Random Projection-Based Anderson-Darling Test for Random Fields

Yasser Al Zaim [1]**, and Mohammad Reza Faridrohani** [1]**.**

[1] Department of Statistics, Shahid Beheshti University, Tehran, Iran.

Abstract. In this paper, we present the Anderson-Darling (AD) and Kolmogorov-Smirnov (KS) goodness of fit statistics for stationary and non-stationary random fields. Namely, we adopt an easy-to-apply method based on a random projection of a Hilbert-valued random field onto the real line $\mathbb{R}$, and then, applying the well-known AD and KS goodness of fit tests. We conclude this paper by studying the behavior of the proposed approach in the wide range of simulation studies and in a case study of autistic and healthy individuals.

Keywords. Goodness of Fit Tests, Multiple Testing, One-dimensional Random Projection, Random Field.

MSC: 60G60, 62M40.

Yasser Al Zaim(alzaim@sbu.ac.ir)

Corresponding Author: Mohammad Reza Faridrohani(faridrohani@sbu.ac.ir)

1 Introduction

The data collected by modern techniques may be considered as some independent and identically distributed realizations of a random field, taking values in a real separable Hilbert space $\mathcal{L}^2(T)$ endowed with a scalar inner product $\langle \cdot , \cdot \rangle$, where T is a rectangle in $\mathbb{R}^N$. In other words, the data for a given phenomenon is a random sample of an N-dimensional random field. Formally, Adler (1981) defines the N-dimensional random field as follows.

Let $G^{N,d}$ denote the set of all $\mathbb{R}^d$-valued functions on $\mathbb{R}^N$, $N, d \geqslant 1$, and $\mathcal{G}^{N,d}$ denote the σ-field consisting of all sets of the form $\left\{ g \in G^{N,d} | g(t_j) \in B_j; \quad j = 1, 2, \cdots , m \right\}$, where m is an arbitrary fixed integer, the t_j's are points of $\mathbb{R}^N$ and $B_j \in \mathcal{B}^d$, with $\mathcal{B}^d$ stands for the Borel σ-field generated by the half-open intervals in $\mathbb{R}^d$.

Definition 1.1. (Adler , 1981, p.13) For a given probability space $(\Omega, \mathcal{A}, \eta)$, an n-dimensional d-valued random field is the measurable mapping $X : (\Omega, \mathcal{A}) \longrightarrow (G^{N,d}, \mathcal{G}^{N,d})$. We call X an (N, d) random field and write it for short $X = \{X_t\}_{t \in \mathbb{R}^N}$.

The random projection technique (Johnson and Lindenstrauss , 1984) has been adapted in different scientific areas; for example, in machine learning, functional regression, data mining and so on. For instance, using randomly chosen projections to random lines, Lejsek et al. (2005) present the nearest neighbor application to study copyright protection for online posted images. Also, based on a univariate kernel smoothing and for the functional regression model with scalar response, Patilea et al. (2012) present a projection-based effect test of a functional covariate.

The logic behind the finite random projection approach is to project, using a random projection matrix R, a set of high-dimensional n points in Euclidean space $\mathbb{R}^p$ onto a subspace R^k, with $k \ll \min(n, p)$, such that the pairwise distances are preserved by a small factor $\epsilon > 0$. (Achlioptas , 2003).In other words, given a matrix $A \in \mathbb{R}^{n \times p}$, we can project the p-dimensional n points of A onto $\mathbb{R}^{n \times k}$ by computing $\sqrt{(1/k)}A \cdot R$, where $R_{p \times k}$ is a random projection matrix. The elements of R can be generated, for example, from $\mathcal{N}(0, 1)$ distribution. In effect, many studies have been carried out to choose the best projection matrix R and to study the statistical properties of the projected dataset. We refer the interested reader to Achlioptas (2003) and the references therein for more details on effective methods for choosing the matrix R.

We are concerned with projecting, at random, high-dimensional points of a dataset on one-dimensional random directions and then applying a univariate test on a few randomly elected marginals. In fact, the one-dimensional random projection approach

allows one to test whether the entire distribution of the process is Gaussian or not, and hence it is not restricted to the marginal distributions, and that is why it is suitable for goodness of fit problems. The one-dimensional random projection approach was developed for the infinite dimensional space by Cuesta-Albertos et al. (2006) and Cuesta-Albertos et al. (2007a). Relying on this approach, Cuesta-Albertos et al. (2009) proposed the one-dimensional KS goodness of fit test to study the Gaussianity of a stationary stochastic processes, by observing one realization of the process. Other applications can be found in Cuevas and Fraiman (2009) and when observations are realizations of a random function (functional data), in Cuesta-Albertos et al. (2007b). On the other hand and by observing just one realization, Di Bernardino et al. (2017) presented a Gaussianity test for an isotropic stationary random field, which is based on computing the Euler characteristic function.

The rest of this article is organized as follows. In Section 2, we point out the main results of the one-dimensional random projection approach and give a review of the one-dimensional projected KS goodness of fit test. We generalize the one-dimensional KS test to random field settings and define the projected AD statistic for stochastic processes and random fields in Section 3. Some series of simulation studies are carried out in Section 4 to study the performance of the proposed method. Finally, an application of real data will be developed to illustrate and compare the performance of the projected tests for three-dimensional random fields in Section 5.

2 One-dimensional Random Projection

We assume that all random elements are defined on the probability space $(\Omega, \mathcal{A}, \eta)$. Suppose that X and Y are two $\mathbb{R}^p$-valued random vectors with probability distribution laws $\mathbb{P}$ and $\mathbb{Q}$, respectively. If $\langle \cdot, \cdot \rangle$ denotes the usual scalar product in $\mathbb{R}^p$ and if we define the set $\mathcal{E}(X, Y) = \{h \in \mathbb{R}^p | \langle h, X \rangle \overset{law}{=} \langle h, Y \rangle\}$, then the main statistical problem is finding a sufficient condition on $\mathcal{E}(X, Y)$ to guarantee that $X \overset{law}{=} Y$, that is, X and Y are identically distributed. Cuesta-Albertos et al. (2007a) give a general answer to this problem in the following theorem. In effect, h is a fixed random element generated according to a non-degenerate probability law ψ independent of $\mathbb{P}$ and $\mathbb{Q}$.

Theorem 2.1. *For two $\mathbb{R}^p$-valued random vectors X and Y, let $\mathbb{P}$ be a Borel probability law of X on $\mathbb{R}^p$ so that $\sum_{n \geqslant 1} (\mathbb{E} \|X\|^n)^{-1/n} = \infty$. Then $X \overset{law}{=} Y$ if and only if the set $\mathcal{E}(X, Y)$ has a positive Lebesgue measure.*

Theorem 2.1 was generalized to infinite-dimensional separable Hilbert space as follows.

Theorem 2.2. *(Cuesta-Albertos et al., 2007b) Let $\mathcal{H}$ be a separable Hilbert space endowed with norm $\|\cdot\|$. Let X and Y be two $\mathcal{H}$-valued random elements such that $\sum_{n \geqslant 1} (\mathbb{E}\|X\|^n)^{-1/n} = \infty$. Take ψ as a non-degenerate Gaussian measure on $\mathcal{H}$. Define the set $\mathcal{E}(X, Y) = \{h \in \mathcal{H} | \langle h, X \rangle \overset{law}{=} \langle h, Y \rangle\}$. Then $X \overset{law}{=} Y$ if and only if $\psi[\mathcal{E}(X, Y)] > 0$.*

Suppose that $\mathbb{P}_{\langle h \rangle}$ is the distribution law of the random variable $\langle X, h \rangle$ and, similarly, $\mathbb{Q}_{\langle h \rangle}$ is the probability distribution of the random variable $\langle Y, h \rangle$.

