Besson, O., Vincent, F., Matteoli, S.

"Adaptive target detection in hyperspectral imaging from two sets of training samples with different means"

(2021) Signal Processing, 181, art. no. 107909

DOI: 10.1016/j.sigpro.2020.107909

https://www.sciencedirect.com/science/article/pii/S0165168420304539

Adaptive target detection in hyperspectral imaging from two sets of training samples with different means

Olivier Besson[a,*], François Vincent[a], Stefania Matteoli[b]

[a] ISAE-SUPAERO, 10 Avenue Edouard Belin, 31055 Toulouse, France.
[b] CNR-IEIIT, via Girolamo Caruso 16, Pisa, Italy.

Abstract

In this paper, we consider local detection of a target in hyperspectral imaging and we assume that the spectral signature of interest is buried in a background which follows an elliptically contoured distribution with unknown parameters. In order to infer the background parameters, two sets of training samples are available: one set, taken from pixels close to the pixel under test, shares the same mean and covariance while a second set of farther pixels shares the same covariance but has a different mean. When the whole data samples (pixel under test and training samples) follow a matrix-variate t distribution, the one-step generalized likelihood ratio test (GLRT) is derived in closed-form. It is shown that this GLRT coincides with that obtained under a Gaussian assumption and that it guarantees a constant false alarm rate. We also present a two-step GLRT where the mean and covariance of the background are estimated from the training samples only and then plugged in the GLRT based on the pixel under test only.

Keywords: Hyperspectral imaging, detection, generalized likelihood ratio test, Student distribution.

1. Introduction

Detecting the presence of a given spectral signature among the pixels of an hyperspectral image serves many purposes, including characterization of soils and vegetation, detection of man-made materials and vehicles, among others [1, 2]. The difficulty of this problem lies in the fact that the signature of interest (SoI) is buried in a background with partly unknown statistics. For instance, the distribution of the background itself is subject to debate. Even if the distribution is known, the parameters describing it (for instance mean and covariance matrix) are not known and must be estimated from the available data. Consequently, detection of the SoI in a pixel under test (PUT) entails using other pixels (so called training samples) to learn the background present in the PUT. This can be done at the global level, where all pixels of the image are used to infer the background statistics, or at a local level where only pixels in the vicinity of the PUT

*Corresponding author
Email addresses: olivier.besson@isae-supaero.fr (Olivier Besson), francois.vincent@isae-supaero.fr (François Vincent), stefania.matteoli@ieiit.cnr.it (Stefania Matteoli)

are used, on the rationale that the background there is more representative while the background in farther pixels may differ from that in PUT [3, 4, 5].

Concerning the choice of the background distribution, the Gaussian assumption prevails, probably due to the huge amount of methods that have been developed for previous applications, such as radar, and its mathematical tractability that enables straightforward derivations and analytical performance evaluations. Thereby, many target detection schemes can be used in this context, such as the adaptive matched filter (AMF) [6] or Kelly's detector [7], to name a few. They correspond to two different approaches to derive the generalized likelihood ratio test (GLRT). Kelly's detector is known as a one-step GLRT, as it has been derived directly from the joint distribution of both the PUT and the training samples, whereas the AMF is its two-step counterpart, namely derived from the PUT distribution, assuming that the background parameters are known and then replaced by their estimates from the training samples.

However, with real hyperspectral data, Gaussian distributions rarely occur, as has been reported in the literature [2, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16], leading to more realistic models. One of the most popular is the Elliptically Contoured (EC) t-distributed model that allows to extend the Gaussian distribution to a broader class of probability density functions (p.d.f.). Different detectors have been derived under this hypothesis, such as the EC-GLRT [17] or the EC-FTMF [18].

Nevertheless, in-depth analysis of real data reveals that these models are strongly related to the number of classes constituting the background (grass, roads, buildings, ...). Indeed, a more representative model for the background is to consider each class of the background as a given distribution with specific parameters. This so-called finite mixture models (FMMs) [10, 12] approach is usually exploited with Gaussian distributions [8, 10, 16, 19, 20], but some authors suggest using EC distributions to better fit the background behaviour of each class [21, 22, 10]. Each class belonging to a different region of the map, the background behaviour is changing with the position within the image. This nonstationarity has been noticed since a long time, when dealing with optical or Infra-Red (IR) images [23]. More precisely, it has been observed that the main difference between the classes is mostly contained in the mean of the distribution rather than in its covariance matrix. As a consequence, the mean of the background varies more rapidly than the covariance matrix along most optical, IR or hyperspectral images. Thereby, many authors derived target detectors under the assumption that only the closest pixels of the PUT are representative of the mean of the background, whereas the covariance matrix can be estimated using a larger area of secondary pixels [24, 25]. One of the most popular anomaly detection scheme, namely the RX detector has been initially derived under such an hypothesis [26].

The fact that the variation rate of the mean is the predominant nonstationarity effect leads to consider two different windows for the training samples to compute separately the mean and the covariance matrix [12, 27]. This situation is illustrated in Figure 1 where $\mathbf{y}$ is the PUT, with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. A first set of pixels $\mathbf{X}$ shares the same statistical properties while a second set of pixels $\mathbf{Z}$ has a mismatched mean, yet the same covariance, unlike in the usual framework where $\mathbf{X}$ only is considered but with a broader size. It has to be noticed that separating mean estimation from covariance estimation is somehow a sort of intuitive processing, where the data are first converted to zero-mean data (demeaning step), using only the more representative training samples, and then a zero-mean detection scheme is computed. The demeaning step just consists of a standard 2-D high-pass filtering where a local blurring of the image is removed [24, 25, 26]. After this demeaning step, the data are more likely to be Gaussian distributed because the main non-stationary parameter has been removed, so that conventional zero-mean detectors can be used. Unfortunately, this widespread and intuitive way to proceed exhibits differences with the correct GLRT formulation, as will be shown in this paper.

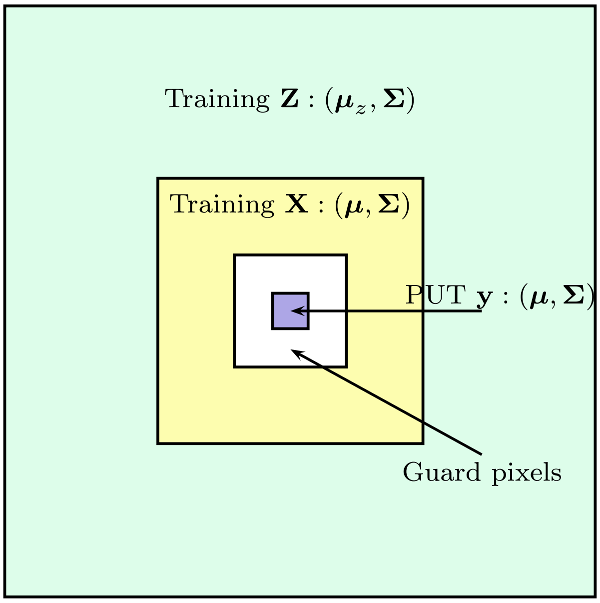

Figure 1: Pixel under test and training samples.

More precisely, in this paper we will derive both the one-step and two-step GLRT for detecting a known signature $\mathbf{t}$, when considering two sets of training samples $\mathbf{X}, \mathbf{Z}$ with the same covariance matrix, where only the closest one $\mathbf{X}$ shares the same mean as the PUT $\mathbf{y}$. We will consider EC t distributed background, the Gaussian p.d.f. being a special case. We will show that this GLRT formulation exhibits two main differences compared to the intuitive detectors consisting in demeaning and applying the standard zero-mean GLRT. These two differences concern the way the covariance matrix is computed and a different scaling factor in the GLR. The GLRT for the conventional case where $\boldsymbol{\mu} = \boldsymbol{\mu}_z$ will be obtained as a special case and also exhibits a difference in a scaling factor with the standard zero-mean detector. Moreover, we will show that the one-step GLRT coincides with its Gaussian counterpart, contrary to the two-step approach. Finally, we also show that these two GLRT possess the constant false alarm rate (CFAR) property.

