# Supporting information

## Contents

# Appendix S1. Further methodological details for modelling woodland susceptibility to rhododendron invasion

## Appendix S1.1 Rationale for modelling rhododendron occurrence, the dataset and selected drivers

### Rationale

Rhododendron is an evergreen shrub that has become a well-established pest of conservation and economic concern in Britain since its introduction as an ornamental in 1763. It inhibits the regeneration of native woodlands, reduces woodland biodiversity and increases the cost of forestry operations if pre-treatment is required. Eradication expenses depend on the site accessibility and clearance method and can cost as much as £10,000 $ha^{-1}$ (Dehnen-Schmutzet al. 2004). Several studies have investigated its establishment requirements, but at small geographic extents thus far (e.g. Thomson et al. 1993; Stephenson et al. 2006). Understanding variation in the effects of factors that facilitate rhododendron invasion at broad extents, such as across Britain, could allow for the spatial targeting of grant aid and monitoring for early removal and management.

### National Forest Inventory

We used the first cycle of the Forestry Commission's National Forest Inventory (NFI; 2010 – 2015) to model rhododendron invasion. This rolling field survey scheme incorporates over 15,000 1ha woodland 'squares' across Great Britain, from which data describing the site's biophysical attributes and human activities are collected using a standardised protocol (https://www.forestry.gov.uk/fr/beeh-a3gf9u#fieldsurvey). These sites were selected using a stratified-random sampling technique. Each 1-ha square is subdivided into sections, areas of woodland within a survey square that are relatively homogenous in terms of management and land use attributes including silvicultural system, age, height, and are at least 0.05ha in extent (S1.1).

## Appendix S1.2 Details of the drivers selected for the analysis of rhododendron occurrence

Table S1.1. Drivers selected for the analysis of rhododendron occurrence

| | Relationship to rhododendron occurrence probability | Variable type | Source, original resolution and processing details |
|---|---|---|---|
| **Region-level** | | | |
| Soil moisture deficit | Negative to negative quadratic – rhododendron requires damp soils and is intolerant to drought and water-logging[37]. | Continuous | Ecological Site Classification (ESC), originally available at 250-m https://www.forestresearch.gov.uk/tools-and-resources/forest-planning-and-management-services/ecological-site-classification-decision-support-system-esc-dss/ |
| Elevation | Negative - suitability decreases with increasing elevation[38]. | Continuous | OS Terrain 50, available at 50-m resolution https://www.ordnancesurvey.co.uk/xml/products/OSTerrain50Grid.xml |
| Soil pH | Negative - rhododendron is found in a range of acidic soil conditions, from pH 3 to 6.4, but growth is generally inhibited below pH 5[37]. | Continuous | Countryside Survey's Model estimates of topsoil pH and bulk density at 1-km resolution https://catalogue.ceh.ac.uk/documents/gemini/waf/ |
| **Landscape-level** | | | |
| Road density | Positive – roads represent potential corridors along which invasive species may spread as a direct result of as well as indirectly through modification of the environment in a way that is favourable to establishment. | Continuous | OS OpenRoads https://www.ordnancesurvey.co.uk/business-and-government/products/os-open-roads.html |
| Distance to woodland edge | Positive or negative – increasing distance to an edge tends towards damper conditions that favour germination, however, it likely also results in a further distance from propagule sources | Continuous | Derived using 2016 Forestry Commission National Forest Inventory (NFI) Map (vector) for Great Britain (http://data-forestry.opendata.arcgis.com/datasets/national-forest-inventory-woodland-gb). Measured using ArcGIS (www.esri.com, v10.2.2), using section centroids. |
| Distance to historic park or garden | Negative – will decline with increasing distance from this propagule source. | Continuous | (Historic England; The Welsh Historic Environment Service (Cadw), Historic Environment Scotland). Measured using ArcGIS (www.esri.com, v10.2.2), using section centroids |
| Woodland amount and configuration metrics | Positive - woodland cover serves a potential source of propagules for establishment. | Continuous | 25-m resolution land cover raster for Britain, LCM2007[53]. Landscape metrics were calculated using the *ClassStat* function provide by the R package SDMTools (VanDerWal et al. 2014). https://digimap.edina.ac.uk/webhelp/environment/data_information/lcm2007.htm |
| **Local-level** | | | |
| Interpreted forest type (IFT) | Rhododendron is typically associated with mixed, rather than monoculture forest stands. | Categorical | NFI field survey. Sites classified as IFT category 'Young trees' (*n*=) were removed due to ambiguity. |
| Stocking density | Negative – a high stocking density of trees, limits both the amount of light reaching the forest floor and water availability, therefore limiting rhododendron establishment. | Continuous | NFI field survey |
| Stand vertical complexity | Positive – seedling establishment is typically associated with bryophytes, which are more abundant under more complex canopies. Rhododendron also associated with deep leaf litter, a correlate of stand vertical complexity[37]. | Ordinal (1-5) | NFI field survey (level 5 was merged with level 4 due to very low sample size) |
| Stand age | . Canopy closure, and therefore the amount of light reaching the forest floor, varies with stand age, so rhododendron establishment may vary with stand age. Quadratic, linear or log relationships are plausible. | | NFI field survey |
| Signs of herbivory | Positive – Disturbances caused by grazing creates 'safe sites' for seedling establishment[76]. | Binary | NFI field survey |
| Aspect | Rhododendron has been shown to favour northerly aspects[36]. | Categorical (NSEW) | Derived using ArcGIS (www.esri.com, v10.2.2), using section centroids |

