Statistical considerations on reducing measurement network size for estimating pollution in rainfall in the UK

Ron Smith*
Centre for Ecology and Hydrology, Edinburgh, UK – ris@ceh.ac.uk

Lindsay Banin
Centre for Ecology and Hydrology, Edinburgh, UK – libanin@ceh.ac.uk

Christine Braban
Centre for Ecology and Hydrology, Edinburgh, UK – chri2@ceh.ac.uk

David Fowler
Centre for Ecology and Hydrology, Edinburgh, UK – dfo@ceh.ac.uk

Abstract

Measurements of the pollutant ion concentrations in rainfall have been made across the UK for the last 28 years. The network originally had more sites (59 sites in 1987) but for the last decade has operated with 38 sites. This study was part of work to quantify the effect of likely budget cuts which would potentially reduce numbers of sites in the UK monitoring networks. The network measurements, taken weekly or fortnightly, are used to construct an annual concentration map for the UK from which is calculated the wet deposition, i.e. the amount of atmospheric pollutant that enters the ecosystem through rain. The study looked at sulphate, nitrate and ammonium in rainfall, as they are drivers of change through acidification and eutrophication, and how the changes in site numbers affected the concentration maps and their associated uncertainty. Concentration maps were routinely produced by conventional kriging. A simulation study, examining the effect of network size, removed sites at random with 100 simulations each of network reductions by 1 up to 25 sites. The results for the 3 ions were different reflecting differences in underlying physical processes. The automated process for the simulations gave an increasing number of failures to produce a concentration map with reduced site numbers, mainly due to difficulty in fitting variograms. Uncertainty increased as expected but there was also a shift, using the median predicted concentration for each set of site reductions, to increase concentrations in the west and decreases in the east. The second phase removed 11 sites by informed choice rather than random selection and also resulted in a fall in mapped mean concentration for all ions by a small amount with a 'tilting' of the map (non-homogeneous change to the mean). Bayesian kriging was implemented to provide a better variogram fit and reasonable estimates of uncertainty, with an increase of around 10% in the coefficient of variation with the reduced numbers of sites. Further work is required to explore the effects of these changes on production of national map estimates of concentrations and deposition.

Keywords: geostatistics; uncertainty; bias.

1. Introduction

The current network of measurements of atmospheric pollutant ion concentrations in rainfall in the UK has existed for 28 years. It originally consisted of 59 sites but for the last decade has operated with 38 sites. Over the latter period there was also a reduction from weekly sampling to the current fortnightly sampling. A likely reduction in funding suggests the future network will be reduced further by about 10 sites. Data collected by the network is used for many purposes, including

- to determine current UK rural atmospheric pollution concentrations and deposition
- to assess trends in air pollution over the last 25 years
- to validate long-range atmospheric emission and transport models

A single quantitative criterion to use in determining an optimal network configuration is difficult to identify. Transport models are often linked to weather models and will use time steps of less than a day while the prediction of likely changes in vegetation composition relates to much longer time periods, for example monthly or annual, but the spatial resolution is more important.

Damage to the ecosystem can be caused by deposition of pollutants from the atmosphere to the landscape, and wet deposition is that component that comes in rainfall. Rainfall ion concentration fields interpolated from the network measurements are combined with a rainfall field from the UK Meteorological Office to produce annual wet deposition at a 5km x 5km scale across UK (Smith and Fowler, 2001). The critical load is "a quantitative estimate of an exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment do not occur according to present knowledge" (Nilsson and Grennfelt 1988). Predicted changes in the area where deposition exceeds the critical load is a measure used in Europe to target emission reduction policies so as to deliver most benefit for given cost. With critical load exceedances used for policy and planning development, consistency in calculation of these statistics over time is important.

This exploratory study focussed on the spatial aspects of possible site reductions and how these affected the uncertainty of predicted annual mean concentrations.

2. Random removal of sites.

The initial work was focused on attempting to define a target size for the network, so used simulation studies to investigate the effect purely of site number on the change in uncertainty in the concentration maps by randomly removing sites from the mapping procedures. The focus was centred on non-seasalt sulphate (xSO_4^{2-}), ammonium (NH_4^+) and nitrate (NO_3^-) in rainfall. Reductions of the network size by more than 10 sites were shown to substantially increase uncertainty estimates, particularly for ammonium and nitrate in rainfall where coefficients of variation of over 50% were not uncommon at a large number of 5km x 5km mapped grid squares. The automated process for the simulations gave an increasing number of failures to produce a concentration map with reduced site numbers, mainly due to difficulty in fitting variograms.

