



Robust inference for parsimonious model-based clustering

Francesco Dotto and Alessio Farcomeni

Sapienza - University of Rome, Rome, Italy

ABSTRACT

We introduce a robust clustering procedure for parsimonious model-based clustering. The classical `mclust` framework is robustified through impartial trimming and eigenvalue-ratio constraints (the `tcclust` framework, which is robust but not affine invariant). An advantage of our resulting `mtclust` approach is that eigenvalue-ratio constraints are not needed for certain model formulations, leading to affine invariant robust parsimonious clustering. We illustrate the approach via simulations and a benchmark real data example. R code for the proposed method is available at <https://github.com/afarcome/mtclust>.

ARTICLE HISTORY

Received 7 July 2018

Accepted 28 November 2018

KEYWORDS

`mclust`; `tcclust`; factor model; groups; affine invariance

1. Introduction

Generally speaking multivariate methods can be inconsistent in presence of contamination. This can be particularly problematic with unsupervised classification methods, where even the difference between a small cluster and a scattered set of outliers may be very subtle. Even a single malicious observation can break down parameter estimates and completely hide the groupwise structure of the data [1]. Formal robustness to departures from model assumptions can be achieved in several ways, one of which being impartial trimming. The observations that are farthest from the bulk of any cluster are discarded simultaneously to cluster (parameter) estimation. We point the reader to Farcomeni and Ventura [2], Farcomeni and Greco [3], part II, and to Dotto [4] for general review of robust methods and especially of robust clustering. In robust clustering impartial trimming is additionally combined with some constraint to avoid spurious maximizers of the objective function [5,6], resulting in the `tcclust` framework. Popular constraints are eigenvalue-ratio ones, which place bounds on the ratio of eigenvalues of the cluster-specific scatter matrices. There are two main limitations of the approaches discussed. First, choice of the tuning parameter is somehow subjective. Secondly, affine invariance is lost. Several other possibilities are listed in Farcomeni and Greco [3], but alternatives are difficult to implement or similarly lead to loss of affine invariance. See also Coretto and Hennig [7] and Cerioli et al. [8].

In this paper we focus on parsimonious model-based clustering methods, which are widely used. These are based on the assumption that data arise from a finite mixture of K parametric distributions. Each of these is often assumed to be multivariate normal, or at

least elliptically-contoured. For general elliptically-contoured mixture components see e.g. McLachlan and Peel [9], Zhang and Liang [10] or Bagnato et al. [11]. Finally, a structure is assumed for the cluster-specific covariance matrices. This is the well known `mclust` framework [12–15]. Robust versions of `mclust` exist. A closely related paper is that of Punzo and McNicholas [16], which deals with the presence of *mild* outliers. Their approach is very flexible but (admittedly) can not deal with gross outliers and has no formal robustness properties. Another approach, that is described in Fraley and Raftery [15], involves specifying a number of outliers which are then modelled by assignment to a uniform mixture component.

We show that commonly used (and easily interpretable) constraints on the covariance matrices in parsimonious model-based clustering can sometimes replace eigenvalue-ratio or similar constraints, thus overcoming the issue of invariance and that of selecting one of the tuning parameters. A clear advantage is that just by trimming one can obtain a model that is robust, affine invariant, parsimonious, and interpretable. Invariance properties with respect to a certain class of transformations exist whenever once data are transformed within the class, the optimal estimates are changed through the same transformation. For example translation invariance means that adding a constant to all measurements leads to add the same constant to the elements of each centroid.

We outline how to estimate each model formulation, by building on the `mclust` and `tclust` frameworks. The resulting `mtclust` procedure can be seen as a trimmed (and formally robust) version of `mclust`, or as a shape-constrained (and hence, more flexible, parsimonious, affine invariant, and interpretable) version of `tclust`. Shape restrictions are readily imposed within the M-step of the estimation algorithm, with little or no computational overhead.

The rest of the paper is as follows: in the next section we recall several different parsimonious model-based clustering formulations from Fraley and Raftery [15], summarizing what constraints are used in each case and the properties of the resulting estimators. In Section 3 we describe the `mtclust` methodology and show how to obtain robust parameter estimates. In Section 4 we report our simulation studies, and an application to a real data set is described in Section 5. Concluding remarks are in Section 6.

R code for the proposed `mtclust` algorithm, and the step-by-step data analysis for our real data example is available at <https://github.com/afarcome/mtclust>.

2. Set up

Let us consider a sample point $\mathbf{x}_i^T \in \mathbb{R}^d$, $i = 1, \dots, n$, with density given by the finite mixture model:

$$f(\mathbf{x} | \boldsymbol{\theta}) = \sum_{j=1}^K \pi_j \phi(\mathbf{x}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), \quad (1)$$

where the number of components K is known and $\boldsymbol{\theta}$ denotes the overall vector of parameters. In (1) π_j denotes the *a priori* component probability, and $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ the density of a d -variate Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For the (soft) robust version of `mclust`, an additional $K+1$ th component is included whose distribution is a uniform in a subset of \mathbb{R}^d .

Celeux and Govaert [13] consider the eigenvalue decomposition given by:

$$\Sigma_j = \lambda_j \mathbf{D}_j \mathbf{A}_j \mathbf{D}_j^T \tag{2}$$

where $\lambda_j = |\Sigma_j|^{1/d}$ is a measure of the volume of the k th cluster, \mathbf{A}_j is an orthogonal matrix with the eigenvalues of Σ_j on the diagonal and it describes the shape of each cluster; and \mathbf{D}_j is a matrix whose columns are given by the eigenvectors of Σ_j , and it summarizes the direction of each cluster. Of course, there is no difference between the completely general heterogeneous clustering model (1) and (2) until constraints are imposed. These usually take the form of *homogeneity* (over K) constraints of some (or even all) of the elements of (2).

Combining all the possible assumptions regarding scale, volume and orientation, Celeux and Govaert [13] describe 14 different models. A shorter list can be found in Fraley and Raftery [15]. We make a summary in Table 1, where we report the model name as commonly referred (and as used in R library `mclust`), the final parameterization of Σ_j , cluster shapes and properties of invariance of the solutions (see the introduction for more details on invariance). The simplest model, EII, involves spherical clusters and its solution corresponds to homogeneous model-based clustering and K -means. The solution is invariant only with respect to isometric transformations (that is, preserving distances). Model VVV, on the other hand, corresponds to the unconstrained case (1) where Σ_j is arbitrary. Note that these models are usually fit using an EM algorithm, while we will use CEM for our robust parsimonious procedure.

It shall be noted that an open issue with model VVV (and also some other models in Table 1) is that spurious solutions may arise. Spurious solutions are cases in which the maximum of the likelihood corresponds to a solution far from the true underlying clustering model. They can happen if singleton clusters can be estimated, or if a subset of the data lie in a lower dimensional space. They correspond to unbounded likelihoods, usually with unbounded cluster volumes. We give an example in Figure 1, panel (a).

Table 1. Models considered in this paper with indication of their requirements of eigenvalue-ratio (ER) constraints and invariance properties.

Model name	Parametrization	ER	Invariance	M step
EII	λI	Not required	Isometric transformations	CF
VII	$\lambda_j I$	Not required	Isometric transformations	CF
EEI	λA	Not required	Scaling	CF
VEI	$\lambda_j A$	Not required	Scaling	IP
EVI	λA_j	Not required	Traslation	CF
VVI	$\lambda_j A_j$	Required	Traslation	CF
EEE	λDAD^T	Not required	Linear transformations	CF
EVE	$\lambda DA_j D_j^T$	Required	Traslation	IP
VEE	$\lambda_j DAD^T$	Not Required	Linear transformations	IP
VVE	$\lambda_j D_j A_j D_j^T$	Not Required	Linear transformations	CF
EEV	$\lambda D_j A_j D_j^T$	Not required	Linear transformations	CF
VEV	$\lambda_j D_j A_j D_j^T$	Not required	Linear transformations	IP
EVV	$\lambda D_j A_j D_j^T$	Required	Traslation	CF
VVV	$\lambda_j D_j A_j D_j^T$	Required	Traslation	CF

Notes: Last column states whether a closed form (CF) is available or an iterative procedure (IP) is needed for computing the associated estimator of the covariance matrix. The subscript j indicates a cluster-specific parameter, its absence that the parameter is uniform over j .

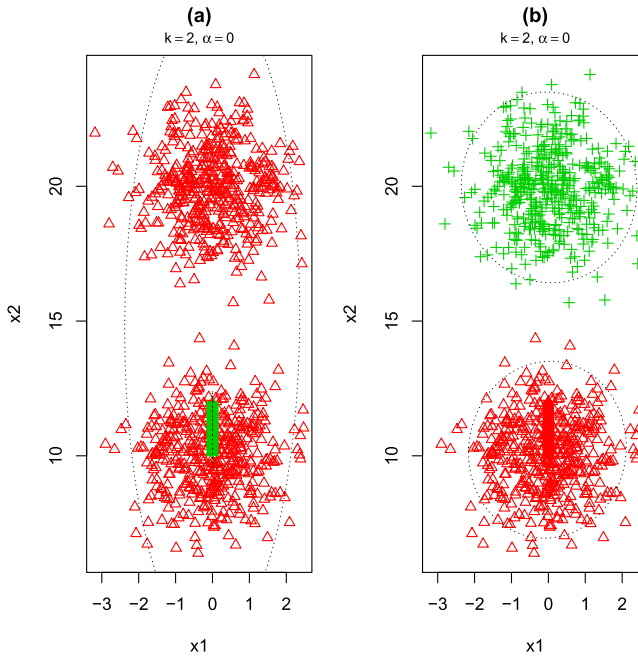


Figure 1. Spurious Maximizers' effect in unconstrained (a), and constrained (b) clustering method.

A solution is to impose (additional) constraints when estimating model parameters. Some constraints involve bounds on the determinant of the covariance matrices, or on the minimal cardinality of each cluster. See for instance Gallegos and Ritter [17–19]. Another popular constraint [5,20] requires that the ratio between the largest and the smallest eigenvalue of the scatter matrices is bounded above by a fixed constant $c \geq 1$:

$$\frac{M_n}{m_n} = \frac{\max_{j=1,\dots,K} \max_{d=1,\dots,p} \lambda_d(\Sigma_j)}{\min_{j=1,\dots,K} \min_{d=1,\dots,p} \lambda_d(\Sigma_j)} \leq c, \quad (3)$$

where $M_n = \max_{j=1,\dots,K} \max_{d=1,\dots,p} \lambda_d(\Sigma_j)$, $m_n = \min_{j=1,\dots,K} \min_{d=1,\dots,p} \lambda_d(\Sigma_j)$ and $\lambda_d(\Sigma_k)$ is the d th largest eigenvalue of Σ_j . By constraining the parameter estimates to (3), spurious solutions can not occur. An example is given in Figure 1, panel (b). An open issue is selection of the constraint limit c , which is common to several robust clustering approaches. This choice is not necessary for certain model specifications we discuss. A formal procedure for the other cases is discussed below.

As stated in Table 1 different parameterizations of the covariance matrix imply different properties in terms of invariance. Invariance is an important property as, in the lack of it, simply by scaling the data or adding a constant might otherwise affect the result in unpredictable and not interpretable ways. A notable drawback of (3) is that all properties of affine invariance are lost, as any affine transformation (except for translations) lead to different eigenvalue ratios. Additionally, a direct interpretation of constraint (3) is not straightforward as soon as $c > 1$. We note here that spurious solutions need not arise under some parsimonious formulations, basically those pooling scale or volume across clusters. Hence

constraints like (3) can be removed as soon as any of these (interpretable and often invariant) parsimonious assumptions is used. A full account on whether (3) is needed to avoid spurious solutions is given in the ER column of Table 1.

All the models above are not resistant to contamination, even under (3) or similar constraints. It can be argued that even a single outlier can completely spoil the solution, making it arbitrarily distant from the true one. Under model VVV, an effective robust clustering approach is given by the `tcclust` procedure [21]. The procedure uses impartial trimming to remove a pre-specified number of $n\alpha$ observations, and (3) to avoid spurious solutions. A central theme of this paper is that assumptions on the cluster shapes has clear advantages in terms of interpretability, and can be seen as a useful alternative to (3) in some cases. In the next section we show how to work with impartial trimming, parsimonious assumptions on the covariance matrix, and (3) *only when needed*.

3. Robust inference for constrained model-based clustering

In the following we adopt the usual (1) plus (2) formulation, leading to parsimonious model-based clustering. We robustly estimate model parameters by developing a CEM algorithm augmented with an impartial trimming step.

Our task is to maximize the classification log-likelihood

$$L(\theta) = \sum_{i=1}^n \left[\ln \sum_{j=1}^K z_{ij} \pi_j \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \right], \tag{4}$$

where z_{ij} is a binary indicator that the i th observation belongs to the j th cluster, with $\sum_j z_{ij} \leq 1$ and $\sum_{ij} z_{ij} = n(1 - \alpha)$, where α is a pre-specified trimming level. Consequently, $n\alpha$ observations are not classified into any cluster, and do not contribute to the objective function (4) or to parameter estimation. The full likelihood is a popular alternative, but in our context the classification approach is often preferred, also considering that impartial trimming is more naturally embedded within it (see Farcomeni and Greco [3] for a detailed discussion on this point). The use of a vector of weights $\pi_j, j = 1, \dots, K$ in connection with the classification likelihood leads to slightly more unbalanced cluster cardinalities with respect to the classical assumption that $\pi_j = 1/K$. This is also known as penalized CEM; see for instance Symons [22] and Bryant [23].

The model is completed by specifying a parsimonious assumption on (2) and, if needed, (3).

3.1. The `mtclust` algorithm

In order to maximize (4) under the constraints we can propose a general algorithm, which needs only minor changes for each specific model formulation.

The classification log-likelihood is maximized through an iterative algorithm, whose general iteration is made of two steps: Expectation-Classification-Trimming (ECT), Constrained Maximization (CM). Each step leads to an increase in the objective function.