| | | | |
|---|---|---|---|
| Woodland patch size | Positive or negative – increasing distance to an edge tends towards damper conditions that favour germination, however, it likely also results in a further distance from propagule sources. | Continuous | NFI map data |

## Appendix S1.3 Defining regional contexts

Regional drivers of rhododendron establishment (identified in Step 2) were harmonised to a common resolution of 1-km and subjected to Spearman's correlation analysis within the variable ranges sampled by the NFI, following standardisation to *z*-scores. Principal component analysis (PCA) was applied to moisture deficit (MD) and elevation as these were highly correlated across Britain as a whole (Spearman's rho = 0.80 within the range sampled by NFI squares). PCA identified two gradients.

## Appendix S1.4 Characterising landscape structure

We used a 25-m resolution land cover raster for Britain, LCM2007 (Morton et al. 2014). Landscape metrics were calculated using the *ClassStat* function provide by the R package SDMTools (VanDerWal et al. 2014). Distance measures were carried out in ArcGIS (www.esri.com, v10.2.2), using section centroids. R code to calculate landscape metrics within multiple buffers for multiple sites can be found in Appendix S4.

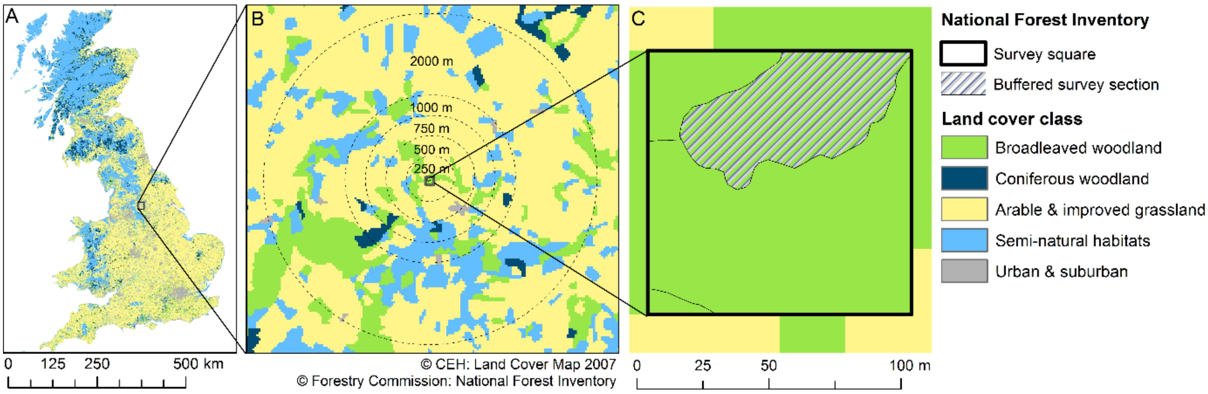


Figure S1.1 Landscape structure metrics hypothesised to influence the probability of rhododendron establishment were quantified within multiple buffers surrounding each NFI section.

