Robust Principal Components for Hyperspectral Data Analysis

María M. Lucini[1,2,⋆] and Alejandro C. Frery[3, ⋆⋆]

[1] CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina)
[2] FaCENA, Universidad Nacional del Nordeste, Av. Libertad 5460
3400 Corrientes, Argentina
magdalenalucini@gmail.com
[3] Instituto de Computação, CPMAT & LCCV, Universidade Federal de Alagoas
BR 104 Norte km 97, 57072-970 Maceió, AL – Brazil
acfrery@pesquisador.cnpq.br

Abstract. Remote sensing data present peculiar features and characteristics that may make their statistical processing and analysis a difficult task. Among them, it can be mentioned the volume of data involved, the redundancy, the presence of unexpected values that arise mainly due to noisy pixels and background objects whose responses to the sensor are very different from those of their neighbours. Sometimes, the volume of data and number of variables involved is so large that any statistical analysis becomes unmanageable if data are not condensed in some way. A commonly used method to deal with this situation is Principal Component Analysis (PCA) based on classical statistics: sample mean and covariance matrices. The drawback in using sample covariance or correlation matrices as measures of variability is their high sensitivity to spurious values. In this work we analyse and evaluate the use of some Robust Principal Component techniques and make a comparison of Robust and Classical PCs performances when applied to satellite data provided by the hyperspectral sensor AVIRIS (Airborne Visible/Infrared Imaging Spectrometer). We conclude that some robust approaches are the most reliable and precise when applied as a data reduction technique before performing supervised image classification.

Keywords: principal component analysis, robust inference, image classification.

1 Introduction

As already stated in the abstract, the difficulty in modelling remote sensing data, the volume of data involved, the redundancy and multivariate nature of observations, and the presence of unexpected values [5,24] make the statistical processing and analysis of this kind of data a difficult task, specially when applying classical statistical methods. These methods are based on the validity of

⋆ ANPCyT - PICT 35776.
⋆⋆ Work partially funded by CNPq and FAPEAL.

M. Kamel and A. Campilho (Eds.): ICIAR 2009, LNCS 5627, pp. 126–135, 2009.

certain assumptions that try to theoretically and formally explain conjectures (or even information previously obtained by other results and experiences) that the user makes about the data under study in order to achieve certain optimal (theoretical) properties. The fact is that all those assumptions are seldom satisfied simultaneously, leading to inaccurate and unreliable results.

One of the aims of Robust Statistics is to improve classical statistical methods and techniques by deriving methods that are not very sensitive to atypical observations or slight departures from model assumptions. Since its beginnings, back in the early 70s, quantitative robustness has been an active research area from the theoretical point of view [3,15,16] and the need of advanced computational methods to deal with practical situations and problems was always clear. With the increased speed of computers some robust techniques have been evaluated in different fields showing, most of the times, better performances than those of classical statistical methods.

Although there exists certain awareness of the impact that spurious values and departures from models assumptions might have on the reliability of classical methods, and in spite of the fact that for every classical statistical technique there is at least one robust version, the evaluation and application of robust methods has not been widely spread in some applied fields, among them image analysis and processing. Moreover, commercial platforms commonly used for those tasks only offer classical methods and, in some cases, poorly implemented [7]. Nevertheless, there are works that highlight the need of using robust procedures in some applications, as for instance, parameter estimation of statistical models for remote sensing data [1], [8], [19], image de-noising [4], [14], [23] and dimensionality reduction [10].

In many situations of interest the number of variables involved in a statistical analysis is so large that makes it difficult and unamanegeable when the number of variables is not condensed in some way. Many processing techniques are useless unless data dimensionality is somehow reduced. There are multivariate statistical techniques that can be used to reduce data dimensionality; some of them concentrate on grouping variables in a reduced number of variables based on the statistical properties of data (i.e. cluster analysis), while others aim to reduce data dimensionality by replacing a large number of correlated variables by a smaller number of linear combinations of those variables. Principal Components (PC) analysis is one of these latter methods [17]. It is widely used in many applications, that range from meteorology to remote sensing, where it is commonly used not only as a dimension reduction technique, but also as a visualisation and analysis tool, specially when displaying multispectral images [26]. Chen and Qian [11] proposed a method for hyperspectral image dimensionality reduction and de-noising based on a combination of PC analysis and wavelet denoising. In this work we show how to do that simultaneously by simply choosing a robust procedure.

