

# Multivariate Normality test for colored data

El Bouch Sara

*Gipsa-Lab*

*Univ. Grenoble Alpes, France*

sara.el-bouch@grenoble-inp.fr

Michel J. J. Olivier

*Gipsa-Lab*

*Univ. Grenoble Alpes, France*

olivier.michel@grenoble-inp.fr

Comon Pierre

*Gipsa-Lab, CNRS,*

*Univ. Grenoble Alpes, France*

pierre.comon@grenoble-inp.fr

**Abstract**—Performances of the Multivariate Kurtosis are investigated when applied to colored data, with or without Auto-Regressive prewhitening, and with or without projection onto a lower-dimensional random subspace. Computer experiments demonstrate the importance of taking into account the possible color of the process in calculating the power of the normality test, in all the scenarios.

**Index Terms**—high-order statistics, multivariate kurtosis, detection of changes, dependence, Gaussianity

## I. INTRODUCTION

Detecting changes in the distribution of a stochastic process is a longstanding problem and a myriad of methods has been proposed; see e.g [1]. Our concern is the detection of non-Gaussian signals in a *Gaussian* background, *i.e.* *Normality tests*. The framework considered here is one in which time-series are recorded on  $d$  sensors (typically  $d = 2$  or  $3$  in many applications). Moreover, we are interested in the *online* problem of reacting to a change as quickly as possible after it occurs, also known as *sequential detection* problem [2], [3].

Normality tests belong to the the class of tests without alternative, unlike other approaches based on the Likelihood Ratio such as the CUSUM test [4], [5]. The case of i.i.d (scalar or  $d$ -variate) processes has received significant attention, see [6], [7] for a detailed survey. On the other hand, few tests concern the case of dealing with time-series from multiple sensors that in practical applications cannot be considered to be i.i.d, a case we will refer to as *n.i.d* or *colored*. Moreover, the proposed approaches for dealing with multivariate colored processes, such as [8] [9], can hardly be used in real-time applications. For this reason, we have derived a joint normality test for multivariate colored data [10] that is implemented with low computational burden.

There has also been interest in testing the normality of *unobserved* regression residuals [11], also referred to as the process of innovations. More precisely, define the Multi-dimensional Auto-Regressive [12] (or Vector AR of order  $p$  denoted VAR( $p$ )) model to describe the statistical behavior of the  $d \times 1$  vector of observation  $\mathbf{x}(i)$  for  $i = 1, \dots, N$ :

$$\mathbf{x}(i) = \sum_{k=1}^p \mathbf{A}_k \mathbf{x}(i-k) + \boldsymbol{\epsilon}(i) \quad (1)$$

where  $\mathbf{A}_k$  is a  $d \times d$  matrix of unknown parameters and  $\boldsymbol{\epsilon}(i)$  is the  $i$ th unobservable residual assumed zero-mean and

This work has been supported by the MIAI chair “Environmental issues underground” of Institut MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

i.i.d. An additional assumption is that residuals are drawn from a normal distribution. The drawbacks of violating the latter assumption has been studied, for instance [13] showed that the ordinary least squares method, which is usually used to estimate  $\{\mathbf{A}_k\}_{1 \leq k \leq p}$ , is sub-optimal for heavy-tailed distributions. Thus, it is important to validate this assumption of normality in this linear model.

Moreover, since the residuals are estimated in practice, we expect that errors in the model specification and estimation will impact the whitening performance of the filter and the estimated innovation process  $\hat{\boldsymbol{\epsilon}}$  could no longer be considered i.i.d. In this case, the normality tests designed for i.i.d processes become biased as shown in [7], highlighting the importance of deriving our joint normality test for variables that are not statistically independent.

Within this framework, we concentrate on the *Multivariate Kurtosis* (MK) defined in (3). In [10], we calculated the power of this test variable in the colored case, and in [14] we studied its performances when the observed multivariate process was projected onto an arbitrary subspace of low dimension (typically  $1-D$  or  $2-D$ ), in particular for time series generated by colored copulas. The main contributions of the present paper are now the following:

- We compare the performances of the MK test with and without linear prewhitening, *i.e.* using  $\mathbf{x}(n)$  or  $\boldsymbol{\epsilon}(n)$ .
- We observe the impact of projecting  $3-D$  observations and their *regression residuals* onto an arbitrary plane ( $2-D$  projection) or direction ( $1-D$  projection).
- For sake of time and memory effectiveness, the test statistics and the regression function are estimated recursively, by using both exponential averaging and the Recursive Least Squares (RLS) algorithm.

