Fuzzy and Crisp Mahalanobis Fixed Point Clusters

Christian Hennig[1]

Fachbereich Mathematik - SPST,
Universität Hamburg, D-20146 Hamburg, Germany,
Department of Statistical Science,
University College London, London WC1e 6BT, United Kingdom

Abstract. Fixed point clusters (FPCs) are based on the idea that local optima of redescending M-estimators can be used to locate clusters. FPCs satisfy a fixed point condition which means that they are data subsets that do not contain any outlier and with respect to that all other points in the data set are outliers. In this paper, outliers are defined in terms of the Mahalanobis distance. Crisp FPCs (where outlyingness is defined with a 0-1 weight function) are compared to fuzzy FPCs where outliers are smoothly downweighted. An algorithm to find substantial crisp and fuzzy FPCs is proposed, the results of a simulation study and a data example are discussed.

1 Introduction

Fixed point clusters (FPCs) are based on the idea that local optima of redescending M-estimators can be used to locate clusters. In this paper, FPCs for p-dimensional data based on the Mahalanobis distance are treated. Assume $p = 1$ for a moment to illustrate the principle. An M-estimator T_M for location in a real-valued data set $\mathbf{x} = (x_1, \ldots, x_n)$ is defined by

$$\sum_{i=1}^{n} \rho\left(\frac{x_i - T_M}{S}\right) \stackrel{!}{=} \min \text{ or } \sum_{i=1}^{n} \psi\left(\frac{x_i - T_M}{S}\right) \stackrel{!}{=} 0, \qquad (1)$$

where S is a scale estimator, ρ is a positive function with $\rho(0) = 0$ (usually nondecreasing in $|x|$) and $\psi = \rho'$ (possibly piecewise). It can be shown easily (Huber (1981), p. 183 ff.) that the second equation in (1) is equivalent to

$$T_M \stackrel{!}{=} \frac{\sum_{i=1}^{n} \omega\left(\frac{(x_i - T_M)^2}{S^2}\right) x_i}{\sum_{i=1}^{n} \omega\left(\frac{(x_i - T_M)^2}{S^2}\right)} \text{ with } \omega(y) = \frac{\psi(\sqrt{y})}{\sqrt{y}}, \qquad (2)$$

i.e., T_M is a weighted mean of the observations with weights depending on T_M itself. An M-estimator is called "redescending" if $\psi(x) = 0$ for x outside some interval $[-c, c]$. The interesting feature for cluster analysis is the invariance of (2) against changes in the data in the areas where $\omega = 0$. If T_M solves (2) and the points with $\omega > 0$ are well separated, i.e., they form a cluster, the solution is independent of the structure of the data far away from this cluster. Thuss, clusters of a certain shape (compatible with ω) can be found if there are similar clusters elsewhere in the data (in which case (2) will have multiple solutions), but also if the rest of the data consists of subsets with different shapes, is heterogeneous or contains extreme outliers.

The scale estimator S is often chosen as a preliminary robust estimator such as the MAD, but this is not adequate for cluster analysis, because such an estimator depends always on more than 50% of the data points and cannot yield a valid estimate for the scale of smaller clusters. The use of multiple local solutions for cluster analysis requires a scale estimation that adapts itself to the local cluster. $w_i = \omega\left(\frac{(x_i - T_M)^2}{S^2}\right)$ can then be interpreted as a cluster membership indicator. The most natural way to define a cluster-adapted scale is

$$S^2 \stackrel{!}{=} \frac{\sum_{i=1}^n \omega\left(\frac{(x_i-T_M)^2}{S^2}\right)(x_i - T_M)^2}{\sum_{i=1}^n \omega\left(\frac{(x_i-T_M)^2}{S^2}\right)}. \tag{3}$$

Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be a data set consisting of p-dimensional points. With $n(\mathbf{w}) = \sum_{i=1}^n w_i$, $T(\mathbf{w}) = \frac{1}{n(\mathbf{w})}\sum_{i=1}^n w_i \mathbf{x}_i$ and $\mathbf{S}(\mathbf{w}) = \frac{1}{n(\mathbf{w})}\sum_{i=1}^n w_i(\mathbf{x}_i - T(\mathbf{w}))(\mathbf{x}_i - T(\mathbf{w}))'$ for any weight vector $\mathbf{w} = (w_1, \ldots, w_n) \in [0,1]^n$, the fixed point condition (2)/(3) can be written as a fixed point condition for the cluster indicator:

Definition 1. An indicator vector $\mathbf{w} \in [0,1]^n$ is called **Mahalanobis FPC indicator** w.r.t. a decreasing weight function ω with $\omega(0) = 1,\ y \geq c \Rightarrow \omega(y) = 0$ for some $\infty > c > 0$, if for $i = 1, \ldots, n$:

$$\mathbf{w} = \left(\omega\left[(\mathbf{x}_i - T(\mathbf{w}))'\mathbf{S}(\mathbf{w})^{-1}(\mathbf{x}_i - T(\mathbf{w}))\right]\right)_{i=1,\ldots,n}. \tag{4}$$

$\frac{(x_i-T_M)^2}{S^2}$ has been replaced by the Mahalanobis distance $(\mathbf{x}_i - T)'\mathbf{S}^{-1}(\mathbf{x}_i - T)$ for $p > 1$. The connection between weight functions and ψ-functions for multivariate M-estimators of location and scatter is a bit more complicated than for $p = 1$, but leads to an analogous fixed point equation, see Hampel et al. (1986, p. 289). The resulting estimators are affine equivariant and the cluster indicators are invariant under linear transformations (Hampel et al. (1986, p. 283)).

FPC analysis has been introduced for linear regression data by Hennig (1997, 2002, 2003). Mahalanobis FPCs have been used by Hennig and Christlieb (2002) to validate the findings of a graphical cluster search. This idea of using local optima of redescending M-estimators for clustering goes back to Hampel (1975) and has become somewhat popular in recent literature on unsupervised pattern recognition, see, e.g., Müller and Garlipp (2005).

The FPC concept is a local cluster concept. The FPC property of a data subset depends on its shape and on its separateness, but not on data points far away from it. FPCs may overlap, and a fixed point clustering does not need to be exhaustive. It may be interesting to compare the FPC concept with more classical concepts of overlapping clustering, which are a research topic of Professor Gaul (e.g., Gaul and Schader (1994), Baier, Gaul and Schader (1997)).

The present paper focuses on two aspects:

- The choice of the weight function (especially if it should be chosen crisp or fuzzy) is discussed in Section 2.
- An algorithm is proposed to find all relevant Mahalanobis FPCs in a data set (Section 3).

In Section 4, the results of a simulation study are given where crisp FPCs have been compared with fuzzy FPCs. In Section 5, the methods are applied to a real data set.

2 Crisp and fuzzy weight functions

In previous papers (Hennig (1997, 2002, 2003), Hennig and Christlieb (2002)), the weight function ω has been chosen as 0-1-valued:

$$\omega_\alpha(y) = 1(y \leq \chi^2_{p;1-\alpha}), \tag{5}$$

$\chi^2_{p;1-\alpha}$ denoting the $1-\alpha$-quantile of the χ^2-distribution with p degrees of freedom. This corresponds to the definition of an α-outlier region w.r.t. the $N_p(\mathbf{a}, \mathbf{\Sigma})$ (normal)-distribution by Becker and Gather (1999). α-outliers w.r.t. some distribution are defined as points lying in low density regions so that the probability of a point to be an outlier under the distribution is not larger than a small α:

$$A_\alpha = \{\mathbf{x} : \ (\mathbf{x} - \mathbf{a})^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{a}) > \chi^2_{p;1-\alpha}\}.$$

If ML estimators are plugged into this definition, crisp FPCs are exactly the data subsets that do not contain any outlier and with respect to that all other points in the data set are outliers. This formalizes two properties necessary to call a data subset a “cluster”, namely homogeneity (no outliers included) and separation (all other points are outliers). In terms of statistical models, an FPC can then be interpreted as a central part of a well separated normal subpopulation, which can be said to be the “cluster prototype model” for this method. The corresponding estimators work well for more general subpopulations that are at least approximately symmetric and unimodal. Differences in tail areas do not matter because of $\omega = 0$.

