Spatial structure in lotic macroinvertebrate communities in England and Wales: relationship with physicochemical and anthropogenic stress variables.

JOHN F. MURPHY and JOHN DAVY-BOWKER

Centre for Ecology and Hydrology Dorset, Winfrith Technology Centre, Winfrith Newburgh, Dorchester, Dorset, UK.

Correspondence: John F. Murphy, CEH Dorset, Winfrith Technology Centre, Winfrith Newburgh, Dorchester, Dorset, DT2 8ZD, UK

Email: jomu@ceh.ac.uk

Abbreviated title: Determinants of macroinvertebrate community structure

Keywords: macroinvertebrates, community structure, spatial structure, physicochemical variables, pollution impacts.

This paper has not been submitted elsewhere in identical or similar form, nor will it be during the first three months after its submission to *Hydrobiologia.*

Abstract

We describe the relationship between macroinvertebrate community composition, the physicochemical environment and anthropogenic impacts, in running water sites across a range of water qualities in England and Wales. We have also investigated the degree of spatial structure present in both the macroinvertebrate community and the measured environment.

Selected explanatory variables could account for 26% of the variation in lotic macroinvertebrate assemblage composition across England and Wales. The explanatory power of the CCA model was based predominantly on a combination of local scale variables (substrate, alkalinity, urban run-off) and regional scale variables (discharge category, northing). The physicochemical gradient associated with changes in stream type from headwaters to estuary dominated assemblage composition. The influence of pollution and habitat modification were of secondary importance. There was a substantial level of spatial structure to both the physicochemical (47% of its explanatory power spatially structured) and anthropogenic stress data (63% of its explanatory power spatially structured), which resulted in a high level of predictable spatial structuring in macroinvertebrate assemblages. Almost 40% of the variation in assemblage composition accounted for by the explanatory model exhibited spatial structure. Positive spatial autocorrelation in macroinvertebrate community composition extended to sites up to 150 km apart. As a consequence, community composition could be described from northing and easting with 75% of the explanatory power of the eight physicochemical variables.

Our study has confirmed the importance of the longitudinal gradient within catchments, as well as the geographical position of the catchment to macroinvertebrate communities. We have also demonstrated how quantifying the spatial structure in the dataset can improve our understanding of the factors influencing macroinvertebrate community structure.

Introduction

Spatial variation in lotic macroinvertebrate community structure is due to a combination of intrinsic biotic community interactions (e.g. M^{c}Auliffe, 1984; Oberndorfer *et al*., 1984; Kohler, 1992; Malmqvist, 1993), environmental conditions (e.g. Edington, 1968; Minshall & Minshall, 1977; Erman & Erman, 1984) and historical factors (Ricklefs & Schluter, 1993; Lake 2000). Whilst there are examples of biotic interactions influencing the regional distribution of species (Holdrick & Reeve, 1991; Kohler & Wiley, 1997) it is more often the case that environmental conditions and historical events play a dominant role in determining the structure of macroinvertebrate communities over broad geographical areas (Allan, 1995).

The importance of physical and hydrochemical conditions in determining lotic macroinvertebrate community composition has been established (Wright *et al*., 1984; Corkum, 1989) and has lead to the development of robust models that predict community composition at un-polluted river sites using a range of such variables e.g. RIVPACS in the UK, AUSRIVAS in Australia and BEAST in Canada (see Wright *et al*., 2000). Initially in the UK, 28 environmental variables were used to predict the fauna at sites but it was found that the accuracy of the prediction was little affected when a subset of 11 variables was used (Wright, 2000). In Australia different sets of predictor variables were used for different regions. In both the Australian and UK models, the variables could be grouped into broad-scale between-catchment variables, stream-size variables describing the position of the site within the catchment and small-scale habitat-specific features (Moss *et al*., 1987; Simpson & Norris, 2000). Similarly in Swedish streams, a combination of broad scale factors (e.g. latitude, longitude and altitude) and local-scale factors (e.g. stream velocity and depth) were found to best predict the macroinvertebrate community at un-polluted sites (Sandin & Johnson, 2000).

A range of anthropogenic stresses (e.g. eutrophication, acidification, canalisation and sedimentation) can alter water quality or modify riparian and instream habitat features with consequences for the macroinvertebrate fauna (Mason, 1991). In England and Wales, 88.8% of sites surveyed by the Environment Agency (EA) during their 1995 General Quality Assessment (GQA) survey were considered to be influenced by some degree of human impact, with sewage treatment works and farming being the most commonly recorded stressors on lotic systems (Davy-Bowker *et al*., 1999). Such disturbances can disrupt the natural physicochemical gradients within rivers, can alter the macroinvertebrate community and can affect the balance of instream ecosystem processes (Vannote *et al*., 1980; Giller & Malmqvist, 1998). To date, most attempts to understand regional or national scale relationships between environmental gradients and macroinvertebrate taxonomic composition have tried to exclude the confounding influence of polluted or physically modified streams (Wright *et al*., 1984; Sandin & Johnson, 2000; Simpson & Norris, 2000; Heino *et al*., 2002). Therefore at a regional scale the relative importance of physicochemical variables and anthropogenic stresses in determining community composition is poorly understood.

The lotic environment is characterised by a definite spatial structure both between and within catchments, which can result in patches or gradients in the values of spatially structured variables (Legendre & Legendre, 1998). The level of spatial structuring within the macroinvertebrate community will depend to a certain extent on how closely it responds to the environment. However, previous studies have demonstrated that taxon distributions can exhibit significant spatial structure even after accounting for the variation due to measured environmental factors (Borcard *et al*., 1992; Magnan *et al*., 1994; Magalhães *et al*., 2002; Potapova & Charles, 2002). This unexplained spatial element may represent the biota responding to unmeasured spatially structured abiotic variables, biotic interactions within the community that

lead to spatial autocorrelation, or it may represent the legacy of past events and biogeographical constraints (Legendre & Legendre 1998). By addressing the degree to which macroinvertebrate communities are spatially structured and the role of the physicochemical environment in determining this pattern, a better knowledge of the forces governing the taxonomic composition at a given site can be obtained.

The development and application of consistent field and laboratory methods and quality assurance systems within the Environment Agency, together with recent improvements in data management practices, have made it possible to undertake investigations of the factors influencing broad scale lotic macroinvertebrate community structure across the whole of England and Wales using extensive and reliable data sets. It is clear that such studies would benefit the more effective implementation of the EU Water Framework Directive (WFD) (European Commission, 2000) and the operation of biomonitoring programmes (Hawkins *et al*., 2000; Logan & Furse, 2002). Therefore, the present study had three objectives:

1. Identify the relative importance of physicochemical variables and anthropogenic stresses in determining macroinvertebrate community composition over a broad spatial extent.
2. Assess the degree to which macroinvertebrate communities are spatially structured.
3. Determine the extent to which the physicochemical and stress variables exhibit spatial structure.

Materials and methods

Macroinvertebrate and associated physicochemical data were acquired for 5752 sites, from the EA quinquennial GQA survey of English and Welsh rivers undertaken in 1995 (Figure 1). During the spring (March-May) and autumn (September-November) of that year sites were sampled using standard Environment Agency methodology (Murray-Bligh, 1999). This involved a 3 minute active kick sample with a 900-µm mesh pond net, where all habitats within the site were sampled in proportion to their occurrence. The samples were collected as part of a national biomonitoring programme and hence the macroinvertebrates were not identified further than to Biological Monitoring Working Party (BMWP) family level (National Water Council, 1981) (Table 1). Specimens not included within the 82 BMWP groups were omitted from the analysis. The log abundance category (*0*: not present, *1*: 1-9 individuals, *2*: 10-99, *3*: 100-999, *4*: 1000-9999, *5*: ≥10000) for each taxon in each season was also recorded and the maximum log abundance over the two seasons was used in the analysis. Eight associated physicochemical variables for each site were either measured at the time of sampling or recorded from maps and long-term datasets (Table 2). Mean alkalinity values were obtained from a parallel EA chemical monitoring programme. Perceived anthropogenic stress data were obtained from a questionnaire circulated to EA staff (Davy-Bowker *et al*., 1999). The severity of 12 major types of anthropogenic stress (Table 2) acting on each site during the sampling period in 1995 was recorded by local EA biologists from their detailed knowledge of sites in their area (*0*: not present, *1*: light, *2*: moderate or *3*: severe). Environment Agency catchment management plans were also consulted by biologists where necessary information was unavailable.