The paper is organized as follows: our recent results [10] are summarized in Section II. The recursive implementation is described in Section III, and the complete algorithm in Section IV. Computer results are eventually reported in Section V.

## II. ASYMPTOTIC STATISTICS OF THE KURTOSIS

Let  $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_d(n)]^T \in \mathbb{R}^d$  be a real-valued  $d$ -variate random variable. Let  $\mathbf{S}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}(n-\tau)^T\}$  be the covariance function whose entries are  $S_{ab}(\tau)$ . Also denote  $\mathbf{S}(0) = \mathbf{S}$ .

Let  $\mathbf{X} \stackrel{\text{def}}{=} \{\mathbf{x}(1), \dots, \mathbf{x}(N)\}$  be a stochastic stationary process of  $N$  random variables  $\mathbf{x}(n)$ . Our aim is to test:

$$H_0 : \mathbf{X} \underset{n.i.d.}{\sim} \mathcal{N}(0, \mathbf{S}) \quad \text{versus} \quad \bar{H}_0 \quad (2)$$

Where  $\mathbf{x}(n)$  are identically but not independently distributed (n.i.d). Following Mardia's definition of the Multivariate Kurtosis (MK) [15], our test statistic reads:

$$\hat{B}_d(N) = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}(n)^T \hat{\mathbf{S}}^{-1} \mathbf{x}(n))^2 \quad (3)$$

with

$$\hat{\mathbf{S}} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}(k) \mathbf{x}(k)^T \quad (4)$$

#### A. Known results: Mardia's MK test for a i.i.d. process

The following results make use of Landau's notations  $o(\frac{1}{N})$  and  $O(\frac{1}{N})$ , to precise that the absolute approximation error is dominated by  $\frac{1}{N}$  or is of the order of  $\frac{1}{N}$ , respectively.

*Theorem 2.1:* [15] Let  $\mathbf{X}$  be an i.i.d. process of dimension  $d$ . If  $\mathbf{X} \underset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{S})$  then  $\hat{B}_d(N) \xrightarrow[N \rightarrow \infty]{} \mathcal{N}(\mu, \sigma^2)$ , with:

$$\mu = d(d+2) + o\left(\frac{1}{N}\right) \quad (5)$$

$$\sigma^2 = \frac{8d(d+2)}{N} + o\left(\frac{1}{N}\right) \quad (6)$$

In [10] we devised a similar theorem for multivariate data that are no longer considered i.i.d. In the following, we sketch the main assumptions to identify the statistics of Mardia's MK test in the multivariate n.i.d. case.

#### B. Joint normality test for a colored process

Proofs of the results presented in this subsection can be found in [10], as well as theoretical details.

1) *Mathematical Preliminaries:* In order to guarantee convergence while relaxing the i.i.d assumption, we make the mixing assumption:  $\sum_{\tau=0}^{\infty} |S_{ab}(\tau)|^2$  converges to a finite limit  $\forall (a, b) \in \{1, \dots, d\}^2$ .

Let  $\hat{\mathbf{S}} = \mathbf{S} + \mathbf{\Delta}$ , where  $\mathbf{\Delta}$  is small compared to  $\mathbf{S}$ , then we have the following lemma:

*Lemma 2.1:* The entries of matrix  $\mathbf{\Delta}$  are of order  $O(1/\sqrt{N})$ . Let  $\mathbf{G} = \mathbf{S}^{-1}$ . This lemma allows the following expansion of the precision matrix  $\hat{\mathbf{G}} \stackrel{\text{def}}{=} \hat{\mathbf{S}}^{-1}$  up to second order in  $\mathbf{\Delta}$ :

$$\hat{\mathbf{G}} = \mathbf{G} - \mathbf{G}\mathbf{\Delta}\mathbf{G} + \mathbf{G}\mathbf{\Delta}\mathbf{G}\mathbf{\Delta}\mathbf{G} + o(1/N). \quad (7)$$

Finally, by injecting (7) in  $\hat{B}_d(N)$  defined in (3), an approximation of the test statistic in  $o(N^{-1})$  may be obtained, as  $\hat{B}_d$  can be shown to converge to a normal variable. Thus we only need to calculate its mean and variance, whose expressions are given below.

