# Discriminatory power of different arthropod data sets for the biological monitoring of anthropogenic disturbance in tropical forests

YVES BASSET[1,*], JACQUES F. MAVOUNGOU[2], JEAN BRUNO MIKISSA[3], OLIVIER MISSA[4], SCOTT E. MILLER[5], ROGER L. KITCHING[6] and ALFONSO ALONSO[7]

[1]*Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Ancon, Panama City, Republic of Panama;* [2]*Institut de Recherche en Ecologie Tropicale I.R.E.T/CENAREST, B.P. 13354, Libreville, Gabon;* [3]*Ecole Nationale des Eaux et Forêts, B.P. 3960, Libreville, Gabon;* [4]*Australian School of Environmental Studies, Griffith University, Brisbane, QLD 4111, Australia;* [5]*Department of Systematic Biology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560-0105, USA;* [6]*Cooperative Research Centre for Tropical Rainforest Ecology and Management, Australian School of Environmental Studies, Griffith University, Brisbane, QLD 4111, Australia;* [7]*Smithsonian Institution/Monitoring and Assessment of Biodiversity Program, 1100 Jefferson Drive, S.W. Suite 3123, Washington, DC 20560-0705, USA;* *Author for correspondence (e-mail: bassety@tivoli.si.edu; fax: +507-212-8148)*

**Abstract.** Arthropods were monitored by local parataxonomists at 12 sites of increasing anthropogenic disturbance (old and young secondary forests, savanna and cultivated gardens) at Gamba, Gabon. We report on the discriminatory power of different data sets with regard to the classification of sites along the disturbance gradient, using preliminary data accounting for 13 surveys and 142 425 arthropods collected by Malaise, pitfall and yellow-pan traps. We compared the performance of different data sets. These were based upon ordinal, familial and guild composition, or upon 22 target taxa sorted to morphospecies and either considered *in toto* or grouped within different functional guilds. Finally we evaluated 'predictor sets' made up of a few families or other target taxa, selected on the basis of their indicator value index. Although the discriminatory power of data sets based on ordinal categories and guilds was low, that of target taxa belonging to chewers, parasitoids and predators was much higher. The data sets that best discriminated among sites of differing degrees of disturbance were the restricted sets of indicator families and target taxa. This validates the concept of predictor sets for species-rich tropical systems. Including or excluding rare taxa in the analyses did not alter these conclusions. We conclude that calibration studies similar to ours are needed elsewhere in the tropics and that this strategy will allow to devise a representative and efficient biotic index for the biological monitoring of terrestrial arthropod assemblages in the tropics.

## Introduction

Since water pollution is often transient and unpredictable, biological monitoring may be more appropriate than traditional chemical evaluation of water quality to assess contamination of aquatic ecosystems (Guérold 2000). Plants, which are traditionally used as measures of habitats as well as disturbance in terrestrial

systems (e.g., Watt 1998), may be less useful than invertebrates for the biological monitoring of aquatic systems. This pragmatic reason has driven the development of community-level analyses of the effects of anthropogenic disturbance on invertebrates in aquatic ecosystems (e.g., Clarke 1993; Rossaro and Pietrangelo 1993; Growns et al. 1997; Thorne and Williams 1997; Guérold 2000). The effects of pollution on freshwater invertebrate communities can be calculated routinely as an 'index of biological integrity' (Karr 1991), and different statistical methods are available for the study of environmental impact on marine communities (Warwick 1993). In contrast, such recipes and consensus are almost non-existent for terrestrial arthropods, particularly in the tropics (see Urzelai et al. (2000) for nematodes and O'Connell et al. (1998) for birds).

It is important to stress that the ultimate goal of deriving a biotic index based on terrestrial arthropods is to monitor the effects of anthropogenic disturbance *per se* on arthropods (Basset et al. 1998). Arthropods represent a substantial proportion of all terrestrial biodiversity and, accordingly, their responses to disturbance are important. Documenting which species reacts, and how, to varying disturbance levels is more significant than assessing and monitoring habitat quality, which is rather trivial and can easily be assessed by vegetation censuses alone (Watt 1998). The sheer number of arthropods, however, leads to major challenges in fulfilling this goal. Specifically, difficulties attendant upon the study of terrestrial arthropods include (a) the high diversity of terrestrial assemblages; (b) the high diversity and complexity of terrestrial habitats (McGeogh 1998), with concomitant logistical problems in sampling (Kitching et al. 2001); and (c) the taxonomic impediment (Kitching 1993a). By way of example, compare the 56 families of macrobenthos collected by Thorne and Williams (1997) at various locations in the tropics, with the 222 insect families encountered by Stork (1991) in a single event using insecticide fogging in the crowns of 10 Bornean trees.

It is not surprising that most studies of anthropogenic disturbance on terrestrial invertebrates in the tropics focus on well-known, less speciose, taxa, usually restricted to the family level and to a specific feeding guild (e.g., Belshaw and Bolton 1993; Eggleton et al. 1996; Hill 1999; Intachat et al. 1999; Vasconcelos 1999; Davis et al. 2001; McGeogh 1998). This approach may be overly restrictive given the absence of any consensus on the appropriate choice of 'indicator' or 'umbrella' species, especially in the tropics (e.g., Landres et al. 1988; Prendergast et al. 1993; Hammond 1994; Lawton et al. 1998; McGeogh 1998; Kotze and Samways 1999; Basset et al. 2001a; Moritz et al. 2001). Kitching (1993b) and Didham et al. (1996) have advocated the use of 'predictor sets' which comprise taxa representative of different functional groups ('guilds') as an alternative to the use of taxa selected on the basis of taxonomic tractability or familiarity (see also Collins and Thomas (1991) and Kremen et al. (1993) for similar arguments). Such 'predictor sets' are properly selected only following statistical analysis of a larger, more or less complete, data set including all taxa and the catches from several complementary sampling methods. Some studies have indeed widened their taxonomic focus to a whole order or a few families of different orders, thereby, often unintentionally, including representatives of different guilds (e.g., Kremen 1992; Didham et al. 1998;

Kotze and Samways 1999; Chung et al. 2000; Kitching et al. 2000). Lawton et al. (1998) went further and included both invertebrate and vertebrate orders in their study of forest disturbance in Cameroon. They noted the huge effort necessary to implement such approaches efficiently.

Recently, a novel approach that relies on the training and input of local paraxonomists has allowed the considerable widening of entomological investigations in the tropics (e.g., Janzen and Gauld 1997; Basset et al. 2000; Novotny et al. 2002). Properly used, this strategy yields higher numbers of statistical replicates that are, accordingly, more representative of the system studied (Basset et al. 2000). Adequate statistical replication represents a significant obstacle in conservation studies of highly complex environments, such as tropical rainforests. The parataxonomist strategy employed in Guyana enabled us to achieve one of the first Before–After/Control-Impact experiments, demonstrating unequivocally the influence of selective logging on rainforest insects, despite the excessively low insect densities in the study system (Basset et al. 2001a).

Presently, about 100 000 valid species of insects are known from the Afrotropical region, but even conservative estimates, such as the scenario of Gaston and Hudson (1994), may see this number increase to about 600 000 species (Miller and Rogo 2002). Basic ecological information on described species of Afrotropical arthropods is fragmentary and often relates to a few localities only. The level of availability of this information also varies greatly from one taxon to another. Gaps in knowledge are evident, even for well-studied taxa (Miller and Rogo 2002).

Entomological studies in Gabon have been few and follow these trends. The few recent checklists available for higher taxa are restricted to groups that are not particularly speciose, such as Mantodea (Roy 1973); Haliplidae and Dytiscidae (Bilardo and Rocchi 2000); Lucanidae (Maes and Pauly 1998); Brentidae (Bartolozzi and Sforzi 1997); and Apoidea (Pauly 1998). There are no recent reviews of agricultural and timber pests in Gabon, although Coccoidea and, in particular, the cassava mealybug and its parasitoids, have been well studied (e.g., Boussienguet et al. 1991). Ecological studies are likewise infrequent and targeted at a few groups such as cockroaches, dung beetles, fig wasps, bees or ants (Walter 1987; Anstett et al. 1995; Grandcolas 1997; Roubik 1999; Wetterer et al. 1999). The need for baseline information on Gabonese arthropods is obvious.

The work presented in this paper is part of a wider project aimed ultimately at providing baseline entomological data and the assessment of anthropogenic disturbance on local arthropod faunas within the Gamba Complex, Gabon. Arthropod activity is being monitored by trained and supervised local parataxonomists in four distinct habitats of increasing degrees of anthropogenic disturbance. The taxonomic scope of the material collected is large and includes representatives of several feeding guilds. The structure and scope of this project have few equivalents to date in tropical Africa.

In the present report, we consider a preliminary (but by no means small) data set which we use to analyse the discriminatory power of different subsets of data (including orders, families and morphospecies representative of different feeding guilds) against the disturbance gradient. We explore the question of whether a few

ecologically selected taxa ('predictor sets') may be suitable for the biological monitoring of anthropogenic disturbance at Gamba, as surrogates for the entire local fauna. In doing so, we also evaluate the need to sort specimens to a specific level (as opposed to sorting the material to familial or ordinal levels) and examine the contribution of rare species to the discriminatory power of different data sets.