Site location was included as an environmental variable as a non-linear function of the geographic site coordinates. The easting (x) and northing (y) of the each site were centred to zero mean (to reduce collinearity) and used with the linear (xy),

quadratic (x^2, x^2y, xy^2, y^2) and cubic terms (x^3, y^3) of a third-order non-linear polynomial equation as spatial variables in the analysis (Legendre & Legendre, 1998). Second and third order terms were calculated to allow for more complex, patchy spatial patterns in assemblage composition to be detected (Legendre & Legendre, 1998).

Taxa occurring at less than 1% of sites were removed from the analyses following preliminary exploration of their influence in an initial detrended correspondence analysis (DCA) and as a result, 71 taxa remained. All multivariate ordinations were carried out using CANOCO 4.5 (ter Braak & Smilauer, 2002). DCA revealed that the rate of turnover of macroinvertebrate taxa across the sites on the first axis of variation was such that a unimodal model assumption would be more appropriate than a linear model assumption for the dataset (DCA axis 1 length = 2.8) (ter Braak, 1995). Therefore, canonical correspondence analysis (CCA) was undertaken between the macroinvertebrate assemblage data and the 29 explanatory variables at the 5752 sites. A forward selection procedure identified the subset of variables that were making a statistically significant contribution to the variation in the macroinvertebrate data ($P < 0.001$). Variables with substantial collinearity (inflation factor > 3) with other more powerful explanatory variables were also excluded from subsequent analyses. This resulted in 6 variables being rejected, the perceived stresses, sedimentation, riparian modifications, consolidated banks and intensive arablisation and the spatial terms x^3 and y^3. Eight physicochemical variables, 8 perceived stress variables and 7 spatial variables were retained.

To assess the relationship between the physicochemical and stress variable groups and macroinvertebrate community composition and the degree to which they exhibited a predictable spatial structure, a series of CCAs and partial CCAs were carried out as described in Bocard *et al*. (1992), Okland & Eilertsen (1994) and Legendre & Legendre (1998). Initially, a CCA was run to measure the total explained

variation attributable to all 23 explanatory variables and conversely the unexplained portion. Then 3 partial CCAs were carried out to calculate the variation uniquely attributable to each explanatory variable group. Next, a further 6 partial CCAs with one explanatory variable group as the variables in the analysis and another as covariables were undertaken. This calculated the variation attributable to each variable group plus that portion due to the interaction with the appropriate other variable group. The interaction terms between the 3 variable groups were calculated by appropriate subtraction of terms calculated in the previous steps (Legendre & Legendre, 1998).

As an alternative approach to quantifying the extent of spatial structure in lotic macroinvertebrate communities a multivariate Mantel correlogram was computed to delineate the 'zone of spatial autocorrelation' (Oden & Sokal, 1986). Biological dissimilarity between the 5752 sites was calculated as their Euclidean distances apart on the DCA ordination axes. Euclidean distances were calculated from DCA ordination space instead of using ecological distances computed directly e.g. Bray-Curtis or Kulczyinski coefficient in order to maintain consistency in the methods between the two alternative approaches to assessing spatial structure. A random sub-sample of 10% of the sites was used to test for spatial autocorrelation in the macroinvertebrate data. The corresponding pair-wise geographical distances between the sub-sampled sites were grouped into 8 classes and each distance class was tested for spatial autocorrelation using a Mantel test with 5000 permutations (Oden & Sokal, 1986; Legendre & Legendre, 1998).

Results

The total extent of variation or total inertia (TI) in lotic macroinvertebrate assemblage composition across England and Wales was equivalent to 1.174 eigenvalues of which the 23 explanatory variables could explain 26% (Table 3). The most powerful individual explanatory variables were substrate composition (10.2%), alkalinity (9.4%), easting (8.5%) and mean water depth (5.1%) (Table 4). The explanatory power of the forward selection model was based predominantly on a combination of substrate, alkalinity, discharge category, northing and urban run-off (Table 4).

The dominant gradient (axis 1) through the macroinvertebrate data distinguished assemblages that contain greater than average abundances of taxa such as Philopotamidae, Perlidae and Cordulegasteridae from those with greater than average abundances of Corixidae, Notonectidae and Coenagrionidae (Figure 2). This axis was primarily a function of the longitudinal position of a site within a catchment and the geographical position of the catchment (Figure 3). Sites situated towards the western half of the survey area with high altitude and steep slopes were found in the negative end of axis 1 (Figure 3). Such sites were also more prone to acidification stress (Figure 3). Sites towards the positive end of axis 1 had finer substrate composition, higher alkalinity and were characterised by problems with canalisation, excessive in-stream macrophyte growth and agrochemical inputs (Figure 3).

The second CCA axis distinguished sites impacted by urban run-off, organic inputs and industrial discharges, with assemblages dominated by Oligochaeta, Chironomidae, Erpobdellidae, Glossiphonidae and Asellidae, from the larger river sites that were relatively un-impacted by pollution and supported diverse assemblages of Odonata, Trichoptera and Unionidae (Figures 2 & 3). On the one hand the second axis is a further extension of the upstream-downstream gradient within catchments as well as the gradient from upland catchments to more lowland

catchments. While on the other hand the second axis represents the gradient of organic and industrial impact on macroinvertebrate assemblages regardless of position along the catchment. The spatial pattern of variation in macroinvertebrate assemblage structure can be clearly seen when the geographical position of each site relative to its position along each axis is plotted (Figure 4). The community composition characteristic of the negative end of axis 1 was generally found at sites in Wales, the northwest and southwest of England while the community composition characteristic of the positive end of axis 1 was found at sites in the centre and east of England (Figure 4a). This underlines the strong influence of physical characteristics and geographical position of each site on macroinvertebrate community structure. The assemblage structure characteristic of the negative end of the second axis was found at clusters of sites in the west and north midlands and north east of England. This coincided with areas of traditional heavy industry and urban development (Figure 4b). The assemblage structure characteristic of the positive end of the second axis was found at sites along lowland, relatively un-impacted rivers such as the rivers Frome, Stour and Avon in southern England and the R. Nene in the midlands (Figure 4b).

Positive spatial autocorrelation in macroinvertebrate community composition extended to sites up to 150 km apart (Figure 5). Sites within this zone of influence had more similar assemblages than would be expected for randomly associated pairs of sites. Negative spatial autocorrelation was also evident between sites greater than 200 km apart, indicating that assemblages at these sites were more similar the further apart they were situated.

When considered independently, the eight physicochemical variables accounted for 19% and the eight stress variables explained 9% of the TI in the taxon data. The total variance explained (TVE) by the two variable groups together with the seven spatial terms was 26%. Thirty three percent of the TVE could be accounted for by

the physicochemical variables alone, whilst 9.4% of the TVE was uniquely attributable to the stress variables (Figure 6). Just under half of the TVE exhibited no predictable spatial structure. A further 39.5 % of the TVE was accounted for by physicochemical and stress variables exhibiting a predictable spatial pattern (Figure 6). The remaining 13% of the TVE was uniquely attributable to the spatial variables (Figure 6).

Nearly half of the contribution of the physicochemical variables towards TVE was spatially structured i.e. it was also accounted for by the spatial variables (Figure 6). The other half did not have an easily defined spatial structure (Figure 6). The stress variables were very strongly spatially structured, sharing over half of their explanatory power with the spatial variables (Figure 6).

In order to examine the nature of the 13% of TVE that could be accounted for by the spatial variables alone, the site scores from the partial CCA of the spatial variables, with the physicochemical and stress variables as covariables, were plotted against each site's geographical position (Figure 7). The gradient of sites along axis 1 is similar to axis 2 of the initial CCA (Figure 4b), suggesting that it represents a further expression of the variation in macroinvertebrate communities along the pollution gradient in English and Welsh streams, but one not previously identified by the measured physicochemical or stress variables (Figure 7). The gradient of sites along axis 2 generally distinguishes the macroinvertebrate assemblage at sites to the south and midlands of the survey area from those on the western, eastern and northern fringes (Figure 7).

