10 Clustering using wavelet transformation

Abdullah Almasri and Ghazi Shukur

This chapter introduces and describes an alternative clustering approach based on the Discrete Wavelet Transform (DWT) which satisfies requirements that other clustering methods, like discriminative-based clustering and model-based clustering approaches, do not satisfy.

The clustering method has been constructed using wavelet analysis that has the ability of decomposing a data set into different scales. Wavelet algorithm is then used to specify the number of the clusters and quality of the clustering results at each scale. The same algorithm can be generalized for more than one-dimensional data. Some examples about how to use this approach are presented in this chapter, using different sample sizes, and where different kinds of noises are imposed on simulated data. These examples show the successfulness and efficiency of this kind of methodology in detecting clusters under different situations.

1 Introduction

Cluster analysis (originally used by Tryon, 1939) combines a number of different classification algorithms that are usually done to join cases or a set of data objects into groups or clusters when the group membership is not known a priori. Hence, it is a technique for linking individuals or objects into unknown groups or clusters such that those within each group or cluster are more closely related to one another than those assigned to other clusters. An observation or object can be explained by a number of measurements or by its relation to other observations or objects. Clustering, generally, is an unsupervised classification where no predefined classes are given. It can be used as a tool to gain insight into data, or as a pre-processing step for other algorithms. Cluster analysis is used in several areas, e.g. in pattern recognition, spatial data analysis, image processing and economic science (especially, market research, labour market and regional economics).

In the last few decades a variety of methods has been proposed for ways to conduct cluster analysis. In general, there are two main categories into which existing clustering approaches can be classified, namely, discriminative and model-based approaches.

In the discriminative based approaches one measures the distance or similarity/dissimilarity between two individual observations, and then join similar samples together into clusters. The most commonly used distance measures are Euclidean or standardized Euclidean distance and Mahalanobis distance. These approaches include hierarchical agglomerative clustering using various between-cluster dissimilarity measures, such as smallest dissimilarity (single linkage, also called the nearest-neighbour technique, that takes the inter-group dissimilarity to be that of the closest or least dissimilar pair); maximum dissimilarity (complete linkage, also called furthest-neighbour technique, takes the inter-group dissimilarity to be that of the furthest or most dissimilar pair); average dissimilarity (average linkage uses the average dissimilarity between the groups); or the k-means algorithm (MacQueen, 1967); and Self-Organizing Maps (Kohonen, 2001). These methods are relatively easy to apply and often give good results in simple cases. At this stage, it is crucial to mention that these methods are highly empirical, and that, in more complex cases, different methods can lead to different clustering, regarding both the number of clusters and the content. This might be due to the facts that these methods are sensitive for outliers; they do not involve statistical tools for choosing the number of clusters and do not pay attention to measurement error in the dissimilarities or to clustering uncertainties.

Parametric model-based approaches, on the other hand, try to find generative models from the data, with each model corresponding to one particular cluster. Note that the type of model here is often specified a priori, e.g. Gaussian or hidden Markov models, and can be processed in both single and multi-level perspectives. In other words, model-based clustering is a framework for putting cluster analysis on a principle statistical footing; for reviews and more discussions, see Fraley and Raftery (2002). It is based on probability models in which objects are assumed to follow a finite mixture of probability distributions such that each component distribution represents a cluster. One of the most important advantages of the model-based clustering over the discriminative based clustering is that the model-based clustering objects and estimates component parameters simultaneously. Processing in this manner leads to avoiding biases that might exist when the clustering and the estimation are done separately. Another advantage is the ability to use statistical model selection methods when specifying the number of components and their probability distributions. Moreover, model-based clustering provides clustering uncertainties, which is important especially for objects close to cluster boundaries. On the other hand, model-based clustering requires object coordinates rather than dissimilarities between objects (as in the discriminative based clustering) as an input. This means that model-based clustering is only applicable when object coordinates are available, and not when dissimilarities are provided.

More recently, Oh and Raftery (2003) developed a model-based clustering method for dissimilarity data. They assume that an observed dissimilarity measure is equal to the Euclidean distance between the objects plus a normal measurement error. They model the unobserved object configuration as a realization of a mixture of multivariate normal distributions, each one of which corresponds to a different cluster. The authors then carried out Bayesian inference for the resulting hierarchical model using Markov Chain Monte Carlo (MCMC). The resulting method combines multidimensional scaling and model-based clustering in a coherent framework.

Note that these and several other methods for clustering are not easy to apply when dealing with large and/or multidimensional data sets. In our opinion, we consider a clustering approach to be good if it is efficient, it detects clusters of arbitrary shape and it is insensitive to the noise (outliers).

In this chapter we introduce another clustering approach based on Discrete Wavelet Transform (DWT) which satisfies all the above requirements. The wavelet methods have been shown to be very useful in different areas of research, such as signal processing, image analysis, geophysics and atmospheric sciences. Recently, the wavelet methods have also been introduced to the subject of economics and, in particular, time series econometrics (see, e.g. Almasri & Shukur, 2003; Percival & Walden, 2000; Ramsey and Lampart, 1998). However, using the multi-resolution property of wavelet transform, we can effectively identify arbitrarily shaped clusters in different degrees of detail. Since the wavelet transform is a very natural tool to detect spatial scales and clusters in an image, the technique can even be applied to spatial patterns. The main advantages of this technique over other methods of spatial analysis are its ability to preserve and display hierarchical information while allowing for pattern decomposition; see Chave and Levin (2003). The wavelet transform can be used in many applications such as Geographic Information Systems (GIS), segmentation of airborne laser scanner data, image database exploration, seismology, etc. It can also be used in economics and other social sciences in a GIS framework which increase the understanding of many sorts of social processes, including patterns of employment and unemployment, crime, economic growth and population change; see, e.g. Morehart et al. (1999) and Vu and Tokunaga (2001). In the two-dimensional cases one can apply our methodology to study real-life data such as the relationship between firms of different sizes and their natural resource bases. Here, the centre of geographically referenced bins/blocks can be considered as the first dimension and the second dimension as the variable of interest.

Since clustering analysis is a vast area of statistical methodology, we previously confined ourselves to a brief description of some of the available methods, and further details are found in standard textbooks and cited references. We will, however, discuss more thoroughly the clustering associated with the wavelet methodology, since this topic is not mentioned in standard textbooks.

The chapter is arranged as follows: in the next section we present the wavelet methodology as an alternative clustering approach. In Section 3, we demonstrate a number of artificial examples, while, in Section 4, we give some concluding remarks.

