

Efficient Kernel Cook's Distance for Remote Sensing Anomalous Change Detection

José Antonio Padfón-Hidalgo¹, Adrián Pérez-Suay, Fatih Nar, Valero Laparra, and Gustau Camps-Valls

Abstract—Detecting anomalous changes in remote sensing images is a challenging problem, where many approaches and techniques have been presented so far. We rely on the standard field of multivariate statistics of diagnostic measures, which are concerned about the characterization of distributions, detection of anomalies, extreme events, and changes. One useful tool to detect multivariate anomalies is the celebrated Cook's distance. Instead of assuming a linear relationship, we present a novel kernelized version of the Cook's distance to address anomalous change detection in remote sensing images. Due to the large computational burden involved in the direct kernelization, and the lack of out-of-sample formulas, we introduce and compare both random Fourier features and Nyström implementations to approximate the solution. We study the kernel Cook's distance for anomalous change detection in a *chronochrome* scheme, where the anomalousness indicator comes from evaluating the *statistical leverage* of the residuals of regressors between time acquisitions. We illustrate the performance of all algorithms in a representative number of multispectral and very high resolution satellite images involving changes due to droughts, urbanization, wildfires, and floods. Very good results and computational efficiency confirm the validity of the approach.

Index Terms—Anomalous change detection (ACD), Cook's distance, efficiency, influential points, kernel methods, Nyström method, random Fourier features, statistical leverage.

I. INTRODUCTION

THE Earth's surface is constantly changing due to natural events and various anthropogenic interventions. Natural events can be repetitive ones such as seasonal changes as well as extreme or rare events such as disasters. Newly constructed man-made structures, urbanization, and agriculture activities can be given as examples of anthropogenic interventions [1]. However, observing such changes in a timely and accurate

manner is challenging since the Earth's surface is very large and complex, and changes are constantly happening, which may be pervasive or anomalous. Change detection (CD) using remote sensing (RS) images is an active field of research with many systematic methods and procedures to capture the changes on the earth surface. CD methods that are applied to RS images can be acquired from satellite or airborne platforms [2].

CD is extremely important because it allows us to improve predictions and our understanding of events occurring over the entire surface of the earth, such as floods and droughts [3], [4], using Landsat 7 images. In addition, the developed CD methods can help us improve designing and implementing urban monitoring [5]. However, factors such as seasonal differences, atmospheric effects, sensor noise, and registration errors create spurious changes that decrease the performance of the CD methods [1]. In addition, the Earth's surface is complex and heterogeneous, while obtained images can be multimodal or multisource with different spectral and temporal resolutions, which further increase the complexity of the development of robust, accurate, and fast CD methods. The recent advances in RS sensors, statistical models, and computational power, as well as an immense amount of data availability, have provided additional possibilities and challenges in RS image processing [6]–[8]. Most importantly, the increasing spatial and temporal resolution of globally available satellites such as Sentinel-2 gives a unique opportunity to monitor regular and extreme events. In addition, the use of very high-resolution satellite imagery (such as Digital Globe QuickBird) is becoming increasingly important for RS applications.

A related field to CD called anomalous change detection (ACD) is concerned about a slightly different problem [9]: the ACD setting differs from standard CD because the objective is to identify only rare (or anomalous) events, ignoring the regular (or pervasive) ones. These pervasive differences may be due to calibration, illumination, look angle, and even the choice of RS satellite. By contrast, the anomalous changes are assumed to be relatively rare and can be highlighted in a minor part of the image. Although there were related studies before, first focused study of ACD was proposed by Theiler and Perkins using a machine learning approach [9] with many other subsequent studies of Theiler and his colleagues [10]–[17]. In the literature, ACD was tackled using distribution-based [9], [12], distance-based [13], classifier-based [11], and reconstruction-based [18] approaches. Note that ACD is also closely related to anomaly detection [16] and novelty detection [19], where all employ similar approaches. However, most ACD methods

Manuscript received May 14, 2020; revised August 10, 2020; accepted August 23, 2020. Date of publication September 1, 2020; date of current version September 25, 2020. 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. The work of José Antonio Padfón-Hidalgo was supported by Generalitat Valenciana under Grisolia Grant GRISOLIA/2016/100. (Corresponding author: José Antonio Padfón-Hidalgo.)

José Antonio Padfón-Hidalgo, Adrián Pérez-Suay, Valero Laparra, and Gustau Camps-Valls are with the Image Processing Laboratory, Universitat de València, 46980 Valencia, Spain (e-mail: joseantoniopadronhidalgo@gmail.com; gustau.camps@uv.es).

Fatih Nar is with the Ankara Yıldırım Beyazıt University, Ankara 06010, Turkey.

This article has supplementary downloadable material available at <https://ieeexplore.ieee.org>, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2020.3020913

are linear, which limit their success and applicability to the real-world problems [20]. Researchers introduced nonlinearity to overcome this limitation, i.e., using neural networks [21] and kernel methods [20], [22], [23]. Although kernel methods bring excellent performance, they are computationally demanding, so efficient approximations are needed [22]. Interested readers can read more about ACD in [10], [17], [24], and [25] that are, respectively, comparison, analysis, review, and tutorial style studies.

The ACD settings can be properly framed in statistical terms. However, the concept of *anomalousness* is elusive and difficult to define concretely. Nevertheless, identifying influential points in multivariate data distributions is an active field of research in statistics, information theory, and machine learning. Main applications involve characterizing distributions, detecting anomalies, extremes and changes, and assessing robustness [14], [15]. Detection of such influential points also have relevant applied implications for climate, health, and social sciences, and in a wide diversity of engineering and computer science problems. It is important to remark that we aim to detect anomalous (extreme) changes, i.e., not pervasive changes related to, for example, illumination conditions. Therefore, we will refer to anomalies among two images, leverage points, or changes interchangeably.

The interest to find anomalous changes is very broad, and many methods have been proposed in the literature, ranging from equalization-based approaches that rely on whitening principles [26] to multivariate methods that extract distinct features out of the change (difference) image [27] and that reinforce directions in feature spaces associated with noisy or rare events [28], [29], as well as regression-based approaches such as in the chronochrome [30], [31], where a regression model approximates the next incoming image and big residuals are associated with anomalies. In this article, we build our nonlinear ACD method on this latter chronochrome approach based on Cook's distance [32], where our initial efforts can be seen in [22]. Among other measures, we preferred Cook's distance since it allows robust fitting despite data are being contaminated by outliers (or anomalies). Many diagnostic measures have been introduced other than the seminal work of Cook such as linear regression [33], [34], penalized (ridge) regression [35], sparse regression models like LASSO [36] as *parametric models*, spline smoothing [37]–[39] and polynomial regression [40] as *non-parametric models*, and longitudinal regression [41], generalized linear, and Cox proportional hazard models [42]–[44] as *semiparametric models*.