Note. Let us highlight some consequences of these two results. If $X \overset{law}{=} Y$ ($\mathbb{P} = \mathbb{Q}$), then clearly, $\forall h$, the marginal distributions $\mathbb{P}_{\langle h \rangle}$ and $\mathbb{Q}_{\langle h \rangle}$ are matched. But if $\mathbb{P} \neq \mathbb{Q}$, then the probability that all one-dimensional marginal distributions are matched, is zero, or equivalently, if the probability that $\mathbb{P}_{\langle h \rangle} = \mathbb{Q}_{\langle h \rangle}$ is positive for some h, then $\mathbb{P} = \mathbb{Q}$. As a result, the null hypothesis testing of $H_0 : \mathbb{P} = \mathbb{Q}$ is equivalent to testing $H_{0,h} : \mathbb{P}_{\langle h \rangle} = \mathbb{Q}_{\langle h \rangle}$ for a given non-degenerate random direction h.

As a result, the main idea of the proposed goodness of fit test is, at first, generating some random directions h's, and then, performing the underlying null hypothesis testing (H_0) according to $H_{0,h}$'s through the multiple testing paradigm. The empirical error and the power of this proposed test are evaluated through the simulation studies.

Now, suppose that we have a random sample $X_1, X_2, \cdots, X_n$ of a random element X with probability law $\mathbb{P}$ on a separable Hilbert space $\mathcal{H}$ and let $\mathbb{P}_0$ be a given probability law on $\mathcal{H}$. To test the null hypothesis $H_0 : \mathbb{P} = \mathbb{P}_0$ against the alternative $H_1 : \mathbb{P} \neq \mathbb{P}_0$, Cuesta-Albertos et al. (2006) defined the following one-dimensional projected KS statistic.

$$D_n(h) := \sup_{u \in \mathbb{R}} \left| F_n^h(u) - F_0^h(u) \right|, \tag{2.1}$$

where

$$F_n^h(u) := \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, u]}(\langle X_i, h \rangle); \qquad (u \in \mathbb{R}), \tag{2.2}$$

and

$$F_0^h(u) := \mathbb{P}_0 \{x \in \mathcal{H} : \langle x, h \rangle \leq u\}; \qquad (u \in \mathbb{R}). \tag{2.3}$$

That is, if we define $\mathcal{E}(\mathbb{P}, \mathbb{P}_0) = \left\{h \in \mathcal{H} | \mathbb{P}_{\langle h \rangle} = \mathbb{P}_{0, \langle h \rangle}\right\}$, then the hypothesis testing of $H_{0,h} : \mathbb{P}_{\langle h \rangle} = \mathbb{P}_{0, \langle h \rangle}$ is equivalent to testing the underlying null hypothesis H_0. They also

proved that the statistic (2.1) has the exact properties of the well-known one-sample KS statistic and that it is independent of the random direction h. The projected null hypothesis is rejected if $D_n(h)$ is large enough. The above hypothesis testing problem is known as the goodness of fit problem.

Similarly, suppose that $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ are two random samples of the two Hilbert-valued random elements X and Y with probability laws $\mathbb{P}$ and $\mathbb{Q}$ on $\mathcal{H}$, respectively. To test $H_0 : \mathbb{P} = \mathbb{Q}$ against the alternative $H_1 : \mathbb{P} \neq \mathbb{Q}$, Cuesta-Albertos et al. (2006) proposed the following one-dimensional projected KS statistic.

$$D_{n,m}(h) := \sup_{u \in \mathbb{R}} \left| F_n^h(u) - G_m^h(u) \right|, \tag{2.4}$$

where

$$F_m^h(u) := \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, u]} \left(\langle X_i, h \rangle \right); \qquad (u \in \mathbb{R}),$$

and

$$G_m^h(u) := \frac{1}{m} \sum_{j=1}^{m} I_{(-\infty, u]} \left(\langle Y_j, h \rangle \right); \qquad (u \in \mathbb{R}).$$

As a result, with probability one, if $H_{0,h} : \mathbb{P}_{\langle h \rangle} = \mathbb{Q}_{\langle h \rangle}$ is rejected, then the underlying null hypothesis $H_0 : \mathbb{P} = \mathbb{Q}$ is rejected. The above hypothesis testing problem is known as the equality of distribution problem.

In fact, from the results presented by Cuesta-Albertos et al. (2007a) and Cuesta-Albertos et al. (2007b), we conclude that, by defining a proper separable Hilbert space endowed with a suitable inner product, the results are still valid for Hilbert-valued random field settings. To our knowledge, there is no published work on the goodness of fit tests for random fields based on observing a sample.

Therefore, in the following section, we will define the one-dimensional projected KS and AD statistics for Hilbert-valued random fields.

3 Goodness of Fit Tests for Random Fields

Let X be an N-dimensional random field defined on $\Omega \to \mathcal{H} = L^2([0,1]^N)$, where $\mathcal{H}$ is endowed with the following inner product for trajectories (non-random sample paths),

$$\langle f, g \rangle = \int_{t \in [0,1]^N} f_t g_t \, dt, \tag{3.1}$$

which is well defined because of the inequality $2|ab| \leqslant |a|^2 + |b|^2$. By considering this Hilbert space $\mathcal{H}$ and the inner product (3.1), the one-dimensional projected KS test is still reliable and valid in random fields settings, and we will propose the one-dimensional projected AD test.

The inner product in (3.1) is well defined for the Brownian sheet according to the following theorem (Adler and Taylor, 2004, Theorem 3.1.2).

Theorem 3.1. *For a centered Gaussian process h with a continuous covariance function, if h is almost surely continuous, then the sum*

$$h_t = \sum_{n=1}^{\infty} \xi_n \varphi_n(t); \quad t \in T,$$

converges uniformly on T with probability one, where the ξ_n are i.i.d. $\mathcal{N}(0,1)$, and the φ_n are certain functions on T determined by the covariance function of h. In general, the convergence is in $\mathcal{L}^2(\mathbb{P})$ for each $t \in T$.

Now, suppose that $X_1, X_2, \cdots, X_n$ is a random sample of a random field X. Using the one-dimensional random projection approach, we are interested in testing $H_0 : \mathbb{P} = \mathbb{P}_0$, where $\mathbb{P}_0$ is a given continuous probability law on $\mathcal{H}$. According to Section 2, it is enough to generate, at random and independently of $\mathbb{P}$, a Hilbert-valued element h according to non-degenerate law ψ, a Brownian sheet for example. Then, we compute the scalar products $X_i(h) = \langle X_i, h \rangle; (i = 1, 2, \ldots, n)$ and apply the KS test for the projected sample.

Similarly, the projected two-sample KS test for independent samples $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ of two independent random fields is defined.

In the next subsection, we will give our main result, the one-dimensional projected AD goodness of fit test. We will see in Section 4 that the projected AD statistic is more powerful than the projected KS one, as is the case in the standard AD and KS tests (Lehmann, 1999, Lemma 5.7.1).

3.1 AD Goodness of Fit Test for Random Fields

Assume that $X_1, X_2, \cdots, X_n$ is a real-valued sample of a random variable X defined on the probability space $(\Omega, \mathcal{A}, \eta)$. Let F be a continuous distribution function of the probability law $\mathbb{P}$. To test the null hypothesis of $H_0 : \mathbb{P} = \mathbb{P}_0$, for a given continuous

$\mathbb{P}_0$, Anderson and Darling (1952) proposed the following test statistic:

$$A_n^2 := n \int_{\mathbb{R}} \frac{[F_n(u) - F_0(u)]^2}{F_0(u)\,[1 - F_0(u)]}\,\mathrm{d}F_0(u), \tag{3.2}$$

where F_0 is the distribution function of $\mathbb{P}_0$, $F_n(u) := \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,u]}(X_i); u \in \mathbb{R}$, is the empirical distribution function and $w(u) = (F_0(u)\,[1 - F_0(u)])^{-1}$ is a weight function. This test statistic is consistent against all alternative hypotheses (DasGupta, 2008, Sec. 26.5) and, under H_0, it is distribution-free (Lehmann, 1999). Stephens (1974) showed that the AD test is the best empirical distribution function (EDF) statistic in detecting departures of F from Gaussianity. The next Theorem gives the asymptotic distribution of (3.2).

