

Using metabolic fingerprinting of plants for evaluating nitrogen deposition impacts on the landscape level

ELEANOR A. GIDMAN*, CARLY J. STEVENS†, ROYSTON GOODACRE‡, DAVID BROADHURST‡, BRIDGET EMMETTS§ and DYLAN GWYNN-JONES*

*Institute of Biological Sciences, University of Wales, Penglais Campus, Cledwyn Building, Aberystwyth SY23 3DA, UK,

†Department of Earth Sciences, Open University, Walton Hall, Milton Keynes MK7 6AA, UK, ‡School of Chemistry,

University of Manchester, Sackville Street, Manchester M60 1QD, UK, §Center for Ecology and Hydrology Bangor,

Deiniol Road, Bangor LL57 2UW, UK

Abstract

Nitrogen emissions and atmospheric deposition are globally significant with the potential to alter ecosystem nutrient balance, provoking changes in vegetation composition. Shifts in plant biochemistry are good indicators of nitrogen pollution and have been used to monitor vegetation health. Fourier transform-infrared (FT-IR) spectroscopy has previously been shown to be a rapid and relatively inexpensive method for evaluating leaf biochemistry. In the present study, FT-IR spectra were collected from *Galium saxatile* samples taken from sites across the United Kingdom. Spectral changes in the tissue samples were correlated with a gradient of N deposition using partial least squares regression analysis. We show that FT-IR analysis of *G. saxatile* leaf tissue is an effective way to evaluate nitrogen deposition across the entire UK landscape.

Key words: ecosystem health, FT-IR, metabolomics, nitrogen deposition, partial least-squares regression analysis, plant biochemistry

Received 4 May 2005; revised version received 2 September 2005; accepted 17 March 2006

Introduction

One of the most challenging and urgent research needs in ecology over the last few decades has been to find inexpensive and effective ways of evaluating nitrogen (N) deposition over a *landscape* level. Globally, N deposition threatens natural ecosystems through eutrophication and potential reductions in plant biodiversity (Vitousek, 1994; Vitousek *et al.*, 1997), thereby adversely affecting ecosystem health. A definition of a healthy ecosystem is one that is resilient, retaining organization over time and when subjected to stress (Costanza, 1992). For example, in the United Kingdom and mainland Europe N deposition is known to reduce terrestrial species richness and impact on ecosystem health in communities such as acid grasslands (Stevens *et al.*, 2004) and heathlands (Heil & Diemont, 1983; Carroll *et al.*, 1999). On a global scale N deposition effects are expected to intensify in the future as the human population continues to grow (Tilman *et al.*, 2001), hence

there is a clear need to monitor the level of N deposition and concurrent impacts on semi-natural communities.

Previous efforts to determine the effects of threshold emissions and N deposition rates on ecosystem health have revolved around the so-called critical loads approach. Critical loads are quantitative estimates of pollutant exposure over which significant harmful effects may occur in a given environment (Nilsson & Grennfelt, 1988). However, exactly what constitutes a harmful effect is the subject of an ongoing debate (Skeffington, 1999). Determining exposure to N pollution for a particular ecosystem, in order to assess whether the critical load has been exceeded, has also been fraught with difficulty. Measuring N deposition is a multifaceted undertaking as N species can occur in various states, such as gaseous forms, in precipitation and as aerosols (Pitcairn *et al.*, 2003). While there have been significant improvements in physical monitoring methods, many of these are costly and particularly inappropriate for monitoring in remote areas (Pitcairn *et al.*, 2003).

