Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Outlier identificability in time series

Francesco Battaglia[1] | **Domenico Cucina**[2]

[1]Dipartimento di Scienze Statistiche, University La Sapienza, Rome, Italy
[2]Dipartimento di Scienze Politiche, University Rome 3, Italy

Correspondence
Francesco Battaglia. Email: francesco.battaglia@uniroma1.it

Present Address
Dipartimento di Scienze Statistiche, Piazzale Aldo Moro 5, 00100 Rome Italy

Summary

The occurrence of undetected outliers severely disrupts the model building procedures and produces unreliable results. This topic has been widely addressed in the statistical literature. However, little attention has been paid to determine how large an outlier has to be for correct detection of both time and magnitude to safely take place. This issue has been the object of research mainly in geodesy. In this paper the minimal detectable bias (MDB) concept is extended to vector time series data, and the risk of accepting an outlier as a clean observation is evaluated according to both the size and power of the statistical tests. This approach seems able to deal with the difficult issues known as masking and swamping. The proposed measure of outlier identifiability helps to determine if any configurations of multiple outliers, also occurring in patches is easily detectable.

KEYWORDS:
Minimal Detectable Bias, Outlier patches, Additive outlier

1 | INTRODUCTION

An additive outlier is a perturbation affecting just one observation of the time series, and may arise from a recording error or from the influence of unexpected events, in any case it is important to determine the exact time of that occurrence, and its relevance. The literature has widely addressed inference on additive outliers, considering both the detection of the occurrence of an outlier, and the estimation of its size. There are essentially two different approaches, one based on interpolators and the other on ARIMA models.

In the first case each observation is compared to its best (in mean square sense) reconstruction based on all the other observations (this is called the linear interpolator), and if the difference is large an outlier is suspected. Thus what is important is the significance of the interpolation error, see Ljung (1993); Peña (1987); Peña and Maravall (1991), among others, for details of this approach in the univariate case, and Baragona, Battaglia, and Poli (2011); Cucina, di Salvatore, and Protopapas (2014) for the multivariate case.

In the second approach we assume an ARIMA model and denote the outlier size by a parameter ω, then all the parameters are jointly estimated and if the estimated ω is significantly different from zero an outlier is suspected. The first proposal by Fox (1972) was later developed by several authors, e. g. Chang, Tiao, and Chen (1988); Chen and Liu (1993); Tsay (1986 1988), extension to multivariate ARIMA models was proposed by Tsay, Peña, and Pankratz (2000). The model may be alternatively written in state space form and estimated using the Kalman filter (e. g. Atkinson, Koopman, & Shepard 1997; De Jong & Penzer 1998). Bayesian estimation is also employed (e. g. Abraham & Box 1979; Box & Tiao 1968).

The two approaches are not independent (see e. g. Peña 1990) because if the linear interpolator is expressed in function of the ARMA parameters, and their estimates are plugged in, the statistics on which they are based coincide. In case of Gaussian data both methods are essentially equivalent to a likelihood ratio test of the null hypothesis of absence of outlier.

More recently a new approach was proposed for multivariate series (Baragona & Battaglia 2007; Galeano, Peña, & Tsay 2006), based on univariate contemporaneous linear combinations on which the univariate detection methods are applied.

Here we adopt a different point of view. We do not propose a new detection method, but try to determine how large an outlier must be to ensure a safe detection by current statistical methods. This topic has not attracted great attention in time series, while it was widely addressed in geodesy and in particular in the study of GPS networks (e. g. Prószynski 2015) which led to the important concept of Minimal Detectable Bias (Baarda 1968).

The problem of determining the size of a detectable outlier is important both in the design of analysis and in the interpretation of the results. The minimal detectable bias concept has been used for evaluating the reliability of GPS networks (Knight, Wang, & Rizos 2010; Teunissen 1998) and in quality control of integrated navigation systems (Salzmann 1991; Teunissen 1990; Teunissen & Montenbruck 2017), applications were proposed also in computer vision (e. g. Förstner 1987). Determining the minimal size of a perturbation that may be safely detected from the data helps to understand the level of ineradicable uncertainty in the results of analysis. On the other side, if the series is associated to a dynamical system, knowing to what extent a given perturbation in the input may be recovered from the output is useful in the design of the system and the measurement tools.

We define a measure of outlier identifiability that relates the size and position of an additive outlier to the probability of detecting it. It will be seen that such measure depends not only on the variability of the time series, but also on its autocorrelation structure.

The plan of the paper is as follows. Section 2 considers the problem of outlier identifiability in univariate time series and Section 3 in multivariate time series. Section 4 is concerned with a formal definition of masking and swamping. Section 5 addresses the problem of outlier detection when many different outliers are found in the same series, and Section 6 considers the detection methods based on linear univariate combinations. Section 7 contains an application, and conclusions are drawn in the last Section. The proofs of Theorems and a simulation study may be found in the Supporting Information.

2 | OUTLIER IDENTIFIABILITY IN UNIVARIATE TIME SERIES

First of all we introduce the idea of Minimal Detectable Bias due to Baarda, and formalize it for a general stationary time series.

Let $\{y_t\}$ denote a second-order stationary stochastic process with mean zero, autocovariances $\gamma(h) = E\{y_t y_{t+h}\}$ and spectral density $f(\lambda)$ positive everywhere. The inverse autocovariances $\gamma i(h)$ are defined by $\gamma i(h) = (2\pi)^{-2}\int_{-\pi}^{\pi} f(\lambda)^{-1}\exp\{i\lambda h\}d\lambda$ and the inverse autocorrelations by $ri(h) = \gamma i(h)/\gamma i(0)$. We shall use the notation $\delta(x) = 1$ if $x = 0$ and zero otherwise.

We suppose that the observed time series $\{z_t, t = 1, \dots, n\}$ is affected by p additive outliers at times $t_1, t_2, \dots, t_p$ with size $\omega_1, \omega_2, \dots, \omega_p$. On denoting $y = (y_1, \dots, y_n)'$, $z = (z_1, \dots, z_n)'$, $\omega = (\omega_1, \omega_2, \dots, \omega_p)'$ we may write $z = y + X\omega$ where the $n \times p$ design matrix has entries $X_{ij} = \delta(i - t_j)$, $i = 1, \dots, n; j = 1, \dots, p$. If the process $\{y_t\}$ is Gaussian, then $y \sim N(0, \Gamma)$ where $\Gamma_{i,j} = \gamma(j - i)$ and $z \sim N(X\omega, \Gamma)$. A standard test of $H_0 : \omega = 0$ against $H_1 : \omega \neq 0$ may be based on the likelihood ratio $L(z|0)/\max_\omega L(z|\omega)$, where

$$L(z|\omega) = (2\pi)^{-n/2}|\Gamma|^{-1/2}\exp\{-\frac{1}{2}(z - X\omega)'\Gamma^{-1}(z - X\omega)\}.$$