The **ECT-step** proceeds by computing the posterior probabilities that each observation belongs to each cluster, as

$$T_{ij} = \frac{\pi_j f(\mathbf{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_{j=1}^K \pi_j f(\mathbf{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \quad (5)$$

Then, classification is performed by assigning each observation to the most likely cluster, that is, maximizing T_{ij} for each i with respect to j . Secondly, trimming is performed by discarding (i.e. setting $z_{ij} = 0 \forall j$) the $\lceil n\alpha \rceil$ observations with the smallest probabilities of belonging to their assigned cluster.

The **CM-step** proceeds by estimating the clusters' centres using the sample mean of each cluster given by:

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{i=1}^n z_{ij} \mathbf{x}_i}{\sum_{i=1}^n z_{ij}}.$$

Note that trimmed observations do not contribute to the estimates of the sample means as $z_{ij} = 0$ for all j . The sample covariance matrix is then estimated under model assumptions. Under model VVV there are no constraints and an initial estimate is given by

$$\hat{\boldsymbol{\Sigma}}_j = \frac{\sum_{i=1}^n z_{ij} (\mathbf{x}_i - \boldsymbol{\mu}_j)(\mathbf{x}_i - \boldsymbol{\mu}_j)'}{\sum_{i=1}^n z_{ij}}$$

This initial estimate is then adjusted to satisfy (3). A closed form for the covariance matrix exists also for most of the possible models. For instance under model EEI we have

$$\hat{\boldsymbol{\Sigma}}_j = \text{diag} \left(\frac{\sum_{j=1}^K \sum_{i=1}^n z_{ij} (\mathbf{x}_i - \boldsymbol{\mu}_j)(\mathbf{x}_i - \boldsymbol{\mu}_j)'}{\sum_{j=1}^K \sum_{i=1}^n z_{ij}} \right).$$

Some models do not admit a closed form for the estimator of the covariance matrix. Such cases, for which an iterative procedure is needed, are discussed in Celeux and Govaert [13]. If ER is needed, we embed the numerical procedure in Dykstra [24] within the covariance estimation step, that is, the initial estimate is updated by iteratively censoring the outliers. Cluster weights are then updated as $\pi_j = \sum_{i=1}^n z_{ij} / n$.

It shall be noted that, as expected, under model VVV our proposal coincides with `tclust` algorithm as proposed by Fritz et al. [25].

As any (C)EM algorithm, our proposal can be trapped into local optima. As usual, in order to reduce the risk of incurring into sub-optimal maximizers, we repeat the estimation procedure from different initial solutions. Unless model VVV is specified, a deterministic initial solution is obtained using `tclust`, applying (2), and pooling the estimates to have a feasible initial solution. The deterministic initial solution is then randomly perturbed to get the other ones.

It is straightforward to check that the proposed `mtclust` procedure inherits several properties of `tclust`, namely robustness, existence of a solution when $n \rightarrow \infty$, and consistency. Robustness of `tclust` has been discussed in detail in a few papers. In particular see Ruwet et al. [26] where it is shown that if clusters are well separated, `tclust` can bear up to $n\alpha$ outliers under fairly general assumptions. Additionally, in Farcomeni [27,28] it is discussed how `tclust` can bear up to $n\alpha/d$ scattered outliers. Since results for `tclust`

precisely correspond to `mtclust` under model VVV, Ruwet et al. [26] directly applies to our case under model VVV. Since all other models are a restriction of VVV, their likelihood can only be smaller or equal to that of VVV, their robustness directly follows. Similarly one can extend results regarding existence and consistency of `tclust`, which are discussed in García-Escudero et al. [21]. As a final remark we point out the reader to Ritter [29] for a detailed discussion regarding the identifiability of mixture models with respect to the different parametrizations available.

In order to choose the best model among the 14 possibilities outlined in Table 1, we propose using robustified versions of AIC and BIC. These are directly obtained penalizing as usual the trimmed likelihood (that is, the likelihood based on the estimated parameters and the non-trimmed observations). We will argue in the simulation study below that this simple approach allows us to compare models belonging to different families and reach a proper compromise between parsimony and goodness of fit.

We conclude this section by discussing choice of c (when needed) and α . While there are several ways to do so, we focus here on one graphical heuristic tool for choosing K and α (the `ctlcurves` method of [30]), and the recent proposal by Cerioli et al. [31] for selecting c conditional on K and α . Note that, as clearly noted in Cerioli et al. [31], a formal choice of the triplet (K, α, c) is very cumbersome as these are clearly interrelated, and hence no solutions are so far available. The procedure of `ctlcurves` involves plotting the objective function at convergence as a function of α , for different values of K . Given that the objective function increases with α , the optimal one can be chosen as the minimal value after which there are no more substantial increases in the objective function. A similar reasoning applies for K . While in our procedure fixing c is usually not needed, in case one is considering models VII or VVV `ctlcurves` must be generated for one (large) value of c . After K and α are selected, Cerioli et al. [31] propose optimization (in c) of a modified information criterion based on minus twice the observed trimmed likelihood (4) plus a penalty parameter:

$$BIC_c = -2L(\theta) + \left(Kd + K - 1 + K \frac{d(d-1)}{2} + (Kd-1)(1-1/c) + 1 \right) \log n. \quad (6)$$

Note that one could in principle iterate the use of `ctlcurves` and minimization of (6) as a function of c , to achieve possibly a better tuning. Finally note that in our experience a careful tuning of α and (when needed) c is anyway not crucial in that similar results are often obtained for a wide range of values of α and c .

4. Simulation study

4.1. Data generation

In order to validate the proposed methodology we carried out an extensive simulation study. We generated the clean part of the data from a p -variate mixture of K Gaussian components with centres equal to $\overbrace{(\mu_0 \cdot j, 0 \dots, 0)}^p$ for $j = 1, \dots, K$ and with $\mu_0 = 10$. For each scenario we generate the data $B = 500$ times and compute the Mean Square Error of the vector means and the associated misclassification rate computed as the proportion of clean observations wrongly flagged as outlying plus outliers wrongly not trimmed.

Different scenarios have been obtained by combining the following choices for data generation: $K = 2, 4$ clusters; $d = 2, 4, 6$ dimensions, different balancing, different outlier generating distributions, different trimming levels $\alpha = 0.05, 0.1$. In particular when $K = 2$ the unbalanced proportions are equal to $\pi_1 = 0.75$ and $\pi_2 = 0.25$ while, when $K = 4$, the clusters proportions are given by $\pi_j = 0.25, 0.40, 0.20, 0.15$ for $j = 1, 2, 3, 4$.

Outliers are generated using the multivariate uniform distribution. We generated two types of *mild outliers*. Let $\underline{\mathbf{X}}_C^j$ be the j th column of the clean part of the data and c be a fixed constant. The two hypercubes used for sampling outliers from the multivariate uniform distribution are fixed as follows (see Figure 2 for an example of the two types of outliers in $d = 2$ dimensions)

- Contamination type 1 ('close' mild outliers):
 $[\min\{\underline{\mathbf{X}}_C^j\}, \max\{\underline{\mathbf{X}}_C^j\} + c]$ for $j = 1, 2, \dots, d$.
- Contamination type 2 ('far' mild outliers):
 $[\min\{\underline{\mathbf{X}}_C^j\} + c, \max\{\underline{\mathbf{X}}_C^j\} + c + 30]$ for $j = 1, 2, \dots, d$.

As an illustrative purpose we report, in Figure 2, the scatterplot of the simulated data in $d = 2$ dimensions with $K = 2$ clusters imposed and the two different type of contaminating points generated. As a final remark we point out that outliers' Mahalanobis distance from the true clusters' parameter is imposed to be greater than the .99 quantile of the χ_p^2 distribution.

We also repeated data generation in each scenario using each of the 14 covariance structures outlined in Table 1. The aim of varying the covariance parametrization is twofold: it allows us to compare the capability of the procedure involved in estimating cluster

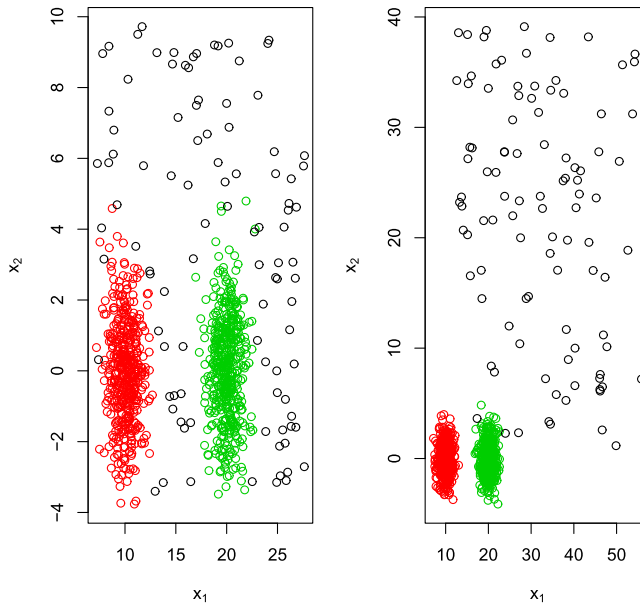


Figure 2. The two types of contamination involved in our simulation study. $\alpha = .10$, $d = 2$ and 'EEI' covariance structure.

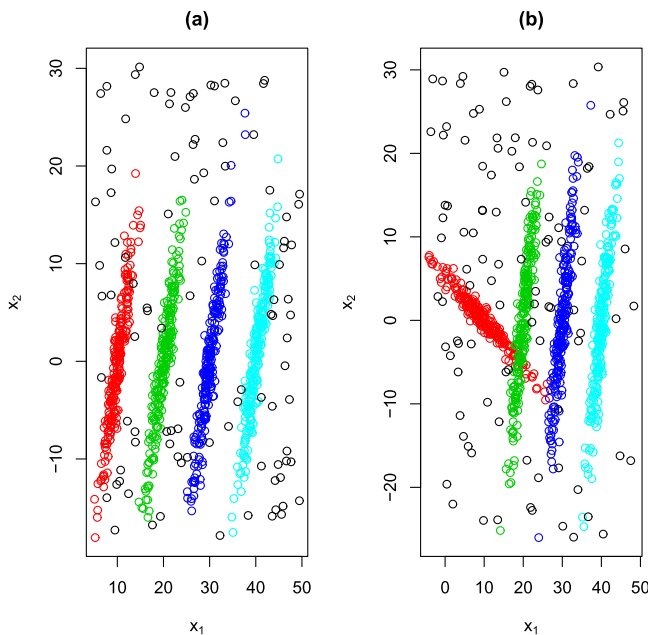


Figure 3. $d = 2, K = 4$. EEE parametrization on Panel (a), EVV parametrization on Panel (b).

parameters and outliers' recognition; secondly, by varying the orientation we also vary the degree the overlap among the clusters. See Figure 3 for a graphical illustration of this phenomenon. By combining all the different parameters listed above we obtained $2 \times 3 \times 14 \times 2 \times 2 = 672$ different simulation scenarios.

In our simulation study we use our newly proposed method in two different versions. In the first one (labeled as `mtclust`) we assume the correct parametrization of the covariance matrix is known and we fix the trimming level equal to the proportion of outliers. In the second one (labeled as `mtclust BIC`) we do not specify the parametrization: the proper model is automatically chosen by picking the one yielding the minimum BIC score. The other procedures involved for comparison are `tclust` with $c = 12$ (default setting in R), robust `mclust` where a $K+1$ th term involves noise as a first order Poisson process. See Fraley et al. [32] for more details. The proportion of outlying points is specified in advance and is set equal to the true contamination level. We also compare with Contaminated Mixtures as proposed by Punzo and McNicholas [16] and implemented in the `ContaminatedMixt` R package [33] and `otrimle` as proposed by Coretto and Hennig [7] where a 'pseudo-model where the noise is represented by an improper constant density over the whole Euclidean space' is used; and `teigen` [16,34], a new family of models based on mixtures of multivariate t distributions. Such approach may be involved against the effect of *mild* outliers due to the heavy tail of the Student t distribution.

In order to present the results we divide our simulation setting in 4 macro scenarios. These have been obtained by combining $K = 2, 4$ and the two type of outliers generated, 'close' outliers and 'far' outliers. Each macro scenario summarizes 168 simulation scenarios for which we computed the Mean squared error and the misclassification rate. Those values

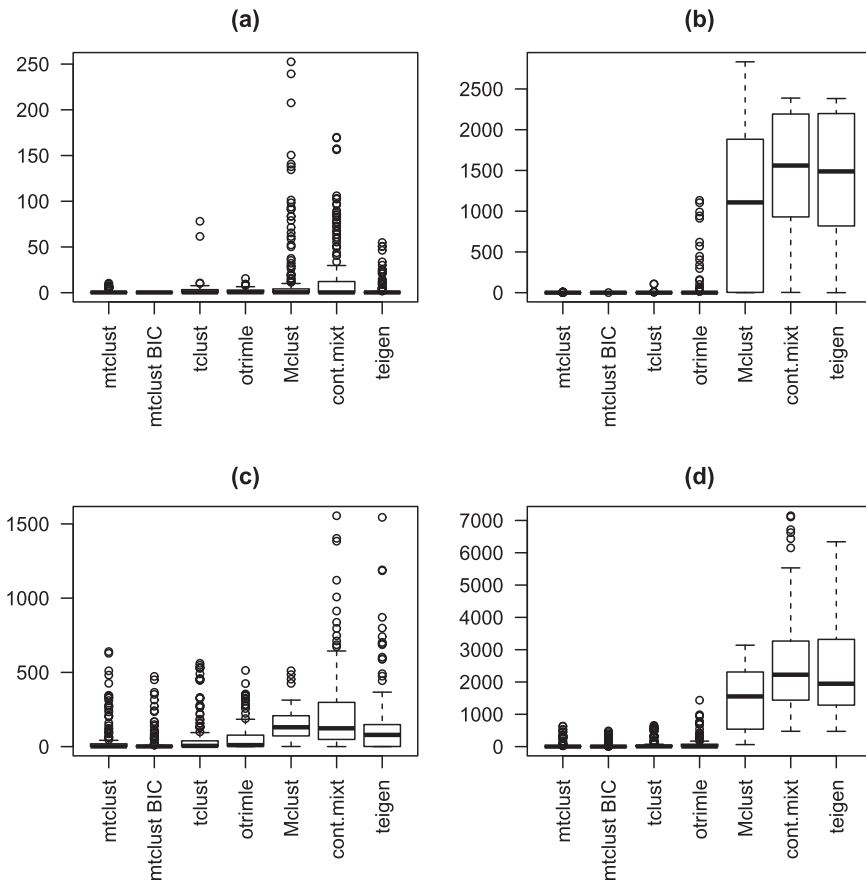


Figure 4. Boxplot of the Mean Squared Error in each macro scenario. $K = 2$ and ‘close’ outliers in Panel (a). $K = 2$ and ‘far’ outliers in Panel (b). $K = 4$ and ‘close’ outliers in Panel (c). $K = 4$ and ‘far’ outliers in Panel (d).

are reported in the boxplot plot in Figures 4 and 5. On the other hand, the precise results associated to each scenario are reported in Tables 2–13.