There has also been considerable interest in the potential for using plant responses as bio-indicators of N deposition (Pitcairn & Fowler, 1995; Huhn & Schulz,

Correspondence: Dylan Gwynn-Jones, tel. +44 1970 622318, fax +44 1970 622307, e-mail: dyj@aber.ac.uk

1996; Hicks *et al.*, 2000; Pitcairn *et al.*, 2003). Foliar amino acids (Huhn & Schulz, 1996) and total tissue N (Pitcairn & Fowler, 1995; Hicks *et al.*, 2000; Pitcairn *et al.*, 2003) have both been shown to correlate positively with increasing N in plant species including bryophytes (Pitcairn & Fowler, 1995; Pitcairn *et al.*, 2003), shrubs (Pitcairn & Fowler, 1995), grasses (Pitcairn & Fowler, 1995) and Scots pine (Huhn & Schulz, 1996). However, the methods required for total tissue N and amino acid measurement are time consuming and costly. We attempted to expand and advance the bio-indicator species concept by investigating responses in tissue metabolic fingerprints of heath bedstraw (*Galium saxatile*) as affected by a UK wide gradient of N deposition (Stevens *et al.*, 2004).

Materials and methods

Metabolic fingerprinting via Fourier transform-infrared (FT-IR) spectroscopy is a relatively rapid technique for screening the global biochemistry of a sample (Goodacre *et al.*, 1998; Fiehn, 2001; Ellis *et al.*, 2003; Gidman *et al.*, 2003; Johnson *et al.*, 2003). Metabolic fingerprinting itself can be considered as unravelling (otherwise hidden) metabolic shifts without attempting to quantify individual biochemical components (Fiehn, 2001). In turn, this allows rapid classification of a sample according to its biological relevance or origin (Fiehn, 2001; Johnson *et al.*, 2003). The data generated give a spectral description of the entire biochemical composition of the sample (Fig. 1). This 'holistic' approach allows high-throughput, unbiased examination of sample biochemistry and, through suitable interpretation, can be used to monitor plant responses to abiotic or biotic influences (Goodacre *et al.*, 1998). Previous work on food quality (Lai *et al.*, 1994; Goodacre & Anklam, 2001; Ellis *et al.*, 2002), bacterial and fungal species (Helm *et al.*, 1991; Goodacre *et al.*, 1998; Timmins *et al.*, 1998), and plant systems (Gidman *et al.*, 2003; Johnson *et al.*, 2003), including those subjected to N application (Gidman *et al.*, 2004, 2005), has already indicated the potential of this approach.

The N gradient examined here involved a survey of 66 acid grassland sites situated across the United Kingdom, with N deposition ranging from 6 to 36 kg N ha⁻¹ yr⁻¹ (Stevens *et al.*, 2004). N deposition measurements for each site were modelled by Centre for Ecology and Hydrology, Edinburgh at a 5 km resolution. Assessment of an N deposition gradient from across the United Kingdom, with recorded environmental variables (Stevens *et al.*, 2004), allowed the examination of plant responses as they were occurring *in situ*, rather than the potential responses identifiable using experimental N applications.

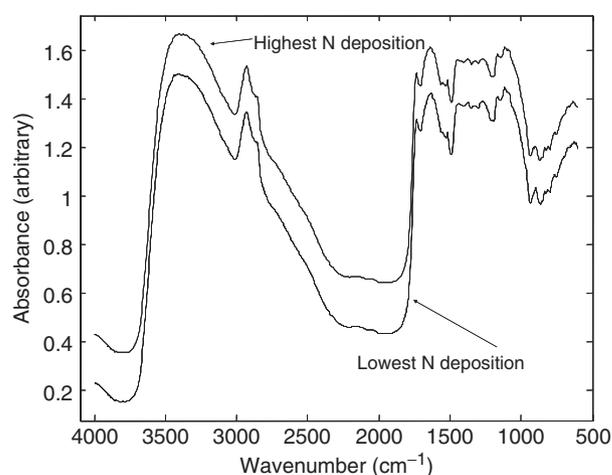


Fig. 1 Examples of *Galium saxatile* Fourier transform-infrared (FT-IR) metabolic fingerprint spectra taken samples obtained from the two sites experiencing the lowest (broken line, 6 kg N ha⁻¹ yr⁻¹) and highest (solid line, 36 kg N ha⁻¹ yr⁻¹) N depositions. For clarity, these FT-IR spectra are offset so that any features can be more readily observed.