The other consequence of a reduction in site numbers was the risk that the mapped concentrations will be altered at important locations. For a site specific indicator derived from the national mapping, this could introduce discontinuities in trends purely as a result of network reorganisation. It appeared for sulphate that this was likely to be an issue only in the English Midlands and down the east coast of the country for reductions of up to 10 sites. However the nitrogen compounds appeared to be less robust to the smaller site reductions (5-10 sites) with the potential for noticeable changes with the 10 site reduction. The general pattern of changes as site numbers reduced was to lower the mapped concentrations in the east of the UK and to increase concentrations in the west. The size of these regional changes was substantial and could affect one of the policy indicators used in Europe, the area of critical load exceedance.

With both the interpolation uncertainty increasing and the possibility of changes in mapped mean concentration values in some parts of the country (particularly for the nitrogen compounds), this initial phase predicted there would be a noticeable loss of quality in estimated concentrations and deposition if a reduction of more than 10 sites were implemented.

3. Targeted site reductions – changes in the mean map

A second phase looked at a specific proposal to close 11 sites from the current network. The sites were chosen partly on cost grounds and partly to avoid likely issues in providing a national map.

An example comparison of rain ion concentration maps for 2012 using all current sites and the proposed reduced site network are shown in Figure 1. A similar pattern of concentrations has been achieved from the 2 sets of maps. However, the average concentration is lower for all ions with the reduced site map compared to the all site map (Table 1), with a 2% drop for NO_3^-, 3% for xSO_4^{2-}, and 5% for NH_4^+. Although there may not be a substantial net change to the wet deposition budget for the UK, there are regional changes as illustrated in Figure 1 with some areas having increases or decreases in concentration of more than 15% (which directly lead to the same percentage change in estimated wet deposition). The spatial patterns for xSO_4^{2-} and NH_4^+ are similar to NO_3^-, though NH_4^+ has less area where concentrations have increased by 30% or more. The differences between ions occur because of differences in the smoothness of the concentration fields, so NH_4^+ is affected by more dispersed local sources of ammonia and is more locally variable, while SO_4^{2-} is generally a longer-range transport issue with a smaller number of sources across Europe resulting in a smoother concentration field.