## Methods

### Study area and sites

The study area was in the Shell–Gabon oil concession of Gamba, within the Gamba Complex of Protected Areas in southeastern Gabon (approximately 2°43′ S, 10°1′ E, 25 m a.s.l.; see Thibault and Blaney (2001) and Doumenge et al. (2001) for background information about the area). The Gamba oil field includes a mosaic of old growth secondary rainforests, younger secondary rainforests and savanna areas. The latter result mainly from anthropogenic action. Primary rainforests are absent from the Gamba oil field, following the selective logging of Okoumé (*Aucoumea klaineana* Pierre), mostly over the past 20 years, but these forests are found elsewhere in the Gamba Complex. Botanical information about the area is available in Prins and Reitsma (1989). The mean annual temperature in the area is 26 °C and annual rainfall ranges between 2000 and 2400 mm per year, with the major wet season from September to December (Prins and Reitsma 1989). The Gamba oil field has been active since 1967 and Gamba has grown from a small village in 1960 to a town of 8000 inhabitants (Bourgeais 2001). The earliest cultivated crop gardens of relative size were established near the town as recently as 1998 (see Bourgeais (2001) for a summary of environmental concerns in the Gamba Complex).

 We considered four distinct habitats of increasing anthropogenic disturbance (i.e., increasing forest clearing and introduction of exotic vegetation) and selected three sites (replicates) within each habitat. The four habitat types were: (a) the interior of old secondary rainforests, 'old forest'; (b) the edge of young secondary rainforests, 'young forest'; (c) area of rainforest cleared to install oil rigs and subsequently invaded by savanna, 'savanna'; and (d) cultivated crop gardens, 'gardens'. Differences between these habitats were obvious and readily noticeable to a non-biologist. The main characteristics of the study sites (coded A–L) are indicated in Table 1. Sites G–H and I were abandoned and active oil wells, respectively. They were established in 1980, 1980 and 1968, respectively. There were no old or recent oil spills at these three sites and none was burned during the study period. As far as possible, sites were spread through the concession, within an area of approximately 13 × 11 km. The northwestern part of the concession, however, is more forested and, accordingly, most forest sites occurred there. The shortest distance between sites was ca. 600 m (sites B and G), whereas the longest distance between sites was ca. 15 km (sites A and F).

*Table 1.* Main characteristics of study sites within the Shell–Gabon Gamba oil field (for gardens, the main crops cultivated during the study period are listed).

| Code | Habitat | Location and coordinates (Shell–Gabon landmark) | Fragment size (ha) | Physiognomy | Vegetation characteristics |
|---|---|---|---|---|---|
| A | Old forest | Near TOT004 02°42'21" S, 09°59'43" E | 700 | Secondary forest, tallest trees 45 m, sandy soil | *Neochevalierodendron stephanii* (A. Chevalier) Léonard dominant, *Diospyros zenkeri* (Gurke) F. White and *D. ?vermoeseni* De Wild common |
| B | Old forest | Near TOT006 02°42'52" S, 09°59'53" E | 84 | Secondary forest, tallest trees 45 m, sandy soils | *Neochevalierodendron stephanii* dominant, *Diospyros zenkeri, D. ?vermoeseni* and *Pialisota ambigua* CB. Clarke common |
| C | Old forest | Between GA031 and GA032 02°44'27" S, 10°00'07" E | 28 | Secondary forest, tallest trees 40 m, but many small trees 10–20 m tall, sandy soil | *Diospyros ?vermoeseni* and *D. conocarpa* Gurke ex K. Schum common, *P. ambigua* and *Trichoscypha acuminata Engler* less common |
| D | Young forest | Between GA022 and GA026 02°45'35" S, 10°01'32" E | 12 | Secondary forest, tallest trees 20 m, many small trees and bushes, sandy soil | *Pialisota ambigua, Aframomum* sp. and *Rauvolfia* sp. common; one pioneer *Musanga cecropioides* R. Br. ex Tedlie present |
| E | Young forest | Near IV034 02°46'02" S, 10°02'15" E | 19 | Secondary forest, very open canopy, tallest trees 30 m, swampy soil | *Xylopia hypolampra* Mildb. and *Xylopia* spp. dominant |
| F | Young forest | Near IV006 02°47'31" S, 10°03'40" E | 166 | Secondary forest, plot at the edge of a thin tongue of forest connected to a large forested area; tallest trees 30 m, important re-growth in the understorey, sandy soil | *Pachypodanthium staudtii* Engl. & Diels, *Diospyros ?vermoeseni, Pialisota ambigua, Leptactina mannii* Hook.f., *Ouratea sulcata* (Van Tiegh.) Keay, *Sacoglottis gabonensis* (Baillon) Urb. and *Bertiera subsessilis* Hiern present |
| G | Savanna | TOT006 02°42'49" S, 09°59'51" E | 2.7 | Surrounded by forest; isolated bushes and trees, sandy soil, bare soil = 50% | *Borreria verticillata* (L.) GFW Mey and two unidentified Poaceae dominant, *Cyperus tenax* Boeck and *Dracaena* sp. present |
| H | Savanna | GA054 02°44'09" S, 10°00'19" E | 3.0 | Surrounded by forest, sandy soil, bare soil = 25% | *Borreria verticillata, Dracaena* sp. and one unidentified Poaceae dominant, *Cyperus halpan* J. Kern and *Heterotis decumbens* (Pal. Beauv.) H. Jacques-Félix present |
| I | Savanna | IV013 02°48'21" S, 10°03'16" E | 2.5 | Surrounded by forest, sandy soil, bare soil = 25% | *Merremia tridentata* Hallier f., *Cyperus tenax* and one unidentified Poaceae dominant |
| J | Garden | School IRAF[a] 02°44'47" S, 10°01'06" E | 2 | Sandy soil fertilised with compost | Amaranth, aubergine, cabbage, carrot, lettuce, pepper, spinach, sweet pepper, tomato and water melon |
| K | Garden | Mr Ebala-Ebala 02°44'06" S, 10°01'59" E | 0.5 | Clayish sand fertilised with compost | Aubergine, banana, maize, manioc, pepper, pineapple, spinach, sugar cane and taro |
| L | Garden | Mr Talla 02°44'38" S, 10°00'59" E | 0.8 | Sandy soil fertilised with compost | Amaranth, aubergine, cabbage, cucumber, gombo, pepper, sorrel, spinach and tomato |

[a]Institut de Recherches en Agronomie Forestière.

*Arthropod collecting and processing*

Each site was equipped with a similar array of traps targeting the flying and epigeic arthropods of the understorey and litter. At each site, these traps included: one ground Malaise trap, four yellow-pan traps set up on the ground and five pitfall traps buried in the ground. In addition, four flight-intercept traps were also set up in forested sites, but for ease of comparison among all sites, these data are not discussed further. The collecting methods were selected because they are simple, inexpensive, and behave more or less consistently across sites with respect to the profile of arthropods collected (Kitching et al. 2001).

The collecting surface of one Malaise trap was 2.7 m$^2$ (model similar to Townes (1972); Santé Traps, Lexington, Kentucky). Collecting fluid was 70% ethanol. Yellow-pan traps were 27 cm in diameter by 8 cm deep and filled with a mixture of water (ca. 80%), 70% ethanol (ca. 20%) and a few drops of liquid detergent to break the surface tension of the water. Pitfall traps were small 0.5-l plastic cups and filled with the same water, ethanol and detergent mixture. At each site, a Malaise trap occupied the center of the set of traps, with four pitfall traps established to the north, south, east and west, 10 m distant from the Malaise trap. Four yellow-pan traps were set up at equal distances between the pitfall traps, again 10 m distant from the Malaise trap. The fifth pitfall trap was established 30 m north of the Malaise trap.

The 120 traps were operated for 3 days and surveyed weekly (= one survey), by a team of six local parataxonomists, who were trained specifically for this project (see Basset et al. (2000) for similar training in Papua New Guinea and Guyana). Although the project is ongoing, we consider here preliminary data from the first 13 surveys obtained from July to November 2001.

The material collected was first sorted to families or higher taxa (see exceptions below) by the parataxonomists in a laboratory equipped with four microscopes and two computers. The material, belonging to 22 target taxa (Table 2), was isolated and pinned (or otherwise dry-mounted), with each individual identified by a unique specimen number. The target taxa were eventually sorted by morphospecies (i.e., unnamed species diagnosed using standard taxonomic techniques, *sensu* Cranston 1990) by the parataxonomists. Formal taxonomic study of this material is ongoing. The 22 target taxa were selected using three main criteria: (a) they were well represented in the samples; (b) they were workable taxonomically and sought after by taxonomists who had expressed interest in the material; and (c) they represented a variety of functional guilds and orders (Table 2). All target taxa of beetles further represented 'focal groups' *sensu* Hammond (1994), that is, they were abundant and easily sorted. Taxa were assigned to the following feeding guilds, following Moran and Southwood (1982) and Stork (1987): chewers, sap-suckers, pollinators, epiphyte grazers, fungal-feeders, insect predators, other predators, parasitoids, wood-eaters, scavengers, ants, 'tourists' (i.e. non-feeding residents), and 'unknown' (i.e. taxa for which the sorting resolution did not allow assignment of the material to a particular guild; these were not retained for analyses). All of the resulting information was data-based using the program BIOTA (Colwell 1997). Voucher specimens will be deposited at the Smithsonian Institution (Washington, DC) and in an institution in Gabon to be determined.

