

# Robust proposals for model based clustering of multivariate data

**Agustín Mayo-Iscar**

*Dpto. Estadística I.O., Instituto de Matemáticas Universidad de Valladolid*

---

Impartial trimming and constrained approaches have been successfully applied to robustify maximum likelihood procedures when estimating clustering and mixture models during the last 20 years. As usual, trimming methods are useful to diminish the influence of anomalous observations that do not follow the model. However, and for robustness purposes, it is also needed to regularize the estimation due to the presence of singularities in the objective function. Our proposal is to also apply constraints that allow us to derive well-defined estimating procedures and to reduce the prevalence of spurious local likelihood maximizers. These robust procedures, known as TCLUST, initially were developed for normal multivariate distributed components. Now, more flexible procedures are available based on skewed distributions. The joint application of trimming and constraints also works for identifying regression models when data belong to a mixture of them. In this setting, the corresponding TCLUST estimators for the cluster-weighted model appear highly competitive. Parsimonious approaches are frequently needed for estimating clusters. Among them are Mixture Factor Analyzers and Celeux and Govaert's collection of models. In a similar fashion, TCLUST proposals are available for estimating clusters robustly. An important issue related with the application of TCLUST methodologies is their input parameters. Users have to provide at least the level of trimming, the strength of the constraints together with the number of clusters/components, which is the classical input parameter in clustering/mixture modelling. There are available exploratory tools and automatized procedures for assist to the users in choosing these parameters. TCLUST methodologies are available via the "tclust" package in CRAN and the "FSDA" toolbox in MATLAB.