Unsupervised Anomaly and Change Detection With Multivariate Gaussianization
José A. Padrón-Hidalgo©, Valero Laparra©, and Gustau Camps-Valls©, Fellow, IEEE

Abstract—Anomaly detection (AD) is a field of intense research in remote sensing (RS) image processing. Identifying low probability events in RS images is a challenging problem given the high dimensionality of the data, especially when no (or little) information about the anomaly is available a priori. While a plenty of methods are available, the vast majority of them do not scale well to large datasets and require the choice of some (very often critical) hyperparameters. Therefore, unsupervised and computationally efficient detection methods become strictly necessary, especially now with the data deluge problem. In this article, we propose an unsupervised method for detecting anomalies and changes in RS images by means of a multivariate Gaussianization methodology that allows to estimate multivariate densities accurately, a long-standing problem in statistics, and machine learning. The methodology transforms arbitrarily complex multivariate data into a multivariate Gaussian distribution. Since the transformation is differentiable, by applying the change of variables formula, one can estimate the probability at any point of the original domain. The assumption is straightforward: pixels with low estimated probability are considered anomalies. Our method can describe any multivariate distribution, makes an efficient use of memory and computational resources, and is parameter-free. We show the efficiency of the method in experiments involving both AD and change detection (CD) in different RS image sets. For AD, we propose two approaches. The first is using directly the Gaussianization transform and the second is using a hybrid model that combines Gaussianization and the Reed–Xiaoli (RX) method typically used in AD. For CD, we take advantage of the Gaussianization transform and attribute the change to pixels with low probability compared to the first image, instead of those with high difference value typically employed in RS. Results show that our approach outperforms other linear and nonlinear methods in terms of detection power in both anomaly and CD scenarios, showing robustness and scalability to dimensionality and sample sizes.

Index Terms—Anomaly detection (AD), change detection (CD), deep learning, extremes, Gaussianization, information, principal component analysis, probability density estimation.

Manuscript received November 14, 2020; revised April 7, 2021 and September 10, 2021; accepted September 16, 2021. Date of publication October 12, 2021; date of current version January 26, 2022. This work was supported in part by the European Research Council (ERC) through the ERC-CoG-2014 SEDAL Project under Grant 647423 and in part by the Spanish Ministry of Economy, Industry and Competitiveness through the Network of Excellence Program under Grant TEC2016-81900-REDT, Grant TEC2016-77741-R, Grant DPI2017-89867-C2-2-R, Grant RTI2018-096765- A-I00, Grant PID2019-109026RB-I00, and Grant PID2020-118071GB-I00. The work of José A. Padrón-Hidalgo was supported by the Grisolía Grant from Generalitat Valenciana (GVA) under Grant GRISOLIA/2016/100. (Corresponding author: José A. Padrón-Hidalgo.)

The authors are with the Image Processing Laboratory (IPL), Universitat de València, 46980 Valencia, Spain (e-mail: joseanthoniopadronhidalg@hotmail.com; gustau.camps@uv.es). This article has supplementary downloadable material available at https://doi.org/10.1109/TGRS.2021.3116186, provided by the authors. Digital Object Identifier 10.1109/TGRS.2021.3116186

I. INTRODUCTION

Remote sensing (RS) has become a powerful tool to develop applications for Earth monitoring [1]–[3]. Earth observation (EO) satellite missions, such as Sentinels-2 and Landsat-8, are able to replace the hard and costly work of the man on the ground. Also, the use of very high-resolution (VHR) satellite imagery (e.g., QuickBird and the Worldview constellation) is becoming increasingly important for RS applications, and it makes possible the detection of dangerous events, such as extreme precipitations, heat waves, latent fires, droughts, floods, or urbanization. The vast amount of data available from different sensors makes it urgent to have automatic methods to detect these events. A good and quite standard approach nowadays to tackle this problem considers statistical models that allow us to detect anomalies and changes on the Earth cover.

