

# Visualization of multivariate data with tail trees

Jussi Klemelä \*

Department of Statistics, Economics Faculty

University of Mannheim

L 7 3-5 Verfügungsgebäude

68131 Mannheim, Germany

Email: klemela@rumms.uni-mannheim.de

Fax +49 621 1811931

February 20, 2007

## Abstract

We introduce graphical tools to visualize the shape, the location and the orientation of a multivariate data set. We define a tree structure among the observations, called a *tail tree*. A tail tree is a tree whose root node corresponds to a center point of the data, and whose branches correspond to the tails of the data. We visualize a tail tree with a *tail tree plot*. Visualizing the tree structure among the observations makes it feasible to detect features from the data. A tail tree may also be used to define and enhance other visualizations. We define a *tail frequency plot* which visualizes the empirical probabilities of the disconnected tails of the point cloud. A tail tree induces a segmentation of the data which may be used to enhance a grand tour, graphical matrices, and parallel coordinate plots. We apply tail tree plots in exploratory data analysis of financial data.

**Key Words:** Anisotropic spread, Clustering, Dependence, Exploratory data analysis, High dimensional data

---

\*Writing of this article was financed by Deutsche Forschungsgemeinschaft under project MA1026/8-2.

# 1 Introduction

We want to visualize the shape, the location, and the orientation of a multivariate point cloud  $x^1, \dots, x^n \in \mathbf{R}^d$ . We interpret the point cloud as realizations of  $n$  random vectors having a common density function. We concentrate on the case where the point cloud does not have clusters (it is connected in the sense of a single linkage hierarchical clustering). We define a tree structure among the observations, which is called a tail tree. A tail tree is visualized with a tail tree plot and with a tail frequency plot, and a tail tree defines a segmentation of the data which may be used to enhance a grand tour, graphical matrices, and parallel coordinate plots.

The method is designed for visualizing dependency among several variables, whereas many of the previously used visualization tools are more directed to finding and visualizing clusters in the data. In some cases one can assume that the data does not contain clusters. This is typically the case for example for data consisting of financial returns. Then one is interested in the dependence structure and not interested in finding clusters. If one cannot exclude the possibility of clusters a priori, then one should explore the existence of clusters with other methods (projections, parallel coordinate plots, graphical matrices), and possibly divide data to clusters before applying tail trees.

A *tail tree* is a tree whose nodes are associated with the observations so that the root node corresponds to a central observation and the nodes in the upper levels of the tree correspond to the observations lying in the tails of the data. If the data is not spherically symmetric but has separated tails, then a tail tree has many branches, each branch corresponding to a separate tail of the data.

A *tail tree plot* consists of  $d$  scatter plots, where the  $x$ -coordinate of the  $i$ th scatter plot corresponds to the  $i$ th coordinate of the observations, and the  $y$ -coordinate of the scatter plots gives the Euclidean distance of the observations from the center point. We visualize in each scatter plot the tail tree by connecting the observations with straight lines, and use colors to identify the branches of the tail tree among the  $d$  scatter plots. A tail tree adds a structure to the data which makes it possible for humans to detect important features in the data. A basic idea behind a tail tree plot is that a tail tree may be considered as a spatial tree (1D curve in the  $d$ -dimensional Euclidean space which has branches) and 1D curves in the  $d$ -dimensional space may be visualized with  $d$  projections to the coordinate axes.

We concentrate in this article on the visualization of the data with tail tree plots, but a tail tree is potentially very useful in defining further visualizations and in enhancing other visualization tools.

We define a *tail frequency plot* which visualizes the heaviness of the tails separately in each tail of the point cloud. This visualization is useful when the underlying distribution has anisotropic tails.

A tail tree defines a segmentation of the data. The segmentation may be used to define a coloring of the observations. The coloring helps the orientation during a grand tour and makes the patterns in a parallel coordinate more visible. In addition, the segmentation of the data may be used to define a permutation of the data which reveals patterns when one uses a graphical matrix to visualize the data.

In Section 2 we define a tail tree. In Section 3 we define a tail tree plot. Section 4 contains examples of tail tree plots. In particular, in Section 4.3 we analyze a financial data set with tail tree plots. Section 5 introduces a tail frequency plot. Section 6 defines the segmentation of the data induced by a tail tree. Section 7 compares tail tree plots to shape trees, parallel level plots, parallel coordinate plots, graphical matrices, and to projections and slices. In addition, tail trees are applied to enhance graphical matrices and parallel coordinate plots. Section 8 contains a discussion. Appendix A defines copulas and marginal distributions.

Computations of the article have been made with an R-package "denpro", which may be downloaded from <http://denstruct.net>. We will use the convention that the individual frames of a figure are labeled with a, b, c, and d, starting from left to right.

## 2 Tail trees

A tail tree is defined with the help of a *tail clustering* of the data. We find tail clusters of the data when we remove central data points and then cluster the remaining observations. However, tails may appear at different distances and there is no distinguished rule for choosing the number of central points to be removed. Instead, we may consider a sequence of balls of increasing radius, and make a tree structure of clusters by clustering at each step the observations in the complement of a ball of the sequence.

Figure 1(a) shows a scatter plot of data of size 2000, generated from the distribution whose copula is the Clayton copula with parameter  $\theta = 4$ , and whose marginals have the Student distribution with degrees of freedom 4 (for definitions, see Appendix A). This data may be said to contain three tails. Figure 1(b) shows the remaining observations when we have removed the observations inside a ball of radius 2. At this stage there are 2 tail clusters. Figure 1(c) shows the three most distant tail clusters. Now we have removed the observations inside a ball of radius 6.5.

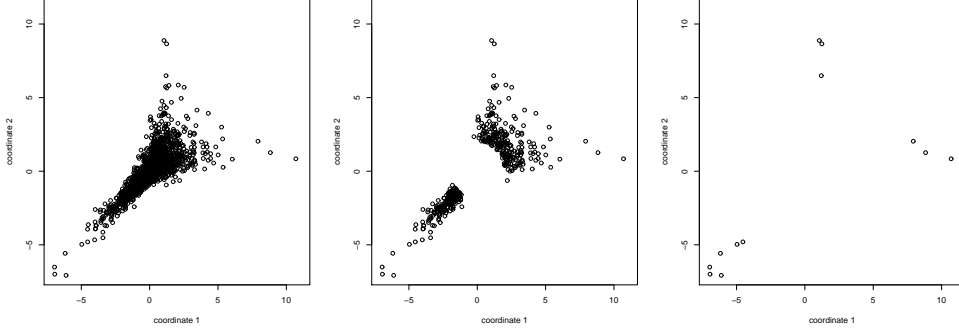


Figure 1: Frame a) shows a data set with three tail clusters, and the other frames show the data after removing central observations.

## 2.1 Connected sets and single linkage clustering

We define the concepts of  $\rho$ -separated sets and a  $\rho$ -connected set for sets of possibly finite cardinality.

**Definition 1** ( $\rho$ -separated sets and a  $\rho$ -connected set.)

1. Sets  $A, B \subset \mathbf{R}^d$  are separated for the resolution threshold  $\rho \geq 0$  ( $\rho$ -separated), if for each  $x \in A$  and  $y \in B$ ,  $\|x - y\| > 2\rho$ , where  $\|\cdot\|$  is the Euclidean norm.
2. Set  $A \subset \mathbf{R}^d$  is connected for the resolution threshold  $\rho \geq 0$  ( $\rho$ -connected), if for every nonempty  $B, C$  such that  $A = B \cup C$ ,  $B$  and  $C$  are not separated for the resolution threshold  $\rho$ .

In other words, a set is connected if it cannot be written as a union of two separated sets. When  $\rho > 0$ , then set  $A$  is connected for the resolution threshold  $\rho$  if for each  $x, y \in A$ , there is a path  $z_1, \dots, z_m \in A$  so that  $z_1 = x$ ,  $z_m = y$  and  $\|z_i - z_{i+1}\| \leq 2\rho$  for  $i = 1, \dots, m - 1$ . A set of finite cardinality is  $\rho$ -connected if the union of balls, centered at the elements, with radius  $\rho$ , is 0-connected. (That is why we use distance  $2\rho$  in the definition of separated sets.) For a set of cardinality  $n$  we have a finite sequence intervals  $[0, \rho_1), \dots, [\rho_{n-1}, \infty)$ , so that when  $\rho \in [\rho_{k-1}, \rho_k)$ ,  $k = 1, \dots, n$ ,  $\rho_0 = 0$ ,  $\rho_n = \infty$ , the set has  $n - k + 1$  connected components which are pairwise separated. We call values  $\rho_1, \dots, \rho_{n-1}$  the critical resolution thresholds.

The single linkage clustering partitions a finite set to separated sets in the sense of Definition 1. The single linkage clustering is an agglomerative hierarchical clustering algorithm, where the distance between clusters is defined

to be the distance between the closest pair of data points in the clusters. In fact, the hierarchical clustering tree (dendrogram) with the single linkage agglomeration gives the sequence  $\rho_1, \dots, \rho_{n-1}$  of the critical resolution thresholds. We build the hierarchical clustering tree by finding the 2 closest observations, and joining them together into a cluster. The distance between the 2 closest data points is equal to  $2\rho_1$ . At step  $k$ ,  $k = 1, \dots, n-1$ , we find the 2 closest clusters among the set of  $n-k+1$  previously found clusters, and join these clusters together. The distance between these 2 closest clusters is equal to  $2\rho_k$ . Finally at step  $n-1$  we merge together the remaining 2 clusters, and the distance between these clusters is equal to  $2\rho_{n-1}$ .