The inverse of the variance-covariance matrix Γ^{-1} may be approximated by the inverse autocovariance matrix Γi with element $\gamma i(j - i)$ at row i and column j. Such an approximation is motivated by the orthogonality property:

$$\sum_{u=-\infty}^{\infty}\gamma(u)\gamma i(u - v) = \delta(v)$$

and was studied by Shaman (1975 1976). Its properties are analyzed below when dealing with the multivariate case. On substituting Γi to Γ^{-1}, the maximum likelihood estimator $\hat\omega$ is: $\hat\omega = (X'\Gamma iX)^{-1}X'\Gamma iz$ and the maximum likelihood is:

$$L(z|\hat\omega) = (2\pi)^{-n/2}|\Gamma i|^{1/2}\exp\{-\frac{1}{2}(z - X\hat\omega)'\Gamma i(z - X\hat\omega)\} = (2\pi)^{-n/2}|\Gamma i|^{1/2}\exp\{-\frac{1}{2}z'\Gamma iz + \frac{1}{2}z'\Gamma iX(X'\Gamma iX)^{-1}X'\Gamma iz\}.$$

The test statistic becomes:

$$-2\log\frac{L(z|0)}{L(z|\hat\omega)} = z'\Gamma iX(X'\Gamma iX)^{-1}X'\Gamma iz = u'u$$

where $u = (X'\Gamma iX)^{-1/2}X'\Gamma iz$. Under H_0, u is unit normal, therefore the statistic $u'u$ follows a central chi square distribution with p degrees of freedom, thus the rejection region is $u'u > \overline{\chi}^2_{p,1-\alpha}$. Under the alternative hypothesis of presence of outliers $z \sim N(X\omega, \Gamma)$, then u has mean $(X'\Gamma iX)^{-1/2}X'\Gamma iX\omega = (X'\Gamma iX)^{1/2}\omega$ and variance I, therefore $u'u$ is a non central chi square with noncentrality parameter $E(u)'E(u) = \omega'(X'\Gamma iX)\omega$. The test statistic has distribution $\chi^2_p[\omega'(X'\Gamma iX)\omega]$ and the power under H_1 is

$$\Pr\{\chi^2_p[\omega'(X'\Gamma iX)\omega] > \overline{\chi}^2_{p,1-\alpha}\}.$$

If the power is denoted by β, then $\overline{\chi}^2_{p,1-\beta}[\omega'(X'\Gamma iX)\omega] = \overline{\chi}^2_{p,1-\alpha}$. In other words, the value of the noncentrality parameter Δ that ensures a test with size α and power β is the solution of

$$\overline{\chi}^2_{p,1-\beta}[\Delta] = \overline{\chi}^2_{p,1-\alpha}.$$

This is the equivalent for time series of the Minimal Detectable Bias of Baarda (1968).

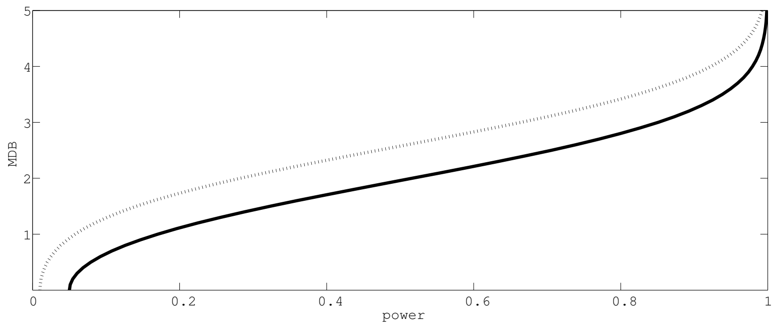

FIGURE 1 Minimal detectable bias in function of the power for test size $\alpha = 0.05$ (continuous line) and $\alpha = 0.01$ (dotted line). The bias is expressed in units of the mean square interpolation error $\gamma\mathrm{i}(0)^{-1/2}$.

Consider the case that only one outlier $(\mathrm{p} = 1)$ at time q affects the time series: the noncentrality parameter is $\Delta = \omega^2\gamma\mathrm{i}(0)$. The maximum likelihood estimator $\hat{\omega}$ equals the interpolation error $\mathrm{z_q} - \mathrm{l_q}$ where $\mathrm{l_t} = -\sum_{\mathrm{j}\neq 0}\mathrm{ri(j)z_{t-j}}$ is the linear interpolator (see e. g. Battaglia & Bhansali 1987; Pourahmadi 2001). The interpolation error has mean ω and variance $1/\gamma\mathrm{i}(0)$, thus $\mathrm{u'u} = \hat{\omega}^2/\mathrm{var}\{\hat{\omega}\}$ and the noncentrality parameter Δ is the square of the mean of the standardized interpolation error. To illustrate the concept of minimal detectable bias, we reported in Figure 1, separately for size 0.05 and 0.01, the power of the test for a single outlier when the outlier size is k times the mean square interpolation error $\gamma\mathrm{i}(0)^{-1/2}$, and k ranges from zero to five. From the figure it may be argued, for example, that a test with size 0.05 will reject the null hypothesis of clean data with probability 0.9 if the outlier size is at least 3.3 times the mean square interpolation error.

When p outliers occur at times $\mathrm{t_1, t_2, \ldots, t_p}$ and $\omega = (\omega_1, \ldots, \omega_\mathrm{p})'$ it follows

$$[X'\Gamma iX]_{ij} = \sum_a\sum_b [X']_{ia}\Gamma i(a,b)[X]_{bj} = \sum_a\sum_b X_{ai}\Gamma i(a,b)X_{bj} = \sum_{a=1}^n\sum_{b=1}^n \delta(a-t_i)\gamma i(b-a)\delta(b-t_j) = \gamma i(t_j - t_i)$$

and the noncentrality parameter becomes

$$\Delta = \sum_{j=1}^p\sum_{k=1}^p \omega_j\omega_k\gamma i(t_k - t_j) = \sum_{j=1}^p \omega_j^2\gamma i(0) + \sum_{j=1}^p\sum_{k\neq j}\omega_j\omega_k\gamma i(t_k-t_j) = |\omega|^2\gamma i(0)\sum_j\sum_k \frac{\omega_j\omega_k}{|\omega|^2} ri(t_k - t_j) \tag{1}$$

therefore it depends on $|\omega|^2$, on $\gamma\mathrm{i}(0)$ and on the quadratic form $\xi'\mathrm{Ri(X)}\xi$ where $\xi = \omega/|\omega|$ and the matrix Ri(X) has entry equal to $\mathrm{ri(t_j - t_i)}$ in row i and column j.

The noncentrality parameter may be alternatively expressed in function of eigenvalues and eigenvectors of the autocovariance matrix Γ: let λ_j denote the eigenvalues and $\mathrm{v_j}$ the associated eigenvectors, then $\Gamma\mathrm{i} = \Gamma^{-1}$ has eigenvalues λ_j^{-1} and the same eigenvectors; thus

$$\Delta = (X\omega)'\Gamma iX\omega = \sum_j \lambda_j^{-1}[\omega'X'v_j]^2 = \sum_j \lambda_j^{-1}\{\sum_k \omega_k v_j(t_k)\}^2$$

and depends on the angles between the outlier size and the eigenvectors.

3 | OUTLIER IDENTIFIABILITY IN MULTIVARIATE TIME SERIES

Let $\{\mathrm{y_t}\}$ with $\mathrm{y_t} = (\mathrm{y_{1,t}, y_{2,t}, \ldots, y_{s,t}})'$ denote a s-variate stochastic process.

Assumption A The process $\{\mathrm{y_t}\}$ is second-order stationary with means zero, autocovariances $\Gamma(\mathrm{h}) = \mathrm{E}\{\mathrm{y_t y'_{t+h}}\}$, spectral density matrix $\mathrm{F}(\lambda) = (2\pi)^{-1}\sum_\mathrm{h}\Gamma(\mathrm{h})\exp\{-\mathrm{i}\lambda\mathrm{h}\}$ positive definite for all λ, and inverse covariance matrices defined by $\Gamma\mathrm{i(h)} = (2\pi)^{-2}\int_{-\pi}^{\pi}\mathrm{F}(\lambda)^{-1}\exp(\mathrm{i}\lambda\mathrm{h})\mathrm{d}\lambda$.