## Appendix S1.5 Details of data exploration and processing

Woodland patch size ("wood.size") was omitted from the global model due to high collinearity with landscape-level woodland cover ("wood.cov"):

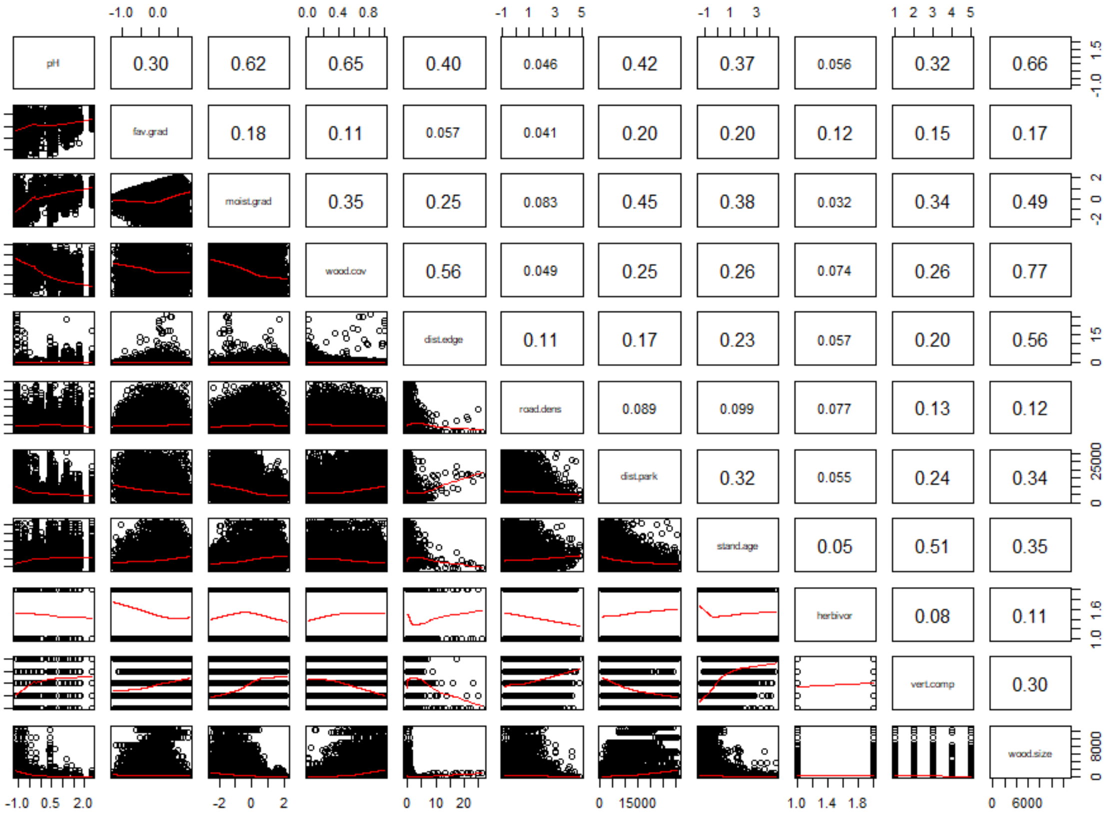

Fig. S1.2. Spearman's rank correlation coefficients of variables considered in the analysis of rhododendron occurrence

## Appendix S1.6 Details of model selection to identify important drivers and scales

To quantify how susceptibility to invasion varied with regional, landscape, local-level drivers and their interactions as hypothesised in Step 5 (Table 1), we fitted generalised linear models against a binomial distribution with a clog-log link function to the rhododendron occurrence data. Sections area varied, so section area was fitted as an offset, thus controlling for effects of section area on the probability of rhododendron occurrence. To account for the spatial non-independence of sections within squares, a single section was randomly selected from each NFI square, leaving a total of 12,473 sections. This subsampling approach was used instead of a mixed-effects modelling framework (with NFI squares identified as random effects), due to insufficient computational power and also because handling random effects in the IT environment is troublesome, particularly when averaging models[82]. All possible combinations of meaningful terms were included in models constructed by maximum likelihood methods with R package MuMIn (Barton, 2013), to allow model comparisons based on AIC with small-sample correction (AICc; Burnham and Anderson, 2004). We included quadratic relationships with regional gradients and log relationships with landscape and local-level metrics. The ΔAIC value between the best and second best model was 1.7.