Theorem 3.2. *(DasGupta, 2008, Theorem 26.3) Suppose that the function g is strictly positive on $(0,1)$. If $\int_0^1 u(1-u)\,[g(u)]^{-1}\,\mathrm{d}u < \infty$, then*

$$n \int_{\mathbb{R}} \frac{[F_n(u) - F_0(u)]^2}{g\,(F_0(u))}\,\mathrm{d}F_0(u) \quad \xrightarrow[n\to\infty]{\eta} \quad \int_0^1 \frac{B^2(t)}{g(t)}\,\mathrm{d}t, \tag{3.3}$$

where, $B(t)$ is a Brownian bridge on $t \in [0,1]$.

As a result, if we take the function $g(u) = u(1-u)$, then (3.2) satisfies this Theorem.

In the following, we present the AD type test statistics for Hilbert-valued random fields and for the one-sample case.

Suppose that $X_1, X_2, \cdots, X_n$ is a random sample of realizations of a Hilbert-valued random field X with a probability law $\mathbb{P}$ on $\mathcal{H} = \mathcal{L}^2([0,1]^N)$. Let h be an $\mathcal{H}$-valued random element with a non-degenerate Gaussian probability law ψ, a Brownian sheet for example, independent of $\mathbb{P}$. Since, with probability one, testing $H_0 : \mathbb{P} = \mathbb{P}_0$ against $H_1 : \mathbb{P} \neq \mathbb{P}_0$ is equivalent to testing $H_{0,h} : \mathbb{P}_{\langle h \rangle} = \mathbb{P}_{0,\langle h \rangle}$ against $H_{1,h} : \mathbb{P}_{\langle h \rangle} \neq \mathbb{P}_{0,\langle h \rangle}$, we define the one-dimensional projected AD goodness of fit test statistic as

$$A_n^2(h) := n \int_{\mathbb{R}} \frac{\left[F_n^h(u) - F_0^h(u)\right]^2}{F_0^h(u)\left[1 - F_0^h(u)\right]}\,\mathrm{d}F_0^h(u), \tag{3.4}$$

where $F_n^h(u)$ and $F_0^h(u)$ are defined as (2.2) and (2.3), respectively. The following Theorem gives the asymptotic distribution of (3.4) under H_0.

Theorem 3.3. *Let $\{X_n\}_{n\geqslant 1}$ be an $\mathcal{H}$-valued sequence of independent and identically distributed random elements with a probability law $\mathbb{P}$ and defined on a probability space $(\Omega, \mathcal{A}, \eta)$. Also, for a given continuous probability law $\mathbb{P}_0$ on $\mathcal{H}$ and for a given random direction $h \in \mathcal{H}$, define*

$$A_n^2(h) := n \int_{\mathbb{R}} \frac{\left[F_n^h(u) - F_0^h(u)\right]^2}{F_0^h(u)\left[1 - F_0^h(u)\right]} \, dF_0^h(u).$$

Now, for every $h \in \mathcal{H} - \{0\}$ and $\forall n \geqslant 1$, the statistic $A_n^2(h)$ has the same distribution as the statistic A_n^2 under the null hypothesis. In particular, its null distribution is independent of the random direction h as follows.

$$A_n^2(h) := n \int_{\mathbb{R}} \frac{\left[F_n^h(u) - F_0^h(u)\right]^2}{g\left(F_0^h(u)\right)} \, dF_0^h(u) \quad \xrightarrow[n\to\infty]{\eta} \quad \int_0^1 \frac{B^2(t)}{g(t)} \, dt,$$

where $g(u) = u(1-u)$.

Proof. First, we know that the null distribution of the AD test statistic is indeed independent of F_0 (Lehmann, 1999, p.343). Now, it is clear that if the common distribution of $\{X_n\}_{n\geqslant 1}$ is $\mathbb{P}_0$, then F_0^h is simply the common null distribution function of the projected random field, that is, F_0^h is the common null distribution function of the real-valued random variables $\{X_n(h)\}_{n\geqslant 1} = \{\langle X_n, h\rangle\}_{n\geqslant 1}$ and it is also continuous. Furthermore, F_n^h is the empirical distribution function of the random variables $X_1(h), X_2(h), \ldots, X_n(h)$. As a result, for the projected sample, and by Theorem 3.2, the statistic (3.4) satisfies (3.3) and it is independent of h. $\qquad\square$

A consequence of this theorem is that, there exists a constant $c_{\alpha,n}$, large enough, such that $\eta(A_n^2(h) > c_{\alpha,n}) = \alpha$, where α is the desired nominal significance level. Also, for ψ-almost every $h \in \mathcal{H}$, we have $\eta\left(\liminf_{n\to\infty} A_n^2(h) > 0\right) = 1$, that is, the proposed test is consistent against every possible alternative hypothesis. In other words,

$$\lim_{n\to\infty} \eta\left(A_n^2(h) > c_{1-\alpha,n}\right) = 1.$$

Also, since (3.4) is independent of h, the hypothesis testing of $H_0 : \mathbb{P} = \mathbb{P}_0$ against $H_1 : \mathbb{P} \neq \mathbb{P}_0$ is equivalent to testing, $H_{0,h} : \mathbb{P}_{\langle h\rangle} = \mathbb{P}_{0,\langle h\rangle}$ against $H_{1,h} : \mathbb{P}_{\langle h\rangle} \neq \mathbb{P}_{0,\langle h\rangle}$.

3.2   AD Test for Equality of Two Independent Random Fields

Before extending the method to the two-sample case of two Hilbert-valued random fields, let us recall the AD goodness of fit test for two independent real-valued samples. Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ be two independent real-valued random samples of two random variables X and Y. Assuming that F and G are two continuous distribution functions of X and Y, respectively, Baumgartner et al. (1998) suggested the two-sample AD type test as follows, which is available in the R package *kSamples*.

Let $r(X_i)$ be the rank of the observation $X_{(i)}$ in the ordered sample $\{X_{(n)}\}_{n \geqslant 1}$. Also, $r(Y_j)$ is the rank of the observation $Y_{(j)}$ in the ordered sample $\{Y_{(m)}\}_{m \geqslant 1}$. The AD test statistic of $H_0 : \mathbb{P} = \mathbb{Q}$ against $H_1 : \mathbb{P} \neq \mathbb{Q}$ is given by

$$B = \frac{1}{2}(B_X + B_Y), \tag{3.5}$$

where

$$B_X = \frac{1}{n} \sum_{i=1}^{n} \frac{\left[r(X_i) - \frac{m+n}{n}i\right]^2}{\frac{i}{n+1}\left(1 - \frac{i}{n+1}\right)\frac{m(m+n)}{n}},$$

and

$$B_Y = \frac{1}{m} \sum_{j=1}^{m} \frac{\left[r(Y_j) - \frac{m+n}{m}j\right]^2}{\frac{j}{m+1}\left(1 - \frac{j}{m+1}\right)\frac{n(m+n)}{m}}.$$

They proved these following properties of the proposed two-sample AD test statistic (3.5).

Theorem 3.4. *Assuming that $n, m \to \infty$ and $n/m \to c < \infty$, and the two random variables X and Y are defined on the probability space $(\Omega, \mathcal{A}, \eta)$. We have that*

$$\sqrt{\frac{1}{m(m+n)}}\left(r(X_i) - \frac{m+n}{n}i\right) \xrightarrow[\min(n,m)\to\infty]{\eta} Z,$$

$$\sqrt{\frac{1}{n(m+n)}}\left(r(Y_j) - \frac{m+n}{m}j\right) \xrightarrow[\min(n,m)\to\infty]{\eta} Z,$$

where $Z = \{Z(t)\}_t$ is a Brownian bridge, $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$. Also,

$$\lim_{\min(n,m)\to\infty} \eta\,(B < b) = \sqrt{\frac{\pi}{2}}\frac{1}{b}\sum_{j=0}^{\infty}\binom{-1/2}{j}(4j+1)\int_0^1 \frac{\exp\left\{\frac{ub}{8} - \frac{\pi^2(4j+1)^2}{8ub}\right\}}{\sqrt{u^3(1-u)}}\,du,$$

with

$$\binom{\frac{-1}{2}}{j} = \frac{(-1)^j \Gamma(j + \frac{1}{2})}{j! \, \Gamma(\frac{1}{2})},$$

is the generalized Euler symbol.