Statistical methods for anomaly detection (AD) focus on detecting small portions of the image, which does not belong to the background of the scene [5]. Anomalies are considered a group of weird (low probability) pixels that significantly differ from their neighbors. AD is a challenging task and many variants have been proposed in the literature, such as neighbor based, clustering, and classification [6]–[11]. However, among all of them, the Reed–Xiaoli (RX) approach [12] is still the most widely used method for AD since the Gaussian distribution assumption is a reasonable approach in several cases, and it is unsupervised, fast, and easy to implement. The RX method allows us to detect the anomalous samples compared to the background using the well-known Mahalanobis distance. RX can be used as either a local or global approach [13]. In this study, all the methods are used in a global setting for the sake of simplicity. However, the proposed method is general enough and could be easily adapted to both local and dual-window approaches. Since the Gaussian assumption is not flexible enough in most cases, variants of the RX have been developed to cope with higher order feature relations. One option that obtains good results is based on the theory of reproducing kernels in Hilbert spaces, which extends the RX approach to the kernel RX (KRX) [14]–[16]. However, the KRX algorithm has not been widely adopted in practice because, being a kernel method, the memory and computational cost increase with the number of pixels cubically and quadratically, respectively, and more importantly, the selection of the kernel parameters is critical to achieve a good performance. While unsupervised approaches to fit the kernel parameter exist, they achieve a suboptimal performance,
and hence, supervised approaches have to be used. In this article, we propose to use a different, more straightforward approach to the problem of AD based on multivariate Gaussianization transformation. The proposed method is able to provide an estimation of the probability of multidimensional data. The basic idea is simple: the method finds an invertible and differentiable transformation that converts the multivariate distribution of the input data into a multivariate Gaussian and applies the change of variable formula under transformations to estimate the density in the input domain. Therefore, the probability of each data point in the original domain can be estimated by using two ingredients: the probability of the data in the transform domain and the determinant of the Jacobian of the transformation applied to the point. It is important to stress the fact that both ingredients are needed to estimate the density in the input domain, that is, the method is not transforming the data to a Gaussian domain to estimate the probability there, but uses the Gaussian domain as a convenient intermediate stage to estimate the density in the original one. Actually, the probability in the Gaussian domain must be multiplied by the determinant of the Jacobian. As we will see, the operations involved to estimate these transformations are easy and fast, the method scales well with data dimensionality and does not require additional information to fit any parameter [17], [18]. The parameters required for rotation-based iterative Gaussianization (RBIG) computation are the ones to fit the marginal density estimation. In our case, we use the simplest case based on computing histograms with the default parameters in the toolbox (see [4]) and no extra fitting is used in any case. Note that the proposed method estimates the data density using all available data. In the AD setting, a naive application would incidentally use the anomalies too, which is not desirable. To address this issue, we propose a two-step procedure that finds a tradeoff between the Gaussian assumption (very rigid) aimed to discard the potential anomalies, and the proposed method (very flexible) aimed to better estimate the background density. This two-step approach improves the results of the two components when used separately and yields more robust density estimates, which ultimately leads to a powerful, automatic, unsupervised, algorithm for AD.

Detecting changes in images automatically is extremely important because it allows us to improve predictions and our understanding of events. The usual approach to change detection (CD) in RS is based on arithmetic operations over the image before and after the possible change [9], [19]. The most usual approach to detect changes is based on thresholding or classifying the difference image. This approach relies on previous (and quite critical) stages, such as atmospheric correction or image coregistration. Here, we instead illustrate our method on the CD approach of finding statistical (probabilistic) differences between the signals before and after the event [20]. Other settings for CD with RBIG could be defined by exploiting differences, ratios, or other convenient transformations of the pair of images. Our approach can be considered as a particular case of the AD problem, where the change class is the target class to be detected.

As for AD, a similar problem occurs when using statistical methods for CD [19], [21]: one aims to learn the distribution of the original image and analyze the statistical differences of the pixels in the new incoming image. Likewise, the RX and KRX extensions have been proposed to deal with CD problems. However, they show the same drawbacks as in AD. Here, we describe how the multivariate Gaussianization can also be used in CD problems. Note that in both cases, AD and CD, the proposed statistical method is used to evaluate the pixel’s probability so that one can classify them as anomalies or changes, respectively.