*Table 2.* The 22 target taxa selected, with the total number of individuals collected and morphospecies sorted for the first 13 surveys.

| Order | Target taxa | Guild | Individuals | Morphospecies |
|---|---|---|---|---|
| Mantodea | Mantodea[a] | Predators | 9 | 3 |
| Orthoptera | Acridoidea[a] | Chewers | 205 | 13 |
| Hemiptera | Fulgoroidea[b] | Sap-suckers | 1401 | 94 |
| Hemiptera | Membracidae | Sap-suckers | 13 | 7 |
| Coleoptera | Cerambycidae | Wood-eaters | 124 | 25 |
| Coleoptera | Chrysomelidae | Chewers | 536 | 85 |
| Coleoptera | Buprestidae | Wood-eaters | 16 | 5 |
| Coleoptera | Scarabaeidae | Chewers + scavengers | 421 + 449 | 32 + 28 |
| Coleoptera | Coccinellidae | Predators | 540 | 18 |
| Coleoptera | Histeridae | Predators | 287 | 10 |
| Coleoptera | Cleridae | Predators | 8 | 4 |
| Coleoptera | Tenebrionidae | Scavengers | 389 | 21 |
| Diptera | Asilidae | Predators | 145 | 22 |
| Diptera | Dolichopodidae | Predators | 3263 | 38 |
| Diptera | Syrphidae | Predators | 110 | 17 |
| Diptera | Tephritidae | Chewers | 72 | 9 |
| Diptera | Pipinculidae | Parasitoids | 6 | 3 |
| Neuroptera | Neuroptera[c] | Predators | 54 | 9 |
| Hymenoptera | Ichneumonidae | Parasitoids | 995 | 80 |
| Hymenoptera | Chalcidoidea[d] | Parasitoids | 1453 | 73 |
| Hymenoptera | Formicidae | Ants | 37 202 | Not yet sorted |
| Hymenoptera | Apoidea[c] | Pollinators | 273 | 25 |
| Totals | – | – | 47 971 | 621 |

[a] One family; [b] 11 families; [c] six families; [d] 12 families.

There were a few exceptions to the sorting and mounting pattern. Non-insect material was mostly sorted to order. Adults of Lepidoptera were not sorted to families, since this wet material was useless. Some individuals, notably within the Diptera (both acalypterate and calypterate) and the Cucujoidea, were also not assigned to families. Chalcidoidea smaller than 2 mm were sorted, but not mounted or morphotyped. Formicidae is the only target taxon that has yet to be sorted to morphospecies.

*Statistical methods*

As sampling effort was identical at each site, we have pooled the results of all traps at each site for the analyses presented here (i.e. we have summed the occurrence of each taxon at each site). We examined the discriminatory power of 11 data sets (the matrix of taxa × sites with number of individuals as the elements in each case) against the classification of the sites along the disturbance gradient. First, we considered higher taxa (orders and families) and guilds. Second, we considered all morphospecies of target taxa that belonged to a particular guild. Data sets which included enough morphospecies for the analyses included: chewers, parasitoids, predators, pollinators, sap-suckers, scavengers and wood-eaters. Subsequently, we considered a data set incorporating all morphospecies from all guilds. Sap-suckers

included only phloem-feeders and predators included only insect predators (see Table 2).

Rare species represent a statistical challenge when comparing different samples, as the information provided by singletons is very low (Novotny and Basset 2000). 'Rare' species may be subsisting at genuinely low population levels or they may be apparently rare because (a) they are poorly collected by the chosen sampling methods; (b) sampling effort has been insufficient; (c) they may be seasonal or have restricted timing of activity; or (d) they may have been collected from marginal habitats ('vagrant species', see Novotny and Basset 2000). We analyzed each of the 11 data sets both with and without the rare species. In the latter case we considered only taxa that were represented by 12 or more individuals (i.e. an average of one individual collected at each site). We present here detailed results for the data sets that exclude rare taxa (see Discussion). There were insufficient data to justify an analysis of the wood-eating guild after the exclusion of rare morphospecies.

We considered four statistics that describe the discriminatory power of the different data sets, and calculated them for data sets both with and without rare taxa:

(a) *Standard deviation of multivariate scores*: Non-metric multi-dimensional scaling has been advocated as a powerful method to study the similarity of samples ('$\beta$-diversity') in species $\times$ samples matrices, as the similarity among samples can be computed using a variety of distances or similarity indices (Clarke 1993). However, the arbitrariness of orientation of the axes extracted is a serious drawback when one is primarily interested in comparing ordinations (Clarke 1993), as is the case here. Accordingly, we have chosen to use a more straightforward approach by quantifying $\beta$-diversity using detrended correspondence analysis (DCA) with Hill's scaling, using untransformed data (Ter Braak and Smilauer 1998). The differences between the scores of any two sites on the first axis of the DCA represent a measure of species turnover between these two sites. The standard deviation of the scores on Axis 1 also represents a measure of the spread of the value within the sites. These analyses were computed using the programme CANOCO (Ter Braak and Smilauer 1998).

(b) *Modified beta diversity of Whittaker*: We calculated $\beta$-diversity ($\beta$-2) (*sensu* Whittaker 1960) for each data set, using the modified formula of Harrison et al. (1992):

$$\beta-2=\{[(S/\alpha_{max})-1]/(N-1)\}\times100$$

where $S$ is the total number of species collected, $\alpha_{max}$ is the maximum within-taxon richness per sample and $N$ is the number of samples per taxon (12 sites in this case). $\beta$-2 is insensitive to species richness and ranges from 0 (no turnover) to 100 (every sample has a unique set of species).

(c) *ANOSIM*: We used a non-parametric analysis of similarity (ANOSIM) to test for differences in the rank similarities of sites grouped by habitats. This is analogous to a one-way analysis of variance (Clarke 1993). ANOSIMs were calculated based upon Bray–Curtis distances and their significance was tested with 5000 random permutations using the programme BioDiversity Pro (McAleece et al. 1997).

(d) *Non-parametric analysis of variance on multivariate scores*: We also performed non-parametric analyses of variance (Kruskal–Wallis tests) on the scores of Axis 1 of the DCA grouped by habitats.

Statistics (a) and (b) reflect the spread of the sites (the higher the better, the decrease of $\beta$-diversity being a measure of the loss of information), whereas statistics (c) and (d) evaluate the classification of sites along the disturbance gradient (the more significant the better, the decrease in significance being also a measure of the loss of information). We compared the first two statistics for each data set both with and without rare species by Wilcoxon signed-rank tests. Since the inclusion of additional taxa in the data sets could increase their discriminatory power, we tested the independence between the number of taxa included in the analyses and the first two statistics by regression, for all data sets.

The final set of analyses evaluated the usefulness of predictor sets (Kitching 1993b) for our data. To identify the taxa to be included in the predictor sets, Kitching et al. (2000) used principal component analysis (PCA) to obtain a linear combination of taxa scores that could be used as an indicator of habitat quality. In the present account we are less concerned about deriving an index of habitat quality than we are about exploring which taxa should be considered in calculating a putative index and, indeed, whether this is a practical strategy at all. The two-way indicator species analysis (TWINSPAN) further represents an alternative method to PCA, free of the assumptions of linearity. Another method, the indicator value index (INDVAL), has also been proposed as an alternative which overcomes several limitations of TWINSPAN (Dufrêne and Legendre 1997; see McGeogh et al. (2002) for an application). In short, this indicator index for a particular species is independent of the relative abundance of other species (comparisons between different species may be impeded by different catchability, activity patterns, behaviors, etc.), and there is no need to use 'pseudospecies'. The significance of each species indicator value is tested by a site randomisation procedure (Dufrêne and Legendre 1997). INDVAL was computed for the two best data sets (see Results: families for higher taxa and all target taxa for morphospecies), with and without rare species. We used PC-ORD, with Monte Carlo permutation tests (1000 permutations; McCune and Medford 1999) for this analysis.

Finally, we combined taxa with significant indicator value indices ($P < 0.05$) in new matrices for families and target taxa, including and excluding rare species, and re-calculated the four statistics (a)–(d) above, for these derived data sets. We have termed these 'indicator families' and 'indicator target taxa', respectively.

## Results

*Overall results*

In total, 142 425 arthropods were collected during the 13 surveys, which included 29 orders and at least 175 families. Different trap types yielded a similar number of individual arthropods (Malaise traps: 55 108, pitfalls: 45 571, yellow pans: 41 746).

*Table 3.* Distribution of guilds (sum of individuals collected) among the different habitats studied, for the first 13 surveys.

| Variable | Old forest | Young forest | Savanna | Gardens |
|---|---|---|---|---|
| Ants | 15751 | 7858 | 3246 | 10347 |
| Chewers | 369 | 445 | 674 | 972 |
| Epiphyte grazers | 327 | 358 | 446 | 671 |
| Fungal-feeders | 979 | 1457 | 822 | 2288 |
| Insect predators | 571 | 907 | 1294 | 4046 |
| Other predators | 662 | 642 | 436 | 329 |
| Parasitoids | 1112 | 961 | 1000 | 2473 |
| Pollinators | 1 | 8 | 70 | 194 |
| Sap-suckers | 1263 | 1740 | 701 | 4380 |
| Scavengers | 6061 | 8109 | 3131 | 10866 |
| Tourists | 8867 | 10588 | 5152 | 13580 |
| Wood-eaters | 261 | 2518 | 229 | 789 |
| Unknown | 412 | 1079 | 423 | 560 |
| Total individuals | 36625 | 36672 | 17624 | 51504 |
| Total morphospecies | 203 | 233 | 163 | 317 |

Since ants, the major catch in pitfall traps, have not yet been sorted to morphospecies, the total number of individuals represented by morphospecies was lower in the pitfall than in other types of traps (Malaise traps: 2768, pitfalls: 626, yellow pans: 3620). The material assigned to morphospecies resulted in the identification of 621 units and represented 7.6% of the total material collected (Table 2). The taxonomic composition of the material collected will be presented elsewhere, but we noted the collection of two specimens of Dilaridae (Neuroptera), representing the first record of this family in Gabon and, indeed, in tropical Africa (one species is known from South Africa: Oswald 1998).

