Developing a Monitoring Program for Invertebrates: Guidelines and a Case Study

JASON R. ROHR,* ‡ CAROLYN G. MAHAN,† AND KE CHUNG KIM*

Penn State Institutes of the Environment and Department of Entomology, The Pennsylvania State University, 501 A.S.I. Building, University Park, PA 16802, U.S.A.

Division of Mathematics and Natural Sciences, Penn State Altoona, 209 Hawthorne Building, Altoona, PA 16601, U.S.A.

Abstract: Invertebrates provide the majority of ecosystem services; thus, it is important that they be inventoried, monitored, and protected. Nevertheless, inventories, monitoring, and management generally focus on vertebrates and flowering plants. Consequently, there are few guidelines or case studies for invertebrates. We present a procedure for developing a monitoring program for species-rich invertebrates that entails (1) characterizing the community; (2) identifying surrogates for biodiversity; and (3) establishing efficient methods to monitor surrogates and any ecologically important or sensitive taxa. We used these procedures, biodiversity-based statistical advances, and a survey of arthropods to develop a monitoring plan for the forests of Shenandoah National Park, Virginia (U.S.A.). Our case study revealed that mixed hardwood and hemlock forests had significantly different compositions of arthropods in their soil and understory strata. Of the 10 orders tested Coleoptera and Hymenoptera were the only two to pass most of the five surrogate tests, and their combination improved predictions of overall arthropod diversity. Because arthropods represent the majority of macroscopic species in most ecosystems, the ability of this assemblage to predict overall arthropod diversity makes it a powerful surrogate. Of the 11 collecting methods used, the beat-sheet method was the most efficient for monitoring this surrogate assemblage. To complement this coarse-filter approach to monitoring at-risk, invasive, or other important taxa (fine filter), we used ordination analyses to match 66 taxa with the methods that most effectively sampled them. Our methods serve as a model for developing an invertebrate monitoring plan and should facilitate linking such monitoring with ecosystem functions and management.

Keywords: arthropods, biodiversity surrogates, complementarity, Convention on Biological Diversity, higher taxa, morphospecies, ordination, rarefaction curves

Desarrollo de un Programa de Monitoreo de Invertebrados: Directrices y un Estudio de Caso

Resumen: Los invertebrados proporcionan la mayoría de los servicios ecosistémicos; por lo tanto, es importante que sean inventariados, monitoreados y protegidos. Sin embargo, los programas de inventarios, monitoreo y gestión generalmente se concentran en vertebrados y plantas con flores. Consecuentemente, hay pocas directrices o estudios de caso para invertebrados. Presentamos un procedimiento para el desarrollo de un programa de monitoreo para invertebrados que implica (1) caracterización de la comunidad; (2) identificación de sustitutos de la biodiversidad y (3) establecimiento de métodos eficientes para el monitoreo de sustitutos y cualesquier taxa sensible o ecológicamente importante. Utilizamos estos procedimientos, avances estadísticos basados en biodiversidad, y un muestreo de artrópodos para desarrollar un plan de monitoreo para los bosques del Parque Nacional Shenandoah (E.U.A.). Nuestro estudio de caso reveló que los bosques mixtos de maderas duras y abetos tenían composiciones de artrópodos significativamente diferentes en el suelo y en el sotobosque. De los 10 órdenes evaluados, Coleoptera e Hymenoptera fueron los únicos 2 que aprobaron la mayoría de las cinco pruebas de sustitutos, y su combinación mejoró las predicciones de la diversidad total de artrópodos. Debido a que los artrópodos representan la mayoría de las especies macroscópicas en casi
Introduction

In an effort to safeguard the ecosystem services that maintain natural life-support processes, the parties to the 2002 Convention on Biological Diversity (CBD) committed themselves "to achieve, by 2010, a significant reduction of the current rate of biodiversity loss at the global, regional and national levels" (UNEP 2002). This target was later endorsed by the leaders of 190 countries at the 2002 Johannesburg World Summit on Sustainable Development (WSSD). To distinguish true biodiversity declines or increases from natural fluctuations, there will need to be a surge in biodiversity inventories and monitoring projects. In most cases inventory and monitoring projects represent good investments. Conservative estimates of benefit-to-cost ratios based on the value of ecosystem functions indicate that detailed biodiversity surveys and monitoring and maintenance of biodiversity reserves are typically cost-effective (Balmford & Gaston 1999).

Most management and monitoring, however, take place on a much smaller scale than the global, regional, and national emphases of CBD and WSSD's entreaty (Balmford et al. 1996a, 1996b; Reid 1998). Furthermore, management and monitoring have conventionally focused on charismatic vertebrates and flowering plants despite the fact that hyperdiverse groups, such as invertebrates, fungi, and microbes, represent a higher proportion of the diversity that drives the fundamental ecosystem processes that the CBD and WSSD seek to protect (e.g., Kremen et al. 1993; Nee 2004). This taxonomic bias might be partly due to the lack of integrative case studies to guide natural area managers in proactively, effectively, and efficiently inventorying and monitoring invertebrates (but see Kremen 1992, 1994).

