16. Future challenges: the importance of benchmarking and intercalibration for integrated application of nitrogen indicators by conservation agencies

M. A. Sutton[a], I. D. Leith[a], C. E.R. Pitcairn[a], P. A. Wolseley[b], N. van Dijk[a] and C. P. Whitfield[c]

a. NERC Centre for Ecology and Hydrology, Edinburgh

b. Department of Botany, The Natural History Museum, London

c. JNCC, Monkstone House, Peterborough

16.1. Introduction

This report has described in detail the results of intensive biomonitoring for nitrogen at 4 test sites plus the application of the moss chemistry and lichen diversity methods at the UK scale. These have provided the basis to consider protocols for the application of nitrogen biomonitoring in example scenarios and to show how biomonitoring can be used to assess the consequences of elevated nitrogen input for the integrity of designated sites.

The analysis shows that, while some biomonitoring methods are directly relatable to designated features of sites, others are only indirectly relatable. For example, application of the Ellenberg scoring method for higher plant species composition is relevant where the CSM target includes sward or woodland understory composition as a designated feature. By contrast, the measurement of epiphytic lichen species composition is only directly relevant at the few sites where lichens are included as part of the site designation. A further difficulty is that the biomonitoring methods most relevant to condition assessment tend to provide the weakest link to the cause of any changes, since observed changes may also be driven or modified by several factors other than air pollution.

In the first instance, it can be considered that improved robustness in biomonitoring for nitrogen can be found by using *several methods in parallel*. If all the methods agree, this provides a more convincing assessment. Such a basic combination of methods, however, does not address the link between monitoring of condition and N deposition as the driver of change. Similarly, it does not answer the question of how the results of a biomonitoring method that is not directly related to site condition can still be used as a quantitative indicator of threat to site condition.

These issues provide the basis to address the future challenges of *benchmarking* and *intercalibration* of N bioindicators. In this section, the concept of the nitrogen "biomonitoring chain" is used as a starting point to develop the basis for benchmarking and intercalibrating bioindicators and to visualize approaches for combining the results of different bioindicator methods. A framework emerges that provides a skeleton for the integrated application of bioindicators, while the key uncertainties in the framework help identify priorities for future development.

16.2. The biomonitoring chain and its implications for linking source attribution and effects on site condition

While an element of robustness in biomonitoring can be found by using the results of several methods, the interpretability of results can be further improved by considering the pathway from air pollution emission to ultimate effect on site condition. Each of the bioindicator methods can be placed at different stages in this pathway, which naturally also includes physico/chemical indicator methods. Figure 16.1 envisages the pathway as a "biomonitoring chain", where measurements can be made at 10 different stages. The first three "links", represent the pollutant threat or exposure, while each of the others represent biological responses of the system and potential effects on site integrity. However, this simple two-way distinction is less helpful than the fact that logical connections apply between each of the stages along the chain.

The biomonitoring chain concept is highly relevant to the practical application of indicator methods. Although the chain is applied here to nitrogen, it is equally relevant for monitoring of other air pollutants and their impacts. Conversely, it must be recognized that not all stages in the chain apply in every context. For example, impacts of N deposition on species composition will often occur without obvious visible injury. In relation to ozone (O_3) impacts, visible injury may occur, but these cannot always be related directly to changes in plant growth. It is important not to get distracted by such obvious exceptions, but instead focus on the wider implications the biomonitoring chain.

The first implication is that, instead of robustness being provided by a random collection of several methods, it is advisable to *select monitoring methods widely distributed along the biomonitoring chain*. Methods toward the top of the chain are better suited to indicate exposure and make the link with source attribution, but provide a weak link to impacts on designated features. Conversely, methods toward the bottom of the chain are better suited to indicate effects on the designated features, but provide a very uncertain link to the cause of change. Hence by selecting several methods distributed along the chain, the logical cause-effect link is strengthened. Such a structured approach to biomonitoring might show at a certain site that:

a) relevant impacts on target species and site condition are occurring,
b) these impacts can be related to biochemical changes reflecting a perturbation of the natural processes at a site,
c) the biochemical changes can be related to exposure of a certain pollutant, and finally that,
d) the high level of a pollutant can be related to emissions from a certain polluting activity.

Of course, such a chain of measured indicators does not prove causality between emission and impact on site condition. Nevertheless, it provides a suitable framework within which causality can most effectively be assessed.