#### 2) Statistics for a colored scalar process:

$$\mathbb{E}\{\hat{B}_1\} = 3 - \frac{6}{N} - \frac{12}{N^2} \sum_{\tau=1}^{N-1} (N-\tau) \frac{S(\tau)^2}{S^2} + o\left(\frac{1}{N}\right) \quad (8)$$

$$\text{Var}\{\hat{B}_1\} = \frac{24}{N} \left[ 1 + \frac{2}{N} \sum_{\tau=1}^{N-1} (N-\tau) \frac{S(\tau)^4}{S^4} \right] + o\left(\frac{1}{N}\right) \quad (9)$$

The dependence between time samples is taken into account in the covariance function terms  $S(\tau)$ . In the case of i.i.d samples, we recover the expressions given in Theorem 2.1 above, for  $d = 1$ .

#### 3) Statistics for a bivariate colored process:

$$\mathbb{E}\{\hat{B}_2\} = 8 - \frac{16}{N} - \frac{4}{N^2} \sum_{\tau=1}^{N-1} \frac{(N-\tau)Q_1(\tau)}{(S_{11}S_{22} - S_{12}^2)^2} + o\left(\frac{1}{N}\right) \quad (10)$$

$$\text{Var}\{\hat{B}_2\} = \frac{64}{N} + \frac{16}{N^2} \sum_{\tau=1}^{N-1} \frac{(N-\tau)Q_2(\tau)}{(S_{11}S_{22} - S_{12}^2)^4} + o\left(\frac{1}{N}\right) \quad (11)$$

where  $Q_i(\tau)$  are linear combinations of  $S_{ab}(\tau)$  thus containing information about the time-dependence between samples. Despite the difficulty to derive such expressions, the computational burden involved by their implementation is low.

### III. TIMES SERIES MODEL: MULTIDIMENSIONAL AUTO-REGRESSIVE MODEL

#### A. Definition

The VAR( $p$ ) model defined in (1) can be written in matrix format as:

$$\mathbf{Y} = \mathbf{Z}\mathbf{w} + \boldsymbol{\epsilon}, \quad (12)$$

where,  $\forall i \in \{p, \dots, N-1\}$ :

$$\begin{aligned} \mathbf{w} &= [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p]^T \in \mathbb{R}^{dp \times d} \\ \mathbf{Y} &= [\mathbf{x}(p+1), \dots, \mathbf{x}(N)]^T \in \mathbb{R}^{(N-p) \times d} \\ \mathbf{Z} &= [\mathbf{z}(p), \mathbf{z}(p+1), \dots, \mathbf{z}(N-1)]^T \in \mathbb{R}^{(N-p) \times dp} \\ \mathbf{z}(i) &= [\mathbf{x}(i); \mathbf{x}(i-1); \dots; \mathbf{x}(i-p+1)] \in \mathbb{R}^{dp \times 1} \\ \boldsymbol{\epsilon} &= [\boldsymbol{\epsilon}(p+1), \boldsymbol{\epsilon}(p+2), \dots, \boldsymbol{\epsilon}(N)]^T \in \mathbb{R}^{(N-p) \times d} \end{aligned}$$

A Least Squares estimation can be performed to find  $\hat{\mathbf{w}}$  that minimizes  $\|\mathbf{Z}\mathbf{w} - \mathbf{Y}\|_2$  assuming that  $\mathbf{\Gamma} = \mathbf{Z}^T \mathbf{Z}$  is non-singular:

$$\hat{\mathbf{w}} = \mathbf{\Gamma}^{-1} \mathbf{Z}^T \mathbf{Y} \quad (13)$$

#### B. Recursive Estimation of VAR( $p$ ) parameters

To estimate the regression parameters, a classical Recursive Least Squares methods (RLS) is proposed. This adaptive strategy is well known to be suitable for data showing some sort of non-stationarity, and has been widely used (see e.g. [16] [17], [18] and [19] to only cite few).