Table S1.2. Parameter estimates of the minimum adequate model explaining variation in the probability of rhododendron establishment at a woodland site. Explanatory variables were centred and scaled prior to analysis to improve interpretability of regression coefficients (Schielzeth, 2010).

| Explanatory variable | | Hierarchy level | Parameter est. | Standard error | *P* |
|---|---|---|---|---|---|
| intercept | | NA | -0.17 | 0.43 | <0.001 |
| log10(woodland cover) | | lands | 0.61 | 0.15 | <0.001 |
| pH | | rgn | -0.08 | 0.08 | 0.003 |
| poly(favourability gradient,2)1 | | rgn | -15.26 | 6.52 | <0.001 |
| poly(favourability gradient,2)2 | | rgn | 9.99 | 3.82 | <0.001 |
| log10(woodland cover):favourability gradient | | rgn: lands | 0.61 | 0.26 | <0.001 |
| log10(woodland cover): pH | | rgn: lands | 0.40 | 0.15 | 0.001 |
| road density | | lands | 0.24 | 0.03 | <0.001 |
| distance to historic garden | | lands | -0.82 | 0.11 | <0.001 |
| vertical structure | | loc | 1.35 | 0.14 | <0.001 |
| IFT | coniferous | loc | 0.05 | 0.11 | 0.042 |
| | broadleaved mixed | | 0.34 | 0.17 | <0.001 |
| | coniferous mixed | | 0.36 | 0.11 | <0.001 |
| stand age | | loc | 0.19 | 0.37 | <0.001 |

Table S1.3. Relative importance values for explanatory variables contained within considerably supported models (ΔAIC ≤10) explaining rhododendron occurrence. All models compared here included landscape-level variables calculated at a 500-m-extent, except road density, which was calculated within a 250-m buffer. Weights calculated by summing up the Akaike weights of models that included the term in question (Burnham and Anderson, 2004).

| Term | Importance value |
|---|---|
| log10(woodland cover): | 1.00 |
| ph | 1.00 |
| stand age | 1.00 |
| vertical structure | 1.00 |
| moisture and elevation gradient | 1.00 |
| road density | 1.00 |
| distance to historic garden | 1.00 |
| IFT | 0.96 |
| log10(woodland cover): ph | 0.96 |
| favourability gradient | 0.91 |
| distance to woodland edge | 0.76 |
| log10(woodland cover):favourability gradient | 0.74 |
| log(stocking density) | 0.71 |
| signs of herbivory | 0.27 |
| woodland origin | 0.14 |

Table S1.4. Generalised variance inflation factors (GVIF) for variables contained within the minimum adequate model, calculated following Fox & Monette (1992). All GVIF values are below 2, suggesting collinearity is not an issue.

| Variable | Df | GVIF |
|---|---|---|
| Stand age | 1 | 1.166 |
| IFT | 3 | 1.083 |
| Vertical structure | 4 | 1.032 |
| Road density | 1 | 1.018 |
| distance to historic garden | 1 | 1.110 |
| Favourability gradient | 2 | 1.659 |
| log10(woodland cover) | 1 | 1.374 |
| Moisture and elevation gradient | 2 | 1.158 |
| PH98 | 1 | 1.910 |
| log10(woodland cover):favourability gradient | 2 | 1.712 |