The remaining of this paper is organized as follows. We first introduce the problem at hand in section 2. Then, we derive the one-step GLRT under t distributed background in Section 3 , we show that it coincides with its Gaussian counterpart and that it has the CFAR property. Moreover the standard case where all the training samples share the same mean is given as a special case. Similarly, the two-step GLRT is derived in section 4. These detectors are then compared to the GLRT assuming the same mean for all the training samples, using a real data benchmarking, in Section 5. Finally concluding remarks end this paper in Section 6.

2. Local target detection with two sets of t-distributed background samples with different means

As said in the introduction, we consider the problem of deciding whether a pixel under test $\mathbf{y}$ contains some spectral signature $\mathbf{t}$ when training samples $\mathbf{X} \in \mathbb{R}^{p\times n_x}$ and $\mathbf{Z} \in \mathbb{R}^{p\times n_z}$ are available to infer the background parameters. However, we assume that the training samples $\mathbf{Z}$, which are farther from the PUT than $\mathbf{X}$, do not share the same mean, as illustrated in Figure 1. Additionally, we assume that the data follow an elliptically contoured (EC) distribution, more precisely we assume a matrix-variate t-distribution (which we will sometimes referred to also as Student distribution) for the whole data matrix $\begin{bmatrix}\mathbf{y} & \mathbf{X} & \mathbf{Z}\end{bmatrix}$. The problem thus amounts to decide between H_0 and H_1 where the two hypotheses are given by

$$
\begin{aligned}
H_0 : \begin{bmatrix}\mathbf{y} & \mathbf{X} & \mathbf{Z}\end{bmatrix} &\overset{d}{=} \mathcal{T}_{p,n+1}\left(\nu, \mathbf{M}_0, \mathbf{\Sigma}, \mathbf{I}_{n+1}\right) \\
H_1 : \begin{bmatrix}\mathbf{y} & \mathbf{X} & \mathbf{Z}\end{bmatrix} &\overset{d}{=} \mathcal{T}_{p,n+1}\left(\nu, \mathbf{M}_1, \mathbf{\Sigma}, \mathbf{I}_{n+1}\right)
\end{aligned} \tag{1}
$$

with

$$
\begin{aligned}
\mathbf{M}_0 &= \begin{bmatrix}\boldsymbol{\mu} & \boldsymbol{\mu}\mathbf{1}_{n_x}^T & \boldsymbol{\mu}_z\mathbf{1}_{n_z}^T\end{bmatrix} \\
\mathbf{M}_1 &= \begin{bmatrix}\alpha\mathbf{t} + \boldsymbol{\mu} & \boldsymbol{\mu}\mathbf{1}_{n_x}^T & \boldsymbol{\mu}_z\mathbf{1}_{n_z}^T\end{bmatrix}
\end{aligned} \tag{2}
$$

where $n = n_x + n_z$ and $\mathbf{1}_q$ is the $q \times 1$ vector whose elements are all equal to 1. Below we derive the one-step and two-step generalized likelihood ratio test (GLRT) for this problem.

3. One-step GLRT

The generalized likelihood ratio based on $(\mathbf{y}, \mathbf{X}, \mathbf{Z})$ is given by

$$
\text{GLR} = \frac{\max\limits_{\alpha,\boldsymbol{\mu},\boldsymbol{\mu}_z,\mathbf{\Sigma}} p_1(\mathbf{y}, \mathbf{X}, \mathbf{Z}|\alpha, \boldsymbol{\mu}, \boldsymbol{\mu}_z, \mathbf{\Sigma})}{\max\limits_{\boldsymbol{\mu},\boldsymbol{\mu}_z,\mathbf{\Sigma}} p_0(\mathbf{y}, \mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\mu}_z, \mathbf{\Sigma})} \tag{3}
$$

where $p_i(.)$ stands for the probability density function of the whole observed data under H_i, $i = 0, 1$. The one-step GLRT thus consists in comparing GLR to a threshold. The next proposition gives the final expression of this GLR, once all maximization problems have been solved.

Proposition 1. *The GLR for the problem* (1) *is given by*

$$\mathrm{GLR}^{2/(n+1)} = \frac{1+\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})}{1+\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})-\frac{n_x}{n_x+1}\frac{[(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}\mathbf{t}]^2}{\mathbf{t}^T\mathbf{S}_{xz}^{-1}\mathbf{t}}} = \frac{1+\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})}{1+\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-T/2}\mathbf{P}^{\perp}_{\mathbf{S}_{xz}^{-1/2}\mathbf{t}}\mathbf{S}_{xz}^{-1/2}(\mathbf{y}-\bar{\mathbf{x}})} \tag{4}$$

where $\bar{\mathbf{x}} = n_x^{-1}\mathbf{X}\mathbf{1}_{n_x}$ *and* $\mathbf{S}_{xz} = \mathbf{X}\mathbf{P}^{\perp}_{n_x}\mathbf{X}^T+\mathbf{Z}\mathbf{P}^{\perp}_{n_z}\mathbf{Z}^T$ *with* $\mathbf{P}^{\perp}_q = \mathbf{I}_q - q^{-1}\mathbf{1}_q\mathbf{1}_q^T$ *the projector onto the subspace orthogonal to* $\mathbf{1}_q$.

Proof. See Appendix A. □

Rewriting this last expression as

$$\mathrm{GLR}^{2/(n+1)} = \frac{1}{1-\frac{\frac{n_x}{n_x+1}[(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}\mathbf{t}]^2}{[1+\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})](\mathbf{t}^T\mathbf{S}_{xz}^{-1}\mathbf{t})}}$$

where the second term of the denominator is shown to be the product of two positive terms less or equal than 1:

$$\frac{\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})}{[1+\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})]}\frac{[(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}\mathbf{t}]^2}{((\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}}))(\mathbf{t}^T\mathbf{S}_{xz}^{-1}\mathbf{t})}$$

Therefore, as $f(x) = \frac{1}{1-x}$ is an increasing function for $0 \leq x \leq 1$, the GLRT amounts to comparing the following test statistic

$$t(\mathbf{y},\mathbf{X},\mathbf{Z}) = \frac{\frac{n_x}{n_x+1}[(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}\mathbf{t}]^2}{[1+\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})][\mathbf{t}^T\mathbf{S}_{xz}^{-1}\mathbf{t}]} \tag{5}$$

to a threshold. This bears a strong resemblance with Kelly's detector except that here the primary data has been replaced by $\sqrt{\frac{n_x}{n_x+1}}(\mathbf{y}-\bar{\mathbf{x}})$ and that the sample covariance matrix of the training samples $(\mathbf{X},\mathbf{Z})$ has been computed after removal of their respective means. Thereby, as stated in the introduction, the GLRT is not exactly the intuitive detector consisting in a demeaning step followed by a zero-mean GLRT.

We now state some important properties of this GLR.

Proposition 2. *If* $\begin{bmatrix}\mathbf{y} & \mathbf{X} & \mathbf{Z}\end{bmatrix}$ *follows a Gaussian distribution, the GLR is still given by* (5).

Proof. See Appendix B for the technical proof. An intuitive way to figure out this equivalence is to realize that the expression of the GLR in (5) does not depend on ν and that, letting ν grow to infinity, one should recover the GLR for Gaussian distributed data. □

Proposition 3. *Under* H_0 *the distribution of* $t(\mathbf{y},\mathbf{X},\mathbf{Z})$ *does not depend on* $\boldsymbol{\mu}$, $\boldsymbol{\mu}_z$ *or* $\boldsymbol{\Sigma}$, *and thus the GLR has a constant false alarm rate (CFAR) with respect to these parameters.*

Proof. See Appendix C where we provide a stochastic representation of GLR for both Student and Gaussian distributions. Note however that the distribution of the GLR under H_0 depends on ν. □

The previous property is important as the threshold of the detector can be set irrespective of $\boldsymbol{\mu}$, $\boldsymbol{\mu}_z$ or $\boldsymbol{\Sigma}$ with a guarantee to have the same probability of false alarm whatever the values of these parameters.