Below, the main RLS equations are given for the present problem. Let  $\lambda_1$  be the decay factor weighting the observations, and  $\hat{\mathbf{w}}_k$  be the  $k^{\text{th}}$  column of  $\hat{\mathbf{w}}$  defined in (13), and  $\mathbf{y}_k(t-1) = (x_k(1), x_k(2), \dots, x_k(t-1))^T$ .

$$\mathbf{y}_k(t-1) = \mathbf{Z} \hat{\mathbf{w}}_k(t-1) + \boldsymbol{\epsilon}_k(t-1) \quad (14)$$

$$\hat{\mathbf{w}}_k(t-1) = \mathbf{\Gamma}^{-1}(t-1) \mathbf{Z}^T \mathbf{y}_k(t-1) \quad (15)$$

Suppose we want to update the model with new observations  $\mathbf{x}(t)$ ; a new row  $\mathbf{z}(t)^T$  is appended to  $\mathbf{Z}$  in (13), and a new observation  $x_k(t)$  is appended to  $\mathbf{y}_k$ . In the RLS algorithm,  $\mathbf{\Gamma}^{-1}(t)$  and  $\hat{\mathbf{w}}_k(t)$  are recursively expressed for  $t > N$  as

$$\Gamma^{-1}(t) = \left( \lambda_1 \Gamma(t-1) + \mathbf{z}(t)\mathbf{z}(t)^T \right)^{-1} \quad (16)$$

$$= \lambda_1^{-1} \Gamma^{-1}(t-1) - \mathbf{b}\mathbf{u}\mathbf{u}^T \quad (17)$$

$$\begin{aligned} \hat{\mathbf{w}}_k(t) &= \Gamma^{-1}(t) \tilde{\mathbf{Z}}^T \tilde{\mathbf{y}}_k \\ &= \hat{\mathbf{w}}_k(t-1) - b(\mathbf{z}^T(\hat{\mathbf{w}}_k(t-1) + x_k(t)\mathbf{u}))\mathbf{u} + x_k(t)\mathbf{u}, \end{aligned} \quad (18)$$

$\mathbf{u} = \lambda_1^{-1} \Gamma^{-1}(t-1)\mathbf{z}(t)$  and  $b = (1 + \mathbf{z}(t)^T \mathbf{u})^{-1}$ . Note that  $\tilde{\mathbf{Z}}$  is  $\mathbf{Z}$  in (13) augmented with the row  $\mathbf{z}(t)^T$ , and  $\tilde{\mathbf{y}}_k = [\mathbf{y}_k(t-1); x_k(t)]$ . Note that we can also vary  $\lambda_1$  and the order  $p$  of the model to allow for more flexibility and adapt to local non-stationarities of the data, see e.g [20], [21]. This is not detailed in this communication.

#### IV. CHANGE DETECTION ALGORITHM

##### A. The test in practice

First we choose the nominal level of the test:

$$\alpha = \mathbb{P}(\text{choose } \tilde{H}_0 | H_0 \text{ is true}).$$

- It has been shown in previous sections that  $z = (\hat{B}_d - \mathbb{E}\{\hat{B}_d\})/\sqrt{\text{Var}\{\hat{B}_d\}}$  is asymptotically normal.
- We reject the null hypothesis  $H_0$  at a significance level  $\alpha$  if:

$$2(1 - \Phi(z)) < \alpha$$

where  $\Phi$  denotes the cumulative distribution function (cdf) of  $\mathcal{N}(0, 1)$ .

- In practice, the time-structure is unknown and the covariance (matrix) function entries  $S_{ab}(\tau)$  are replaced by their sample counterparts at time  $t$ .

##### B. Recursive normality test on regression residuals

The proposed test statistic is computationally efficient i.e it can be easily computed over a sliding window or by using the exponential weighting technique to test the normality of the estimated regression residuals available at time  $t$  assumed to follow a Gaussian distribution  $\mathcal{N}(0, \mathbf{V}(t))$ . Let  $0 < \lambda_2 < 1$ :

$$\mathbf{V}(t) = \lambda_1 \mathbf{V}(t-1) + (1 - \lambda_1) \hat{\boldsymbol{\epsilon}}(t) \hat{\boldsymbol{\epsilon}}(t)^T \quad (19)$$

$$\hat{B}_d(t) = \lambda_2 \hat{B}_d(t-1) + (1 - \lambda_2) (\hat{\boldsymbol{\epsilon}}(t)^T \hat{\mathbf{V}}^{-1}(t) \hat{\boldsymbol{\epsilon}}(t))^2. \quad (20)$$

The algorithm for sequentially detecting changes reads:

**Require:**  $p \geq 1, 0 < \lambda_1, \lambda_2 < 1, \alpha, \delta$

**Initialization:**  $\Gamma(p) \leftarrow \delta \mathbf{I}_{dp}, \hat{B}_d(p) \leftarrow 0, \mathbf{V}(p) \leftarrow \mathbf{I}_d$