During the study period, gardens yielded more individuals and morphospecies than, notably, savanna habitats (Table 3). The yields from old and young forests were similar, but with apparent differences in the guild structure between these two habitats (Table 3). Differences among the various types of habitat studied in terms of productivity, species richness and guild structure are indisputable, but will be discussed elsewhere, with more appropriate data.

*Data sets including higher taxa*

Among higher taxa (orders, families and guilds), the data set with the most discriminant power was that at the family level, with or without rare taxa (Figure 1, Table 4; DCA plots of data sets including rare taxa are not shown). The standard deviation of the DCA scores on Axis 1 (35.1% of the total inertia) was relatively high for this data set and its Kruskal–Wallis test was significant. However, this data set, with or without rare families, did not effectively classify the sites along the disturbance gradient (Figure 1), its β-diversity was low and its ANOSIM not significant (Table 4). The data set based on guilds was the worst of the three data sets and, in general, the inclusion or exclusion of rare taxa identified few differences within these data sets (Table 4).
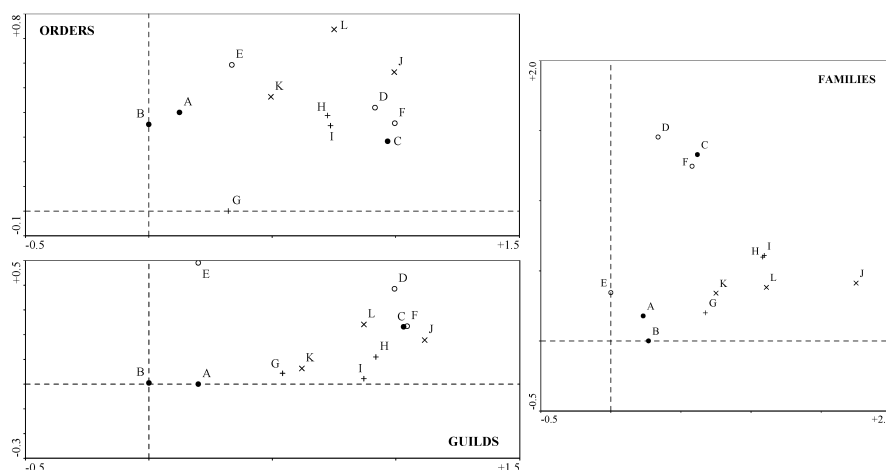
*Figure 1.* Plot of study sites within the plane formed by the first two axes of DCA for higher taxa and groups (orders, families and guilds), excluding rare taxa. Sites are identified by their codes (Table 1) and corresponding habitats are indicated by the following symbols: ● old secondary forest; ○ young secondary forest; + savanna; and × gardens.

*Table 4.* Discriminatory power of the different data sets, with and without rare taxa: number of taxa included in the analyses (*n*), standard deviation of the site scores on Axis 1 of the DCA (SD), results of non-parametric tests for score differences between habitats (Kruskal–Wallis tests: *W* and *P*), $\beta$-diversity ($\beta$-2) and results of ANOSIM (*R* and *P*).

| Data set | Excluding rare taxa | | | | | | | Including rare taxa | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | n | SD | W | P | $\beta$-2 | R | P | n | SD | W | P | $\beta$-2 | R | P |
| Orders | 20 | 0.350 | 2.69 | 0.442 | 0.48 | 0.40 | 2.680 | 29 | 0.350 | 2.69 | 0.442 | 2.37 | 0.40 | 2.320 |
| Families | 108 | 0.488 | 8.64 | 0.034 | 0.83 | 0.49 | 1.340 | 175 | 0.489 | 8.64 | 0.034 | 3.64 | 0.49 | 1.220 |
| Guilds[a] | 14 | 0.383 | 1.26 | 0.740 | 0.00 | 0.46 | 1.079 | – | – | – | – | – | – | – |
| Target taxa | 95 | 1.834 | 9.97 | 0.019 | 5.06 | 0.79 | 0.039 | 621 | 2.074 | 9.97 | 0.019 | 16.92 | 0.81 | 0.039 |
| Chewers | 14 | 1.413 | 9.46 | 0.024 | 4.55 | 0.74 | 0.039 | 140 | 2.226 | 9.97 | 0.019 | 16.36 | 0.83 | 0.019 |
| Parasitoids | 18 | 1.252 | 9.15 | 0.027 | 5.30 | 0.73 | 0.019 | 155 | 1.464 | 9.67 | 0.022 | 15.63 | 0.79 | 0.019 |
| Pollinators | 6 | 0.742 | 6.33 | 0.096 | 1.52 | 0.50 | 2.220 | 25 | 0.841 | 6.33 | 0.096 | 7.14 | 0.50 | 1.739 |
| Predators | 33 | 1.414 | 9.46 | 0.024 | 1.62 | 0.80 | 0.039 | 121 | 1.508 | 9.46 | 0.024 | 11.66 | 0.81 | 0.039 |
| Sap-suckers | 15 | 1.220 | 9.15 | 0.027 | 6.06 | 0.60 | 0.059 | 101 | 1.429 | 9.05 | 0.029 | 18.73 | 0.64 | 0.079 |
| Scavengers | 9 | 2.572 | 8.95 | 0.030 | 4.55 | 0.58 | 0.400 | 49 | 2.643 | 9.05 | 0.029 | 15.66 | 0.59 | 0.299 |
| Wood-eaters[b] | – | – | – | – | – | – | – | 30 | 1.186 | 11.00 | 0.012 | 21.21 | 0.41 | 1.739 |
| Indicator families | 26 | 0.766 | 10.39 | 0.016 | 0.76 | 0.78 | 0.019 | 27 | 0.761 | 10.39 | 0.016 | 1.14 | 0.78 | 0.019 |
| Indicator target taxa | 32 | 1.254 | 10.39 | 0.015 | 1.68 | 0.90 | 0.019 | 44 | 1.458 | 10.39 | 0.016 | 1.72 | 0.92 | 0.019 |

[a] Data set identical with or without rare guilds. [b] Insufficient data for the analyses excluding rare wood-eating morphospecies.

## Data sets including morphospecies

Data sets that included morphospecies belonging to different guilds had different discriminatory power. They ranked as follows (from highest to lowest): chewers, parasitoids, predators, sap-suckers, scavengers, wood-eaters and pollinators, again regardless of whether or not rare morphospecies were included (Figure 2, Table 4).

In spite of the relatively low number of morphospecies ($n = 14$), the data set on chewers classified the sites along the disturbance gradient effectively (Figure 2). The Spearman rank correlation coefficient between the scores of Axis 1 of the DCA (without rare species, 29.9% of total inertia) and habitats, coded as a categorical variable, was 0.907 ($P < 0.001$) and its ANOSIM was significant (Table 4). The data set for parasitoids and predators also performed well, with slightly more inconsistencies in the ordering of the different sites. Scavengers isolated forested from non-forested sites remarkably well, with a concomitant high standard deviation, but fared less well in classifying sites at a finer scale (Figure 2, Table 4). This outcome was similar to that for wood-eaters when rare morphospecies were included (Table 4; DCA plot not shown). No bees at all were collected at several forested sites and this greatly diminished the discriminatory power of the pollinator data set. In general, the discriminatory power of the data sets was affected little by the inclusion or exclusion of rare morphospecies (Table 4).

*Indicator value index and predictor sets*

Not surprisingly, the discriminatory power of the data set which included all target taxa was the greatest of all data sets examined, with or without rare morphospecies, and its ANOSIM was significant (Figure 3, Table 4; Spearman correlation between the scores of Axis 1 of the DCA (excluding rare species, 22.5% of total inertia) and habitats, $r_s = 0.950$, $P < 0.001$).

Within the two categories of data sets (based on higher and lower taxa, respectively) the two data sets with the greatest discriminating power were those based on 'families' and 'all target taxa'. Accordingly, we calculated indicator value indices separately for each of these two data sets (Table 5). Again, data sets excluding or including rare taxa showed few differences. Most indicator taxa (families or morphospecies) had their optima in the garden habitat, whereas the savanna habitat was more difficult to characterise. This habitat could be distinguished by the occurrence of the family Geometridae (geometrid larvae were sorted but adults were not, cf. Methods) and of one morphospecies of Dolichopodidae (predator). Seven families were shared by our initial list of target taxa and by indicator lists (compare Tables 2 and 5). Many indicator morphospecies belonged to the predator or parasitoid guilds, as those guilds included the highest number of morphospecies sorted from our samples (Table 5, including rare species). However, no particular guild had indicator morphospecies across all habitats studied (Table 5). We note that several popular indicator taxa are conspicuously missing from Table 5, either because (a) the sampling methods did not target them (i.e., adult butterflies, moths and dragonflies); (b) they were not included in the sorting process (Araneae, with the exception of Salticidae); or (c) the family-level information that they provided was not very informative for the scale of disturbance studied (Formicidae, dung-feeding Scarabaeidae). Table 5 nevertheless included other popular indicator taxa, such as Collembola, Carabidae, Mycetophilidae and Apoidea (see review of indicator taxa in, e.g., Collins and Thomas 1991).