It is straightforward to see that `mtclust` and `mtclust-BIC` are both more resistant to outliers with respect to the robustified version of `mclust` and to the contaminated `mixt` and `teigen` models; even though we did contaminate clusters with *mild* outliers. On the other hand the performance with respect to `tclust` and to `otrimle` is often quite similar. It shall be noted though that a minor MSE decrease is seen both for `mtclust`, and more importantly for `mtclust-BIC`. This small improvement is mostly linked to the decreased standard errors that one obtains by restricting appropriately the parameter space. Additionally, in our opinion the motivation of using `mtclust` with respect to `tclust` and `otrimle` is that the former is more interpretable and, when not using (3), also affine invariant. A further limitation of (3) is that sometimes the working constraint for the eigenvalue ratio is more restrictive than the true eigenvalue ratio. In our simulations we have used an upper bound of 12 (the default value in `tclust` R package). In our data generating scenarios the true eigen value ratio is sometimes larger, which has

Table 2. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.01	0.01	0.01	0.01	0.01	0.02	0.02
EII	0.05	0.01	0.01	0.01	0.01	0.01	0.01	0.01
VII	0.1	0.03	0.03	0.03	0.03	0.03	0.13	0.06
VII	0.05	0.02	0.02	0.02	0.02	0.02	0.03	0.03
EEI	0.1	0.02	0.02	0.02	0.02	0.02	0.02	0.03
EEI	0.05	0.02	0.02	0.02	0.02	0.02	0.02	0.02
VEI	0.1	0.03	0.03	0.03	0.03	0.03	0.04	0.05
VEI	0.05	0.02	0.02	0.02	0.02	0.02	0.03	0.03
EVI	0.1	0.02	0.02	0.02	0.02	0.02	0.02	0.02
EVI	0.05	0.01	0.01	0.01	0.01	0.01	0.01	0.02
VVI	0.1	0.02	0.02	0.02	0.02	0.02	0.07	0.04
VVI	0.05	0.02	0.02	0.02	0.02	0.02	0.02	0.02
EEE	0.1	0.24	0.24	0.29	0.28	0.26	157.44	24.60
EEE	0.05	0.21	0.21	0.24	0.21	0.25	55.32	5.51
VEE	0.1	0.50	0.49	0.60	0.81	0.90	169.98	50.83
VEE	0.05	0.44	0.44	0.49	0.56	0.51	156.27	30.10
EVE	0.1	0.46	0.37	0.46	0.36	0.41	102.35	25.68
EVE	0.05	0.37	0.32	0.37	0.32	0.39	49.43	9.30
VVE	0.1	0.75	0.54	0.74	0.53	1.60	6.78	3.57
VVE	0.05	0.59	0.48	0.59	0.47	0.65	6.40	0.51
EEV	0.1	0.42	0.42	4.19	0.24	2.13	2.26	0.33
EEV	0.05	0.38	0.38	3.74	0.22	2.28	0.22	0.27
VEV	0.1	0.30	0.30	3.39	1.28	2.85	0.19	0.22
VEV	0.05	0.27	0.27	3.72	0.64	1.95	0.16	0.19
EVV	0.1	6.56	0.55	6.67	0.36	2.54	14.46	2.98
EVV	0.05	5.62	0.51	5.71	0.44	3.39	0.33	0.41
VVV	0.1	6.80	0.43	6.88	0.25	4.76	0.26	0.32
VVV	0.05	6.36	0.39	6.46	0.22	4.06	0.23	0.28
Unbalanced clusters								
EII	0.1	0.02	0.02	0.02	2.14	0.02	0.07	0.03
EII	0.05	0.01	0.01	0.01	0.01	0.02	0.02	0.02
VII	0.1	0.02	0.02	0.02	0.02	0.02	0.04	0.04
VII	0.05	0.02	0.02	0.02	0.02	0.02	0.02	0.02
EEI	0.1	0.02	0.02	0.02	4.80	0.03	0.12	0.04
EEI	0.05	0.02	0.02	0.02	3.25	0.02	0.04	0.03
VEI	0.1	0.03	0.03	0.03	0.03	0.03	0.03	0.04
VEI	0.05	0.02	0.02	0.02	0.02	0.02	0.03	0.03
EVI	0.1	0.02	0.02	0.02	0.02	0.02	0.03	0.03
EVI	0.05	0.02	0.02	0.02	0.02	0.02	0.02	0.02
VVI	0.1	0.02	0.02	0.02	0.02	0.02	0.02	0.03
VVI	0.05	0.02	0.02	0.02	0.01	0.02	0.02	0.02
EEE	0.1	0.31	0.31	0.38	37.49	0.34	169.29	3.22
EEE	0.05	0.28	0.28	0.32	32.17	0.33	82.94	0.33
VEE	0.1	0.61	0.63	0.79	0.61	0.71	2.26	5.33
VEE	0.05	0.57	0.58	0.64	0.56	0.66	0.57	0.65
EVE	0.1	0.52	0.44	0.52	0.43	0.50	3.51	5.21
EVE	0.05	0.44	0.40	0.45	0.39	0.45	0.96	0.46
VVE	0.1	0.67	0.51	0.67	0.49	9.47	50.78	8.77
VVE	0.05	0.55	0.47	0.55	0.45	3.45	1.72	0.52
EEV	0.1	0.74	0.76	4.91	15.90	3.21	3.22	0.45
EEV	0.05	0.69	0.69	5.06	14.76	2.45	0.35	0.40
VEV	0.1	0.51	0.51	2.86	0.31	2.89	0.35	0.39
VEV	0.05	0.48	0.48	3.56	0.40	1.29	0.29	0.33
EVV	0.1	10.36	1.06	10.63	0.49	4.82	4.09	0.68
EVV	0.05	9.32	0.98	10.06	0.46	4.67	0.48	0.59
VVV	0.1	6.71	0.73	6.81	0.42	6.16	1.54	0.54
VVV	0.05	7.59	0.66	7.69	0.38	3.63	0.61	0.46

Notes: $K = 2, d = 2$. Contamination type 1.

Table 3. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.02	0.02	0.02	0.02	0.02	0.03	0.03
EII	0.05	0.02	0.02	0.02	0.02	0.02	0.02	0.02
VII	0.1	0.05	0.05	0.05	0.05	0.05	0.11	0.11
VII	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06
EEI	0.1	0.06	0.06	0.06	0.05	0.06	0.07	0.08
EEI	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06
VEI	0.1	0.08	0.08	0.08	0.08	0.08	0.10	0.13
VEI	0.05	0.07	0.07	0.07	0.07	0.07	0.08	0.09
EVI	0.1	0.04	0.04	0.04	0.04	0.04	0.05	0.06
EVI	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04
VVI	0.1	0.05	0.05	0.05	0.05	0.05	0.10	0.10
VVI	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06
EEE	0.1	0.23	0.23	0.25	29.27	0.52	43.70	46.31
EEE	0.05	0.22	0.22	0.23	0.22	0.36	23.96	2.54
VEE	0.1	0.47	0.47	0.52	78.97	7.98	71.69	54.90
VEE	0.05	0.44	0.44	0.46	3.73	5.64	77.36	11.70
EVE	0.1	0.39	0.35	0.39	6.38	1.75	97.24	50.40
EVE	0.05	0.35	0.32	0.35	0.32	0.81	60.57	2.25
VVE	0.1	0.62	0.53	0.61	0.52	0.59	33.90	2.14
VVE	0.05	0.56	0.49	0.56	0.49	0.68	5.72	0.54
EEV	0.1	0.42	0.42	3.44	19.03	2.50	0.27	0.32
EEV	0.05	0.40	0.40	3.19	8.06	2.71	0.24	0.29
VEV	0.1	0.24	0.24	3.53	2.29	3.14	0.18	0.23
VEV	0.05	0.22	0.22	3.31	1.09	1.92	0.16	0.20
EVV	0.1	5.31	0.53	5.37	12.50	3.67	17.81	1.31
EVV	0.05	4.89	0.50	4.94	11.39	5.06	0.34	0.41
VVV	0.1	5.83	0.33	5.97	0.72	5.15	0.87	0.31
VVV	0.05	5.17	0.31	5.28	0.63	4.28	0.23	0.28
Unbalanced clusters								
EII	0.1	0.03	0.03	0.03	0.03	0.03	0.04	0.05
EII	0.05	0.03	0.03	0.03	0.03	0.03	0.03	0.04
VII	0.1	0.04	0.04	0.04	0.04	0.04	0.06	0.06
VII	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04
EEI	0.1	0.08	0.08	0.08	0.08	0.08	0.12	0.12
EEI	0.05	0.07	0.07	0.07	0.17	0.07	0.08	0.09
VEI	0.1	0.09	0.09	0.09	0.09	0.09	0.10	0.11
VEI	0.05	0.08	0.08	0.08	0.08	0.08	0.08	0.09
EVI	0.1	0.05	0.05	0.05	0.05	0.05	0.06	0.07
EVI	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
VVI	0.1	0.04	0.04	0.04	0.04	0.04	0.06	0.07
VVI	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.05
EEE	0.1	0.32	0.32	0.36	0.49	0.35	85.06	14.66
EEE	0.05	0.30	0.30	0.32	0.30	0.36	63.64	0.35
VEE	0.1	0.65	0.64	0.74	4.56	3.51	67.09	22.10
VEE	0.05	0.61	0.61	0.67	0.60	1.06	13.87	0.70
EVE	0.1	0.50	0.44	0.50	0.71	1.89	27.32	12.83
EVE	0.05	0.45	0.41	0.45	0.41	0.60	3.31	0.47
VVE	0.1	0.61	0.52	1.63	0.52	0.64	96.18	5.25
VVE	0.05	0.55	0.49	1.96	0.48	0.57	29.72	0.53
EEV	0.1	0.72	0.72	4.53	1.26	3.48	0.40	0.45
EEV	0.05	0.66	0.67	4.12	0.31	2.80	0.33	0.42
VEV	0.1	0.44	0.44	3.62	0.30	3.44	0.35	0.40
VEV	0.05	0.40	0.40	3.74	0.26	1.21	0.27	0.35
EVV	0.1	7.40	0.98	7.65	0.76	6.06	8.59	0.66
EVV	0.05	6.60	0.91	6.76	0.43	8.37	0.45	0.61
VVV	0.1	6.87	0.61	7.05	0.39	6.55	1.67	0.53
VVV	0.05	6.48	0.55	6.70	0.36	3.29	0.37	0.47

Notes: $K = 2$, $d = 4$. Contamination type 1.

