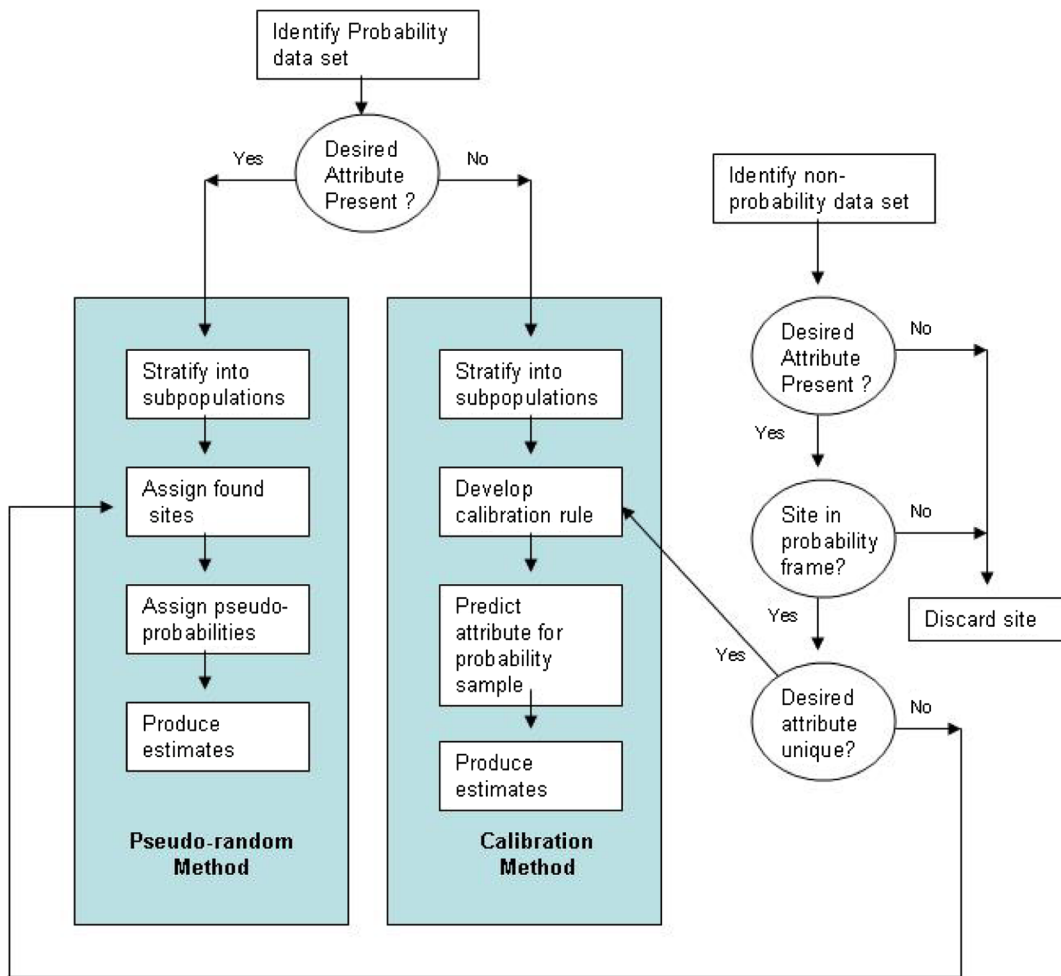
Fig. 1 A schematic of the Overton et al. (1993) process for combining nonprobability-based data with a probability-based dataset. Reproduced with kind permission from Springer Science+Business Media

the kriged values; this error is then used to calculate measures of bias and variance. Brus and de Gruijter (2003) consider the resulting estimators to be fairly unbiased, even when the nonprobability sample is very biased.

Finally, a preferable approach to combining data after they are collected is to design a monitoring framework that incorporates known and future probabilistic and nonprobability sites from the beginning. Stein and Bernstein (2008) demonstrated how to construct a hybrid sampling design to incorporate fixed targeted monitoring sites to monitor discharge permit compliance and fixed sites located at unique areas of interest with random ambient water quality sites to better assess conditions of the entire San Gabriel River watershed, CA. The hybrid design resulted in a more complete assessment of

contamination impacts and patterns; then, either of the two sampling designs could achieve alone.

Data aggregation

In statistics, data aggregation refers to data that have been summarized to provide information at a broader level than the sampling sites. The summary statistics can then be used for further analyses. Fundamentally, the statistical appropriateness of data aggregation is a function of the properties of the data as determined by the underlying sampling design.

A problematic aspect of aggregation is that inferences about relationships in the data can change as the level of aggregation changes. The challenge for the

analyst then becomes using inference procedures that are relatively invariant to such changes or that vary in a controllable and predictable way. Aggregation is most straightforward with data that can be summarized with totals or averages, for example, the total number of salmon spawning in Oregon coastal streams. In this case, aggregation can be as simple as summing fine-scale data, perhaps using weights and confidence limits that reflect the size of the spatial unit associated with the fine-scale data relative to the size of the population.

Data collected within spatial domains can be highly correlated because samples collected near or adjacent to one another are typically more similar to one another than samples taken further away. A continuous spatial domain can be a conterminous area or region, for example, a political jurisdiction (e.g., city, county, or state), a natural feature (e.g., lake, estuary, watershed, ecotype), or a management unit (e.g., a state forest, ranch, agricultural field; Stehman and Overton 1996). Distorted conclusions from aggregated spatial data can arise from several sources, including the sampling designs themselves, alternative ways of combining the data, the process by which data are spatially "scaled" up or down, changing the geographical boundaries represented by the data, and hidden influences in the environment that are not taken into account in sampling. This section briefly describes some of the problems, specifically change of support problems, encountered when grouping spatial data (Table 1).