The remainder of this article is organized as follows. Section II summarizes the Gaussianization transformation in general and how to adapt it to anomaly and CD. In Section III, we illustrate its performance in three experiments, involving simulated anomalies, real AD, and CD examples with a database of multispectral and hyperspectral images. Results show that the proposed approach is robust and flexible enough to be applied in different AD and CD scenarios and obtains better performance than other simple and robust methods (such as the RX) and more flexible and adaptable ones, such as the KRX. Section IV concludes this article with some remarks and further work.
II. Multivariate Gaussianization for Detection

In this section, we use the RBIG, a nonparametric method for density estimation of multivariate distributions proposed in [17] and [18]. RBIG is rooted in the idea of Gaussianization, introduced in the seminal work in [22] and further developed in [23], which consists of seeking for a transformation $G_x$ that converts a multivariate dataset $X \in \mathbb{R}^{d \times d}$ in domain $X$ to a domain where the mapped data $Y \in \mathbb{R}^{d \times d}$ follow a multivariate normal distribution in domain $Y$, i.e., $p_Y(y) \sim \mathcal{N}(0, I)$:

$$G_x : x \in \mathbb{R}^d \mapsto y \in \mathbb{R}^d$$

$$\sim p_X(x)p_Y(y) \sim \mathcal{N}(0, I) \quad (1)$$

where inputs and mapped data points have the same dimensionality $x, y \in \mathbb{R}^d$, $0$ is a vector of zeros (for the means), and $I_d$ is the identity matrix for the covariance of dimension $d$. Using the change of variable formula, one can estimate the probability of a point $x$ in the original domain

$$p_X(x) = p_Y(y)|J_{G_x}(y)|$$

where $p_X(x)$ is the probability distribution of the original data point $x$, $p_Y(y)$ is the probability of the data point transformed to the Gaussian domain (note that by definition, $p_Y \sim \mathcal{N}(0, I_d)$ and can be easily computed), and $|J_{G_x}(y)|$ is the determinant of the Jacobian of the transformation $G_x$ in the point $y$.

It is important to stress the important role of the term $|J_{G_x}(y)|$. Note that one would be tempted to see the Gaussianization as a sort of preprocessing step that converts data to a convenient (Gaussian) domain where one could then apply methods that assume the Gaussian distribution. This is, however, incorrect since the term $|J_{G_x}(y)|$ has to be included in the density estimation, and it is computed for each data point. For this formula to work, $G_x$ has to be differentiable, i.e., $|J_{G_x}(y)| > 0, \forall y$. The Gaussianization method we propose in this article, RBIG, obtains a transformation $G_x$ that fulfills this property, cf. [17]. Therefore, RBIG can be easily applied to estimate the probability of data points in the original domain, $p_X(x)$.

RBIG is an iterative algorithm, where in each iteration, $n$, two steps are applied: 1) a set of $d$ marginal Gaussianizations to each of the variables, $\Psi = [\Phi_1, \ldots, \Phi_d]$, and 2) a linear rotation, $R \in \mathbb{R}^{d \times d}$.

$$x[i + 1] = R[i] \cdot \Psi[i](x[i]), \quad i = 1, \ldots, M \quad (3)$$

where $M$ is the number of steps (iterations) in the sequence, $i = 1, \ldots, M$. The final transformation $G_x$ is the composition of all performed transformations through iterations. This procedure is illustrated with a toy example in Fig. 1. In [17], we showed that with enough iterations $M$, the method converges and the transformed data ultimately follows a standardized Gaussian, i.e., $p_Y(y) \sim \mathcal{N}(0, I_d)$, taking $y = x[M]$.

The computational load can be divided into two steps: the first one when the method is trained on a representative training set and the second when is applied to unseen test dataset. On the one hand, the training time will depend on the number of iterations and the employed type of linear transformation (see [17] for a discussion on the different options). A reasonable number of iterations cannot be set beforehand since it will depend on the nongaussianity of the data. However, it can be automatically found with the stopping criteria proposed in [17]. On the other hand, when applying the method, the computation of the Jacobian will depend only on the number of iterations. It is important to note that, since the linear transformations are restricted to be rotations, the determinant of the Jacobian for them is always equal to 1. Therefore, the only part to be computed is the Jacobian of the marginal transformation, which is equal to the probability of the data before applying the transformation [17]. Therefore, the application of the method is in principle computationally easier than the training step. However, note that if one wants to obtain the full Jacobian, not only the determinant as it is needed to compute the probability, the amount of data to be store is huge, which can make the method difficult to apply. The use of RBIG in the framework of normalizing flows or density destructors can be useful in this sense since it is devoted to find transformation with easily computable Jacobians [24]. An implementation of RBIG in this framework can be found in [4].