2 Methodology

In this section we define the wavelet filter and DWT in one and two dimensions. The basic idea of wavelet analysis is to imitate the Fourier analysis, but with functions (wavelets) that are better suited to capture the local behaviour of data sets. The wavelet transform utilizes a basic function (called the mother wavelet), then dilates and translates it to capture features that are local in time and local in frequency. The wavelet function, say $\psi(.)$, should satisfy the following two basic properties:

- The integral of the real-valued function $\psi(.)$ is zero:

$$\int_{-\infty}^{\infty} \psi(u)du = 0. \tag{10.1}$$

- The square of $\psi(.)$ integrates to unity:

$$\int_{-\infty}^{\infty} \psi^2(u)du = 1. \tag{10.2}$$

The oldest and simplest wavelet which satisfies (10.1) and (10.2) is called the Haar wavelet, and is given by

$$\psi^H(u) \equiv \begin{cases} \frac{-1}{\sqrt{2}}, & -1 < u \leq 0 \\ \frac{1}{\sqrt{2}}, & 0 < u \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

This function is a basis for other wavelets by means of two operations, dyadic dilation and integer translation, producing an orthonormal basis for $L_2(R)$:

$$\psi^H_{j,k}(u) = 2^{j/2}\psi(2^j u - k),\ j,k \in Z = \{0, \pm 1, \pm 2, \ldots\},$$

where j denotes the dilation index and k represents the translation index. In this section we describe the Haar wavelet and other wavelets in discrete terms.

Daubechies (1992) derives a class of wavelets defined by two filters of positive integer width L:

- The high-pass filter (wavelet filter): $\{h_l\} = \{h_0, \ldots, h_{L-1}\}$.
- The low-pass filter (scaling filter): $\{g_l\} = \{g_0, \ldots, g_{L-1}\}$ which is defined via the quadrature mirror relationship $\{h_l = (-1)^l g_{L-1-l}; l = 0, \ldots, L-1\}$.

Fundamental properties of the continuous wavelet functions, such as integration to zero and unit energy in (10.1) and (10.2), respectively, have discrete counterparts. A discrete wavelet filter must satisfy the following three properties:

$$\sum_{l=0}^{L-1} h_l = 0;$$

$$\sum_{l=0}^{L-1} h_l^2 = 1;$$

and

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0,$$

for all nonzero integers n, and where we define $h_l = 0$ for $l < 0$ and $l \geq L$ so that $\{h_l\}$ is an infinite sequence with at most L nonzero values. This means that a wavelet filter must sum to zero, must have unit energy and must be orthogonal to its even shifts.

The transfer functions for $\{h_l\}$ and $\{g_l\}$ are given by

$$H(f) \equiv \sum_{l=0}^{L-1} h_l e^{-i2\pi f l} \tag{10.3}$$

and

$$G(f) \equiv \sum_{l=0}^{L-1} g_l e^{-i2\pi f l}, \tag{10.4}$$

where f is the Fourier frequencies. Further, useful functions are the squared gain function for $\{h_l\}$ and $\{g_l\}$:

$$\mathcal{H}(f) \equiv |H(f)|^2 \text{ and } \mathcal{G}(f) \equiv |G(f)|^2.$$

The Daubechies wavelets are most easily defined through the squared gain function of their scaling filter:

$$\mathcal{G}^{(D)}(f) \equiv 2\cos^{L}(\pi f)\sum_{l=0}^{\frac{L}{2}-1}\binom{\frac{L}{2}-1+l}{l}\sin^{2l}(\pi f),$$

where L is a positive even integer. By using the relationship $\mathcal{H}^{(D)}(f) = \mathcal{G}^{(D)}(f+\frac{1}{2})$, we find that the corresponding Daubechies wavelet filters have squared gain function satisfying

$$\mathcal{H}^{(D)}(f) \equiv 2\sin^{L}(\pi f)\sum_{l=0}^{\frac{L}{2}-1}\binom{\frac{L}{2}-1+l}{l}\cos^{2l}(\pi f).$$

Daubechies (1992) introduces two types of wavelets, the extremal phase D(L) and the least asymmetric LA(L). The difference between them lies only in their phase functions, i.e., $\theta^{(G)}(.)$ in the polar representation

$$G(f) = [\mathcal{G}^{(D)}(f)]^{1/2}e^{i\theta^{(G)}}(f).$$

The Haar wavelet or D(2), is a filter of width $L = 2$, that can be defined through its wavelet filter,

$$h_0 = \frac{1}{\sqrt{2}} \text{ and } h_1 = \frac{-1}{\sqrt{2}},$$

or, equivalently, through its scaling filter,

$$g_0 = g_1 = \frac{1}{\sqrt{2}}.$$

The wavelet filter coefficients for the D(4) wavelet, at unit scale, are defined by

$$h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}},\ h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}},\ h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}} \text{ and } h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}.$$

The scaling coefficients for the LA(8) and other Daubechies wavelets are given in Daubechies (1992). The wavelet filter $\{h_l\}$ approximates an ideal high-pass filter, and the scaling filter $\{g_l\}$ approximates an ideal low-pass filter. The accuracy of the approximation increases as L increases, so that the squared gain functions for $\{h_l\}$ and $\{g_l\}$ converge to the squared functions for an ideal high-pass and an ideal low-pass, respectively.

2.1 The discrete wavelet transform

The key idea of the discrete wavelet transform (DWT) is to decompose a data set orthogonally into different new data sets. In this section we are going to introduce the one-dimensional DWT (1D DWT) and two-dimensional DWT (2D DWT).

2.1.1 The 1D DWT Let $\mathbf{X}=(X_0, \ldots, X_{N-1})$ be a data vector of length N, where we assume that N is an integer divisible by 2^J, where J is a positive integer. The wavelet and scaling filters are used in parallel to define the DWT, i.e., we have two types of coefficients in the DWT based on these two types of filter:

- The scaling coefficients which represent the smoothed version of the original data. These coefficients can help us in detecting the number of clusters.
- The wavelet coefficients.

The 1D DWT is calculated using Mallat's algorithm, introduced by Mallat (1989), which uses linear filtering operations. The transform coefficients, $V_{j,k}$ and $W_{j,k}$, at different scales, are calculated using the following convolution-like expressions. There are $J - 1$ subsequent stages of the pyramid algorithm. The scaling coefficients for level j ($j, \ldots, J$) are given by

$$V_{j,k} = \sum_{l=0}^{L-1} g_l V_{j-1,2k+1-l \bmod N_{j-1}} \text{ for } k = 0,\ldots,N_j - 1, \tag{10.5}$$

and the wavelet coefficients for level j are given by

$$W_{j,k} = \sum_{l=0}^{L-1} h_l V_{j-1,2k+1-l \bmod N_{j-1}} \text{ for } k = 0,\ldots,N_j - 1, \tag{10.6}$$

where $\mathbf{V}_0 \equiv \mathbf{X}$ and $N_j \equiv N2{-}j$. The modulus operator in (10.5) and (10.6) is required in order to deal with the boundary of a finite length vector of observations. This operator circularly filters the data, by using a fast filtering algorithm of order $O(N)$. We see from (10.5) and (10.6) that at each step we filter the previous level scaling coefficients using either the scaling or wavelet filter, and then subsample the resulting sequence. The DWT can be defined also by matrix calculation. Let $\mathbf{W}_j \equiv [W_{j,0}, W_{j,1}, \ldots, W_{j,N_j-1}]^T, j = 1, 2, \ldots, J$ and $\mathbf{V}_J \equiv [V_{J,0}, V_{J,1}, \ldots, V_{J,N_{J-1}}]^T$. The elements of the subvectors $\mathbf{W}_j$ correspond to those in (10.6) and the subvector $\mathbf{V}_j$ correspond to those in (10.5). We then have the analysis equation $\mathbf{W} = \mathcal{W}\mathbf{X}$, where $\mathbf{W}$ contains the DWT coefficients, i.e.,

$$\mathbf{W} = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_j \\ \vdots \\ W_J \\ V_J \end{pmatrix}, \tag{10.7}$$

and $\mathcal{W}$ is an orthonormal $N \times N$ real-valued matrix whose rows depend on the wavelet filter h_l, i.e., $\mathcal{W}^{-1} = \mathcal{W}^T$, so $\mathcal{W}^T\mathcal{W} = \mathcal{W}\mathcal{W}^T = \mathbf{I}_N$ (Percival & Walden, 2000, Ch. 4). A partial DWT will be obtained by stopping the algorithm after $j_0 < J$ repetitions. The partial DWTs are more commonly used in practice than the full DWT, owing to the flexibility they offer in specifying a scale beyond which a wavelet analysis into individual large scales is no longer of real interest.