Plant tissue samples were collected from five 2 m × 2 m quadrats within a 1 ha sampling area from each site, washed in deionized water, and then oven-dried at 40 °C. Aliquots (10 μL) of sample slurries prepared at 100 mg mL⁻¹ in MilliQ H₂O were then applied in triplicate to the wells of a FT-IR sample carrier plate. The carrier plate was then dried at 40 °C in order to remove extraneous moisture before FT-IR analysis. Spectral collection was fully automated using a motorised stage (Bouffard *et al.*, 1994) and the FT-IR apparatus (Bruker Ltd, Coventry, UK), was controlled using an IBM-PC operating under a standard protocol as described elsewhere (Goodacre *et al.*, 1998; Timmins *et al.*, 1998).

Before analysis, the spectral data were randomly split into a training (50 sites) and a validation set (16 sites); while keeping the triplicate FT-IR spectra from each site together. Analysis was by partial least-squares regression (PLSR; Martens & Næs, 1989), a supervised method of multivariate linear regression (Winson *et al.*, 1997), which was on processed spectral data as per Ellis *et al.* (2002). Sixteen PLSR models, each examining different environmental parameters, were then calculated. These included; total inorganic N deposition (kg N ha⁻¹ yr⁻¹), SO₄²⁻ deposition (kg N ha⁻¹ yr⁻¹), altitude (m), slope (°), aspect (categorical) sub- and topsoil pH, mean annual temperature (°C) and precipitation (mm), actual (AE) and potential (PE) evapotranspiration (mm day⁻¹), soil moisture deficit (mm), topsoil C:N ratio, mean number of species, and total plant and litter cover (%). Consequently, PLSR models were

Table 1 Regression r^2 and F -ratio values, line of best fit (LOBF) equations and PLSR error for significant PLSR models that gave training gradients of >0.5

Environmental parameter predicted	Training data			Test data				Error (RMSEP)
	r^2	F ratio	LOBF	Error	r^2	F ratio	LOBF	
Total N deposition ($\text{kg ha}^{-1} \text{yr}^{-1}$)	0.681	315.46	$x = 7.24 + 0.681y$	4.940	0.388	29.16	$x = 7.24 + 0.681y$	6.627
Total S deposition ($\text{kg ha}^{-1} \text{yr}^{-1}$)	0.675	307.76	$x = 4.84 + 0.675y$	3.780	0.481	42.61	$x = 9.36 + 0.352y$	6.667
Mean annual precipitation (mm)	0.809	627.37	$x = 20.20 + 0.809y$	19.812	0.138	7.35	$x = 91.21 + 0.250y$	48.682
Topsoil C:N	0.723	307.55	$x = 4.68 + 0.723y$	3.553	0.566	51.32	$x = 7.74 + 0.459y$	6.329
Topsoil pH	0.670	300.58	$x = 1.40 + 0.670y$	0.198	0.323	21.93	$x = 3.02 + 0.296y$	0.462
Actual evapotranspiration	0.704	351.46	$x = 13.03 + 0.704y$	1.824	0.111	5.74	$x = 33.65 + 0.224y$	3.492
Mean number of species	0.843	791.95	$x = 2.19 + 0.843y$	1.816	0.536	53.11	$x = 8.64 + 0.442y$	3.747

Each model trained and tested using the same FT-IR spectral data, but with a different environmental variable as a prediction parameter.

PLSR, partial least-squares regression.

individually trained to predict a selection of environmental parameters using FT-IR spectra taken from *G. saxatile* leaf tissue. The validation spectra were used to examine the predictive capability of these models and agreement between training and validation predictions were analysed using linear regression analysis of variance (ANOVA; Sokal & Rohlf, 1969). The outcome of PLSR analysis was also assessed using a 'perfect model' assumption, where a regression of expected vs. predicted scores would give an intercept of '0' and a gradient of '1'. Under our perfect model assumption, any training data regression models that gave gradients of <0.5 were discarded, along with the relevant test data, as being unsuitable.

