Environmental Modelling & Software 15 (2000) 161–167

Environmental Modelling & Software

www.elsevier.com/locate/envsoft

A fuzzy model of heavy metal loadings in Liverpool bay

L.I. Kuncheva [a,*], J. Wrench [b], L.C. Jain [c], A.S. Al-Zaidan [a]

[a] School of Mathematics, University of Wales, Bangor, Gwynedd LL57 1UT, UK
[b] School of Ocean Sciences, University of Wales, Bangor, Gwynedd LL57 1UT, UK
[c] University of South Australia, Adelaide, Mawson Lakes, S.A. 5095, Australia

Received 1 June 1999; accepted 18 August 1999

Abstract

We design a fuzzy model of the loadings of 10 heavy metals in Liverpool bay. Each metal concentration is associated with a fuzzy set "contaminated", defined over the set of 70 sampling sites. The higher the concentration, the higher the degree of membership of the site. Six overall loading indices are calculated using aggregation connectives between fuzzy sets. The loading indices are then interpolated and plotted on a map. A visual inspection shows that: (1) product aggregation is most indicative for the locations of the disposal grounds; (2) mean aggregation reflects sediment movement in the bay well; and (3) maximum aggregation indicates all highly contaminated sites. The proposed fuzzy model is easy to implement and the results are directly interpretable.

Keywords: Fuzzy aggregation connectives; Environmental modelling; Liverpool bay; Heavy metal concentration; Index of spatial distribution

1. Introduction

The dynamic nature of coastal waters presents a severe challenge to environmental assessment of disposal activities in near-shore waters. Surface sediments lying in shallow water act as a primary sink for pollutants delivered through rivers and groundwater flow and are most often studied in an attempt to identify and assess contaminant sources. However, sediments are subjected to strong oscillating tidal forces resulting in bed load transport; storms may result in mixing of contaminated sediment with material from a different source; and organisms in sediments may re-work deeper consolidated material to the surface layer. The overall effect of these uncontrollable environmental variables is to obscure spatial data.

The problem is to find an overall distribution of metal concentrations (or contamination) given that metals have different concentration scales and the way of combining the concentrations is not prescribed. Large data sets are being collected and stored, awaiting processing and analysis. Typically, principal component analysis (PCA) or cluster analysis is used for such data, possibly because these techniques are available in most statistical software packages. The results of both methods are difficult to interpret, unless the data have favourable structure and characteristics. Fuzzy set modelling is a straightforward option for this kind of problem.

We study the loadings of 10 heavy metals in Liverpool bay and design *loading indices* to represent the overall metal concentration. Section 2 describes the environmental problem and the data set. In Section 3 we briefly introduce fuzzy sets and give the fuzzy aggregation operators used to design six loading indices. The results are shown and discussed in Section 4 along with PCA and cluster analysis results.

2. Liverpool bay data

The situation in Liverpool bay is complicated by multiple sources of contaminants. The bay receives heavy metals from continuous sources (Mersey and Dee Estuaries), point sources (offshore disposal ground) and a higher-than-normal input of some heavy metals (notably arsenic) through erosion of natural mineral sources (Camacho-Ibar et al., 1992). Although water cir-

* Corresponding author. Tel.: +44-1248-383661; fax: +44-1248-383663.
E-mail address: l.i.kuncheva@bangor.ac.uk (L.I. Kuncheva).

1364-8152/00/$ - see front matter
PII: S1364-8152(99)00031-6

culation in the bay depends on tides, winds, freshwater inputs, etc., it has been found that there is a well established estuarine-like circulation induced by the low-density freshwater inputs from the rivers (the Mersey in particular) and by higher-density sea water from the Irish Sea (Camacho-Ibar et al., 1992). Low-density water moves offshore through the surface and high-density water moves inshore near the bottom. This density-driven inshore movement of bottom water, coupled with tidal asymmetry producing stronger flood than ebb tides, induces a net sediment transport directed east and south-east towards the Mersey. The combined effect of environmental factors and multiple sources of contamination is to generate complex and changing patterns in the distribution of metal contaminants in surface sediments. The problem from the point of view of environmental management is to develop patterns from metal data which reflect the current status of sediments. The effect of metal contamination on the biota in sediments is complex; exposure of organisms to high levels of more than one metal introduces further environmental stress. It is therefore appropriate that an approach to developing patterns of contaminant distribution should include all metals. Identification of distribution patterns is therefore relevant to:

- monitoring the effect of waste disposal in Liverpool bay;
- locating regions which may have unacceptably high metal loading;
- detecting changes of loading pattern over the long term; and
- using metal loading patterns to disclose the sediment transport regime.