Our goal was to provide natural areas managers with a heuristic model and guidelines for developing an effective and efficient monitoring program for hyperdiverse taxa, with a focus on terrestrial arthropods. The guidelines we present stem directly from Noss (1990) and Kremen et al.'s (1993) frameworks for biodiversity monitoring. To bridge the gap between research and practice (Prendergast et al. 1999), we provide a cohesive case study in which we implement recent statistical advances that can make monitoring more efficient.

Our aim of providing a general blueprint for developing a monitoring program for species-rich groups is intentionally broad in an attempt to link local monitoring with CBD and WSSD's interest in maintaining natural life-support processes, to encourage ecosystem rather than species management (Noss 1996), and because of recent emphases on coarse-filter approaches for monitoring invertebrates (Wilcove & Master 2005). Nevertheless, monitoring projects often have more-restricted objectives, such as providing early warnings against anthropogenic threats (Kremen 1994). More-specific goals can be met with our general guidelines in combination with the monitoring recommendations provided by others. We also defer to others for guiding principles on monitoring-program implementation and the evaluation and dissemination of monitoring and management results (e.g., Stork & Samways 1995; Elzinga et al. 2001; Yoccoz et al. 2001; Green et al. 2005).

In our case study, we developed a monitoring program for the forests of Shenandoah National Park, Virginia (U.S.A.), partly because U.S. national parks are mandated to monitor their natural resources (Boone et al. 2005). This program was based on a structured, nearly complete inventory of arthropods in all strata (sensu Longino & Colwell 1997) of mixed hardwood and eastern hemlock forests, two of the park's primary forest types. We focused our survey on terrestrial arthropods because they play vital roles in fundamental ecosystem processes, possess enormous species and functional diversity, are often sensitive to environmental change, are in dire need of monitoring, and, in some areas, represent the demonstrable majority of the most threatened species (e.g., Kellert 1993; Kim 1993; Goldstein 2004). Indeed, most past and predicted extinctions are of insects (Dunn 2005). Although monitoring terrestrial arthropods has not traditionally been considered cost-effective (Margules & Austin 1991), recent advances have now made monitoring arthropods more efficient (e.g., Hammond 1994; Fisher 1999; Andersen & Majer 2004).
Methods

Three Steps to Developing a Monitoring Program

We propose that developing a monitoring program should entail three steps: (1) characterizing the community (inventory); (2) identifying valid surrogates for biodiversity; and (3) establishing efficient methods to monitor the surrogates and ecologically important or sensitive taxa.

A baseline characterization of biodiversity (inventory) is essential to document its temporal change, the fundamental goal of environmental monitoring. Although there are often existing biodiversity databases for vertebrates (e.g., natural heritage programs; see www.natureserve.org), these data rarely exist for invertebrates. Thus, tracking invertebrate taxa often requires generating baseline data through surveys (Kremen et al. 1993). The goal of these surveys should be community characterization (sensu Colwell & Coddington 1994) or the estimation of species richness, complementarity, and the distribution of species abundances (Kremen et al. 1993; Longino & Colwell 1997). That is, when possible, baseline data should be acquired through a structured inventory in which relative abundances are emphasized, which contrasts with a strict inventory (sensu Longino & Colwell 1997) in which relative abundances may be of minor importance. In a structured inventory replicated samples are stratified with respect to a set of variables, such as method, habitat, or time and attempts are made to identify specimens to species (Longino & Colwell 1997). Ideally, samples should be collected using a variety of methods and should span a diversity of habitats, taxa, and seasons.

Once a thorough community characterization has been conducted, these data can be used to identify surrogates of biodiversity. Surrogates are necessary because complete species enumeration of speciose taxa is often impractical (Kremen et al. 1993). Although the reliability of surrogates for biodiversity should be validated by demonstrating a positive correlation between the surrogate and a larger group of interest or an environmental gradient, more often than not surrogates go untested (Noss 1999). Surrogates can be assessed at a variety of organizational levels (Noss 1990). For example, richness of higher taxonomic ranks can be used to predict richness of lower ranks (reviewed by Gaston 2000), richness of one group can be used to predict richness of another (Oliver & Beattie 1996; Fisher 1999), and surrogates can be used as indicators of environmental change (Kremen 1992, 1994).

Few attempts have been made to identify surrogates for overall arthropod diversity because few researchers have conducted surveys of all arthropods (see exceptions in Oliver & Beattie 1996). A taxon that is correlated with overall arthropod diversity would likely capture more macroscopic biodiversity than any other surrogate because, in most ecosystems, arthropods represent the majority of macroscopic species (Groombridge 1992; Samways 2005). Our study represents one of the few nearly complete inventories of arthropods, which allowed us to rigorously test for surrogates of arthropod diversity.