An illustration of how RBIG can be adapted to describe the distribution of RS data is shown in Fig. 2. In this example, we take data from the Sentinel-2 image Australia (see Table I for details), which has $d = 12$ bands, and use RBIG to Gaussianize its pixel’s distribution. We can see that the Gaussianized data follow a Gaussian distribution. Besides, we apply the inverse of the learned Gaussianization transformation to randomly generated Gaussian points obtaining synthetic new data that follow a deemed similar distribution as the original one. This illustrates the invertibility property of RBIG, which allows us to estimate densities in the original domain and use the well-known relation between probability and anomaly to derive unsupervised density-based anomaly and change detectors. Now, we describe how Gaussianization can be applied to both AD and CD problems. The pointwise density estimate can be trivially linked to a degree of anomaly in a particular image; the higher the probability the lower the anomalousness. Likewise, the pointwise density estimate can be used to assess the probability of change between images; one uses the first image to estimate the background and applies the transform to the second image.

A. RBIG for Detection of Anomalies

One of the most successful methods applied to the problem of AD is the RX method [12], a successful type of matched filter. The idea behind the RX method can be interpreted in probabilistic terms [15]: intuitively, a data point is more anomalous when it has less probability to appear in probabilistic terms [15]; actually, this relation defines the RX method anomaly detector, i.e., $A(x) \sim \mathcal{N}(\mu, \Sigma)$. Actually, $A_{RX}(x)$ is equivalent to the Mahalanobis distance between the data point and the mean, i.e., $A_{RX}(x) = (x - \mu)^T \Sigma^{-1} (x - \mu)$, where $p(x) \sim \mathcal{N}(0, \Sigma)$.

While RX has been widely used, it has the limitations inherent to the Gaussian distribution assumption. The use
of kernel methods has been proposed to generalize the RX method to the nonlinear and non-Gaussian case [14], [15]. Conceptually, the KRX method can be understood as a way to propose a Gaussian distribution where the covariance is defined in a higher dimensional Hilbert feature space. Taking as a reference the RX method, this translates into replacing the covariance matrix with a kernel matrix that estimates the similarities between samples [25], [26]. In practice, this implies that correlation is substituted by a nonlinear (kernel) similarity measure. The anomaly detected using the KRX method can be formulated as

$$A_{\text{KRX}}(x) \propto \frac{1}{p_K(x)}$$

(5)

where $p_K(x)$ is the distribution induced by using the kernel function instead of the covariance. The KRX is an elegant extension of the RX, yet it has the problem of fitting kernel parameters and the high computational cost (as one has to invert a kernel matrix, which has a cubic cost with the number of points $\ell$). While some heuristics exist in the literature to fit the kernel parameters, in practice, one only achieves the full potential of the KRX approach by fitting the parameters after cross validation [15]. This requires having access to labeled data as anomalous versus nonanomalous classes, which is not a very realistic and not even practical setting. In this work, we approach the more useful and practical, yet more challenging, problem of unsupervised AD (i.e., no labeled data available), and therefore, in our comparisons, we will fit the kernel method parameter using the most successful (and sensible) heuristic to set the Gaussian kernel lengthscale $\sigma$ as the average of all distances among $X$.

As an alternative to linear measures of anomalousness like in RX or nonlinear yet implicit feature transformations with parameters to tune like in KRX, we here propose a more straightforward approach to estimate the probability density function with RBIG (Section II). Here, we stick to the idea of assuming a Gaussian distribution in a transform domain. In particular, RBIG forces the true distribution in the transform domain to be a multivariate Gaussian. However, note that the distribution in the original domain is estimated not only assuming Gaussian distribution in the transform domain but also including the determinant of the Jacobian of the transformation [see (2)]. This gives us a nonparametric, parameter-free, and efficient estimation of the data distribution in the original domain. RBIG is an unsupervised method by construction, so it does not require labeled data and scales linearly with the data. By using RBIG to compute $p_X$, we obtain the method proposed in this work

$$A_{\text{RBIG}}(x) \propto \frac{1}{p_{\text{RBIG}}(x)}$$

(6)