Environmental management is based on yearly analysis of heavy metals in surface sediments from a sampling grid approximately 20 km^2. Fig. 1 shows the Liverpool

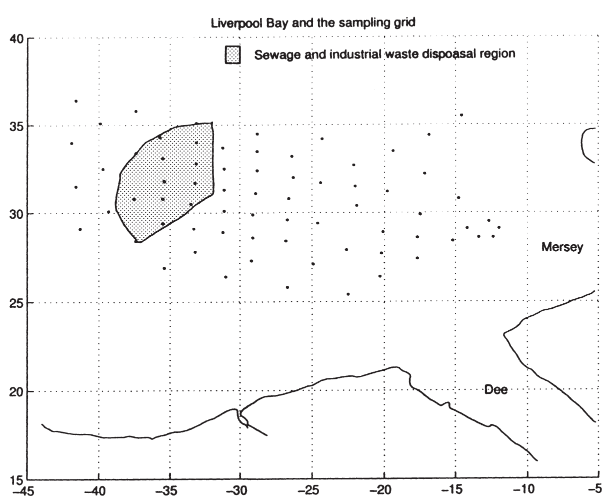

Fig. 1. Liverpool bay area, the disposal ground and the sampling grid.

bay area, the disposal ground and the sampling sites (stations). Interpretation of the sampled data is limited by the factors described above and there is a need to develop a rational protocol whereby a realistic picture may be produced from spatial data.

The data set in this study consists of the concentrations of 10 heavy metals relative to aluminium content, measured between 14 and 16 September 1988 at the 70 sites on the sampling grid. The metals are: mercury (Hg), cadmium (Cd), chromium (Cr), copper (Cu), nickel (Ni), lead (Pb), zinc (Zn), arsenic (As), manganese (Mn) and iron (Fe). The data set is available as an ASCII file at http://www.bangor.ac.uk/~mas00a/lb1988.txt. The first two columns are the latitude and longitude of the sampling sites, and the remaining 10 columns are the metal concentrations in the above order.

The same stations have been sampled every year. A database has been collected over the period 1986–1993. Our aim in this pilot study was to develop a mathematical tool for analysing metal distributions using one data set. We chose the 1988 data set because a thorough analysis of the processes in Liverpool bay for the year 1988 can be found elsewhere (Camacho-Ibar, 1991; Camacho-Ibar et al., 1992). Changes in the metal distribution pattern over the years using the loading indices proposed in this paper is a topic of our forthcoming research.

3. Loading indices design by fuzzy aggregation

3.1. Fuzzy sets

Lotfi Zadeh introduced the simple and intuitive concept of a fuzzy set in his seminal paper in 1965 (Zadeh, 1965). Since then fuzzy sets have been applied to a vast number of areas (Bezdek and Kuncheva, 1999) including the environmental sciences: soil, forest and air pollution, meteorology, water resources, etc.

Let U be an ordinary set with elements $u_1, \ldots, u_m$. A *fuzzy set A* on U is defined by assigning a *degree of membership* between 0 and 1 to each $u_i \in U$, usually with regard to a linguistic term. For example, let U be the set of integers from 1 to 100 denoting the age of a person, and let A be "middle aged". We can define a (subjective!) function that assigns to each u_i a degree of membership $\mu_A(u_i) \in [0, 1]$. Degree 0 denotes non-membership, degree 1 full membership, and any other value is partial membership. A plausible model of "middle aged" will be obtained by using a function (*membership function*) that yields high values between, say, 40 and 55 and gradually decreases towards the two edges of the scale. Thus, the degree of membership of 37, $\mu_A(37)$, can be 0.75, and of 82, $\mu_A(82)$=0.1. A fuzzy set is determined by its membership function, so the two notions will be used interchangeably.