Now, we turn to the two-sample case in random field settings. Assuming that $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ are two independent random samples of two $\mathcal{H}$-valued random fields X and Y on $(\Omega, \mathcal{A}, \eta)$. We are interested in testing $H_0 : \mathbb{P} = \mathbb{Q}$ against $H_1 : \mathbb{P} \neq \mathbb{Q}$ using the one-dimensional random projection method. So, we generate an $\mathcal{H}$-valued random element h independently of $\mathbb{P}$ and $\mathbb{Q}$ according to a non-degenerate probability law ψ on $\mathcal{H}$, and then we test the projected null hypothesis

$$H_{0,h} : \mathbb{P}_{\langle h \rangle} = \mathbb{Q}_{\langle h \rangle}, \tag{3.6}$$

against

$$H_{1,h} : \mathbb{P}_{\langle h \rangle} \neq \mathbb{Q}_{\langle h \rangle}. \tag{3.7}$$

Therefore, the random samples of X and Y are randomly projected onto the one-dimensional subspace generated by h to get the real-valued random variables $X_i(h) = \langle X_i, h \rangle$ and $Y_j(h) = \langle Y_j, h \rangle$ for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$. Now, let $r(X_i(h))$ be the rank of the observation $\langle X_i, h \rangle$ in the ordered sample $\{\langle X, h \rangle_{(n)}\}$ and, similarly, $r(Y_j(h))$ is the rank of the observation $\langle Y_j, h \rangle$ in the ordered sample $\{\langle Y, h \rangle_{(m)}\}$.

Now, we define the two-sample one-dimensional projected AD statistic as follows.

$$B(h) = \frac{1}{2} \left(B_X(h) + B_Y(h) \right), \tag{3.8}$$

where

$$B_X(h) = \frac{1}{n} \sum_{i=1}^{n} \frac{\left[r(X_i(h)) - \frac{m+n}{n} i \right]^2}{\frac{i}{n+1} \left(1 - \frac{i}{n+1} \right) \frac{m(m+n)}{n}},$$

and

$$B_Y(h) = \frac{1}{m} \sum_{j=1}^{m} \frac{\left[r(Y_j(h)) - \frac{m+n}{m} j \right]^2}{\frac{j}{m+1} \left(1 - \frac{j}{m+1} \right) \frac{n(m+n)}{m}}.$$

Obviously, the statistic (3.8) is the test statistic of testing $H_{0,h}$, and as a result, it satisfies Theorem 3.4, and the proof of this result is similar to the proof of Theorem 3.3. In other words, $B(h)$ is independent of the choice of the random direction h.

In the next section, we will show that the projected AD statistic has a better performance than the projected KS statistic for one-sample and two-sample cases.

4   Simulation Studies

In this section, we present results from extensive series of simulation studies that performed by considering two-dimensional Hilbert-valued random fields observed on equispaced points in the unit square $[0, 1]^2$. Also, we generate the random projection direction, h, from a Brownian sheet. All simulation results were obtained from 1000 repetitions and then computing the empirical error rates, which are the power estimations of the projected test. In each repetition, we generate the sample and h, and hence there is no need to adjust the p-values derived from these simulations.

4.1   One-sample Case

Suppose that $X_1, X_2, \cdots, X_n$ is a random sample of realizations of a random field X with probability law $\mathbb{P}$. We are interested in testing $H_0 : \mathbb{P} = \mathbb{P}_0$ against $H_1 : \mathbb{P} \neq \mathbb{P}_0$, where $\mathbb{P}_0$ is a given Gaussian distribution law on $\mathcal{H}$. As mentioned before and after computing the scalar products $X(h) = \langle X, h \rangle$, testing H_0 is equivalent to test

$$H_{0,h} : X_1(h), \ldots, X_n(h) \overset{i.i.d}{\sim} \mathcal{N}(m_{X(h)}, \sigma^2_{X(h)}), \tag{4.1}$$

where

$$m_{X(h)} = \frac{1}{n} \sum_{i=1}^{n} X_i(h),$$

and

$$\sigma^2_{X(h)} = \frac{1}{n-1} \sum_{i=1}^{n} \left\{ X_i(h) - m_{X(h)} \right\}^2,$$

against its alternative. In order to study the performance of the proposed method, we consider multiple choices of $\mathbb{P}$ and test the Gaussianity of the projected samples.

4.1.1   First Scenario

In this scenario, we assume that X is a Gaussian random field with different covariance structures. In Table 1, we listed the covariance structures for the Gaussian random fields that we consider in our simulation studies i.e. *exponential*, *spherical*, *wave*, *circular* and *generalized Cauchy* (GC) covariance structures.

Table 1: Stationary isotropic covariance structures used in simulation studies to test H_0 : ($\mathbb{P}$ is a Gaussian random field). The amount $z \geqslant 0$ refers to the distance between points of the random field.

model	formula
exponential	$\exp(-z)$
spherical	$\left(1 - \frac{3}{2}z + \frac{1}{2}z^3\right) \mathbb{1}_{[0,1]}(z)$
wave	$\mathbb{1}_{\{0\}}(z) + \frac{\sin(z)}{z} \mathbb{1}_{[0,\infty)}(z)$
circular	$\left[1 - \frac{2}{\pi}\left(z\sqrt{1-z^2} + \arcsin(z)\right)\right] \mathbb{1}_{[0,1]}(z)$
GC	$(1 + z^{\alpha_X})^{-\beta_X/\alpha_X}$

Table 2: The empirical rejection rates for testing $H_{0,h} : X(h) \sim \mathcal{N}\left(m_{X(h)}, \sigma^2_{X(h)}\right)$ against $H_{1,h} : X(h) \nsim \mathcal{N}\left(m_{X(h)}, \sigma^2_{X(h)}\right)$. The p-values averages are recorded in parentheses.

n		G-exp	G-sph	G-wave	G-circ	B-sheet
25	rej(AD)	.000	.000	.000	.000	.000
		(.863)	(.865)	(.865)	(.856)	(.855)
	rej(KS)	.000	.000	.000	.000	.000
		(.794)	(.804)	(.796)	(.783)	(.792)
50	rej(AD)	.000	.000	.000	.000	.000
		(.856)	(.854)	(.862)	(.857)	(.854)
	rej(KS)	.000	.000	.001	.000	.000
		(.791)	(.792)	(.792)	(.782)	(.786)
100	rej(AD)	.000	.000	.000	.000	.000
		(.851)	(.851)	(.854)	(.850)	(.855)
	rej(KS)	.000	.000	.001	.000	.000
		(.800)	(.807)	(.804)	(.803)	(.803)
200	rej(AD)	.000	.000	.000	.000	.000
		(.844)	(.847)	(.850)	(.854)	(.854)
	rej(KS)	.000	.000	.000	.000	.000
		(.791)	(.788)	(.794)	(.800)	(.797)

In Table 2, rej(AD) and rej(KS) stand for the rates of rejections of projected AD and projected KS statistics, where $\alpha = .05$, and the p-values averages over 1000 repetitions (1000 independent trials) are listed in the parentheses. For sample sizes $n = 25, 50, 100$ and 200, Table 2 shows that, if $\mathbb{P}$ is a Gaussian probability law on $\mathcal{H}$, then the projected sample is a real Gaussian sample and this is true for all considered covariance structures. Also, when X is a Brownian sheet, the projection-based test statistics assure that the projected sample $X(h)$ is a set of observations of a Gaussian random variable.