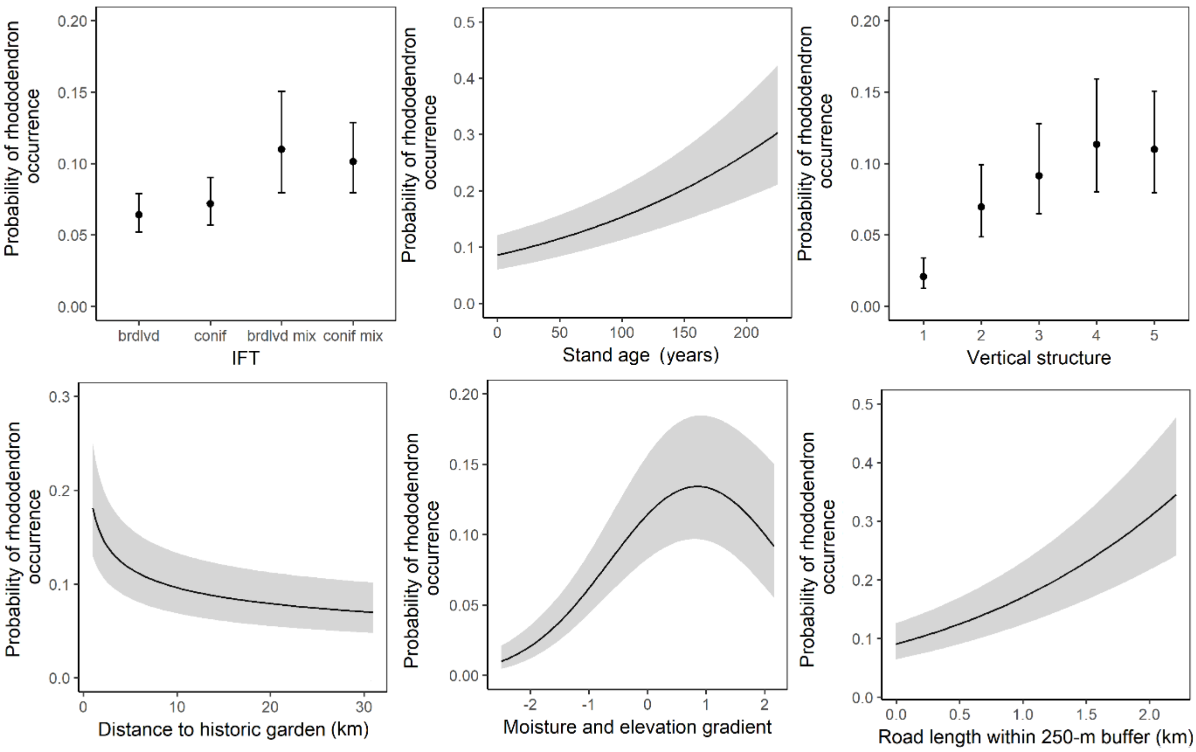


Figure S1.3. Conditional plots of variation in the probability of rhododendron occurrence in woodland sites in relation to local and landscape-level drivers. Results were graphed using coefficients from the minimum adequate model. Error bars represent 95% confidence intervals.

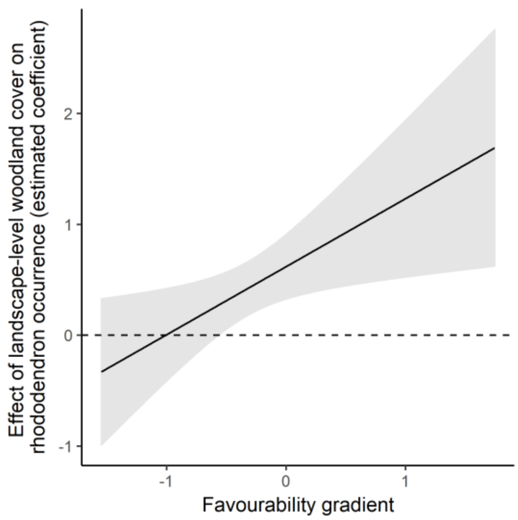


Figure S1.4. Cross-scale interactions of the effects of landscape-level woodland cover (within 500-m of a woodland site) on woodland susceptibility to invasion depends on the favurability gradient. Relationship graphed using coefficients from the minimum adequate model with conditional variables held at their mean.

# Appendix S2. Further methodological details for modelling pond water quality

## Appendix S2.1 Details of the drivers selected for the analysis of soluble reactive phosphorus concentrations (SRP) in ponds

Table S2.1. Drivers selected for the analysis of pond water quality

| Driver | Relationships with SRP | Variable type | Source, original resolution and processing details |
|---|---|---|---|
| *Region-level* | | | |
| Precipitation | Can increase runoff and carry agricultural runoff to ponds in agricultural areas. Could also serve to 'flush out' ponds, which are generally shallow in the UK. | Continuous | UK Met office<br>Average of monthly values across 2006-2007 / mm<br>https://www.metoffice.gov.uk/climate/uk/data/ukcp09/datasets |
| Temperature | Many nutrient cycling processes, such as microbially mediated reactions, are affected by temperature. | Continuous | UK Met office<br><br>Average of monthly values across 2006-2007 / °C<br>https://www.metoffice.gov.uk/climate/uk/data/ukcp09/datasets |
| Slope | High slopes can exacerbate agricultural runoff. | Continuous | OS Terrain 50, available at 50-m resolution<br>https://www.ordnancesurvey.co.uk/xml/products/OSTerrain50Grid.xml |
| Soil | SRP is more susceptible to leaching in light sandy soils, and so is more likely to flow into ponds that are surrounded by intensive agriculture. | Continuous | Soil Parent Material Model (PMM) gives a soil classification at a resolution at 1-km. Proportion of 10-km square covered by soils classified as 'light' by the Soil Parent Material Model (PMM)<br>https://www.bgs.ac.uk/products/onshore/soilPMM.html |
| Atmospheric nitrate deposition | Atmospheric deposition can be a significant source of nitrogen, which can affect N:P, ratios and so the concentration of SRP. | Continuous | Deposition maps for the UK are available at a resolution of 5km from UK Pollution Deposition. 2006-2007 average values were used for wet nitrates. Available online http://www.pollutantdeposition.ceh.ac.uk/ |
| *Landscape-level* | | | |
| Source cover landscape metrics | Intensive land covers including arable and improved grasslands are sources of P. | Continuous | LCM2007. Calculated in buffers of varying extents surrounding ponds.<br>https://digimap.edina.ac.uk/webhelp/environment/data_information/lcm2007.htm |
| Sink cover landscape metrics | Can reduce agricultural runoff by interception. | Continuous | LCM2007<br>https://digimap.edina.ac.uk/webhelp/environment/data_information/lcm2007.htm |
| Slope (topographic position index) | Ponds with low (negative) TPI values were predicted to be more susceptible to agricultural runoff, if surrounded by intensive land uses. | Continuous | OS Terrain 50, available at 50-m resolution<br>https://www.ordnancesurvey.co.uk/xml/products/OSTerrain50Grid.xml<br>A pond's TPI equals the elevation of the pond minus the mean elevation of the surrounding area - this value can be calculated at multiple extents of the surrounding area.<br>SRP exhibited a Laplace distribution in relation to TPI, peaking at 0 TPI, so values were converted to absolute values for modelling, and the variable was referred to as 'slope'. |
| *Local-level* | | | |