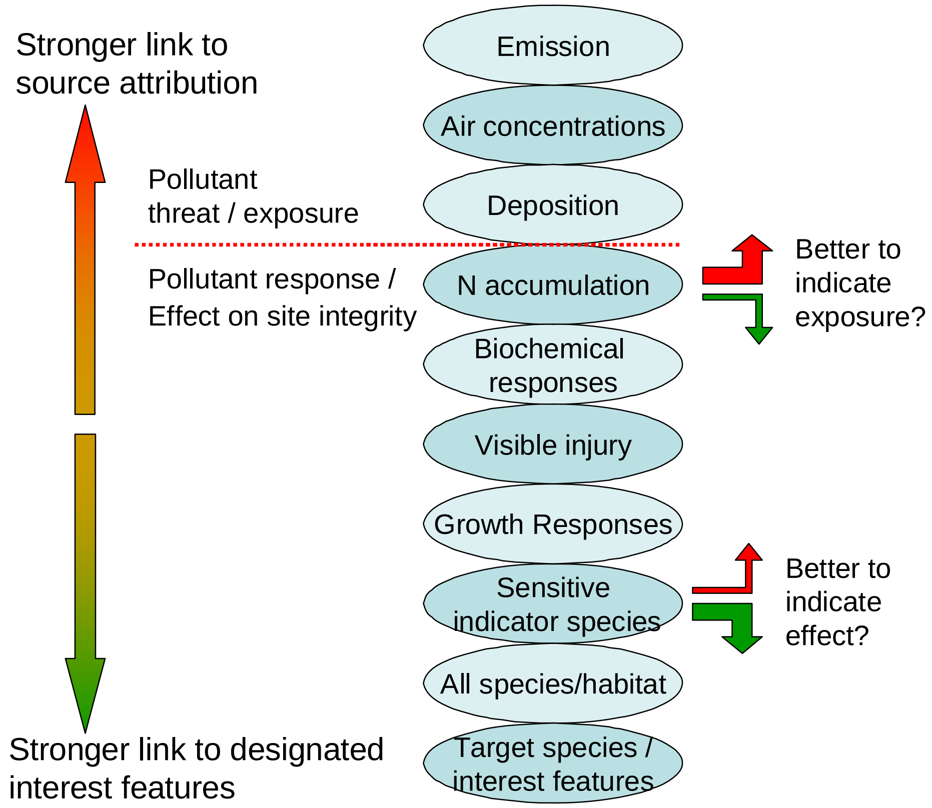

Figure 16.1. Overview of the "biomonitoring chain" showing how different indicator measurements may be ordered from pollutant source to ultimate pollutant impacts. Measurements closer to emission show a stronger link to source attribution, but weaker link to effects on designated interest features. Conversely, species-based measurements show a close link to the designated interest features, but a weak link to source attribution. A comprehensive robust program of biomonitoring should therefore combine measurements distributed along the biomonitoring chain. The dark shaded ellipses show measurements that are, in general, the most practicable.

The second practical point to note about the biomonitoring chain is that some links are harder to measure than others. Fortuitously, the easiest parameters to measure are well distributed along the chain. Thus it is difficult to monitor emissions from many types of air pollution source (e.g. road, farm, complex factory sources etc), but it is straightforward to measure atmospheric concentrations. Similarly, measurement of atmospheric deposition is a complex issue with many uncertainties, but estimates can be directly derived from or related to atmospheric pollutant concentrations.

Continuing down the chain, it may be noted that measurement of pollutant accumulation is in many cases much easier than determining the biochemical response. In the context of nitrogen, the measurements of total foliar N content and soluble NH_4^+ concentration have both been shown to be straight-forward (Sections 8 and 10), and these are easier to measure than responses such as the activity of the enzyme nitrate reductase. The biochemical responses also tend to provide a poorer measure of pollutant exposure than simple biochemical accumulation (Sutton *et al.* 2004a). This is because they are affected both by pollutant accumulation and other constraining factors, such as light, soil type and water availability.

Where relevant, visible injury is straight-forward to record, given appropriate training, however, it does not always represent the possible extent of impact and in addition many impacts to chronic exposure can induce change without visible injury occurring. Although injury is not always an issue at sites affected by chronic nitrogen deposition, injury can occur in the field, for example with high NH_3 concentrations leading to characteristic visible injury to mosses and lichens (I.D. Leith pers. comm.). By contrast, measurement of growth responses of in situ vegetation is very labour intensive and not suited to practical application in support of CSM. Assessment of growth rate changes can be extremely useful, however due to the resources required such measurements are limited to situations where an extremely detailed assessment is justified (e.g. Mitchell *et al.* 2004). An exception is the measurement of growth responses and N inventories of standardised grass 'biomonitors' placed in the field (Section 5 and Appendix II), which can provide a rapid demonstration of impact particularly useful for regulators (i.e. environment agencies) and local stakeholders.

Recording the response of sensitive 'indicator species' can provide a very useful and cost-effective method for assessing the impacts of N air pollution on a given site. By definition, in using the most sensitive species, such responses do not necessarily imply an adverse effect on the designated features of a site. However, they do provide the clearest signal of actual species change. Examples include the use of epiphytic lichen species composition as an indicator of ammonia (Section 7 and 11), or the use of acidophyte and nitrophyte species in higher plant ground flora (Section 4).

Complete monitoring of all species in a designated site is, of course, labour intensive, and this is why CSM tends to focus on a more rapid assessment of target species and/or key interest features. The last link in the chain is thus closely related to CSM, although additional measurements of the target species/ interest features may be relevant in the context of assessing air pollution impacts. For this purpose, starting point may simply be to emphasize the known links between existing CSM monitoring and signals of air pollution impact.

16.3. The need for benchmarking and intercalibation of bioindicator methods

The biomonitoring chain thus provides a useful framework for linking source attribution and effects on site condition. By contrast, it has been noted that being able to make these linkages does not necessarily imply causality between emissions and effects on site condition. Similarly, there remains the question of how quantitatively to relate signals recorded in one part of the biomonitoring chain to signals in another part of the chain. This second problem is well understood in relation to the limitations of physical monitoring. For example, high NO_x concentrations or critical load exceedance do not automatically imply that deterioration of a certain site is currently detectable. However, the problem is not limited to this comparison. Rather, the problem applies to the relationships between all the stages in the biomonitoring chain. Hence, the question can be asked how a high value of foliar N quantitatively relates to both N deposition and to changes in site condition. Similarly, the lichens most sensitive to NH_3 may disappear from a site, but how does this relate to changes in woodland ground flora, the health of oak trees or some other designated feature?

These issues point to the need for benchmarking the results of each of the indicator methods and for improving the intercalibration between their results.

Benchmarking refers to comparing a given bioindicator result against a standard, in particular, defining the *critical values* for each of the indicators in different contexts. These critical values need to be known for each indicator, including how they vary by context. The benchmarks for air concentrations and deposition of N are well known (critical levels and critical loads), but much more effort is needed to develop the benchmarks for different bioindicator methods.

Intercalibration refers to the setting of the benchmarks so that the results of different indicators can be quantitatively compared. This has two parts:

a) *Setting the zero point*: the benchmarks for different indicator parameters should be set so that they reflect a *common standard of effect* on the habitat. This is to say that the point of exceedance of the critical values for the different indicators should occur at the same position on the dose-response function.

b) *Setting the range scale*: the scales for the indicators should be set so that a certain degree of exceedance is comparable between different methods. For example, if several of the bioindicator methods show 150% of the critical values at a site, each 150% should ideally relate to a common point in the dose-response function.

Benchmarking and intercalibration between indicators provide the essential *basis to relate the results of one indicator to another*. Through benchmarking and intercalibration to a common standard, the signals recorded by, for example, epiphytic lichens could be quantitatively related to foliar N concentrations of mosses, atmospheric N concentrations or deposition and vegetation change in woodland ground flora.

Establishing the benchmarks and intercalibration is equally a prerequisite to address the *causality between emission and effect* on site integrity and condition. It was noted above that applying indicator methods well-distributed along the biomonitoring chain does not necessarily imply a causal link (even if there is both a known loss of site condition and known pollutant source). However, if the same suite of indicators is applied, where these have previously been intercalibrated, then consistent exceedance along the biomonitoring chain provides a much stronger evidence base to assess causality.