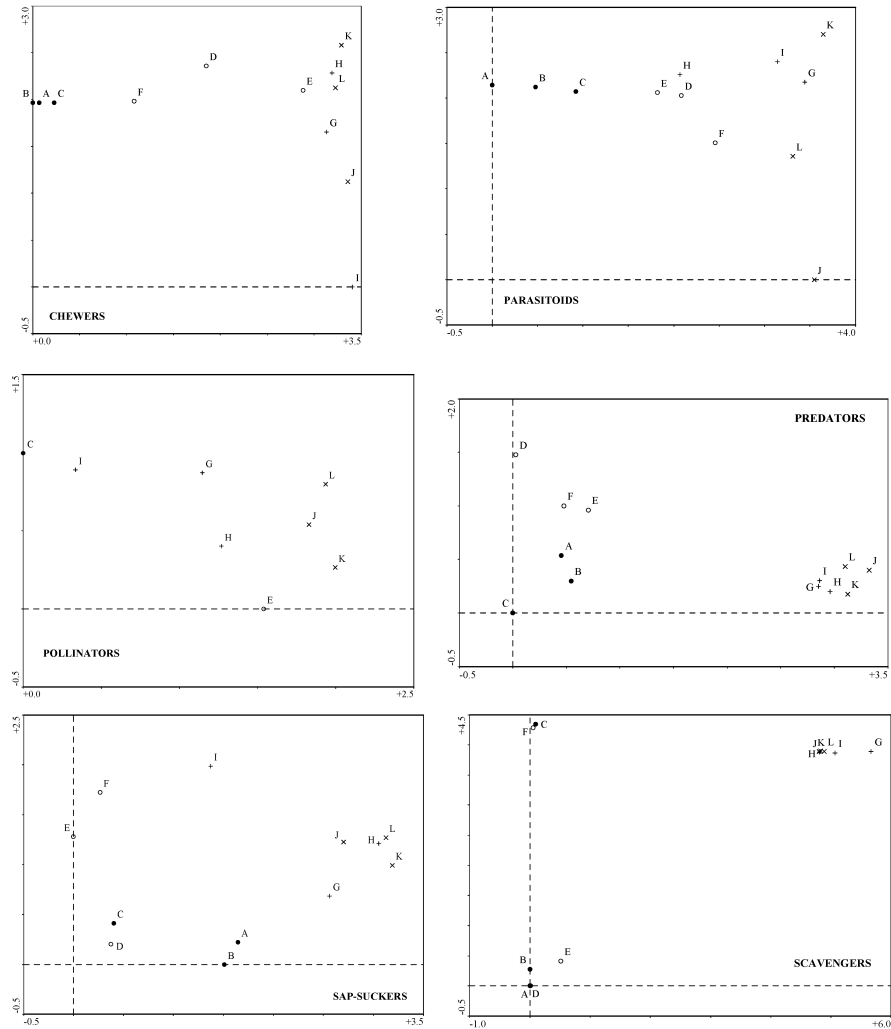The newly combined data sets of indicator families and target taxa proved to have

*Figure 2.* Plot of study sites within the plane formed by the first two axes of DCA for different guilds (chewers, parasitoids, pollinators, predators, sap-suckers and scavengers), excluding rare morphospecies. Presentation follows Figure 1. No pollinators were collected in sites A, B, D and F.

as much discriminating power as those which included all target taxa, and their ANOSIMs were also significant (Figure 3, Table 4). Overall, we regard the 'best' data set as being the indicator target taxa (Spearman correlation between the scores of Axis 1 of the DCA (excluding rare species, 33.6% of total inertia) and habitats, $r_s$ = 0.972, $P < 0.001$). However, the discriminatory power of the data set based on indicator families was also surprisingly good (excluding rare families: Axis 1 = 50.6% of total inertia, correlation between scores and habitats, $r_s = 0.777$, $P < 0.01$).
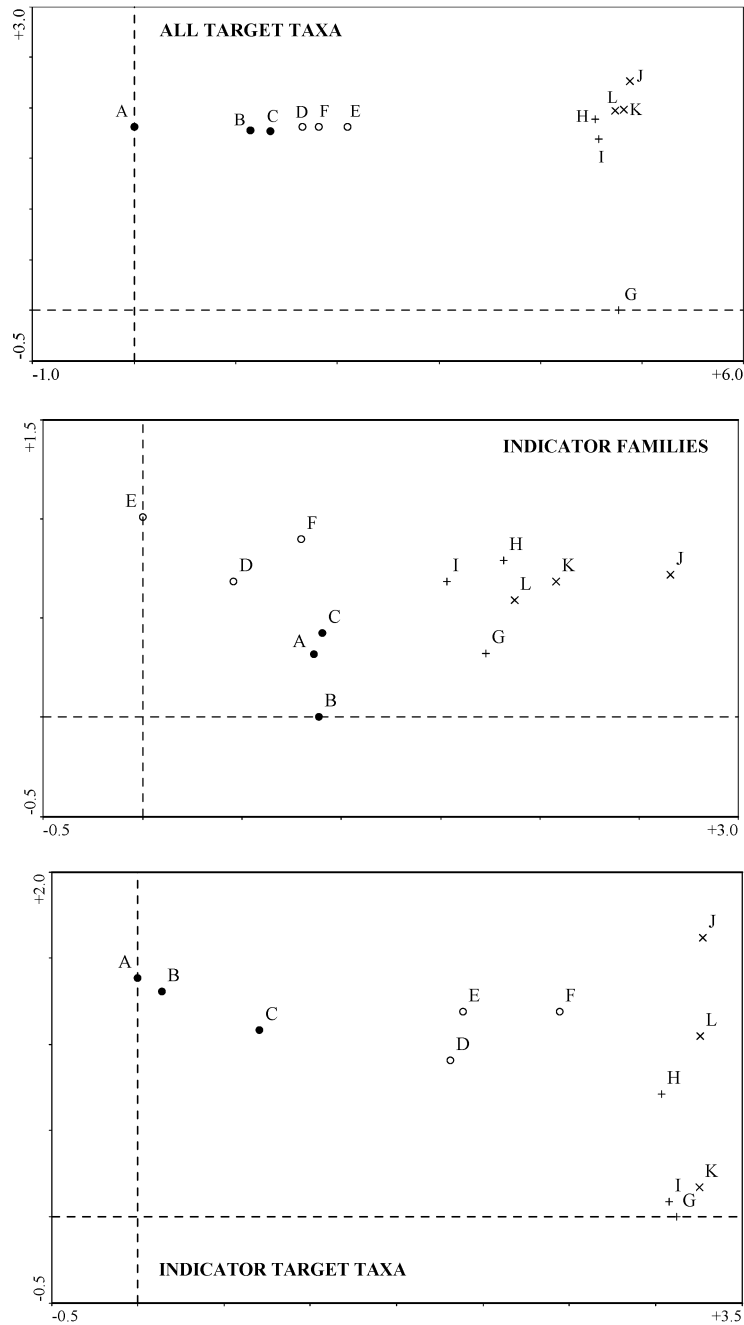
722



*Figure 3.* Plot of study sites within the plane formed by the first two axes of DCA for all target taxa, indicator families and indicator target taxa, excluding rare taxa. Presentation follows Figure 1.

*Table 5.* Summary of analyses of the indicator value index for (a) families and (b) all target taxa, excluding and including rare taxa. For (a) significant ($P < 0.05$) families are indicated for each habitat. For (b), significant ($P < 0.05$) indicator morphospecies are counted for each habitat and grouped by guilds.

| Variable | Excluding rare taxa | | | | Including rare taxa | | | |
|---|---|---|---|---|---|---|---|---|
| | Old forest | Young forest | Savanna | Gardens | Old forest | Young forest | Savanna | Gardens |
| *(a) Families* | | | | | | | | |
| Family codes[a] | Endo Myce Ptil Salt | Cory Curc Ento Micr Trix | Geom | Aley Anth Aphi Beth Clam Cocc Delp Doli Empi Ency Gryl Leio Lyga Scia Syrp Trid | Endo Myce Ptil Salt | Cory Curc Ento Micr Trop | Geom | Aley Anth Aphi Apoi Beth Cara Clam Cocc Delp Doli Empi Ency Gryl Leio Lyga Scia Trid |
| Total no. | 4 | 5 | 1 | 16 | 4 | 5 | 1 | 17 |
| *(b) All target taxa* | | | | | | | | |
| No. of chewers | 1 | 0 | 0 | 2 | 1 | 2 | 0 | 3 |
| No. of parasitoids | 3 | 1 | 0 | 2 | 2 | 1 | 0 | 4 |
| No. of pollinators | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 3 |
| No. of predators | 0 | 0 | 1 | 13 | 1 | 1 | 1 | 18 |
| No. of sap-suckers | 0 | 2 | 0 | 2 | 0 | 2 | 0 | 4 |
| No. of scavengers | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| No. of wood-eaters[b] | – | – | – | – | 0 | 0 | 0 | 0 |
| Total no. | 4 | 3 | 1 | 24 | 4 | 6 | 1 | 33 |