Discussion

The present study considered macroinvertebrate assemblage structure as a function of physicochemical variables in combination with anthropogenic impacts across a range of sites from high quality to severely impacted. Many national surveys of equivalent sampling intensity, investigating relationships between the biota and their environment, have examined only un-impacted sites (Wright *et al*., 1984; Potapova & Charles, 2002; Heino *et al*., 2003; Sandin 2003). Although the survey of anthropogenic impacts conducted by EA biologists provided data with inherent subjectivity, it still provided a good indication of the spatial distribution and intensity of the major types of stream pollution and modification in England and Wales. Urban run-off and organic pollution were identified as being the two most widespread and influential anthropogenic impacts disrupting the ecological integrity of streams and rivers. Other stresses were less influential on a national scale, though the incidence and intensity of canalisation and inputs of agrichemicals increased to the east of the study area, while acidification was more prevalent in the west. The stress variables contributed relatively little extra explanatory power (9.4%) to the model over that provided by the physicochemical variables. However the considerable amount of interaction between the spatial and stress components illustrated the strong spatial structuring in the distribution of the various impacts e.g. the impact of industrial discharges and run-off tended to be more common and intense at sites in the north of the region.

The importance of quantifying the spatial structure in correlative studies such this one has been emphasised by Borcard *et al*. (1992) and Legendre & Legendre (1998). The pure spatial component of the variance may act as a synthetic descriptor of unmeasured but spatially structured physicochemical variables, of biotic factors such as aggregation or dispersal, and of historical factors that have left a detectable spatial pattern. By including spatial variables in the multivariate analysis the spatial

context or structure of other variables and the biotic data may be described. The physical landscape of England & Wales exhibits a definable spatial structure. The east-west gradient in macroinvertebrate assemblage composition is five times as long as that found in a north-south direction i.e. there is five times more taxa turnover from east to west than north to south. This reflects the more diverse nature of physicochemical conditions from the Cornish, Welsh and Cumbrian uplands to the Humberside and East Anglian lowlands than that found between the north and south of the region. We have shown that the spatial structure of the stream macroinvertebrate assemblages is very closely associated with this pattern. As a consequence, the community composition at a site can be described from its northing and easting with 75% of the explanatory power of the eight physicochemical variables. The Mantel correlogram also confirms the considerable extent of spatial structure in macroinvertebrate assemblages in that the zone of influence extended for up to 150 km from a site. Perhaps a more ecologically relevant measure of geographic distance between sites would have been linear distance along the river network as opposed to simple geometric distance (e.g. in Magalhães et al., 2002) however this would have been computationally very difficult, even for the 10% of the 5752 sites analysed. Furthermore, geometric distance, while it ignored the river network, preserved some features of catchments in that sites at similar altitude in adjacent streams can be as biologically similar as equidistant sites on the same waterway.

The inclusion of spatial variables also detected otherwise unaccounted for patterns in the macroinvertebrate assemblages that seemed to be related to urbanisation/industrialisation. The concentrations of various pollutants at each site were not included in the analysis. The spatial variables may well have been acting as surrogate variables for infrequently measured or poorly understood pollutants associated with urban waterways e.g. heavy metals or petrochemical concentrations.

Also a second gradient perhaps related to temperature was detected by the spatial variables (Figure 7b). In RIVPACS, longitude and latitude are incorporated as variables, along with mean air temperature and air temperature range, as predictors of community composition at undisturbed sites (Moss *et al*., 1987). However, quadratic and cubic functions of the spatial variables are not included so only linear trends can be detected. Nevertheless these variables contribute to the understanding of broad scale regional changes in assemblage structure, even over the relatively short latitudinal range of Great Britain, over and above those accounted for by the other physicochemical variables (Moss *et al*., 1987). The current findings also support the approach of stratifying bioassessment programmes into stream types based on either the biotic assemblages as in RIVPACS or environmental characteristics as adopted by the WFD (European Commission, 2000). The WFD will provide the impetus and mandate for a more coordinated collection of comprehensive sets of site-matched biological and physicochemical data at regional and national levels within the EU in the future. Such data will assist future investigations of the factors determining variation in macroinvertebrate assemblages across the European landscape.

The current study found that a combination of variables measured at a range of scales (substrate composition, alkalinity, urban run-off, discharge category and northing) were the most important descriptors of community composition across England & Wales. Previous studies of macroinvertebrate communities in undisturbed UK streams have found substrate composition, alkalinity and slope to be the most powerful explanatory variables, with distance from source and discharge category also important (Wright *et al*., 1984). In this previous work the dominant gradient described the broad scale differences in assemblage composition between catchments across the UK while the secondary gradient was related more to the position of a site within a catchment (Wright *et al*., 1984). The length of the

ordination gradients in the present study were not as long as that found by Wright *et al*. (1984) due to the greater range of physical conditions and more detailed level of taxonomic resolution incorporated in their study.

Johnson & Goedkoop (2002) assessed the relative importance of variables acting on a range of scales on littoral macroinvertebrate communities in Swedish lakes. Their study found that local habitat variables were better predictors of community composition than regional or geographical scale variables. However the best predictive model incorporated variables from a range of spatial scales. The spatial distribution of river invertebrate assemblages across the northwest of North America was also best described using a combination of biogeographical (drainage basin, bedrock geology) and in-site hydrological (current velocity, stream depth) variables (Corkum, 1989). Other studies in catchments draining into Lake Huron, USA, streams in Finland and in Sweden have also shown that a combination of locally measured factors (substrate composition, channel width, pH, nitrate concentration, in-stream moss cover, depth, riparian vegetation) and regional factors (latitude, air temperature, precipitation) offered the best explanation for the variation in benthic macroinvertebrate community structure (Richards *et al*., 1993; Heino *et al*., 2003; Sandin, 2003). Therefore such a grouping of variables acting or measured at a hierarchical range of scales seems to be a powerful determinant of lotic community structure (Johnson & Goedkoop, 2002). The present study has shown that while there is considerable interaction between different variable groups acting at different scales, they still contribute significant amounts of unique explanatory power to the model and hence merit inclusion.

The initial CCA indicated that 26% of the variation in the taxon data could be explained by the 23 variables used in the analysis. This is towards the lower end of the range (20-50%) for TVE found in most CCA (Okland, 1999). The large proportion of residual variation in the model means that the relationships between the abiotic

and biotic variables need to be interpreted with caution. However, Okland (1999) has emphasised that we should not be overly concerned with the ratio of TVE to TI because of the unknown relative influence of unmeasured physicochemical variables, lack-of-fit of the data to the unimodal model, polynomial distortions and stochastic biological processes on the unexplained portion of the TI. The proportion of TVE in this study would probably have been increased had further important physicochemical and biological variables been measured directly across all sites e.g. phosphates, suspended sediment, macrophyte cover. Also at the taxonomic level used in the current study it is likely that some groups did not completely satisfy the unimodal model assumption for CCA e.g. Chironomidae. At a lower taxonomic level the response curves would have been better defined along gradients leading to a less noisy biota-environment association. The logistical difficulties of undertaking such a detailed survey of the region within a single year preclude the identification of samples to generic or species level. Nonetheless, the interpretation of relative values for variation attributable to different variable groups as used in the present study is still valid (Okland, 1999) and provides useful information for the broad scale management of lotic systems in terms of the regional scale distribution of macroinvertebrate taxa and spatial trends in the impact of anthropogenic stresses on lotic systems. It should be stressed that the current analysis is based on a correlative approach and no causal links have been established, however the substantial number of sites and comprehensive coverage of all catchments within the study area adds considerable weight to the relationships described.

In conclusion this analysis has confirmed that physicochemical gradients dominate assemblage composition in England & Wales and that the impacts of pollution and habitat modification are secondary to this gradient. It has also quantified the degree of spatial structuring in macroinvertebrate assemblages across the region due to the spatial predictability in both the measured physicochemical environment and

unmeasured but spatially structured variables e.g. temperature, and in the occurrence of anthropogenic impacts. This highlights the importance of including spatial terms in such analyses, first so that an understanding of the spatial structure of the macroinvertebrate data can be gathered but also because it improves the fit of the model by acting as a surrogate for unmeasured spatially structured variables.

Acknowledgments

The authors would like to thank the Environment Agency for use of the 1995 GQA data set and for dealing with various queries. We would also like to thank John Wright and Ralph Clarke for very helpful comments on earlier drafts.

References

Allan, J.D., 1995. Stream Ecology: Structure and Function of Running Waters. Chapman and Hall, London, 379 pp.

Borcard, D., P. Legendre & P. Drapeau, 1992. Partialling out the spatial component of ecological variation. Ecology 73: 1045-1055.

Corkum, L.D., 1989. Patterns of benthic invertebrate assemblages in rivers of northwestern North America. Freshwater Biology 21: 191-205.