**for**  $p+1 \leq t \leq N$  **do**

Update  $\Gamma^{-1}(t)$  ▷ using (16)

Update  $\{\hat{\mathbf{A}}_k(t)\}_{1 \leq k \leq p}$  ▷ using (18)

Compute  $\hat{\boldsymbol{\epsilon}}(t) = \mathbf{x}(t) - \sum_{k=1}^p \hat{\mathbf{A}}_k \mathbf{x}(t-k)$

Compute  $z = (\hat{B}_d(t) - \mathbb{E}\{\hat{B}_d(t)\})/\sqrt{\text{Var}\{\hat{B}_d(t)\}}$

**if**  $2(1 - \Phi(z)) < \alpha$  **then**

Change is detected

**else**

No change

**end if**

**end for**

Forgetting factors  $\lambda_1$  and  $\lambda_2$  (for 2nd and 4th order statistics respectively) are usually chosen by a rule of thumb, depending on the time-scale of the change. To give a better intuition of this factor, one can calculate the length  $N_{\text{eff}}$  of a uniform sliding window that would yield the same estimator variance, when data are i.i.d. normal. This leads to  $\lambda \propto 1/N_{\text{eff}}$ . More details will be provided in a subsequent paper.

For the RLS algorithm, a common choice is  $\lambda_1 = 0.99$  which corresponds to using a virtual window's size  $N_{\text{eff}} = \frac{2}{1-\lambda_1}$ . As for the recursive estimation of the Kurtosis, to prevent the problem of slow convergence, we choose  $N_{\text{eff}} = 1000$ .

#### V. COMPUTER RESULTS

A set of Monte Carlo simulations is presented to compare the power of the proposed normality test when applied directly on data or on regression residuals. Then, we study the performance of the test statistic on a low-dimensional (2-D or 1-D) projection of the initial multivariate data. As a final illustration of the effectiveness of our method, we apply the change detection Algorithm presented in subsection IV-B on synthetic colored data undergoing an abrupt change in its distribution.

##### A. Applying the test on colored data

1) *Directly on data:*  $M = 2000$  simulations are considered, each being based on a sequence of  $N = 1000$  samples. First, we simulate 1-D AR( $p$ ) processes (for  $p \in \{4, 14, 20\}$ ). The AR coefficients are computed such that the equivalent filter is low-pass, with band pass equal to .25 (normalized freq).

For each simulation a 2-D Gaussian (or Uniform) AR( $p$ ) process is constructed by time embedding:  $\mathbf{x}(t) = \{x(2t), x(2t+1)\}$ . Then, both Mardia's test derived for i.i.d. samples (denoted  $\hat{B}_{1,i.i.d}$ ) and the test for colored samples, whose statistics are defined by equations (8, 9) for colored samples (denoted  $\hat{B}_1$ ) are applied on the marginals of the 2-D process, and compared. Finally, the statistics derived for n.i.d. bi-variate data  $\hat{B}_2$ , described by equations (10, 11) is also applied to the 2-D process.

The obtained empirical rejection rates of hypothesis  $H_0$  are computed as  $\frac{\# \text{Rejections}}{M}$ , for a test level  $\alpha = 5\%$ . The results are summarized in Tables I and II.

TABLE I  
EMPIRICAL REJECTION RATES FOR 2000 SIMULATIONS FOR  $\alpha = 5\%$  SIGNIFICANCE LEVEL WITH THE  $\hat{B}_{1,i.i.d}, \hat{B}_1, \hat{B}_2$  TEST APPLIED DIRECTLY ON AR( $p$ ) DATA WITH  $p = 4, 14$

Test Statistic	AR(4)		AR(14)	
	Gaussian	Uniform	Gaussian	Uniform
$\hat{B}_{1,i.i.d}$	0.067	1.	0.123	0.512
$\hat{B}_1$	0.052	0.99	0.045	0.456
$\hat{B}_2$	0.06	1.	0.065	0.88

- It seems that  $\hat{B}_{1,i.i.d}$  has a better detection power than  $\hat{B}_1$ . As a matter of fact, by comparing the formulas

TABLE II  
EMPIRICAL REJECTION RATES FOR 2000 SIMULATIONS FOR  $\alpha = 5\%$  SIGNIFICANCE LEVEL WITH THE  $\hat{B}_{1,i.i.d}$ ,  $\hat{B}_1$ ,  $\hat{B}_2$  TEST APPLIED DIRECTLY ON AR(20) PROCESS

Test Statistic	AR(20)	
	Gaussian	Uniform
$\hat{B}_{1,i.i.d}$	0.228	0.430
$\hat{B}_1$	0.047	0.399
$\hat{B}_2$	0.06	0.688

of their variance in (5, 9), we can see that that the variance of  $\hat{B}_{1,i.i.d}$  is underestimated, consequently, the latter over-rejects the hypothesis of Gaussianity. This is noticeable even for Gaussian AR( $p$ ) process (Under  $H_0$ ), with rejection rates that surpass the nominal level 5%.