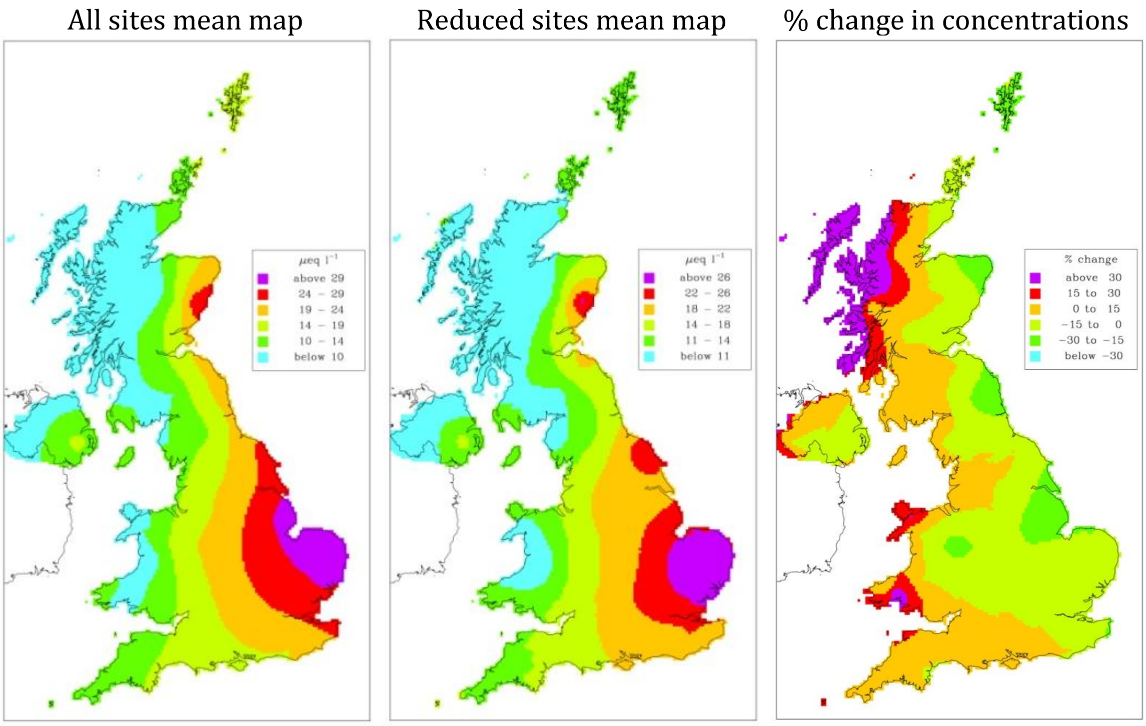

Figure 1. Rain ion concentration maps for 2012 for nitrate (NO3-) in rainfall showing the mapping with all sites, the mapping with the proposed reduction in sites, and the percentage change in concentrations.

Table 1. UK annual mean mapped concentrations ($\mu eq\ l^{-1}$) for ammonium, nitrate, and non-seasalt sulphate from the all sites and reduced sites mapping procedures for 2008-2012 using conventional kriging.

Ion	Year	All sites	Reduced sites	% change	average % change
NH_4^+	2008	22.01	21.04	-4.4	
	2009	21.14	20.46	-3.2	
	2010	27.19	26.17	-3.8	
	2011	25.68	24.22	-5.7	
	2012	21.34	20.62	-3.4	-4.1
NO3-	2008	16.38	15.98	-2.4	
	2009	17.45	17.17	-1.6	
	2010	19.32	19.03	-1.5	
	2011	16.95	16.54	-2.4	
	2012	15.31	15.02	-1.9	-2

xSO_4^{2-}	2008	14.05	13.62	-3.1	
	2009	11.82	11.53	-2.5	
	2010	14.59	14.14	-3.1	
	2011	14.04	13.57	-3.3	
	2012	11.88	11.65	-1.9	-2.8

4. Updating the methodology – changes in uncertainty

For the last 25 years UK concentration maps of rainfall ions have been produced using kriging. This method is in many ways optimal and provides an estimate of the uncertainty introduced by the interpolation process. Conventional kriging first fits an empirical variogram to estimate the variance of the difference between observed values at any two spatial locations – making the assumption that locations close together in space are more similar than those far apart. The parameters of the fitted variogram are then using in the kriging stage which uses a Gaussian process to predict intermediate values.

While the variogram fitted reasonably well to data in the earlier years of the network, as time progressed there were years when fitting the variogram was more difficult. One weakness of the conventional kriging approach is that the uncertainty in fitting the variogram is not carried forward to be included in the estimate of mapping uncertainty, and when the variograms are fitted less well that becomes a more important consideration. Model-based geostatistics (Diggle et al, 1998), based on linear mixed and generalised linear mixed models, leads to a Bayesian kriging with an estimation error that includes variogram uncertainty. In this study with further site reductions proposed, it was clear that conventional kriging was no longer an appropriate way to estimate the predicted uncertainty, especially for a comparison of networks with different numbers of observation sites. Bayesian kriging does not produce a single concentration map, as is the case in conventional kriging, but instead generates a distribution of possible maps given (a) the measurement data and (b) a number of assumptions on the distribution of parameters that are included as priors.

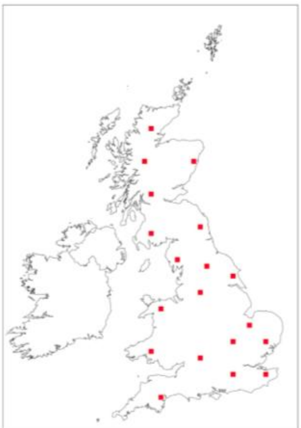

Figure 2. Map of the locations chosen for illustrating the outcome of the change of site numbers.

A Bayesian kriging method was implemented and results analysed for a limited number of point locations across the UK (shown in Figure 2), using the same assumptions about prior distributions for fits from the full and reduced sets of sites. The results are summarised as the mean and the variance of a sample of the distribution of possible predicted maps for each of the 19 locations. The mean of the predictions and the average percentage change in the coefficient of variation for the 19 locations are given in Tables 2 and 3 respectively.

