Disparities between plant community responses to nitrogen deposition and critical loads in UK semi-natural habitats

(the late) Richard J Payne[1,2*], Claire Campbell[2], Carly J Stevens[3], Robin J Pakeman[4], Louise C. Ross[5], Andrea J Britton[4], Ruth J Mitchell[4], Laurence Jones[6], Christopher Field[7], Simon JM Caporn[7], Jacky Carroll[7], Jill L Edmondson[8], Ed J Carnell[9], Sam Tomlinson[9], Anthony Dore[9], Ulrike Dragosits[9], Nancy Dise[9**]

1. Environment and Geography, University of York, York YO105DD, UK.

2. Scottish Environmental Protection Agency, Strathallan House, Stirling FK94TF, UK.

3. Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster LA14YQ, UK.

4. The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK.

5. Institute of Biological and Environmental Sciences, University of Aberdeen, St Machar Drive, Aberdeen AB243UU, UK.

6. Centre for Ecology & Hydrology, Environment Centre Wales, Bangor LL572UW, UK.

7. School of Science and the Environment, Manchester Metropolitan University, Chester St, Manchester M15GD, UK.

8. Department of Animal and Plant Sciences, Alfred Denny Building, University of Sheffield, Sheffield S10 2TN, UK.

9. Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian EH260QB, UK.

** Author for correspondence.

ABSTRACT

Empirical critical loads are widely used to quantify and manage the ecological impacts of reactive nitrogen (N) deposition. Critical load values aim to identify a level of N deposition below which significant harmful effects do not occur according to present knowledge. Critical loads have been primarily based on experiments but these are few in number and have well-known limitations, so there is a strong imperative to test and validate values with other forms of evidence. We assembled data on the spatial variability in vegetation communities in the United Kingdom, and used Threshold Indicator Taxa Analyses (TITAN) to investigate linkages between species changes and modelled current and cumulative N deposition. Our analyses focused on five datasets from acid grasslands, alpine habitats, coastal fixed dunes, dune slacks, and wet grassland. In four of these habitats there was evidence for a significant decline in the cover of at least one species (a 'species-loss change-point') occurring below the critical load, and often at very low levels of N deposition. In all of the habitats there was evidence for clustering of many individual species-loss change-points, implying a community change-point analogous to an ecological threshold. Three of these community change-points occurred below the critical load and the remaining two overlapped with the critical load range. Studies using similar approaches are now increasingly common, with similar results. Across 19 similar analyses there has been evidence for plant species loss change-points below the critical load in 18 analyses, and community-level species loss change-points below the critical load in 13 analyses. None of these analyses has shown community change-points above the critical load. Field data increasingly suggest that many European critical loads are too high to confidently prevent loss of sensitive species.

KEYWORDS: Air pollution, Ammonia, Biodiversity, Nitrogen deposition, Threshold responses.

HIGHLIGHTS:

- We analyse plant cover changes along N deposition gradients for five UK habitats.
- Our study shows both species and community changes below the current critical load.
- Current critical loads may be too high to prevent biodiversity impacts.

1. INTRODUCTION

Reactive nitrogen deposition (N deposition) derived from intensive agriculture, industry and transport emissions is recognised as an important threat to global biodiversity (Baron et al., 2014; Sutton et al., 2011). In terrestrial ecosystems, N deposition is associated with eutrophication, acidification and increased susceptibility to secondary stressors (Dise et al., 2011). N deposition can lead to changed assemblage composition and reduced diversity in plant communities, which may lead to knock-on impacts at higher trophic levels (Nijssen et al., 2017; Payne et al., 2012b; Stevens et al., 2018). These changes may ultimately have significant impacts (both positive and negative) on ecosystem services (Jones et al., 2018; Jones et al., 2014), ultimately imposing a significant societal cost (Sutton et al., 2011).

In many nations the key policy instruments used for the management of air pollution impacts are critical levels (for gaseous pollutants) and critical loads (for pollutant deposition). A critical load is defined as ‘a quantitative estimate of an exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment do not occur according to present knowledge’ (Nilsson and Grennfelt, 1988). Critical loads were originally proposed following the UN ECE Convention on Long-Range Transboundary Air Pollution and have been developed and applied in Europe over more than thirty years (Nilsson and Grennfelt, 1988). More recently, critical loads have been developed for the USA (Pardo et al., 2011) and the approach has also been trialled in many countries around the world (Kuylenstierna et al., 2001; Liu et al., 2011; Reinds et al., 2008). Four different values are relevant to nitrogen pollution in Europe: the critical levels for gaseous ammonia and NOx, and the critical loads for acid deposition and nutrient N; here we focus on the empirical critical load for nutrient N. These critical load values form empirically-based ‘impact floors’ below which the negative consequences of pollution are not expected.

Critical loads are used for two main purposes: policy and permitting. Critical loads are used in policy to understand the large-scale impacts of current air pollution and the potential consequences of future pollution scenarios. Critical loads are also used to make decisions concerning the permitting of new pollution sources. In the United Kingdom this involves the modelling of additional N deposition from a proposed development (the ‘process contribution’) and then the appropriate agencies making judgements of potential harm to conservation-designated sites, based largely on critical load exceedance.

One of the key limitations of the critical load concept is that - as strictly defined - it is a binary: dividing locations at risk of impacts from those which are not. This is simple and easy to understand, but can often be unhelpful in practice. For instance, the empirical critical loads ranges for nine UK habitats, including widespread ecosystems such as blanket bogs, begin at just 5 kg N ha^{-1} yr^{-1} (Bobbink & Hettelingh, 2011). However, modelling data show that more than 96% of the UK receives N deposition above 5 kg N ha^{-1} yr^{-1} [CBED model, 2014 (Smith et al., 2000)]. Most permit applications will be for sites at which the critical load value is already exceeded. Economic imperatives mean that prohibiting all additional N deposition in these sites is often unrealistic and the critical load values offer no direct information on the consequences of additional loading in sites where the critical load is already exceeded. In practice, the degree to which N deposition exceeds the critical load value (the ‘exceedance’) is often taken as an index of harm. However this usage goes beyond the original definition in making the implicit assumption that impacts develop linearly: every additional kg of N produces the same degree of environmental harm. This assumption is rarely tested and there is some evidence that N deposition impacts often do not develop incrementally but rather show 'threshold-like' responses whereby the rate of change in a biological assemblage varies with the intensity of a driver. Previous studies have shown non-linear species richness responses to N deposition (Tipping et al., 2013) and disproportionate changes in individual species responses at particular levels of N deposition (Payne et al., 2013; Wilkins and Aherne, 2016; Wilkins et al., 2016).

Empirical critical loads are primarily based on experiments, with other forms of evidence largely restricted to a supporting role. Experiments are ideal for identifying cause-effect relationships, and testing the impacts of N deposition while controlling for other factors, but are poorly suited to the identification of full response patterns because each experiment will rarely have more than a small number of treatment levels. Many experiments also have limitations including small scale, high treatment levels, infrequent treatments, and in many experiments even the 'control' plots have experienced substantial background N deposition. An alternative is to use field data from sites spanning gradients of N deposition. These spatial data are more complicated to analyse and interpret, with a lower signal-to-noise ratio, but offer a better representation of the range of real-world situations with no experimental artefacts and a greater N deposition range. Both experimental and field data have strengths and weaknesses and both have their roles. A useful analogy is medical science where randomised controlled trials provide causal evidence of effects, but epidemiology is essential to understand the real world consequences of external factors.

Over the last 15 years numerous spatial datasets have been analysed to identify N deposition effects on vegetation (Field et al., 2014; Maskell et al., 2010; Payne et al., 2014; Payne et al., 2011; Stevens et al., 2004). These have primarily considered impacts at the level of the community or functional group, although recently effects at the species level have also been considered (Payne et al. 2013, Clark et al. 2019). There is a need to further understand pollution effects on the individual species in a community, since this can pinpoint more exactly the conditions leading to a decline in species of high conservation value (ecologically, economically, or culturally), or an increase in undesirable invasive species. Combining species-specific responses can also allow one to calculate the community-level response to a pollutant, and compare this to existing critical load values; neither of which has been done previously.

In this study we use datasets spanning N deposition gradients in UK semi-natural vegetation to test the critical loads for those vegetation communities. We aim to assess how plant assemblages change with increasing N deposition, pin-point levels at which species and communities show change, and relate these points to critical loads. We simultaneously consider both current N deposition and cumulative N deposition in order to understand differing responses to current conditions and long-term N exposure.

2. METHODS

2.1 Vegetation and N deposition data

We first compiled a pool of vegetation data for UK semi-natural habitats. Full details of these datasets and their compilation are presented in Payne et al. (2019) and summarised in Supplementary Table 1. We considered two metrics of N deposition: current and cumulative deposition. Current deposition was estimated for each survey site using data from the CBED model for the year of data collection, or the latest year in the case of surveys conducted over multiple years (Smith et al., 2000). Current annual N deposition is the metric that is the basis of critical load values and is most widely used in air pollution management and policy. However, there is extensive evidence for long-term N accumulation in ecosystems (Rowe et al., 2016) and 30-year cumulative deposition generally explains greater significant variance in species cover (Payne et al., 2019). Available evidence suggests that this 30-year cumulative metric may be more ecologically meaningful, but it is also less widely used, making results more difficult to place in the context of previous research. We therefore conducted parallel analyses for both current and cumulative deposition. Past N deposition was calculated using the FRAME model with historic data on N deposition sources (Tipping et al., 2017), with 30-year cumulative deposition calculated based on linear interpolation between fixed time points and the trapezoidal area method (Payne et al., 2019).

2.2 TITAN

Threshold Indicator Taxa ANalysis (TITAN) was used to identify species and community changes in relation to N deposition (Baker and King, 2010). TITAN focuses on the identification of *change-points* in taxon abundance in relation to environmental gradients, quantification of the uncertainty in these values and, by combining the multiple individual taxon responses, change-points in overall community response. Underlying TITAN is the Indicator Value (IndVal) method of Dufrêne and Legendre (1997); a technique for the identification of taxa which typify groups of an *a priori* sample classification. A taxon with a high IndVal score will have a high concentration of abundances and high fidelity to a single group (Dufrêne and Legendre, 1997; Podani and Csányi, 2010). A taxon with a maximal IndVal score would be found in all samples of a group and only in that group. In TITAN, IndVal scores are calculated for all taxa for all possible change-points along the environmental gradient (excluding very rare taxa and the very ends of the gradient) with permutation tests to assess the uncertainty in these scores.

To assess overall community response, permuted IndVal scores are standardised as z-scores and summed for positive (sum(z+)) and negative (sum(z-)) responses for each possible change-point. Sum(z) peaks highlight values of the environmental variable around which many taxa exhibit strong directional changes in abundance. Uncertainty in these maxima is assessed by boot-strapping and quantiles of the boot-strapped maxima are used as confidence intervals. For each taxon response TITAN also returns measures of purity (the proportion of boot-strap replicates matching group assignment in the original data) and reliability (the proportion of boot-strap replicates with maximum IndVal reaching a specified P-value). Key advantages of the technique are the ability to differentiate individual taxon responses and separate community responses in taxa responding positively and negatively (Baker and King, 2010; King and Baker, 2010).

We applied TITAN to vegetation cover data with species present in fewer than five sampling sites excluded from each dataset. We conducted separate TITAN analyses using both single-year current N deposition and thirty-year cumulative deposition. TITAN was implemented using the TITAN2 package (Baker et al., 2015) in R (R Development Core Team, 2014) with the five most extreme candidate change-points from either end of the gradient excluded from the analysis. TITAN is computer-power intensive, so for speed we conducted initial screening analyses with 250 IndVal permutations and 500 boot-strap replicates but increased this to 1000 IndVal permutations and 1000 boot-strap replicates for the final analyses of selected datasets presented below. Results are presented as sum(z) plots, taxon change-point plots with associated uncertainties, and aggregated community-level change-points. Results are compared to currently-accepted critical load values (Bobbink and Hettelingh, 2011) for each habitat based on accepted conversions between the UK National Vegetation Classification and EUNIS classes, using our best judgement where there was ambiguity in this assignment.