Assumption B The observed series is denoted by $\{\mathrm{z_t} = (\mathrm{z_{1,t}, z_{2,t}, \ldots, z_{s,t}})', \mathrm{t} = 1, \ldots, \mathrm{n}\}$ and is contaminated by p additive outliers at times $\mathrm{t_1, t_2, \ldots, t_p}$ with sizes $\omega_1, \omega_2, \ldots, \omega_\mathrm{p}$. In vector notation $\mathrm{z} = (\mathrm{z'_1, z'_2, \ldots, z'_n})'$ and analogue for $\mathrm{y_t}$, we may write again $\mathrm{z} = \mathrm{y} + \mathrm{X}\omega$.

When there is only one outlier at time τ and size $\omega = (\omega_1, \omega_2, \ldots, \omega_\mathrm{s})'$, the design matrix X has ns rows and s columns defined by

$$X_{ij} = 0 \text{ for } i < (\tau-1)s,\ \ i > \tau s \quad ; \quad X_{ij} = 1 \text{ for } i = (\tau-1)s + j$$

and may be written $X = [0_{s,(\tau-1)s}, I_{s,s}, 0_{s,(n-\tau)s}]'$. Under Gaussian assumption $z \sim N(X\omega, \mathbf{\Gamma})$ where $\mathbf{\Gamma}$ denotes a block matrix, with each block a $(s \times s)$ autocovariance matrix, and a similar convention for $\mathbf{\Gamma i}$:

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(n-1) \\ \Gamma(1)' & \Gamma(0) & \dots & \Gamma(n-2) \\ \dots & \dots & \dots & \dots \\ \Gamma(n-1)' & \Gamma(n-2)' & \dots & \Gamma(0) \end{pmatrix}. \quad \mathbf{\Gamma i} = \begin{pmatrix} \Gamma i(0) & \Gamma i(1) & \dots & \Gamma i(n-1) \\ \Gamma i(1)' & \Gamma i(0) & \dots & \Gamma i(n-2) \\ \dots & \dots & \dots & \dots \\ \Gamma i(n-1)' & \Gamma i(n-2)' & \dots & \Gamma i(0) \end{pmatrix}.$$

We use the approximation $\mathbf{\Gamma}^{-1} = \mathbf{\Gamma i}$, it was analyzed by Shaman (1975 1976) in the univariate case and by Bhansali (1990) for multivariate time series, and is motivated by the orthogonality property:

$$\sum_{u=-\infty}^{\infty} \Gamma(u)\Gamma i(u-v)' = \delta(v)I.$$

The approximation is exact if the doubly infinite vector $z = [\dots, z'_{-2}, z'_{-1}, z'_0, z'_1, z'_2, \dots]'$ is considered, while in the finite case $\mathbf{\Gamma\Gamma i} - I$ is positive semi-definite. A more precise result may be obtained for vector autoregressive processes, where inverse covariances vanish at a finite lag. In that case the error is confined in the submatrices in the upper left corner and lower right corner.

Theorem 1. Let $\{y_t\}$ be a stationary vector autoregressive process of order q and Assumption A hold. The only non zero blocks of $\mathbf{\Gamma}^{-1} - \mathbf{\Gamma i}$ are those for $1 \le i \le q, 1 \le j \le q$ and $n-q \le i \le n, n-q \le j \le n$.

If there is just one outlier $X'\mathbf{\Gamma i}X = \Gamma i(0)$ and the noncentrality parameter becomes: $\Delta = \omega'\Gamma i(0)\omega$ where $\Gamma i(0)$ is the inverse variance matrix of the process. The test statistic is $(X'\mathbf{\Gamma i}z)'(X'\mathbf{\Gamma i}X)^{-1}X'\mathbf{\Gamma i}z$ and under absence of outliers has a distribution χ^2_s. Note that under the hypothesis of the Theorem, if the outlier position τ is larger than q and smaller than $n-q$ then $X'\mathbf{\Gamma}^{-1}X = X'\mathbf{\Gamma i}X$.

If there are p outliers at times $t_1, t_2, \dots, t_p$ the size of each of them is denoted $\omega_1 = (\omega_{1,1}, \omega_{2,1}, \dots, \omega_{s,1})' \dots, \omega_p = (\omega_{1,p}, \omega_{2,p}, \dots, \omega_{s,p})'$ and the vector ω is defined by $\omega = [\omega'_1, \omega'_2, \dots, \omega'_p]'$. The design matrix X is a $s \times s$ blocks matrix (with n row blocks and p column blocks): the first column block has a $s \times s$ identity matrix at the t_1-th row block, the second column block has an identity matrix at the t_2-th row block, and so on. So the matrix $(X'\mathbf{\Gamma i}X)$ is also a $(s \times s)$-blocks matrix, and its block (i,j) equals the inverse covariance matrix $\Gamma i(t_j - t_i)$. Again we note that under the assumptions of Theorem 1., if $q < t_1 < \dots < t_p < n-q$ then $X'\mathbf{\Gamma i}X = X'\mathbf{\Gamma}^{-1}X$.

The test statistic is also in this case equal to $(X'\mathbf{\Gamma i}z)'(X'\mathbf{\Gamma i}X)^{-1}X'\mathbf{\Gamma i}z$ and in absence of outliers is a chi square with sp degrees of freedom, while otherwise the noncentrality parameter is

$$\Delta = \omega' X' \mathbf{\Gamma i} X\omega = \sum_{i=1}^{p}\sum_{j=1}^{p} \omega'_i \Gamma i(t_j - t_i)\omega_j. \tag{2}$$

It follows that for multivariate series the value of the noncentrality parameter Δ that ensures a test with size α and power β is the solution of

$$\overline{\chi}^2_{sp,1-\beta}[\Delta] = \overline{\chi}^2_{sp,1-\alpha}.$$

In a similar way like the univariate case, the noncentrality parameter may be expressed in function of the eigenvalues λ_j and eigenvectors v_j of the matrix $\mathbf{\Gamma}$:

$$\Delta = \omega' X' \mathbf{\Gamma i} X\omega = \sum_j \lambda_j^{-1}(\omega' X' v_j)^2.$$

The analysis of the presence of p outliers at times $t_1, t_2, \dots, t_p$ simplifies when the inverse covariance matrices die out and the outliers are far from each other in the sense that $\Gamma i(t_j - t_i) = 0,\ i \neq j$. In such a case we say that the outliers are separate. Let us denote by X_i the design containing only the outlier at t_i: the matrix X_i has n row blocks $(s \times s)$ and just one column block, with the identity matrix at the t_i-th row block and zero elsewhere. Then

$$X = (X_1, X_2, \dots, X_p) = \begin{pmatrix} \dots & \dots & \dots & \dots \\ I & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & I & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I \\ \dots & \dots & \dots & \dots \end{pmatrix} \quad ; \quad X'\mathbf{\Gamma i}X = \begin{pmatrix} \Gamma i(0) & \Gamma i(t_2 - t_1) & \dots & \Gamma i(t_p - t_1) \\ \Gamma i(t_1 - t_2) & \Gamma i(0) & \dots & \Gamma i(t_p - t_2) \\ \dots & \dots & \dots & \dots \\ \Gamma i(t_1 - t_p) & \Gamma i(t_2 - t_p) & \dots & \Gamma i(0) \end{pmatrix}$$

and since $\Gamma i(t_j - t_i) = 0\ (i \neq j)$:

$$X'\mathbf{\Gamma i}X = \mathrm{diag}\{\Gamma i(0), \Gamma i(0), \dots, \Gamma i(0)\}.$$