For ACD, an adequate model assumption and specification is crucial and has many theoretical and applied implications. The main problem is to select a flexible model that can capture nonlinear relations while also providing high detection power and computational efficiency. All these are relevant aspects to consider for the diagnostic measure, for which many methods have been proposed. However, Cook's distance models only linear relations, which limit its applicability to complex real-world data. In recent years, kernel methods have been widely adopted as an appropriate framework for nonlinear model development in machine learning for classification, regression, hypothesis

testing, and dimensionality reduction [45], [46]. Kernel methods allow one to derive flexible nonlinear and nonparametric models, are intrinsically regularized, and are endorsed with solid mathematical properties. This has allowed us to define diagnostics based on leveraging the kernel ridge regression (KRR) method [47]. However, despite the excellent modeling performance of KRR, the direct definition of leverage scores based on KRR implies a huge computational cost and the lack of a practical out-of-sample estimates [48]–[50]. This hampers its adoption and usefulness in real practice.

In this article, we introduce the Cook's distance for the KRR model in a reproducing kernel Hilbert space (RKHS) for ACD. Noting the high computational cost, we introduce random Fourier features [51] and the Nyström method [52], [53] for improved efficiency. Both approaches allow us to compute residuals [54] and leverage the KRR *explicitly* in RKHS, while the Nyström method also provides implicit regularization capabilities. Essentially, the Nyström method approximates the large kernel matrix by a much smaller low-rank matrix. Although the best low-rank approximation is obtained by singular value decomposition (SVD), it is computationally expensive. On the contrary, the Nyström method achieves low-rank approximation with considerably higher computational efficiency [55], [56]. The proposed methods are simple, computationally very efficient in both memory and processing costs, and achieve improved detection compared to standard approaches. We show results in a set of real ACD problems with pairs of large-scale multispectral satellite images acquired by different sensors (Quickbird, Sentinel-2) and involving different changes of interest (floods, wildfires, urbanization, and droughts).

The remainder of this article is organized as follows. Section II sets the notation, introduces the Cook's distance, briefly reviews the concept of influential points and leveraging in statistics, and introduces the direct kernel Cook's distance. Section III elaborates further on our proposed fast implementations and provides a comparison of space and time complexity in all methods. Section IV presents the performance of the proposed fast Cook's chronochrome method for ACD on synthetic and real-world data. Finally, we conclude in Section V with some remarks and prospective future work.

II. KERNELIZED COOK'S DISTANCE

A. Notation and the Chronochrome Approach

Let us define two consecutive d -band multispectral images in matrix form $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times d}$ composed of n pixels $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^d$, $i = 1, \dots, n$. Assume that a set of changes have occurred in between, and that such changes do not alter the image distribution significantly. The “chronochrome” approach [30] builds on this idea and fits a model to predict the second image \mathbf{Y} from the first one \mathbf{X} and decides that a point is anomalous (i.e., it has changed) if, for instance, the corresponding residual is significantly large. The prediction function $f: \mathbf{x} \rightarrow \mathbf{y}$ is learned from the observations. The task is now to assess the significance of the obtained residuals, $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$, that is to derive a sensible diagnostic measure.

B. Cook's Distance

Cook's distance comes from the definition of *leverage*, which measures how distant are the independent variable values (of a particular observation) from those of the other observations. The highest leveraged points are those observations that could be considered as extreme or outlying values of the independent variables. Cook's distance measures the effect of removing a given observation. Therefore, the aim is to find out which elements from the sample set are more relevant to the model.

The standard Cook's distance assumes a linear model for prediction of the second image from the first one, i.e., $\hat{\mathbf{Y}} = \tilde{\mathbf{X}}\mathbf{W}$, where $\mathbf{W} \in \mathbb{R}^{(d+1) \times d}$, and $\tilde{\mathbf{X}}$ is the augmented design matrix with a column of ones to account for the bias term, $\tilde{\mathbf{X}} = [\mathbf{X} | \mathbf{1}_n]$. The solution to this least squares problem is given by the Wiener-Hopf normal equations, $\mathbf{W} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \mathbf{Y}$. The predictions can be expressed as $\hat{\mathbf{Y}} = \tilde{\mathbf{X}}\mathbf{W} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \mathbf{Y} = \mathbf{H}\mathbf{Y}$, where \mathbf{H} is known as the *projection matrix*, and we define the *leverage score* of the i th observation as

$$h_i = \mathbf{x}_i^\top (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \mathbf{x}_i. \quad (1)$$

Similarly, the i th element of the residual vector $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$ is denoted as e_i . The Cook's distance D_i for observation \mathbf{x}_i , $i = 1, \dots, n$, is defined as the sum of all the changes in the regression model when the i th observation is deleted

$$D_i = \frac{\sum_{j=1}^n (\hat{\mathbf{y}}_j - \hat{\mathbf{y}}_{j \setminus i})^2}{d \text{MSE}^2} \quad (2)$$

where $\hat{\mathbf{y}}_j$ means to predict the j th sample through the model trained with all the samples and $\hat{\mathbf{y}}_{j \setminus i}$ is the fitted response value obtained when i is excluded, and MSE is the mean square error of the regression model with all samples, i.e., $\text{MSE} = \frac{1}{N} \sum_{j=1}^n (\hat{\mathbf{y}}_j - \mathbf{y}_j)^2$. Cook's distance can be equivalently expressed using the leverage

$$D_i = \frac{e_i^2 h_i}{d \text{MSE}^2 (1 - h_i)^2}. \quad (3)$$

Cook showed that this estimation can be obtained using incremental rank-1 updates of covariances, without even needing to recompute each model when the i th sample is removed [32].

C. Kernel Cook's Distance

The kernel Cook's distance can be easily derived by departing from (3). For that, we need to compute both the errors and the leverage scores as a function of the input data only. Let us first recall the KRR prediction formula, $\mathbf{y} = \mathbf{K}(\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}$, where λ is a regularization parameter, and \mathbf{K} is the kernel matrix. The residuals are thus $\mathbf{e} = (\mathbf{I} - \mathbf{H}^{\mathcal{H}})\mathbf{y}$, where the (kernel) projection matrix $\mathbf{H}^{\mathcal{H}} = \mathbf{K}(\mathbf{K} + \lambda \mathbf{I})^{-1}$, and the (kernel) leveraging scores become

$$h_i^{\mathcal{H}} = \text{diag}(\mathbf{H}^{\mathcal{H}}), \quad i = 1, \dots, n. \quad (4)$$

From here, one can readily compute $e_i^{\mathcal{H}}$ and the kernel Cook's distance as

$$D_i^{\mathcal{H}} = \frac{(e_i^{\mathcal{H}})^2}{d \text{MSE}^2 (1 - h_i^{\mathcal{H}})^2}. \quad (5)$$

Note that the inversion of a large \mathbf{K} matrix in $\mathbf{H}^{\mathcal{H}}$ has a cost of cubic time complexity and quadratic space (memory) complexity. One could think of computing the leverage scores using an SVD, but the exact computation is as costly as solving the original problem since the cost is also cubic. Unlike the linear case, the recursive solution of (5) is cumbersome, and one has to recompute each model after sample deletion, thus involving a cascade of costly inverse operations.

