Reassessing the observational evidence for nitrogen deposition impacts in acid grassland: Spatial Bayesian linear models indicate small and ambiguous effects on species richness

Oliver L. Pescott [Corresp., 1] , **Mark Jitlal** [2]

[1] UK Centre for Ecology & Hydrology, Wallingford, Oxfordshire, United Kingdom

[2] Queen Mary University of London, Wolfson Institute of Preventative Medicine, London, UK

Corresponding Author: Oliver L. Pescott
Email address: olipes@ceh.ac.uk

Nitrogen deposition (Ndep) is considered a significant threat to plant diversity in grassland ecosystems around the world. The evidence supporting this conclusion comes from both observational and experimental research, with "space-for-time" substitution surveys of pollutant gradients a significant portion of the former. However, estimates of regression coefficients for Ndep impacts on species richness, derived with a focus on causal inference, are hard to locate in the observational literature. Some influential observational studies have presented estimates from univariate models, overlooking the effects of omitted variable bias, and/or have used *P*-value-based stepwise variable selection (PSVS) to infer impacts, a strategy known to be poorly suited to the accurate estimation of regression coefficients. Broad-scale spatial autocorrelation has also generally been unaccounted for. We re-examine two UK observational datasets that have previously been used to investigate the relationship between Ndep and plant species richness in acid grasslands, a much-researched habitat in this context. One of these studies (Stevens et al. 2004, *Science*, 303:1876–1879) estimated a large negative impact of Ndep on richness through the use of PSVS; the other reported smaller impacts (Maskell et al. 2010, *Glob. Change Biol.*, 16:671–679), but did not explicitly report regression coefficients or partial effects, making the actual size of the estimated Ndep impact difficult to assess. We reanalyse both datasets using a spatial Bayesian linear model estimated using integrated nested Laplace approximation (INLA). Contrary to previous results, we found similar-sized estimates of the Ndep impact on plant richness between studies, both with and without bryophytes, albeit with some disagreement over the most likely direction of this effect. Our analyses suggest that some previous estimates of Ndep impacts on richness from space-for-time substitution studies are likely to have been over-estimated, and that the evidence from observational studies could be fragile when confronted with alternative model

specifications, although further work is required to investigate potentially nonlinear responses. Given the growing literature on the use of observational data to estimate the impacts of pollutants on biodiversity, we suggest that a greater focus on clearly reporting important outcomes with associated uncertainty, the use of techniques to account for spatial autocorrelation, and a clearer focus on the aims of a study, whether explanatory or predictive, are all required.

1 Reassessing the observational evidence for nitrogen deposition impacts
2 in acid grassland: Spatial Bayesian linear models indicate small and
3 ambiguous effects on species richness

4 O.L. Pescott[1]* & M. Jitlal[2]
5 1. UK Centre for Ecology & Hydrology, Wallingford, Oxfordshire, United Kingdom
6 2. Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, United Kingdom
7 * Corresponding author, oliver.pescott@ceh.ac.uk

Abstract

Nitrogen deposition (Ndep) is considered a significant threat to plant diversity in grassland ecosystems around the world. The evidence supporting this conclusion comes from both observational and experimental research, with "space-for-time" substitution surveys of pollutant gradients a significant portion of the former. However, estimates of regression coefficients for Ndep impacts on species richness, derived with a focus on causal inference, are hard to locate in the observational literature. Some influential observational studies have presented estimates from univariate models, overlooking the effects of omitted variable bias, and/or have used *P*-value-based stepwise variable selection (PSVS) to infer impacts, a strategy known to be poorly suited to the accurate estimation of regression coefficients. Broad-scale spatial autocorrelation has also generally been unaccounted for. We re-examine two UK observational datasets that have previously been used to investigate the relationship between Ndep and plant species richness in acid grasslands, a much-researched habitat in this context. One of these studies (Stevens et al. 2004, *Science*, 303:1876–1879) estimated a large negative impact of Ndep on richness through the use of PSVS; the other reported smaller impacts (Maskell et al. 2010, *Glob. Change Biol.*, 16:671–679), but did not explicitly report regression coefficients or partial effects, making the actual size of the estimated Ndep impact difficult to assess. We reanalyse both datasets using a spatial Bayesian linear model estimated using integrated nested Laplace approximation (INLA). Contrary to previous results, we found similar-sized estimates of the Ndep impact on plant richness between studies, both with and without bryophytes, albeit with some disagreement over the most likely direction of this effect. Our analyses suggest that some previous estimates of Ndep impacts on richness from space-for-time substitution studies are likely to have been over-estimated, and that the evidence from observational studies could be fragile when confronted with alternative model specifications, although further work is required to investigate potentially nonlinear responses. Given the growing literature on the use of observational data to estimate the impacts of pollutants on biodiversity, we suggest that a greater focus on clearly reporting important outcomes with associated uncertainty, the use of techniques to account for spatial autocorrelation, and a clearer focus on the aims of a study, whether explanatory or predictive, are all required.

Introduction

Nitrogen deposition (Ndep) is a significant threat to the plant diversity of various habitat types, both in north-western Europe (UK National Ecosystem Assessment, 2011) and around the world (Phoenix et al., 2006). The evidence for this position comes from a variety of sources, including correlative analyses of observational data (e.g. Maskell et al., 2010), typically conducted across relatively large areas as "space-for-time" gradient studies, and small scale experiments (e.g. Van der Eerden et al., 1991), although the latter have also frequently been pooled across larger areas through meta-analyses or other approaches to evidence synthesis (Clark et al., 2007; Phoenix et al., 2012; Soons et al., 2017). Reviews of Ndep impacts on plant biodiversity have typically drawn on all of this evidence (Bobbink et al., 2010; UK National Ecosystem Assessment, 2011; Stevens et al., 2011c; RoTAP, 2012; Rowe et al., 2017), in addition to other types of studies, such as Before-After surveys of historic plots (e.g. Britton et al., 2009). Observational and experimental studies are therefore both generally considered useful ways of understanding pollutant-driven biodiversity change in terrestrial ecosystems.

Different inferential approaches are often considered complementary, with large-scale, observational methods potentially allowing access to "treatment" effects across pre-existing gradients, with levels of replication that would likely be challenging to resource via an experimental route (but see Fraser et al., 2013). One cost of this approach is that the effect of interest is likely to be crossed in various complex ways with numerous other variables, including historic drivers for which data are likely to be inaccessible, leaving one with a large choice of covariates that could potentially be included in a model, including some which will be unknown, or suspected to be of importance but impossible to access. Furthermore, spatially autocorrelated variables that are not captured by the covariates included, or other processes causing spatial structure such as dispersal, may also need to be accounted for to ensure accurate estimation of regression coefficients (Beale et al., 2010; Crase et al., 2014). The ultimate purpose of a statistical model must also be taken into account in making analytical decisions: does one primarily wish to make predictions, or is the focus on unbiased effect estimation to develop causal understanding (Mac Nally, 2000; Stephens, Buskirk & del Rio, 2007; Shmueli, 2010)? Although causal inference may imply predictive success, models that are constructed using methods that solely seek to maximise predictive accuracy will not necessarily capture causal processes accurately. Even when explicitly aiming for causal explanations through regression modelling in non-experimental situations, estimated coefficients may still only have a weak claim to be viewed as causal effects (Gelman & Hill, 2007; Young, 2018). Statistical issues then, in addition to domain-specific understanding, must also be at the forefront when attempting to make statements about cause and effect from observational data (Rubin, 2004; Young, 2018).

The work on Ndep presented here arose from a desire to use information from existing studies to inform the analysis of new data through the use of informative priors in a Bayesian framework (Lemoine, 2019). However, causality-focused estimates of regression coefficients from observational studies of Ndep impacts on plant richness proved hard to find in the current literature. For example, several studies have presented "headline" estimates from univariate models after multiple regression modelling (Stevens et al., 2004; Maskell et al., 2010; Field et al., 2014), implying regression coefficient values for Ndep that do not necessarily have any causal meaning, and/or have used P-value-based stepwise variable selection (PSVS) to derive final models, a strategy long known to be poorly suited to the accurate estimation of regression coefficients for inferential purposes (Greenland & Neutra, 1980; Mac Nally, 2000; Whittingham et al., 2006; Faraway, 2014; Harrell, 2015; Heinze, Wallisch & Dunkler, 2018). Re-analysis

79   of existing studies with the primary aim of developing models that focus on causal inference should
80   therefore be valuable in exploring the dependence of earlier conclusions on modelling choices.
81   Fortunately, data focusing on UK acid grasslands from two such studies (Stevens et al., 2004; Maskell et
82   al., 2010) were available for re-analysis; but note also that very similar work has been done across other
83   habitats (e.g. Maskell et al., 2010; Field et al., 2014), and across larger areas (e.g. Dupré et al., 2010).

