On methods of biodiversity data collection and monitoring

How can a policy be evaluated without measurements? Biodiversity measurements are difficult, but necessary. This difficulty is due essentially to the nature of biodiversity itself (diversity of species and environments, diversity of evaluation methods, diversity of monitoring systems, etc.). The author presents here a few ideas on the topic.

The importance of measurement quality has long been recognised in industry where an entire certification sector is devoted to it and this recognition is spreading to other sectors including, increasingly, the environmental field. An example of standardisation in biodiversity is the biotic index which provides an indication on the hydrobiological quality of a river based on the composition of aquatic invertebrate species. The index has been standardised in France as AFNOR NF T 90-350.

However, the trend toward standardisation is not prevalent in the field of ecology. A number of reasons explain this delay.

- Ecologists are perhaps reluctant to shift from "naturalist" work, where they feel freer and probably entertain the illusion of exhausting the topic, to a more quantitative and formalised approach, seen as more restrictive.
- Decision-makers know little about biodiversity and the means to measure it.
- Neither ecologists nor decision-makers are fully aware of the value, even the necessity of standardisation in biodiversity measurements.
- Finally, measurements are a complex topic. Biodiversity cannot be measured similar to a physical or chemical quantity. There are many sources of imprecision, some of which are difficult to control. This complexity is not favourable to the emergence of a consensus on the best methods.

The purpose of this article is to discuss a number of crucial factors for biodiversity measurement, i.e. the lack of methods standardisation, taxonomic representativeness, the limits to measurement quality and the proposed improvements, all points that are often neglected in biodiversity monitoring (Yoccoz *et al.*, 2001). By "measurement" we mean all biodiversity quantification processes (sampling, protocols) and not only the measurements themselves. In the biodiversity field, it is common to distinguish genetic diversity (within a species), specific diversity (between species) and ecosystem diversity (between sets of species). In this article, we will focus on specific diversity, even if most of the issues addressed concern the two other levels as well. In addition, we will look exclusively at direct biodiversity monitoring and not indirect monitoring which attempts to measure a resource or potential habitat, e.g. surface of limestone swards, volume of dead wood, i.e. not biodiversity itself.

Insufficient standardisation in biodiversity measurements

Biodiversity measurements are relatively recent in France and started in earnest for fauna and flora communities in the 1960s and 1970s. Until then, naturalists essentially carried out limited observations (presence of a species in a given place) or formulated theories based on a set of observations, where some leeway was left to the subjectivity of the author. These initiatives gradually resulted in regional or national atlases showing the distribution of fauna and flora. Notions such as the sampling effort, the representativeness of the sampled habitats, repeatability and statistical power were generally absent from this work and would have been difficult to control at any rate. On the other hand, biodiversity monitoring, i.e. repeated samples from a same place over time, is more recent, but took care from the start to include all the above elements, an example being the STOC-ESP programme (temporal monitoring of common breeding birds).

Unfortunately, there is no consensus in the scientific community on methods and there exist almost as many sampling protocols as there are monitoring programmes. The most remarkable case was the ICP-Forest network launched in 1985 by UNECE (United Nations Economic Commission for Europe). This international cooperation programme was launched to study the effects of cross-border pollution on forest ecosystems. The network comprises

❶ Mushrooms on dead woods.

© M. Carrouée (Cemagref).

two subsets of sites. The first (level II) is made up of 800 sites on which a large number of physical and chemical measurements are carried out according to standardised protocols. Monitoring of flora was set up on all the sites in 1995. Because there was no consensus and no decision, the basic size of the plots for the floristic surveys varied from 4 000 to 5 500 square metres, occasionally with variations in a single country! Quite logically, the number of species increases with the size of the plot (there is ample literature on the species-area relationship). The result is that the data could never be analysed as a whole. Ten years passed before all participants agreed to a common protocol and a basic plot size (400 m² in this case) to monitor changes in flora over time across the entire network.