Change of support problem

The change of support problem (COSP) arises when inferences are made from spatially transformed data; i.e., observations are made at one spatial scale, but the physical or environmental process of interest is operating at a different spatial scale (Gotway and Young 2002; Table 2). Here, support refers to the geometric size (or volume), shape, and spatial orientation of the area associated with a measurement (Gotway and Young 2002; Crawford and Young 2005). Aggregation changes the underlying two- or three-dimensional space represented by a variable, creating a new variable with different spatial and statistical properties (Gotway and Young 2002; Crawford and Young 2005). In mineral surveys, COSPs receive considerable attention in calculating volumes of material over large areas from core samples. Meteorological and snow pack data are also subject to COSPs, where a continuum (for example, of temperature, precipitation, or snow water content) must be inferred from point data. COSPs can be addressed and various geostatistical solutions are available (e.g.,

Table 1 Common problems encountered when aggregating data

Problem encountered	What it is	Selected references
Change of support problem	Occurs when observations are made on one spatial scale but the process of interest is operating at a different spatial scale. Can create inference problems	Gotway and Young (2002); Crawford and Young (2005)
Modifiable areal unit problem	Occurs when changes in the size, configuration, and number of groupings of data alter the apparent relationships. May obscure actual relationships	Openshaw and Taylor (1979); Openshaw (1983); Jelinski and Wu (1996); Dark and Bram (2007); Alexandridis et al. (2010)
Ecological fallacy	Occurs when the relationships between group means is inferred to individuals, leading to false conclusions about individuals	Johnson and Chess (2006)
Ecological correlation	Correlations occur between group means as opposed to individual means. Assuming that the correlations at the group level are equal to those at the individual level is incorrect	Robinson (1950); Clark and Avery (1976)
Simpson's paradox	Relationships between attributes appear to change (or even reverse) depending on how a population and its attributes are stratified. Occurs with discrete data in descriptive statistical analyses	Wagner (1982); Cohen (1986); Thomas and Parresol (1989); Piñeiro et al. (2006)

The modifiable areal unit problem and ecological fallacy are specific change of support problems. Ecological correlation and Simpson's paradox are specific types of ecological fallacies

Gotway and Young 2002). Gelfand et al. (2001) were able to address spatial and temporal aspects of COSPs when they determined ozone levels over different areas of Atlanta, GA. Ravines et al. (2008) also addressed spatial and temporal aspects of COSPs in rainfall and runoff data from the Rio Grande basin, Brazil. Specific types of COSPs also exist including the modifiable areal unit problem and ecological fallacy, discussed below.

Modifiable areal unit problem

In the absence of variability, the unit of aggregation has no impact on the value of a quantity expressed as a per unit value (e.g., velocity expressed as m/s, density as g/m^3, or species richness as number of species/km^2). The result is the same regardless of the size of the measurement unit. However, real systems always have some variation, so the result of aggregation can be highly influenced by the measurement unit size and the variation encompassed therein. Yule and Kendall (1950) noted that correlations between variables measured on *modifiable* units such as field plots or geographical areas depend on the size of the unit in contrast to variables measured on *nonmodifiable* units such as

Table 2 Examples of change of support problems

We observe or analyze	But the nature of the process is	Examples
Point	Point	Point kriging; prediction of under-sampled variables
Area	Point	Ecological inference; quadrat counts
Point	Line	Contouring
Point	Area	Use of areal centroids; spatial smoothing; block kriging
Area	Area	Modifiable areal unit problem; areal interpolation; incompatible/misaligned zones
Point	Surface	Trend surface analysis; environmental monitoring; exposure assessment
Area	Surface	Remote sensing; multiresolution images; image analysis

Support refers to the size, shape, and spatial orientation associated with each data value. Table reproduced from Gotway and Young (2002) with permission. Reprinted with permission from Taylor & Francis Ltd. (http://www.tandf.co.usk/journals)

persons, automobiles, or trees. Openshaw and Taylor (1979) described this issue of variability in a geographical context as the *modifiable areal unit problem* (MAUP). The MAUP is a specific type of COSP (Crawford and Young 2005) and a potential source of error that can affect analyses that aggregate spatial data; that is, if relationships between variables change with selection of different areal units, then the reliability of the results decreases. For example, estimates of fish species richness at sites vary with site size (Hughes et al. 2002; Kanno et al. 2009; Terra et al. 2013a), meaning that combining such estimates into regional means will be more variable or will tend to underestimate richness if site size is variable or small, respectively.

The MAUP arises because spatial units are modifiable (in the sense that they can be aggregated to form other units or change configuration) and are often arbitrarily determined (Jelinski and Wu 1996). There are two components to the MAUP, the scale (aggregation) effect, and the zonation (grouping) effect. The scale effect describes the inconsistency of statistical results from various levels of aggregation (Openshaw 1983; Amrhein 1995; Wong 1996). Aggregation decreases variances and smooths the resulting values such that information is lost (Wong 1996). Smoothing applies to all variables or attributes associated with spatial observations, but the amount varies with the level of aggregation (Wong 1996). The zonation effect refers to the variability of statistical values when areal units vary in size and shape while the number of units remains the same (Openshaw and Taylor 1979; Openshaw 1983; Jelinski and Wu 1996; Wong 1996).

Svancara et al. (2002) examined how the MAUP affected the statistical relationship between elk (*Cervus elaphus*) recruitment and three independent variables (forest productivity, the proportion of nonbatholith land across the summer range, and mature bull elk density) when game management units were aggregated to three different levels in three different configurations. Svancara et al. (2002) found inconsistencies in variances, correlation coefficients, regression parameters, and regression model fit (coefficient of determination) across aggregations. Differences were not only dependent upon the unit configuration and level of aggregation but on the variable of interest.