16.4. Dealing with different units between indicators in benchmarking and intercalibration

One of the major challenges in setting benchmarks and intercalibrating between indicators is that each indicator is measured using different units. The units of atmospheric deposition (e.g. kg N ha^{-1} y^{-1}) differ to those for air concentrations (e.g. $\mu g\ m^{-3}$), while even the different N accumulation parameters use different scales, such as foliar N content of foliar NH_4^+ concentration as $\mu g\ NH_4\text{-}N\ g^{-1}$ FW. The intercomparability appears even harder if the species-based assessments are considered, such as Ellenberg score or acidophyte-nitrophyte values.

Potentially, this could give rise to a multi-dimensional comparison between all the different parameters, which substantially complicates the intercalibration process. Ideally, therefore, the intercalibration should be based on as few unit scales as possible. The unit scales should be naturally quantitative and ideally be relatable to source activity and its mitigation. From this perspective, the natural choice for such a “common currency” is to, as far as possible, base the benchmarking and intercalibration on the units of atmospheric N deposition and air concentrations.

There are different advantages of each of these scales:

- *Atmospheric N deposition* has the advantage that it integrates all forms of N input. However, it has the disadvantage that it is complex and uncertain to estimate, while different N forms may vary in their effects.
- *Atmospheric reactive N concentrations* have the advantage that they are simple to measure accurately and can be related to specific dose-responses. However, they have the disadvantage that effects are often the consequence of more than one form of N including both dry and wet deposition.

Both scales are in general well suited because the scientific experiments that assess impacts are based on responses to deposition or air concentrations. This also facilitates inter-comparability between the practical application of biomonitoring methods and the critical loads/levels approach.

The use of deposition and air concentrations as “common currency” may superficially appear to blur the distinction between the critical loads/levels and the biomonitoring approaches. The two approaches do, however, remain fundamentally different:

- the critical loads/levels approach provides a risk assessment for expected impacts (which are not actually measured), even if it is applied using site-relevant estimates of concentrations/deposition and critical levels/loads.
- the bioindicator approach provides measurements of biological response that give evidence of actual change in site condition.

Having noted this, both bioindicator measurements and the use of site-based critical loads/levels provide complementary components in an integrated assessment using the indicator chain approach.

Some bioindictors respond most closely to N deposition, while others respond more closely to air concentrations of N gases. Additionally, the two scales are not consistently proportional, since the partitioning of wet versus dry deposition varies widely between sites. For this reason, while air concentrations and deposition may provide the units for intercalibration, it is helpful to envisage results in a unit less dimension, e.g. as percentage of the critical value. Hence hypothetical values for all indicators at 100% would imply a site exactly at the critical point. Where values for all the indicators are <100%, this would imply a site in favourable condition/ not under threat, while values consistently >100% would imply a site under threat and with demonstrable impacts on site integrity.

Although air concentrations and N deposition can provide common scales for many of the indicators, they are increasingly difficult to apply further down the biomonitoring chain, where the link between indicator and source is most distant. In these cases, such as Ellenberg score or CSM results, different units are needed, although the aim is to apply the same standard of effect as for the other indicators. In the intercalibration of such scales, setting the critical value for the benchmark is (conceptually at least) straightforward. For this purpose, the concept of the critical load as the deposition below which effects are not observed (according to current knowledge) provides a well understood standard. By contrast, it is a significant challenge to intercalibrate the range-scales to give comparable percentage exceedance values.

16.5. Relevance of spatial and temporal analysis for intercalibration of indicators

There are four main ways in which bioindicators may be used to assess impacts of N on conservation sites, and consideration of these provides further understanding of the need for benchmarking and intercalibration.

The four main assessment types are:

1. **Single point assessment**: Comparison of a measured (single or replicated) indicator value against one or more benchmarked critical values.
2. **Local spatial assessment**: Comparison of indicator measurements at different nearby locations, where conditions are otherwise similar, apart from a known gradient in N exposure.
3. **Regional spatial assessment**: Comparison of indicator measurements at widely separated locations in relation to known regional differences in N exposure, but also including other inter-site differences, such as climate, soils and other pollutant exposure.
4. **Temporal assessment**: Comparison of repeated indicator values measured over time as 'biomonitoring' in relation to temporal changes in N exposure, as well as any other changes that may have occurred.

The temporal assessment, may of course also be applied to each of the 3 assessments other assessment scales. The single point assessment at one time can only be interpreted in relation to established benchmark values. Conversely, benchmarks are not essential to examine relationships with the other three assessments, although they are still needed to determine when and where there is a significant adverse effect. These different assessment approaches are visualized in Figure 16.2, which compares the value of a given nitrogen indicator (I_N) against exposure to reactive N.

Comparison of different values with either space or time is obviously more powerful than assessment of a single value at a single time or location. Space/time assessments provide a more accurate assessment of the extent to which I_N exceeds $I_{N(critical)}$, and information on the shape of the dose-response function. Such assessments provide the basis for setting empirical critical loads for nitrogen (Sutton *et al.* 2004a), but are equally applicable for benchmarking and intercalibrating the other indicators.

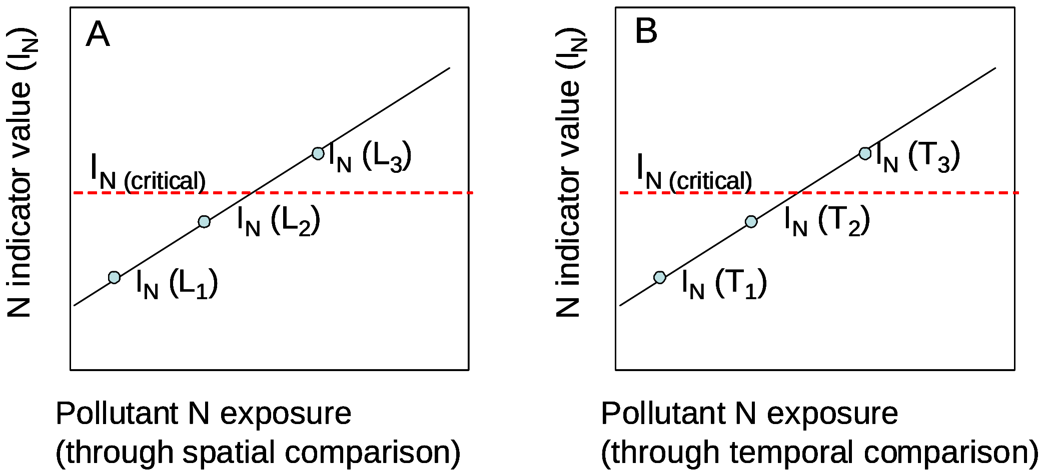

Figure 16.2. Conceptual relationship between the response of a nitrogen indicator (I_N) to pollutant N exposure. $I_{N\ (critical)}$ is the benchmark for significant effects on the ecosystem for the particular indicator used. **A**. the range of pollutant exposure is obtained through local or regional spatial comparison of different locations (L_1..L_n). **B**, the range of pollutant exposure is obtained by biomonitoring over time (T_1..T_n).