Now, as we see in Table 1, the GFGCC is a Gaussian random field with the following covariance function

$$C(z, \alpha_X, \beta_X) = (1 + z^{\alpha_X})^{-\beta_X/\alpha_X} , \qquad (4.2)$$

where $\beta_X > 0$ and $\alpha_X \in (0, 2]$. In fact, the GFGCC has so many applications in geostatistics and many other areas (Cressie and Huang, 1999; Berizzi et al., 2004; Stein, 2005; Gneiting et al., 2006; Tscheschel et al., 2005; Mateu et al., 2007; Morariu et al., 2006; Stein, 2007). As stated in Lim and Teo (2009), the GFGCC is long-range dependence (LRD) if and only if $0 < \alpha_X \beta_X \leqslant N$, and otherwise it is short-range dependence (SRD) ($N = 2$ in our simulation studies). As a result, we are interested here in testing (4.1) by considering X with SRD and LRD. In other words, we consider cases where X is a non-stationary random field, and we want to evaluate the projected AD and KS test statistics in non-stationary settings.

As we note in Table 3, the one-dimensional projected test statistics show that the projected null hypothesis $H_{0,h}$ is supported, confirming that the random field X is a Gaussian random field.

As a result, the proposed one-dimensional projected tests are powerful in determining whether an underlying distribution law is Gaussian or non-Gaussian (as shown in the next scenario), in both LRD and SRD cases. All the empirical rejection rates are near zero and the p-value averages are close to 1 over 1000 repetitions.

Table 3: The empirical rejection rates for testing $H_{0,h}$ against $H_{1,h}$ at a level of significance $\alpha = .05$. X is a GFGCC with covariance structure $C(z, \alpha_X, \beta_X)$. The parentheses contain p-values averages.

n		$\alpha_X = .5$ &		$\alpha_X = 1$ &		$\alpha_X = 2$ &	
		$\beta_X = .5$	$\beta_X = 2$	$\beta_X = 1$	$\beta_X = 2$	$\beta_X = 2$	$\beta_X = 3$
50	rej(AD)	.000	.000	.000	.000	.000	.000
		(.852)	(.859)	(.854)	(.862)	(.856)	(.850)
	rej(KS)	.000	.000	.000	.000	.000	.000
		(.786)	(.793)	(.790)	(.792)	(.791)	(.784)
100	rej(AD)	.000	.000	.000	.000	.000	.000
		(.852)	(.858)	(.852)	(.849)	(.845)	(.859)
	rej(KS)	.000	.000	.000	.001	.000	.000
		(.800)	(.810)	(.806)	(.797)	(.798)	(.809)
200	rej(AD)	.000	.000	.000	.000	.000	.000
		(.843)	(.858)	(.847)	(.856)	(.845)	(.849)
	rej(KS)	.000	.000	.000	.000	.000	.000
		(.790)	(.805)	(.793)	(.803)	(.784)	(.800)

4.1.2 Second Scenario

In the current scenario, we test the Gaussianity of the two-dimensional random field X by testing (4.1) when the actual sample is generated from a Gaussian-based random field, namely, the t-student random field. Let t_ν denote the t-student random field with ν degrees of freedom. We will consider that t_ν has stationary isotropic exponential and Gaussian covariance structures, where the Gaussian covariance structure is given by $C(z) = \exp(-z^2); z \geqslant 0$.

Across 1000 repetitions, the results of the simulations are listed down in Table 4. As we see in this table, the empirical power estimation of testing (4.1) for the $t_{\nu=1}$ random field reaches its maximum value of 1. As ν gets larger and larger, the test power becomes lower and lower, and this is due to the fact that the t-student distribution converges to the Gaussian distribution for large degrees of freedom. Note that, the maximum power estimation value of the projection-based AD test statistic equals to .123, which is somewhat low, for $n = 200$ and $\nu = 5$.

Table 4: The empirical rejection rates for testing (4.1) by assuming that X is a t-student random field with $v = 1$ and 5 degrees of freedom, and by considering the exponential and Gaussian covariance structures. The results are rounded to three decimal places.

n		t_1-*exp*	t_1-*Gauss*	t_5-*exp*	t_5-*Gauss*
25	rej(AD)	.635	.619	.005	.004
		(.108)	(.115)	(.731)	(.758)
	rej(KS)	.650	.633	.007	.007
		(.101)	(.106)	(.684)	(.712)
50	rej(AD)	.942	.943	.017	.017
		(.012)	(.012)	(.669)	(.674)
	rej(KS)	.927	.940	.016	.017
		(.014)	(.013)	(.640)	(.643)
100	rej(AD)	1.000	.999	.044	.035
		(.000)	(.000)	(.516)	(.536)
	rej(KS)	.998	.999	.035	.025
		(.000)	(.000)	(.533)	(.547)
200	rej(AD)	1.000	1.000	.123	.112
		(.000)	(.000)	(.353)	(.358)
	rej(KS)	1.000	1.000	.084	.066
		(.000)	(.000)	(.397)	(.405)

4.1.3 Third Scenario

In the last scenario, X is considered as a χ^2 random field with generalized Cauchy, exponential and Gaussian covariance structures. Namely, we study the performance of the proposed projection-based test statistics in testing (4.1) when X is a χ^2 random field with $f = 4$ degrees of freedom and the covariance structure (4.2), for both SRD and LRD. Also, we test (4.1) when X is a χ^2 random field with stationary isotropic exponential and Gaussian covariance structures. Since X is not a Gaussian random field, that is, $\mathbb{P} \neq \mathbb{P}_0$, we expect that most empirical rejection rates tend to 1 as the sample size n gets larger.

Table 5 shows that, the values of the parameters α and β in (4.2) are the most

influencing factors in supporting or rejecting H_0. Also, as n increases, the estimations of the empirical power of the projected tests tend to 1 and the averages of p-values to 0.

Table 5: The empirical rejection rates for testing $H_{0,h}$ against $H_{1,h}$ at a level of significance .05. X is a $\chi^2_{f=4}$ random field with covariance structure $C(z, \alpha, \beta)$. The parentheses contain p-values averages.

n		$\alpha_X = .5$ &		$\alpha_X = 1$ &		$\alpha_X = 2$ &	
		$\beta_X = .5$	$\beta_X = 2$	$\beta_X = 1$	$\beta_X = 2$	$\beta_X = 2$	$\beta_X = 3$
50	rej(AD)	.024	.003	.046	.016	.045	.048
		(.445)	(.595)	(.397)	(.454)	(.348)	(.355)
	rej(KS)	.032	.009	.054	.019	.048	.049
		(.447)	(.575)	(.412)	(.450)	(.372)	(.375)
100	rej(AD)	.183	.041	.235	.151	.357	.303
		(.231)	(.402)	(.190)	(.250)	(.140)	(.169)
	rej(KS)	.150	.042	.160	.117	.232	.206
		(.275)	(.430)	(.238)	(.295)	(.193)	(.218)
200	rej(AD)	.658	.251	.721	.646	.857	.820
		(.087)	(.202)	(.087)	(.100)	(.045)	(.053)
	rej(KS)	.503	.191	.562	.491	.720	.657
		(.115)	(.247)	(.113)	(.126)	(.067)	(.077)

Finally, Table 6 shows that, for different sample sizes n, the underlying null hypothesis H_0 is always rejected, and the empirical power estimation reaches 1 in most cases. Note that the projection-based AD test statistic outperforms the KS one, especially for χ^2 with $f = 4$ degrees of freedom and large sample sizes ($n \geqslant 100$).

4.2 Two-sample Case

Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ (we take $m = n$) be random samples of two independent two-dimensional Hilbert-valued random fields X and Y, with distribution laws $\mathbb{P}$ and $\mathbb{Q}$, respectively. In these settings, we are interested in testing $H_0 : \mathbb{P} = \mathbb{Q}$ against $H_1 : \mathbb{P} \neq \mathbb{Q}$ using the one-dimensional projected AD and KS statistics.