2.1.2 The 2D DWT Similarly, the 2D DWT is computed using 2D filtering and downsampling operations. Let $\mathbf{Y}_{\mathbf{x,z}}$ be an image with two dimension. The image will be transformed through the two stages of analysis filters h and g and subsampled by two. Analogous to the 1D DWT, the 2D DWT is decomposed into a sum of fine to coarse resolution detail coefficients and sum of coarse resolution smooth coefficients. The coefficients for the first scale are given by the following formulas:

$$W^{(d)}_{z,x,1} = \sum_{k=0}^{L-1}\sum_{l=0}^{L-1} h^0 Y_{2z+1-k \bmod M, 2x+1-l \bmod N} \tag{10.8}$$

$$W^{(v)}_{z,x,1} = \sum_{k=0}^{L-1}\sum_{l=0}^{L-1} h^1 Y_{2z+1-k \bmod M, 2x+1-l \bmod N} \tag{10.9}$$

$$W^{(h)}_{z,x,1} = \sum_{k=0}^{L-1}\sum_{l=0}^{L-1} h^2 Y_{2z+1-k \bmod M, 2x+1-l \bmod N} \tag{10.10}$$

$$V_{z,x,1} = \sum_{k=0}^{L-1}\sum_{l=0}^{L-1} h^3 Y_{2z+1-k \bmod M, 2x+1-l \bmod N}, \tag{10.11}$$

where the filters h^0, h^1, h^2 and h^3 are formed by vector outer products of the quadrature mirror filters h and g.

$$h^0 = h.h^T,\ h^1 = h.g^T,\ h^2 = g.h^T,\ h^3 = g.g^T \tag{10.12}$$

where $W^{(d)}_{z,x,1}$, $W^{(v)}_{z,x,1}$, $W^{(v)}_{z,x,1}$ and $V_{z,x,1}$ are coefficients from the diagonal, vertical, horizontal and scaling.

We can obtain the second scale of coefficients by replacing $Y_{2z,2x}$ with $V_{z,x,1}$ in (10.8)–(10.11). In summary, the 2D DWT consist of four sub images:

- $W^{(h)}_{z,x,1}$ is associated with horizontal features in the image.
- $W^{(v)}_{z,x,1}$ is associated with vertical features in the image.
- $W^{(d)}_{z,x,1}$ is associated with diagonal features in the image.
- $V_{z,x,1}$ is the smoothed version of the original image.

More details can be found in Mallat (1989). A major goal in this chapter is to study the smoothed version of the original image to get an idea about the number of clusters.

2.2 Maximal overlap DWT (MODWT)

An alternative wavelet estimation is achieved by using the maximal overlap DWT (MODWT): see Percival and Walden (2000). The MODWT is also called the undecimated or translation invariant or shift invariant DWT: see Nason and Silverman (1995). The MODWT either for one dimensional or two dimensional data is computed by the same filtering steps as the ordinary 1D DWT or 2D DWT, respectively, but without subsampling the filtered output. The MODWT gives up orthogonality in order to gain features which the DWT does not possess.

2.3 Multi-resolution analysis

The concept of multi-resolution analysis (MRA) was first introduced by Mallat (1989). The multi-resolution analysis of the data leads to a better understanding of wavelets. The idea behind multi-resolution analysis is to express the right-hand side of (10.7) as the sum of several new data sets, each of which is related to variations in the data at a certain scale. Now, since the matrix is orthonormal, we can reconstruct our data sets from the wavelet coefficients **W** by using $\mathbf{X} = \mathcal{W}^T\mathbf{W}$.

We partition the columns of **W** commensurate with the partitioning of **W** to obtain

$$\mathcal{W}^T = [\mathcal{W}_1\mathcal{W}_2 \ldots \mathcal{W}_j \ldots \mathcal{W}_J\mathcal{V}_J],$$

where $\mathcal{W}_j$ is an $N \times N/2^j$ matrix and $\mathcal{V}_J$ is an $N \times N/2^J$ matrix. Thus, we can define the multi-resolution analysis of a data set by expressing $\mathcal{W}^T\mathbf{W}$ as a sum of several new data sets, each of which is related to variations in X at a certain scale:

$$\mathbf{X} = \mathcal{W}^T\mathbf{W} = \sum_{j=1}^{J}\mathcal{W}_j\mathbf{W}_j + \mathcal{V}_J\mathbf{V}_J = \sum_{j=1}^{J}D_j + S_J. \tag{10.13}$$

The terms in (10.13) constitute a decomposition of X into orthogonal data sets components D_j (detail) and S_J (smooth) at different scales and the length of D_j and S_J coincides with the length of **X** ($\mathrm{N}\times$vector). Because the terms at different scales represent components of **X** at different resolutions the approximation is called 'multi-resolution decomposition' (Percival and Walden, 2000). The smooth scale S_J gives a smooth approximation to **X**. Adding the detail scale, D_j, yields S_{J-1}, a scale 2^{J-1} approximation to **X**. The S_{J-1} approximation is a refinement of the S_J approximation. Similarly, we can refine further to obtain the scale 2^{J-1} approximations. The collection S_J, S_{J-1} and S_1 provides a set of multi-resolution approximations of X. Analogous to kernel smoothing, which has a parameter called bandwidth or smoothing parameter, the index J is called a wavelet smoothing parameter. Increasing the smoothing parameter J allows less detail in the smooth approximation of **X**, while a small J allows additional detail in the smooth approximation of **X** (see Ogden, 1997). The principle of the one-dimensional MRA can be extended to two or several dimensional MRA.

2.4 Wavelet algorithm

The DWT decomposes the data into two parts, the high frequency part and the low frequency part. The high frequency part consists of the detail coefficients and represents the boundaries of clusters, where the low frequency part consists of the smooth coefficients and represents the clusters. The goal of DWT-based clustering (DWTBC) is to detect different smooth versions of the data at different scales by removing the background noise in the wavelet coefficients. By reconstructing the DWTBC at different scales, we get a multi-resolution approximation of our data set that represents the original data set at different levels of smoothing. We can summarize the wavelet clustering algorithm as follows:

- Apply the DWT of the data by starting with $j = 1$.
- Increase the value of j until getting a clear picture of the number of clusters.
- Reconstruct the smooth coefficients at suitable j.