Generally speaking, greater standardisation in biodiversity monitoring would provide all study sites (nature reserves, public forests, etc.) with external references (e.g. from all reserves in France). In addition, compiling of comparable data makes it possible to answer questions that isolated data cannot, at no extra cost other than concatenating the databases and running the analysis. Standardisation of methods is facilitated by the emergence of shared databases and above all by the creation of national monitoring systems, e.g. the various Vigie-Nature programmes managed by the National museum on natural history in Paris. Protocol standardisation is always positive for the stated reasons, but a further necessity is critical evaluation of measurement quality. A standardised method producing poor-quality data would be of little use.

Biodiversity and the taxonomic representativeness of samples

Biodiversity is a generic term. It is never actually measured, only a small part of it is measured. For genetic diversity, the diversity is often measured within a species or even a population, and on a limited portion of the genome. Given the diversity of species, biodiversity measurements often deal with an order, family, occasionally only a genus or an ecological group, e.g. organisms found in dead wood.

That would not be important if study results did not depend on the choices made. But the response (to a climate, habitat or management-intensity gradient, etc.) of different taxonomic groups is rarely consistent between groups. For example, the variety of vascular plants tends to drop with the age of forest stands, while that of saproxylic organisms (dependent on dead or dying wood) increases, similar to many mushrooms, insects and vertebrates (see photo ❶). Even within a taxonomic group, responses can be contradictory. We must always remember that the response of a taxonomic group is not valid for biodiversity as a whole.

However, the issue of sample representativeness goes beyond the taxonomic choice made. In many cases, only a part of the targeted family is effectively sampled. For example, among birds, birds of prey and ducks are poorly sampled by listening stations. Similarly, glass traps for flying insects trap only those species capable of reaching the height of the trap.

Completeness, identification and practical risks

Measurement bias and precision

Measurements should be as precise and unbiased as possible. The two are not exclusive, i.e. all measurements are somewhat imprecise and somewhat biased (see figure ❶). But they produce different effects for data analysis (Archaux et Bergès, 2008). In classic statistics (probabilistic or Popperian referring to the Austrian philosopher of science Karl Popper) and without entering into unnecessary detail, measurement imprecision results in a loss of statistical power, i.e. it limits the capacity of a study to prove that biodiversity levels are different between habitats, management practices or over time. For decision-makers, the main risk lies in delaying corrective action because the problem is not identified. Biased measurements can similarly mask real differences, but they can also signal differences in biodiversity that do not in fact exist. The risk then lies in taking unnecessary (even disastrous) corrective measures. For example, listening stations are often used to estimate numbers of common birds. But the more closed the environment, the less the bird sounds travel, due to reverberation. An ornithologist will thus underestimate bird abundance in a closed environment as compared to an open one and conclude that differences exist when in fact there is a systemic error.

Advantages and disadvantages of species richness

One of the most commonly used measurements in biodiversity concerns the number of species, also called species richness. All the studies addressing this measurement, highly appreciated for its simplicity, concluded that a non-negligible percentage of species is not detected during biodiversity surveys. On average, one out of five plants is missed during floristic surveys. The percentage is similar for bird surveys using listening stations. The most worrisome is not that the surveys are not complete, but that the degree of completeness varies between the compared aspects (e.g. different environments, similar environments but different years, etc.). It is possible to estimate using simulations the risk of erroneously concluding that one aspect is more diverse due to differences in the probability of detection. The risk is far from negligible even when the probability of detection differs only by a few percentage points between aspects.

There is every reason to think that such differences in detectability are very common in biodiversity studies. Whatever the taxonomic group (vertebrates, invertebrates, plants), the differences in detectability between individuals or species are often approximately 10%. There are many factors likely to cause these differences. For flora, they include plant cover (a moss is less likely to be noted than an oak tree), the experience level of the botanist (the percentage of missed species can vary from 10 to 40%), the number of persons carrying out the survey (two persons together, even if one is not a botanist, will produce more complete results than a single botanist), the time spent (the number of detected species increases exponentially with the time spent), fatigue and the experience level of the team as a whole (Archaux *et al.*, 2009). The weather and period of year also impact on the results. These detection errors are occasionally compounded by identification errors (probably less than 1% of detected species), though the latter generally decrease in step with the increase in the experience of the botanist. The effect of personnel is a particularly worrisome problem for long-term monitoring in that people progressively gain in experience and new personnel replace the old over time.