Davy-Bowker, J., M.T. Furse, R.T. Clarke & M.J. Gravelle, 1999. Analysis of 1995 survey data. Phase 2 post-survey appraisal. Unit I: Taxon distribution studies. R&D Technical Report E103, Environment Agency, Bristol, 575 pp.

Edington, J.M., 1968. Habitat preferences in net-spinning caddis larvae with special reference to the influence of water velocity. Journal of Animal Ecology 37: 675-692.

Erman, D.C. & N.A. Erman, 1984. The response of stream macroinvertebrates to substrate size and heterogeneity. Hydrobiologia 108: 75-82.

European Commission, 2000. Directive 2000/60/EC, Establishing a framework for community action in the field of water policy. European Commission PE-CONS 3639/1/100 Rev 1, Luxembourg.

Giller, P.S. & B. Malmqvist, 1998. The Biology of Streams and Rivers. Oxford University Press, 296 pp.

Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson & R.J. Stevenson, 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19: 541-556.

Heino, J., T. Muotka, R. Paavola, H. Hämäläinen & E. Koskenniemi, 2002. Correspondence between regional delineations and spatial patterns in macroinvertebrate assemblages of boreal headwater streams. Journal of the North American Benthological Society 21: 397-413.

Heino, J., T. Muotka & R. Paavola, 2003. Determinants of macroinvertebrate diversity in headwater streams: regional and local influences. Journal of Animal Ecology 72: 425-434.

Holdrick, D.M. & I.D. Reeve, 1991. Distribution of freshwater crayfish in the British Isles, with reference to the crayfish plague, alien introductions and water quality. Aquatic Conservation: Marine and Freshwater Ecosystems 1: 139-158.

Johnson, R.K. & W. Goedkoop, 2002. Littoral macroinvertebrate communities: spatial scale and ecological relationships. Freshwater Biology 47: 1840-1854.

Kohler, S.L., 1992. Competition and structure of a benthic stream community. Ecological Monographs 62: 1749-1761.

Kohler, S.L. & M.J. Wiley, 1997. Pathogen outbreaks reveal large-scale effects of competition in stream communities. Ecology 78: 2164-2176.

Lake, P.S., 2000. Disturbance, patchiness and diversity in streams. Journal of the North American Benthological Society 19: 573-592.

Legendre, P. & L. Legendre, 1998. Numerical Ecology. Elsevier, Amsterdam, 853 pp.

Logan, P & M.T. Furse, 2002. Preparing for the European Water Framework Directive – making the links between habitat and aquatic biota. Aquatic Conservation: Marine and Freshwater Ecosystems 12: 425-437.

Magalhães, M.F., D.C. Batalha & M.J. Collares-Pereira, 2002. Gradients in stream fish assemblages across a Mediterranean landscape: contributions of environmental factors and spatial structure. Freshwater Biology 47: 1015-1031.

Magnan, P., M.A. Rodríguez, P. Legendre & S. Lacasse, 1994. Dietary variation in a freshwater fish species: relative contribution of biotic interactions, abiotic factors and spatial structure. Canadian Journal of Fisheries and Aquatic Sciences 51: 2856-2865.

Malmqvist, B., 1993. Interactions in stream leaf packs: effects of a stonefly predator on detritivores and organic matter processing. Oikos 66: 454-462.

Mason, C.F., 1991. Biology of Freshwater Pollution. Longman Scientific & Technical, Harlow, UK, 351 pp.

M^{c}Auliffe, J.R., 1984. Competition for space, disturbance, and the structure of a benthic stream community. Ecology 65: 894-908.

Minshall, G.W. & J.N. Minshall, 1977. Microdistribution of benthic invertebrates in a Rocky Mountain (USA) stream. Hydrobiologia 55: 231-249.

Moss, D., M.T. Furse, J.F. Wright & P.D. Armitage, 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17: 41-52.

Murray-Bligh, J.A.D., 1999. Procedure for Collecting and Analysing Macroinvertebrate Samples. Environment Agency, Bristol, UK.

National Water Council, 1981. River Quality: the 1980 Survey and Future Outlook. National Water Council, London.

Oberndorfer, R.Y., J.V. McArthur, J.R. Barnes & J. Dixon, 1984. The effects of invertebrate predators on leaf litter processing in an alpine stream. Ecology 65: 1325-1331.

Oden, N.L. & R.R. Sokal, 1986. Directional autocorrelation: an extension of spatial correlograms to two dimensions. Systematic Zoology 35: 608-617.

Okland, R.H., 1999. On the variation explained by ordination and constrained ordination axes. Journal of Vegetation Science 10: 131-136.

Okland, R.H. & O. Eilertsen, 1994. Canonical correspondence analysis with variation partitioning: some comments and an application. Journal of Vegetation Science 5: 117-126.

Potapova, M.G. & D.F. Charles, 2002. Benthic diatoms in USA rivers: distributions along spatial and environmental gradients. Journal of Biogeography 29: 167-187.

Richards, C., G.E. Host & J.W. Arthur, 1993. Identification of predominant environmental factors structuring stream macroinvertebrate communities within a large agricultural catchment. Freshwater Biology 29: 285-294.

Ricklefs, R.E. & D. Schluter, 1993. Species diversity: regional and historical influences. In Ricklefs, R.E. & D. Schluter (ed.), Species Diversity in Ecological Communities: Historical and Geographical Perspectives. University of Chicago: 350-363.

Sandin, L. & R.K. Johnson, 2000. Ecoregions and benthic macroinvertebrate assemblages of Swedish streams. Journal of the North American Benthological Society 19: 462-474.

Sandin, L., 2003. Benthic macroinvertebrates in Swedish streams: community structure, taxon richness, and environmental relations. Ecography 26: 269-282.

Simpson, J.C. & R.H. Norris, 2000. Biological assessment of river quality: development of AusRivAS models and outputs. In Wright, J.F., D.W. Sutcliffe & M.T. Furse (ed.), Assessing the Biological Quality of Fresh Waters. Freshwater Biological Association, Ambleside, UK: 125-142.

ter Braak, C.J.F., 1995. Ordination. In Jongman, R.H.J., C.J.F. ter Braak & O.F.R. van Tongeren (ed.), Data Analysis in Community and Landscape Ecology. Cambridge University Press, Cambridge, UK: 91-173.

ter Braak, C.J.F. & P. Smilauer, 2002. CANOCO reference manual and User's guide to Canoco for windows software for canonical community ordination (version 4.5). Wageningen, Netherlands, 500 pp.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell & C.E. Cushing, 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37: 130-137.

Wright, J.F., 2000. An introduction to RIVPACS. In Wright, J.F., D.W. Sutcliffe & M.T. Furse (ed.), Assessing the Biological Quality of Fresh Waters. Freshwater Biological Association, Ambleside, UK: 1-24.

Wright, J.F., D. Moss, P.D. Armitage, & M.T. Furse, 1984. A preliminary classification of running water sites in Great Britain based on macro-invertebrate species and prediction of community type using environmental data. Freshwater Biology 14: 221-256.

Wright, J.F., D.W. Sutcliffe & M.T. Furse (ed.), 2000. Assessing the Biological Quality of Fresh Waters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, UK, 373 pp.

Table 1 The 82 macroinvertebrate taxonomic groups which occurred in the initial dataset. Those with an asterisk were excluded from the analysis due to rarity (<1% frequency of occurrence).