| | | | |
|---|---|---|---|
| Inflow | Inflows directly link ponds to stream drainage systems and so are inlets of stream-borne pollutants. | Binary | CS; Presence/absence of wet or dry inflow into pond |
| Area | Larger area corresponds to larger perimeter for intercepting pollutants in ground water or surface runoff | Continuous | CS measured in $m^2$ |
| Buffer | Will intercept runoff and therefore reduce SRP concentrations | Continuous | CS; % land within 100-m zone occupied by sink land covers (woodland or scrubby vegetation). |

## Appendix S2.2 Defining regional contexts

Regional drivers of pond water quality (identified in Step 2) were harmonised to a common resolution of 10-km and subjected to PCA following standardisation to *z*-scores.

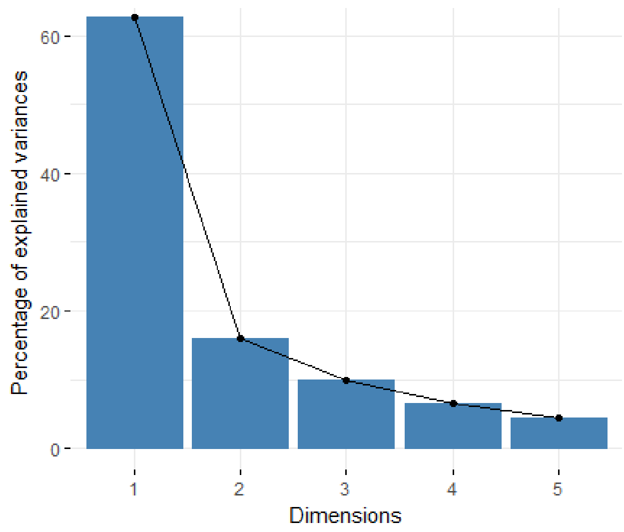


Figure S2.1. Percentage of variance explained by each axis obtained by the PCA of regional variables across the UK.

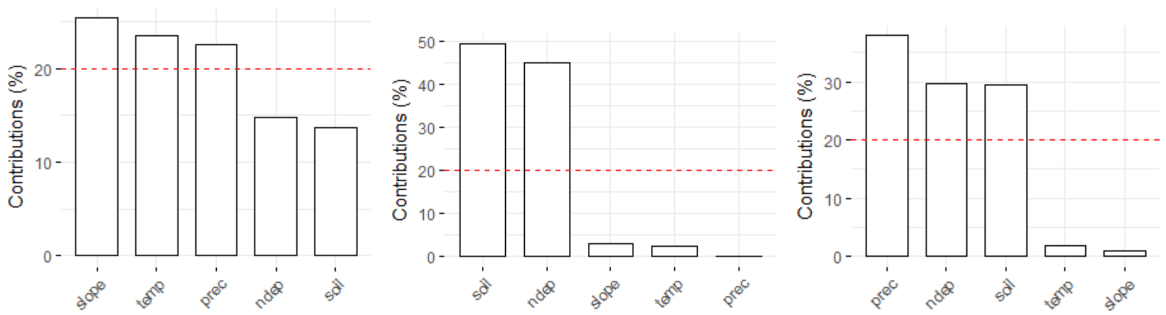


Figure S2.2. Variable contributions to three regional gradients identified by PCA. The red dashed indicates the expected average contribution if the contributions of the variables were uniform; variables with a contribution larger than this value are considered the most important.