Biomonitoring of N impacts can be related broadly to three main situations:

1. **Local impacts in the vicinity of NO_x sources** and associated N emissions (including particles and NH_3). This mainly refers to roads as line sources. NO_x emissions from combustion plant may also have local direct impacts, though this has not clearly been demonstrated (due to the more dispersed nature of the emission).
2. **Local impacts in the vicinity of NH_3 sources** (in particular farms as point sources, or in transects away from agricultural land).
3. **Impacts related to diffuse sources of N deposition**, from wet deposition, plus aerosol inputs of both oxidized and reduced N. This is most relevant at locations distant from sources, so that any local gradients relate to topography rather than the emission sources.

In the first two examples, pollutant exposure varies strongly with distance from source, allowing transect studies ("local spatial assessment") to demonstrate the link between increased exposure and impacts. In these cases, it is possible to perform the dose response assessment, shown in Figure 16.2 A. This allows responses to be clearly detected and related to the pollutant source of origin. In addition, because pollutant levels may be high near sources, such studies can provide a large range of pollutant exposure, allowing clearer responses to be seen. This is particularly the case for NH_3 emissions from farms, due to the much higher local rates of N deposition this can generate compared with NO_x emissions from roads.

By contrast, for diffuse sources of N deposition, it is not possible to make a local spatial assessment in relation to nearness to source. Spatial comparisons may be made by comparing sites locally (e.g. in relation to local topographic effects on deposition) or regionally (in relation to country-scale differences in pollution levels). In such assessments for impacts of diffuse N deposition, it is however, harder to detect clear I_N signals that can be related to N exposure. This is because both topographical and regional comparisons also lead to differences in site climate, management, soils and other pollutant exposure. In addition, because diffuse sources of N deposition are necessarily already dispersed in the environment, such comparisons tend to provide a smaller range of N deposition, than in near-source assessments, making it harder to detect the I_N response. Given these increased uncertainties, the use of time as a variable for changing N exposure is particularly important for assessing the impact of diffuse atmospheric N deposition.

16.6. Defining benchmarks for different nitrogen indicators

The importance of benchmarking nitrogen indicators means that the potential to set such values is a key additional criterion in selecting practical indicator methods. Using the concept of the biomonitoring chain, Figure 16.3 suggests benchmarks in the form of different critical parameters for each indicator. The figure summarizes the status of each critical parameter on the left, while on the right, the status of model or measurement approaches to estimate the indicator is also summarized.

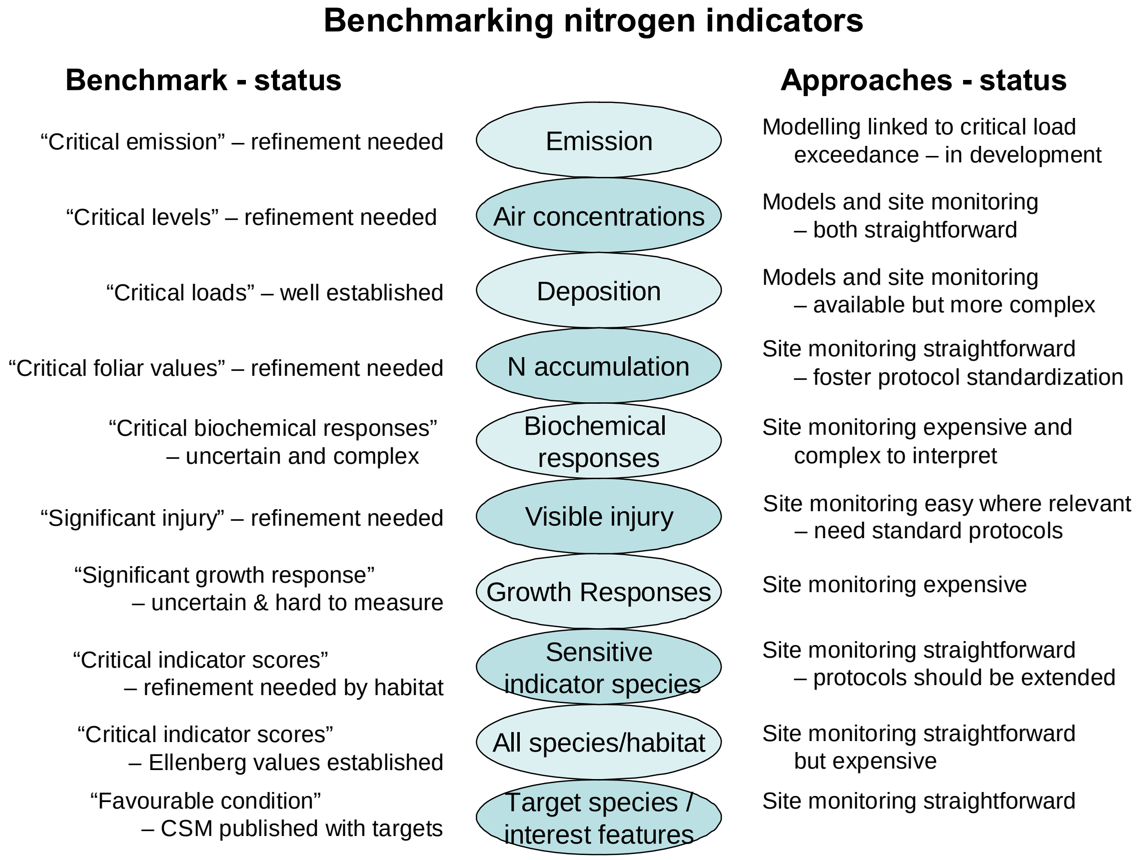

Figure 16.3. Summary of potential and existing benchmarks in relation to nitrogen indicators along the biomonitoring chain. The status of each benchmark is summarized, noting where further development of the benchmark is needed. In addition, the availability of measurement or modelling approaches is summarized for each indicator, noting the status and key development needs. Dark shaded ellipses indicate indicators that are most practical to measure/monitor.