## 2.2 Definition

A tail tree is defined for sets of finite cardinality. We do not grow a sequence of balls with continuously increasing radius as in Figure 1: since the data has a finite cardinality we may remove observations from the data set one in time. We start at a center point of the data, and remove observations in the order determined by the closeness to the center point. At each step we cluster the remaining observations.

**Definition 2** (Tail tree.) *Assume that set  $A = \{x^1, \dots, x^n\} \subset \mathbf{R}^d$  is connected for resolution threshold  $\rho > 0$ . A tail tree of data set  $A$ , associated with a resolution threshold  $\tau \geq \rho$ , and with center point  $\mu \in \mathbf{R}^d$ , is a tree where each node is associated with a subset of  $A$  and with one observation.*

1. *The tree has a single root node, and the root node is associated with set  $A$  itself, and with the observation closest to the center point  $\mu$ .*
2. *Let node  $m$  be associated with set  $B \subset A$ .*
  - (a) *If  $B$  contains only one observation  $x \in A$ , then node  $m$  is a leaf node, associated with set  $B = \{x\}$  and observation  $x$ .*
  - (b) *Otherwise, let  $x = \operatorname{argmin}_{y \in B} \|y - \mu\|$  be the closest point in  $B$  to center point  $\mu$ . Write*

$$B \setminus \{x\} = C_1 \cup \dots \cup C_M \tag{1}$$

*where sets  $C_i$  are pairwise separated, and each is connected, with the resolution threshold  $\tau$ . Then node  $m$  has  $M$  children, which are associated with sets  $C_i$  and observations  $c_i = \operatorname{argmin}_{y \in C_i} \|y - \mu\|$ ,  $i = 1, \dots, M$ .*

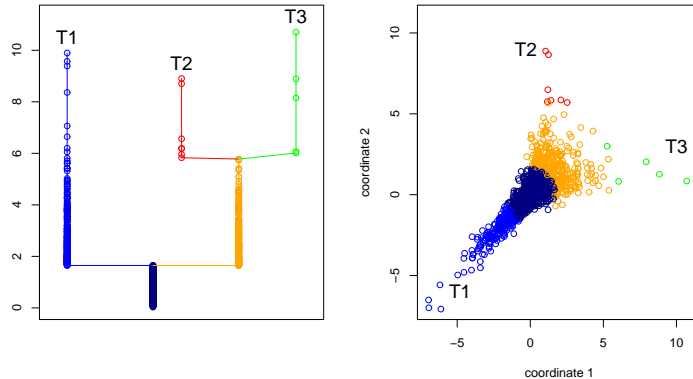


Figure 2: Frame a) shows a tail tree of the data in frame b). The coloring of the nodes helps to find the corresponding observations in the scatter plot.

We choose in this article the center point to be the mean:  $\mu = n^{-1} \sum_{i=1}^n x^i$ . An other possibility is to use some concept of a multivariate median, defined in terms of a data depth, see Liu, Parelius and Singh (1999).

Figure 2(a) shows a tail tree of the data in frame b), which is the same data as in Figure 1(a). The resolution threshold is  $\rho = 1.1$  and the center point is  $(0.026, 0.031)$ . The nodes in the tail tree are plotted at the height which is equal to the Euclidean distance of the corresponding observation from the center point.

Figure 3 gives a step by step illustration of Definition 2. We have drawn the data points as colored bullets. In the first window we have drawn a circle of radius  $\rho = 0.75$  around each observation, and one sees that the set of observations is connected for resolution threshold  $\rho$ . The center point is chosen to be the origin. In the second window we have removed the circles around the 2 closest observations to the center, and one sees that the set of the remaining observations consists of 3 separated components. The rightmost component consists of a single red observation and this observation corresponds to the leaf node T4. In the third window we have removed the circle around the other but the two furthest observations. These two furthest observations correspond to the leaf nodes T1 and T2. In the 4-th window we show the tail tree of the data set.

To calculate a tail tree we may apply the LeafsFirst algorithm with the bounding box enhancement. This algorithm was introduced in Klemelä (2006) for calculating shape trees, and we can modify the algorithm to calculate a tail tree. The algorithm starts by ordering the observations according

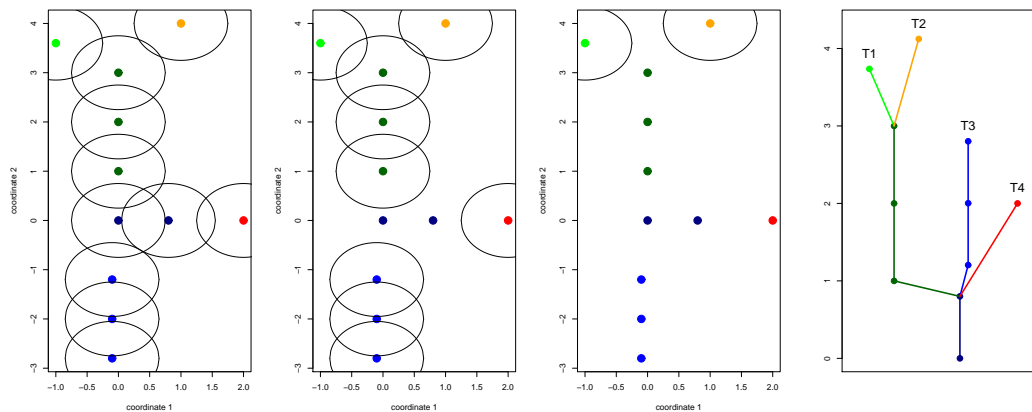


Figure 3: An illustration of Definition 2 of a tail tree. Frames a-c) show the data and frame d) shows a tail tree of the data set.

to the distance from the center point. Then one goes through the observations starting from the most distant observation, and at each step one checks whether the current observation touches some previously found connected components. If it touches several previously found connected components, then one joins together those components. One keeps record of the bounding boxes of the components. When an observation does not touch the bounding box of a component, then it does not touch any observations inside the bounding box, and only when the observation touches the bounding box, one has to go deeper in the nested sequence of bounding boxes to find out whether the observation touches some of the observations belonging to the component. The number of flops in the worst case is  $O(dn^2)$ , but with the bounding box enhancement one achieves considerable savings in typical cases. When we apply the LeafsFirst algorithm, we do not have to find beforehand a resolution threshold which would make the data set connected. When the data set is not connected the algorithm gives as an output a tree with several root nodes.

### 3 Tail tree plot

We have defined a tail tree as a tree whose nodes are associated with observations. We may consider a tail tree as a spatial tree (discrete curve, having branches, in the  $d$ -dimensional Euclidean space). Thus we may visualize this tree by plotting each coordinate separately.

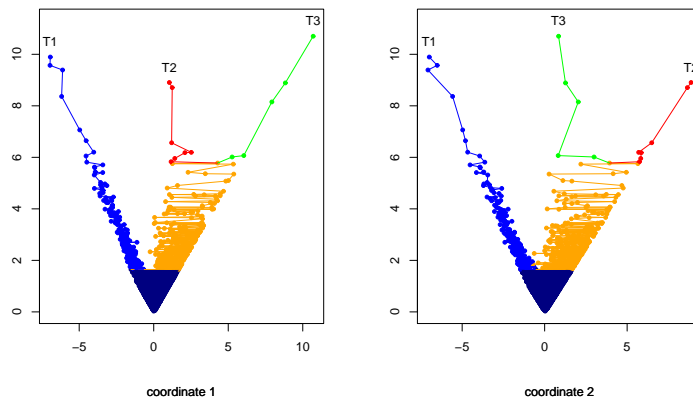


Figure 4: A tail tree plot of the data in Figure 1(a).

**Definition 3** (Tail tree plot.) A tail tree plot of data  $A = \{x^1, \dots, x^n\} \subset \mathbf{R}^d$  is a visualization of a tail tree of data, as defined in Definition 2. A tail tree plot consists of  $d$  windows. The nodes of the tree are drawn as bullets.

1. The horizontal position of a node in the  $i$ -th window,  $i = 1, \dots, d$ , is equal to the  $i$ -th coordinate of the observation associated with the node.
2. The vertical position of a node is equal to the distance of the observation associated with the node to the center point,
3. The parent-child relations are expressed by lines joining the parent with its children.

We add colors to the plot to identify the branches across different windows. We choose first distinct colors for the leaf nodes and then travel towards the root node, changing the color always when two branches are merging. We color also the lines joining a child and a parent. The color of a line will be the same as the color of the child node. As an additional help in identifying the nodes across the windows one may also label the leaf nodes with numbers or letters.

Figure 4 shows a tail tree plot of the data in Figure 2(b) and in Figure 1(a). The tail tree plot shows the three tails of the data and one sees that the south-west tail (blue branch) is much thinner than the north-east tail (orange branch). The red and green branch have different positions in the two windows and this visualizes the different locations of these tails. Figure 5 shows a tail tree plot of the data in Figure 3.



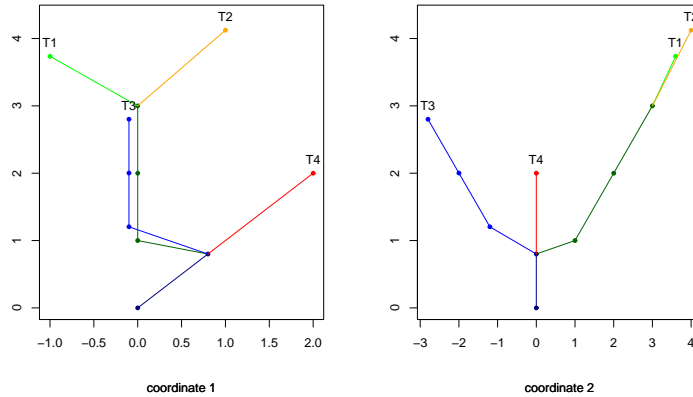


Figure 5: A tail tree plot of the data in Figure 3.