Let $S=\{s_1, \ldots, s_{70}\}$ be the set of 70 sites in Liverpool bay. Let $A_1, \ldots, A_{10}$ be fuzzy sets on S, one for each metal, with membership functions:

$$\mu_{A_i}: S \to [0, 1], \quad i=1, \ldots, 10.$$

The higher the ith metal concentration at site s_j, the higher the degree of membership $\mu_{A_i}(s_j)$. We choose simple rescaling to devise the 10 membership functions from data. Let $x_i(s_j)$ be the concentration of the ith metal measured at site s_j. Then:

$$\mu_{A_i}(s_j)=\frac{x_i(s_j)-\min_k\{x_i(s_k)\}}{\max_k\{x_i(s_k)\}-\min_k\{x_i(s_k)\} k=1,\ldots 70}. \quad (1)$$

Fig. 2 plots the membership function of mercury over the two-dimensional space (sampling ground) spanned by the 70 sites in Liverpool bay. Referring the plot to the original geographical problem, we observe a high concentration of mercury in the area of river estuaries. The contamination with mercury in that area is higher than that at the disposal ground (the second highest peak).

Another way of representing the membership functions (adopted here) is to use colour or a contour plot and overlay the scatterplot of the sampling sites.

Clearly, the concentration pattern would be the same if we did not scale the concentration between 0 and 1. Although individual metal distribution is an interesting topic on its own, it was argued above that an index of overall loading (contamination) is needed.

3.2. Fuzzy aggregation connectives

Dubois and Prade (1997) point out that, although fuzzy membership functions have numerous possible interpretations, fuzzy mathematics has gone a long way disregarding fuzzy sets semantics: "The risk is to leave the user with no guidelines about how to apply fuzzy set theory…" They distill three main semantics: similarity, preference and uncertainty. The interpretation of the fuzzy sets used here fits best in the second category: preference (in a broad sense), because the membership functions do not measure a similarity to some "ideal" prototypes nor do they express any type of uncertainty. For example, we *prefer* (to call "contaminated") a site with degree of membership 0.7 to a site with degree 0.4. Taking this interpretation, a fuzzy decision-analysis approach seems reasonable. Fuzzy aggregation connectives (aggregation operators) will be used to define overall loading indices.

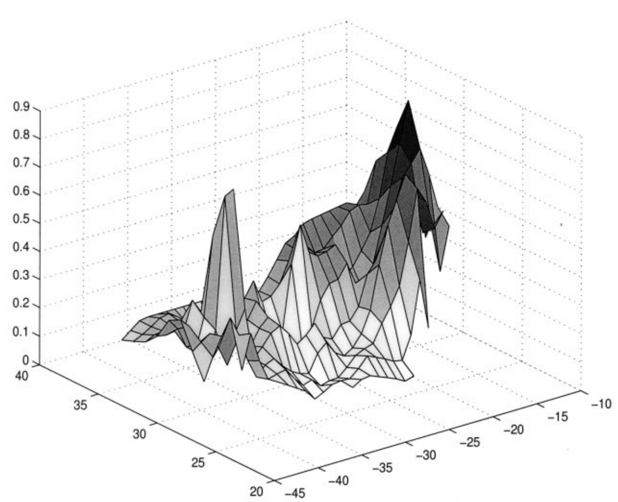

Fig. 2. Hg membership function for the fuzzy set "contaminated".

An m-place *aggregation operator* $\mathcal{A}$ is defined as:

$$\mathcal{A}:[0: 1]^m \to [0, 1],$$

satisfying the following properties.

1. *Limit conditions*:
   $\mathcal{A}(0, 0, \ldots, 0)=0, \quad \mathcal{A}(1, 1, \ldots, 1)=1.$
2. *Commutativity*:
   $\mathcal{A}(a_1, a_2, \ldots, a_m)=\mathcal{A}(b_1, b_2, \ldots, b_m)$, where $b_1, \ldots, b_m$ is any permutation of $a_1, a_2, \ldots, a_m$, $a_i \in [0, 1]$, $i=1, \ldots, m$.
3. *Monotonicity*:
   $\mathcal{A}(a_1, a_2, \ldots, a_m) \geq \mathcal{A}(b_1, b_2, \ldots, b_m)$, for any $a_1, a_2, \ldots, a_m$, $b_1, \ldots, b_m \in [0, 1]$, such that $a_i \geq b_i$, $\forall i=1, \ldots, m$.

There are a great variety of fuzzy connectives and aggregation operators (Bloch, 1996; Dubois and Prade, 1985; Grabisch, 1995b; Yager and Filev, 1994). Since this is a pilot study, here we use perhaps the simplest six aggregation operators.

