

Article (refereed)

Evans, Chris D.; Cooper, David M.; Monteith, Donald T.; Helliwell, Rachel C.; Moldan, Filip;
Hall, Jane; Rowe, Edwin C.; Cosby, Bernard J.. 2010 Linking monitoring and modelling: can
long-term datasets be used more effectively as a basis for large-scale prediction?

Biogeochemistry, 101. 211-227. [10.1007/s10533-010-9413-x](https://doi.org/10.1007/s10533-010-9413-x)

© Springer Science+Business Media 2010

This version available <http://nora.nerc.ac.uk/8706/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the authors and/or other rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

This document is the author's final manuscript version of the journal article, incorporating any revisions agreed during the peer review process. Some differences between this and the publisher's version remain. You are advised to consult the publisher's version if you wish to cite from this article.

The original publication is available at www.springerlink.com

Contact CEH NORA team at
noraceh@ceh.ac.uk

Linking monitoring and modelling: can long-term datasets be used more effectively as a basis for large-scale prediction?

Chris D Evans^{1*}, David M Cooper¹, Donald T Monteith², Rachel C Helliwell³, Filip Moldan⁴, Jane Hall¹, Edwin C Rowe¹, and Bernard J. Cosby⁵

¹) CEH Bangor, Environment Centre Wales, Deiniol Road, Bangor, Gwynedd LL57 2UP, UK, cev@ceh.ac.uk

²) CEH Lancaster, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4AP, UK

³) Macaulay Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK.

⁴) IVL Swedish Environmental Research Institute, P.O. Box 5302, 400 14 Gothenburg, Sweden

⁵) Department of Environmental Sciences, University of Virginia, Charlottesville, VA22901 USA

Abstract

Data from long-term monitoring sites are vital for biogeochemical process understanding, and for model development. Implicitly or explicitly, information provided by both monitoring and modelling must be extrapolated in order to have wider scientific and policy utility. In many cases, large-scale modelling utilises little of the data available from long-term monitoring, instead relying on simplified models and limited, often highly uncertain, data for parameterisation. Here, we propose a new approach whereby outputs from model applications to long-term monitoring sites are upscaled to the wider landscape using a simple statistical method. For the 22 lakes and streams of the UK Acid Waters Monitoring Network (AWMN), standardised concentrations (Z scores) for Acid Neutralising Capacity (ANC), dissolved organic carbon, nitrate and sulphate show high temporal coherence among sites. This coherence permits annual mean solute concentrations at a new site to be predicted by back-transforming Z scores derived from observations or model applications at other sites. The approach requires limited observational data for the new site, such as annual mean estimates from two synoptic surveys. Several illustrative applications of the method suggest that it is effective at predicting long-term ANC change in upland surface waters, and may have wider application. Because it is possible to parameterise and constrain more sophisticated models with data from intensively monitored sites, the extrapolation of model outputs to policy relevant scales using this approach could provide a more robust, and less computationally demanding, alternative to the application of simple generalised models using extrapolated input data.

Keywords: Modelling, Long-term monitoring, Catchments, Upscaling, Acid Neutralising Capacity, Dissolved Organic Carbon, Sulphate, Nitrate

Introduction

Headwater catchments incorporate a major proportion of the total stream length of many river systems (Bishop et al., 2008), and are intimately linked to the landscapes they drain. Monitoring in headwater systems thus provides a uniquely long-term insight into the biogeochemical and ecological processes that govern natural ecosystems, their condition, and their response to anthropogenically imposed change. Long-term monitoring is, however, challenging to maintain in the face of uncertain funding, changing policy priorities, and reliance on the sustained commitment of small numbers of individuals (Nisbet, 2007). While the importance of monitoring is increasingly recognised by scientists and policymakers, there is often little connection between information obtained from detailed catchment studies, and the whole-landscape (national or international) scale at which policy decisions must be taken. This problem is accentuated where catchment studies have developed on an individual, *ad hoc* basis, often at scientifically interesting but atypical locations, and without consistency of measurements or methods between sites. Recognition of this issue has led, in some countries, to the establishment of centrally coordinated national monitoring networks. In the UK, 22 existing and new monitoring catchments were combined into the UK Acid Waters Monitoring Network (Monteith and Evans, 2005), with standardised measurements and protocols, consistent sampling dates, and centralised analyses providing a greater degree of representivity and comparability (Patrick et al, 1991). Similar networks have been established in other countries, such as the 14 sites of the GEOMON network in the Czech Republic (Fottová, 1995). At the international level, data from sites within national monitoring programmes have been combined into international programmes, with the aim of providing scientific support to decision-making at this scale. For example, the UNECE Convention on Long-Range Transboundary Air Pollution (CLRTAP, Sliggers and Kakebeeke, 2004), which is responsible for international agreements to abate atmospheric sulphur, nitrogen, ozone and particulates, established a number of International Cooperative Programmes (ICPs) to collate national monitoring data, including (for catchments) the ICP on Integrated Monitoring and ICP Waters (e.g. Skjelkvåle et al., 2005).

Despite these drives towards integration and aggregation of monitoring, the gap between monitoring and decision-making remains. Taking both the UK and CLRTAP examples, decisions on pollution abatement are ultimately made on the basis of mapped estimates of ecosystem sensitivity and exposure to damaging pollutant levels. These are derived from

calculated (static) critical loads models, or more recently from dynamic biogeochemical models (Hettelingh et al., 2008). In either case, the approach used is to combine limited measurement data with extrapolated parameter values, and then to apply the model to very large numbers of locations. In the UK, for example, freshwater critical loads are calculated for 1752 surface waters where a chemical measurement exists, and terrestrial critical loads are calculated for 1-8 habitats within approximately 240,000 1 km grid squares, based on default parameter sets (Hall et al., 2004). Dynamic model applications have followed the same general structure, using the MAGIC model (Cosby et al., 2001) for surface waters, and the VSD model (Posch and Reinds, 2009) for soils.

Thus, policy decisions are made based on very limited and often highly extrapolated input parameters for critical loads or dynamic models. Opportunities for evaluating predictions against observations are limited, and even where (as for surface water MAGIC simulations) models have been calibrated against observations, these are often single time points (e.g. from synoptic surveys), so that it is not possible to gauge whether the simulated trajectory of historic change adequately reproduces observations.

In this context, long-term monitoring performs several roles. Firstly, intensive catchment studies have in many cases provided the process insight and detailed measurements to support model development (e.g. Cosby et al., 1985; Tipping, 1996; Gbondo-Tugbawa et al., 2001). Secondly, they have been used to test the outputs from models applied at large spatial scales against the small subset of sites with long-term observations (e.g. Forsius et al., 1998; Jenkins et al., 2001; Ouhlehle et al., 2007; Reinds et al., 2009). Thirdly, monitoring sites have been used to undertake more complex simulations, for example incorporating additional environmental drivers and scenarios for land management and climate (e.g. Evans, 2005; Beier et al., 2003; Aherne et al., 2008; Sjøeng et al., 2009; Futter et al., 2009); detailed carbon and nitrogen cycles (e.g. Belyazid et al., 2006; Futter et al., 2009); and more sophisticated methods for parameter and uncertainty estimation such as Bayesian calibration (Larssen et al., 2007; Reinds et al., 2008). While much of this work has appeared in the scientific literature, and may indirectly have supported the development and application of models at larger scales, the higher data demands of more complex model applications preclude their parameterisation to a wider set of data-poor locations. In other words, there is currently little direct linkage between the detailed time series datasets obtained from catchment monitoring

studies, and the simplified catchment models that ultimately underpin national and international-level policy development.

We propose that this disconnection between small-scale monitoring and large-scale modelling could potentially be removed, via an alternative approach to the use of monitoring data in models. Rather than attempt to use monitoring data to estimate the parameters of detailed process models, recent approaches have empirically related monitoring data to simple readily available regional scale data, such as catchment characteristics (Cooper et al., 2004). Simple relationships between water quality and landscape class have then been used for spatial extrapolation. This approach relies on the spatial coherence of hydrochemical response for locations with the same landscape. A similar approach may be taken to temporal interpolation; just as hydrochemical conditions are dependent on spatially varying catchment characteristics, they are also dependent on temporally varying external drivers such as climate and deposition. Long term monitoring datasets provide local measurements of temporal variability over time, but extrapolation to regional scales represents a significant challenge.