## 4 Examples

### 4.1 Ellipsoids

**Orientation.** A ball looks like a triangle in a tail tree plot. Ellipsoids have two tails and thus tail trees of ellipsoidal point clouds have typically two branches. Figures 6 - 8 show a series of rotated 2D ellipses. We have rotated a sample of size  $n = 2000$  which is generated from the 2D Gaussian distribution whose marginal standard deviations are 1 and 2, and the correlation is 0. The tail tree plot in Figure 6 shows that the observations in the blue branch have large values in the first coordinate and small values in the second coordinate, which means that this branch lies in the south-east direction. The tail tree plot in Figure 7 shows that the observations in the blue branch have small values in both coordinates which means that this branch lies in the south-west direction. In Figure 8 the first window of the tail tree plot shows that the two branches have the same location with respect to the 1st coordinate (we distinguish the branches only through the coloring of the nodes). The second window of the tail tree plot shows that the locations of the branches differ in the second coordinate, the other lying in the south and the other in the north.

**Four dimensional ellipsoids.** Figure 9 shows a tail tree plot of a 4D Gaussian point cloud of size  $n = 2000$ , when the marginal standard deviations are 2.5, 2, 1.5, and 1, and the correlations are 0.25. The first direction shows the 2 extensions of the ellipsoid, and in the other windows one sees how the triangular shape is slowly vanishing, and in the 4th window it has almost

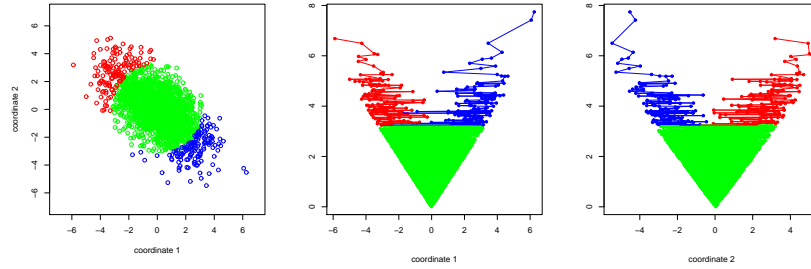


Figure 6: A tail tree plot of a 2D ellipsoidal point cloud; from the south-east to the north-west orientation.

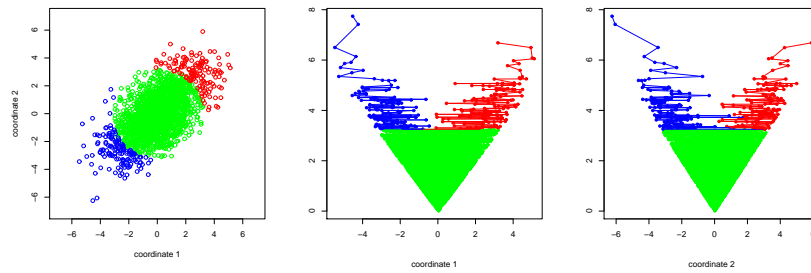


Figure 7: A tail tree plot of a 2D ellipsoidal point cloud; from the south-west to the north-east orientation.

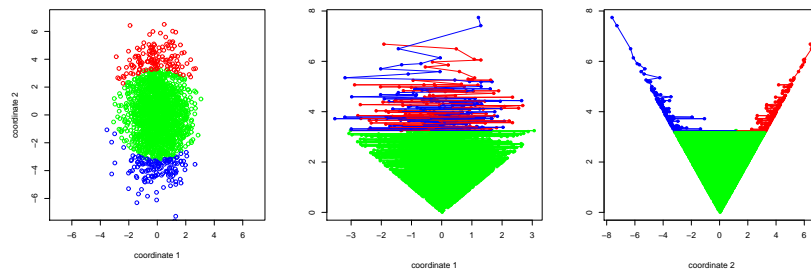


Figure 8: A tail tree plot of a 2D ellipsoidal point cloud; from the south to the north orientation.

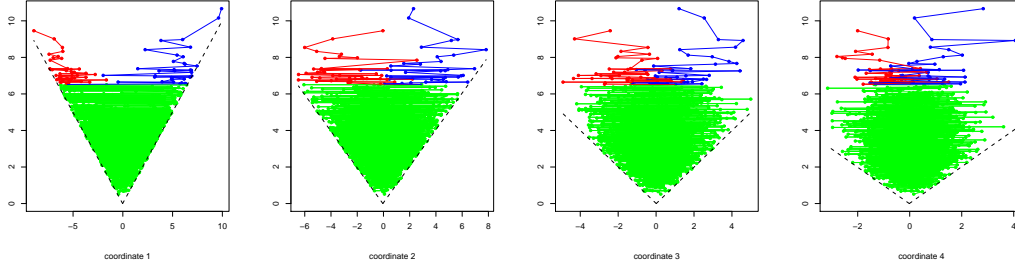


Figure 9: 4D Gaussian point cloud when the marginal standard deviations are 2.5, 2, 1.5, and 1, the correlations are 0.25, and the sample size is  $n = 2000$ .

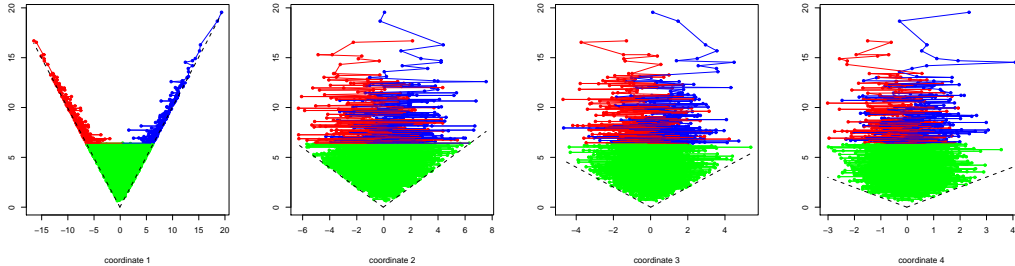


Figure 10: 4D Gaussian point cloud when the marginal standard deviations are 5, 2, 1.5, and 1, the correlations are 0.25, and the sample size is  $n = 2000$ .

completely vanished.

Figure 10 shows a tail tree plot of a 4D Gaussian point cloud of size  $n = 2000$ , when the marginal standard deviations are 5, 2, 1.5, and 1, and the correlations are 0.25. The standard deviation is much larger in the first coordinate, and the triangular shape is not clearly visible in the coordinates 2 – 4: the ellipsoid is so thin that it is cutted to two halves by a ball of small radius.

## 4.2 Copulas

Copulas and marginal distributions are defined in Appendix A.

**Elliptical copulas.** Figure 11 shows a sample of size  $n = 1000$  from the distribution whose marginals are two independent Student distributions with degrees of freedom 4, that is, the distribution has the Gaussian copula with shape parameter (correlation coefficient)  $r = 0$ , and Student marginals. We

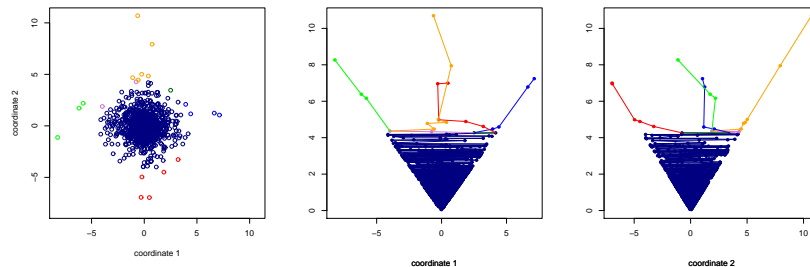


Figure 11: Gaussian copula with shape parameter  $r = 0$ , Student marginals with degrees of freedom 4,  $n = 1000$  observations, the resolution threshold of the tail tree is  $\rho = 1.5$ .

show a scatter plot and a tail tree plot. The data has 4 tails along each coordinate half-axis. The red and orange tails have the same x-coordinate, and the green and blue tails have the same y-coordinate, and the tails are distinguished with the coloring in the tail tree plot.

Figure 12 shows a sample of size  $n = 1000$  from the distribution which has the Gaussian copula with shape parameter  $r = 0.5$ , and Student marginals with degrees of freedom 3. The data has 4 tails which are not exactly along the coordinate axis, due to the positive shape parameter of the Gaussian copula.

Figure 13 shows a sample of size  $n = 1000$  from the distribution which has the Student copula with shape parameter  $r = 0.6$ , and standard Gaussian marginals. The data has 4 sharp tails, which lie along the diagonals between the coordinate axis. The tails lie close together in the tail tree plot, since the red and the orange tails have the same x-coordinate, and so have the blue and the green tails, whereas the red and the blue tails have the same y-coordinate, similarly as the orange and the green tails. The colors distinguish the separate tails in the tail tree plot.

**Archimedean copulas.** Figure 14 shows 1000 observations from the distribution with the Clayton copula with parameter 4, and the standard Gaussian marginals. The dependence is greater in the blue tail, and the tail tree plot shows this tail as a narrower tail than the red tail.