A last comment concerns the usual case where all *training samples share the same average value*. The one-step GLRT in this case is obtained by replacing $\mathbf{X}$ by $\begin{bmatrix}\mathbf{X} & \mathbf{Z}\end{bmatrix}$ and by considering that $\mathbf{Z}$ does no longer exist. Doing so, one obtains the following test statistic, whether for Gaussian or Student distributions:

$$t'(\mathbf{y},\mathbf{X},\mathbf{Z}) = \frac{\frac{n}{n+1}[(\mathbf{y}-\overline{\mathbf{x}+\mathbf{z}})^T\mathbf{S}_{x+z}^{-1}\mathbf{t}]^2}{[1+\frac{n}{n+1}(\mathbf{y}-\overline{\mathbf{x}+\mathbf{z}})^T\mathbf{S}_{x+z}^{-1}(\mathbf{y}-\overline{\mathbf{x}+\mathbf{z}})][\mathbf{t}^T\mathbf{S}_{x+z}^{-1}\mathbf{t}]} \tag{6}$$

where $\overline{\mathbf{x}+\mathbf{z}} = n^{-1}\begin{bmatrix}\mathbf{X} & \mathbf{Z}\end{bmatrix}\mathbf{1}_n$ is simply the mean of the training samples and $\mathbf{S}_{x+z} = \begin{bmatrix}\mathbf{X} & \mathbf{Z}\end{bmatrix}\mathbf{P}_n^{\perp}\begin{bmatrix}\mathbf{X} & \mathbf{Z}\end{bmatrix}^T$ its standard sample covariance matrix.

This last expression corresponds to Kelly's detector in case of non-zero mean data. Again, we see that the popular Kelly's detector has to be slightly corrected by a different factor, namely $\frac{n}{n+1}$ in place of 1, when considering non-zero mean data. Moreover, as already noticed in [28], this expression is the same for both Gaussian and Student distributed background, giving it an optimality for a broader class of distributions than initially expected.

4. Two-step GLRT

We investigate here a *two-step procedure*, similarly to the AMF detector. First, let us assume that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known and let us consider the GLR for the problem

$$\begin{aligned} H_0 &: \mathbf{y} \overset{d}{=} \mathcal{T}_p(\nu, \boldsymbol{\mu}, (\nu-2)\boldsymbol{\Sigma}) \\ H_1 &: \mathbf{y} \overset{d}{=} \mathcal{T}_p(\nu, \alpha\mathbf{t}+\boldsymbol{\mu}, (\nu-2)\boldsymbol{\Sigma}) \end{aligned} \tag{7}$$

This problem has been solved in [17] where it is shown that

$$\mathrm{GLR}(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma}) \equiv \frac{[(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\mathbf{t}]^2}{[(\nu-2)+(\mathbf{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})][\mathbf{t}^T\boldsymbol{\Sigma}^{-1}\mathbf{t}]} \tag{8}$$

The second step consists in estimating $\boldsymbol{\mu}$, $\boldsymbol{\mu}_z$ and $\boldsymbol{\Sigma}$ from $(\mathbf{X},\mathbf{Z})$. These estimates, say $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, are then plugged in (8) in place of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to yield the two-step GLR. Again, we choose to estimate the unknown parameters using a maximum likelihood approach. Mimicking the derivations of Appendix A, it is straightforward to show that $\boldsymbol{\mu}_{\mathrm{ML}} = \bar{\mathbf{x}}$ and $\boldsymbol{\Sigma}_{\mathrm{ML}} = \frac{(\nu+p-1)}{(\nu-2)n}\mathbf{S}_{xz}$. Using these values in (8), one obtains the two-step GLR as

$$\mathrm{GLR}_{2s}(\mathbf{y},\mathbf{X},\mathbf{Z}) \equiv \frac{[(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}\mathbf{t}]^2}{[1+\frac{n_x+n_z}{\nu+p-1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})][\mathbf{t}^T\mathbf{S}_{xz}^{-1}\mathbf{t}]} \tag{9}$$

Note that the previous test statistic bears strong resemblance with its one-step counterpart in (5), they only differ by a scaling factor in one of the terms. This resemblance allows one to show, with the same derivations as in Appendix C, that the two-step GLRT is also CFAR with respect to $\boldsymbol{\mu}$, $\boldsymbol{\mu}_z$ and $\boldsymbol{\Sigma}$.

Before closing this section, we note that the two-step GLR obtained by assuming that $\begin{bmatrix}\mathbf{y} & \mathbf{X} & \mathbf{Z}\end{bmatrix}$ is Gaussian distributed is given by

$$\mathrm{GLR}_{2\mathrm{s}}^{\mathrm{G}}(\mathbf{y}, \mathbf{X}, \mathbf{Z}) \equiv \frac{[(\mathbf{y}-\bar{\mathbf{x}})^T \mathbf{S}_{xz}^{-1}\mathbf{t}]^2}{\mathbf{t}^T \mathbf{S}_{xz}^{-1}\mathbf{t}} \tag{10}$$

which can be derived either from the Gaussian matched filter or by letting ν grow to infinity in (9). Hence, in contrast to one-step GLRTs where the Student and Gaussian distributions lead to the same test statistics, the two-step GLRTs are different.

Table 1 summarizes the detectors available in the literature and those derived in the present paper, as a function of the scenario concerning the training samples and the data distribution. The color and line-type given between brackets refer to the plots of next section.

	Gaussian		Student	
	2-step GLRT ($\diamond$)	1-step GLRT (...)	1-step GLRT (...)	2-step GLRT (+)
$\boldsymbol{\mu} = \boldsymbol{\mu}_z$ (red)	AMF [6]	Eqn. (6)		EC-GLRT [17]
$\boldsymbol{\mu} \neq \boldsymbol{\mu}_z$ (blue)	Eqn. (10)	Eqn. (5)		Eqn. (9)

Table 1: Summary of detectors as a function of scenario (one set of training samples or two sets of training samples with different means) and distribution (Gaussian or Student).

5. Performance evaluation

In order to assess the benefits of considering two different training windows, we now conduct a Monte-Carlo simulation based on a real experiment, namely the airborne Viareggio 2013 trial [29]. This benchmarking hyperspectral detection campaign took place in Viareggio (Italy) in May 2013 with an aircraft flying at 1200 meters. The open data consist in a $[450 \times 375]$ pixels map composed of 511 samples in the Visible Near InfraRed (VINR) band $(400 - 1000nm)$. The spatial resolution of the image is about 0.6 meters.
Different kinds of vehicles as well as coloured panels served as known targets. For each of these targets, a spectral signature obtained from ground spectroradiometer measurements is available. Moreover, a black and a white cover, serving as calibration targets, were also deployed. As can be seen on Fig. 2, the scene is composed of parking lots, roads, buildings, sport fields and pine woods.

As for the majority of hyperspectral detection schemes, the first step of the processing aims at converting the raw measurements into a reflectance map, namely removing all atmospheric effects and non-uniform sun illumination. To this end, we use the Empirical Line Method (ELM) [30] [31], considering the black and white calibration panels. Then a spectral binning [32] is performed to reduce the vector size dimension to $N = 32$.

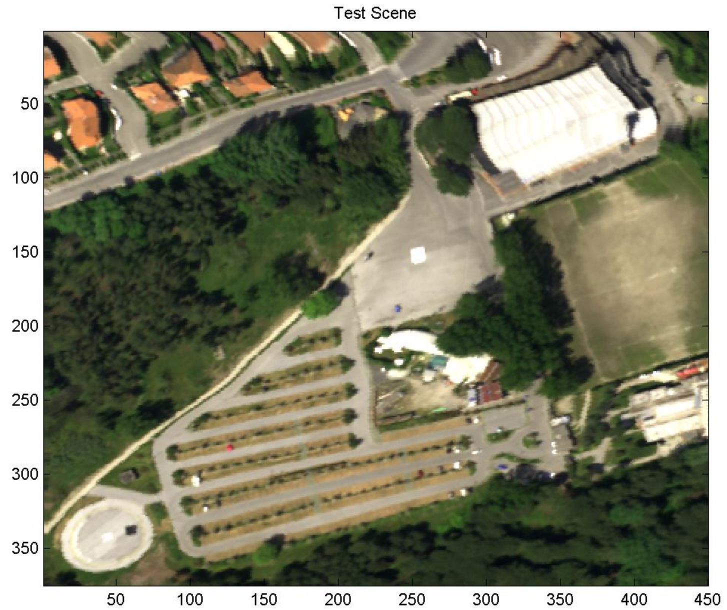

Figure 2: Complete RGB view of the Viareggio test scene

In order to obtain statistical results to compare the different detectors, we conduct a Monte-Carlo experiment where we randomly insert a target that does not initially exist in the map. For each target, we can then estimate a probability of detection P_d. The total image without target serves as reference to compute the probability of false alarm P_{fa}. Changing the threshold position, we can plot the so-called receiver operation characteristics (ROC) as represented on Figs. (3), (4), (5), (6) respectively for the so-called V_5, V_6 V_3 and P_2. In each case the target amplitude used in the simulation, α is indicated in the title of the plot.