Results and discussion

Examples of FT-IR spectra that were subjected to the PLSR models can be found in Fig. 1. Interrogation of the resulting expected vs. predicted scores from each model showed that only seven environmental variables gave statistically significant ($P < 0.05$) models that displayed training data gradients greater than >0.5 (Table 1). The environmental variables were N deposition, S deposition, precipitation, topsoil C:N, topsoil pH, AE and the mean number of plant species (Fig. 2). This suggests that within the measured ranges of the available environmental parameters examined *G. saxatile* foliar biochemistry only responds, either directly or indirectly, to these particular variables.

A previous study on this experimental system examined variability in species richness (Stevens *et al.*, 2004). There it was shown that reductions in species richness were significantly correlated with increasing N deposition and decreasing topsoil pH. Precipitation and altitude also seemed to affect biodiversity across the sites,

but this may have been due to the effect of seven high-altitude grasslands in Scotland skewing the model. Removal of these sites from the regression in that study left topsoil pH and N deposition as the only significant influences on species richness, with N deposition explaining 55% of the total variation (Stevens *et al.*, 2004). Thus, from previous work we can ascertain that a level of correlation exists between N deposition, topsoil pH and mean number of species, being a crude measure of species richness. Indeed, our model produces predictive PLSR models based on *G. saxatile* biochemistry responses to all three of these variables (Fig. 2). As there is evidence that species richness is correlated with N deposition and topsoil pH in the studied sites (Stevens *et al.*, 2004), it can be assumed that the model based on mean number of species is liable to be affected by cross-correlation with these variables. There may also be cross-correlations between S deposition and topsoil pH. Additionally, N deposition is currently recognized as a greater cause of acidification than S pollution in the United Kingdom (NEG-TAP, 2001) and, thereby, will also influence topsoil pH.

Concerning the remaining environmental parameters, it is not surprising that precipitation produces a significant PLSR model as wet N deposition will be strongly correlated with this factor (Asman *et al.*, 1998). The other model showing a correlation with AE is likely to be confounded by causality with precipitation (Zhang *et al.*, 2004). From the environmental data taken it also seems that N deposition shows a strong correlation with AE (LOBF: $y = 36.286 + 0.3102x$, $r^2 = 0.483$, F -ratio = 54.28), suggesting further cross-correlation. However, both precipitation and AE gave the worst test correlations (F -ratios of 7.35 and 5.74, respectively) suggesting that any correlations between *G. saxatile* foliar chemistry fluctuations and these factors are the