1. *Pessimistic-type aggregation*:
   - Minimum
     $\mathcal{A}_1(a_1, a_2, \ldots, a_m)=\min\{a_1, a_2, \ldots, a_m\}$.
   - Product
     $\mathcal{A}_2(a_1, a_2, \ldots, a_m)=a_1 \cdot a_2 \cdot \ldots \cdot a_m$.
   - Geometric mean
     $\mathcal{A}_3(a_1, a_2, \ldots, a_m)=(a_1 \cdot a_2 \cdot \ldots \cdot a_m)^{1/m}$.
2. *Optimistic-type aggregation*:
   - Maximum
     $\mathcal{A}_4(a_1, a_2, \ldots, a_m)=\max\{a_1, a_2, \ldots, a_m\}$.
3. *Indifferent-type aggregation*:
   - Arithmetic mean
     $$\mathcal{A}_5(a_1, \ldots, a_m)=\frac{1}{m}(a_1+\ldots+a_m).$$
   - "Competition jury" — this is an operator where we discard the highest and the lowest values from the set $a_1, a_2, \ldots, a_m$ and average the remaining $m-2$ values:
     $$\mathcal{A}_6(a_1, a_2, \ldots, a_m)=\frac{1}{m-2}(a_1+\ldots+a_m-\max_k a_k-\min_k a_k).$$

Table 1
Correlations between the metal concentrations (×100)

	Cd	Cr	Cu	Ni	Pb	Zn	As	Mn	Fe
Hg	0	−16	−23	−34	−31	−23	−27	−31	−27
Cd		55	54	63	71	70	51	55	48
Cr			77	83	76	85	70	64	87
Cu				70	77	80	60	57	78
Ni					78	84	65	73	79
Pb						95	81	82	78
Zn							88	88	88
As								90	87
Mn									76

Replacing a_i with $\mu_{A_i}(s_j)$, each of these six aggregation operators defines a *loading index* as a fuzzy set on *S*. The indices are denoted respectively LI_1, …, LI_6.

4. Results

Table 1 shows the correlation coefficients (×100) between the metal concentrations.

Cluster analysis is often used in environmental studies to find spatial areas corresponding to values "low", "medium" and "high" of a certain parameter indicative of the concentration (Markus and McBratney, 1996). When this parameter is based on more than one variable, e.g., contamination with heavy metals, clustering might not be appropriate. Fig. 3 shows the difference between "favourable data" (where the clusters correspond to "low" and "high" contamination) and "unfavourable data" (where such correspondence does not exist). Shown on the left plot are two clusters found by the hard c-means (HCM) clustering procedure for zinc (Zn) and lead (Pb) from our data. Their concentrations correlate well (0.95), so the two cluster centres can be labelled "low" and "high" as shown. The plot on the right shows the results from the same procedure applied to mercury (Hg) and lead (Pb). The metal concentrations have a negative correlation coefficient (−0.31), so the cluster centres (circled) do not correspond to "low" and "high" in any order.

When more than two components are involved such inconsistencies are more difficult to resolve and likely to be obscured, and the result can be misinterpreted.

As Table 1 shows, most of the correlations are positive (due to high concentrations around the waste disposal ground and low elsewhere) except for mercury, and therefore cluster analysis can be expected to produce sensible results. For three suspected clusters the hard c-means algorithm on the scaled data [Eq. (1)] found the centres shown in Table 2.

All centre components but the first one (Hg) are ordered so that the three clusters correspond to "high", "low" and "medium", respectively. Fig. 4 shows the result as the spatial distribution of loading on the geographical sampling grid. Dark regions correspond to high metal loading, and bright regions to low loading. Hard c-means clustering uses random initialization, hence different sets of centres can be obtained. The two plots show the results from clustering the original metal concentrations (left) and the scaled data (right). Clustering results are influenced by data transformation (Duda and Hart, 1973), which is demonstrated by the differences in the two plots. Then, should we scale the data or not? If not, the metals with concentrations that are orders of

Table 2
Three cluster centres from one HCM run

	Cluster 1	Cluster 2	Cluster 3
Hg	0.1259	0.3011	0.2448
Cd	0.4159	0.0810	0.1679
Cr	0.7228	0.1739	0.5275
Cu	0.6875	0.0663	0.3147
Ni	0.7823	0.1846	0.4810
Pb	0.5663	0.0322	0.1836
Zn	0.5707	0.0674	0.2413
As	0.4846	0.0252	0.1312
Mn	0.4184	0.0092	0.0863
Fe	0.7141	0.1339	0.3602