Though based only on these 19 locations, the changes broadly confirm the observations made from the maps generated by the conventional kriging procedure. The average rain ion concentrations are generally reduced as a consequence of reducing site numbers while the change in average coefficient of variation is increased in most cases. On average there is an increase in uncertainty of the order of 10% for NH_4^+, NO_3^- and xSO_4^{2+}.

Table 2. Average percentage change in the mean of the predictions for the 19 locations for major ions between 2008 and 2012 using Bayesian kriging.

				Year			
%change in mean		2008	2009	2010	2011	2012	All years
Ions	NH_4^+	-6.7	-2.2	1.3	-2.6	-0.8	-2.2
	NO_3^-	-1.4	-0.3	0.8	-1.3	1.3	-0.2
	xSO_4^{2+}	-2.8	-1.8	-1.6	-0.5	0.7	-1.2

Table 3. Average percentage change for 19 locations in the coefficient of variation for major ions between 2008 and 2012 using Bayesian kriging.

				Year			
%change in CV		2008	2009	2010	2011	2012	All years
Ions	NH_4^+	4.5	14.5	5.8	5.9	13.0	8.8
	NO_3^-	19.3	17.3	12.8	10.1	15.7	15.0
	xSO_4^{2+}	11.7	1.3	10.0	-0.1	20.5	8.7

5. Conclusions

Three separate elements of uncertainty have been identified as relevant to assessing the impact of reducing the number of sites in the network. The first is the interpolation uncertainty which is estimated by the kriging variance, and shows how good the interpolation is given the fixed set of site measurements. The uncertainty produced by the Bayesian kriging includes the uncertainty from fitting the variogram, and so provides a better comparison of the changes in mapping uncertainty related to site reductions. A general increase in uncertainty is shown by the Bayesian kriging procedures.

The second element reflects the possibility that the predicted (mapped) concentration may vary depending on the number of sites used in the network. The conventional kriging showed a reduction in average concentrations across the whole map for all ions. The Bayesian kriging with its predictions at the 19 chosen locations also showed reduced concentrations on average, but these are not directly comparable to the mean mapped concentrations. These changes are all relatively small percentages of the mean concentrations.

The third element of uncertainty has not been quantified in this work. The reduced set of sites may not be capturing the spatial variability that is actually present in concentrations measured over the UK, and the proposed site reduction data sets seem to have lost some sites with higher measured concentrations. Therefore there is a potential issue of how well the sites sample the concentration field.

While this study considered a change in uncertainty, that should be put in the context of the current uncertainty estimates. There is little observational data to confirm an uncertainty estimate on total deposition, the value eventually used in calculating a critical load exceedance. In Smith and Fowler (2001) an estimate of uncertainty in wet deposition was given as around ±35% (i.e. a CV of 17.5%) for 5km x 5km grid squares. However this estimate ignored the kriging error in the concentration maps. Later work (Smith, 2001) on how good the national maps were at estimating deposition for specific vegetation fieldwork studies looked at kriging errors. The kriging CV increased from below 20% where the location was less than 15km from a national network monitoring site to nearer 30% at 40km from a monitor. Heywood et al (2006) estimating uncertainty in critical load exceedances used a CV on the deposition estimates (which included both wet and dry deposition, the latter not considered in this paper) of 25%.

In conclusion, the increase in uncertainty with reduction in site numbers is sufficiently large for its effects to be investigated further.

References

Diggle, P., Moyeed, R.A. and Tawn, J.A. (1998) Model-based geostatistics (with discussion). Journal of the Royal Statistical Society: Series C (Applied Statistics), 47, 299-350.

Heywood, E., Hall, J. and Smith, R.I. (2006) Uncertainty in mass balance critical loads and exceedance: Application to a UK national data set. Atmospheric Environment, 40, 6146-6153.

Nilsson, J. and Grennfelt, P. (1988) Report from a workshop held at Skokloster, Sweden, 19-24 March 1988. NORD Miljørapport 15: Nordic Council of Ministers, 418pp.

Smith, R.I. & Fowler, D. (2001) Uncertainty in estimation of wet deposition of sulphur. Water, Air & Soil Pollution: Focus 1, 341-353.

Smith, R.I. (2001) Sampling for estimation of air pollution levels and air pollution environmental effects. Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001. (http://www.amstat.org/sections/srms/proceedings/y2001/proceed/00512.pdf)