III. EFFICIENCY IN KERNEL COOK

In this article, we will exploit both random Fourier features and Nyström approximation of the leverage scores and the errors for Cook's distance approximation.

A. Randomized Cook's Distance

Let us first approximate the kernel matrix with random Fourier features [51]. Formally, we now use a linear regression model expressed on data explicitly projected onto q random Fourier features. Let us define a feature map $\mathbf{z}(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{C}^q$, explicitly constructed as $\mathbf{z}(\mathbf{x}) := [\exp(i\mathbf{w}_1^\top \mathbf{x}), \dots, \exp(i\mathbf{w}_q^\top \mathbf{x})]^\top$, where $i = \sqrt{-1}$, and $\mathbf{w}_q \in \mathbb{R}^d$ is randomly sampled from a data-independent distribution [51]. The prediction model is now defined as $\hat{\mathbf{Y}} = \Re\{\mathbf{Z}\mathbf{W}\}$, where $\mathbf{Z} = [\mathbf{z}_1 \dots \mathbf{z}_n]^\top \in \mathbb{R}^{n \times q}$, with the weight matrix $\mathbf{W} \in \mathbb{R}^{q \times d}$. The *randomized leverage* of a particular sample is now expressed

$$h_i^R = \Re\{\mathbf{z}(\mathbf{x}_i)^\top (\mathbf{Z}^\top \mathbf{Z} + \lambda \mathbf{I})^{-1} \mathbf{z}(\mathbf{x}_i)\} \quad (6)$$

which is then plugged into (3) owing to the linearity of the model where $\mathbf{e}^R = (\mathbf{I} - \mathbf{H}^R)\mathbf{y}$ and then leads to

$$D_i^R = \frac{(e_i^R)^2}{d \text{MSE}^2 (1 - h_i^R)^2}. \quad (7)$$

This allows us to control the memory and computational complexity explicitly through q , as one has to store matrices of $n \times q$ and invert matrices of size $q \times q$ only. It is worth noting that, in practice, a low number of random Fourier features are needed, $q \ll n$. This is not only beneficial in computation time and memory savings but also has a regularization effect in the solution.

B. Nyström Cook's Distance

The Nyström method selects a small set of $r \ll n$ samples to make a low-rank approximation of an $n \times n$ kernel matrix $\mathbf{K} \approx \mathbf{K}_{rn}^\top \mathbf{K}_{rr}^{-1} \mathbf{K}_{rn}$ [52], where $\mathbf{K}_{rn} \in \mathbb{R}^{r \times n}$ contains the kernel similarities between $\tilde{\mathbf{X}} \in \mathbb{R}^{r \times d}$ and $\mathbf{X} \in \mathbb{R}^{n \times d}$, and $\mathbf{K}_{rr} \in \mathbb{R}^{r \times r}$ is a kernel matrix containing data similarities between the points in $\tilde{\mathbf{X}}$. By exploiting the Nyström method in the Woodbury-Morrison formula, we obtain

$$(\mathbf{K} + \lambda \mathbf{I})^{-1} = \lambda^{-1} (\mathbf{I} - \mathbf{K}_{nr} (\lambda \mathbf{K}_{rr} + \mathbf{K}_{nr}^\top \mathbf{K}_{nr})^{-1} \mathbf{K}_{nr}^\top) \quad (8)$$

and now defining $\mathbf{Q} = \lambda \mathbf{K}_{rr} + \mathbf{K}_{nr}^\top \mathbf{K}_{nr}$, the projection matrix approximation is defined as

$$\mathbf{H}^N = \lambda^{-1} \mathbf{K} (\mathbf{I} - \mathbf{K}_{nr} \mathbf{Q}^{-1} \mathbf{K}_{nr}^\top) \quad (9)$$

TABLE I
SPACE AND TIME COMPLEXITY FOR ALL METHODS

Method	T	C	C ⁻¹	W	L	ACD	O(.)
Space							
L-Cook	—	d^2	d^2	d^2	n	n	$\mathcal{O}(nd)$
R-Cook	nq	q^2	q^2	q^2	n	n	$\mathcal{O}(nq)$
N-Cook	n^2	r^2	r^2	—	n	n	$\mathcal{O}(n^2)$
K-Cook	n^2	n^2	n^2	—	n	n	$\mathcal{O}(n^2)$
Time							
L-Cook	—	nd^2	d^3	nd^2	nd^2	nd^2	$\mathcal{O}(nd^2)$
R-Cook	nqd	nq^2	q^3	nq^2	nq^2	nq^2	$\mathcal{O}(nq^2)$
N-Cook	n^2d	nr^2	r^3	—	n^2r	n^2d	$\mathcal{O}(n^2r)$
K-Cook	n^2d	n^3	n^3	—	n^3	n^2d	$\mathcal{O}(n^3)$

T is transformation of image into a nonlinear space, C is for covariance/kernel matrix, W is for regression weight, L is for leverage, ACD is the Cook's distance, and O(.) is the overall complexity.

with Nyström leverage scores

$$h_i^N = \text{diag}(\mathbf{H}^n) \quad (10)$$

and $\mathbf{e}^N = (\mathbf{I} - \mathbf{H}^N)\mathbf{y}$; thus, the Nyström Cook's distance becomes

$$D_i^N = \frac{(e_i^N)^2}{d \text{MSE}^2} \frac{h_i^N}{(1 - h_i^N)^2}. \quad (11)$$

C. Memory and Computational Cost

Space (memory) and time (computational) efficiency of the linear and nonlinear versions are presented in Table I. In this study, the linear version is named as L-Cook, while the nonlinear versions are named Randomized Cook (R-Cook), Nyström Cook (N-Cook), and Kernel Cook (K-Cook). Note that d is the spectral dimension, and it is around 10 for multispectral images and around 100 for hyperspectral images. Although q and r can have similar values, generally $q < r$. Since large images are used, n is much larger than r , q , and d . Therefore, in general, $d < q < r \ll n$.

As can be seen in Table I, the L-Cook method provides superior space and time efficiency. However, the L-Cook method is only limited to rare linear scenarios, where the real-world nonlinear transformation between multitemporal images are formed due to various reasons. However, space and time complexity of the K-Cook method is proportional to the number of pixels in the image, respectively, quadratic in space and cubic in time. Thus, the use of the K-Cook method is not feasible for large images, which is the common scenario nowadays. Note that, for the N-Cook method, kernel matrix \mathbf{K} is still used in (9), but there is no inversion operation on it. Therefore, N-Cook has same space complexity with the K-Cook method as we need to store kernel matrix \mathbf{K} . However, time complexity of N-Cook is still superior to the K-Cook method, since only an $r \times r$ matrix is inverted.