From a series of controlled statistical simulations, Amrhein (1995) concluded that the effects of MAUP

on aggregation depend on the statistics calculated (e.g., means, variances, regression coefficients, or Pearson correlation coefficients). Amrhein found that means and variances were resistant to aggregation effects. Regression coefficients and Pearson correlation statistics exhibited dramatic aggregation effects. Based on these simulations, Amrhein concluded that the MAUP in spatial analysis does not appear to be as pervasive or unpredictable as described in earlier literature, and aggregation effects may be more easily identified and dealt with than once thought.

Ecological fallacy

Assuming what holds true for the group also holds true for an individual is an inappropriate extrapolation or *ecological fallacy* (Johnson and Chess 2006). Ecological fallacy is comprised of *aggregation bias* caused by the grouping of individuals and *specification bias* caused by the differential distribution of confounding variables created by grouping (Gotway and Young 2005). Aggregation and specification biases are analogous to the scale and zoning effect in the MAUP (Gotway and Young 2005). Similarly, what holds true for a region does not necessarily hold for an area or site within the region. For example, the relationship between years of schooling and support of environmental issues on a state-wide basis may be quite different from the relationship between average years of schooling and support of environmental issues on an individual basis. Two types of ecological fallacy of concern for environmental assessments are ecological correlation and Simpson's paradox.

Ecological correlation

Ecological correlation was originally used by sociologists to refer to correlations between variables that are group means (e.g., the correlation between salmonid/seafood consumption rates and per capita income) as opposed to individuals. Clark and Avery (1976, p. 429) stated that a significant "disadvantage of using aggregate data is the inherent difficulty of making valid multilevel inferences based on a single level of analysis." Variables used in individual correlations (such as weight, age, or length) are descriptive properties of individuals, while the statistical objects in an ecological correlation are properties of groups (e.g., rates, percentages, or means; Robinson 1950).

Ecological correlations between aggregated individual properties can therefore be misleading (Robinson 1950). For example, Schooley (1994) found that black bear (*Ursus americanus*) habitat selection varied by year but was similar between two study areas in individual years. However, when data from individual years were aggregated, selection at the two sites appeared to differ. In this case, the annual variation was lost in the aggregation, leading to incorrect inferences about habitat selection between the two sites.

Simpson's paradox

Simpson's paradox is an ecological fallacy in which the apparent associations of variables seem to reverse when they are grouped. It is often illustrated with contingency tables reporting frequency data and marginal totals. It occurs because there can be more than one way to stratify the variables; for example, Pacific salmon counts may be stratified by watershed or by hatchery versus wild origin. When linear operators such as simple summations or means are used to examine the data, no apparent distortions occur as a result of grouping; the aggregate of mean values is the mean of aggregate values. However, nonlinear operators, such as ratios or rates, do not have this characteristic. The ratio of aggregated values is mathematically and conceptually not the same as the aggregated value of the ratios in the stratum used for grouping, which makes a difference in the outcome. Therefore, it is critical to determine the parameter of interest before the data are collected.

Several examples of Simpson's paradox exist for environmental data. In Thomas and Parresol (1989), a previous analysis of loblolly pine (*Pinus taeda*) plantations had shown that recent radial growth rates had decreased when rates were compared diameter class by diameter class, implying that the stand-level wood volume growth rates were declining. However, individual tree growth rates did not typically show this trend. This led Thomas and Parresol (1989) to weight diameter class means by the number of trees in each class and change the measure of growth to basal area growth. They then found that overall growth rates were increasing, not declining.

Allison and Goldberg (2002) also observed Simpson's paradox in a comparison of species-level versus community-level responses to arbuscular mycorrhizal fungi across a gradient of phosphorus availability. Several individual species

showed a declining response to the fungi as phosphorous increased, but when species were grouped into communities, the relationship of declining response to phosphorus weakened significantly.

Piñeiro et al. (2006) found Simpson's paradox occurred when the whole-soil C/N ratio decreased after long-term grazing, but the C/N ratios of all soil organic matter pools increased. They concluded that whole-soil C/N ratios can erroneously assess the impact of disturbance on soil organic matter quality and estimation of nitrogen mineralization rates.

Recommendations for, and examples of, improved surveys of environmental status and trends

Our goal with this review has been to balance the presentation of approaches to combine data from multiple sources for regional environmental assessments coupled with cautions about statistical complexities of doing so. Heightened interest in evaluating the success of policies for managing natural resources and protecting the environment makes it increasingly likely that disparate information will be used in assessing status and trends of species and ecosystems. We hope that by raising awareness about the difficulties inherent in combining data and aggregating data for local- and regional-level analyses, it will increase the likelihood that future monitoring efforts will be modified and/or planned to accommodate data from multiple sources. We remind analysts that the techniques that we described in this paper (e.g., pseudorandom and stratified calibration techniques) should be used as a basis for statistical consultations because they may require modifications before use. Each situation will be unique, and the services of a statistician with experience in data combining methods should be obtained when planning data integration projects. We strongly encourage managers and monitoring practitioners to consult with applied statisticians at all levels of environmental monitoring including planning, implementation, and analysis. This will help ensure that objectives are achievable, the population(s) properly identified, sampling schemes and protocols are rigorous, data are accurate, and results are valid (Gitzen and Millspaugh 2012; Reynolds 2012). Finally, we encourage analysts to publish or, otherwise, make available their

detailed protocols for combining data and/or addressing COSPs. By doing so, more techniques can be developed or modified to increase the use of data from multiple sources.