84   The main focus of the current work is on deriving estimates of the effect sizes of nitrogen deposition on
85   plant species richness, rather than on their statistical significance (Amrhein, Greenland & McShane,
86   2019), whilst accounting for previously unmodelled broad-scale spatial autocorrelation. The model
87   forms investigated were specified *a priori*; in general these were delimited by the full sets of covariates
88   previously investigated by the original studies, given that these all have good ecological reasons for
89   inclusion. We do not make the (unprovable) claim that our "full" models are fully correct with respect to
90   reality, merely that the inclusion of as many plausible "pre-treatment" covariates as possible is likely to
91   help avoid omitted variable bias, given that the ignorability assumption (i.e. all confounders are
92   measured) is more likely to be satisfied (Gelman & Hill, 2007; Young, 2018), and that even our worst
93   case ratio of response data points to predictors meets rules of thumb put forward by statisticians to help
94   ensure that coefficients can be estimated accurately (Harrell, 2015, p. 72; Heinze, Wallisch & Dunkler,
95   2018). In addition, we also include a model for the Stevens et al. (2004) data that focuses on a reduced
96   set of covariates chosen for their expected similarity to the ecological impacts of the covariates used by
97   Maskell et al. (2010) for comparative purposes. Two additional models for the Maskell et al. (2010) data
98   were specified *post hoc*, due to a desire to investigate differences between our results and those of the
99   original paper.

100   Note that some covariates of interest (e.g. pH, Al, C:N, and soil %N) could themselves be influenced by
101   Ndep, but will also have independent impacts on the dependent variable (species richness). These types
102   of partly intermediate variables are sometimes distinguished from fully post-treatment variables as
103   "proxy variables" (Angrist & Pischke, 2009, pp. 64–68) whose inclusion is often better than their
104   omission when a causal interpretation is desired. We assume here that, due to their pre-treatment
105   importance for richness, adjusting for these covariates is more likely to result in accurate estimates of
106   the effect of Ndep that not adjusting for them (Angrist & Pischke, 2009); indeed, in some cases, their
107   inclusion will be essential for avoiding bias (Rosenbaum, 1984). Ultimately, we assume that these
108   potential post-treatment (or proxy) variables are "plausible surrogate[s] for … clearly relevant but
109   unobserved pretreatment variable[s]" (Rosenbaum, 1984).

Materials and Methods

Datasets

112   These are discussed chronologically by date of the field survey that created the original dataset, rather
113   than the date of publication of the analysis. The data analysed by Maskell et al. (2010), "MEA10"
114   hereafter, are described in that paper, and were originally collected as a part of the 1998 UK
115   Countryside Survey (UKCS; http://www.countrysidesurvey.org.uk). Briefly, the data analysed by MEA10
116   were selected from the 1998 UKCS on the basis of matches between plant communities in 2 x 2 m plots
117   and particular National Vegetation Classification syntaxa; for acid grasslands, the chosen plots had a best
118   fit to the acid grassland types U1-9 (Rodwell, 1992). For more details on the UKCS sampling strategy see
119   Maskell et al. (2010) and Smart et al. (2003a); we only note here that the UKCS is a stratified sample of
120   so-called UK land classes (Firbank et al., 2003), with systematic random sampling within strata. We take

121   the opportunity here to clarify some points relating to the datasets used in MEA10: the number of acid
122   grassland plots used in analyses was 883, not 895 as reported; total nitrogen deposition from the
123   modelled dataset of Smith et al. (2000) was specifically the estimated deposition over moorland, where
124   moorland was defined according to the Land Cover Map 1990 (Fuller et al., 1993), and so is generally
125   considered a better match to acid grassland than the grassland category of Smith et al. (2000); and,
126   finally, the sulphur dioxide deposition covariate was not absolute deposition for a particular year, but
127   the difference between the modelled values over moorland for 1998 and the modelled peak in 1970 at a
128   5 x 5 km resolution (all L.C. Maskell & S.M. Smart, pers. comms). This approach attempts to ensure that
129   the covariate effectively measures the recent and substantial reduction in acidifying sulphur deposition
130   across Britain.

131   Descriptions of the data collected by Stevens et al. (2004), "SEA04" hereafter, can be found in that
132   paper, and several others (e.g. Stevens et al., 2011b). The dataset reanalysed here was archived by
133   Stevens, Dupré, Dorland et al. (2011a; Ecological Archives deposit E092-128). Briefly, SEA04 surveyed a
134   random sample of 68 sites (after applying size and accessibility filters) from a larger database of acid
135   grasslands collated by British national conservation agencies (note that the grasslands in this database
136   may not be representative of the total national habitat resource). The random sample was stratified
137   across a nitrogen deposition gradient, the gradient being again the deposition model of Smith et al.
138   (2000), as for MEA10. Within sites, five 2 x 2 m plots were recorded (i.e. $n = 68 \times 5 = 340$) within a larger
139   100 x 100 m area chosen to contain at least 50% of NVC type U4 (*Festuca ovina - Agrostis capillaris -*
140   *Galium saxatile*; Rodwell, 1992) acid grassland (Stevens et al. 2004). Note, however, that Stevens et al.
141   (2011a) archived data for 320 of the plots from Stevens et al. (2004), and it is this dataset that we
142   reanalyse here.

Data preparation
144   Within each dataset, some covariates were re-scaled to allow for more direct comparisons between
145   regression coefficients, and to allow for their more intuitive ecological interpretation (Gelman & Hill,
146   2007). For example, altitude, scaled in metres in the original datasets, were divided by 100 to produce
147   regression coefficients that estimated the change in richness per 100 m. This produces a range of 9.75
148   (0.00–9.75), comparable to the other covariates, and a more ecologically-interpretable regression
149   coefficient (see Table 1 for all covariates used and their ranges). Although standardisation to unit
150   variance is also often recommended for improving the comparability of coefficients in regression
151   modelling (Gelman & Hill, 2007), it can also make direct comparisons between studies more difficult
152   (Baguley, 2009), and so we focus here on models estimated using our rescaled covariates (where
153   deemed necessary; see Table 1). MEA10 reported little difference in conclusions with respect to the
154   analysis of vascular plant richness only or vascular plant plus bryophyte richness. SEA04 included
155   bryophytes in their analyses. We focus primarily on vascular plant richness responses only (given that
156   they are more likely to be accurately estimated; for example, MEA10 note that UKCS surveys only
157   include "a selected range of the more easily identifiable bryophytes"), but we also report results from
158   analyses including bryophytes in Supplementary Information 2.

159   SEA04 considered a larger number of covariates than MEA10, therefore we present two reanalyses of
160   the SEA04 dataset here: model 1, using a smaller set of covariates chosen to match those of MEA10 in
161   terms of their likely ecological effects; and model 2, using a larger set of covariates, matching those
162   considered by SEA04 as closely as possible. Note that the full set of covariates considered by SEA04
163   contains some that are perfect linear combinations of each other (e.g. total acid deposition is given by

164   SEA04 as total N plus total S; likewise, total N is normally calculated as reduced N plus oxidised N), for
165   this reason, we only consider total N deposition and total S deposition as pollutant covariates in our
166   models (Table 1).

167   For SEA04 model 2, four of the available climate variables chosen to match the analysis of Stevens et al.
168   (2004) had very high pairwise linear correlations (all with r ≥ 0.78); these variables (mean annual
169   potential evapotranspiration (PET), mean annual daily maximum and minimum temperatures, 1996-
170   2006, all from the MARS dataset (see Stevens et al., 2011b); and mean annual potential
171   evapotranspiration from Tanguy et al. (2018)) were combined using PCA and the first two principle
172   components of this ordination used in their place (Table 1). PET data were unavailable for 3 sites (15
173   plots) in SEA04 (all on Lundy Island, England, 51° 10′ 57.82″ N, 4° 40′ 11.46″ W), and these values were
174   imputed as the mean PET value across the remainder of the SEA04 dataset prior to PCA. Imputation of
175   the missing values by predicting the missing values of PET using the other highly correlated climate
176   variables changed the imputed value of the missing data (all 15 plots had the same values for the
177   predictive climate variables), but did not result in any substantive change to the final regression
178   coefficients of the two climate principle components in the spatial Bayesian models described in the
179   next section, nor to those of other variables.