IV. EXPERIMENTAL RESULTS

This section analyzes the performance of the proposed linear and nonlinear Cook's distance methods for ACD. In order to test the robustness of the proposed methods, we performed tests in

both simulated and real scenes with changes. We evaluate the detection performance of the methods quantitatively through the area under the curve (AUC) of the receiver operating characteristic (ROC) and qualitatively by inspection of the detection maps. We have performed two experiments with different complexities of difficulty while controlling the analyzed changes. The first experiment is designed over a real scenario and synthetic changes. The second set of experiments deals with both real scenes and natural changes related to floods, fires, and urbanization. In order to ease the reproducibility, we provide MATLAB implementations of the methods. Moreover, we made available a database with the labeled images used in the second experiment.¹

A. Experiment 1: Real Scene With Simulated Changes

The aim of this experiment is to show and analyze the performance of the proposed methods when the change between images is nonlinearly distributed. In this example, we can analyze how nonlinear methods fit the regression model to the data well and how they detect the influential points in the Cook's distance approach. The experiment involves representing a nonlinear relation between two images in order to demonstrate the limitations of the linear algorithms in this situation.

Fig. 1(a) and (b) shows an aerial scene taken over the Image Processing Laboratory from Google Earth in the R band. Fig. 1(a) represents the image at time t_1 (no change class), while Fig. 1(b) represents the image at time t_2 (change class). All the values of the second image (t_2) were modified by applying a soft nonlinear function (an inverted parabola) to simulate nonanomalous changes. In order to introduce the anomalous changes, we interchanged square patches of 4×4 pixels randomly selected.

Since kernel Cook's distance is computationally very demanding, we have selected a portion of the full image in order to have a comparison of all proposed methods together. In particular, we used the region of interest shown in Fig. 1(c) and marked in a red box in Fig. 1(b); the anomalies are highlighted in a black rectangle and the anomalous class represent 0.016%. Fig. 1(d) represents the scatter of original image x -axis against transformed image y -axis; the points in yellow color are the change pixels, but the points in blue color ideally would not be detected as an anomalous change pixels. Fig. 1(e) illustrates how a linear model does not fit the distribution well and the inferred values lead to false-positive errors (in the tails) and true negative errors (green color). Fig. 1(f) shows how a nonlinear model over distribution fits well and both avoid the false positives and detect the changed pixels in the images. These results are confirmed visually through the prediction maps in Fig. 2, where the kernel Cook's distance excels in detection.

In contrast, Table II showcases how efficient the proposed efficient methods can be, achieving better values of AUC compared to the kernel one in less time. Therefore, because of the huge computational cost involved in its calculation, one cannot use it in standard images (even as small as the one in Fig. 1), so efficient algorithms for computing Cook's distances in nonlinear kernel settings are strictly necessary.

¹[Online]. Available: <http://isp.uv.es/code/kcook>

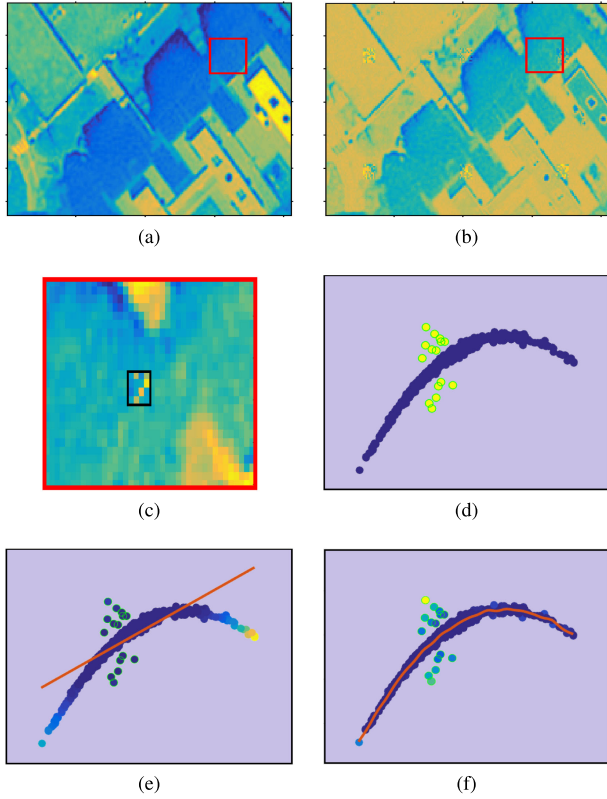


Fig. 1. (a) Image (R band) at time t_1 and the region of interest (red box). (b) Image (R band) at time t_2 and the region of interest (red box). We apply background color distortion and added square patches of 4×4 over t_2 simulating the anomalies, (c) region of interest (red box in t_2) and the corresponding label is surrounded and highlighted in black, (d) scatter plots between t_1 and t_2 pixels in R band, blue dots represent the non-change class and the yellow dots correspond to change class. Panel (e) shows how mis-specification of the linear regression model cannot detect anomalies, while a nonlinear Cook's distance can in (f). In both (e) and (f), the dots color specify how much anomalous the point is for the model (blue less, yellow more).

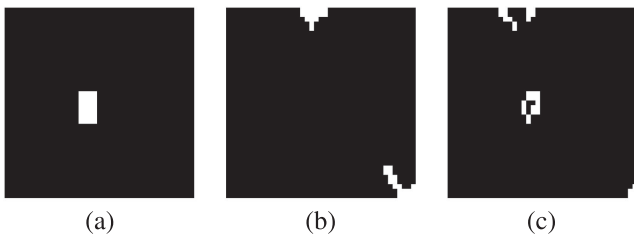


Fig. 2. (a) Prediction map (labels). (b) Change prediction map detected by the linear method. (c) Change prediction map detected by the nonlinear Cook's distance.

B. Experiment 2: Real and Natural Changes

In this section, we report experiments in several real satellite images. We aim to detect changes that can be found naturally in a real environment. The dataset is composed of five different scenes with natural changes, including urbanization, wildfires, droughts, and flooding.

1) *Data Collection*: We collected pairs of multispectral images acquired at different times over the same location. We selected the images in such a way that a noticeable change happened between the two acquisition times. We photo-interpreted

TABLE II
AUC AND THEIR RESPECTIVE TIME VALUES (IN SECONDS) PER METHOD

Methods	L-Cook	R-Cook	N-Cook	K-Cook
AUC	0.55	0.93	0.93	0.92
Time	0.01	0.03	2.64	6.32

TABLE III
IMAGE ATTRIBUTES USED IN THE EXPERIMENTATION DATASET

Images	Sensor	Size	Bands	SR
Argentina	Sentinel-2	381 x 500	12	10-60m
Denver	Quickbird	101 x 101	4	0.6-2.4m
Arizona	Cross-Sensor	201 x 201	7	30m
Texas	Cross-Sensor	301 x 201	7	30m
Australia	Sentinel-2	201 x 501	12	10-60m

and manually labeled all the image pixels affected by a change of interest. This step is critical and delicate since we could fall into many false alarms due to, for instance, shadows, illumination changes, or natural changes in the vegetation. All images contain changes of a different nature, which allows us to study how the different Cook's distance algorithms perform in a diversity of realistic scenarios.