Examples of current spatially extensive status and trend monitoring programs for aquatic ecosystems

Because of the difficulties in combining and aggregating existing and disparate data from multiple sources, several agencies have developed their own status and trend monitoring programs for aquatic ecosystems, but they also use existing geographic data layers to help interpret results. The USA states of California, Maryland, Minnesota Ohio, and Oregon have implemented multi-year monitoring programs that now include samples from hundreds to thousands of sites for relating physical and chemical habitat conditions to fish or macroinvertebrate assemblage condition (e.g., Yoder et al. 2005; Mulvey et al. 2009; Anlauf et al. 2011; Stranko et al. 2012; MDNR 2014; May et al. 2015). Stanfield (2012) reported on a cooperative fish and macroinvertebrate monitoring program based on hundreds of sample sites for Ontario, Canada, tributaries draining into Lake Ontario. Thirteen institutions with responsibilities for salmon and steelhead recovery in the lower Columbia River Basin have initiated a project to improve monitoring by developing standardized sampling frames, field methods, and data sharing protocols (Puls et al. 2014). Callisto et al. (2014) described a monitoring program for streams and reservoirs of four hydropower basins in southeastern Brazil in which they relate fish and macroinvertebrate assemblage condition at hundreds of sites to land use and physical and chemical habitat condition. In the Pacific Northwest states of the USA, the US Forest Service, US Bureau of Land Management, and Bonneville Power Administration are using such programs to relate status and trends in stream physical habitat structure with macroinvertebrate (e.g., Lanigan et al. 2012; Irvine et al. 2014) and fish (CHaMP 2014) assemblage condition at hundreds of sites. Australia has developed a Sustainable Rivers Audit for monitoring river health biannually at hundreds of sites in 23 catchments across five states (Davies et al. 2010). At a national scale, the USEPA has implemented a status and trend ecological monitoring program for lakes, reservoirs, streams, and rivers based on thousands of sites (USEPA 2009, 2013; Kaufmann et al. 2014).

The key challenges for such expansive and expensive biological monitoring programs are data management, creating reports that are understandable to the general public, and maintaining funding levels. However, inadequate status and trend monitoring means that we are ignorant of aquatic ecosystem condition and changes, ignorant of the biological effects of anthropogenic pressures and stressors, and ignorant of the biological effectiveness of rehabilitation measures. Also, protecting aquatic ecosystems costs much less than attempting to rehabilitate them after they are degraded (e.g., Woody et al. 2010; Hughes et al. 2014).

References

Alexandridis, T. K., Katagis, T., Gitas, I. Z., Silleos, N. G., Eskridge, K. M., & Gritzas, G. (2010). Investigation of aggregation effects in vegetation condition monitoring at a national scale. *International Journal of Geographical Information Science, 24*(4), 507–521.

Allison, V. J., & Goldberg, D. E. (2002). Species-level versus community-level patterns of mycorrhizal dependence on phosphorous: an example of Simpson's paradox. *Functional Ecology, 15*(3), 346–352.

Amrhein, C. G. (1995). Searching for the elusive aggregation effect: evidence from statistical simulations. *Environment & Planning A, 27*(1), 105–119.

Anlauf, K. J., Gaeuman, W., & Jones, K. K. (2011). Detection of regional trends in salmonid habitat in coastal streams, Oregon. *Transactions of the American Fisheries Society, 140*, 52–66.

Astin, L. E. (2006). Data synthesis and bioindicator development for nontidal streams in the interstate Potomac River basin, USA. *Ecological Indicators, 6*(4), 664–685.

Bataineh, A. L., Oswald, B. P., Bataineh, M., Unger, D., Hung, I.-K., & Scognamillo, D. (2006). Spatial autocorrelation and pseudoreplication in fire ecology. *Fire Ecology, 2*(2), 107–118.

Battin, J., & Lawler, J. (2006). Cross-scale correlation and the design and analysis of avian habitat selection studies. *Condor, 108*(1), 59–70.

Boyce, D., Judson, B., & Hall, S. (2006). Data sharing—a case of shared databases and community use of on-line GIS support systems. *Environmental Monitoring and Assessment, 113*(1–3), 385–394.

Brown, L. R., Gregory, M. B., & May, J. T. (2009). Relation of urbanization to stream fish assemblages and species traits in nine metropolitan areas of the United States. *Urban Ecosystems, 12*, 391–416.

Brus, D. J., & de Gruijter, J. J. (2003). A method to combine non-probability sample data with probability sample data in estimating spatial means of environmental variables. *Environmental Monitoring and Assessment, 83*(3), 303–317.

Callisto, M., Hughes, R. M., Lopes, J. M., & Castro, M. A. (Eds.). (2014). *Ecological conditions in hydropower basins. Série Peixe Vivo 2.* Belo Horizonte: Companhia Energética de Minas Gerais.

Canfield, D. E., Jr., Brown, D. C., Bachmann, R. W., & Hoyer, M. V. (2002). Volunteer lake monitoring: testing the reliability of data collected by the Florida LAKEWATCH program. *Lake and Reservoir Management, 18*(1), 1–9.

Cao, Y., Larsen, D. P., Hughes, R. M., Angermeier, P. M., & Patton, T. M. (2002). Sampling effort affects multivariate comparisons of stream assemblages. *Journal of the North American Benthological Society, 21*(4), 701–714.

Chaves, L. F. (2010). An entomologist guide to demystify pseudoreplication: data analysis of field studies with design constraints. *Journal of Medical Entomology, 47*, 291–298.

Clark, W. A. V., & Avery, K. L. (1976). The effects of data aggregation in statistical analysis. *Geographical Analysis, 8*(4), 428–438.

