Hyperspectral Anomaly Detectors Using Robust Estimators

Joana Frontera-Pons, Miguel Angel Veganzones, *Member, IEEE*, Frédéric Pascal, *Senior Member, IEEE*, and Jean-Philippe Ovarlez, *Member, IEEE*

***Abstract*—Anomaly detection methods are devoted to target detection schemes in which no a priori information about the spectra of the targets of interest is available. This paper reviews classical anomaly detection schemes such as the widely spread Reed–Xiaoli detector and some of its variations. Moreover, the Mahalanobis distance-based detector, rigorously derived from a Kelly's test-based approach, is analyzed and its exact distribution is derived when both mean vector and covariance matrix are unknown and have to be estimated. Although, most of these techniques are based on Gaussian distribution, we also propose here ways to extend them to non-Gaussian framework. For this purpose, elliptical distributions are considered for background statistical characterization. Through this assumption, this paper describes robust estimation procedures (M-estimators of location and scale) more suitable for non-Gaussian environment. We show that using them as plug-in estimators in anomaly detectors leads to some great improvement in the detection process. Finally, the theoretical contribution is validated through simulations and on real hyperspectral scenes.**

***Index Terms*—Anomaly detection, elliptical distributions, hyperspectral imaging, *M*-estimation.**

I. Introduction

TARGET detection and anomaly detection of multidimensional signals have proved to be valuable techniques in a wide range of applications, including search-and-rescue, surveillance, rare mineral, and land mines detection (e.g., [1]–[3]). Target detection aims to discover the presence of a specific signal of interest among a set of signals. Statistical target detection is based on the Neyman–Pearson (NP) criterion, which maximizes the probability of detection (PD) for a given probability of false alarm (PFA).

Classical target detection methods require the knowledge of the spectra of the desired targets. One could be interested in a large number of possible targets each with different signatures. Thus, variety of sought spectra corresponding to the different kinds of targets and the difficulties due to the atmospheric compensation for the measured spectral signatures (used as steering vectors) have led to the derivation of new algorithms that intend to distinguish unusual materials in a scene without reference to target signatures. In this work, we are focused on anomaly detection (e.g., [4] and references therein). It can be interpreted as a particular case of target detection in which no *a priori* information about the spectra of the targets of interest is available. Hence, the aim of anomaly detection is to locate objects in the image that are anomalous with respect to (w.r.t.) the background. The type of interesting targets can differ significantly from one application to another, e.g., in forestry applications, infected trees are the anomalies of interest, whereas in defense and intelligence applications, the anomalies to be detected are usually vehicles. Note that, since anomaly detectors do not use any *a priori* knowledge, they cannot distinguish between true targets and detections of bright pixels of the background or targets that are not of interest. This fact makes extremely difficult to define a false alarm rate for the detectors as highlighted in [5].

Anomalies are defined with reference to a model of the background. As for the previous target detection methods, the background model is developed adaptively using reference data (see, e.g., [6] for a complete survey in anomaly detection methods). Most of these methods rely on the classical Gaussian distribution assumption and need for the statistical characterization of the background usually through first- and second-order parameters (i.e., the mean vector and the covariance matrix). In this case, the reference data are taken either from a local neighborhood around the observation vector or using all the pixels in the image. Both approaches have their benefits (e.g., [7]). Local strategy provides more realistic scenario for the background characterization. However, it can be sensitive to the presence of false alarms due to isolated anomalies. While the global approach is not likely to generate this kind of false alarms, it will decrease the detection capability for isolated targets. From here on, local procedures will be considered for different detection schemes.

A. Related Work

The actual distribution of the hyperspectral background pixels differs from the theoretically predicted under Gaussian hypothesis (see [8] for a recent overview on background modeling for HSI). In fact, as stated in [9] and [10], the empirical distribution usually has heavier tails compared to the Gaussian distribution, and these tails strongly influence the observed false-alarm rate of the detector. Therefore, the class of elliptical distributions is assumed for background statistics characterization. The family of elliptical distributions was originally introduced by Kelker [11] and widely studied in [12].

Manuscript received September 15, 2014; revised May 22, 2015; accepted June 19, 2015. Date of publication July 22, 2015; date of current version February 09, 2016.

J. Frontera-Pons and F. Pascal are with SONDRA, Supélec, F-91190 Gif-sur-Yvette, France (e-mail: joana.fronterapons@supelec.fr; frederic.pascal@supelec.fr).

M. A. Veganzones is with GIPSA-Lab, CNRS, F-38402 Saint Martin d'Hres, France (e-mail: miguel-angel.veganzones@gipsa-lab.fr).

J.-P. Ovarlez is with ONERA, DEMR/TSI, F-91120 Palaiseau, France (e-mail: jean-philippe.ovarlez@onera.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2015.2453014

They account for non-Gaussianity providing a long-tailed alternative to multivariate normal model. They are proved to represent a more accurate characterization of HSI than models based on Gaussian assumption [9]. However, when considering elliptical distributions, the classical Gaussian-based estimators do not provide optimal performance anymore. Complementary to this paper, [8] focused on the non-Gaussian background modeling with elliptical distributions; this work aims at studying appropriate robust parameters estimates.

B. Contributions

We consider on the first part of this paper the most popular Gaussian-based anomaly detectors, and we provide a rigorous derivation of the Mahalanobis distance through a Kelly's test-based approach. Moreover, one of the theoretical contributions is the derivation of the exact distribution for the classical Mahalanobis-based anomaly detector when both the mean vector and the covariance matrix are unknown in Gaussian environment.

Furthermore, robust estimation methods are considered in classical anomaly detection schemes for non-Gaussian-distributed background assumption, and the improvement brought in most scenarios is pointed out. The fixed point (FP) estimators (also known as Tyler's estimators [13]) are proposed for the parameters estimation. These can then be used as plug-in estimators in place of the unknown mean vector or/and of the covariance matrix in the detection scheme (e.g., [14] and [15]). This is a simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption. One of the contributions of this work is to extend the results presented in [16].

C. Outline and Notation

This paper is organized as follows. Section II revisits classical anomaly detection schemes and provides the theoretical contribution of this paper by deriving the distribution of one of the detectors. Section III describes the family of elliptical distributions and the robust estimation methods studied in this paper for anomaly detection purposes. Section IV illustrates the theoretical analysis through simulations and Section V reveals the theoretical improvement over real hyperspectral images. Finally, Section VI concludes this work.

In the following, vectors (matrices, respectively) are denoted by bold-faced lowercase letters (uppercase letters, respectively). ^{T} represents the transpose operator. $|\mathbf{A}|$ represents the determinant of the matrix $\mathbf{A}$ and $\mathrm{Tr}(\mathbf{A})$ its trace. j is used to denote the unit imaginary number. $\sim$ means "distributed as." $\Gamma(\cdot)$ denotes the gamma function. Eventually, $||\mathbf{x}||$ represents the Euclidean norm of the vector $\mathbf{x}$.

II. Adaptive Anomaly Detection Methods

Before detailing the analysis of the corresponding detectors, let us recap the most common Gaussian-based estimators. Along with their well-known properties and their simplicity of analysis, the sample covariance matrix (SCM) and the sample mean vector (SMV) are the most extended estimates since they are the maximum likelihood estimators (MLEs) for Gaussian case, as shown in [17]

$$\hat{\boldsymbol{\mu}}_{SMV} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{1}$$

$$\hat{\boldsymbol{\Sigma}}_{SCM} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})^T. \tag{2}$$

Further, we shall denote the centered SCM (CSCM) as

$$\hat{\boldsymbol{\Sigma}}_{CSCM} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \tag{3}$$

where N denotes the number of secondary data. However, such widespread techniques are suboptimal when the noise is a non-Gaussian stochastic process. Section III reviews some robust procedures particularly suited for estimating the covariance matrix and the mean vector of elliptical populations.

Let us now detail the most popular Gaussian-based anomaly detectors.

