Applied Condition Monitoring

Fakher Chaari
Jacek Leskow
Antonio Napolitano
Radoslaw Zimroz
Agnieszka Wylomanska *Editors*

Cyclostationarity: Theory and Methods III

Contributions to the 9th Workshop on Cyclostationary Systems and Their Applications, Grodek, Poland, 2016

Springer

Applied Condition Monitoring

Volume 6

About this Series

The book series Applied Condition Monitoring publishes the latest research and developments in the field of condition monitoring, with a special focus on industrial applications. It covers both theoretical and experimental approaches, as well as a range of monitoring conditioning techniques and new trends and challenges in the field. Topics of interest include, but are not limited to: vibration measurement and analysis; infrared thermography; oil analysis and tribology; acoustic emissions and ultrasonics; and motor current analysis. Books published in the series deal with root cause analysis, failure and degradation scenarios, proactive and predictive techniques, and many other aspects related to condition monitoring. Applications concern different industrial sectors: automotive engineering, power engineering, civil engineering, geoengineering, bioengineering, etc. The series publishes monographs, edited books, and selected conference proceedings, as well as textbooks for advanced students.

More information about this series at http://www.springer.com/series/13418

Fakher Chaari · Jacek Leskow
Antonio Napolitano · Radoslaw Zimroz
Agnieszka Wylomanska
Editors

Cyclostationarity: Theory and Methods III

Contributions to the 9th Workshop on Cyclostationary Systems and Their Applications, Grodek, Poland, 2016

Springer

Editors
Fakher Chaari
National School of Engineers of Sfax
Sfax
Tunisia

Jacek Leskow
Institute of Mathematics
Cracow University of Technology
Cracow
Poland

Antonio Napolitano
Department of Engineering
University of Napoli "Parthenope"
Naples
Italy

Radoslaw Zimroz
Institute of Mining Engineering
Wrocław University of Technology
Wrocław
Poland

Agnieszka Wylomanska
Institute of Mathematics and Computer Science
Wrocław University of Technology
Wrocław
Poland

ISSN 2363-698X ISSN 2363-6998 (electronic)
Applied Condition Monitoring
ISBN 978-3-319-51444-4 ISBN 978-3-319-51445-1 (eBook)
DOI 10.1007/978-3-319-51445-1

Library of Congress Control Number: 2017934545

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction

It is our pleasure to present already third volume (time flies!) of the series dedicated to the cyclostationarity and applications. In recent years, there is a growing interest in modelling nonstationary signals and processes with cyclic patterns and our book is a response to that. As in the previous two volumes, we have included a wide range of papers: from quite theoretical dedicated to cyclostationary models to very applied papers studying various mechanical cyclical signals.

More theoretically oriented papers are dedicated to such topics as: heavy tailed time series and processes, PAR models, rational spectra for PARMA processes, covariance invariant analysis, change point problems, subsampling for time series, fraction-of-time approach, GARMA models and weak dependence. In the present volume, there are published fifteen papers that can be divided into three groups. It should be mentioned that the groups are not separable and some articles could be classified to two or even three classes. The first group is devoted to strictly periodically correlated processes and time series. The reader can find theoretical investigations of PC models as well as their possible applications in mechanical systems. The special attention is paid to periodically autoregressive model (PARMA) and their special case PAR systems (A. Makagon and H. Hurd); however, one can find also the papers related to other processes which exhibit periodically correlated behaviour, like vectorial PC processes (I. Javors'Kyj et al.). To this group, we also include the application-oriented articles where PC methodology is taken under consideration (A. Parey et al.). In the second group of papers, the reader can find the analysis of general class of stochastic processes and methods dedicated to them. In two papers, authors consider the heteroscedastic models (A. Wylomanska et al., P. Doukhan et al.); in the paper of B. Silveira de Andrade et al., the reader can find the examination of generalized autoregressive model (GARMA) with the inverse Gaussian distribution, while in M. Polak et al. paper, the autoregressive model is used in the de-noising process for seismic signals. The next two papers which belong to the second group contain statistical methods used in the analysis of nonstationary processes (J. Leskow et al., E. Gajecka-Mirek et al.). The third group includes applied studies of various mechanical systems by using stochastic and statistical methods, especially in the

context of damage detection. We highlight here the analysis of vibration signals (P. Kruczek et al., R. Burdzik et al., L. Konieczny et al.) and temperature data analysis (P. Kruczek, J. Sokolowski et al., J. Wodecki et al.). The number of papers dedicated to methods (8) and the number of papers dedicated to applications (7) represent a perfect balance we attempt to maintain throughout our meetings.

Our annual meetings in a lovely small town of Grodek (80 km south-east from Krakow, Poland) are usually organized around two main topics: methods and applications. The meetings are attended by specialists that come from various countries and continents. In this volume, we present results of researchers from Europe (Poland, Ukraine and France), India, USA and Brazil.

In our opinion, our series on cyclostationarity and applications is quite unique. We have succeeded in putting together mathematicians, statisticians and engineers working on cyclostationary models. And it works! We have been able to develop common language, similar approaches, complementary models and interesting applications. We hope that the reader will share our views and enthusiasm for that kind of an approach.

Weak Dependence: An Introduction Through Asymmetric ARCH Models

P. Doukhan and N. Mtibaa

Abstract With the simple example of an asymmetric ARCH(1) model as a pretext, we introduce some of the main tools for weak dependence conditions introduced in [7]. The power of weak dependence is shown up for this very elementary model. This a special case of the infinite memory models in [8]. Asymptotic properties of a moment based parametric estimation do not need any regularity assumption over innovations. In a last section we address a subsequent estimation of residuals: then model based bootstrap is rapidly derived as well as the estimation of innovations density based on those fitted innovations.

Keywords AMS 2010 classification · 60J05 Discrete-time Markov processes on general state spaces · 62M05 Markov processes: estimation · 62G20 Asymptotic properties · 62D05 Sampling theory · Sample surveys

1 Introduction

The paper presents some basic concepts related with weak dependence. We use a simple model in order to demonstrate the efficiency of those techniques.

Namely we consider the equation

$$X_t = \sqrt{(aX_{t-1} + b)^2 + c^2} \cdot \xi_t, \qquad a > 0 \tag{1}$$

The *innovations* (ξ_t) constitute an independent and identically distributed (iid) sequence. In the following we thus assume that ξ_t is independent of X_{t-1}. The properties and the estimation of the parameters in the above model will rely on the observation of a sample from a stationary solution of Eq. (1).

P. Doukhan (✉) · N. Mtibaa
University Cergy-Pontoise, Cergy, France
e-mail: paul.doukhan@gmail.com

P. Doukhan
French Universitary Institute, Cergy, France

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_1

This is a very special case of the model with infinite memory

$$X_t = F(X_{t-1}, X_{t-2}, \dots; \xi_t) \tag{2}$$

introduced in [8].

The paper is organized as follows. The following section recalls some definitions of weak dependence conditions and we especially point out their heredity properties which are not immediate, as this the case for the standard mixing properties (see [4]). After this, a section addresses stationarity conditions and the dependence structure of infinite memory models.

We then specialize to the case of asymmetric ARCH(1) models.

A first section makes precise the stationarity and the dependence structure for this case.

The parametric moment estimators are detailed in a following section.

Finally we point out the asymptotic properties of fitted residuals with applications to bootstrap, in a last section; we especially focus on the estimation of the density of residuals providing a weak version of [9] results (which dedicates to functional autoregressive models).

2 Weak Dependence

The whole material of this section is a selection from the lecture notes [2], readers should refer to this volume for complements and details.

Definition 2.1 [7] Assume that there exist classes of functions

$$\mathcal{F}, \mathcal{G} : \bigcup_{u \geq 1} \mathbb{R}^u \to \mathbb{R},$$

and a function $\psi : \mathcal{F} \times \mathcal{G} \to \mathbb{R}$ (which depends on f, g and on the number of their arguments (u, v) and a sequence $\epsilon_r \downarrow 0$ as $r \uparrow \infty$.

A random process $(X_t)_{t \in \mathbb{Z}}$ is said to be $(\mathcal{F}, \mathcal{G}, \psi, \epsilon)$—weakly dependent in case

$$\left| \text{Cov} \left(f(X_{i_1}, \dots, X_{i_u}), g(X_{j_1}, \dots, X_{j_v}) \right) \right| \leq \epsilon_r \psi(u, v, f, g) \tag{3}$$

for functions f, g belonging respectively to classes $\mathcal{F}, \mathcal{G}$, and

$$i_1 \leq \cdots \leq i_u \leq j_1 - r \leq j_1 \leq \cdots \leq j_v.$$

Remark 2.1 The definition above deals with real valued models for simplicity, anyway nothing needs to be changed in case $\mathbb{R}$ is replaced by any Banach space, and in fact a Polish space is even enough in order to define Lipschitz conditions with very simple changes.

The index set is also a matter of simplicity since any (unbounded) metric space $(\mathbb{T}, d)$ may replace the set $\mathbb{Z}$ of integers leading either to discrete random fields $\mathbb{T} = \mathbb{Z}^d$ or random processes $\mathbb{T} = \mathbb{R}^d$, and so on…

Example 2.1 (some coefficients). References [1, 7] suggest:

- θ—weak dependence coefficients $\theta_r = \epsilon_r$ which correspond to

$$\psi(f, g) = v\mathrm{Lip}\, g\|f\|_\infty.$$

- η—weak dependence coefficients $\eta_r = \epsilon_r$, here

$$\psi(f, g) = u\mathrm{Lip} f\|g\|_\infty + v\mathrm{Lip}\, g\|f\|_\infty.$$

- κ—weak dependence coefficients $\kappa_r = \epsilon_r$, here

$$\psi(f, g) = uv\mathrm{Lip} f \cdot \mathrm{Lip}\, g.$$

- λ—weak dependence coefficients $\lambda_r = \epsilon_r$, now

$$\psi(f, g) = u\mathrm{Lip} f\|g\|_\infty + v\mathrm{Lip}\, g\|f\|_\infty + uv\mathrm{Lip} f \cdot \mathrm{Lip}\, g.$$

Remark 2.2 (Heredity under non-Lipschitz functions) Here $(X_n)_{n\in\mathbb{N}}$ and we denote $Y_n = h(X_n)$, for some function h.

Then heredity of weak dependence for $(Y_n)_{n\in\mathbb{Z}}$ is simple in case of a Lipschitz function h.

We consider useful cases for which h is not Lipschitz.

- Polynomial functions

Lemma 2.1 *Let $(X_n)_{n\in\mathbb{Z}}$ be a sequence of $\mathbb{R}^k$-valued random variables. Let $p > 1$. We assume that there exists some constant $C > 0$ such that $\max_{1\le i\le k} \|X_i\|_p \le C$. Let h be a function from $\mathbb{R}^k$ to $\mathbb{R}$ such that $h(0) = 0$ and for $x, y \in \mathbb{R}^k$, there exist a in $[1, p[$ and $c > 0$ such that*

$$|h(x) - h(y)| \le c|x - y|(|x|^{a-1} + |y|^{a-1}) .$$

Then,

- if $(X_n)_{n\in\mathbb{Z}}$ is η-weakly dependent, then $(Y_n)_{n\in\mathbb{Z}}$ also, and

$$\eta_Y(r) = \mathcal{O}\left(\eta(r)^{\frac{p-a}{p-1}}\right);$$

- if $(X_n)_{n\in\mathbb{Z}}$ is λ-weakly dependent, then $(Y_n)_{n\in\mathbb{Z}}$ also, and

$$\lambda_Y(r) = \mathcal{O}\left(\lambda(r)^{\frac{p-a}{p+a-2}}\right).$$

- Discontinuous functions (indicator functions)

The empirical cumulative distribution needs $h_t(x) = \mathbb{I}_{\{x \le t\}}$ and this case extends to:

Lemma 2.2 *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of r.v's. We assume that h is a step-function with finitely many steps on $\mathbb{R}$ Suppose that, for some positive real constants C, α, λ*

$$\sup_{x\in\mathbb{R}} \sup_{i\in\mathbb{N}} \mathbb{P}\left(x \le X_i \le x + \lambda\right) \le C\lambda^{\alpha}. \tag{4}$$

(i) If the sequence (X_n) is η-dependent, then $(Y_n)_{n\in\mathbb{N}}$ is $(\mathcal{I}, c)$-dependent with $\epsilon(r) = \eta(r)^{\frac{\alpha}{1+\alpha}}$, $c(u, v) = 2(8C)^{\frac{1}{1+\alpha}}(u + v)$

(ii) If the sequence (X_n) is κ-dependent, then $(Y_n)_{n\in\mathbb{N}}$ is $(\mathcal{I}, c)$-dependent with

$$\epsilon(r) = \kappa(r)^{\frac{\alpha}{2+\alpha}}, c(u, v) = 2(8C)^{\frac{2}{2+\alpha}}(u + v)^{\frac{2(1+\alpha)}{2+\alpha}}$$

(iii) If the sequence (X_n) is θ-dependent, then $(Y_n)_{n\in\mathbb{N}}$ is $(\mathcal{I}, c)$-dependent with

$$\epsilon(r) = \theta(r)^{\frac{\alpha}{1+\alpha}}, c(u, v) = 2(8C)^{\frac{1}{1+\alpha}}(u + v)^{\frac{1}{1+\alpha}}$$

(iv) If the sequence (X_n) is λ-dependent (with $\lambda(r) \le 1$), then $(Y_n)_{n\in\mathbb{N}}$ is $(\mathcal{I}, c)$-dependent with

$$c(u, v) = 2\left((8C)^{\frac{1}{1+\alpha}} + (8C)^{\frac{2}{2+\alpha}}\right)(u + v)^{\frac{2+2\alpha}{2+\alpha}}, \epsilon(r) = \lambda(r)^{\frac{\alpha}{2+\alpha}}$$

An appropriate notion of weak dependence for the model (2) was introduced in [3]. It is based on the concept of the coefficient τ defined below. Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-subalgebra of $\mathcal{C}$ and X a random variable with values in E. Assume that $\|X\|_1 < \infty$ and define the coefficient τ as

$$\tau(\mathcal{M}, X) = \left\| \sup_{f\in\Lambda_1(E)} \left\{ \left| \int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx) \right\} \right\|_1 .$$

An easy way to bound this coefficient is based on a coupling argument:

$$\tau(\mathcal{M}, X) \le \|X - Y\|_1$$

for any Y with the same distribution as X and independent of $\mathcal{M}$. Moreover, if the probability space $(\Omega, \mathcal{C}, \mathbb{P})$ is rich enough (we always assume so in the sequel) there exists an X^* such that $\tau(\mathcal{M}, X) = \|X - X^*\|_1$. Using the definition of τ, the dependence between the past of the sequence $(X_t)_{t\in\mathbb{Z}}$ and its future k-tuples may be assessed: Consider the norm $\|x - y\| = \|x_1 - y_1\| + \cdots + \|x_k - y_k\|$ on E^k, set $\mathcal{M}_p = \sigma(X_t, t \le p)$ and define

$$\tau_k(r) = \max_{1\le l\le k} \frac{1}{l} \sup\left\{\tau(\mathcal{M}_p, (X_{j_1}, \dots, X_{j_l}))|\ p + r \le j_1 < \cdots < j_l\right\},$$
$$\tau_\infty(r) = \sup_{k>0} \tau_k(r).$$

For the sake of simplicity, $\tau_\infty(r)$ is denoted by $\tau(r)$. Finally, the time series $(X_t)_{t\in\mathbb{Z}}$ is τ-*weakly dependent* when its coefficients $\tau(r)$ tend to 0 as r tends to infinity.

3 Models with Infinite Memory

This section refers to results from [8].

3.1 Assumptions

The existence of a solution to (2) is proved under a Lipschitz-type condition. It is expressed in terms of Orlicz functions in order to handle general moments functions, see Theorem 3.1. These moments will be needed to establish the asymptotic results of Theorem 3.2.

Assume some Orlicz function Φ is such that for $x, y \in E^{\mathbb{N}}$:

$$\left\|F(x;\xi_0) - F(y;\xi_0)\right\|_\Phi \le \sum_{j=1}^{\infty} a_j \|x_j - y_j\|, \tag{5}$$

with $(a_j)_{j\ge1}$ a sequence of nonnegative real numbers such that

$$a = \sum_{j=1}^{\infty} a_j < 1, \tag{6}$$
$$\mu_\Phi = \left\|F(0,0,\dots;\xi_0)\right\|_\Phi < \infty. \tag{7}$$

The Lipschitz property of F and the moment assumption (7) induce that $\|F(c;\xi_0)\|_\Phi < \infty$ for any constant $c \in E^{\mathbb{N}}$. We set $c = (0, 0, 0, \dots)$ in condition (7) for convenience.

As a reminder, Orlicz spaces are convenient extentions of the classical $\mathbb{L}^m$-spaces. Let Φ be an Orlicz function, defined on $\mathbb{R}^+$, convex, increasing and satisfying $\Phi(0) = 0$. For any E-valued random variable X, the norm $\|X\|_\Phi$ is defined by the equation

$$\|X\|_\Phi = \inf\left\{u > 0, \quad \mathbb{E}\Phi\left(\frac{\|X\|}{u}\right) \le 1\right\},$$
$$\mathbb{L}^\Phi = \left\{E\text{-valued random variables } X \text{ such that } \|X\|_\Phi < \infty\right\}$$

is a Banach space equipped with the norm $\|\cdot\|_\Phi$. For $m \geq 1$ and $\Phi(x) = x^m$, notice that $\mathbb{L}^\Phi$ is the usual $\mathbb{L}^m$-space. Recall:

Lemma 3.1 *if the Orlicz function* Φ *satisfies* $\Phi(xy) \leq \Phi(x)\Phi(y)$, $\forall x, y \in \mathbb{R}_+$ *then* $\|X\|_1 \leq \|X\|_\Phi$ *for any E-valued random variable.*

3.2 Properties of the Stationary Solution

The following theorem settles the existence of a solution to (2). It also states that the Φ-th moment of this solution is finite.

Theorem 3.1 *Assume that conditions* (6) *and* (7) *hold for some Orlicz function* Φ *satisfying* $\Phi(xy) \leq \Phi(x)\Phi(y)$ $\forall x, y \in \mathbb{R}_+$. *Then there exists a* τ*-weakly dependent stationary solution* $(X_t)_{t\in\mathbb{Z}}$ *of* (2) *such that* $\|X_0\|_\Phi < \infty$ *and*

$$\tau(r) \leq 2\frac{\mu_1}{1-a} \inf_{1\leq p\leq r} \left(a^{r/p} + \frac{1}{1-a} \sum_{k=p+1}^{\infty} a_k \right) \to 0 \quad as \quad r \to \infty.$$

Remark 3.1 This is also proved that there exists some measurable function H such that $X_t = H(\xi_t, \xi_{t-1}, \ldots)$. The process $(X_t)_{t\in\mathbb{Z}}$ is thus a causal Bernoulli shift. For those processes, conditions (6) and (7) together imply the Dobrushin uniqueness condition. Thus $(X_t)_{t\in\mathbb{Z}}$ is the unique causal Bernoulli shift solution to (2). Moreover, as a causal Bernoulli shift, the solution $(X_t)_{t\in\mathbb{Z}}$ is automatically an ergodic process. Under the conditions of Theorem 3.1, the solution to (2) has finite Φ-th moment. From Lemma 3.1, $(X_t)_{t\in\mathbb{Z}}$ has also finite first order moments. The ergodic theorem yields the SLLN for any *chain with infinite memory* under the assumptions of Theorem 3.1.

Corollary 3.1 *[8] Under the assumptions of Theorem 3.1, and in case* $a_j = 0$ *for* $j > p$ *there exists a* τ*-weakly dependent stationary solution* $(X_t)_{t\in\mathbb{Z}}$ *to* (2) *such that* $\|X_0\|_\Phi < \infty$ *and* $\tau(r) \leq 2\mu_1(1-a)^{-1}\, a^{r/p}$ *for* $r \geq p$.

Dedecker and Prieur [3] stated that there exists $0 < \rho < 1$ and $C > 0$ such that $\tau(r) \leq C\rho^r$. Applying Corollary 3.1, we get the bound $\rho \leq a^{1/p}$. The bounds of the weak dependence coefficients in Theorem 3.1 come from an approximation with Markov chains of order p and from the result of Corollary 3.1.

In Theorem 3.1, the τ-weak dependence property is linked to the choice of the parameter p and then to the rate of decay of the Lipschitz coefficients a_j. For example, if $a_j \leq ce^{-\beta j}$, we choose p as the largest integer smaller than $\sqrt{-\ln(a)r/\beta}$ to derive the bound $\tau(r) \leq Ce^{-\sqrt{-\ln(a)\beta r}}$ for some suitable constant $C > 0$. If $a_j \leq cj^{-\beta}$, we choose the largest integer p such that $p \ln p(1-\beta)/\ln a \leq r$. Then there exists $C > 0$ such that $\tau(r) \leq Cp^{1-\beta}$. Notice that $\ln r$ is smaller than $\ln p + \ln\ln p$ up to a constant and that $\ln r/r$ is proportional to $1/p(1 + \ln\ln p/\ln p)$ and then equivalent to $1/p$

as p tends to infinity with r. From these equivalences, we achieve thus there exists $C > 0$ such that $\tau(r) \leq C\,(\ln r/r)^{\beta-1}$.

3.3 Asymptotic Results

In this section $E = \mathbb{R}$ and conditions (8a, 8b) yield the results proved in [1, 3].

Theorem 3.2 *Assume that conditions (6) and (7) hold for some Orlicz function Φ with $\Phi(xy) \leq \Phi(x)\Phi(y), \forall x, y \in \mathbb{R}_+$, and assume that there exists $c_0 > 0$ such that*

$$\sum_{k\geq 1} a_k \widetilde{\Phi}_q(c_0 k) < \infty \quad \text{if } \sum_{j>p} a_j = 0 \text{ for some } p \geq 1, \tag{8a}$$

$$\sum_{k\geq 1} a_k \widetilde{\Phi}_q\left(-c_0 k \ln\Big(\sum_{j\geq k} a_j\Big)\right) < \infty \quad \text{otherwise,} \tag{8b}$$

where $\widetilde{\Phi}_q$ is defined by $\widetilde{\Phi}_q(x) = \sup_{y>0}\{(xy)^{q-1} - \Phi(y)/y\}$.

Then:

SLLN *If $q \in]1, 2[$ then: $n^{-1/q}\sum_{i=1}^{n}(X_i - \mathbb{E}X_0) \to_{n\to\infty} 0$, a.s.*

CLT *If $q = 2$, then: $\frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}(X_i - \mathbb{E}X_0) \overset{D[0,1]}{\longrightarrow} \sigma W(t)$ as $n \to \infty$, where $\sigma^2 = \sum_{i=-\infty}^{\infty} Cov\,(X_0, X_i)$ is finite and $W(t)$ is the standard Wiener process.*

SIP *If $q = 2$ and if the underlying probability space is rich enough then there exist independent $\mathcal{N}(0, \sigma^2)$-distributed random variables $(Y_i)_{i\geq 1}$ such that*

$$\sum_{i=1}^{n}(X_i - Y_i) = o(\sqrt{n \ln\ln n}) \qquad a.s. \tag{9}$$

Note that $x^2\ln(1 + x)$-th moments are necessary to get the CLT for such weakly dependent processes. Note also that approximations by martingale difference or projective criterion and yields the CLT under weaker assumptions.

Condition (8a) is relevant for the Markov solution $(X_t)_{t\in\mathbb{Z}}$ to (2) i.e. when $\sum_{j>p} a_j = 0$. For the other cases, we rewrite assumption (8b) for various rates of decay of the Lipschitz coefficients a_j. Let a, b, c be some positive real numbers then there exists $c_0 > 0$ such that if

$$a_k \leq ck^{-a}, \qquad \text{then } \sum_{k\geq 1} a_k \widetilde{\Phi}_q\left(c_0 k \ln k\right) < \infty, \tag{8a'}$$

$$a_k \leq ce^{-ak^b}, \qquad \text{then } \sum_{k\geq 1} a_k \widetilde{\Phi}_q\left(c_0 k^{1+b}\right) < \infty. \tag{8a''}$$

For instance, condition (8a′) holds if $\Phi(x) = x^m$ for $m > q$ and $a > 1 + (q-1)(m-1)(m-q)^{-1}$. Condition (8a″) holds for $\Phi(x) = x^q(1+\ln(1+x))^{(1+b)(q-1)}$. Applying Theorem 3.2, the CLT and the SIP (9) both hold for sub-geometric rates of decay of the Lipschitz coefficients as in (8a″) under the existence of $x^2(1+\ln(1+x))^{1+b}$-moment.

4 Asymmetric ARCH(1)-model

General variants of those models, ARCH(∞) are carefully studied in [6] and even give rise to long range memory models. This section derive specific properties of such very simple models.

4.1 Existence of a Stationary Solution

This is easy to derive from the above results that $a \cdot \|\xi_0\|_p < 1$ for $p \geq 1$, is the condition for the existence of a stationary solution in $\mathbb{L}^p$, moreover:

Lemma 4.1 *for each* $\alpha \in]0,1[$*, and* $x \in \mathbb{R}$*, there exists* $A > 0$ *such that*

$$|(ax+b)^2 + c^2|^{\frac{p}{2}} \geq (1-\alpha)|ax|^p - A.$$

Proof Write

$$ax = (1-\epsilon)\frac{ax+b}{1-\epsilon} + \epsilon\frac{-b}{\epsilon}, \text{ for each } \epsilon \in]0,1[,$$

then as $p \geq 1$ the function $x \mapsto |x|^p$ is convex, thus

$$\begin{aligned} |ax|^p &\leq (1-\epsilon)\left(\frac{|ax+b|}{1-\epsilon}\right)^p + \epsilon\left(\frac{|b|}{\epsilon}\right)^p \\ (1-\epsilon)^{p-1}|ax|^p &\leq |ax+b|^p + \left(\frac{1}{\epsilon}-1\right)^{p-1}|b|^p \\ &\leq |(ax+b)^2 + c^2|^{\frac{p}{2}} + \left(\frac{1}{\epsilon}-1\right)^{p-1}|b|^p \end{aligned}$$

The required inequality holds with:

$$\alpha = 1-(1-\epsilon)^{p-1}, \qquad A = \left(\frac{1}{\epsilon}-1\right)^{p-1}|b|^p. \qquad \square$$

Proposition 4.1 *Let* $1 \leq q < p$*. If* $a \cdot \|\xi_0\|_p > 1$ *and* $a \cdot \|\xi_0\|_q < 1$*, then the* $\mathbb{L}^q$*-stationary solution is not in* $\mathbb{L}^p$*.*

Proof First quote that 0 is not solution the above equation. Choose α small enough so that $\beta = (1-\alpha)a^p\mathbb{E}|\xi_0|^p > 1$, we derive:

$$\begin{aligned}\mathbb{E}|X_1|^p &= \mathbb{E}|\xi_1|^p(\mathbb{E}(aX_0+b)^2+c^2)^{\frac{p}{2}}\\ &\geq \beta\mathbb{E}|X_0|^p - A\mathbb{E}|\xi_1|^p\end{aligned}$$

Then $\mathbb{E}|X_1|^p - B \geq \beta(\mathbb{E}|X_0|^p - B)$, for $B = A\mathbb{E}|\xi_1|^p/(\beta-1)$.