2.3 Inclusion criteria and testing

TITAN was originally developed and tested using datasets from freshwater systems with a single, dominant anthropogenic gradient leading to major assemblage change. The signal of N deposition in large-scale vegetation datasets can be complicated as there are likely to be other drivers of change (other pollutants, climate, land-use etc). The TITAN method does not directly account for co-variates and there is a risk of misleading results if the method is inappropriately applied to datasets where N deposition impacts are absent, weak or confounded by other variables. In this study we adopted a strictly precautionary approach to ensure that TITAN was only applied and results interpreted in situations where it was appropriate to do so.

We first screened out datasets where N deposition was not a significant driver of plant assemblage change, when accounting for co-variates. To identify potentially significant co-variates we assembled a large pool of environmental variables comprising a consistent set of data on mean annual temperature, precipitation (Hijmans et al., 2005), altitude (Farr et al., 2007) and historic peak S deposition (CBED 86-88: (Smith et al., 2000)) along with other relevant environmental data where available for the individual datasets (Payne et al., 2019). We used partial redundancy analysis (RDA) on Hellinger transformed data to test the explanatory power of alternative combinations of explanatory variables (Borcard et al., 1992; Legendre and Gallagher, 2001). From the total pool of environmental data – excluding N deposition variables – we constructed an optimum model using the ordistep function in the vegan R package (Oksanen et al., 2007). Variables selected in this model were then introduced as co-variates in analyses with each of current N deposition and 30 year cumulative N deposition as explanatory variables (Payne et al., 2019). Datasets were taken forward for further analysis if N deposition explained significant variance at $P<0.01$ in Monte Carlo testing. As one of our aims was to compare responses to current and cumulative N deposition, we required that both of these N deposition metrics were significant in these tests.

In datasets where N deposition variables explained significant variance independent of other large-scale drivers of environmental change we conducted TITAN analyses. However, a few of these analyses yielded a relatively small proportion of taxon change-points with high purity and reliability in boot-strap testing. We defined a conservative criterion for adequate characterisation of a species change-point of at least 95% of boot-strap replicates matching original group assignment and P-value. We excluded datasets where at least 30% of taxa did not meet this criterion for both current and cumulative N deposition (Table 1; Supplementary Table 1). In datasets failing this test it is likely that only a small proportion of taxa are unambiguously responsive to N, complicating the quantification of community responses.

The datasets which passed these tests are those on which we based our main analysis. Focussing solely on those datasets where the signal of N deposition is highly significant when accounting for co-variates, and where a large proportion of taxa show pure and reliable change-points along the N deposition gradient greatly reduces the possibility of spurious results. As an additional test of the potential influence of co-variates, we also conducted tests in which we identified and eliminated taxa where N deposition change-points correlated with change-points for co-varying environmental variables (Payne et al., 2013). For each of the co-variates identified in the RDA model-building we conducted a TITAN analysis and identified change-points. We then regressed each of these co-variates against the N deposition variable. We used these regression equations to calculate 'N deposition equivalent' values for each co-variate change-point for each species. Where a species change-point in the N deposition TITAN analysis lay between the 10th and 90th boot-strap percentile of the 'N deposition equivalent' change-point for any co-variate we eliminated this species from the dataset and conducted a further TITAN analysis. The removal of species in these tests does not imply that the change-points are spurious but does suggest that these should be treated with greater caution. The comparison of these results to the original analyses allows us to assess the potential consequences of a scenario in which species change-points reflect co-variates rather than N deposition.

All analyses were conducted with both current and 30 year cumulative N deposition. To quantitatively compare results, for each change-point based on current N deposition we calculated an equivalent cumulative N deposition change-point value based on a linear regression between current and cumulative N deposition in each dataset (Supplementary Fig. 2). We then compared these values and calculated the proportions which were higher or lower than those based on the cumulative N deposition TITAN analysis (Supplementary Fig. 3). This analysis is used to provide insight into the relative position of change-points in terms of current and cumulative N deposition. The change points were compared with the latest version of the empirical critical loads for European habitats (Bobbink & Hettelingh, 2011).

3. RESULTS

3.1 Data selection and screening

We ultimately focused our study on five of the candidate datasets. A large proportion of the datasets (28 of 36) were eliminated at the first screening stage as one or both N deposition metrics failed to explain significant variance in redundancy analysis with co-variates partialled out (Supplementary Table 1). Many of these datasets did meet $P<0.05$ but not the more conservative $P<0.01$ we opted to use as a screening threshold. In 11 of these cases the lack of significance related solely to current N deposition, with cumulative N deposition explaining significant variance. This is not unexpected given that previous analyses of these data have shown that cumulative deposition is a better predictor of assemblage composition (Payne et al., 2019). However, given the aim to compare TITAN results between deposition metrics and to critical load values which are defined solely in terms of current deposition it was considered important that TITAN could be meaningfully applied based on both current and cumulative deposition. A further three datasets were excluded based on a high proportion of taxa with low purity and reliability change-points in initial TITAN analyses. In two of these cases the purity and reliability criteria were not met for both current and cumulative deposition, while in one dataset these criteria were not met only for current deposition (Supplementary Table 1).

The exclusion of datasets in this filtering exercise does not imply that they contain no evidence of nitrogen deposition impacts, and certainly not that these vegetation types are insensitive to N deposition. On the contrary, N deposition variables are significant in most ordination analyses

(Payne et al., 2019) and plausible change-points are often identified for individual species. However, the lower significance of N deposition in initial redundancy analyses and lower proportion of pure and reliable indicator taxa means that community-level responses are less likely to be robustly identified and there is a greater risk of results being confounded by co-varying environmental factors. Following this filtering we focused on the five datasets which met our criteria: the acid grasslands dataset of Stevens et al. (2004) and Stevens et al. (2006); the 'alpine' habitats dataset of Ross et al. (2012) and the wet grassland, fixed dune and dune slack components of the Scottish Coastal Resurvey dataset (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017) (Table 1). These five datasets are from a range of habitats with a variety of levels and ranges of N deposition (Table 1; Supplementary Fig. 1). Compared to the datasets which did not meet our inclusion criteria these five are notable for relatively large sample sizes and relatively high species numbers: attributes which are clearly likely to aid the identification of community responses (Supplementary Table 1).

3.2 Community changes

3.2.1 Wet grasslands

In the wet grasslands fourteen species showed high reliability/purity negative responses to current N deposition but only two showed positive responses (Fig. 1A). The first negative responses to current N deposition were seen from less than 3 kg N ha^{-1} yr^{-1} in species including the forbs *Veronica arvensis* and *Daucus carota*. Responses to cumulative N were generally similar but with notable differences in species ordering. For instance, the forb *Euphrasia officinalis* had the second lowest high purity/reliability negative response change-point in the cumulative N analysis but the highest in the current N deposition analysis. Positive response change-points were apparent in *Cirsium arvense* and *Cirsium vulgare* in both analyses, and in *Urtica dioca* in the cumulative N analysis.

In terms of the overall assemblage, there was a strongly 'peaked' response in negative-responding species (sum z-), typical of an ecological threshold (Figure 2A). There was little trend in aggregated positive responses (sum(z+)) due to the small number of positive-responding taxa. The response was more peaked – indicative of a more abrupt and 'threshold-like' response – in the cumulative than the current N deposition analysis. The current N deposition sum(z-) peak was centred on 3.9 kg N ha^{-1} yr^{-1} and was tightly constrained in boot-strapping (3.5-4.9 kg N ha^{-1} yr^{-1}) while the cumulative deposition sum(z-) peak was centred at 86 (80-98) kg N ha^{-1} (Table 2). In both analyses the bootstrap confidence intervals of the sum(z+) peak spanned a large proportion of the total deposition range. Few change-points (none high purity/reliability in the current N analysis) were identified as potentially affected by co-varying variables, and the exclusion of these taxa made little difference to the results (Fig. 1, Table 2). The current N deposition sum(z-) peak (3.9 kg N ha^{-1} yr^{-1}) was well below the existing critical load for the habitat (10-20 kg N ha^{-1} yr^{-1}).

3.2.2 *Acid grasslands*

The acid grasslands showed a large number of negative (z-) change-points with these clustered at the lower end of the current N deposition gradient (Fig. 1B). There were far fewer positive response change-points, and these typically occurred at higher levels of N deposition. Species responding positively to N included the moss *Hypnum cupressiforme* and the grass *Nardus stricta*. The UK acid grasslands dataset is a subset of the European-scale acid grasslands dataset (Stevens et al., 2010) previously analysed using TITAN by Payne et al. 2013, and the pattern of species responses was generally similar. In the current N deposition analysis, six high purity/reliability change-points occurred below the critical load, ten within the critical load range and seven above the critical load range.

Results based on cumulative deposition showed a similar pattern, with some difference in species ordering. The most marked difference was that in the analysis based on cumulative deposition many species change-points clustered at the level of the lowest change-point. Sum(z-) results for current N deposition showed a peaked response while there was little strong trend in sum(z+) response (Fig. 2B). Results for cumulative N deposition were similar, but the sum(z-) peak was more elongated. The current N deposition sum (z-) peak was centred at 13.2 kg N ha^{-1} yr^{-1}, within the critical load range, and the cumulative N deposition peak was centred at 216 kg N ha^{-1}. Both were tightly constrained in boot-strapping (Table 2). Both values were modestly affected by the exclusion of taxa which were potentially affected by co-varying environmental factors, but confidence intervals extensively overlapped (Table 2). There was no evidence for a sum(z+) threshold in either dataset, with peak location very variable under boot-strapping.

3.2.3 Alpine habitats

Seventeen negative response and eleven positive response change-points were identified in the alpine dataset (Figure 1C). The negative-responding taxa clustered into two groups responding around 5-7 and 10-14 kg N ha^{-1} yr^{-1}, with the positive-responding taxa more widely distributed along the gradient (Fig. 1C). The pattern of species responses in the cumulative N analysis was broadly similar. Species showing negative responses in both sets of analyses included the lichen *Cladonia uncialis* and moss *Racomitrium lanuginosum*, while positive responses were present in taxa including the grass *Festuca ovina* and moss *Pleurozium schreberi*. Most current N deposition species change-points occurred within the critical load range, which in this dataset spanned most of the total N deposition range (5-15 kg N ha^{-1} yr^{-1}).

Sum(z) plots for both current and cumulative N deposition, with both positive and negative species responses, showed abrupt increases at the lower end of the deposition range and declines at the upper end of the deposition range but with peaks rather broad indicating a community response which was more gradual than for some of the other datasets (Figure 2C). For current deposition both sum(z-) and sum(z+) change-points were located towards the upper end of critical load range. In bootstrapping, the sum(z+) community change point was relatively consistent across iterations while the sum(z-) change-point was less tightly constrained. All change points were relatively robust to the exclusion of taxa potentially affected by co-variates (Table 2).

3.2.4 Dune slacks

Thirty negative and four positive response change-points were identified in the current N analysis of the dune slacks dataset (Fig. 1D). Results of current and cumulative N analyses were generally similar, with the cumulative N analysis showing more evidence for clustering of change-points (at around 80 kg N ha^{-1}). In both datasets positive-responding taxa included the forbs *Cirsium palustre* and *Salix repens* and negative-responding taxa included the forbs *Centaurea nigra* and *Succisa pratensis*. With both current and cumulative deposition, the sum(z-) plots show marked peaks while there were not strong trends in the sum(z+) plots, presumably due to the limited numbers of taxa showing positive responses (Fig. 2). The sum(z-) peaks were relatively tightly constrained in boot-strapping at around 4.8 kg N ha^{-1} yr^{-1} and 90 kg N ha^{-1} and were not affected by species filtering (Table 2). The sum(z+) peaks were rather variable in boot-strapping with both N metrics. All species and community change-points were below the critical load range (10-15 kg N ha^{-1} yr^{-1}).