Table S2.2. Variable correlation coefficients with the three principle components representing regional gradients across the UK. Variance values indicate the percentage of the total variance in regional heterogeneity accounted for by each principle component.

| | Axis1 | Axis2 |
|---|---|---|
| prec | 0.84 | 0.02 |
| temp | -0.86 | -0.14 |
| soil | 0.56 | 0.63 |
| slope | 0.89 | -0.16 |
| ndep | 0.68 | -0.60 |
| Variance / % | 62.74 | 16.13 |

## Appendix S2.3 Details of model selection to identify important drivers and scales

To investigate drivers of pond water quality, we fitted generalised linear models against a negative binomial distribution with a log link function to soluble reactive phosphorous concentrations after conversion to integer values (multiplication by 1000). The model section procedure followed that for rhododendron establishment (see main text).

The minimum adequate model explaining variation in pond water quality contained intensive land cover within 250-m of the ponds and its interaction with the soil gradient (Axis 2), a main effect of the precipitation gradient (Axis 1) and the presence of inflows (Tables S2.4-S2.7).

Table S2.3. Parameter estimates of the minimum adequate model that explained pond soluble reactive phosphorous concentration. Explanatory variables were centred and scaled prior to analysis to improve interpretability of regression coefficients (Schielzeth, 2010).

| Explanatory variable | Hierarchy level | Parameter estimate | Standard error | *P* | Importance value* |
|---|---|---|---|---|---|
| Intercept | NA | 5.10 | 0.23 | <0.001 | NA |
| Intensive cover | Landscape | 1.35 | 0.187 | <0.001 | 1.00 |
| Soil gradient (Axis 2) | Regional | -0.14 | 0.17 | 0.488 | 1.00 |
| Inflow | Local | -0.67 | 0.33 | 0.040 | 0.70 |
| Soil gradient: Intensive cover | Regional: landscape | 0.54 | 0.19 | 0.003 | 1.00 |

Table 2.4. Relative importance values for explanatory variables contained within considerably supported models (ΔAIC ≤10) explaining pond soluble reactive phosphorous concentrations. All models compared here included landscape-level variables calculated at a 250-m-extent. Weights calculated by summing up the Akaike weights of models that included the term in question (Burnham and Anderson, 2004).

| Term | Importance value |
|---|---|
| intensive cover | 1.00 |
| soil gradient | 0.95 |
| soil gradient: intensive cover | 0.84 |
| inflow | 0.70 |
| slope | 0.50 |
| precipitation gradient | 0.47 |
| buffer | 0.34 |
| inflow: intensive cover | 0.19 |
| precipitation gradient: intensive cover | 0.13 |
| buffer: intensive cover | 0.09 |