The test statistic for the design X_j is $u_j'u_j = (X_j'\mathbf{\Gamma i}z)'(X_j'\mathbf{\Gamma i}X_j)^{-1}X_j'\mathbf{\Gamma i}z$ where $X_j'\mathbf{\Gamma i}X_j = \Gamma i(0)$. It follows that the test statistic for the complete design X is:

$$u'u = (X'\mathbf{\Gamma i}z)'(X'\mathbf{\Gamma i}X)^{-1}X'\mathbf{\Gamma i}z = z'\mathbf{\Gamma i}\begin{pmatrix} X_1' \\ X_2' \\ \dots \\ X_p' \end{pmatrix}'\begin{pmatrix} \Gamma i(0) & 0 & \dots & 0 \\ 0 & \Gamma i(0) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Gamma i(0) \end{pmatrix}^{-1}\begin{pmatrix} X_1' \\ X_2' \\ \dots \\ X_p' \end{pmatrix}\mathbf{\Gamma i}z = \sum_{j=1}^{p}(X_j'\Gamma iz)'\Gamma i(0)^{-1}X_j'\Gamma iz = \sum_{j=1}^{p}u_j'u_j$$

equal to the sum of the statistics for the X_j tests.

This happens also for the noncentrality parameters: from (2) when $\Gamma i(t_j - t_i) = 0$ $(i \neq j)$ we obtain

$$\Delta = \sum_{j=1}^{p}\omega_j'\Gamma i(0)\omega_j$$

equal to the sum of the noncentrality parameters of the single designs X_j. Thus the test statistic $u_j'u_j$ for the design X_j has a $\chi_s^2[\omega_j'\Gamma i(0)\omega_j]$ distribution, while the test statistic for the complete design X is the sum of $u_j'u_j$ and its distribution is $\chi_{ps}^2[\sum_j \omega_j'\Gamma i(0)\omega_j]$.

4 | MASKING AND SWAMPING

When the outliers are not separate, there may be interactions among them, that lead to the concepts of masking and swamping. Masking happens when an existing outlier is not detected owing to the existence of other outliers in surrounding observations. Swamping happens when a clean observation is recognized as an outlier due to the effect of other outliers in surrounding observations. Through the analysis of noncentrality parameters the concepts of masking and swamping may be made more precise and formal.

We start with a general result concerning the case that a wrong outlier configuration is tested.

Theorem 2. Let Assumptions A and B hold. Suppose that we test the design C stating the presence of q outliers at times $\tau_1, \tau_2, \dots, \tau_q$. The noncentrality parameter is $\Delta_C = \omega'X'\mathbf{\Gamma i}C(C'\mathbf{\Gamma i}C)^{-1}C'\mathbf{\Gamma i}X\omega$. and if $\tau_1, \tau_2, \dots, \tau_q$ are separated, i.e. $\Gamma i(\tau_j - \tau_i) = 0$, $i \neq j$, then

$$\Delta = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{q}\omega_i'\Gamma i(t_i - \tau_k)\Gamma i(0)^{-1}\Gamma i(\tau_k - t_j)\omega_j.$$

Let us consider in particular the univariate case. The matrix $C'\Gamma iC$ of dimension $q \times q$ has element (i,j) equal to $\gamma i(\tau_i - \tau_j)$, while $C'\Gamma iX\omega$ is a q-vector with k-th entry equal to $\sum_j \gamma i(\tau_k - t_j)\omega_j$; if the τ_j are separated we obtain

$$\Delta_C = \frac{1}{\gamma i(0)}|C'\Gamma iX\omega|^2 = \frac{1}{\gamma i(0)}\sum_k\{\sum_j \omega_j\gamma i(\tau_k - t_j)\}^2 \tag{3}$$

and in this case if there is only one τ equal to an outlying time t_j, and all the other true outliers are separated, the noncentrality parameter of the single outlier at t_j is obtained. In conclusion, if the true outliers all are separated and in the design C there are only some of them, the noncentrality parameter Δ_C equals the sum of the terms $\gamma i(0)\omega_j^2$ for the outliers contained in design C. The other times τ contained in C, provided that they are separated both from the t_j's and from the other τ_j's, do not contribute to the noncentrality parameter.

If on the contrary there is not complete separation, the easiest case is that C contains just one time τ, a clean observation. Then C is a column of zeros except an one in position τ and

$$C'\Gamma iC = \gamma i(0) \ , \ \ C'\Gamma iX = [\gamma i(\tau - t_1), \gamma i(\tau - t_2), \dots, \gamma i(\tau - t_p)]$$

from which $C'\Gamma iX\omega = \sum_j \omega_j\gamma i(\tau - t_j)$ and the noncentrality parameter becomes

$$\Delta_{\{\tau\}} = \frac{1}{\gamma i(0)}\{\sum_j \omega_j\gamma i(\tau - t_j)\}^2 = \gamma i(0)\{\sum_{j=1}^{p}\omega_j ri(\tau - t_j)\}^2 \tag{4}$$

that may be large and induce to identify observation at time τ as contaminated also if it is not. For example if X contains just an outlier at time t_1 then $\Delta_{\{\tau\}} = \gamma i(0)\omega_1^2 ri(\tau - t_1)^2$.

For multivariate series the analogue of (4) is

$$\Delta_{\{\tau\}} = \sum_{i=1}^{p}\sum_{j=1}^{p}\omega_i'\Gamma i(t_i - \tau)\Gamma i(0)^{-1}\Gamma i(\tau - t_j)\omega_j.$$

The above analysis suggests the following definition.

Definition 1. The outlier configuration (X, ω) α-*swamps* the unperturbed time τ if $\Delta_{\{\tau\}} > \bar{\chi}^2_{\mathsf{s},1-\alpha}$.

To analyze masking, suppose that Assumption B holds, but we test the presence only of the outlier at time t_k. In this case the design C contains only t_k and the noncentrality parameter is

$$\Delta_{\{t_k\}} = \sum_{i=1}^{p}\sum_{j=1}^{p} \omega_i' \Gamma i(t_i - t_k)\Gamma i(0)^{-1}\Gamma i(t_k - t_j)\omega_j.$$

For a univariate time series, C is a column with only 1 at position t_k and zero elsewhere, and the noncentrality parameter is

$$\Delta_{\{t_k\}} = \gamma i(0)\{\sum_{j=1}^{p} \omega_j ri(t_k - t_j)\}^2.$$

If the true design X includes only the outlier at $t = t_k$ (i.e., $p = 1$) then Δ correctly equals $\gamma i(0)\omega_k^2$, while for $p > 1$:

$$\Delta_{\{t_k\}} = \gamma i(0)\{\omega_k + \sum_{j\neq k} \omega_j ri(t_k - t_j)\}^2$$

that may be larger or smaller than $\gamma i(0)\omega_k^2$, and if it is smaller a masking occurs. For example in the simple case of a univariate series with two outliers at times t_1 and t_2 if we test only the presence of an outlier at t_1, we get $\Delta_{\{t_1\}} = \gamma i(0)\{\omega_1 + \omega_2 ri(t_1 - t_2)\}^2$ and the true noncentrality parameter is altered by the term $\omega_2 ri(t_1 - t_2)$; in the extreme case that $\omega_2 = -\omega_1/ri(t_1 - t_2)$ the noncentrality parameter would vanish.