[a] Codes: Aley – Aleyrodidae; Anth – Anthicidae; Aphi – Aphididae; Apoi – Apoidea; Beth – Bethylidae; Cara – Carabidae; Clam – Clambidae; Cocc – Coccinellidae; Cory – Corylophidae; Curc – Curculionidae; Delp – Delphacidae; Doli – Dolichopodidae; Empi – Empididae; Ency – Encyrtidae; Endo – Endomychidae; Ento – Entomobryidae; Geom – Geometridae; Gryl – Gryllotalpidae; Leio – Leiodidae; Lyga – Lygaeidae; Micr – Micropezidae; Myce – Mycetophilidae; Ptil – Ptiliidae; Salt – Salticidae; Scia – Sciaridae; Syrp – Syrphidae; Trid – Tridactylidae; Trix – Trixagidae; Trop – Tropiduchidae. [b] Insufficient data for the analyses excluding rare wood-eating morphospecies.
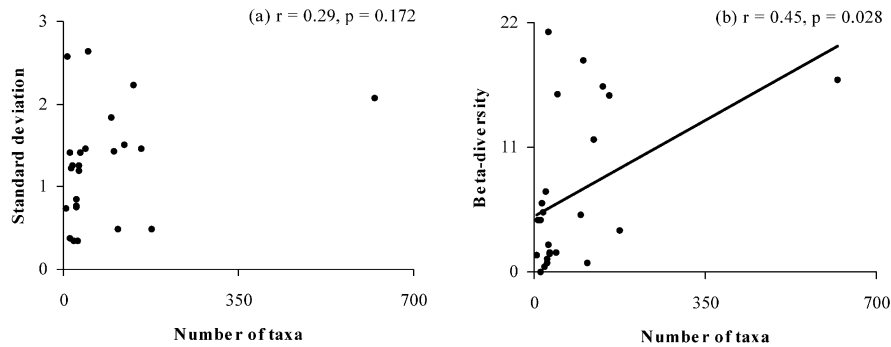
*Figure 4.* Relationships between the number of taxa included in data sets and (a) the standard deviation of scores on Axis 1 of the DCA; and (b) $\beta$-diversity.

Although the standard deviations and $\beta$-diversities of all data sets were significantly higher when including rare taxa than when excluding them (Wilcoxon signed-rank test, $Z = 2.667$, $P = 0.008$ and $Z = 2.934$, $P = 0.003$, respectively), this did not translate into a notably better classification of the sites along the disturbance gradient (comparison between Spearman's coefficients relating scores of Axis 1 of the DCA to habitats for data sets including and excluding rare taxa: $Z = 0.921$, $P = 0.357$).

For the different data sets examined, the relationships between the number of taxa included in the analysis and the standard deviation of the DCA scores on Axis 1 or the calculated $\beta$-diversity were not obvious and weak, respectively (Figure 4a and b; for the latter, the correlation was still weakly significant after removing the outlier, $r = 0.42$, $P = 0.047$). This suggests that the discriminatory power of the data sets was only marginally influenced by the number of taxa included in the data sets.

## Discussion

### Sampling protocol and methodological remarks

Our protocol (as all others) did not enable us to collect all arthropod species present at the study sites. The canopy habitat of forested sites was not sampled and other sampling methods, such as light trapping, might have yielded a different fauna. Some target taxa, in particular, may well have been better sampled. In comparing the performance of smaller data sets we used the data matrix which included all target taxa as a surrogate for the entire arthropod fauna present at the study sites. This approach is necessarily simplistic. However, the taxonomic and functional scopes of these preliminary data were much broader than the taxon-based approach of most studies addressing the conservation of terrestrial invertebrates in the tropics

(cf. Introduction) and, as such, offer a stimulating viewpoint for commenting on the discriminatory power of different data sets.

The present project also indicates that training of, and working with local parataxonomists (e.g., Basset et al. 2000) is a promising strategy in the monitoring of invertebrates in tropical systems. Lawton et al. (1998) commented on the high costs involved with biodiversity surveys in tropical systems, but did not consider the advantages of training of local parataxonomists in their protocols.

All of the traps used in this study have specific advantages and limitations, as discussed, for example, in Adis (1979) and Basset et al. (1997). In short, they measure the 'density activity' of relatively active arthropods. Sedentary arthropods are less likely to be collected and, thus, similarity measurements derived from our data are likely to be higher than if sedentary arthropods had been targeted.

Several important guilds, such as ants, fungal-feeders, other predators and tourists, have not been analyzed in the present account. However, we note that, with the exception of fungal-feeders (Endomychidae, Corylophidae, Clambidae, etc.), few of these guilds are represented by the indicator families identified in Table 5. A potentially more severe problem with any guild analysis is the sensitivity of the guild assignment (Stork 1987). The taxonomic study of the material collected is ongoing, so that changes in the assignment of morphospecies may be expected. Factors such as better taxonomy, a wider spectrum of target taxa (including sedentary species) and additional collecting methods will contribute to a better differentiation among sites. Our measurement of similarity among sites should be regarded as conservative.

Two observations are of particular note. First, many indicator taxa were added as a result of their occurrence in the garden habitat. This may reflect the higher catches in the gardens (i.e. the information on the garden fauna was higher than for other habitats) and also the very distinctive (i.e. weedy or invasive) fauna of gardens. Second, the species richness of gardens also appears superficially to be higher than in other habitats. This in turn may be explained by (a) the higher catches in gardens (i.e. rarefaction techniques will be needed to compare habitats); (b) the canopy habitat of forested sites was not surveyed and it may well account for a significant part of diversity, since faunal turnover between the understorey and the canopy is usually considerable (e.g., Basset et al. 2001b); and (c) insects may be more seasonal in forests than in gardens and hence temporal turnover may account for a substantial amount of diversity in forested habitats.

*Taxonomic resolution of the data sets*

There has been much debate, especially related to aquatic systems, as to what taxonomic level (either family- or species-level) is most suited for biological monitoring (e.g., Guérold 2000; Lenat and Resh 2001). The consensus is that, whenever possible, sorting to species is better. However, in some conditions, sorting to families may be acceptable (Bailey et al. 2001; Lenat and Resh 2001). Not surprisingly, analyses using higher taxa appear to be better suited to studies at broader geographic scales (Hewlett 2000). The present report suggests that for

studies of terrestrial arthropods distributed along a disturbance gradient confined to a small geographical area, such as the Gamba oil concession, data sets based on orders or guilds achieve only a poor discrimination of sites. Ordinal signatures have been found useful for terrestrial invertebrates only when encompassing broad geographic areas, such as latitudinal transects (e.g., Kitching et al. 1993).

Our family data sets, either including or excluding rare families, performed better than orders or guilds, but could not, for example, distinguish between sites situated in old and young secondary forests (Figure 1). However, filtering and retaining only indicator families resulted in a better discrimination of sites (Figure 3). This result is reassuring and suggests that counting the individuals within a set of a few arthropod families ($n = 26$ in our case) may be a possible strategy for biological monitoring of terrestrial systems in the tropics (cf. Kitching et al. 2000). This task can be performed easily by local parataxonomists trained beforehand, as in this study. The choice of indicator families may be guided by Table 5, for study systems similar to the present one. However, one must bear in mind that the indicator value of the families listed in Table 5 depended on collecting methods and on the regional species pool available. Further studies of similar scope evaluating other ecosystems in the tropics may eventually allow a consensus 'predictor set' to be reached.

Nevertheless, the data sets based on morphospecies performed better than those based on any higher taxa. Our two most discriminating data sets were those which included all morphospecies sorted from target taxa, and from the concomitant restricted set of indicator target taxa. Sorting only the indicator families identified above to morphospecies represents an additional strategy for biological monitoring. However, the current taxonomy of some families listed in Table 5 is difficult or requires particular techniques (e.g., Aleyrodidae, Entomobryidae, Sciaridae, etc.). It is doubtful whether a good correspondence could ever be achieved between a set of indicator families and the availability of taxonomic expertise. Choice will ultimately always be influenced by taxonomic expertise. However, as the present analyses show, considering morphospecies belonging to different guilds (i.e. analyses with all target taxa) was a better strategy than considering morphospecies belonging to a particular guild, especially since no guild yielded indicator taxa for all habitats studied (Table 5). We suggest, therefore, that the choice of target taxa should be guided by (1) available taxonomic expertise, (2) the inclusion of representatives from different guilds, and (3) considering as priorities taxa belonging to the following guilds (Tables 4 and 5): chewers, parasitoids, predators and sap-suckers. We also note that including more taxa in the data sets does not ensure that these data sets will have greater discriminatory power (Figure 4). The quality and choice of the information included in the data sets are more important in this regard.

*The importance of rare species in the analyses*

This debate may be considered from either a statistical or biological viewpoint. We focus on the latter and distinguish between 'true' rarity (species with genuinely low population levels, occurring at the limit of their geographical distribution, etc.) and 'apparent' rarity (sampling methods not appropriate, sampling of marginal habitats, seasonality, or timing of activity, etc.). Rare species are perceived as important in

aquatic systems and are usually retained for the analyses, since their preservation is often the ultimate aim of biological monitoring (Lenat and Resh 2001). The situation is not as evident in terrestrial systems, where the higher and more complex habitat diversity means that a greater proportion of 'apparent' rare species is also to be expected (i.e. vagrant or poorly sampled species are probably much more common in terrestrial than in aquatic systems). Work with local parataxonomists, by increasing the scope of the protocol to include other sampling methods and obtaining more spatial and temporal replicates, may reduce the proportion of 'apparent' rare species, but not beyond a certain threshold, which relates to the sampling of marginal habitats (Novotny and Basset 2000). Since, in the tropics, the proportion of 'apparent' rare species may be relatively high in terrestrial data sets obtained with low sampling effort, common sense dictates the exclusion of all rare species from these data sets. In contrast, data sets obtained with intensive sampling effort should be analyzed with their rare species, since these data are more likely to include 'true' rare species, as defined above.