This paper is organised as follows. In section 2 we present Principal Components in their classical and robust versions. In section 3 we show and application of the proposed methods for dimensionality reduction and image classification

applied to satellite data provided by the hyperspectral sensor AVIRIS (Airborne Visible/Infrared Imaging Spectrometer), whereas in section 4 we present the main conclusions.

2 Principal Components

The main purpose of PC analysis is to reduce data dimensionality by choosing p variables in such a way that the p linear combinations of the original n variables retain as much as possible of the variability present in the data set. The original variables are transformed to a new set of variables, the Principal Components, that happen to be uncorrelated. Furthermore, they can be ordered in such a way that the first p explain most of the variability in the data set. Therefore, PCs represent the directions of maximum variability in the data space.

2.1 Classical Principal Components

In the classical approach [25] data variability is measured by analysing a sample covariance or correlation matrix, and the directions of maximum variability are the first p eigenvectors of the sample covariance (correlation) matrix, when the corresponding eigenvalues are decreasingly ordered.

That is, let $\mathbf{x} = (x_1, \ldots, x_n)^t$, be an n-dimensional random vector, with mean $E(\mathbf{x}) = \mu$ and covariance matrix $\mathbf{\Sigma}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\mathbf{\Sigma}$ decreasingly ordered and let $\mathbf{b}_k$ be the eigenvector corresponding to eigenvalue λ_k, subject to $||\mathbf{b}_k|| = 1$. Then $\mathrm{Var}(\mathbf{b}_k^t\mathbf{x}) = \lambda_k$, $\mathbf{b}_k^t\mathbf{x}$ and $\mathbf{b}_k$ are the k-th principal component and k-th principal direction, respectively. Moreover, $\mathbf{b}_j^t\mathbf{b}_k = 0$ for any $j \neq k$.

Given a data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ the sample principal components and directions are computed by replacing μ and $\mathbf{\Sigma}$ by the sample mean and sample covariance matrix, respectively.

Depending on the final application the user can choose the number of components that he or she wants to work with. For instance, for image visualisation the user normally chooses the first three principal components and displays them in R (red), G (green) and B (blue) image bands, whereas in some other applications the user needs to achieve a certain "proportion of explained variance". This "proportion of explained variance" for the, let us say, first k principal components is given by

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \quad . \tag{1}$$

One of the drawbacks of using the sample covariance matrix as a measure of data variability is its high sensitivity to outliers or spurious values. If those values are influencing observations they might have distorting effects on the resulting directions and, if not aware of that, lead to results dominated by those observations.

2.2 Robust Principal Components

As already mentioned, for each classical statistical technique there is, at least, one robust version. Methods to find robust PCs have been proposed in the last 25 years. The "natural" robust version of classical PC would be obtained replacing the covariance matrix by a robust scatter matrix, and then proceed as in the "classical" version by finding the eigenvalues and eigenvectors of this matrix. This has been done in, among others, references [9,12,13,22]

Other type of approach, called "projection pursuit", aims to find the directions of maximum (minimum) variability by maximising (minimising) a robust dispersion estimate (that is not the variance). Maronna [20] extended this idea by proposing a technique to find a p-linear manifold that minimises a robust scale of the orthogonal distances of the data to that manifold.

That is, if $\mathbf{x}_i$, $i = 1, \ldots, m$ is an n-dimensional data set given $p < n$, the goal is to find a p-linear manifold with $\mathbf{Bx} = \mathbf{a}$, where $\mathbf{B}$ is a $q \times n$ matrix and $\mathbf{a}$ a q-vector ($q = n - p$), that minimises the function $\sigma(r(\mathbf{B}, \mathbf{a}))$, where σ is a robust scale of the orthogonal distances

$$r_i(\mathbf{B}, \mathbf{a}) = \|\mathbf{B}\mathbf{x}_i - \mathbf{a}\|^2 \tag{2}$$

of the data to the manifold, with the condition that $\mathbf{B}$ also satisfies $\mathbf{B}\mathbf{B}' = I_q$ (further details in [20,21]).

A simple and fast approach, called *Spherical Principal Components* (SPC), that works very well when dealing with large datasets was proposed by Locantore et al. [18].