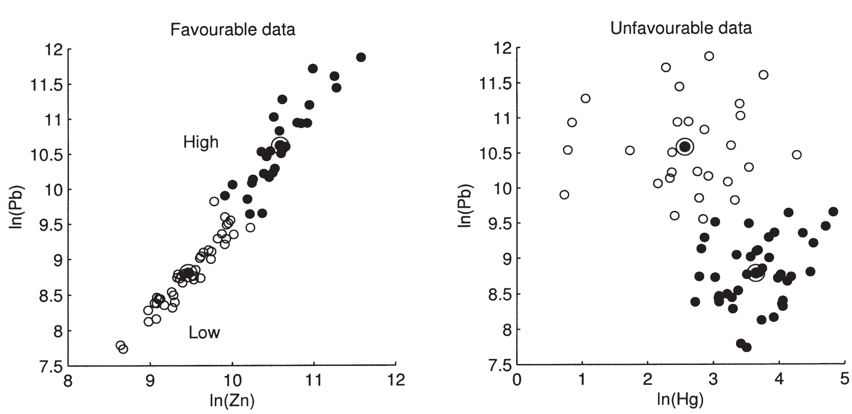

Fig. 3. Scatterplot of "favourable" and "unfavourable" data.

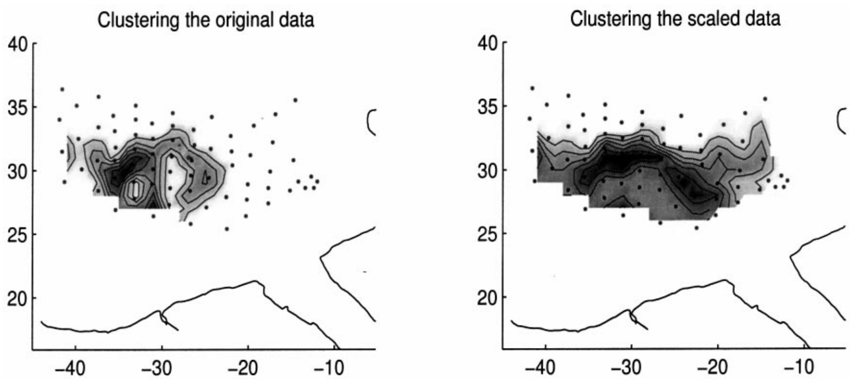

Fig. 4. Overall distribution of the 10 heavy metals in Liverpool bay calculated by hard c-means clustering.

magnitude higher than the others will dominate and determine the clustering result. On the other hand, if we decide to scale the data, we need to choose a scaling method [e.g., taking the logarithm, z-normalization, scaling as in Eq. (1), etc.]. Each of these methods might lead to a different clustering result.

Fuzzy c-means has also been used for clustering purposes in spatial data analysis (Markus and McBratney, 1996). It is debatable, however, what the added value of using fuzzy c-means is over that of the hard c-means.

Principal component analysis (PCA) gives results that are not easily interpretable in the general case. Here, the variance of most metals is along the same axis: high values at the disposal site, which are an order of magnitude higher than the values elsewhere. In this case the first principal component should follow the pattern found by the HCM clustering. Fig. 5 shows the distribution defined by the first and the second principal components. While the first component might correspond to metal loading, the second component is not easy to name. The top two plots show the results with the original data, and the bottom two plots with the scaled data. As the figure shows, similarly to clustering, PCA is sensitive to the type of data scaling. The difference is especially clear in the second principal component which makes its interpretation even more obscure.

Contour maps of the loading indices $LI_1, \ldots, LI_6$ for the 10 fuzzy sets $A_1, \ldots, A_{10}$ are plotted in Fig. 6. The horizontal axis is longitude between $-4.00°$ and $-3.00°$ (west of Greenwich); the vertical axis is 53° and decimal minutes north.

The LI results identify clearly a number of important patterns in the spatial data. The feature common to all six methods is the area of low metal contamination in the north west (top left) section of the sampling grid. The product method of calculation efficiently resolves the disposal ground whilst most of the other methods identify a residual south easterly movement of material from the disposal ground towards the estuary mouth. The smaller high area in the extreme eastern sector (seen most clearly in the Maximum plot) identifies a widening area at the estuary mouth where there is net deposition of suspended material delivered from industrial sources higher up the River Mersey.