3.2.5 *Fixed dunes*

More high purity/reliability species change-points were identified in the fixed dune dataset than the other four datasets considered, with 39 negative and 16 positive response change-points in the current N deposition analysis (Figure 1E). The first negative responses occurred below 3 kg N ha^{-1} yr^{-1} and the first positive responses below 4 kg N ha^{-1} yr^{-1}. High purity and reliability change points were all below 8 kg N ha^{-1} yr^{-1}, compared to the critical load range of 8-15 kg N ha^{-1} yr^{-1}. The lowest negative response to current N deposition was in the forb *Arenaria serpyllifolia* and the lowest positive response in the forb *Cirsium vulgare*. The sum(z) community response plots showed a peak which was more distinct for negative than positive responses and more defined in the cumulative than current N deposition analysis (Fig. 2). Positive and negative change-points were relatively tightly constrained in boot-strapping and only modestly affected by species exclusion (Table 2). Current N deposition community change-points (sum(z-): 5.0 kg N ha^{-1} yr^{-1} and sum(z+): 4.8 kg N ha^{-1} yr^{-1}) were both well below the critical load range.

4. DISCUSSION

4.1 Species responses

Our analysis provides a large-scale assessment of the response of individual plant species to N deposition in a range of habitats. The vast majority of the species responses identified here agree with the known ecology of the species. Species which have well-characterised tolerance or sensitivity to N deposition are typically shown to have positive or negative change-points, respectively. For instance, positive responses are identified in *Urtica dioica*, a well-known nitrophile, and *Cirsium arvense*, a competitive forb (Hamdoun, 1970; Hogg et al., 1995; Pitcairn et al., 2003). Similarly, the pollution-sensitive moss *Racomitrium lanuginosum*, the forb *Plantago lanceolata* and hemi-parasite *Euphrasia officinalis* show negative change-points in response to N deposition (Armitage et al., 2014; Maskell et al., 2010).

A few species show negative responses in some datasets and positive responses in some others (e.g. *Hypnum cupressiforme*, *Carex panicea*), and in many others the value and relative position of change-points varies (e.g. *Luzula campestris, Succisa pratensis*). These remain plausible responses to N deposition. It is possible that plants may respond differently in different habitats with different competitors and environmental pressures: a species of intermediate resilience to N deposition might increase in abundance when competing against N-sensitive taxa and decrease when competing against more resilient taxa. Change-point values are also context-dependent and only reflect the range of N deposition covered by the dataset; values can be expected to differ in datasets covering differing parts of the N deposition gradient. In any correlative analysis it is likely that some changes along the environmental gradient may be coincidental, and individual change-points should not be over-interpreted. It is further possible that results might be affected by inconsistent taxonomy or cryptic diversity. However, overall the results present a convincing representation of plant communities responding to N deposition.

4.2 Plant community responses

In all analyses, many more taxa were identified that responded negatively than positively to N deposition. This result supports the general observation that N deposition is associated with a loss of diversity in many (but not all) semi-natural habitats (Field et al., 2014; Stevens et al., 2004). Across most analyses, there was also evidence for the clustering of species change-points at particular levels of N deposition, implying disproportionate community change. There was stronger evidence for non-linear community change when considering species decreasing in occurrence and abundance than species increasing. Sum(z-) scores often showed distinctly different profiles to sum(z+) scores, being more likely to show a ‘peaked’ response. Sum(z-) change-points were typically similar to or at lower N deposition levels than sum(z+) change-points and were often more tightly constrained in bootstrapping. Evidence for a tightly defined sum(z-) peak, implying an ecological threshold in species loss, was strongest in dune slacks and wet- and acid grasslands, and weakest in the ‘alpine’ habitats. Only a small proportion of all species showed change-points at corresponding positions on co-varying environmental gradients, and exclusion of these species made little difference to community-level results. Overall, these analyses imply that the assumption of impacts generally developing at a consistent rate with increasing N loading may be misplaced; instead there is more evidence for threshold-like effects.

4.3 Relationship to critical loads

In all analyses there is evidence for change in species composition with N deposition and a loss of diversity. There can be little doubt that these datasets demonstrate a ‘significant harmful effect’, but results show little correspondence with critical load values. In three of the datasets all (wet grassland) or most (fixed dunes, slacks) of the sites had N deposition below the critical load. Nevertheless, N deposition explained significant variance, and numerous species change points were identified representing plausible responses to N deposition.

In four of the five datasets negative species change-points in response to N deposition were identified below the critical load value. The sole exception was the ‘alpine’ habitats dataset, where it would be impossible to identify change-points below the critical load because of the lack of very low deposition sites. There has been a difference amongst previous studies in how TITAN results should be related to the ‘significant harmful effects’ principle of the critical load concept. Whereas Payne et al. (2013) focussed on comparing the lowest individual species change-point to the critical load, Wilkins et al. (2016) focussed on the community-level sum(z) change points. The former is arguably more in keeping with the underlying principles, but the latter focusses on the community-level results which are likely to be more robust than the species-level results. In practise, this distinction only matters to one of the five sets of analyses presented here: the acid grasslands dataset in which the community change point was within the critical load range but several individual species change-points were below it. In the other datasets, both the community and lowest species change-points were consistently either within (alpine) or below (other habitats) the critical load range. Considered in aggregate, the results strongly suggest that N deposition has ecological impacts at levels below the currently defined European critical load values for many habitats.

In the first study to adopt the TITAN approach for the identification of N deposition impacts on plant communities, Payne et al. (2013) found that the lowest change-points coincided with the lowest point at which any change-point could possibly be recorded. This could be taken to imply that not only is the critical load value for this community too high, but also that it might not be realistic to identify a level of N deposition at which there is no evidence for negative consequences. Of the analyses presented in this study, only one (fixed dunes with current N deposition) showed the lowest species change-point at the lowest possible level of N deposition. However in several others (e.g. alpine habitats with current N deposition) the lowest change-point was only marginally above this lowermost point or the boot-strap confidence intervals extended to the lowermost point. Whether it is fundamentally possible to identify a level of N deposition below which there are *absolutely* no impacts remains unclear.

Findings of species change-points and sum(z) peaks below current critical load values are now increasingly well-replicated. TITAN results considered robust by the authors are available for nineteen different vegetation datasets spanning a substantial range of semi-natural habitat types in northwest Europe (Payne et al., 2013; Wilkins and Aherne, 2016; Wilkins et al., 2016). In all but one of these datasets, high purity and reliability change-points were identified for some species below the critical load. In twelve of the nineteen datasets, community-level sum(z-) change-points were also identified below the critical load and in none of the datasets was a sum(z-) change-point identified above the critical load range. In seventeen of the nineteen analyses, more species were identified with negative than positive responses, usually by a substantial margin. Correlative studies can never prove cause-effect relationships and all statistical methods have limitations, which in the case of TITAN include sensitivity to co-variates and to the distribution of samples along the gradients (Baker and King, 2013). While these factors could have had some influence on some of these analyses, overall, the weight of evidence is becoming increasingly strong: TITAN applied to large-scale vegetation datasets spanning N deposition gradients suggests that the majority of critical loads are currently too high to prevent all species change, and in many cases may also fail to prevent habitats passing points of disproportionate impact. Similar results have also recently been presented for ectomycorrhizal fungi in European forests (van der Linde et al., 2018), implying that critical loads may similarly fail to prevent harm in other important elements of the ecosystem.

4.4 Current and cumulative deposition

A recognised limitation of empirical critical load-based N deposition management is that critical loads are a steady-state concept and provide no information on the timescales for damage or recovery (Hall et al., 2015). Critical loads are based on current deposition only and do not allow for the possibility of chronic impacts with gradual accumulation of N in ecosystems, despite this being probable (Rowe et al., 2016). There are several implications of this. First, a community in which current N deposition falls below the critical load may still retain extensive damage due to historical N deposition which is largely retained within the plant-soil system. The current concept of critical loads does not distinguish between this community and one with a much lower load of accumulated N due to lower historical N deposition. Second, vegetation communities that have previously shifted due to chronically elevated N deposition will not necessarily shift back to their initial state, even if the excess N is removed from the system. This is especially the case for perennial species. In these cases, active management such as biomass removal and turf cutting may be necessary to restore communities to a desired species composition. Clearly, the less accumulated N in the system, the less likely it is that a community will shift to an altered species composition and the less intervention would be needed to restore that community.

For experimental studies, the critical load-setting process specifies a minimum treatment duration of 1 year for inclusion in assessment (Bobbink and Hettelingh, 2011). A number of long-term N deposition experiments are now available, and results from these studies are given greater weight. However, even long-running experiments have a relatively short duration compared with exposure time-scales of many decades of elevated nitrogen deposition. For this reason it has been stated that ‘critical loads cannot guarantee to offer protection to ecosystems over longer timescales’ (Hall et al., 2015).

The simultaneous analysis of changes based on both cumulative and current N deposition metrics gives the opportunity to compare species and community responses. Patterns of community response were generally similar between current and cumulative N deposition (Fig. 2) but in many of the cumulative analyses more N-responsive species were identified. There were some differences in species ordering which are to be expected, as different species, and particularly different functional groups, may be sensitive to N deposition on differing time-scales. If change-points based on cumulative N deposition were equivalent to those based on current N deposition it would be expected that a similar proportion would be higher and lower when compared by regressing values (Supplementary Fig. 2). However, the actual distribution appears skewed (Supplementary Fig. 3). In all five datasets there were more change-points which were higher when based on current N deposition and in all but one of these datasets the difference was substantial. This suggests the possibility that for an equivalent level of current N deposition more species change-points may be passed in sites which have received higher cumulative N deposition. This possibility has significant implications for the management of N deposition implying that pollution management needs to account for long-term deposition history. This possibility will require further testing.

5. CONCLUSIONS

Our results demonstrate that plant species across a range of habitats respond to N deposition from low levels of deposition which are widely exceeded in the UK and across the developed world. Across habitats there is evidence for non-linear community responses, assuming the nature of an ecological threshold and challenging current assumptions of iterative development of impacts. Numerous individual species change-points and most community change-points lie below the critical load values which are widely used in management, science, and policy, suggesting that these values may be currently set too high. With European critical load values soon to undergo periodic review, there is an urgent need to adopt more systematic approaches to the synthesis of experimental evidence, and make better use of field data for validation and testing (Banin et al., 2014).

ACKNOWLEDGEMENTS

This study was primarily funded by the Natural Environment Research Council, part of UK Research and Innovation, through grant NE/R00546X/1. N deposition modelling was funded by the Natural Environment Research Council through grant NE/J011533/1 under the NERC Macronutrient Cycles Programme. AB, RJM and RoJP were supported by the 2011-2016 and 2016-2021 Strategic Research Programmes of the Scottish Government. We acknowledge support and advice from Dr Willie Duncan (SEPA).

Author contributions: RiJP conceived the study, conducted data analysis and wrote the first draft of the manuscript. UD, ST, AD and EJC conducted N deposition modelling. AJB, RJM, RoJP, MLJ, LR, CJS, CF, SJMC, JC, JLE and RiJP designed and conducted vegetation surveys. All authors contributed interpretation and had the opportunity to comment on the manuscript.

*The authors of this paper report with great sadness that our lead author Richard Payne suffered a fatal climbing accident in the Himalaya in late May, between submission and acceptance of this paper. Richard was an outstanding environmental scientist, an inspiring mentor to his students, a well-respected colleague, and a kind and thoughtful friend. He will be greatly missed.

REFERENCES

Aggenbach, C.J.S., Kooijman, A.M., Fujita, Y., van der Hagen, H., van Til, M., Cooper, D., Jones, L., 2017. Does atmospheric nitrogen deposition lead to greater nitrogen and carbon accumulation in coastal sand dunes? Biological Conservation 212, 416-422.

Armitage, H.F., Britton, A.J., van der Wal, R., Woodin, S.J., 2014. The relative importance of nitrogen deposition as a driver of Racomitrium heath species composition and richness across Europe. Biological Conservation 171, 224-231.

Baker, M., King, R., Kahle, D., 2015. TITAN2: threshold indicator taxa analysis. R package.

Baker, M.E., King, R.S., 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods in Ecology and Evolution 1, 25-37.