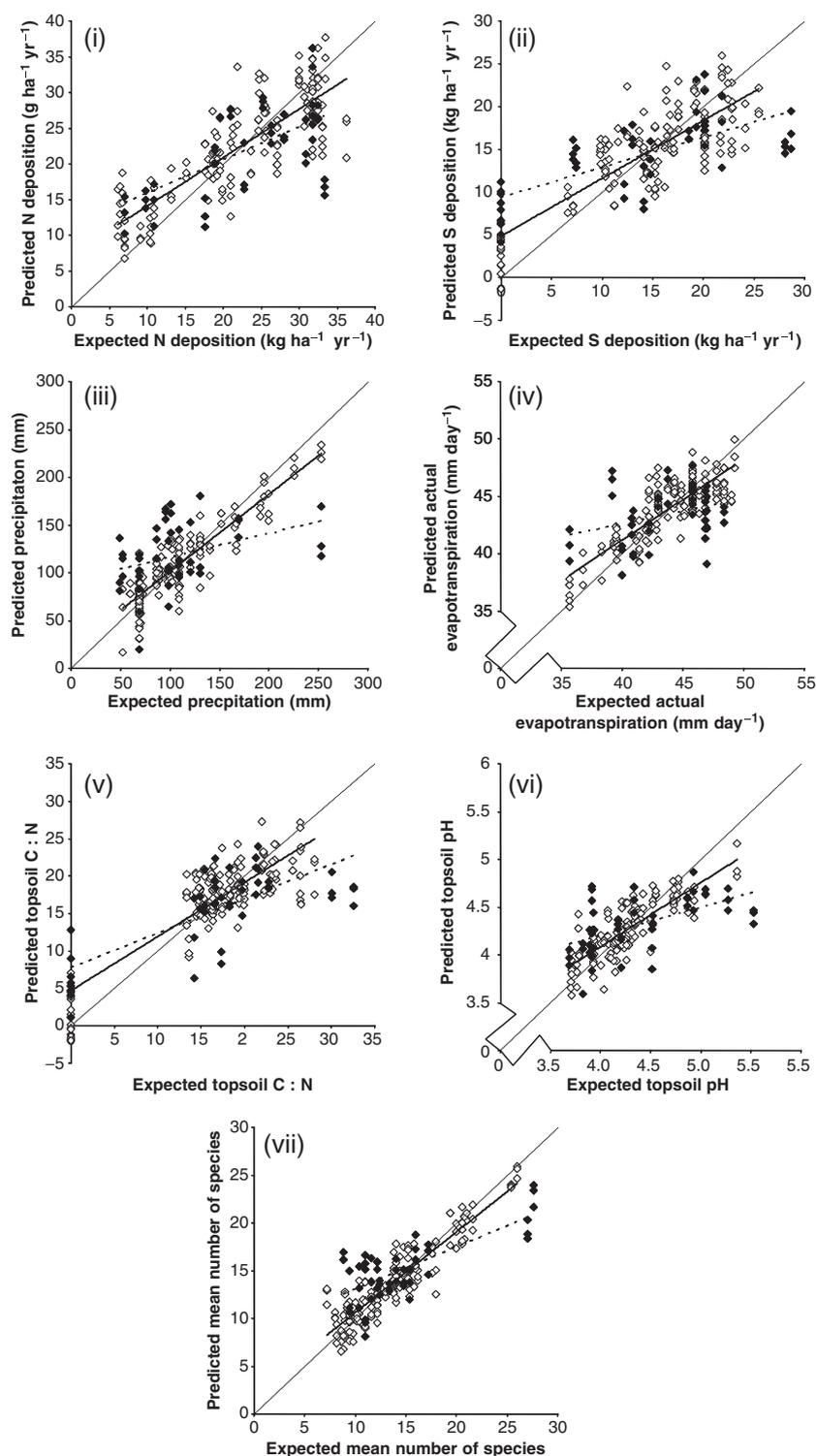


Fig. 2 Plot of expected vs. predicted values for six environmental variables: (i) N deposition kg ha⁻¹ yr⁻¹, (ii) S deposition kg ha⁻¹ yr⁻¹, (iii) topsoil pH, (iv) subsoil pH, (v) topsoil C : N and (vi) mean number of species as calculated by a partial least-squares regression model using Fourier transform-infrared spectra. Predictions for both the training (◇) and validation (◆) data sets are given. Also, the regression line of best fit for the training (solid line) and validation (broken line) data is displayed on each plot. The presumed 'perfect model' line, where the gradient is 1 and the intercept 0, is also given as a faint broken line through the origin of each plot.

most variable (Table 1). For comparison, the test-set correlation F -ratios for N and S deposition were 29.16 and 42.61, respectively (Table 1).

The model examining correlations between topsoil C:N and foliar metabolic fingerprints of *G. saxatile* gave the best test-set r^2 value (0.566) and the second best test-set gradient (0.459) and F -ratio (51.32). Interestingly, examinations of correlations between topsoil C:N and other environmental factors gave no significant regressions (data not shown). This suggests that from the measured parameters no single factor correlates with topsoil C:N, yet variation in topsoil C:N can still be correlated with changes in *G. saxatile* biochemistry.

Loading plots (data not shown) for the N deposition PLSR model were examined in order to assess if any particular regions of the spectra were correlated with this factor. From this three regions in particular, at 1600, ~1493 and ~1430 cm^{-1} were suggested to be responsible for the successful validation of the model. Amide II absorption bands (N-H, C-N and protein vibrations) and C-H bending in CH_2 groups is associated with these regions (Schmitt & Flemming, 1998). Loading plots for the other models suggested that fluctuations in the amide I region could also be partly responsible for any correlations seen (data not shown).