The same algorithm can also be used by applying MODWT instead of the DWT. The same algorithm could be generalized for two-dimensional data.

3 Examples

In this section we give some examples about using the DWTBC and the multi-resolution approximation-based clustering (MRABC). Figure 10.1a shows two simulated normal distributed clusters with means equal to 3 and 5, and with standard deviations equal to 0.40 and 0.40, respectively; and with sample size equal to 1024 observations each. Figure 10.1b shows the two clusters with additive normal noise with mean equal to 4 and standard deviation equal to 2. We apply the D(2) wavelet in our wavelet algorithm to the simulated data sets in order to detect the number and quality of existing clusters. Figures 10.1c–10.1h show the first, second and third level of the DWT and MRA. These results show a clear recovery of the two clusters.

In Figure 10.2 we present the last example with larger sample size (4096 observations). We increase the number of clusters from two to four to check the ability of DWT-based

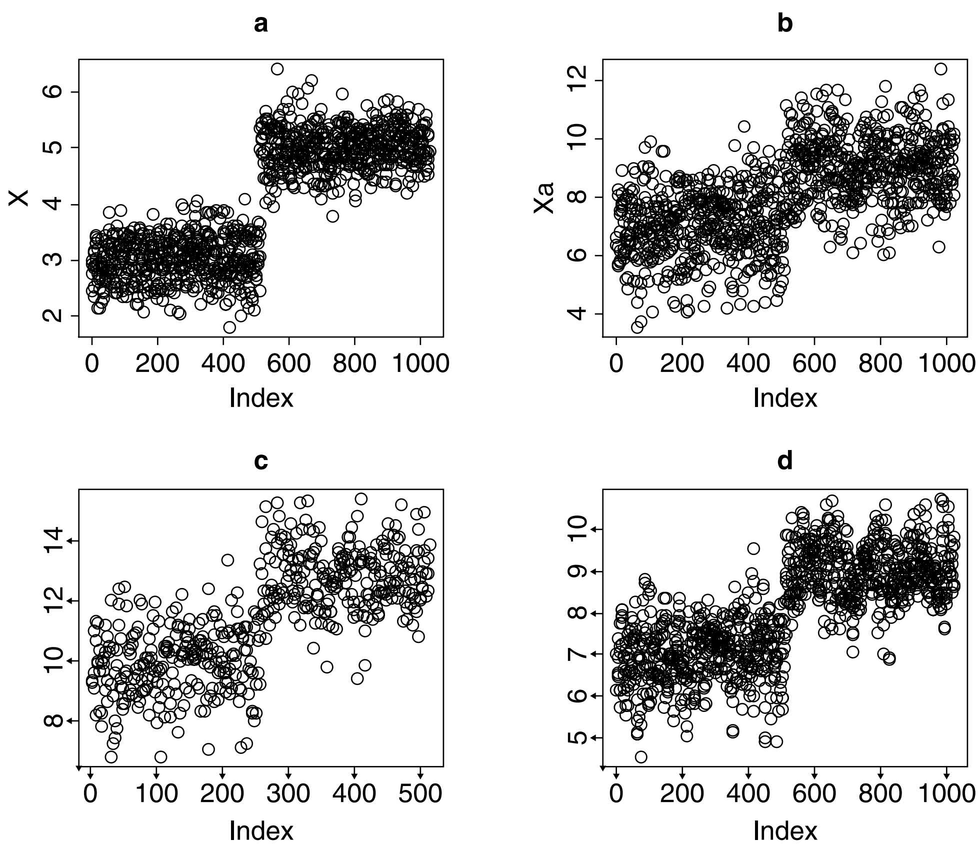

Figure 10.1a–d (a) data set with 2 clusters (1024 obs.), (b) the data set with normal noise, (c) first scale DWTBC, (d) first scale MRABC

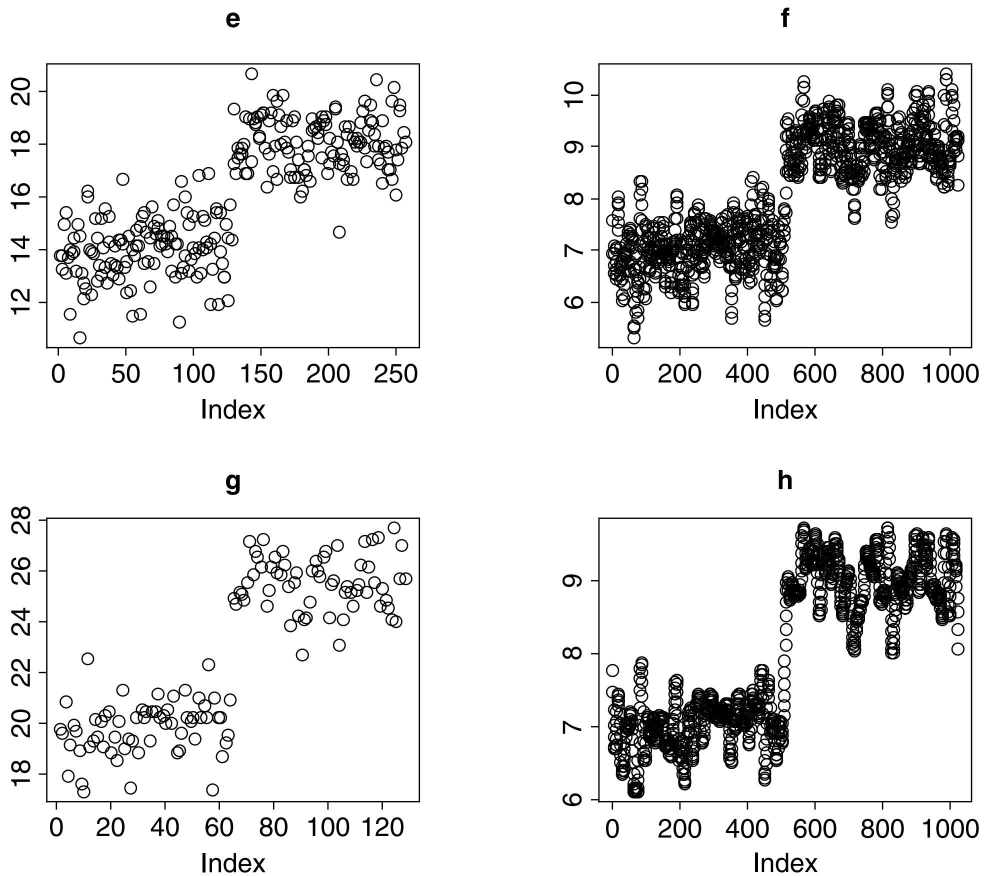

Figure 10.1e–h (e) second scale DWTBC, (f) second scale MRABC, (g) third scale DWTBC, (h) third scale MRABC