A. Reed–Xiaoli Detector

The original Reed–Xiaoli detector (RXD) proposed in [18] is commonly considered as the benchmark anomaly detector for hyperspectral data. The considered signal model can be written as

$$\begin{cases}\mathcal{H}_0 : \mathbf{x}_i = \mathbf{b}_i, & i = 1, \ldots, N\\ \mathcal{H}_1 : \mathbf{x}_i = \mathbf{p}\,\alpha_i + \mathbf{b}_i, & i = 1, \ldots, N\end{cases}$$

where $\mathbf{x}_i$ are the N available data vectors on the image of dimension m. $\mathbf{b}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ represents the residual background; $\mathbf{p}$ is the spectral signature of the possible anomalous material assumed to be unknown; and α_i stands for the amplitude of the intended targets through the N available data, i.e., it is a known vector $\boldsymbol{\alpha}$ of dimension N that indicates the strength and position of the sought targets over the image. Remark that each vector from the available data can potentially contain an anomaly, whereas in classical detection problem, secondary data are assumed to be signal free. Thus, one can arrange the vector data into a matrix as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$, and the detection scheme derived in [18] takes the form

$$\Lambda(\mathbf{X}) = \frac{(\mathbf{X}\boldsymbol{\alpha}^T)^T(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\boldsymbol{\alpha}^T)}{\boldsymbol{\alpha}\boldsymbol{\alpha}^T} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda.$$

Since hyperspectral data are not zero mean, let us now consider that the background $\mathbf{b}_i$ is distributed according to $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the mean vector $\boldsymbol{\mu}$ is supposed to be known. In the case just one anomaly in the data under test is intended to be detected, the corresponding amplitude vector can be written as $\boldsymbol{\alpha}_i = [0 \ldots 0\,1\,0 \ldots 0]^T$, where 1 is at the ith position, and the previous detector, whatever $i \in [1, N]$, takes the form

$$\Lambda_{RXD} = (\mathbf{x}_i - \boldsymbol{\mu})^T\hat{\boldsymbol{\Sigma}}_{CSCM}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda\,.$$

Finally, since the mean vector is usually unknown, it can be replaced on the detector by its estimate $\hat{\boldsymbol{\mu}}_{SMV}$. The resulting detector, commonly known as two-step generalized likelihood ratio test (GLRT), yields

$$\Lambda_{ARXD} = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda\,. \quad (4)$$

The covariance matrix estimation $\hat{\boldsymbol{\Sigma}}_{SCM}$ in (4) is performed over all the data set, i.e., including the vector $\mathbf{x}_i$ under test. In the following, the test in (4) will be referred to as the adaptive RXD (ARXD) to underline the fact that the unknown mean vector is replaced by its estimate.

B. Kelly Anomaly Detector

We detail here a classical anomaly detector often mistakenly referred to as the RXD. Following the development proposed in [19], let us now assume the following signal model

$$\begin{cases} \mathcal{H}_0 : \mathbf{x} = \mathbf{b}, & \mathbf{x}_i = \mathbf{b}_i, \quad i = 1, \ldots, N \\ \mathcal{H}_1 : \mathbf{x} = \alpha\,\mathbf{p} + \mathbf{b}, & \mathbf{x}_i = \mathbf{b}_i\,, \quad i = 1, \ldots, N \end{cases}$$

and the $\mathbf{b}_1, \ldots, \mathbf{b}_N$ are assumed to be an independent identically distributed (IID) sample from a Gaussian distribution $\mathbf{b}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. As in classical Kelly detector, the covariance matrix $\boldsymbol{\Sigma}$ is unknown and the mean vector $\boldsymbol{\mu}$ is supposed to be known. However, for anomaly detector derivation, the amplitude of the signal α is supposed to be known and the unknown parameter is now the steering vector $\mathbf{p}$. Therefore, $N+1$ m-dimensional vectors are observed under each hypothesis and the joint probability density function (pdf) of the the N secondary data and the observation vector $\mathbf{x}$ under the two hypotheses $\mathcal{H}_i$ can be written as

$$f_i(\mathbf{x}) = \left(\frac{1}{2\pi^{\frac{m}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \mathrm{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{T}_i) \right] \right)^{N+1} \quad (5)$$

where $\mathbf{T}_i$ is the composite sample covariance matrix constructed from both the secondary data and the observation vector

$$\mathbf{T}_0 = \frac{1}{N+1} \left((\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T + \hat{\mathbf{W}} \right)$$

$$\mathbf{T}_1 = \frac{1}{N+1} \left((\mathbf{x} - (\alpha\mathbf{p} + \boldsymbol{\mu}))(\mathbf{x} - (\alpha\mathbf{p} + \boldsymbol{\mu}))^T + \hat{\mathbf{W}} \right)$$

and $\hat{\mathbf{W}} = N\,\hat{\boldsymbol{\Sigma}}_{CSCM}$. The first step is to maximize w.r.t. the unknown covariance matrix $\boldsymbol{\Sigma}$. Thus, the matrix maximizing the pdf f_i is simply $\mathbf{T}_i$. When this estimator is replaced in the pdf, one obtains

$$\max_{\boldsymbol{\Sigma}} f_i = \left(\frac{1}{(\pi e)^m |\mathbf{T}_i|} \right)^{\frac{N+1}{2}}. \quad (6)$$

and the GLRT, neglecting the exponent $(N+1)/2$ is given by

$$\Lambda(\mathbf{x}, \mathbf{p}) = \frac{|\mathbf{T}_0|}{|\mathbf{T}_1|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta\,. \quad (7)$$

It remains to maximize this expression over the unknown spectral signature $\mathbf{p}$ and the resulting MLE takes the form

$$\hat{\mathbf{p}} = \frac{\mathbf{x} - \boldsymbol{\mu}}{\alpha}. \quad (8)$$

After replacing $\mathbf{p}$ by (8) in (7), it is easy to show that the resulting GLRT test is equivalent to

$$\Lambda_{KellyAD\,\hat{\boldsymbol{\Sigma}}} = (\mathbf{x} - \boldsymbol{\mu})^T \hat{\boldsymbol{\Sigma}}_{CSCM}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda\,. \quad (9)$$

The quadratic form in (9) corresponds to the Mahalanobis distance detailed in [20]. It performs statistically as an outlier detector. When Gaussian assumption is valid, the quadratic form $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ follows a χ^2-distribution with m degrees of freedom for $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ perfectly known. In case the parameter $\boldsymbol{\Sigma}$ is replaced by its MLE, the CSCM, the distribution of the quadratic form follows a Hotelling T^2 distribution (see [21])

$$\Lambda^{(N)}_{KellyAD\,\hat{\boldsymbol{\Sigma}}} \sim T^2 \quad (10)$$

and thus, it can be approximated by

$$\frac{N - m + 1}{m\,N} \Lambda^{(N)}_{KellyAD\,\hat{\boldsymbol{\Sigma}}} \sim F_{m, N-m+1} \quad (11)$$

where $F_{m,N-m+1}$ is the noncentral F-distribution with m and $N - m + 1$ degrees of freedom [22] and the superscript $^{(N)}$ is used to stress the dependence on the number of secondary data N. For high values of N, $(N > 10\ \mathrm{m})$, the distribution can be approximated by the χ^2-distribution.

As discussed above, when the mean vector is unknown, it can be replaced on the detector (two-step GLRT) by its MLE leading to

$$\Lambda^{(N)}_{KellyAD\,\hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\mu}}} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T \hat{\boldsymbol{\Sigma}}_{SCM}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda. \quad (12)$$

Remark II.1: Interestingly, note that Λ_{RXD} (Λ_{ARXD}, respectively) and the $\Lambda_{KellyAD\,\hat{\boldsymbol{\Sigma}}}$ ($\Lambda^{(N)}_{KellyAD\,\hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\mu}}}$, respectively) differ only on the fact that the vector $\mathbf{x}$ under test is also present in the covariance matrix estimation in (4). Therefore, in Λ_{RXD}, the N secondary data are not assumed to be signal free and the proposed detector aims to compare every sample to the covariance matrix over all the samples. While in the second approach $\Lambda_{KellyAD\,\hat{\boldsymbol{\Sigma}}}$, one intends to differentiate the observation vector from the background statistically characterized using N samples. Hence, $N+1$ vectors are available in the latter and $\Lambda_{KellyAD\,\hat{\boldsymbol{\Sigma}}}$ does not represent anymore a benchmark structure. Often, the local Kelly detector is mistakenly referred to as the local RXD when the users either remove the vector $\mathbf{x}_i$ from the secondary data or prevent it to be part of this set using a guard window.

The distribution of this detection test is given in the next Proposition.

Proposition II.1: The distribution of the detector under Gaussian assumption is given by

$$\frac{N - m}{m\,(N+1)} \Lambda^{(N)}_{KellyAD\,\hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\mu}}} \sim F_{m, N-m} \quad (13)$$

with $F_{m,N-m}$ is the noncentral F-distribution with m and $N-m$ degrees of freedom.

Proof: For simplicity matters, the following notations are used: $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}_{SCM}$ and $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_{SMV}$. Let us set $\forall i = 1, \ldots, N$, $\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, and $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, where all these vectors are independent. Now, let us denote

$$\hat{\mathbf{W}}_{N-1} = \sum_{i=1}^{N} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T = N\,\hat{\mathbf{\Sigma}}_{SCM}.$$

Note that as an application of the Cochran theorem (e.g., [23]), one has

$$\hat{\mathbf{W}}_{N-1} \overset{dist.}{=} \sum_{i=1}^{N-1} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T = (N-1)\,\hat{\mathbf{\Sigma}}_{CSCM}$$

where $\overset{dist.}{=}$ means *is distributed as*.