180   All other pairwise correlations in the SEA04 dataset were ≤ |0.57|, except for total Ndep and total Sdep,
181   which were also highly correlated (r = 0.83); a similar situation applied to the MEA10 data, where total
182   Ndep and Sdep change had a correlation of r = -0.70 (all others were ≤ |0.68|). However, in both cases
183   these pollutant variables were retained in our models due to their intrinsic interest to our causal
184   question. Collinearity should not bias regression coefficient estimates, but can lead to higher variance
185   (Harrell, 2015; Fox, 2016); variance inflation factors (VIFs) calculated for both datasets using standard
186   Poisson generalised linear mixed effects models (with a random intercept for each 1 kilometre square
187   containing plots; see below) indicated that all VIFs were below 5.4. The square-root of this VIF is 2.3,
188   below the range that Fox (2016, p. 343) indicates can present serious issues for the precision with which
189   regression coefficients are estimated.

Statistical models

191   As discussed by Blangiardo & Cameletti (2015), models for point-referenced data (i.e. those with
192   measurements of some outcome across a set of specific locations, where the locations are indexed by a
193   two (or three) dimensional vector) are one type of spatial data that can be modelled within a Bayesian
194   framework (such point-referenced data are also known as geostatistical data). Hierarchical approaches
195   to regression modelling, where unstructured random effects are incorporated within the model, are
196   often implemented using a Bayesian approach (Gelman & Hill, 2007; Blangiardo & Cameletti, 2015).
197   Such models can be extended to include structured random effects that allow analysts to account for
198   similarities based on temporal or spatial neighbourhoods. The integrated nested Laplace approximation
199   (INLA) method of approximate Bayesian inference is particularly well-suited to this area of modelling
200   given its speed and relative ease of implementation (Blangiardo & Cameletti, 2015).

201   We modelled both datasets using hierarchical Poisson regressions in R-INLA (Rue, Martino & Chopin,
202   2009; www.r-inla.org). All models included a set of covariates and an additional random spatial field, to
203   account for broad-scale spatial autocorrelation, as independent variables. The spatial field was
204   evaluated using the stochastic partial differential equation (SPDE) approach developed by Lindgren et al.
205   (2011), and was specified as a mesh constructed using a triangulation based on the 1 × 1 kilometre

206   squares of the British national grid (EPSG identifier 27700). Our models also included a random effect for
207   the 1 kilometre squares containing the plots (as per Maskell et al., 2010), because in both datasets there
208   were instances of multiple plots being recorded within a single square. Therefore the random spatial
209   field models broader scale spatial autocorrelation between squares, whilst the 1 km square random
210   effect captures smaller scale autocorrelation between plots within squares (Maskell et al., 2010). For the
211   MEA10 dataset, such models had considerably lower values of the Deviance Information Criterion (DIC)
212   than models omitting a random effect of 1 km square (ΔDIC = 54.6 or 31.8, without and with bryophyte
213   data respectively); for SEA04 models 1 and 2 the inclusion of the 1 km square random effect was less
214   important, but still generally improved the model (model 1: ΔDIC = 1.3 or -3.7; model 2: ΔDIC = 2.6 or
215   2.0; both without and with bryophyte data respectively). The 1 km square random effect was therefore
216   included for all models. The same check was made on the inclusion of the random spatial field based on
217   excluding this term from a model including the 1 km square random effect. The results indicated that
218   the spatial random field was required in five out of six models (MEA10: ΔDIC = 7.4 or 10.8, without and
219   with bryophyte data respectively); for SEA04: model 1: ΔDIC = 7.4 or -1.1; model 2: ΔDIC = 11.1 or 10.6;
220   both without and with bryophyte data respectively). The form of our final model can therefore be
221   described as follows:

222    $y_{ij} \sim Poisson(\lambda_{ij})$

223    $\lambda_{ij} = \exp(\beta_0 + \Sigma_{m=1}^{M}\beta_m X_{mij} + \beta_j Q_{j[i]} + \omega_{ij})$

224   Where y_{ij} is an observation of plant species richness in any plot/square combination, i is a unique plot
225   identifier, j is a unique 1 km square identifier, β_0 is a global intercept, β_m is the regression coefficient for
226   covariate X_m (m = 1,...,M) and β_j is the hierarchical coefficient for each 1 km square, with $\beta_j \sim N(0,\theta^2)$
227   and $Q_{j[i]}$ indicates the association of the ith observation with the jth square (Gelman & Hill, 2007). The
228   ω_{ij} term represents an additional spatial effect assumed to be a zero mean Matern Gaussian Markov
229   random field (GMRF) evaluated by the SPDE solution estimated over the specified triangulated mesh
230   (Figure A1.7). A detailed overview of geostatistical models estimated using SPDEs and GMRFs can be
231   found in Blangiardo & Cameletti (2015). Priors for all parameters were left at the INLA defaults (see
232   www.r-inla.org).
233

234   ## Results

235   The inclusion of both the 1 km square random effect and the spatial random field improved almost all
236   the models of the effects of N deposition on plant species richness in acid grassland explored here (as
237   judged by DIC, see "Statistical models" above). The one model (SEA04 model 1 with bryophyte data)
238   where ΔDIC was negative when either the 1 km square random effect or the spatial random field was
239   dropped (indicating that the simplified model was preferred) still required a spatial random effect at
240   some scale: dropping both random effects gave a DIC value of 1732.7 compared to a value of 1712.7 for
241   the full model; the preferred model in this case was that with the spatial random field only (DIC =
242   1709.0). Estimates of the overall random effects variance explained by the structured spatial field varied
243   between 87.4% (MEA10, vascular plants only) to 99.7% (SEA04 Model 1, vascular plant and bryophyte
244   richness combined) across the models including both random effects.

245   The addition of an extra kilogram of total N deposition on plant species richness per hectare per year
246   was consistently amongst the smallest effect estimated by our models, with low uncertainty in

247   comparison to other covariates (Figs 1a-c, A2.1, A2.2, A2.3). Exponentiating the fixed effects indicated
248   that the impact of increasing total N deposition is equivocal across these datasets and models (Table 2):
249   six models estimated a drop of around 1% in species richness as the most likely impact (Figs 1b-c, A2.2,
250   A2.3, A3.1, A3.2), although two of these models (Supplementary Information 3; SI3) were formulated
251   *post hoc* and were not favoured as the best models for the dataset by their DIC estimates. Two models,
252   estimated from MEA10 using the DIC-favoured model, suggested that an increase in species richness
253   had more support (Figs 1a, A2.1; Table 2). The reduced covariate set used for the SEA04 data (model 1)
254   for closer comparison to the covariate set used in the MEA10 analysis made no difference to the
255   estimated effect of N deposition calculated from the larger set of covariates from SEA04 (model 2; Figs
256   1b-c, A2.2, A2.3; Table 2). Figure 2 illustrates the estimated partial effect of N deposition (with other
257   covariates set to zero) on vascular plant species richness for two of our models (MEA10 and SEA04
258   model 1); these demonstrate both the relatively small predicted mean impact on species richness, the
259   disagreement in impact direction between the datasets, and the uncertainty attached to these
260   predictions.

261   The results presented in SI3 were included as exploratory, *post hoc*, analyses after it was found that our
262   DIC-favoured models of the MEA10 dataset indicated that the most likely overall influence of Ndep on
263   species richness was more likely to be positive, contrary to the conclusions of Maskell et al. (2010). Our
264   additional analyses (SI3), each with only one of the spatial random effects used in the full model, did
265   show a reversal of average sign, with the analysis matching that of Maskell et al. (2010), i.e. using a 1 km
266   square random effect, indicating the strongest negative effect of N deposition on vascular plant richness
267   (a mean of a 2% loss of species richness per kg ha^{-1} yr^{-1} of total N deposition; Table 2, Fig. A3.1).
268   However, this model had a DIC value 7.4 units higher than the favoured model.

Discussion

270   Previous work on the estimation of the effects of nitrogen deposition on plant species richness from
271   space-for-time substitution studies is likely to have had a significant impact on both science and policy
272   (assuming that the number of citations received can be used as an index to this). Therefore being clear
273   about the reliability and accuracy of estimates from observational work is important, not only for
274   fundamental reasons of improving ecological understanding and analysis, but also because policy
275   decisions must be made concerning the utility of funding work in this area relative to other ways of
276   assessing environmental change (although field experimental work may have issues of its own; Peters,
277   1991; Phoenix et al., 2012). Clear and accurate estimates of the likely sizes of effects are invaluable for
278   informing future studies, for example, through power analyses, or the use of informative priors for
279   Bayesian approaches. Our results suggest that only small effects of nitrogen deposition on species
280   richness may be detectable in these observational datasets.