Tricladida	Gomphidae*
Dendrocoelidae	Cordulegastridae
Planariidae/Dugesiidae	Aeshnidae
Mollusca	Corduliidae*
Neritidae	Libellulidae
Viviparidae	Hemiptera
Valvatidae	Mesoveliidae*
Hydrobiidae/Bithyniidae	Hydrometridae
Physidae	Gerridae
Lymnaeidae	Nepidae
Planorbidae	Naucoridae
Ancylidae/Acroloxidae	Aphelocheiridae
Unionidae	Notonectidae
Sphaeriidae	Pleidae*
Oligochaeta	Corixidae
Oligochaeta	Coleoptera
Hirudinea	Haliplidae
Piscicolidae	Hygrobiidae*
Glossiphoniidae	Gyrinidae
Hirudinidae*	Dytiscidae/Noteridae
Erpobdellidae	Hydrophilidae/Hydraenidae
Crustacea	Scirtidae
Astacidae	Dryopidae*
Asellidae	Elmidae
Corophiidae	Neuroptera
Gammaridae/Crangonyctidae/Niphargidae	Sialidae
Ephemeroptera	Trichoptera
Siphlonuridae*	Hydroptilidae
Baetidae	Rhyacophilidae/Glossosomatidae
Heptageniidae	Philopotamidae
Leptophlebiidae	Polycentropodidae
Potamanthidae*	Hydropsychidae
Ephemeridae	Psychomyiidae/Ecnomidae
Ephemerellidae	Phryganeidae
Caenidae	Brachycentridae
Plecoptera	Lepidostomatidae
Taeniopterygidae	Limnephilidae
Nemouridae	Goeridae
Leuctridae	Beraeidae
Capniidae*	Sericostomatidae
Perlodidae	Odontoceridae
Perlidae	Molannidae
Chloroperlidae	Leptoceridae
Odonata	Diptera
Platycnemidae	Tipulidae
Coenagrionidae	Simuliidae
Lestidae*	Chironomidae
Calopterygidae	

Table 2 Physicochemical variables and anthropogenic stresses recorded for each site

Time invariant map-derived variables	Abbreviation
Altitude at site (m asl)	ALT
Distance from source (km)	DFS
Slope (m km^{-1})	SLO
Long-term historical data	
Discharge category (1-10) (1= <0.31, 2= 0.31-0.62, 3= 0.62-1.25, 4= 1.25-2.5, 5= 2.5-5.0, 6= 5-10, 7= 10-20, 8= 20-40, 9= 40-80, 10= 80-160 m^3s^{-1})	DIS
Measured during site visits and averaged over the year	
Stream width (m) (mean of 3 seasonal measurements)	WW
Stream depth (cm) (mean of 3 seasonal measurements)	AvD
Substrate composition (% cover of clay/silt, sand, gravel/pebbles, cobbles/boulders converted to a mean particle size phi score) (mean of 3 seasonal measurements)	Substrat
Alkalinity (mg l^{-1} $CaCO_3$) (mean of 12 monthly measurements)	ALK
Anthropogenic stress variables	
Organic inputs	OrgInp
Acidification	Acidific
Reduced discharge	ReduDisc
Canalisation	Channeli
Agricultural chemical inputs	AgriChem
Sedimentation	Sediment
Riparian habitat modifications	RipaModi
Industrial discharge and run-off	IndDisRu
Urban run-off	UrbanRun
Excessive instream plant growth	PlantGro
Consolidated banks	ConsBank
Intensive arablisation	IntsArab

Table 3 Summary results of CCA for 71 taxa against 23 explanatory variables.

Axes	1	2	3	4	Total Inertia
					1.174
Eigenvalues	0.190	0.044	0.024	0.015	
Species-environment correlations	0.893	0.707	0.663	0.615	
Cumulative percentage variance:					
of taxa data	16.2	19.9	22.0	23.3	
of taxa-explanatory variables relation	61.8	76.1	84.0	89.0	
Sum of all eigenvalues					1.174
Sum of all canonical eigenvalues					0.307

Table 4 The individual explanatory power of each variable (marginal effect) and the additional contribution of each successive variable to the forward selected model (conditional effect). Variables are ranked in importance by their conditional effect. All variables were statistically significant (*P*<0.001, Monte Carlo permutation test)

Variable	Marginal Effects	Conditional Effects	
	λ_1	λ_a	cumulative λ_a
Substrate	0.12	0.12	0.12
Alkalinity	0.11	0.05	0.17
Discharge Category	0.02	0.02	0.19
Y (northing)	0.02	0.02	0.21
Urban Run-off	0.03	0.02	0.23
Altitude	0.05	0.01	0.24
Depth	0.06	0.01	0.25
X (easting)	0.10	0.01	0.26
X^2	0.01	0.01	0.27
X^2Y	0.01	0.01	0.28
XY^2	0.05	0.01	0.29
Organic Inputs	0.04	0.01	0.30
Reduced Discharge	<0.01	0.01	0.31
Distance from source	0.03	<0.01	0.31
Slope	0.05	<0.01	0.31
Width	0.02	<0.01	0.31
XY	0.02	<0.01	0.31
Y^2	0.02	<0.01	0.31
Acidification	0.02	<0.01	0.31
Canalisation	0.02	<0.01	0.31
Agri-chemical inputs	0.01	<0.01	0.31
Industrial discharge and run-off	0.01	<0.01	0.31
Excessive plant growth	0.01	<0.01	0.31

Figure Legends

Figure 1 The location of the 5752 sites from the Environment Agency 1995 GQA survey analysed in this study.

Figure 2 Position of taxa in ordination space as defined by axes 1 and 2 of the canonical correspondence analysis for 71 taxa and 23 variables. The CANOCO short code is the first 8 letters of each taxon name (see Table 1 for full taxon names).

Figure 3 The relative importance and direction of influence of the 23 variables in the canonical correspondence analysis ordination space (see Table 2 for key to abbreviations).

Figure 4 Site scores from canonical correspondence analysis with 23 explanatory variables illustrating the geographical position of sites and their position along the (a) axis 1 and (b) axis 2 ordination gradients.

Figure 5 Mantel correlogram for spatial autocorrelation in macroinvertebrate communities. Dark circles represent significant correlations between biological distance and geographical distance (P<0.00625, following Bonferroni correction for multiple testing).

Figure 6 Partitioning of the total variance explained (TVE) into its constituent variable groups using canonical correspondence analysis.

Figure 7 Site scores from a 'pure' spatial canonical correspondence analysis with 7 spatial variables (physicochemical and stress variables as covariables), illustrating the geographical position of sites and their position along the (a) axis 1 and (b) axis 2 ordination gradients.