Figure 15 shows 1000 observations from the distribution with the Gumbel copula with parameter 3, and the standard Gaussian marginals. As for the Clayton copula the dependence is greater in the blue tail. However, the red tail is not spreading so wide at the extreme tail as in the case of the Clayton copula. In the tail tree plot one sees how the red tail has an arrow-like shape,

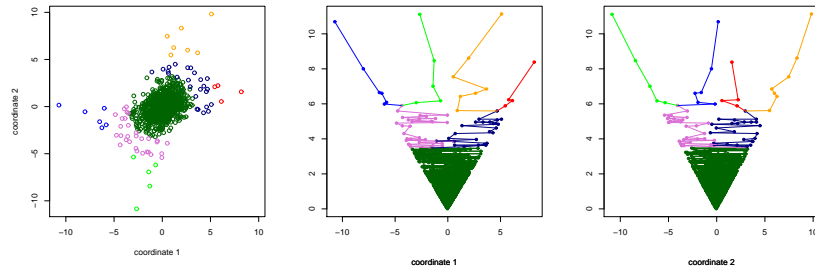


Figure 12: Gaussian copula with shape parameter  $r = 0.5$ , Student marginals with degrees of freedom 3,  $n = 1000$  observations, the resolution threshold of the tail tree is  $\rho = 1.6$ .

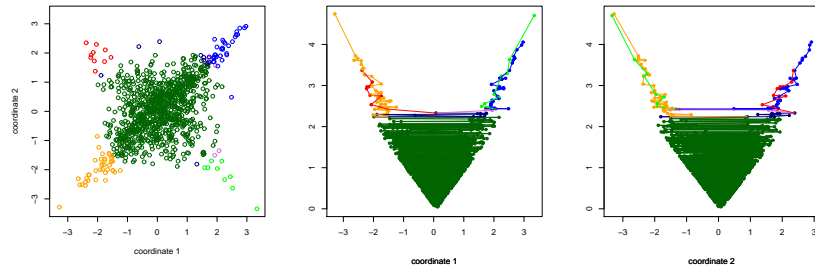


Figure 13: Student copula with shape parameter  $r = 0.6$ , standard Gaussian marginals,  $n = 1000$  observations, the resolution threshold of the tail tree is  $\rho = 1$ .

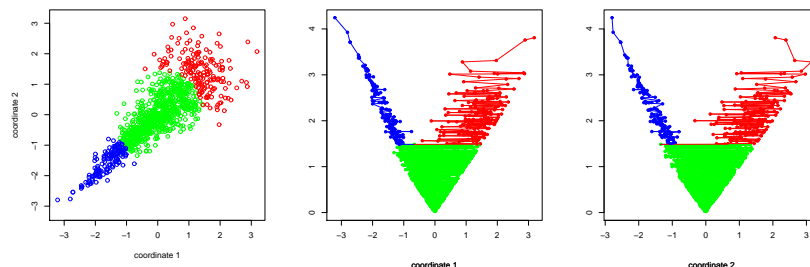


Figure 14: Clayton copula with parameter 4, standard Gaussian marginals, 1000 observations, the resolution threshold of the tail tree is  $\rho = 1$ .

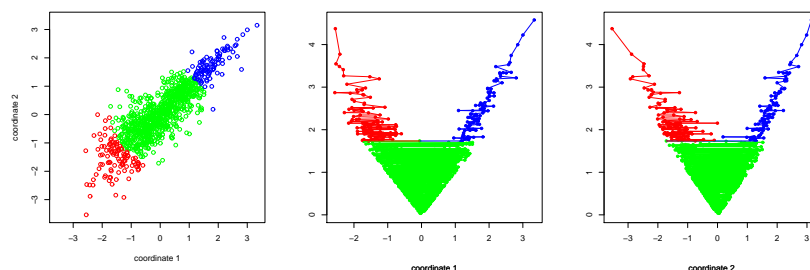


Figure 15: Gumbel copula with parameter 3, standard Gaussian marginals, 1000 observations, the resolution threshold of the tail tree is  $\rho = 1$ .

whereas the red tail in the tail tree plot of the Clayton copula has almost uniform wideness.

Figure 16 shows 2000 observations from the distribution with the Frank copula with parameter 5, and the standard Gaussian marginals. The two tails of the Frank copula are wider than in an ellipse, as can be seen by comparing Figure 16 with the ellipse in Figure 7.

### 4.3 Currencies

Figure 17 shows a tail tree plot of relative changes of exchange rates of the Danish Kroner, Swedish Kronor, Norwegian Kroner, and British Pound. The daily exchange rates are measured with respect to one U.S. Dollar and converted to returns ( $r_i \mapsto (r_i - r_{i-1})/r_{i-1}$ ). The data covers the period from 1971-01-04 to 2006-04-15. There are  $n = 8854$  observations and the dimension is  $d = 4$ . The data is from Federal Reserve Economic Data (<http://research.stlouisfed.org>).

A useful technique in exploratory data analysis is to compare visualiza-

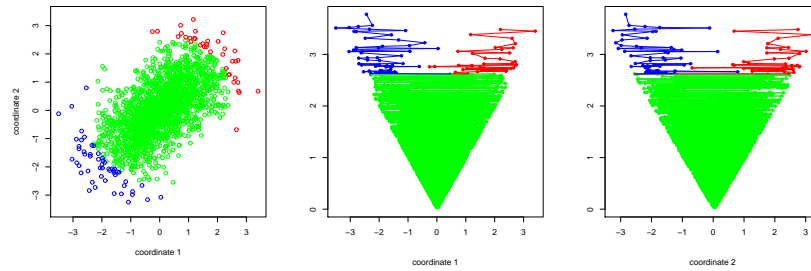


Figure 16: Frank copula with parameter 5, standard Gaussian marginals, 2000 observations, the resolution threshold of the tail tree is  $\rho = 1.1$ .

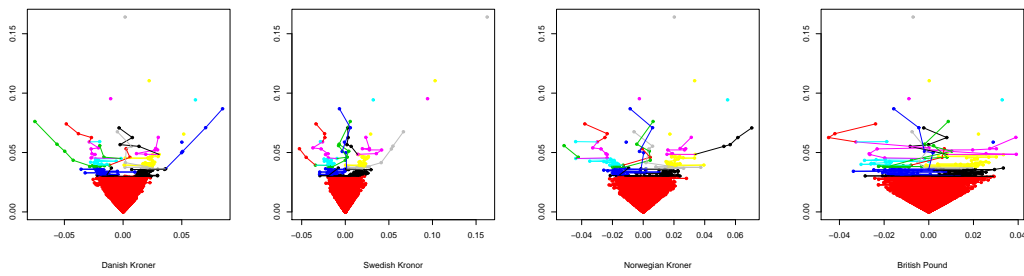


Figure 17: A tail tree plot of relative changes of exchange rates of Danish Kroner, Swedish Kronor, Norwegian Kroner, and British Pound during the period from 1971-01-04 to 2006-04-15. (Some colors are repeated.)

tions of real data to visualizations of simulated data. Figure 17 has similarities with tail tree plots of data whose distribution has Gaussian copula with the correlation parameter  $r \approx 0.5$ , and with Student marginals. Simulated data of this type is visualized in Figure 12. The data of this type has a tail in each coordinate direction and two tails due to dependence. In the 4 dimensional case there would be 10 tails (8 tails in the coordinate directions and 2 tails due to dependence).

In Figure 17 one sees the tails due to dependence (blue+turquoise tail is at the negative orthant and the black+yellow tail is at the positive orthant). One sees the tails along the coordinate axis for Danish Kroner (green and blue), for Swedish Kroner (red and gray), and for Norwegian Kroner (green and black). For the British Pound one sees rudiments of the tails, although these are not present in the tail tree structure.

The tail tree plot of Figure 17 has the resolution threshold  $\rho = 0.01$ , and for this resolution threshold there are 6 outliers (observations which are not  $\rho$ -connected to the other observations). Three of the outliers are large moves in the Swedish Kroner.

In conclusion, the visualization with a tail tree plot indicates that the Gaussian copula with Student marginals is a promising model for the relative changes in exchange rates.

## 5 Tail frequency plot

A *tail frequency plot* visualizes the heaviness of the tails of the underlying distribution. In the multivariate case the tails of the distribution may have anisotropic heaviness: the tails decrease at different rates in different directions. For the elliptical distributions the density has isotropic tails, determined by the 1D generator function, but in the general case the tails are anisotropic.

The nodes of a tail tree are associated with subsets of the data. A tail frequency plot visualizes a tail tree so that each node of the tree is drawn as a line whose length is proportional to the number of observations in the set associated with the node. In fact, we identify the lines as level sets of a 1D function.

**Definition 4** (Tail frequency plot.) *A tail frequency plot of data  $A = \{x^1, \dots, x^n\} \subset \mathbf{R}^d$  is a visualization of a tail tree of data, as defined in Definition 2. A tail frequency plot is a plot of a 1D piecewise constant function, which is defined by associating a separated component of a level set to each node of the tail tree: (1) the length of the separated component of a level*



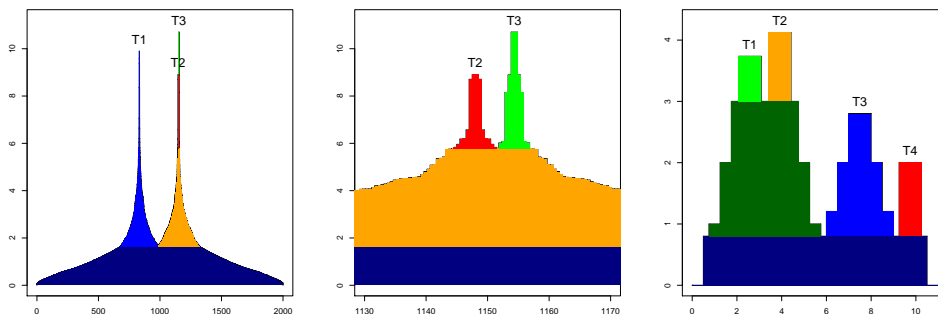


Figure 18: Frame a) shows a tail frequency plot of the tail tree in Figure 2(b), and in Figure 1(a). Frame b) zooms into the function. Frame c) shows a tail frequency plot of the tail tree in Figure 3.