Baker, M.E., King, R.S., 2013. Of TITAN and straw men: an appeal for greater understanding of community data. Freshwater Science 32, 489-506.

Banin, L.F., Bealey, W.J., Smith, R.T., Sutton, M., Campbell, C., Dise, N.B., 2014. Quantifying uncertainty in critical loads. CEH Report to SEPA. Scottish Environment Protection Agency, Stirling.

Baron, J.S., Barber, M., Adams, M., Agboola, J.I., Allen, E.B., Bealey, W.J., Bobbink, R., Bobrovsky, M.V., Bowman, W.D., Branquinho, C., 2014. The effects of atmospheric nitrogen deposition on terrestrial and freshwater biodiversity, Nitrogen deposition, critical loads and biodiversity. Springer, Dordrecht, pp. 465-480.

Beaumont, N.J., Jones, L., Garbutt, A., Hansom, J.D., Toberman, M., 2014. The value of carbon sequestration and storage in coastal habitats. Estuarine, Coastal and Shelf Science 137, 32-40.

Bobbink, R., Hettelingh, J.P., 2011. Review and revision of empirical critical loads and dose-response relationships : Proceedings of an expert workshop, Noordwijkerhout, 23-25 June 2010. RIVM, The Netherlands.

Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the Spatial Component of Ecological Variation. Ecology 73, 1045-1055.

Britton, A.J., Beale, C.M., Towers, W., Hewison, R.L., 2009. Biodiversity gains and losses: Evidence for homogenisation of Scottish alpine vegetation. Biological Conservation 142, 1728-1739.

Britton, A.J., Hester, A.J., Hewison, R.L., Potts, J.M., Ross, L.C., 2017a. Climate, pollution and grazing drive long-term change in moorland habitats. Applied Vegetation Science 20, 194-203.

Britton, A.J., Hewison, R.L., Mitchell, R.J., Riach, D., 2017b. Pollution and climate change drive long-term change in Scottish wetland vegetation composition. Biological Conservation 210, 72-79.

Britton, A.J., Mitchell, R.J., Fisher, J.M., Riach, D.J., Taylor, A.F.S., 2018. Nitrogen deposition drives loss of moss cover in alpine moss–sedge heath via lowered C : N ratio and accelerated decomposition. New Phytologist 218, 470-478.

Caporn, S.J., Carroll, J.A., Dise, N.B., Payne, R.J., 2014. Impacts and indicators of nitrogen deposition in moorlands: Results from a national pollution gradient study. Ecological Indicators 45, 227-234.

Clark, C. M., S. M. Simkin, E. B. Allen, W. D. Bowman, J. Belnap, M. L. Brooks, S. L. Collins, L. H. Geiser, F. S. Gilliam, S. E. Jovan, L. H. Pardo, B. K. Schulz, C. J. Stevens, K. N. Suding, H. L. Throop, and D. M. Waller. 2019. Potential vulnerability of 348 herbaceous species to atmospheric deposition of nitrogen and sulfur in the United States. Nature Plants 5:697-705.

Dise, N.B., Ashmore, M.R., Belyazid, S., Bobbink, R., De Vries, W., Erisman, J.W., Spranger, T., Stevens, C., van den Berg, L., 2011. Nitrogen as a threat to European terrestrial biodiversity, in: Sutton, M. (Ed.), The European nitrogen assessment: sources, effects and policy perspectives. Cambridge University Press.

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

Edmondson, J., Terribile, E., Carroll, J.A., Price, E.A.C., Caporn, S.J.M., 2013. The legacy of nitrogen pollution in heather moorlands: Ecosystem response to simulated decline in nitrogen deposition over seven years. Science of the Total Environment 444, 138-144.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. Reviews of Geophysics 45, n/a-n/a.

Field, C.D., Dise, N.B., Payne, R.J., Britton, A.J., Emmett, B.A., Helliwell, R.C., Hughes, S., Jones, L., Lees, S., Leake, J.R., 2014. The role of nitrogen deposition in widespread plant community change across semi-natural habitats. Ecosystems 17, 864-877.

Hall, J., Curtis, C., Dore, A.J., Smith, R.T., 2015. Methods for the calculation of critical loads and their exceedences in the UK. Report to Defra AQ0826. Defra, London.

Hamdoun, A.M., 1970. The effects of different levels of nitrogen upon Cirsium arvense (L.) scop. plants grown from seeds and root fragments. Weed Research 10, 121-125.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965-1978.

Hogg, P., Squires, P., Fitter, A., 1995. Acidification, nitrogen deposition and rapid vegetational change in a small valley mire in Yorkshire. Biological Conservation 71, 143-153.

Jones, L., Milne, A., Hall, J., Mills, G., Provins, A., Christie, M., 2018. Valuing Improvements in Biodiversity Due to Controls on Atmospheric Nitrogen Pollution. Ecological Economics 152, 358-366.

Jones, L., Provins, A., Holland, M., Mills, G., Hayes, F., Emmett, B., Hall, J., Sheppard, L., Smith, R., Sutton, M., Hicks, K., Ashmore, M., Haines-Young, R., Harper-Simmonds, L., 2014. A review and application of the evidence for nitrogen impacts on ecosystem services. Ecosystem Services 7, 76-88.

Jones, M.L.M., Sowerby, A., Williams, D.L., Jones, R.E., 2008. Factors controlling soil development in sand dunes: evidence from a coastal dune soil chronosequence. Plant and Soil 307, 219-234.

Jones, M.L.M., Wallace, H.L., Norris, D., Brittain, S.A., Haria, S., Jones, R.E., Rhind, P.M., Reynolds, B.R., Emmett, B.A., 2004. Changes in Vegetation and Soil Characteristics in Coastal Sand Dunes along a Gradient of Atmospheric Nitrogen Deposition. Plant Biology 6, 598-605.

King, R.S., Baker, M.E., 2010. Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. Journal of the North American Benthological Society 29, 998-1008.

Kuylenstierna, J.C., Henning, R., Cinderby, S., Hicks, K., 2001. Acidification in Developing Countries: Ecosystem Sensitivity and the Critical Load Approach on a Global Scale. Ambio 30, 20-28.

Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129, 271-280.

Lewis, R.J., Marrs, R.H., Pakeman, R.J., Milligan, G., Lennon, J.J., 2016. Climate drives temporal replacement and nested-resultant richness patterns of Scottish coastal vegetation. Ecography 39, 754-762.

Liu, X., Duan, L., Mo, J., Du, E., Shen, J., Lu, X., Zhang, Y., Zhou, X., He, C., Zhang, F., 2011. Nitrogen deposition and its ecological impact in China: An overview. Environmental Pollution 159, 2251-2264.

Maskell, L.C., Smart, S.M., Bullock, J.M., Thompson, K., Stevens, C.J., 2010. Nitrogen deposition causes widespread loss of species richness in British habitats. Global Change Biology 16, 671-679.

Mitchell, R.J., Hewison, R.L., Britton, A.J., Brooker, R.W., Cummins, R.P., Fielding, D.A., Fisher, J.M., Gilbert, D.J., Hester, A.J., Hurskainen, S., Pakeman, R.J., Potts, J.M., Riach, D., 2017. Forty years of change in Scottish grassland vegetation: Increased richness, decreased diversity and increased dominance. Biological Conservation 212, 327-336.

Nijssen, M.E., WallisDeVries, M.F., Siepel, H., 2017. Pathways for the effects of increased nitrogen deposition on fauna. Biological Conservation 212, 423-431.

Nilsson, J., Grennfelt, P., 1988. Critical loads for sulphur and nitrogen. UNECE/Nordic Council workshop report, Skokloster, Sweden. March 1988. Nordic Council of Ministers, Copenhagen.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J., Suggests, M., 2007. The vegan package. Community ecology package 10.

Pakeman, R.J., Alexander, J., Beaton, J., Brooker, R., Cummins, R., Eastwood, A., Fielding, D., Fisher, J., Gore, S., Hewison, R., Hooper, R., Lennon, J., Mitchell, R., Moore, E., Nolan, A., Orford, K., Pemberton, C., Riach, D., Sim, D., Stockan, J., Trinder, C., Lewis, R., 2015. Species composition of coastal dune vegetation in Scotland has proved resistant to climate change over a third of a century. Global Change Biology 21, 3738-3747.

Pakeman, R.J., Alexander, J., Brooker, R., Cummins, R., Fielding, D., Gore, S., Hewison, R., Mitchell, R., Moore, E., Orford, K., Pemberton, C., Trinder, C., Lewis, R., 2016. Long-term impacts of nitrogen deposition on coastal plant communities. Environmental Pollution 212, 337-347.

Pakeman, R.J., Hewison, R.L., Lewis, R.J., 2017. Drivers of species richness and compositional change in Scottish coastal vegetation. Applied Vegetation Science 20, 183-193.

Pardo, L.H., Fenn, M.E., Goodale, C.L., Geiser, L.H., Driscoll, C.T., Allen, E.B., Baron, J.S., Bobbink, R., Bowman, W.D., Clark, C.M., 2011. Effects of nitrogen deposition and empirical nitrogen critical loads for ecoregions of the United States. Ecological Applications 21, 3049-3082.

Payne, R.J., Campbell, C., Britton, A.J., Mitchell, R.J., Pakeman, R.J., Jones, L., Ross, L.C., Stevens, C.J., Field, C., Caporn, S.J.M., Carroll, J., Edmondson, J.L., Carnell, E.J., Tomlinson, S., Dore, A.J., Dise, N., Dragosits, U., 2019. What is the most ecologically-meaningful metric of nitrogen deposition? Environmental Pollution 247, 319-331.

Payne, R.J., Caporn, S.J., Field, C.D., Carroll, J.A., Edmondson, J.L., Britton, A., Dise, N.B., 2014. Heather Moorland Vegetation and Air Pollution: A Comparison and Synthesis of Three National Gradient Studies. Water, Air, & Soil Pollution 225, 1-13.

Payne, R.J., Dise, N.B., Stevens, C.J., Gowing, D.J., Duprè, C., Dorland, E., Gaudnik, C., Bleeker, A., Diekmann, M., Alard, D., 2013. Impact of nitrogen deposition at the species level. Proceedings of the National Academy of Sciences 110, 984-987.

Payne, R.J., Stevens, C.J., Dise, N.B., Gowing, D.J., Pilkington, M.G., Phoenix, G.K., Emmett, B.A., Ashmore, M.R., 2011. Impacts of atmospheric pollution on the plant communities of British acid grasslands. Environmental Pollution 159, 2602-2608.

Payne, R.J., Thompson, A.M., Standen, V., Field, C.D., Caporn, S.J.M., 2012b. Impact of simulated nitrogen pollution on heathland microfauna, mesofauna and plants. European Journal of Soil Biology 49, 73-79.

Pitcairn, C., Fowler, D., Leith, I., Sheppard, L., Sutton, M., Kennedy, V., Okello, E., 2003. Bioindicators of enhanced nitrogen deposition. Environmental Pollution 126, 353-361.

Podani, J., Csányi, B., 2010. Detecting indicator species: some extensions of the IndVal measure. Ecological Indicators 10, 1119-1124.

R Development Core Team, 2014. R: A language and environment for statistical computing. R foundation for Statistical Computing.

Reinds, G.J., Posch, M., de Vries, W., Slootweg, J., Hettelingh, J.P., 2008. Critical Loads of Sulphur and Nitrogen for Terrestrial Ecosystems in Europe and Northern Asia Using Different Soil Chemical Criteria. Water, Air, and Soil Pollution 193, 269.

Ross, L.C., Woodin, S.J., Hester, A.J., Thompson, D.B.A., Birks, H.J.B., 2012. Biotic homogenization of upland vegetation: patterns and drivers at multiple spatial scales over five decades. Journal of Vegetation Science 23, 755-770.

Rowe, E., Jones, L., Dise, N., Evans, C., Mills, G., Hall, J., Stevens, C., Mitchell, R., Field, C., Caporn, S., 2016. Metrics for evaluating the ecological benefits of decreased nitrogen deposition. Biological Conservation.