Cohen, J. E. (1986). An uncertainty principle in demography and the unisex issue. *The American Statistician, 40*(1), 32–39.

Columbia Habitat Monitoring Program (CHaMP). (2014). *Scientific protocol for salmonid habitat surveys within the Columbia Habitat Monitoring Program. Columbia Habitat Monitoring Program.* Portland: Bonneville Power Administration.

Council, N. R. (1995). *Finding the forest in the trees: the challenge of combining diverse environmental data.* Washington, DC: National Academy Press.

Cox, L. H., & Piegorsch, W. W. (1994). *Combining environmental information: environmetric research in ecological monitoring, epidemiology, toxicology, and environmental data reporting.* Technical Report Number 12. Research Triangle Park, North Carolina, USA: National Institute of Statistical Sciences. https://www.niss.org/sites/default/files/pdfs/technicalreports/tr12.pdf. Accessed 24 April 2014.

Cox, L. H., & Piegorsch, W. W. (1996). Combining environmental information I: Environmental monitoring, measurement and assessment. *Environmetrics, 7*(3), 299–308.

Crawford, C. A. G., & Young, L. J. (2005). Change of support: an inter-disciplinary challenge. In P. Renard, H. Demougeot-Renard, & R. Froidevaux (Eds.), *Geostatistics for Environmental Applications* (pp. 1–13). Berlin: Springer-Verlag.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography, 31*(5), 471–479.

Darwall, W. R. T., & Dulvy, N. K. (1996). An evaluation of the suitability of non-specialist volunteer researchers for coral reef fish surveys Mafia Island, Tanzania—a case study. *Biological Conservation, 78*(3), 223–231.

Davies, P. E., Harris, J., Hillman, T., & Walker, K. (2010). The Sustainable Rivers Audit: assessing river ecosystem health in the Murray-Darling Basin, Australia. *Marine and Freshwater Research, 61*, 764–777.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography, 30*(5), 609–628.

Evans, S. M., Foster-Smith, J., & Welch, R. (2001). Volunteers assess marine biodiversity. *Biologist, 48*(4), 168–172.

Fayram, A. H., Miller, M. A., & Colby, A. C. (2005). Effects of stream order and ecoregion on variability in coldwater fish index of biotic integrity scores within streams in Wisconsin. *Journal of Freshwater Ecology, 20*(1), 17–25.

Ficetola, G. F., Manenti, R., De Bernardi, F., & Padoa-Schioppa, E. (2012). Can patterns of spatial autocorrelation reveal population processes? An analysis with the fire salamander. *Ecography, 35*(8), 693–703.

Fortin, M.–. J., Drapeau, P., & Legendre, P. (1989). Spatial autocorrelation and sampling design in plant ecology. *Vegetatio, 83*(1–2), 209–222.

Frissell, C. W., & Bayles, D. (1996). Ecosystem management and the conservation of aquatic biodiversity and ecological integrity. *Water Resources Bulletin, 32*, 229–240.

Fuller, W. A. (2009). *Sampling statistics*. Hoboken: Wiley.

Gelfand, A. E., Zhu, L., & Carlin, B. P. (2001). On the change of support problem for spatio-temporal data. *Biostatistics, 2*(1), 31–45.

Gitzen, R. A., & Millspaugh, J. J. (2012). Ecological monitoring: the heart of the matter. In R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, & D. S. Light (Eds.), *Design and analysis of long-term ecological monitoring studies* (pp. 3–22). UK: Cambridge University Press.

Goffredo, S., Piccinetti, C., & Zaccanti, F. (2004). Volunteers in marine conservation monitoring: a study of the distribution of seahorses carried out in collaboration with recreational scuba dives. *Conservation Biology, 18*(6), 1492–1503.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association, 97*(458), 632–648.

Gouveia, C., Fonseca, A., Câmara, A., & Ferreira, F. (2004). Promoting the use of environmental data collected by concerned citizens through information and communication technologies. *Journal of Environmental Management, 71*(2), 135–154.

Hanson, S. (2006). Volunteer vs. agency comparison: *E. coli* monitoring. *The Volunteer Monitor, 18*(1), 7. & 12.

Harding, J. S., Benfield, E. F., Bolstad, P. V., Helfman, G. S., & Jones, E. B. D., III. (1998). Stream biodiversity: the ghost of land use past. *Proceedings of the National Academy of Science of the United States of America, 95*, 14843–14847.

Heffner, R. A., Butler, M. J., & Keelan, R. C. (1996). Pseudoreplication revisited. *Ecology, 77*(8), 2558–2562.

Herlihy, A. T., Hughes, R. M., & Sifneos, J. C. (2006). Landscape clusters based on fish assemblages in the conterminous USA and their relationship to existing landscape classifications. In R. M. Hughes, L. Wang, & P. W. Seelback (Eds), *Landscape influences on stream habitat and biological assemblages* (pp. 87–112). Bethesda, Maryland: American Fisheries Society.

Hughes, R. M., & Peck, D. V. (2008). Acquiring data for large aquatic resource surveys: the art of compromise among science, logistics, and reality. *Journal of the North American Benthological Society, 27*(4), 837–859.

Hughes, R. M., Kaufmann, P. R., Herlihy, A. T., Intelmann, S. S., Corbett, S. C., Arbogast, M. C., & Hjort, R. C. (2002). Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. *North American Journal of Fisheries Management, 22*, 1229–1240.

Hughes, R. M., Dunham, S., Maas-Hebner, K. G., Yeakley, J. A., Schreck, C. B., Harte, M., Molina, N., Shock, C. C., & Kaczynski, V. W. (2014). A review of urban water body challenges and approaches: (2) Mitigation and research needs. *Fisheries, 39*, 30–40.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs, 54*(2), 187–211.