So that if we suppose that the $\mathbb{L}^q$-stationary solution is also in $\mathbb{L}^p$ we derive $\mathbb{E}|X_0|^p = 0$. This contradiction proves the result.

An alternative proof follows. Set $\gamma_t = \mathbb{E}|X_t|^p - B$, the above inequalities yield $\gamma_t \geq \beta\gamma_{t-1}$ and thus $\gamma_t \geq \beta^t\gamma_0$ thus $\gamma_t = +\infty$ if $\mathbb{E}|X_0|^p \neq 0$ leading again to a contradiction. □

5 Moment Estimators

5.1 Explicit Moments

Proposition 5.1 *Assume that $p \geq 1$ then the previous stationary solution satisfies $\mathbb{E}X_0 = 0$. If $p \geq 2$, $\mathbb{E}\xi_0 = 0$, and $\mathbb{E}\xi_0^2 = 1$, then*

$$\mathbb{E}X_0^2 = \frac{b^2+c^2}{1-a^2}.$$

Proof The first part follows from the independence of ξ_t and X_{t-1}. For the second part squaring the ARCH equation, quote that $M = \mathbb{E}X_0^2$ satisfies: $M = a^2M + b^2 + c^2$. □

Proposition 5.2 *In case $p = 2$ and $a^2 \cdot \mathbb{E}|\xi_0|^2 = 1$, and $a\|\xi_0\|_q < 1$ for some $q < 2$, then $\mathbb{E}X_0^2 = \infty$, if moreover $b^2 + c^2 \neq 0$.*

Proof In case $p = 2$ and $a^2\mathbb{E}|\xi_0|^2 = 1$, we have

$$\mathbb{E}|X_1|^2 = \mathbb{E}|\xi_1|^2(\mathbb{E}(aX_0+b)^2+c^2) = \mathbb{E}|X_1|^2 + \mathbb{E}|\xi_1|^2(b^2+c^2)$$

Hence $b^2 + c^2 > 0$ entails $\mathbb{E}X_0^2 = \infty$. □

Remark 5.1 Hence this again proves that Proposition 4.1 is essentially tight.

Proposition 5.3 *If $p = 3$ and ξ_0 admits a symmetric distribution, then $\mathbb{E}X_1^3 = 0$.*

Proof If $p = 3$ and $\mathbb{E}\xi_0 = 0$, we derive

$$\mathbb{E}X_1^3 = \mathbb{E}\xi_1^3\mathbb{E}((aX_0+b)^2+c^2)^{\frac{3}{2}}.$$

Thus $\mathbb{E}X_0^3 = 0$ since $\mathbb{E}\xi_0^3 = 0$. □

Proposition 5.4 *Set $\ell_t = Cov\,(X_0, X_t^2)$, the leverage for $t > 0$. It measures the asymmetry properties of a random process. If again $p = 3$ and if ξ_0 admits a symmetric distribution with $\mathbb{E}\xi_0^2 = 1$, then for $t \geq 1$:*

$$\ell_t = \frac{2a^{2t-1}b(b^2+c^2)}{1-a^2}.$$

Proof Here $\mathbb{E}\xi_0^3 = 0$, thus $\mathbb{E}X_0^3 = 0$, and

$$\ell_1 = \mathrm{Cov}\,(X_0, X_1^2) = a^2\mathbb{E}X_0^3 + \frac{2ab(b^2+c^2)}{1-a^2} = \frac{2ab(b^2+c^2)}{1-a^2}$$

Taking into account that $\mathbb{E}\xi_0^2 = 1$ and $\mathbb{E}X_0 = 0$ yields

$$\begin{aligned}\ell_2 = \mathbb{E}X_0X_2^2 &= \mathbb{E}X_0((aX_1+b)^2+c^2)\\ &= a^2\mathbb{E}X_0X_1^2 + 2ab\mathbb{E}X_0X_1 = \frac{2a^3b(b^2+c^2)}{1-a^2}\end{aligned}$$

Then a recursion entails:

$$\begin{aligned}\ell_t = \mathbb{E}X_0X_t^2 &= \mathbb{E}X_0((aX_{t-1}+b)^2+c^2)\\ &= a^2\mathbb{E}X_0X_{t-1}^2 + 2ab\mathbb{E}X_0X_{t-1} + c^2\mathbb{E}X_0\\ &= a^2\mathbb{E}X_0X_{t-1}^2\\ &= \frac{2a^{2t-1}b(b^2+c^2)}{1-a^2}.\end{aligned}$$

□

5.2 Moment Based Estimation

The previous triplet writes $(M, \ell_1, \ell_2) = F(a, b, c)$. From those relations $\ell_2/\ell_1 = a^2$ and $\ell_1 = 2abM$ imply $a = \sqrt{\ell_2/\ell_1}$, $b = \ell_1^{3/2}/2M\ell_2^{1/2}$. Finally $c^2 = M(1-a^2) - b^2 = M(1-\ell_2/\ell_1) - \ell_1^3/4M^2\ell_2$.

This writes

$$(a,b,c) = \left(\sqrt{\frac{\ell_2}{\ell_1}}, \frac{\sqrt{\ell_1^3}}{2M\sqrt{\ell_2}}, \sqrt{M\left(1-\frac{\ell_2}{\ell_1}\right) - \frac{\ell_1^3}{4M^2\ell_2}}\right) \tag{10}$$

Since the above relation defines the inverse of the function F, an accurate estimate of the model writes:

$$(\widehat{a},\widehat{b},\widehat{c}) = F^{-1}\left(\widehat{M},\widehat{\ell}_1,\widehat{\ell}_2\right)$$

Here we consider the standard empirical estimates:

$$\widehat{M} = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

$$\widehat{\ell}_1 = \frac{1}{n-1}\sum_{i=1}^{n-1} X_i X_{i+1}^2,$$

$$\widehat{\ell}_2 = \frac{1}{n-2}\sum_{i=1}^{n-2} X_i X_{i+2}^2.$$

Quote that the above (ℓ_1,ℓ_2)-based estimators may be replaced by (ℓ_i,ℓ_j)-based estimators for $0 < i < j$, but even considering all of them by aggregation would not improve on convergence rates.

5.3 Asymptotic Considerations

By using the previous transforms the only asymptotic to be considered is for empirical estimates of M,ℓ_1,ℓ_2. The *a.s.* consistency of those estimates follows from the ergodic theorem. Geometric θ-weak dependence holds. Now in order to get all asymptotic results assume the moment condition $a\|\xi_0\|_{6+\epsilon} < 1$, for some $\epsilon > 0$ ([1]).

Denote $\widehat{\alpha} = (\widehat{a},\widehat{b},\widehat{c})$, then

$$\sqrt{n}(\widehat{\alpha}-\alpha) \to^{\mathcal{D}}_{n\to\infty} \mathcal{N}(0,\Sigma_\alpha),$$

yielding asymptotic confidence intervals.

[1] In fact (X_i) only needs general moments with $\Phi(x) = x^6\log^3 x$, unfortunately the conditions in Sect. 3.2 do not write simply in this case.

Now (9) applies with $\bar{X}_i = X_i^2, X_i X_{i+1}^2$ and $X_i X_{i+2}^2$ since the heredity Lemma 2.1 holds here. We thus also conclude that $\widehat{\alpha}$ is a strongly consistent estimator for α and fulfills a LIL. More precisely this implies that whatever is the norm $\|\cdot\|$ on $\mathbb{R}^3$:

$$\limsup_{n\to\infty} \sqrt{\frac{n}{\ln\ln(n)}} \cdot \|\widehat{\alpha} - \alpha\| < \infty, \qquad a.s. \tag{11}$$

6 Estimating the Density of the Residuals

We shall write for simplicity

$$X_t = \sigma(X_{t-1}) \cdot \xi_t, \quad \text{with} \quad \sigma(x) = \sqrt{(ax+b)^2 + c^2}.$$

The previous section is devoted to fit $\alpha = (a, b, c)$, when a n-sample is observed. We now let $\widehat{\sigma}(x) = \sqrt{(\widehat{a}x + \widehat{b})^2 + \widehat{c}^2}$.

We estimate the density f of the (non-observed) residuals ξ_i. For this assume that this density exists and that it is bounded. We want to estimate the density f by means of the data $X_{i+1}, \ldots, X_{i+N}$ for some convenient i. A main problem is that we do not observe the residuals ξ_i. It is addressed below.

6.1 Fitted Residuals

Above we conducted the estimation of the function σ by using a sample $X_1, \ldots, X_n$, in order to fit N residuals $(\xi_i)_{i\in A_N}$ we shall let $n = n(N) \uparrow \infty$ as $N \uparrow \infty$, depend on $N \uparrow \infty$. Now we need to make the sample $(X_i)_{i\in A_N}$ asymptotically independent of the sample $X_1, \ldots, X_n$ thus we introduce another sequence $k = k(N) \uparrow \infty$ as $N \uparrow \infty$ so that $A_N = \{i,\ 1 \le i \le n + k + N\}$.

Below we assume that a sample $(X_i)_{1\le i\le n_0}$ is observed and we thus separate into 3 parts, the first part is devoted to parametric estimation of the coefficients, then k-random variables are discarded in order to infer asymptotic independence, the last part of the sample is dedicated to fit the residuals and then the density of innovations, and $n_0 = n + k + N$.

We assume for simplicity that $k(N) = [N^\epsilon]$ for some $0 < \epsilon < 1$.

Theorem 6.1 *Set*

$$\Delta_N = \frac{1}{N} \sum_{i\in A_N} |\widetilde{\xi}_i - \xi_i|, \quad \textit{with } \widetilde{\xi}_i = \frac{X_i}{\widehat{\sigma}(X_{i-1})}.$$

Then under the above assumptions

$$\limsup_{N\to\infty} \sqrt{\frac{N}{\ln\ln n(N)\cdot \ln\ln N}}\cdot \Delta_N < \infty, \qquad a.s.$$

Proof First assume that the set A_N is such that $\widehat{\sigma} \perp\!\!\!\perp (X_i, i \in A_N)$ is independent of the estimation sample, then the SIP entails

$$\Delta_N = \mathcal{O}_{a.s}\left(\sqrt{\frac{\ln\ln n\cdot \ln\ln N}{n}}\right)$$

Write $\sigma(\alpha, x) = \sqrt{(ax+b)^2 + c^2}$ then its differential writes

$$D_\alpha \sigma(\alpha, x)(u, v, w) = \frac{1}{\sigma(\alpha, x)}\cdot (uax^2 + vb + wc),$$

so that

$$|\widehat{\sigma}(x) - \sigma(x)| \le \frac{1}{\sigma(x)}\cdot (|a-\widehat{a}||a|x^2 + |b-\widehat{b}||b| + |c-\widehat{c}||c|),$$

and $|\widehat{\sigma}(x) - \sigma(x)| \le \text{cst.}\sqrt{\frac{\ln\ln(n)}{n}}(x^2+1)$ from (11). Thus $\sigma(x)$ is differentiable as a function of α with partial derivatives $\mathcal{O}(x^2+1)$. Hence

$$\begin{aligned}
|\widetilde{\xi}_i - \xi_i| &= \frac{|X_i|}{\widehat{\sigma}(X_{i-1})}(\sigma(X_{i-1}) - \widehat{\sigma}(X_{i-1})) \\
&\le \text{cst.}\ \frac{|X_i|(X_{i-1}^2+1)}{c\widehat{c}}\max\{|\widehat{a}-a|, |\widehat{b}-b|, |\widehat{c}-c|\} \\
&= \mathcal{O}_{a.s}\left(\sqrt{\frac{\ln\ln(n)}{n}}\right)|X_i|(X_{i-1}^2+1)
\end{aligned}$$

The SIP (9) for $X_i(X_{i-1}^2+1)$ and an independence argument together entail the result in this case since $Y_i = |X_i|(X_{i-1}^2+1)$ satisfies SIP thus LIL entails $\sum_{i=1}^N Y_i = \mathcal{O}_{a.s}(N\ln\ln N)$.
We use the coupling argument in [3] under τ-weak dependence in the general case:

$$\mathbb{E}|X_i - X_i^*| \le \tau(k)$$

where $X_i^* \perp\!\!\!\perp \mathcal{F}((\xi_i)_{1\le i\le n})$ and where $\tau(k) = \tau(k(N))$ is here a geometrically decaying sequence. We obtain $\widetilde{\xi}_i^*$ from X_i^* and now $\Delta_N \le \Delta_N^* + \delta_N$ for $\delta_N \ge 0$ with $\mathbb{E}\delta_N \le \tau(k)$, thus: $\mathbb{P}(\delta_N \ge t) \le \frac{\tau(k(N))}{t}$ implies $\delta_N = \mathcal{O}_{a.s}(1/n)$ (from Borel-Cantelli lemma

and from the geometric decay of the sequence $N \mapsto \tau(k(N))$ and we obtained above (case $(X_i)_{1\le i\le n} \perp\!\!\!\perp (X_i)_{j\in A_N}$)

$$\Delta_N^* = \mathcal{O}_{a.s}(1)\sqrt{\frac{\ln\ln(n)\cdot\ln\ln(N)}{n}}$$

The result follows. □

Remark 6.1 (*Bootstrap*) Consider the discrete measure

$$\nu^* = \frac{1}{N}\sum_{j\in A_N}\delta_{\widehat{\xi}_i}, \quad \widehat{\xi}_i = \widetilde{\xi}_i - \frac{1}{N}\sum_{j\in A_N}\widetilde{\xi}_j$$

then a bootstrapped version of our ARCH(1) models is the solution of the equation $X_t^* = \widehat{\sigma}(X_{t-1}^*)\cdot\xi_i^*$ for an iid sequence $\xi_i^* \sim \nu^*$.

Quote that the above distribution is centered conditionally wrt observations. The attractive feature of weak dependence conditions relies on the fact that the $\mathbb{L}^p$-stationarity of the above solution only relies on $a\cdot\|\xi_0\|_p < 1$. Indeed the empirical counterpart of the above expression again satisfies $\widehat{a}\cdot\|\xi_0^*\|_p < 1$ for N large enough.

6.2 Density Estimation

On the basis of the estimator $\widehat{\alpha}$, we compute the estimated residuals $\widetilde{\xi}_i$ for $i \in A_N$. As suggested in [9] the density f is estimated by the kernel estimator $\widetilde{f}_N$:

$$\widetilde{f}_N(x) = \frac{1}{Nb(N)}\sum_{i\in A_N} K\left(\frac{\widetilde{\xi}_i - x}{b(n)}\right), \qquad \forall x\in\mathbb{R}.$$

Suppose that the bandwidth $b(n)$ satisfies $\lim_{N\to\infty} b(N) = 0$. We assume that for some integer $r \ge 2$ the kernel fulfills the following condition:

CONDITION *K(r)* : K is a symmetric Lipschitz function which vanishes outside the interval $[-1, 1]$. Assume $\int_{-1}^{1} K(t)dt = 1$ and

$$\int_{-1}^{1} t^r K(t)dt \neq 0, \quad \int_{-1}^{1} t^k K(t)dt = 0, \text{ for each integer } 0 < k < r.$$

A standard literature for density estimation of the iid sequence (ξ_i) is e.g. the nice monograph, [10], see also [5]. Relying on ideas from [9] we derive:

Theorem 6.2 *Suppose that, for some even integer $r \ge 2$, the derivative $f^{(r)}$ exists and is continuous on $\mathbb{R}$. Let $D \subset \mathbb{R}$ be some compact interval we assume that $b(N) =$*

$C(N/\ln N)^{-\beta}$, $k(N) = N^{\epsilon}$ and $n(N) = N^{\gamma}$ for $0 < \beta < 1, 0 < \epsilon < 1$ and $\gamma > 4\beta$. Then, under the assumption $K(r)$, we have

$$\sup_{x\in D} |\widetilde{f}_N(x) - f(x)| = \mathcal{O}_{a.s}(\sqrt{\ln(N)}N^{-1/2}b(N)^{-1/2} + b^r(N))$$

Optimizing those expressions yields

$$\sup_{x\in D} |\widetilde{f}_N(x) - f(x)| = \mathcal{O}_{a.s}((\ln(N)/N)^{r/(2r+1)}),$$

if $b(N) = C_1 \cdot (\ln(N)/N)^{1/(2r+1)}$ and $\gamma > 4/(2r+1)$.

Theorem 6.3 *Assume that f is Lipschitz continuous in a neighborhood of $x \in \mathbb{R}$ and $f(x) > 0$. Then under the assumptions in Theorem 6.2 with now $b(N) = CN^{-\beta}$, and $\gamma > 1 + 3\beta$, and under the assumptions $K(r)$:*

$$\sqrt{Nb(N)}(\widetilde{f}_N(x) - \mathbb{E}\widehat{f}_N(x)) \to^{\mathcal{D}} \mathcal{N}(0, \sigma^2(x))$$

where $\sigma^2(x) = f(x)\int_{-1}^{1} K^2(t)dt$.

Corollary 6.1 *Assume that $b(N) = C_2 \cdot N^{-1/(2r+1)}$ for some positive constant C_2, if $f^{(r)}$ exists and is continuous at $x \in \mathbb{R}$ for some even $r \geq 2$. Suppose that $f(x) > 0$ then, under the assumptions $K(r)$:*

$$\sqrt{Nb(N)}(\widehat{f}_n(x) - f(x)) \to^{\mathcal{D}} \mathcal{N}(\mu(x), \sigma^2(x))$$

with $\mu(x) = \frac{C_2^{r+1/2}}{r!}\int_{-1}^{1} t^r K(t)dt \cdot f^{(r)}(x)$.

Remark 6.2 Among others conditions [9] also assumes that the second derivative K'' exists on $[-1, 1]$ and moreover assume that $C_2 = \sup_{t\in[-1,1]} |K''(t)| < +\infty$. His proof is really technical and relies on a second order Taylor expansion. However he does not assume any restriction on $n(N)$ and he chooses $n(N) = N$ and $A_N = \{1, \ldots, N\}$. The frame is thus much more restrictive.

In our case we assume $n \sim N^{\gamma}$ and we need $\gamma > 4\beta$ for the uniform law of large numbers (Theorem 6.2; this is not a really restrictive condition since the right exponent is $\beta = 1/(2r+1)$, and the condition reduces to $\gamma > 4/5$ and includes $\gamma = 1$ as in [9]. For very large values this is a weak assumption $\gamma > 4/(2r+1) \downarrow 0$ as $r \uparrow \infty$. The central limit Theorem 6.3 needs $\gamma > 1 + 3\beta$ which is always more than 1 (essentially $\gamma > 1$ for high order differentiable densities f). The latter condition is more restrictive.

Proof If $(\xi_i)_{1\leq i\leq n}$ are observed, then the standard kernel density estimator of f writes

$$\widehat{f}_N(x) = \frac{1}{Nb(N)} \sum_{i\in A_N} K\left(\frac{\xi_i - x}{b(n)}\right), \qquad \forall x \in \mathbb{R}.$$

we have classically

$$\mathbb{E}\widehat{f}(x) - f(x) \sim b^r \cdot \frac{f^{(r)}(x)}{r!} \int_{-1}^{1} t^r K(t)dt$$

Since $\ln u = o(u)$ as $u \uparrow \infty$ the powers of ln ln's are $\mathcal{O}(\ln N)$.

Now using the above Theorem 6.1 and assumptions a Lipschitz argument entails

$$|\widetilde{f}(x) - \widehat{f}(x)| \leq \Delta_N \frac{\text{Lip } K}{b(N)^2} = \mathcal{O}_{a.s}(\ln N \cdot N^{2\beta-\gamma/2})$$

In order to make the substitution ξ_i by $\widetilde{\xi}_i$ negligible, we assume $\gamma > 4\beta$ in order to derive Theorem 6.2. For the optimized case this means $\beta = 1/(2r+1)$ and thus $\gamma > 4/(2r+1)$ (for $r = 2$ this is only $\gamma > 4/5$ and for large values of r this essentially means $\gamma > 0$).

Theorem 6.3 thus needs $\lim_N \sqrt{Nb(N)}\Delta_N \text{Lip } K/b(N)^2 = 0$. Quote that $\sqrt{Nb(N)}$ $\Delta_N/b(N)^2 = \mathcal{O}_{a.s}(\ln N \cdot N^{1/2+3\beta/2-\gamma/2})$ and we need here $\gamma > 1 + 3\beta$. □

Acknowledgements This work has been developed within the MME-DII center of excellence (ANR-11-LABEX-0023-01). We also wish to thank Pr. Jacek Leskow for his hospitality and his scientific generosity.

References

1. Dedecker, J., & Doukhan, P. (2003). A new covariance inequality and applications. *Stochastic Processes and their Applications*, *106*, 63–80.
2. Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S., & Prieur, C. (2007). Weak dependence: With examples and applications. Lecture Notes in Statistics (Vol. 190). New York: Springer.
3. Dedecker, J., & Prieur, C. (2004). Coupling for τ-dependent sequences and applications. *Journal of Theoretical Probability*, *17*, 861–885.
4. Doukhan, P. (1994). *Mixing: properties and examples*. Lecture Notes in Statistics. Springer.
5. Doukhan, P. (2016). *Stochastic models for time series*. Mathématiques et Application: Springer.
6. Doukhan, P., Grublyté, I., & Surgailis, D. (2016). A nonlinear model for long memory conditional heteroscedasticity. *Lithuanian Journal of Mathematic*, *56*.
7. Doukhan, P., & Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, *84*, 313–342.
8. Doukhan, P., & Wintenberger, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications*, *118*, 1997–2013.
9. Liebscher, E. A. (1999). Estimating the density of residuals in autoregressive models. *Statistical Inference for Stochastic Processes*, *2*, 105–117.
10. Rosenblatt, M. (1991). *Stochastic curve estimation*. NSF-CBMS Regional Conference Series in Probability and Statistics (Vol. 3).

Subsampling for Non-stationary Time Series with Long Memory and Heavy Tails Using Weak Dependence Condition

Elżbieta Gajecka-Mirek and Oskar Knapik

Abstract Statistical inference for unknown distributions of statistics or estimators may be based on asymptotic distributions. Unfortunately, in the case of dependent data the structure of such statistical procedures is often ineffective. In the last three decades we can observe an intensive development the of so-called resampling methods. Using these methods, it is possible to directly approximate the unknown distributions of statistics and estimators. A problem that needs to be solved during the study of the resampling procedures is the consistency. Their consistency for independent or stationary observations has been extensively studied. Resampling for time series with a specific non-stationarity, i.e. the periodic and almost periodic strong mixing dependence structure also been the subject of research. Recent research on resampling methods focus mainly on the time series with the weak dependence structure, defined by Paul Doukhan, Louhichi et al. and simultaneously Bickel and B*ü*hlmann. In the article a time series model with specific features i.e.: long memory, heavy tails (with at least a fourth moment, e.g.: GED, t-Student), weak dependence and periodic structure is presented. and the estimator of the mean function in the above-mentioned time series is investigated. In the article the necessary central limit theorems and consistency theorems for the mean function estimator (for one of the resampling techniques—the subsampling) are proven.

E. Gajecka-Mirek (✉)
Departament of Economics, State Higher Vocational School in Nowy Sącz,
ul. Jagiellonska 61, Nowy Sącz, Poland
e-mail: egajecka-mirek@pwsz-ns.edu.pl; egajecka@gmail.com

O. Knapik
CREATES, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark
e-mail: oknapik@econ.au.dk

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_2

1 Introduction

Possibility to construct sampling distributions of estimators for time series is very important in statistical studies. Traditional statistical inference based on asymptotic distributions does not always lead to effective statistical procedures. There are several reasons for this, e.g.:

- the convergence of the estimator to the asymptotic distribution is slow and often requires a large collection of observations. In practice, there is not always the possibility to receive enough data because of the costs or technical restrictions.
- The asymptotic distribution is often very complicated and depends on the unknown parameters, which in the case of dependent data is difficult to estimate.

In such situations, the so-called resampling methods (Efron [14]) are helpful. Moreover, in many cases these methods are the only effective technique. These methods allow us to approximate the unknown distributions (or characteristics) of the statistics and estimators without a reference to the form of the distribution. These approximations are used to construct the confidence intervals for the parameters and testing statistical hypothesis. (One can compute the confidence intervals and critical values from the resampling distributions instead of the asymptotic distributions.)

In the nineties of the last century the research was focused on stationary time series. At the present time, the efforts of researchers are concentrated on the non-stationary series, with discrete and continuous time. See, e.g. [13, 16, 26, 29] or [35] among many others.

One of the specific form of non-stationarity is periodicity. Gladyshev [18] initiated the development of research on periodicity in time series and stochastic process.

In 2006 Gardner et al. [17] have provided a general overview of research on periodicity and time series, considering over 1500 published papers on the topic. It was shown that the models with periodic structure are widely applicable e.g. in communication, signal processing, vibromechanics, econometrics, climatology and biology.

The resampling methods for periodic time series is an open research area, where many fundamental properties have yet to be proven.

In the article, we will deal only with one form of resampling—subsampling, since we will work with long memory time series.

The attention will be focused on the class of time series which simultaneously deals with three features: periodic structure, heavy tails (finite at least fourth moment) and long memory.

The heavy-tailed random variables are variables with distributions whose extreme values are "more probable than normal". Examples of such distributions are the Generalized Error Distribution (GED) distributions or stable distributions. Both classes will be discussed in this thesis. Additionally, in real data sets one has to deal with long range dependence as well.

The presence of long range dependence in time series means that there exists dependence between observations which are distant in time from each other.

It is obvious that among the observations of the time series there is a relationship—the dependence. Over the years, the most popular way for studying this dependence

have been the mixing conditions. The fact is that under natural restrictions on the process parameters, many processes of interest fulfill mixing conditions [11]. On the other hand, there is a large class of processes for which mixing properties do not hold.

Simple example of such a time series is a stationary $AR(1)$ process:

$$X_t = aX_{t-1} + \varepsilon_t,$$

where the innovations are independent with $P(\varepsilon_t = 1) = P(\varepsilon_t = -1) = 1/2$ and $0 < |a| \leq 1/2$.