Smith, R., Fowler, D., Sutton, M., Flechard, C., Coyle, M., 2000. Regional estimation of pollutant gas dry deposition in the UK: model description, sensitivity analyses and outputs. Atmospheric Environment 34, 3757-3777.

Stevens, C.J., David, T.I., Storkey, J., 2018. Atmospheric nitrogen deposition in terrestrial ecosystems: Its impact on plant communities and consequences across trophic levels. Functional Ecology 32, 1757-1769.

Stevens, C.J., Dise, N.B., Gowing, D.J., Mountford, J.O., 2006. Loss of forb diversity in relation to nitrogen deposition in the UK: regional trends and potential controls. Global Change Biology 12, 1823-1833.

Stevens, C.J., Dise, N.B., Mountford, J.O., Gowing, D.J., 2004. Impact of nitrogen deposition on the species richness of grasslands. Science 303, 1876-1879.

Stevens, C.J., Duprè, C., Dorland, E., Gaudnik, C., Gowing, D.J., Bleeker, A., Diekmann, M., Alard, D., Bobbink, R., Fowler, D., 2010. Nitrogen deposition threatens species richness of grasslands across Europe. Environmental pollution 158, 2940-2945.

Sutton, M.A., Howard, C.M., Erisman, J.W., Billen, G., Bleeker, A., Grennfelt, P., van Grinsven, H., Grizzetti, B., 2011. The European nitrogen assessment: sources, effects and policy perspectives. Cambridge University Press.

Tipping, E., Davies, J.A.C., Henrys, P.A., Kirk, G.J.D., Lilly, A., Dragosits, U., Carnell, E.J., Dore, A.J., Sutton, M.A., Tomlinson, S.J., 2017. Long-term increases in soil carbon due to ecosystem fertilization by atmospheric nitrogen deposition demonstrated by regional-scale modelling and observations. Scientific Reports 7, 1890.

Tipping, E., Henrys, P., Maskell, L., Smart, S., 2013. Nitrogen deposition effects on plant species diversity; threshold loads from field data. Environmental pollution 179, 218-223.

van der Linde, S., Suz, L.M., Orme, C.D.L., Cox, F., Andreae, H., Asi, E., Atkinson, B., Benham, S., Carroll, C., Cools, N., De Vos, B., Dietrich, H.-P., Eichhorn, J., Gehrmann, J., Grebenc, T., Gweon, H.S., Hansen, K., Jacob, F., Kristöfel, F., Lech, P., Manninger, M., Martin, J., Meesenburg, H., Merilä, P., Nicolas, M., Pavlenda, P., Rautio, P., Schaub, M., Schröck, H.-W., Seidling, W., Šrámek, V., Thimonier, A., Thomsen, I.M., Titeux, H., Vanguelova, E., Verstraeten, A., Vesterdal, L., Waldner, P., Wijk, S., Zhang, Y., Žlindra, D., Bidartondo, M.I., 2018. Environment and host as large-scale controls of ectomycorrhizal fungi. Nature 558, 243-248.

Wilkins, K., Aherne, J., 2016. Vegetation community change in Atlantic oak woodlands along a nitrogen deposition gradient. Environmental Pollution 216, 115-124.

Wilkins, K., Aherne, J., Bleasdale, A., 2016. Vegetation community change points suggest that critical loads of nutrient nitrogen may be too high. Atmospheric Environment 146, 324-331.

FIGURES AND TABLES

Table 1. Details of the five datasets ultimately utilised in this study showing dataset references, NVC communities considered, lower limit of the critical load range, dataset codes used in this study, number of data points, number of species, full current N deposition range and key details of survey design. For full details of survey methodologies and habitat definitions, refer to cited publications.

Dataset	**Code**	**N**	**Species**	**Current N dep range (kg ha^{-1} yr^{-1})**	**Details**
Scottish Coastal - Wet grasslands (Pakeman et al., 2015)	SC.WGRASS	57	224	2.9-9.0	A minimum of five, 5m x 5 m quadrats were recorded for each site. Vascular plant cover was estimated by species and lichen and bryophyte cover was estimated collectively. NVC communities: MG8, MG11, MG13, OV28, OV29
Stevens - Acid grasslands (Stevens et al., 2006; Stevens et al., 2004)	CS.AGRASS	64	181	7.7-40.9	Acid grasslands were sampled to span the N deposition gradient. Five sampling points were randomly selected within a 100 m x 100 m area. At each point a 2 mx 2 m quadrat was surveyed and species cover estimated. NVC communities: U4
McVean - Alpine (Ross et al., 2012)	MCV.ALP	91	191	4.9-19.4	Surveys on Domin scale in 1 m x 1 m or 2 m x 2 m quadrats (the latter most frequent), recording all species including bryophytes and lichens. Surveys based on percentage cover. NVC communities: H13, H14, H17, H19, H20, U7, U8, U10, U12
Scottish Coastal - Dune slacks (Lewis	SC.SLAC	65	246	2.7-11.8	A minimum of five, 5 m x 5 m quadrats were recorded for each site. Vascular plant cover was estimated by species and

et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)					lichen and bryophyte cover was estimated collectively. NVC communities: SD13, SD16, SD17
Scottish Coastal - Fixed dunes (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.FDU	121	310	2.7-11.8	A minimum of five, 5 m x 5 m quadrats were recorded for each site. Vascular plant cover was estimated by species and lichen and bryophyte cover was estimated collectively. NVC communities: CG10, CG11, CG13, SD8, SD9, SD11, SD12

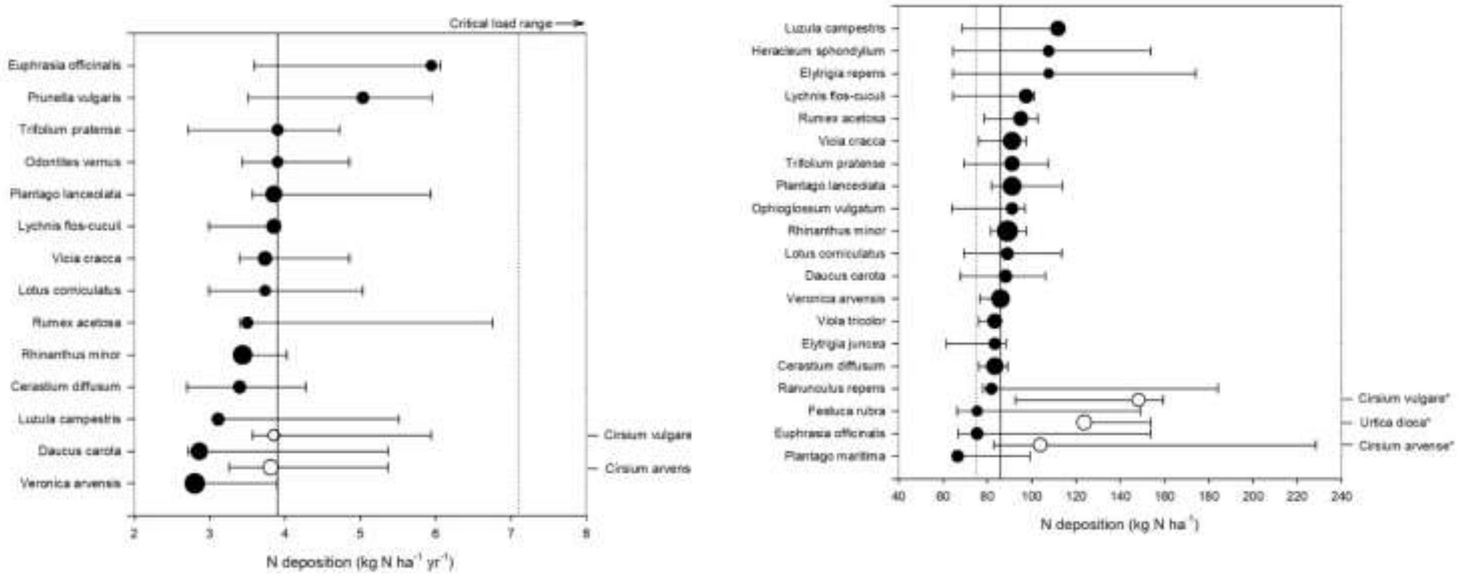
Table 2. Current empirical critical load (Bobbink & Hettelingh, 2011) and community-level change points in response to N deposition for the five datasets meeting inclusion criteria (see Table 1 for codes and details). Results show sum(z-) and sum(z+) change points along with 5th and 95th bootstrap percentiles. Separate results are presented for analyses based on current and cumulative (30 year moving window) N deposition, and for results based on all taxa and only on taxa where the influence of co-variates could be excluded. See text for full details. Co-variates used in taxon exclusion are listed in Supplementary Table 2 and identified taxa are highlighted with "*" in Figure 1.

Dataset		All taxa				Selected taxa excluded			
	Current CL, kg N ha^{-1} yr^{-1} (ecosystem response)	Current N deposition (kg N ha^{-1} yr^{-1})		Cumulative N deposition (kg N ha^{-1})		Current N deposition (kg N ha^{-1} yr^{-1})		Cumulative N deposition (kg N ha^{-1})	
		Sum (z-)	Sum (z+)	Sum (z-)	Sum (z+)	Sum (z-)	Sum (z+)	Sum (z-)	Sum (z+)
SC.WGRASS (wet grassland)	20-30 (↑ graminoids; ↓ diversity)	3.9 (3.5-4.9)	7.1 (3.0-7.2)	86 (80-98)	75 (68-163)	3.9 (3.5-4.9)	7.1 (3.0-7.2)	86 (82-97)	70 (68-229)
CS.AGRASS (acid grassland)	10-15 (↑ graminoids; ↓ diversity)	13.2 (9.0-13.7)	23.4 (13.2-30.6)	216 (205-514)	858 (399-884)	9.0 (8.7-12.8)	30.6 (11.9-31.3)	230 (196-313)	883 (221-931)
MCV.ALP (alpine)	5-15 (↓ moss & lichen cover)	12.3 (6.3-12.3)	11.2 (11.0-13.5)	233 (136-242)	238 (233-251)	6.3 (6.0-12.3)	12.3 (11.2-14.0)	241 (130-247)	238 (236-251)
SC.SLAC (slacks)	10-20 (↑ graminoid biomass)	4.8 (2.9-5.3)	7.5 (3.4-7.6)	90 (82-163)	245 (68-270)	5.0 (2.9-5.8)	7.5 (3.6-7.6)	90 (82-160)	245 (68-270)
SC.FDU (fixed dunes)	8-15 (↑ graminoids; ↓ diversity)	5.0 (3.9-6.8)	4.8 (3.9-5.8)	131 (98-168)	132 (102-247)	5.3 (3.9-6.8)	4.8 (3.9-5.9)	131 (87-171)	132 (98-282)