Since $\hat{\boldsymbol{\mu}} \sim \mathcal{N}\left(\boldsymbol{\mu}, \frac{1}{N}\mathbf{\Sigma}\right)$, one has $\mathbf{x} - \hat{\boldsymbol{\mu}} \sim \mathcal{N}\left(\mathbf{0}, \frac{N+1}{N}\mathbf{\Sigma}\right)$. This can be equivalently rewritten as

$$\mathbf{y} = \sqrt{N/(N+1)}(\mathbf{x} - \hat{\boldsymbol{\mu}}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}).$$

As we jointly estimate the mean and the covariance matrix, a degree of freedom is lost, compared with the only covariance matrix estimation problem.

Let us now consider $\Lambda^{(N-1)}_{KellyAD\,\hat{\mathbf{\Sigma}}}$ (i.e., $\boldsymbol{\mu}$ known) built from $N-1$ secondary data, rewritten in terms of $\hat{\mathbf{W}}_{N-1}$

$$\mathbf{\Lambda}^{(\mathbf{N-1})}_{\mathbf{KellyAD}\,\hat{\mathbf{\Sigma}}} = (\mathbf{N}-\mathbf{1})\left((\mathbf{x}-\boldsymbol{\mu})^{\mathbf{T}}\hat{\mathbf{W}}^{-\mathbf{1}}_{\mathbf{N-1}}(\mathbf{x}-\boldsymbol{\mu})\right)$$

where $(x - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and whose distribution is given by (11), where N is replaced by $N-1$.

Now, for the joint estimation problem, the $\Lambda_{KellyAD\,\hat{\mathbf{\Sigma}},\hat{\boldsymbol{\mu}}}$ can be rewritten as

$$\begin{aligned}\Lambda^{(N)}_{KellyAD\,\hat{\mathbf{\Sigma}},\hat{\boldsymbol{\mu}}} &= N\left((x-\hat{\boldsymbol{\mu}})^T\hat{\mathbf{W}}^{-1}_{N-1}(x-\hat{\boldsymbol{\mu}})\right)\\ &= N\,\frac{N+1}{N}\left(y^T\hat{\mathbf{W}}^{-1}_{N-1}y\right)\\ &\overset{dist.}{=} \frac{N+1}{N-1}\Lambda^{(N-1)}_{KellyAD\,\hat{\mathbf{\Sigma}}}\end{aligned}$$

This concludes the proof. ■

The "PFA-threshold" relationship is easily obtained as the complementary cumulative density function (cdf) of the detector distribution.

It is worth pointing out from (12) that $\Lambda_{KellyAD\,\hat{\mathbf{\Sigma}},\hat{\boldsymbol{\mu}}}$ performs similar to a matched filter structure applied to $\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}$

$$\Lambda(\mathbf{x}) = c\,\mathbf{H}^T(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}) \tag{14}$$

where $\mathbf{H}^T$ is the matched signal and c a constant that can be also a function on $\mathbf{x}$. The expression in (14) is completely characterized by the matched signal $\mathbf{H}^T$ and the scale constant c. Hence, one can identify from (12) the matched signal $\mathbf{H}^T = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T\hat{\mathbf{\Sigma}}^{-1}_{SCM}$ and $c = 1$.

C. Normalized-RXD and Uniform Target Detector

Following the same approach as in (14), one can derive many different anomaly detection schemes. We recall here two popular variants of the Mahalanobis distance described in [4]: 1) the normalized-RXD (N-RXD) and 2) the uniform target detector (UTD).

The N-RXD takes the form

$$\Lambda_{N-RXD} = \frac{(\mathbf{x}-\hat{\boldsymbol{\mu}}_{SMV})^T}{||\mathbf{x}-\hat{\boldsymbol{\mu}}_{SMV}||}\hat{\mathbf{\Sigma}}^{-1}_{SCM}\frac{(\mathbf{x}-\hat{\boldsymbol{\mu}}_{SMV})}{||\mathbf{x}-\hat{\boldsymbol{\mu}}_{SMV}||}\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}}\lambda \tag{15}$$

where $||\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}||^2 = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})$ stands for the Euclidean norm of the vector. The detection test in (15) can be immediately identified as the normalized version of $\Lambda_{KellyAD}$. In addition, Λ_{N-RXD} takes also the form of a matched filter specified in (14) with matched signal $\mathbf{H}^T = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})^T\hat{\mathbf{\Sigma}}^{-1}_{SCM}$ the same as in (9) and a different scale constant $c = ||\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV}||^{-2}$.

The UTD is another widespread anomaly detection test. It was first introduced in [24] and can be defined as

$$\Lambda_{UTD} = (\mathbf{1} - \hat{\boldsymbol{\mu}}_{SMV})^T\hat{\mathbf{\Sigma}}^{-1}_{SCM}(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}}\lambda \tag{16}$$

with $\mathbf{1} = [1, \ldots, 1]^T$ is the m-dimensional unity vector. Once again, the detector in (16) can be interpreted as a matched filter where $\mathbf{H}^T = (\mathbf{1} - \hat{\boldsymbol{\mu}}_{SMV})^T\hat{\mathbf{\Sigma}}^{-1}_{SCM}$ is the matched signal. If there is no *a priori* information about the target spectra, the nonprior approach is the one that does not introduce any information into the detector and consists on assuming uniform distribution for the spectra over all the bands.

D. Generalized Kelly Anomaly Detector

In the previous detection schemes, it has not been taken into account in the derivation of the test that both mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$ were unknown. One simply replaced the mean vector by a plug-in estimate in the detector (two-step GLRT). In case both covariance matrix and mean vector are unknown, we need to derive a new detector. This strategy is similar to the one proposed in [25] for the generalized Kelly detection test. The likelihood functions under $\mathcal{H}_0$ and $\mathcal{H}_1$ are given in (5). Under $\mathcal{H}_0$ and $\mathcal{H}_1$, the maxima are achieved at

$$\max_{\mathbf{\Sigma},\boldsymbol{\mu}} f_i = \left(\frac{1}{(\pi\boldsymbol{e})^{\mathbf{m}}|\mathbf{T_i}|}\right)^{\frac{N+1}{2}}, \quad \text{for } i = 0, 1$$

where

$$(N+1)\mathbf{T}_0 = (\mathbf{x}-\boldsymbol{\mu}_0)(\mathbf{x}-\boldsymbol{\mu}_0)^T + \sum_{i=1}^{N}(\mathbf{x}_i-\boldsymbol{\mu}_0)(\mathbf{x}_i-\boldsymbol{\mu}_0)^T$$

$$\begin{aligned}(N+1)\mathbf{T}_1 &= (\mathbf{x}-\alpha\mathbf{p}-\boldsymbol{\mu}_1)(\mathbf{x}-\alpha\mathbf{p}-\boldsymbol{\mu}_1)^T\\ &\quad+\sum_{i=1}^{N}(\mathbf{x}_i-\boldsymbol{\mu}_1)(\mathbf{x}_i-\boldsymbol{\mu}_1)^T\end{aligned}$$

and

$$\boldsymbol{\mu}_0 = \frac{1}{N+1}\left(\mathbf{x} + \sum_{i=1}^{N} x_i\right) \tag{17}$$

$$\boldsymbol{\mu}_1 = \frac{1}{N+1}\left(\mathbf{x} - \alpha\mathbf{p} + \sum_{i=1}^{N} x_i\right). \tag{18}$$

Following the same lines as in [25], one has to maximize the LR in (7) w.r.t. $\mathbf{p}$. This is obtained by taking

$$\hat{\mathbf{p}} = \frac{N+1}{N}\frac{(\mathbf{x} - \boldsymbol{\mu_0})}{\alpha}. \tag{19}$$

Hence, the resulting detector can be written according to

$$\Lambda_{G-KellyAD} = (\mathbf{x} - \boldsymbol{\mu}_0)^H \, \mathrm{S}_0^{-1} \, (\mathbf{x} - \boldsymbol{\mu}_0) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda \tag{20}$$

where $\mathbf{S}_0 = \sum_{i=1}^{N}(x_i - \boldsymbol{\mu}_0)(x_i - \boldsymbol{\mu}_0)^H$ and $\boldsymbol{\mu}_0 = \frac{1}{N+1}(x + \sum_{i=1}^{N} x_i)$. Once again, the mean vector estimate $\boldsymbol{\mu}_0$ and the covariance matrix $\mathbf{S}_0$ depend on the data under test $\mathbf{x}$. Hence, $\mathbf{x} - \boldsymbol{\mu}_0$ and $\mathbf{S}_0$ are not independent. Remark that one can write $(\mathbf{x} - \boldsymbol{\mu}_0) = \frac{N}{N+1}(\mathbf{x} - \hat{\boldsymbol{\mu}}_{SMV})$. Neglecting the multiplicative constants, the test in (20) appears to be equivalent to the classical Λ_{RXD} obtained throughout a different approach but built with $N+1$ available data.