Figure 16.3 includes some very well known benchmarks for which the approaches are well established (e.g. critical loads, CSM). However, there are many components which are uncertain and require further development.

Key points include:

- The concept of critical emission is relevant for environmental agencies. Refinement of this approach may come from the application of simple screening models, such as SCAIL (Theobald and Sutton 2003).
- While the concept of critical levels is well established, the intercalibration to critical loads and other indicators contains major uncertainties. For example, the lichen data (Section 11) and air monitoring results (Burkhardt *et al.* 1998) show that the current NH_3 critical level is not consistent with the N critical load.
- Further refinement is needed of the critical foliar values, although the emerging datasets provide a good basis for such values (e.g. Section 12).
- Biochemical responses are considered rather less certain and more complex to benchmark, given the importance of other limiting factors.
- Significant injury is an area where much has been done for other pollutants, such as O_3. Standard protocols need to be developed for the known injury signals to N.
- Growth responses of native species in the field require significant resources. Conversely, practical protocols are already available for the application of standardized grass biomonitors. Pending the refinement of benchmark values, these are best suited to local spatial assessments.
- The protocols for scoring indicator species are available, but efforts need to be placed in standardizing the critical indicator scores and range scales, both for sensitive species and for habitats.

16.7. Integrated visualisation of indicator results in relation to benchmarks

16.7.1. Integrated visualisation for hypothetical data

The integrated assessment of N indicators can support a site assessment in relation to given policy objectives, such as a review of permit for an emitting source adjacent to a European designated site. As explained above, in order to ensure a strong link between source attribution and impact, methods should be selected well-distributed along the biomonitoring chain. Given the need to be cost effective in the monitoring, the methods shown with dark shaded ellipses in Figure 16.3 provide a sound focus for measurements.

Once each indicator method is benchmarked with a critical value for a given context, it is important to compare the different indicator results in relation to the common critical values. This is illustrated by four conceptual examples in Figure 16.4, which show the percentage of the critical value for each of a series of practical indicators.

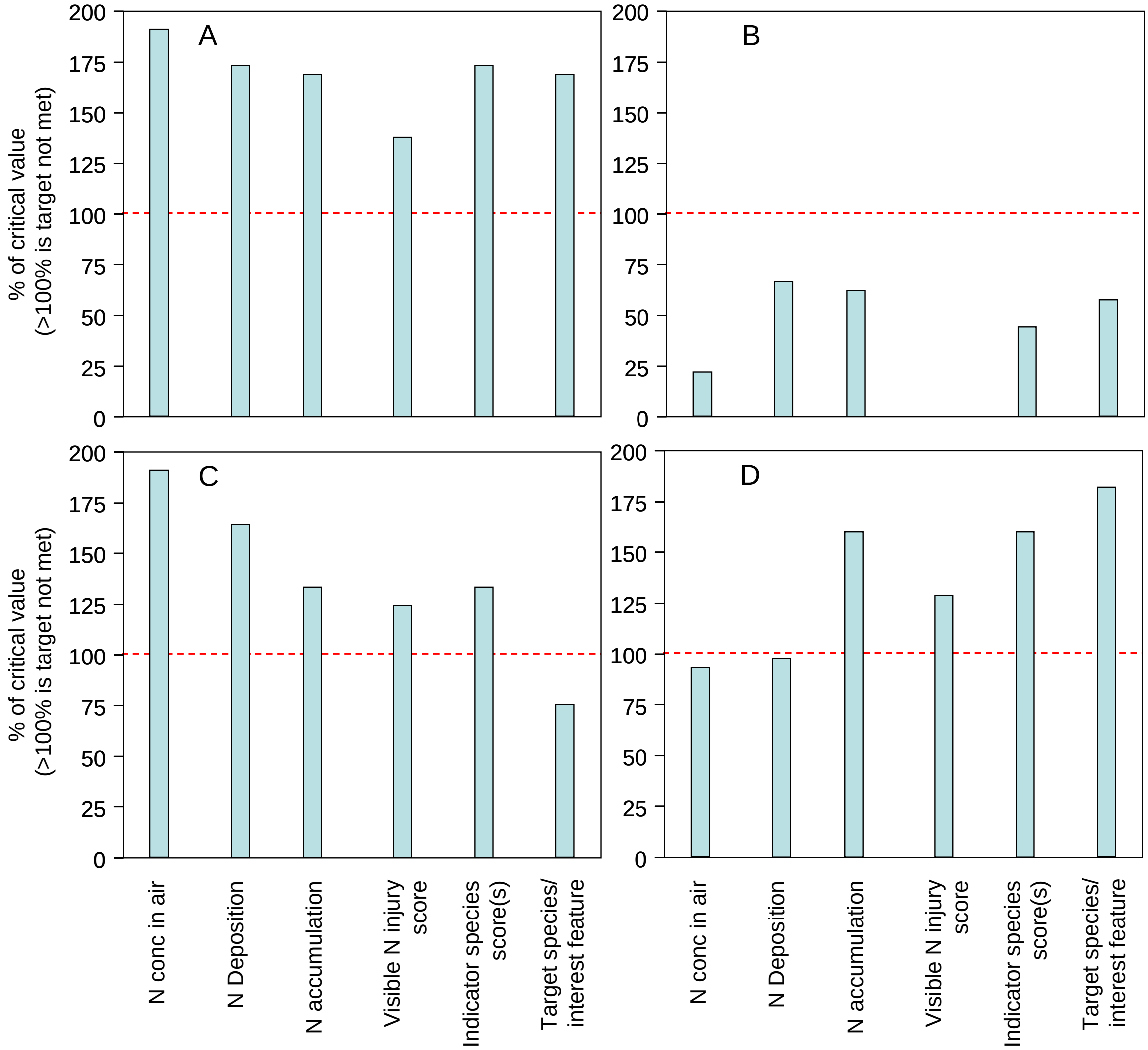

Figure 16.4. Conceptualization of how benchmarked N indicator values measured in four different contexts may be compared in relation to the critical values for each indicator.