From the above, using the known measurements, it can be hypothesised that by using FT-IR spectroscopy it is possible to correlate *G. saxatile* biochemistry with a range of environmental variables. Whether each model is correlating on biochemistry fluctuations that occur as a direct response to the effects of a given environmental variable or on causal effects related to another factor, which may be unmeasured here, is admittedly uncertain. Variation in models will also be influenced by variation in measurements, especially where those measurements, themselves such as N deposition, are estimates. However, the correlations shown in the presented PLSR models, using data used here has been gathered across a spatial and environmental scale that spans the United Kingdom. They strongly suggests that this technique could be used for predictive purposes.

For the first time, we have shown that metabolic fingerprinting by FT-IR is capable of correlating subtle variations in *G. saxatile* leaf biochemistry that are related to increasing levels of N deposition from sites surveyed across the *landscape level*. Using the same data the technique can also effectively correlate on other environmental parameters that also influence *G. saxatile* metabolism, such as fluctuations in topsoil C:N, S deposition and changes in community biodiversity. This valuable approach is able to interrogate complex ecological systems at the landscape level providing a rapid and sensitive indicator of ecosystem health. The results also confirm previous evidence that N deposi-

tion is a strong influence on acid grasslands in the United Kingdom. We propose to develop these methods further to allow predictive modelling of environmental perturbations.

Acknowledgements

D. Gwynn-Jones and R. Goodacre thank the BBSRC and NERC for financial support to develop this area of metabolomics. C. Stevens was funded by an Open University/NERC CASE studentship on analysis of N gradients in the United Kingdom. C. S. would like to thank N. Dise, D. Gowing and J. O. Mountford. E. A. Gidman thanks the Centre for Ecology and Hydrology Bangor and the NERC for PhD support on N deposition and metabolomics as part of the United Kingdom GANE programme. Modelled nitrogen and Sulphur deposition data were gratefully provided CEH Edinburgh. H. Johnson and D. Causton are thanked for statistical assistance.

References

- Asman WAH, Sutton MA, Schjorring JK (1998) Ammonia: emission, atmospheric transport and deposition. *New Phytologist*, **139**, 27–48.
- Bouffard SP, Katon JE, Sommer AJ *et al.* (1994) Development of microchannel thin-layer chromatography with infrared micro-spectroscopic detection. *Analytical Chemistry*, **66**, 1937–1940.
- Carroll JA, Caporn SJM, Cawley L *et al.* (1999) The effect of increased atmospheric nitrogen on *Calluna vulgaris* in upland Britain. *New Phytologist*, **141**, 423–431.
- Costanza R (ed.) (1992) *Ecosystem Health: New Goals for Environmental Management*. Island Press, Washington, DC.
- Ellis DI, Broadhurst D, Kell DB *et al.* (2002) Rapid and quantitative detection of the microbial spoilage of meat by Fourier transform infrared spectroscopy and machine learning. *Applied and Environmental Microbiology*, **68**, 2822–2828.
- Ellis DI, Harrigan GG, Goodacre R (2003) Metabolic fingerprinting with fourier transform infrared spectroscopy. In: *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis* (eds Harrigan GG, Goodacre R), pp. 111–124. Kluwer Academic Publishers, Dordrecht.
- Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics*, **2**, 155–168.
- Gidman E, Goodacre R, Emmett B *et al.* (2003) Investigating plant–plant interference by metabolic fingerprinting. *Phytochemistry*, **63**, 705–710.
- Gidman E, Goodacre R, Emmett B *et al.* (2004) Applying metabolic fingerprinting to ecology: the use of Fourier-transform infrared spectroscopy for the rapid screening of plant responses to N deposition. *Water, Air and Soil Pollution Focus*, **4**, 251–258.
- Gidman EA, Goodacre R, Emmett B *et al.* (2005) Metabolic fingerprinting for bio-indication of nitrogen responses in *Calluna vulgaris* heath communities. *Metabolomics*, **1**, 279–285.
- Goodacre R, Anklam E (2001) Fourier transform infrared spectroscopy and chemometrics as a tool for the rapid detection of