Figure 1

A) Wet Grassland

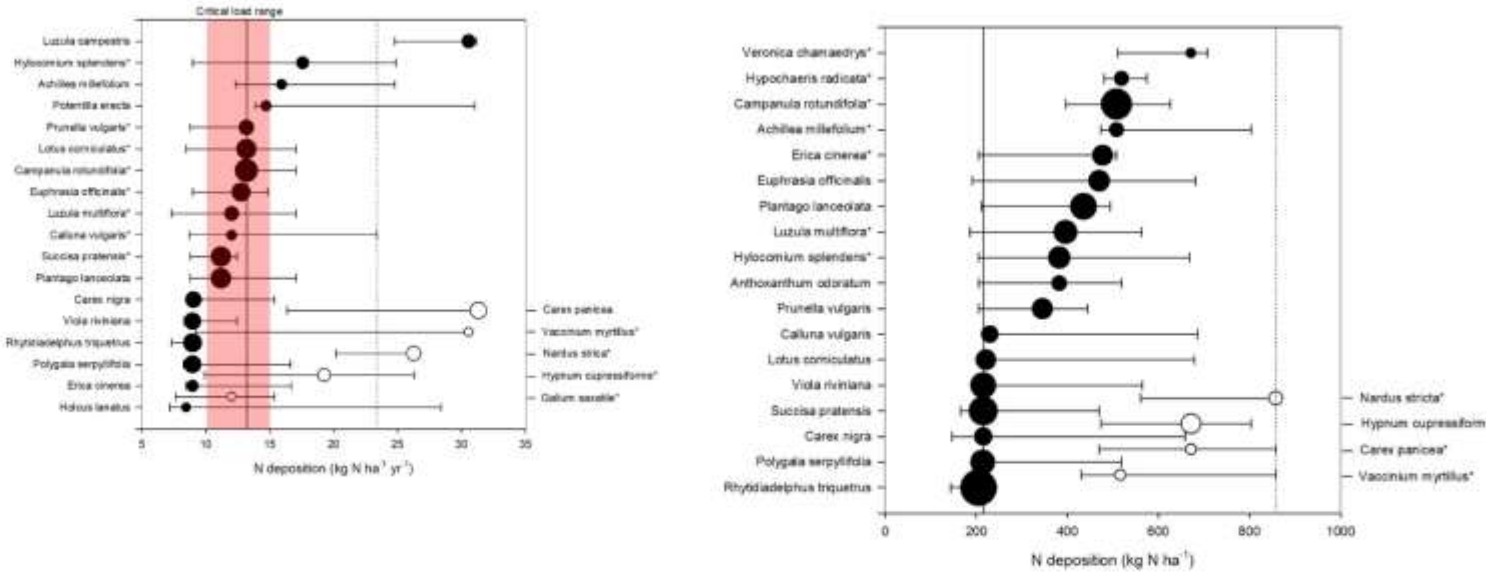
Current N **Cumulative N**



B) Acid Grassland

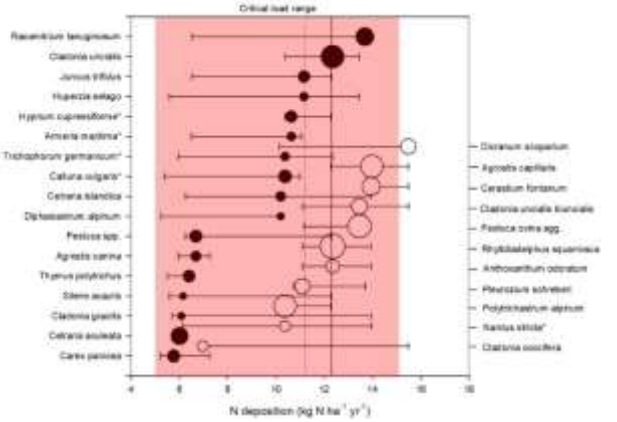
Current N

Cumulative N

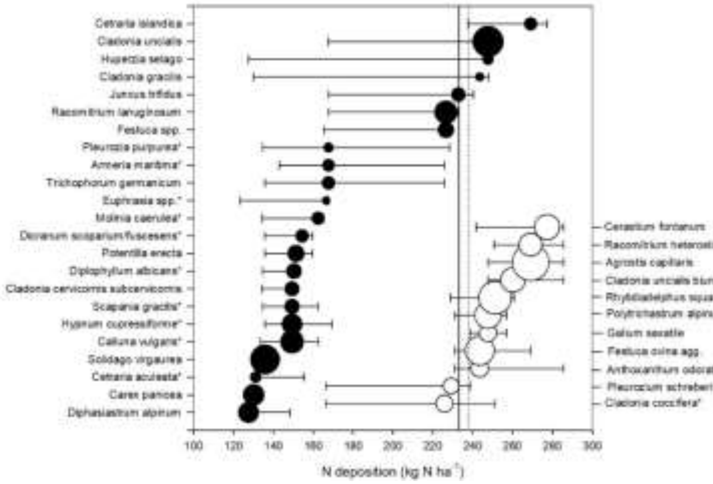


C) Alpine

Current N

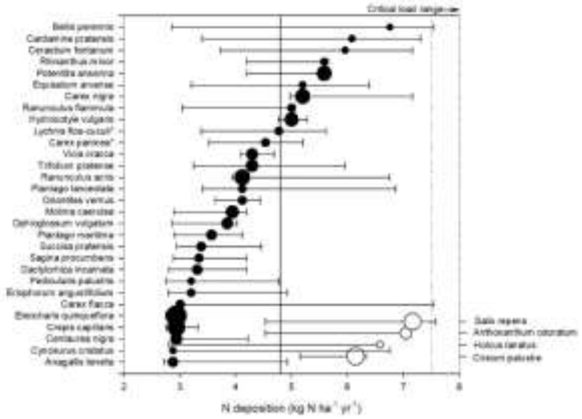


Cumulative N

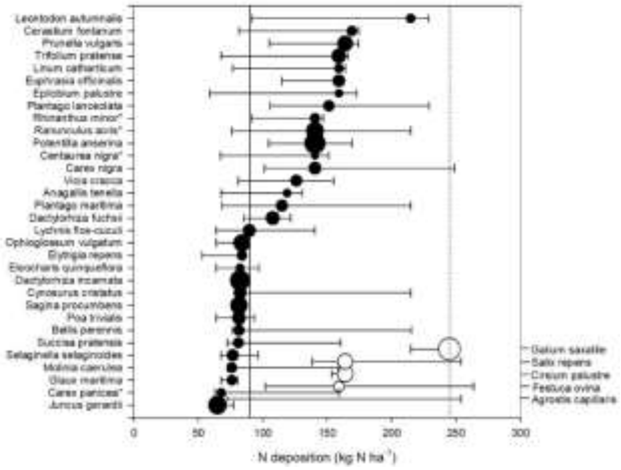


D) Dune slack

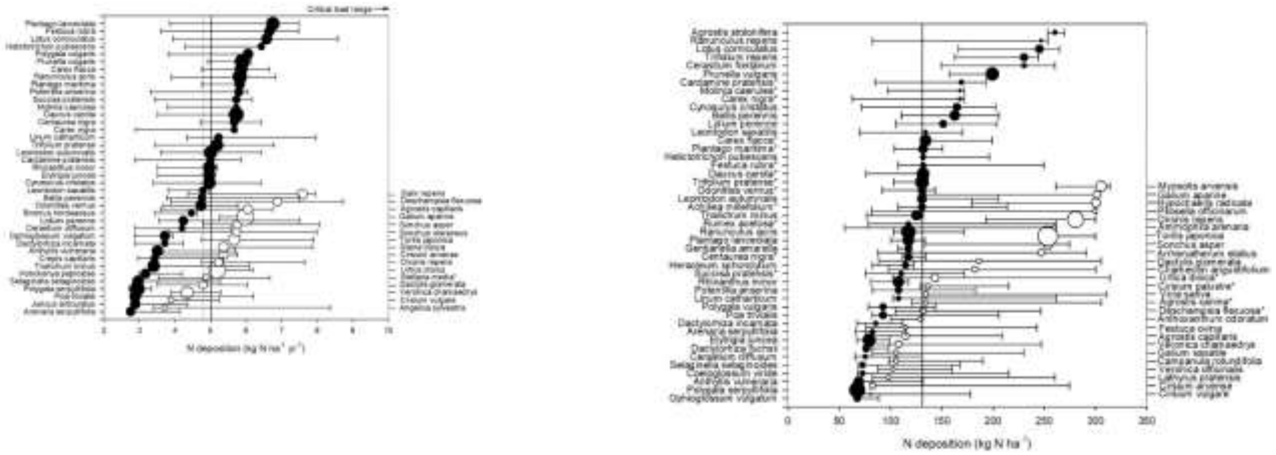
Current N



Cumulative N



E) Fixed Dune



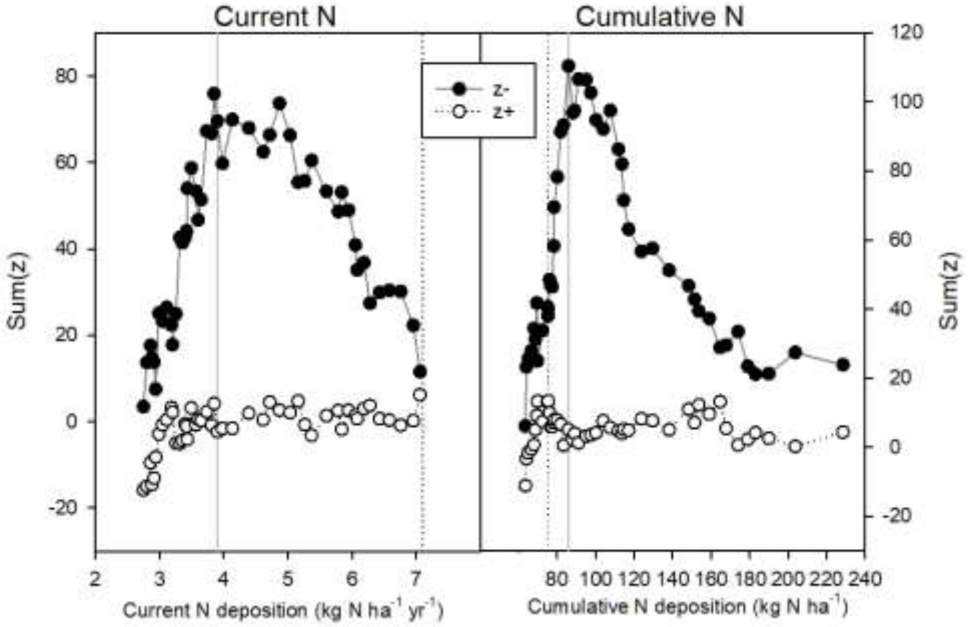
Current N

Cumulative N

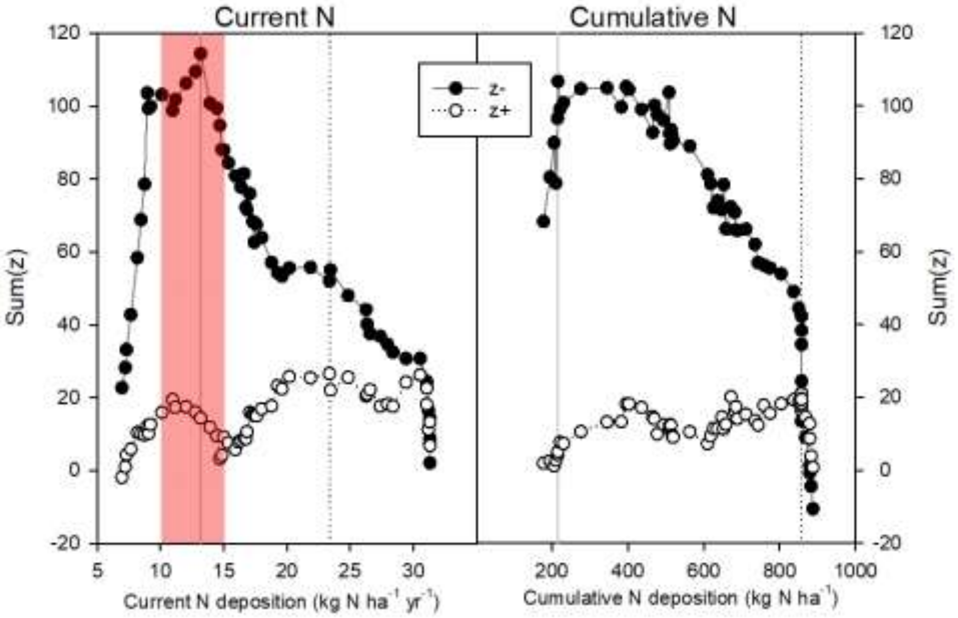
Figure 1. Species change-points for five vegetation datasets from UK semi-natural habitats. Separate results are presented for analyses based on current annual (left), and thirty year cumulative deposition (right). A) Wet Grassland (SC.WGRASS); B) Acid Grassland (CS.AGRASS); C) Alpine (MCV.ALP); D) Dune Slack (SC.SLAC); E) Fixed Dune (SC.FDU). See Table 1 for details. Plots show species showing high purity and reliability negative (black circles) and positive (white circles) change points in response to nitrogen deposition and boot-strap 5% and 95% quantiles. Vertical lines show overall community sum(z-) [solid line] and sum(z+) [dotted line] change-points. Red shaded bands show critical load ranges; where not shown the critical load lies at higher deposition levels out-with the plotted range. Species highlighted with "*" have change-points at equivalent positions on other environmental gradients.