A brief summary of the images and change events is as follows. The Argentina dataset represents an area burned between the months of July and August 2016. Denver Region Urbanized Project Area describes the stereo-compiled building roofprints feature of Denver Regional Council of Government. The Texas wildfire dataset is composed of a set of four images acquired by different sensors over Bastrop County, Texas (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 postevent images. This phenomenon is considered the most destructive wildland-urban interface wildfire in Texas history. The Arizona dataset corresponds to the decline of Lake Powell in the USA. The first image was taken by Landsat-5 and shows its highest water level. The second was taken by Landsat-8 following a period of drought that began in 2000. When the water volume was measured five months later, it was less than half of the maximum lake capacity. The Australia dataset 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 that produced major riverine floods. Table III gives some descriptors of the images in the database, while Fig. 3 shows the RGB composites of the pairs of images and the corresponding reference map.

2) *Numerical Comparison*: We selected the hyperparameters using 1000 randomly selected pixels for cross-validation. Each method implies a different set of parameters. For both the randomized and Nyström methods, we have cross-validated the r and q parameters by exploring values between 1 and 400, particularly $r, q \in \{1, 5, 10, 25, 50, 100, 200, 300, 400\}$. In this work, we used the standard radial basis function (RBF) kernel function, $k(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / (2\sigma^2))$. The RBF kernel shows good theoretical properties (universal kernel, smoothness, and robustness), convenience (only the lengthscale parameter σ needs to be tuned), and good performance in practice. The RBF

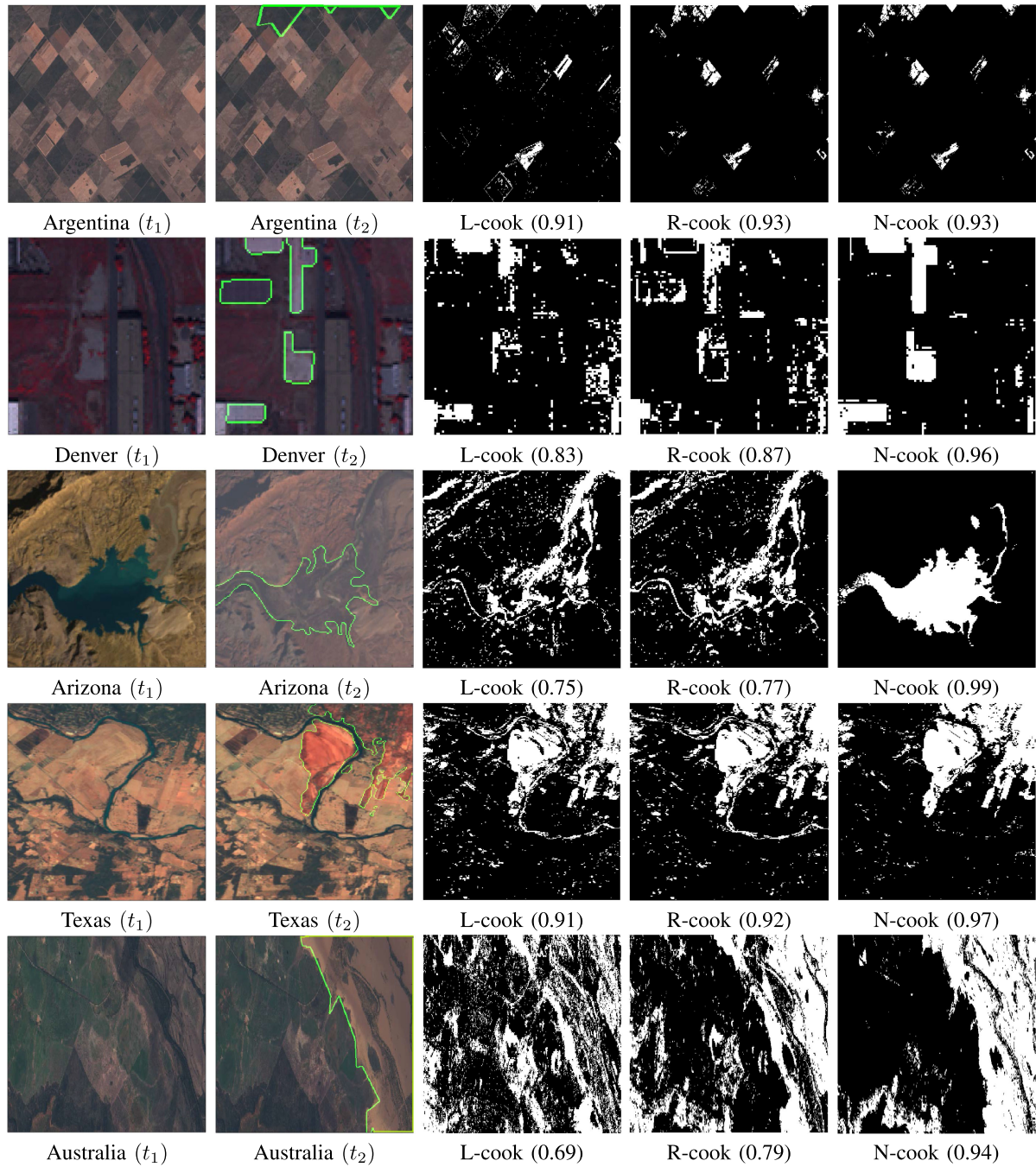


Fig. 3. Images with natural changes and predictions maps. First row: area burned in Argentina between the months of July and August 2016, anomalous samples represent 2.7%. Second row: urbanization area over Denver city correspond to roofprints (extension of anomalous pixels represents the 11.5% of the image). Third row: decline of the Lake Powell in Arizona, USA (16.35%). Fourth row: the most destructive wildland-urban interface wildfire in Texas history (19.5%). Last row: natural floods caused by Cyclone Debbie in Australia (34%). First column: images without changes, first time of acquisition (t_1). Second column: images with the anomalous changes and their corresponding labels are surrounded and highlighted with green color, second time of acquisition (t_2). Third column: prediction map of linear method. Fourth column: prediction map of random Fourier features method. Last column: prediction map of Nyström approximation method. AUC value in parentheses.

kernel is used to perform kernel regression, which incorporates a regularization parameter λ . We searched both σ and λ logarithmic grids between 10^{-4} and 10^{20} .