For these four plots, we compare the three GLRT derived in this paper, for two sets of training samples, namely the one-step GLRT (eq. (5)) and the 2 versions of the two-step GLRT derived under Student and Gaussian distributions (eq. (9) and (10) resp.). In the case of the Student distribution, we have chosen $\nu = 3$ in order to have a large difference from the Gaussian distribution. In order to avoid problems related to ill-conditionning of the sample covariance matrix, we consider a large outer window of size 25×25 (green part in fig. (1)). By contrast, as the mean is supposed to move rapidly, we consider the smallest possible inner window (yellow part in fig. (1)), namely a 3×3 pixels window. This configuration corresponds to $n_x = 8$, $n_z = 616$ and $n = 624$. It can be noticed that no guard window is necessary in this experiment, as the target is only inserted in a single pixel. These 3 detectors assuming $\boldsymbol{\mu} \neq \boldsymbol{\mu}_z$ are also compared with those based on $\boldsymbol{\mu} = \boldsymbol{\mu}_z$ hypothesis. The $\boldsymbol{\mu} = \boldsymbol{\mu}_z$ cases are represented in red in the following curves, whereas the $\boldsymbol{\mu} \neq \boldsymbol{\mu}_z$ cases correspond to the blue plots. The line-styles are also indicated in table (1) for a better readability.

We can first observe that, for all the different targets, there is a noticeable improvement in considering two windows, namely assuming that the mean is more representative in the vicinity of the PUT. The gain can reach a P_{fa} reduction by 5 for a given P_d. Secondly, we can see that in both cases of single or two windows, the one-step and two-step GLRT exhibit approximately the same performance. A slightly better behavior for the two-step GLRT under Student hypothesis can be noticed for very low P_{fa}, suggesting that the background has a heavy tailed distribution on these real data.

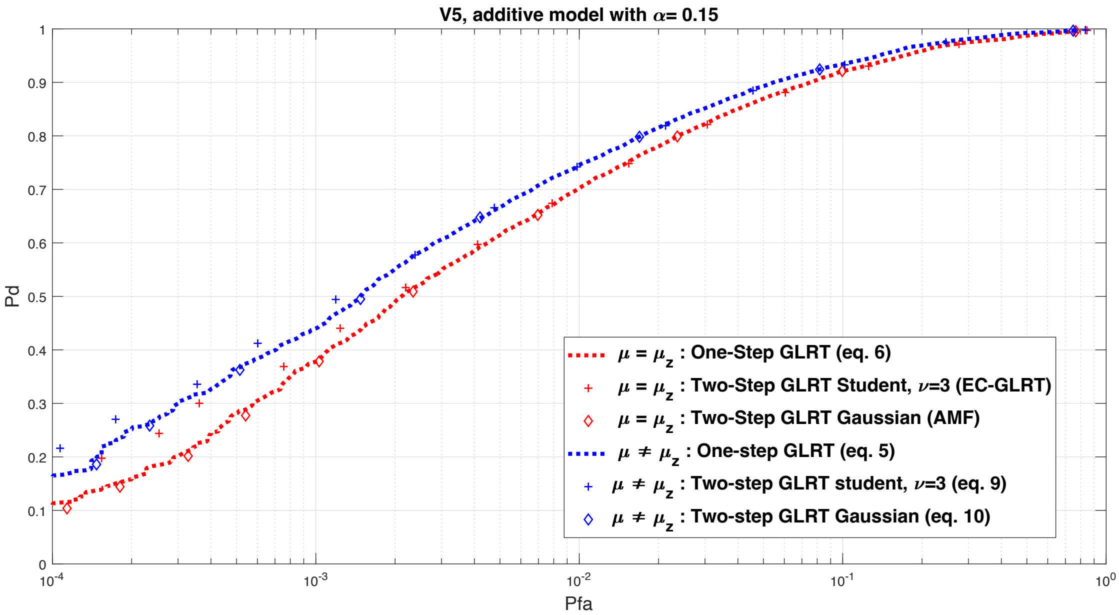

Figure 3: Receivers operation characteristics for V_5

6. Conclusions

In this paper, we considered target detection taking into account that mean is the main nonstationary parameter in hyperspectral imaging,. Rather than considering two different window sizes for mean and covariance estimation as is usually done, we addressed the problem under the more theoretical framework of generalized likelihood ratio test. We derived the one-step and two-step GLRT for the problem at hand, under EC t-distributed background, and showed some differences compared to usual, more intuitive techniques. Moreover, we showed that these GLRT posses the desirable CFAR property. Real data experiments illustrated the gain associated with the use of two training sample sets.

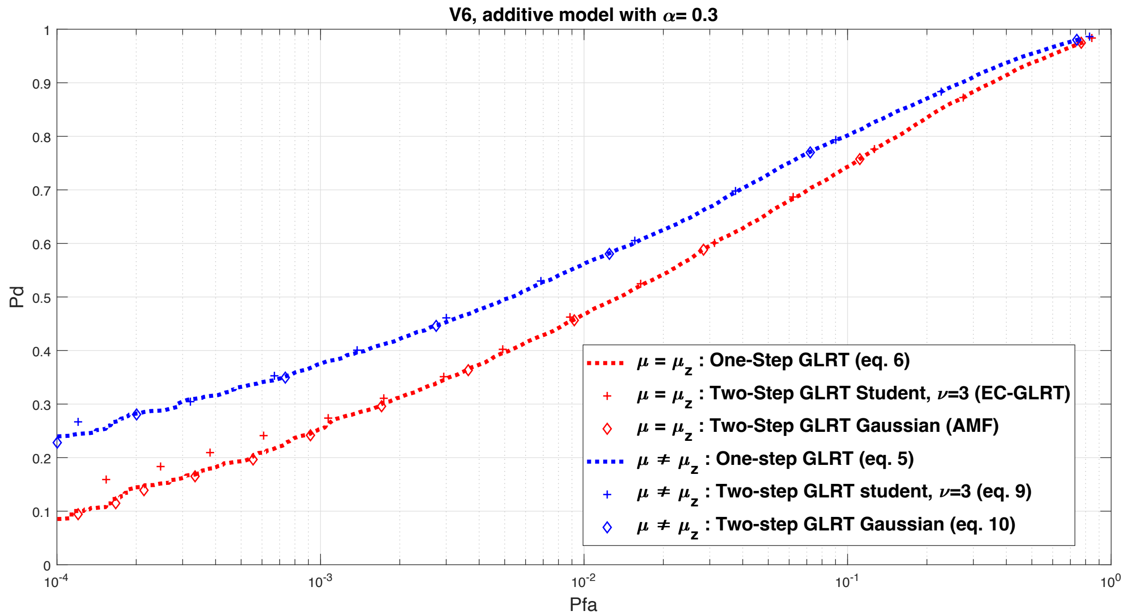

Figure 4: Receivers operation characteristics for V_6

Appendix A. Proof of Proposition 1

In this appendix, we derive the generalized likelihood ratio for the composite hypotheses testing problem in (1). For the sake of notational convenience, let us note $\mathbf{T} = \begin{bmatrix} \mathbf{y} & \mathbf{X} & \mathbf{Z} \end{bmatrix}$ the whole data matrix. The probability density function (p.d.f.) of $\mathbf{T}$ under each hypothesis is thus given by