Let $\mathbf{x}_i$, $i = 1, \ldots m$ be a n-dimensional data set, and let $\hat{\mu}$ be a robust multivariate location estimate, for instance:

j

$$\hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{m} ||\mathbf{x}_i - \mu|| \ , \tag{3}$$

and let

$$\mathbf{y}_i = \begin{cases} (\mathbf{x}_i - \hat{\mu})/||\mathbf{x}_i - \hat{\mu}|| & \text{if } \mathbf{x}_i \neq \hat{\mu},\ i = 1, \ldots, m \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Let $\mathbf{b}_i, i = 1, \ldots, m$ be the eigenvectors of the covariance matrix of the $\mathbf{y}_i$ and let $\hat{\lambda}_i = \hat{\sigma}(\mathbf{x}^t \mathbf{b}_i)^2$ where σ is a robust dispersion estimate (for instance the median absolute deviation $\text{MAD}(\mathbf{x}) = \text{median}\{|\mathbf{x} - \text{median}(\mathbf{x})\}$). Let $\hat{\lambda}_{(1)} \geq \cdots \geq \hat{\lambda}_{(n)}$ and $\mathbf{b}_{(i)}$ be the corresponding eigenvectors. Boente and Freiman [6] showed that $\mathbf{b}_{(i)}$ coincide with the eigenvectors of the covariance matrix of the $\mathbf{x}_{(i)}$ and the $\hat{\lambda}_{(i)}$ are proportional to its eigenvalues. Moreover, the "proportion of explained variance" can be computed replacing λ_i by $\lambda_{(i)}$ in (1).

In the following section we show results obtained by applying classical and some robust PCs to hyperspectral satellite image analysis.

3 Results

In this section we evaluate the use of some Robust and Classical Principal Components techniques and make a comparison of their performances when applied to the analysis of satellite data provided by the hyperspectral sensor AVIRIS (Airborne Visible/Infrared Imaging Spectrometer).

This sensor provides up to 224 spectral channels with spectral bands ranging from 0.4 μm to 2.4μm in wavelength, and it is aimed at identifying, measuring, and monitoring constituents of the Earth's surface and atmosphere based on molecular absorption and particle scattering signatures. The spectral bandwidth is approximately 0.01μm, overlapping and redundancy in contiguous bands is not rare.

The data chosen for the examples shown here are from Moffet Field reflectance images, available online from [28]. Before performing any statistical analysis we removed all bands that contained no data. Some other bands were also removed based on the following criteria:

Case 1: Clean data – Remove all bands that contained at least one value larger than a threshold set to 30000.
Case 2: 1% Contamination – Remove all bands that contained more than 1% of pixels with values larger than the threshold.
Case 3: 5% Contamination – Remove all bands that contained more than 5% of pixels with values larger than the threshold.

For each case we computed classical and robust PCs. For the latter we used the following approaches, already described in the previous section:

1. Replace the covariance matrix by a robust scatter matrix: We used the Minimum Volume Ellipsoid (MVE) and Minimum Covariance Determinant (MCD) [21],
2. Spherical Principal Components (SPC),
3. the method proposed by Maronna [20] (PCM).

All the codes and computing were performed in the `R` platform [27], whose excellent numerical properties have been reported by [2]. The approach based on replacing the covariance matrix by the MVE estimate did not converge for different subsets of the original data. When replacing it by the MCD estimate we obtained results, but after several hours of processing in a 2.40 GHz Core 2 Duo processor with 4 MB of RAM. The example we show here was chosen because all the proposed methods converged. The results obtained with other samples by computing the different versions of PCs (except MVE) had similar performances to the ones presented here.

The example shown in this work is an area of 117×171 pixels from Moffet Field reflectance images. We used 174 bands for Case 1, 199 for Case 2 and 200 for Case 3.

For each case we kept the three first PCs from each method. These components were assigned to the R, G and B channels (first, second and third principal components, respectively) for visualisation purposes.

Five different classes were identified, each representing a different type of ground cover. A careful exploratory analysis was performed in order to assess the quality and separability of the training samples. We performed Maximum Likelihood Gaussian classification, using ENVI 4.5, on the original data and on the transformed datasets using both the same number of classes and the same training and test samples.

The original sample and graphical display of the classification obtained using clean data (Case 1) by each method, along with its Overall Accuracy (OA) and κ coefficients of agreement are shown in Fig. 1. The proportion of variance explained by the three first principal components was about 0.98 for each method used.