This suggests the following definition.

Definition 2. The outlier configuration (X, ω) α-*masks* the outlier at time t_k if $\omega_k' \Gamma i(0)\omega_k > \bar{\chi}^2_{\mathsf{s},1-\alpha}$ and $\Delta_{\{t_k\}} < \bar{\chi}^2_{\mathsf{s},1-\alpha}$.

Example 1. For a univariate AR(1) process $\gamma i(0) = (1+\phi^2)/\sigma^2$, $\gamma i(1) = -\phi/\sigma^2$, $ri(1) = -\phi/(1+\phi^2)$, $\gamma i(h) = 0$, $|h| > 1$. If there is only an outlier at time q with size ω, swamping may occur at times $q-1$ and $q+1$. The noncentrality parameter for testing only q is $(\omega_q^2/\sigma^2)(1+\phi^2)$, while when testing an outlier at only $q-1$ or only $q+1$ the noncentrality parameters are equal and given by $(\omega_q^2/\sigma^2)(1+\phi^2)[-\phi/(1+\phi^2)]^2 = (\omega_q^2/\sigma^2)\phi^2/(1+\phi^2)$, i. e. equal to the previous one multiplied by $ri(1)^2 = \phi^2/(1+\phi^2)^2$ that ranges from zero to 1/4. As far as masking is concerned, here the outlier at time q could be masked by those at $q-1$ or $q+1$: if both of them are present the noncentrality parameter for the outlier at q alone is

$$\Delta = \gamma i(0)\{\omega_q - \frac{\phi}{1+\phi^2}(\omega_{q-1} + \omega_{q+1})\}^2.$$

In this case also Δ may be small, and even vanish if for example there are outliers only at times q and $q+1$, and $\omega_{q+1} = \omega_q(1+\phi^2)/\phi$.

An other interesting particular case for univariate series is that all the outliers but one are detected. Here Assumption B holds but we test the configuration C containing only the first $p-1$ outliers. Thus $X = [C, a_p]$ where a_p is a vector with all zero entries except a 1 at position t_p.The noncentrality parameter of the true design is $\Delta_X = \omega' X' \Gamma i X \omega$ while that for the design C is $\Delta_C = \omega' X' \Gamma i C (C' \Gamma i C)^{-1} C' \Gamma i X \omega$, we consider the difference $\Delta_X - \Delta_C$.

Theorem 3. Under Assumptions A and B let C be the design specifying only the first $p-1$ outliers at times $t_1, t_2, \ldots, t_{p-1}$. The difference between the non centrality parameter of the true design X and that of the design C is

$$\Delta_X - \Delta_C = \omega_p^2\{\gamma i(0) - [\Gamma i C(C'\Gamma i C)^{-1}C'\Gamma i]_{p,p}\}.$$

If the first $p-1$ outliers are separated (non necessarily that at t_p), then $(C'\Gamma i C)^{-1} = I/\gamma i(0)$ and $[\Gamma i C]_{j,k} = \gamma i(t_j - t_k)$; moreover

$$[\Gamma i C(C'\Gamma i C)^{-1}C'\Gamma i]_{p,p} = \sum_{k=1}^{p-1} \gamma i(t_p - t_k)^2/\gamma i(0)$$

and it follows

$$\Delta_X - \Delta_C = \omega_p^2 \gamma i(0)\{1 - \sum_{k=1}^{p-1} ri(t_p - t_k)^2\}. \tag{5}$$

The following conclusion may be drawn: if all the outliers but one have been detected, and they are separated, the identifiability of the last outlier is determined by the increase in noncentrality given by $\omega_p^2\gamma i(0)\{1 - \sum_{k=1}^{p-1} ri(t_p - t_k)^2\}$. If t_p is separate from the others, the increase is correctly $\omega_p^2\gamma i(0)$, but if t_p is not separate the increase is smaller due to inverse correlations $ri(t_p - t_j)^2$ that induce a masking effect.

For multivariate series the difference reads:

$$\Delta_X - \Delta_C = \omega_p'\{\Gamma i(0) - \sum_{k=1}^{p-1} \Gamma i(t_p - t_k)\Gamma i(0)^{-1}\Gamma i(t_k - t_p)\}\omega_p.$$

5 | MULTIPLE OUTLIERS DETECTION

The concept of noncentrality parameter is also useful for clarifying the problem of multiple outliers detection. When there are several outliers, and not completely separated, the search for the correct configuration is known to be a difficult task. The correct identification is especially hard when sequences of consecutive outliers (often called outlier patches) occur, and several ad hoc methods for discovering patches may be found in the literature (for example Bruce & Martin 1989; Justel, Peña, & Tsay 2001; McCulloch & Tsay 1994; Penzer 2007; Proietti 2003). From a decision theoretic point of view, the choice of the outlier configuration presents several possible decisions that must all be evaluated in principle to find the best one, therefore it is important to define exactly the form of the loss function. The noncentrality parameter is an obvious choice for the loss function, because it is maximized in correspondence of the true outlier configuration, as detailed in the next theorem.

Theorem 4. Under Assumptions A and B, for any outlier configuration C we have $\Delta_X \geq \Delta_C$.

The proof suggests that the noncentrality parameter for the design C is the squared norm of the projection of the vector $X\omega$ on the linear space spanned by the columns of C, equipped with the inner product $(x, y) = x'\Gamma i y$. It follows easily that for any design C containing X, we have $\Delta_C = \Delta_X$. Moreover, if C specifies some but not all the existing outliers, i. e., $X = [C, A]$ with $C'A = 0$, then $\Delta_X > \Delta_C$.

However, this is not enough to ensure that a sequential detection procedure based on the noncentrality parameter leads to a correct identification. More precisely, consider any proper subset of X, say X_1 and the addition to the design X_1 of a single new observation indicated by the column C. A correct procedure should ensure that the noncentrality parameter for the design $[X_1, C]$ is larger when C is a column of X, i. e. we are adding a true outlier to those already detected. Such property does not hold if the outliers are not all separated. With X_1 equal to the empty set the implication is that the test statistic for a single outlier does not necessarily reach its maximum at a contaminated observation. An example will clarify the point.

Example 2. Assume that there are two outliers with equal size ω at times $a-1$ and $a+1$ in a second-order moving average univariate process $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$. The noncentrality parameter for testing the existence of a single outlier at time q is $\Delta_{\{q\}} = \gamma i(0)\{\omega ri(q-a+1) + \omega ri(q-a-1)\}^2$ and in particular

$$\Delta_{\{a\}} = \gamma i(0)\omega^2\{2ri(1)\}^2 \quad ; \quad \Delta_{\{a-1\}} = \Delta_{\{a+1\}} = \gamma i(0)\omega^2\{ri(2)+1\}^2.$$

Choosing $\theta_1 = -1.125, \theta_2 = 0.875$ we obtain $ri(1) = -\theta_1/(1+\theta_2) = 0.6$, $ri(2) = -\theta_1 ri(1) - \theta_2 = -0.2$; it follows $\Delta_{\{a-1\}} = \Delta_{\{a+1\}} = 0.64\,\gamma i(0)\omega^2$ and $\Delta_{\{a\}} = 1.44\,\gamma i(0)\omega^2$, and time a is largely more significant than the perturbed times $a-1$ and $a+1$.