Since critical loads based methods to support pollution abatement policies first evolved in the 1980s, long-term monitoring datasets have increased greatly in number and duration. Furthermore, comparisons of these datasets have often shown remarkable temporal coherence among sites across wide gradients of geographical location, catchment characteristics, exposure to air pollutants, climate and land management (Evans and Monteith, 2001; Watmough et al, 2004; Davies et al., 2005; Monteith et al., 2007). The implication is that headwater catchments are responding in a consistent fashion to a set of external environmental drivers, whose effect can be inferred from responses even if not explicitly identified. Here, we utilise this observed temporal coherence as the basis of a method for analysing and predicting chemical change at one site, based on observed change at other sites. We then extend this approach, in order to extrapolate modelled changes from a small number of monitoring sites to predict changes at unmonitored sites, as an alternative to modelling these sites directly. We present examples of the application of this approach, discuss its advantages and disadvantages compared to direct modelling, and consider its potential utility for upscaling from a limited set of monitoring catchments to larger and more policy-relevant spatial scales.

Methods

Site description

We based our study on the 22 catchments of the UK Acid Waters Monitoring Network (AWMN, Figure 1). The network comprises 11 lakes and 11 streams, distributed across upland areas of the UK, which was initiated in 1988. The sites span a wide range of anthropogenic and marine ion deposition, land cover (5 catchments contain plantation conifer forest, the remainder comprise acid grassland and/or heathland and blanket bog), soil type (including peats, podzols, gleys and rankers), climate (range of mean temperatures 3 to 10 °C, mean annual rainfall 0.87 to 3.5 m), and topographic position (altitude of sampling site 10 to 785 m) (Patrick et al., 1991; Monteith and Evans, 2000). Lakes are sampled at a consistent quarterly frequency, and streams monthly to reflect their greater short-term variability. Analyses follow strict protocols to ensure intra- and inter-site comparability (Patrick et al., 1991) and have been subjected to rigorous analytical quality control comparisons (Gardner, 2008). All samples are now analysed at either the Centre for Ecology and Hydrology, Lancaster, or at the Freshwater Fisheries Laboratory, Pitlochry. Analysis is based on 19 years of full chemical data, from 1989 to 2007.

Time series standardisation

Previous time series analyses of the AWMN dataset (e.g. Evans and Monteith, 2001; Davies et al., 2005) have shown that, despite substantial differences in the mean and variance of most chemical variables between sites, underlying patterns of temporal variation in rescaled (or standardised) data often show strong similarities. To demonstrate these underlying patterns, measured concentrations at each sampling site s and date t (C_{st}) were standardised over time, by subtracting the site mean for the whole time series (\bar{C}_s) and dividing by the site standard deviation (σ_s) so that

$$Z_{st} = \frac{C_{st} - \bar{C}_s}{\sigma_s} \quad (1)$$

The time series of standardised concentrations (Z_{st} referred to as ‘Z scores’) thus has a mean of zero and a standard deviation of 1. Exploratory analysis suggests that this standardisation leads to new series which show temporal consistency following an approximate model

$$Z_{st} = \mu_s + x_t + \varepsilon_{st} \quad (2)$$

where x_t is a temporal effect which is independent of location and ε_{st} is a zero mean uncorrelated error term with variance independent of both location and time. Such a model implies that temporal effects are essentially multiplicative, so that sites with generally high concentrations will tend to have high variability over time in absolute but not in relative terms. This effect is to some extent moderated by the presence of a sample mean in the definition of the standardised variable.

The behaviour of x_t might be further modelled as, for example

$$x_t = \mu_x + \beta t + \eta_t \quad (3)$$

Equation (3) might be appropriate if the time effect were following a linear trend. The combination of equations (2) and (3) constitutes a hierarchical linear model which could be formally analysed as such. The model description presented here is intended to provide a general view of methodology, rather than focusing on a detailed statistical analysis. For this study, we simplified the chemical time series further, by first taking an annual mean of all samples collected within each year (so that t in equation (2) represents years), and then undertaking the standardisation procedure as above based on annual rather than individual sample values. The resulting time series of annual mean Z scores permits direct comparison between quarterly-sampled lakes and monthly-sampled streams, and is directly comparable to the annual time step at which many biogeochemical models, such as MAGIC and VSD, typically operate.

Finally, we calculated median annual Z scores for each year across a set of sites (either the all sites, or a defined subset). Again, this approach was used previously by Evans and Monteith (2001) and Davies et al. (2005), but based on individual sampling dates and for lake and

stream subsets. 10th and 90th percentile Z scores were also calculated, providing evidence of the stability of the variance of ε between years.

For this study, we focused on a key water quality variable, Acid Neutralising Capacity (ANC), which is widely used for modelling and monitoring assessments. While alternative formulations exist, the preferred method for calculating ANC in the AWMN is based on measured (Gran) alkalinity (Alk), inorganic Al concentration (Al_{inorg}) and an estimate organic acid concentration (OA) based on a conversion from dissolved organic carbon (DOC) concentration (Evans et al., 2001; Equation (4)).

$$ANC = \text{Gran Alkalinity} + OA - Al_{inorg} \quad (4)$$

The same methodology was also applied to calculate Z scores for DOC, sulphate (SO_4), and nitrate (NO_3). Median annual Z scores were calculated for all 22 AWMN sites, and for subsets of lakes/streams, and moorland/forest catchments.

Extrapolating from monitored to unmonitored sites

The fundamental assumption in extrapolating to a new site is that it follows a coherent temporal pattern to another site, or set of sites as indicated in equation (2). The time series of Z scores for this site, or group median Z score for a set of sites, can thus be directly back-transformed to give estimates of C_{st} for the new site according to equation (5), and following the notation of equation (1).

$$C_{st} = \bar{C}_s \times Z + \sigma \quad (5)$$

In the case that \bar{C}_s and σ are known, this calculation is straightforward (where Z scores are calculated for annual mean data, \bar{C}_s and σ are the mean and standard deviation of the annual means, rather than the raw data). At most unmonitored sites, however, it is unlikely that these statistics will be known, and they must therefore be estimated. At many sites in the UK Freshwater Critical Loads dataset, water quality data exist from two or more separate time points: for example, the Welsh Acid Water Survey of over 100 surface waters was repeated in

1985 and 1995 (Stevens et al 1997), while a set of 48 lochs in Galloway, Southwest Scotland, has been sampled on seven occasions between 1979 and 1998 (Ferrier et al. 2001). For sites such as these for which C_{st} and Z_{st} are available at two or more time points $t = \dots$ a sequence of equations of type (5) can be used to solve for \bar{C}_s and σ , either using two simultaneous equations, or statistically. We examined heuristically whether we could accurately predict long-term ANC mean and standard deviation from samples collected at two time points based on the AWMN sites. In general, we found that a good estimate of long-term mean could be obtained from the mean of any two sampling years (e.g. Figure 2a, for the years 2007 and 1992), and that a reasonable estimate of the standard deviation could be obtained from a regression of the variance against the squared difference between the two samples (e.g. Figure 2b). We assumed that the regression equations obtained by this method were applicable to other sites where samples had been collected in the same years.

Extrapolating from modelled to unmodelled sites

The method described permits the extrapolation of measured water quality data from a monitored site to a site at which two sets of synoptic survey data are available. In principle, the same approach may be applied to modelled data, by substituting measured annual means with output values from a biogeochemical model applied to the monitored site. The potential advantages of this approach over direct application of the same model to the unmonitored site are i) that data rich monitoring sites can support a more accurate and/or sophisticated model parameterisation, and ii) that model outputs can be tested or constrained (e.g. Larssen et al., 2007) against observed time series, rather than simply calibrated to a single point in time.

For this study, we undertook two illustrative model output extrapolations based on previous, well-parameterised and tested applications of the MAGIC model to monitoring sites. For the first example, we used a calibration of MAGIC to the Afon Gwy, a Mid-Wales stream that has been monitored by the Centre for Ecology and Hydrology since 1980. MAGIC was applied with a two soil layer structure, detailed parameterisation using local measurements, and calibration against both soil solution and stream chemical data (Evans et al., 2008). Outputs from the model were used to predict ANC change at Llyn Llagi, an AWMN lake catchment located 65 km to the north, in the Snowdonia region of North Wales. Predicted ANC changes at Llyn Llagi were compared both to observations from the site, and to

predictions from a simpler direct calibration of MAGIC to Llyn Llagi as part of a regional suite of MAGIC applications. This application was based on a single soil layer, parameterisation data extrapolated from national or regional datasets and calibration to a single year of water chemistry data (Evans et al., 2007).