$$p_0(\mathbf{T}) \propto |\mathbf{\Sigma}|^{-\frac{n+1}{2}} \left| \mathbf{I}_p + \frac{\mathbf{\Sigma}^{-1}}{\nu - 2} (\mathbf{T} - \mathbf{M}_0)(\mathbf{T} - \mathbf{M}_0)^T \right|^{-\frac{\nu+n+p}{2}}$$
$$p_1(\mathbf{T}) \propto |\mathbf{\Sigma}|^{-\frac{n+1}{2}} \left| \mathbf{I}_p + \frac{\mathbf{\Sigma}^{-1}}{\nu - 2} (\mathbf{T} - \mathbf{M}_1)(\mathbf{T} - \mathbf{M}_1)^T \right|^{-\frac{\nu+n+p}{2}} \tag{A.1}$$

where $\propto$ means proportional to. It can be readily verified that $|\mathbf{\Sigma}|^{-\frac{n+1}{2}} \left| \mathbf{I}_p + (\nu - 2)^{-1} \mathbf{\Sigma}^{-1} \mathbf{S} \right|^{-\frac{\nu+n+p}{2}}$ achieves its maximum at

$$\mathbf{\Sigma}_* = \frac{(\nu + p - 1)\mathbf{S}}{(\nu - 2)(n + 1)} \tag{A.2}$$

and is given by

$$\max_{\mathbf{\Sigma}} |\mathbf{\Sigma}|^{-\frac{n+1}{2}} \left| \mathbf{I}_p + \mathbf{\Sigma}^{-1} \mathbf{S} \right|^{-\frac{\nu+n+p}{2}} \propto |\mathbf{S}|^{-\frac{n+1}{2}} \tag{A.3}$$

It follows that

$$\max_{\mathbf{\Sigma}} p_0(\mathbf{T}) \propto \left| (\mathbf{T} - \mathbf{M}_0)(\mathbf{T} - \mathbf{M}_0)^T \right|^{-\frac{n+1}{2}}$$
$$\max_{\mathbf{\Sigma}} p_1(\mathbf{T}) \propto \left| (\mathbf{T} - \mathbf{M}_1)(\mathbf{T} - \mathbf{M}_1)^T \right|^{-\frac{n+1}{2}} \tag{A.4}$$

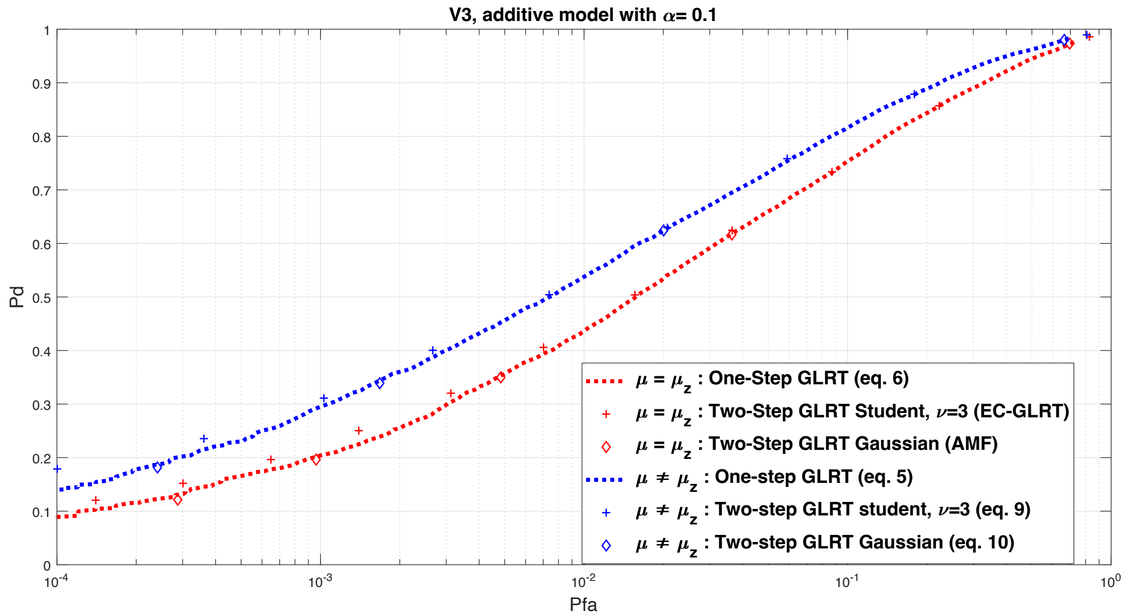

Figure 5: Receivers operation characteristics for V_3

Let $\mathbf{y}_i = \mathbf{y} - i\alpha\mathbf{t}$ for $i = 0, 1$ and note that

$$(\mathbf{T} - \mathbf{M}_i)(\mathbf{T} - \mathbf{M}_i)^T = (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^T + (\mathbf{X} - \boldsymbol{\mu}\mathbf{1}_{n_x}^T)(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}_{n_x}^T)^T + (\mathbf{Z} - \boldsymbol{\mu}_z\mathbf{1}_{n_z}^T)(\mathbf{Z} - \boldsymbol{\mu}_z\mathbf{1}_{n_z}^T)^T \tag{A.5}$$

It is straightforward to check that

$$(\mathbf{Z} - \boldsymbol{\mu}_z\mathbf{1}_{n_z}^T)(\mathbf{Z} - \boldsymbol{\mu}_z\mathbf{1}_{n_z}^T)^T = n_z(\boldsymbol{\mu}_z - \bar{\mathbf{z}})(\boldsymbol{\mu}_z - \bar{\mathbf{z}})^T + \mathbf{Z}\mathbf{P}_{n_z}^{\perp}\mathbf{Z}^T \tag{A.6}$$

where $\bar{\mathbf{z}} = n_z^{-1}\mathbf{Z}\mathbf{1}_{n_z}$ and $\mathbf{P}_q^{\perp} = \mathbf{I}_q - q^{-1}\mathbf{1}_q\mathbf{1}_q^T$ the orthogonal projector on the null space of $\mathbf{1}_q$. Similarly

$$\begin{aligned}
&(\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^T + (\mathbf{X} - \boldsymbol{\mu}\mathbf{1}_{n_x}^T)(\mathbf{X} - \boldsymbol{\mu}\mathbf{1}_{n_x}^T)^T \\
&= (n_x + 1)\left[\boldsymbol{\mu} - \frac{\mathbf{y}_i + \mathbf{X}\mathbf{1}_{n_x}}{n_x + 1}\right]\left[\boldsymbol{\mu} - \frac{\mathbf{y}_i + \mathbf{X}\mathbf{1}_{n_x}}{n_x + 1}\right]^T \\
&+ \begin{bmatrix}\mathbf{y}_i & \mathbf{X}\end{bmatrix}\mathbf{P}_{n_x+1}^{\perp}\begin{bmatrix}\mathbf{y}_i & \mathbf{X}\end{bmatrix}^T
\end{aligned} \tag{A.7}$$

Consequently, if we define $\mathbf{T}_i = \begin{bmatrix}\mathbf{y}_i & \mathbf{X} & \mathbf{Z}\end{bmatrix}$, then after maximization with respect to $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_z$, we have

$$\max_{\boldsymbol{\mu}, \boldsymbol{\mu}_z, \boldsymbol{\Sigma}} p_i(\mathbf{T}) \propto \left|\mathbf{T}_i \begin{pmatrix}\mathbf{P}_{n_x+1}^{\perp} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{n_z}^{\perp}\end{pmatrix} \mathbf{T}_i^T\right|^{-\frac{n+1}{2}} \tag{A.8}$$