An important aspect to consider is the intrinsic characteristics of the data used to estimate the density, which has implications in the quality of the estimation. When the distribution contains even a moderate number of anomalies, an accurate density estimate will cast anomalies as regular points, i.e., nonanomalous. This vastly depends on the flexibility of the class of models used. When the model is rigid like in the RX case, this is not a problem since it cannot be adapted to the anomalies. For the KRX, one can control this effect by tuning the kernel lengthscale and the regularization term, but as explained before requires labeled data. This is an important aspect to consider mostly in the AD scenario, where all data (included the anomalous samples) are used to estimate the density. Therefore, we propose to use a hybrid model that combines the (too rigid) RX model with the (too flexible) RBIG model. First, we apply RX to discard data more likely to be anomalous. The remaining points are then used to apply RBIG. This tries to avoid using anomalous data in RBIG, which after all is intended to learn the background or pervasive data distribution. The number of data points selected as nonanomalous is an accurate density estimate will cast anomalies as regular points, i.e., nonanomalous. This vastly depends on the flexibility of the class of models used. When the model is rigid like in the RX case, this is not a problem since it cannot be adapted to the anomalies. For the KRX, one can control this effect by tuning the kernel lengthscale and the regularization term, but as explained before requires labeled data. This is an important aspect to consider mostly in the AD scenario, where all data (included the anomalous samples) are used to estimate the density. Therefore, we propose to use a hybrid model that combines the (too rigid) RX model with the (too flexible) RBIG model. First, we apply RX to discard data more likely to be anomalous. The remaining points are then used to apply RBIG. This tries to avoid using anomalous data in RBIG, which after all is intended to learn the background or pervasive data distribution. The number of data points selected as nonanomalous.

### B. RBIG for CD

RS CD algorithms have mainly relied on thresholding or classifying the difference (or ratio) between a pair of previously coregistered images [9]. Multitemporal registration and radiometric and atmospheric corrections are thus very important aspects that strongly impact the detection performance. A minimum error in image coregistration, for example, leads to false detections. In this article, we illustrate and motivate the use of RBIG following an alternative probabilistic approach. Basically, the idea is to characterize statistically the
original image probability and attribute the change to pixels in the second image with low probability. When applied pixelwise (as we do here), this has the drawback of not detecting some changes that are not statistically noticeable. For instance, in two images with the same car in different locations of a road, the method would not detect the car as a change. Such unfortunate cases could be, however, addressed by extending the idea easily by working in a sliding window instead of a pixelwise approach. The statistical approach has the advantage of reducing the importance of perfect coregistration since the method operates in the geometric space defined by the image, not in the spatial domain explicitly. Thus, small errors of coregistration would not penalize the performance of the proposed approach.

The idea to exploit RBIG for CD is using data coming from the first image \(X_1\) only to estimate the probability model and then evaluating the probability (or change score, \(C\)) for each point in the second image \(X_2\) as follows:

\[
C(x_2) \propto \frac{1}{p_{X_1}(x_2)}. \tag{7}
\]

As for the AD case, we can use different models to estimate \(p_{X_1}\). The most widely used is the Gaussian model. As in Section II-A, when assuming a Gaussian distribution for the input data, the RX method can be used here too, i.e., \(C_{RX}(y)\).

Likewise, kernel methods have been proposed to alleviate the strict assumption of Gaussian distribution [15]. While different configurations were proposed in order to consider only the anomalous changes, here, we use the configuration designed for CD. Following the idea in (7), the data of the first image \(X_1\) is used to estimate the kernel, and then, the method is evaluated in the second image

\[
C_{KRX}(x_2) \propto \frac{1}{p_{X_2}(x_1|x_2)\propto p_{X_1}(x_2)}. \tag{8}
\]

Equivalently, we can use RBIG to estimate the probability of the first image and evaluate the probability in the second one

\[
C_{RBIG}(x_2) \propto \frac{1}{p_{RBIG}(x_2|x_1)}. \tag{9}
\]

It is important to note that, in this case, the data used to estimate the probability density do not contain anomalies (changes in this setting), so the hybrid model is not needed here.

III. EXPERIMENTAL RESULTS

This section analyzes the performance of the proposed RBIG method for anomaly and CD. In order to assess the robustness, we performed tests in both simulated and real scenes of varying dimensionality and sample size. We evaluate the detection power of the methods quantitatively through the receiver operating characteristic (ROC) and precision–recall (PR) curves, along with the area under the curve (AUC) scores. Besides, we provide examples of detection maps of each method to evaluate their quality by visual inspection.

We have performed three experiments. The first experiment is designed to illustrate the effect of the evaluation in an AD toy example. The second experiment deals with AD problem in different real scenarios: detection of airplanes, latent fires, vehicles, and urbanization (roofs). The third experiment is related to evaluate the methods in CD problems involving floods, fires, and droughts. Table I summarizes the different datasets used in the experiments.