- other vegetable fats mixed in cocoa butter. *Journal of the American Oil Chemists Society*, **78**, 993–1000.
- Goodacre R, Timmins EM, Burton R *et al.* (1998) Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology*, **144**, 1157–1170.
- Heil GW, Diemont WH (1983) Raised nutrient levels change heathland into grassland. *Vegetation*, **53**, 113–120.
- Helm D, Labischinski H, Schallehn G *et al.* (1991) Classification and identification of bacteria by Fourier-transform infrared-spectroscopy. *Journal of General Microbiology*, **137**, 69–79.
- Hicks WK, Leith ID, Woodin SJ *et al.* (2000) Can the foliar nitrogen concentration of upland vegetation be used for predicting atmospheric nitrogen deposition? Evidence from field surveys. *Environmental Pollution*, **107**, 367–376.
- Huhn G, Schulz H (1996) Contents of free amino acids in Scots pine needles from field sites with different levels of nitrogen deposition. *New Phytologist*, **134**, 95–101.
- Johnson HE, Broadhurst D, Goodacre R *et al.* (2003) Metabolic fingerprinting of salt-stressed tomatoes. *Phytochemistry*, **62**, 919–928.
- Lai YW, Kemsley EK, Wilson RH (1994) Potential of fourier transform infrared-spectroscopy for the authentication of vegetable oils. *Journal of Agricultural and Food Chemistry*, **42**, 1154–1159.
- Martens H, Næs T (1989) *Multivariate Calibration*. John Wiley & Sons, UK.
- NEGTA (2001) *Transboundary Air Pollution: Acidification, Eutrophication and Ground-Level Ozone in the UK*. National Expert Group on Transboundary Air Pollution, Edinburgh.
- Nilsson J, Grennfelt P (1988) *Critical loads for Sulphur and Nitrogen. Report of the Skokloster workshop*. Miljörapport 15, Nordic Council of Ministers, Copenhagen.
- Pitcairn CER, Fowler D (1995) Deposition of fixed atmospheric nitrogen and foliar nitrogen content of bryophytes and *Calluna vulgaris* (L.) Hull. *Environmental Pollution*, **88**, 193–205.
- Pitcairn CER, Fowler D, Leith ID *et al.* (2003) Bioindicators of enhanced nitrogen deposition. *Environmental Pollution*, **126**, 353–361.
- Schmitt J, Flemming HC (1998) FTIR-spectroscopy in microbial and material analysis. *International Journal on Biodeterioration and Biodegradation*, **41**, 1–11.
- Skeffington RA (1999) The use of critical loads in environmental policy making: a critical appraisal. *Environmental Science and Technology*, **33**, 245–252.
- Sokal RR, Rohlf FJ (1969) *Bioometry*. WH Freeman & Co., San Francisco.
- Stevens CJ, Dise NB, Mountford JO *et al.* (2004) Impact of nitrogen deposition on the species richness of grasslands. *Science*, **303**, 1876–1879.
- Tilman D, Fargione J, Wolff B *et al.* (2001) Forecasting agriculturally driven global environmental change. *Science*, **292**, 281–284.
- Timmins ÉM, Howell SA, Alsberg BK *et al.* (1998) Rapid differentiation of closely related *Candida* species and strains by pyrolysis mass spectrometry and Fourier transform-infrared spectroscopy. *Journal of Clinical Microbiology*, **36**, 367–374.
- Vitousek PM (1994) Beyond global warming: ecology and global change. *Ecology*, **75**, 1861–1876.
- Vitousek PM, Aber JD, Howarth RW *et al.* (1997) Human alteration of the global nitrogen cycle: sources and consequences. *Ecological Applications*, **7**, 737–750.
- Winson MK, Goodacre R, Timmins ÉM *et al.* (1997) Diffuse reflectance absorbance spectroscopy taking in chemometrics (DRASTIC). A hyperspectral FT-IR-based approach to rapid screening for metabolite overproduction. *Analytica Chimica Acta*, **348**, 273–282.
- Zhang L, Hinkel K, Dawes WR (2004) A rational function approach for estimating mean annual evapotranspiration. *Water Resources Research*, **40**, W02502.