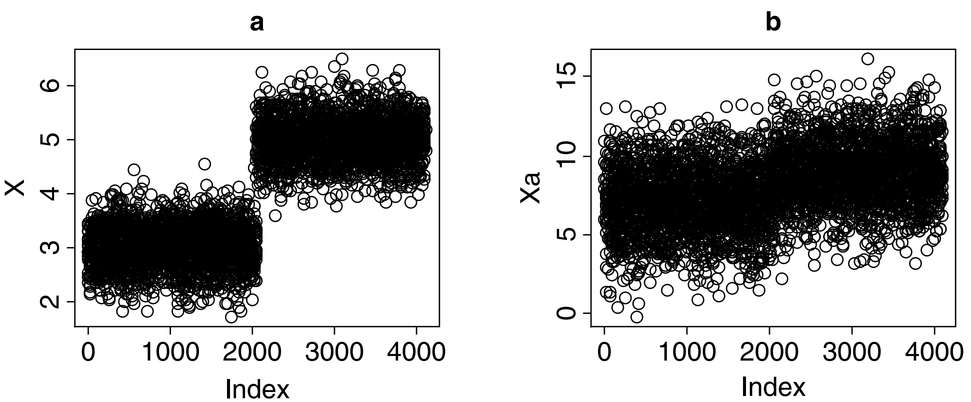

Figure 10.2a–b (a) data set with 2 clusters (4096 obs.), (b) the data set with normal noise

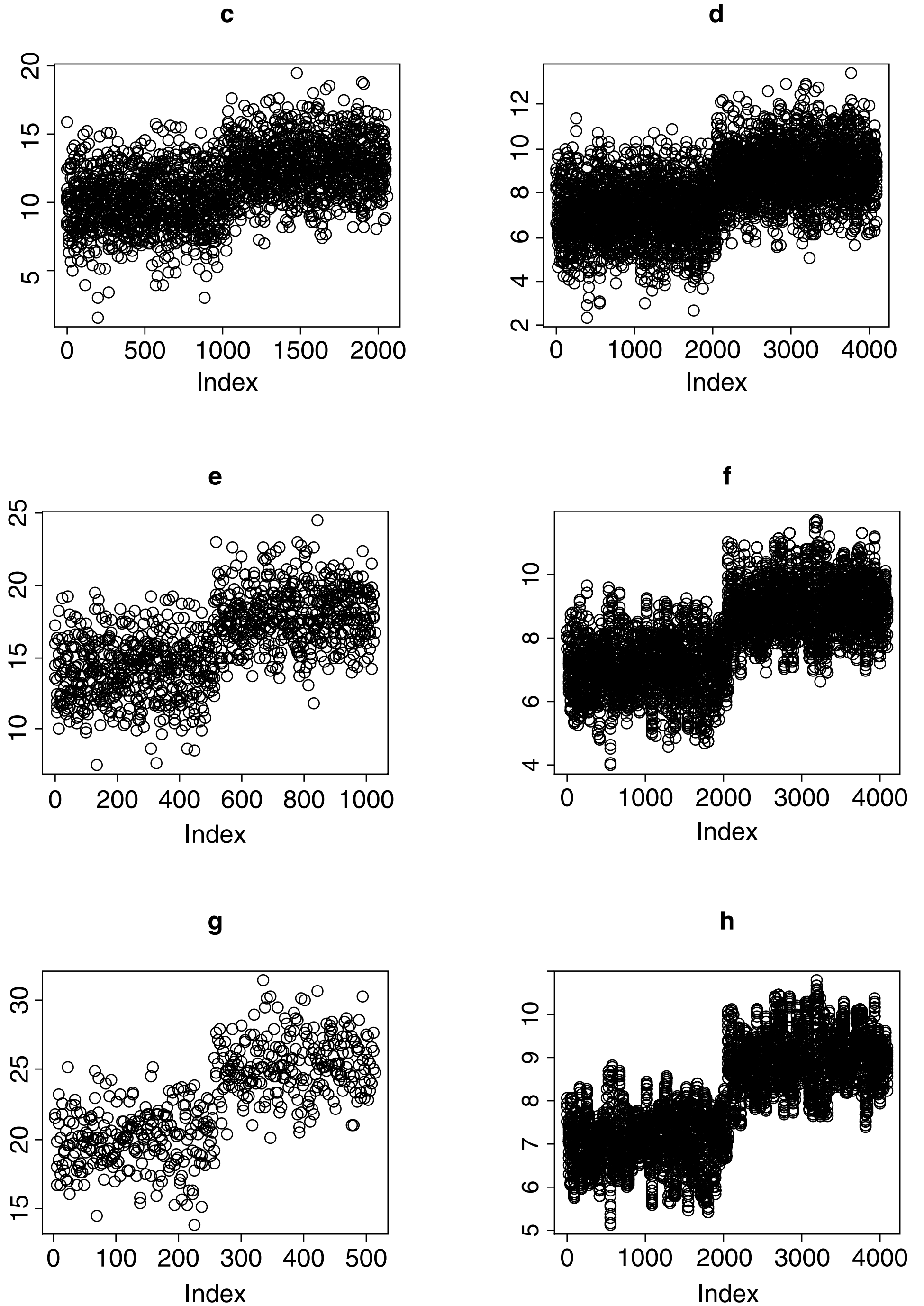

Figure 10.2c–h (*c*) *first scale DWTBC,* (*d*) *first scale MRABC* (*e*) *second scale DWTBC,* (*f*) *second scale MRABC,* (*g*) *third scale DWTBC,* (*h*) *third scale MRABC*