We added two standard methods in the experimental results for comparison: 1) a kernel-based method known as the support vector domain descriptor (SVDD) [27], where one seeks to embrace all data points into a hypersphere in Hilbert space, and 2) the classical kernel density estimation (KDE), where a Gaussian kernel is used to define the distribution. All methods were used in the same way to the AD [see (4)] and the CD problems [see (7)]. In order to ease the reproducibility, we provide MATLAB code implementations of all the methods, as well as database with the labeled images used in the second and third experiments in [4].

A. Experiment 1: Simulated Anomalies

The aim of this experiment is to illustrate the behavior of the proposed methods in challenging distributions exhibiting highly nonlinear feature relations. We designed a 2-D dataset where the nonanomalous data are in a circumference and the anomalous data in the middle. Fig. 3 shows the performance of the different methods. The RX method assumption does not hold (the data are clearly non-Gaussian), and hence, it shows poor performance. The performance of KRX is better than RX, but some false detections emerge in the outer circle, mainly related to the difficulty to select a reasonable kernel parameter. The direct application of RBIG easily identifies the anomalous points since they are far from the more dense (most probable) region. The proposed hybrid model further refines the detection since the density is estimated from pervasive data yielded by RX only.
TABLE I

| Images  | Sensor       | Size       | Bands | SR [m] |
|---------|--------------|------------|-------|--------|
| AD      |              |            |       |        |
| Cat-Island | AVIRIS      | 150×150   | 188   | 17.2   |
| WTC     | AVIRIS      | 200×200   | 224   | 17.7   |
| Texas-Coast | AVIRIS     | 100×100   | 204   | 17.2   |
| GulfPort | AVIRIS      | 100×100   | 191   | 3.4    |

CD

| Images | Sensor       | Size       |       |       |
|--------|--------------|------------|-------|--------|
| Texas  | Cross-Sensor | 301×201   | 7     | 30     |
| Argentina | Sentinel-2 | 1237×964  | 12    | 10-60  |
| Chile  | Landsat-8   | 201×251   | 12    | 10-60  |
| Australia | Sentinel-2 | 1175×2031 | 12    | 10-60  |

B. Experiment 2: AD in Real Scenarios

We performed tests in four real examples. Table I summarizes relevant attributes of the datasets, such as sensors, spatial resolution, and spectral resolution.

1) Data Collection: We collected multispectral and hyperspectral images acquired by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) and ROSIS-03 sensors. Fig. 4 shows the scenes used in the experiments. The AD scenarios consider anomalies related to a diversity of problems: airplane, latent fires, urbanization, and vehicle detection [28]–[30].

The Cat-Island dataset corresponds to the airplane captured flying over the beach and it is considered a strange object when compared to the rest of the image (a white spot in the middle of a beach) and the percentage of anomalies represents 0.09% of the scene. The World Trade Center (WTC) image was collected by the AVIRIS over the WTC area in New York on September 16, 2001 (after the collapse of the towers in NY). The dataset covered the hot spots corresponding to latent fires at the WTC, which can be considered as anomalies, and it represents 0.23% of the scene. In the Texas Coast dataset, the anomalies represent 0.67% of the scene and the image contains roofs built on a wooded site and bright spots that reflect light, which can be considered an anomaly. The GulfPort dataset corresponds to a battery of airplanes taxied on the runway and the percentages of anomalies represent 0.60% of the scene.

2) Numerical and Visual Comparison: It is important to take into consideration that KRX requires the selection of some hyperparameters, being the kernel parameter the most critical one. In order to perform a fair comparison while staying in an unsupervised learning setting, we use the standard RBF kernel function, \( k(a, b) = \exp(-||a-b||^2/(2\sigma^2)) \), and set the lengthscale parameter \( \sigma \) to the median distance between all examples.

A visual comparison of the results in terms of activation maps for all methods is given in Fig. 4. They display the predictions given to each sample. The prediction maps show a binary representation between change and nonchange samples obtained from the model subject to a threshold. Results in all scenes demonstrate that: 1) RX is a competitive method for detection; 2) KRX struggles to obtain reasonable results mainly due to the problem of hyperparameter tuning; and 3) RBIG alone excels in all cases, while the hybrid approach (i.e., RX followed by RBIG) refines the results and yields clearer activation maps with sharper spatial detections.