We optimized the hyperparameters of different methods to maximize the cross-validation AUC. We compared the ROCs and precision-recall curves in terms of AUCs for all methods and images in Fig. 4. In general, all methods can cope with

the large dimension of the images and can provide reasonable results, $\text{AUC} > 0.70$ (see Table IV).

The nonlinear versions (randomized and Nyström approximations) improve the results of the linear Cook's distance, revealing nonlinear changes in all scenes, yet differences are minor for the Texas scene. The Nyström Cook's distance achieves consistently the best results in all the scenarios, and false or positive rates

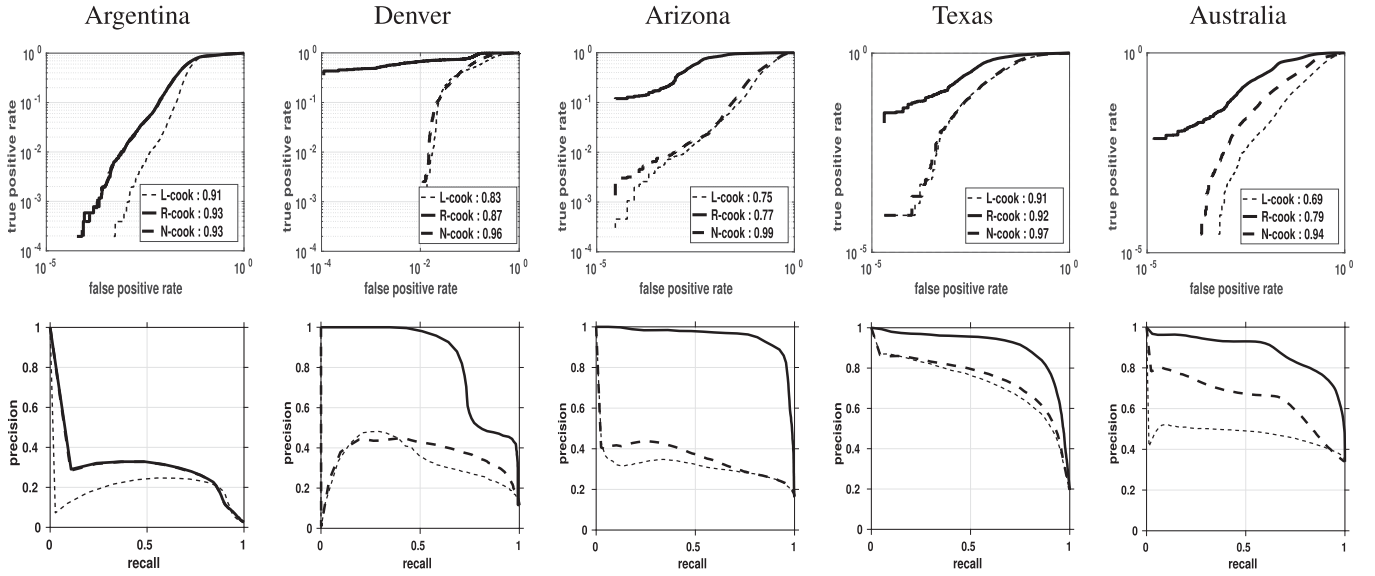


Fig. 4. ROC curves and precision–recall for all images by columns. First row showcases the ROC curves in logarithmic scale. Numbers in legend display the AUC values for each method. Second row showcases the precision–recall following the ROC curves legend.

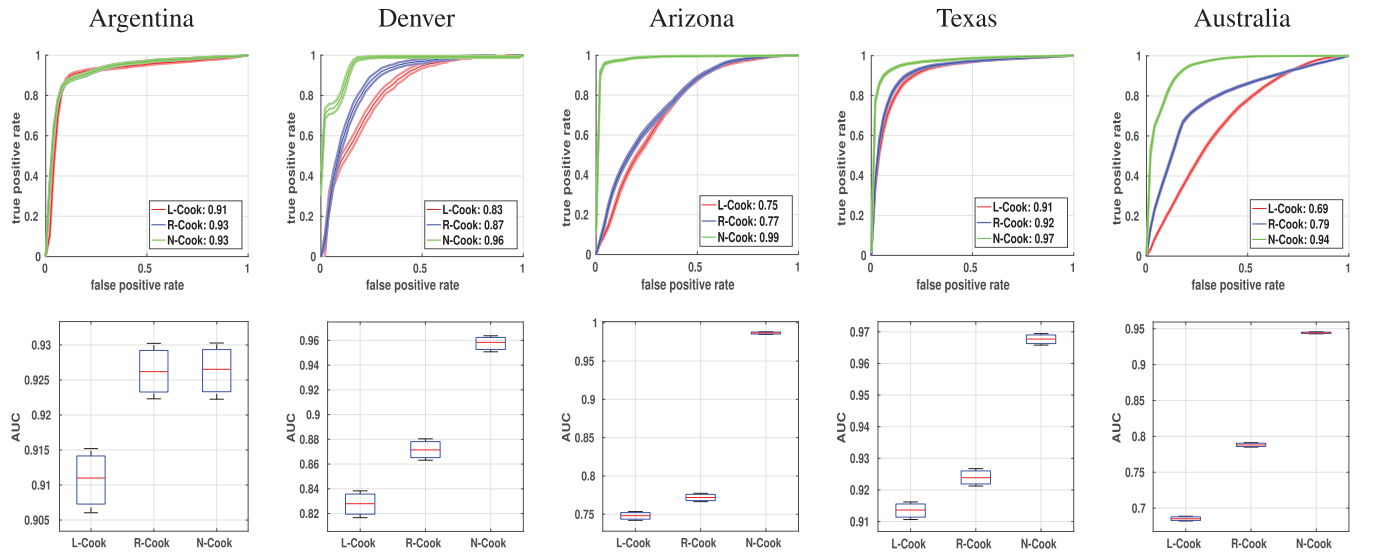


Fig. 5. Bootstrap experiment. In the top row, mean value of the 1000 experiments is plotted as ROC curves with the standard deviation of each detection algorithm represented by the shaded region. In the bottom row, AUC values and standard deviation for each method are shown as boxplot.

TABLE IV
AUC PER METHOD AND SCENE

Methods	Argentina	Denver	Arizona	Texas	Australia
L-Cook	0.91	0.83	0.75	0.91	0.69
R-Cook	0.93	0.87	0.77	0.92	0.79
N-Cook	0.93	0.96	0.99	0.97	0.94

The best results are bold faced.

regimes. A average gain of +15.6% over the linear approach and of +11.8% over the randomized approach, along with the computational efficiency, justify the adoption of this approach. The double logarithmic plot aims to better appreciate the differences in very low false positive rate regimes. In addition, precision and

recall are an understanding and measure of relevance. Here, it becomes clear that the Nyström approach excels in all images.