A. Site where all indicators exceed critical values: significant effects on site integrity linked to exceedance of N exposure thresholds.
B. Site where all indicators are less than critical values: no evidence of site integrity under threat from excess N.
C. Site where interest features not recorded as under threat (e.g. CSM extent of unfavourable condition smaller than critical threshold), but all other indicators indicate a significant threat to site integrity. The decreasing profile from left to right may indicate a site of with a recent increase in N exposure, and where effects on interest features are expected in the future.
D. Site where interest features are in unfavourable condition, in agreement with results of sensitive indicator species injury and N accumulation, but where N exposure indicators are less than critical thresholds. This may reflect a site formerly exposed to high N pollution exposure, which has not yet fully recovered to favourable condition, or a site which has been subject to another stress factor (e.g. agricultural fertilization, over grazing).

Values larger than 100% reflect measurements showing that the critical value is exceeded. Should all indicators exceed 100% (Figure 16.5, Case A), this provides very strong overall evidence that a significant adverse effect is both apparent (e.g. from monitoring of interest features) and can be related to unsustainably high pollutant levels (e.g. from air concentration measurements). In addition, the intermediate indicators in the chain support the connection between source and impact. By contrast in Case B, all of the indicators are less than 100%. Hence, based on these indicator results, there is no evidence of air pollution threat and impacts in Case B.

Cases C and D illustrate how measurements distributed along the biomonitoring chain may provide additional information. In Case C, although a significant problem is not identified in the site interest feature, all the other indicators are exceeded, particularly those most closely related to N exposure. This indicates a significant adverse effect to site integrity, which may be expected to be manifested as loss of site condition in the future. Finally, case D, represents an example where exceedance of critical values for the indicators most close to the interest features could be due to either historic levels of N pollution or causes other than high nitrogen deposition. For example, such a result might be obtained if mineral N fertilizer were applied to a site.

While the four cases A to D are hypothetical and idealised, they show how the results from an integrated framework for nitrogen indicators may be visualized and interpreted. The indicator scores are presented sequentially in the order of the biomonitoring chain, while the use of intercalibrated percentage values facilitates the intercomparison between different indicators. This framework provides the basis for an operational assessment using indicator methods.

16.7.2. Preliminary benchmarking and intercalibration of nitrogen indicators

Based on the visualisation framework, the next step is to actually assign the critical indicator values and intercalibrate range scales. In a large part, this must be a major challenge for ongoing work. Nevertheless, there is already a substantial basis to set values, which allows a preliminary application of the integrated indicator framework.

The following approaches are used here to set the benchmarks and intercalibration of a selection of the most applicable indicators. These are applied in the following Section (16.7.3) to example intensive and extensive sites of this study.

Air concentrations:

Critical value: Apply the critical level for the habitat or receptor. For NO_x the annual critical level of 30 $\mu g\ m^{-3}$ as NO_2 is used here. However, for NH_3 it is known that the critical level for NH_3 (8 $\mu g\ m^{-3}$; annual mean) is not well intercalibrated with the critical load (Burkhardt *et al.* 1998). This is acceptable if application of the critical level is solely to reflect direct concentration effects, as compared to the consequences of NH_3 on N deposition. However in the present context, the aim is to intercalibrate a baseline of minimum significant effect. Therefore, it is relevant here to consider the NH_3 air concentration that would give rise also to indirect effects (which may be much lower than 8 $\mu g\ m^{-3}$). Additionally, the lichen data from the present study (Sections 7 and 11) and those of Sutton *et al.* (2004a) point to direct effects of NH_3 at mean concentrations of 1-2 $\mu g\ m^{-3}$. For the present application, the critical value for NH_3 is set conservatively at 2 $\mu g\ m^{-3}$.

Range scale: For concentrations (C_N), this is straightforward and linear, with the % value being derived as $C_{N\ measured} / C_{N\ critical\ value} * 100$.

Deposition:

Critical value: This is the best established nitrogen indicator, through the use of critical loads. Empirical values are available for a wide range of habitats (Achermann and Bobbink 2003) and can be applied directly. In the preliminary examples below, the results are calculated using just two nominal values of the critical load: 15 kg N $ha^{-1}\ y^{-1}$ for woodland sites and 10 kg N $ha^{-1}\ y^{-1}$ for sites containing both woodland and bog habitats (Brown Moss, Castle Ennigan).

Range scale: For deposition (D_N), the range scale is straightforward and linear, with the % value being derived as $D_{N\ estimated} / D_{N\ critical\ load} * 100$.

Moss total foliar nitrogen content:

Critical value: This is probably the best established measured bioindicator for N impacts. Significant attention has already been given to establishing critical values (e.g. Pitcairn *et al*. 2002) and these are also estimated in previous chapters of this report both as an average (1.3% tissue N) or habitat specific values. In the preliminary examples below, a standard critical value of 1.3% is applied.

Range scale: The response of total foliar nitrogen (F_{TN}) to atmospheric nitrogen deposition is not 1:1 proportional and therefore the range scale needs to be calibrated. An attractive approach is to apply the results of the regression between foliar N and deposition, thereby providing an estimated N deposition input based on the bioindicator. This has the advantage that the result can be scaled directly against the critical load. However, a reasonable proportionality can be obtained for F_{TN} with the following simpler scaling, which is used in the examples, the % critical value being $(F_{TN\ observed})^2/(F_{TN\ critical})^2 *100$.

Moss soluble ammonium concentration:

Critical value: Based on the revised extraction protocol reported in Appendix I, an average critical value of this parameter has been assessed as 6 µg NH_4^+-N g^{-1} FW (Section 13). Although, Section 13 investigates the use of habitat specific values, to main simplicity in the following preliminary examples, this single value is applied.

Range scale: The response of soluble foliar ammonium content ($F_{NH4\text{-}N}$) to atmospheric N deposition has been clearly shown to have a much stronger slope than that to F_{TN} (Sutton *et al*. 2004a, this report (Section 6). Again the regression between deposition and $F_{NH4\text{-}N}$ might be used as a predictor of atmospheric N deposition with results scaled against critical loads, but this requires further development. For the following examples, a simpler scaling is used whereby the % indicator value is $(F_{NH4\text{-}N\ observed})^{0.5}/(F_{NH4\text{-}N\ critical})^{0.5} *100$, which is found to give broad intercalibration with the F_{TN} values.