III. Robust Anomaly Detection

In this section, the class of elliptical distributions and robust estimation procedures are reviewed.

A. Elliptical Distributions

Hyperspectral data have been proved not to be multivariate normal but long-tailed distributed [9]. In order to take into account these features, the class of elliptical distributions (see, e.g., [12] and [26] for a complete survey on elliptical distributions) is considered to describe the statistical behavior of the hyperspectral background. An m-dimensional random real vector $\mathbf{x}$ has a multivariate elliptical distribution if its characteristic function is of the form

$$\Phi_{\mathbf{x}}(\mathbf{c}) = \exp\left(j\mathbf{c}^T\boldsymbol{\mu}\right)\, \phi\left(\frac{1}{2}\mathbf{c}^T\,\boldsymbol{\Sigma}\,\mathbf{c}\right) \tag{21}$$

for some function $\phi : \mathbb{R}^+ \to \mathbb{R}$, called characteristic generator, a positive semidefinite matrix $\boldsymbol{\Sigma}$, called *scatter* matrix, and $\boldsymbol{\mu} \in \mathbb{C}^m$ the *location* vector. We shall write $\mathbf{x} \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. From (21), it does not follow that $\mathbf{x}$ has a pdf $f_{\mathbf{x}}(\cdot)$, but if exists, it has the form

$$f_{\mathbf{x}}(\mathbf{x}) = c_{m,h}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\, h_m\left(\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T\,\boldsymbol{\Sigma}^{-1}\,(\mathbf{x} - \boldsymbol{\mu})\right) \tag{22}$$

where $c_{m,h}$ is a normalization constant and $h_m(\cdot)$ is any function, such as (22) that defines a pdf in $\mathbb{R}^m$. The function h_m is usually called density generator and it is assumed to be only approximately known. In this case, we shall write $\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_m)$ instead of $\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. Remark that the pdf in (22) depends on $\mathbf{x}$ only through the quadratic form $(\mathbf{x} - \boldsymbol{\mu})^T\,\boldsymbol{\Sigma}^{-1}\,(\mathbf{x} - \boldsymbol{\mu})$. Thus, the level sets of the density $\mathbf{f}_{\mathbf{x}}(\mathbf{x})$ are ellipsoids in the Euclidean m-space.

If the second-order moment exists, then $\boldsymbol{\Sigma}$ reflects the structure of the covariance matrix of the elliptically distributed random vector $\mathbf{x}$, i.e., the covariance matrix is equal to the scatter matrix up to a scalar constant. It serves to characterize the correlation structure existing within the spectral bands. It is worth pointing out that the family of elliptical distributions includes a large number of distributions, notably the Gaussian distribution, multivariate t-distribution, K-distribution, or multivariate Cauchy. Thus, it allows for heterogeneity of the background power with the texture.

In order to improve the parameter estimation, the objective is to find an appropriate model and to use the corresponding MLEs. Therefore, if the density generator h_m is perfectly known, once could obtain the optimal MLEs for such h_m. This method leads to asymptotically efficient estimators but not necessarily robust. Indeed, the robust estimator is rather one that is still fairly reliable, regardless of the data departure, failing to be optimal in some scenarios. In real-life applications, although elliptical distributions offer a great deal of possible distributions, the risk that the data do not follow the model considered still remains. Thus, the models used always correspond to simplifications of the reality. The fact that a slight deviation between reality and the model assumed has little or no influence on the parameter estimates is precisely the robustness of the estimator.

B. Robust Parameters Estimation

We detail in this section robust estimation procedures suitable for estimating the mean vector and scatter matrix within the class of elliptical distributions.

1) MLEs: When the density generator $h_m(.)$ is perfectly known, i.e., the pdf of the underlying distribution is explicit, then the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be derived and they are given by

$$\hat{\boldsymbol{\mu}}_{MLE} = \frac{\sum_{i=1}^{N} \psi(t_i)\,\mathbf{x}_i}{\sum_{i=1}^{N} \psi(t_i)}$$

$$\hat{\boldsymbol{\Sigma}}_{MLE} = \sum_{i=1}^{N} \psi(t_i)\,(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T\,(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{MLE})$$

where $t_i = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{MLE})^T\,\hat{\boldsymbol{\Sigma}}_{MLE}^{-1}\,(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{MLE})$ and $\psi(t) = -2\,h_m'(t)/h_m(t)$.

Note that the two quantities $\hat{\boldsymbol{\mu}}_{MLE}$ and $\hat{\boldsymbol{\Sigma}}_{MLE}$ appear on both sides of these equations, characterizing FP equations $\hat{\boldsymbol{\mu}}_{MLE} = f(\hat{\boldsymbol{\mu}}_{MLE}, \hat{\boldsymbol{\Sigma}}_{MLE})$ and $\hat{\boldsymbol{\Sigma}}_{MLE} = g(\hat{\boldsymbol{\mu}}_{MLE}, \hat{\boldsymbol{\Sigma}}_{MLE})$. Note that, due to the classical ML theory, solutions of such equations exist.

2) FP Estimators: Generalizing these MLEs leads to the class of M-estimators, introduced in hyperspectral community

in [27]. More precisely, the main idea is to define a class of estimates that are not directly related to the underlying pdf: $\psi(.)$ is not anymore a function of $h_m(\cdot)$. However, the asymptotical distributions of these M-estimates are very close to those obtained in the Gaussian context, i.e., for $\hat{\mathbf{\Sigma}}_{SCM}$ and $\hat{\boldsymbol{\mu}}_{SMV}$. In the elliptically distributed background context, they appear to be more appropriate and robust to potential outliers present in the data. For consequence, this implies that these M-estimates can replace conventional Gaussian estimates in all detection schemes without degrading their performance in Gaussian context but with enhancing their performance in non-Gaussian context. Among the large class of M-estimators, the FP estimators, according to the definition proposed by Tyler in [13], appear to be the most robust estimates satisfying the following implicit equations:

$$\hat{\boldsymbol{\mu}}_{FP} = \frac{\displaystyle\sum_{i=1}^{N} \frac{\mathbf{x}_i}{\left((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\mathbf{\Sigma}}_{FP}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})\right)^{1/2}}}{\displaystyle\sum_{i=1}^{N} \frac{1}{\left((\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\mathbf{\Sigma}}_{FP}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})\right)^{1/2}}} \tag{23}$$

$$\hat{\mathbf{\Sigma}}_{FP} = \frac{m}{N} \sum_{i=1}^{N} \frac{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T}{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\mathbf{\Sigma}}_{FP}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})}. \tag{24}$$

The joint solutions can be obtained using the recursive algorithm given by

$$\begin{aligned} &\hat{\boldsymbol{\mu}}_{FP}^{(0)} = \hat{\boldsymbol{\mu}}_{SMV} \qquad \hat{\mathbf{\Sigma}}_{FP}^{(0)} = \hat{\mathbf{\Sigma}}_{SCM} \\ &\begin{cases} \hat{\boldsymbol{\mu}}_{FP}^{(n+1)} = f\left(\hat{\boldsymbol{\mu}}_{FP}^{(n)}, \hat{\mathbf{\Sigma}}_{FP}^{(n)}\right) \\ \hat{\mathbf{\Sigma}}_{FP}^{(n+1)} = g\left(\hat{\boldsymbol{\mu}}_{FP}^{(n)}, \hat{\mathbf{\Sigma}}_{FP}^{(n)}\right) \end{cases} \end{aligned} \tag{25}$$

The FP estimates have been widely investigated in statistics and signal processing literature. We refer to [28] for a detailed performance analysis. It is worth pointing out that $\hat{\mathbf{\Sigma}}_{SCM}$ and $\hat{\mathbf{\Sigma}}_{FP}$ have the same asymptotic Gaussian distribution which differs from their second-order moment by a factor $\frac{m+1}{m} N$, i.e., for N sufficiently large, $\hat{\mathbf{\Sigma}}_{FP}$ behaves as a Wishart matrix with $\frac{m}{m+1} N$ degrees of freedom. Indeed, these estimators belong to the wider class of robust M-estimators [29].

3) Shrinkage Estimators: We present now shrinkage methods that are suitable for high-dimensional problems with small number of samples (large m small N). In these "large m small N" problems, classical estimators suffer from a distorted eigen-structure and improved estimators are required.