For each experiment, 1000 runs were made for testing the significance of the methods based on the ROC profiles. The mean value of the experimental runs is plotted with the standard deviation of each detection algorithm represented by the shaded region in Fig. 5. In addition, a boxplot is shown in the same figure to illustrate the standard deviation of each methods with a better precision. As seen in Fig. 5, N-Cook has always superior or equivalent performance compared to L-Cook and R-Cook, i.e., higher detection rate and lower false alarm rate, and higher AUC value and lower standard deviation.

3) *Visual Comparison*: A visual comparison of the results is given in Fig. 3. Differences between the L-Cook and the

R-Cook are not visually significant either. In general, N-Cook yields clear and sharper detection maps (last column), especially in large spatial structures (see, e.g., roofs in Denver, lake in Arizona) but also exhibits a much lower false alarm rate (see, e.g., a less amount of spurious detections in Texas wildfires). This is, however, sometimes compensated with sensitivity to subtle reflectance changes and misclassified pixels in Australia due to imperfect labeling of pixels. This is why this problem is so difficult to solve in an automatic way.

V. CONCLUSION

We introduced the kernel Cook's distance for ACD settings, with the particular focus on RS image CD problems. The key in the proposed methodology is to redefine the ACD problem in an RKHS where the data are mapped to. This endorses the methods with improved capacity and flexibility, since nonlinear feature relations (and hence outliers) can be better identified. However, the obtained kernelized method encounters huge computational problems in practice, which hampers its applicability and wider adoption. To resolve this problem, we proposed computationally efficient techniques based on random Fourier features and low-rank Nyström approximations and compared their capabilities in a wide range of both simulated and real changes. The Nyström approximation excelled over the rest of the implementations in both simulated and real scenarios and in terms of accuracy and efficiency. Future work will study other related kernel diagnostic measures. Extension to online and multichange problems are also topics of further research.

REFERENCES

- [1] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.
- [2] A. Asokan and J. Anitha, "Change detection techniques for remote sensing applications: A survey," *Earth Sci. Inf.*, vol. 12, pp. 1–18, 2019.
- [3] S. Pouyanfar *et al.*, "Unconstrained flood event detection using adversarial data augmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 155–159.
- [4] P. Mishra, T. Feller, M. Schmuck, A. Nicol, and A. Nordon, "Early detection of drought stress in *Arabidopsis thaliana* utilising a portable hyperspectral imaging setup," in *Proc. Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2019, pp. 1–5.
- [5] P. Du, S. Liu, P. Gamba, K. Tan, and J. Xia, "Fusion of difference images for change detection over urban areas," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1076–1086, Aug. 2012.
- [6] 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.
- [7] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. Suppl. 1, pp. S110–S122, 2009.
- [8] G. Camps-Valls, D. Tuia, L. Gmez-Chova, S. Jimnez, and J. Malo, *Remote Sensing Image Processing*, 1st ed. San Rafael, CA, USA: Morgan & Claypool, 2011.
- [9] J. Theiler and S. Perkins, "Proposed framework for anomalous change detection," in *Proc. ICML Workshop Mach. Learn. Algorithms Surveillance Event Detection*, Jan. 2006, pp. 7–14.
- [10] J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Appl. Opt.*, vol. 47, no. 28, pp. F12–F26, 2008.
- [11] I. Steinwart, J. Theiler, and D. Llamocca, "Using support vector machines for anomalous change detection," *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 3732–3735.
- [12] 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.
- [13] J. Theiler and A. M. Matsekh, "Total least squares for anomalous change detection," *Proc. SPIE*, vol. 7695, pp. 507–518, 2010.
- [14] J. Theiler and B. Wohlberg, "Local coregistration adjustment for anomalous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3107–3116, Aug. 2012.
- [15] J. Theiler, "Spatio-spectral anomalous change detection in hyperspectral imagery," in *Proc IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 953–956.
- [16] J. Theiler and A. Ziemann, "Background estimation in multispectral imagery," in *Proc. Opt. Sens. Sens. Congr.*, 2019, Paper HW6B.1.
- [17] J. Simmons, L. Drummy, C. Bouman, and M. Graef, *Statistical Methods for Materials Science: The Data Science of Microstructure Characterization*. Boca Raton, FL, USA: CRC Press, 2019.
- [18] C. Wu, B. Du, and L. Zhang, "Hyperspectral anomalous change detection based on joint sparse representation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 146, pp. 137–150, 2018.
- [19] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Process.*, vol. 99, pp. 215–249, 2014.
- [20] N. Longbotham and G. Camps-Valls, "A family of kernel anomaly change detectors," in *Proc. 6th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2014, pp. 1–4.
- [21] Chris Clifton, "Change detection in overhead imagery using neural networks," *Appl. Intell.*, vol. 18, pp. 215–234, 2003.
- [22] J. A. P. Hidalgo, A. Pérez-Suay, F. Nar, and G. Camps-Valls, "Nonlinear Cook distance for anomalous change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 5025–5028.
- [23] J. A. P. Hidalgo, V. Laparra, N. Longbotham, and G. Camps-Valls, "Kernel anomalous change detection for remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7743–7755, Oct. 2019.
- [24] Y. Elhadad, S. R. Rotman, and D. Blumberg, "Analysis of hyperspectral anomaly change detection algorithms," in *Proc. Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2016, pp. 1–5.
- [25] N. Acito, M. Diani, G. Corsini, and S. Resta, "Introductory view of anomalous change detection in hyperspectral images within a theoretical Gaussian framework," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 32, no. 7, pp. 2–27, Jul. 2017.
- [26] R. Mayer, F. Bucholtz, and D. Scribner, "Object detection by using 'whitening/dewhitenning' to transform target signatures in multitemporal hyperspectral and multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1136–1142, May 2003.
- [27] J. Arenas-Garcia, K. B. Petersen, G. Camps-Valls, and L. K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 16–29, Jul. 2013.
- [28] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
- [29] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection MAD and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.
- [30] A. Schaum and A. Stocker, "Long-interval chronochrome target detection," in *Proc. Int. Symp. Spectral Sensing Res.*, 1997.
- [31] A. Schaum and A. Stocker, "Linear chromodynamics models for hyperspectral target detection," in *Proc. IEEE Aerosp. Conf. Proc.*, 2003, vol. 4, pp. 1879–1885.
- [32] R. D. Cook, "Detection of influential observation in linear regression," *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.
- [33] R. D. Snee, "Regression diagnostics: Identifying influential data and sources of collinearity," *J. Qual. Technol.*, vol. 15, no. 3, pp. 149–153, 1983.
- [34] R. D. Cook and S. Weisberg, "Characterizations of an empirical influence function for detecting influential cases in regression," *Technometrics*, vol. 22, no. 4, pp. 495–508, 1980.
- [35] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [36] Tibshirani R, "Regression shrinkage and selection via the Lasso," *J. Royal Statist. Soc., Ser. B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.