Figure 2

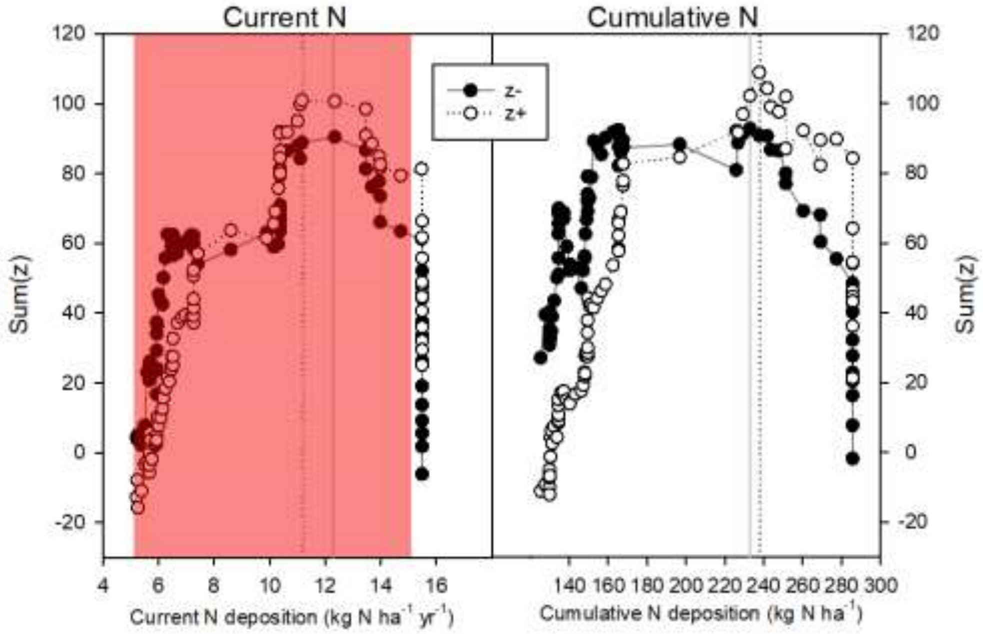
A) Wet Grassland



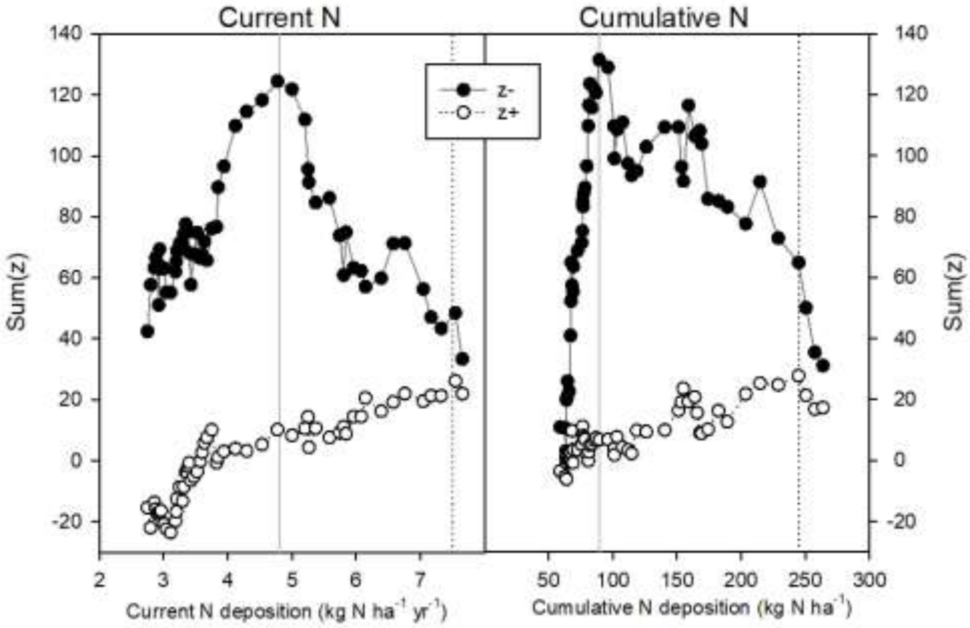
B) Acid Grassland



C) Alpine



D) Dune slacks



E) Fixed Dunes

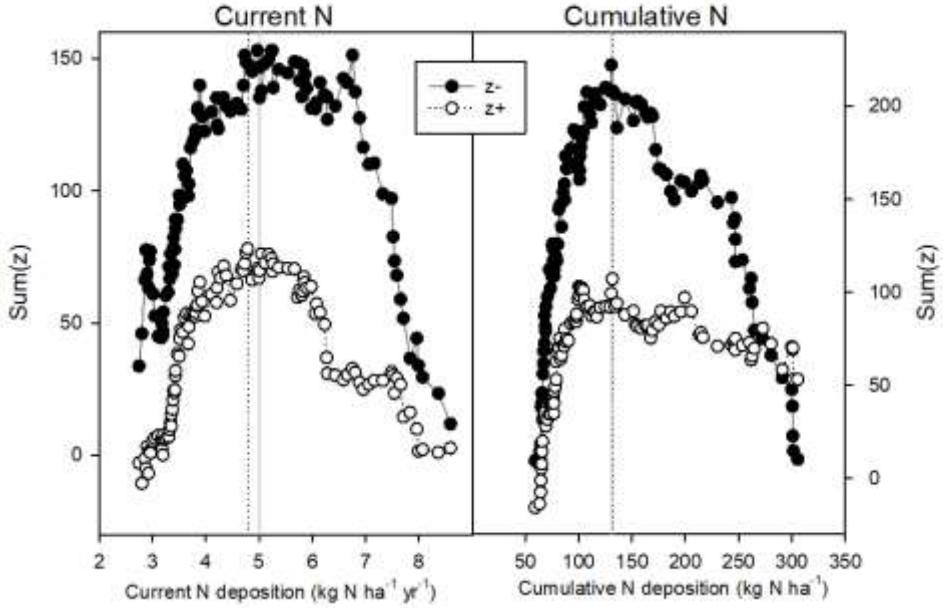


Figure 2. Sum(z) plots for five vegetation datasets from UK semi-natural habitats, see Table 1 for details. Plots show sum(z-) [filled circles] and sum(z+) [open circles] scores for all possible change-points along the N deposition gradient. Separate results are presented for analyses based on current annual (left), and thirty year cumulative deposition (right). Vertical lines show overall community sum(z-) [solid line] and sum(z+) [dotted line] change-points. Red shaded bands show critical load ranges; where not shown the critical load lies at higher deposition levels outwith the plotted range. Critical loads are based only on current deposition.

SUPPLEMENTARY MATERIAL

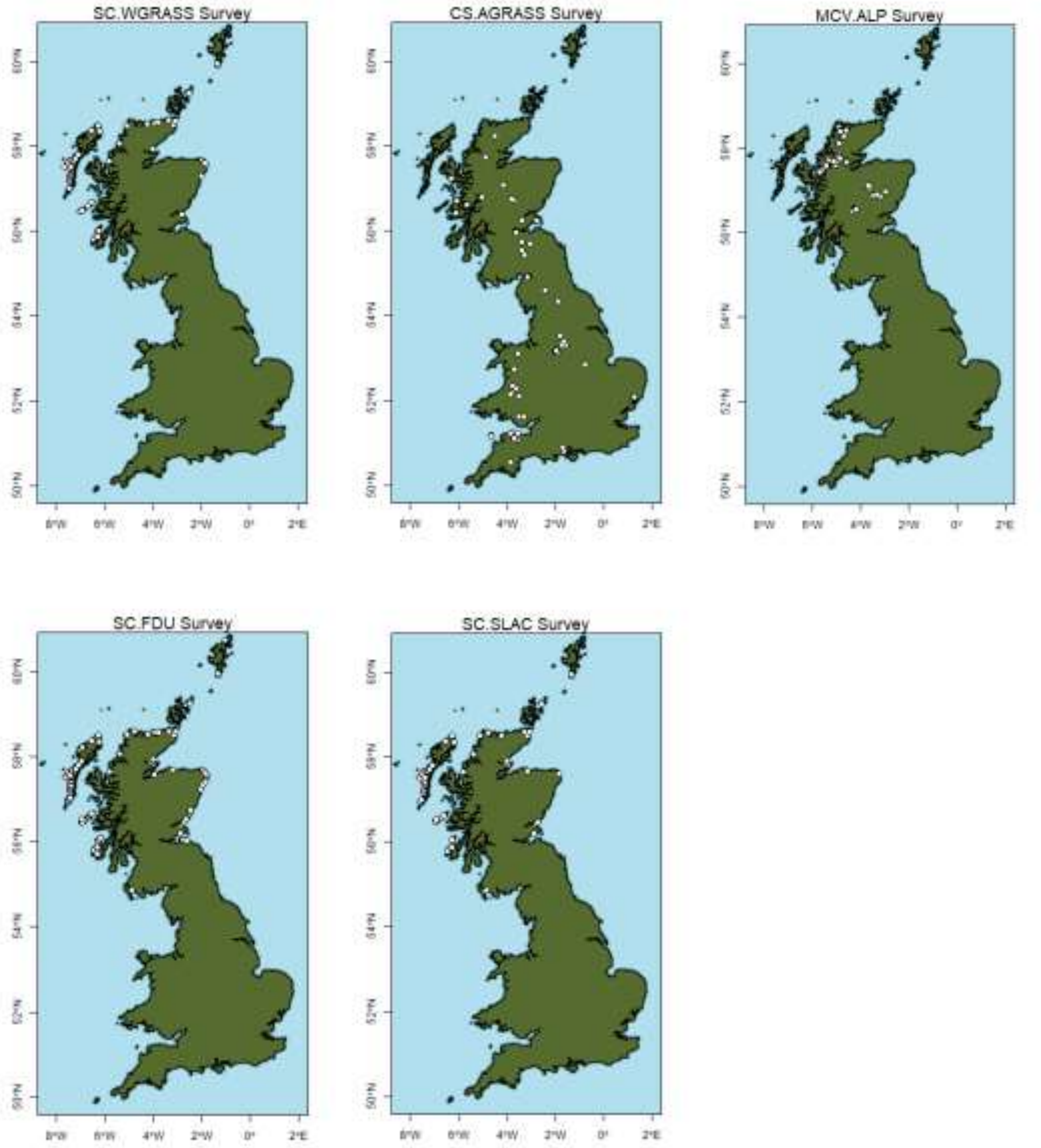
Supplementary Table 1. Details of all datasets considered in this study and rationale for inclusion/exclusion.