A common regularization approach has been widely studied, the shrinkage-SCM approach introduced in [30] and [31]. Thus, in Gaussian context, the regularized SCM takes the form

$$\hat{\mathbf{M}}_{Shr-SCM}(\beta) = \frac{1-\beta}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{SMV})^T + \beta \mathbf{I}_m. \tag{26}$$

In the presence of non-Gaussian, impulsive background, the estimate in (26) suffers from the same drawbacks as the SCM and the class of robust estimates are more appropriate. Yet, the

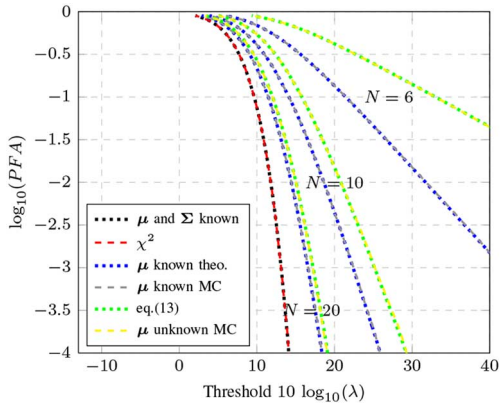

Fig. 1. PFA versus threshold for the $\Lambda_{Kelly\,AD}$ ($m = 5$) when (a) $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known (Mahalanobis) (red and black curves), (b) only $\boldsymbol{\mu}$ is known (gray and blue curves), and (c) Proposition II.1: both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown (yellow and green curves).

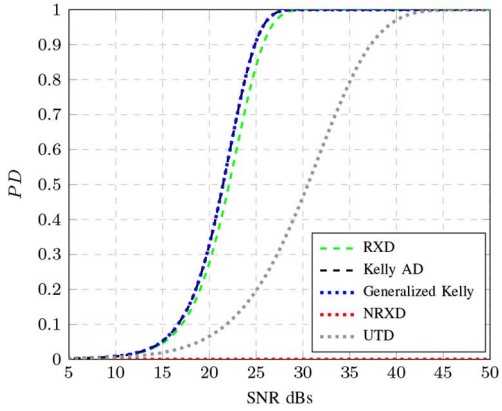

Fig. 2. PD for different SNR values, $m = 5$, $N = 10$, and $PFA = 10^{-3}$ in Gaussian case.

FP estimators described above exhibit important shortcomings in high-dimensional context and they cannot be computed for the undersampling case when $m > N$.

Moreover, we extend here FP covariance matrix estimator to the high-dimensional setting using shrinkage regularization. Let us consider now the shrinkage FP introduced in [32] and defined as the solution of the following FP equation:

$$\hat{\mathbf{M}}_{Shr-FP}(\beta) = (1-\beta)\frac{m}{N} \times \sum_{i=1}^{N} \frac{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T}{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})^T \hat{\mathbf{M}}_{Shr-FP}^{-1}(\beta)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{FP})} + \beta \mathbf{I}_m \tag{27}$$

for $\beta = \in (0, 1]$ and $\hat{\boldsymbol{\mu}}_{FP}$ given in (23).

It was shown in [32] that when β tends to 0, the proposed shrinkage estimator in (27) tends to the FP estimator in (24) whose inverse has its trace equal to m. A different approach that introduces a normalization constraint in the algorithm for the shrinkage FP estimates is found in [33]. Moreover, in [34]–[36], this estimator has been used within the expected likelihood framework. The optimization of the shrinkage parameter β has been discussed in [37].

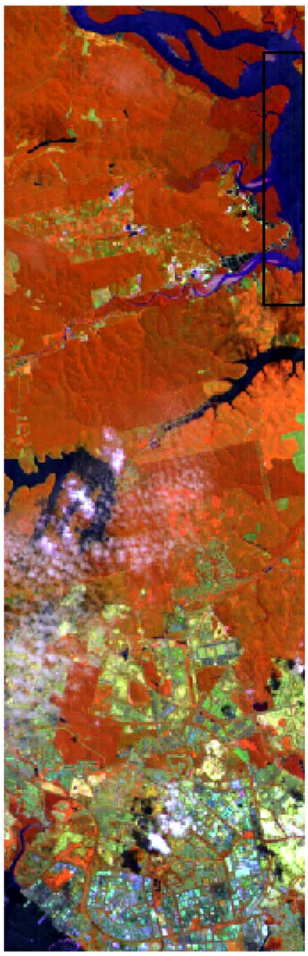

Fig. 3. True color composition of the Hyperion scene.

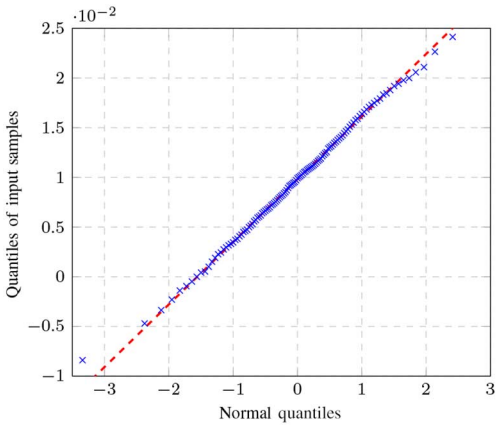

Fig. 4. Q–Q plot of the data sample versus the normal theoretical distribution.

The basis of the proposed method are the FP estimators. However, the approach presented here could be extended to other M-estimators.

C. Robust Kelly Anomaly Detector

All the detection schemes explained in Section II are derived under Gaussian assumption. In this section, we explore the use of robust estimation methods presented above for anomaly

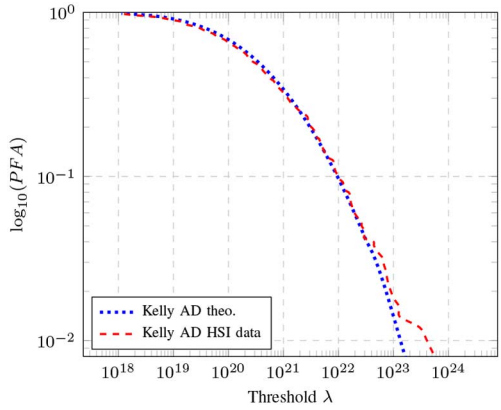

Fig. 5. Kelly AD complementary CDF of the Mahalanobis distance for a real HSI image ($m = 6$, $N = 24$).

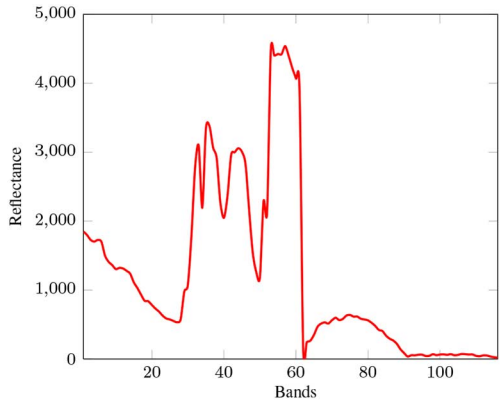

Fig. 6. Endmember used in the experiment which corresponds to land materials.

detection. These can then be used as plug-in estimators in place of the unknown mean vector and/or of the covariance matrix in the detection scheme. This is a simple but often efficient method to obtain robust properties for signal processors derived under the Gaussian assumption.

The Kelly anomaly detector has the advantage that the mean vector and the covariance matrix are independent to each other and to the observation vector, which is not the case for the generalized Kelly or the classical RXD. This allows replacing the unknown parameters by a robust FP estimators or shrinkage estimators and the detector can be written as

$$\Lambda_{KellyAD\,\hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\mu}}} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_{Robust})^T \hat{\boldsymbol{\Sigma}}_{Robust}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{Robust}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda \tag{28}$$

where $\hat{\boldsymbol{\mu}}_{Robust}$ and $\hat{\boldsymbol{\Sigma}}_{Robust}$ are those described in Section III-B. It is important to highlight that the distribution of this detector is still an open question, as far as the authors are aware. In fact, it will surely depend on the underlying particular CE distribution, i.e., the distribution will change with the choice of $h_m(\cdot)$.

IV. Simulations

In this section, we validate the theoretical analysis on simulated data. First, we validate through Monte-Carlo simulations

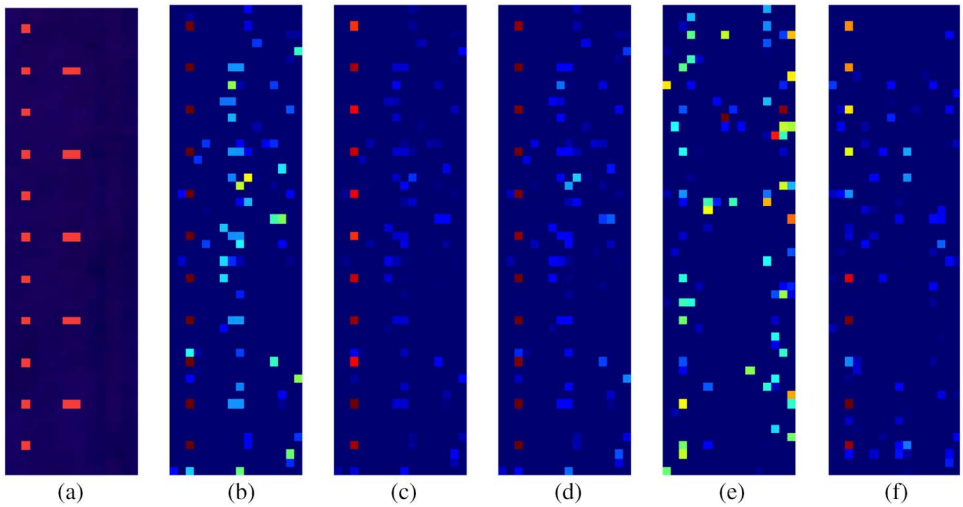

Fig. 7. Conventional anomaly detection for artificial targets in real HSI ($m = 6$, $N = 24$, same PFA = 10^{-1}). The pixels with values below the threshold λ have been set to zero. (a) Original. (b) RXD. (c) Kelly AD. (d) G-Kelly. (e) N-RXD. (f) UTD.