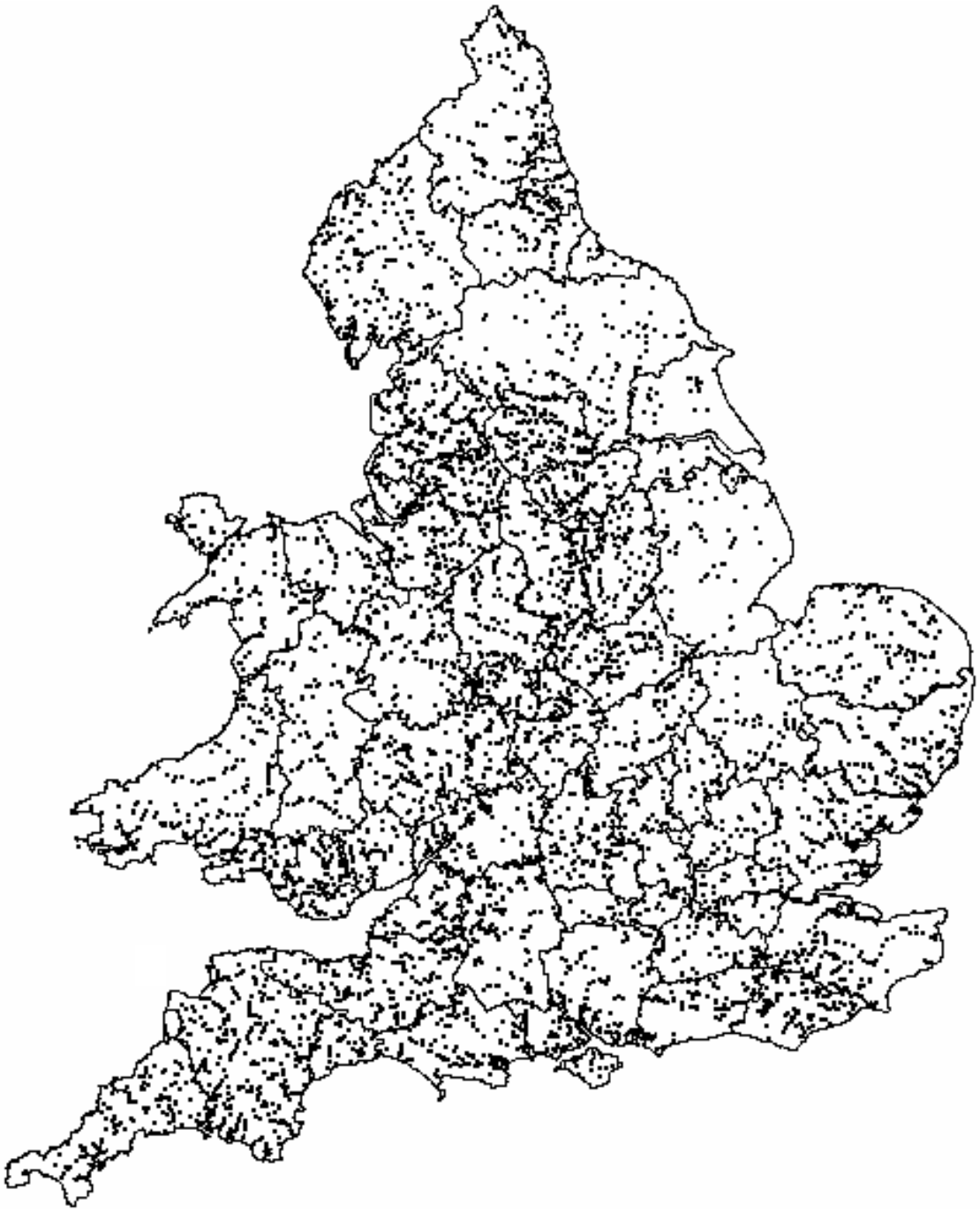

Figure 1

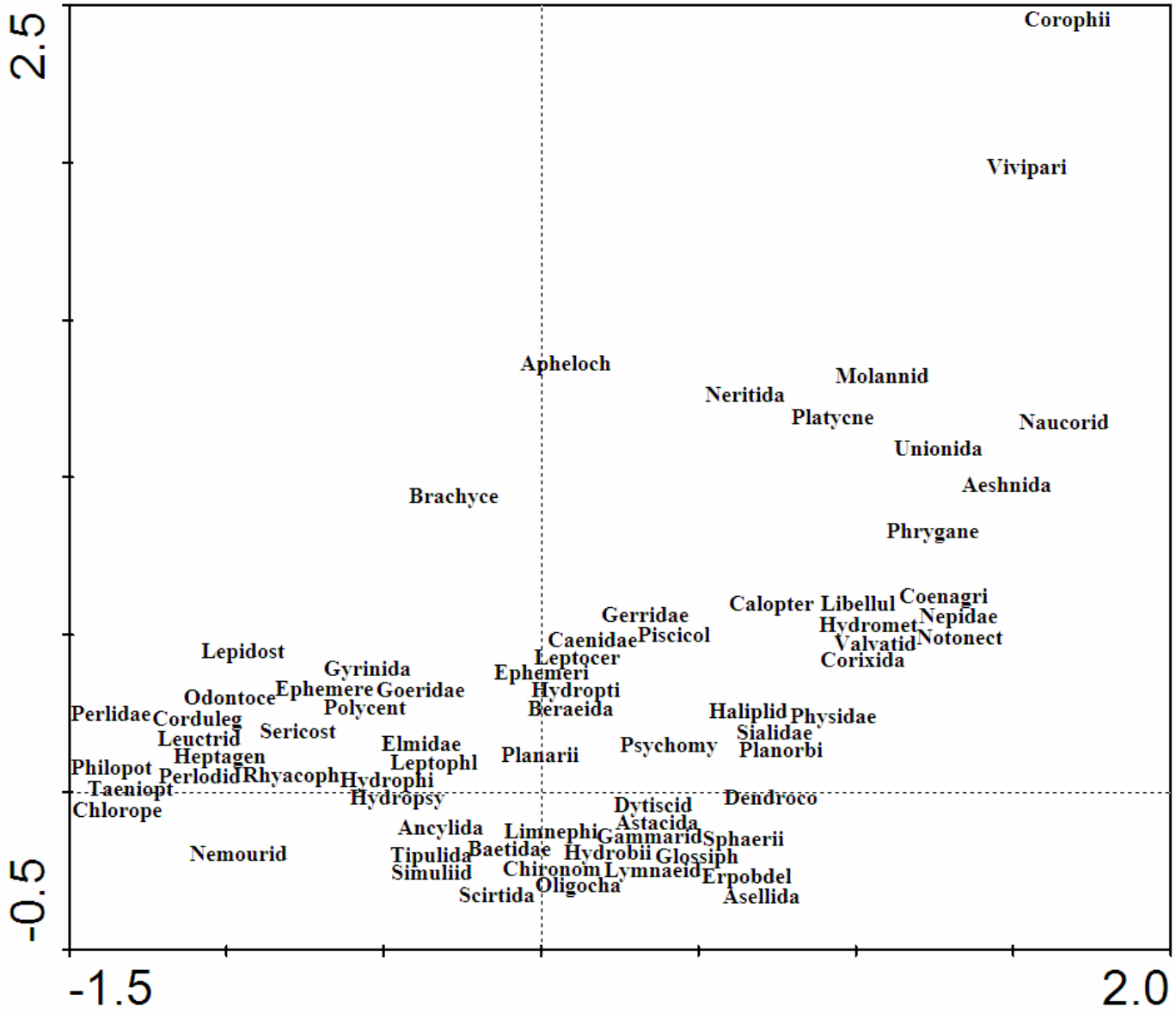
2.5
-0.5
-1.5
2.0
Corophii
Vivipari
Apheloch
Molannid
Neritida
Platycne
Naucorid
Unionida
Aeshnida
Brachyce
Phrygane
Calopter
Libellul
Coenagri
Gerridae
Nepidae
Hydromet
Caenidae
Piscicol
Notonect
Valvatid
Lepidost
Leptocer
Corixida
Gyrinida
Ephemeri
Ephemere
Goeridae
Hydropti
Odontoce
Polycent
Beraeida
Haliplid
Physidae
Perlidae
Corduleg
Sericost
Sialidae
Leuctrid
Elmidae
Psychomy
Planorbi
Heptagen
Planarii
Philopot
Leptophl
Perlodid
Rhyacoph
Hydrophi
Taeniopt
Hydropsy
Dendroco
Chlorope
Dytiscid
Astacida
Ancylida
Limnephi
Gammarid
Sphaerii
Nemourid
Tipulida
Baetidae
Hydrobii
Glossiph
Simuliid
Chironom
Lymnaeid
Erpobdel
Oligocha
Scirtida
Asellida

Figure 2

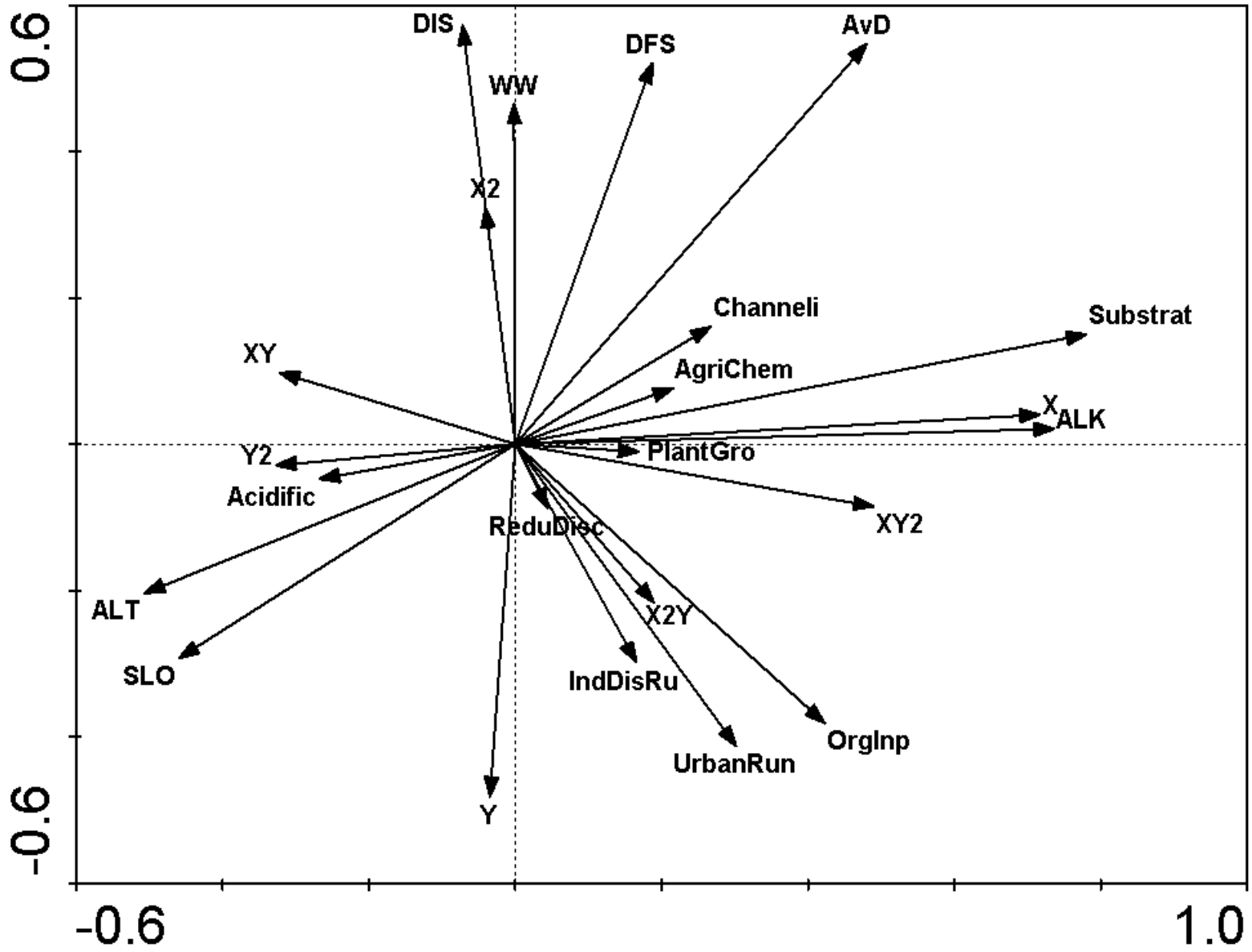
0.6
-0.6
-0.6
1.0
DIS
DFS
AvD
WW
X2
Channeli
Substrat
XY
AgriChem
X
ALK
Y2
PlantGro
Acidific
XY2
ReduDisc
ALT
X2Y
SLO
IndDisRu
OrgInp
UrbanRun
Y