In addition, for a quantitative assessment of the results, it is customary to provide the ROC curves and to derive scores like the AUC from it. Fig. 5 shows the ROC curves and Table II summarizes all AUC values for all images and methods. For each experiment, we performed 1000 runs for testing the significance of the methods based on the ROC profiles. The results are shown in Fig. 6. Although the RBIG model achieves good results, the RX model is able to compete and achieve results as good as RBIG for some images. The HYBRID model is able to keep the properties of the above-mentioned models obtaining results equal or better than any other method. While KRX obtains a reasonable performance in some images, it clearly fails in some situations, such as the Cat-Island image. The low standard deviations show that all methods but the KRX are clearly robust with a little bit bigger standard deviation for the RX method in most cases.

C. Experiment 3: Real and Natural Changes

This section reports an experiment to analyze the performance of the proposed methods in CD problems. The database is composed of different scenes with natural changes, whose characteristics are summarized in Table I.

1) Data Collection: We collected pairs of multispectral images in such a way that they coincide at the same spatial resolution, but at different acquisition times, the images are coregistered. We selected the images in such a way that an anomalous change happened between the two acquisition times. We manually labeled all the images finding the changed pixels. Labeling considered avoiding shadows, changes in lighting, and natural changes in vegetation, which could compromise results evaluation. All images contain changes of different nature, which allows us to analyze and study how

TABLE II

| METHODS | RX | K-RX | SVDD | KDE | RBIG | HYBRID |
|---------|----|------|------|-----|------|--------|
| Cat-Island | 0.96 | 0.70 | 0.70 | 0.97 | 0.99 | 0.99 |
| WTC     | 0.95 | 0.82 | 0.87 | 0.95 | 0.95 | 0.95 |
| Texas-Coast | 0.99 | 0.85 | 0.77 | 0.99 | 0.94 | 0.99 |
| GulfPort | 0.90 | 0.95 | 0.83 | 0.90 | 0.95 | 0.95 |

TABLE III

| METHODS | RX | K-RX | K-SVDD | KDE | RBIG |
|---------|----|------|--------|-----|------|
| Texas   | 0.91 | 0.80 | 0.55  | 0.91 | 0.98 |
| Argentina | 0.94 | 0.93 | 0.87 | 0.77 | 0.97 |
| Chile   | 0.64 | 0.66 | 0.61  | 0.65 | 0.72 |
| Australia | 0.85 | 0.88 | 0.33 | 0.83 | 0.33 |
Fig. 4. AD predictions in four images (one per row). First column: Cat-Island, WTC, Texas Coast, and Pavia original datasets with anomalies outlined in green. Second column: reference maps of each image. From the third column to the last column, activation maps and the AUC values (in parenthesis) for the RX, the best performance method among KRX, KDE and SVDD, RBIG, and the HYBRID model.

Fig. 5. AD ROC curves in linear scale for all scenes. Numbers in legend display the AUC values for each method. (a) Cat-Island. (b) WTC. (c) Texas Coast. (d) GulfPort.

The algorithms perform in heterogeneous realistic scenarios. The Texas wildfire dataset is composed of a set of four images acquired by different sensors over Bastrop County, TX, USA, and is composed of a Landsat 5 TM as the pre-event image and a Landsat 5 TM plus an EO-1 ALI and a Landsat 8 as post-event images. This phenomenon is considered the most destructive wildland-urban interface wildfire in Texas history and the interest region represents 19.54%. The Argentina image represents an area burned between the months of July and August 2016 due to the high temperatures in these crop areas, the change region representing 7.5% of the whole scene. The Chile dataset represents the Aculeo lake in the central part of this country, which has now dried up completely. These images contrast the lake in 2014 when it...
still contained substantial water and in 2019 when it consisted of dried mud and green vegetation. Scientists attribute the lake’s decline to an unusual decade-long drought, coupled with increased water consumption from a growing population, and the changed region represents a relevant 10.81% of the whole scene. The last dataset labeled as Australia shows the natural floods caused by Cyclone Debbie in Australia 2017. Storm damage resulted from both the high winds associated with the cyclone, and the very heavy rain produced major riverine floods. The change samples represent an important portion of the scene, and 17.35% of pixels affected. Since our RBIG approach only takes the time $t_1$ image, these big changes do not have a critical impact on the method’s performance.
TABLE IV
LOW FALSE ALARM RATE REGIME. FIRST ROW CORRESPONDS TO FPR AND COLUMNS REPRESENT THE AUC VALUES FROM RIGOROUS TO MODERATE FPR