281   The headline result of Stevens et al. (2004) reported "a reduction of one species per 4-m^2 quadrat for
282   every 2.5 kg N ha^{-1} yr^{-1} of chronic nitrogen deposition". An estimated reduction in species richness of
283   23% (based on a change from deposition of 5 kg N ha^{-1} yr^{-1} to 17 kg N ha^{-1} yr^{-1}) was also highlighted as a
284   key message using a regression coefficient estimated from a model including only Ndep as a covariate.
285   The estimate from our SEA04 model 1 implied a loss of around 1% of richness per kg N ha^{-1} yr^{-1};
286   however, this ignores the uncertainty associated with this particular coefficient estimate, which included
287   an estimate of no effect at the 97.5% percentile, and the larger uncertainty associated with alternative
288   model specifications (including ones that we have not explored here). Indeed, our estimates from the

289 larger, random-stratified national survey of Maskell et al. (2010) suggested that a small average increase
290 in richness along the total nitrogen deposition gradient was more plausible under the favoured model
291 for that dataset. Furthermore, the partial effects calculated from our full models (Fig. 2) indicated that
292 the average effect on richness across Britain may only approach a difference of around one species in
293 either direction. This is also notable because previous commentaries (e.g. Tipping et al. 2013) have
294 suggested that differences in the apparent effect sizes of Ndep between the two observational studies
295 reanalysed here were likely due to study design, whereas our work suggests that the differences are
296 actually minimal, and that those found previously may simply be artefacts of the different statistical
297 modelling procedures adopted by the original studies. The different median signs found in our analyses
298 could also be due to differences in the statistical populations targeted, namely the differential sampling
299 of grassland types, with the greater plant community coverage of the dataset of Maskell et al. (2010)
300 potentially covering more sites on the left of a net primary productivity/diversity unimodal curve,
301 resulting in an average increase in richness under fertilisation (see below for more discussion of the
302 potential impacts of non-linear responses and other issues on the current work).

303 Our results should cause others to re-evaluate their approaches to observational data. For example,
304 Field et al. (2014) also used PSVS, and other P-value based selection techniques, to formulate models in
305 their space-for-time analyses of Ndep impacts for a number of semi-natural habitats across Britain, as
306 well as highlighting simple univariate relationships through scatter plots. These authors, however, did
307 note in their methods section that their results should be "interpreted with caution", although this
308 statement of uncertainty was not clearly carried through to other parts of their paper, nor to other
309 research. For example, Payne et al. (2017) used data from both SEA04 and Field et al. (2014) to forecast
310 the impacts of Ndep on plant species richness under future nitrogen deposition scenarios. Ndep and its
311 polynomial transformations were the only covariates in these models. Payne et al. (2017) noted in their
312 online supplementary material (their Web Panel S1) that such forecasting makes the strong assumption
313 that covariances between variables remain the same at the future time point (i.e. the distribution of
314 nitrogen deposition amounts will vary with, e.g., mean maximum July temperature in the same way in
315 2030 as they did when originally modelled). However, they did not clearly acknowledge that the Ndep-
316 only models in their forecasting exercise had received no validation in support of their status as the best
317 predictive (or explanatory, for that matter) models by the original studies or elsewhere. Ndep may be a
318 useful covariate for producing good predictive models of species richness in certain habitats, but this
319 should be demonstrated using measures of out-of-sample predictive ability, and the resulting partial
320 effects of Ndep in such models have no intrinsic claim to be reliable estimates of its causal relationship
321 with richness (cf. Payne et al. 2017).

322 These issues from the literature should not detract from the fact that there are also several ways in
323 which the work presented here could potentially be improved upon. Our efforts are linear models (as for
324 SEA04, MEA10, Field et al. 2014, among others), but there is evidence that the response of species
325 richness to Ndep may be better modelled as unimodal (Tipping et al., 2013; Simkin et al., 2016; Clark et
326 al., 2019). Simkin et al. (2016) still, however, reported declines in the plant species richness of open
327 habitats above 8.7 kg N ha^{-1} yr^{-1}, very similar to the 7.9 kg N ha^{-1} yr^{-1} richness decline threshold identified
328 for acid grassland by Tipping et al. (2013), suggesting that a large part of the gradient studied by SEA04
329 and MEA10 may still be well-approximated by a linear relationship (although we note that the approach
330 of Tipping et al. 2013 also makes the assumption that there are no omitted linear or non-linear variables
331 correlated with Ndep that might change their estimated univariate breakpoint relationship). Indeed, the

332   linear component of a non-linear trend is often considered the most policy-relevant summary by those
333   who routinely produce ecological indicators (e.g. Soldaat et al., 2017). Different approaches to
334   accounting for this likely non-linearity in multivariable models, such as the inclusion of interactions or
335   smoothers, could however be further explored, particularly given that novel methods focused on causal
336   inference that can account for these issues continue to be developed (Dorie et al., 2019).

337   Other uncertainties relating to our conclusions pertain to the fact that the broad-scale spatial field used
338   here may be accounting for information that, if known, would change the size of the Ndep impact
339   regression coefficient. This could be in the form of additional covariates, or more highly resolved
340   estimates of the Ndep load that a location has actually received. The Ndep estimates used here (and by
341   the original studies) are resolved to a grain size of 5 x 5 km, and this additional uncertainty could have
342   attenuated our estimate of the regression coefficient (the measurement error in explanatory variables
343   problem; e.g. see Fox, 2016), even if there is no systematic bias relating to the association of 5 x 5 km-
344   estimated Ndep levels with particular types of vegetation. This argument is likely to apply to many of the
345   variables used here and in other studies, particularly given that many are estimated from other
346   modelling exercises at relatively large spatial grain sizes (Table 1). Conversely, measurement error
347   coupled with conditioning on statistical significance, as happens through PSVS (e.g. Stevens et al., 2004;
348   Field et al., 2014), is likely to result in the overestimation of effects (Gelman & Carlin, 2014; Loken &
349   Gelman, 2017). Whether or not the potential for these biases is more serious for inference than the
350   absence of covariates that are unavailable, such as historic land management events (e.g. Rackham,
351   1986), is difficult to say.

352   The new observational results presented here are in line with much of the experimental literature on
353   Ndep impacts. For example, Phoenix et al. (2012) reviewed a group of nine experiments conducted
354   across the UK, with N treatments which had been running for lengths of time between 7 and 22 years at
355   the time of review. These studies aimed to examine the impacts of "modest treatment doses and
356   avoid[ed] single dose or solid form applications" (Phoenix et al., 2012) in order to overcome previous
357   criticisms relating to (potentially unrealistic) very high experimental loadings. These nine studies
358   reported no effects of their experimental N treatments on higher plant richness (although these
359   conclusions of no effect were all based on thresholding *P*-values); two sites indicated increases in
360   richness using the calculated accumulated Ndep dose over the duration of an experiment, although
361   these cases were discounted by Phoenix et al. (2012) as being of either minor ecological significance or
362   transitory. Phoenix et al. (2012) put forward several reasons why the results from this series of
363   experiments may not reflect the true impact of chronic Ndep on plant richness: for example, sensitive
364   species might already have been lost prior to the establishment of an experiment, and/or an experiment
365   might not have been running long enough for the impacts to have been fully realised. Of course,
366   important changes in community structure can also occur without species loss (Hillebrand et al., 2018).

367   Longer running studies, such as the Park Grass experiment at Rothamsted, England (1856 to the present
368   day; Silvertown et al., 2006), avoid some of these criticisms. Although the Park Grass plots that have
369   received experimental N addition have received very high doses, with annual N fertiliser doses starting
370   at 48 kg N ha^{-1} yr^{-1} (Storkey et al., 2015), control plots having received only atmospheric deposition may
371   be a useful comparator for some habitats in the wider landscape. Plot 3 of the Park Grass experiment,
372   for example, has in theory only ever received ambient Ndep (Storkey et al., 2015); Lawes & Gilbert
373   (1880) reported an average of around 48.5 plant species (including bryophytes) in the 0.5 acre (~2000
374   m^2) neutral grassland plot between 1862–1877; the average between 1939–1948 was 34 (Brenchley &

375   Warington, 1958), indicating a decline over this period that preceded the local increase in Ndep (see
376   the first figure of Storkey et al. 2015 for the local Ndep trend at this location over the 20th century).
377   Unfortunately, the local increase in Ndep at Rothamsted coincided with the decision to split the plots
378   into different liming treatments (Williams, 1974; Storkey et al., 2015) meaning that subsequent richness
379   estimates cannot be unambiguously compared to the earlier numbers (there is confounding of the Ndep
380   increase with the application of lime, and the area that could be directly compared shrinks to a quarter
381   of the original plot, introducing the need for species-area adjustments). Over the period 1991–2012,
382   however, survey data indicate that there may have been a very slight recovery in richness in Plot 3
383   (although Storkey et al. present no statistics for this trend), coinciding with a reduction in the N
384   composition of the plot herbage and increases in Simpson's diversity index (Storkey et al., 2015).