Management and monitoring projects are ultimately constrained by budgets, and thus it is crucial that these projects be cost-effective (Margules et al. 2002). Identifying efficient monitoring designs may be even more important for hyperdiverse taxa due to the sheer number of individuals and species, the difficulty of identifying specimens to species, the multitude of methods that can be used for their sampling, and the paltry funds available for their inventory and monitoring. For invertebrates, specimen processing and identification can be more time consuming and expensive than specimen collecting. Therefore, it is important to identify methods that collect predominantly surrogates so that time is minimized in processing nonsurrogates. By identifying the method that most efficiently samples validated surrogates, researchers or managers might also increase their statistical power to detect temporal changes because each sample will have proportionally more surrogates (see Gibbs et al. 1998 for discussion of power).

Efficiency of lone or combined sampling methods can be evaluated by comparing the predicted rates of taxa accrual as a function of the number of individuals sampled, relationships that are referred to as rarefaction curves (Gotelli & Colwell 2001). More efficient sampling methods collect many taxa but few individuals and thus have steeper rarefaction curves. In addition to comparing sampling methods, rarefaction curves can be used to evaluate inventory completeness and to assess whether inventory efficiency can be improved by stratifying sampling methods in space or time (Longino & Colwell 1997). We demonstrate how sample-based rarefaction curves can identify efficient collecting methods and sampling designs for monitoring surrogates.

It might be judicious to complement surrogate monitoring, which is intended to provide a pulse on overall biodiversity, with more-specific monitoring of ecologically or economically important taxa, such as invasive, keystone, threatened, or endangered species (Noss 1999 provides other examples). Monitoring would then occur at multiple hierarchical levels using both coarse- and fine-filter approaches (Noss 1990; Kremen et al. 1993; Wilcove & Master 2005). We demonstrate how ordination techniques can be used to develop sampling-method maps that match taxa with the methods that most effectively sample them. These maps can be used to design monitoring plans for important or sensitive species.

Case Study

We studied a mixed hardwood and hemlock forest in Shenandoah National Park in the southern Appalachians.
of northern Virginia. Each forest contained a single first-order stream and had similar elevation, slope, aspect, stand size, and perimeter (Mahan et al. 1998). Eight 20 × 20 m plots were placed on each side of the stream 10 m from its edge and spaced 25 m apart (16 plots in each forest; Mahan et al. 2004).

Parataxonomists, trained in entomological techniques, conducted arthropod field sampling and specimen identification to the family and morphospecies levels. Acari, Pseudoscorpiones, and Protura were the only arthropod orders not identified to family, and these three orders plus classes Collembola, Julida, Lithobiomorpha, Neuroptera, and family Cecidomyiidae (order Diptera) were the only taxa not identified to morphospecies. Genus- and species-level identifications were conducted by taxonomic specialists (list in Mahan et al. 2004). All specimens were deposited in The Pennsylvania State University Frost Entomological Museum.

Field Sampling Methods

We only briefly describe our sampling methodology because it is detailed in Mahan et al. (1998, 2004) and the implementations of most of our methods are described by Southwood and Henderson (2000). Types, numbers, and strata distribution of our 11 collecting methods are described in Table 1. Collecting methods were distributed randomly among the plots. Within each plot soil-based sampling was conducted 5 m from the plot center in a randomly selected direction; vegetation-based methods were conducted on a randomly selected tree or shrub of the dominant species in each plot. Understory and canopy were defined as 0.5 to 2 m and >7 m above ground, respectively.

We buried pitfall traps (9-cm diameter) flush with the soil surface. Each contained 2 cm of salt water and was left open for 4 days but was emptied daily. Soil cores (5 × 5 cm) and leaf-litter samples (areas from 0.25 m²) were placed in sealed bags and extracted using a modified Tullgren funnel technique. Substrate searches were conducted by the same experimenters in each plot and entailed collecting arthropods while turning rocks, logs, and leaves in a 5-m radius for 10 minutes. Sweep-net sampling was conducted by sweeping a net (37.5-cm diameter) through low-lying vegetation along a 10-m transect. Branch clips entailed enclosing 0.5 m of foliage-bearing understory (2-m high) and canopy (8-m high) branches in a 60-L plastic bag, clipping the branch, and later enumerating the arthropods in the bag. Beat-sheet sampling entailed placing a 1-m² sheet below vegetation, beating the vegetation 10 times, and collecting all arthropods that fell on the sheet. Trunk traps were aluminum flashing sealed around tree trunks that funneled crawling arthropods into collecting jars. Malaise traps, open-sided tent-like structures, concentrated arthropods in collecting jars, were open for 48 hours, and were placed in potential arthropod flyways.

All collecting occurred from 19 August 1997 to 22 August 1997 and the same type and number of collecting methods were used each day in each forest to control for temporal confounders. The extensive taxonomic scope of the survey (focus on the majority of arthropod taxa rather than a small subset) and the limited budget precluded extensive temporal breadth. This can affect the appropriate selection of surrogates. Thus, when possible, preparation for monitoring programs should involve sampling across seasons.