This process has a stationary distribution on $[-2, 2]$ and X_t has always the same sign as ε_t. It is possible to recover $X_{t-1}, X_{t-2}, \ldots$ from X_t, it means that the process $\{X_t\}_{t\in\mathbb{Z}}$ is purely deterministic going backwards in time, so it cannot be strong mixing (proof is given in [1]).

More examples of weakly dependence sequences can be found in the research of Doukhan et al. [8, 21].

In 1999 Doukhan and Louhichi [12] and simultaneously Bickel and Bühlmann [6] proposed an alternative condition for the dependence in time series called weak dependence and ν-dependence, respectively.

The definition of weak dependence in comparison to, for example, mixing is very general. It includes very general data sets and models like causal, non causal linear, bilinear, strong mixing processes, dynamical systems or Markov processes driven by discrete innovations.

The main objective of this article is to introduce the theoretical and practical results describing the consistency of subsampling method and to show how to use them in statistical inference for time series with periodic behavior.

Three specific features of time series will be studied: heavy tails, long memory and periodic behavior. The construction of described in the thesis process entails the weak dependence property.

The central limit theorem for the mean estimator will be given. The subsampling method to estimate the mean vector will be presented and the applications of the central limit theorem to prove the consistency of subsampling method will be shown.

The structure of the article is as follows. In the second section the definitions, the examples and the main ideas will be introduced. The third section contains the construction of long memory, heavy tailed and periodically stationary model and its properties. The fourth section contains the central limit theorems. In the next section the consistency of the subsampling is shown.

2 Basic Concepts and Definitions

In this section the basic concepts and definitions will be presented. Some of them will be illustrated by examples.

Many real life phenomena are characterized by a seasonal behavior which, obviously, is not non-stationary. Seasonal data appear in such fields as: economics, biology, climatology, telecommunications and many others. If seasonality is not easily removable it means that we are dealing with a particular type of non-stationarity, for example the periodic structure. In such cases it is not just the mean that has a periodic rhythm. A periodic rhythm also describes the behavior of covariance.

Popular models used for describing such phenomena are periodically nonstationary processes.

Below are introduced the formal definitions of periodicity of the time series.

For the time series $\{X_t\}_{t\in\mathbb{Z}}$ we define the autocovariance of the pair (X_t, X_{t+h}) to be

$$\gamma_X(t,h) = Cov(X_t, X_{t+h}).$$

Definition 1 ([25], p. 5) Time series $\{X_t\}_{t\in\mathbb{Z}}$ is periodically correlated (PC) in Gladyshev sense, if the mean $\mu_X(t)$ is periodic ($\mu_X(t) = \mu_X(t+T)$) and the autocovariance function $\gamma_X(t,h)$ is periodic in t for all $h \in \mathbb{Z}$ ($\gamma_X(t,h) = \gamma_X(t+T,h)$).

If there is no ambiguity, we will write $\gamma(t,h)$ (or $\gamma(h)$ if we deal with classical weak stationarity) instead of $\gamma_X(t,h)$ for the autocovariance function of time series $\{X_t\}_{t\in\mathbb{Z}}$.

In many periodic phenomena the existence of the long range dependence is observed [5, 24, 31]. The presence of long range dependence in time series means that there exists a relationship between observations which are far away from each other in time. Long memory occurs in the sense that a hyperbolic behavior of the autocorrelations holds for almost all lags and frequencies respectively.

The long range dependence can be defined as long memory. Note that the definition of long memory introduced below is one of many possible definitions.

Definition 2 ([7], p. 520) A stationary, time series $\{X_t\}_{t\in\mathbb{Z}}$ has long memory if its autocovariance function γ satisfies the following formulas:

$$\sum_{0<|h|<n} \gamma(h) \sim Cn^{\beta}$$

where $\beta \in [0,1)$, and $C \neq 0$.

Definition 3 A PC time series $\{X_t\}_{t\in\mathbb{Z}}$ has a long memory if the autocovariance function $\gamma(s)(h) = Cov(X_{s+qT}, X_{s+(q+h)T})$ for each $q \in \mathbb{Z}$ satisfies the following formula

$$\sum_{0<|h|<n} \gamma(s)(h) \sim C(s)n^{\beta}, \quad s \in \{1,\dots,T\}$$

where $\beta \in [0,1)$. For each $s \in \{0,\dots,T-1\}$ $C(s)$ is the finite constant such that

$$C(s) = 2 \cdot \lim_{n\to\infty} \frac{\sum_{h=1}^{n-1} \gamma(s)(h)}{n^{\beta}} > 0.$$

Let us assume that the notation for the long memory with parameter $\beta \in [0, 1)$ will be $LM(\beta)$.

Suitable tool to describe the long memory behavior are so-called Gegenbauer processes [15, 23]. This kind of process will be used in modeling the long range dependence in this article.

Definition 4 ([19]) Let us assume that ε_t is i.i.d. innovation process. The process $\{G_t\}_{t\in\mathbb{Z}}$ defined by the equation:

$$\Pi_{1\le i\le k}(I - 2\nu_i B + B^2)^{d_i} G_t = \varepsilon_t, \tag{1}$$

is the k-factor Gegenbauer process.
$0 < d_i < 1/2$ if $|\nu_i| < 1$ or $0 < d_i < 1/4$ if $|\nu_i| = 1$ for $i = 1, ..., k$ and I is identity operator, B is backshift operator.

Theorem 1 ([21]) *Process defined by the Definition 4 is long memory, stationary, causal and invertible and has a moving average representation:*

$$G_t = \sum_{j\ge 0} \psi_j(d, \nu)\varepsilon_{t-j},$$

with $\sum_{j=0}^{\infty} \psi_j^2(d, \nu) < \infty$, *where* $\psi_j(d, \nu), j \ge 0$, *is defined by:*

$$\psi_j(d, \nu) = \sum_{\substack{0 \le l_1, ..., l_n \le j \\ l_1 + ... + l_n = j}} C_{l_1}(d_1, \nu_1) \cdot ... \cdot C_{l_k}(d_k, \nu_k),$$

where $C_{l_i}(d_i, \nu_i)$ *are the Gegenbauer polynomials defined as follows:*

$$(1 - 2\nu z + z^2)^{-d} = \sum_{j\ge 0} C_j(d, \nu) z^j, \ |z| \le 1, \ |\nu| \le 1.$$

Moreover, if $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *in the Definition 4 is the Gaussian white noise, then* $\{G_t\}_{t\in\mathbb{Z}}$ *is Gaussian time series.*

The Gegenbauer processes are stationary, seasonal fractional models; see: [19, 20]. The form of the $\psi_j(d, \nu)$ also can be found in [19, 20].

The time series is a sequence of dependent observations. In 1999 Doukhan and Louhichi [12] and simultaneously Bickel and Bühlmann [6] introduced a new way of describing data—the weak dependence.

Below, the definition of weak dependence is introduced.

Let $(E, \|\cdot\|)$ be a normed space and $u \in \mathbb{N}^*$. We assume that a function $h : E^u \longrightarrow \mathbb{R}$ belongs to the class $\mathscr{L} = \{h : E^u \to \mathbb{R},\ \| h \|_\infty \leq 1, Lip(h) < \infty\}$, where $Lip(h) = \sup_{x \neq y} \frac{|h(x)-h(y)|}{\|x-y\|_1}$ and $\| x \|_1 = \sum_{i=1}^{u} \| x_i \|$.

Definition 5 ([4]) A sequence $\{X_t\}_{t\in\mathbb{Z}}$ of random variables taking values in $E = \mathbb{R}^d$ $(d \in \mathbb{N}^* = \mathbb{N} \setminus \{0\})$ is $(\varepsilon, \mathscr{L}, \Psi)$-weakly dependent if there exists $\Psi : \mathscr{L} \times \mathscr{L} \times \mathbb{N}^* \times \mathbb{N}^* \to \mathbb{R}$ and a sequence $\{\varepsilon_r\}_{r\in\mathbb{N}}$ $(\varepsilon_r \to 0)$ such that for any $(f, g) \in \mathscr{L} \times \mathscr{L}$, and $(u, v, r) \in \mathbb{N}^{*2} \times \mathbb{N}$

$$|Cov(f(X_{i_1}, ..., X_{i_u}), g(X_{j_1}, ..., X_{j_v}))| \leq \Psi(f, g, u, v)\varepsilon_r$$

whenever $i_1 < i_2 < ... < i_u \leq r + i_u \leq j_1 < j_2 < ... < j_v$.

The weak dependence notions are related to the initial time series and are measured in terms of covariance of the functions.

The asymptotic behavior of the covariance shows us the independence between "past" and "future". Intuitively, the weak dependence is "forgetting" in time series.

There are several concepts of weakly dependent processes e.g.:. λ-, θ-, η-weak dependence. The form of the Ψ function corresponds to particular cases of weak dependence.

The coefficient λ corresponds to:

$$\Psi(f, g, u, v) = uvLip(f)Lip(g) + uLip(f) + vLip(g),$$

the coefficient η corresponds to:

$$\Psi(f, g, u, v) = uLip(f) + vLip(g),$$

the coefficient θ corresponds to:

$$\Psi(f, g, u, v) = vLip(g).$$

Note there are other cases of weakly dependent, that are not quoted here.

In the definition of weak dependence we denote respectively λ_r, η_r or θ_r instead of ε_r.

The heavy-tailed random variables are variables with distributions whose extreme values are "more probable than normal". The heavy tail phenomena occur frequently in real life. In contradiction to Gaussian phenomena which do not allow for large fluctuations, "heavy tails" can be used to describe high variability. The data with "heavy tails" appear in such different fields as economics, telecommunications, meteorology, physics and signal processing.

If we define kurtosis as μ_4/σ^4 and μ_4 is the fourth central moment (if it exists), while σ is the standard deviation then we can say that heavy-tailed variables are those

with kurtosis greater than three, and whose tails go to zero slower than in the normal distribution.

One of the example of heavy tailed distributed random variables are GED—random variables.

The Generalized Error Distribution (GED) is a parametric model of a heavy tailed distribution. All moments of the GED are finite and the GED has a relatively simple form of a probability density function.

The Generalized Error Distribution is a symmetric unimodal member of the exponential family. The domain of the probability distribution function is $(-\infty, \infty)$.

The original concept of the GED was introduced by Subbotin in 1923 [34], so it is known as *"Subbotin's family of distributions"*. However, Subbotin proposed a two-parameters GED model:

$$f(x; h, m) = \frac{mh}{2\Gamma(1/m)} exp\{-h^m |x|^m\}, \tag{2}$$

where $x \in \mathbb{R}$ and $h > 0$ and $m \geq 1$ are scale and shape parameters, respectively.

In 1963 Lunetta ([30]) has defined a three-parameters GED class, as follows:

$$f(x; \mu, \sigma, \alpha) = \frac{1}{2\sigma\alpha^{1/\alpha}\Gamma(1 + 1/\alpha)} exp\{-\frac{1}{2}|\frac{x-\mu}{\sigma}|^\alpha\}, \tag{3}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\tau > 0$ is the scale and $\alpha > 0$ is the shape (power).

Of course, the m in the Eq. (2) is equal to the α in the Eq. (3) while $h = (\alpha^{1/\alpha}\tau)^{-1}$. Taking into account the fact that the Euler gamma function Γ satisfies the formula $r\Gamma(r) = \Gamma(r + 1)$, the Eqs. (2) and (3) are equivalent, whenever the location parameter μ in (3) is equal to zero.

The GED is also called the generalized normal class. The reason is that for the random variable X with density function as in the formula (3) we have the following equation:

$$\tau = \{E|X - \mu|^\alpha\}^{1/\alpha},$$

which for $\alpha = 2$, gives the standard deviation in the normal case. Note that if $\alpha \neq 2$, τ must not be confused with the standard deviation of X.

Below we give the definition in which τ stands for the standard deviation:

Definition 6 ([36]) The random variable X has GED distribution ($X \sim G(\mu, \tau, \alpha)$) if the density function, $f(x)$, of X is given by the equation:

$$f(x; \mu, \tau, \alpha) = (2\Gamma(1 + 1/\alpha)A(\tau, \alpha))^{-1} exp\{-|\frac{x-\mu}{A(\tau, \alpha)}|^\alpha\} \tag{4}$$

with $A(\tau, \alpha) = \tau\sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}$.

In the sequel of this dissertation it is enough to consider the case $\tau = 1$ in the Eq. (4). Therefore, we will be considering the density function

$$f(x;\mu,\alpha) = \frac{\alpha}{2A(\alpha)\Gamma(1/\alpha)} exp\{-|\frac{x-\mu}{A(\alpha)}|^{\alpha}\}, \tag{5}$$

where $A(\alpha) = \sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}$, $\alpha > 0$, $\mu \in (-\infty, \infty)$, and $x \in \mathbb{R}$.

Our definition of the GED is as follows:

Definition 7 The random variable X has a GED distribution $X \sim G(\mu, 1, \alpha)$ if the density function, $f(x)$, of X follows the Eq. (5).

The rth central moment of a random variable $X \sim G(\mu, 1, \alpha)$ can be calculated as

$$E(X - EX)^r = \frac{1}{\sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)}\Gamma(1 + 1/\alpha)} \int_{-\infty}^{\infty} (x - EX)^r e^{-\frac{1}{2}|x-EX|^{\alpha}} dx,$$

where $r \in \mathbb{N}$. When r is odd then the rth moments are equal to zero, by symmetry. For r even the rth moments are as follow:

$$EX^r = (\sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)})^r \frac{\Gamma(1/\alpha(r+1))}{\Gamma(1/\alpha)}.$$

Notice that in the consequence of the Definition 7 the first four moments of GED distribution are: mean $= \mu$, variance $= 1$, skewness $= 0$, kurtosis $= \frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{\Gamma^2(3/\alpha)}$.

Recall that the GED distribution has heavy tails, if $\alpha < 2$. When $\alpha > 2$ we get tails lighter than normal.

It is clear to see that the normal distribution is GED with $\alpha = 2$. Below are other examples of the GED distribution.

Example 1 If we choose $\alpha = 1$ in the Definition 6 then the GED distribution is so-called Double Exponential, or Laplace, distribution, i.e. $G(\mu, \tau^2, 1) = L(\mu, 4\tau^2)$.

Note that the Subbotin's model (2) does not allow for the tails heavier than those in the Laplace distribution. Unlike the formula (3), where the tails heavier than those in the Laplace distribution are allowed.

Example 2 If we consider the Definition 6 and $\alpha \to \infty$, then the GED distribution tends to the uniform distribution $U(\mu - \tau, \mu + \tau)$.

3 The Model and Its Properties

In many applications of time series analysis one is confronted separately with heavy tailed and long memory behavior. The non-stationarity of the time series, and its special case—the periodicity is also a feature that researchers are dealing with.

Below we present a model which will simultaneously be dealing with three features: periodicity, long memory and heavy tails. We build it by adjusting a long memory and a heavy tailed stationary model to the T-variate process.

Let the time series $\{X_t\}_{t\in\mathbb{Z}}$ be defined as:

$$X_t = \sigma_t GG_t + \eta_t, \tag{6}$$

where

A1 The volatility time series σ_t and the Gaussian-Gegenbauer time series GG_t are independent

A2 The sequence of random variables σ_t is i.i.d and its marginal distribution comes from a heavy tail family, with finite at least fourth moment, for example.

A3 GG_t is periodic Gaussian-Gegenbauer time series. We put that $GG_t = f_t \cdot G_t$, where G_t is Gaussian-Gegenbauer mean zero time series with $k = 1, |\nu| \leq 1$, $LM(\beta)$ with $\beta \in [0, 1)$. The function f_t is a periodic, deterministic, bounded with a known period T. The autocovariance of G_t is γ_G.

A4 The deterministic function η_t is periodic with the same period T as f_t.

$\beta = 2d$, where d is a memory parameter from the Definition 4.

Properties of the model

Lemma 1 *The process $\{X_t\}_{i\in\mathbb{Z}}$ defined by the Eq. (6) is a long memory process in the sense of Definition 3, with $\beta \in [0, 1)$.*

Proof

$$\sum_{0<|h|<n} \gamma(s)(h) = \sum_{0<|h|<n} (E\sigma)^2 f_h^2 \gamma_G(h) \sim C(s) n^{\beta}.$$

The last asymptotic equivalence follows from Theorem 2 [10]. □

Moreover:

Lemma 2 ([21], Theorem 2—[11]) *The long memory stationary Gaussian-Gegenbauer time series is not strong mixing.*

Lemma 3 *Assume* **A1** *through* **A4**. *Then X_t defined by the Eq. (6) is λ-weakly dependent.*

Proof It follows from Lemma 2 that the Gaussian-Gegenbauer time series $G_{t\in\mathbb{Z}}$, is not strong mixing.

We already know that Gegenbauer (in the sense of the Definition 4) time series has a long memory. And from the [4], p. 8 follows that the stationary Gaussian long memory time series has the λ-weak dependence properties.

Finally the λ-weak dependence of X_t is implied from the Proposition 1 in [27].

Lemma 4 *Assume* **A1** *through* **A4**. *Then the weak dependence coefficients of the model defined by the Eq. (6) satisfy the following relationship:*

$$\lambda_r = O(r^{\beta-1}), \quad \beta \in [0, 1).$$

Proof of above lemma follows from the Lemma 4 [9].

The Lemma 3 provide clear motivation to study weakly dependent structures.

GED volatility or other heavy tailed distribution with finite at least fourth moment

Assume that the volatility process σ_t in the model (6) comes from a GED distribution, as in Definition 7, i.e. $\sigma_t \sim G(\mu, 1, \alpha)$).

The following fact provides the information about the heaviness of X_t tails in the GED case:

Lemma 5 *Assume* **A1** *through* **A4**. *Then X_t defined by the Eq. (6) has a heavy tailed marginal distribution for $\alpha > 0$.*

We do not have information about the marginal distribution of X_t, but since all the moments of both distribution GG_t and σ_t exist the model X_t is $(2 + \delta)$-order, with $\delta > 0$.

In particular X_t is second-order, so it is periodically correlated in the sense of Gladyshev. The formalization of this statement is as follows:

Lemma 6 *X_t defined by the Eq. (6) with σ_t coming from the GED has a periodic mean, a periodic variance and a periodic autocovariance. Moreover, the autocovariance of X_t has a form:*

$$\gamma(t, h) = (Cov(\sigma_t, \sigma_{t+h}) + \varphi^2)|f_t||f_{t+h}|\gamma_G(h).$$

The estimator and its properties

For the model defined by the Eq. (6) one of the resampling method—subsampling—is considered to approximate an asymptotic distribution of the seasonal trend components, the overall mean and the vector of the seasonal trend components.

We start with the definition.

Definition 8 We define the estimator of the seasonal trend components $\eta(s)$ as follows:

$$\hat{\eta}_N(s) = \frac{1}{N}\sum_{p=0}^{N-1} X_{s+pT}, \quad s = 1, 2, \ldots, T, \tag{7}$$

where T is the known period.

Definition 9 We define the estimator of the overall mean $\bar{\eta} = \frac{1}{T}\sum_{s=1}^{T}\eta(s)$ as follows:

$$\hat{\bar{\eta}} = \frac{1}{T}\sum_{s=1}^{T}\hat{\eta}_N(s) \tag{8}$$

where T is the known period.

4 Central Limit Theorems in the GED Case

In this subsection, the central limit theorems for the seasonal and overall mean for the variables from the GED are introduced.

Let us define $\Sigma_s = Var(X_s)$. In the GED case $\Sigma_s = |f_s|^2\gamma_G(h), \;\; s = 1, ..., T$.

Let us define:

$$B_N(s) = N^{1/2}(\hat{\eta}_N(s) - \eta(s)).$$

For a sequence $\{m_N\}_{N\in\mathbb{N}}$ such that $m_N \to \infty$, *if* $N \to \infty$ let us define $\{k_N\}_{N\in\mathbb{N}}$ such that

$$k_N = [\frac{N}{m_N}] \to \infty, \; \textit{if } N \to \infty$$

and for λ-weakly dependent model $\{X_t\}_{t\in\mathbb{Z}}$ defined by the Eq. (6) following condition holds

$$\lambda_{m_N} k_N^{\frac{3}{2}} \to 0, \qquad N \to \infty.$$

Let us consider a subsample $(X_{m_N}, \ldots, X_{k_N m_N})$ of $(X_s, \ldots, X_{s+(N-1)T})$.

Theorem 2 (Central Limit Theorem—for the seasonal means) *Assume* **A1** *through* **A4** *and the volatility process* σ_t *is as in the Definition 7. Then, for a sequence* $\{m_N\}_{N\in\mathbb{N}}$ *such that* $m_N \to \infty$ *and* $k_N = [\frac{N}{m_N}] \to \infty$, *if* $N \to \infty$ *and for each* $s = 1, \ldots, T$ *following convergence holds:*

$$B_{k_N}(s) \to \mathcal{N}(0, \Sigma_s), \qquad N \to \infty.$$

The proof follows from Proposition 4.1, [4].

Since the model defined by the Eq. (6) satisfies property $LM(\beta)$ (Definition 2), for $\beta \in [0, 1)$ the time series X_t defined by the Eq. (6) does not satisfy any known Central Limit Theorem. The reason for this fact is that it has the long memory property. But the Central Limit Theorem will be satisfied if we choose a subsample of the observation with the appropriate asymptotic step of sampling. In the model (6) the subsampled time series $\{Y_{sm_N} = X_{sm_N} - \eta_{sm_N}\}$, with a "subsampling" step m_N such that

$$o(m_N) = N^{3/(2\beta+1)}, \tag{9}$$

satisfies the Central Limit Theorem with a convergence rate $o(N^{(1-\beta)/(4\beta+2)})$.

The Eq. (9) gives us the restriction for β it means: $\beta \in (0, 1/2)$.

Following [4] there are objections to this "subsampling" method: only a part of the sample is used and the choice of the convergence rate of the "subsampling" implies the knowledge of the convergence rate of λ_r. But the convergence rate of λ_r, in long memory processes is connected with the the long memory parameter β. It could give us a step of "subsampling".

Let us define:

$$B_{k_N,N} = k_N^{1/2}(\hat{\eta} - \eta).$$

Theorem 3 (Central Limit Theorem—for the vector of the means) *Assume* **A1** *through* **A4**. *Moreover, assume that*

$$\lambda_{m_N} k_N^{\frac{3}{2}} \to 0, \qquad N \to \infty.$$

Let $Z_N = (Y_{s+pT}, ..., Y_{s+(p+1)T})$, $p = 0, ..., N-1$ *be a sequence of zero mean random variables with values in* $\mathbb{R}^T$, *where* T *is the period. Then, for a sequence* $\{m_N\}_{N\in\mathbb{N}}$ *such that* $m_N \to \infty$ *and* $k_N = [\frac{N}{m_N}] \to \infty$, *if* $N \to \infty$

$$B_{k_N,N} = \frac{1}{\sqrt{k_N}} \sum_{i=1}^{k_N} Z_{im_N} \to \mathcal{N}_T(0, Cov(X_0)), \qquad N \to \infty.$$

The proof of Theorem 3 implies from the Proposition 4.1, [4], but applied to the vectors.

Theorem 4 (Central Limit Theorem—for the overall mean) *Assume* **A1** *through* **A4**. *Moreover assume that*

$$\lambda_{m_N} k_N^{\frac{3}{2}} \to 0, \qquad N \to \infty.$$

Then, for a sequence $\{m_N\}_{N\in\mathbb{N}}$ *such that* $m_N \to \infty$ *and* $k_N = [\frac{N}{m_N}] \to \infty$, *if* $N \to \infty$

$$B = \frac{1}{\sqrt{k_N T}} \sum_{s=1}^{T} \Big(\sum_{p=0}^{k_N-1} (X_{sm_N+pT} - \eta(s)) \Big) \to \mathcal{N}(0, \Sigma), \qquad N \to \infty.$$

The proof of the Theorem 4 implies from the Proposition 4.1, [4] for the vectors and the Cramer–Wald theorem.

5 Consistency of the Subsampling Method for the Mean

From research of Hall and Lahiri [22, 28] we know that, for long-range dependent process, the bootstrap (MBB, CBB, PBB, SBB, GSBB) do not work, whereas subsampling still works asymptotically.

All we need to know to use subsampling is if there exists a non-degenerated asymptotic distribution of the statistic (we do not have to know the form of the asymptotic distribution)—we need to have the Central Limit Theorem.

Recall that our sample size n is equal $n = NT$.

The idea of subsampling in our model for the seasonal components is as follows:

1. For each $s = 1, \dots, T$ the estimator $\hat{\eta}_N(s)$ is recomputed from the $(X_s, \dots, X_{s+(N-1)T})$ over "short" overlapping blocks of length b_s (b_s depends on N-the length of the sample)
2. From Step 1 $N - b_s + 1$ statistics are obtained for each s. In our context those will be $a_{b_s}(\hat{\eta}_{N,b_s,i}(s) - \hat{\eta}_N(s))$ where $\hat{\eta}_{N,b_s,i}(s)$ is subsampling version of the estimator $\hat{\eta}_N(s)$ and a_{b_s} is the normalize sequence.
3. then the empirical distributions:

$$L_{N,b_s}(x,s) = \frac{1}{N-b_s+1} \sum_{i=1}^{N-b_s+1} 1_{\{a_{b_s}(\hat{\eta}_{N,b_s,i}(s)-\hat{\eta}_N(s))\le x\}}$$

are used to approximate the asymptotic distribution $L(s)(x)$ of the estimator $a_N(\hat{\eta}_N(s) - \eta(s))$.

The idea of subsampling in our model for the vector of seasonal components is as follows:

1. For each $s = 1, \dots, T$ the estimator $\hat{\eta_N}$ is recomputed from the $(X^1, \dots, X^N)$, where $X^i = (X_{1+(i-1)T}, ..., X_{iT})$ over "short" overlapping blocks of length b (b depends on N-the length of the sample)
2. From Step 1 $N - b + 1$ statistics are obtained. In our context those will be $a_b(\hat{\eta}_{N,b} - \hat{\eta}_N)$ where $\hat{\eta}_{N,b}$ is subsampling version of the estimator $\hat{\eta}_N$ and a_b is the normalizing sequence.
3. then the empirical distributions: $L_{N,b}(x) = \frac{1}{N-b+1} \sum_{i=1}^{N-b+1} 1_{\{a_b(\hat{\eta}_{N,b}-\hat{\eta}_N)\le x\}}$ are used to approximate the asymptotic distribution $L(x)$ of the estimator $a_N(\hat{\eta}_N - \eta_N)$.

The main problem with the subsampling procedure is its consistency. We need to prove that the finite sample quantiles generated by the subsampling procedure converge asymptotically ($N \to \infty$) to the quantiles of the asymptotic distribution.