Lichen acidophyte-nitrophyte scores

Critical value: As with the moss chemistry methods, two approaches may be considered, either using critical lichen values directly, or by relating a critical lichen value to the critical air concentration. In the former (not used in the following examples), the data from the UK extensive survey (Section 11) indicate a critical value for the difference between the acidophyte and nitrophyte scores (AV-NV) of +5 for trunks and 0 for twigs, with smaller values indicating exceedance. This approach may be useful where there is no data available on bark pH. In the approach applied in the following examples, the critical value is related through regression from the UK survey results (Section 11) to a critical NH_3 concentration, which is taken in this instance as 2 µg m^{-3}.

Range scale: Where only the AV-NV score is available, without data on bark pH, the following range scales provide a reasonable intercomparability between the twig and trunk lichen data (acid barked species, such as oak, only): For trunks the % indicator result may be calculated as (NV-AV+17)/7 * 100. For twigs comparable values may be calculated as (NV-AV+5)/5 *100.

It has been shown however, in Section 11 that there is a significant interaction between bark pH, atmospheric ammonia concentration and the AV-NV score. To account for this the multiple regression of AV-NV vs. bark pH and NH_3 concentration may be transformed to provide a 'calculator' for NH_3 concentration (µg NH_3 m^{-3}). Based on the lichen estimates of NH_3 concentration ($C_{NH3(lichen)}$) the % indicator can be taken as $C_{NH3(lichen)\ observed} / C_{NH3\ critical} * 100$.. The same formula can be applied for twigs and trunks:

$$C_{NH3(lichen)} = 23.63 - 0.615(AV\text{-}NV) - 4\ (\text{bark pH})$$

This approach is used in the following examples for lichen indicators. It is important to note that this function is limited to low NH_3 concentrations (<8 µg m^{-3}), since at very high NH_3, the positive effect of NH_3 on bark pH becomes important compared with the differences in bark pH due to different tree species. This limitation may be assessed by using a critical bark pH value appropriate for different tree species and specific for twigs and trunks. Hence uncharacteristically large bark pH values provide a flag for very high NH_3 concentrations, showing that this equation is not applicable.

Ellenberg scores for higher plants and mosses:

Critical value: There currently seems to have been little focus on benchmarking Ellenberg scores and much more effort is required. For the present preliminary application in the following section only a woodland example is considered. In this case, from the transect studies away from point sources, a critical unweighted Ellenberg mean value of 5 is applied, and this would be expected to vary between habitats.

Range scale: Setting the range scale for mean Ellenberg score (E_N) is equally challenging, particularly since the relationship with atmospheric deposition is so variable. For the present preliminary estimate, the following approach to calculate the Ellenberg % critical indicator is broadly comparable with the other indicators: ($E_{N\ recorded} + 1 - E_{N\ critical\ value}$) * 100. Using a critical value of 5 means that a recorded mean E_N score of 5 gives 100%, while a recorded value of 7 gives 300%.

16.7.3. Integrated visualisation for actual data from this study

The results of applying these benchmark and intercalibrations to example results from this study are shown in Figure 16.5.

The clearest pictures emerge for the Castle Ennigan and Ariundle sites. For Castle Enigan, all the nitrogen indicators indicate exceedance of the benchmark values. From the available parameters, it can be fairly concluded that NH_3 is a major threat to the Castle Ennigan site with altered integrity in terms of moss chemistry and lichen species composition. Actual change in the species composition of the designated features was not assessed, but significant adverse effects would be expected. By contrast, for Ariundle, none of the nitrogen indicators exceed the benchmark values, indicating that this site is not under significant threat from nitrogen deposition.

Of the examples in Figure 16.5, Brown Moss is also shown to be under substantial threat from nitrogen deposition. All the indicators exceed the benchmarks, with the exception of just one (moss ammonium content). Again, although actual change in designated species composition was not assessed, significant adverse effects are expected.

While Castle Enigan and Brown Moss show sites with significant problems due to NH_3, these may be contrasted with Borrowdale and Yarner Wood, where NH_3 levels are low, but N deposition still exceeds critical loads, implying sites with high wet N deposition. It is interesting to note that at both Borrowdale and Yarner Wood the N deposition is the only indicator that exceeds the critical values. This does not mean that significant adverse effects are not expected, but simply that the bioindicators used have failed to demonstrate such effects, and that a more comprehensive monitoring program might still detect effects (e.g. through assessments methods 2, 3 or 4, Section 16.5). Borrowdale provides a good example of this: although the moss foliar N values do not exceed the critical values, a detailed study of growth and foliar N using bryophytes reciprocally transplanted between Ariundle and Borrowdale showed substantial significant N effects at Borrowdale (increased foliar N and reduced growth rates) (Mitchell *et al.* 2004).

It may be noted that at Yarner Wood and Ariundle, some negative indicator values are shown, which is theoretically not possible. This is simply due to scatter in the results, particularly where the indicator value is derived by regression with atmospheric deposition or concentration.