385   Overall, then, experimental data using realistic applications of Ndep appear to support our finding that
386   richness is a relatively insensitive metric of such impacts (see also Hillebrand et al., 2018). Finally, and to
387   avoid any misunderstanding as to the thesis being presented in this paper, we note that we are *not*
388   stating that overall eutrophication from all sources (e.g. livestock, local fertiliser drift etc.) is
389   unimportant for the conservation of plant biodiversity. General signals of eutrophication in plant
390   communities are widespread and beyond doubt (e.g. Smart et al., 2003b; Willi, Mountford & Sparks,
391   2005). We are forced to conclude, however, that the contribution of Ndep to this phenomenon appears
392   to be smaller, and more uncertain, than many previous analyses of space-for-time Ndep gradients have
393   concluded (cf. Stevens et al., 2004; Field et al., 2014; Payne et al., 2017). It is possible that this is due to
394   the fact that earlier losses of species due to accumulated historic deposition means that the remaining
395   opportunity to detect effects in recent space-for-time studies is limited; this suggests, however, that
396   richness should not be used as an indicator of Ndep impacts, and previous work taking this approach
397   (particularly where highly significant statistical concerns abound) should no longer be cited in support of
398   general conclusions regarding these impacts.

399   ## Conclusions
400   The evidence for large negative impacts of nitrogen deposition on plant species richness put forward
401   through analyses of observational data appears to have been overstated. We estimate a possible decline
402   in richness of around 1% per kg ha^{-1} yr^{-1} of total N deposition from two spatially and temporally
403   separated British space-for-time gradient studies, considerably less than the estimates implied
404   previously by described and/or plotted relationships from primary studies (Stevens et al., 2004; Maskell
405   et al., 2010). Moreover, even this estimate appears uncertain, and our favoured model for the acid
406   grassland data of Maskell et al. (2010) suggests that an average increase in richness of a similar
407   magnitude may be more likely. The previous lack of presented regression coefficients derived with
408   causal inference as their main focus, and of models that account for broad-scale spatial autocorrelation,
409   is important to note, because scientists wishing to use the estimated effects of Ndep for the design of
410   future studies, or for the construction of informative priors in new analyses, may be misled as to the size
411   of effect that is expected to be detectable in datasets of this type. The various models presented here
412   could be thought of as a small section of the "multiverse" of potential approaches to these data, an
413   approach that has been put forward as an additional route to transparency and reproducibility in
414   science, and which can provide insights into the fragility or robustness of particular conclusions (Steegen
415   et al., 2016). As such this work is unlikely to be the last word on these datasets, or in the general area of
416   observational studies of nitrogen deposition impacts. We hope that our re-analyses inspire further
417   efforts to accurately extract the maximum available knowledge from these valuable datasets, whether

418   for explanatory or predictive purposes, and that evidence synthesis in this area takes these uncertainties
419   and methodological issues into greater account going forwards.

420

421 References

422 Amrhein V, Greenland S, McShane B. 2019. Scientists rise up against statistical significance. *Nature*
423   567:305. DOI: 10.1038/d41586-019-00857-9.

424 Angrist JD, Pischke J-S. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ:
425   Princeton University Press.

426 Baguley T. 2009. Standardized or simple effect size: What should be reported? *British Journal of*
427   *Psychology* 100:603–617.

428 Beale CM, Lennon JJ, Yearsley JM, Brewer MJ, Elston DA. 2010. Regression analysis of spatial data.
429   *Ecology Letters* 13:246–264. DOI: 10.1111/j.1461-0248.2009.01422.x.

430 Blangiardo M, Cameletti M. 2015. *Spatial and Spatio-temporal Bayesian Models with R - INLA*.
431   Chichester, West Sussex: John Wiley & Sons.

432 Bobbink R, Hicks K, Galloway J, Spranger T, Alkemade R, Ashmore M, Bustamante M, Cinderby S,
433   Davidson E, Dentener F, Emmett B, Erisman J-W, Fenn M, Gilliam F, Nordin A, Pardo L, De Vries
434   W. 2010. Global assessment of nitrogen deposition effects on terrestrial plant diversity: a
435   synthesis. *Ecological Applications* 20:30–59. DOI: 10.1890/08-1140.1.

436 Brenchley WE, Warington K. 1958. *The Park Grass Plots at Rothamsted 1856-1949*. Harpenden:
437   Rothamsted Experimental Station.

438 Britton AJ, Beale CM, Towers W, Hewison RL. 2009. Biodiversity gains and losses: Evidence for
439   homogenisation of Scottish alpine vegetation. *Biological Conservation* 142:1728–1739. DOI:
440   10.1016/j.biocon.2009.03.010.

441 Clark CM, Cleland EE, Collins SL, Fargione JE, Gough L, Gross KL, Pennings SC, Suding KN, Grace JB. 2007.
442   Environmental and plant community determinants of species loss following nitrogen
443   enrichment. *Ecology Letters* 10:596–607. DOI: 10.1111/j.1461-0248.2007.01053.x.

444 Clark CM, Simkin SM, Allen EB, Bowman WD, Belnap J, Brooks ML, Collins SL, Geiser LH, Gilliam FS, Jovan
445   SE, Pardo LH, Schulz BK, Stevens CJ, Suding KN, Throop HL, Waller DM. 2019. Potential
446   vulnerability of 348 herbaceous species to atmospheric deposition of nitrogen and sulfur in the
447   United States. *Nature Plants* 5:697–705. DOI: 10.1038/s41477-019-0442-8.

448 Crase B, Liedloff A, Vesk PA, Fukuda Y, Wintle BA. 2014. Incorporating spatial autocorrelation into
449   species distribution models alters forecasts of climate-mediated range shifts. *Global Change*
450   *Biology* 20:2566–2579. DOI: 10.1111/gcb.12598.

451 Dorie V, Hill J, Shalit U, Scott M, Cervone D. 2019. Automated versus Do-It-Yourself Methods for Causal
452   Inference: Lessons Learned from a Data Analysis Competition. *Statistical Science* 34:43–68. DOI:
453   10.1214/18-STS667.

454 Dupré C, Stevens CJ, Ranke T, Bleeker A, Peppler-Lisbach C, Gowing DJ, Dise NB, Dorland EDU, Bobbink
455   R, Diekmann M. 2010. Changes in species richness and composition in European acidic
456   grasslands over the past 70 years: the contribution of cumulative atmospheric nitrogen
457   deposition. *Global Change Biology* 16:344–357.

458 Faraway JJ. 2014. *Linear models with R*. Boca Raton, FL: CRC Press.

459 Field CD, Dise NB, Payne RJ, Britton AJ, Emmett BA, Helliwell RC, Hughes S, Jones L, Lees S, Leake JR,
460   Leith ID, Phoenix GK, Power SA, Sheppard LJ, Southon GE, Stevens CJ, Caporn SJM. 2014. The
461   role of nitrogen deposition in widespread plant community change across semi-natural habitats.
462   *Ecosystems* 17:864–877. DOI: 10.1007/s10021-014-9765-5.

463 Firbank LG, Barr CJ, Bunce RGH, Furse MT, Haines-Young R, Hornung M, Howard DC, Sheail J, Sier A,
464   Smart SM. 2003. Assessing stock and change in land cover and biodiversity in GB: an
465   introduction to Countryside Survey 2000. *Journal of Environmental Management* 67:207–218.