Data Analysis

By displaying the most salient relationships among taxa, samples, and environmental gradients in a limited number of dimensions, ordination represents one of many powerful multivariate statistical techniques that can facilitate distillation of the inherent complexity of most management and monitoring data sets. Ordination can be dichotomized into indirect and direct gradient analyses. Indirect gradient analyses sequentially identify hypothetical environmental gradients (axes) that give the best fit to

<table>
<thead>
<tr>
<th>Collection methods (abbrev)</th>
<th>Strata</th>
<th>Samples/forest type</th>
<th>No. of specimens</th>
<th>No. of families</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitfall trap (PF)</td>
<td>soil</td>
<td>11</td>
<td>455</td>
<td>48</td>
</tr>
<tr>
<td>Leaf-litter sample (LL)</td>
<td>soil</td>
<td>15</td>
<td>3005</td>
<td>71</td>
</tr>
<tr>
<td>Soil core (SC)</td>
<td>soil</td>
<td>15</td>
<td>1365</td>
<td>37</td>
</tr>
<tr>
<td>Substrate search (SS)</td>
<td>soil</td>
<td>10</td>
<td>210</td>
<td>27</td>
</tr>
<tr>
<td>Beat sheet (BS)</td>
<td>understory</td>
<td>5</td>
<td>202</td>
<td>47</td>
</tr>
<tr>
<td>Malaise trap (GM)</td>
<td>understory</td>
<td>6</td>
<td>2601</td>
<td>81</td>
</tr>
<tr>
<td>Lower branch clip (LB)</td>
<td>understory</td>
<td>6</td>
<td>50</td>
<td>19</td>
</tr>
<tr>
<td>Sweep net (SN)</td>
<td>understory</td>
<td>3</td>
<td>150</td>
<td>39</td>
</tr>
<tr>
<td>Tree trunk trap (TT)</td>
<td>understory</td>
<td>5</td>
<td>465</td>
<td>54</td>
</tr>
<tr>
<td>Malaise trap (CM)</td>
<td>canopy</td>
<td>2</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>Upper branch clip (UB)</td>
<td>canopy</td>
<td>6</td>
<td>58</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>—</td>
<td>84</td>
<td>8636</td>
<td>167</td>
</tr>
</tbody>
</table>
the remaining variation in the distribution of taxa among samples. Direct gradient analyses use the same approach but constrain the analyses to quantified environmental variables. The two approaches compliment one another (Lepš & Šmilauer 2003) because the degree of congruence between the direct and indirect approach provides insight into whether important environmental variables were not quantified.

Ordination methods, such as eigen analysis-, distance-, or nonmetric-based approaches, are invaluable but are also complex. There are user-friendly, menu-driven statistical programs available (e.g., CANOCO and PC-ORD) and user guides that are written for neophytes (e.g., McCune & Grace 2002; Lepš & Šmilauer 2003) and the more statistically savvy (e.g., ter Braak & Šmilauer 2002). We encourage resource managers that are hesitant to use or learn these tools to collaborate with individuals who might be more experienced.

We specifically used ordination to reveal taxa that might be unique to, or indicators of, particular habitat types. We first used detrended correspondence analysis (DCA; selected because the average response to the hypothetical environmental gradients was unimodal) to identify hypothetical gradients that accounted for the greatest proportion of variation among our samples. The DCA was followed by redundancy analysis (RDA; selected because the average response to environmental variables was monotonic), a direct gradient analysis constrained by the following categorical environmental variables: forest type (hemlock, hardwood), strata (soil, understory, and canopy), and their statistical interaction (and blocked by collection method). We tested the significance of each of these variables with a Monte Carlo simulation (999 iterations with a forward stepwise selection procedure; ter Braak & Šmilauer 2002). We applied a variant of the Hellinger transformation to the relative abundance data before conducting the RDA on mean abundance scores (Legendre & Galaghei 2001) and conducted a log (+1) transformation before the DCA. All ordination analyses were conducted in CANOCO (version 4.5; ter Braak & Šmilauer 2002).

The goal of the surrogate analyses was to determine whether “shortcuts” could be made in assessing arthropod diversity. For example, could identifying specimens to family or genus accurately predict species richness or could quantifying the richness of a portion of arthropods predict total arthropod diversity? We used a series of hierarchical regression analyses (with collecting methods as samples) to identify potential surrogates. First we tested whether family and genus richness of major arthropod orders was predictive of their species and morphospecies richness. Relationships between species and higher taxonomic ranks were only tested for the four orders thoroughly identified to species by taxonomic specialists (Aranea, Coleoptera, Diptera, Plecoptera). Second, we tested whether the family richness of arthropod orders was predictive of the cumulative family richness of all other orders. Third, we tested whether the family richness of arthropod orders was predictive of the pooled morphospecies richness of all other orders, an analysis across both orders and taxonomic ranks.