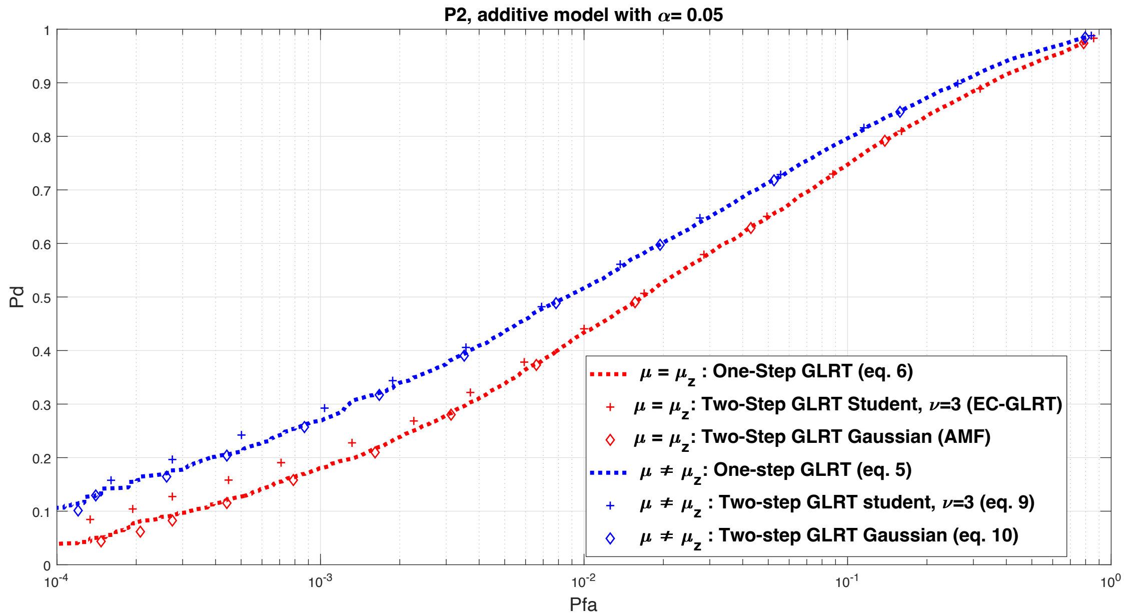

Figure 6: Receivers operation characteristics for P_2

Next, note that

$$\begin{aligned}\mathbf{Q} &= \begin{pmatrix}\mathbf{P}^{\perp}_{n_x+1} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}^{\perp}_{n_z}\end{pmatrix}\\ &= \begin{pmatrix}1-\frac{1}{n_x+1} & -\frac{\mathbf{1}^T_{n_x}}{n_x+1} & \mathbf{0}\\ -\frac{\mathbf{1}_{n_x}}{n_x+1} & \mathbf{I}_{n_x}-\frac{\mathbf{1}_{n_x}\mathbf{1}^T_{n_x}}{n_x+1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{P}^{\perp}_{n_z}\end{pmatrix}\\ &= \begin{pmatrix}\frac{n_x}{n_x+1} & -\frac{\mathbf{1}^T_{n_x}}{n_x+1} & \mathbf{0}\\ -\frac{\mathbf{1}_{n_x}}{n_x+1} & \mathbf{P}^{\perp}_{n_x}+\frac{\mathbf{1}_{n_x}\mathbf{1}^T_{n_x}}{n_x(n_x+1)} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{P}^{\perp}_{n_z}\end{pmatrix}\end{aligned} \tag{A.9}$$

so that

$$\begin{aligned}\mathbf{T}_i\mathbf{Q}\mathbf{T}_i^T &= \begin{bmatrix}\mathbf{y}_i & \mathbf{X} & \mathbf{Z}\end{bmatrix}\begin{pmatrix}\frac{n_x}{n_x+1} & -\frac{\mathbf{1}^T_{n_x}}{n_x+1} & \mathbf{0}\\ -\frac{\mathbf{1}_{n_x}}{n_x+1} & \mathbf{P}^{\perp}_{n_x}+\frac{\mathbf{1}_{n_x}\mathbf{1}^T_{n_x}}{n_x(n_x+1)} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{P}^{\perp}_{n_z}\end{pmatrix}\begin{bmatrix}\mathbf{y}_i^T\\ \mathbf{X}^T\\ \mathbf{Z}^T\end{bmatrix}\\ &= \frac{n_x}{n_x+1}\mathbf{y}_i\mathbf{y}_i^T - \frac{n_x}{n_x+1}\bar{\mathbf{x}}\mathbf{y}_i^T - \frac{n_x}{n_x+1}\mathbf{y}_i\bar{\mathbf{x}}^T\\ &+ \mathbf{X}\mathbf{P}^{\perp}_{n_x}\mathbf{X}^T + \frac{n_x}{n_x+1}\bar{\mathbf{x}}\bar{\mathbf{x}}^T + \mathbf{Z}\mathbf{P}^{\perp}_{n_z}\mathbf{Z}^T\\ &= \frac{n_x}{n_x+1}(\mathbf{y}_i-\bar{\mathbf{x}})(\mathbf{y}_i-\bar{\mathbf{x}})^T + \mathbf{S}_{xz}\end{aligned} \tag{A.10}$$

with $\bar{\mathbf{x}} = n_x^{-1}\mathbf{X}\mathbf{1}_{n_x}$ and $\mathbf{S}_{xz} = \mathbf{X}\mathbf{P}_{n_x}^{\perp}\mathbf{X}^T + \mathbf{Z}\mathbf{P}_{n_z}^{\perp}\mathbf{Z}^T$. It then follows that

$$\max_{\boldsymbol{\mu},\boldsymbol{\mu}_z,\boldsymbol{\Sigma}} p_i(\mathbf{T}) \propto |\mathbf{S}_{xz}|^{-\frac{n+1}{2}} \left[1 + \frac{n_x}{n_x+1}(\mathbf{y}_i - \bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}_i - \bar{\mathbf{x}})\right]^{-\frac{n+1}{2}} \tag{A.11}$$

and therefore

$$\begin{aligned}\mathrm{GLR}^{2/(n+1)} &= \frac{1 + \frac{n_x}{n_x+1}(\mathbf{y} - \bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y} - \bar{\mathbf{x}})}{1 + \frac{n_x}{n_x+1}\min_{\alpha}(\mathbf{y} - \bar{\mathbf{x}} - \alpha\mathbf{t})^T\mathbf{S}_{xz}^{-1}(\mathbf{y} - \bar{\mathbf{x}} - \alpha\mathbf{t})} \\ &= \frac{1 + \frac{n_x}{n_x+1}(\mathbf{y} - \bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y} - \bar{\mathbf{x}})}{1 + \frac{n_x}{n_x+1}(\mathbf{y} - \bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y} - \bar{\mathbf{x}}) - \frac{n_x}{n_x+1}\frac{[(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}\mathbf{t}]^2}{\mathbf{t}^T\mathbf{S}_{xz}^{-1}\mathbf{t}}}\end{aligned} \tag{A.12}$$

which concludes the proof.

Appendix B. One-step GLRT for Gaussian distributed background

In this appendix, we show that the the GLRT for Gaussian distributed background is the same as for Student distributed background. We thus consider the following detection problem

$$\begin{aligned} H_0 &: \mathbf{T} \overset{d}{=} \mathcal{N}_{p,n+1}\left(\mathbf{M}_0, \boldsymbol{\Sigma} \otimes \mathbf{I}_{n+1}\right) \\ H_1 &: \mathbf{T} \overset{d}{=} \mathcal{N}_{p,n+1}\left(\mathbf{M}_1, \boldsymbol{\Sigma} \otimes \mathbf{I}_{n+1}\right)\end{aligned} \tag{B.1}$$

The p.d.f. of $\mathbf{T}$ is in this case

$$\begin{aligned} p_0(\mathbf{T}) &\propto |\boldsymbol{\Sigma}|^{-\frac{n+1}{2}} \operatorname{etr}\left\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{T} - \mathbf{M}_0)(\mathbf{T} - \mathbf{M}_0)^T\right\} \\ p_1(\mathbf{T}) &\propto |\boldsymbol{\Sigma}|^{-\frac{n+1}{2}} \operatorname{etr}\left\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{T} - \mathbf{M}_1)(\mathbf{T} - \mathbf{M}_1)^T\right\}\end{aligned} \tag{B.2}$$

It is well-known that $|\boldsymbol{\Sigma}|^{-\frac{n+1}{2}} \operatorname{etr}\left\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{S}\right\}$ achieves its maximum at $\boldsymbol{\Sigma}_* = (n+1)^{-1}\mathbf{S}$, and hence

$$\max_{\boldsymbol{\Sigma}} |\boldsymbol{\Sigma}|^{-\frac{n+1}{2}} \operatorname{etr}\left\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{S}\right\} \propto |\mathbf{S}|^{-\frac{n+1}{2}} \tag{B.3}$$

It follows that

$$\begin{aligned} \max_{\boldsymbol{\Sigma}} p_0(\mathbf{T}) &\propto \left|(\mathbf{T} - \mathbf{M}_0)(\mathbf{T} - \mathbf{M}_0)^T\right|^{-\frac{n+1}{2}} \\ \max_{\boldsymbol{\Sigma}} p_1(\mathbf{T}) &\propto \left|(\mathbf{T} - \mathbf{M}_1)(\mathbf{T} - \mathbf{M}_1)^T\right|^{-\frac{n+1}{2}}\end{aligned} \tag{B.4}$$

But this is exactly (A.4) which holds for Student distributions. From there, everything follows and the GLRs for Student or Gaussian distributions are the same and are given by (4).