Other factors add to the difficulty for families that are detected by their behaviour, such as song birds and moving insects. That is notably the case for weather conditions. On cold, rainy days, birds sing less, bats hunt less (they are identified essentially by their ultrasound emissions), insects travel less. Concerning insects, even when traps theoretically eliminate the effect of personnel (interception or lure traps), other error factors include the type of trap, its height and openness to the environment (which determine its effectiveness), the lures used and, still, the competency of personnel in laying the traps. An experienced entomologist has a better chance of identifying the preferred travel corridors of the entomofauna and should thus trap more insects (see photo ❷).

❶ Measurement bias and precision.

If a measurement is carried out repeatedly on a given sample and produces a number of similar results, it is precise, as shown in the two top graphs. A measurement is not biased if, on average, it is equal to the actual value of the measured sample items, as shown in the two graphs on the left where the line of best fit between the measurement and the actual value of the sample items is identical to line y = x (shown as a solid line in the two graphs on the right). In the two graphs on the right, the measurement is negatively biased, i.e. on average, the measurement is less than the actual value. Precision and bias are not exclusive.

Advantages and disadvantages of "averaged" indices

This sensitivity of species-richness measurements to a wide array of factors has led some authors to prefer "averaged" indices covering all the detected species. In principle, all that is needed is a sample of species representing the community to calculate the averaged index without bias. The average degree of specialisation of bird communities is one of these indices (Devictor *et al.*, 2009). The underlying assumption is that a community dominated by specialist species is preferable to one dominated by generalist species. A specialisation index, assigned nationally to each species of common bird, made it possible to organise species along a gradient ranging from specialist species (requiring specific habitats) to generalist species that can tolerate a wider range of habitats. It is then necessary to calculate the index average for all the detected species in a given place to determine the average degree of specialisation in the community. This approach is however open to the criticism expressed for species richness if the probability of detection varies with the degree of specialisation of species. For example, a specialist species may be less frequently detected than a generalist species. There are similar averaged indices for invertebrates (biotic index) and flora (Ellenberg's indicator values). However, they provide more information on habitat quality than on biodiversity itself.

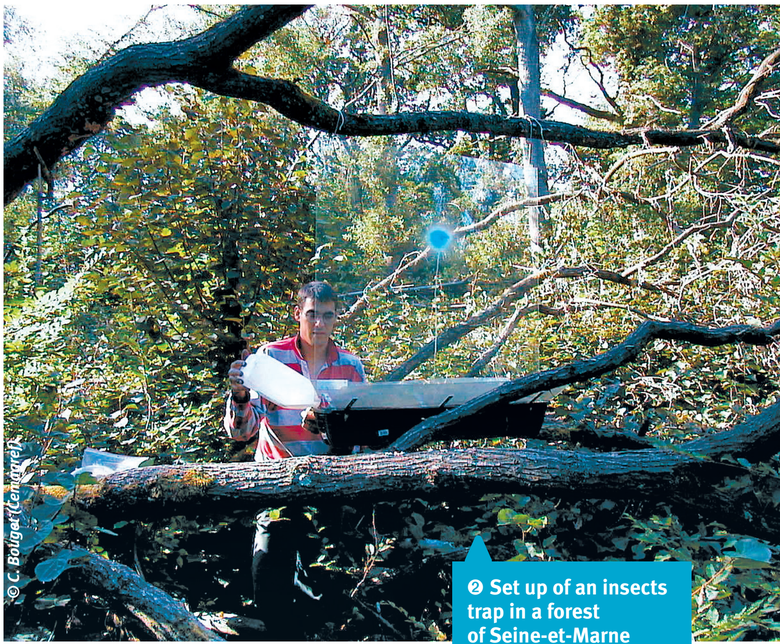
These averaged indices cannot replace absolute indices such as species richness. Consider the following two situations. In the first, a community has gained in species, but more in generalist than in specialist species. In the second, the community has lost a similar proportion of both specialist and generalist species. In the first case, the specialisation index of the community dropped whereas it remained stable in the second. On the basis of this single index, one would conclude that changes in the first community are more worrisome than in the second, whereas obviously, the reverse is true. Averaged and absolute indices should therefore be used together, not exclusively, particularly given that averaged indices are not necessarily free of sampling bias.