Independent Multidisciplinary Science Team (2010). *Letter to Mike Carrier*, Governor's Natural Resource Office and Tom Byler, Oregon Watershed Enhancement Board, Salem, Oregon. Feb. 4, 2010. http://www.fsl.orst.edu/imst/reports/monitoring/Carrier&Byler2-4-10.pdf. Accessed 19 May 2014.

Independent Multidisciplinary Science Team (2011a). *Letter to Doug Decker*, Oregon Department of Forestry and Louise Solliday, Oregon Department of State Lands, Salem, Oregon. Aug. 29, 2011. http://www.fsl.orst.edu/imst/reports/monitoring/ODF&DSL_Elliott_recs_8-29-11.pdf. Accessed 19 May 2014.

Independent Multidisciplinary Science Team (2011b). *Letter to Roy Elicker*, Oregon Department of Fish and Wildlife, Salem, Oregon. Dec. 1, 2011. http://www.fsl.orst.edu/imst/reports/monitoring/Carrier&Byler2-4-10.pdf. Accessed 19 May 2014.

Independent Multidisciplinary Science Team (2013). *Letter to Richard Whitman*, Governor's Natural Resource Office, Salem, Oregon. Nov. 6, 2013. http://www.fsl.orst.edu/imst/reports/monitoring/Whitman-GNRO_11-16-13.pdf. Accessed 19 May 2014.

Irvine, K. M., Miller, S. W., Al-Chokhachy, R. K., Archer, E. K., Roper, B. B., & Kershner, J. L. (2014). Empirical evaluation of the conceptual model underpinning a regional aquatic long-term monitoring program using causal modeling. *Ecological Indicators, 50*, 8–23.

Jelinski, D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology, 11*(3), 129–140.

Johnson, B. B., & Chess, C. (2006). Evaluating public responses to environmental trend indicators. *Science Communication, 28*(1), 64–92.

Kanno, Y., Vokoun, J. C., Dauwalter, D. C., Hughes, R. M., Herlihy, A. T., Maret, T. R., & Patton, T. M. (2009). Influence of rare species on electrofishing distance–species

richness relationships at stream sites. *Transactions of the American Fisheries Society, 138*, 1240–1251.

Kaufmann, P. R., Peck, D. V., Paulsen, S. G., Seeliger, C. W., Hughes, R. M., Whittier, T. R., & Kamman, N. C. (2014). Lakeshore and littoral physical habitat structure in a national lakes assessment. *Lake and Reservoir Management, 30*, 192–215.

Kautza, A., & Sullivan, S. M. P. (2012). Relative effects of local- and landscape-scale environmental factors on stream fish assemblages: evidence from Idaho and Ohio, USA. *Fundamental and Applied Limnology, 180*(3), 259–270.

Lanigan, S. H., Gordon, S. N., Eldred, P., Isley, M., Wilcox, S., Moyer, C., & Andersen, H. (2012). *Northwest Forest Plan— the first 15 years (1994–2008): watershed condition status and trend*. PNW-GTR-856. Portland: U.S. Forest Service.

Larsen, D. P., Olsen, A. R., Lanigan, S. H., Moyer, C., Jones, K. K., & Kincaid, T. M. (2007). Sound survey designs can facilitate integrating stream monitoring data across multiple programs. *Journal of the American Water Resources Association, 43*(2), 384–397.

Larsen, D. P., Olsen, A. R., & Stevens, D. L., Jr. (2008). Using a master sample to integrate stream monitoring programs. *Journal of Agricultural, Biological, and Environmental Statistics, 13*(3), 243–254.

Lawler, J., & Edwards, T. C., Jr. (2006). A variance-decomposition approach to investigating multiscale habitat associations. *Condor, 108*(1), 47–58.

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs, 72*(3), 445–463.

Macedo, D. R., Hughes, R. M., Ligeiro, R., Ferreira, W. R., Castro, M., Junqueira, N. T., Silva, D. R. O., Firmiano, K. R., Kaufmann, P. R., Pompeu, P. S., & Callisto, M. (2014). The relative influence of multiple spatial scale environmental predictors on fish and macroinvertebrate assemblage richness in cerrado ecoregion streams, Brazil. *Landscape Ecology, 29*, 1001–1016.

Marzin, A., Verdonschot, P. F. M., & Pont, D. (2013). The relative influence of catchment, riparian corridor, and reach-scale anthropogenic pressures on fish and macro- invertebrate assemblages in French rivers. *Hydrobiologia, 704*(1), 375–388.

May, J.T., Brown, L.R., Rehn, A.C., Waite, I.R., Ode, P.R., Mazor, R.D., & Schiff, K.C. (2015). Correspondence of biological condition models of California streams at statewide and regional scales. *Environmental Monitoring and Assessment, 187*. doi:10.1007/s10661-014-4086-x.

McDonald, T. (2012). Spatial sampling designs for long-term ecological monitoring. In R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, & D. S. Light (Eds.), *Design and analysis of long-term ecological monitoring studies* (pp. 102–125). UK: Cambridge University Press.

McGarvey, D. J., & Hughes, R. M. (2008). Longitudinal zonation of Pacific Northwest (USA) fish assemblages and the species-discharge relationship. *Copeia, 2008*(2), 311–321.

MDNR (Minnesota Department of Natural Resources) (2014). *Fisheries lake surveys*. www.dnr.state.mn.us/lakefind/ surveys.html. Accessed 1 March 2015.

Millar, R. B., & Anderson, M. J. (2004). Remedies for pseudoreplication. *Fisheries Research, 70*(2–3), 397–407.