Simulation results presented in the following two scenarios show that, the proposed projection-based test statistics have nice performance and can discriminate successfully

between H_0 and the alternatives.

Table 6: The empirical rejection rates for testing (4.1) by assuming that X is a χ^2 random field with $f = 1$ and 4 degrees of freedom, and by considering exponential and Gaussian covariances.

n		χ_1^2-exp	χ_1^2-Gauss	χ_4^2-exp	χ_4^2-Gauss
25	rej(AD)	.176	.233	.008	.008
		(.225)	(.199)	(.574)	(.573)
	rej(KS)	.164	.214	.015	.012
		(.239)	(.215)	(.551)	(.551)
50	rej(AD)	.688	.728	.053	.066
		(.066)	(.054)	(.368)	(.342)
	rej(KS)	.596	.633	.058	.061
		(.084)	(.067)	(.386)	(.371)
100	rej(AD)	.950	.964	.317	.369
		(.011)	(.007)	(.169)	(.147)
	rej(KS)	.930	.955	.226	.260
		(.016)	(.010)	(.222)	(.200)
200	rej(AD)	.993	.997	.824	.828
		(.001)	(.001)	(.057)	(.048)
	rej(KS)	.988	.992	.656	.686
		(.002)	(.001)	(.081)	(.070)

4.2.1 First Scenario

Suppose that $X_1, X_2, \cdots, X_n$ is a random sample from a centered Gaussian random field with isotropic covariance structure $C_1(z, \gamma) = \exp(-z^\gamma)$, where $\gamma \in (0, 2]$ is the stability parameter and $z \geqslant 0$ is the distance between two points of the field. Also, let $Y_1, Y_2, \cdots, Y_n$ be a random sample of an isotropic stationary Gaussian random field Y with covariance structure $C_2(z) = \exp(-z^2)$. Obviously, for $\gamma = 2$ the underlying null hypothesis is true, whereas for $0 < \gamma < 2$, the alternative hypothesis is true.

Figure 1 shows a plot of the covariance function $C_1(z, \gamma)$ for different values of γ. Note that $C_1(z, \gamma)$ is an increasing function in γ and $C_1(z, 2)$ looks like a one half of a bell-shaped function.
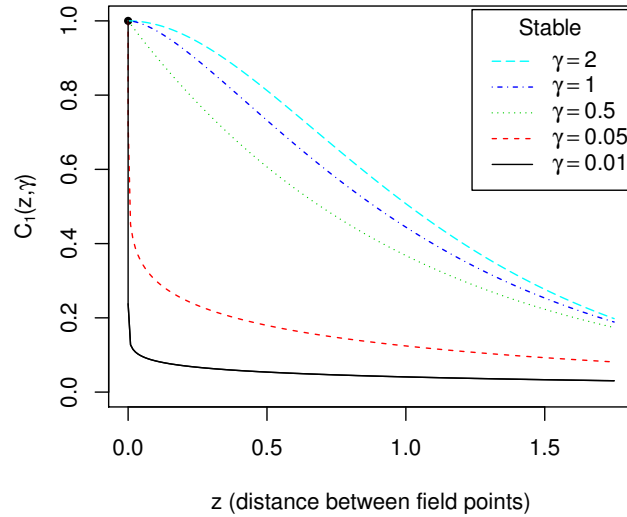
Figure 1: Plot of $C_1(z, \gamma)$ for different values of γ where z refers to the distance between two points of the field.

Now, under the null hypothesis (3.6), the rates of rejection should be close to $\alpha = .05$ and the projected p-values mean should be around .50. Likewise, away from H_0, the empirical test power should tend to 1 and the p-values mean to zero.

The results shown in Table 7 support our previous expectations over 1000 repetitions. As results in Table 7 reflect, the projected AD and KS test statistics discriminate successfully between the null and the alternative hypotheses. Obviously, the projected AD test outperforms the KS one.

Table 7: The empirical rejection rates for testing $H_{0,h} : \mathbb{P}_{\langle h \rangle} = \mathbb{Q}_{\langle h \rangle}$ against $H_{1,h} : \mathbb{P}_{\langle h \rangle} \neq \mathbb{Q}_{\langle h \rangle}$ and p-values average for different values of γ and multiple sample sizes n. The parentheses contain p-values averages.

n		0.01	0.05	0.08	0.1	$\gamma =$ 0.5	1	1.5	2
30	rej(AD)	.246	.176	.180	.176	.055	.069	.056	.046
		(.251)	(.295)	(.292)	(.292)	(.442)	(.475)	(.492)	(.517)
	rej(KS)	.149	.097	.096	.084	.041	.040	.038	.034
		(.374)	(.411)	(.413)	(.415)	(.543)	(.556)	(.563)	(.589)
50	rej(AD)	.351	.289	.295	.273	.096	.046	.036	.057
		(.174)	(.203)	(.207)	(.219)	(.373)	(.491)	(.503)	(.494)
	rej(KS)	.217	.175	.164	.149	.069	.037	.031	.039
		(.294)	(.325)	(.329)	(.338)	(.466)	(.560)	(.561)	(.558)
100	rej(AD)	.618	.540	.532	.512	.136	.064	.069	.044
		(.073)	(.105)	(.095)	(.108)	(.299)	(.443)	(.487)	(.504)
	rej(KS)	.359	.287	.292	.283	.076	.043	.042	.033
		(.166)	(.202)	(.199)	(.212)	(.403)	(.499)	(.537)	(.539)
200	rej(AD)	.926	.857	.857	.845	.319	.090	.044	.058
		(.016)	(.024)	(.029)	(.027)	(.174)	(.385)	(.489)	(.493)
	rej(KS)	.601	.539	.515	.505	.141	.060	.036	.039
		(.067)	(.090)	(.094)	(.094)	(.290)	(.448)	(.524)	(.523)

4.2.2 Second Scenario

In the current scenario, we take X and Y to be two-dimensional centered Gaussian random fields which differ in covariance structures. More specifically, we assume that X and Y are GFGCC's (4.2) with parameters α_X, β_X, α_Y and β_Y, respectively.

To assess the performance of the proposed AD and KS tests under H_0 and H_1, we considered cases when H_0 is true and others to take into account the alternative.

The results of the simulations (across 1000 repetitions) are shown in Table 8, Table 8, where the sample size is equal to $n = 100$. Under the null hypothesis, when $\alpha_X = \alpha_Y \in \{0.5, 2\}$ and $\beta_X = \beta_Y \in \{2, 3\}$, the empirical power is close to the significance level 0.05 and p-values average is close to 0.5, as expected.

Note that, for the test of equality of distribution of SRD ($\alpha_X = 2$, $\beta_X = 3$) and

LRD ($\alpha_Y = 0.5$ and $\beta_Y = 3$) random fields, the empirical power of AD test reaches its maximum value (0.997).

Table 8: The empirical rejection rates of testing the underlying hypothesis H_0, where X and Y are assumed to be two independent GFGCC's.

H_0		$\alpha_Y = 2$ & $\beta_Y = 2$	$\beta_Y = 3$	$\alpha_Y = 0.5$ & $\beta_Y = 2$	$\beta_Y = 3$
$\alpha_X = 2$	rej(AD)	.039	.048	.959	.989
		(.501)	(.466)	(.008)	(.003)
& $\beta_X = 2$	rej(KS)	.036	.042	.890	.954
		(.549)	(.512)	(.026)	(.010)
$\alpha_X = 2$	rej(AD)		.060	.979	.997
			(.494)	(.006)	(.002)
& $\beta_X = 3$	rej(KS)		.032	.896	.974
			(.538)	(.023)	(.005)
$\alpha_X = 0.5$	rej(AD)			.052	.290
				(.486)	(.178)
& $\beta_X = 2$	rej(KS)			.044	.125
				(.531)	(.301)
$\alpha_X = 0.5$	rej(AD)				.046
					(.509)
& $\beta_X = 3$	rej(KS)				.038
					(.547)

As expected, the projection-based statistic AD has more power than the projection-based KS statistic and, in total, each has a nice performance in discriminating completely and successfully between the null hypothesis and the alternatives.