The choice (5) for the (1-dimensional) location estimator T_M, corresponds to the skipped mean, which is among the optimal B-robust redescending M-estimators (minimum variance in the normal model under a bound on the gross error sensitivity, Hampel et al. (1986), pp. 87, 158). Nevertheless, it

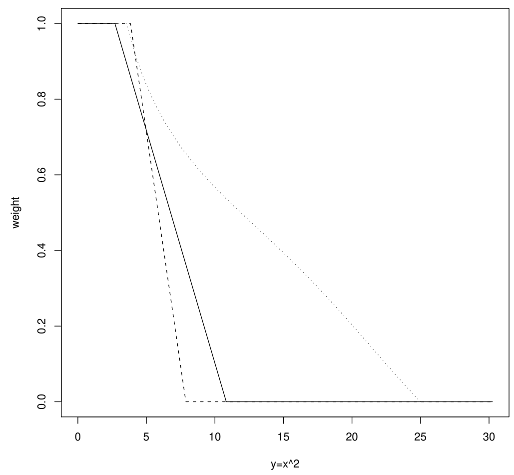

Fig. 1. Weight functions $\omega_{1,0.9,0.999}$ (solid), $\omega_{1,0.95,0.995}$ (dashed) and corresponding to a tanh-estimator (dotted) recommended by Hampel et al. (1986).

has been demonstrated by theoretical considerations and extensive simulations (Andrews et al. (1972), Hampel et al. (1986), p. 166, Hampel (2002)) that smooth rejection of outliers as done with weight functions that decrease continuously to 0 leads to more stable and efficient location estimators. FPC analysis may therefore be improved by replacing the crisp weight function ω_α by a continuously decreasing weight function. This is recommended by Hampel (2002). Nevertheless, the situation is more complicated in cluster analysis than for the estimation of the parameters of a homogeneous majority of the data. Good choices of ω for the estimation problem of a single population have a wide region of decrease of the weight function to include information not only from central points, but also from points in a more or less large distance from the center. The dotted line in Figure 1 is the weight function corresponding to one of the recommended ("optimal V-robust") M-estimators in Hampel et al. (1986). But if the center a_1 of a cluster should be estimated in the presence of another cluster with a center a_2 with, say, $|a_1 - a_2|^2 = 25S^2$, S^2 being the within-cluster variance for both clusters, about half of the points of the second cluster affect the estimation of the first one if the shown tanh-estimator is used, and the two clusters are not even separable by the fixed point condition (4). Therefore, the separation of clusters that are not extremely far away from each other requires much faster decreasing weight functions. Weight functions of the form

$$\omega_{p,\beta,\gamma}(y) = \begin{cases} 1 & : y < c_1 = \chi^2_{p;1-\beta} \\ 0 & : y > c_2 = \chi^2_{p;1-\gamma} \\ \frac{c_2-y}{c_2-c_1} : & c_1 \leq y \leq c_2 \end{cases} \tag{6}$$

are a simple alternative. $\omega_{1,0.9,0.999}$ and $\omega_{1,0.95,0.995}$ are shown in Figure 1. The simulations summarized in Section 4 show that these are reasonable choices, and that the decrease of $\omega_{1,0.9,0.999}$, though much faster than that of the tanh-estimator, is still a bit too slow to separate close, but clearly visible clusters.

3 An algorithm for Mahalanobis fixed point clusters

It is impossible to check (4) for all possible $\mathbf{w}$. But FPCs can be found by a fixed point algorithm defined by

$$\mathbf{w}^{k+1} = \left(\omega\left[(\mathbf{x}_i - T(\mathbf{w}^k))'\mathbf{S}(\mathbf{w}^k)^{-1}(\mathbf{x}_i - T(\mathbf{w}^k))\right]\right)_{i=1,\ldots,n}. \tag{7}$$

This algorithm is shown to converge toward an FPC in a finite number of steps for crisp Mahalanobis FPCs in Hennig and Christlieb (2002). Up to now, no such proof exists for fuzzy FPCs. However, using a weight function of the form (6), I have carried out about a million of runs of this algorithm on real and simulated data sets, and I have never seen any convergence problems, though I know of artificial counterexamples with a particular different continuous weight function. The remainder of this section applies to crisp as well as to fuzzy Mahalanobis FPCs.