Secondly, we took a recent two soil layer calibration of MAGIC to the River Etherow catchment in the South Pennines, England (Helliwell et al., in prep.) and used this to predict ANC change for a suite of 27 sites within the region which were sampled in 1998 (Evans et al., 2000) and again in 2001-02 (Helliwell et al., 2007). Predictions were again compared against direct MAGIC simulations to all sites, which formed part of the regional model application described above.

Results

Standardised annual time series

Untransformed annual mean ANC, SO₄, NO₃ and DOC time series for the individual AWMN sites (Figure 3) highlight the extent of between-site heterogeneity in terms of mean concentrations, rates of change, and short-term variability. Standardisation, on the other hand (Figure 4), highlights the degree of underlying temporal coherence across the network. For SO₄, there was a very pronounced decline in concentrations during the late 1990s, and high coherence between sites (mean annual 10th-90th percentile range 1.03 standardised units). ANC showed a consistent increase, with the fastest rate of change coinciding with the period of SO₄ decrease, and a reasonable level of coherence (mean 10th-90th percentile range 1.54). In contrast, there has been no clear trend in NO₃ concentrations, and relatively high variability among sites (mean 10th-90th percentile range 2.04). However, all sites showed a clear NO₃ peak in 1996. DOC has increased at all sites, with low variability between sites (mean 10th-90th percentile range 1.25). There was some indication of a reversal in the rising trend from 2001-04, but concentrations subsequently rose to unprecedented levels during 2006-07.

To a large extent, these observations reinforce previous inferences from analyses of monthly stream and quarterly lake data, namely that reduced S inputs to UK upland surface waters have led to a large reduction in runoff SO₄ concentrations, partial recovery from acidification

as evidenced by rising ANC (e.g. Davies et al., 2005), and an apparently associated increase in the leaching of DOC (e.g. Evans et al., 2006). Temporal variations in NO_3 leaching appear to be mainly linked to climate fluctuations, such as occurrence of soil freezing, rather than directly to deposition (Monteith et al., 2000; Cooper, 2005). The relatively low temporal coherence in NO_3 between sites suggests that local factors such as land-management (notably tree maturation, felling and replanting in forested catchments), as well as N saturation status, may also have influenced NO_3 leaching.

Reducing time series to annual mean values tends to highlight underlying trends rather than short-term variations, particularly for streams (cf. Figure 1 of Davies et al. 2005). It also permits direct comparisons among sites sampled at different frequencies, and between different site types. Here, we compared lakes and streams, and forested and moorland catchments. Studies in North American and Scandinavian lakes have shown a substantial role of lakes in the processing of carbon and nutrients (e.g. Dillon and Molot, 2005; Algesten et al., 2003). UK lakes in high rainfall upland areas typically have shorter residence times than those in continental areas, and here their impact on long-term trends is less clear. On the other hand, there is strong evidence that catchment afforestation of parts of the UK (primarily exotic conifer plantation on moorland during the post-war period) significantly amplified the effects of acid deposition, due to increased base cation uptake and pollutant scavenging from the atmosphere by the forest canopy (NEG-TAP, 2001). Evidence of intensified current damage has led to predictions that afforested catchments may not respond to reduced acid deposition in the same way as moorland catchments, resulting in suppressed recovery or even ongoing acidification (e.g. Jenkins et al., 1990; Helliwell et al., 2001).

Our analysis shows, for all variables analysed, a tight correlation between median annual Z scores for the two pairs of catchment subsets (Figure 5). No clear deviation is apparent between ANC Z scores for forested and moorland catchments that might support the conclusion that forests are stopping recovery from acidification. However, because Z scores do not permit comparison of the magnitude of change at individual sites, this observation does not exclude the possibility that moorland and forest catchments are recovering at different rates. Similarly, neither the presence of lakes or forests appears to have influenced the trajectory of DOC or NO_3 leaching, although differences in rate of change again cannot be discounted on this basis alone. Overall, the data do not support the subdivision of the AWMN sites according to catchment or surface water type. We were also unable to detect

sufficient evidence of geographical clustering in the pattern of Z score variations to justify subdividing data into regional subgroups (see below).

Extrapolating from monitored to unmonitored sites

Based on the analysis above, suggesting no strong geographic or typological clustering of Z score variations among the sites, we henceforth treated all AWMN sites as a single group. For ANC, we then attempted to predict annual ANC concentrations at each site, based on back-transformation of median annual Z scores for the remaining 21 sites. Equation (5) was applied, firstly based on known mean and standard deviation values, and then on values estimated from only two years of data, 1992 and 2007 (Figure 2).

Predictions of ANC based on known mean and standard deviation had root mean square error (RMSE) values ranging from 3.3 to 22.6 $\mu\text{eq l}^{-1}$ (Figures 6,7). With the exception of the few sites showing little overall ANC trend (e.g. Loch Coire nan Arr, Narrator Brook), annual ANC predictions captured the underlying trend at most sites. The method was most successful at capturing inter-annual trends at sites in Southwest Scotland (Round Loch of Glenhead, Loch Grannoch, Dargall Lane), Northwest England (Scoat Tarn, Burnmoor Tarn), North Wales (Llyn Llagi) and the Eastern part of Northern Ireland (Beaghs Burn, Bencrom River). These sites, all of which are clustered around the Irish Sea (Figure 1), show particularly coherent behaviour and thus dominate the median Z score calculation for the dataset as a whole. Prediction of inter-annual variations becomes poorer with distance from this central cluster, as might be expected given the increasing heterogeneity of short-term climatic and depositional patterns across larger spatial scales. We were unable to identify other clear regional clusters, although it is likely that these would emerge given additional monitoring datasets in (for example) Northern Scotland or Southern England.

The second method, in which mean and standard deviation were estimated from two observation years, provides a more meaningful test of the utility of this method for extrapolating from monitored to unmonitored sites. ANC predictions based on 1992 and 2007 'survey' years (Figure 6) resulted in only small increases in RMSE at most sites (Figure 7). With few exceptions, this method thus remained effective in reproducing the trend in annual mean ANC.

Extrapolating from modelled to unmodelled sites

Upscaling to a single site: For the Llyn Llgi AWMN site, ANC was modelled directly in MAGIC based on regional-scale parameter data (Figure 8a), and indirectly based on annual Z scores derived from a site-specific MAGIC application to the Afon Gwy (Figure 8b), which were back-transformed using known (Figure 8c) and estimated (Figure 8d) mean and standard deviation values. The direct regional MAGIC application to Llyn Llgi captures the approximate trend in ANC, but (since the model application did not incorporate short-term variations in driving variables) does not reproduce the short-term variation. RMSE over the period of observations is $7.0 \mu\text{eq l}^{-1}$. Modelling the site indirectly using the Z score approach gave no discernable reduction in model performance (RMSE 7.3 and $6.4 \mu\text{eq l}^{-1}$ based on known and estimated mean and standard deviation respectively), and appeared to capture both the trend and some of the between-year variation in ANC. Comparing long-term simulations, the indirect simulations suggest a slightly greater responsiveness of the site to changing atmospheric deposition, with pre-industrial 1850 ANC values of around $70 \mu\text{eq l}^{-1}$ compared to $56 \mu\text{eq l}^{-1}$ in the direct application, and forecast 2050 ANC of around $45 \mu\text{eq l}^{-1}$ compared to $37 \mu\text{eq l}^{-1}$ in the direct application.

Upscaling to multiple sites: Simulated ANC for the South Pennine reservoirs is shown in Figure 9, modelled directly by multiple MAGIC applications to all sites, and indirectly by Z score transformation from a single site-specific MAGIC application to the River Etherow AWMN site. Agreement between the two modelling methods is generally good, apart from at the 1970 peak of acid deposition, when the indirect simulation suggests a greater degree of acidification. Without historic data against which to compare, the relative accuracy of the two hindcast simulations cannot be evaluated. However in forecast, it is notable that the two simulations converge; for 2050, median predicted ANC for the regional dataset is $51 \mu\text{eq l}^{-1}$ based on the Z score approach (11 sites having ANC below the critical limit of $20 \mu\text{eq l}^{-1}$), compared to a median of $57 \mu\text{eq l}^{-1}$ based direct MAGIC application (9 sites with ANC $< 20 \mu\text{eq l}^{-1}$). Thus, in this example, the two methods would lead to similar interpretation of future regional water quality status.