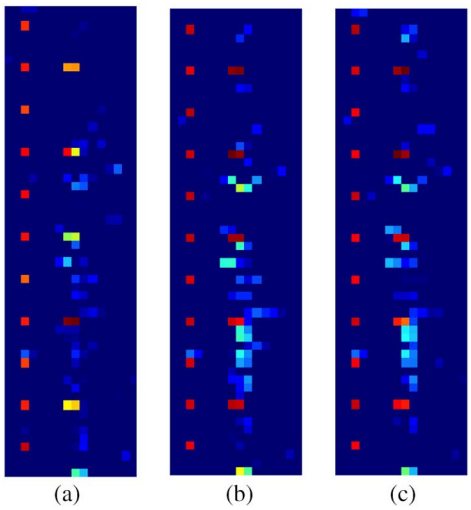

Fig. 8. Extended Kelly AD detectors built with robust estimates for artificial targets in real HSI ($m = 6$, $N = 24$, same PFA=10^{-1}). (a) FP. (b) Shr-SCM. (c) Shr-FP.

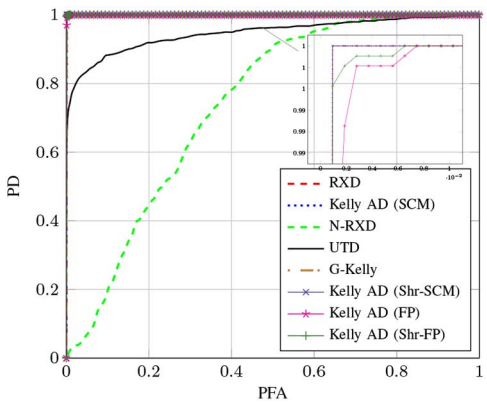

Fig. 9. ROC curve comparing different detectors in Gaussian environment ($m = 6$, $N = 24$, and SNR = 10 dB).

the distribution of $\Lambda_{Kelly\,AD}$ detailed above. The experiments have been conducted on Gaussian vectors of dimension $m = 5$ and for different values of N. The computations have been made through 10^6 Monte-Carlo trials. The true covariance is chosen as a Toeplitz matrix whose entries are $\Sigma_{i,j} = \rho^{|i-j|}$ and where $\rho = 0.4$. The mean vector is arbitrary set to have all entries equal to 3. Exceedance plot shows the fraction of points in the data set whose Mahalanobis distance is larger than the indicated value. This is essentially a cumulative histogram of Mahalanobis distance values which correspond to the "PFA-threshold" relationship. Remark that the definition of false alarms is not unique and it depends on the application. Thus, we will rather refer to the distribution of the detector in target absent hypothesis.

Fig. 1 illustrates the distribution of the detector under null hypothesis. The case where both covariance matrix and mean vector are perfectly known corresponds to the χ^2-distribution and the adaptive versions of the quadratic form become a T^2 Hotelling. The perfect agreement of the green and yellow curves bears out the results of Proposition II.1. Furthermore, we compare, in Fig. 2, the five proposed anomaly detectors in terms of PD for different values of the signal-to-noise ratio (SNR). The experiments were on Gaussian vectors of dimension $m = 5$, for $N = 10$ and the artificial targets signature used for the simulations is the unity vector $\mathbf{p} = [1, \ldots, 1]^T$. On a first step, the threshold is determined to ensure exactly the same $PFA = 10^{-3}$ for all detectors. The best results are obtained for the Mahalanobis-based detectors, i.e., the classical RXD, Kelly AD, and generalized Kelly AD. The two detectors derived according to Kelly's approach perform fundamentally the same and slightly better than the RXD. This improvement may be due to the fact that $N + 1$ data are available for the Kelly's strategies, while only N samples are used in the classical RXD. The matched filter-based detectors deliver poor performance in the case of the UTD, as the matching signal is the unity vector, which is shown to be not optimal even in the case the artificial targets signature used for the simulations is the unity vector.

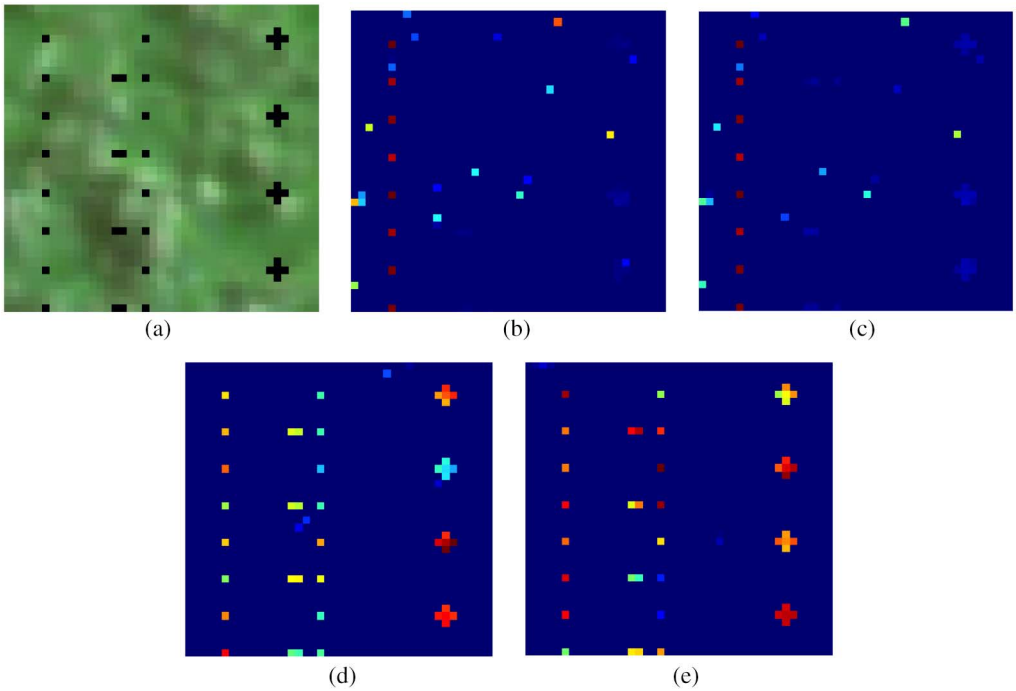

Fig. 10. Extended Kelly AD built with conventional and robust estimates for artificial targets in real HSI with all the bands ($m = 9$, $N = 80$, same PFA = 0.03). (a) Original. (b) SCM. (c) Shr-SCM. (d) FP. (e) Shr-FP.

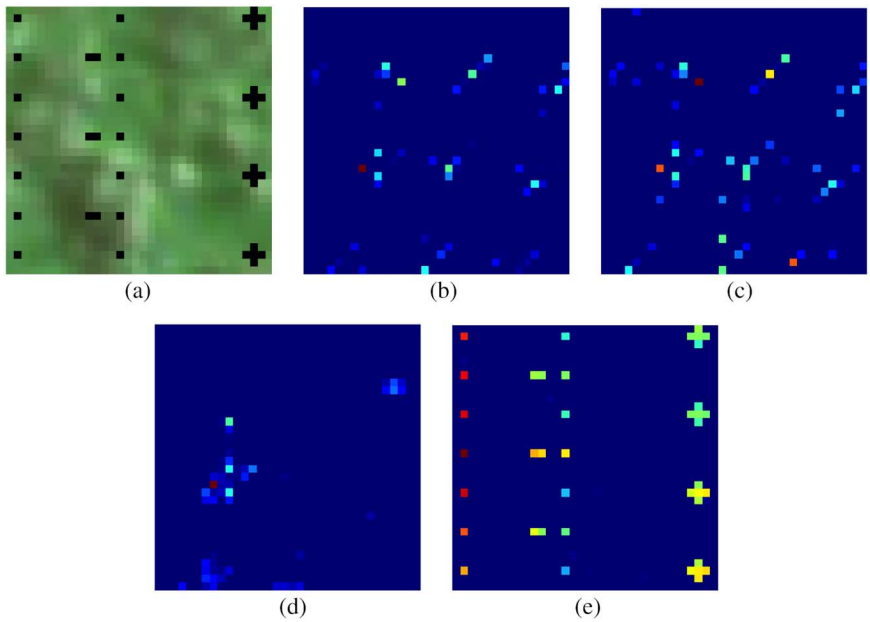

Fig. 11. Extended Kelly AD built with conventional and robust estimates for artificial targets in real HSI ($m = 126$, $N = 288$, same PFA = 10^{-1}).