Table 4. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.03	0.03	0.03	0.03	0.03	0.04	0.05
EII	0.05	0.03	0.03	0.03	1.70	0.03	0.03	0.04
VII	0.1	0.07	0.07	0.07	0.07	0.07	0.13	0.15
VII	0.05	0.07	0.07	0.07	0.07	0.07	0.08	0.09
EEI	0.1	0.11	0.11	0.11	52.23	0.11	0.13	0.16
EEI	0.05	0.11	0.11	0.11	2.10	0.11	0.11	0.12
VEI	0.1	0.16	0.16	0.16	0.16	0.16	0.20	0.25
VEI	0.05	0.15	0.15	0.15	0.15	0.16	0.16	0.18
EVI	0.1	0.07	0.07	0.07	3.62	0.07	0.09	0.11
EVI	0.05	0.07	0.07	0.07	1.22	0.07	0.07	0.08
VVI	0.1	0.10	0.10	0.10	0.10	0.10	0.15	0.18
VVI	0.05	0.09	0.09	0.09	0.09	0.10	0.11	0.12
EEE	0.1	0.24	0.24	0.25	140.59	1.08	40.87	20.69
EEE	0.05	0.23	0.23	0.24	150.45	0.51	16.40	0.38
VEE	0.1	0.50	0.49	0.52	239.30	15.54	79.68	33.71
VEE	0.05	0.47	0.46	0.48	252.48	9.14	85.21	1.84
EVE	0.1	0.39	0.36	0.39	93.60	2.87	87.42	23.21
EVE	0.05	0.37	0.35	0.37	91.22	0.92	76.26	0.36
VVE	0.1	0.62	0.54	0.62	0.54	0.62	58.85	0.60
VVE	0.05	0.55	0.51	0.55	0.50	0.67	27.75	0.53
EEV	0.1	0.43	0.43	3.11	98.29	2.46	0.28	0.35
EEV	0.05	0.42	0.42	3.02	101.31	2.86	0.25	0.30
VEV	0.1	0.22	0.22	3.12	20.50	2.95	0.19	0.25
VEV	0.05	0.22	0.22	2.98	28.65	1.94	0.17	0.22
EVV	0.1	4.76	0.53	4.78	91.23	3.73	27.71	0.49
EVV	0.05	4.64	0.53	4.68	83.70	5.53	2.96	0.42
VVV	0.1	4.88	0.31	5.02	8.72	5.19	12.31	0.34
VVV	0.05	4.61	0.30	4.75	19.09	4.34	0.24	0.30
Unbalanced clusters								
EII	0.1	0.04	0.04	0.04	26.36	0.04	0.55	0.08
EII	0.05	0.04	0.04	0.04	12.32	0.04	0.05	0.06
VII	0.1	0.06	0.06	0.06	0.06	0.06	0.08	0.09
VII	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.07
EEI	0.1	0.16	0.16	0.16	0.51	0.16	0.21	0.23
EEI	0.05	0.15	0.15	0.15	0.15	0.15	0.16	0.17
VEI	0.1	0.17	0.17	0.18	10.07	0.18	0.93	0.23
VEI	0.05	0.16	0.16	0.16	0.16	0.16	0.17	0.18
EVI	0.1	0.10	0.09	0.09	3.13	0.09	0.11	0.13
EVI	0.05	0.09	0.09	0.09	2.25	0.09	0.09	0.10
VVI	0.1	0.08	0.08	0.08	0.08	0.09	0.11	0.12
VVI	0.05	0.08	0.08	0.08	0.08	0.08	0.09	0.09
EEE	0.1	0.33	0.33	0.35	137.73	0.35	72.95	4.43
EEE	0.05	0.31	0.31	0.32	71.45	0.37	40.72	0.33
VEE	0.1	0.68	0.68	0.71	207.56	3.53	103.15	10.74
VEE	0.05	0.63	0.63	0.66	134.39	1.30	67.62	0.67
EVE	0.1	0.51	0.47	0.51	83.35	0.87	50.42	4.94
EVE	0.05	0.46	0.44	0.46	61.52	0.53	7.77	0.46
VVE	0.1	0.61	0.54	61.52	19.37	0.67	106.00	1.31
VVE	0.05	0.55	0.51	78.11	2.08	0.61	89.67	0.52
EEV	0.1	0.67	0.67	3.71	65.45	3.07	1.34	0.49
EEV	0.05	0.65	0.65	3.57	58.87	2.92	0.36	0.44
VEV	0.1	0.38	0.38	3.45	0.29	3.29	0.39	0.42
VEV	0.05	0.36	0.36	3.23	0.28	1.15	0.30	0.35
EVV	0.1	5.81	0.92	5.95	50.20	5.37	12.12	0.68
EVV	0.05	5.57	0.91	5.70	35.31	6.25	0.49	0.65
VVV	0.1	5.75	0.52	6.07	0.39	5.79	0.53	0.57
VVV	0.05	5.30	0.50	5.55	0.37	3.04	0.40	0.49

Notes: $K = 2, d = 6$. Contamination type 1.

Table 5. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.01	0.01	0.01	0.01	0.01	839.05	669.11
EII	0.05	0.01	0.01	0.01	0.01	0.01	864.65	671.40
VII	0.1	0.02	0.02	0.02	0.02	0.03	808.46	664.74
VII	0.05	0.02	0.02	0.02	0.02	0.02	164.99	645.35
EEI	0.1	0.02	0.02	0.02	2.75	0.02	889.26	701.08
EEI	0.05	0.01	0.01	0.01	0.01	0.02	928.18	706.41
VEI	0.1	0.02	0.02	0.02	0.02	0.02	541.82	723.66
VEI	0.05	0.02	0.02	0.02	0.02	0.02	150.89	727.85
EVI	0.1	0.01	0.01	0.01	0.90	0.01	798.19	686.69
EVI	0.05	0.01	0.01	0.01	0.01	0.01	838.24	688.15
VVI	0.1	0.02	0.02	0.02	0.02	0.02	630.16	676.25
VVI	0.05	0.02	0.02	0.02	0.02	0.02	125.73	669.55
EEE	0.1	0.23	0.23	0.25	385.25	0.24	1635.63	897.75
EEE	0.05	0.21	0.20	0.22	70.98	0.25	1558.83	942.06
VEE	0.1	0.47	0.47	0.52	824.75	0.49	712.36	911.89
VEE	0.05	0.42	0.42	0.44	130.59	0.50	748.76	850.51
EVE	0.1	0.38	0.35	0.38	236.13	0.36	1857.72	919.32
EVE	0.05	0.33	0.31	0.33	2.51	0.38	1792.18	907.39
VVE	0.1	0.62	0.51	0.62	28.71	150.84	724.31	846.05
VVE	0.05	0.51	0.45	0.51	0.44	2.72	720.50	706.28
EEV	0.1	0.40	0.40	3.60	12.85	447.32	1467.74	433.46
EEV	0.05	0.37	0.37	3.24	0.21	58.26	698.39	182.97
VEV	0.1	0.28	0.28	3.64	14.42	159.91	672.81	597.05
VEV	0.05	0.26	0.26	3.28	3.71	31.86	547.53	232.84
EVV	0.1	5.51	0.53	5.58	3.48	339.81	1774.74	214.30
EVV	0.05	5.02	0.48	5.08	0.32	33.22	862.09	53.51
VVV	0.1	6.01	0.40	6.13	2.01	403.81	433.84	570.34
VVV	0.05	5.19	0.38	5.26	0.22	26.82	202.28	161.83
Unbalanced clusters								
EII	0.1	0.01	0.01	0.01	0.14	0.01	947.52	714.16
EII	0.05	0.01	0.01	0.01	0.01	0.01	982.82	725.73
VII	0.1	0.02	0.02	0.02	435.43	0.02	874.76	830.66
VII	0.05	0.02	0.02	0.02	110.58	0.02	288.02	847.42
EEI	0.1	0.02	0.02	0.02	54.03	0.02	959.98	749.46
EEI	0.05	0.02	0.02	0.02	0.02	0.02	984.36	766.47
VEI	0.1	0.02	0.02	0.02	68.82	0.02	590.68	802.34
VEI	0.05	0.02	0.02	0.02	0.02	0.02	262.13	821.67
EVI	0.1	0.02	0.02	0.02	1.75	0.02	906.67	739.25
EVI	0.05	0.02	0.02	0.02	0.02	0.02	931.31	752.51
VVI	0.1	0.02	0.02	0.02	47.62	0.02	751.17	801.45
VVI	0.05	0.01	0.01	0.01	2.10	0.01	236.30	816.37
EEE	0.1	0.29	0.29	0.32	90.98	912.97	1700.36	964.09
EEE	0.05	0.27	0.27	0.29	32.79	46.16	1723.84	842.95
VEE	0.1	0.61	0.60	0.67	186.34	941.88	734.50	949.42
VEE	0.05	0.56	0.56	0.60	0.55	570.22	767.15	256.57
EVE	0.1	0.44	0.42	0.45	58.19	990.61	1812.69	997.20
EVE	0.05	0.40	0.38	0.40	0.38	296.41	975.91	525.99
VVE	0.1	0.55	0.47	0.55	727.37	1104.39	890.17	1027.83
VVE	0.05	0.47	0.45	0.47	25.43	1133.41	816.98	542.28
EEV	0.1	0.69	0.70	4.91	2.21	2.92	1715.97	0.63
EEV	0.05	0.65	0.66	4.53	6.36	2.44	990.45	5.18
VEV	0.1	0.48	0.48	3.87	0.34	8.48	15.94	0.66
VEV	0.05	0.45	0.45	3.94	0.28	1.33	4.29	3.00
EVV	0.1	8.94	0.99	9.76	0.49	3.99	1944.25	0.79
EVV	0.05	7.92	0.93	8.53	0.45	4.59	530.91	0.56
VVV	0.1	7.61	0.68	7.72	0.40	6.34	41.10	0.79
VVV	0.05	7.10	0.63	7.24	0.37	3.69	3.08	6.61

Notes: $K = 2$, $d = 2$. Contamination type 2.

Table 6. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.02	0.02	0.02	1558.43	0.02	1511.12	1454.98
EII	0.05	0.02	0.02	0.02	1609.64	0.02	1529.42	1472.77
VII	0.1	0.05	0.05	0.05	0.05	0.05	1539.11	1572.37
VII	0.05	0.04	0.04	0.04	0.04	0.04	767.09	1618.88
EEI	0.1	0.05	0.05	0.05	1579.24	0.05	1545.52	1464.40
EEI	0.05	0.05	0.05	0.05	1645.55	0.05	1600.13	1484.09
VEI	0.1	0.07	0.07	0.07	1688.10	0.07	1496.82	1492.72
VEI	0.05	0.07	0.07	0.07	1766.79	0.07	820.44	1515.85
EVI	0.1	0.04	0.04	0.04	1558.11	0.04	1518.46	1469.44
EVI	0.05	0.03	0.03	0.03	1623.16	0.04	1553.41	1488.31
VVI	0.1	0.05	0.05	0.05	4.13	0.05	1514.95	1540.97
VVI	0.05	0.05	0.05	0.05	0.05	0.05	773.08	1574.29
EEE	0.1	0.23	0.23	0.23	1753.20	0.79	1624.85	1462.44
EEE	0.05	0.22	0.22	0.22	1769.13	0.46	1678.05	1486.38
VEE	0.1	0.46	0.46	0.46	1773.08	8.21	1483.97	1478.76
VEE	0.05	0.44	0.44	0.44	1839.46	5.92	1515.54	1510.78
EVE	0.1	0.32	0.33	0.32	1611.13	1.73	1544.88	1471.99
EVE	0.05	0.34	0.34	0.34	1717.33	0.41	1601.60	1512.75
VVE	0.1	0.51	0.50	0.51	1797.77	3.84	1528.34	1523.72
VVE	0.05	0.49	0.49	0.49	1864.86	101.42	1563.77	1540.39
EEV	0.1	0.40	0.40	2.80	1600.76	2.02	1583.52	1345.21
EEV	0.05	0.39	0.39	2.79	1622.96	2.71	1379.02	1371.30
VEV	0.1	0.23	0.23	2.68	1434.99	2.43	1345.66	1342.18
VEV	0.05	0.22	0.22	2.67	1483.03	1.94	1354.76	1364.95
EVV	0.1	4.35	0.49	4.34	1471.06	2.84	1450.94	1345.67
EVV	0.05	4.40	0.47	4.43	1165.73	5.07	1319.85	1347.25
VVV	0.1	4.11	0.32	4.15	1450.48	3.61	1342.73	1347.16
VVV	0.05	4.08	0.31	4.12	1509.38	4.34	1366.26	1373.31
Unbalanced clusters								
EII	0.1	0.03	0.03	0.03	0.03	0.03	1574.87	1492.45
EII	0.05	0.03	0.03	0.03	0.03	0.03	1603.70	1509.72
VII	0.1	0.04	0.04	0.04	1682.64	0.04	1585.45	1576.97
VII	0.05	0.03	0.03	0.03	1801.00	0.04	1017.56	1598.04
EEI	0.1	0.07	0.07	0.07	972.78	0.07	1581.02	1503.05
EEI	0.05	0.07	0.07	0.07	820.08	0.07	1634.69	1522.05
VEI	0.1	0.08	0.08	0.08	1733.11	0.08	1541.66	1540.66
VEI	0.05	0.08	0.08	0.08	1831.45	0.08	1103.90	1561.17
EVI	0.1	0.05	0.05	0.05	700.79	0.05	1560.13	1508.72
EVI	0.05	0.04	0.04	0.04	826.36	0.04	1594.69	1527.10
VVI	0.1	0.04	0.04	0.04	1662.60	0.04	1558.22	1562.45
VVI	0.05	0.04	0.04	0.04	1752.31	0.04	982.09	1582.77
EEE	0.1	0.31	0.31	0.31	1754.36	0.31	1664.06	1504.79
EEE	0.05	0.30	0.30	0.30	1778.50	0.35	1711.39	1528.60
VEE	0.1	0.63	0.63	0.63	1788.31	0.91	1528.55	1526.23
VEE	0.05	0.59	0.59	0.60	1889.10	0.70	1560.33	1548.03
EVE	0.1	0.43	0.43	0.44	1694.65	1.06	1593.79	1516.54
EVE	0.05	0.43	0.42	0.43	1669.10	0.49	1626.02	1543.36
VVE	0.1	0.51	0.51	4.54	1876.21	0.54	1573.93	1574.65
VVE	0.05	0.48	0.48	4.59	2000.51	620.59	1602.80	1494.60
EEV	0.1	0.68	0.69	3.53	776.51	2.77	1660.56	1395.78
EEV	0.05	0.64	0.65	3.47	606.02	2.82	1709.01	1422.11
VEV	0.1	0.42	0.42	2.94	156.91	2.86	1358.31	1412.21
VEV	0.05	0.38	0.38	2.86	34.23	1.22	1307.83	1441.31
EVV	0.1	6.11	0.96	6.19	314.16	4.54	1451.13	1396.90
EVV	0.05	5.99	0.92	6.08	86.13	8.78	1490.27	1420.13
VVV	0.1	4.72	0.58	4.78	144.52	4.55	1329.55	1417.92
VVV	0.05	4.63	0.53	4.69	21.69	3.32	1301.10	1449.83

Notes: $K = 2, d = 4$. Contamination type 2.