Mulvey, M., Leferink, R., & Borisenko, A. (2009). *Willamette Basin rivers and streams assessment*. DEQ 09-LAB-016. Portland: Oregon Department of Environmental Quality.

Network, N. E. D. (2005). *Final white papers and recommenda- tions from beyond ad-hoc: organizing, administrating, and funding a Northwest Environmental Data Network*. Portland: Northwest Power and Conservation Council.

Nicholas, J. W. (1997). *The Oregon Plan for Salmon and Watersheds: Oregon Coastal Salmon Restoration Initiative*. Salem: State of Oregon.

Olsen, A. R., & Peck, D. V. (2008). Survey design and extant estimates for the Wadeable Streams Assessment. *Journal of the North American Benthological Society, 27*(4), 822–836.

Olsen, A. R., Sedransk, J., Edwards, D., Gotway, C. A., Liggett, W., Rathbun, S., Reckhow, K. H., & Young, L. J. (1999). Statistical issues for monitoring ecological and natural re- sources in the United States. *Environmental Monitoring and Assessment, 54*(1), 1–45.

Olsen, A. R., Kincaid, T. M., & Payton, Q. (2012). Spatially balanced survey designs for natural resources. In R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, & D. S. Light (Eds.), *Design and analysis of long-term ecological monitor- ing studies* (pp. 126–150). UK: Cambridge University Press.

Openshaw, S. (1983). *The modifiable areal unit problem: concepts and techniques in modern geography No. 38*. Connecticut: GeoBooks.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences* (pp. 127–144). London: Pion Limited.

Oregon Plan. (1997). *Oregon Plan for Salmon and Watersheds (consisting of the Oregon Coastal Salmon Restoration Initiative, March 10, 1997 and as amended with the Steelhead Supplement, December 1997)*. Salem: Governor's Natural Resources Office, State of Oregon.

Overton, J., Young, T., & Overton, W. S. (1993). Using 'found' data to augment a probability sample: procedure and case study. *Environmental Monitoring and Assessment, 26*(1), 65– 83.

Pacific Northwest Aquatic Monitoring Partnership (2014). http:// www.pnamp.org/. Accessed 5 May 2014.

Piñeiro, G., Oesterheld, M., Batista, W., & Paruelo, J. M. (2006). Opposite changes of whole-soil vs. pools C:N ratios: a case of Simpson's paradox with implications on nitrogen cycling. *Global Change Biology, 12*(5), 804–809.

Pinto, B. C. T., Araujo, F. G., Rodriguez, V. D., & Hughes, R. M. (2009). Local and ecoregion effects on fish assemblage struc- ture in tributaries of the Rio Paraíba do Sul, Brazil. *Freshwater Biology, 54*, 2600–2615.

Plafkin, J. L., Barbour, M., Porter, K., Gross, S., & Hughes, R. (1989). *Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish*. EPA 440-4-89-001. Washington, DC: US Environmental Protection Agency, Office of Water Regulations and Standards.

Pont, D., Hugueny, B., Beier, U., Goffaux, D., Melcher, A., Noble, R., Rogers, C., Roset, N., & Schmutz, S. (2006). Assessing river biotic condition at the continental scale: a European approach using functional metrics and fish assemblages. *Journal of Applied Ecology, 43*, 70–80.

Pont, D., Hughes, R. M., Whittier, T. R., & Schmutz, S. (2009). A predictive index of biotic integrity model for aquatic-vertebrate assemblages of western U.S. streams. *Transactions of the American Fisheries Society, 138*, 292–305.

Puls, A., Dunn, K. A., & Hudson, B. G. (2014). *Evaluation and prioritization of stream habitat monitoring in the lower Columbia salmon and steelhead recovery domain as related to the habitat monitoring needs of ESA recovery plans*. PNAMP Series 2104-003. Portland: Pacific Northwest Aquatic Monitoring Partnership.

Ravines, R. R., Schmidt, A. M., Migon, H. S., & Rennó, C. D. (2008). A joint model for rainfall-runoff: the case of the Rio Grande Basin. *Journal of Hydrology, 353*(1–2), 189–200.

Reeves, G. H., Hohler, D. B., Larsen, D. P., Busch, D. E., Kratz, K., Reynolds, K., Stein, K. F., Atzet, T., Hays, P., & Tehan, M. (2004). *Effectiveness monitoring for the aquatic and riparian component of the Northwest Forest Plan: conceptual framework and options*. PNW-GTR-577. Portland: USDA Forest Service, PNW Research Station.

Reynolds, J. H. (2012). An overview of statistical consideration in long-term monitoring. In R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, & D. S. Light (Eds.), *Design and analysis of long-term ecological monitoring studies* (pp. 23–53). UK: Cambridge University Press.

Rieman, B. E., Dunham, J. D., & Peterson, J. T. (1999). *Development of a database to support a multiscale analysis of the distribution of westslope cutthroat trout*. Final report to the US Geological Survey, Agreement 1445-HQ-PG-01026BRD, Reston, VA.

Robinson, A. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*(3), 351–357.

Roni, P. (2005). Overview and background. In P. Roni (Ed.), *Monitoring stream and watershed restoration* (pp. 6–11). Bethesda: American Fisheries Society.

Roper, B. B., Bennett, S., Lanigan, S. H., Archer, E., Downie, S. T., Faustini, J., Hillman, T. W., Hubler, S., Jones, K., Jordan, C., Kaufmann, P. R., Merritt, G., Moyer, C., & Pleus, A. (2010). A comparison of the performance of protocols used by seven monitoring groups to measure stream habitat in the Pacific Northwest. *North American Journal of Fisheries Management, 30*(2), 565–587.