- Both scalar tests  $\hat{B}_{1,i.i.d}$  and  $\hat{B}_1$  perform poorly compared to the joint normality test statistic  $\hat{B}_2$ .
- When the correlation tails last longer (Table II), the overall performance of the test statistics tends to decrease.

2) *On Regression residuals*: We now study the power of the normality tests on *estimated* regression residuals. We utilize the Least Squares method to obtain  $\hat{\epsilon}$ . The simulation procedure and the tests are the same as those described in the previous paragraph V-A1. We study the case where the order  $p$  of the generated AR( $p$ ) process is known (we choose  $p = 20$ ), and the case where the order is misspecified ( $\hat{p} = 9$ ). The empirical rejection rates are summarized in Table III.

TABLE III  
EMPIRICAL REJECTION RATES FOR 2000 SIMULATIONS FOR  $\alpha = 5\%$  SIGNIFICANCE LEVEL WITH THE TEST STATISTICS  $\hat{B}_{1,i.i.d}$ ,  $\hat{B}_1$ ,  $\hat{B}_2$  APPLIED ON ESTIMATED REGRESSION RESIDUALS USING OLS METHOD, WITH KNOWN AND MISSPECIFIED ORDER

Test Statistic	On Residuals of AR(20) $p = 20$		On Residuals of AR(20) $\hat{p} = 9$	
	Gaussian	Uniform	Gaussian	Uniform
$\hat{B}_{1,i.i.d}$	0.06	1.	0.064	0.582
$\hat{B}_1$	0.05	1.	0.051	0.429
$\hat{B}_2$	0.055	1.	0.06	0.850

- If the model is perfectly known, then all test statistics perform well on the well estimated regression residuals. However, if the model order is under estimated, some important time correlation remain, and scalar tests perform poorly compared to the joint normality test, as expected.

### B. Random Projections

We simulate a 3- $D$  process VAR( $p$ ),  $p \in \{5, 20\}$ , of length  $N = 1000$  following equation (1), where the inputs  $\epsilon(t)$  are i.i.d with distributions either multivariate standard normal  $\mathcal{N}(0, 1)$  or multivariate  $\mathcal{U}(-2, 2)$ .

Then  $M = 2000$  different projections on an arbitrary plane (2- $D$  projection) going through the origin of the 3- $D$  space are computed, corresponding to as many 2- $D$  time series. For comparison, the same set of observations is also projected  $M$  times on an arbitrary direction (1- $D$  projection). Eventually,

we run the same set of experiments on estimated regression residuals estimated by ordinary least squares method (OLS). The results are reported below.

TABLE IV  
EMPIRICAL REJECTION RATES OF THE TEST STATISTICS APPLIED TO A LOW REPRESENTATION OF 3- $D$  VAR( $p$ ) PROCESS

2- $D$ projection	3- $D$ VAR(5)		1- $D$ projection	3- $D$ VAR(5)	
	Gaussian	Uniform		Gaussian	Uniform
$\hat{B}_2$	0.051	0.986	$\hat{B}_1$	0.056	0.529

TABLE V  
EMPIRICAL REJECTION RATES OF THE TEST STATISTICS APPLIED TO A LOW REPRESENTATION OF 3- $D$  VAR(20) PROCESS WITH UNIFORM INPUTS AND ITS ESTIMATED REGRESSION RESIDUALS WITH A VAR(10) MODEL

1- $D$ proj., $\hat{B}_1$	3- $D$ VAR(20)	On residuals $\hat{p} = 10$
	2- $D$ proj., $\hat{B}_2$	0.250
	0.580	0.9

- The test  $\hat{B}_1$  applied directly on 1- $D$  projections of either the observations or its residuals computed by OLS, performs poorly. This is in accordance with our observations from the preceding experiments.
- Even with a misspecified order, the joint test statistic performs best (empirical power of 90%) on the 2- $D$  low representation of regression residuals, as it is able to account for both temporal and spatial (between coordinates) dependences.