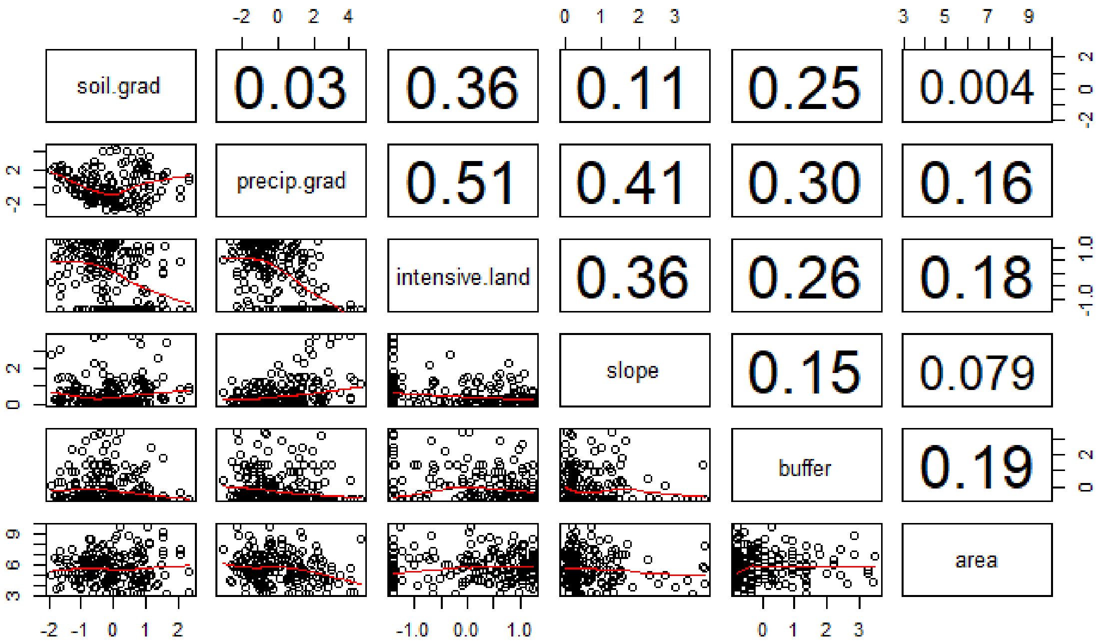


Fig. S2.3. Spearman’s rank correlation coefficients of variables considered in the analysis of pond water quality.

Table S2.4. Generalised variance inflation factors (GVIF) for variables contained within the minimum adequate model, calculated following Fox & Monette (1992). All GVIF values are below 2, suggesting collinearity is not an issue.

| Variable | Df | GVIF |
|---|---|---|
| Soil gradient | 1 | 1.18 |
| Intensive cover | 1 | 1.17 |
| soil gradient: intensive cover | 2 | 1.03 |
| inflow | 1 | 1.02 |

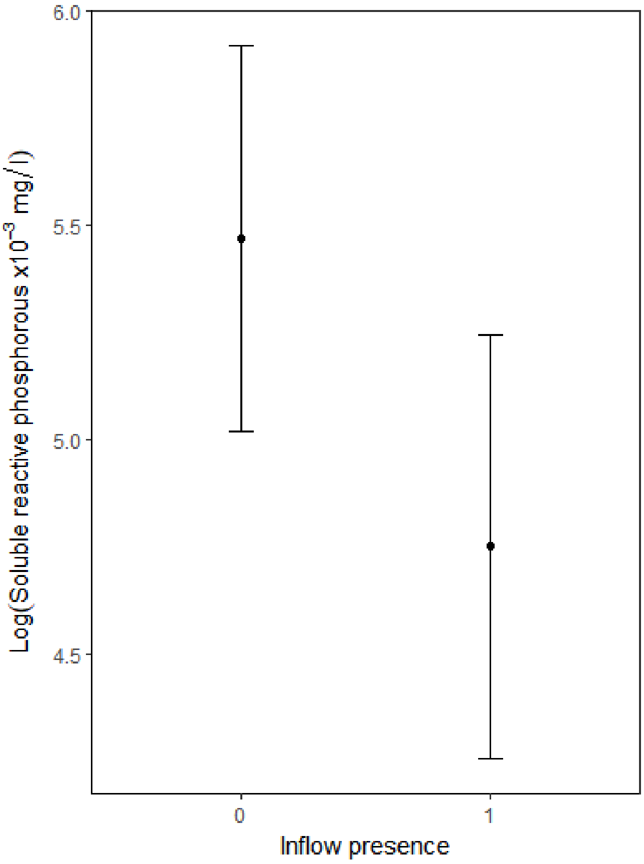


Figure S2.4 Conditional plot of variation in pond soluble reactive phosphorus concentration in relation to inflow presence. Results were graphed using coefficients from the minimum adequate model. Error bars represent 95% confidence intervals.

## Supplementary Information References


Bartoń, K. MuMIn: model selection and model averaging based on information criteria (AICc and alike) R package version. (2018).

Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed)*. *Ecological Modelling* **172,** (2002).

Dehnen-Schmutz, K., Perrings, C. & Williamson, M. Controlling Rhododendron ponticum in British Isles: An economic analysis. J. Environ. Manage. 70, 323–332 (2004).

Forestry Commission (2017). National Forest Inventory field survey data, first cycle. https://www.forestresearch.gov.uk/tools-and-resources/national-forest-inventory/

Fox, J. and Monette, G. Generalized collinearity diagnostics. JASA, **87**, 178--183. (1992).

Morton, R. D. et al. Land Cover Map 2007 (25m raster, GB) v1.2. (2014).

Thomson, A. G., Radford, G. L., Norris, D. A. & Good, J. E. G. Factors affecting the distribution and spread of Rhododendron in North Wales. J. Environ. Manage. 39, 199–212 (1993).

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., & Storlie, C. SDMTools: Species distribution modelling tools: Tools for processing data associated with species distribution modelling exercises. R package version 1.1-221. (2014).