Now we consider the problem of consistency of the Subsampling.

To prove the consistency of the subsampling procedure we need to know if there exists a non-degenerated asymptotic distribution of the statistic which means that we need to have the central limit theorem.

In both the stable and the GED case we have obtained a weakly convergence to the limit random variables.

Denote the cumulative distribution functions of this limit random variables by $L(s)$ and L for the seasonal means and vector of the seasonal means, respectively.

To fulfill the Step 3 of the subsampling procedure the empirical distribution functions $L_N(s)(x) = P(P_N(s) \leq x)$ are computed from the subsamples $(X_s, ..., X_{s+(N-1)T})$, for each $s = 1, \ldots, T$ where X_t is defined by the Eq. (6).

The conclusion from the central limit theorems is that the empirical distribution functions converge weakly to the cumulative distribution functions of the limit random variables

$$L_N(s)(x) \to L(s)(x) \qquad if \qquad N \to \infty, \qquad s = 1, \ldots, T.$$

Denote the density of the limit distribution by $L'(s)$. It is obvious that in the GED case $\| L'(s) \|_\infty < \infty$.

Theorem 5 (Consistency theorem for seasonal means in GED cases) *Assume that CLT holds and assume* **A1** *through* **A4**, *then consistency of the subsampling method holds for the "subsampling" step (9):*

1. *If x is the point of the continuity of $L(s)$, then $L_{k_N,b_s}(s)(x) \xrightarrow{P} L(s)(x)$.*
2. *If L is continuous then $\sup_x |L_{k_N,b_s}(s)(x) - L(s)(x)| \xrightarrow{P} 0$.*
3. *If $L(s)$ is continuous in $c(1-q)$ (where $c(1-q)$ is a q-quantile) then if $k_N \to \infty$*

$$P[k_N^{-1/2}(\hat{\eta}_{k_N}(s) - \eta(s)) \leq c_{k_N,b}(1-q)] \to 1 - q$$

in GED case.

Where $\alpha \in (0,1)$ and

$$c_{k_N,b}(1-\alpha) = inf\{x : L_{k_N,b}(s)(x) \geq 1-\alpha\},$$

$$c(1-\alpha) = inf\{x : L(s)(x) \geq 1-\alpha\}.$$

The $\xrightarrow{P}$ denotes convergence in probability.

Theorem 6 (Consistency theorem for vector of the seasonal means in GED cases) *Assume that CLT holds and assume* **A1** *through* **A4**, *then consistency of the subsampling method holds for the subsequence (9):*

1. *If x is the point of the continuity of L, then $L_{k_N,b}(x) \xrightarrow{p} L(x)$.*
2. *If L is continuous then $\sup_x |L_{k_N,b}(x) - L(x)| \xrightarrow{p} 0$.*

6 Conclusions

In the paper the model of the periodically correlated, long memory and heavy tailed time series was described. Estimator of the vector sample means of this particular model was introduced and Central Limit Theorem was proved. Furthermore, consistency of chosen method of estimation, the subsampling, was investigated. The question to be answered in further research is the extension of the model (6) to the class almost periodically correlated processes.

Appendix

In the appendix the proofs of some results are presented.

Proof Lemma 5
From direct calculations we can obtain the strict formula for the kurtosis of the model X_t, which is:

$$\frac{E(X_t - EX_t)^4}{(E(X_t - EX_t)^2)^2} = \frac{E\sigma_t^4 EG_t^4}{(E\sigma_t^2)^2 (EG_t^2)^2} = 3\frac{\Gamma(5/\alpha)\Gamma(1/\alpha)}{\Gamma^2(3/\alpha)}.$$

If we use the Stirling's formula for Γ function we will obtain the approximation as follows:

$$kurtosis \approx 3 \cdot 1.4 \cdot 4.3^{1/\alpha}.$$

The kurtosis is more than 3 for all $\alpha > 0$. □

Proof Lemma 6
The mean of X_t is η_t, so it is periodic. The variance is periodic:

$$\gamma(t+T,0) = (f_{t+T})^2(1+\varphi^2)\gamma_G(t+T,0) =$$

$$= (f_{t+T})^2(1+\varphi^2)\gamma_G(0) = (f_t)^2(1+\varphi^2)\gamma_G(t,0) = \gamma(t,0).$$

The autocovariance also:

$$\gamma(t+T,h) = |f_{t+T}f_{t+T+h}|\varphi^2\gamma_G(t+T,h) =$$

$$= |f_{t+T}f_{t+T+h}|\varphi^2\gamma_G(h) = |f_t f_{t+h}|\varphi^2\gamma_G(t,h) = \gamma(t,h).$$

The form of the variance and autocovariance follows from the form of the variance of variable with the GED distribution. □

Proof Theorem 5
Let us consider a sequence of statistics $B_{k_N}(s)$, for fixed $s = 1, 2, ..., T$ and k_N as in the Theorem 2.
$L_{k_N}(s)(x) = P(B_{k_N}(s) \leq x)$ is cumulative distribution function of $B_{k_N}(s)$.

From the assumptions

$$sup_{x\in\mathbb{R}}|L_{k_N}(s)(x) - L(s)(x)| \longrightarrow 0, \quad k_N \to \infty$$

For overlapping samples the number of subsamples:

$Y_{b,q}(s) = (X_{s+qT}, X_{s+(q+1)T}..., X_{s+(q+b-1)T})$, $q = 0, 1, ..., k_N - b$ and the number of subsampling statistics:

$B_{k_N,b,q}(s) = \sqrt{b}(\hat{\eta}_{k_N,b,q}(s) - \hat{\eta}_{k_N}(s))$ is $k_N - b + 1$.

Above statistics are used to approximate the distributions $L_{k_N}(s)(x)$ by empirical distribution functions: $L_{k_N,b,q}(s)(x) = \frac{1}{k_N-b+1}\sum_{q=0}^{k_N-b} \mathbb{I}_{\{B_{k_N,b,q}(s)\leq x\}}$.

Let us define subsampled distribution:

$$U_{k_N,b,q}(s)(x) = \frac{1}{k_N - b + 1}\sum_{q=0}^{k_N-b} \varphi(\sqrt{b}(\hat{\eta}_{k_N,b,q}(s) - \eta_{k_N}(s))\varepsilon_n).$$

The sequence ε_n is decreasing to zero and φ is the non-increasing continuous function such that $\varphi = 1$ or 0 according to $x \leq 0$ or $x \geq 1$ and which is affine between 0 and 1.

From [32] it is known that

$$\forall x \in \mathbb{R}\ |L_{k_N,b,q}(s)(x) - U_{k_N,b,q}(s)(x)| \xrightarrow{p} 0.$$

It follows that it is enough to investigate only the variance of $U_{k_N,b,q}(s)$, $s = 1, ..., T$

It is enough to investigate the variance of $U_{k_N,b_s,p}(s)$, $s = 1, ..., T$ (Theorem 3.2.1 ([32])).

$$VarU_{k_N,b_s,p}(s)(x) = (k_N - b_s + 1)^{-2}(\sum_{|h|<k_N-b_s+1}(k_N - b_s + 1 - |h|)\gamma(h))$$

here $\gamma(h) = Cov(\varphi(B_{k_N,b,p}(s)/\varepsilon_n), \varphi(B_{k_N,b,p+h}(s)/\varepsilon_n))$.

From the assumption that we have λ-weak dependence and under condition **A1** through **A4** and the Lemma 3.1 in ([2])

$$Cov(\varphi(B_{k_N,b,p}(s)/\varepsilon_n), \varphi(B_{k_N,b,p+h}(s)/\varepsilon_n)) \leq \sqrt{b_s}\lambda_{h-b+1}/\varepsilon_n$$

It implies that $Var(U_{k_N,b_s,p}(s)(x))$ tends to zero, it proves point 1. of the Theorem 5.

The proof of point 2. and 3. if the 1. holds and under assumption of the model (6) is the same as the proof of 3. in the Theorem 3.2.1 ([32]). □

Proof Theorem 6
For any vector of constants $c \in \mathbb{R}^T$ we have the equation for the subsampling version of the characteristic functions of the distributions:

$$\phi^*_{B_{k_N,b,q}}(c) = \phi^*_{c^T B_{k_N,b,q}}(1) \text{ in GED case}$$

Let $Z_{s+pT} = c_s X_{s+pT}$, where $p = 0, \ldots, k_N - 1$ and $s = 1, \ldots, T$. The series $\{Z_t\}$ fulfills the assumptions of Theorem 6, which means that subsampling is consistent for the mean $(\eta_N)_Z$. By Theorem A in Athreya [3] we have:
in GED case

$$\phi^*_{c^T B_{k_N,b,q}}(1) \xrightarrow{p} \phi_{k_N(\eta, c^T \Sigma c)}(1) = \phi_{k_N(\eta,\Sigma)}(c).$$

Moreover

$$P^*(B_{k_N,b,q} \leq x) \xrightarrow{p} F_{k_N(\eta,\Sigma)}(x),$$

for any $x \in \mathbb{R}^T$, where $F_{k_N(\eta,\Sigma)}(x)$ is the cumulative distribution function of $k_N(\eta, \Sigma)$. The second point of the thesis of the Theorem 6 follows then from Pólya's theorem ([33], p. 447). □

References

1. Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability*, *21*, 930–934.
2. Ango-Nze, P., Dupoiron, S., & Rios, R. (2003). Subsampling under weak dependence condition, Technical Report DT2003-42. CREST
3. Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *The Annals of Statistics*, *15*, 724–731.
4. Bardet, J.-M., Doukhan, P., Lang, G., & Ragache, N. (2008). Dependent Lindeberg central limit theorem and some applications. *ESAIM: Probability and Statistics*, *12*, 154–172
5. Beran, J., Feng, T., Ghosh, S., & Kulik, R. (2013). *Long-memory processes. Probabilistic properties and statistical methods*. New York: Springer.
6. Bickel, P., & Bühlmann, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli*, *5–3*, 413–446.
7. Brockwell, P., & Davis, R. (1991). *Time series: Theory and methods*. New York: Springer.
8. Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S., & Prieur, C. (2008). *Weak dependence: With examples and applications*. Lecture notes 190 in statistics. Springer
9. Doukhan, P., & Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, *84*, 313–342
10. Doukhan, P., Prohl, S., & Robert, C. Y. (2011). Subsampling weakly dependent times series and application to extremes. *TEST*, *20*(3), 487–490.
11. Doukhan, P. (1994). *Mixing: Properties and examples*. Lecture notes in statistics 85. Springer
12. Doukhan, P., & Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, *84*, 313–342.

13. Dudek, A. E., Leśkow, J., Politis, D., & Paparoditis, E. (2014). A generalized block bootstrap for seasonal time series. *Journal of Time Series Analysis*, *35*, 89–114.
14. Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*, 1–26.
15. Ferrara, L., & Guégan, D. (2001). Forecasting with k-factor Gegenbauer processes: theory and applications. *Journal of Forecasting*, *20*(8), 581–601.
16. Gajecka-Mirek, E. (2014). Subsampling for weakly dependent and periodically correlated sequences. In *Cyclostationarity: Theory and methods*. Lecture notes in mechanical engineering. Springer.
17. Gardner, W., Napolitano, A., & Paura, L. (2006). Cyclostationarity: Half a century of research. *Signal Processing*, *86*
18. Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Mathematics*, *2*
19. Gray, H. L., Zhang, N.-F., & Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, *10*, 233–257.
20. Gray, H. L., Zhang, N.-F., & Woodward, W. A. (1994). On generalized fractional processesa correction. *Journal of Time Series Analysis*, *15*(5), 561–562.
21. Guégan, D., Ladoucette S.(2001). Non-mixing properties of long memory sequences. *Comptes Rendus de l' Academie des Sciences Paris*, *333*, 373–376
22. Hall, P., Jing, B.-Y., & Lahiri, S. (1998). On the sampling window method for long-range dependent data. *Statistica Sinica*, *8*, 1189–1204.
23. Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, *68*, 165–176.
24. Hui, Y. V., & Li, W. K. (1995). On Fractionally differenced periodic processes. *The Indian Journal of Statistics*, *57*, Series B, Pt. 1, 19–31.
25. Hurd, H. L., & Miamee, A. G. (2007). *Periodically correlated random sequences: Spectral. Theory and practice*. Wiley
26. Hurd, H., & Leśkow, J. (1992). Strongly consistent and asymptotically normal estimation of the covariance for almost periodically correlated processes. *Statistics and Decisions*, *10*, 201–225.
27. Jach, A., McElroy, T., & Politis, D. N. (2012). Subsampling inference for the mean of heavy-tailed long memory time series. *Journal of Time Series Analysis*, *33*(1), 96–111.
28. Lahiri, S. (1993). On the moving block bootstrap under long-range dependence. *Statistics and Probability Letters*, *18*, 405–413.
29. Leśkow, J., & Synowiecki, R. (2010). On bootstrapping periodic random arrays with increasing period. *Metrika*, *71*, 253–279.
30. Lunetta, G. (1963). Di una generalizzazione dello schema della curva normale. *Annali della Facoltá di Economia e Commercio di Palermo*, *17*, 237–244.
31. Philippe, A., Surgailis, D., & Viano M. C. (2006). *Almost periodically correlated processes with long-memory*. Lecture notes in statistics 187. Springer
32. Politis, D. N., Romano, J. P., & Wolf, M. (1999). *Subsampling*. New York: Springer.
33. Shao, J., & Tu, D. (1995). *The jacknife and bootstrap*. New York: Springer.
34. Subbotin, M. T. (1923). On the law of frequency of error. *Matematicheskii Sbornik*, *31*(2), 296–301.
35. Synowiecki, R. (2007). Consistency and application of moving block bootstrap for non-stationary time series with periodic and almost periodic structure. *Bernoulli*, *13*, 1151–1178.
36. Varanasi, M. K., & Aazhang, B. (1989). Parametric generalized gaussian density estimations. *Journal of the Acoustical Society of America*, *86*(4), 1404–1415.

Change-Point Problem in the Fraction-Of-Time Approach

Jacek Leśkow and Bartosz Stawiarski

Abstract We examine the applicationally vital change-point problem using the Fraction-of-Time (FOT) approach. Specifically, in our extensive simulations, two techniques are proposed in order to estimate an unknown (possibly multiple) time τ, at which the dynamics of a given process undergoes an abrupt, structural change. Both the 3-D empirical FOT distribution profile calculated upon a moving time window, and differenced quantile process detect the change-points satisfactorily well in case of simulated as well as real time series. The paper lays the foundations for tackling the problem numerically in the nonstochastic framework, whereas some future research concerning theory helpful in optimizing the proposed techniques calls for separate development.

Keywords Fraction-Of-Time · Change-point detection · Structural breaks · Empirical distribution

1 Introduction

Quite frequently real-life data stemming from mechanical, environmental, financial, medical sciences, exhibit some peculiar nonstationary types of behavior. One of such vital nonstationarity patterns is a sudden change in the recorded quantity dynamics at an unknown moment τ after which the signal (treated as a stochastic process) is

The research is supported by scientific grant No. *UMO-2013/10/M/ST1/00096*, National Centre of Science, Poland.

J. Leśkow · B. Stawiarski (✉)
Institute of Mathematics, Cracow University of Technology, Kraków, Poland
e-mail: bstawiarski@pk.edu.pl

J. Leśkow
e-mail: jleskow@pk.edu.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_3

driven by another regime. The nature of these dynamics shifts can vary, e.g. change in mean (location), regression slope, structural breaks in variance, sudden change of coefficients in the underlying model—to name just a few. Change-point analysis in the stochastic framework has been an object of research for several decades in a variety of contexts. Throughout this time-span vast scope of models featuring such change-point(s) has been proposed, allowing for either sudden change or so-called smooth transition. Beside proposing appropriate models, multitude of detection techniques and statistical tests has been developed.

Stochastic approach is however invariably associated with underlying distributions of processes in question, which can be highly complicated. Adding dependency within the data and accompanying measures of dependence (mixing, weak dependence) the task becomes extremely challenging. Moreover, in practice, we have just one past record of a given phenomenon. These adversities motivated establishing alternative insight into the data at hand, namely from a nonstochastic perspective. In early 2000s, Leśkow and Napolitano proposed a so-called Fraction-Of-Time (FOT) approach [2], allowing to infer from single signal record without resorting to stochastic calculus. In following years further results have been derived, including versions of limit theorems.

The paper is organized as follows. In Sect. 2 we first recall basic concept of FOT and elementary definitions used in the sequel, then propose two techniques employed for change-point detection in this set-up. Section 3 contains description of data sets (three simulated signals and one real financial time series) and immediately afterwards extensive simulation study concerning the change-point detection is provided. Final conclusions are drawn in Sect. 4, which is followed by References.

2 Change-Point Problem in the FOT Approach

2.1 The Nonstochastic Idea of FOT

First, recall basic concepts of the Fraction-Of-Time approach. Given a real function $x(u)$, defined on an interval $[t, t+T]$ with a measure μ (here Lebesgue), we study the fraction of time during which the function does not exceed a certain barrier ξ over the given interval, namely

$$F_{T,x}(t,\xi) = \frac{\mu(u \in [t, t+T] : x(u) \leq \xi)}{\mu(u \in [t, t+T])} = \frac{1}{T}\int_t^{t+T} 1_{\{x(u)\leq\xi\}} du \tag{1}$$

The above time-averaging is locally associated with the time location t therefore further theoretical research of (1) focused on asymptotics. In the early papers, the behavior of F_T with T growing to ∞ had been of crucial interest:

$$\lim_{T\to\infty} F_{T,x}(t,\xi) = \lim_{T\to\infty} \frac{1}{T}\int_{t}^{t+T} 1_{\{x(u)\le\xi\}}\,du \tag{2}$$

It was shown by Leśkow and Napolitano [2] that the limit in (2) exists for a wide class of functions, called relatively measurable (RM), and in that case it is independent of t. For further details and theory concerning e.g. RM sets, FOT versions of central limit theorems and constructing asymptotic confidence intervals for the functions in question see [2], [3].

In empirical practice, the quantity in (1) is estimated from a discrete sample of the function x, represented as an input signal $\{x(u_i) : i = 1, \ldots, N\}$ for $t \le u_i \le t + T$. Then, quite intuitively, like in the stochastic approach, we can define an empirical FOT distribution as

$$\hat{F}_{N,x}(\xi) = \frac{card\{i \in \{1, 2, \ldots, N\} : x(u_i) \le \xi\}}{N} \tag{3}$$

Let us state here that for a RM function $x(u)$ the empirical distribution consistently estimates the limiting $F(\xi)$. Henceforward, we will simplify the notation, defining just $u_i = i$ for $1 \le i \le N$ so that the signal x is recorded N times in equidistant, prespecified time units (daily, hourly, etc.).

2.2 Change-Point Detection Methodology

Let us now focus on the change-point detection problem within the signal x. Therefore we are rather interested in local behavior of F_N given by (3) under fixed N. We need to have more counterparts of F_N to carry out detection procedures. These quasi-replications can be obtained by estimating sample FOT distributions over a moving window $\{t, t + 1, \ldots, t + T - 1\}$, for the window size T considerably smaller than N, where $1 \le t \le N - T + 1$:

$$\hat{F}_{T,N,x}(t,\xi) = \frac{card\{i \in \{t, t + 1, \ldots, t + T - 1\} : x(i) \le \xi\}}{T} \tag{4}$$

which yields $N - T + 1$ empirical distributions based on the subsamples. Notice that now T denotes the length of the subsample extracted from the whole N-element signal at hand.

For further convenience, denote the empirical quantity in (4) by $F_{T,N}$. One can expect that it will exhibit systemic downward or upside movement whenever the moving window encompasses a change-point τ, provided the signal values change substantially from then on. In the estimation process the choice of an individual ξ is vital for the evolution of $F_{T,N}$ with t. For ξ chosen too large, the distribution would be near 1, whereas for ξ too small, the distribution would reach near-zero values. In both cases results do not have any explanatory value. Therefore to handle this deficiency, we admit ξ to change within a prespecified interval $[\xi_L, \xi_U]$ which enables us to

obtain a three-dimensional profile of $\hat{F}_{T,N,x}(t,\xi)$. In the simulations below, lower and upper quartile, Q_1, Q_3, of the whole signal x were proposed. In many situations, the change-point moment τ can be accurately estimated in two ways: either in advance or post-hoc with respect to t. In case of lasting, structural upward change within the signal the profile starts to decline with increasing t on the substantial portion of the admissible scope $[\xi_L, \xi_U]$ whenever the moving window $\{t, t+1, \ldots, t+T-1\}$ in (4) starts to contain τ. This occurs around T moments in advance. The descent continues until τ drops out of the moving window. The moment the gradient alongside the time axis flattens out can be defined as an estimator of τ. In case of a downward regime-shift symmetric behavior can be observed: ascending profile in advance (T time units prior to τ) followed by its plateau-type stabilization post-hoc. The technique is suitable for detecting multiple change-points, provided the width T exceeds the distance between successive change-point moments. Less pronounced changes translate into rather fuzzy evolutions of the 3-D profile and have to be examined more carefully otherwise.

Alternative and more rigid change-point detection procedure can be carried out using an empirical quantile processes. For $1 \le t \le N-T+1$ define a quantile process of order $q \in (0,1)$

$$\phi_q(t) = Y^{(t)}_{[qT]:T} \tag{5}$$

where $Y^{(t)} = \{x(t), \ldots, x(t+T-1)\}$ is the moving T-subsample of the original signal x, $Y^{(t)}_{[qT]:T}$ denotes the q-th order sample quantile of $Y^{(t)}$ and $[w]$ stands for the integer part of w. The quantile process itself somehow mimics the behavior of the signal, but only to an extent allowed by robustness of quantile estimators. Insensitivity to outliers is helpful here as it prevents the procedure from false labeling these outliers as change-points. Using (5) we construct a differenced quantile process: $\nabla\phi_q(t) = \phi_q(t) - \phi_q(t-1)$. This process will enable more precise inference about τ, namely sudden and sharp spikes of $\nabla\phi_q(t)$ can serve as a preliminary estimator of the change-point moment. Due to the quantile process properties, this preliminary estimator has to be corrected adequately to the quantile order q, providing the appropriate estimator of τ. Indeed, $\phi_q(t)$ reacts to the change-point at an earlier moment τ^*, roughly $[(1-q)T]$ time units prior to τ. Therefore, both $\phi_q(t)$ and $\nabla\phi_q(t)$ behave like leading indicators for the change-point moment, which will be shown later in simulations. Below, with no detriment to the general case, we will focus on medians, setting $q = 0.5$.

Several words of comment need to be added about the choice of T. It should not be too small because in that case $Y^{(t)}_{[qT]:T}$ would be unduly volatile and chaotic (due to small subsample size), which could lead to numerous erroneous signals. It should not be too large, either, to avoid the *oversmoothing* risk that might cause blindness to real change-points. Therefore, some kind of data-driven thumb rule should be developed in further research, eg. $T \sim N^r$ for some $r \in (0,1)$. *Data-driven* means it would also be desirable to incorporate the nonstationarity type of the data (e.g.: multiple level-shifts, polynomial spline trends, cyclostationarity or structural breaks in volatility) into the selection procedure for T.

3 Simulation Study

3.1 Data Sets Description

Now we proceed to the simulational part, in which we will employ the introduced change-point detection techniques based on sample FOT's. First, let us describe the signals (data sets) subject to our change-point analysis. We will consider three different, simulated data-generating processes: one exhibiting single change-point, another with multiple change-points and the third one with no change-point but with transient regimes modeled as short-lasting heavy-tailed clusters. In these three cases throughout the simulations we set $N = 1000$. The last signal will be a real financial time series, describing daily quotes of a tradable asset measuring volatility of S&P500 stock index, consisting of 1006 points.

Model 1
A shift-in-mean model noised with a lower-order ARMA sequence

$$X_t = \begin{cases} 25 + \varepsilon_t, & 1 \le t < \tau \\ 40 + \varepsilon_t, & \tau \le t \le N \end{cases} \tag{6}$$

where ε_t follows a zero-mean ARMA(1,1) model $\varepsilon_t = \alpha\varepsilon_{t-1} + \beta\eta_{t-1} + \eta_t$ with $\alpha = 0.2, \beta = 0.3$ and *iid* gaussian noise $\eta_t \sim N(0, 5^2)$. The single change-point fixed at $\tau = 500$.

Model 2
Second model we consider is a multiple change-point model at three moments $\tau_1 < \tau_2 < \tau_3$ with static ARMA(1, 1) noise ε_t as in Model 1.

$$X_t = \begin{cases} 50 + \varepsilon_t, & 1 \le t < \tau_1 \\ 35 + \varepsilon_t, & \tau_1 \le t < \tau_2 \\ 25 + \varepsilon_t, & \tau_2 \le t < \tau_3 \\ 45 + \varepsilon_t, & \tau_3 \le t \le N \end{cases} \tag{7}$$

Here the location change-point moments are $\tau_1 = 200, \tau_2 = 500, \tau_3 = 750$.

Model 3
Another time series we will study has no lasting change-point but instead exhibits two transient regimes modelled by heavy-tailed clusters. The first one follows t_3 distribution with quadrupled variance, and encompasses 10 time units, starting at 250. The second cluster follows Laplace distribution (symmetrized exponential with $\lambda = 1$) with time-span 20, starting at 600. Specifically,

$$X_t = \begin{cases} iid & N(0,4^2), & 1 \le t < 250 \\ iid & 2 * t_3, & 250 \le t < 259 \\ iid & N(0,4^2), & 260 \le t < 599 \\ iid & Laplace, & 600 \le t < 619 \\ iid & N(0,4^2), & 620 \le t \le N \end{cases} \tag{8}$$

where ε_t is again ARMA(1, 1) noise as in Model 1.

Real Data Set
The real signal we deal with contains daily VIX index quotes. This is a synthetic asset measuring the annualized volatility of the S&P500 stock index, calculated from implied volatilities of several call and put European options on S&P500, expiring within about one month. For further details see [1]. The considered series consists of $N = 1006$ entries, starting on 1st Sept 2006 through 31st Aug 2010, encompassing the 2008 market crash.

3.2 Change-Point Detection—Simulations Results

In all four cases we chose $T = 100$ which equals roughly $N^{2/3}$ (optimized, data-driven choice of T as a function of N is not addressed here in detail, cf. brief discussion in Sect. 2.2). We discuss the results separately for successive data sets and present accompanying figures.

Model 1
Figure 1 shows a simulated trajectory of the model in question. Structural change-in-mean moment is clearly visible half way through. The estimated quartile bounds for ξ range are $\xi_L = Q_1 = 20.21, \xi_U = Q_3 = 40.38$.