The main problem is the choice of reasonable starting configurations $\mathbf{w}^0$. While there are many very small FPCs, which are not very meaningful (e.g., all sets of p or fewer points yield FPCs), FPC analysis aims at finding all substantial FPCs, where "substantial" means all FPCs corresponding to well separated, not too small data subsets which give rise to an adequate description of the data as a central part of a normal (or at least symmetric and unimodal) distribution. For clusterwise regression, this problem is discussed in depth in Hennig (2002). For Mahalanobis FPCs, the following strategy is suggested.

For every point of the dataset, one initial configuration is chosen, so that there are n runs of the algorithm (7). Initial configurations are always chosen as crisp indicators. For every point, the p nearest points in terms of the Mahalanobis distance w.r.t. $\mathbf{S}(1,\ldots,1)$ are added, so that there are $p+1$ points. Because such configurations often lead to too small clusters, the initial configuration is enlarged to contain n_{start} points. To obtain the $(p+2)$nd to the n_{start}th point, the covariance matrix of the current configuration is computed (new for every added point) and the nearest point in terms of the new Mahalanobis distance is added.

The stability or "significance" of an FPC $\mathbf{w}$ can be measured by $r(\mathbf{w}) = t(\mathbf{w})/n(\mathbf{w})$, where $t(\mathbf{w})$ is the number of times this FPC has been found with the algorithm described above (for fuzzy FPCs, there has to be some numerical convergence criterion and an FPC can only be reproduced up to a prespecified accuracy). Usually some of the FPCs are unstable, and more than one FPC may correspond to the same significant pattern in the data. Therefore the number of FPCs is reduced. A similarity matrix is computed between FPCs. Similarity between two FPC indicators $\mathbf{v}, \mathbf{w}$ is defined as follows:

$$s(\mathbf{v}, \mathbf{w}) = \frac{n(i(\mathbf{v}, \mathbf{w}))}{n(u(\mathbf{v}, \mathbf{w}))},$$

where $i(\mathbf{v}, \mathbf{w})$ denotes the indicator of the (fuzzy) intersection (elementwise minimum) and $u(\mathbf{v}, \mathbf{w})$ denotes the indicator of the (fuzzy) union (elementwise maximum). In the literature, s is often called "Jaccard coefficient". Then, the single linkage groups of FPCs of level s_0 are computed, i.e. the connectivity components of the graph where edges are drawn between FPCs with similarity larger than s_0. This is the most easily computable clustering based on similarities. Because groups of FPCs usually seem to be well separated in terms of s, the choice of the clustering method is not too crucial for this task. Groups of FPCs whose members are found often enough (which means that the sum of the $r(\mathbf{w})$ of the members of the group is larger than some cutoff value r_0) are considered as "stable". A representative FPC is chosen for every stable group of FPCs by maximizing $r(\mathbf{w})$.

Some tuning constants have to be chosen for this algorithm: α, β and γ for the weight functions, n_{start}, s_0, and r_0. Suggestions for these constants have been assessed by the simulation study that is described in the next section. The algorithm is implemented in the add-on package "fpc" for the statistical software system R (`www.R-project.org`).

4 Simulation results

A simulation study has been carried out to compare crisp and fuzzy FPCs and different choices of the tuning constants.

The algorithm described above has been applied to a setup with $p = 6$ and four simulated subpopulations. The separation of the clusters has been governed by a parameter d, by which all cluster centers have been multiplied while the covariance structure remained unchanged. d has been chosen as 1.5, 1, 0.85, 0.7. Variable 5 has been homogeneous normal noise, variable 6 has been homogeneous noise from a t_2-distribution. With respect to variable 1-4, subpopulation 1 consisted of 100 points from an $N_4(\mathbf{a}, \mathbf{\Sigma})$-distribution with $\mathbf{a} = d * (0, 2, 0, 2)$ and $\mathbf{\Sigma} = 0.1\mathbf{I}_4$ ($\mathbf{I}_4$ is the unit matrix). Subpopulation 2 consisted of 200 points from an $N_4(\mathbf{a}, \mathbf{\Sigma})$-distribution with $\mathbf{a} = d * (3, 3, 3, 3)$. The variances have been 0.5, the off-diagonal entries of $\mathbf{\Sigma}$ have been 0.25.