Discussion

Despite considerable typological and geographical variation, the 22 AWMN sites show high temporal coherence in a range of key water quality variables, including SO_4 , DOC, ANC and - to a lesser extent - NO_3 . Standardisation of data to annual mean Z scores appears to be an effective method for comparing the behaviour of sites with contrasting absolute concentrations, short-term variability and sampling frequency. Applications of this approach include comparison and differentiation of sites according to typology or location, potentially providing insight into the processes controlling water quality variation across the landscape. Our analysis suggested little differentiation in water quality variations between sites according to catchment type; we found no impact of lake presence on long-term DOC variations, or of plantation forests on ANC variations. The Z score approach also has potential utility for identifying individual sites showing atypical behaviour (e.g. due to local management or pollution factors), and for identifying individual analytical outliers. The method has been tested here on data from semi-natural upland ecosystems, and it is doubtful whether the same approach would be as effective in more heterogeneous agricultural or urban-influenced areas, where diffuse and point-source pollution are likely to be more catchment-specific. Nevertheless, analysis of temporal coherence in these systems might help in the identification of underlying signals, such as from climate or large-scale land-management change, which may not be discernable from the analysis of individual sites.

There are several prerequisites to using the Z score approach for model upscaling. Firstly, it is vital that long-term monitoring data are available for a sufficient number of sites to identify temporal coherence, derive average Z scores, and to allow subdivision according to typological or geographical factors where these lead to differing biogeochemical behaviour. This is greatly facilitated by consistently operated networks such as the AWMN, although by reducing data to annual means it may be possible to combine data from independently operated sites and/or multiple networks. Secondly, the method will only be applicable for areas or ecosystems within which coherent behaviour can be demonstrated; it will become progressively poorer with increasing geographical distance, or typological difference, from the sites used to calculate median Z scores. Thirdly, to predict change at an unmonitored site it is necessary to have sufficient measurements from which to estimate long-term solute concentration means and standard deviations for that site. In this study, we were able to derive reasonable estimates of both parameters from two years of 'survey' data, a minimum

data requirement likely to be available for large numbers of surface waters. Finally, long-term biogeochemical model simulations are required for representative sites in order to derive representative Z scores.

In this respect, this modelling approach has significant potential advantages over previous large-scale biogeochemical model applications, which have in general involved repeated model application to large numbers of individual sites (e.g. Evans et al., 1998; Moldan et al., 2004; Wright et al., 2005; Chen and Driscoll, 2005; Hettelingh et al., 2008). At the regional and national scale, modelling inevitably has to be undertaken using relatively simple models, parameterised with limited and/or extrapolated input data, and making little or no use of available long-term datasets. On the other hand, if biogeochemical variations are coherent across a range of sites, it may be possible to obtain a more accurate simulation of long-term change by extrapolating outputs from more detailed model applications at the smaller number of sites for which direct measurements are available. While the examples in this study were based on two soil layer site-based MAGIC applications, the upscaling method described could equally be based on models with more sophisticated process representations, for example taking account of climatic variations. Where long-term data are available, this approach could be combined with modelling techniques such as Bayesian parameter estimation (Larssen et al., 2005; Reinds et al., 2008) which can be used to constrain simulations, and to generate uncertainty ranges on future predictions.

A number of caveats apply to these conclusions. Firstly, in the method presented, individual water quality variables can only be extrapolated individually. If applied to the full suite of dissolved solutes, this would not ensure maintenance of charge balance constraints. Extrapolation to new (survey) sites also assumes that these sites are comparable to those for which long-term data exist, and caution is therefore required to ensure that upscaling is restricted to sites with similar catchment characteristics, and within a reasonable geographical range. Outputs from highly complex site-specific models, for example those that integrate biogeochemical processes with vegetation competition and succession (e.g. Wamelink et al., 2009) are unlikely to be appropriate for upscaling using this method. Finally, application of the method in the current study has been limited to a small number of illustrative examples, and more rigorous testing is required prior to larger-scale application. Nevertheless, given clear evidence of coherence in biogeochemical trends, we believe that there are significant potential advantages in terms of both model performance and computational time

requirement, in extrapolating outputs from a smaller number of more detailed model applications to monitoring sites, compared with extrapolating inputs in order to run simple models at a larger number of poorly characterised sites.

Acknowledgements

This study was supported by the UK Department of Environment, Food and Rural Affairs, under Contract AQ0801. We are grateful to the many people and organisations who have helped to maintain the UK Acid Waters Monitoring Network through the many challenges of its 20 years of operation.

References

- Aherne A., Futter, M.N., and Dillon, P.J. 2008. The impacts of future climate change and sulphur emission reductions on acidification recovery at Plastic Lake, Ontario. *Hydrol. Earth Syst. Sci.* 12: 383-392.
- Beier C., Moldan F. and Wright R.F. 2003. Terrestrial ecosystem recovery – Modelling the effects of reduced acidic inputs and increased inputs of sea-salts induced by global change. *Ambio* 32: 275–282.
- Belyazid S., Westling O., and Sverdrup H. 2006. Modelling changes in forest soil chemistry at 16 Swedish coniferous forest sites following deposition reduction. *Environ. Pollut.* 144: 596-609.
- Bishop K., Buffam I., Erlandsson M., Fölster J., Laudon H., Seibert S., Temnerud J. 2008. Aqua Incognita: the unknown headwaters. *Hydrol. Process.* 22: 1239-1242.
- Chen L. and Driscoll, C.T. 2005. Regional assessment of the response of the acid-base status of lake watersheds in the Adirondack region of New York to changes in atmospheric deposition using PnET-BGC. *Environ. Sci. Technol.* 39: 787-794.
- Cooper, D.M., Helliwell, R.C. And Coull, M.C., 2004. Predicting acid neutralising capacity from landscape classification: application to Galloway, south-west Scotland. *Hydrological Processes*, 18, 455-471.
- Cooper D.M. 2005. Evidence of sulphur and nitrogen deposition signals at the United Kingdom Acid Waters Monitoring Network sites. *Environ. Pollut.* 137: 41-54.
- Cosby B.J., Hornberger G.M., Galloway J.N., and Wright R.F. 1985. Modeling the effects of acid deposition: assessment of a lumped parameter model of soil and streamwater chemistry. *Water Resour. Res.* 21, 51-63.
- Cosby B.J., Ferrier R.C., Jenkins A. and Wright R.F. 2001. Modelling the effects of acid deposition: refinements, adjustments and inclusion of nitrogen dynamics in the MAGIC model. *Hydrol. Earth Sytem Sci.* 5: 499-517.
- Davies J.J.L., Jenkins A., Monteith D.T., Evans C.D, and Cooper DM. 2005. Trends in surface water chemistry of acidified UK Freshwaters, 1988–2002, *Environ. Pollut.* 137: 27-40.
- Dillon P.J., and Molot L.A. 2005. Long-term trends in catchment export and lake retention of dissolved organic carbon, dissolved organic nitrogen, total iron, and total phosphorus: The Dorset, Ontario, study, 1978-1998. *J. Geophys. Res. – Biogeosci.* 110: G01002.