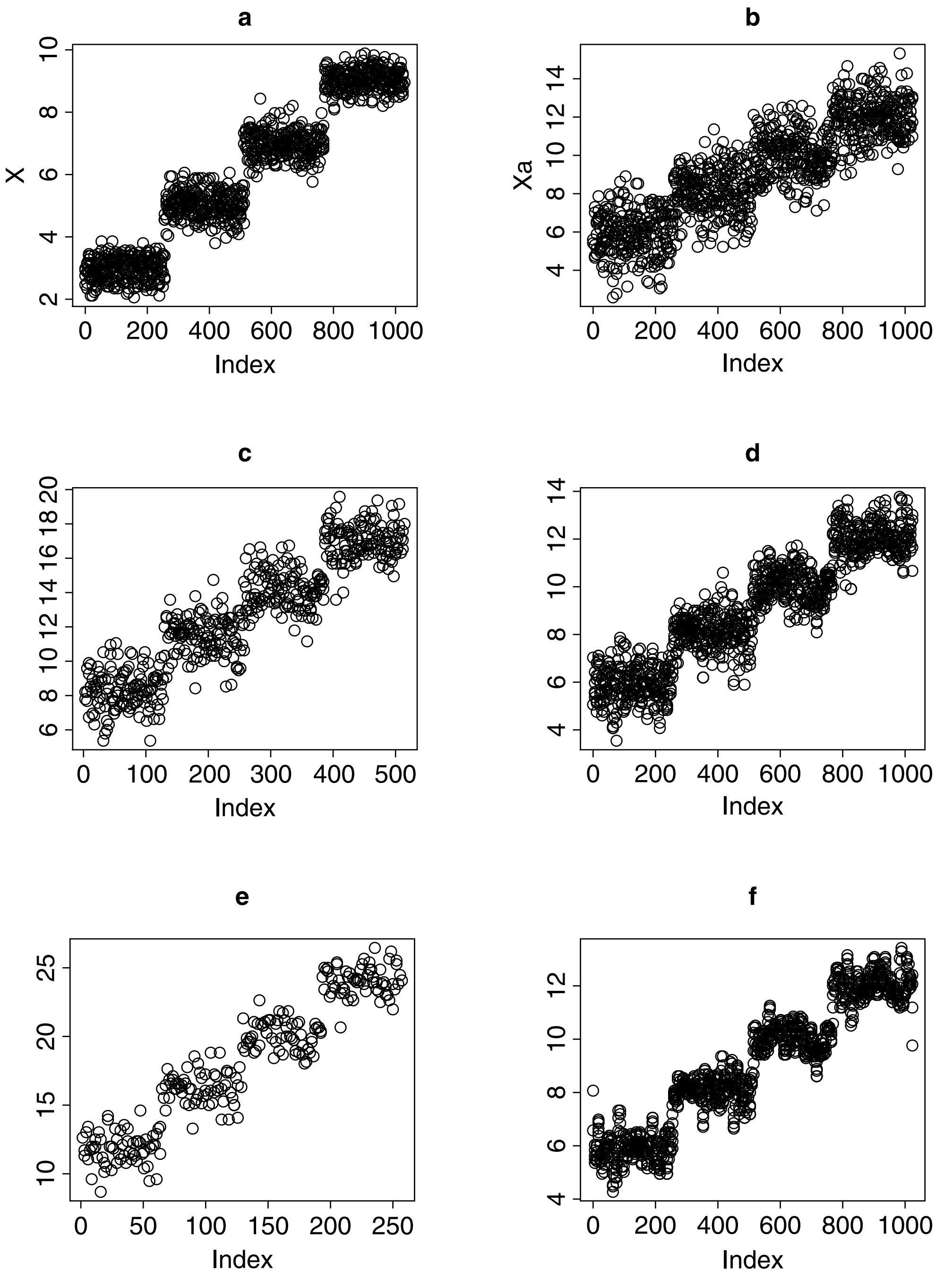

Figure 10.3a–f *(a) data set with 4 clusters (1024 obs.), (b) the data set with normal noise, (c) first scale DWTBC, (d) first scale MRABC, (e) second scale DWTBC, (f) second scale MRABC*

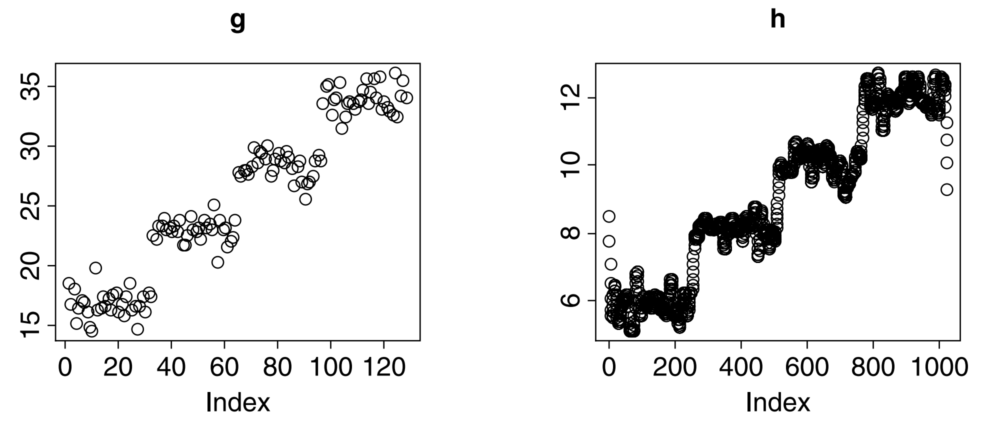

Figure 10.3g–h (g) third scale DWTBC, (h) third scale MRABC

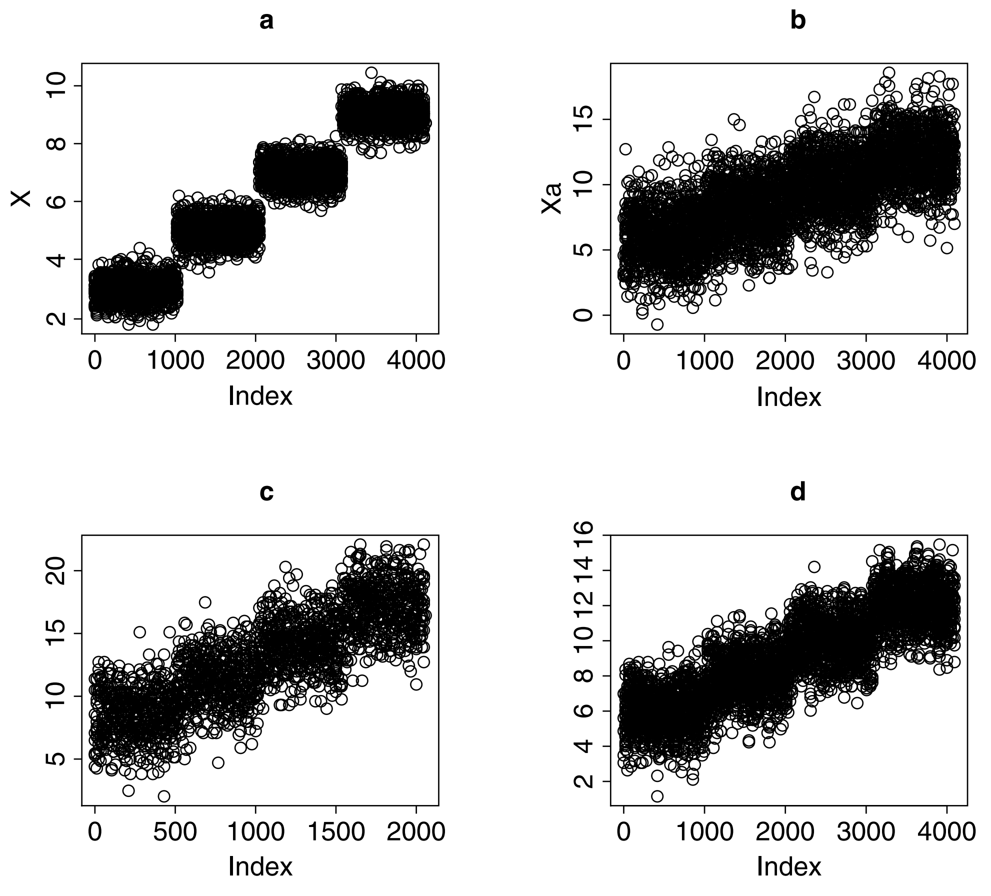

Figure 10.4a–d (a) data set with 4 clusters (4096 obs.), (b) the data set with normal noise, (c) first scale DWTBC, (d) first scale MRABC