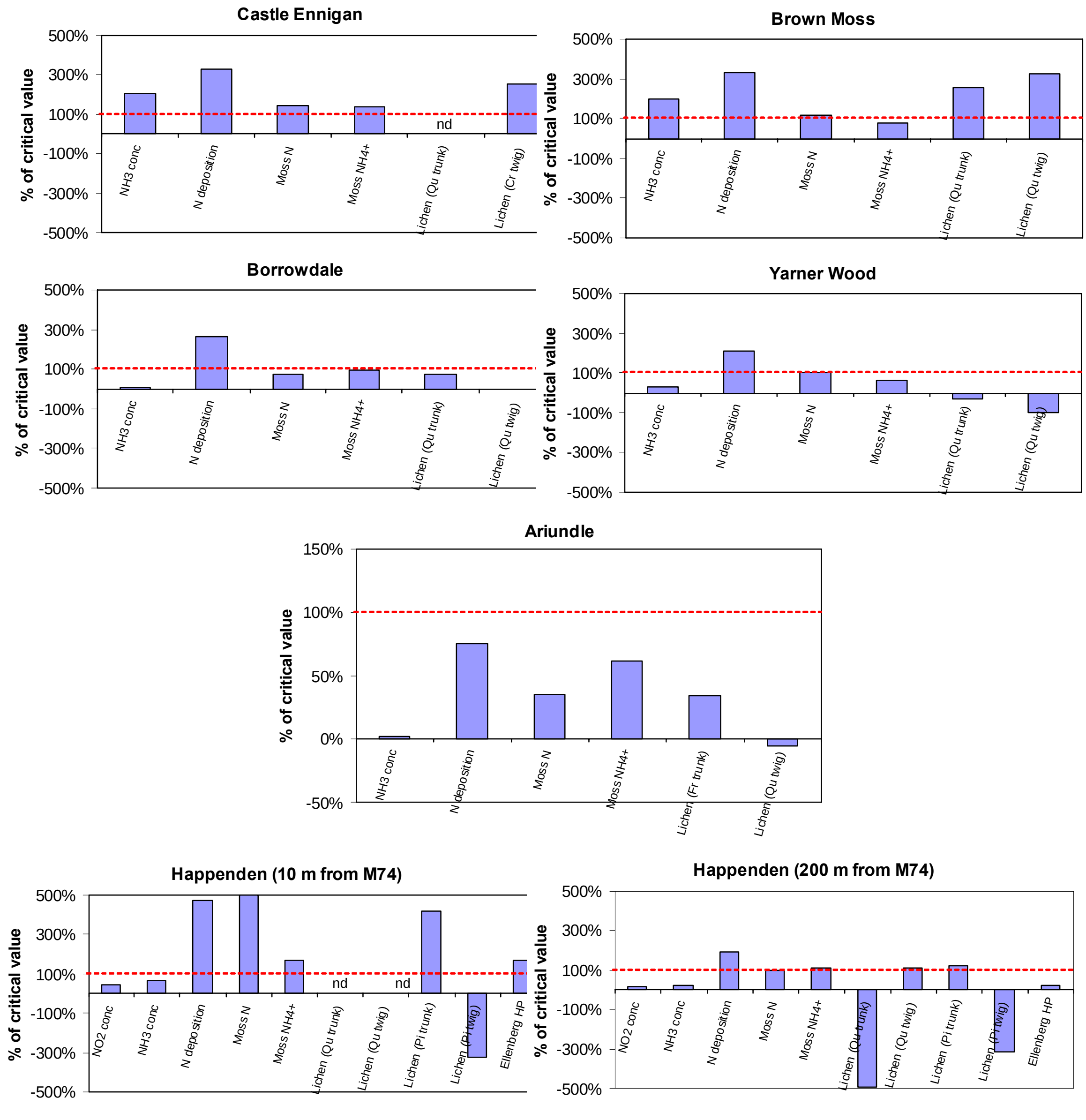

Figure 16.5. Examples of benchmarking nitrogen indicators at sites from this study. Note the different scale for Ariundle; nd = not determined.

The last example of Happendon Wood provides the most complex picture to interpret, and for this purpose the local spatial comparison between results 10 m from the M74 motorway with those 200 m from the M74 is useful. Overall, higher values are shown at the site closest to the motorway. Thus the impact of the motorway on the 10 m site is clear. However, there is a large amount of scatter in the bioindicator results and wide variation between the different indicators. This almost certainly reflects the fact that responses to road pollution are different to those from NH_3 from agriculture. This point was emphasised by the standardised grass biomonitor results (see Section 5), which recorded a *decrease* in growth adjacent to the motorway, despite increased N inputs. Another explanation is that the response to the road is modest compared with the very high wet deposition inputs at this site (shown by the distinction between NH_3 and NO_x concentrations versus total N deposition). Even at the 200 m site, moss chemical parameters are already at the critical limit, although these are increased substantially by the presence of the road at the 10 m site for both total N and foliar ammonium. At the two Happendon Wood sites the lichen assessment gives rather unclear results, and this may be explained by three reasons: a) data for oak were not available for the 10 m site and the 200 m site shows significant scatter, but no clear effect, b) the remaining data are for pine, for which the lichen methodology is much more uncertain, c) the influence of roads and associated N on lichens is significantly different to that of NH_3 from agriculture.

A detailed Ellenberg assessment of the woodland ground flora was made at the Happendon Wood sites. This showed that while the site most distant from the motorway did not exceed the Ellenberg benchmark, the site adjacent to the motorway exceeded it substantially.

16.8. Conclusions

It is obvious that there is significant work needed to refine the intercalibration of nitrogen indicators. In addition to benchmarking the indicators, with critical thresholds, it is essential to ensure that the quantitative responses are comparable. This may seem obvious, but it actually requires careful design of the indicator scales to ensure comparability between such a wide range of different data types (e.g. concentration, deposition, foliar concentrations, injury, species diversity scores etc) and responses.

The choice of how to set such scales can be informed by considering the purpose of site assessments with bioindicators (c.f. Figures 16.4 and 16.5). Firstly, these focus on establishing an integrated quantification of the status of the system in relation to the benchmarks. Secondly, they provide a framework to monitor the success of policy and management actions to reduce atmospheric N deposition to sustainable levels. Given the importance of this second objective, the indicator scales should, as far as possible be set in proportion to the extent of pollutant exposure. Hence, in the example of Figure 16.4a, a reduction of current pollutant exposure of at least 75/175 = 43% would be required to bring the indicators within the critical values in the long term. As noted, scaling indicators by exposure also has the practical advantage in that the scales for air concentrations and N deposition are already well established ($\mu g\ m^{-3}$ and $kg\ N\ ha^{-1}\ y^{-1}$, respectively), with the critical values (critical levels and critical loads) being widely recognized.

Two elements to indicator the range calibration need to be addressed. The first is the *response intercalibration* of the indicators as normally measured. In this case, the objective is to relate quantitatively the response of one indicator to another, e.g. a lichen diversity indicator score to given NH_3 air concentrations. The second is the *unit normalization* of the indicator scores, so that measured values for different indicators can be interpreted on a common basis. In this case, the basic indicator value is transformed to a related measure and presented as a percentage of the critical indicator value in a way that aims broadly to maintain the proportionality between indicators for values above and below 100%.

Recommendations for future work

The following key recommendations are identified:

- The current study found that the selected bioindicator methods were robust at sites with defined N point sources. However, for sites with diffuse long-range N deposition it was more difficult to quantify the impacts of N. Therefore, further development of bioindcator methods needs to be targeted at sites with long-range and often wet dominated N deposition.