- Evans C.D., Jenkins A., Helliwell R.C., and Ferrier R.C. (1998). Predicting regional recovery from acidification; the MAGIC model applied to Scotland, England and Wales. *Hydrol. Earth Syst. Sci.*, 2: 543-554.
- Evans C.D., Jenkins A. and Wright R.F. 2000. Surface water acidification in the South Pennines. 1. Current status and spatial variability. *Environ. Pollut.* 109: 11-20.
- Evans C.D., Monteith D.T., Harriman, R. and Jenkins, A. 2001. Assessing the suitability of Acid Neutralising Capacity as a measure of long-term trends in acidic waters based on two parallel datasets. *Water, Air Soil Pollut.* 130: 1541-1546.
- Evans C.D. and Monteith, D.T. 2001. Chemical trends at lakes and streams in the UK Acid Waters Monitoring Network, 1988-2000: Evidence for recent recovery at a national scale. *Hydrol. Earth Syst. Sci.* 5: 351-366.
- Evans C., Hall J., Rowe E., Aherne J., Helliwell R., Jenkins A., Cosby J., Smart S., Howard D., Norris D., Coull M., Bonjean M., Broughton R., O'Hanlon S., Heywood E., and Ullyett J. 2007. Critical Loads and Dynamic Modelling, Final Report. Report to the Department of the Environment, Food and Rural Affairs under Contract No. CPEA 19. 53 pp. (http://critloads.ceh.ac.uk/contract_reports.htm).
- Evans C.D., Reynolds B., Hinton C., Hughes S., Norris D., Grant S. and Williams B. 2008. Effects of decreasing acid deposition and climate change on acid extremes in an upland stream. *Hydrol. Earth Syst. Sci.*, 12: 337-351.
- Evans C.D. 2005. Modelling the effects of climate change on an acidic upland stream, *Biogeochemistry* 74: 21-46.
- Ferrier R.C., Helliwell, R.C., Cosby B.J., Jenkins, A. and Wright R.F. 2001. Recovery from acidification of lochs in Galloway, south-west Scotland, UK: 1979-1998. *Hydrol. Earth Syst. Sci.* 5: 421-431
- Forsius M., Alveteg M., Jenkins A., Johansson M., Kleemola S., Lukewille A., Posch M., Sverdrup H., Walse C. MAGIC, SAFE and SMART model applications to integrated monitoring sites: Effects of emission reduction scenarios. *Water Air Soil Pollut.* 105: 21-30.
- Fottová D. Regional evaluation of mass element fluxes – GEOMON network of small catchments. *Environ. Monitor. Assess.* 34: 215-221.
- Futter M.N., Helliwell R.C., Hutchins M., Aherne J. 2009. Modelling the effects of changing climate and nitrogen deposition on nitrate dynamics in a Scottish mountain catchment. *Hydrol. Res.* 40: 153-166.
- Gardner M. 2008. Long-term proficiency testing for the UK acid waters monitoring network. *Accred. Qual. Assur.* 13: 255-260.
- Gbondo-Tugbawa S.S., Driscoll C.T., Aber J.D. and Likens G.E. 2001. Evaluation of an integrated biogeochemical model (PnET-BGC) at a northern hardwood forest ecosystem. *Water Resour. Res.* 37: 1057-1070.
- Hall, J., Ullyett, J., Heywood, L., Broughton, R. & Fawehinmi, J. 2004b. UK National Focal Centre report. In: Critical Loads and Dynamic Modelling results. CCE Progress Report 2004. ICP M&M Coordination Centre for Effects (eds. J.-P. Hettelingh, M. Posch, J. Slootweg). RIVM Report No. 259101014/2004. RIVM, Bilthoven, Netherlands. pp114-134.

- Helliwell R.C., Ferrier R.C., Johnston L., Goodwin J., and Doughty R. Land use influences on acidification and recovery of freshwaters in Galloway, south-west Scotland. *Hydrol. Earth Syst. Sci* 5: 451-458.
- Helliwell R.C., Davies J.J.L., Jenkins A., Evans C.D., Coull M.C., Reynolds B., Norris D. and Ferrier R.C. 2007. Spatial and seasonal variations in nitrogen leaching and acidity across four acid-impacted regions of the UK. *Water Air Soil Pollut.* 185: 3-19.
- Hettelingh J-P., Posch M., Slootweg J. 2008. Critical load, dynamic modelling and impact assessment in Europe: CCE Status Report 2008. Coordination Centre for Effects, Report No. 500090003, ISBN No. 978-90-6960-211-0.
- Jenkins A., Cosby B.J., Ferrier R.F., Walker T.A.B., Miller J.D. 1990. Modelling stream acidification in afforested catchments: An assessment of the relative effects of acid deposition and afforestation. *J. Hydrol.* 120: 163-181.
- Jenkins A., Ferrier R.C., and Helliwell R.C. 2001. Modelling nitrogen dynamics at Lochnagar, N.E. Scotland. *Hydrol. Earth Syst. Sci.* 5: 519-528.
- Larssen T., Høgåsen T. and, Cosby B.J. 2007. Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model. *Environ. Model.* 207: 22-33.
- Moldan F., Kronnas V., Wilander A., Karlun E. and Cosby B.J. 2004. Modelling acidification and recovery of Swedish lakes. *Water Air Soil Pollut: Focus.* 4: 139-160.
- Monteith D.T., Evans C.D. and Reynolds B. 2000. Evidence for a link between temporal variations in the nitrate content of UK upland freshwaters and the North Atlantic Oscillation. *Hydrol. Process.* 14, 1745-1749.
- Monteith D.T., Stoddard J.L., Evans C.D., de Wit H., Forsius M., Høgåsen T., Wilander A., Skjelkvåle B.L., Jeffries D.S., Vuorenmaa J., Keller B., Kopáček J. and Vesely J. 2007. Rising freshwater dissolved organic carbon driven by changes in atmospheric deposition. *Nature* 450: 537-540.
- Monteith D.T., and Evans C.D. (Eds). 2000. UK Acid Waters Monitoring Network: 10 year report. Analysis and interpretation of results April 1988 – March 1998. ENSIS Publishing, London, ISBN 1 871275 26 1.
- Monteith D.T., and Evans C.D. 2005. The United Kingdom Acid Waters Monitoring Network: a review of the first 15 years and introduction to the special issue. *Environ. Pollut.* 137: 3-13.
- NEG-TAP. 2001. Transboundary Air Pollution: Acidification, eutrophication and ground-level ozone in the UK. Report of the National Expert Group on Transboundary Air Pollution. Department for Environment, Food and Rural Affairs, London. ISBN 1 870393 61 9 (www.nbu.ac.uk/negtap).
- Nisbet E. 2007. Cinderella science. *Nature* 450: 789-790.
- Oulehle F., Hofmeister J., Hruška J. 2007. Modeling of the long-term effect of tree species (Norway spruce and European beech) on soil acidification in the Ore Mountains. *Ecol. Model.* 204, 359-371.
- Patrick, S., Waters, D., Juggins, S., and Jenkins, A. 1991. The United Kingdom Acid Waters Monitoring Network: site descriptions and methodology report. Report to the Department of the

- Environment and Department of the Environment (Northern Ireland). ENSIS Ltd, London. ISBN 1 871275 04 0.
- Posch M. and Reinds GJ. 2009. A very simple dynamic soil acidification model for scenario analyses and target load calculations. *Environ. Model. Soft.* 24: 329-340.
- Reinds G.J., Posch M., and de Vries W. 2009. Modelling the long-term soil response to atmospheric deposition at intensively monitored forest plots in Europe. *Environ. Pollut.* 157: 1258-1269.
- Reinds G.J., van Oijen M., Heuvelink G.B.M., Kros, H. 2008. Bayesian calibration of the VSD soil acidification model using European forest monitoring data. *Geoderma* 146: 475-488.
- Sjøeng A.M.S., Wright, R.F., and Kaste Ø. 2009. Modelling seasonal nitrate concentrations in runoff of a heathland catchment in SW Norway using the MAGIC model: I. Calibration and specification of nitrogen processes. *Hydrol Res.* 40: 198-216.
- Skjelkvåle B.L., Stoddard J., Jeffries D., Tørseth K., Høgåsen T. Bowman J., Mannio J., Monteith D.T., Mosello R., Rogora M., Rzychon D., Vesely J., Wieting J., Wilander A., Worsztynowicz A., 2005. Regional scale evidence for improvements in surface water chemistry 1990-2001. *Environ. Pollut.* 137: 165-176.
- Sliggers J. and Kakebeeke W. (Eds). 2004. Clearing the air: 25 years of the Convention on Long-Range Transboundary Air Pollution. ECE/EB.AIR/84, United Nations, Geneva, ISBN 92-1-116910-0.
- Stevens P.A., Ormerod S.J. and Reynolds B. 1997. Final report of the acid waters survey for Wales. Volume 1, Main Text. Centre for Ecology and Hydrology, Bangor.
- Tipping E. 1996. CHUM: A hydrochemical model for upland catchments. *J. Hydrol.* 174: 305-330.
- Wamelink G.W.W., van Dobben H.F. and Berendse F. 2009. Vegetation succession as affected by decreasing nitrogen deposition, soil characteristics and site management: A modelling approach. *Forest Ecol. Manage.* 258: 1762-1773.
- Wright, R.F., Larssen T., Camarero L., Cosby B.J., Ferrier R.C., Helliwell R., Forsius M., Jenkins A., Kopacek J., Majer V., Moldan F., Posch M., Rogora M., and Schopp W. 2005. Recovery of acidified European surface waters. *Environ. Sci. Technol.* 39: 64A-72A.