Sály, P., Takács, P., Kiss, I., Biró, P., & Erös, T. (2011). The relative influence of spatial context and catchment- and site-scale environmental factors on stream fish assemblages in a human-modified landscape. *Ecology of Freshwater Fish, 20*, 251–262.

Sandel, B., & Smith, A. B. (2009). Scale as a lurking factor: incorporating scale-dependence in experimental ecology. *Oikos, 118*(9), 1284–1291.

Schooley, R. L. (1994). Annual variation in habitat selection: patterns concealed by pooled data. *Journal of Wildlife Management, 58*(2), 367–374.

Smith, K. L., & Jones, M. L. (2008). Allocation of sampling effort to optimize efficiency of watershed-level ichthyofaunal inventories. *Transactions of the American Fisheries Society, 137*(5), 1500–1506.

Smith, D. R., & Michels, S. F. (2006). Seeing the elephant: importance of spatial and temporal coverage in a large-scale volunteer-based program to monitor horseshoe crabs. *Fisheries, 31*(10), 485–491.

Stanfield, L. W. (2012). Reporting on the condition of stream fish communities in the Canadian tributaries of Lake Ontario, at various spatial scales. *Journal of Great Lakes Research, 38*, 196–205.

Stehman, S. V., & Overton, W. S. (1996). Spatial sampling. In S. L. Arlinghaus (Ed.), *Practical handbook of spatial statistics* (pp. 31–63). Boca Raton: CRC Press, Inc.

Stein, E. D., & Bernstein, B. (2008). Integrating probabilistic and targeted compliance monitoring for comprehensive watershed assessment. *Environmental Monitoring and Assessment, 144*(1–3), 117–129.

Stevens, D. L., Jr., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association, 99*(465), 262–428.

Stoddard, J. L., Peck, D. V., Paulsen, S. G., Van Sickle, J., Hawkins, C. P., Herlihy, A. T., Hughes, R. M., Kaufmann, P. R., Larsen, D. P., Lomnicky, G., Olsen, A. R., Peterson, S. A., Ringold, P. L., & Whittier, T. R. (2005). *An ecological assessment of western streams and rivers*. EPA 620/R-05/005. Washington, DC: U.S. Environmental Protection Agency.

Stranko, S. A., Hilderbrand, R. H., & Palmer, M. A. (2012). Comparing the fish and macroinvertebrate diversity of restored urban streams to reference streams. *Restoration Ecology, 20*, 747–755.

StreamNet (2014). StreamNet. *Pacific States Marine Fisheries Commission*. https://www.streamnet.org/. Accessed 5 May 2014

Svancara, L. K., Garton, E. O., Chang, K.-T., Scott, J. M., Zager, P., & Gratson, M. (2002). The inherent aggravation of aggregation: an example with elk aerial survey data. *Journal of Wildlife Management, 66*(3), 776–787.

Terra, B. D. F., Hughes, R. M., & Araujo, F. G. (2013a). Sampling sufficiency for fish assemblage surveys of Atlantic Forest streams, southeastern Brazil. *Fisheries, 38*, 150–158.

Terra, B. D. F., Hughes, R. M., Francelino, M. R., & Araujo, F. G. (2013b). Assessment of biotic condition of Atlantic Rain Forest streams: a fish-based multimetric approach. *Ecological Indicators, 34*, 136–148.

Thomas, C. E., & Parresol, B. R. (1989). Comparing basal area growth rates in repeated inventories: Simpson's paradox in forestry. *Forest Science, 35*(4), 1029–1039.

USEPA (U.S. Environmental Protection Agency). (2009). *National Lakes Assessment: a collaborative survey of the nation's lakes*. EPA 841/R-09/001.Washington, D.C.: Office of Water and Office of Research and Development.

USEPA (U.S. Environmental Protection Agency). (2013). *National Rivers and Streams Assessment 2008–2009: a collaborative survey*. EPA/841/D-13/001, Washington, D.C.: Office of Wetlands, Oceans and Watersheds and Office of Research and Development.

Van Sickle, J., & Hughes, R. M. (2000). Classification strengths of ecoregions, basins and geographic clusters for aquatic vertebrates in Oregon. *Journal of the North American Benthological Society, 19*, 370–384.

Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician, 36*(1), 46–48.

Walter, R. C., & Merritts, C. J. (2008). Natural streams and the legacy of water-powered mills. *Science, 319*, 299–304.

Washington State Department of Ecology. (2006). *Status and trends monitoring for watershed health and salmon recovery: quality assurance monitoring plan. Ecology Publication No. 06-03-203*. Olympia: Washington State Department of Ecology.

Whittaker, J. (1984). Model interpretation from the additive elements of the likelihood function. *Applied Statistics, 33*(1), 52–64.

Wong, D. (1996). Aggregation effects in geo-referenced data. In S. L. Arlinghaus (Ed.), *Practical handbook of spatial statistics* (pp. 83–106). Boca Raton: CRC Press, Inc.

Woody, C. A., Hughes, R. M., Wagner, E. J., Quinn, T. P., Roulsen, L. H., Martin, L. M., & Griswold, K. (2010). The U.S. General Mining Law of 1872: change is overdue. *Fisheries, 35*, 321–331.

Yoder, C. O., Rankin, E. T., Smith, M. A., Alsdorf, B. C., Altfater, D. J., Boucher, C. E., Miltner, R. J., Mishne, D. E., Sanders, R. E., & Thoma, R. E. (2005). In J. N. Rinne, R. M. Hughes, & B. Calamuss (Eds.), Historical changes in large river fish assemblages of the Americas (pp. 399–429). Bethesda, Maryland: American Fisheries Society.

Yule, G. U., & Kendall, M. G. (1950). *An introduction to the theory of statistics*. New York: Hafner Publishing Company.