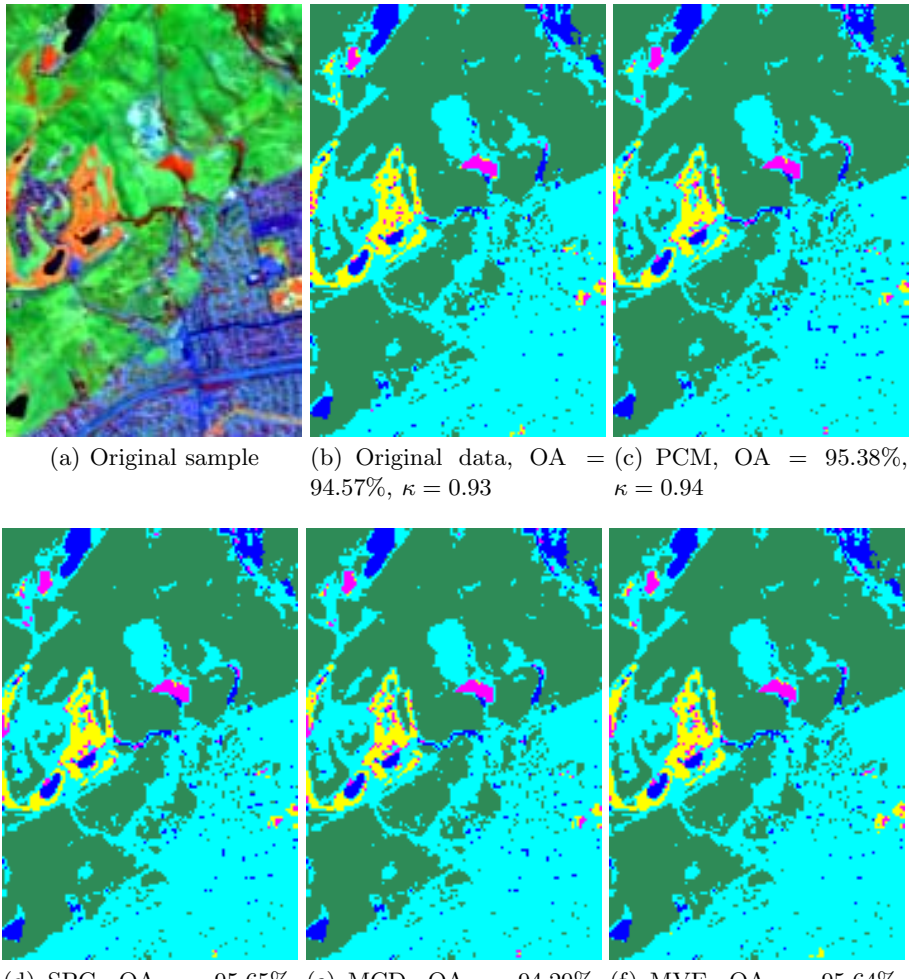
As can be seen from the results shown in Fig. 1, robust methods perform as well as, or even better, than the classical approach regarding visualization, overall accuracy and kappa values. The SPC and PCM approaches were as fast to compute as the classical version, having a better performance. On the other hand, the MCD and MVE approaches were extremely slow, as it was expected, and their use in the other two cases was discarded.

We followed the same steps as before, but using data with 1% (Case 2) and with 5% (Case 3) of contamination. By doing this we evaluate the performance of the different methods when huge unexpected values are present in the data, values whose occurrence is not rare in AVIRIS images. The effect of these large values affects the dispersion measures, as well as the location parameters involved in the computation of the PCs. As before, we kept the three first PCs from each method to compose the RGB image before performing the maximum likelihood gaussian classification; the number of classes as well as training and test samples are also the same as before.

Results of the classification for Case 2 are shown in Table 1, whereas Fig. 2 shows the classified images, OA and κ values for Case 3. The residual improvement (RI) of the accuracy with respect to the worst case, classification with classical principal components in this case, is also reported.

From Table 1 it can be seen that robust approaches are less sensitive to noisy data than the classical approach, as it was expected. Classical PCs are noticeably affected by this small proportion of spurious values, its OA goes down to 89.40% while the two others, specially the SPC approach, are not that sensitive to a contamination of at most 1% in certain bands. The proportion of variance explained by the first three principal components of each method ranged from 0.61 (Classical) to 0.96 (SPC), being 0.94 for PCM.