Appendix C. CFAR property of the GLRT

In this appendix, we show that the distribution of the GLR under H_0 does not depend on $\boldsymbol{\mu}$, $\boldsymbol{\mu}_z$ or $\boldsymbol{\Sigma}$. Let us first recall the expression of the test statistic

$$t = \frac{\frac{n_x}{n_x+1}[(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}\mathbf{t}]^2}{[1+\frac{n_x}{n_x+1}(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{S}_{xz}^{-1}(\mathbf{y}-\bar{\mathbf{x}})][\mathbf{t}^T\mathbf{S}_{xz}^{-1}\mathbf{t}]}$$

and let us rewrite $\mathbf{S}_{xz} = (\mathbf{X}\mathbf{H}_{n_x})(\mathbf{X}\mathbf{H}_{n_x})^T + (\mathbf{Z}\mathbf{H}_{n_z})(\mathbf{Z}\mathbf{H}_{n_z})^T$ where $\mathbf{H}_q$ is a $q|q-1$ matrix whose columns form an orthonormal basis for the hyperplane orthogonal to $\mathbf{1}_q$, i.e., $\mathbf{H}_q^T\mathbf{H}_q = \mathbf{I}_{q-1}$ and $\mathbf{P}_q^{\perp} = \mathbf{H}_q\mathbf{H}_q^T$. Let $\tilde{\mathbf{y}} = \sqrt{\frac{n_x}{n_x+1}}(\mathbf{y}-\bar{\mathbf{x}})$, $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{H}_{n_x}$ and $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{H}_{n_z}$ so that t can be rewritten as

$$t = \frac{[\tilde{\mathbf{y}}^T\tilde{\mathbf{S}}^{-1}\mathbf{t}]^2}{[1+\tilde{\mathbf{y}}^T\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{y}}][\mathbf{t}^T\tilde{\mathbf{S}}^{-1}\mathbf{t}]} \tag{C.1}$$

with $\tilde{\mathbf{S}} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$. Now, one can write that

$$\begin{aligned}\tilde{\mathbf{T}} &= \begin{bmatrix}\tilde{\mathbf{y}} & \tilde{\mathbf{X}} & \tilde{\mathbf{Z}}\end{bmatrix} = \begin{bmatrix}\sqrt{\frac{n_x}{n_x+1}}(\mathbf{y}-\bar{\mathbf{x}}) & \mathbf{X}\mathbf{H}_{n_x} & \mathbf{Z}\mathbf{H}_{n_z}\end{bmatrix} \\ &= \mathbf{T}\mathbf{A} \stackrel{d}{=} \mathcal{T}_{p,n-1}\left(\nu, \mathbf{M}_i\mathbf{A}, \boldsymbol{\Sigma}, \mathbf{A}^T\mathbf{A}\right)\end{aligned} \tag{C.2}$$

where

$$\mathbf{A} = \begin{pmatrix} \sqrt{\frac{n_x}{n_x+1}} & \mathbf{0} & \mathbf{0} \\ -\sqrt{\frac{n_x}{n_x+1}}n^{-1}\mathbf{1}_{n_x} & \mathbf{H}_{n_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{n_z} \end{pmatrix} \tag{C.3}$$

It can be easily verified that $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{n-1}$ and that

$$\mathbf{M}_i\mathbf{A} = \begin{bmatrix}\sqrt{\frac{n_x}{n_x+1}}i\alpha\mathbf{t} & \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{C.4}$$

which implies that

$$\tilde{\mathbf{T}} \stackrel{d}{=} \mathcal{T}_{p,n-1}\left(\nu, \begin{bmatrix}\sqrt{\frac{n_x}{n_x+1}}i\alpha\mathbf{t} & \mathbf{0} & \mathbf{0}\end{bmatrix}, \boldsymbol{\Sigma}, \mathbf{I}_{n-1}\right)$$

Next, let $\boldsymbol{\Sigma} = \mathbf{G}\mathbf{G}^T$ and let $\mathbf{U}$ be the unitary matrix such that $\mathbf{U}^T\mathbf{G}^{-1}\mathbf{t} = (\mathbf{t}^T\boldsymbol{\Sigma}^{-1}\mathbf{t})^{1/2}\mathbf{e}_1$ with $\mathbf{e}_1 = \begin{bmatrix}1 & 0 & \dots & 0\end{bmatrix}^T$. Let us make the change of variables

$$\begin{aligned}\breve{\mathbf{T}} &= \mathbf{U}^T\mathbf{G}^{-1}\tilde{\mathbf{T}} \\ &\stackrel{d}{=} \mathcal{T}_{p,n-1}\left(\nu, \begin{bmatrix}\sqrt{\frac{n_x}{n_x+1}}i\alpha(\mathbf{t}^T\boldsymbol{\Sigma}^{-1}\mathbf{t})^{1/2}\mathbf{e}_1 & \mathbf{0} & \mathbf{0}\end{bmatrix}, \mathbf{I}_p, \mathbf{I}_{n-1}\right)\end{aligned} \tag{C.5}$$

It follows that the distribution of $\breve{\mathbf{T}}$ does not depend on $\boldsymbol{\mu}$ or $\boldsymbol{\mu}_z$, and that, under H_0, it does neither depend on $\boldsymbol{\Sigma}$ since then $\alpha = 0$. Moreover, since $\tilde{\mathbf{y}} = \mathbf{G}\mathbf{U}\breve{\mathbf{y}}$, $\tilde{\mathbf{X}} = \mathbf{G}\mathbf{U}\breve{\mathbf{X}}$ and $\tilde{\mathbf{Z}} = \mathbf{G}\mathbf{U}\breve{\mathbf{Z}}$, it is readily verified that t can be written as

$$t = \frac{[\breve{\mathbf{y}}^T\breve{\mathbf{S}}^{-1}\mathbf{e}_1]^2}{[1+\breve{\mathbf{y}}^T\breve{\mathbf{S}}^{-1}\breve{\mathbf{y}}][\mathbf{e}_1^T\tilde{\mathbf{S}}^{-1}\mathbf{e}_1]} \tag{C.6}$$

where $\breve{\mathbf{S}} = \breve{\mathbf{X}}\breve{\mathbf{X}}^T + \breve{\mathbf{Z}}\breve{\mathbf{Z}}^T$. Therefore, the distribution of t under H_0 is independent of $\boldsymbol{\mu}$, $\boldsymbol{\mu}_z$ or $\boldsymbol{\Sigma}$, which proves the CFAR property of the GLRT. It means that the threshold of the detector can be set irrespective of $\boldsymbol{\mu}$, $\boldsymbol{\mu}_z$ and $\boldsymbol{\Sigma}$.