Figure 3

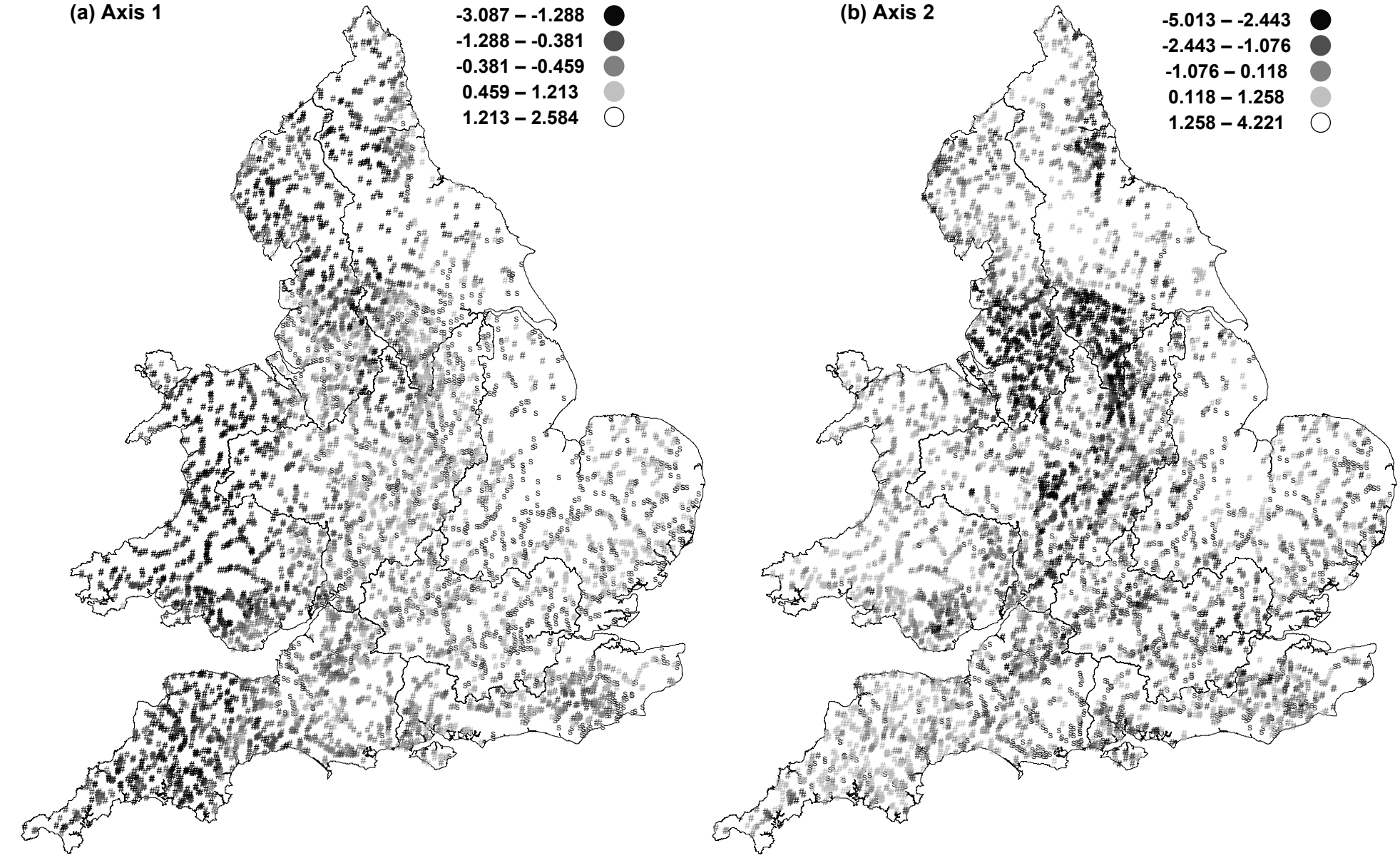
(a) Axis 1
-3.087 – -1.288
-1.288 – -0.381
-0.381 – -0.459
0.459 – 1.213
1.213 – 2.584
(b) Axis 2
-5.013 – -2.443
-2.443 – -1.076
-1.076 – 0.118
0.118 – 1.258
1.258 – 4.221

Figure 4

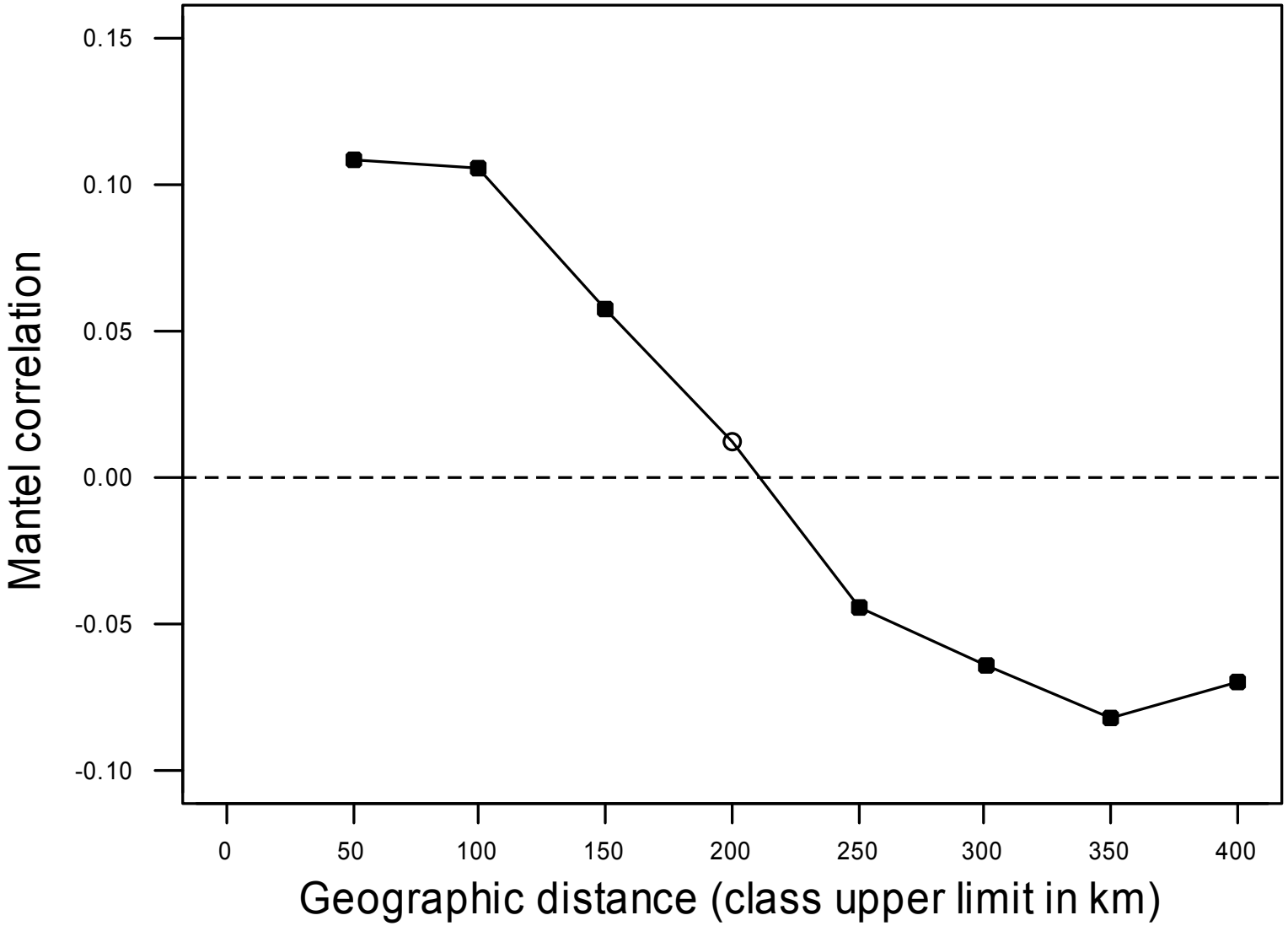
0.15
0.10
0.05
0.00
-0.05
-0.10
Mantel correlation
0
50
100
150
200
250
300
350
400
Geographic distance (class upper limit in km)