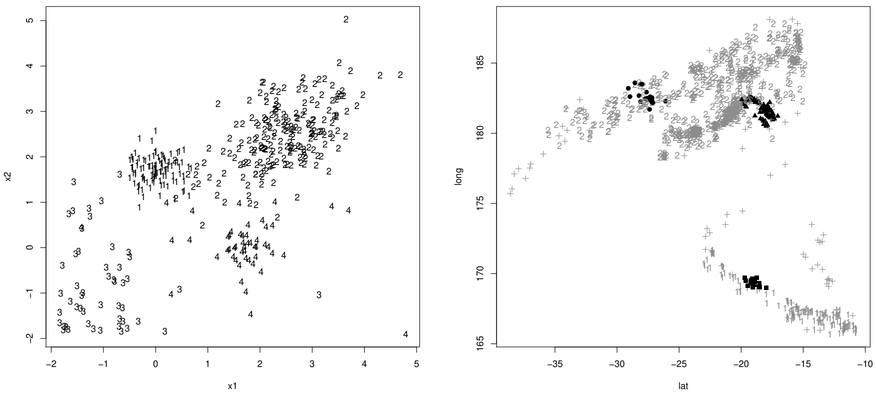

Fig. 2. Left: first two dimensions of simulated data set ($d = 0.85$). Numbers denote cluster memberships. Right: first two dimensions of quakes data with fuzzy FPCs ($\beta = 0.9,\ \gamma = 0.999$). Points denoted by "+" are not part of any FPC. The black squares, triangles and circles denote FPCs no. 3, 4 and 5, which are subsets of FPC no. 1 or 2.

Subpopulation 3 (50 points) has been generated independently from 4 exponential distributions with $\lambda = 1$ shifted to a center of $-d$. Subpopulation 4 (50 points) has been generated by a multivariate t_2-distribution with center $d * (2, 0, 2, 0)$ and the same scatter matrix as for cluster 1. Variables 1 and 2 of such a data set ($d = 0.85$) are shown on the left side of Figure 2.

A desirable solution for a simulated data set consists of 2 FPCs for the two normal clusters no. 1 and 2 and one FPC containing the non-outlying points (but not all) of the t_2-cluster no. 4. There may or may not be an FPC corresponding to some core points of subpopulation 3, of which the shape deviates strongly from the cluster prototype model (subsamples with non-normal shape occur often in complex data), and there should not be other FPCs except of reasonable unions of the clusters.

Additionally there have been simulations of two different types of homogeneous data sets, namely from a single $N_p(\mathbf{0}, \mathbf{I}_p)$-distribution and a single multivariate t_2-distribution with $p = 2, 5, 10$ and $n = 50, 200, 500$ ($n = 100$ instead of $n = 50$ has been used for $p = 10$). Simulations on homogeneous data show the tendency of the methods to generate meaningless FPCs (there should be always only one FPC in these situations). Because FPCs are invariant against addition and deletion of points in areas where the weight function is 0, the results generalize to data sets with further points and subpopulations that are outlying w.r.t. the original FPC. For every combination of parameters and tuning constants there have been 100 simulation runs.

α for crisp FPCs has been chosen as 0.95 and 0.99 ($\alpha = 0.95$ yielded always too many FPCs). (β, γ) for fuzzy FPCs have been chosen as $(0.9, 0.999)$,