Results for Case 3 show the effect of an intense contamination on the results of a classification procedure when a Classical PCA is used in the preprocessing step as a data dimensionality reduction tool. As can be see from Fig. 2, some classes are confused and even undetected when using the Classical approach; the OA and κ values go down to 76.09% and 0.69 respectively, whereas they are slightly affected in the robust approaches. For this case, the residual improvement when using SPC instead of Classical PCs was 65.9%, while the residual improvement



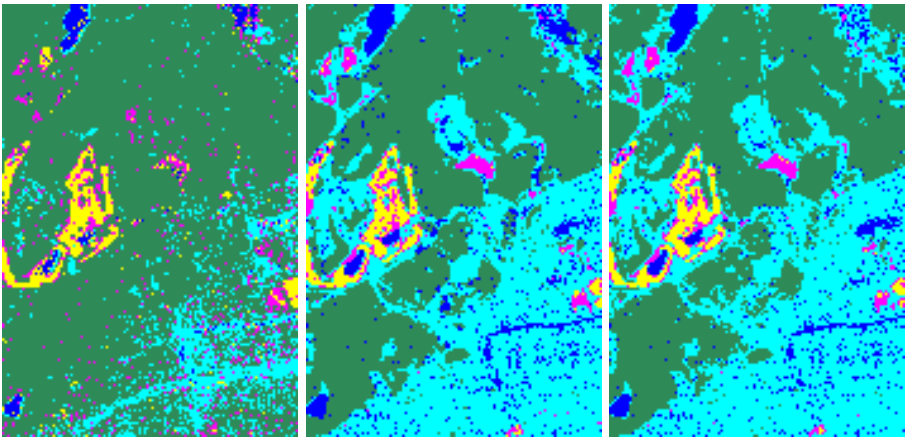
(a) Original sample (b) Original data, OA = 94.57%, $\kappa = 0.93$ (c) PCM, OA = 95.38%, $\kappa = 0.94$

(d) SPC, OA = 95.65%, $\kappa = 0.94$ (e) MCD, OA = 94.29%, $\kappa = 0.93$ (f) MVE, OA = 95.64%, $\kappa = 0.94$

Fig. 1. Original sample and classified images

Table 1. κ coefficient, % of Overall Accuracy (OA) and residual improvement (RI) for each method in presence of less than 1% of spurious values

PC Method	κ	OA%	RI%
Classical	0.87	89.40	
PCM	0.91	92.93	33.3
SPC	0.93	94.56	48.68



(a) Classical PC, OA = 76.09%, $\kappa = 0.69$ (b) PCM, OA = 92.12%, $\kappa = 0.90$ (c) SPC, OA = 91.85%, $\kappa = 0.90$

Fig. 2. Classified images of contaminated data (contamination $\leq 5\%$)

for PCM was 67%. In this case the proportion of variance explained by the three first PCs was 0.73 for the classical version, 0.93 for SPC and 0.91 for PCM.

4 Comments and Conclusions

In this work we propose the use of robust versions of the commonly used "Classical" Principal Components, which are based on classical covariances (correlation) matrices, as a tool for dimensionality reduction in the pre-processing step of image analysis and processing. Covariance (correlation) matrices are higly influenced by the presence of outliers or noisy data as can be seen in the examples shown before.

We evaluated the use of three different robust procedures, one based on the replacement of the covariance matrix for a robust scatter matrix, the second one by an extension of a projection pursuit method and the last one by a procedure called "Spherical Principal Components" on hyperspectral data and compared their performances with the results obtained by a classical Principal Component Analysis.

After some evaluations we discarded the use of the procedure that replaces the covariance matrix by a robust scatter matrix for being too slow and not always convergent with this type of large datasets.

The other robust procedures showed very similar results to those obtained by the classical method when data are completely free of spurious values being, as it was expected, more accurate in presence of a small percentage of those values (noise). It should be remarked that only one band was added from Case 2 to Case 3, this single band has at most 5% of pixels whose values are larger than the threshold used. It should be highlighted the large (bad) effect that a single

band can have on this kind of applications. The classification results obtained when using robust approaches were not as much affected in noisy situations.

As we do not know how "clean" or reliable our data are, we suggest the use of a robust approach when PCA is used as a data dimensionality reduction tool in the pre-processing stage.