The N-RXD presents an almost flat curve as the normalization factor grows as the SNR gets higher. The outcome of this detector can be assimilated to the residual background level and its use in Gaussian environment should be avoided.

V. Real Hyperspectral Data

A. Gaussian Background

The same experiments as in simulations have been conducted on a real hyperspectral image. The scene analyzed is the NASA Hyperion sensor data set displayed in Fig. 3. The image is constituted of 798×253 pixels and 116 spectral bands after water absorption bands have been removed. The analysis has been done on a homogenous part of the image corresponding to the water region on the top right of the image. The part extracted consists of 60×20 pixels. In order to ensure the validity of the proposed methods, we show in Fig. 4 the outcome of a classical Gaussianity test "Q–Q plot" for the selected region over the band 42. Even if this allows to "validate" the Gaussianity of each band, it cannot ensure the Gaussianity of the corresponding multivariate vector.

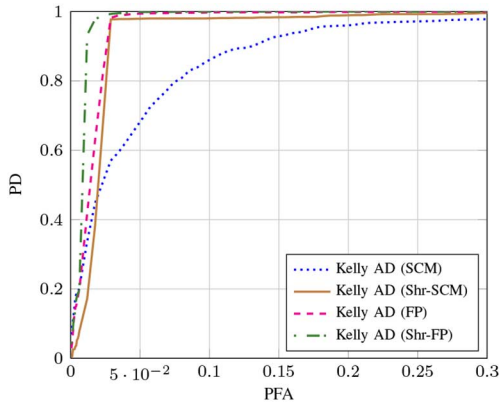

Fig. 12. ROC curve comparing different estimation methods in non-Gaussian environment ($m = 9$, $N = 80$ and SNR = 10 dB). (a) Original. (b) SCM. (c) Shr-SCM. (d) FP. (e) Shr-FP.

To avoid the well-known problem due to high dimensionality, we have chosen sequentially $m = 6$ bands. In this approach, both covariance matrix and mean vector are estimated using a sliding window of size 5×5, having $N = 24$ secondary data.

Fig. 5 shows the distribution of the $\Lambda_{KellyAD\,\hat{\boldsymbol{\Sigma}},\hat{\boldsymbol{\mu}}}$ in real hyperspectral data (red curve). We also plot the theoretical relationship defined by (13). The results obtained on real HSI data on a Gaussian-distributed region agree with the theoretical relationships presented above.

Finally, we illustrate the detection capability of the proposed methods when artificial anomalies with known spectral signature are inserted on the real hyperspectral image. For this purpose, we extract the spectral signature from ground materials in Fig. 3 and the anomaly spectra is depicted in Fig. 6. Fig. 7(a) details the position and the shape of the targets. For the same fixed value of FA $PFA = 0.1$, we present in Fig. 7 the outcome of different conventional detectors for $m = 6$ and $N = 24$ secondary data. Note that the edges of the image are not processed and the images are trimmed in function of the window size. The detectors based on the Mahalanobis distance deliver best results for detection purposes as expected and the matched filter-based detectors do not detect properly the artificial targets. These detection maps are in agreement with the SNR figure detailed above. Remark that the two-pixel targets are not detected by any of the detection schemes. This problem is due to the presence of a strong target in the secondary data that pollutes the covariance matrix estimation. Its occurrence has a significant impact on the detection process and it degrades the performances of conventional detectors. In summary, Fig. 7 shows that 1) the generalized Kelly anomaly detector (original contribution) performs reasonably well compared to the RXD and Kelly AD; 2) the classical N-RXD (e) and UTD (f) are not suited for anomaly detection in Gaussian-distributed background; and 3) the two-pixel targets are not detected when the detectors are built with SCM-SMV estimates. The latter point supports the use of robust estimators.

Fig. 8 shows the results of the extended $\Lambda_{Kelly\,AD}$ detector ($m = 6$, $N = 24$) built now with FP estimators and with shrinkage estimators, both SCM and FP. The same value of FA $PFA = 10^{-1}$ is considered for the three detectors. Remark that all the anomalies of interest are now detected even those bigger than one pixel. Thereafter, due to their robustness, the proposed estimation methods allow for better detection results in Gaussian case.

Receiver operating characteristics (ROC) curves are widely used in signal processing to evaluate the performances of the detectors [38]. ROC curves depict the outcome for a set of thresholds instead of showing the misclassification for only one. The x-axis represents the PFA and the y-axis the PD. A good detector presents high PD values at low PFA, i.e., the curve is closer to the upper left corner. Fig. 9 shows the results of the ROC curves for the different detectors presented above ($m = 6$, $N = 24$) computed on the Gaussian water region of the Hyperion image. For a fixed value of the SNR = 10 dB, RXD, Kelly AD built with the different estimators, and generalized Kelly AD exhibit perfect classification, while N-RXD and UTD provide worse results.

B. Non-Gaussian Background

Let us now present some results on a real hyperspectral image in which the background cannot be characterized with Gaussian distribution and artificial targets were introduced as anomalies. The original data set consists of 50×50 pixels with 126 bands, from which we have chosen sequentially $m = 9$ bands; see Fig. 10(a). For this example, both covariance matrix and mean vector are estimated using a sliding window of size 9×9 having $N = 80$ secondary data. The results for the $\Lambda_{Kelly\,AD}$ built with classical SMV–SCM estimates, FP estimates, and shrinkage estimators are shown in Fig. 10, the FA being fixed at the same value of $PFA = 0.03$. In this case, FP estimators and notably shrinkage FP estimates are capable of locating all the artificial targets and exhibit a lower number of false alarms. This improvement is due to the fact that FP estimates treat the outliers and impulsive samples in order for them to have a smaller contribution to the background characterization process, while the SMV–SCM estimates (and its respective shrinkage version) suffer from the presence of strong reflectance pixels in the secondary data. Remark that the shrinkage FP estimates lead to a better detection compared to FP estimates.

Let us now consider the data set in Fig. 10(a) with all 126 bands available. In high-dimensionality problems, the SMV–SCM and the FP estimators suffer from distorted eigenstructure. This fact motivates the use of shrinkage estimators. As the background is shown to be non-Gaussian, the shrinkage FP estimators are the most appropriate solutions when dealing with all the bands. We show in Fig. 11 the results when using all the bands, $m = 126$, and the sliding window has been increased to 17×17, $N = 288$. Remark that as the edges of the image are not processed, the original image and the detection maps are cropped in function of the window size. The shrinkage FPs are still capable of detecting all the targets, while all the other estimation techniques lead to poor detection results.

Fig. 12 displays different ROC curves in non-Gaussian background for the image in Fig. 10(a), with $m = 9$, $N = 80$, and a fixed value of SNR = 10 dB. Classical SMV–SCM provides worse results than their shrinkage counterpart. Moreover,

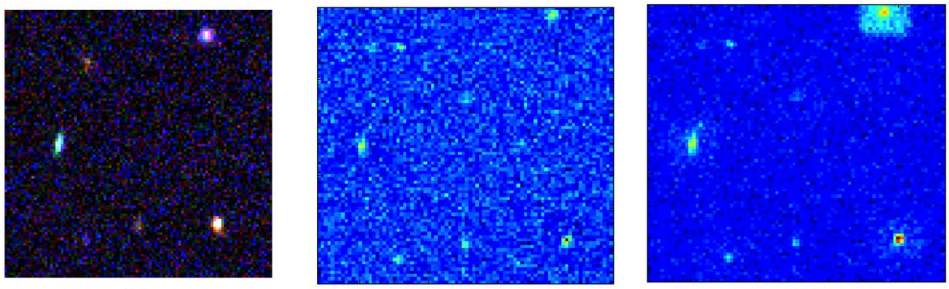

Fig. 13. Anomaly detection in a MUSE hyperspectral image (size 300×300) with $m = 36$ channels and $N = 120$ secondary data. (a) Original image (MUSE data cube). (b) Kelly AD built with conventional estimates (SMV–SCM). (c) Kelly AD built with FP estimates (SMV–SCM).

in non-Gaussian case, FP estimators and the corresponding shrinkage FP estimators allow for better detection while keeping the false alarm low. This is translated in ROC curves closer to the upper left corner and improved detection performances.