466 Fox J. 2016. *Applied Regression Analysis and Generalized Linear Models*. California, USA: SAGE
467   Publications.

468 Fraser LH, Henry HA, Carlyle CN, White SR, Beierkuhnlein C, Cahill JF, Casper BB, Cleland E, Collins SL,
469   Dukes JS, Knapp AK, Lind E, Long R, Luo Y, Reich PB, Smith MD, Sternberg M, Turkington R. 2013.
470   Coordinated distributed experiments: an emerging tool for testing global hypotheses in ecology
471   and environmental science. *Frontiers in Ecology and the Environment* 11:147–155. DOI:
472   10.1890/110279.
473 Fuller RM, Groom GB, Jones AR, Thomson AG. 1993. *Countryside Survey 1990. Mapping the land cover of*
474   *Great Britain using Landsat imagery: a demonstrator project in remote sensing. Final report.*
475   NERC/Institute of Terrestrial Ecology.
476 Gelman A, Carlin J. 2014. Beyond power calculations: assessing type S (sign) and type M (magnitude)
477   errors. *Perspectives on Psychological Science* 9:641–651.
478 Gelman A, Hill J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York,
479   NY: Cambridge University Press.
480 Greenland S, Neutra R. 1980. Control of confounding in the assessment of medical technology.
481   *International Journal of Epidemiology* 9:361–367. DOI: 10.1093/ije/9.4.361.
482 Harrell FE. 2015. *Regression Modelling Strategies*. Switzerland: Springer International Publishing.
483 Heinze G, Wallisch C, Dunkler D. 2018. Variable selection – A review and recommendations for the
484   practicing statistician. *Biometrical Journal* 60:431–449. DOI: 10.1002/bimj.201700067.
485 Hillebrand H, Blasius B, Borer ET, Chase JM, Downing JA, Eriksson BK, Filstrup CT, Harpole WS, Hodapp D,
486   Larsen S, Lewandowska AM, Seabloom EW, Waal DBV de, Ryabov AB. 2018. Biodiversity change
487   is uncoupled from species richness trends: Consequences for conservation and monitoring.
488   *Journal of Applied Ecology* 55:169–184. DOI: 10.1111/1365-2664.12959.
489 Lawes JB, Gilbert JH. 1880. Agricutural, botanical and chemical results of experiments on the mixed
490   herbage of permanent meadow, conducted for more than twenty years in succession on the
491   same land. Part I. The agricultural results. *Philosophical Transactions of the Royal Society (A and*
492   *B)* 171:289–415.
493 Lemoine NP. 2019. Moving beyond noninformative priors: why and how to choose weakly informative
494   priors in Bayesian analyses. *Oikos* 128:912–928. DOI: 10.1111/oik.05985.
495 Lindgren F, Rue H, Lindström J. 2011. An explicit link between Gaussian fields and Gaussian Markov
496   random fields: the stochastic partial differential equation approach. *Journal of the Royal*
497   *Statistical Society: Series B (Statistical Methodology)* 73:423–498.
498 Loken E, Gelman A. 2017. Measurement error and the replication crisis. *Science* 355:584–585.
499 Mac Nally R. 2000. Regression and model-building in conservation biology, biogeography and ecology:
500   the distinction between–and reconciliation of–'predictive'and 'explanatory'models. *Biodiversity*
501   *& Conservation* 9:655–671.
502 Maskell LC, Smart SM, Bullock JM, Thompson K, Stevens CJ. 2010. Nitrogen deposition causes
503   widespread loss of species richness in British habitats. *Global Change Biology* 16:671–679. DOI:
504   10.1111/j.1365-2486.2009.02022.x.
505 Payne RJ, Dise NB, Field CD, Dore AJ, Caporn SJ, Stevens CJ. 2017. Nitrogen deposition and plant
506   biodiversity: past, present, and future. *Frontiers in Ecology and the Environment* 15:431–436.
507   DOI: 10.1002/fee.1528.
508 Peters RH. 1991. *A critique for ecology*. Cambridge, UK: Cambridge University Press.
509 Phoenix GK, Emmett BA, Britton AJ, Caporn SJM, Dise NB, Helliwell R, Jones L, Leake JR, Leith ID,
510   Sheppard LJ, Sowerby A, Pilkington MG, Rowe EC, Ashmore MR, Power SA. 2012. Impacts of
511   atmospheric nitrogen deposition: responses of multiple plant and soil parameters across
512   contrasting ecosystems in long-term field experiments. *Global Change Biology* 18:1197–1215.
513   DOI: 10.1111/j.1365-2486.2011.02590.x.
514 Phoenix GK, Hicks WK, Cinderby S, Kuylenstierna JCI, Stock WD, Dentener FJ, Giller KE, Austin AT, Lefroy
515   RDB, Gimeno BS, Ashmore MR, Ineson P. 2006. Atmospheric nitrogen deposition in world

516   biodiversity hotspots: the need for a greater global perspective in assessing N deposition
517   impacts. *Global Change Biology* 12:470–476. DOI: 10.1111/j.1365-2486.2006.01104.x.
518 Rackham O. 1986. *The History of the Countryside*. London: Dent.
519 Rodwell JS (ed.). 1992. *British Plant Communities Volume 3. Grasslands and montane communities*.
520   Cambridge, UK: Cambridge University Press.
521 Rosenbaum PR. 1984. The Consquences of Adjustment for a Concomitant Variable That Has Been
522   Affected by the Treatment. *Journal of the Royal Statistical Society. Series A (General)* 147:656–
523   666. DOI: 10.2307/2981697.
524 RoTAP. 2012. *Review of Transboundary Air Pollution: Acidification, Eutrophication, Ground Level Ozone*
525   *and Heavy Metals in the UK*. Centre for Ecology & Hydrology.
526 Rowe EC, Jones L, Dise NB, Evans CD, Mills G, Hall J, Stevens CJ, Mitchell RJ, Field C, Caporn SJM,
527   Helliwell RC, Britton AJ, Sutton MA, Payne RJ, Vieno M, Dore AJ, Emmett BA. 2017. Metrics for
528   evaluating the ecological benefits of decreased nitrogen deposition. *Biological Conservation*
529   212:454–463. DOI: 10.1016/j.biocon.2016.11.022.
530 Rubin DB. 2004. Teaching statistical inference for causal effects in experiments and observational
531   studies. *Journal of Educational and Behavioral Statistics* 29:343–367.
532 Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using
533   integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B*
534   *(Statistical Methodology)* 71:319–392.
535 Shmueli G. 2010. To explain or to predict? *Statistical science*:289–310.
536 Silvertown J, Poulton P, Johnston E, Edwards G, Heard M, Biss PM. 2006. The Park Grass Experiment
537   1856–2006: its contribution to ecology. *Journal of Ecology* 94:801–814.
538 Simkin SM, Allen EB, Bowman WD, Clark CM, Belnap J, Brooks ML, Cade BS, Collins SL, Geiser LH, Gilliam
539   FS, Jovan SE, Pardo LH, Schulz BK, Stevens CJ, Suding KN, Throop HL, Waller DM. 2016.
540   Conditional vulnerability of plant diversity to atmospheric nitrogen deposition across the United
541   States. *Proceedings of the National Academy of Sciences* 113:4086–4091. DOI:
542   10.1073/pnas.1515241113.
543 Smart SM, Clarke RT, van de Poll HM, Robertson EJ, Shield ER, Bunce RGH, Maskell LC. 2003a. National-
544   scale vegetation change across Britain; an analysis of sample-based surveillance data from the
545   Countryside Surveys of 1990 and 1998. *Journal of Environmental Management* 67:239–254. DOI:
546   10.1016/S0301-4797(02)00177-9.
547 Smart SM, Robertson JC, Shield EJ, Van De Poll HM. 2003b. Locating eutrophication effects across British
548   vegetation between 1990 and 1998. *Global Change Biology* 9:1763–1774.
549 Smith RI, Fowler D, Sutton MA, Flechard C, Coyle M. 2000. Regional estimation of pollutant gas dry
550   deposition in the UK: model description, sensitivity analyses and outputs. *Atmospheric*
551   *Environment* 34:3757–3777.
552 Soldaat LL, Pannekoek J, Verweij RJT, van Turnhout CAM, van Strien AJ. 2017. A Monte Carlo method to
553   account for sampling error in multi-species indicators. *Ecological Indicators* 81:340–347. DOI:
554   10.1016/j.ecolind.2017.05.033.
555 Soons MB, Hefting MM, Dorland E, Lamers LPM, Versteeg C, Bobbink R. 2017. Nitrogen effects on plant
556   species richness in herbaceous communities are more widespread and stronger than those of
557   phosphorus. *Biological Conservation* 212:390–397. DOI: 10.1016/j.biocon.2016.12.006.
558 Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. 2016. Increasing Transparency Through a Multiverse
559   Analysis. *Perspectives on Psychological Science* 11:702–712. DOI: 10.1177/1745691616658637.
560 Stephens PA, Buskirk SW, del Rio CM. 2007. Inference in ecology and evolution. *Trends in Ecology &*
561   *Evolution* 22:192–197. DOI: 10.1016/j.tree.2006.12.003.
562 Stevens CJ, Dise NB, Mountford JO, Gowing DJ. 2004. Impact of nitrogen deposition on the species
563   richness of grasslands. *Science* 303:1876–1879.