References

1. Allende, H., Frery, A.C., Galbiati, J., Pizarro, L.: M-estimators with assymetric influence functions: the GA0 distribution case. Journal of Statistical Computation and Simulation 76(11), 941–956 (2006)
2. Almiron, M.G., Almeida, E.S., Miranda, M.: The reliability of statistical functions in four software packages freely used in numerical computation. Brazilian Journal of Probability and Statistics, Special Issue on Statistical Image and Signal Processing (in press), http://www.imstat.org/bjps
3. Andrews, D., Bickel, P., Hampel, F., Huber, P.K., Rogers, W., Tuckey, J.: Robust Estimates of Location: Survey and Advances. Princeton University Press, New Jersey (1972)
4. Aysal, T.C., Barner, K.E.: Generalized mean-median filtering for robust frequency-selective applications. IEEE Transactions on Signal Processing 55, 937–958 (2007)
5. Barret, H.H., Myers, K.J.: Foundations of Image Science. Pure and Applied Optics. Wiley Interscience, New Jersey (2004)
6. Boente, G., Fraiman, R.: Discussion of "Robust Principal Components for Functional Data" by Locantore et al. Test 8, 28–35 (1999)
7. Bustos, O.H., Frery, A.C.: Statistical Functions and Procedures in IDL 5.6 and 6.0. Computational Statistics and Data Analysis 50, 301–310 (2005)
8. Bustos, O.H., Lucini, M.M., Frery, A.C.: M-estimators of Roughness and Scale for GAo -modelled SAR Imagery. EURASIP Journal on Applied Signal Processing 1, 105–114 (2002)
9. Campbell, N.A.: Robust procedure in multivariate analysis I: Robust covariance estimators. Applied Statistics 29, 231–237 (1980)
10. Chang, H., Yeung, D.Y.: Robust locally linear embedding. Pattern Recognition 39(6), 1053–1065 (2006)
11. Chen, G., Qian, S.: Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal components analysis: Can. J. Remote Sensing 34(5), 447–454 (2008)
12. Croux, C., Haesbroeck, G.: Principal Component Analysis Based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies. Biometrika 87, 603–618 (2000)
13. Devlin, S.J., Gnanadesikan, R., Kettenring, J.R.: Robust Estimation of Dispersion Matrices and Principal Components. Journal of the American Statistical Association 76, 354–362 (1981)
14. Hamza, B., Krim, H.: Image Denoising: A Nonlinear Robust Statistical Approach. IEEE Trans. Signal Processing 49(12), 3045–3054 (2001)
15. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: Robust Statistics: The Approach Based on Influence Functions. John Wiley and Sons, New York (1986)
16. Huber, P.J.: Robust Statistics. John Wiley and Sons, New York (1981)
17. Jolliffe, I.T.: Principal Component Analysis. Springer, New York (2002)

18. Locantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T., Cohen, K.L.: Robust principal components for functional data. Test 8, 1–28 (1999)
19. Lucini, M.M., Ruiz, V.F., Frery, A.C., Bustos, O.H.: Robust Classification of SAR Imagery. In: IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing, vol. VI, pp. 557–561. IEEE Press, Los Alamitos (2003)
20. Maronna, R.A.: Principal components and orthogonal regression based on robust scales. Technometrics 47(3), 264–273 (2005)
21. Maronna, R.A., Martin, D., Yohai, V.J.: Robust Statistics: Theory and Methods. John Wiley and Sons, Chichester (2006)
22. Naga, R., Antille, G.: Stability of robust and non-robust principal components analysis. Computational Statistics and Data Analysis 10, 159–174 (1990)
23. Oh, H., Nychka, D.W., Lee, T.C.M.: The role of pseudo data for robust smoothing with application to wavelet regression. Biometrika 94(4), 893–904 (2007)
24. Oliver, C., Quegan, S.: Understanding Synthetic Aperture Radar Images. Artech-House (1998)
25. Seber, G.A.F.: Multivariate observations. John Wiley and Sons, Chichester (1985)
26. Tyo, J.S., Konsolakis, A., Diersen, D.I., Olsen, R.C.: Principal-Components-Based Display Strategy for Spectral Analysis. IEEE Transactions on Geoscience and Remote Sensing 41(3), 708–718 (2003)
27. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, `http://www.r-project.org`; ISBN: 3-900051-07-0
28. AVIRIS, `http://aviris.jpl.nasa.gov/html/aviris.freedata.html`