Collecting methods that catch many arthropods are likely to be rich in both species and higher taxa, whereas methods that catch few arthropods are likely to be poor in both (Gaston 2000). In an effort to control for this confounding effect of capture rates, we regressed log$_{10}$-transformed richness on log$_{10}$-transformed arthropod captures and used the residuals from these analyses when conducting the surrogate analyses described above (Balmford et al. 1996a).

Surrogates should also ideally identify natural and anthropogenic gradients so that they might serve as bioindicators or early warnings of system change (Noss 1990; Pearson 1994). We redid the RDA analyses described above on 10 major arthropod orders to assess the utility of each order as a bioindicator of the known environmental and biotic gradients identified by the entire arthropod data set. Because the significance and percentage of variation accounted for by axes or predictors can be viewed as a measure of information content, we used these criteria to evaluate the quality of a surrogate in identifying the known environmental gradient (Kremen 1994). Although it is beyond the scope of this study, these analyses can and should be applied to anthropogenic gradients to identify surrogates that can warn of anthropogenic change (e.g., Kremen 1994).

We used rarefaction curves to assess inventory completeness and sampling method efficiency. All rarefaction curves were smoothed by averaging 50 random reorderings of the samples (as in Colwell & Coddington 1994) in Estimates 7.0 (Colwell 2005). We extrapolated the rarefaction curves with a logarithmic model (Soberón & Llorente 1993) to estimate how many additional families would have been provided by a doubling of sampling effort. We then compared the efficiency of lone collecting methods and each possible combination of soil and understory methods for the surrogates selected with the steps above. Efficiency was compared at the estimated “true” or “asymptotic” richness of each method (with the Chao 2 richness estimator provided by EstimateS) and at the arthropod abundance collected by the method that caught the fewest arthropods. In the efficiency analyses nonsurrogate taxa were lumped into a single hypothetical family. Thus, after the first nonsurrogate was collected, each additional nonsurrogate captured reduced efficiency by increasing the number of arthropods, but not the number of families, collected. Because substantial time is required to sort surrogates from nonsurrogates, this approach intentionally penalizes methods that caught many nonsurrogates. In all efficiency analyses we excluded branch clips and canopy malaise traps because they caught few arthropods (Table 1) and substrate searches because they can be biased by collectors.
If the taxonomic composition of closely spaced samples is more similar than those that are far apart, inventory and monitoring efficiency might be improved by increasing the distance between samples (Colwell & Coddington 1994). We therefore tested for a relationship between distance among plots and their taxonomic similarity. This was done for each of our four soil-based collecting methods because they offered the greatest number of samples per method (Table 1). The Morisita-Horn index of complementarity for all pairs of samples was calculated using EstimateS 7.0 (Colwell 2005), and the average of these indexes for each distance class was regressed against distance between plots (n = 6, the two greatest distance classes were pooled due to few sample pairs). These analyses can also be conducted with the more computationally intensive Mantel’s test.

To match collecting methods with the arthropod families they best targeted, we conducted a second RDA with the same methods described above except that forests were our replicates and the analyses were constrained to our 11 collecting methods. This analysis therefore associates sampling methods with the arthropod families they most frequently collected. Only the 66 families with at least 10 specimens in the database were included in this analysis.

Results

We collected 8636 arthropods (Table 1). Twenty-seven of the 30 arthropod orders collected were identified to the family level, and individuals in 156 of the 167 identified families (Table 1) were assigned to morphospecies. Genus and species identification is incomplete predominantly due to a lack of available taxonomic specialists. Detailed taxonomic lists and abundance data are available in Mahan et al. (2004). Extrapolating the family rarefaction curves with the logarithmic model ($R^2 > 0.9983$) revealed that a 100% increase in sampling effort (84 additional samples) would provide an estimated 22% (180 total) and 19% (132 total) increase in the number of families found in hardwood and hemlock forests, respectively.

The DCA showed there were detectable differences in the taxonomic composition of samples generated by different collecting methods. In DCA distance among samples approximates their taxonomic dissimilarity. Thus, soil-based methods were more complementary to (farther apart from) one another than the understory methods (Fig. 2a). The first DCA axis appears to be a strata axis because samples collected from soil are on the right-hand side of the figure followed by samples collected from the understory and then the canopy (Fig. 2a). Within understory collecting methods there were frequently observable differences in community composition between forest types. This pattern was less evident for soil-collecting methods. The detectable differences between forests and strata in the DCA suggest these are both important environmental variables worthy of further study in RDA.

In the RDA biplot (Fig. 2b) distance from the origin indicates the relative importance of the environmental variables and taxa, and distance among the nominal environmental variables approximates the taxonomic dissimilarity of their associated samples. The rank abundance of species with respect to environmental variables and correlations among species are approximated by projecting the environmental variables or species arrowheads perpendicularly onto the vector of the focal species. For instance, if one extends the “Le Noct” taxon arrow in Fig. 2b in both directions and perpendicularly projects the nominal environmental variables and species onto this vector, one discovers that this taxon was most frequently caught in the understory of the hardwood forest followed by the soil and canopy layers of this forest and was negatively correlated with the abundance of “Ps Peri Peri” (because it projected far away from the “Le Noct” tip on the opposite side of the origin).