We conclude that for difficult non-separate outlier configurations a sequential search does not guarantee the correct solution, but every possible design should be evaluated in principle, as is done by methods based on evolutionary computation like the genetic algorithm proposed by Baragona, Battaglia, and Calzini (2001); Cucina et al. (2014).

6 | OUTLIER IDENTIFIABILITY THROUGH UNIVARIATE LINEAR COMBINATIONS

An alternative method for outlier detection (Baragona & Battaglia 2007; Galeano et al. 2006) is based on univariate (contemporaneous) linear combinations of the multivariate series:

$$w_t = \sum_{j=1}^{s} d_j z_{j,t}.$$

The outliers are searched on the series w_t by univariate techniques. Under Assumption B, this series has p outliers at times $t_1, \dots, t_p$ and sizes $d'\omega_1, d'\omega_2, \dots, d'\omega_p$, the corresponding design will be denoted by the matrix X_* of dimension $n \times p$ and sizes $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)'$ with $\alpha_k = d'\omega_k$. Let us consider the vector $w = (w_1, \dots, w_n)'$ that has mean zero under absence of outliers, and mean $X_*\alpha$ under assumption B, its dispersion matrix will be denoted by Σ_w, with $\Sigma_w(i, j) = d'\Gamma(j-i)d$. On applying the results for univariate series in the previous Sections, the test statistic is $w'\Sigma_w^{-1}X_*(X_*'\Sigma_w^{-1}X_*)^{-1}X_*'\Sigma_w^{-1}w$ and its distribution is chi square with p degrees of freedom and non centrality parameter $\Delta = \alpha'(X_*'\Sigma_w^{-1}X_*)\alpha$. If we let as before $\omega = (\omega_1', \omega_2', \dots, \omega_p')'$, and $d = (d_1, \dots, d_s)'$ and define the $(n \times ns)$ matrix D by

$$D = \begin{pmatrix} d' & 0 & \dots & 0 \\ 0 & d' & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d' \end{pmatrix}$$

we have $w = Dz$ and $E(w) = DX\omega$, $\Sigma_w = D\mathbf{\Gamma}D'$. Thus, on substituting $DX\omega$ to $X_*\alpha$ and $D\mathbf{\Gamma}D'$ to Σ_w the noncentrality parameter may be rewritten

$$\Delta = \omega' X' D' (D\mathbf{\Gamma}D')^{-1} DX\omega = \sum_{i=1}^{p}\sum_{j=1}^{p} (d'\omega_i)(d'\omega_j)(D\mathbf{\Gamma}D')^{-1}_{i,j}.$$

Note that the matrix $(D\mathbf{\Gamma}D')^{-1}$ has the meaning of inverse autocovariance matrix of w_t.

When the design X specifies just one outlier at time q, the noncentrality parameter becomes:

$$\Delta = (d'\omega)^2 (D\mathbf{\Gamma}D')^{-1}_{q,q} \simeq (d'\omega)^2 \gamma i_w(0).$$

The linear combination weights d may be chosen in order to maximize the above noncentrality parameter, but the maximization of Δ with respect to d is not simple involving the inverse of $D\mathbf{\Gamma}D'$.

When the process is white noise $\Gamma(h) = 0, h \neq 0$, and $D\mathbf{\Gamma}D'$ is diagonal with elements $d'\Gamma(0)d$, therefore $(D\mathbf{\Gamma}D')^{-1} = I/d'\Gamma(0)d$ and

$$\Delta = \frac{(d'\omega)^2}{d'\Gamma(0)d}.$$

The above quantity is maximized (Rao 1973, p. 60) when $d \propto \Gamma(0)^{-1}\omega$, and if the components are also contemporaneously uncorrelated ($\Gamma(0)$ diagonal) the optimum vector d has entries $d_j = \omega_j/\Gamma(0)_{j,j}$, and if further the z_j have equal variance then the best d is proportional to ω. Baragona and Battaglia (2007) use the Independent Component Analysis (Hyvarinen, Karhunen, & Oja 2001) and obtain that the first component has coefficients $d = \Gamma(0)^{-1}\omega/\sqrt{\omega'\Gamma(0)^{-1}\omega}$, therefore it maximizes the noncentrality parameter. Galeano et al. (2006) search for the vector that maximizes the kurtosis, and working on standardized data ($\Gamma(0) = I$) obtain d parallel to ω. But since the kurtosis is affine equivariant, this amounts to $d \propto \Gamma(0)^{-1}\omega$. Note that with such a choice we have $\Delta = (\omega'\Gamma(0)^{-1}\omega)^2/\omega'\Gamma(0)^{-1}\omega = \omega'\Gamma(0)^{-1}\omega$, coinciding with the noncentrality parameter for the multivariate process, but relating to a χ^2 with one degree of freedom rather than s, therefore the test is more powerful (as already noted by Galeano et al. 2006) and the outlier is more identifiable.

When the process is not white noise, w_t is also autocorrelated, therefore $d'\Gamma(0)d = \mathrm{var}(w_t) > 1/\gamma i_w(0)$, it follows that the noncentrality parameter is larger than in the white noise case:

$$(d'\omega)^2\gamma i_w(0) > \frac{(d'\omega)^2}{d'\Gamma(0)d}.$$

The choice $d \propto \Gamma(0)^{-1}\omega$ does not maximize here the noncentrality parameter, the vector d that maximizes $(d'\omega)^2\gamma i_w(0)$ may only be obtained by numerical optimization, but a simple lower bound is derived in the next Theorem.

Theorem 5. Let Assumptions A and B hold with $p = 1$ (only one outlier present). Then

$$(d'\omega)^2\gamma i_w(0) \geq \frac{(d'\omega)^2}{d'\Gamma i(0)^{-1}d}$$

and the right hand side is maximized by $d \propto \Gamma i(0)\omega$.

As an example we consider the case of a bivariate moving average process.

Example 3. Let $\{z_t\}$ be a bivariate MA(1) process $z_t = \varepsilon_t + \Theta\varepsilon_{t-1}$ with $E\{\varepsilon_t\varepsilon_t'\} = I$. Then $\Gamma(0) = I + \Theta\Theta', \Gamma(1) = \Theta', \Gamma(h) = 0, |h| > 1$ and $\Gamma i(0) = (I - \Theta\Theta')^{-1}$. The linear combination $w_t = d'z_t$ has variance $d'\Gamma(0)d$ and autocovariance $\gamma_w(1) = d'\Theta'd, \gamma_w(h) = 0, |h| > 1$. Let us consider normalized vectors such that $d'\Gamma(0)d = 1$, it follows that $\gamma_w(0) = 1$ and

$$|\gamma_w(1)| = |d'\Theta'd\,| \leq |d|\,|\Theta'd| \leq \frac{1}{2}\{|d|^2 + |\Theta'd|^2\} = \frac{1}{2}d'\Gamma(0)d = \frac{1}{2}$$