Figure 5

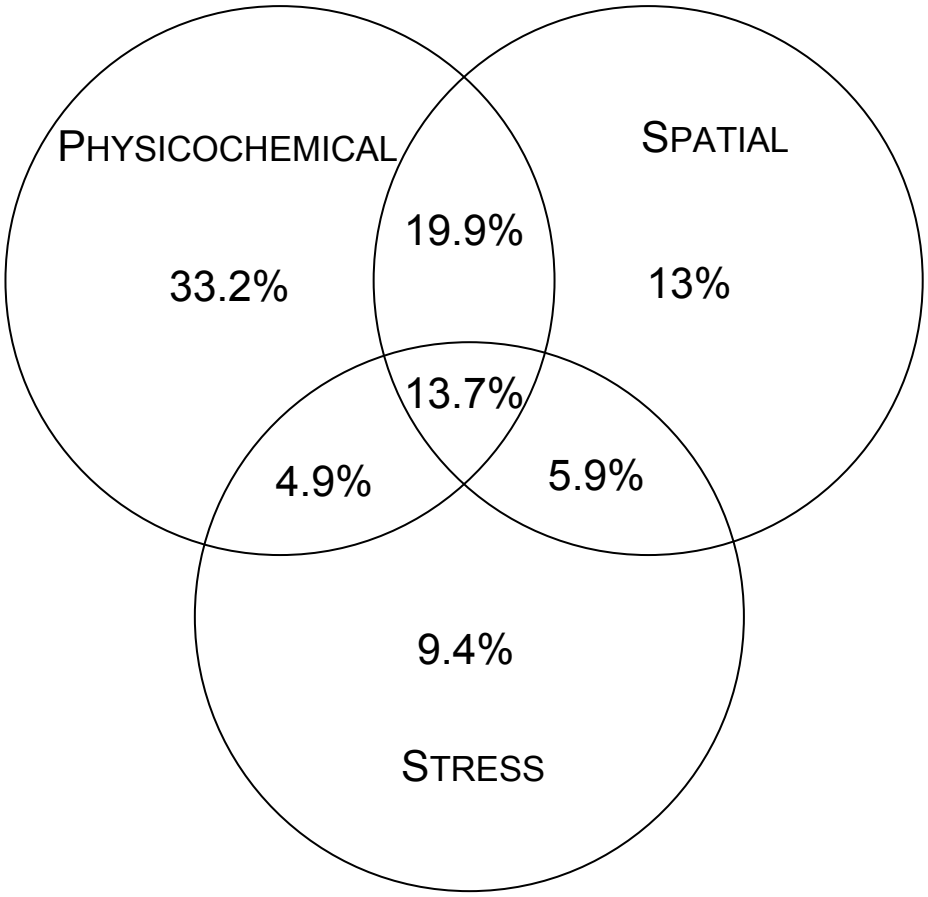

Figure 6

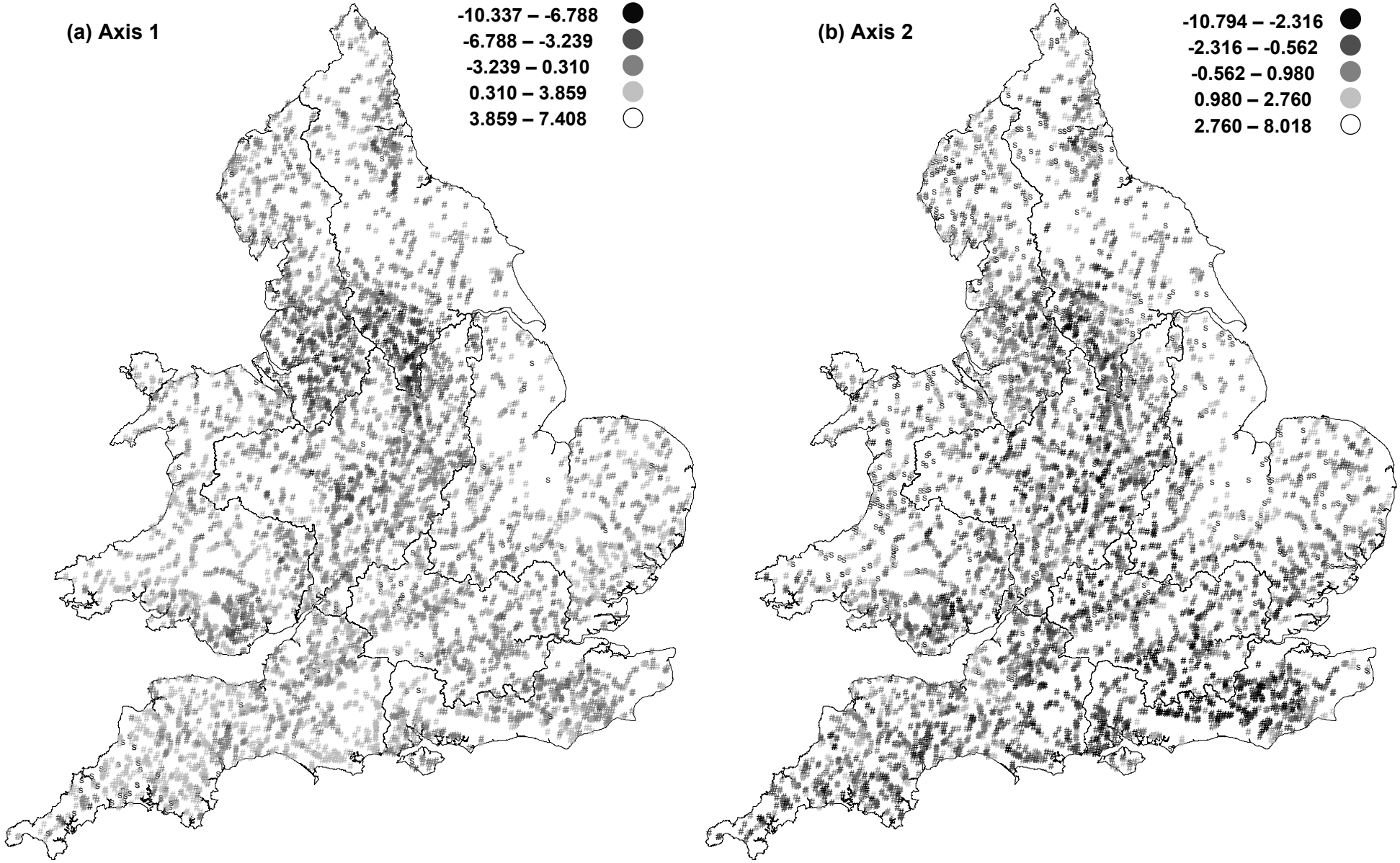
(a) Axis 1
-10.337 – -6.788
-6.788 – -3.239
-3.239 – 0.310
0.310 – 3.859
3.859 – 7.408
(b) Axis 2
-10.794 – -2.316
-2.316 – -0.562
-0.562 – 0.980
0.980 – 2.760
2.760 – 8.018

Figure. 7