Dataset	**Code**	**N**	**Species**	**Current N dep range (kg ha^{-1} yr^{-1})**	**Exclusion stage.**
Birse - Calluna heaths (Britton et al., 2009; Britton et al., 2017a)	B.CHEATH	67	233	4.5-26.3	RDA significance testing P>0.01 (current N only).
Birse - *Vaccinium* heaths (Britton et al., 2009; Britton et al., 2017a)	B.VHEATH	33	152	7.9-26.3	RDA significance testing P>0.01 (current N only).
Edmondson- heather moorlands (Edmondson et al., 2013)	EDM	14	19	20.2-28.7	RDA significance testing P>0.01 (current N only).
McVean - moorlands (Ross et al., 2012)	MCV.MOOR	79	200	3.9-19.6	RDA significance testing P>0.01 (current N only).
Moorland Regional Survey- heaths (Caporn et al., 2014)	MRS	22	50	6.9-33.7	RDA significance testing P>0.01 (current N only).
Scottish Coastal - heathlands (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.HEATH	36	173	2.7-11.8	RDA significance testing P>0.01 (current N only).
Scottish Coastal - wet heathlands (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.WHEATH	38	174	2.9-10.7	RDA significance testing P>0.01 (both N variables).
Terrestrial Umbrella - lowland heaths (Field et al., 2014)	TU.LH	27	87	4.8-18.1	RDA significance testing P>0.01 (both N variables).
Terrestrial Umbrella - upland heaths (Field et al., 2014)	TU.UH	24	78	5.6-29.5	RDA significance testing P>0.01 (both N variables).
Birse - acid grasslands (Britton et al., 2009; Britton et al., 2017a; Britton et al., 2017b; Mitchell et al., 2017)	B.AGRASS	42	192	4.6-21.8	RDA significance testing P>0.01 (both N variables).
Birse - calcareous grasslands (Mitchell et al., 2017)	B.CGRASS	41	209	5.8-21.6	RDA significance testing P>0.01 (current N only).
Birse - Lolium grasslands (Mitchell et al., 2017)	B.LGRASS	46	96	4.6-19.0	RDA significance testing P>0.01 (both N variables).
Birse - mesotrophic grasslands (Mitchell et al., 2017)	B.MGRASS	73	178	4.0-23.3	RDA significance testing P>0.01 (both N variables).
Birse - wet grasslands (Mitchell et al., 2017)	B.WGRASS	56	248	3.3-31.1	RDA significance testing P>0.01 (current N only).
McVean – grassland (Ross et al., 2012)	MCV.GRASS	56	218	5.1-18.8	RDA significance testing P>0.01 (both N variables).
Scottish Coastal - acid grasslands (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.AGRASS	53	230	2.7-11.2	TITAN purity/reliability test (>30% low purity or reliability, both N variables).
Scottish Coastal - cliffs (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.CLIFF	38	175	2.8-10.7	RDA significance testing P>0.01 (both N variables).
Scottish Coastal - unimproved grasslands (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.UGRASS	76	296	2.7-9.0	RDA significance testing P>0.01 (both N variables).
Scottish Coastal - wet grasslands (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.WGRASS	57	224	2.9-9.0	**Included.**
Stevens - acid Grasslands (Stevens et al., 2006; Stevens et al., 2004)	CS.AGRASS	64	181	7.7-40.9	**Included.**
Birse - springs (Britton et al., 2009; Britton et al., 2017b)	B.SPRI	25	191	5.3-20.4	RDA significance testing P>0.01 (both N variables).

Birse - swamps (Britton et al., 2017b)	B.SWAM	33	160	3.6-20.9	RDA significance testing P>0.01 (both N variables).
McVean - wetlands (Ross et al., 2012)	MCV-WETL	28	170	5.1-15.8	RDA significance testing P>0.01 (both N variables).
Payne - bogs (Payne, unpublished)	PAYN	33	81	3.4-29.2	TITAN purity/reliability test (>30% low purity or reliability, current N only).
Scottish Coastal - tall grass mire (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.TGM	51	233	2.7-10.7	RDA significance testing P>0.01 (both N variables).
Terrestrial Umbrella - bogs (Field et al., 2014)	TU.BOG	29	97	4.8-26.7	RDA significance testing P>0.01 (current N only).
Armitage - *Racomitrium* heaths (Armitage et al., 2014)	ARM.RHE	26	58	8.9-47.9	RDA significance testing P>0.01 (current N only).
Birse - *Racomitrium* heaths (Britton et al., 2009)	B.RHE	77	214	5.8-31.2	TITAN purity/reliability test (>30% low purity or reliability, both N variables).
Britton - *Racomitrium* heaths (Britton et al., 2018)	BRI.RHE	15	66	6.0-34.7	RDA significance testing P>0.01 (both N variables).
McVean - alpine (Ross et al., 2012)	MCV.ALP	91	191	4.9-19.4	**Included.**
CEH dune grasslands (Aggenbach et al., 2017; Beaumont et al., 2014; Jones et al., 2008; Jones et al., 2004)	CEH.DUGR	34	345	3.4-13.1	RDA significance testing P>0.01 (both N variables).
CEH dune slacks (Aggenbach et al., 2017; Beaumont et al., 2014; Jones et al., 2008; Jones et al., 2004)	CEH.SLAC	29	362	2.8-11.4	RDA significance testing P>0.01 (both N variables).
Scottish Coastal - dune slacks (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.SLAC	65	246	2.7-11.8	**Included.**
Scottish Coastal - fixed dunes (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.FDU	121	310	2.7-11.8	**Included.**
Scottish Coastal - mobile dunes (Lewis et al., 2016; Pakeman et al., 2015; Pakeman et al., 2016; Pakeman et al., 2017)	SC.MDU	60	136	2.7-11.8	RDA significance testing P>0.01 (both N variables).
Terrestrial Umbrella - sand dunes (Field et al., 2014)	TU.SD	24	190	3.9-12.5	RDA significance testing P>0.01 (current N only).

Supplementary Table 2. Co-variates identified in RDA and used in change-point filtering. Co-variates are listed in order of selection in model-building.

Dataset	Co-variates identified	
	Current N deposition	Cumulative N deposition
SC.WGRASS	Altitude, mean annual precipitation, mean annual temperature	Altitude, mean annual precipitation, mean annual temperature
CS.AGRASS	S deposition, mean maximum temperature, mean annual precipitation, altitude, management index.	S deposition, mean maximum temperature, mean annual precipitation, altitude, management index.
MCV.ALP	Altitude, S deposition, aspect, mean annual precipitation, mean annual temperature, slope.	Altitude, S deposition, mean annual precipitation, aspect, mean annual temperature, slope.
SC.SLAC	Mean annual precipitation, mean annual temperature, S deposition.	Mean annual precipitation, mean annual temperature.
SC.FDU	Mean annual precipitation, mean annual temperature, altitude.	Mean annual precipitation, mean annual temperature.


SC.WGRASS Survey

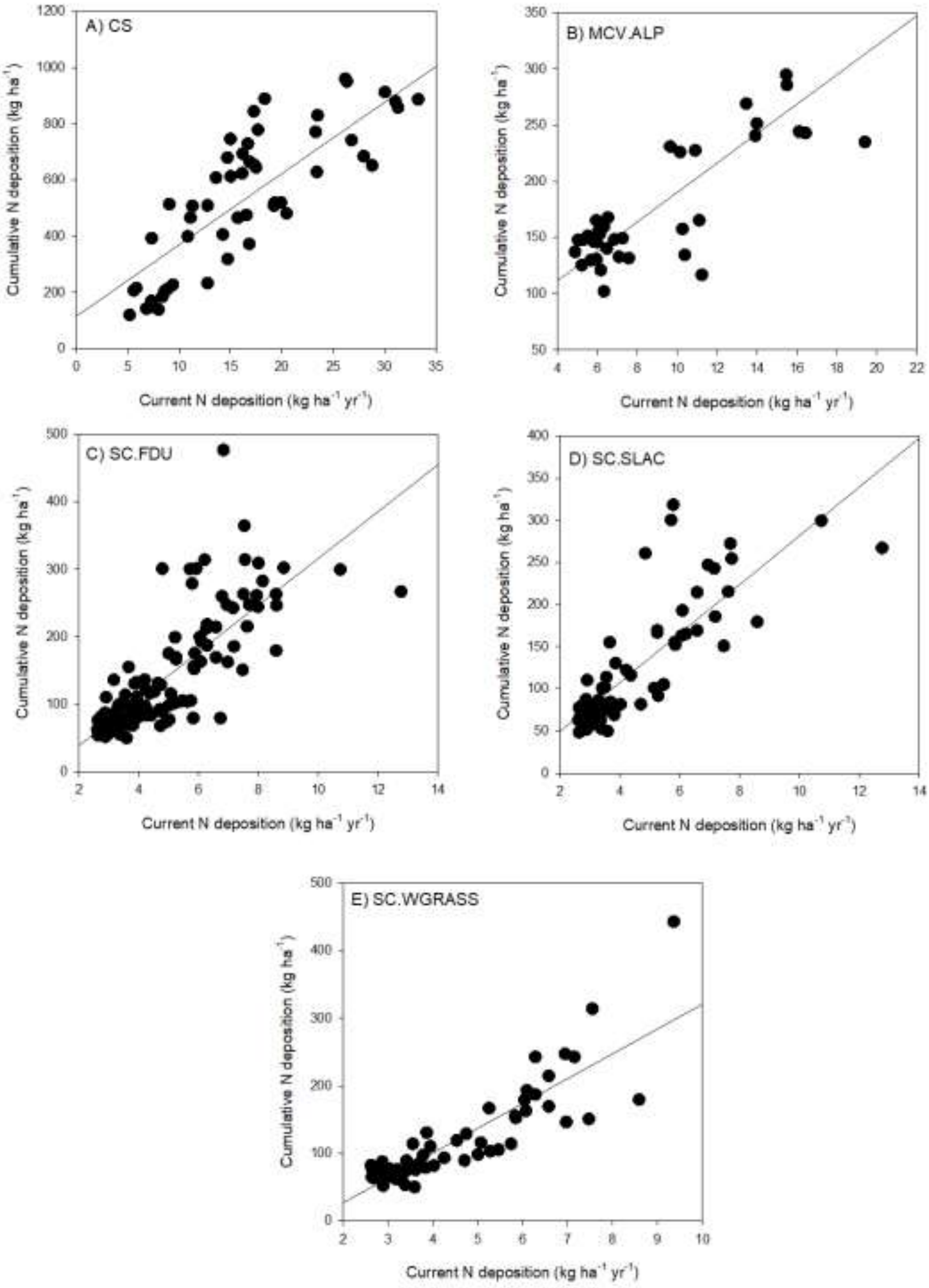
CS.AGRASS Survey

MCV.ALP Survey

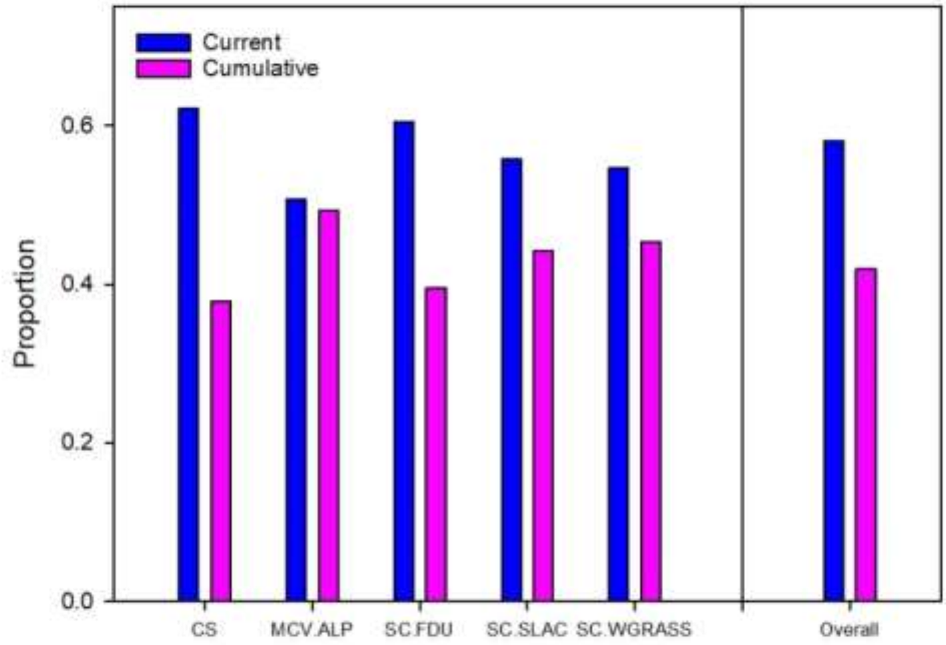
SC.FDU Survey

SC.SLAC Survey

Supplementary Figure 1. Locations of sites in the five focal datasets.



Supplementary Figure 2. Correlations between current and 30-year cumulative N deposition in the five focal datasets. Plots showing linear regressions used to produce Supplementary Figure 3.



Supplementary Figure 3 Proportion of change-points higher in either a TITAN analysis based on cumulative deposition or a TITAN analysis based on current deposition with values subsequently converted to cumulative deposition based on the overall correlation between cumulative and current deposition in each dataset (Supplementary Figure 2).