*set is equal to the number of observations in the the node, (2) the height of the separated component of a level set is equal to the distance of the closest observation from the center point, among all observations associated with the node, and (3) the separated components of level sets are nested according to the parent-child relations.*

A more formal definition may be given in terms of function generating trees, defined in Klemelä (2006).

The modes of a tail frequency function correspond to the branches in a tail tree plot. We may show the correspondance by labeling the modes and by coloring the graph of the tail frequency function.

Figure 18(a) shows a tail frequency plot of the data of size  $n = 2000$  in Figure 2(b). The corresponding tail tree plot is shown in Figure 4. The tail frequency function has three modes, and the size of the modes visualizes the number of observations in each tail. Frame b) zooms into the function showing details of the red and green mode. Figure 18(c) shows a tail frequency plot of the tail tree in Figure 3(d) and in Figure 5. There are 11 data points. The tail frequency plot shows the four tails of the data.

## 6 Segmentation of data

Tail trees may be used to define a partition of the data. One may apply the partition to enhance parallel coordinate plots, graphical matrices, and a grand tour. This is illustrated in Sections 7.3-7.4.

A segmentation of the data, corresponding to a tail tree, is such that the observations corresponding to the nodes which are in the same branch belong

to the same segment: a child of node  $m$  belongs to the same segment as  $m$  if it does not have siblings (it is the only child), otherwise the children of  $m$  are the starting nodes of new branches and thus each is a starting node of a new segment.

The segmentation induced by a tail tree is illustrated in Figure 2, where the 5 segments are colored in the tree structure and in the scatter plot. Also, the scatter plots in Sections 4.1-4.2 illustrate the segmentation induced by a tail tree: the segments are colored with separate colors in these scatter plots.

## 7 Comparison with other methods

### 7.1 Shape trees

The concept of a tail tree is related to the concept of a shape tree as defined in Klemelä (2006). A shape tree was defined for a connected set, when the connectedness was defined by taking the resolution threshold  $\rho = 0$  in Definition 1. We have modified that concept of a shape tree in order to visualize sets of finite cardinality. Our approach is analogous to the approach of computational topology, see Robins, Abernethy, Rooney and Bradley (2004).

A shape tree may be applied to visualize level sets of a density estimate, but with tail trees we may visualize the raw data, avoiding the density estimation step. Since density estimation is difficult in high dimensional cases we may be able to analyze higher dimensional data with tail trees than with shape trees. In addition, sometimes a part of the data arises from a discrete distribution so that there does not exist an underlying density for the complete data.

**Tail probability plot.** It is of interest to define a population concept corresponding to a tail frequency plot, defined in Definition 4. The population concept may be defined with the help of a shape tree. A shape tree was visualized in Klemelä (2006) with a radius plot, which visualizes a shape tree so that a node of a shape tree is identified with a separated component of a level set of a 1D function, and the length of the separated component of a level set is proportional to the volume of the set associated with the node. We get a *tail probability plot* when we choose instead the length of a separated component of a level set to be equal to the probability content of the set associated with the node. When a shape tree corresponds to a low-level level set (or the support) of the underlying density, then a tail frequency plot is an empirical version of the tail probability plot.

## 7.2 Parallel level plots

In *parallel level plots* one introduces a new dimension to the point cloud, called “level”. Each observation is attached with a level and one makes  $d$  parallel scatter plots from each coordinate together with the level. Thus we identify the observations across the scatter plots through the value of the level of the observations.

**Definition 5** Parallel level plot of data  $x^1, \dots, x^n \in \mathbf{R}^d$ , associated with a level mapping  $\text{level}(x^i) \in \mathbf{R}$ , consist of  $d$  windows, where the  $k$ -th window,  $k = 1, \dots, d$ , shows a scatter plot of points  $(x_k^i, \text{level}(x^i))$ ,  $i = 1, \dots, n$ .

Typically the level mapping should satisfy  $\text{level}(x^i) \neq \text{level}(x^j)$ , when  $i \neq j$ ,  $i, j = 1, \dots, n$ , which makes the unique identification of the observations across the windows possible. For multivariate time series data parallel level plots have been common: one takes the level of data points to be their time index. Also, one dimensional curves in the  $d$ -dimensional Euclidean space may be visualized naturally with parallel level plots by taking the parameter of the curve to be the level.

When visualizing general point clouds the usefulness of a parallel level plot depends on the choice of the level. Figure 19(a) shows a scatter plot of a data with 4 tails and frames b-c) show the same data with a parallel level plot where we have attached a unique random level in  $\{1, \dots, 1000\}$  to each observation (the level is the index of the observation). The scatter plot and the parallel level plot of contain the same information but the scatter plot lets a human eye to detect the pattern of 4 tails in the data. A natural choice for the level is the distance from a center point, say, from the arithmetic mean. Figure 20(a)-(b) shows a parallel level plot of the data in Figure 19 when we choose the level to be the distance from the arithmetic mean. Figure 20(a)-(b) indicates that there are at least 2 tails in the data, but the fact that there are 4 tails is only revealed when we draw a tail tree plot, which is shown in frames c-d). The 4 tails of the data are shown as 4 branches of the tail tree, colored with black, red, blue, and green. The tail tree plot adds a sufficient structure to the visualization making it possible for humans to detect the 4 tails in the data.

## 7.3 Graphical matrices

A graphical matrix is a  $n \times d$  matrix of graphical elements; a  $n \times d$  data matrix of real numbers is transformed by representing each real number by a graphical element. Graphical matrices were studied by Bertin (1967), Bertin (1981). There are many ways to code real numbers with graphical elements.

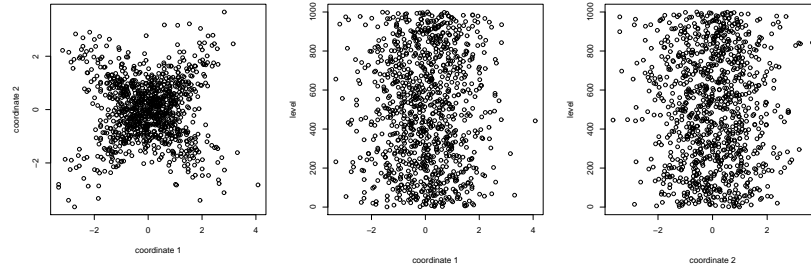


Figure 19: Window a) shows a scatter plot of a data of size 1000 generated from a mixture of two Gaussian distributions. Windows b-c) show a parallel level plot of the data in frame a), with the indices of the observations as the level.

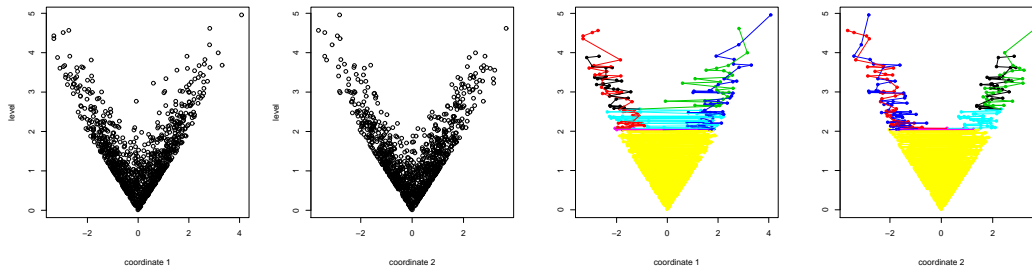


Figure 20: Windows a-b) show a parallel level plot of the data in Figure 19 with the distance from the arithmetic mean as the level. Windows c-d) show a tail tree plot of the data. The resolution threshold is  $\rho = 0.65$ .

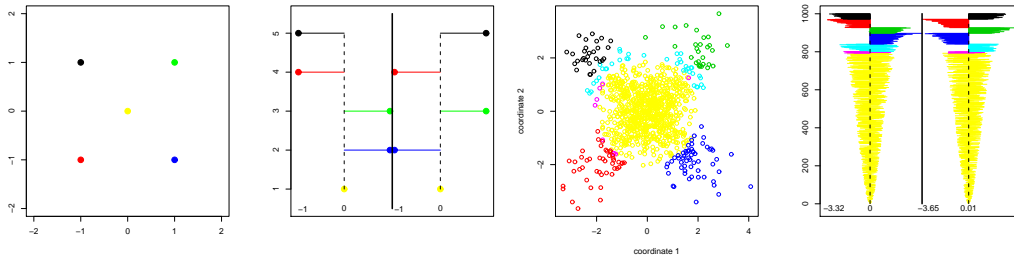


Figure 21: Window a) shows a scatter plot of 5 points, and window b) shows a graphical matrix of the data. Window c) shows a scatter plot with 7 segments colored, and window d) shows a graphical matrix of the data, when the permutation is given by the tail tree of Figure 20(c)-(d).

Figure 21(a) shows a scatter plot of 5 observations and frame b) shows one type of graphical matrix of this data. We have drawn each observation as a circle which is connected with a line to the arithmetic mean of the coordinate. This graphical matrix is identical with a parallel level plot, where the level is the index of the observation, and when we delete the lines and keep only the circles.