therefore w_t follows a univariate invertible MA(1) process $w_t = e_t + \theta e_{t-1}$ with $E(e_t^2) = 1/(1+\theta^2)$ and $\gamma_w(1) = d'\Theta'd = \theta/(1+\theta^2)$. For this process the inverse covariances are given by $\gamma i_w(h) = (-\theta)^h(1+\theta^2)/(1-\theta^2)$ and the inverse variance by $\gamma i_w(0) = (1+\theta^2)/(1-\theta^2)$. The value of the univariate parameter θ may be expressed in function of $\gamma_w(1) = d'\Theta'd$ as $\theta = [1 - \sqrt{1-4\gamma_w(1)^2}]/\{2\gamma_w(1)\}$. Since the inverse variance is increasing with θ^2 its minimum value is one and its maximum is attained when $d'\Theta'd$ is maximized, i. e., if d is proportional to the eigenvector associated to the maximum eigenvalue of $(\Theta + \Theta')/2$. On the other hand, obviously $(d'\omega)^2$, under the constraint $d'\Gamma(0)d = 1$, ranges from zero to $(\omega'\omega)^2/\omega'\Gamma(0)\omega$. Since the noncentrality parameter is the product of those two functions, its maximum value will be attained at an intermediate vector (unless ω is exactly the maximal eigenvector of Θ, in which case the maxima are attained at the same point).

Assume for example $\Theta = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.4 \end{pmatrix}$ and $\omega = (5, -2)'$. The largest eigenvalue is 0.885 and the associated eigenvector is $(1, 0.618)'$, therefore the best d equals

$$d = (1, 0.618)'\{(1, 0.618)\Gamma(0)(1, 0.618)'\}^{-1/2} = (0.637, 0.393)'$$

that corresponds to $\gamma_w(1) = d'\Theta d = 0.496$ and $\theta = 0.885$, thus $\gamma i_w(0) = 8.257$ while $(d'\omega)^2 = 5.74$ and the noncentrality parameter is 47.44. On the other side, the normalized ω is $(0.812, -0.325)'$, therefore putting $d \propto \Gamma(0)^{-1}\omega$ the first factor equals $(\omega'\omega)^2/\omega'\Gamma(0)\omega = 22.19$ and the inverse variance of the related linear combination is 1.384, then the noncentrality parameter is 30.71. The lower bound derived in Theorem 5, since $\Gamma i(0) = (I - \Theta\Theta')^{-1}$, is maximized by d proportional to $\Gamma i(0)\omega$, or $d = (0.737, 0.193)'$ and equals 40.04. However, it may be easily seen that the maximum of $(d'\omega)^2\gamma i_w(0)$ is 51.26 and is obtained with $d = (0.660, 0.354)'$.

TABLE 1 Significant outlier statistics for the gas furnace series

t=	42	43	54	55	113	198	199	235	264
$\hat{\omega}_1$	-.41	0.43	0.40	-.36	-.26	-.04	-.04	-.01	0.18
$\hat{\omega}_2$	-.05	-.06	.01	0.1	.,08	-.51	0.61	0.45	-.53
$u'u$	32.8	37.2	31.8	26.3	13.2	18.1	25.8	13.8	25.9

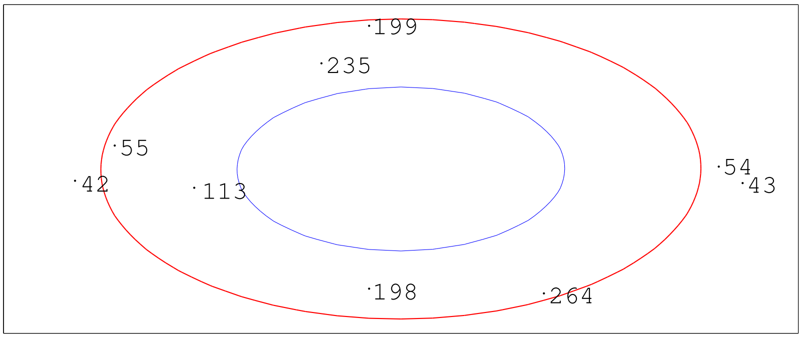

FIGURE 2 Plot of the estimated outlier sizes for gas furnace data.

7 | AN APPLICATION

The noncentrality measure depends on parameters (the inverse covariances and the autoregressive coefficients) that need to be estimated from the time series and are subject to bias due to sample error. The results of a simulation study (appearing in the Supplementary Information) suggest that the distortion induced by estimation is generally negligible.

We consider the well-known gas furnace series analyzed, among others, by Tiao and Box (1981), who propose a bivariate sixth-order autoregressive model. With their estimates of the autoregressive parameters, the estimated inverse covariance estimates are as follows:

$\Gamma i(0)$		$\Gamma i(1)$		$\Gamma i(2)$		$\Gamma i(3)$		$\Gamma i(4)$		$\Gamma i(5)$		$\Gamma i(6)$	
191	-0.7	-134	7.8	53	-9	-29	5.9	32.9	-4.1	-23	2.7	6.1	-0.8
-0.7	67	-.05	-42	-10	4.3	13	6	-8.7	-.22	8.4	-2.1	-4.1	0.7

This series was considered by Tsay et al. (2000) who identified anomalous observations at times 43, 55, 113, 199, 236, 265, 287, 288. Also Cucina et al. (2014) analyzed the gas furnace data with the method based on genetic algorithm, and found outliers at times 43, 54, 113, 199, 235, 264. We have computed the test statistics and their squares, and found that only the times listed in Table 1 were significant at 1%. These values are plotted in Figure 2 where the two ellipses denote the rejection regions with size 0.1 and 0.001. The results are rather similar for the three methods, but there are inconsistencies on some pairs of consecutive observations that could be subject to swamping or masking (due also to the large values of the entries of $\Gamma i(1)$). Such pairs: (42,43), (54,55), (198, 199), (235, 236), (264, 265), (287, 288) are well separated from each other ($\Gamma i(h)$ vanishes for $h > 6$) and are considered in Table 2. For any pair of observations at times $(t, t+1)$ we report the significance of the test statistics for a single outlier at t or $t+1$ and for the joint configuration $(t, t+1)$ ($*$ at 5 %, $**$ at 1 %, - not significant) and if the following definitions apply: t swamps the clean observation at $t+1$; $t+1$ swamps the clean observation at t ; the configuration $(t, t+1)$ masks the outlier at t; the configuration $(t, t+1)$ masks the outlier at $t+1$. The results suggest that the outlier at $t = 199$ swamps time 198 and the outlier at $t = 264$ swamps time 265, while the perturbation at time 288 does not appear strongly significant, and the observation at $t = 236$ appear as an isolated outlier. For the last two pairs (42, 43) and (54,55) we see that the swamping definition applies in both directions and the test statistics are both significant for the first and the second observation, and for the pair. Thus, to have an indication we look at the values of $u'u(42) = 32.8$ and $u'u(43) = 37.2$ from Table 1: since the observation at $t = 43$ is more significant we propend to conclude that the outlier at $t = 43$ swamps time 42. A similar reasoning suggests that the outlier at $t = 54$ swamps time 55.

TABLE 2 Analysis of the apparently aberrant pairs of observations for gas furnace series.