Figure captions

Figure 1. Location of the UK Acid Waters Monitoring Network sites (site names corresponding to the numbers shown are given in Figure 6).

Figure 2. Relationship between a) mean annual ANC (1989-2007) and the mean of two years (2007 and 1992); and b) standard deviation of mean annual ANC and the absolute difference in ANC between 2007 and 1992, for all AWMN sites.

Figure 3. Untransformed annual mean time series for ANC, sulphate, nitrate and DOC concentrations for all individual sites in the UK Acid Waters Monitoring Network

Figure 4. Standardised annual mean (Z score) time series for ANC, sulphate, nitrate and DOC concentrations for all sites in the UK Acid Waters Monitoring Network. Bold central line shows median annual Z score for all sites, outer lines show 10th and 90th percentile values.

Figure 5. Comparison of median annual Z scores between lakes and streams, and forested and moorland catchments, for ANC, nitrate and DOC.

Figure 6. Observed annual mean ANC for all AWMN sites, compared to predictions derived by back-transforming median Z scores from the remaining 21 sites.

Figure 7. Root mean squared errors of annual mean ANC predictions obtained by extrapolation from mean Z scores at other sites based on 1) known and 2) estimated mean and standard deviation values.

Figure 8. Simulated long-term ANC (line) and observed annual means (circles) for a) Afon Gwy based on a previous site-specific model application (Evans et al., 2008); b) Llyn Llagi based on a previous regional-scale model application; and for Llyn Llagi based on a Z score transformation of the MAGIC simulation for the Afon Gwy for c) known and d) estimated ANC mean and standard deviation.

Figure 9. Simulated ANC at four time points for a set of 27 South Pennine reservoirs modelled directly using MAGIC, and indirectly by Z score transformation from a site-specific application of MAGIC to the River Etherow AWMN site. Boxes show median, lower and upper quartiles, whiskers 10th and 90th percentile simulated values for all sites.