564   Stevens CJ, Dupré C, Dorland E, Gaudnik C, Gowing DJG, Diekmann M, Alard D, Bobbink R, Corcket E,
565   Mountford JO, Vandvik V, Aarrestad PA, Muller S, Dise NB. 2011a. Grassland species
566   composition and biogeochemistry in 153 sites along environmental gradients in Europe. *Ecology*
567   92:1544–1544. DOI: 10.1890/11-0115.1.
568   Stevens CJ, Dupré C, Gaudnik C, Dorland E, Dise N, Gowing D, Bleeker A, Alard D, Bobbink R, Fowler D,
569   Vandvik V, Corcket E, Mountford JO, Aarrestad PA, Muller S, Diekmann M. 2011b. Changes in
570   species composition of European acid grasslands observed along a gradient of nitrogen
571   deposition. *Journal of Vegetation Science* 22:207–215. DOI: 10.1111/j.1654-1103.2010.01254.x.
572   Stevens CJ, Smart SM, Henrys P, Maskell LC, Walker KJ, Preston CD, Crowe A, Rowe E, Gowing DJ,
573   Emmett BA. 2011c. *Collation of evidence of nitrogen impacts on vegetation in relation to UK*
574   *biodiversity objectives*. Peterborough, UK: JNCC.
575   Storkey J, Macdonald AJ, Poulton PR, Scott T, Köhler IH, Schnyder H, Goulding KWT, Crawley MJ. 2015.
576   Grassland biodiversity bounces back from long-term nitrogen addition. *Nature* 528:401–404.
577   DOI: 10.1038/nature16444.
578   Tanguy M, Prudhomme C, Smith K, Hannaford J. 2018. Historical gridded reconstruction of potential
579   evapotranspiration for the UK. *Earth System Science Data* 10:951–968. DOI:
580   https://doi.org/10.5194/essd-10-951-2018.
581   Tipping E, Henrys PA, Maskell LC, Smart SM. 2013. Nitrogen deposition effects on plant species diversity;
582   threshold loads from field data. *Environmental Pollution* 179:218–223. DOI:
583   10.1016/j.envpol.2013.04.008.
584   UK National Ecosystem Assessment. 2011. *The UK National Ecosystem Assessment Technical Report*.
585   Cambridge: UNEP-WCMC.
586   Van der Eerden LJ, Dueck TA, Berdowski JJM, Greven H, Van Dobben HF. 1991. Influence of NH_3 and
587         $(NH_4)_2SO_4$ on heathland vegetation. *Acta Botanica Neerlandica* 40:281–296.
588   Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP. 2006. Why do we still use stepwise
589   modelling in ecology and behaviour? *Journal of Animal Ecology* 75:1182–1189. DOI:
590   10.1111/j.1365-2656.2006.01141.x.
591   Willi JC, Mountford JO, Sparks TH. 2005. The Modification of Ancient Woodland Ground Flora at Arable
592   Edges. *Biodiversity & Conservation* 14:3215–3233. DOI: 10.1007/s10531-004-0443-3.
593   Williams ED. 1974. Changes in Yield and Botanical Composition Caused by the New Liming Scheme on
594   Park Grass. In: *Rothamsted Experimental Station Report for 1973 Part 2*. Rothamsted Research,
595   67–73.
596   Young C. 2018. The Difference Between Causal Analysis and Predictive Models: Response to "Comment
597   on Young and Holsteen (2017)." *Sociological Methods & Research*:0049124118782542. DOI:
598   10.1177/0049124118782542.
599

600   Acknowledgements

601   We thank the authors of the original studies reanalysed for publishing their data or otherwise making it
602   available to us. We also thank A. Britton, M. Grainger, P. Henrys, S.M. Smart, L.C. Maskell for useful
603   comments on the manuscript.

Figure 1

Regression coefficient plots

(a) Estimated regression coefficients for the reanalysis of Maskell et al. (2010); (b) estimated regression coefficients for the reanalysis of Stevens et al. (2004) using a reduced set of covariates chosen for their similar ecological status to the covariates used by Maskell et al. (2010), referred to in this paper as SEA04 model 1; (c) estimated regression coefficients for the reanalysis of Stevens et al. (2004) using a set of covariates designed to match the original analysis of that paper as closely as possible, referred to in this paper as SEA04 model 2. The dependent variable was vascular plant species richness in all cases. White circles represent the posterior median estimate, black bars the posterior 50% credible interval, grey bars the posterior 95% credible interval. All covariates are described in Table 1.

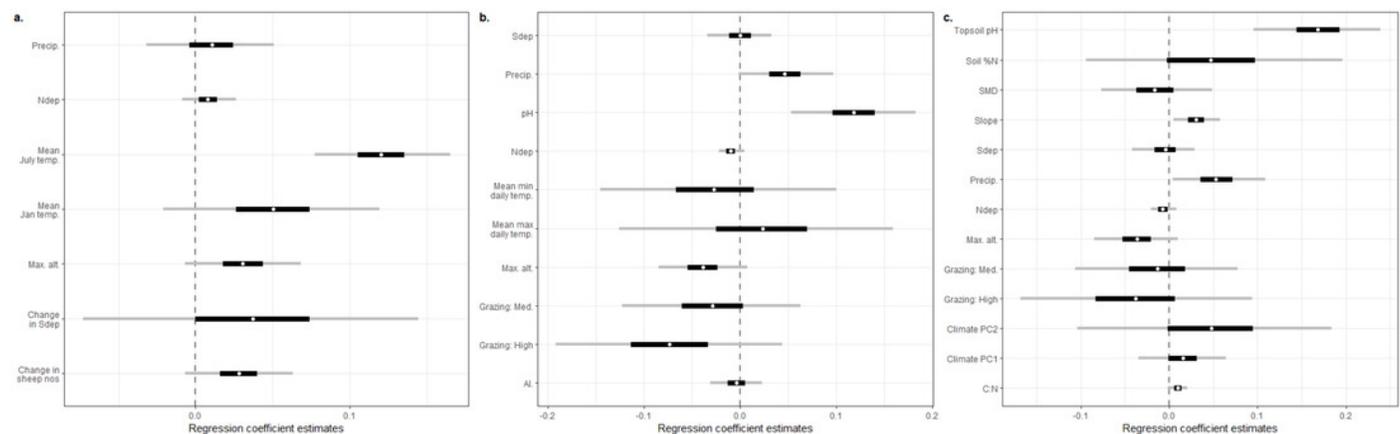
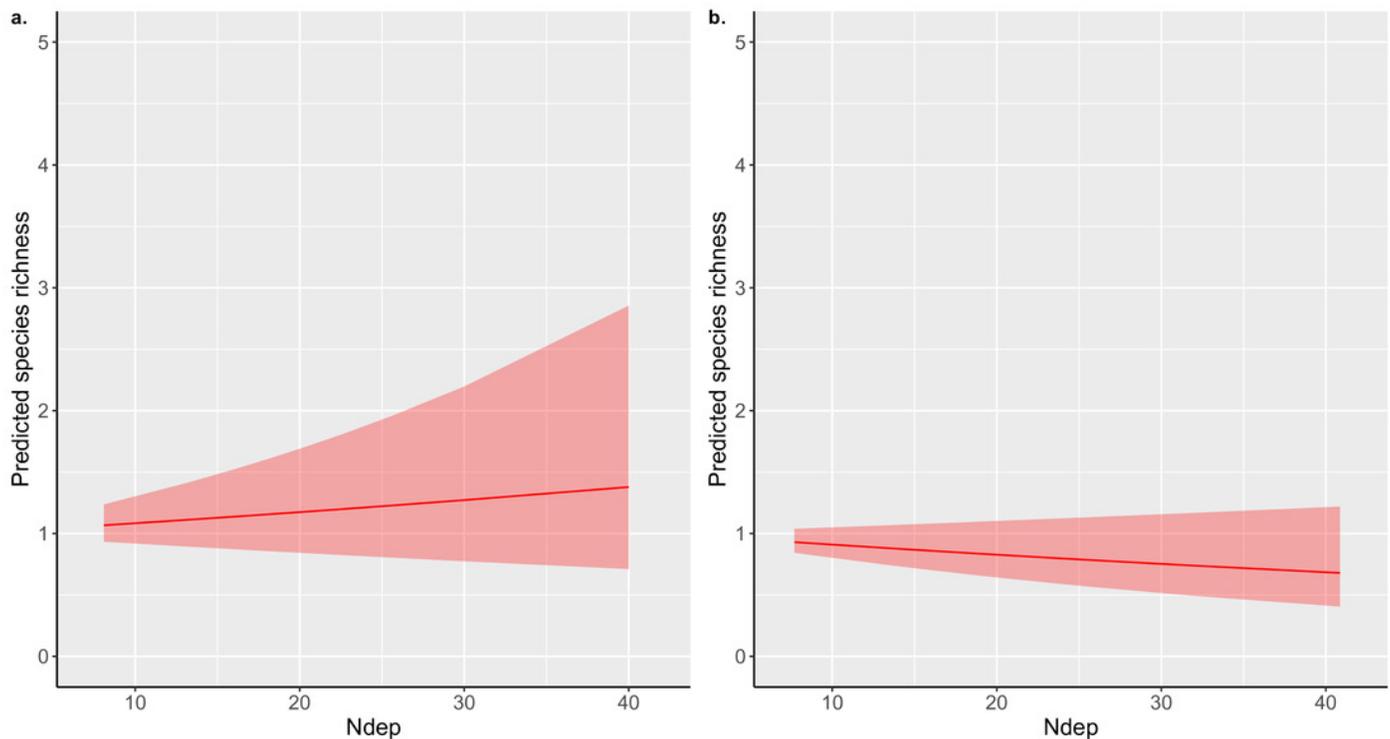