- [37] R. L. Eubank, "Diagnostics for smoothing splines," *J. Royal Statist. Soc. Ser. B (Methodol.)*, vol. 47, no. 2, pp. 332–341, 1985.
- [38] K. Choongrak and E. S. Barry, "Reference values for Cook's distance," *Commun. Statist.—Simul. Comput.*, vol. 25, no. 3, pp. 691–708, 1996.
- [39] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986.
- [40] C. Kim, Y. Lee, and B. U. Park, "Cook's distance in local polynomial regression," *Statist. & Probab. Lett.*, vol. 54, no. 1, pp. 33–40, 2001.
- [41] W. Bae, S. Hwang, and C. Kim, "Influence diagnostics in the varying coefficient model with longitudinal data," *Comput. Statist.*, vol. 23, pp. 185–196, 2008.
- [42] H. Zhu, J. G. Ibrahim, and M.-H. Chen, "Diagnostic measures for the Cox regression model with missing covariates," *Biometrika*, vol. 102, no. 4, pp. 907–923, 2015.
- [43] H. Zhu, J. G. Ibrahim, and X. Shi, "Diagnostic measures for generalized linear models with missing covariates," *Scand. J. Statist.*, vol. 36, no. 4, pp. 686–712, 2009.
- [44] H. Zhu, S. Lee, B. Wei, and J. Zhou, "Case-deletion measures for models with incomplete data," *Biometrika*, vol. 88, no. 3, pp. 727–737, 2001.
- [45] B. Schölkopf and A. Smola, *Learning with Kernels—Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
- [46] J. L. Rojo-Álvarez, M. Martínez-Ramón, J. Muñoz Marí, and G. Camps-Valls, *Digital Signal Processing with Kernel Methods*. Hoboken, NJ, USA: Wiley, 2017.
- [47] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, MA, USA: Cambridge Univ. Press, 2004.
- [48] A. Alaoui and M. W. Mahoney, "Fast randomized kernel ridge regression with statistical guarantees," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2015, pp. 775–783.
- [49] Alessandro Rudi, Daniele Calandriello, Luigi Carratino, and Lorenzo Rosasco, "On fast leverage score sampling and optimal learning," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2018, pp. 5672–5682.
- [50] S. McCurdy, "Ridge regression and provable deterministic ridge leverage score sampling," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2018, pp. 2463–2472.
- [51] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1177–1184.
- [52] C. K. I. Williams and M. Seeger, "Using the Nyström method to speed up kernel machines," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2000, pp. 682–688.
- [53] C. Zhao, G. Zhao, and X. Jia, "Hyperspectral image unmixing based on fast kernel archetypal analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 331–346, Jan. 2017.
- [54] R. Touati, M. Mignotte, and M. Dahmane, "Anomaly feature learning for unsupervised change detection in heterogeneous images: A deep sparse residual model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 588–600, 2020.
- [55] T. Yang, Y.-F. Li, M. Mahdavi, R. Jin, and Z.-H. Zhou, "Nyström method vs random fourier features: A theoretical and empirical comparison," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2012, pp. 476–484.
- [56] N. K. Kumar and J. Schneider, "Literature survey on low rank approximation of matrices," *Linear Multilinear Algebra*, vol. 65, no. 11, pp. 2212–2244, 2017.



José Antonio Padfón-Hidalgo received the B.Sc. degree in telecommunications and electronics from the University of Pinar del Río, Pinar del Río, Cuba, in 2015. He is currently working toward the Ph.D. degree in electronics with the Universitat de València, València, Spain.

His current research interests include developing algorithms in order to detect anomalous and extreme changes for remote sensing imagery.



ods, and causal inference for remote sensing data analysis.

Adrián Pérez-Suay received the B.Sc. degree in mathematics, the master's degree in advanced computing and intelligent systems, and the Ph.D. degree in computational mathematics and computer science from the Universitat de València, València, Spain, in 2007, 2010, and 2015, respectively.

He is currently an Assistant Professor with the Department of Mathematics Education, Universitat de València, where he is also a Researcher with the Image and Signal Processing Group. His research interests include dependence estimation, kernel methods, and causal inference for remote sensing data analysis.



Fatih Nar received the Ph.D. degree in information systems from the Middle East Technical University, Ankara, Turkey.

From 1996 to 2016, he was a Software Developer, Database Administrator, Trainer, Technical Leader, and Researcher with many software development and research companies. From 2017 to 2018, he was a visiting Postdoctoral Researcher with the Image Processing Laboratory, Universitat de València, València, Spain. From 2016 to 2020, he was an Assistant Professor with the Department of Computer Engineering, Konya Food and Agriculture University, Konya, Turkey. In 2020, he joined the Department of Computer Engineering, Ankara Yıldırım Beyazıt University, Ankara, as an Assistant Professor. His main research interests include image processing, machine learning, and remote sensing.



Valero Laparra was born in València, Spain, in 1983. He received the B.Sc. degree in telecommunications engineering and the B.Sc. degree in the electronics engineering from the Universitat de València, València, in 2005 and 2007, respectively, the B.Sc. degree in mathematics from the Universidad Nacional de Educación a Distancia, Madrid, Spain, in 2010, and the Ph.D. degree in computer science and mathematics from the Universitat de València, in 2011.

He is currently an Assistant Professor with the Escuela Técnica Superior de Ingeniería, Universitat de València, where he is also a Researcher with the Image Processing Laboratory.



Gustau Camps-Valls received the Ph.D. degree in physics from the Universitat de València, València, Spain, in 2002.

He is currently a Full Professor of Electrical Engineering and a Coordinator of the Image and Signal Processing Group, Universitat de València. He is involved in the development of machine learning algorithms for geoscience and remote sensing data analysis. He has authored 200 journal papers, more than 200 conference papers, and 20 international book chapters. He holds a Hirsch's index, $h = 60$ (source: Google Scholar), entered the ISI list of Highly Cited Researchers, in 2011, and Thomson Reuters ScienceWatch identified one of his papers on kernel-based analysis of hyperspectral images as a fast moving front research. He is the Editor for the books entitled *Kernel Methods in Bioengineering, Signal and Image Processing* (Hershey, PA, USA: IGI, 2007), *Kernel Methods for Remote Sensing Data Analysis* (Hoboken, NJ, USA: Wiley, 2009), *Remote Sensing Image Processing* (San Rafael, CA, USA: Morgan & Claypool, 2011), and *Digital Signal Processing with Kernel Methods* (Hoboken, NJ, USA: Wiley, 2018).

Dr. Camps-Valls was a recipient of the Prestigious European Research Council Consolidator Grant on Statistical Learning for Earth Observation Data Analysis, in 2015. He is/has been an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and IEEE SIGNAL PROCESSING LETTERS. He was the Invited Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, in 2012 and *IEEE Geoscience and Remote Sensing Magazine*, in 2015.