Figure 1

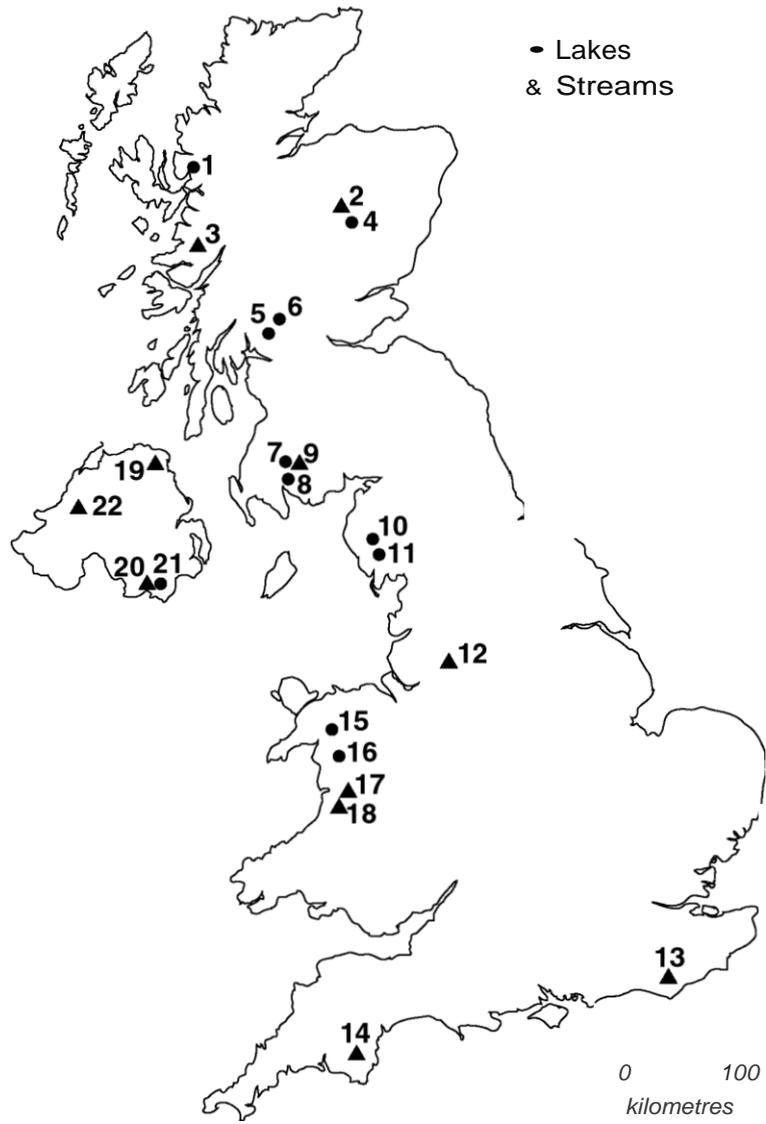


Figure 2

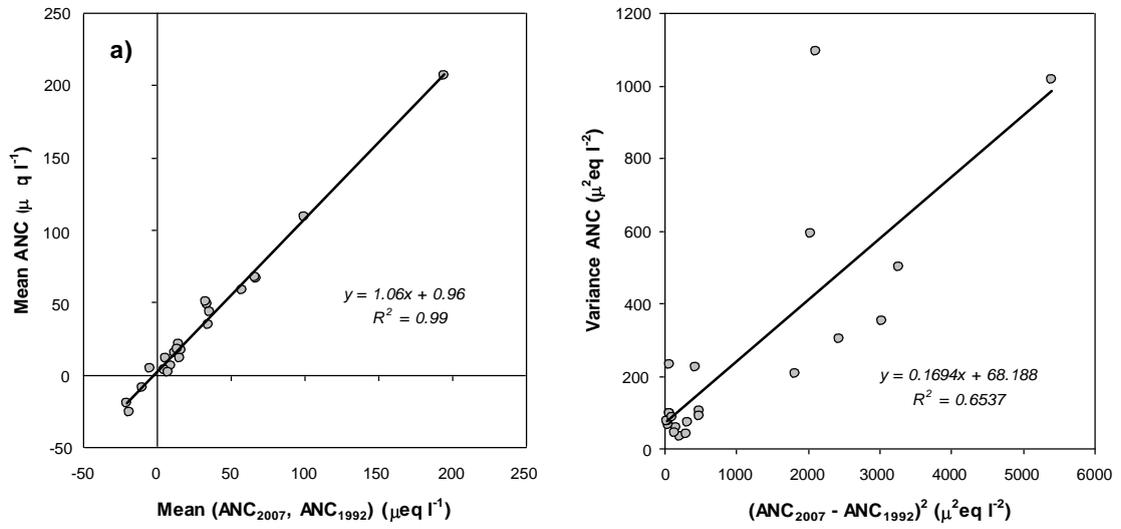


Figure 3

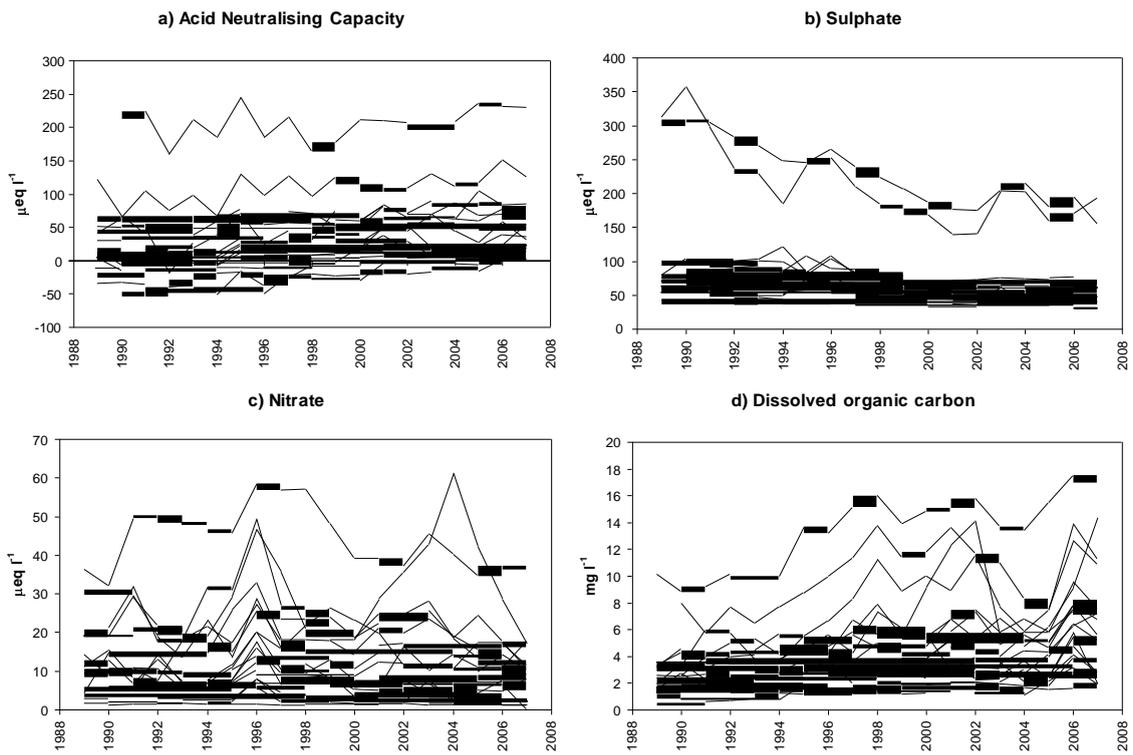


Figure 4

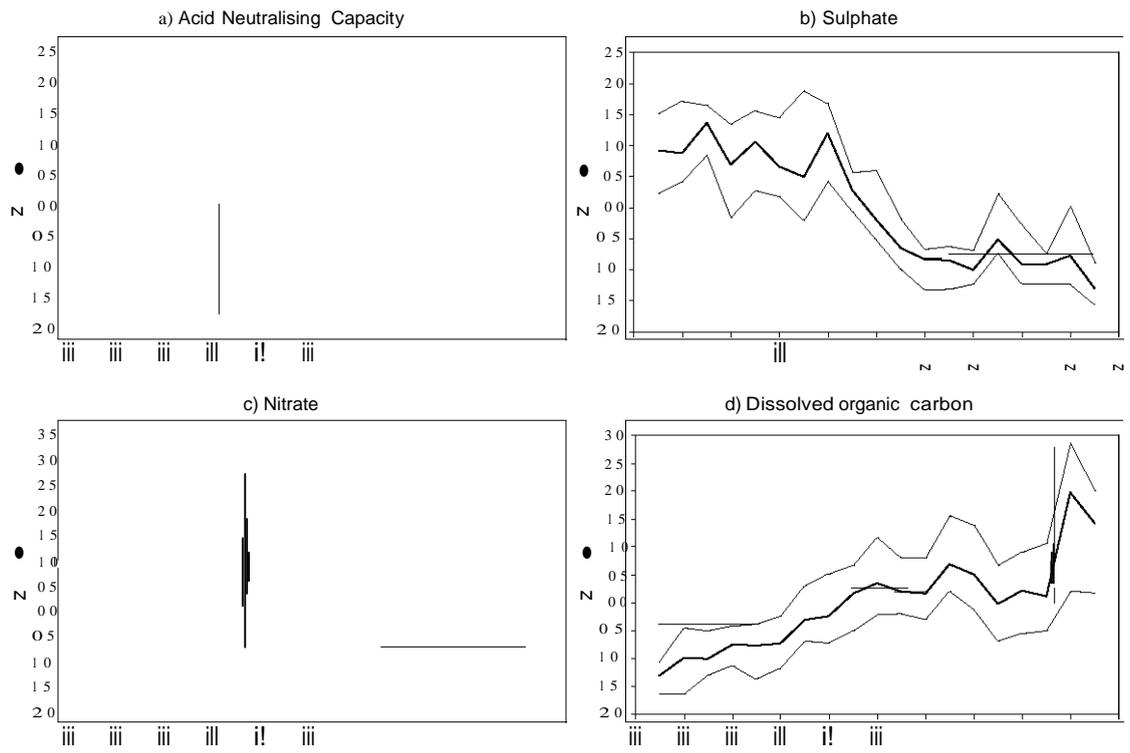