The RDA revealed significant differences in the taxonomic composition of these hardwood and hemlock forests. Twenty-one taxa were disproportionately found in the hardwood forest as opposed to four in the hemlock forest (Fig. 2b). This is consistent with the significantly greater number of families sampled in the hardwood than hemlock forest (Fig. 1). The first canonical axis in the RDA was represented by differences in the understory of these forests, consistent with the strong separation of these forests in the understory methods of the DCA (Fig. 2a). The second axis represented the significant differences between the forests in taxa collected from the soil stratum (Fig. 2b).

![Figure 1](image-url)  
*Figure 1. Sample-based rarefaction curves for the accumulation of arthropod families in hardwood and hemlock forests of Shenandoah National Park, Virginia, U.S.A. To reduce clutter confidence bands are not provided. Nevertheless, the 95% intervals are provided for 84 samples, the maximum number for each forest alone.*
There were only minor differences in surrogate (Tables 2 & 3) and efficiency analyses (available upon request from J.R.R.) between forests, so we highlight the results for the two forests combined. Genus and family richness were, in general, predictive of species richness for the four arthropod orders thoroughly identified to species (Table 2). In each case the predictive power of genus richness was greater than family richness. The family richness of seven out of eight of the tested orders was a significant predictor of their morphospecies richness (Table 3).

Coleoptera and Hymenoptera were the only two orders with significant positive correlations between their family richness and the pooled family richness of all other arthropod orders (Table 3). The family richness of Coleoptera plus Hymenoptera was a considerably better predictor of the pooled family richness of all other arthropod orders than either order alone. Hymenoptera and Psocoptera were the only two orders to pass the more rigorous test of predicting diversity across both taxa and ranks. In each case their family richness was a significant, but weak, predictor of the pooled morphospecies richness of all other orders (Table 3).

Surrogates should also be capable of detecting known environmental heterogeneities, which might suggest that they are capable of serving as early warnings of system change (Noss 1990; Pearson 1994). Of the 10 orders tested (those in Table 3), Coleoptera and Hymenoptera were the only two capable of revealing the known compositional differences (gradients) between the soil and understory of hemlock and hardwood forests that were identified by the RDA conducted on the entire data set (Fig 2b). Coleoptera increased the variance accounted

---

**Figure 2.** Results from (a) partial detrended correspondence analysis (controlling for total number of arthropods collected in each sample) for the distribution of the invertebrate samples from each collecting method (closed symbols, hardwood forest; open symbols, hemlock forest) and (b) from redundancy analysis displaying taxa disproportionately found in the understory and soil of hardwood and hemlock forests. In both panels, taxa scores are centered and standardized and post-transformed. In (a) polygons enclose the majority of samples for the collecting methods, and only some of the branch clips are shown (X’s, crosses, and striped squares). For panel (b) environmental variables were multiplied by 0.30 to fit in the coordinate system, axes one and two account for 20% and 12% of the taxa abundances and 61% and 32% of the fitted taxa abundances, respectively, and only significant probability values (p < 0.05) and the 25 arthropod taxa with the best fit to the displayed axes (with ≥10 specimens in the database) are shown. The first two letters of the taxa arrows represent the first two letters of their order, the next four are the first four letters of their family, followed by the first four letters of their genus, followed by the first four letters of their species. Full names of the orders and families are in Fig. 4, and full names of genera and species are in Maban et al. (2004).

---

### Table 2. The significance of genus and family richness as predictors of species richness for the four orders thoroughly identified to species by taxonomic specialists.

<table>
<thead>
<tr>
<th>Order</th>
<th>( r(p)^a )</th>
<th>( n_b )</th>
<th>no. of genera</th>
<th>no. of families</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aranea</td>
<td>0.96 (0.001)</td>
<td>10</td>
<td>0.70 (0.015)</td>
<td></td>
</tr>
<tr>
<td>Coleoptera</td>
<td>0.97 (0.001)</td>
<td>9</td>
<td>0.61 (0.082)</td>
<td></td>
</tr>
<tr>
<td>Diptera</td>
<td>0.99 (0.001)</td>
<td>6</td>
<td>0.85 (0.033)</td>
<td></td>
</tr>
<tr>
<td>Plecoptera</td>
<td>1.00</td>
<td>6</td>
<td>0.89 (0.019)</td>
<td></td>
</tr>
</tbody>
</table>

---

*aResults are presented for the two forest types combined because there were no substantial differences between forests.*

*bCollecting methods served as the replicates; methods that did not collect any arthropods within the focal order were excluded from the analyses, accounting for the different sample sizes.*
for between the soil strata of hemlock and hardwood forest by 29%, and Hymenoptera more than doubled the chances of detecting compositional differences in both the soil and understory of these forests (increased variance accounted for by 229% and 136%, respectively). This was despite Coleoptera and Hymenoptera only comprising 17% and 25% of the families and 3% and 7% of the specimens, respectively.