The 3-dimensional empirical FOT profile is presented in Fig. 2 togehter with the true τ. Substantial trend emerges T moments prior to τ, i.e. around 400, which is followed by a plateau quite uniformly over the scope of ξ, starting from ca. 500. This provides an accurate change-point detection as by that moment the effect of regime-change has been fully incorporated into $F_{N,T}$. We also present the bird's eye view of this profile to get a better insight into gradients behavior alongside time axis, see Fig. 3. Rapid slopes suggesting the change-point occurence can be distinctly spotted.

Turning now to the FOT-quantile-based detection technique, Fig. 4 shows the empirical median process and Fig. 5 presents the differenced median process. We denote the moment when the huge peak occurs as $\hat{\tau}^*$ and hence the change-point estimator is $\hat{\tau} = \hat{\tau}^* + [T/2] + 1$, cf. discussion at the end of Sect. 2. Adding a single time-unit accounts for differencing, so in this case finally $\hat{\tau} = 504$. We missed the true change-point just by 4 time units which corresponds to 2% error.

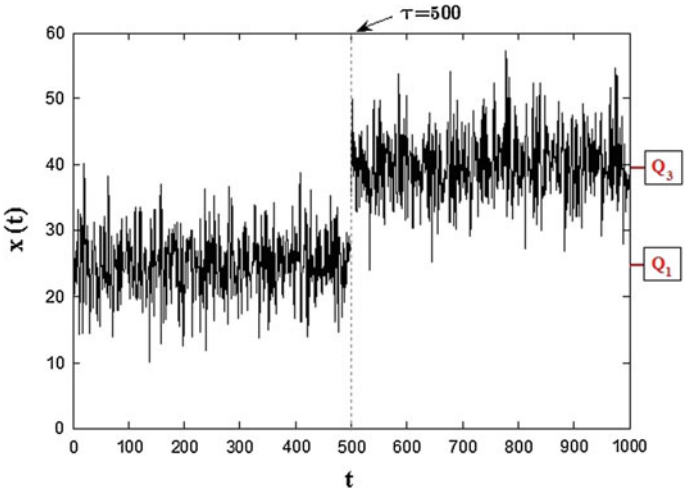


Fig. 1 Model 1 trajectory with one change-point

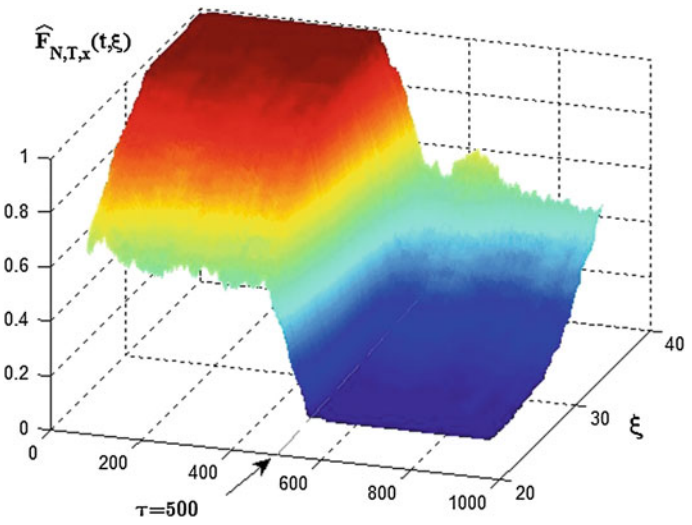


Fig. 2 3-D FOT empirical distribution profile for Model 1

Model 2

The trajectory of the model with multiple change-points is presented in Fig. 6, where distinct level-shifts are evident. Here sample quartiles are $\xi_L = Q_1 = 29.81, \xi_U = Q_3 = 46.73$.

The magnitude of respective change-points was deliberately made variable so as to show the sensitivity of our detection techniques with respect to the intensity of level jumps. With the passage of time t plateaus and valleys alternately appear on

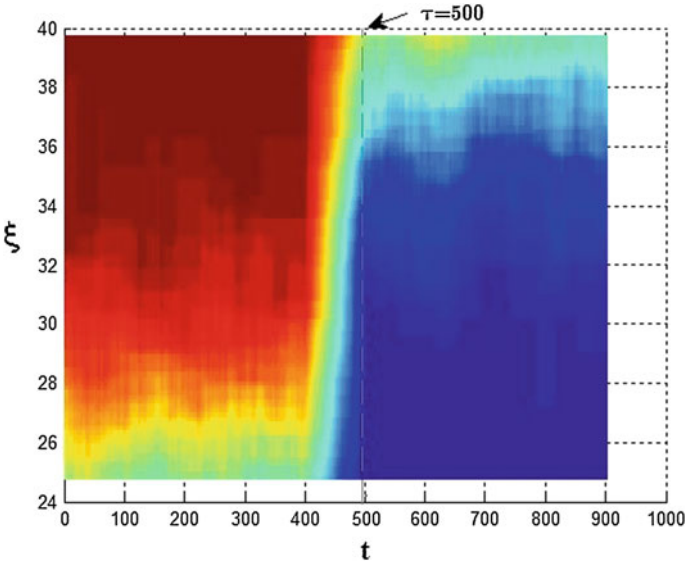


Fig. 3 3-dimensional FOT empirical profile for Model 1—bird's eye view

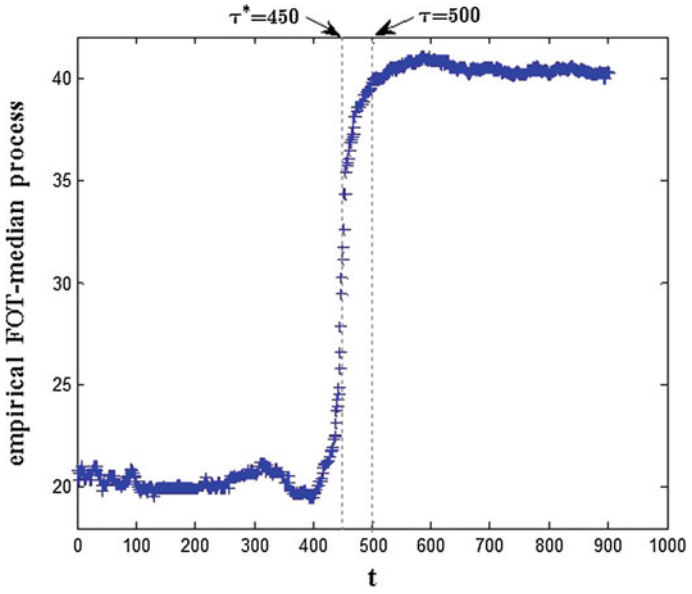


Fig. 4 FOT median process—Model 1

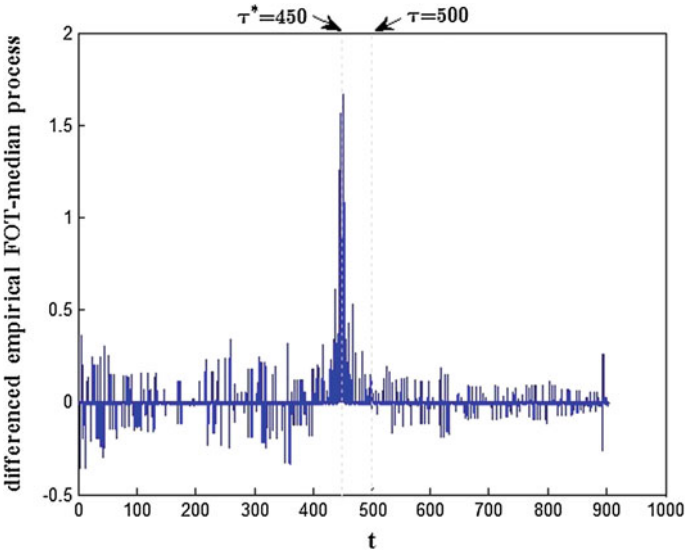


Fig. 5 Differenced FOT median process—Model 1

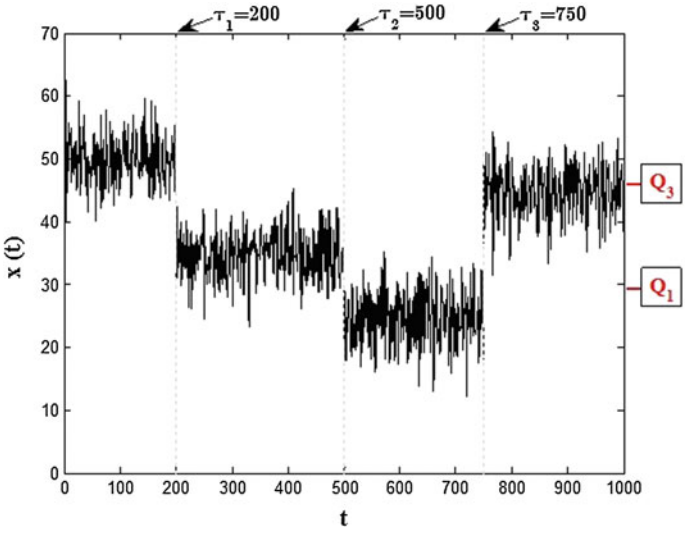


Fig. 6 Model 2 trajectory—multiple change-points

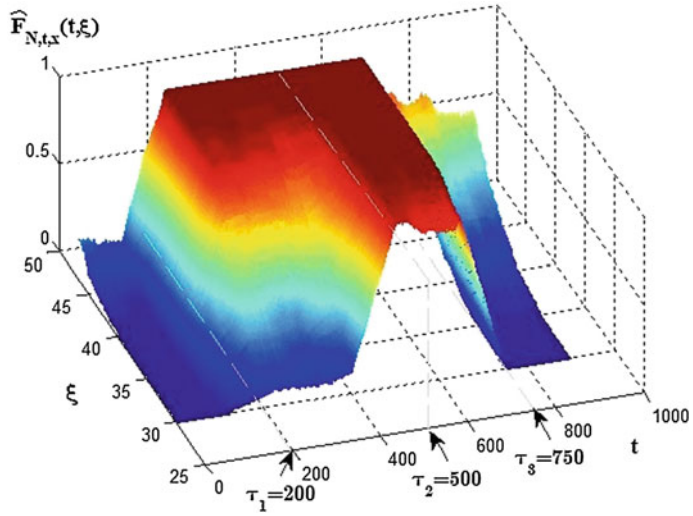


Fig. 7 3-D FOT empirical distribution profile for Model 2

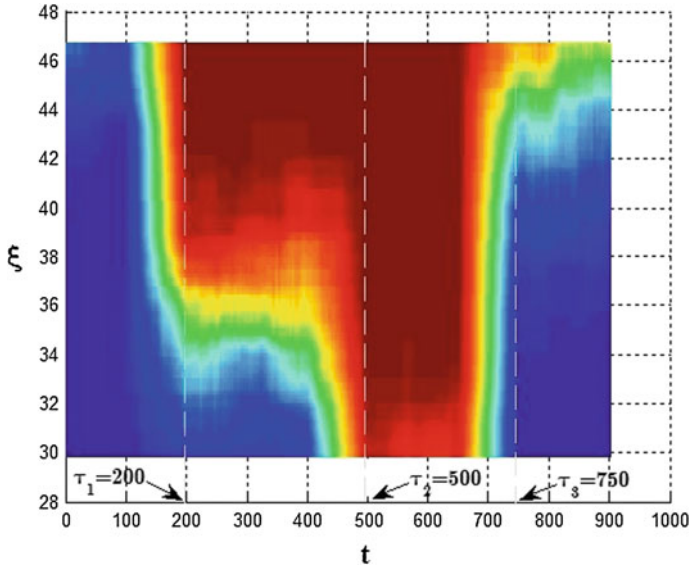


Fig. 8 3-D FOT empirical profile for Model 2—bird's eye view

3-D FOT profiles, indicating at possible change-points as can be seen in Figs. 7 and 8. Recall that every τ_j can be signaled either in advance or post-hoc, hence six distinct gradient changes are present.

As we had stated previously, the median process to some extent mimics the behavior of the signal itself, so rapid upward and downward trends are visible in Fig. 9.

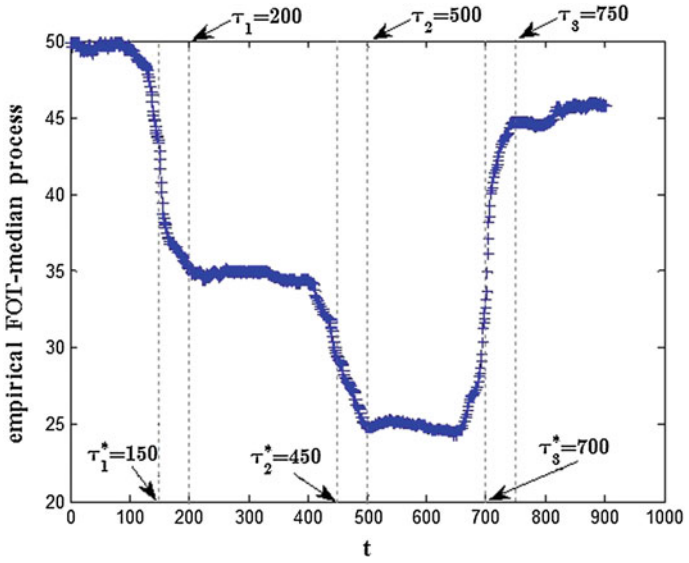


Fig. 9 FOT median process—Model 2

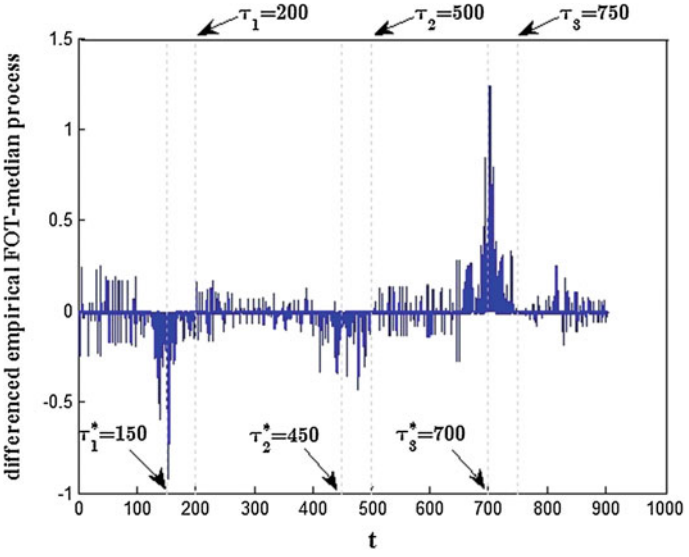


Fig. 10 Differenced FOT median process—Model 2

Differenced process plotted on Fig. 10 shows less or more pronounced peaks, the magnitude of which corresponds with respective level change intensities. The true change-point moments τ_j together with their leading versions τ_j^* are also marked.

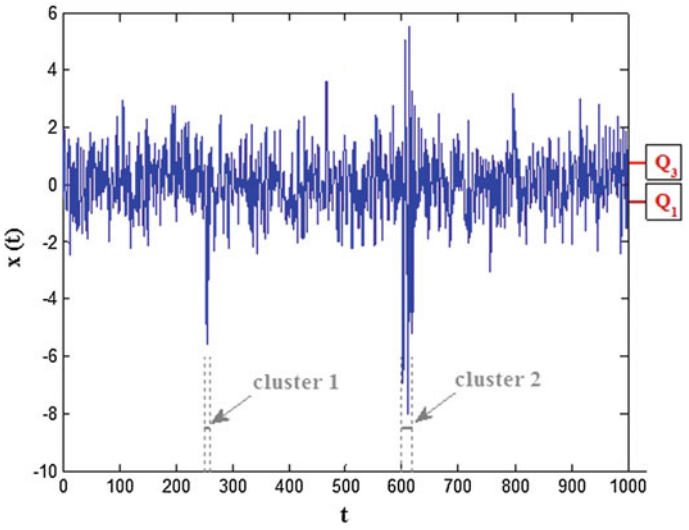


Fig. 11 Model 3 trajectory—transient clusters instead of change-points

Two large spikes show that the change-point moments τ_1 and τ_3 have been distinctly recorded, whereas τ_2 passed nearly unnoticed. Their respective estimators are $\hat{\tau}_1 = \hat{\tau}_1^* + 50 + 1 = 204, \hat{\tau}_2 = \hat{\tau}_2^* + 50 + 1 = 529, \hat{\tau}_3 = \hat{\tau}_3^* + 50 + 1 = 753$. The errors are therefore: 4, 29 and 3, respectively.

Model 3

The signal following this model is plotted in Fig. 11 with the two short-lasting clusters denoted and clearly visible. The sample quartile bounds here are $\xi_L = Q_1 = -0.62, \xi_U = Q_3 = 0.715$. Similar patterns can be found in logreturns of stock quotes, where calm market is sometimes disturbed by (usually transient) nervous episodes identified as clusters (Fig. 11).

As the 3-D profile suggests, the two short-lasting regimes correctly have not been classified as structural change-points (Fig. 12). Neither this device, nor the differenced quantile process, looking just noisy—see Fig. 13—detects any change-points. This proves the robustness of our detection techniques against outliers.

Real Data Set

The VIX historical daily quotes we consider span over 4 years' horizon and are plotted on Fig. 14. Throughout that period we have $\xi_L = Q_1 = 17.58, \xi_U = Q_3 = 28.46$ however one huge volatility spike appears roughly in mid-time, which coincides with the 2008 financial crash. In the following weeks the spike tapers off with VIX trending downwards. One more spike of smaller magnitude occurs towards the end of the time scope, which accounts for the spring 2010 mini-crash.

The 3-D FOT distribution profile (see Fig. 15) seems to indicate that regime changes occur at least on these three aforementioned occasions. This is confirmed by behavior of the differenced quantile process in Fig. 16. First, huge spike occurs

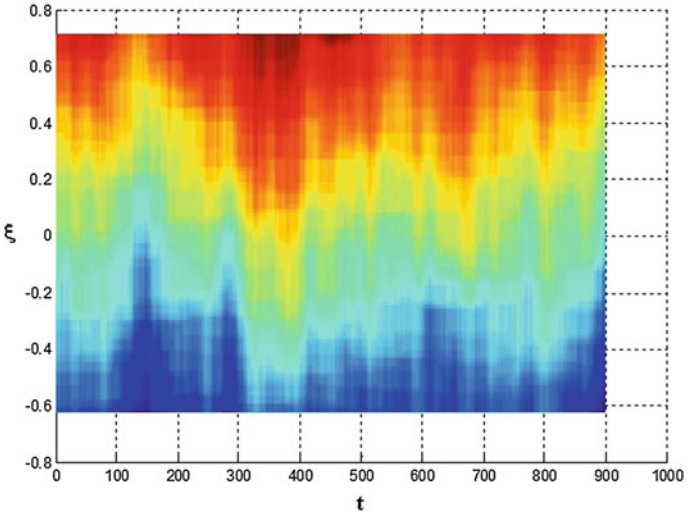


Fig. 12 3-D FOT empirical profile for Model 3—bird's eye view

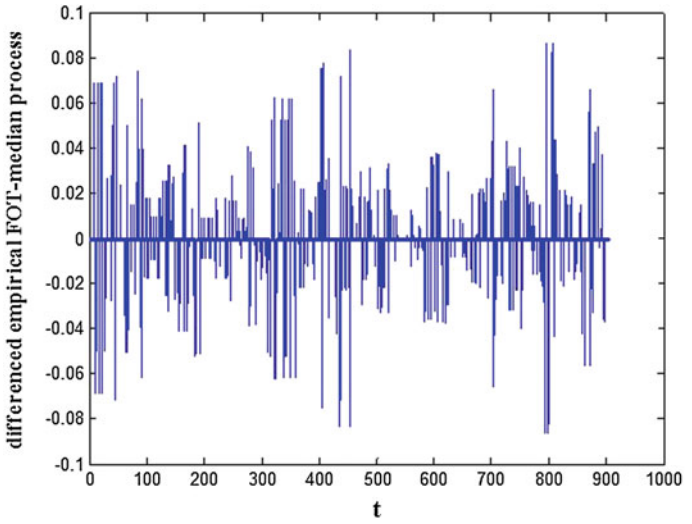


Fig. 13 Differenced FOT median process—Model 3

at $\hat{\tau}_1^* = 474$ which gives $\hat{\tau}_1 = \hat{\tau}_1^* + 50 + 1 = 525$. Next, after the crash volatility reverses downwards and this regime-change is detected at $\hat{\tau}_2^* = 530$ so that $\hat{\tau}_2 = 581$. Minor volatility spike is detected also in spring 2010 and in that case we have $\hat{\tau}_3^* = 873$ and accordingly $\hat{\tau}_3 = 924$. Looking at the VIX trajectory we can argue that these major shifts have been captured with a satisfactory accuracy.

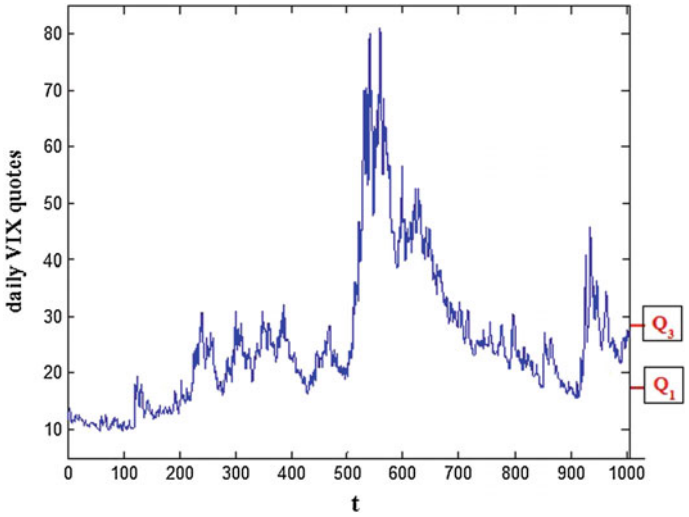


Fig. 14 Historical VIX daily quotes between Sept 1, 2006 and Aug 31, 2010

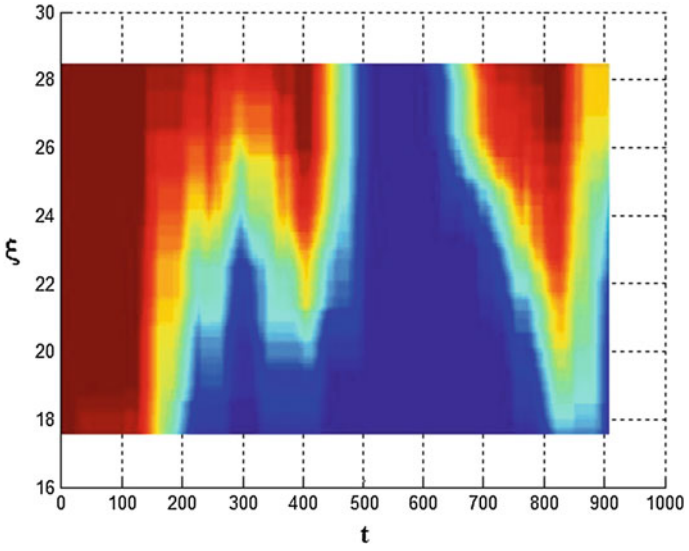


Fig. 15 3-D FOT empirical profile for VIX quotes—bird's eye view

4 Final Conclusions

In this paper we presented two change-point detection techniques within the non-stochastic Fraction-Of-Time framework. One technique is based on empirical FOT profiles considered as 3-dimensional surfaces evolving with t and ξ and the other uses

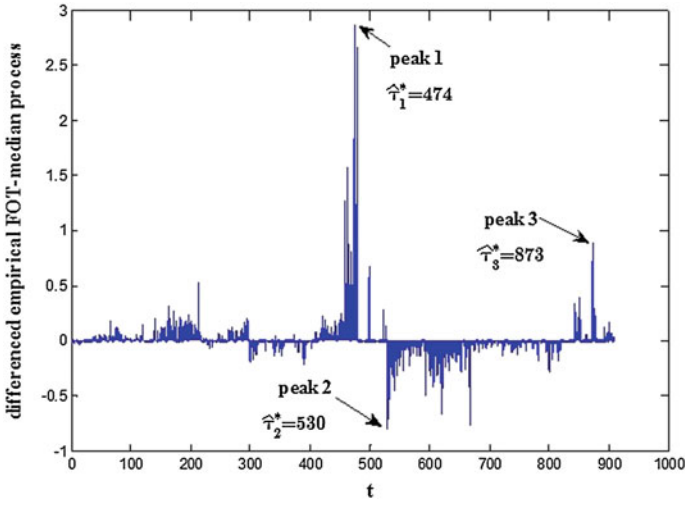


Fig. 16 Differenced FOT median process for VIX quotes

differenced median processes based on the empirical FOT distribution. For the three simulated models and one real data set the results look convincing. The true change point moments have been detected rather accurately (Model 1 and 2), whereas clusters and outliers have been correctly ignored in the sense that no change-point has been reported. The techniques have proved also useful in analyzing real data, here the daily VIX quotes. Within the considered 4-year time span three major change-points have been detected. Overall, further research focuses on optimizing the choice of the window width T depending on the sample size N as well as on applications for mechanical—possibly cyclostationary—signals.

Acknowledgements The authors thank an anonymous Referee for remarks and comments that improved the final version of the paper.

References

1. Chicago Board of Options Exchange. www.cboe.com.
2. Leśkow, J., & Napolitano, A. (2002). Prediction for time series in the fraction-of-time probability framework. *Signal Processing*, *82*(11), 1727–1741.
3. Leśkow, J., & Napolitano, A. (2005). *Fraction-Of-Time approach in predicting value-at-risk*. Berlin: Springer.

Seismic Signal Enhancement via AR Filtering and Spatial Time-Frequency Denoising

Marta Polak, Jakub Obuchowski, Agnieszka Wyłomańska and Radosław Zimroz

Abstract In this paper two-step algorithm based on autoregressive filtering and spatial time-frequency denoising is proposed. Firstly, we remove the discrete-spectrum components that contaminate the seismic signals using autoregressive (AR) filter. In contrast to widely used notch filter, the proposed scheme does not need to specify frequency of each component found as contamination. AR filter requires only order of model to enhance the signal. In the second step of the algorithm we remove the background noise of the AR model residuals using a filter based on the Short-Time Fourier Transform (STFT). Specifically, STFT amplitudes lower than specified cutoff are removed. We motivate our analysis by seismic signals analysis acquired using a seismic monitoring network in a copper-ore mine.