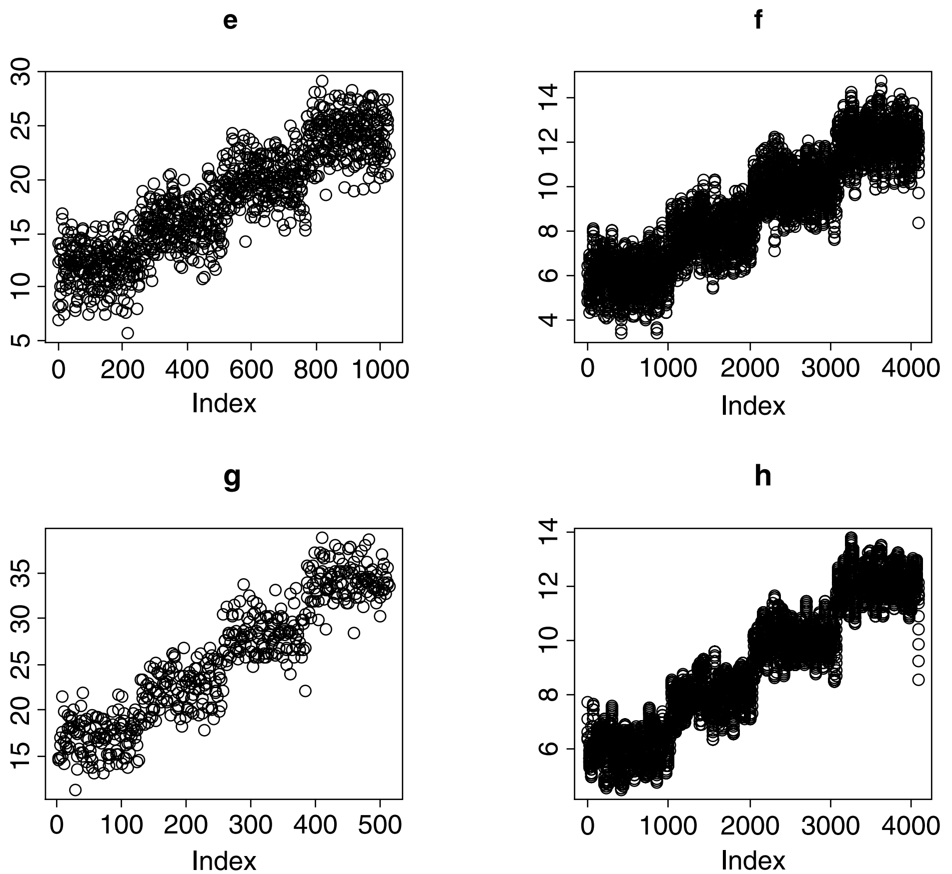

Figure 10.4e–h (e) second scale DWTBC, (f) second scale MRABC, (g) third scale DWTBC, (h) third scale MRABC

clustering in detecting them. Figures 10.3 and 10.4 show the data and the results for sample sizes equal to 1024 and 4096 observations, respectively.

In addition to the normal noise, we also imposed a poisson and lognormal noises to check the robustness of the clustering process. The poisson noise is with mean equal to 6 and the lognormal noise whose logarithm has mean equal to 0 and standard deviation equal to 1. The DWT clustering results at scale three for these two noises (added to the data sets in Figure 10.1a) are shown in Figure 10.5. In the poisson noise case, the results have been shown to be very similar to the normal noise. In the lognormal case, however, we might need some handling like using wavelet shrinkage method: see Donoho and Johnstone (1994).

Moreover, we apply the same algorithm to two-dimensional data by using two-dimensional DWT and MRA. This example consists of two dimensional images (1024 $\times$ 2) that have two columns; each column has two clusters, and they have the normal distribution. The image has two clusters which can be shown in Figure 10.6a. In Figure 10.6b

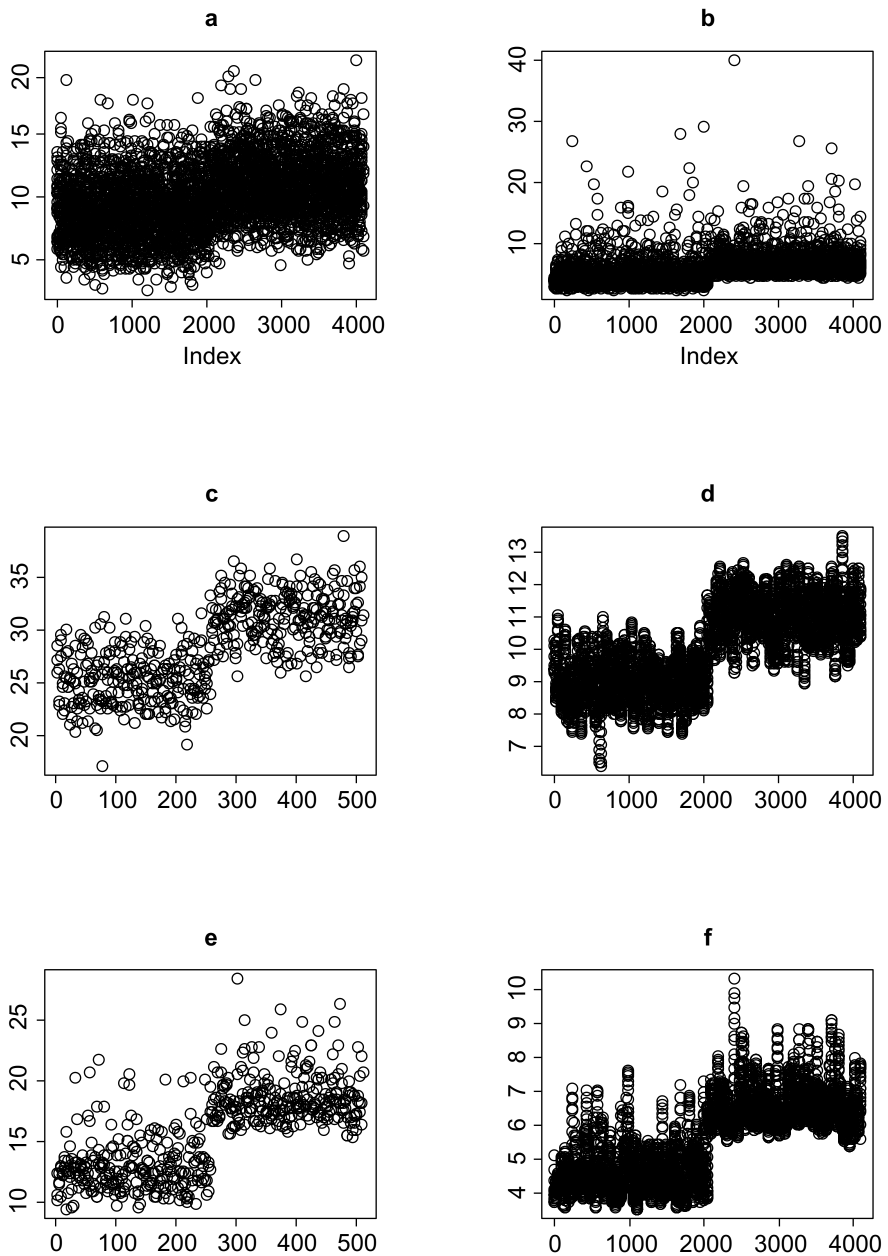

Figure 10.5 (a) the data set in Figure 10.1a with poisson noise, (b) the data set with lognormal noise, (c) third scale DWTBC of a, (d) third scale MRABC of a, (e) third scale DWTBC of b, (f) third scale MRABC of b