The beat-sheet method was the most efficient method for sampling the Coleoptera–Hymenoptera assemblage, and there was an efficiency reduction (see “true” richness) when this method was combined with any of the soil methods (Fig. 3). These results were generally consistent whether we used interpolation or extrapolation techniques (Fig. 3). Ground malaise traps also appeared efficient at sampling the surrogate assemblage if paired with a soil method, and efficiency rankings of methods often depended on the number of arthropods collected (Fig. 3). There was no significant relationship between the distance between samples and the similarity of their taxonomic composition for any of the four soil methods in either forest ($-0.35 < r < 0.67$, $n = 6$, $p > 0.142$).

The RDA analysis (Fig. 4) indicated that leaf-litter samples in hemlock forests might improve sampling efficiency and help monitor taxa that might have narrow habitat requirements (Fig. 4).

### Discussion

We have built on the biodiversity monitoring framework of others (Noss 1990; Kremen et al. 1993) in an attempt to facilitate monitoring invertebrates, a group that often contributes greatly to ecosystem processes. Community characterization, surrogate validation, and efficient sampling design are crucial components to developing a program for monitoring invertebrates, but they have not been synthesized previously within a monitoring context. Consequently, we integrated these three key concepts with a case study and statistical advances that make monitoring design more efficient. Our monitoring framework and case study included coarse- and fine-filter approaches, provided methods for identifying bioindicators of natural and anthropogenic gradients, emphasized efficiency, and should be applicable to any invertebrate group (and perhaps vertebrates), aspects of monitoring that are championed by many (e.g., Noss 1990; Kremen 1994; Margules et al. 2002; Wilcove & Master 2005).

Our results suggest that future monitoring of the summer arthropod diversity in the forests of Shenandoah National Park could benefitted by evaluating the efficiency and effectiveness of collecting methods.
Repeated sampling of this assemblage at the family level or below should facilitate detecting compositional change. Gibbs et al. (1998) offer insights on the frequency and intensity of sampling needed to detect a given percent change, and the use of geographic information systems with time series of environmental layers should further facilitate monitoring efforts. Focusing on surrogates might be supplemented by placing leaf-litter samples in hemlock forests to target species unique to this habitat or by using Fig. 4 to identify methods efficient at monitoring other ecologically important taxa.

Only four arthropod orders in our data set were thoroughly identified to species by taxonomic specialists (despite having funds to pay specialists), perhaps owing to the lack of prestige and resources for taxonomy that is crippling the discipline worldwide (Godfray 2002). Consequently, the majority of our analyses could only be conducted at the family level. But the family level might actually be most relevant for management and monitoring purposes because, for most invertebrates, the funds and expertise are not presently available to readily identify individuals below this taxonomic level (exceptions might be at-risk, invasive, or pest species). Thus, it might be most feasible to monitor invertebrates at family or morphospecies levels. For the four orders that were identified to species, both family and genus richness were, in general, significant predictors of species richness, but genus richness was a stronger predictor. These results are consistent with tests of the higher-taxon approach (using higher taxonomic ranks to predict lower ranks) in plants and vertebrates (Balmford 1996a, 1996b; Gaston 2000) and with the only other study that tested the approach on a highly speciose group, the macromycete fungi (Balmford et al. 2000). More tests of the higher-taxon approach might help compensate for the taxonomic impediments associated with monitoring hyperdiverse groups at the genus or species levels.

The use of multiple, hierarchical surrogate tests conducted on numerous taxa is important because it can reveal both the strengths and weaknesses of many prospective surrogates (Pressey 2005). Coleoptera and Hymenoptera had few limitations based on the battery of surrogate tests we conducted. In fact, the predictive abilities of Hymenoptera for these forests and collection times crossed ranks and taxa, the first reported case of which we are aware. Although Coleoptera and Hymenoptera alone provided little precision in predicting the family richness of the remaining orders, combining these two orders considerably increased the predictive power (Table 3), which supports the use of taxonomic assemblages as recommended by others (Noss 1990; Kremen et al. 1993). These two orders also satisfy many additional requirements of bioindicators: they are abundant, functionally diverse, taxonomically rich, and have higher taxa broadly distributed over a breadth of habitat types and lower taxa that are specialized and sensitive to environmental change (Noss 1990; Kremen et al. 1993; Pearson 1994).
Most importantly, because the majority of multicellular biodiversity are composed of arthropods (Groombridge 1992), this surrogate assemblage has the potential to be predictive of the majority of macroscopic, summer biodiversity for these forests. Perhaps not surprisingly families of both Coleoptera and Hymenoptera have been the focus of many monitoring and conservation projects (Pearson 1994; Oliver & Beattie 1996; Niemela 2000; Andersen & Majer 2004). Although the desirable qualities of an indicator assemblage depend on the goals of a management project (e.g., inventory, monitoring, reserve network design; see Kremen et al. 1993), our results are part of growing evidence suggesting that Coleoptera and Hymenoptera have the indicator properties necessary to meet many of these goals.