The usefulness of a graphical matrix depends on whether one can find an informative permutation of the rows/observations (and of the columns/variables). Similarly to the parallel level plot in Figure 19(b)-(c), a random permutation leads typically to useless graphical matrices. A suggestion of Bertin (1981) is to order indices so that the values of one variable are in a monotonic order.

Tail trees may be used to find a useful permutation with the help of the segmentation defined in Section 6. One chooses a permutation where the observations which belong to the same segment are grouped together in the graphical matrix.

Figure 21(c) shows a colored scatter plot of the data in Figure 19. The segments induced by the tail tree in Figure 20(c)-(d) are colored. Figure 21(d) shows a graphical matrix where we have used a permutation corresponding to the segmentation in frame c). The observations belonging to the same segment are grouped together, and the observations are ordered inside a group according to the distance to the arithmetic mean of the complete data. The four tails are well visualized. For example, the black “flag” in the graphical matrix points to the left in the column of the first coordinate, and to the right in the column of the second coordinate. Thus we know that this tail lies in the north-west direction from the center.

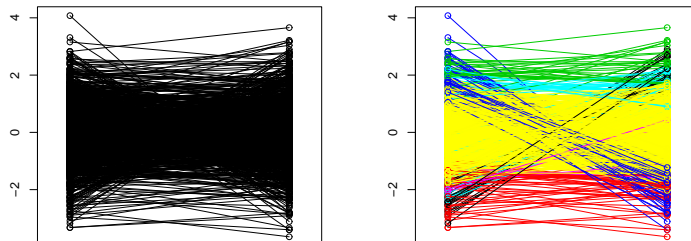


Figure 22: Frame a) shows a parallel coordinate plot of the data in Figure 19. Frame b) shows a parallel coordinate plot where the lines and points are colored according to the segmentation of the tail tree in Figure 20(c)-(d).

## 7.4 Parallel coordinate plots

Parallel coordinate plots were introduced by Inselberg (1985) and Wegman (1990). Parallel coordinate plots resemble parallel level plots: in both cases one visualizes 1D projections of the data at the coordinate axis, and uses a method to identify the observations among the  $d$  projections. In the case of parallel coordinate plots one identifies the observations among the projections by joining the observations with straight lines.

The segmentation of the data induced by a tail tree, defined in Section 6, may be used to enhance parallel coordinate plots. Figure 22(a) shows a parallel coordinate plot of the data in Figure 19. Frame b) shows a colored parallel coordinate plot, where the lines and points are colored according to the segmentation induced by the tail tree in Figure 20(c)-(d).

Parallel coordinate plots provide a more compact visualization of the data than parallel level plots. Consequently, parallel coordinate plots suffer more from overplotting. Indeed, the lines connecting the observations introduce more ink to the plot. In contrast, parallel level plots introduce a new variable to the plot (level), and spread the observations with the help of the new variable, thus diminishing the problem of overplotting. One may also decrease the problem of overplotting with parallel coordinate density plots, see Miller and Wegman (1991).

## 7.5 Projections and slices

The segmentation induced by a tail tree, defined in Section 6, may be used to enhance projections of multivariate point clouds: one colors the observations

so that the observations in the same segment have the same color. This coloring technique has been used in the scatter plots of Sections 4.1-4.2. When one takes a grand tour along the projections of the data this kind of coloring may be extremely helpful. A grand tour is defined in Asimov (1985).

Projections may be excellent in visualizing clusters in the data. When we are interested in the dependence between the variables, then tail tree plots may provide good dimension insensitive alternatives to projections.

## 7.6 Spatial trees

A tail tree is a spatial tree (a tree which generates a branching curve). These are trees whose every node is associated with a vector of  $d$  real numbers and with a real valued height, so that the height of a child is larger than the height of the parent. Also a minimal spanning tree and a shortest path tree are spatial trees, when we define the height of the nodes properly. Thus we may visualize these trees with similar plots as a tail tree plot. Friedman and Rafsky (1981) propose to visualize a minimal spanning tree by plotting the data in 2D so that the  $n - 1$  edge lengths are preserved, and showing in the plot the lines connecting the observations.

## 7.7 Database exploration

Database exploration in Keim and Kriegel (1994) has some similarities with tail tree plots. Their center point is determined by the user (as a query to the data base), whereas we calculate the center point with an algorithm (as the arithmetic mean). They sort the data points according to a distance to the center point. The center point is drawn at the center of the display and the other data points are arranged on a rectangular spiral originating from the center point. The distances to the center point are visualized with colors. In addition, an own display is dedicated for each coordinate. The position of the data points in the coordinate displays is the same as in the main display, but the coloring is determined by the univariate distance in the given coordinate from the observation to the center point. (A coordinate is also called a “dimension” or “predicate”.)

A weakness of database exploration might come from the fact that colors are not well suited for expressing quantitative information like distances. We have expressed distances by using scatter plots where the y-coordinate gives the overall distance and the x-coordinate gives the univariate distance in the given coordinate, and colors are used to express nominal information (the labeling of the tails).

## 8 Discussion

A tail tree is a tree structure among the observations. It is a useful data structure which can be applied to enhance many visualization tools: it can be applied with scatter plots (grand tour), parallel coordinate plots, and with graphical matrices. A tail tree plot is a visualization technique which is tailored for the visualization of tail trees. It has some advantages over other methods: it is a dimension insensitive tool, it may be used to visualize dependence and anisotropic tails, and it is not severely affected by overplotting.

Graphs may be used in two different ways: to communicate information to an audience and to analyze data. Graphs used for communications (presentation graphs) tend to be simple and are mainly intended for the display of discovered patterns. Graphs used for analysis may be quite detailed, complicated, or esoteric, containing most or even all of the original data, and are tools for the detection of important or unusual features in the data. (Spence and Lewandowsky (1990), pages 19-20) Tail tree plots may be used to analyze data. We believe that tail tree plots have intuitive appeal which makes them useful also for presentation purposes.

## References

- Asimov, D. (1985), ‘The grand tour: a tool for viewing multidimensional data’, *SIAM J. of Scient. and Statist. Comp.* **6**, 128–143.
- Bertin, J. (1967), *Semiologie Graphique*, Gauthier Villars, Paris.
- Bertin, J. (1981), *Graphics and graphic information-processing*, de Gruyter.
- Friedman, J. H. and Rafsky, L. C. (1981), ‘Graphics for the multivariate two-sample problem’, *J. Amer. Statist. Assoc.* **7**(374), 277–287.
- Inselberg, A. (1985), ‘The plane with parallel coordinates’, *The Visual Computer* **1**, 69–91.
- Keim, D. A. and Kriegel, H.-P. (1994), ‘VisDB: database exploration using multidimensional visualization’, *IEEE Computer Graphics Applic.* **14**(5), 40–49.
- Klemelä, J. (2004), ‘Visualization of multivariate density estimates with level set trees’, *J. Comput. Graph. Statist.* **13**(3), 599–620.



- Klemelä, J. (2006), ‘Visualization of multivariate density estimates with shape trees’, *J. Comput. Graph. Statist.* **15**(2), 372–397.
- Liu, R. Y., Parelius, J. M. and Singh, K. (1999), ‘Multivariate analysis by data depth: descriptive statistics, graphics and inference’, *Ann. Statist.* **27**(3), 783–858.
- Miller, J. J. and Wegman, E. J. (1991), Construction of line densities for parallel coordinate plots, *in* A. Buja and P. Tukey, eds, ‘Computing and Graphics in Statistics’, Springer, New York, pp. 107–123.
- Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer.
- Robins, V., Abernethy, J., Rooney, N. and Bradley, E. (2004), ‘Topology and intelligent data analysis’, *Intelligent Data Analysis* **8**(5), 505–515.
- Spence, I. and Lewandowsky, S. (1990), Graphical perception, *in* J. Fox and J. Scott Long, eds, ‘Modern methods of data analysis’, Sage Publications, pp. 13–57.
- Wegman, E. J. (1990), ‘Hyperdimensional data analysis using parallel coordinates’, *J. Amer. Statist. Assoc.* **85**(411), 664–675.

## A Copulas and marginal distributions

We define shortly the mathematical concepts of copula and marginal distribution. Distribution function  $F$  may be decomposed into a part which describes the dependency and into a part which describes the marginal distributions. We call a *copula* the part which describes the dependency. The word “copula” is a latin word which means a “link”, “tie”, or “bond”. The basic idea is that any distribution function  $F : \mathbf{R}^d \rightarrow \mathbf{R}$  of a random vector  $(X_1, \dots, X_d)$  may be written as  $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$ , where  $C$  is the copula and  $F_i$ ,  $i = 1 \dots, d$ , are the marginal distribution functions:  $F_i(x_i) = P(X_i \leq x_i)$ . A copula is a function  $C : [0, 1]^d \rightarrow [0, 1]$  which links univariate marginals to their multivariate distribution. We may construct parametric and semiparametric families of multivariate distributions by giving separately a model for the dependency and separately a model for the marginal distributions. Typical copulas include the Gaussian copula, Student copula, Clayton copula, Gumbel copula, and Frank copula. The definitions of the copulas can be found for example in Nelsen (1999). Typical marginal distributions include the Gaussian and Student distributions. The standard Gaussian distribution has density  $c \exp\{-t^2/2\}$  and the Student distribution

has density  $c(1 + t^2/\nu)^{-(\nu+1)/2}$ , where  $c$  is a normalization constant and  $\nu$  is a parameter (degrees of freedom). The Student distribution has heavier tails than the exponentially decreasing tails of a Gaussian distribution, and is suited for describing phenomena where extreme events happen with high probability.