The algorithm has also been applied for galaxy detection on the MUSE data cube. The Multi Unit Spectroscopic Explorer (MUSE) project (see [39]) aims to provide astronomers with a new generation of optical instrument, capable of simultaneously imaging the sky [in two-dimensional (2-D)] and measuring the optical spectra of the light received at a given position on the sky. MUSE was installed on the VLT telescope and operational in 2013, and its performances are expected to allow observation of far galaxies up to 100 times fainter than those presently detectable. MUSE will deliver a three-dimensional (3-D) data-cube made of a stack of images recorded at 3578 different wavelengths over the range 465–930 nm. Each monochromatic image represents a field of view of 60×60 arcsec, recorded with a spatial sampling of 0.2 arcsec. Each record results in a data cube of size 1570 MB encoding 3578 images of 300×300 pixels, possibly containing thousands of objects (galaxies) existing over different subsets of wavelengths.

An example of MUSE data cube image is displayed in Fig. 13(a), from the 3578 available bands, we have chosen one band of each 100. The results for anomaly detection are presented in Fig. 13 for the same imposed FA value $PFA = 10^{-3}$. Note that detection with FP estimators (c) provides better and clearer results than the classical ones (b).

These examples illustrate the robust behavior of FP estimators in non-Gaussian environments or for close targets detection problems.

VI. Conclusion

The classical RXD test is explored and compared to other four detectors. The different advantages and drawbacks for the different detection schemes are commented. Furthermore, the comparison is performed through Monte-Carlo simulations in Gaussian context and extended to real hyperspectral data with simulated anomalies. The family of elliptical distributions is considered for impulsive background characterization in hyperspectral imaging. In this context, robust estimation methods for mean vector and covariance matrix are used to overcome the non-Gaussianity of the background and the presence of outliers or strong scatters in the secondary data. Moreover the robust methods presented in this work significantly outperform the classical Gaussian-based SMV–SCM. Therefore, robust estimators offer a versatile alternative to Gaussian estimates. They allow to obtain better performances in impulsive environments while keeping good results in Gaussian background. The theoretical improvement provided by the robustness of the estimators is borne out through two real hyperspectral images.

References

[1] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[2] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *Signal Process. Mag.*, vol. 31, no. 1, pp. 24–33, 2014.

[3] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, Jun. 2009.

[4] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.

[5] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.

[6] N. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.

[7] D. W. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[8] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, Jun. 2014.

[9] D. Manolakis and D. Marden, "Non gaussian models for hyperspectral algorithm design and assessment," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS'02)*, 2002, vol. 3, pp. 1664–1666.

[10] J. Theiler, C. Scovel, B. Wohlberg, and B. R. Foy, "Elliptically contoured distributions for anomalous change detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 271–275, Apr. 2010.

[11] D. Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *Sankhyā Indian J. Stat. Ser. A*, pp. 419–430, 1970.

[12] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*. London, U.K.: Chapman and Hall, 1990.

[13] D. Tyler, "A distribution-free m-estimator of multivariate scatter," *Ann. Stat.*, vol. 15, no. 1, pp. 234–251, 1987.

[14] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, Dec. 2002.

[15] J. Frontera-Pons, M. Mahot, J. Ovarlez, F. Pascal, S. Pang, and J. Chanussot, "A class of robust estimates for detection in hyperspectral images using elliptical distributions background," in *Proc. IEEE Geosci. Remote Sens. Symp. (IGARSS'12)*, 2012, pp. 4166–4169.
[16] J. Frontera-Pons *et al.*, "Robust anomaly detection in hyperspectral imaging," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS'14)*, 2014, pp. 4604–4607.
[17] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman and Hall/CRC, 2000.
[18] I. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
[19] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 2, pp. 115–127, Mar. 1986.
[20] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proc. Nat. Inst. Sci.* 1936, vol. 2, pp. 49–55.
[21] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*. New York, NY, USA: Springer, 1999.
[22] E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics*. Boca Raton, FL, USA: CRC Press, 2010.
[23] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. Hoboken, NJ, USA: Wiley, 1984.
[24] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, Dept. Electr. Eng., Univ. Maryland, Baltimore, MD, USA, 1993.
[25] J. Frontera-Pons, F. Pascal, and J.-P. Ovarlez, "Adaptive non-zero mean gaussian detection and application to hyperspectral imaging," arXiv:1404.2977, 2014.
[26] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.
[27] J. Ovarlez, S. Pang, F. Pascal, V. Achard, and T. Ng, "Robust detection using the SIRV background modelling for hyperspectral imaging," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS'11)*, 2011, pp. 4316–4319.
[28] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2206–2217, Jun. 2008.
[29] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Stat.*, vol. 35, no. 1, pp. 73–101, 1964.
[30] Y. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR," *Radio Eng. Electron. Phys.*, vol. 26, no. 3, pp. 87–95, 1981.
[31] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, Jul. 1988.
[32] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to stap detection problem," arXiv:1311.6567, 2013.
[33] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, Sep. 2011.
[34] Y. I. Abramovich and O. Besson, "Covariance matrix estimation in complex elliptic distributions using the expected likelihood approach," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'13)*, 2013, pp. 6476–6480.
[35] Y. Abramovich and O. Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 1: The over-sampled case," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5807–5818, Dec. 1, 2013.
[36] O. Besson and Y. Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach—Part 2: The under-sampled case," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5819–5829, Dec. 2013.
[37] R. Couillet and M. R. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *J. Multivar. Anal.*, vol. 131, pp. 99–120, 2014.
[38] J. P. Egan, *Signal Detection Theory and ROC Analysis*. New York, NY, USA: Academic Press, 1975.
[39] "Official website of the muse project" [Online]. Available: http://muse.univ-lyon1.fr

Joana Frontera-Pons was born in Mallorca, Spain, in 1986. She received the telecommunications engineering degree from the Technical University of Catalonia (UPC), Barcelona, Spain, in 2011, and the Ph.D. degree in signal processing from Supélec, Gif-sur-Yvette, France, in 2014.

Her research interests include estimation and target detection in statistical signal processing and hyperspectral imaging.

Miguel Angel Veganzones (M'12) received the M.Sc. and Ph.D. degrees in computer science from the Basque Country University (UPV/EHU), Donostia, Spain, in 2005 and 2012, respectively.

Since 2012, he is a Posdoctoral Researcher with the Images-Signal Department, GIPSA-Lab, Grenoble, France. His research interests include analysis of hyperspectral images by means of multiway analysis, computational intelligence, and statistical techniques.

Dr. Veganzones is a Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND REMOTE SENSING, and the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING.

Frédéric Pascal (S'04–M'06–SM'14) received the Master's degree (Probabilities, Statistics and Applications: Signal, Image et Networks) with merit, in applied statistics from the University of Paris VII–Jussieu, Paris, France, in 2003, and the Ph.D. degree in signal processing from the University of Paris X–Nanterre, Paris, France, advised by Pr. Philippe Forster: "Detection and Estimation in Compound Gaussian Noise," in 2006. The Ph.D. thesis was in collaboration with the French Aerospace Lab (ONERA), Palaiseau, France.

From November 2006 to February 2008, he was in a Postdoctoral Position with the Signal Processing and Information Team, Laboratory SATIE, CNRS, ENS Cachan, France. Between March 2008 and December 2011 (Jan. 2012–Dec. 2013, respectively), he was an Assistant Professor (Associate Professor, respectively) with SONDRA, Gif-sur-Yvette, France, the French–Singaporean Laboratory of SUPELEC. In 2012, he obtained a Research Directorship Habilitation (HDR) thesis in signal processing from the University of Paris-Sud, Orsay, France. From January 2014, he is a Full Professor with L2S Laboratory, SUPELEC. His research interests include estimation, detection, and classification in statistical signal processing and radar processing.

Dr. Pascal is an elected senior member of the IEEE Signal Processing Society SAM technical committee, from January 2015.

Jean-Philippe Ovarlez (M'06) was born in Denain, France, in 1963. He received jointly the Engineering degree from Ecole Suprieure d'Electronique Automatique et Informatique (ESIEA), Paris, France, and the Diplme d'Etudes Approfondies degree in signal processing from the University of Paris XI, Orsay, France, and the Ph.D. degree in physics from the University of Paris VI, Paris, France, in 1987 and 1992, respectively.

In 2011, he obtained a Research Directorship Habilitation (HDR) thesis in signal processing from the University of Paris-Sud, Orsay, France, and his qualification to the University Professor position. In 1992, he joined the Electromagnetic and Radar Division, French Aerospace Lab (ONERA), Palaiseau, France, where he is currently a Chief Scientist and Member of the Scientific Committee of the ONERA Physics Branch. Since 2008, he has been working part time with Centrale-Supélec SONDRA Lab in charge of Signal Processing activities supervision. His research interests include topic of statistical signal processing for radar and SAR applications such as time–frequency, imaging, detection, and parameters estimation.