The present results suggest that inclusion or exclusion of rare species *per se* in the analyses is not as important as the primary information included in the data matrices (i.e. which taxa are included). Analyses including rare taxa had slightly more discriminating power than those excluding rare species, but the fine discrimination of study sites was not improved notably. This conclusion was confirmed when we repeated the analyses (not presented here) of DCAs including rare taxa but with the raw data $\log(x + 1)$ transformed, to downscale the effect of abundant species. This observation is reassuring and confirms the prospect of reaching a consensus and establishing a 'recipe' for biological monitoring of species-rich terrestrial eco-systems in the tropics.

*Predictor sets and indicator taxa*

No single taxon was useful in accurately classifying our study sites along the disturbance gradient, but a restricted set of families and morphospecies was. Accordingly, this study validates the concept of predictor sets for biological monitoring of terrestrial tropical systems and provides a starting point for identify-ing taxa that may be included in such predictor sets (see also Kotze and Samways (1999), for a similar argument). It is germane to ask what would have been the results of our study if we had considered only a popular indicator taxon such as, say, dung beetles (e.g., Davis et al. 2001; McGeogh et al. 2002). Certainly re-directing our whole protocol and sampling effort to increase spatial replicates and using different traps (including baits) would have resulted in better representation of dung beetles within our data sets, and those may have been able to discriminate the study sites more accurately. That is not our argument. Our protocol allowed comparison of the relative discriminatory power of similar data sets based either on chewers (including 536 individuals and 85 morphospecies of Chrysomelidae; Table 2), parasitoids (including 200 individuals and 22 morphospecies of Chalcididae) or scavengers (including 449 individuals and 28 morphospecies of dung beetles; Table 2) obtained with a similar sampling effort. The data sets for chewers and parasitoids proved to be more discriminating than the scavenger data set, which could differen-

tiate only between forested and non-forested sites. These results were similar when restricting the data set to dung beetles (analyses not presented here). We do not suggest that dung beetles should be excluded from biological monitoring, rather that they should be analyzed conjointly with other taxa, as predictor sets. Dung beetles are undoubtedly useful indicators of structural differences between ecosystems, in contrast to insect herbivores, such as chrysomelids, that reflect plant-feeding specialisations (Davis et al. 2001).

In choosing taxa for inclusion in predictor sets, one must consider that the occurrence of some taxa may be related, and hence redundant. For example, Coccinellidae, Dolichopodidae and Syrphidae are likely to be predators of Aphididae in gardens (Table 5). Accordingly it will be more informative to consider but one of these families, redirecting effort to other families that may be of greater indicator value for old forests (such as Mycetophilidae; Table 5, and see the study of Økland (1994) in Norway). Although Table 5 provides baseline information to identify predictor sets for forest–savanna ecosystems in Africa, we remain reluctant to propose specific families and morphospecies for predictor sets before studies similar to ours are carried out elsewhere in the tropics to build up a minimum level of information.

## Conclusions

This study validated the usefulness of arthropod predictor sets in biological monitoring at a taxonomic scope not investigated before in the tropics. Kitching et al. (2000) identified 17 subfamilies of moths from light-trapping data that may be used as predictor sets of the quality of rainforest remnants in Australia. Work should now proceed to gather data similar to those presented here but originating from different tropical systems, and seeking a consensus as to which taxa should be included in predictor sets of wider taxonomic scope. Such a consensus will allow the derivation of a biotic index that will be representative, efficient and applicable to terrestrial invertebrate assemblages in the tropics, unlike that based on a few putative indicator taxa, such as butterflies or dung beetles. Reaching such a consensus will not be easy. Proposed indicators are legion. They can indicate different effects. Their validation requires studies at different disturbance and geographical scales, and they should also be representative of major important microhabitats (McGeogh 1998; Andersen 1999; Taylor and Doran 2001). However, the calibration studies (with a statistically based, relatively objective, analysis of the kind presented here) required to achieve this ambitious task could be reasonably quickly performed (in less than half a year, as in the present study) with the help of local parataxonomists, with adequate eutaxonomic support.

## Acknowledgements

## References

Adis J. 1979. Problems of interpreting arthropod sampling with pitfall traps. Zoologischer Anzeiger, Jena 202: 177–184.

Andersen A.N. 1999. My bioindicator or yours? Making the selection. Journal of Insect Conservation 3: 61–64.

Anstett M.-C., Michaloud G. and Kjellberg F. 1995. Critical population size for fig/wasp mutualism in a seasonal environment: effect and evolution of the duration of female receptivity. Oecologia 103: 453–461.

Bailey R.C., Norris R.H. and Reynoldson T.B. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. Journal of the North American Benthological Society 20: 280–286.

Bartolozzi L. and Sforzi A. 1997. Contribution to the knowledge of the Brentidae from Gabon (Coleoptera Brentidae). Bollettino della Societa Entomologica Italiana 29: 79–86.

Basset Y., Charles E.L., Hammond D.S. and Brown V.K. 2001a. Short-term effects of canopy openness on insect herbivores in a rain forest in Guyana. Journal of Applied Ecology 38: 1045–1058.

Basset Y., Aberlenc H.-P., Barrios H., Curletti G., Béranger J.-M., Vesco J.-P. et al. 2001b. Stratification and diel activity of arthropods in a lowland rainforest in Gabon. Biological Journal of the Linnean Society 72: 585–607.

Basset Y., Springate N.D., Aberlenc H.-P. and Delvare G. 1997. A review of methods for collecting arthropods in tree canopies. In: Stork N.E., Adis J.A. and Didham R.K. (eds), Canopy Arthropods. Chapman & Hall, London, pp. 27–52.

Basset Y., Novotny V., Miller S.E. and Springate N.D. 1998. Assessing the impact of forest disturbance on tropical invertebrates: some comments. Journal of Applied Ecology 35: 461–466.

Basset Y., Novotny V., Miller S.E. and Pyle R. 2000. Quantifying biodiversity: experience with parataxonomists and digital photography in New Guinea and Guyana. BioScience 50: 899–908.

Belshaw R. and Bolton B. 1993. The effect of forest disturbance on the leaf litter ant fauna in Ghana. Biodiversity and Conservation 2: 656–666.

Bilardo A. and Rocchi S. 2000. Haliplidae e Dytiscidae (Coleoptera) del Gabon (Parte terza). Atti della Societa Italiana di Scienze Naturali e del Museo Civico di Storia Naturale di Milano 140: 215–236.

Bourgeais J. 2001. Impact de l'exploitation pétrolière dans le complexe de Gamba: le pire est à venir. Canopée 21: 5–7.

Boussienguet J., Neuenschwander P. and Herren H.R. 1991. Essais de lutte biologique contre la cochenille du manioc au Gabon: I. – Etablissement, dispersion du parasite exotique *Epidinocarsis lopezi* [Hym.: Encyrtidae] et déplacement compétitif des parasites indigènes. Entomophaga 36: 455–469.

Chung A.Y.C., Eggleton P., Speight M.R., Hammond P.M. and Chey V.K. 2000. The diversity of beetle assemblages in different habitat types in Sabah, Malaysia. Bulletin of Entomological Research 90: 475–496.

Clarke K.R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18: 117–143.

Collins N.M. and Thomas J.A. 1991. The Conservation of Insects and Their Habitats. Academic Press, New York.

Colwell R.K. 1997. Biota: The Biodiversity Database Manager. Sinauer Associates, Sunderland, Massachusetts.

Cranston P.S. 1990. Biomonitoring and invertebrate taxonomy. Environmental Monitoring and Assessment 14: 265–273.

Davis A.J., Holloway J.D., Huijbregts H., Krikken J., Kirk-Spriggs A.H. and Sutton S.L. 2001. Dung beetles as indicators of change in the forests of northern Borneo. Journal of Applied Ecology 38: 593–616.

Didham R.K., Ghazoul J., Stork N.E. and Davis A.J. 1996. Insects in fragmented forests: a functional approach. Trends in Ecology and Evolution 11: 255–260.

Didham R.K., Hammond P.M., Lawton J.H., Eggleton P. and Stork N.E. 1998. Beetle species richness responses to tropical forest fragmentation. Ecological Monographs 68: 295–323.

Doumenge C., Garcia Yuste J.-E., Gartlan S., Langrand O. and Ndinga A. 2001. Conservation de la biodiversité forestière en Afrique Centrale Atlantique: le réseau d'aires protégées est-il adéquat? Bois et Forêts des Tropiques 268: 5–27.

Dufrêne M. and Legendre P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs 67: 345–366.

Eggleton P., Bignell D.E., Sands W.A., Mawdsley N.A., Lawton J.H., Wood T.G. et al. 1996. The diversity, abundance and biomass of termites (Isoptera) under differing levels of forest disturbance in the Mbalmayo Forest Reserve, southern Cameroon. Philosophical Transactions of the Royal Society of London, Series B 351: 51–68.

Gaston K.J. and Hudson E. 1994. Regional patterns of diversity and estimates of global insect species richness. Biodiversity and Conservation 3: 493–500.

Grandcolas P. 1997. Habitat use and population structure of a polyphagine cockroach, *Ergaula capensis* (Saussure 1893) (Blattaria Polyphaginae) in Gabonese rainforest. Tropical Zoology 10: 215–222.

Growns J.E., Chessman B.C., Jackson J.E. and Ross D.G. 1997. Rapid assessment of Australian rivers using macroinvertebrates: cost and efficiency of 6 methods of sample processing. Journal of the North American Benthological Society 16: 682–693.