Table 7. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
EII	0.1	0.03	0.03	0.03	2441.73	0.03	2275.30	2265.89
EII	0.05	0.03	0.03	0.03	2493.36	0.03	2288.07	2266.20
VII	0.1	0.07	0.07	0.07	9.81	0.07	2342.82	2347.32
VII	0.05	0.07	0.07	0.07	0.07	0.07	2232.99	2363.98
EEI	0.1	0.11	0.11	0.11	2589.44	0.11	2300.82	2273.89
EEI	0.05	0.11	0.11	0.11	2594.34	0.11	2329.68	2276.57
VEI	0.1	0.16	0.16	0.16	2684.02	0.16	2304.90	2303.94
VEI	0.05	0.15	0.15	0.15	2703.01	0.15	2166.37	2310.93
EVI	0.1	0.07	0.07	0.07	2511.02	0.07	2290.41	2281.37
EVI	0.05	0.07	0.07	0.07	2533.76	0.07	2303.93	2281.57
VVI	0.1	0.09	0.09	0.09	2552.20	0.10	2321.09	2335.48
VVI	0.05	0.09	0.09	0.09	120.23	0.09	2216.99	2349.81
EEE	0.1	0.24	0.24	0.24	2706.57	0.50	2305.50	2274.98
EEE	0.05	0.23	0.23	0.23	2685.52	0.40	2299.58	2272.52
VEE	0.1	0.48	0.48	0.48	2753.87	16.37	2283.45	2283.54
VEE	0.05	0.46	0.46	0.46	2784.20	8.99	2288.87	2288.87
VVE	0.1	0.53	0.53	0.53	2616.54	0.54	2335.79	2333.80
VVE	0.05	0.50	0.50	0.50	2630.19	61.10	2341.26	2332.43
EEV	0.1	0.42	0.42	2.81	2479.10	6.90	2202.77	2153.05
EEV	0.05	0.41	0.41	2.82	2583.74	2.85	1865.94	2154.94
VEV	0.1	0.22	0.22	2.64	2284.41	2.61	2147.67	2149.37
VEV	0.05	0.22	0.22	2.66	2299.76	1.95	2142.61	2151.41
EVV	0.1	4.93	0.78	4.93	851.93	3.58	2243.06	2245.13
VVV	0.1	4.04	0.31	4.11	2280.91	19.39	2151.98	2153.31
VVV	0.05	4.06	0.30	4.12	2254.45	4.35	2154.79	2156.32

Notes: $K = 2$, $= 6$. Unbalanced clusters. Contamination type 2.

lead `tclust` and `otrimle` to a slightly worse performance as compared to the other settings. On the other hand, no such problem exists with `mtclust` and in particular with the model automatically chosen by selecting the parametrization having the best performance in terms of BIC.

5. Real data analysis

We apply the proposed method to the `thyroid` dataset available in the `mclust` package [15].

The dataset is made of $n = 215$ patients and $p = 6$ variables. As an exploratory tool we report, in Figure 6, the pairs plot of data. Each point is coloured according to the value of the first categorical variable indicating whether the subject had euthyroidism (safe, coloured with green), hypothyroidism (colored in red) and hyperthyroidism (colored with blue).

The last two conditions are clinically relevant. The remaining variables are quantitative and record T3-resin uptake test (in percentage), T4 levels, T3 levels, TSH and the maximal absolute difference of TSH after injection of 200 micro grams of thyrotropin-releasing hormone. The aim of clustering, in this case, is to differentiate subjects with respect to their level of activity of the thyroid.

In order to proceed we need to select a trimming level. There are several ways to select this parameter now in the literature [3,29,30,35]. One approach is that of using the `ctlcurves` method proposed in García-Escudero et al. [30]: in Figure 7 the objective function at convergence of the model VVV is plot for several choices of K and α . This is closely connected to the idea of monitoring [36,37].

Table 8. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.05	0.05	0.05	42.59	3.16	10.61	0.06
EII	0.05	0.04	0.04	0.04	21.82	0.60	12.62	0.05
VII	0.1	1.89	1.79	2.67	0.80	1.85	20.16	61.77
VII	0.05	1.88	1.74	2.37	0.79	1.36	4.39	17.87
EEI	0.1	0.07	0.07	0.07	97.03	3.52	37.00	0.10
EEI	0.05	0.06	0.06	0.06	77.19	0.06	48.36	0.07
VEI	0.1	0.18	0.18	0.19	30.16	2.28	23.19	0.28
VEI	0.05	0.16	0.16	0.16	35.25	0.83	21.13	0.19
EVI	0.1	0.09	0.09	0.09	72.16	10.30	13.41	0.12
EVI	0.05	0.07	0.07	0.07	105.47	0.07	9.22	0.09
VVI	0.1	0.49	0.40	0.49	59.76	0.48	33.38	8.36
VVI	0.05	0.49	0.38	0.48	39.74	0.38	20.71	0.62
EEE	0.1	0.97	0.97	1.93	111.40	39.14	253.07	46.71
EEE	0.05	0.85	0.85	2.07	108.96	21.92	78.23	45.77
VEE	0.1	342.34	271.87	458.94	126.26	424.67	1402.84	1184.60
VEE	0.05	201.85	138.14	561.43	125.36	184.21	747.61	474.62
EVE	0.1	2.43	1.56	22.73	167.39	1.78	206.42	82.27
EVE	0.05	1.88	1.36	16.27	144.92	1.63	56.98	70.30
VVE	0.1	65.73	9.01	137.16	154.40	57.48	1007.70	870.95
VVE	0.05	42.50	5.81	130.84	144.71	22.60	368.74	181.27
EEV	0.1	1.25	1.25	5.07	90.46	3.22	161.79	150.06
EEV	0.05	1.13	1.13	5.07	86.19	3.02	70.91	130.06
VEV	0.1	1.09	1.12	4.43	121.42	3.25	94.88	119.48
VEV	0.05	1.00	1.03	4.51	113.43	2.86	39.28	73.44
EVV	0.1	9.95	1.94	12.48	76.15	5.38	309.44	223.38
EVV	0.05	9.64	1.73	11.24	81.28	5.13	102.64	238.89
VVV	0.1	8.91	1.73	9.87	83.81	5.62	155.64	196.76
VVV	0.05	8.82	1.55	9.06	78.40	5.16	110.52	200.55
Unbalanced clusters								
EII	0.1	2.26	1.08	0.06	94.17	115.05	0.20	0.08
EII	0.05	0.05	0.05	0.05	139.57	0.05	51.83	0.06
VII	0.1	2.76	2.68	3.17	35.54	1.76	36.53	114.19
VII	0.05	2.58	2.51	2.89	71.89	1.27	21.78	56.97
EEI	0.1	0.08	0.08	0.08	151.46	0.09	5.19	0.12
EEI	0.05	0.07	0.07	0.07	162.50	0.07	118.70	0.09
VEI	0.1	0.35	0.35	0.25	74.45	58.44	5.29	0.72
VEI	0.05	0.21	0.21	0.29	98.90	0.22	39.90	0.28
EVI	0.1	0.12	0.11	0.11	75.43	65.15	0.67	0.21
EVI	0.05	0.10	0.09	0.09	100.44	0.10	2.94	0.12
VVI	0.1	0.66	0.62	0.68	107.62	0.56	77.94	68.79
VVI	0.05	0.68	0.57	0.66	124.47	0.47	48.59	16.89
EEE	0.1	1.11	1.11	2.13	203.92	171.18	212.65	115.32
EEE	0.05	1.00	1.00	2.41	198.79	143.81	205.16	98.71
VEE	0.1	426.18	314.32	490.17	301.81	512.96	1555.14	1544.25
VEE	0.05	345.12	235.31	533.74	293.77	275.58	1120.71	740.99
EVE	0.1	2.50	1.83	26.15	155.26	2.10	349.63	266.31
EVE	0.05	1.88	1.60	16.00	141.75	1.91	183.36	151.11
VVE	0.1	104.95	63.29	179.27	202.46	94.24	1383.66	1190.08
VVE	0.05	62.41	46.47	184.63	185.35	45.42	673.17	366.87
EEV	0.1	1.49	1.49	5.57	83.44	3.74	287.07	175.12
EEV	0.05	1.36	1.36	5.37	95.34	3.33	223.38	196.45
VEV	0.1	1.31	1.33	4.72	129.47	3.65	179.55	165.07
VEV	0.05	1.18	1.21	4.79	123.05	3.06	161.69	202.83
EVV	0.1	11.08	2.38	14.53	61.63	6.62	412.01	271.98
EVV	0.05	10.65	2.14	12.56	64.17	6.08	271.00	278.92
VVV	0.1	9.76	2.09	10.84	111.04	6.66	262.74	268.34
VVV	0.05	9.64	1.88	10.95	100.16	5.91	256.24	257.94

Notes: $K = 4, d = 2$. Contamination type 1.

Table 9. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
EII	0.1	0.09	0.09	0.09	37.90	0.09	31.85	0.11
EII	0.05	0.08	0.08	0.08	42.99	16.59	23.66	0.09
VII	0.1	2.06	2.00	2.97	2.92	2.29	44.37	45.14
VII	0.05	1.94	1.87	2.83	1.57	1.93	7.03	7.43
EEI	0.1	0.23	0.23	0.23	113.49	18.04	78.56	0.29
EEI	0.05	0.21	0.21	0.21	116.90	11.58	92.32	0.23
VEI	0.1	1.08	1.05	9.65	57.79	0.90	31.28	0.74
VEI	0.05	1.76	1.62	7.32	84.03	4.54	43.24	0.57
EVI	0.1	0.21	0.20	0.20	111.52	41.33	33.00	0.27
EVI	0.05	0.19	0.19	0.19	128.45	28.46	28.98	0.21
VVI	0.1	0.91	0.75	1.26	20.08	1.10	60.99	4.84
VVI	0.05	0.83	0.68	0.91	9.48	0.70	37.69	1.07
EEE	0.1	0.98	0.98	4.75	209.85	101.12	153.39	105.69
EEE	0.05	0.91	0.91	3.87	220.77	89.57	105.85	40.50
VEE	0.1	272.66	241.91	447.25	222.69	162.36	379.88	693.04
VEE	0.05	204.45	167.98	446.84	224.35	115.57	358.91	308.05
EVE	0.1	46.21	3.14	85.55	210.57	41.73	304.59	147.69
EVE	0.05	19.71	1.42	89.88	230.96	15.58	220.70	65.65
VVE	0.1	261.44	10.60	264.09	166.59	256.51	913.27	798.67
VVE	0.05	246.25	7.99	270.53	164.22	223.91	473.02	139.23
EEV	0.1	1.27	1.27	5.23	128.55	3.93	253.37	101.76
EEV	0.05	1.18	1.18	4.77	136.91	3.52	95.07	63.42
VEV	0.1	1.11	1.14	4.74	149.12	3.88	76.31	80.59
VEV	0.05	1.02	1.04	4.46	165.79	3.17	57.78	31.02
EVV	0.1	18.00	3.28	38.72	135.11	15.63	502.62	147.02
EVV	0.05	10.38	1.73	39.69	139.12	8.02	333.63	119.43
VVV	0.1	9.22	1.70	14.21	205.47	7.69	128.96	134.80
VVV	0.05	8.45	1.59	12.34	219.08	6.25	102.41	98.91

Notes: $K = 4$, $d = 4$. Unbalanced clusters. Contamination type 1.

From Figure 7 we first of all note that there are three groups, as the lines for $K = 2$ and $K = 3$ are well separated while the lines for $K = 3$ and $K = 4$ are not. Additionally, for $\alpha > 10\%$ the lines for $K = 2$ and $K = 3$ get closer, indicating that with those trimming level the smallest group is being emptied (and it is empty for $\alpha > 12.5\%$, approximately, as the two lines are not separated anymore). To formally validate our choices we report in Table 14 the BIC (or minimal BIC_c over $c \leq 50$, for models with an ER constraint), when $\alpha = 0$. The formal information criterion approach confirms that there are $K = 3$ clusters and leads us to select model VEV. Since there is no formal procedure yet for selecting α , we tentatively fix it heuristically based on the `ctlcurves` as $\alpha = 5\%$, and then provide some evidence by proceeding along the lines of Hardin and Rocke [38].

Using a 5% trimming level we end up by selecting the EVI model with $K = 3$ clusters (Table 15), whose performance are reported in Table 16.

Being aware of the fact that these are somehow subjective considerations we formally provide some evidence by proceeding along the lines of Hardin and Rocke [38]: we compute the MCD estimator [39] in each cluster. We then compute the Mahalanobis distance of each observation from the MCD parameters and discard the ones exceeding the opportune quantile (i.e. .99) of its target distribution (the χ^2 distribution with d d.f.). By doing so we end up estimating a percentage of outlying observations equal to 7% (a value very close to the one chosen by looking at the `ctlcurves`).

We then compute BIC (and $\inf_{c \leq 50} BIC_c$ for models under an ER constraint) as in Table 14 and end up selecting once again the EVI model, with $K = 3$ clusters and $\alpha = 7\%$.