Keywords Seismic signals denoising · Autoregressive filter · Time-varying filter · Short-Time fourier transform

1 Introduction

Monitoring of seismic events induced by mining activity is a crucial issue in terms of safety in an underground mine. Moreover, seismic activity is one of the key factors taken into account during excavation planning. The most basic parameters

M. Polak · J. Obuchowski (✉) · R. Zimroz
KGHM CUPRUM Ltd., CBR, Sikorskiego 2-8, 53-659 Wrocław, Poland
e-mail: jobuchowski@cuprum.wroc.pl

M. Polak
e-mail: mpolak@cuprum.wroc.pl

R. Zimroz
e-mail: rzimroz@cuprum.wroc.pl

A. Wyłomańska
Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wroclaw University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: agnieszka.wylomanska@pwr.edu.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_4

monitored in underground mines include location and energy of seismic events. More advanced analysis include determination of source size, stress drop, seismic moment tensor, focal mechanism (normal fault, strike-slip, thrust), etc. [1, 2]. One can also discriminate seismic events using several other features based on P and S waveforms and their spectra, instantaneous frequency, group delay, correlation coefficient, etc. that might be found in the literature [3–7]. Such parameters might be further classified using neutral networks, Fisher classifiers, support vector machines or statistical methods [3, 8–10]. The classification error might be increased by choice of inappropriate feature, weak classifier or noise that contaminates seismic signals. In this paper we investigate the issue of seismic signal denoising. The signals are acquired by a seismic network installed in an underground mine. The noise that contaminates these signals is of two natures. One source of the noise is a wideband additive noise which is common also in fields other than seismic. Another kind of the noise is related to transmission from seismic sensors to the database and its spectrum is rather discrete. Such noise might influence seismic signal features based on the spectrum, thus denoising is a crucial preprocessing step for clustering. Moreover, noise might cause difficulties in P-wave picking, which affects accuracy of localization [11–13].

Classic method for removing discrete-spectrum component from the signal is a notch filter, which belongs to a group of stop-band filters [14]. Its characteristic property is narrow stopband, thus we can remove specific component from the signal. In order to use the notch filter the frequency band that contains interfering signal component should be specified. That is a significant disadvantage, especially in case of real seismic data where two or more discrete-spectrum components are present. Firstly one should specify each of the frequency band to filter it out, based on time-frequency representation of the signal (e.g. spectrogram or ARgram) and next apply set of notch filters.

The motivation of research on discrete-spectrum component removal is occurrence of this kind of noise in real seismic signals from a copper-ore mine. Noise which is visible through the entire length of seismogram might be also noticed on the spectrogram as high-amplitude ridges remaining throughout duration of the seismic signal. Moreover, the components responsible for noise have similar amplitude as the seismic event. Hence, there is need to apply an algorithm which reduces noise without elimination important information of the signal. In simple methods such as a high-pass filter the main difficulty is to choose the appropriate cutoff frequency to remove noise and leave seismic event without significant modification.

In this article we propose a method for removing one or more discrete-spectrum components without a priori specification of each frequency responsible for noise in the registered signal. Such a method is autoregressive model based filtering. It requires only order of model to delete noise in examined signals automatically, even for seismic recording contaminated by series of narrowband components.

The model's parameters estimation is based on initial part of signal which contains noise only, thus the process does not need additional time-frequency analysis.

In the next step the signal is denoised using a time-varying filter designed upon a spectrogram. Such a filter minimizes level of the background noise. As a result, it is easier to notice arrival of the signal related to the seismic event. Moreover proposed method of denoising can be useful in seismic signal segmentation [15] or clustering process [16] as pre-processing of examined data.

The paper is organized as follows. The algorithm of signal enhancement via autoregressive filter and spatial time-frequency denoising is described in Sect. 2. Evaluation of proposed method on simulated signals is presented in Sect. 3. Section 4 contains results of algorithm applied to the real data. Summary of the research is included in Sect. 5.

2 Methodology

In this section the algorithm of enhancement via AR filtering and spatial time-frequency denoising is introduced. It is suitable for the denoising problem when the entire seismic recording is contaminated by discrete-spectrum components (with amplitudes constant in time) and signal contains a few hundred samples before the first break phase (seismic signal arrival). Moreover the energy, carrier frequency and the number of narrowband components contaminating the entire signal is unknown and there is an additive noise with spectral amplitudes lower than these of pure seismic signal related to the seismic event.

First part of this Sect. 2.1 concerns the basic definitions related to time series analysis and ARMA modeling. Then motivation of using autoregressive filter will be presented in Sect. 2.2. The proposed algorithm consists of two steps. In the first one the autoregressive model is fitted to the reference noise. Then, the inverse filter is applied to the signal in order to remove discrete-spectrum component of noise. After preliminary noise reduction we use spatial time-frequency denoising in Sect. 2.3. Section 2.4 contains discussion on algorithm's parameters calibration.

2.1 Autoregressive Filter

Definition 1 [17] ARMA(p, q) (autoregressive—moving-average) model of orders p and q is defined as

$$X_t = \varepsilon_t + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} b_j \varepsilon_{t-j}, \tag{1}$$

where $a_1, \ldots, a_p$ are parameters of autoregressive model of order p, $b_1, \ldots, b_q$ are parameters of moving-average model of order q and $\{\varepsilon_t\}$ is white noise [17]. ARMA model can be interpreted as infinite impulse response filter (IIR). The transfer function for IIR filter is defined as:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{j=0}^{q} b_j z^{-j}}{1 + \sum_{i=1}^{p} a_i z^{-i}}, \tag{2}$$

where $X(z)$—z-transform of input signal and $Y(z)$—z-transform of output signal, q—order of moving average model for input, p—order of autoregressive model for output, $\{a_i\}$—coefficients for autoregressive model, $\{b_i\}$—coefficients for moving average model and z^{-i} is time-shifting operator.

In this paper we concentrate only on the autoregressive model, i.e. model defined as

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_t. \tag{3}$$

In terms of the transfer function for AR(p), the numerator in Eq. (2) reduces to 1.

2.2 Algorithm of Enhancement via AR Filtering

After analysis of the spectrogram of the real signal we noticed that there are some characteristic high–value ridges during all the record at specific frequency bins. Moreover, at the beginning of the seismogram there is no seismic event. In view of that first n-samples of signal might be treated as a reference noise, however the noise is not considered as the sequence of independent identically distributed random variables, in contrast to [18]. In order to select the appropriate model, the autocorrelation (ACF) and partial autocorrelation function (PACF) [17, 19] were examined (Fig. 1). Moreover the assumption of the stationarity is fulfilled, since the amplitude spectrum of the reference noise can be considered as constant. Therefore we can fit coefficients of the autoregressive model of order p to the reference noise. For estimation of AR model coefficients the standard method based on Yule-Walker equations [17] was applied. The coefficients received due to model fitting will be used to create the autoregressive filter which removes discrete-spectrum components present in the entire signal. Filter was applied not only to first n-samples (i.e. reference noise), but also to the rest of the record. Thus, the expected residual signal should not contain discrete–spectrum components which contaminate the signal.

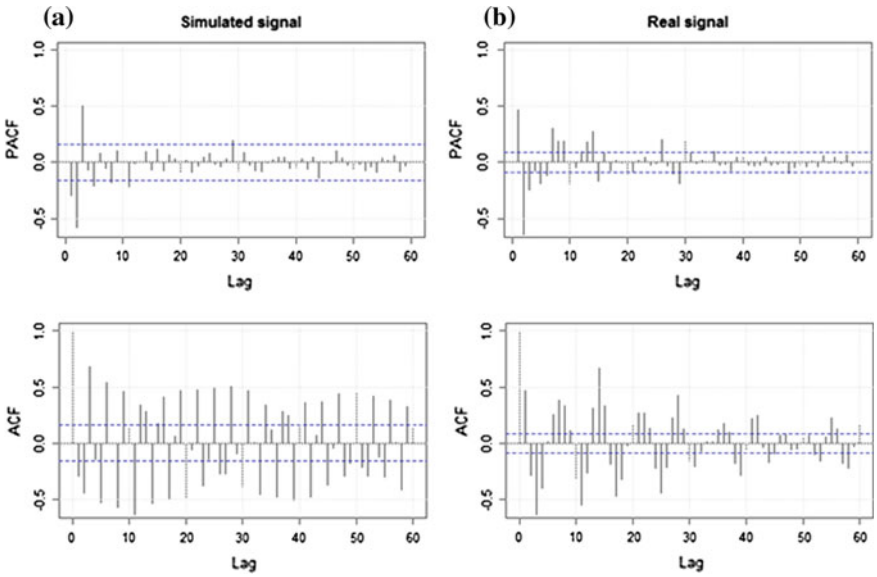


Fig. 1 PACF and ACF for **a** simulated and **b** real signal

2.3 Algorithm of Spatial Time-Frequency Denoising

Second step of the proposed algorithm is spatial time-frequency denoising. After AR filtering we calculate the Short-Time Fourier Transform from the filtered signal.

Definition 2 [20] Discrete Short-Time Fourier Transform (STFT) for observations $X_1, \ldots X_N$ is defined as:

$$\mathrm{STFT}(t,f) = \sum_{k=0}^{N-1} X_k w(t-k) e^{-\frac{2i\pi fk}{N}}, \quad t \in T, f \in F, \tag{4}$$

where t—time point, f—frequency, X_t—input signal and $w(t-k)$—shifted window.

Definition 3 [20] Squared magnitude of the STFT is defined as the spectrogram:

$$\mathrm{spect}[t,f] = |\mathrm{STFT}[t,f]|^2, \tag{5}$$

where t—time and f—frequency.

Detailed algorithm of spatial time-frequency denoising is presented below. The cutoff level for each frequency (c_{f_i}) is determined as q-quantile of absolute value of STFT for reference noise only (t_{ref}). The formula for cutoff level is described as:

$$c_{f_i} = \text{quantile}(|\text{STFT}[t,f_i]|, q), \quad f_i \in F, \quad t \in \{1, \ldots, t_{ref}\}, \tag{6}$$

The algorithm puts 0 to values less than the cutoff and creates denoised STFT described by equation:

$$\text{STFT}_d[t,f_i] = \begin{cases} \text{STFT}[t,f_i] & \text{if} |\text{STFT}[t,f_i]| > c_{f_i} \\ 0 & \text{if} |\text{STFT}[t,f_i]| \leq c_{f_i}, \end{cases} \tag{7}$$

where $f_i \in F$—frequency, $t \in \{1, \ldots, t_N\}$—time and N—length of the signal.

Last step of the algorithm uses Inverse Short-Time Fourier Transform for recovering signal from denoised STFT:

$$\text{STFT}^{-1}(\text{STFT}_d[t,f_i]) \quad f_i \in F, \quad t \in \{1, \ldots, t_{ref}\}, \tag{8}$$

where t_{ref}—length of the reference noise.

In implementation of STFT and STFT^{-1} one can use overlap–add method (OLA) [21]. All steps of the proposed algorithm are presented in block diagram in Fig. 2.

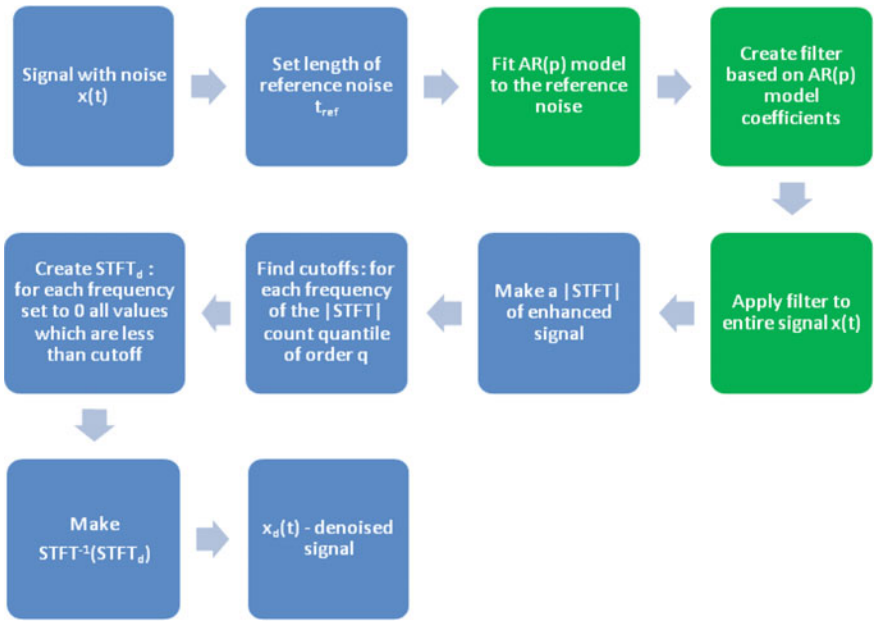


Fig. 2 Block diagram of the algorithm of enhancement via AR filtering (*green blocks*) and spatial time-frequency denoising (*blue blocks*)

2.4 AR Order and Reference Noise Length Selection Based on AIC

In this section a detailed methodology of choosing length of reference noise and order of autoregressive model will be described. As it was mentioned, first step of removing discrete-spectrum component is to fit an autoregressive model to reference noise. We can examine model's goodness of fit using Akaike Information Criterion (AIC), which for AR(p) is defined as:

$$\mathrm{AIC}(p) = -2\mathcal{L}(\theta|x) + 2p, \tag{9}$$

where p—model order, $\mathcal{L}(\theta|x)$—log-likelihood function. In our analysis we assume the ε_t in Eq. (1) is Gaussian then the log-likelihood function is based on the density of multidimensional Gaussian distribution. The best model in set of different AR models is that which minimizes AIC with respect to the order.

An autoregressive model is fitted to reference noise of different length. For each of them we examine models of order in range from 2 up to 60 and their goodness of fit. Figure 3a shows relation between AIC values and order of autoregressive model for simulated seismic signal from Fig. 6. Colors correspond to different lengths of reference noise, which are also marked at the seismogram of the signal in Fig. 3b. The longer the reference noise is, the more information contains. The character of the signal is reflected in order of the best AR model, which was indicated by AIC.

It should be mentioned that during the selection of length of reference noise we should be careful not to take too many samples which will contain information of the event additionally. Thus choice of the reference noise should be preceded by detailed analysis of signal's seismogram.

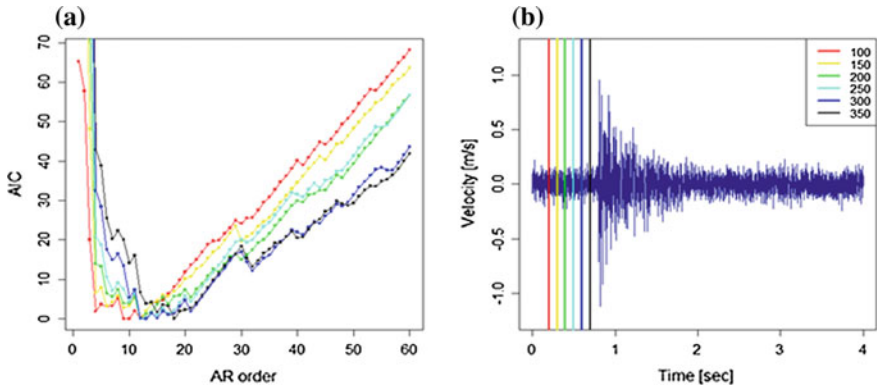


Fig. 3 **a** Relationship between order of autoregressive model and Akaike Information Criterion for different length of reference noise for simulated seismic signal (corresponding to colors of lines), **b** simulated signal with marked different length of reference noise

3 Simulation Results

In this section we will apply our denoising algorithm to simulated signal. In this situation we can control signal and character of additive noise. In Fig. 4 we can see each component of the noisy signal (Fig. 4c): pure signal (Fig. 4a) and noise (Fig. 4b). We simulate noise similar to one, which can be found in real seismic signals what can be noticed at the signal's spectrogram (Fig. 5).

First step of the proposed algorithm is enhancement via AR filtering. As it was described in Sect. 2.4 we select order of autoregressive model and length of the reference noise using AIC. Different lengths of the reference noise are presented on the seismogram of the simulated signal (Fig. 5). AIC is minimized for value 12 of

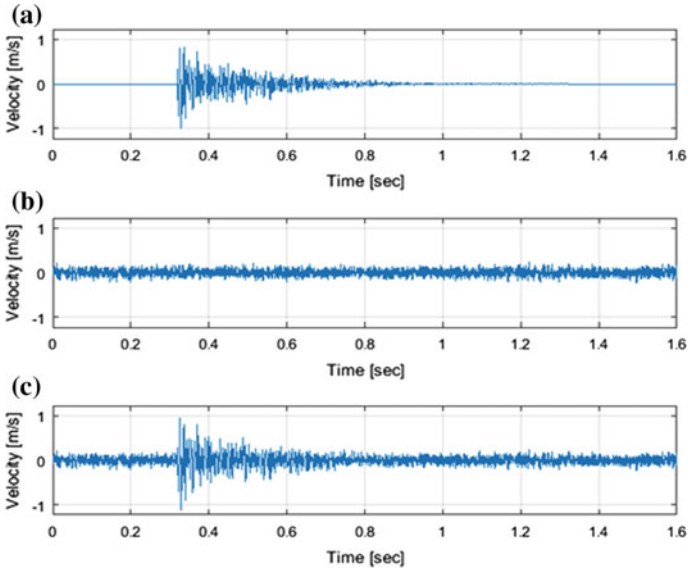


Fig. 4 Simulations of the seismic signal (**a**), noise (**b**) and noisy signal (**c**)

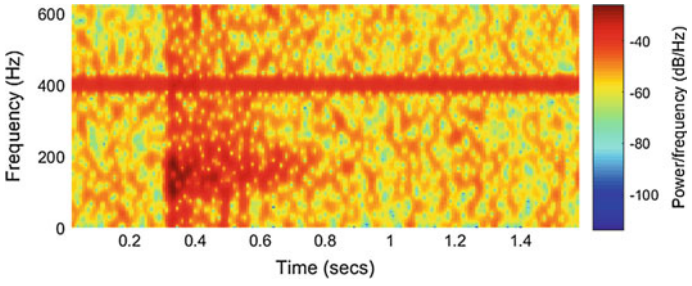


Fig. 5 Spectrogram of the simulated noisy signal

AR order for lengths of reference noise equal 150, 200, 250 and 300 (Fig. 3a). For the further analysis we take the reference noise of length 150. That choice will be concluded at the end of this section.

After choice of length of reference noise and order of AR filter, we apply the first step of the algorithm, namely enhancement. To notice changes in properties of the record we should examined time-frequency structure of signal. As an alternative to classical spectrogram we use more precise time-frequency map called ARgram [22, 23].

Definition 4 [23] The gliding power spectrum density of AR model (ARgram) of signal $x(t)$ is defined as

$$S_{AR}(t,f) = \frac{P_w}{\left|1 + \sum_{k=1}^{p} a_{k,t} e^{-2i\pi fk}\right|^2}, \tag{10}$$

where P_w—power of the white noise, p—model order and $a_{k,t}$ are coefficients of model for part of signal in consecutive, overlapping window began at time t.

As we can see in Fig. 6a there is clearly visible ridge caused by discrete-spectrum component. During the enhancement via AR filtering the ridge is removed what one can notice in Fig. 7b. Reduction of noise is visible also at the seismogram of the signal (compare Figs. 6a and 7a), but still noise is not removed from the record.

Second step of the proposed algorithm should clean off rest of the noise. The order of quantile for each frequency component was set to 0.95. Results of spatial

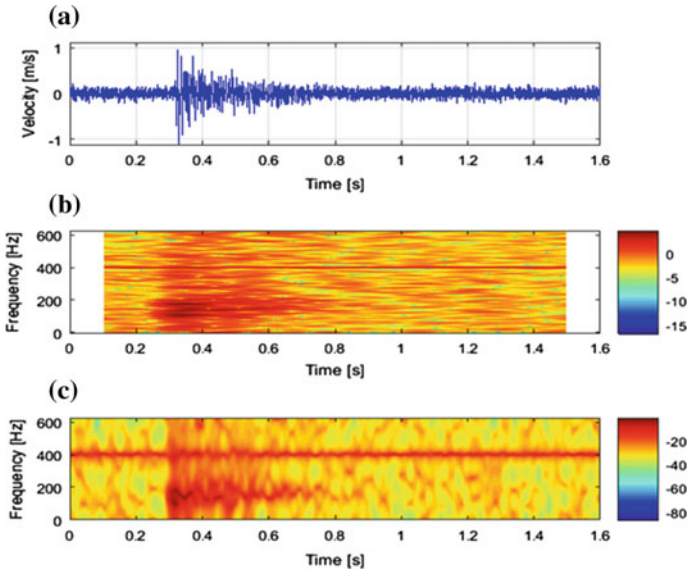


Fig. 6 Seismogram of simulated noisy signal (**a**), its spectrogram (**b**) and ARgram (**c**)

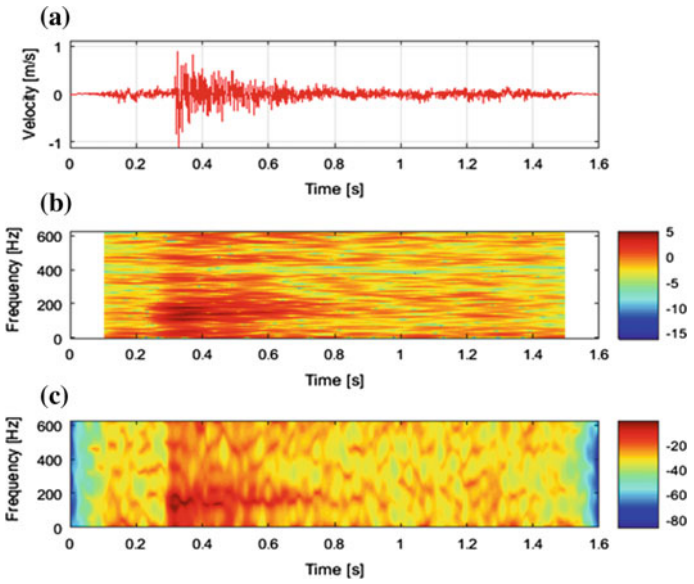


Fig. 7 Seismogram of simulated signal after enhancement via AR filtering (**a**), its spectrogram (**b**) and ARgram (**c**)

time-frequency denoising can be found in Fig. 8. We can notice noise cancellation at the beginning and the end of the signal. Moreover amplitude of the noisy area in the middle of the record is decreased, but seismic event does not change the structure.

For comparison, in Fig. 9 we present results of spatial time-frequency denoising without prior AR filtering. It is worth noting that there are residues of discrete-spectrum component (marked in Fig. 9b). This proves that for seismic signal with discrete-spectrum component only cut off values below specific level at the |STFT| are not sufficient. In view of that we should apply two-step algorithm of denoising for this kind of records.

In case where we control component of pure signal and noise we can check effectiveness of proposed algorithm of enhancement via AR filtering and spatial time-frequency denoising by analysis of Signal-to-Noise Ratio.

Definition 5 [24, 25] Signal-to-Noise Ratio (SNR) is defined as:

$$\mathrm{SNR} = \frac{P_{signal}}{P_{noise}} = \left(\frac{A_{signal}}{A_{noise}}\right)^2, \tag{11}$$

where P_{signal}—power of a signal, P_{noise}—power of noise, A_{signal}—root mean square of the signal amplitude, A_{noise}—root mean square of the noise amplitude.