Developing cost-effective monitoring designs for surrogates requires comparing the efficiency of available sampling methods and determining the optimal distribution of sampling over space and time. Our analyses revealed that monitoring efficiency, here defined as surrogate families accumulated per specimen collected, can differ among habitats and surrogates, increase, decrease, or remain unchanged when methods are combined, and depend on the number of organisms sampled (Figs. 1 & 3). These general conclusions stem from the fact that sampling methods accumulate taxa at different rates in different habitats and that complementarity of sampling methods can depend on the habitat and targeted taxa (Longino & Colwell 1997).

The spatial distribution of monitoring methods represents a compromise between the degree of independence among samples, the time and costs associated with spacing sampling stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our sampling stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats. We detected no relationship between taxonomic similarity and distance for our soil-sampling methods, which suggests that having stations far apart, and the efficiency of stratifying samples among habitats.

Figure 4. Results of a redundancy analysis matching 66 invertebrate taxa (table) represented by at least 10 specimens in the database with the collection method that best targets each taxa. The biplot also depicts correlations among taxa. Only significant probability values (p < 0.05) are shown (calculated with a Monte Carlo permutation test). Scale marks along the axes apply to the collecting methods. The family scores were multiplied by 0.49 to fit in the coordinate system. The biplot displays 40% of the variance in the family abundances and 56% of the variance in the fitted family abundances. Diamonds, squares, and triangles represent canopy, understory, and soil-collecting methods, respectively, and arthropod taxa are depicted by crosses. See Table 1 for an explanation of the collecting-method abbreviations and text for how to interpret RDA biplots.
Although we did not incorporate a seasonal component into our survey (e.g., Rohr et al. 2002, 2003, 2005), determining the efficiency of temporal stratification would be analogous to the assessment for spatial stratification (Oliver & Beattie 1996; Longino & Colwell 1997). Nevertheless, it should be noted that the temporal sampling in this study is likely not representative of what would be needed to meet many monitoring goals. Guidelines for determining temporal sampling intensity to detect specific levels of change in population sizes or diversity are provided by Gibbs et al. (1998). Meaningful levels of change in populations or diversity are likely best determined on a case-by-case basis.

Most of our efficiency improvements were identified by visually inspecting rarefaction curves, which appears to be the predominant method for evaluating efficiency of inventory and monitoring designs (Longino & Colwell 1997; Fisher 1999). Analyses based on rarefaction curves assume equal sorting and identification time for each taxon collected and the same purchase and implementation costs for each sampling method used, assumptions that are rarely met (Lawton et al. 1998). Furthermore, it would simply be too time intensive to develop all the rarefaction curves necessary to determine the best combination of our 11 collecting methods in space and time. There is clearly a need for the development of a computer model to more accurately and efficiently identify optimal monitoring designs based on a definition of efficiency that more explicitly addresses cost-effectiveness.

The more efficient monitoring becomes, the more time and money there will be for including greater taxonomic breadth into our monitoring programs, for improving our management practices (Kremen et al. 1993), and for better understanding important species interactions and community processes (e.g., Rohr & Crumrine 2005). Ultimately, there needs to be a stronger link between monitoring biodiversity, monitoring ecosystem services, and managing ecosystems (Kremen 2005; Pressey 2005). Changes in biodiversity and services should inform and improve management practices with an emphasis on prediction and prescription, the cornerstones of adaptive management (Holling 1978). To successfully reduce biodiversity losses by, or soon after, 2010, and to better understand drivers of ecosystem functions, we will have to efficaciously monitor invertebrates and the services they provide and integrate this monitoring with proactive ecosystem management (Balmford et al. 2005). Our case study and three-step blueprint for designing a monitoring program for invertebrates should help reach these goals.

Acknowledgments

We thank B. Black, J. Boone, R. Byers, J. Gream, A. Schrot, K. Sullivan, and R. P. Withington for research assistance. This project was partially funded by grants to C.M. from the National Park Service (NPS), to K.K. and J.R.R. from NPS and the Department of Military and Veteran Affairs, and to J.R.R. from the National Science Foundation (DEB 0516227) and U.S. Department of Agriculture (2006–01370). We are especially grateful for the cooperation of J. Akerson, J. Hovis, G. Hunt, J. Karish, and R. Kremenaker, and the helpful comments of L. Byrne, R. Colwell, D. Johnson, C. Kremen, E. Main, and two anonymous reviewers.

Literature Cited


Godfray, H. C. J. 2002. Challenges for taxonomy—the discipline will have to reinvent itself if it is to survive and flourish. Nature 417:17–19.


Green, R. E., A. Balmford, P. R. Crane, G. M. Mace, J. D. Reynolds, and R.