### C. Change detection on synthetic data

Consider the case where a process  $\epsilon$  is constituted of i.i.d samples following a standard normal distribution. The process undergoes a change at  $n_c = 5000$  in its distribution: samples in the interval  $[5000, 10000]$  are now following a uniform distribution  $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ . The change ends at  $n = 10000$  and the samples are again normally distributed. The process  $\epsilon$  is then (low pass) filtered using an AR(5) model and is now denoted  $x$ .

One realization of this process is given in Fig. 1. It is clear from this figure that the change in the distribution is unbeknownst to the human eye.

The change detection algorithm presented in subsection IV-B is applied to this realization by setting  $p = 5$ ,  $\lambda_1 = 0.99$ ,  $\lambda_2 = 0.998$ ,  $\alpha = 5\%$  and  $\delta = 1$ . A 2- $D$  process is obtained by taking  $\mathbf{x}(t) = \{x(2t), x(2t + 1)\}$ .

For comparison, the CUSUM algorithm [2] is applied on the regression residuals by using its recursive form. The instantaneous log-likelihood ratio is computed as:

$$L(t) = -\ln(2\sqrt{3}) + \frac{1}{2} \ln(2\pi) + \frac{1}{2} \hat{\epsilon}^2(t) \quad (21)$$

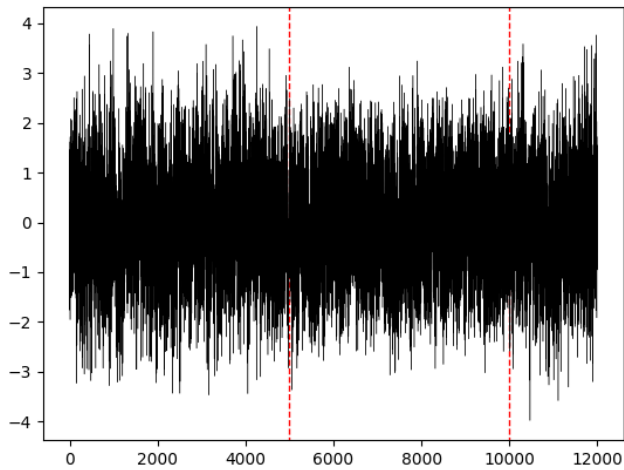


Fig. 1. One realization of a Gaussian AR(5) process that undergoes an abrupt change in the distribution of its excitation  $\epsilon$  (from  $\mathcal{N}(0, 1)$  to  $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ ). Affected samples are between red dashed lines.

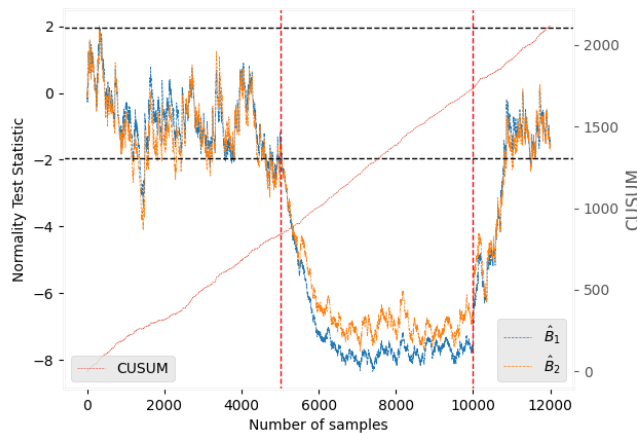


Fig. 2. Evolution of the normalized test statistics  $\hat{B}_1$  (in blue) and  $\hat{B}_2$  (in orange). In red, the evolution of the cumulative sum  $s(n) = \sum_{t=1}^n L(t)$ . Black horizontal dashed lines are the critical values  $\pm 1.96$  corresponding to a test power of 95%. Red vertical dashed lines are the beginning and end of an abrupt change in the excitation statistics of an AR(5) process.