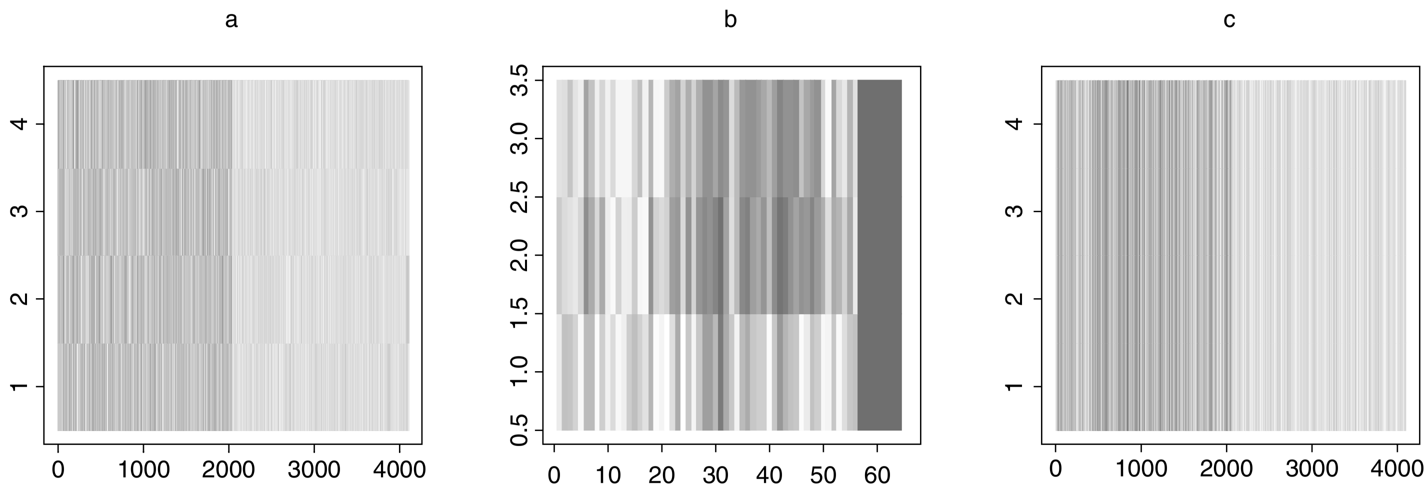

Figure 10.6 *(a) 2 dimensional data set with 2 clusters, (b) the data set with normal noise, (c) first scale 2MRABC*

we have the original image with additive normal noise with means (3, 5); and with standard deviations (0.40, 0.40), respectively. From Figure 10.6c, we can see that the wavelet algorithm can easily detect the two clusters in this two-dimensional data set.

From our results, we can conclude that the wavelet algorithm can successfully and effectively reveal the clusters in the data both in the case of one-dimensional data sets (with different sample sizes, different numbers of clusters and different noises) and the two-dimensional data sets.

4 Concluding remarks

The goal of cluster analysis is to partition the observations into clusters so that the pairwise dissimilarities between those observations assigned to the same cluster tend to be smaller than those in different clusters. Discriminative and model-based clustering are two main categories into which clustering approaches can be classified. Each has its own advantages and disadvantages regarding, e.g., detecting the correct number of clusters, measurement error in the dissimilarities and sensitivity to outliners, requiring the number of clusters before the analysis.

In this chapter we introduce another clustering approach based on the discrete wavelet transform (DWT) which satisfies the requirements that other clustering methods do not satisfy and is more efficient and easy to apply using standard statistical computer packages like R and S+. Moreover, the approach can easily be applied to large data sets and it can also be generalized to be applicable to higher dimension data. Under these conditions other clustering methods are either inapplicable or inefficient.

The wavelet methodology has previously been used in many fields, such as signal processing, image analysis, geophysics, atmospheric sciences, geology, climatology and time series econometrics.

To demonstrate our approach, a number of artificial examples have been conducted under different situations where the sample size, the noise imposed on the data and the number of clusters have been varied. These examples have shown the effectiveness of the DWT clustering approach in detecting the correct number of clusters.

References

Almasri, A. and G. Shukur (2003), 'An illustration of the causality relation between government spending and revenue using wavelets analysis on Finnish Data', *Journal of Applied Statistics*, **30**, 571–84.

Chave, J. and S. Levin (2003), 'Scale and scaling in ecological and economic systems', *Environmental and Resource Economics*, **26**, 527–57.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia, PA: SIAM.

Donoho, D.L. and I.M. Johnstone (1994), 'Ideal spatial adaption via wavelet shrinkage', *Biometrika*, **81**, 425–55.

Fraley, C. and A.E. Raftery (2002), 'Model based clustering, discriminant analysis, and density estimation', *Journal of the American Statistical Association*, **97**(458), 611–31.

Kohonen, T. (2001), *Self Organising Maps*, Heidelberg: Springer Verlag.

MacQueen, J. (1967), 'Some methods for classification and analysis of multivariate observations', in L.M.L. Cam and J. Neyman (eds), *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley, CA: University of California Press, pp. 281–97.

Mallat, S. (1989), 'A theory for multiresolution signal decomposition: the wavelet representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–93.

Morehart, M., F. Murtagh and J. Starck (1999), 'Multiresolution spatial analysis', *Proceedings GeoComp 99*, (http://www.geovista.psu.edu/sites/geocomp 99/Gc99/088/gc_088.htm).

Nason, G.P. and B.W. Silverman (1995), 'The stationary wavelet transform and some statistical applications', in A. Antoniadis and G. Oppenheim (eds), *Lecture Notes in Statistics*, 103, New York: Springer Verlag, pp. 281–300.

Ogden, R.T. (1997), *Essential Wavelets for Statistical Applications and Data Analysis*', Boston, MA: Birkhuser.

Oh, M-S. and A.E. Raftery (2003), 'Model based clustering with dissimilarities: a Bayesian approach', *Technical Report*, no. 441, Department of Statistics, University of Washington.
Percival, D.B. and A.T. Walden (2000), *Wavelet Methods for Time Series Analysis*, Cambridge, UK: Cambridge University Press.
Ramsey, J.B. and C. Lampart (1998), 'Decomposition of economic relationships by timescale using wavelets, money and income', *Macroeconomic Dynamics*, **2**, 49–71.
Tryon, R. (1939), *Cluster Analysis*, Ann Arbor, MI: Edwards Brothers.
Vu, T.T. and M. Tokunaga (2001), 'Wavelet and scale–space theory in segmentation of airborne laser scanner data', *Proceedings of the 22nd Asian Conference on Remote Sensing*,176–80, November, Singapore.