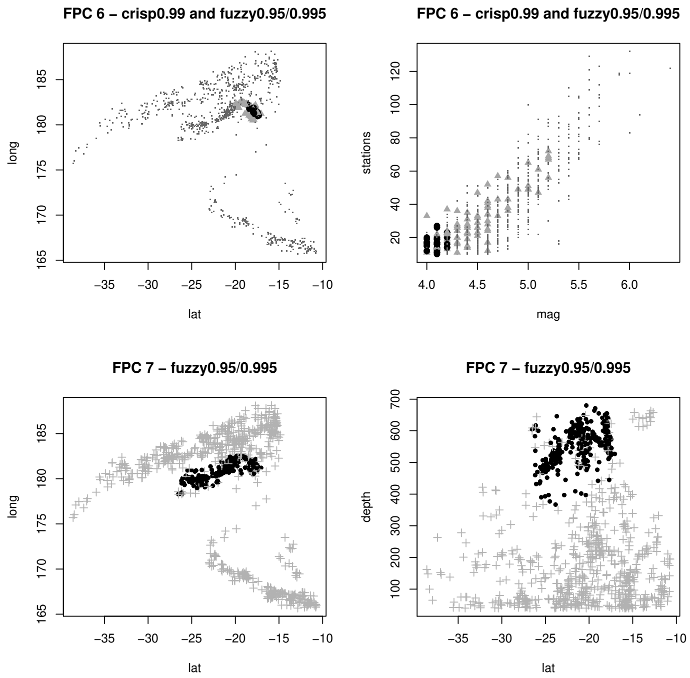

Fig. 3. Above: scatterplots with FPC no. 6 found by crisp0.99 and by fuzzy0.95/0.995 (black). Gray triangles denote FPC no. 4 from Fig. 2, of which FPC no. 6 is a subset. Below: scatterplots with FPC no. 7 only found by fuzzy0.95/0.995.

$(0.9, 0.995)$, $(0.9, 0.99)$, $(0.95, 0.995)$ (only the first and the last choice did not almost always find too many FPCs). r_0 has been chosen as 0.1 and 0.3 ($r_0 = 0.3$ proved necessary to find a reasonable small number of FPCs). n_{start} has been chosen as $6 + 2p, 12 + 2p, 18 + p$ (the latter choice delivered always the best results). s_0 has been chosen as $17/23 = 0.739$ always, which means that two sets with 20 points both are regarded as "similar" if they have at least 17 points in common.

The full simulation results will be published elsewhere. Here are the most interesting results with the best three methods ("crisp0.99", "fuzzy0.95/0.995" and "fuzzy0.9/0.999" with $r_0 = 0.3$, $n_{start} = 18 + p$):

- For the homogeneous populations, the number of found FPCs is sometimes too high for $p = 2$: fuzzy0.9/0.999 delivers on average 1.02 FPCs for $n = 50$, 1.05 FPCs for $n = 200$ and 1.13 FPCs for $n = 500$ (normal distribution; the t_2 results are very similar). For $n = 500$, this is by far the best value: crisp0.99 yields 2.10 FPCs and fuzzy0.95/0.995 leads to 1.67 FPCs. For $p = 5$ and $p = 10$, the fuzzy methods work well, but crisp0.99 delivers still too many FPCs for $p = 5, n < 500$ and often not

even one of them for $p = 10, n < 500$. It may be worthwhile to choose α, β and γ dependent of p or n (namely larger than used here in the cases where problems arose).
- In the clustered data set, the quality of the recovery of the clusters has been measured by the maximum Jaccard similarity and the minimum mean squared error of the center and covariance matrix estimation among the found FPCs. It turned out that all three methods were able to find cluster 1 for $d = 1.5, 1, 0.85$, but for $d = 0.7$, fuzzy0.9/0.999, which requires the strongest simulation, has been seriously biased. Cluster 2 has been successfully separated from the rest of the data for $d = 1.5$ by all methods (fuzzy0.9/0.999 yields the best results if the patterns are separated enough) and for $d = 1$ by crisp0.99 and fuzzy0.95/0.995. For cluster 4, all methods work for $d = 1.5, 1$. For $d = 0.85$, fuzzy0.9/0.999 fails in most cases while the other methods often work and crisp0.99 is a bit better than fuzzy0.95/0.995.

As a conclusion, fuzzy0.9/0.999 is the most stable method, but requires the strongest separation. In terms of the ability to find clusters that are not strongly separated, crisp0.99 is a bit better than fuzzy0.95/0.995, but the latter fuzzy method seems to be a good compromise because it delivers clearly better and more stable parameter estimators on well enough separated data sets, as could be expected from the experience with redescending M-estimators.