Table 10. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.13	0.13	0.13	31.56	17.83	39.23	0.17
EII	0.05	0.12	0.12	0.12	30.51	18.93	31.04	0.14
VII	0.1	2.13	2.12	3.15	7.21	2.63	68.81	49.05
VII	0.05	1.91	1.90	2.91	9.33	2.30	18.04	5.66
EEI	0.1	0.46	0.46	0.68	129.70	21.00	108.68	0.57
EEI	0.05	0.43	0.43	0.47	149.09	19.16	103.66	0.47
VEI	0.1	13.75	11.04	33.13	73.21	8.58	56.64	8.29
VEI	0.05	18.13	18.09	22.43	88.23	7.00	56.91	1.18
EVI	0.1	0.42	0.36	1.57	131.08	34.48	58.23	0.47
EVI	0.05	0.34	0.34	0.42	138.67	18.75	45.58	0.38
VVI	0.1	1.48	1.15	1.61	30.44	1.42	77.76	5.82
VVI	0.05	1.26	1.07	1.45	25.31	1.97	68.88	1.86
EEE	0.1	0.97	0.97	9.79	247.72	100.19	158.16	98.93
EEE	0.05	0.93	0.93	6.34	259.22	88.32	99.95	32.40
VEE	0.1	233.67	147.84	453.15	425.89	175.48	356.52	587.83
VEE	0.05	141.04	90.31	451.22	484.76	118.54	291.49	279.80
EVE	0.1	117.40	2.66	131.02	222.05	107.65	512.26	140.98
EVE	0.05	86.40	1.40	127.80	226.19	71.53	333.61	56.12
VVE	0.1	278.82	11.05	277.87	282.66	280.01	837.49	686.64
VVE	0.05	273.82	4.63	277.89	313.16	279.77	437.55	105.81
EEV	0.1	1.21	1.21	5.13	191.65	3.91	252.66	104.52
EEV	0.05	1.17	1.17	5.00	201.25	3.53	106.22	70.07
VEV	0.1	1.05	1.06	4.38	212.78	3.89	92.15	77.32
VEV	0.05	1.01	1.02	4.32	218.84	3.15	75.79	34.91
EVV	0.1	60.73	8.92	99.67	173.72	56.00	644.04	150.37
EVV	0.05	28.94	2.61	94.44	166.70	25.56	373.27	125.29
VVV	0.1	19.21	3.31	39.27	222.03	17.39	165.38	139.49
VVV	0.05	9.86	1.55	30.97	227.44	7.45	128.14	107.86
Unbalanced clusters								
EII	0.1	0.15	0.15	0.25	7.11	107.34	8.10	0.21
EII	0.05	0.14	0.14	0.14	17.56	0.14	148.01	0.16
VII	0.1	4.66	4.80	4.57	8.89	14.88	61.52	90.54
VII	0.05	3.20	3.19	4.49	12.48	3.21	33.17	28.48
EEI	0.1	0.78	0.64	1.23	135.19	110.54	75.28	0.68
EEI	0.05	0.49	0.49	0.86	152.21	108.67	258.15	0.56
VEI	0.1	7.33	7.74	26.58	31.22	61.52	52.12	24.42
VEI	0.05	4.54	4.47	19.99	50.04	47.90	58.76	2.01
EVI	0.1	0.63	0.43	0.53	80.57	55.42	12.04	0.68
EVI	0.05	0.42	0.42	0.47	80.82	55.30	136.21	0.50
VVI	0.1	3.45	2.32	2.21	40.09	40.82	90.07	61.95
VVI	0.05	2.04	1.59	2.04	30.98	1.97	86.33	9.24
EEE	0.1	5.22	4.12	19.10	244.28	155.60	151.19	97.41
EEE	0.05	1.05	1.05	16.93	256.81	1.25	201.52	71.85
VEE	0.1	481.76	351.70	531.14	454.48	340.74	313.60	591.58
VEE	0.05	510.54	365.94	527.39	511.20	354.33	275.13	348.43
EVE	0.1	207.69	87.88	222.94	226.74	188.67	711.91	122.84
EVE	0.05	158.91	112.53	216.25	206.58	144.65	510.52	93.02
VVE	0.1	329.49	57.86	325.41	249.61	326.72	686.30	491.40
VVE	0.05	331.08	106.34	327.11	257.13	319.06	341.74	96.34
EEV	0.1	1.42	1.42	8.01	226.78	4.25	339.33	107.27
EEV	0.05	1.35	1.35	6.34	241.52	3.79	279.03	104.95
VEV	0.1	1.22	1.22	4.60	232.91	4.12	144.36	104.84
VEV	0.05	1.16	1.16	4.62	250.85	3.30	147.37	127.98
EVV	0.1	123.48	47.05	157.61	182.37	116.12	631.29	152.68
EVV	0.05	97.15	75.34	157.03	166.92	90.02	429.18	149.14
VVV	0.1	39.77	16.75	78.09	221.53	36.79	203.00	152.32
VVV	0.05	30.64	19.78	69.83	223.86	26.76	173.46	145.22

Notes: $K = 4, d = 6$. Contamination type 1.

Table 11. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.04	0.04	0.34	205.87	0.04	1418.94	1458.22
EII	0.05	0.04	0.04	0.04	288.76	0.04	1135.24	618.33
VII	0.1	2.08	1.97	3.28	444.10	2.26	1074.29	1896.14
VII	0.05	1.88	1.81	2.84	539.86	1.49	582.04	697.68
EEI	0.1	0.07	0.07	0.14	252.80	0.07	1724.02	1555.56
EEI	0.05	0.06	0.06	0.06	589.83	0.06	1438.26	654.99
VEI	0.1	0.17	0.17	0.96	713.05	0.17	908.24	1866.91
VEI	0.05	0.15	0.15	0.48	759.98	1.08	534.68	697.80
EVI	0.1	0.08	0.08	0.26	176.11	0.08	2870.56	1634.34
EVI	0.05	0.07	0.07	0.07	545.47	10.85	2234.56	632.30
VVI	0.1	0.48	0.40	0.80	586.26	0.52	1082.41	1858.75
VVI	0.05	0.50	0.38	0.51	672.15	0.40	627.95	646.75
EEE	0.1	0.92	0.92	3.91	618.12	104.71	3267.37	1833.28
EEE	0.05	0.82	0.82	2.37	471.92	36.07	2934.80	791.37
VEE	0.1	460.48	386.14	569.85	298.94	946.87	974.12	789.60
VEE	0.05	253.44	179.09	632.53	123.35	418.43	884.27	763.06
EVE	0.1	4.23	1.48	34.50	626.90	17.23	3799.01	1096.09
EVE	0.05	1.45	1.32	20.28	330.15	108.27	3121.60	881.91
VVE	0.1	92.92	26.44	150.90	398.27	697.50	1082.07	945.05
VVE	0.05	53.49	3.03	146.08	171.64	104.43	779.63	877.96
EEV	0.1	1.20	1.20	4.93	317.17	714.88	3096.75	1581.78
EEV	0.05	1.11	1.11	4.83	211.84	6.12	2352.77	537.23
VEV	0.1	1.05	1.08	4.56	182.53	685.92	1146.96	1633.38
VEV	0.05	0.96	1.00	4.52	166.85	2.89	842.23	535.60
EVV	0.1	9.81	1.86	11.05	124.93	951.35	3458.45	1000.72
EVV	0.05	9.37	1.71	9.70	84.31	66.03	2397.56	640.31
VVV	0.1	8.99	1.67	9.89	589.58	984.71	972.29	1127.09
VVV	0.05	8.76	1.52	8.83	292.13	21.51	604.33	540.71
Unbalanced clusters								
EII	0.1	0.05	0.05	0.05	131.09	0.05	1079.48	1310.63
EII	0.05	0.05	0.05	0.05	331.85	0.05	885.97	473.17
VII	0.1	2.91	2.83	3.18	350.08	2.10	694.86	1781.68
VII	0.05	2.64	2.55	3.23	379.48	1.30	520.86	683.24
EEI	0.1	0.08	0.08	0.52	378.41	0.08	1832.56	1447.28
EEI	0.05	0.07	0.07	0.07	525.41	0.07	1480.74	510.21
VEI	0.1	0.23	0.23	0.23	675.62	83.74	526.80	1788.86
VEI	0.05	0.20	0.20	0.20	731.80	0.20	526.17	597.23
EVI	0.1	0.10	0.10	0.54	376.47	0.10	2572.43	1486.95
EVI	0.05	0.09	0.09	0.09	542.73	0.09	2336.38	509.79
VVI	0.1	0.68	0.61	0.81	585.91	0.62	619.67	1838.73
VVI	0.05	0.70	0.55	0.67	640.56	0.48	577.52	626.03
EEE	0.1	1.07	1.07	2.73	519.01	546.47	3152.01	1268.59
EEE	0.05	0.97	0.97	2.10	353.28	1.13	2564.15	713.31
VEE	0.1	520.14	375.58	523.81	445.16	1435.84	1044.34	1127.18
VEE	0.05	351.79	245.00	558.48	281.28	699.64	916.41	737.47
EVE	0.1	6.98	3.21	20.45	583.75	19.11	4062.71	748.46
EVE	0.05	2.68	1.95	20.24	260.25	402.57	3345.16	752.70
VVE	0.1	118.34	58.90	194.61	378.40	775.77	1438.11	1227.98
VVE	0.05	66.47	50.84	188.31	204.37	163.34	887.09	812.90
EEV	0.1	1.42	1.42	5.40	210.97	192.76	3028.44	1311.17
EEV	0.05	1.31	1.32	5.29	95.89	3.32	2039.35	490.32
VEV	0.1	1.24	1.27	4.90	130.22	203.26	971.88	1327.57
VEV	0.05	1.15	1.18	4.86	132.38	5.34	524.53	491.10
EVV	0.1	10.80	2.26	13.25	95.96	107.75	3102.89	746.07
EVV	0.05	10.38	2.09	13.46	62.74	7.31	2117.40	513.25
VVV	0.1	9.80	2.00	10.22	579.68	147.94	576.45	906.03
VVV	0.05	9.57	1.81	9.85	188.94	7.29	474.37	472.13

Notes: $K = 4$, $d = 2$. Balanced clusters. Contamination type 2.

Table 12. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.05	0.08	0.08	1.79	1520.52	0.08	1570.68	2134.49
VII	0.1	2.15	2.15	6.89	1590.34	10.34	2304.46	3491.03
VII	0.05	1.76	1.72	3.02	1750.74	1.98	1524.76	2025.45
EEI	0.1	0.32	0.21	8.68	1537.38	0.21	2253.13	3211.33
EEI	0.05	0.20	0.20	0.31	1579.15	0.20	1706.61	2081.45
VEI	0.1	5.09	4.12	37.08	1567.22	1.72	2187.44	3345.04
VEI	0.05	1.76	1.54	12.57	1731.09	0.71	1449.31	2165.61
EVI	0.1	0.36	0.19	20.23	1472.22	0.19	5029.66	3260.73
EVI	0.05	0.18	0.18	0.50	1542.12	0.18	4684.12	2175.96
VVI	0.1	3.05	2.50	4.48	1535.49	9.95	2320.16	3382.50
VVI	0.05	0.94	0.77	1.34	1695.07	0.70	1457.24	1997.05
EEE	0.1	1.31	0.95	15.08	1747.53	0.95	2059.37	3497.33
EEE	0.05	0.90	0.90	6.59	1776.04	1.06	1869.72	1723.43
VEE	0.1	318.83	304.96	513.33	2085.40	202.08	1704.26	2159.45
VEE	0.05	239.21	185.97	468.55	2275.64	170.34	1540.22	1482.72
EVE	0.1	85.20	12.01	110.35	1686.11	74.37	5255.11	3410.14
EVE	0.05	24.41	1.34	92.86	1731.01	16.44	4544.88	1476.56
VVE	0.1	282.50	15.01	289.88	1839.52	264.94	1653.01	1698.43
VVE	0.05	265.95	8.55	282.17	1903.53	245.10	1411.65	1418.97
EEV	0.1	1.24	1.24	6.72	1646.63	3.95	5209.79	3188.69
EEV	0.05	1.15	1.15	5.30	1634.60	3.51	4929.54	1568.68
VEV	0.1	1.08	1.11	4.90	1534.83	4.09	1838.22	3145.55
VEV	0.05	1.01	1.03	4.50	1617.54	3.18	1519.42	1641.35
EVV	0.1	27.08	3.44	57.29	1521.70	50.92	5237.72	3068.23
EVV	0.05	10.30	1.86	54.18	1423.82	6.83	4847.56	1321.49
VVV	0.1	10.11	1.66	23.95	1514.67	8.51	1632.58	3231.34
VVV	0.05	8.64	1.56	17.23	1563.56	6.29	1461.47	1389.60
Unbalanced clusters								
EII	0.1	0.10	0.10	13.57	1356.77	0.10	1508.01	3183.22
EII	0.05	0.09	0.09	0.61	1418.44	0.09	1364.09	1528.32
VII	0.1	5.10	5.89	8.06	1542.92	7.18	1934.30	3279.74
VII	0.05	2.91	2.89	4.18	1654.64	2.32	1358.08	1875.35
EEI	0.1	0.25	0.25	3.16	1456.10	0.25	2033.43	3215.63
EEI	0.05	0.23	0.23	0.24	1481.17	224.56	2187.62	1540.25
VEI	0.1	0.80	0.84	13.22	1571.14	0.75	1874.93	3318.36
VEI	0.05	0.64	0.64	3.55	1699.74	0.69	1339.29	1779.42
EVI	0.1	0.24	0.24	7.19	1420.02	0.24	5031.90	3261.97
EVI	0.05	0.22	0.22	0.31	1487.24	0.22	4185.36	1720.92
VVI	0.1	3.01	2.34	1.53	1509.09	28.01	1937.40	3292.23
VVI	0.05	1.25	0.93	1.16	1635.84	1.09	1346.11	1830.46
EEE	0.1	1.37	1.07	18.72	1662.67	1.07	2405.58	3481.64
EEE	0.05	1.80	1.38	8.87	1646.22	1.55	2510.33	1329.91
VEE	0.1	634.43	484.98	580.31	2001.28	411.15	1965.02	3657.29
VEE	0.05	624.40	436.71	552.80	2185.42	748.22	1522.27	1499.41
EVE	0.1	94.42	54.21	167.54	1574.13	79.45	5273.21	2917.90
EVE	0.05	56.61	48.65	144.89	1668.16	51.26	3535.41	1308.71
VVE	0.1	343.09	27.33	336.90	1712.20	322.59	2230.23	3093.50
VVE	0.05	306.75	128.00	341.01	1808.11	291.76	1368.68	1359.45
EEV	0.1	1.44	1.44	6.82	1601.90	4.30	5043.30	3161.95
EEV	0.05	1.33	1.33	5.06	1630.58	3.68	5067.01	1469.66
VEV	0.1	1.23	1.26	5.27	1466.38	4.34	1597.94	3119.61
VEV	0.05	1.15	1.17	4.91	1542.26	3.30	1403.95	1480.74
EVV	0.1	31.26	6.47	66.72	1425.68	28.26	5307.92	3124.21
EVV	0.05	17.16	10.44	62.89	1417.37	14.87	4758.91	1284.15
VVV	0.1	10.93	2.35	27.77	1439.96	9.37	1511.17	3180.05
VVV	0.05	9.91	2.65	16.76	1494.42	7.48	1430.32	1339.83

Notes: $K = 4, d = 4$. Contamination type 2.

Table 13. Mean squared error.