Taking into account the sources of error

There are methods to adjust the data to take into account the incompleteness of surveys. Historically, these methods were initially developed in the 1930s for bird ringing. Ringing consists of capturing, identifying and recapturing individuals on different occasions in order to estimate biological parameters such as life expectancy, attachment to a site and population size. Not capturing a bird on a given occasion does not necessarily mean the bird is dead, it may simply have escaped capture. A series of statistical methods was developed to distinguish between the probability of survival and that of detection. To apply these tools, it is necessary to multiply the local surveys (i.e. the chances of capture). These methods were first used for community ecology studies at the end of the 1990s by drawing parallels between individuals and species, i.e. estimate not the number of individuals but of species and other parameters such as the local extinction and colonisation rates (Yoccoz *et al.*, 2001). It is thus possible to increase the chances of capture by returning several times to the same site, laying a larger number of traps or calling on several naturalists at the same time (on the condition they not communicate because their surveys must be independent). These methods are useful if, during a visit, there is a sufficiently high average probability that species are detected (> 30%). They are, however, very sensitive to:

- determination errors, which are not always east to identify in a data set;
- heterogeneity in the probability of detection between individuals and species.

In spite of these difficulties, this field of research is growing rapidly. It can take into account increasing numbers of variables and any possible identification errors, and should continue to progress quickly.



❷ Set up of an insects trap in a forest of Seine-et-Marne (France).

For birds, another strategy lies in estimating the distance of singing males from the central survey point. Models linking the probability of detection to the distance are then adjusted to the distance data to estimate bird densities. Different functions may be used and the parameters adjusted such that the detection probability decreases with the distance from the central point, in a linear or non-linear manner, rapidly or slowly. This method can also be used for line-transect sampling of plants by measuring the distance separating each specimen from the transect line.

However, it is necessary to keep in mind that all these methods are purely stopgap measures and it is always preferable to reduce as much as possible detection and identification errors during surveys.

Analysis of biodiversity data

Just as there are advantages to homogenising data collection, there are similar advantages to standardising statistical analysis of the data. Very often, a researcher wants to know if the changes observed also occur in neighbouring territories. But it is very difficult to make quantitative comparisons, to say nothing of qualitative, if different statistical methods are used. The scientific community is increasingly aware of the need to progress toward standardisation in data analysis and NGOs, such as the European Bird Census Council in Denmark, offer software for analysis of monitoring data free of cost online.

The point here is not to supply a list of recommended methods. The frequently complex nature of biodiversity data is a good reason to maintain some diversity in methods. Univariate analysis (e.g. hierarchical general models) and multivariate analysis (e.g. canonical analysis) should be seen as partners rather than as competitors.

However, these methods must comply with precise technical specifications.

- The statistical model must take into account the spatial and temporal structure of the data set.

- The underlying assumptions (e.g. concerning the statistic distribution of variables) must be compatible with the data and, if that is not the case, analysis results must be robust to resist non-compliance with the assumptions.
- It must be possible to interpret the results without ambiguity.
- Analysis must be easy to carry out.

Though, generally speaking, it would appear that the statistical methods available are sufficient, progress is still required on a certain number of points.

Conclusion on the repercussions for public policy

It would appear more important than ever to set up a structured biodiversity-monitoring programme sufficiently consistent to enable judgements on the impact of various public policies and the effectiveness of compensation measures. Given the difficulties mentioned above, it might be tempting to simply give up on biodiversity monitoring for the formulation and evaluation of public policies.

Indirect monitoring of the resources used by biodiversity is a credible alternative to direct monitoring of the taxa themselves, particularly if, in addition to the purely technical difficulties mentioned above for direct monitoring, the necessary costs and skills are factored in as well. That being said, indirect monitoring cannot in itself replace direct monitoring because the causal relation between resource levels and biodiversity levels is not always clear and is probably not constant over time and space. The best solution would certainly be to combine large-scale indirect monitoring and more targeted direct monitoring. What is more, managers and politicians are certainly more receptive to direct biodiversity indicators (e.g. a decline in woodpecker populations) than indirect indicators (e.g. a drop in the volume of dead wood in a forest).