In the Gaussian case, derivations follow along the same lines. More precisely,

$$\begin{aligned} \mathbf{T} &\stackrel{d}{=} \mathcal{N}_{p,n+1}\left(\mathbf{M}_i, \boldsymbol{\Sigma} \otimes \mathbf{I}_{n+1}\right) \\ \Rightarrow \tilde{\mathbf{T}} = \mathbf{T}\mathbf{A} &\stackrel{d}{=} \mathcal{N}_{p,n-1}\left(\begin{bmatrix}\sqrt{\frac{n_x}{n_x+1}} i\alpha\mathbf{t} & \mathbf{0} & \mathbf{0}\end{bmatrix}, \boldsymbol{\Sigma} \otimes \mathbf{I}_{n-1}\right) \\ \Rightarrow \breve{\mathbf{T}} = \mathbf{U}^T\mathbf{G}^{-1}\tilde{\mathbf{T}} &\stackrel{d}{=} \mathcal{N}_{p,n-1}\left(\begin{bmatrix}\sqrt{\frac{n_x}{n_x+1}} i\alpha(\mathbf{t}^T\boldsymbol{\Sigma}^{-1}\mathbf{t})^{1/2}\mathbf{e}_1 & \mathbf{0} & \mathbf{0}\end{bmatrix}, \mathbf{I}_p \otimes \mathbf{I}_{n-1}\right) \end{aligned} \tag{C.7}$$

It follows that $\breve{\mathbf{y}}$ and $(\breve{\mathbf{X}}, \breve{\mathbf{Z}})$ are independent with

$$\begin{aligned} \breve{\mathbf{y}} &\stackrel{d}{=} \mathcal{N}_p\left(\sqrt{\frac{n_x}{n_x+1}} i\alpha(\mathbf{t}^T\boldsymbol{\Sigma}^{-1}\mathbf{t})^{1/2}\mathbf{e}_1, \mathbf{I}_p\right) \\ \breve{\mathbf{S}} &\stackrel{d}{=} \mathcal{W}_p\left(n-2, \mathbf{I}_p\right) \end{aligned} \tag{C.8}$$

Therefore Kelly's analysis directly applies to this detector. If η denotes the threshold then the probability of false alarm is given by $P_{\text{fa}} = (1-\eta)^{n-2-p+1}$. The probability of detection depends only on the signal to noise ratio which is now defined as $\frac{n_x}{n_x+1}\alpha^2\mathbf{t}^T\boldsymbol{\Sigma}^{-1}\mathbf{t}$.

[1] M. T. Eismann, Hyperspectral remote sensing, SPIE, 2012.

[2] D. G. Manolakis, R. B. Lockwood, T. W. Cooley, Hyperspectral Imaging Remote Sensing, Cambridge University Press, 2016.

[3] D. Manolakis, G. Shaw, Detection algorithms for hyperspectral imaging applications, IEEE Signal Processing Magazine 19 (1) (2002) 29–43.

[4] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, M. Brueggeman, Detection Algorithms in Hyperspectral Imaging Systems: An Overview of Practical Algorithms, IEEE Signal Processing Magazine 31 (1) (2014) 24–33.

[5] N. M. Nasrabadi, Hyperspectral Target Detection : An Overview of Current and Future Challenges, IEEE Signal Processing Magazine 31 (1) (2014) 34–44.

[6] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, R. Nitzberg, A CFAR Adaptive Matched Filter Detector, IEEE Transactions Aerospace Electronic Systems 28 (1) (1992) 208–216.

[7] E. Kelly, An Adaptive Detection Algorithm, IEEE Transactions Aerospace Electronic Systems 22 (2) (1986) 115–127.

[8] M. J. Carlotto, A cluster-based approach for detecting man-made objects and changes in imagery, IEEE Transactions Geoscience Remote Sensing 43 (2) (2005) 374387.

[9] D. B. Marden, D. Manolakis, Using elliptically contoured distributions to model hyperspectral imaging data, in: Proceedings of SPIE - The International Society for Optical Engineering, vol. 5425, 558572, 2004.

[10] S. Matteoli, T. Veracini, M. Diani, G. Corsini, Models and methods for automated background density estimation in hyperspectral anomaly detection, IEEE Transactions Geoscience Remote Sensing 51 (5) (2013) 28372852.

[11] D. G. Manolakis, D. Marden, J. P. Kerekes, G. A. Shaw, Statistics of hyperspectral imaging data, in: Proceedings SPIE 4381 Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII, vol. 4381, 308–316, 2001.

[12] S. Matteoli, M. Diani, J. Theiler, An Overview of Background Modeling for Detection of Targets and Anomalies in Hyperspectral Remotely Sensed Imagery, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (6) (2014) 2317–2336.

[13] J. S. Tyo, J. Robertson, J. Wollenbecker, R. C. Olsen, Statistics of Target Spectra in HSI Scenes, in: Proceedings of SPIE - The International Society for Optical Engineering, 2000.

[14] S. Niu, V. K. Ingle, D. G. Manolakis, T. W. Cooley, Tests for the elliptical symmetry of hyperspectral imaging data, in: Proceedings of SPIE - The International Society for Optical Engineering, vol. 7812, 2010.

[15] D. Manolakis, D. Marden, G. Shaw, Hyperspectral image processing for automatic target detection applications, Lincoln Laboratory Journal 14 (1) (2003) 79–116.

[16] S. Matteoli, M. Diani, G. Corsini, A Tutorial Overview of Anomaly Detection in Hyperspectral Images, IEEE Aerospace Electronics Systems Magazine 25 (7) (2010) 5–27.

[17] J. Theiler, B. R. Foy, EC-GLRT: Detecting Weak Plumes in Non-Gaussian Hyperspectral Clutter Using an Elliptically-Contoured Generalized Likelihood Ratio Test, in: Proceedings IGARSS, vol. 1, Boston, MA, 221–224, 2008.

[18] J. Theiler, B. Zimmer, A. K. Ziemann, Closed-form detector for solid sub-pixel targets in multivariate T-distributed background clutter, in: Proceedings IGARSS, Valencia, Spain, 2773–2776, 2018.

[19] C. C. Funk, J. Theiler, D. A. Roberts, C. C. Borel, Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery, IEEE Transactions Geoscience Remote Sensing 39 (7) (2001) 1410–1420.

[20] S. Matteoli, M. Diani, G. Corsini, Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images, Optical Engineering 49 (4) (2010) 046201–046215.

[21] N. Acito, G. Corsini, M. Diani, Statistical analysis of hyper-spectral data: A non-Gaussian approach, EURASIP J. Adv. Signal Process. 2007.

[22] D. B. Marden, D. G. Manolakis, Modeling hyperspectral imaging data, in: Proceedings SPIE 5093 Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX, vol. 5093, 253–262, 2003.

[23] B. R. Hunt, T. M. Cannon, Nonstationary Assumptions for Gaussian Models of Images, IEEE Transactions on Systems, Man, and Cybernetics (1976) 876–882.

[24] A. Margalit, I. Reed, R. Gagliardi, Adaptive Optical Target Detection Using Correlated Images, IEEE Transactions Aerospace Electronic Systems 21 (3) (1985) 394–405.

[25] J. Chen, I. Reed, A Detection Algorithm for Optical Targets in Clutter, IEEE Transactions Aerospace Electronic Systems 23 (1) (1985) 46–59.

[26] I. Reed, X. Yu, Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution, IEEE Transactions Acoustics Speech Signal Processing 38 (10) (1990) 1760–1770.

[27] D. Borghys, I. Kasen, V. Achard, C. Perneel, Hyperspectral Anomaly Detection: Comparative Evaluation in Scenes with Diverse Complexity, Journal of Electrical and Computer Engineering, Hindawi Publishing Corporation 2012.

[28] O. Besson, F. Vincent, Sub-pixel detection in hyperspectral imaging with elliptically contoured t-distributed background, arXiv, 2003.11780, 2020.

[29] N. Acito, S. Matteoli, A. Rossi, M. Diani, G. Corsini, Hyperspectral Airborne Viareggio 2013 Trial Data Collection for Detection Algorithm Assessment, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (6) (2016) 2356–2376.

[30] G. Ferrier, Evaluation of apparent surface reflectance estimation methodologies, International Journal of Remote Sensing 16 (1995) 2291–2297.

[31] G. M. Smith, E. J. Milton, The use of the empirical line method to calibrate remotely sensed data to reflectance, International Journal of Remote Sensing 20 (1999) 2653–2662.

[32] M. Shi, G. Healey, Hyperspectral Texture Recognition Using a Multiscale Opponent Representation, IEEE Transactions Geoscience Remote Sensing 41 (5) (2003) 1090–1095.