Guérold F. 2000. Influence of taxonomic determination level on several community indices. Water Research 34: 487–492.

Hammond P.M. 1994. Practical approaches to the estimation of the extent of biodiversity in speciose groups. Philosophical Transactions of the Royal Society of London, Series B 345: 119–136.

Harrison S., Ross S.J. and Lawton J.H. 1992. Beta diversity on geographic gradients in Britain. Journal of Animal Ecology 61: 151–158.

Hewlett R. 2000. Implications of taxonomic resolution and sample habitat for stream classification at a broad geographic scale. Journal of the North American Benthological Society 19: 352–361.

Hill J.K. 1999. Butterfly spatial distribution and habitat requirements in a tropical forest: impacts of selective logging. Journal of Applied Ecology 36: 364–372.

Intachat J., Holloway J.D. and Speight M.R. 1999. The impact of logging on geometroid moth populations and their diversity in lowland forests of Peninsular Malaysia. Journal of Tropical Forest Science 11: 61–78.

Janzen D.H. and Gauld I.D. 1997. Patterns of use of large moth caterpillars (Lepidoptera: Saturniidae and Sphingidae) by ichneumonid parasitoids (Hymenoptera) in Costa Rican dry forest. In: Watt A.D., Stork N.E. and Hunter M.D. (eds), Forests and Insects: 18th Symposium of the Royal Entomological Society. Chapman & Hall, London, pp. 251–271.

Karr J.R. 1991. Biological integrity: a long-neglected aspect of water resource management. Ecological Applications 1: 66–84.

Kitching R.L. 1993a. Biodiversity and taxonomy: impediment or opportunity? In: Moritz C. and Kikkawa J. (eds), Conservation Biology in Australia and Oceania. Surrey Beatty and Sons, Chipping Norton, Australia, pp. 253–268.

Kitching R.L. 1993b. Towards rapid biodiversity assessment – Lessons following studies of arthropods of rainforest canopies Rapid Biodiversity Assessment.: Proceedings of the Biodiversity Workshop, Macquarie University – 1993. Research Unit for Biodiversity and Bioresources, Macquarie University, Sydney, Australia, pp. 26–30.

Kitching R.L., Bergelson J.M., Lowman M.D., McIntyre S. and Carruthers G. 1993. The biodiversity of

arthropods from Australian rainforest canopies: general introduction, methods, sites and ordinal results. Australian Journal of Ecology 18: 181–191.

Kitching R.L., Orr A.G., Thalib L., Mitchell H., Hopkins M.S. and Graham A.W. 2000. Moth assemblages as indicators of environmental quality in remnants of upland Australian rain forest. Journal of Applied Ecology 37: 284–297.

Kitching R.L., Li D.Q. and Stork N.E. 2001. Assessing biodiversity 'sampling packages': how similar are arthropod assemblages in different tropical rainforests? Biodiversity and Conservation 10: 793–813.

Kotze D.J. and Samways M.J. 1999. Support for the multi-taxa approach in biodiversity assessment, as shown by epigaeic invertebrates in an Afromontane forest archipelago. Journal of Insect Conservation 3: 125–143.

Kremen C. 1992. Assembling the indicator properties of species assemblages for natural areas monitoring. Ecological Applications 2: 203–217.

Kremen C., Colwell R.K., Erwin T.L., Murphy D.D., Noss R.F. and Sanjayan M.A. 1993. Terrestrial arthropod assemblages: their use in conservation planning. Conservation Biology 7: 796–808.

Landres P.B., Verner J. and Thomas J.W. 1988. Ecological use of vertebrate indicator species: a critique. Conservation Biology 2: 316–328.

Lawton J.H., Bignell D.E., Bolton B., Bloemers G.F., Eggleton P., Hammond P.M. et al. 1998. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. Nature 391: 72–76.

Lenat D.R. and Resh V.H. 2001. Taxonomy and stream ecology – The benefits of genus- and species-level identification. Journal of the North American Benthological Society 20: 287–298.

Maes J.-M. and Pauly A. 1998. Lucanidae (Coleoptera) du Gabon. Bulletin et Annales de la Société Royale Belge d'Entomologie 134: 279–285.

McAleece N., Lambshead J., Paterson G. and Gage J. 1997. BioDiversity Professional Beta Release 1. The Natural History Museum and the Scottish Association for Marine Science, at http://www.sams.ac.uk/dml/projects/benthic/bdpro/index.htm.

McCune B. and Medford M.J. 1999. Multivariate Analysis of Ecological Data Version 4.10. MjM Software, Gleneden Beach, Oregon.

McGeogh M.A. 1998. The selection, testing and application of terrestrial insects as bioindicators. Biological Reviews 73: 181–201.

McGeogh M.A., Van Rensburg B.J. and Botes A. 2002. The verification and application of bioindicators: a case study of dung beetles in a savanna ecosystem. Journal of Applied Ecology 39: 661–672.

Miller S.E. and Rogo L. 2002. Challenges and opportunities in understanding and utilisation of African insect diversity. Cimbebasia 17: 197–218.

Moran C.V. and Southwood T.R.E. 1982. The guild composition of arthropod communities in trees. Journal of Animal Ecology 51: 289–306.

Moritz C., Richardson K.S., Ferrier S., Monteith G.B., Stanisic J., Williams S.E. et al. 2001. Biogeographical concordance and efficiency of taxon indicators for establishing conservation priority in a tropical rainforest biota. Proceedings of the Royal Society of London, Series B 268: 1875–1881.

Novotny V. and Basset Y. 2000. Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. Oikos 89: 564–572.

Novotny V., Basset Y., Miller S.E., Weiblen G.D., Bremer B., Cizek L. et al. 2002. Low host specificity of herbivorous insects in a tropical forest. Nature 416: 841–844.

O'Connell T.J., Jackson L.E. and Brooks R.P. 1998. A bird community index of biotic integrity for the mid-Atlantic highlands. Environmental Monitoring and Assessment 51: 145–156.

Økland B. 1994. A comparison of clearcut, managed and semi-natural spruce forests in southern Norway. Biodiversity and Conservation 3: 68–85.

Oswald J.D. 1998. Annotated catalogue of the Dilaridae (Insecta: Neroptera) of the world. Tijdschrift voor Entomologie 141: 115–128.

Pauly A. 1998. Hymenoptera Apoidea du Gabon. Musée Royal de l'Afrique Centrale Tervuren, Belgique Annales Sciences Zoologiques 282: 1–121.

Prendergast J.R., Quinn R.M., Lawton J.H., Eversham B.C. and Gibbons D.W. 1993. Rare species, the coincidence of diversity hotspots and conservation strategies. Nature 365: 335–337.

Prins H.H.T. and Reitsma J.M. 1989. Mammalian biomass in an African equatorial forest. Journal of Animal Ecology 58: 851–861.

Rossaro B. and Pietrangelo A. 1993. Macroinvertebrate distribution in streams: a comparison of CA ordination with biotic indices. Hydrobiologia 263: 109–118.

Roubik D.W. 1999. The foraging and potential outcrossing pollination ranges of African honey bees (Apiformes: Apidae; Apini) in Congo Forest. Journal of the Kansas Entomological Society 72: 394–401.

Roy R. 1973. Premier inventaire des mantes du Gabon. Biologia Gabonica 8: 235–290.

Stork N.E. 1987. Arthropod faunal similarity of Bornean rain forest trees. Ecological Entomology 12: 219–226.

Stork N.E. 1991. The composition of the arthropod fauna of Bornean lowland rain forest trees. Journal of Tropical Ecology 7: 161–180.

Taylor R.J. and Doran N. 2001. Use of terrestrial invertebrates as indicators of the ecological sustainability of forest management under the Montreal Process. Journal of Insect Conservation 5: 221–231.

Ter Braak C.J.F. and Smilauer P. 1998. CANOCO Reference Manual and User's Guide to CANOCO for Windows: Software for Canonical Community Ordination (version 4). Microcomputer Power, Ithaca, New York.

Thibault M. and Blaney S. 2001. Sustainable human resources in a protected area in Southwestern Gabon. Conservation Biology 15: 591–595.

Thorne R.S.J. and Williams W.P. 1997. The response of benthic macroinvertebrates to pollution in developing countries: a multimetric system of bioassessment. Freshwater Biology 37: 671–686.

Townes H. 1972. A light-weight Malaise trap. Entomological News 83: 239–247.

Urzelai A., Hernandez A.J. and Pastor J. 2000. Biotic indices based on soil nematode communities for assessing soil quality in terrestrial ecosystems. Science of the Total Environment 247: 253–261.

Vasconcelos H.L. 1999. Effects of forest disturbance on the structure of ground-foraging ant communities in central Amazonia. Biodiversity and Conservation 8: 409–420.

Walter P. 1987. Contribution à la connaissance des Scarabeides coprophages du Gabon [Col.]. 5. Nidification et morphologie larvaire de *Tiniocellus panthera* (Boucomont). Annales de la Société Entomologique de France 23: 309–314.

Warwick R.M. 1993. Environmental impact studies on marine communities. Australian Journal of Ecology 18: 63–80.

Watt A.D. 1998. Measuring disturbance in tropical forests: a critique of the use of species-abundance models and indicator measures in general. Journal of Applied Ecology 35: 467–469.

Wetterer J.K., Walsh P.D. and White L.J.T. 1999. *Wasmannia auropunctata* (Roger) (Hymenoptera: Formicidae), a destructive tramp-ant, in wildlife refuges of Gabon. African Entomology 7: 292–294.

Whittaker R.M. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30: 279–338.