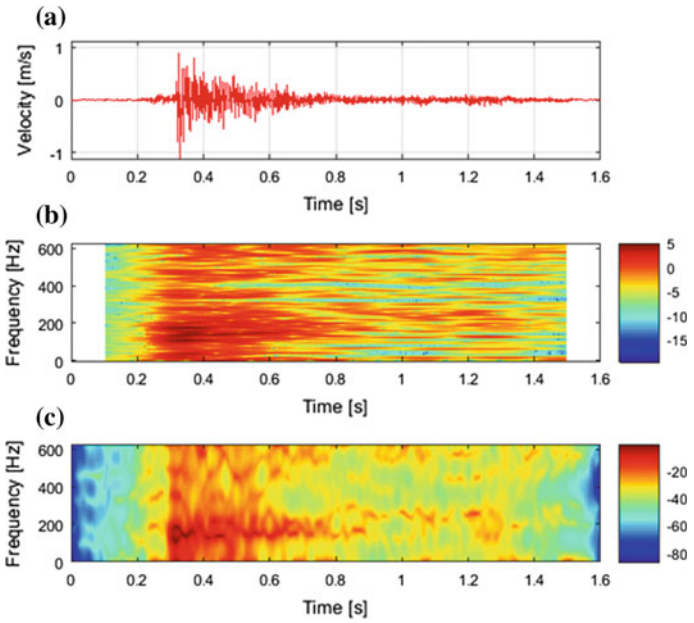


Fig. 8 Seismogram of simulated signal after enhancement via AR filtering and spatial time-frequency denoising (**a**), its spectrogram (**b**) and ARgram (**c**)

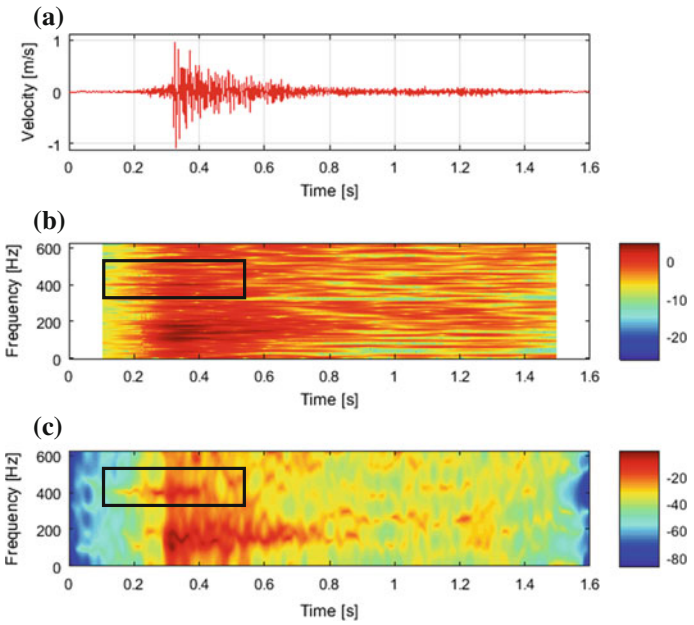


Fig. 9 Seismogram of simulated signal after just spatial time-frequency denoising (**a**), its spectrogram (**b**) and ARgram (**c**)

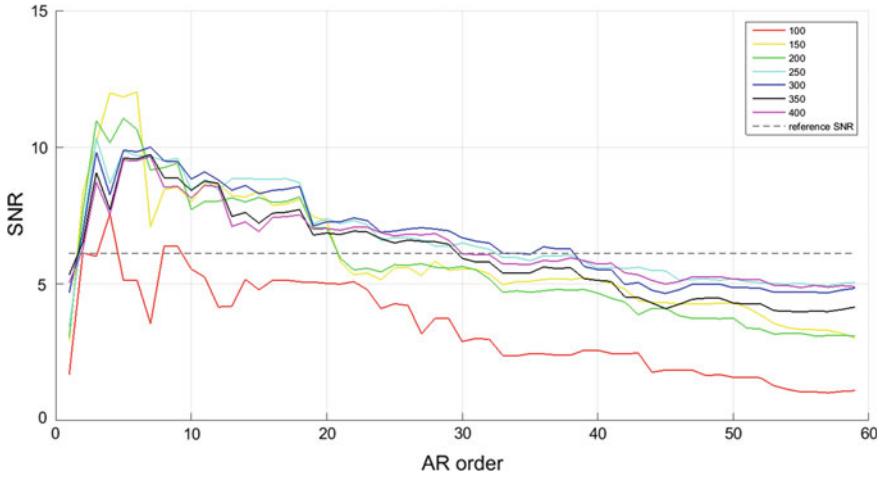


Fig. 10 Signal-to-Noise Ratio for simulated signal after enhancement via AR filtering for different AR order and different length of reference noise (corresponding to colors of lines)

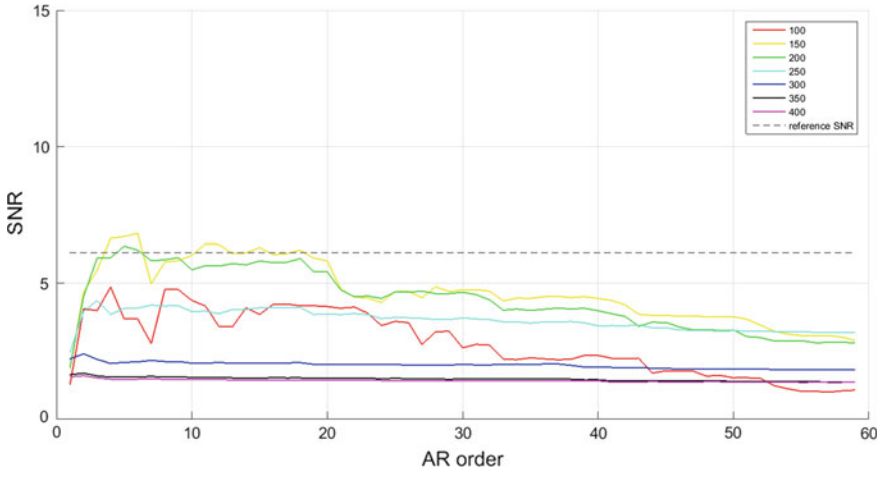


Fig. 11 Signal-to-Noise Ratio for simulated signal after enhancement via AR filtering and spatial time-frequency denoising for different AR order and different length of reference noise (corresponding to colors of lines)

Firstly we present results of SNR analysis just after enhancement via AR filtering for simulated seismic signal (Fig. 10). For the initial components of noisy signal we can calculate reference SNR which is marked in Figs. 10 and 11 by gray dashed line. Other lines in Fig. 10 correspond to different lengths of reference noise taken for algorithm of enhancement via AR filtering. Analysis were carried out for different orders of autoregressive model. We can see that removing of

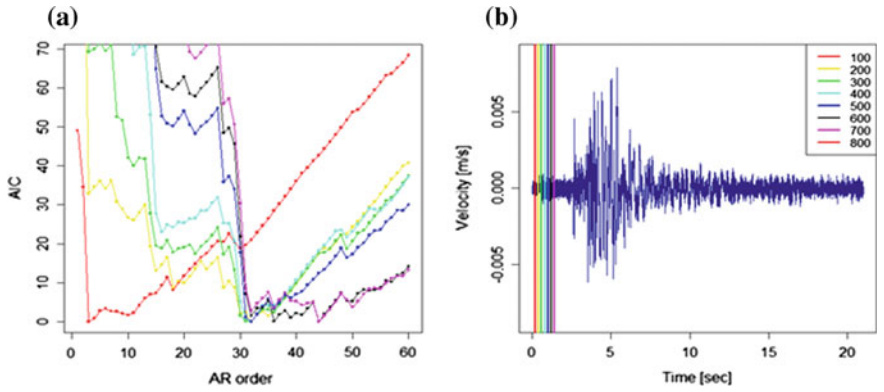


Fig. 12 **a** relationship between order of autoregressive model and Akaike Information Criterion for different length of reference noise for real seismic signal presented in Fig. 13, **b** real signal with marked different length of reference noise

discrete-spectrum component from the signal does not suppress all noise in examined record (what can be observed also at seismogram of the signal in Fig. 7), because values of SNR are not close to the reference SNR.

After applying second step of the algorithm which was spatial time-frequency denoising we also analyzed behavior of SNR (Fig. 11). Comparing plots in Figs. 10 and 11 we noticed decrease of SNR which means that noise which we removed from signal is larger when we applied both steps of the algorithm. However we should notice that in case when SNR for specific parameters is smaller than reference SNR we could remove information of seismic event simultaneously with noise cancellation. The best approximation of reference SNR is observed in case when we use 150-samples reference noise and autoregressive filter with order in the range 10-20. For longer reference noise SNR is smaller than reference SNR, what indicates that signal is denoised "too much". It is explanation why 150-samples reference noise was taken for the analysis.

4 Real Data Results

This section contains results of two-step denoising algorithm applying for real seismic signals, which was registered in copper-ore mine. In Fig. 13 it is presented the seismogram and ARgram of the exemplary record, which we examined in this paper. One can found two ridges at the ARgram, which are responsible for noise in analyzed seismic signal.

In real data the level of noise is unknown thus we cannot use SNR as indicator of appropriate length of reference noise. In Fig. 12a we can see that for the reference

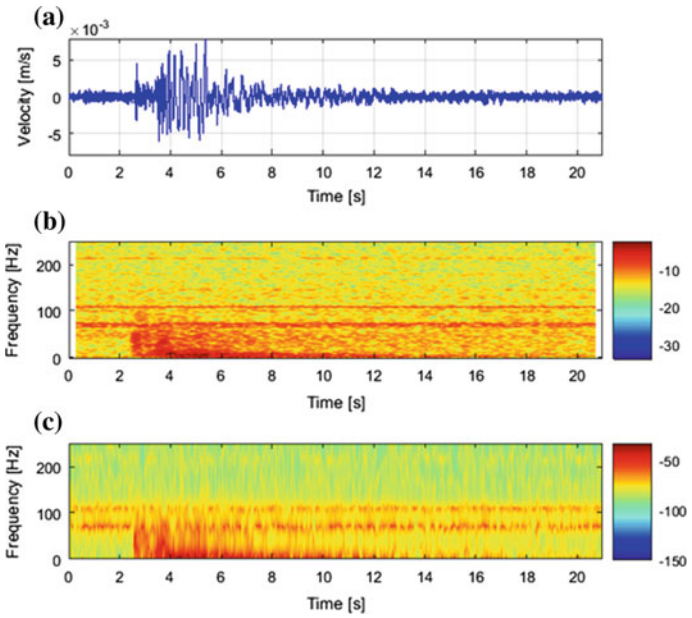


Fig. 13 Seismogram of real noisy signal (**a**), its spectrogram (**b**) and ARgram (**c**)

noise of length 100 the AIC indicates AR(4). From analysis of amplitude spectrum structure of the signal we should reject such a low order of model, because it does not correspond to local maxima of the amplitude spectrum. In view of that we chose the longest reference noise which minimized AIC to better estimation of autoregressive coefficients. For the longer reference noises (for example of length 200–500 samples) Akaike Information Criterion is minimized for the same order of autoregressive model and returns AR(32). Parameters of autoregressive model are more accurate if estimation is basing on larger number of samples. In view of that we take longer reference noise for the analysis in case when the same AR order was indicated by AIC. Summarizing, further analysis was carried out for filter AR (32) and 500-samples reference noise.

As we can see on the ARgram (Fig. 14b) AR filtering removed discrete-spectrum components from the signal, but still (as in case of simulated signal) vibrations disrupted signal are visible on the seismogram. After applying second step of algorithm, the noise is dumped (Fig. 15). Application of the proposed algorithm to real seismic signals gives satisfying results, which are better than denoising only by spatial time-frequency denoising without enhancement via AR filtering. Comparing ARgrams in Figs. 15b and 16b one can found the remains of the ridges.

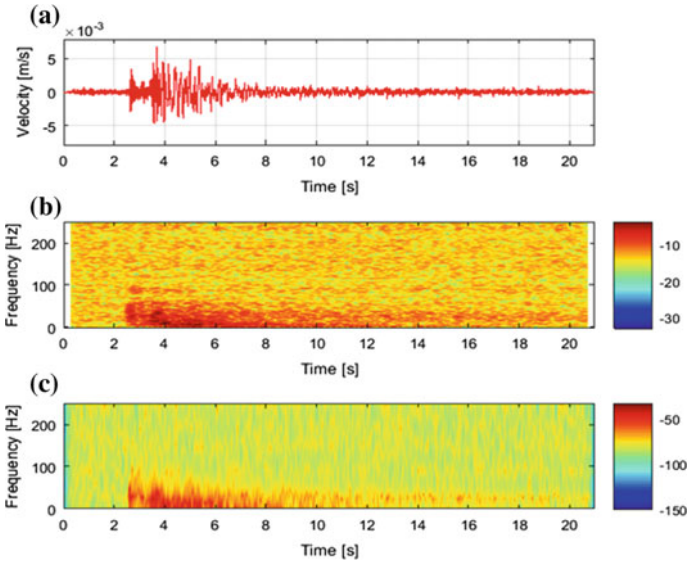


Fig. 14 Seismogram of real signal after enhancement via AR filtering (**a**), its spectrogram (**b**) and ARgram (**c**)

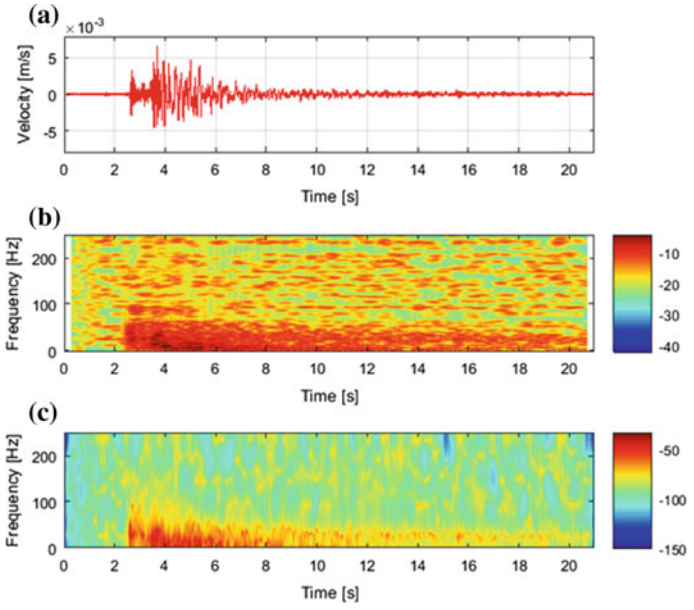


Fig. 15 Seismogram of real signal after enhancement via AR filtering and spatial time-frequency denoising (**a**), its spectrogram (**b**) and ARgram (**c**)

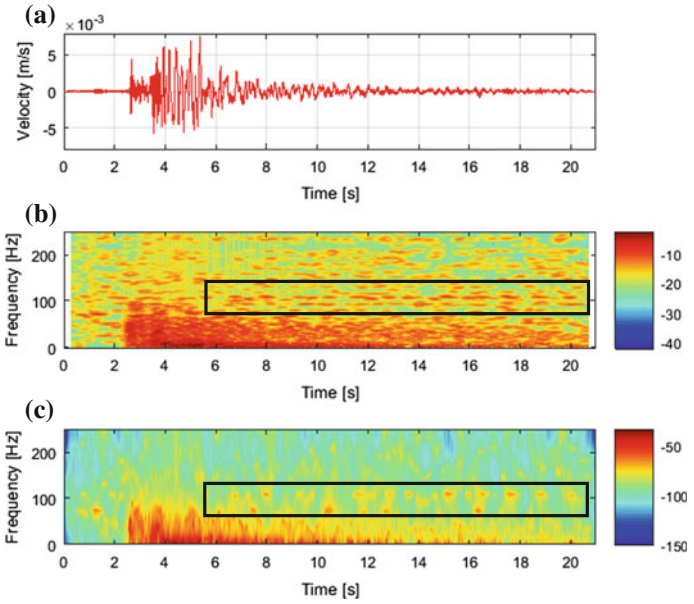


Fig. 16 Seismogram of real signal after just spatial time-frequency denoising (**a**), its spectrogram (**b**) and ARgram (**c**)

5 Conclusions

In this paper we considered the problem of denoising of seismic signals from discrete-spectrum component noise. The proposed two-step algorithm using autoregressive filtering and spatial time-frequency transform of STFT was applied to the simulated and real seismic signals. In each case the high value ridges on the ARgram (responsible for the noise) were disappeared, what verify the effectiveness of the proposed method. The choice of parameters of the AR filter (model order and length of the reference noise) is crucial, thus they were selected based on Akaike Information Criterion, which is classic measure of goodness-of-fit for autoregressive models. Moreover in case of simulated data the Signal-to-Noise-Ratio was examined to evaluate the effectiveness of denoising. The results of analysis proved that seismic signal enhancement via AR filtering before spatial time-frequency denoising has positive influence for removing discrete-spectrum components from the noisy signal.

References

1. Gibowicz, S. J., & Kijko, A. (2013). *An introduction to mining seismology* (Vol. 55). Elsevier.
2. Beresnev, I. A. (2001). What we can and cannot learn about earthquake sources from the spectra of seismic waves. *Bulletin of the Seismological Society of America, 91*(2), 397–400.
3. Malovichko, D. (2012). Discrimination of blasts in mine seismology. In *Deep mining* (pp. 161–171).
4. Ma, J., Zhao, G., Dong, L., Chen, G., & Zhang, C. (2015). A comparison of mine seismic discriminators based on features of source parameters to waveform characteristics. *Shock and Vibration*, Article ID 919143.
5. Polak, M., Obuchowski, J., Madziarz, M., Wyłomańska, A., & Zimroz, R. (2016). Time–varying group delay as a basis for clustering and segmentation of seismic signals. *Journal of Vibroengineering, 18*(1), 267–275.
6. Sokołowski, J., Obuchowski, J., Madziarz, M., Wyłomańska, A., & Zimroz, R. (2016). Features based on instantaneous frequency for seismic signals clustering. *Journal of Vibroengineering, 18*(3), 1654–1667.
7. Shumway, R. H. (1996). Statistical approaches to seismic discrimination. In *Monitoring a comprehensive test ban treaty* (pp. 791–803). Netherlands: Springer.
8. Kushnir, A. F., Troitsky, E. V., Haikin, L. M., & Dainty, A. (1999). Statistical classification approach to discrimination between weak earthquakes and quarry blasts recorded by the Israel seismic network. *Physics of the Earth and Planetary Interiors, 113*(1), 161–182.
9. Kortström, J., Uski, M., & Tiira, T. (2016). Automatic classification of seismic events within a regional seismograph network. *Computers & Geosciences, 87,* 22–30.
10. Vallejos, J. A., & McKinnon, S. D. (2013). Logistic regression and neural network classification of seismic records. *International Journal of Rock Mechanics and Mining Sciences, 62,* 86–95.
11. Leśniak, A., & Pszczoła, G. (2008). Combined mine tremors source location and error evaluation in the Lubin Copper Mine (Poland). *Tectonophysics, 456*(1), 16–27.
12. Hafez, A. G., Rabie, M., & Kohda, T. (2013). Seismic noise study for accurate P-wave arrival detection via MODWT. *Computers & Geosciences, 54,* 148–159.
13. Parolai, S. (2009). Denoising of seismograms using the S transform. *Bulletin of the Seismological Society of America, 99*(1), 226–234.
14. Kanasewich, E. R. (1990). *Seismic noise attenuation* (Vol. 7). Pergamon Press.
15. Zimroz, R., Madziarz, M., Żak, G., Wyłomańska, A., & Obuchowski, J. (2015). Seismic signal segmentation procedure using time frequency decomposition and statistical modelling. *Journal of Vibroengineering, 17*(1), 3111–3121.
16. Leśniak, A., & Isakow, Z. (2009). Space–time clustering of seismic events and hazard assessment in the Zabrze-Bielszowice coal mine, Poland. *International Journal of Rock Mechanics and Mining Sciences, 46*(5), 918–928.
17. Brockwell, P. J., Davis, R. A. (2013). *Time series: Theory and methods*. Springer Science & Business Media.
18. Kucharczyk, D., Wyłomańska, A., Obuchowski, J., Zimroz, R., Madziarz, M. (2016). Stochastic modelling as a tool for seismic signals segmentation. *Shock and Vibration*, Article ID 8453426.
19. Wyłomańska, A. (2012). How to identify the proper model? *Acta Physica Polonica B, 43*(5), 1241–1253.
20. Boashash, B. (2003). *Time frequency signal analysis and processing: A comprehensive reference*. Oxford: Elsevier.
21. Dutoit, T., & Marqur̈es, F. (2009). *Applied signal processing: A MATLAB-based proof of concept*. New York: Springer.
22. Padovese, L. R., Martin, N., & Millioz, F. (2009). Time–frequency and time–scale analysis of Barkhausen noise signals. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 223*(5), 577–588.

23. Martin, N. (1986). An AR spectral analysis of non-stationary signals. *Signal Processing, 10* (1), 61–74.
24. Tucker, D. G., & Gazey, B. K. (1966). Applied underwater acoustics. Elsevier Science & Technology.
25. Carlson, A. B., Crilly, P. B., & Ruttledge, J. C. (1986). *Communication systems. An introduction to signals and noise in electrical engineering*.

Transformed GARMA Model with the Inverse Gaussian Distribution

Breno Silveira de Andrade, Jacek Leśkow and Marinho G. Andrade

Abstract We propose an extension of our previous work [10] regarding GARMA models. The task of this paper is to study transformations leading to GARMA models. We present a simulation study and real data analysis using the Transformed GARMA (TGARMA) models with the inverse Gaussian distribution. We assume the same properties as in [10] however we apply to the inverse Gaussian distribution. We show an application to a real data related to number of hypertension cases in São Paulo.

Keywords Generalized ARMA model · Box Cox transformation · Inverse Gaussian distribution

1 Introduction

In our previous paper [10] we have proposed a model that is based on Box Cox transformation and, subsequently, on the GARMA model. Thus we have obtained TGARMA model. Therefore, the proposed model was TGARMA (Transformed Generalized Autoregressive Moving Average). The simulation study and real data analysis presented in that paper was related to the gamma distribution. This work focused in the inverse Gaussian distribution which is positive asymmetric and can be used when the gamma distribution cannot be fitted to data. Therefore, the novelty of our paper as compared to the previous paper [10] is in using the inverse Gaussian distribution. The inverse Gaussian distribution is also a positive asymmetric distribution that can be useful when gamma distribution cannot be fitted to the data.

B.S. de Andrade
Programa de Pós-Graduação Interinstitucional UFSCar/ICMC-USP, São Carlos, Brazil

J. Leśkow (✉)
Institute of Mathematics, Cracow Technological University, Kraków, Poland
e-mail: jleskow@pk.edu.pl

M.G. Andrade
Departamento de Matemática Aplicada e Estatística - ICMC-USP, São Carlos, Brazil

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_5

The inverse Gaussian distribution is an useful alternative for non-negatives time series, this distribution can reach up high values and admit some extreme values. The inverse Gaussian distribution allows maximum likelihood estimation of the parameter, facilitating the incorporation of the autoregressive and moving average terms. An alternative distribution with this general shape is the three parameter Weibull distribution, which is more difficult to estimate.

The use of transformations have been shown a good alternative to reduce these kind of problems. Hamasaki and Kim [13] described a Box and Cox power-transformation to confined and censored nonnormal responses in regression. da Silva et al. [9] proposed the use of Box-Cox transformation and regression models to deal with fecal egg count data. Gillard [12] presented a study using Box-Cox family of transformations, the paper comment about problems with asymmetry in the transformed data. [5] commented about many fields where the Box-Cox transformation can be used, and also proposed a method to improve the foresting models. Ahmad et al. [1] combined Box-Cox and bootstrapping idea in one algorithm, the Box-Cox is to ensure the data is normally distributed and bootstrap to deal with small and limited sample size data.

The Box Cox transformation parameter λ is estimated using the profile likelihood (PL). The method was introduced by Cox [7] and is based entirely on the conditional distribution of the current response, given past responses, and past covariates information and functions thereof can be used for inference. Zhu and Ghodsi [20] presented a procedure to dimensionally selection maximizing a profile likelihood function. Huang et al. [14] proposed an efficient equation for estimating the index parameter and unknown link function using Adaptive profile-empirical-likelihood inferences. Cole et al. [6] provide a primer on maximum likelihood, Profile Likelihood and Penalized Likelihood which have proven useful in epidemiologic research.

Our paper is organized as follows. In Sect. 2 we provide the idea of transformations that constitute an important motivation to introduce TGARMA model and resampling techniques. Section 3 defines the TGARMA model in general and with continuous distributions. This section also contains the model fitting for TGARMA models present Likelihood estimate and Partial Likelihood. Section 4 describes all the simulation study. The software [19] was used to the simulation study. The real data application was presented in Sect. 5 applying the methodology on number of hypertension cases in São Paulo data. Section 6 gives concluding remarks.

2 Transformations

Box and Cox [4] commented that many important results in statistical analysis follow from the assumption that the population being sampled or investigated is normally distributed with a common variance and additive error structure. For this reason, these authors presented a transformation called Box-Cox power transformation that has generated a great deal of interests, both in theoretical work and in practical applications.

This family has been modified [8] to take account of the discontinuity at $\lambda = 0$, such that

$$y_t^{(\lambda)} = \begin{cases} \frac{(Y_t^{\lambda}-1)}{\lambda}; \lambda \neq 0 \\ \log(Y_t); \lambda = 0 \end{cases}$$

Draper and Cox [11], Manly [17], Sakia [18] discuss others transformations which have the same aim, reduce anomalies in the data. The literature recommend the use of Box-Cox power transformation as a general transformation. The next section present the TGARMA approach using the Box-Cox power transformation.

3 TGARMA Model Fitting

3.1 Model Definition

The TGARMA model specifies the conditional distribution of each transformed observation $y_t^{(\lambda)}$, for $t = 1, \ldots, n$ given the previous information set, defined by $F_{t-1}^{(\lambda)} = (y_1^{(\lambda)}, \ldots, y_{t-1}^{(\lambda)}, \mu_1, \ldots, \mu_{t-1})$. The conditional density belongs to exponential family and is given by

$$f(y_t^{(\lambda)}|F_{t-1}^{(\lambda)}) = \exp\left(\frac{y_t^{(\lambda)}\alpha_t - b(\alpha_t)}{\varphi} + d(y_t^{(\lambda)}, \varphi)\right), \tag{3.1}$$

where α_t and φ are canonical and scale parameters, respectively. Moreover $b(\cdot)$ and $d(\cdot)$ are specific functions that define the particular exponential family. The conditional mean and conditional variance of y_t given $F_{t-1}^{(\lambda)}$ are represented as:

$$\begin{aligned} \mu_t = b'(\alpha_t) = E(y_t^{(\lambda)}|F_{t-1}^{(\lambda)}) \\ Var(y_t^{(\lambda)}|F_{t-1}^{(\lambda)}) = \varphi b''(\alpha_t), \end{aligned} \tag{3.2}$$

with $t = 1, \ldots, n$.

Following the Generalized Linear Models (GLM) approach the parameter μ_t is related to the predictor η_t by a twice differentiable one-to-one monotonic function g, called *link function*. In general, we can also include set of covariates $\mathbf{x}$ into our model. Moreover, we can we add an additional component allowing autoregressive moving average terms to be included. In such a case our model will have a form:

$$g(\mu_t) = \eta_t = x_t'\beta + \sum_{j=1}^{p} \phi_j\{g(y_{t-j}^{(\lambda)}) - x_{t-j}'\beta\} + \sum_{j=1}^{q} \theta_j\{g(y_{t-j}^{(\lambda)}) - \eta_{t-j}\}. \tag{3.3}$$

The parameters p and q are identified using the classical BIC or AIC criteria. For more information the reader is referred to [15, 16].

The TGARMA(p,q) model is defined by the Eqs. (3.1) and (3.3). For certain functions g, it may be necessary to replace $y_t^{(\lambda)}$ with $y_t^{(\lambda^{new})}$ in (3.3) to avoid the non-existence of $g(y_t^{(\lambda)})$ for certain values of y_t. The form $y_t^{(\lambda^{new})}$ depends on the particular function g and is defined for specific cases later.

In our approach we will not include covariates. We will consider two important continuous GARMA models: gamma and inverse Gaussian. We will present each one with a density and a respective predictor. The simulation study and real data analysis were done for each of the distributions.

3.2 Inverse Gaussian TGARMA Model

Suppose that $y_t^{(\lambda)}|F_{t-1}^{(\lambda)}$ follows a inverse Gaussian distribution with the mean μ_t, thus

$$f(y_t^{(\lambda)}|F_{t-1}^{(\lambda)}) = \exp\left\{\frac{1}{\sigma^2}\left[-\frac{2y_t^{(\lambda)}}{\mu_t^2}+\frac{1}{\mu_t}\right]-\frac{1}{2}\log(2\pi\sigma^2 y_t^{(\lambda)^3})-\frac{1}{2\sigma^2 y_t^{(\lambda)}}\right\}. \quad (3.4)$$

The canonical link function for this model is the logarithmic function. The linear predictor is given by

$$\log(\mu_t) = \beta_0 + \sum_{j=1}^{p}\phi_j\left\{\log(y_{t-j}^{(\lambda)})\right\} + \sum_{j=1}^{q}\theta_j(\log(y_{t-j}^{(\lambda)}) - \log(\mu_{t-j})), \quad (3.5)$$

The inverse Gaussian TGARMA model is defined by the Eqs. (3.4) and (3.5).

3.3 Model Fitting

The TGARMA model-fitting procedure described herein performs the maximum likelihood estimation (see [3]). The estimation method is based on the standard GLM.