Parameters	Alpha	MSE mtclust	MSE mtclust BIC	MSE tclust	MSE Mclust	MSE otrimle	MSE cont.mixt	MSE teigen
Balanced clusters								
EII	0.1	0.13	0.13	23.40	2366.55	0.13	4180.67	4765.52
EII	0.05	0.12	0.12	3.07	2427.59	0.12	2307.17	3787.11
VII	0.1	1.98	1.97	7.62	2645.35	2.53	3785.66	5149.45
VII	0.05	1.83	1.82	3.37	2747.60	2.31	2527.45	4018.65
EEI	0.1	2.84	0.84	10.05	2518.52	0.69	3702.45	4813.40
EEI	0.05	0.42	0.42	2.48	2566.34	0.42	2446.34	3910.91
VEI	0.05	20.38	15.69	33.88	2687.02	10.08	2400.03	4312.52
EVI	0.1	2.98	0.35	18.51	2404.95	0.54	6622.51	4849.42
EVI	0.05	0.35	0.33	2.30	2465.66	0.33	5531.72	3948.69
VVI	0.1	7.57	6.83	10.93	2546.02	1.80	3744.67	5018.10
VVI	0.05	1.93	1.22	2.66	2584.11	1.28	2489.74	4050.78
EEE	0.1	0.98	0.98	24.29	2747.32	0.98	2609.44	5123.25
EEE	0.05	0.92	0.92	9.97	2746.73	1.08	2338.08	3182.95
VEE	0.1	281.59	226.64	580.83	3103.85	200.48	2765.18	6131.60
VEE	0.05	170.90	98.10	482.67	3138.65	129.34	2318.71	2306.30
VVE	0.1	289.98	11.61	300.65	2559.33	275.63	2698.93	5291.45
VVE	0.05	280.48	7.78	286.45	2593.64	261.10	2251.85	2214.63
EEV	0.1	1.20	1.20	11.23	2441.47	3.92	7108.35	4797.67
EEV	0.05	1.16	1.16	6.24	2599.36	3.54	6150.97	3094.00
VEV	0.1	1.03	1.04	5.15	2333.24	4.02	2740.82	4760.49
VEV	0.05	1.01	1.01	4.36	2366.26	3.15	2258.74	3317.92
VVV	0.1	31.06	4.27	52.48	2305.27	27.98	2503.46	4906.46
VVV	0.05	11.37	1.55	41.93	2289.31	7.97	2219.04	2701.35
Unbalanced clusters								
EII	0.1	0.15	0.15	9.28	2256.63	0.15	3493.68	4777.27
EII	0.05	0.14	0.14	0.89	2309.75	0.14	2090.65	2806.97
VII	0.1	5.24	5.47	4.83	2601.54	12.91	3571.62	5052.22
VII	0.05	3.01	3.01	4.67	2680.59	3.21	2213.76	3760.46
EEI	0.1	1.41	0.87	19.22	2426.16	1.06	3802.32	4817.90
EEI	0.05	0.49	0.49	4.88	2453.66	0.49	2724.92	2878.22
VEI	0.1	5.58	5.16	42.83	2557.98	4.42	3609.20	4894.60
VEI	0.05	8.58	8.84	21.35	2672.51	45.82	2163.43	3676.45
EVI	0.1	0.80	0.42	16.30	2362.78	0.42	6715.21	4856.52
EVI	0.05	0.41	0.41	2.57	2380.92	0.41	5213.34	3074.83
VVI	0.1	6.16	3.93	2.84	2496.12	39.85	3588.46	4933.41
VVI	0.05	10.45	8.00	1.98	2521.77	41.45	2210.43	3678.57
EEE	0.1	10.10	5.31	44.05	2642.97	4.92	2659.34	5132.35
EEE	0.05	1.05	1.05	23.59	2662.96	1.26	2397.41	2317.46
VEE	0.1	533.49	370.51	656.94	3052.92	371.80	3270.68	6339.31
VEE	0.05	450.86	322.53	591.71	3132.13	384.01	2247.20	2628.74
VVE	0.1	336.05	32.18	337.91	2491.84	322.69	3337.85	5809.76
VVE	0.05	332.71	97.60	331.25	2496.62	306.47	2140.81	2252.39
EEV	0.1	1.74	1.75	10.05	2410.07	4.61	7145.73	4806.39
EEV	0.05	1.35	1.35	5.93	2568.11	3.81	6436.74	2889.18
VEV	0.1	1.20	1.21	6.48	2269.71	4.26	2698.88	4787.08
VEV	0.05	1.16	1.16	4.65	2299.83	3.31	2148.29	2972.38
VVV	0.1	35.45	13.28	74.32	2227.03	33.00	2426.98	4913.99
VVV	0.05	29.66	19.58	58.27	2208.40	26.27	2146.12	2669.26

Notes: $K = 4$, $d = 6$. Contamination type 2.

This is particularly interesting as (i) `tclust` corresponds to `VVV`, which might be overparameterized as suggested by BIC and (ii) a similar reasoning applies to the optimal model with `mclust`, which is `VVI`, slightly more complex than `EVI`. By trimming as few as $n\alpha = 11$ observations we were able to work with a more parsimonious model than `mclust`. A referee kindly pointed out that by trimming some information might be lost. This is indeed true but the trimmed observations could then be assigned to the

closest robustly estimated centroid, and bias reduction almost always compensates variance increase due to trimming. See below for more on this point. (Table 17)

In Table 18 we report the `mtclust` EVI estimated centroids, together with the measurements associated to the observations flagged as outlying. It is straightforward to see that the clusters are well separated and, at the same time, trimmed observations are far from clusters' centroids: the Mahalanobis distance of each outlying observation has a p -value smaller than the Bonferroni adjusted 0.05/11 level [40].

Finally, we compare the grouping results with the actual classification based on clinical definitions of hypothyroidism, normal activity and hyperthyroidism. In Table 16 we report the misclassification rate and the modified Rand index for each method.

Table 16 shows good performance in terms of classification for all procedures. It can be seen that `mtclust` can achieve the same performance of `mclust`, but with a more parsimonious model. Surprisingly, `tclust` achieves a slightly worse performance in terms of agreement between clinical and unsupervised classification. We believe this is due to the fact that it forcibly uses a VVV model, and the geometric constraints of

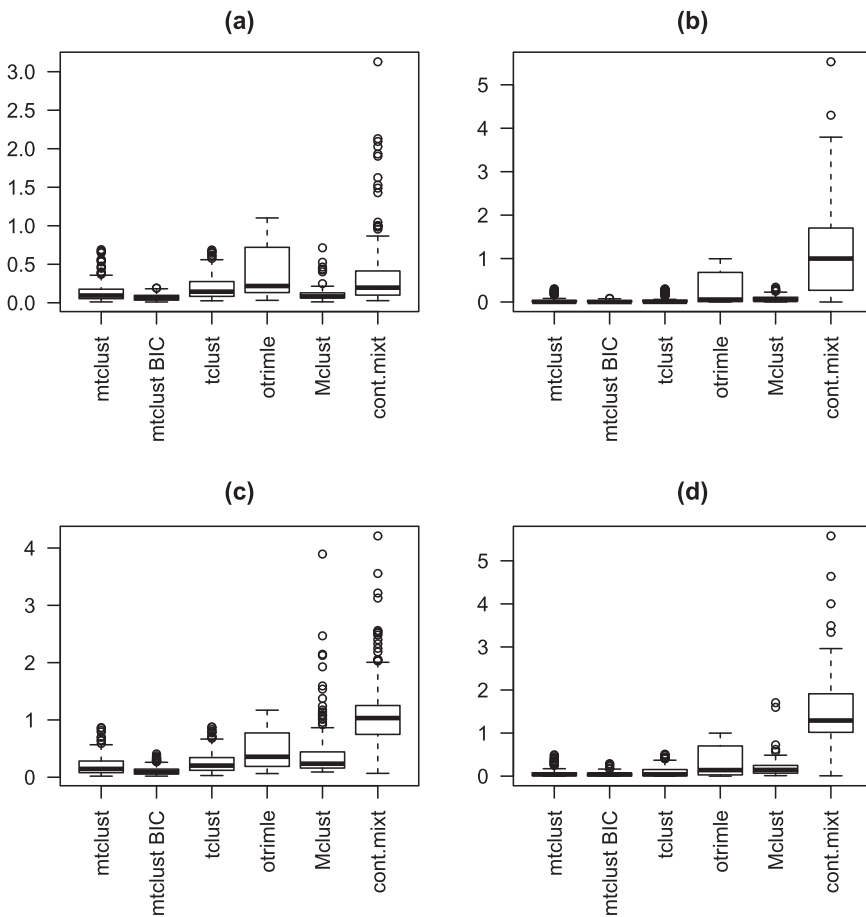


Figure 5. Boxplot of the misclassification rate in each macro scenario. $K = 2$ and ‘close’ outliers in Panel (a). $K = 2$ and ‘far’ outliers in Panel (b). $K = 4$ and ‘close’ outliers in Panel (c). $K = 4$ and ‘far’ outliers in Panel (d).

Table 14. Thyroid data: values of BIC different parametrizations, different number of clusters, and $\alpha = 0$.

Model	Number of clusters				
	1	2	3	4	5
EII	7477.37	7283.95	6977.55	6763.95	6636.21
VII	7477.37	7275.31	6560.44	6213.14	6180.81
EVI	6699.73	6404.44	6096.77	6072.92	5872.48
VEI	6699.73	5476.55	5305.03	5358.89	5309.72
EVI	6699.73	6091.96	5118.44	5155.05	5027.95
VVI	6715.19	6301.50	5307.53	5269.70	5280.75
EEE	6388.42	6241.26	6009.34	5925.01	5852.41
VEE	6388.42	5432.11	5288.98	5273.24	5300.15
EVE	6506.92	6266.89	5341.10	5346.93	5287.89
EEV	6388.42	5985.98	5192.35	5256.83	5082.96
VVE	6506.92	6248.03	5286.56	5350.62	5276.29
VEV	6388.42	5977.12	4933.49	4992.29	4980.72
EVV	6506.92	6325.90	5389.00	5477.26	5499.43
VVV	6506.92	5551.87	5328.39	5427.81	5432.75

Table 15. Thyroid data: values of BIC for different parametrizations, different number of clusters, and $\alpha = 5\%$.

Model	Number of clusters				
	1	2	3	4	5
EII	6654.84	6373.62	6066.99	5871.06	5808.32
VII	6654.84	6302.30	5903.04	5601.51	5514.12
EVI	5646.19	5262.72	4940.56	4844.41	4786.34
VEI	5646.19	4762.53	4634.55	4641.85	4614.18
EVI	5646.19	4851.07	4325.41	4542.28	4538.60
VVI	5660.10	4941.31	4666.64	4678.73	4681.10
EEE	5345.77	5079.39	4923.19	4769.03	4808.11
VEE	5345.77	4703.23	4630.81	4633.13	4685.93
EVE	5446.91	4952.77	4752.38	4714.38	4687.91
EEV	5345.77	4685.60	4475.64	4498.26	4534.86
VVE	5446.91	4891.15	4697.06	4701.60	4721.40
VEV	5345.77	4627.00	4426.13	4490.35	4510.14
EVV	5446.91	4943.30	4768.71	4803.97	4837.64
VVV	5446.91	4911.62	4748.29	4781.89	4838.69

Table 16. Thyroid data.

	Misclassification rate	Rand index
Mclust (VVI)	0.03	0.86
mtclust (EVI)	0.03	0.87
tclust (VVV)	0.08	0.68

Notes: Misclassification Rate and modified Rand index as calculated on observations flagged as clean by all procedures. In parentheses, for reference, the covariance matrix parameterization. tclust and mtclust are based on $\alpha = 5\%$ trimming. Mclust miscalsification rate is computed on the observations flagged as clean by tclust and Mclust.

parsimonious models may be useful for guiding towards the correct underlying clusters' shapes.

As a referee kindly noted, trimming based methodologies may lead to loss of efficiency in terms of parameter estimation. This is indeed true but in the presence of contamination bias reduction can make robust methods more advantageous. In order to support

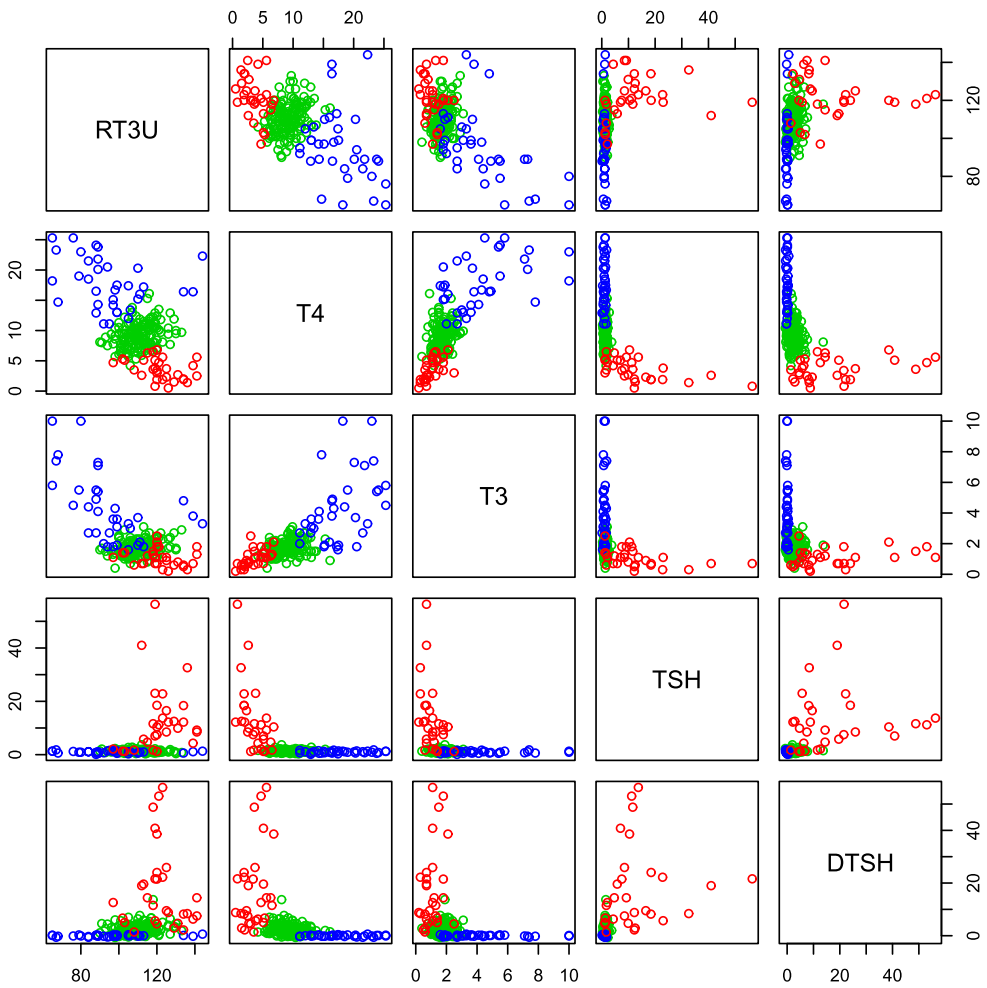


Figure 6. Thyroid data: Pairs plots based on the true classification labels.

this statement, we report that even after assignning each observation initially flagged to its closest cluster, misclassification rates favoured our proposed approach, with `mtclust` having an overall misclassification rate of 3%, `mclust` 3%, and `tclust` 6%. In conclusion, by trimming few observations we obtain the same classification results of the `mclust` methodology and, at the same time, a more parsimonious model is obtained (which might, as intuitive, have a better performance on new data).