## B Supplementary material

### B.1 Tail tree

**Decoration of a tail tree.** A node of a tail tree is associated both with an observation and with a subset of the observations. We may always define the other decoration in the terms of the other. Let us denote with  $\text{set}(m)$  the set associated with node  $m$  in a tail tree of the data, and let us denote with  $\text{obs}(m)$  the observation associated with the node. We may write a set associated with node  $m$  as

$$\text{set}(m) = \bigcup \{\text{obs}(c) : c = m \text{ or } c \text{ is a descendant of } m\}.$$

We may write the observation associated with a node  $m$  as:

$$\text{obs}(m) = \begin{cases} \text{set}(m) \setminus \bigcup \{\text{set}(c) : c \text{ is a child of } m\}, & m \text{ a non-leaf node,} \\ \text{set}(m), & m \text{ a leaf node,} \end{cases}$$

where we use the convention  $\{x\} = x$ , where  $x \in \mathbf{R}^d$ . The decoration with observations is used to define a tail tree plot and the decoration with subsets of observations is used to define a tail frequency plot.

**Other clustering distances.** The definition of separated sets in Definition 1 may be rephrased in the following way: sets  $A, B \subset \mathbf{R}^d$  are separated for the resolution threshold  $\rho \geq 0$ , if  $D(A, B) > 2\rho$ , where  $D(A, B) = \inf\{\|x - y\| : x \in A, y \in B\}$  is the single linkage distance between two sets. The concept of separated sets is the fundamental underlying concept in the definition of a tail tree, and it is natural to ask whether the single linkage distance may be replaced by some other distance between sets, to get other versions of a tail tree. Indeed, each hierarchical agglomerative clustering algorithm is based on some concept of a distance  $D(A, B)$  between sets  $A, B \subset \mathbf{R}^d$ , where the distance need to be defined only for sets of finite cardinality. For example, the complete linkage clustering uses the distance  $D(A, B) = \sup\{\delta(x, y) : x \in A, y \in B\}$ , where  $\delta$  is typically the Euclidean distance, and the average linkage clustering uses the distance  $D(A, B) = (\#A \cdot \#B)^{-1} \sum_{x \in A, y \in B} \delta(x, y)$ , where  $A$  and  $B$  have finite cardinality. The essential condition for the definition of a tail tree to be meaningful is the following.

$$\begin{aligned} &\text{When } A, B \subset \mathbf{R}^d \text{ are } \rho\text{-separated and } A_0 \subset A, B_0 \subset B, \\ &\text{then sets } A_0 \text{ and } B_0 \text{ are } \rho\text{-separated.} \end{aligned} \tag{2}$$

Indeed, when we build a tail tree, we are at each step removing observations from the original set. If the condition (2) is violated it may happen that after two branches have been separated, they may later join again, after more observations have been removed. In this case we are building a graph which is not a tree. These kinds of graphs may indeed be useful in visualization and segmentation, but we don't study these graphs in this article. The distances of the complete linkage clustering and the average linkage clustering do not satisfy condition (2), when  $\delta$  is the Euclidean distance.

**Tail clustering.** In density based clustering, or mode clustering, clusters are defined in terms of the underlying density as regions of high density separated from other such regions by regions of low density. *Tail clusters* may be defined as regions of low density separated from other such regions after removing the regions of high density and regions of zero density (or regions of extreme low density). Tail clustering may be used in the cases where the underlying distribution is unimodal. For example, one might want to cluster a data base of customers in order to concentrate marketing efforts to specific groups, which may lie in the tails of the distribution.

A conceptually natural approach to density based clustering finds a tree structure of nested clusters, corresponding to the level set tree of the underlying density, see Stuetzle (2003) Klemelä (2004), and also Ankerst, Breunig, Kriegel & Sander (1999). Tail trees provide an analogous approach to tail clustering, since they find a nested sequence of tail clusters.

Tail tree clustering has similarities with divisive hierarchical clustering (as opposed to agglomerative hierarchical clustering), since both start with the complete data and extract clusters recursively. However, divisive clustering makes a partition of the data, but in a tail tree the number of observations is decreasing at each step. Dendrograms (hierarchical clustering trees) in general make a recursive partition of the observations, so that the union of the sets associated with children is always equal to the set associated with the parent, and thus dendrograms have a different structure than tail trees.

In this article we are interested in tail clustering as a tool to visualize data, and are not interested in other applications of tail clustering.

## B.2 Tail tree plot

**Tail tree plots versus scatter plots.** Figure 23 shows how points in a scatter plot correspond to nodes in a tail tree plot. Figure 23(a) shows points inside a unit ball, Figure 23(b) shows images of these points under the mapping  $(x_1, x_2) \mapsto (x_1, (x_1^2 + x_2^2)^{1/2})$ . The points on the x-axis (the green points in the 1st window) are mapped on a wedge in the x-window of

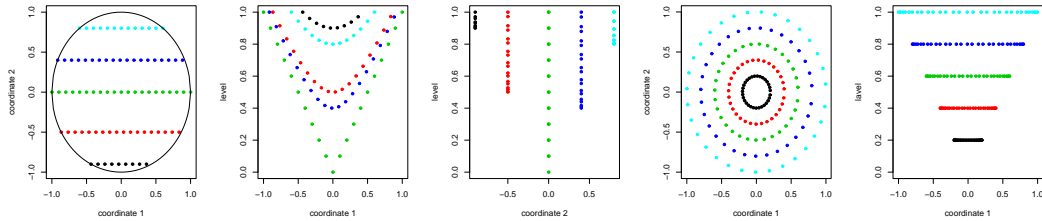


Figure 23: Frames a-b) show that horizontal lines are mapped to (smoothed) wedges in the x-window of a tail tree plot. Frame c) shows that horizontal lines are mapped to vertical lines in the y-window of a tail tree plot. Frames d-e) show that spheres are mapped to horizontal lines.

the tail tree plot, and the points parallel to the x-axis are mapped to lifted smoothed wedges. Figure 23(c) shows images under the mapping  $(x_1, x_2) \mapsto (x_2, (x_1^2 + x_2^2)^{1/2})$ . Figure 23(d)-(e) shows that the points on spheres are mapped to horizontal lines in a tail tree plot. Sometimes we plot the wedge  $\{(x_i, \lambda) : |x_i - \mu_i| = \lambda, \lambda \geq 0\}$  as a dashed line in a tail tree plot, see Figures 9, 10, 26.

**Overplotting.** Any pixel based visualization (a scatter plot, parallel level plot, parallel coordinate plot, graphical matrix, tail tree plot) suffers from overplotting when the size of the data is large. Tail tree plots suffer from overplotting in the center of the data, but in the tails of the data single observations are distinguishable. Usually it is these “outliers” which are most interesting and in need to be identified, whereas central observations are not individually interesting.

One may reduce overplotting in a tail tree plot by plotting only the bullets corresponding to the observations and leaving out the lines connecting the observations. Often one can visualize the tree structure sufficiently well only with the coloring of the observations.

An additional way to reduce the overplotting is subsetting. For example, in the *likelihood subsetting* one plots only the points  $x^i$  for which  $\hat{f}(x^i) \geq \lambda$ , where  $\lambda > 0$  is a level and  $\hat{f}$  is a density estimate.

**Colored parallel level plots.** An idea behind the definition of a tail tree plot is that 1D curves in the d-dimensional Euclidean space may be visualized with coordinate functions (projections to the coordinate axes), and 1D branching curves may be visualized with projections when we add colors to the plot to identify different branches. Figure 24 visualizes a hyperbolic

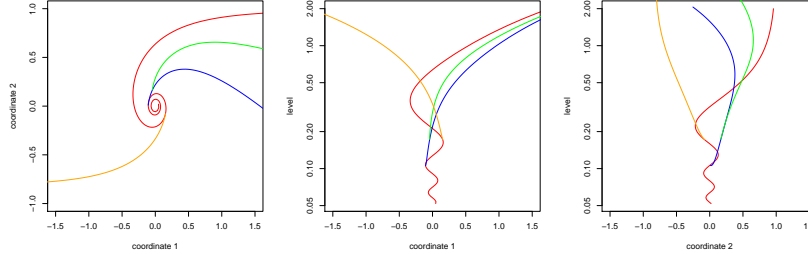


Figure 24: A hyperbolic spiral with 3 branches. Window a) shows the hyperbolic spiral with red. The blue branch is evolving first, then the green branch is evolving from the blue branch, and finally the orange branch is evolving from the red main spiral. Windows b-c) show the colored parallel level plot; the x-axes correspond to the coordinates 1 and 2, and the y-axes give the distance from the origin. The red curve is defined by  $r \mapsto (r \cos(1/r), r \sin(1/r))$ ,  $r \geq 0.05$ .

spiral with 3 branches. Frame a) shows a scatter plot and frames b-c) show a colored parallel level plot. A parallel level plot is defined in Definition 5 and a tail tree plot is closely related to a colored parallel level plot.

**Spherically symmetric data.** We show how spherically symmetric data sets look in tail tree plots. In this case the shape of the point cloud is always the same, but the spread and the location of the data may vary. Figure 25 shows three scatter plots of spherically symmetric data. We generated a sample of size 1000 from the uniform distribution on the 2D ball, from the 2D standard Gaussian distribution, and from the 2D Student distribution with degrees of freedom 3. Figure 26 shows the first coordinates of tail trees of the data in Figure 25. We have chosen the resolution thresholds so that the data sets are connected, and this required to choose thresholds 0.6, 3, and 18.