Figure 5

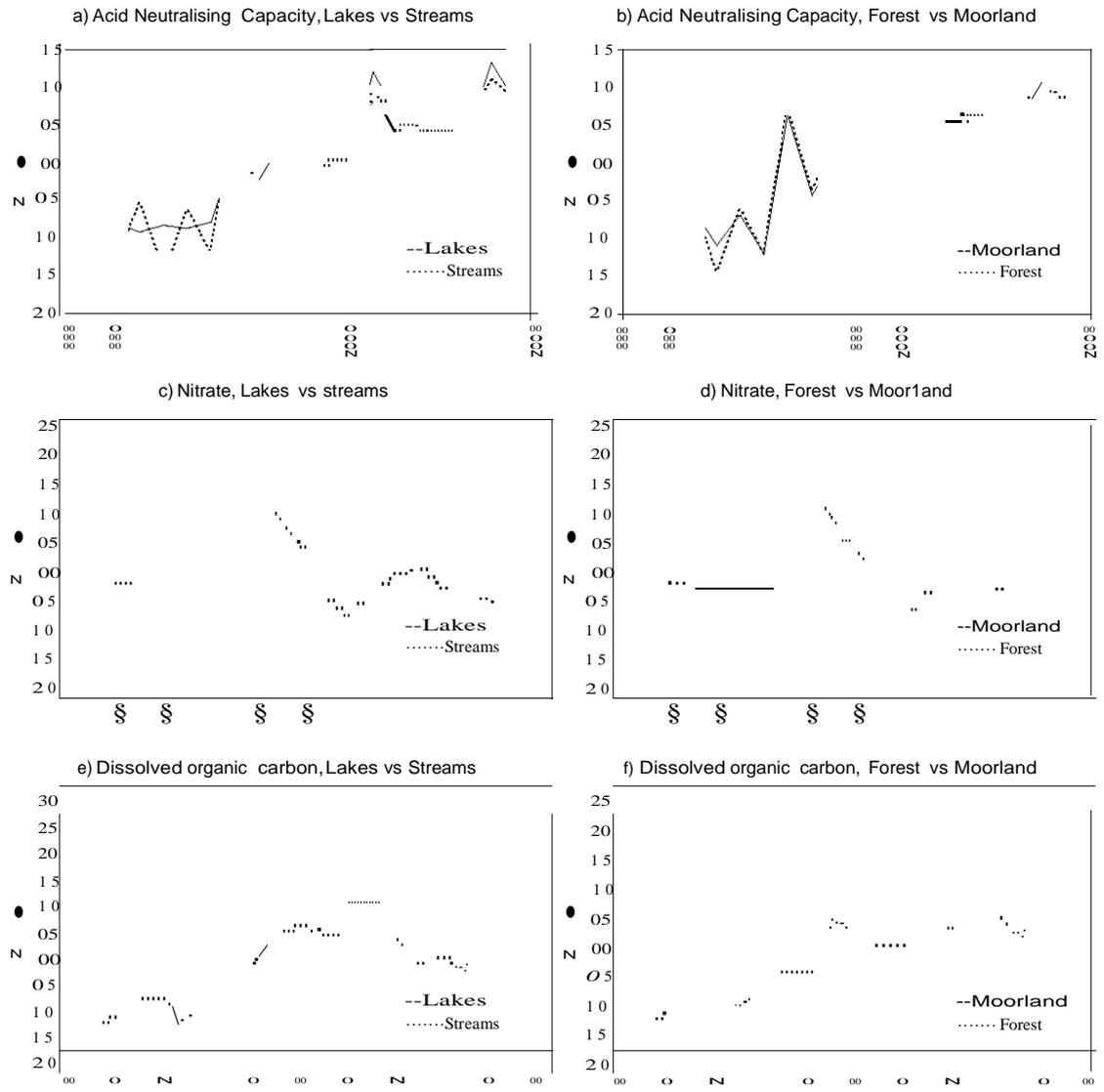


Figure 6

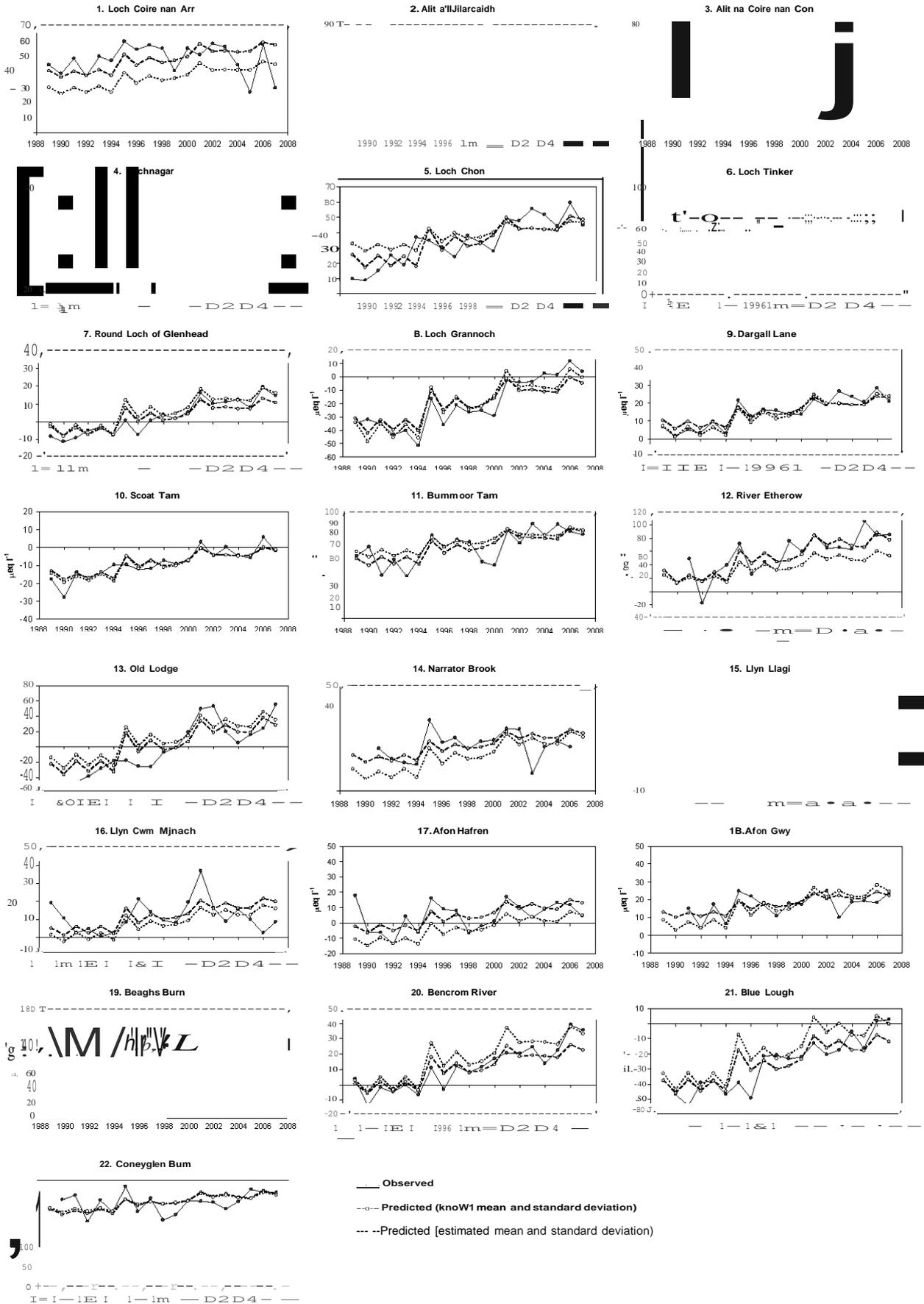


Figure 7

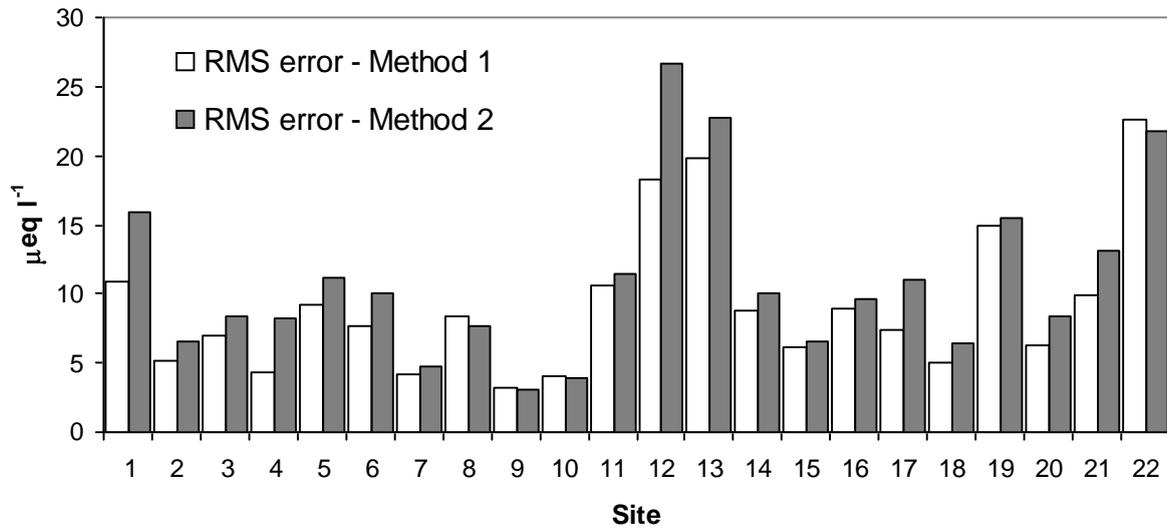


Figure 8

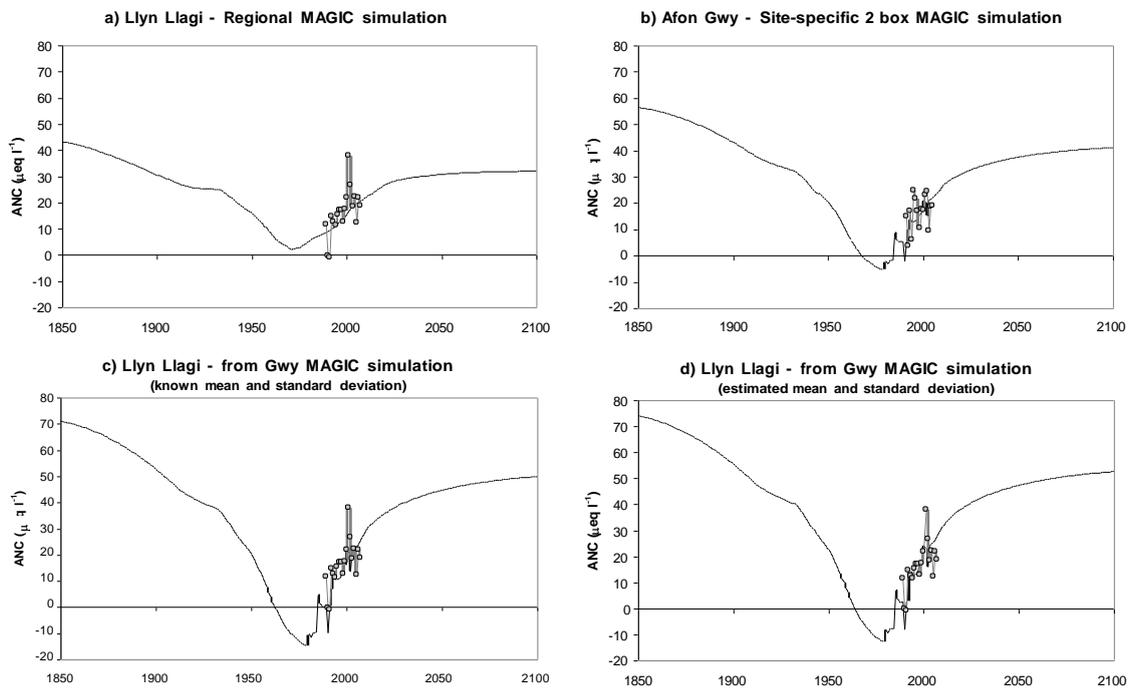


Figure 9