In the next Section, we show how one-dimensional random projected AD and KS tests are applicable to a real dataset of autism spectrum disorder case-control study. Strictly speaking, we test Gaussianity of the autistic and Gaussianity of the healthy individuals, and then, we test the hypothesis that the autistic and healthy samples came from the same probability distribution.

5 Case Study

Statistical techniques have contributed significantly in medical studies. In effect, since early diagnosis is a very challenging task, researchers are seeking outperforming approaches to enhance the whole-brain classification (Song et al., 2015; Dorocic et al., 2014; Li et al., 2018; Tejwani et al., 2017; Hsu et al., 2015). These difficulties arise in neurodevelopmental, autism for example, and neuropsychiatric researches.

The autism spectrum disorder (ASD) is a brain-caused disorder diagnosed on the basis of social and iterative behaviors, where extensive psychological and scientific studies have engaged with determining the diagnostic criteria since the first recognition of ASD (Kanner, 1943).

The Autism Brain Imaging Data Exchange II (ABIDE II) dataset is an international neuroimaging data-sharing initiative that contains brain activation patterns of autistic and healthy individuals (controls). The sample dataset that we study includes a group of 15 autistic individuals and a group of 39 controls using high-resolution anatomical images, that is, using structural magnetic resonance (*MRI*) images.

The ABIDEII–GU_1 dataset contains three-dimensional *MRI* images or equivalently three-dimensional random fields, each with resolution $176 \times 256 \times 256$ pixels. But in order to reduce the computational burden, we will deal only with selected three-dimensional *MRI* images of resolution $106 \times 131 \times 71$ for each individual in the studied sample.

To study the performance of the proposed projection-based test statistics AD and KS, we randomly divided the control sample into two independent groups, namely, Hea1 and Hea2 groups with 19 and 20 healthy individuals, respectively.

At first, we test the Gaussianity of each group by testing the Gaussianity of the projected samples in the subspace generated by an isotropic stationary Gaussian random direction h with a Gaussian covariance structure $C(z) = \exp(-z^2)$. So, for each group Aut, Hea1 and Hea2, we test the null hypothesis $H_0^{(j)} : (\mathbb{P}_j$ is a Gaussian law of a three-dimensional random field) with $j = 1, 2, 3$. According to the previous discussions, this test is equivalent to testing the projected null hypothesis $H_{0,h}^{(j)} : (\mathbb{P}_{j,\langle h \rangle}$ is a Gaussian law of a real random variable) for $j = 1, 2, 3$. The results were obtained using four AMD Opteron processor 6386 SE (16 Core), total 64 core, running at 2.8 GHz and 128 GB of RAM and operating system CentOS 6.3. To check the robustness of these results, we repeated the projected Gaussianity tests 200 times and computed the rejection rates and the p-values average over these repetitions. Actually, the parallel

computing system we used failed to run Gaussianity tests with high repetitions, so we just repeated the calculations 200 times. Note that, in each repetition, the same dataset of each group is projected on a random direction h, that is, we are performing 200 comparison tests simultaneously.

As a result, we have to control the false discovery rate (FDR) using the Benjamini-Hochberg (BH) adjusted p-values of the projection-based AD statistic, and this is the same case for the projection-based KS statistic. The BH method (Benjamini and Hochberg, 1995) is one of the most popular FDR controlling procedures that is very useful in multiple testing studies. We summarize this controlling procedure here as follows:

Assuming that in $k = 200$ repetitions we generate, at random, k independent Brownian sheets $h_1, h_2, \cdots, h_k$ and compute the projected *p-value*, p_i, of testing the projected null hypothesis H_{0,h_i} for $i = 1, 2, \cdots, k$. Now, sort the *p-values* of these k independent one-dimensional tests as $p_{(1)} \leqslant p_{(2)} \leqslant \cdots \leqslant p_{(k)}$ and reject the underlying null hypothesis H_0 if the set $\left\{ i \in \{1, 2, \cdots, k\} \mid p_{(i)} \leqslant (.05i)/\sum_{i=1}^{k} \frac{k}{i} \right\}$ is not empty. In this way, the power of the univariate test of the underlying null hypothesis is increased while maintaining the significance level at $\alpha = .05$. This decision rule, which requires the knowledge of null distribution, is given by Benjamini and Hochberg (1995) and we used the correction formula mentioned in Theorem 1.3 of Benjamini and Yekutieli (2001) which is referred as the BY method.

So, we use the BY procedure to carry out these projection-based AD tests simultaneously and control the rate of false rejections at a level of significance of $\alpha = .05$ and this is also the case for KS. For comprehensiveness, we also included the results of applying the Cramér-von Mises test (CvM) to verify the Gaussianity of the projected data.

The results listed in Table 9 support the underlying null hypothesis that each group j is a sample of a three-dimensional Gaussian random field, for $j = 1, 2, 3$.

Now, the main goal of this section was to apply the one-dimensional random projection approach to test the null hypothesis of coincidence of the distributions of any two groups from {Aut, Hea1, Hea2}. That is, we want to test the null hypothesis $H_0^{(ij)} : \mathbb{P}_i = \mathbb{Q}_j$ against $H_1^{(ij)} : \mathbb{P}_i \neq \mathbb{Q}_j$, for $i \neq j$. Here, $\mathbb{P}_i$ is the probability distribution law of the three-dimensional random field which the sample i came from, and similarly define $\mathbb{Q}_j$; $i \neq j \in \{1, 2, 3\}$. All results are obtained by repeating the procedure of testing $H_{0,h}^{(ij)}$ for 200 times. The two-sample Cramér-von Mises test (CvM) is available in the R package

twosamples, so we also presented its one-dimensional projected results in Table 10.

In every cell of Table 10, we have listed down the empirical rejection rates of testing $H_{0,h}^{(ij)}$ using the one-dimensional projected AD, KS and CvM test statistics over 200 repetitions, where the p-values were adjusted and compared to $\alpha = .05$.

Table 9: Rejection rates of testing H_0 : (the individuals in a group are a sample of a Gaussian random field) according to the projected AD, KS and CvM statistics, and BY-adjusted p-value averages over 200 repetitions. The results are rounded to four decimal places.

	autistic sample	control sample	
	Aut	Hea1	Hea2
rej(AD)	.0000	.0000	.0000
	(.6792)	(.7294)	(.5796)
rej(KS)	.0000	.0000	.0000
	(.6207)	(.6470)	(.5562)
rej(CvM)	.0000	.0000	.0000
	(.6547)	(.6858)	(.5800)

Table 10 shows that, the control samples are identically distributed since the rejection rates of the projected AD or the projected KS and CvM tests are close to zero. Also, the autistic sample differs in distribution significantly from the other healthy samples. For example, the empirical power of testing H_0 :(Hea2 and Aut are identically distributed), equals .360 using the projected KS statistic and it reaches .370 according to the projected AD statistic. This confirms that both one-dimensional AD and one-dimensional CvM tests outperform the one-dimensional KS test in detecting differences between distributions.

Figure 2 shows an autistic and a controls brains, plotted using the *BrainNet Viewer* visualization tool (Xia et al., 2013).

Table 10: The BY-adjusted empirical rejection rates of testing $H_0^{(ij)}$: (the distribution that generated individuals dataset of i and j groups is the same), over 200 repetitions with $\alpha = .05$. The parentheses contain the adjusted p-values averages.