- Our proposed test statistics  $\hat{B}_{1,2}$  stay in between  $\pm 1.96$ ; in other words they do not reject the null hypothesis of Gaussianity with a false alarm rate of 5%. They grow continuously in absolute value after the change time  $n_c = 5000$  and keep rejecting  $H_0$  until the end of the change at  $n = 10000$ .
- There is no clear-cut in the cumulative sum algorithm that indicates a change in the distribution of the residuals. In fact, the likelihood ratio in (21) is derived under the hypothesis that the process  $\epsilon$  is i.i.d. As the latter's values are estimated recursively, they are more likely to have residual correlations between them and the hypothesis of *independence* no longer holds, explaining why the cumulative sum of  $L(t)$  keeps increasing.

## VI. CONCLUDING REMARKS

A novel approach to testing the Gaussianity of colored data is proposed. In a first stage, data are whitened recursively.

Then the MK test is applied in a second stage on available regression residuals in an *online* manner. Computer experiments have been performed to assess the impact of the preprocessing stage on the power of the test. They have evidenced the fact that the bivariate MK statistics we recently obtained (assuming both *spatial* and *temporal* dependence) perform always better than their scalar counterparts. The latter are indeed severely affected by serial correlation in residuals due to imperfect whitening in the first stage. The same conclusions can be drawn if data are randomly projected onto a low-dimensional subspace.

Future work will concern the validation of the change detection algorithm on real-world phenomena recorded by a network of  $d$ -variate sensors within a multiple hypothesis testing framework.

## REFERENCES

- [1] M. Basseville, "Detecting changes in signals and systems—a survey," *Automatica*, vol. 24, no. 3, pp. 309–326, 1988.
- [2] M. Basseville, I. V. Nikiforov *et al.*, *Detection of abrupt changes: theory and application*. prentice Hall Englewood Cliffs, 1993, vol. 104.
- [3] H. V. Poor and O. Hadjiladis, *Quickest detection*. Cambridge University Press, 2008.
- [4] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [5] D. V. Hinkley, "Inference about the change-point from cumulative sum tests," *Biometrika*, vol. 58, no. 3, pp. 509–523, 1971.
- [6] R. Henze, "Invariant tests for multivariate normality: a critical review," *Statistical papers*, vol. 43, pp. 467–506, 2002.
- [7] D. S. Moore, "The effect of dependence on chi squared tests of fit," *The Annals of Statistics*, vol. 10, no. 4, pp. 1163–1171, 1982.
- [8] M. Hinich, "Testing for Gaussianity and linearity of a stationary time series," *Jour. Time Series Analysis*, vol. 3, no. 3, pp. 169–176, 1982.
- [9] E. Moulines, K. Choukri, and M. Charbit, "Testing that a multivariate stationary time series is Gaussian," in *Sixth SSAP Workshop on Stat. Signal and Array Proc.*, Oct. 1992, pp. 185–188.
- [10] S. ElBouch, O. Michel, and P. Comon, *A Normality Test for Multivariate Dependent Samples*, Sep. 2021, hal-03344745. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-03344745>
- [11] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *Int. Statistical Review*, pp. 163–172, 1987.
- [12] H. Lütkepohl, *Introduction to multiple time series analysis*. Springer Science & Business Media, 2013.
- [13] R. V. Hogg, "Statistical robustness: One view of its use in applications today," *The American Statistician*, vol. 33, no. 3, pp. 108–115, 1979.
- [14] S. ElBouch, O. Michel, and P. Comon, "Joint Normality Test Via Two-Dimensional Projection," in *ICASSP*, Singapore, May 2022. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-03369151>
- [15] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, pp. 519–530, 1970.
- [16] R. L. Plackett, "Some theorems in least squares," *Biometrika*, vol. 37, no. 1/2, pp. 149–157, 1950.
- [17] T. S. Rao, "The fitting of non-stationary time-series models with time-dependent parameters," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 32, no. 2, pp. 312–322, 1970.
- [18] M. Hallin, "Mixed autoregressive-moving average multivariate processes with time-dependent coefficients," *Journal of Multivariate Analysis*, vol. 8, no. 4, pp. 567–572, 1978.
- [19] Y. Grenier, "Time-dependent arma modeling of nonstationary signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 4, pp. 899–911, 1983.
- [20] C. D. Giurcaneanu and S. A. Razavi, "AR order selection in the case when the model parameters are estimated by forgetting factor least-squares algorithms," *Signal Processing*, vol. 90, no. 2, pp. 451–466, 2010.
- [21] Y. Cho, S. B. Kim, and E. J. Powers, "Time-frequency analysis using ar models with variable forgetting factors," in *ICASSP*, 1990, pp. 2479–2482.