Figure 2

Total nitrogen deposition partial effect predictions.

INLA model predictions of the partial effect of total nitrogen deposition (Ndep; kg ha^{-1} yr^{-1}) on acid grassland species richness estimated for (a) MEA10 and (b) SEA04 model 1; both the models used for these predictions used vascular plant richness only as the dependent variable. Predictions were estimated for each 10th-percentile of the respective Ndep ranges covered by each study using the linear combinations option of INLA. Linear combinations were estimated for each point in the Ndep range with other covariates set to zero, hence the low values of the predicted richnesses for different levels of total Ndep.

Summary information for all covariates used in our reanalyses.

All covariates with their original spatial grain sizes, original ranges, and re-scalings as used in the reanalyses of Maskell et al. (2010) (MEA10) and Stevens et al. (2004) (SEA04) presented here.

1   Table 1. All covariates with their original spatial grain sizes, original ranges, and re-scalings as used in the
2   reanalyses of Maskell et al. (2010) (MEA10) and Stevens et al. (2004) (SEA04) presented here.

Covariate	Supporting refs or sources	Original data grain size	Original range	Original units	Rescaled range where relevant	Rescaled units where relevant	Relevant reanalysis
Total Ndep estimated over moorland	Maskell et al. (2010); Smith et al. (2000)	5 × 5 km	4.9–40.0	kg ha^{-1} yr^{-1}	-	-	MEA10
Change (1970 to 1998) from peak Sdep estimated over moorland	Smith et al. (2000); L. Maskell (pers. comm.)	5 × 5 km	-5.36–0.00	Δ kg S ha^{-1} yr^{-1}	-	-	
Max. altitude	Maskell et al. (2010)	1 × 1 km	0–975	m	0.00–9.75	100 m	
Mean min. Jan. temp (1961-1999)	Maskell et al. (2010)	5 × 5 km	-8.16–0.08	°C	-	-	
Mean max. Jul. temp (1961-1999)	Maskell et al. (2010)	5 × 5 km	14.11–26.67	°C	-	-	
Mean annual precipitation (1961-1999)	Maskell et al. (2010)	5 × 5 km	554.33–3305.80	mm	2.22–13.22	250 mm	
Change in sheep numbers	Maskell et al. (2010)	2 × 2 km	-11.19–88.47	Δ sheep per year (1969 to 2000)	-1.12–8.85	Δ 10 sheep per year (1969 to 2000)	
Mean annual precipitation (1996-2006)	Stevens, Dupré, Dorland et al. (2011)	25 × 25 km	604.9–1773.3	mm	0.42–7.09	250 mm	SEA04: Model 1
Mean annual daily max. temp. (1996-2006)	Stevens, Dupré, Dorland et al. (2011)	25 × 25 km	11.5–14.6	°C	-	-	
Mean annual daily min. temp. (1996-2006)	Stevens, Dupré, Dorland et al. (2011)	25 × 25 km	4.2–8.1	°C	-	-	
Topsoil aluminium	Stevens, Duprè, Gaudnik et al. (2011)	Empirical plot data	11.60–1318.75	mg kg^{-1} dry soil	0.06–6.59	200 mg kg^{-1} dry soil	
Total Sdep estimated over grassland	Stevens et al. (2004); Smith et al. (2000)	5 × 5 km	3.20–13.44	kg ha^{-1} yr^{-1}	-	-	SEA04: Models 1 & 2
Total Ndep estimated over grassland	Stevens et al. (2004); Smith et al. (2000)	5 × 5 km	7.70–40.86	kg ha^{-1} yr^{-1}	-	-	
Topsoil pH	Stevens et	Empirical	3.69–5.37	pH unit	7.38–10.74	0.5 pH unit	

	al. (2004); Stevens, Duprè, Gaudnik et al. (2011)	plot data					
Max. altitude	Stevens et al. (2004); Stevens, Duprè, Gaudnik et al. (2011)	Empirical plot data	15–500	m	0.15–5.00	100 m	
Grazing intensity	Stevens et al. (2004); Stevens, Duprè, Gaudnik et al. (2011)	Empirical plot data	Coded as low/ medium/ high	-	-	-	
Climate PC1	See methods	See methods	-4.90–3.43	-	-	-	
Climate PC2	See methods	See methods	-1.47–1.20	-	-	-	
C:N	Stevens et al. (2004); Stevens, Duprè, Gaudnik et al. (2011)	Empirical plot data	13.34–30.58	topsoil mass ratio	-	-	
Slope	Stevens et al. (2004); Stevens, Duprè, Gaudnik et al. (2011)	Empirical plot data	0–60	°	0–6	10°	SEA04: Model 2 only
Soil %N	Stevens et al. (2004); Stevens, Duprè, Gaudnik et al. (2011)	Empirical plot data	0.12–1.57	topsoil %N	-	-	
Soil moisture deficit (SMD)	Stevens et al. (2004); Hough & Jones (1997)	40 × 40 km	1.66–48.94	mm	0.17–4.89	10 mm	

Estimated regression coefficients for total nitrogen deposition across models.

Exponentiated total nitrogen deposition regression coefficient medians, 2.5 and 97.5% quantiles for all models (all given to 2 decimal places). For example, for model MEA10, vascular plants only, a median value of 1.01 implies that a 1% gain of species richness per kg ha^{-1} yr^{-1} of total N deposition is highly compatible with the data; however, the 2.5% quantile value of 0.99 for this model also suggests that losses up to and around 1% are plausible (all values being conditional on model assumptions and data accuracy).

1   Table 2. Exponentiated total nitrogen deposition regression coefficient medians, 2.5 and 97.5% quantiles
2   for all models (all given to 2 decimal places). For example, for model MEA10, vascular plants only, a
3   median value of 1.01 implies that a 1% gain of species richness per kg ha^{-1} yr^{-1} of total N deposition is
4   highly compatible with the data; however, the 2.5% quantile value of 0.99 for this model also suggests
5   that losses up to and around 1% are plausible (all values being conditional on model assumptions and
6   data accuracy).

Model, dependent richness variable	Spatial error structure	2.5% quantile	Median	97.5% quantile	Model location
MEA10, vascular plants only	Mesh + square	0.99	1.01	1.03	Fig. 1a
MEA10, vascular plants and bryophytes	Mesh + square	0.99	1.00	1.02	Fig. A2.1
SEA04 model 1, vascular plants only	Mesh + square	0.98	0.99	1.00	Fig. 1b
SEA04 model 2, vascular plants only	Mesh + square	0.98	0.99	1.01	Fig. 1c
SEA04 model 1, vascular plants and bryophytes	Mesh + square	0.98	0.99	1.01	Fig. A2.2
SEA04 model 2, vascular plants and bryophytes	Mesh + square	0.98	0.99	1.01	Fig. A2.3
MEA10, vascular plants only	Square only	0.97	0.98	1.00	Fig. A3.1
MEA10, vascular plants only	Mesh only	0.98	0.99	1.01	Fig. A3.2