Figure 26 shows that a ball looks like a triangle in a tail tree plot. The bullets are always on the upper side of the wedge which is shown in Figure 26 as dashed lines. For every data set a tail tree plot has the triangle shape at the bottom. The triangle is not filled at every sector if the data is not spherical. For spherical data the triangle is not filled at the upper levels, except in the uniformly distributed case. Tail tree plots visualize the spread by showing how the observations become more sparse in the upper levels of the tail tree plot.

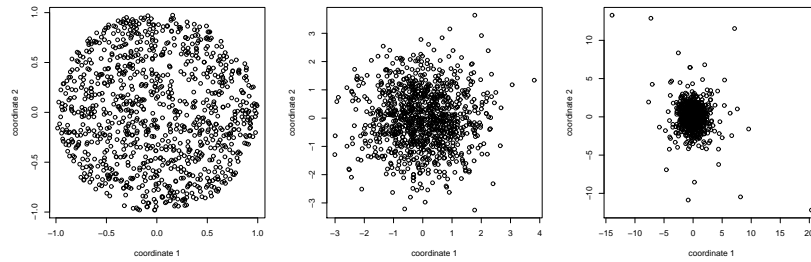


Figure 25: Three scatter plots of 1000 observations: a) uniformly distributed data on the unit ball, b) standard Gaussian data, c) Student distribution with degrees of freedom 3.

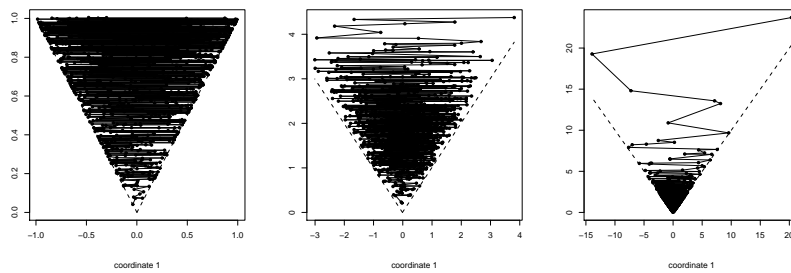


Figure 26: The first coordinates of tail tree plots of the data in Figure 25

### B.3 Tail frequency plot

**Coloring of a tail frequency plot.** By labeling the modes of the function in a tail frequency plot we may indicate the correspondence between the modes of the function and the tails of the data as shown in a tail tree plot. We may use colors to enhance the visualization of the correspondence. In a tail tree plot every node (observation) was colored. A tail frequency plot is a plot of a piecewise constant function, which has a finite number of distinct level sets. These distinct level sets correspond to the nodes of the tail tree, and we may color these level sets with the same color with which the corresponding observation was colored. We color the whole area under the graph of the function in such a way that a color is changed when two separated level sets are merging.

**Scale curves.** In multivariate cases it is useful to study 1D spherical distribution function  $G_n : [0, \infty) \rightarrow [0, 1]$ ,

$$G_n(r) = n^{-1} \#\{x^i : x^i \in B_r(\mu), i = 1, \dots, n\},$$

where  $B_r(\mu) = \{z \in \mathbf{R}^d : \|z - \mu\| \leq r\}$ , and  $\mu \in \mathbf{R}^d$  is a given center point. Function  $G_n$  gives the empirical probabilities of an increasing sequence of balls, and thus  $G_n$  visualizes the heaviness of the tail of the underlying density. Liu et al. (1999) replace the balls  $B_r(\mu)$  with depth regions and call the related function a “scale curve” or a “scalar form of dispersion”.

When a tail tree does not have branches, then a tail frequency plot of the tail tree visualizes exactly the same information as function  $G_n$ . However, when a tail tree has many leaf nodes, then a tail frequency plot visualizes more information: it visualizes the heaviness of the tails separately for each disconnected tail of the underlying density. Spherical distribution function may be used for spherically symmetric densities but with the help of tail frequency plots we may visualize anisotropic spread.

**Examples of tail frequency plots.** Figure 27 shows tail frequency plots for the data sets coming from uniform, Gaussian, and Student distributions. The data is the same as in Figure 25 and the corresponding tail tree plots are shown in Figure 26.

Figure 28 shows tail frequency plots for Gaussian ellipsoids. Frame a) shows a tail frequency plot of the 2D ellipses in Figures 6-8. The orientation of the data does not affect the tail frequency plots. Frame b) shows a tail frequency plot of the 4D Gaussian data where the standard deviation of each marginal is different. The corresponding tail tree plot is shown in



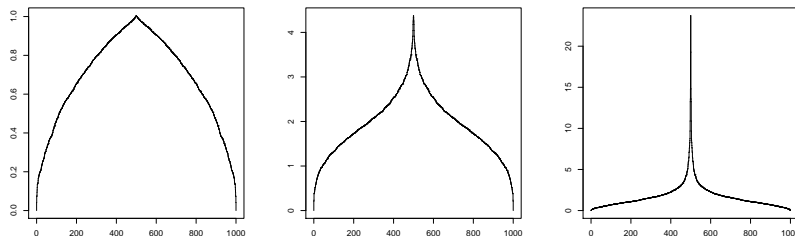


Figure 27: Tail frequency plots of spherically symmetric data; uniform, Gaussian, and Student. The tail frequency plots correspond to the tail tree plots in Figure 26.

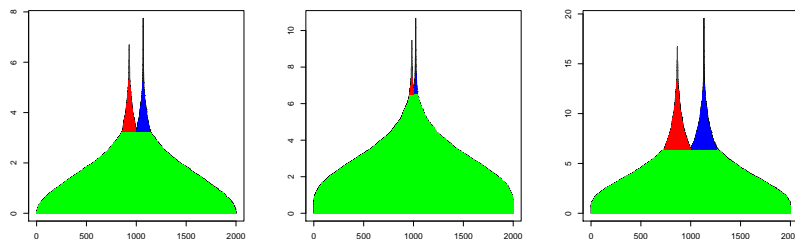


Figure 28: Tail frequency plots of Gaussian ellipsoids; a) a 2D ellipse, b) a 4D ellipsoid with varying standard deviations, and c) a 4D ellipsoid with a dominating standard deviation. The tail frequency plots correspond to the tail tree plots in Figures 6-8, Figure 9, and Figure 10.

Figure 9. Frame c) shows a tail frequency plot of the 4D Gaussian data with a dominating standard deviation. The corresponding tail tree plot is shown in Figure 10.

Figure 29 shows tail frequency plots corresponding to the tail tree plots in Figures 11-13 (independent Student marginals, a Gaussian copula with Student marginals, and a Student copula with Gaussian marginals). The modes are of so small size that we need to zoom into the graph of the function to show the modes, and this is done in Figure 30.

## B.4 Comparison to other methods

**Parallel level plots versus tail tree plots.** Figure 31 shows that a parallel level plot cannot distinguish the cases where there are 4 tails from the cases where there are 3 tails: the parallel level plots in Figure 20(a)-(b) and

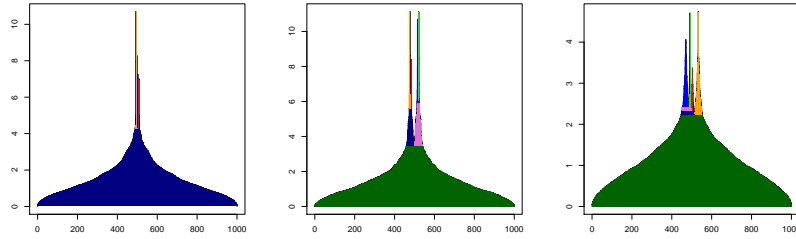


Figure 29: Independent Student marginals, a Gaussian copula with Student marginals, and a Student copula with Gaussian marginals; tail frequency plots corresponding to the tail tree plots in Figures 11-13.

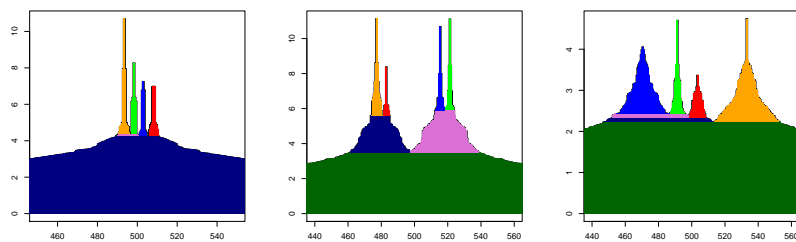


Figure 30: A detailed view of the modes in the tail frequency plots of Figure 29.

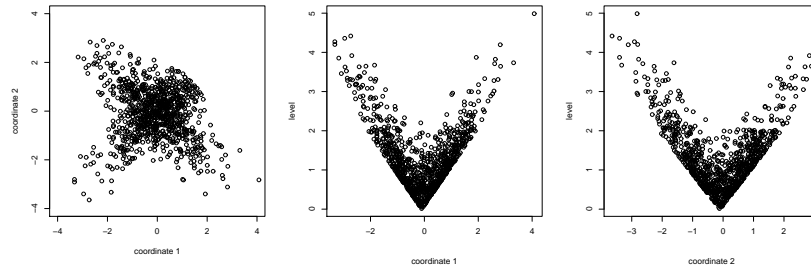


Figure 31: The data in Figure 19 when one tail is cutted away. Window a) shows a scatter plot and windows b-c) show a parallel level plot when the distance from the arithmetic mean is the level.

in Figure 31(b)-(c) look similar although the latter has one tail less.