5 A data example

The methods crisp0.99, fuzzy0.9/0.999 and fuzzy0.95/0.995 have been applied to a 5-dimensional data set of 1000 seismic events on Fiji ("quakes"). The variables are latitude, longitude, depth, Richter magnitude and number of stations reporting. The data set comes with the R base package.

The right side of Figure 2 shows a scatterplot of latitude and longitude, where some clear patterns are visible. The scatterplot shows the five FPCs that were found by fuzzy0.9/0.999 (points with $w_i > 0.5$ are indicated). The data seem to be separated into two clearly visible parts at longitude= 175. The cores of these two parts form the FPCs no. 1 and 2. The further three FPCs are more concentrated subsets of these clusters. Patterns which are not identical, but have a Jaccard similarity of $s > 0.9$ with these FPCs, are found by crisp0.99 and fuzzy0.95/0.995 as well. The latter two methods find a further FPC, which is plotted on the top line of Figure 3. It is a subset of FPC no. 4 of fuzzy0.9/0.999, but it is much more homogeneous in terms of the magnitude and the number of stations reported (right scatterplot). fuzzy0.95/0.995 even yields a seventh FPC, which can be seen on the bottom line of Figure 3. This FPC corresponds to a pattern that is clearly visible in the latitude/longitude scatterplot, is fairly homogeneous also in terms of depth (right scatterplot), and is geographically meaningful. This pattern has a nonlinear structure and deviates strongly from the normal cluster prototype. fuzzy0.9/0.999 seems to converge to the full cluster no. 2 if the algorithm is started from most of the member points of this pattern, while crisp0.99 seems to deliver one of the subclusters no. 4 and 6 in most cases. Thus, in this data set, fuzzy0.95/0.995 yields the most interesting results, at least from an exploratory point of view.

References

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972): *Robust Estimates of Location: Survey and Advances.* Princeton University Press, Princeton NJ.

BAIER, D., GAUL, W. and SCHADER, M. (1997): Two-Mode Overlapping Clustering With Applications to Simultaneous Benefit Segmentation and Market Structuring. In: R. Klar and O. Opitz (Eds.): *Classification and Knowledge Organization*, Springer, Berlin, 557–566.

BECKER, C. and GATHER, U. (1999): The masking breakdown point of multivariate outlier identification rules. *Journal of the American Statistical Association, 94, 947–955.*

GAUL, W. and SCHADER, M. (1994): Pyramidal Classification Based on Incomplete Dissimilarity Data. *Journal of Classification, 11, 171-193.*

HAMPEL, F. R. (1975): Beyond location parameters: Robust concepts and methods, *Bulletin of the International Statistical Institute 46 - Proceedings of the 40th Session*, 375–382.

HAMPEL, F. R. (2002): Some Thoughts about Classification. In: K. Jajuga, A. Sokolowski, H.-H. Bock (Eds.): *Classification, Clustering, and Data Analysis*, Springer, Berlin, 5–26.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986): *Robust Statistics.* Wiley, New York.

HENNIG, C. (1997): Fixed Point Clusters and their Relation to Stochastic Models. In: R. Klar and O. Opitz (Eds.): *Classification and Knowledge Organization*, Springer, Berlin, 20–28.

HENNIG, C. (2002): Fixed point clusters for linear regression: computation and comparison. *Journal of Classification 19, 249–276.*

HENNIG, C. (2003): Clusters, Outliers, and Regression: Fixed Point Clusters. *Journal of Multivariate Analysis, 86, 183–212.*

HENNIG, C. and CHRISTLIEB, N. (2002): Validating visual clusters in large datasets: fixed point clusters of spectral features. *Computational Statistics and Data Analysis, 40, 723–739.*

HUBER, P. J. (1981): *Robust Statistics.* Wiley, New York.

MÜLLER, C. H. and GARLIPP, T. (2005): Simple consistent cluster methods based on redescending M-estimators with an application to edge identification in images. *Journal of Multivariate Analysis, 92, 359–385.*