		Hea1	Hea2
Aut	AD	.565	.370
		(.083)	(.084)
	KS	.495	.360
		(.096)	(.103)
	CvM	.545	.365
		(.086)	(.090)
Hea1	AD		.000
			(.558)
	KS		.000
			(.568)
	CvM		.000
			(.563)

6   Conclusion

In this paper, we defined the Anderson-Darling projection-based test statistic for the stationary and non-stationary random fields. We also compared its performance with the Kolmogorov-Smirnov projection-based test statistic. Based on a variety of simulation studies and analyses of a real data sets, we conclude that the projected AD outperforms relatively than the projected KS as expected. It is also straightforward to define the Cramér-von Mises projection-based test statistic for random fields, as well as all other EDF tests, analogous to the projected AD test. Moreover, in the future work, we plan to define the projected Koziol-Green statistic for randomly censored data for random fields. Future studies are required to meet more information about the behavior of our approach in the functional random fields.
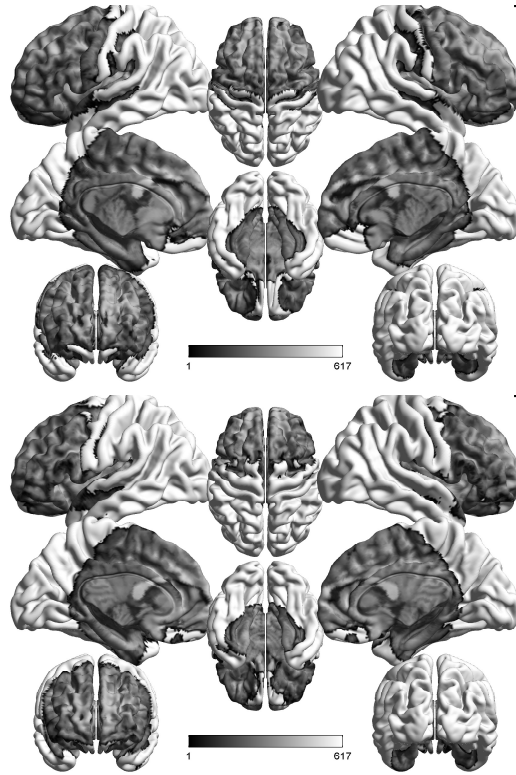
Figure 2: Visualizing brain figures of an autistic individual in the first row and a control in the second.

Acknowledgments

References

Achlioptas, D. (2003), Database-friendly random projections, Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, **66**(4), 671-687.

Adler, R. J. (1981). *The Geometry of Random Fields*. Wiley & Sons, New York.

Adler, R. J., and Taylor, J. E. (2004), *Random Fields and Geometry*. Springer.

Anderson, T. W., and Darling, D. A. (1952), Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, **23**(2), 193-212.

Baumgartner, W., Weiß, P., and Schindler, H. (1998), A nonparametric test for the general two-sample problem. *Biometrics*, **54**(3), 1129-1135.

Benjamini, Y., and Hochberg, Y. (1995), Controlling the false discovery rate, A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, **57**(1), 289-300.

Benjamini, Y., and Yekutieli, D. (2001), The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, **29**, 1165-1188.

Berizzi, F., Dalle Mese, E., and Martorella, M. (2004), A sea surface fractal model for ocean remote sensing. *International Journal of Remote Sensing*, **25**, 1265-1270.

Cressie, N. A. C., and Huang, H. C. (1999), Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**(448), 1330-1340.

Cuesta-Albertos, J. A., Fraiman, R., and Ransford, T. (2006), Random projections and goodness-of-fit tests in infinite dimensional spaces. *Bulletin of the Brazilian Mathematical Society*, **37**, 477-501.

Cuesta-Albertos, J. A., Fraiman, R., and Ransford, T. (2007a), A sharp form of the Cramér-Wold theorem. *Journal of Theoretical Probability*, **20**, 201-209.

Cuesta-Albertos, J. A., del Barrio, E., Fraiman, R., and Matrán, C. (2007b), The random projection method in goodness-of-fit for functional data. *Computational Statistics and Data Analysis*, **51**, 4814-4831.

Cuesta-Albertos, J. A., Gamboa, F., and Nieto-Reyes, A. (2009), A random-projection based procedure to test if a stationary process is Gaussian. *arXiv,* **1**,1-31.

Cuevas, A., and Fraiman, R. (2009), On depth measures and dual statistics. a methodology for dealing with general data. *Journal of Multivariate Analysis,* **100**, 753-766.

DasGupta, A. (2008), *Asymptotic theory of statistics and probability.* Springer Series in Statistics.

Di Bernardino, E., Estrade, A., and León, J. (2017), A test of Gaussianity based on the Euler characteristic on excursion sets. *Electronic journal of statistics, Shaker Heights, OH, Institute of Mathematical Statistics,* **11**(1), 843-890.

Dorocic, I.P., Fürth, D., Xuan, Y., Johansson, Y., Pozzi, L., Silberberg, G., Carlén, M., and Meletis, K. (2014), A whole-brain atlas of inputs to serotonergic neurons of the dorsal and median raphe nuclei. *Neuron,* **83**(3), 663-678.

Gneiting, T., Genton, M. G., and Guttorp, P. (2006), Geostatistical space-time models, stationarity, separability and full symmetry. *Monographs in Statistics and Applied Probability,* **107**. Chapman and Hall/CRC Press.

Hsu, C. Y. , Schneller, B., Ghaffari, M., Alaraj, A., and Linninger, A. (2015), Medical image processing for fully integrated subject specific whole brain mesh generation. *Technologies,* **3**, 126-141.

Johnson, B. W., and Lindenstrauss, J. (1984), Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics,* **26**, 189-206.

Kanner, L. (1943), Autistic disturbances of affective contact. *Nervous Child,* **2**, 217-250.

Lehmann, E. L. (1999), *Elements of large-sample theory*. Springer Texts in Statistics.

Lejsek, H., Ásmundsson, F. H., Jónsson, B., and Amsaleg, L. (2005), *Efficient and effective image copyright enforcement.* 21e journées Bases de données avancées.

Li, H., Parikh, N. A., and He, L. (2018), A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. *Frontiers in Neuroscience,* **12**, 491.

Lim, S. C., and Teo, L. P. (2009), Gaussian fields and gaussian sheets with generalized cauchy covariance structure. *Stochastic Processes and their Applications,* **119**(4), 1325-1356.

Mateu, J., Porcu, E., Christakos, G., and Bevilacqua, M. (2007), Fitting negative spatial covariances to geothermal field temperatures in nea kessani (greece). *Envoronmetrics,* **18**, 759-773.

Morariu, V. , Craciun, C., Neamtu, S., Iarinca, L., and Mihali, C. (2006), A fractal and longe-range correlation analysis of plant nucleus ultrastructure. *Romanian Journal of Biophysics,* **16**, 243-252.

Patilea, V., Sánchez-Sellero, C., and Saumard, M. (2012), Projection-based nonparametric testing for functional covariate effect. *arXiv:*1205.5578.

Song, X., Meng, L., Shi, Q., and Lu, H. (2015), Learning tensor-based features for whole-brain fMRI classification. *Medical Image Computing and Computer-Assisted Intervention (MICCAI),* 613-620.

Stein, M. L. (2005), Space-time covariance function. *Journal of the American Statistical Association,* **100**, 310-321.

Stein, M. L. (2007), Seasonal variations in the spatial-temporal dependence of total column ozone. *Envirometrics,* **18**, 71-86.

Stephens, M. A. (1974), Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association,* **69**, 730-737.

Tejwani, R. , Liska, A., You, H., Reinen, J., and Das, P. (2017), Autism classification using brain functional connectivity dynamics and machine learning. *NIPS workshop BigNeuro.*

Tscheschel, A., Lacayo, J., and Stoyan, D. (2005), Statistical characterization of TEM images of silica-filled rubber. *Journal of Microscopy,* **217**, 75-82.

Xia, M., Wang, J., and He, Y. (2013), Brain Net Viewer, a network visualization tool for human brain connectomics. *PLoS ONE,* **8**(7)(e68910).