We conclude this section by repeating the analysis after artificially adding further contamination to the thyroid data. We do so by adding $n_0 = 10$ (mild) outliers. We generate them from a rectangular distribution similarly to the simulation study. For `tclust` and `mtclust` we set the trimming level at 9.3%, so that 21 observations are trimmed (as 10 new observations were added and there were already 11 anomalous ones in the data set). In Table 19 we report misclassification rate and modified Rand index of the three procedures with the same parameters chosen in the previous section. It can be seen that a larger fraction of outliers leads to a slight worsening of the performance of

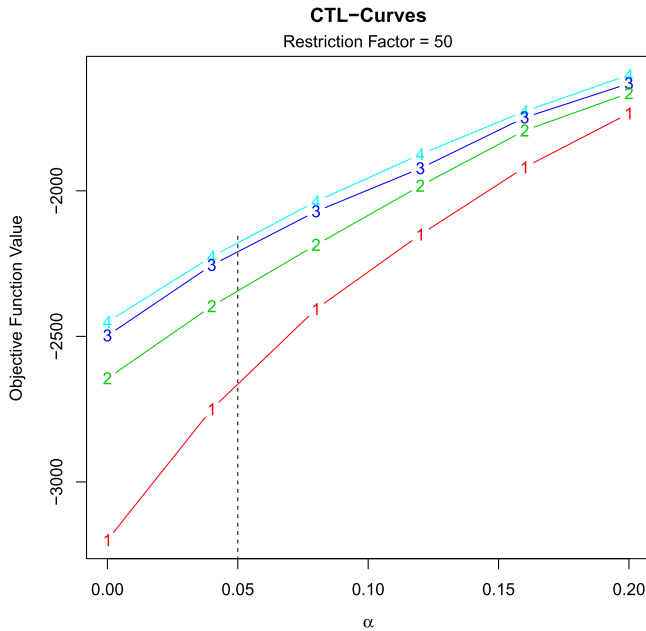


Figure 7. Thyroid data: objective function at convergence of t_{clust} for several values of K and α .

Table 17. Thyroid data: values of BIC for different parametrizations, different number of clusters, and $\alpha = 7\%$.

Model	Number of clusters				
	1	2	3	4	5
EII	6338.46	6047.01	5766.04	5544.30	5518.27
VII	6338.46	6000.79	5634.92	5357.33	5278.68
EEI	5237.00	4866.10	4547.79	4487.11	4443.35
VEI	5237.00	4501.46	4402.64	4413.66	4453.77
EVI	5237.00	4585.34	4224.64	4282.33	4235.90
VVI	5250.58	4673.90	4457.33	4451.40	4452.35
EEE	4936.29	4691.67	4510.04	4500.23	4468.52
VEE	4936.29	4439.11	4403.21	4404.27	4472.11
EVE	5057.18	4711.00	4518.99	4480.51	4466.90
EEV	4936.29	4474.69	4257.11	4305.43	4336.54
VVE	5057.18	4659.91	4493.08	4493.22	4490.93
VEV	4936.29	4407.50	4243.40	4318.43	4387.65
EVV	5057.18	4703.16	4549.75	4579.83	4631.83
VVV	5057.18	4684.95	4552.73	4596.61	4606.78

R-mclust. On the other hand, formally robust procedures are not affected much by the higher contamination level and the modified Rand index of t_{clust} is even slightly increased.

6. Conclusions and further direction of research

We proposed a new clustering methodology based on trimming and different geometric constraints. There are, in our opinion, many advantages in using our proposed methodology. First of all, the procedure is robust thanks to trimming. The extensive

Table 18. Thyroid data.

	RT3U	T4	T3	TSH	DTSH	Distance
C1	127.50	3.50	1.00	9.10	13.70	
C2	95.40	18.10	4.40	1.00	0.00	
C3	110.20	9.00	1.70	1.30	2.60	
O1	120.00	6.80	2.10	10.40	38.60	49.42
O2	119.00	3.80	1.10	23.00	5.70	55.64
O3	118.00	3.60	1.50	11.60	48.80	42.08
O4	120.00	1.90	0.70	18.50	24.00	31.84
O5	119.00	0.80	0.70	56.40	21.60	604.58
O6	123.00	5.60	1.10	13.70	56.30	60.32
O7	121.00	4.70	1.80	11.20	53.00	55.73
O8	136.00	1.40	0.30	32.60	8.40	160.31
O9	123.00	1.90	0.30	22.80	22.20	61.32
O10	112.00	2.60	0.70	41.00	19.00	280.33
O11	134.00	1.90	0.60	18.40	8.20	29.71

Notes: Cluster profiles and measurements for the outlying observations after clustering with `mtclust` based on model EVI, $K = 3$ and $\alpha = 5\%$. C1, C2, and C3 represent the value of the vector mean in each cluster. The remaining rows represent the values of the 11 outlying observation in each variable. Last column report the value of the Mahalanobis distance of each outlying observation from its closest cluster.

RT3U: T3-resin uptake test (percentage). *T4*: Total Serum thyroxin as measured by the isotopic displacement method. *T3*: Total serum triiodothyronine as measured by radioimmuno assay. *TSH*: Basal thyroid-stimulating hormone (TSH) as measured by radioimmuno assay. *DTSH*: Maximal absolute difference of TSH value after injection of 200 micro grams of thyrotropinreleasing hormone as compared to the basal value

Table 19. Contaminated thyroid data.

	Misclassification rate	Rand index
R-mclust (VVI)	0.09	0.80
mtclust (EVI)	0.04	0.88
tclust (VVV)	0.11	0.73

Notes: Misclassification Rate and modified Rand index as calculated on observations flagged as clean by all procedures. In parentheses, for reference, the covariance matrix parameterization as selected by BIC. All procedures are based on $\alpha = 9.3\%$ trimming.

simulation study and the application to the real data clearly showed that there are benefits in terms of both robustness and efficiency when compared with the other existing methodologies. A detailed discussion regarding the formal robustness properties of trimming based clustering methodologies is provided in García-Escudero et al. [41]. Secondly, the usage of geometric constraints has several advantages from both a theoretical and applied point of view. The researcher is simultaneously able to use interpretable constraints involved and, at the same, for some parametrization the estimators are equivariant.

As oftentimes happens in robust methodologies, tuning is needed. Improving both the heuristic tools and formal criteria for selecting the tuning parameters is an important route for further work.

Acknowledgments

The authors are grateful to one referee for kind suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- [1] Garcia-Escudero LA, Gordaliza A, Matrán C. Trimming tools in exploratory data analysis. *J Comput Graph Stat.* 2003;12(2):434–449.
- [2] Farcomeni A, Ventura L. An overview of robust methods in medical research. *Stat Methods Med Res.* 2012;21:111–133.
- [3] Farcomeni A, Greco L. Robust methods for data reduction. Boca Raton: Chapman & Hall / CRC Press; 2015.
- [4] Dotto F. Advances in robust clustering methods with applications [Ph.D. thesis]. Ph.D School of Statistiscal Sciences. “La Sapienza” University of Rome; 2017.
- [5] García-Escudero LA, Gordaliza A, Matrán C, et al. Avoiding spurious local maximizers in mixture modeling. *Stat Comput.* 2015;25:619–633.
- [6] Ingrassia S, Rocci R. Degeneracy of the EM algorithm for the MLE of multivariate Gaussian mixtures and dynamic constraints. *Comput Stat Data Anal.* 2011;55:1715–1725.
- [7] Coretto P, Hennig C. Robust improper maximum likelihood: tuning, computation, and a comparison with other methods for robust Gaussian clustering. *J Am Stat Assoc.* 2016;111:1648–1659.
- [8] Cerioli A, Farcomeni A, Riani M. Wild adaptive trimming for robust estimation and cluster analysis. *Scand J Statist.* 2018. DOI:10.1111/sjos.12349
- [9] McLachlan G, Peel D. Finite mixture models. New York: John Wiley & Sons; 2004.
- [10] Zhang J, Liang F. Robust clustering using exponential power mixtures. *Biometrics.* 2010;66:1078–1086.
- [11] Bagnato L, Punzo A, Zoia MG. The multivariate leptokürtic-normal distribution and its application in model-based clustering. *Can J Stat.* 2017;45:95–119.
- [12] Banfield JD, Raftery AE. Model-based Gaussian and non-Gaussian clustering. *Biometrics.* 1993;49:803–821.
- [13] Celeux G, Govaert G. Gaussian parsimonious clustering models. *Pattern Recognit.* 1995;28:781–793.
- [14] Fraley C, Raftery AE. Model based clustering, discriminant analysis and density estimation. *J Am Stat Assoc.* 2002;97:611–631.
- [15] Fraley C, Raftery AE. Model based methods of classification: using the `mclust` software in chemometrics. *J Stat Softw.* 2007;6:1–13.
- [16] Punzo A, McNicholas PD. Parsimonious mixtures of multivariate contaminated normal distributions. *Biometric J.* 2016;58:1506–1537.
- [17] Gallegos MT, Ritter G. Trimmed ML estimation of contaminated mixtures. *Sankhya.* 2009a;71:164–220.
- [18] Gallegos MT, Ritter G. Trimming algorithms for clustering contaminated grouped data and their robustness. *Adv Data Anal Classif.* 2009b;3:135–167.
- [19] Gallegos MT, Ritter G. Using combinatorial optimization in model-based trimmed clustering with cardinality constraints. *Comput Statit Data Anal.* 2010;54:637–654.
- [20] Hathaway RJ. A constrained EM algorithm for univariate normal mixtures. *J Stat Comput Simul.* 1986;23:211–230.
- [21] García-Escudero LA, Gordaliza A, Matrán C, et al. A general trimming approach to robust cluster analysis. *Ann Stat.* 2008;36:1324–1345.
- [22] Symons M. Clustering criteria and multivariate normal mixtures. *Biometrics.* 1981;37:35–43.
- [23] Bryant P. Large-sample results for optimization-based clustering methods. *J Classif.* 1991;8:31–44.
- [24] Dykstra RL. An algorithm for restricted least squares regression. *J Am Stat Assoc.* 1983;78:837–849.
- [25] Fritz H, García-Escudero LA, Mayo-Iscar A. `tclust`: an R package for a trimming approach to Cluster Analysis. *J Stat Softw.* 2012;47. Available from: <http://www.jstatsoft.org/v47/i12>
- [26] Ruwet C, Garcia-Escudero LA, Gordaliza A, et al. On the breakdown behavior of TCLUST clustering procedure. *TEST.* 2013;22:466–487.

- [27] Farcomeni A. Robust constrained clustering in presence of entry-wise outliers. *Technometrics*. 2014a;56:102–111.
- [28] Farcomeni A. Snipping for robust k-means clustering under component-wise contamination. *Stat Comput*. 2014b;24:909–917.
- [29] Ritter G. Robust cluster analysis and variable selection. New York: CRC Press; 2014.
- [30] García-Escudero LA, Gordaliza A, Matrán C, et al. Exploring the number of groups in robust model-based clustering. *Stat Comput*. 2011;21(4):585–599.
- [31] Cerioli A, García-Escudero LA, Mayo-Isacar A, et al. Finding the number of normal groups in model-based clustering via constrained likelihoods. *J Comput Graph Stat*. 2018a; available online.
- [32] Fraley C, Raftery AE, Murphy B, et al. `mclust` version 4 for r: normal mixture modeling for model-based clustering, classification, and density estimation. Washington, DC: Department of Statistics, University of Washington; 2012. (Tech. rep.).
- [33] Punzo A, Mazza A, McNicholas P. ContaminatedMmixt: an R package for fitting parsimonious mixtures of multivariate contaminated normal distributions. *J Stat Softw*. 2018;85:1–25.
- [34] Andrews JL, Wickins JR, Boers NM, et al. `teigen`: an R package for model-based clustering and classification via the multivariate T distribution. *J Stat Softw*. 2018; available online.
- [35] Dotto F, Farcomeni A, García-Escudero LA, et al. A reweighting approach to robust clustering. *Stat Comput*. 2018;28:477–493.
- [36] Cerioli A, Riani M, Atkinson AC, et al. The power of monitoring: how to make the most of a contaminated multivariate sample. *Stat Methods Appt*. 2018b; available online.
- [37] Farcomeni A, Dotto F. The power of (extended) monitoring in robust clustering. *Stat Methods Appt*. 2018; available online.
- [38] Hardin J, Rocke DM. Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. *Comput Stat Data Anal*. 2004;44:625–638.
- [39] Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. New York: John Wiley & Sons; 2005; Vol. 589.
- [40] Cerioli A, Farcomeni A. Error rates for multivariate outlier detection. *Comput Stat Data Anal*. 2011;55:544–553.
- [41] García-Escudero LA, Gordaliza A, Matrán C, et al. A review of robust clustering methods. *Adv Data Anal Classif*. 2010;4:89–109.