$(t, t+1)$	u_t	u_{t+1}	$u_{(t,t+1)}$	t swamps $t+1$	$t+1$ swamps t	t masked	$t+1$ masked
(42,43)	$**$	$**$	$**$	yes	yes	no	no
(54,55)	$**$	$**$	$**$	yes	yes	no	no
(198,199)	$**$	$**$	$**$	no	yes	no	no
(235,236)	$**$	-	-	no	nos	no	no
(264,265)	$**$	-	$**$	yes	no	no	no
(287,288)	-	$*$	-	no	no	no	no

8 | CONCLUSIONS

Starting from an analogy with the concept of minimal detectable bias widely employed in geodesy, we have defined a measure of outlier identifiability that allows to determine, for a second-order stationary stochastic process, what size an anomaly should reach for ensuring a safe detection by means of standard outlier detection methods. This quantifies the precision and uncertainty level associated with the conclusions that may be drawn from a statistical analysis of a time series generated by that process.

The proposed measure, a noncentrality parameter, is useful for clarifying what configurations of multiple outliers, also occurring in patches, are easily detectable and what are not. Moreover it allows to define more precisely, and evaluate, the concepts of masking and swamping.

We have assumed a general linear contamination model $z = y + X\omega$ and have considered only additive outliers, where each column of the design matrix X contains only one entry equal to one, and all other entries are equal to zero. But the same model allows for several other outlier types addressed in literature (for example level changes, temporary changes, innovation outliers, see e. g. Tsay et al. 2000), that may be analyzed by simply changing the appropriate columns of the design matrix X.

Finally, we observe that the proposed quantities may be considered also in the univariate non-stationary case, when dealing with integrated series (i.e. generated by a process $\{y_t\}$ such that there exists a positive integer d for which $(1 - B)^d y_t$ is second-order stationary). For such series the inverse covariances and the linear interpolator may be defined and estimated (Baragona & Battaglia 1995) and the noncentrality parameter in the form (1) may be computed.

ACKNOWLEDGEMENTS

We wish to thank Roberto Baragona for constant help and advice. The present research was supported by the national fund FFABR2017.

DATA AVAILABILITY STATEMENT

The Matlab code for reproducing simulation data and results is available from the corresponding author upon request. The gas furnace series is found in Box and Jenkins (1970).

References

Abraham, B., & Box, G. E. P. (1979). Bayesian analysis of some outlier problems in time series. *Biometrika*, *66*, 229–236.

Atkinson, A. C., Koopman, S. J., & Shepard, N. (1997). Detecting shocks: outliers and breaks in time series. *Journal of Econometrics*, *80*, 387–422.

Baarda, W. (1968). *A test procedure for use in geodetic models* (Vol. 2) (No. 5). Netherlands Geodetic Commission Publications on Geodesy.

Baragona, R., & Battaglia, F. (1995). Linear interpolators and the inverse correlation function of non-stationary time series. *Journal of Time Series Analysis*, *16*, 531–538.

Baragona, R., & Battaglia, F. (2007). Outlier detection in multivariate time series by independent component analysis. *Neural Computation*, *19*, 1962–1984.

Baragona, R., Battaglia, F., & Calzini, C. (2001). Genetic algorithms for the identification of additive and innovation outliers in time series. *Computational Statistics & Data Analysis*, *37*, 1–12.

Baragona, R., Battaglia, F., & Poli, I. (2011). *Evolutionary statistical procedures*. Heidelberg: Springer.

Battaglia, F., & Bhansali, R. J. (1987). Estimation of the interpolation error variance and an index of linear determinism. *Biometrika*, *74*, 771–779.

Bhansali, R. J. (1990). On a relationship between the inverse of a stationary covariance matrix and the linear interpolator. *Journal of Applied Probability*, *27*, 156–170.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis forecasting and control*. S. Francisco: Holden-Day.

Box, G. E. P., & Tiao, G. C. (1968). A Bayesian approach to some outlier problems. *Biometrika*, *55*, 119–129.

Bruce, A. G., & Martin, R. D. (1989). Leave-k-out diagnostics for time series. *Journal of the Royal Statistical Society Ser. B*, *51*, 363–424.

Chang, I., Tiao, G. C., & Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, *30*, 193–204.

Chen, C., & Liu, L. M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, *88*, 284–297.

Cucina, D., di Salvatore, A., & Protopapas, M. K. (2014). Outliers detection in multivariate time series using genetic algorithms. *Chemometrics and Intelligent Laboratory Systems*, *132*, 103–110.

De Jong, P., & Penzer, J. (1998). Diagnosing shocks in time series. *Journal of the American Statistical Association*, *93*, 796–806.

Förstner, W. (1987). Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision. *Computer Vision Graphics and Image Processing*, *40*, 273–310.

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society Ser. B*, *34*, 350–363.

Galeano, P., Peña, D., & Tsay, R. S. (2006). Outlier detection in multivariate time series by projection pursuit. *Journal of the American Statistical Association*, *101*, 654–669.

Hyvarinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: Wiley.

Justel, A., Peña, D., & Tsay, R. S. (2001). Detection of outlier patches in autoregressive time series. *Statistica Sinica*, *11*, 651–673.

Knight, N., Wang, J., & Rizos, C. (2010). Generalised measures of reliability for multiple outliers. *Journal of Geodesy*, *84*, 625–635.

Ljung, G. M. (1993). On outlier detection in time series. *Journal of the Royal Statistical Society Ser. B*, *55*, 559–567.

McCulloch, R. E., & Tsay, R. S. (1994). Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis*, *15*, 235–250.

Peña, D. (1987). Measuring the importance of outliers in ARIMA models. In M. L. Puri (Ed.), *New perspectives in theoretical and applied statistics* (pp. 109–118). New York: Wiley.

Peña, D. (1990). Influential observations in time series. *Journal of Business and Economic Statistics*, *8*, 235–241.

Peña, D., & Maravall, A. (1991). Interpolations, outliers and inverse autocorrelations. *Communications in Statistics Theory and Methods*, *20*, 3175–3186.

Penzer, J. (2007). State space models for time series with patches of unusual observations. *Journal of Time Series Analysis*, *28*, 629–645.

Pourahmadi, M. (2001). *Foundations of time series analysis and prediction theory*. New York: Wiley.

Proietti, T. (2003). Leave-k-out diagnostics in state space models. *Journal of Time Series Analysis*, *24*, 221–236.

Prószynski, W. (2015). Revisiting Baarda's concept of minimal detectable bias with regard to outlier identifiability. *Journal of Geodesy*, *89*, 993–1003.

Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

Salzmann, M. (1991). MDB: A design tool for integrated navigation systems. *Bulletin géodésique*, *65*, 109–115.

Shaman, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. *The Annals of Statistics*, *3*, 532–538.

Shaman, P. (1976). Approximation for stationary covariance matrices and their inverses with applications to ARMA processes. *The Annals of Statistics*, *4*, 292–301.

Teunissen, P. (1998). Minimal detectable biases of GPS data. *Journal of Geodesy*, *72*, 236–244.

Teunissen, P. J. G. (1990). Quality control in integrated navigation systems. *IEEE Aerospace and Electronic Systems Magazine*, *5*, 35–41.

Teunissen, P. J. G., & Montenbruck, O. (Eds.). (2017). *Springer Handbook of Global Navigation Satellite Systems*. Springer.

Tiao, G. C., & Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, *76*, 802–816.

Tsay, R. S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, *81*, 132–141.

Tsay, R. S. (1988). Outliers, level shifts and variance changes. *Journal of Forecasting*, *7*, 1–20.

Tsay, R. S., Peña, D., & Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika*, *87*, 789–804.

SUPPORTING INFORMATION

Additional information for this article is available. The proofs of theorems and the results of a simulation study are given in the Supporting Information, filename supplem2.pdf.

$\square$