| FPR | 0.1 | 0.2 | 0.3 | 0.1 | 0.2 | 0.3 |
|-----|-----|-----|-----|-----|-----|-----|
|     | Texas |        |     |            |        |     |
| RX  | 0.75 | 0.67 | 0.61 | 0.91 | 0.96 | 0.99 |
| KRX | 0.46 | 0.55 | 0.76 | 0.83 | 0.92 | 0.99 |
| SVDD| 0.90 | 0.93 | 0.95 | 0.97 | 0.94 | 0.97 |
| KDE | 0.78 | 0.86 | 0.90 | 0.80 | 0.98 | 0.73 |
| RBIG| 0.94 | 0.98 | 0.99 | 0.95 | 0.96 | 0.97 |
|     | Chile |        |     |            |        |     |
| RX  | 0.21 | 0.17 | 0.49 | 0.76 | 0.51 | 0.85 |
| KRX | 0.09 | 0.14 | 0.50 | 0.79 | 0.83 | 0.86 |
| SVDD| 0.07 | 0.08 | 0.45 | 0.84 | 0.89 | 0.93 |
| KDE | 0.06 | 0.23 | 0.47 | 0.77 | 0.76 | 0.78 |
| RBIG| 0.38 | 0.72 | 0.73 | 0.84 | 0.89 | 0.93 |
|     | Australia |        |     |            |        |     |
| RX  | 0.21 | 0.17 | 0.49 | 0.76 | 0.51 | 0.85 |
| KRX | 0.09 | 0.14 | 0.50 | 0.79 | 0.83 | 0.86 |
| SVDD| 0.07 | 0.08 | 0.45 | 0.84 | 0.89 | 0.93 |
| KDE | 0.06 | 0.23 | 0.47 | 0.77 | 0.76 | 0.78 |
| RBIG| 0.38 | 0.72 | 0.73 | 0.84 | 0.89 | 0.93 |

2) Numerical and Visual Comparison: Fig. 7 shows the RGB composites of the pairs of images, the corresponding reference map, and activation maps obtained. RBIG obtains clearly better results than the other methods in all cases, very good performance in three out of the four scenarios and a clear advantage in the most difficult one (Chile image).

When dealing with highly skewed datasets, PR curves give a more informative picture of an algorithm’s performance compared to ROC. Fig. 8 shows both the ROC and the PR curves results for all methods and all the images. In all cases, RBIG outperforms the other methods largely, thus suggesting the suitability of adopting a more direct approach of density estimation in the CD problems too. A summary of the AUC values of all methods and scenarios is shown in Table III. The RBIG approach is able to estimate the change samples with a high accuracy overtaking in 7%, 3%, 6%, and 5% with respect to the second best method. While AUC returns an overall measure of method’s robustness, in (RS) AD problems, one typically cares about the low false alarm rate regime. We study the performance of the methods looking at different false positive rates (FPRs) in Table IV. The proposed RBIG approach consistently reports the best performance, especially in very low FPR regimes.

IV. Conclusions

We introduced a novel detector based on multivariate Gaussianization. The methodology copes with anomaly and CD problems in RS image processing and meets all requirements of the problems: is an unsupervised method with no parameters to fit, it can deal with large amount of data, and it is more accurate to competing approaches. The model assumption is based on detecting anomalies by estimating probabilities of pixels. The proposed method excelled quantitatively (AUC, ROC, and PR curves) and qualitative based on visual inspection over the rest of the implementations, in both anomaly and CD. The evaluation considered a wide range of RS images, in a diversity of problems, dimensionality, and number of examples. We also suggested a hybrid approach where the Gaussianization method is applied after a regular anomaly detector: this facilitates the density estimation and improves the results notably. Future work will consider exploiting the information-theoretic properties of RBIG [31], which opens alternatives to identify changes in image time series.

REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, “Advances in hyperspectral image classification: Earth monitoring with statistical learning methods,” IEEE Signal Process. Mag., vol. 31, no. 1, pp. 45–54, Jan. 2014.
[2] A. Plaza, J. A. Benediktsson, and J. W. Boardman, “Recent advances in techniques for hyperspectral image processing,” Remote Sens. Environ., vol. 113, pp. S110–S122, Sep. 2009.
[3] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez, and J. Malo, Remote Sensing Image Processing, 1st ed. San Rafael, CA, USA: Morgan & Claypool, 2011.
[4] J. A. Padrón-Hidalgo, V. Laparra, and G. Camps-Valls. (2021). RBIG for Anomaly and Change Detection in Remote Sensing. [Online]. Available: https://isp.uv.es/RBIG4AD.html