The difficulty in assessing results such as the ones in this study comes from the fact that there is no benchmark against which the new solution can be matched. The *plausibility* of the results can be judged only by eye. Yet, the expert's appreciation will be biased by their attitude. *Product aggregation* clearly indicates where the highest contamination is. This loading index may be favoured by the user responsible for waste disposal. *Maximum aggregation* shows all contaminated sites even if the contamination is due to just one of the components. This loading index may be picked by the user concerned about, say, the fish diversity in the region. The bottom line is that there is no *true* loading distribution nor is there a single one that can be "useful" from all points of view. What the proposed fuzzy approach offers is a collection of indices, each one with comprehensible interpretation, thereby giving the user a chance to make an educated choice. In this respect the fuzzy sets approach to spatial data analysis has an advantage over clustering and PCA where the interpretation is not straightforward and the results are at the mercy of the data.

An Index of Toxicity can also be designed by weighting the fuzzy sets $A_1, \ldots, A_{10}$ with respect to the toxicity of the metals and then applying a proper fuzzy aggregation, if the relative toxicities are known. There are many fuzzy set connectives that can incorporate individual weights for the fuzzy sets but the problem here is more complex. Different combinations of metals could have different implications on the biota. Besides, these implications could be specific for different groups of species. Therefore, a more complex coefficient scheme has to be considered where each combination of metals has its own weight. For example, for Pb, Mn and Hg the aggregation rule should be able to account for seven toxicity coefficients: for (Pb), (Mn), (Hg), (Pb, Mn), (Pb, Hg), (Mn, Hg) and (Pb, Mn, Hg). An apt fuzzy model for this type of weighted aggregation is the fuzzy integral

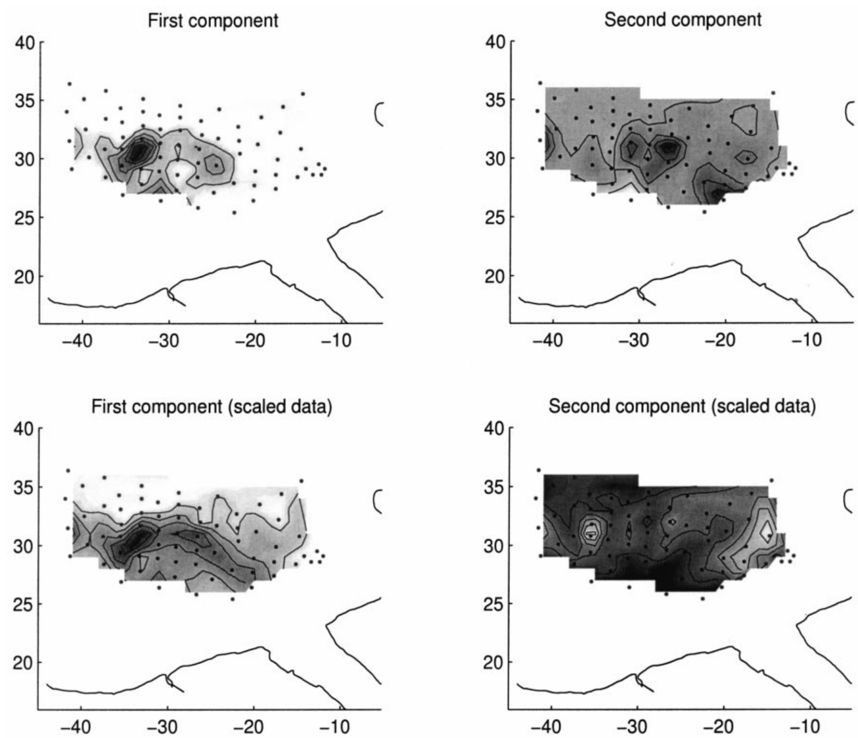

Fig. 5. Overall distribution of the 10 heavy metals in Liverpool bay by the first and second principal components of PCA.