Direct biodiversity monitoring implies that minimum goals must be assigned, e.g. the capacity to detect changes greater than 10% of the average species richness. Starting with those minimum goals, it is possible to optimise monitoring to meet the goals, in terms of the number of sites, visits, traps, etc. (Archaux et Bergès, 2008). It is absolutely necessary to establish a rigorous protocol capable of severely limiting drift in procedures. For flora, important criteria are the surface studied, the time spent and the period of the year. Generally speaking, it is better to use existing protocols so that they gradually become a reference for subsequent studies in a progressive process of standardisation. This effort in favour of consistency must be carried out on the European scale, not simply the national. Some scientific monitoring programmes, e.g. for forest flora in the Renecofor network (National network for long-term monitoring of forest ecosystems), include calibration procedures to limit the effect of personnel, in particular for visual estimates of plant cover and for assignment of plants to different vegetation strata. It would be worthwhile to standardise biodiversity-monitoring protocols. There are already coordinated initiatives to harmonise methods (Nageleisen, 2010), for which effective standardisation would be the ultimate step.

Citizen-monitoring programmes obviously have a role to play in this process. In spite of certain limits inherent to programmes based on the participation of volunteers, the STOC-EPS national monitoring programme for common birds has proven its usefulness both as a warning network (e.g. the reduction in the numbers of birds in agricultural environments) and as a source of new knowledge, notably concerning the effects of climate change and fragmentation of natural environments, to say nothing of its training value for the hundreds of amateur ornithologists that participate each year. It is, however, not very probable that similar citizen programmes will be set up for flora and insects in the near future, even though national structures with the means to conduct scientific monitoring on these two groups exist and in fact do so at least partially for certain species (e.g. the regional botanic conservatories, the National forest inventory, the Forest Health department).

In a country such as France where naturalist activities are not particularly developed, at least compared to neighbouring countries such as the U.K., not many people are capable of carrying out biodiversity monitoring and their number would even tend to drop for certain taxonomic groups such as mosses (bryophytes), mushrooms (mycetes) and many insect families. If we want to train future national experts capable of managing monitoring programmes, it is very important to encourage naturalist activities in France, in schools and universities. ■

Authors

Frédéric Archaux
Cemagref, centre de Nogent-sur-Vernisson,
UR EFNO, Écosystèmes forestiers
Domaine des Barres,
45290 Nogent-sur-Vernisson
frederic.archaux@cemagref.fr

Key bibliographical references...

ARCHAUX, F., BERGÈS, L., 2008, Optimising vegetation monitoring. A case study in French lowland forests, *Environmental Monitoring and Assessment*, n° 141, p. 19-25.

ARCHAUX, F., CAMARET, S., DUPOUEY, J.-L., ULRICH, E., CORCKET, E., BOURJOT, L., BRÊTHES, A., CHEVALIER, R., DOBREMEZ, J.-F., DUMAS, Y., DUMÉ, G., FORÊT, M., FORGEARD, F., LEBRET-GALLET, M., PICARD, J.-F., RICHARD, F., SAVOIE, J., SEYTRE, L., TIMBAL, J., 2009, Can we reliably estimate species richness with large plots? An assessment through calibration training, *Plant Ecology*, n° 203, p. 303-315.

DEVICTOR, V., ROBERT, A., 2009, Measuring community responses to large-scale disturbance in conservation biogeography, *Diversity and Distributions*, n° 15, p. 122-130.

NAGELEISEN, L.-M., BOUGET, C., 2010, *Inventaires Entomologiques en Forêt : éléments essentiels pour une standardisation des méthodes*, ONF.

YOCCOZ, N.G., NICHOLS, J.D., BOULINIER, T., 2001, Monitoring of biological diversity in space and time, *Trends in Ecology and Evolution*, n° 16, p. 446-453.

Floristic surveys suppose a long experience and practice of botanic.