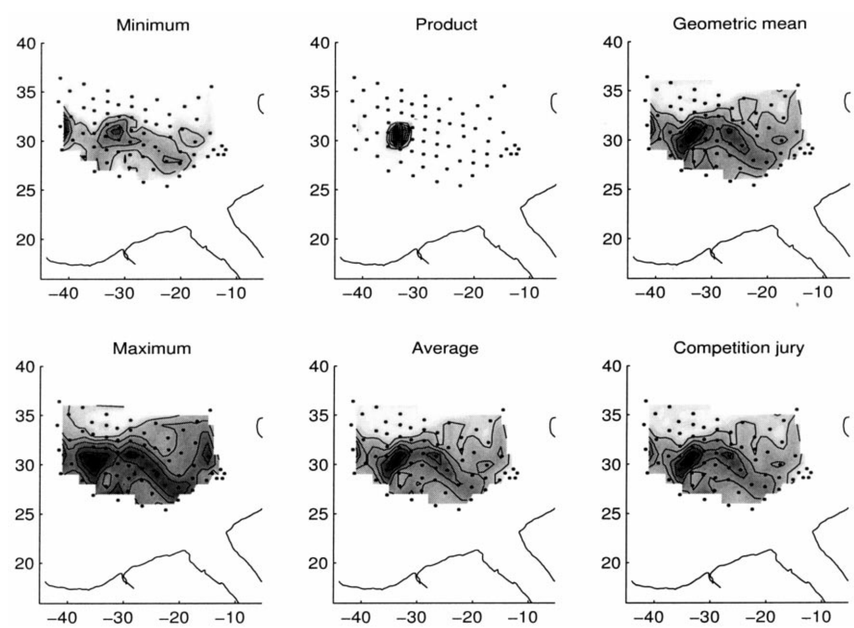

Fig. 6. Overall distribution of the 10 heavy metals in Liverpool bay calculated by the six loading indices.

(cf. Grabisch, 1995a). However, determining a set of coefficients that assess the joint effect of two or more metals on a variety of species is not a trivial task.

5. Conclusions

We show how fuzzy set theoretic aggregation operators can be used for modelling the spatial distribution of a set of variables in environmental problems, thereby providing the non-mathematical user with a simple and effective modelling tool. We applied six different fuzzy aggregation techniques to a set of heavy metal concentrations sampled from Liverpool bay, and assessed the results (geographical scatterplots of metal loading) visually. Unlike PCA and HCM, fuzzy aggregation offers a variety of plots with different information in them. On the two edges of the scale are the product aggregation, which resolves the disposal ground clearly, and the maximum aggregation, which identifies all possible sites with high contamination. The main advantage of our fuzzy model over PCA and clustering is that the results are directly interpretable in the domain context.

References

Bezdek, J., Kuncheva, L., 1999. Fuzzy pattern recognition. In: Webster, J. (Ed.), Wiley Encyclopedia of Electrical and Electronics Engineering. John Wiley & Sons, Inc, New York, pp. 173–181.

Bloch, I., 1996. Information combination operators for data fusion: a comparative review with classification. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 26, 52–67.

Camacho-Ibar, V., 1991. Trace elements and polychlorinated biphenyls (PCB) congeners in Liverpool bay sediments. Ph.D. thesis. University of Wales, Bangor.

Camacho-Ibar, V., Wrench, J., Head, P., 1992. Contrasting behaviour of arsenic and mercury in Liverpool bay sediments. Estuarine, Coastal and Shelf Science 41 (3), 241–263.

Dubois, D., Prade, H., 1985. A review of fuzzy set aggregation connectives. Information Sciences 36, 85–121.

Dubois, D., Prade, H., 1997. The three semantics of fuzzy sets. Fuzzy Sets and Systems 90, 141–150.

Duda, R., Hart, P., 1973. Pattern Classification and Scene Analysis. John Wiley & Sons, Inc, New York.

Grabisch, M., 1995a. A new algorithm for identifying fuzzy measures and its application to pattern recognition. In: Proceedings of FUZZ/IEEE'95, Yokohama, Japan., pp. 145–150.

Grabisch, M., 1995b. On equivalence classes of fuzzy connectives. IEEE Transactions on Fuzzy System 3 (1), 96–109.

Markus, J., McBratney, A., 1996. An urban soil study: heavy metals in Glebe, Australia. Australian Journal of Soil Research 34, 453–465.

Yager, R., Filev, D., 1994. Essentials of Fuzzy Modeling and Control. John Wiley & Sons, Inc, New York.

Zadeh, L., 1965. Fuzzy sets. Information and Control 8, 338–353.