Let $\{y_t^{(\lambda)}\}$ be a time series where the Eqs. (3.1) and (3.3) are satisfied. The parameter vector is $\gamma' = (\beta', \phi', \theta')$, where $\beta = (\beta_0 \ldots \beta_m)'$, $\phi = (\phi_1 \ldots \phi_p)'$ and also $\theta = (\theta_1 \ldots \theta_q)'$. For the estimation procedure the conditional likelihood function is used where the $F_r^{(\lambda)} = \{y_1^{(\lambda)} \ldots y_r^{(\lambda)}\}$, for $r = \max(p, q)$. The partial likelihood function can be constructed considering that $y_{t-1}^{(\lambda)}$ and $y_t^{(\lambda)}$ are conditionally independent. Thus

$$L(\beta, \phi, \theta | F_n^{(\lambda)}) \propto \prod_{t=r+1}^{n} f(y_t^{(\lambda)} | F_t^{(\lambda)})$$
$$\propto \prod_{t=r+1}^{n} \exp\left\{ \frac{y_t^{(\lambda)} g(\mu_t) - b(g^{-1}(\mu_t))}{\varphi} + d(y_t^{(\lambda)}, \varphi) \right\}, \quad (3.6)$$

where $g(\mu_t)$ is the link function given by

$$g(\mu_t) = x_t'\beta + \sum_{j=1}^{p} \phi_j \{ g(y_{t-j}^{(\lambda)}) - x_{t-j}' \} + \sum_{j=1}^{q} \theta_j \{ g(y_{t-j}^{(\lambda)}) - g(\mu_{t-j}) \}. \quad (3.7)$$

In the above equations $t = r+1, \ldots, n$. The Eqs. (3.6) and (3.7) do not have a closed form solution therefore a numerical optimization routine will be used. Due to the results of [3] the asymptotic distribution of the partial likelihood estimator is known. However, the information matrix of the corresponding asymptotic normal law is quite difficult to estimate. This gives us a strong motivation to consider resampling techniques in the GARMA models. As it also known, resampling techniques provide a better alternative to construct confidence intervals in the time series field [2].

The profile likelihood for the parameter vector γ and scalars ϕ and λ is expressed in terms of the transformed series $y^{(\lambda)} = (y_{r+1}^{(\lambda)}, \ldots, y_n^{(\lambda)})^T$ conditioned on the first r transformed observations, where $r = \max\{p, q\}$, is

$$PL(\beta, \phi, \theta, \lambda) = \prod_{t=r+1}^{n} L(\beta, \phi, \theta | F_t^{(\lambda)}) J(\lambda, y_t), \quad (3.8)$$

where $J(\lambda, y_t)$ is the Jacobian of the transformation from y_t to $y_t^{(\lambda)}$. The inference method is made selecting a range $\lambda^{(1)}, \ldots, \lambda^{(k)}$, and evaluating the profile likelihood on each $\lambda^{(i)}$, $i = 1, \ldots, k$.

$$\hat{\lambda} = \max\left(PL(\hat{\gamma}^{(\lambda^i)}, \hat{\phi}^{(\lambda^i)}, \lambda^i) \right), i = 1, \ldots, k \quad (3.9)$$

The profile likelihood depends on the Jacobian of the transformation, so different transformation provide different profiles likelihood. The Box-Cox power transformation was used, thus

$$J(\lambda, y_t) = \begin{cases} \sum_{i=1}^{n} \left\{ (\lambda - 1) \log(y_i) \right\}; \lambda \neq 0 \\ \sum_{i=1}^{n} \left\{ -\log(y_i) \right\}; \lambda = 0 \end{cases}$$

4 Simulation Study

The performance of the Maximum Likelihood estimation was evaluated using two metrics: the corrected bias (CB), the corrected error (CE). These metrics are defined as,

$$CB = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\theta - \hat{\theta}^{(i)}}{\theta}\right|,$$

$$CE^2 = \frac{1}{Var}\frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}^{(i)} - \theta)^2$$

Table 1 Inverse Gaussian GARMA(p,q) real values

Order	β_0	ϕ_1	ϕ_2	θ_1	θ_2	γ
(1,1)	0.70	0.50	–	0.30	–	0.5
(1,2)	−0.10	0.50	–	0.20	−0.30	2
(2,1)	−0.10	0.50	0.30	−0.40	–	0.5
(2,2)	−0.10	−0.30	0.20	0.25	−0.35	2

Table 2 TGARMA(1,1) with inverse Gaussian distribution and Box-Cox power transformation

Parameter ($\lambda = 0.3$)	CB	CE	Coverage
γ	0.0464	0.9407	0.9662
β_0	0.1611	1.1144	0.9226
ϕ_1	0.0868	1.0896	0.9274
θ_1	0.0636	0.9431	0.9560
Parameter ($\lambda = 0.5$)	CB	CE	Coverage
γ	0.0506	1.0143	0.9538
β_0	0.1588	1.0988	0.9242
ϕ_1	0.0844	1.0671	0.9290
θ_1	0.0610	0.8985	0.9632
Parameter ($\lambda = 0.7$)	CB	CE	Coverage
γ	0.0522	1.0524	0.9488
β_0	0.1627	1.1269	0.9154
ϕ_1	0.0844	1.0759	0.9312
θ_1	0.0600	0.8725	0.9678
Parameter ($\lambda = 0.9$)	CB	CE	Coverage
γ	0.0584	1.1635	0.9260
β_0	0.1722	1.1897	0.9026
ϕ_1	0.0874	1.1000	0.9296
θ_1	0.0586	0.8403	0.9766

Table 3 Proportions of correct model using BIC with inverse Gaussian TGARMA(p,q) model

$\lambda = 0.3$				
Size	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
200	0.6929	0.2881	0.4110	0.4041
500	0.7312	0.3372	0.5144	0.4251
1000	0.7815	0.3515	0.5752	0.4554
$\lambda = 0.5$				
Size	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
200	0.6963	0.3174	0.3549	0.4450
500	0.7155	0.3615	0.3882	0.4672
1000	0.7469	0.3855	0.4398	0.4740
$\lambda = 0.7$				
Size	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
200	0.6404	0.3730	0.3260	0.4462
500	0.6980	0.4016	0.3955	0.4593
1000	0.7088	0.4222	0.4439	0.4737
$\lambda = 0.9$				
Size	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	TGARMA(2,2)
200	0.6981	0.2735	0.3894	0.4198
500	0.7430	0.2912	0.4341	0.4321
1000	0.7775	0.3176	0.4707	0.4437

where $\hat{\theta}^{(i)}$ are the estimate of parameter θ for the i-th replication, $i = 1, \ldots, m$. The variance term (Var) that appears in the definition of CE is the sample variance of $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}$. Also the Coverage was presented that represent the proportion of simulated series where the confidence intervals contain the true simulated value.

The estimation results appear in Tables 2 and 4 where the aforementioned metrics are shown for each model and parameter. These results indicate good properties with relatively small values of the corrected bias (CB), values of the corrected error (CE) around 1 and also coverage around 95%.

In Table 3 we present results for the simulation experiment that was checking proportion of times when the correct model was selected. The BIC criterion was used to select the best model, the criterion was used also with different sizes of series.

The experiment was repeated 5000 times for each value of λ, with 500 observations in each of artificial time series. The simulation study for TGARMA models was conducted with Gamma with the orders (1,1), (2,1), (1,2) and (2,2). The results of the orders (1,2) and (2,1) was omitted, but is very similar of the results of orders (1,1) and (2,2) that follow.

The simulation study was carried out to verify the influence of selecting the value of λ on the TGARMA model estimates. Table 2 presents results for the TGARMA model with one autoregressive and one moving average terms, the metrics proposed above showed that for different values of λ, the CB, CE and coverage still indicating good properties. Table 3 presents the proportions of correct model selected by using

Table 4 TGARMA(2,2) with inverse Gaussian distribution and Box-Cox power transformation

Parameter ($\lambda = 0.3$)	CB	CE	Coverage
γ	0.0129	1.0424	0.9437
β_0	1.2730	1.2350	0.8851
ϕ_1	0.3920	1.0544	0.9155
ϕ_2	0.3540	1.0315	0.9292
θ_1	0.4504	1.0254	0.9250
θ_2	0.1868	1.0323	0.9300
Parameter ($\lambda = 0.5$)	CB	CE	Coverage
γ	0.0132	1.0560	0.9477
β_0	1.3470	1.2749	0.8819
ϕ_1	0.3980	1.0688	0.9155
ϕ_2	0.3465	1.0273	0.9283
θ_1	0.4576	1.0444	0.9141
θ_2	0.1920	1.0487	0.9247
Parameter ($\lambda = 0.7$)	CB	CE	Coverage
γ	0.0143	1.1206	0.9378
β_0	1.3070	1.2363	0.9205
ϕ_1	0.3863	1.0248	0.9234
ϕ_2	0.3345	0.9841	0.9401
θ_1	0.4520	1.0016	0.9279
θ_2	0.1908	1.0246	0.9327
Parameter ($\lambda = 0.9$)	CB	CE	Coverage
γ	0.0135	1.0535	0.9513
β_0	1.3920	1.3200	0.8893
ϕ_1	0.4070	1.1115	0.8936
ϕ_2	0.3545	1.0292	0.9179
θ_1	0.4688	1.0767	0.8996
θ_2	0.2028	1.1007	0.9270

the BIC criterion, four different models were proposed, (1,1), (2,1), (1,2) and (2,2). The results presented on Table 2 shows that the higher proportions indicate the corrected model, unrelated to the λ value. The Table 4 presents results for the TGARMA model with two autoregressive and two moving average terms, the results follow the same aspect of the Table 1.

5 Real Data Analysis

Hypertension (HTN or HT), also known as high blood pressure, is a long term medical condition in which the blood pressure in the arteries is persistently elevated. High blood pressure usually does not cause symptoms. Long term high blood

pressure; however, is a major risk factor for coronary artery disease, stroke, heart failure, peripheral vascular disease, vision loss, and chronic kidney disease.

The dataset presents the number of hypertension cases in São Paulo, one of the most important cities in South America. The original series was divided by 1000 reducing the order of the data. The data presents monthly number of hypertension cases between December of 2003 until January of 2013. The data divided by 1000 presents non-negatives values and possible extreme values, thus the inverse Gaussian is a good option to adequate the data (Figs. 1, 2 and Table 5).

The range of the transformation parameter λ was selected from $0 \leq \lambda < 1$, with step 0.001. The Partial Likelihood selected the value $\hat{\lambda} = 0.0530$ (Table 6).

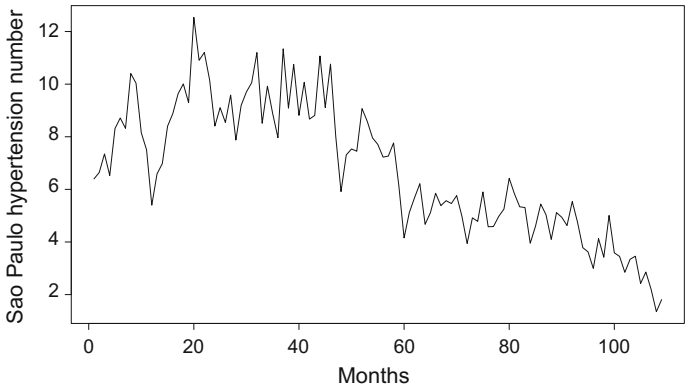


Fig. 1 Graph of number of hypertension in Sao Paulo

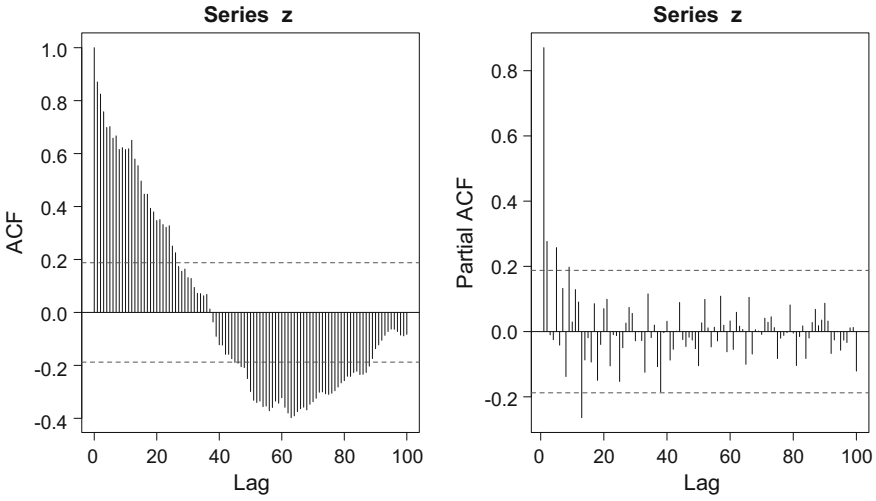


Fig. 2 Auto correlation function and partial auto correlation function of number of hypertension in Sao Paulo

Table 5 Criterion selection using number of hypertension in Sao Paulo

Inverse Gaussian	TGARMA(1,1)	TGARMA(1,2)	TGARMA(2,1)	**TGARMA(2,2)**
AIC	232.23	132.60	141.84	**129.46**
BIC	224.16	121.83	131.08	**116.01**

Table 6 Estimates of number of hypertension in Sao Paulo series with TGARMA(2,2) inverse Gaussian

Parameter	Estimate	Inferior bound	Superior bound
β_0	−0.0945	−0.2180	0.0598
ϕ_1	0.6223	−0.6299	1.6096
ϕ_2	0.5014	−0.5358	1.7646
θ_1	−0.2514	−1.3361	1.0464
θ_2	−0.1847	−0.9962	0.5095
ν	12.3145	8.2165	14.1992

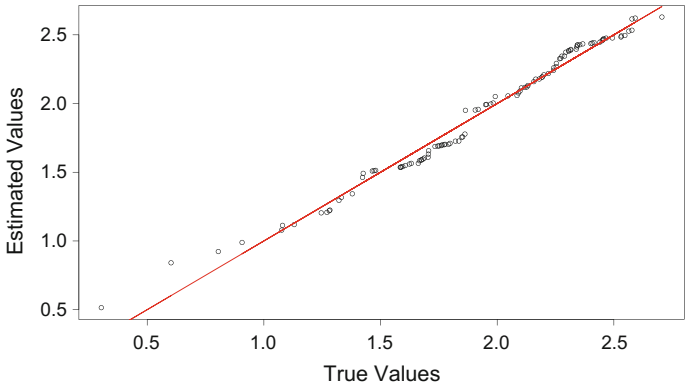


Fig. 3 Graph of true values versus the estimated values of the residuals of rate of number of hypertension in Sao Paulo series

The goodness of fit can be checked using the graph of true values versus the estimated values (Fig. 3).

The Fig. 3 presents a quantile plot with the true values on x axis and the estimated values on y axis. The line represents the perfect model with real values on axis x and y.

Quantile residuals are based on the idea of inverting the estimated distribution function for each observation to obtain exactly standard normal residuals. In the case of discrete distributions, such as the binomial, negative binomial and Poisson, some randomization is introduced to produce continuous normal residuals. The residuals is given by $r_t = \Phi^{-1}(\mathbf{F}_{y_t}(y_t|F_{t-1}))$ where $\mathbf{F}_{y_t}$ represent the cumulative distribution function for the respective density (Fig. 4).

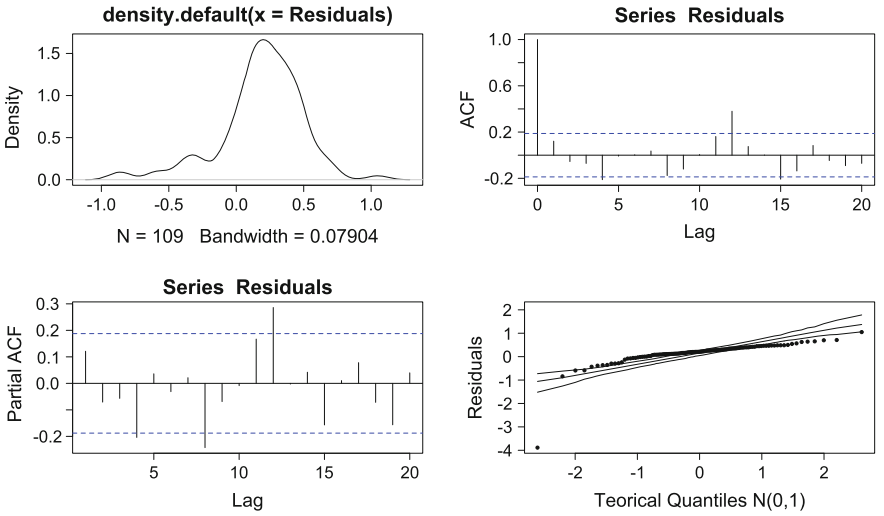


Fig. 4 Autocorrelation function and partial autocorrelation function of the residuals of rate of number of hypertension in Sao Paulo rates series

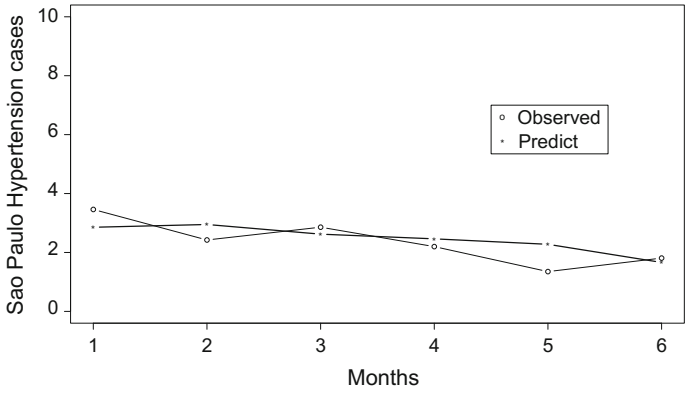


Fig. 5 Original predictions with GARMA(2,2) Inverse Gaussian model with number of hypertension in Sao Paulo series

The prediction were made by the median. Only the fist term of Taylor expansion were used. Using the estimate, predictions of 6 steps ahead of the original series can be made. The 6 last values of the series were removed and fitted the model without them. Prediction one step ahead for 6 months values, thus the predicted value be compared with the true value. The MAPE was calculated to quantify the quality of predictions (Fig. 5).

6 Conclusion

The model under study is a GARMA model that was recently introduced by Benjamin et al. [3]. There are many potential uses of GARMA model in time series analysis. For example, while analyzing discrete time series, GARMA structures allow embedding continuous type ARMA model in the parametrization of the discrete distribution (see [2]). Due to increased popularity of GARMA models, there is a need to find transformations letting the model correspond to data more flexibly. We propose Box-Cox transformation. Our research shows that the Partial Likelihood method can be successfully applied in finding optimal value of λ for the Box-Cox transformation. Such value, say $\hat{\lambda}$, can then be used to transform the data and use the GARMA model for statistical inference. Jointly with [10] work good properties can be evaluated from the TGARMA models, showing a useful class of models.

Acknowledgements Jacek Leśkow acknowledges the financial support from the Polish National Center for Science under the contract: UMO-2013/10/M/ST1/00096. Breno Andrade acknowledges the financial support from Brazilian research agency CAPES, process number: 99999.003692/2015-09 that has allowed him to spend an academic year at CTU in Cracow. The technical assistance obtained from the Institute of Telecommunication of Cracow Technical University is also gratefully acknowledged here.

References

1. Ahmad, W. M. A. W., Zakaria, S. B., Aleng, N. A., Halim, N. A., & Ali, Z. (2015). Box-cox transfromation and bootstrapping approach to one sample t-test. *World Applied Sciences Journal*, *33*(5), 704–708.
2. Andrade, S. B., Leśkow, J., & Andrade, M. G. (2016). Garma models and moving block bootstrap: Simulations and applications (Submitted).
3. Benjamin, M., Rigby, R., & Stasinopoulos, D. (2003). Generalized autoregressive moving average models. *Journal of American Statistical Association*, *98*, 214–223.
4. Box, G., & Cox, D. (1964). An analysis of transformation. *Journal of the Royal Statistical Society Series B*, *26*, 211–252.
5. Castillo, J., & Benitez F. G. (2013). Improving trip forecasting models by means of the box-cox transformation. *Transportmetrica A: Transport Science*, *9*, 653–674.
6. Cole, S. R., Chu, H., & Greenland, S. (2014). Maximum likelihood, profile likelihood, and penalized likelihood: A primer. *American Journal of Epidemiology*, *179*(2), 252–260.
7. Cox, D. (1975). Partial likelihood. *Biometrika*, *62*, 69–76.
8. Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, *8*, 93–115.
9. da Silva, M. V. G. B., Van Tassell, C. P., Sonstegard, T. S., Cobuci, J. A., & Gasbarre, L. C. (2011). Box-cox transformation and random regression models for fecal egg count data. *Frontiers in Genetics*, *2*.
10. de Andrade, B. S., Leśkow, J., & Andrade, M. G. (2016). Transformed garma model: Properties and simulations. *Communications in Statistics—Simulation and Computation*.
11. Draper, N., & Cox, D. (1969). On distributions and their transformation to normality. *Journal of the Royal Statistical Society Series B*, *31*, 472–476.
12. Gillard, J. (2012). A generalised box-cox transformation for the parametric estimation of clinical reference intervals. *Journal of Applied Statistics*, *39*, 2231–2245.

13. Hamasaki, T., & Kim, S. (2007). Box and cox power-transformation to confined and censored nonnormal responses in regression. *Computational Statistics and Data Analysis*, *51*, 3788–3799.
14. Huang, Z., Pang, Z., & Zhang, R. (2013). Adaptive profile-empirical-likelihood inferences for generalized single-index models. *Computational Statistics and Data Analysis*, *62*, 70–82.
15. Kedem, B., & Fokianos, K. (2002). *Regression models for time series analysis*. New Jersey, USA: Wiley.
16. Li, W. K. L. (1994). Time series model based on generalized linear models: Some further results. *Biometrics*, *50*, 506–511.
17. Manly, B. (1976). Exponential data transformation. *The Statistician*, *25*, 37–42.
18. Sakia, R. (1992). The Box-Cox transformation technique: A review. *The Statistician*, *41*, 168–178.
19. Team, R. C. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
20. Zhu, M., & Ghodsi, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics and Data Analysis*, *51*, 918–930.

GARCH Process with GED Distribution

Małgorzata Wiśniewska and Agnieszka Wyłomańska

Abstract Analysis of volatility is a key problem especially in economics and related issues. The volatility can help to measure the risk or the error sizes obtained in modeling several financial variables. But many financial models assume the volatility is constant over time, which contradicts the reality. Therefore in this paper we consider two models that were proposed as a description of real time series with heteroscedastic behavior, namely ARCH and GARCH systems. Moreover, because many observed phenomena do not exhibit Gaussian law therefore we extend the classical econometric models to the case with non-Gaussian distribution. As the extension we propose to use the Generalized Error distribution that is more adequate to examined time series. In this paper we give the main properties of considered models and present testing and estimation procedures. We illustrate the theoretical results with real financial data analysis and simulation study.

Keywords ARCH model · GARCH model · Generalized error distribution · Testing · Prediction · Calibration

1 Introduction

Volatility is one of the most important parameter used in economics. It can help to measure the risk or the error sizes obtained in modeling several financial variables. Furthermore, it is essential for making forecasts. Very often, approaches to data modeling are based on the assumption that the volatility is homoscedastic (i.e. it is constant in time). However, it was observed that in many cases the average size of volatility fluctuates over time. Moreover, it still can be predictable, provided that the accurate model was chosen. The classical models, such as ARMA (autoregressive

M. Wiśniewska · A. Wyłomańska (✉)
Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center,
Wrocław University of Science and Technology, Wroclaw, Poland
e-mail: agnieszka.wylomanska@pwr.wroc.pl; agnieszka.wylomanska@pwr.edu.pl

M. Wiśniewska
e-mail: mal.ag.wisniewska@gmail.com

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_6

moving average) process, assume the homoscedasticity of the variance and hence they are not useful in modeling those economic phenomena characterized by the heteroscedastic nature of the volatility.

In this paper we concentrate on the heteroscedastic approach in the modeling issue. More precisely, we analyze the Autoregressive Conditional Heteroscedasticity (ARCH) process which was first introduced by Engle in 1982 [6] and its extension i.e. the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) that Bollerslev [2] proposed in 1986. Those processes have been successfully applied to model, among others, interest rates, exchange rates, stock returns. Moreover, they were used in VaR calculation or option pricing, [7, 24]. Apart from heteroscedasticity, the original idea is based on the normality of the unconditional distribution of the innovation terms. However this assumption can be generalized since the distribution of the financial asset returns are not always consistent with Gaussian law. We can make such conclusion for example by simply examining the empirical distribution of the asset returns which very often exhibit thicker tails than those from a normal distribution. That is why, we consider conditionally non-Gaussian (G)ARCH models, namely with Generalized Error innovations, [8, 16, 19].

The paper is organized as follows. In the Sect. 2 some basic notations and definitions with remarks concerning classical ARCH and GARCH models are provided. We also indicate at the strong relation between examined processes and classical ARMA models. Moreover we discuss the main properties of the Generalized Error distribution that is taken into account as the candidate for the distribution of innovations.

Next, in the Sect. 3, some well-known tests like Engle ARCH or Ljung-Box are mentioned in order to examine the correlations between data. Then we go through the theory of parameter estimation based on the maximizing the likelihood function corresponding to Generalized Error distribution.

In Sect. 4 we evaluate the analytical forecast simultaneously with empirical methods of prediction, i.e. Monte Carlo and bootstrap.

The Sect. 5 concerns the analysis of real data from Polish financial market, i.e. the closing prices of PKN stocks. Moreover we check the efficiency of the proposed estimation and testing methods by using Monte Carlo simulations. Last section contains conclusions.

2 Theory

As it was mentioned, there are strong arguments for using Autoregressive Conditionally Heteroscedastic models in describing financial time series. The main reason for the success of ARCH/GARCH is that they take into account many observed features of the data that classical models cannot capture. Namely, many time series exhibit some well-known characteristics that can be called the ARCH effects. The reader can find in [24] some stylized facts concerning financial data. First, we deal with heteroscedasticity which has been already introduced at the beginning. Another

significant feature is volatility clustering. According to the Mandelbrot [15] that means situation when "the large changes tend to be followed by large changes and small changes tend to be followed by small changes" or in other words when the volatility tends to form clusters. Moreover, many economic series display time varying volatility in such a way that the variance of the process can be modeled by AR or ARMA. Although ARCH/GARCH successfully model all above properties, they however have problems with capturing leptokurticity. More precisely, when data are leptokurtic then we can observe that its returns have fat tails. What is more, this phenomenon is commonly observed in many financial time series concerning prices, rates or returns. In particular, this implies the higher probability for extreme events than in normally distributed variables. This motivated seeking for distributions that would better model the fat-tailed property of the data.

In this section we recall the definitions of ARCH and GARCH models as well as their relation with classical ARMA systems. Moreover we also consider non-Gaussian versions of the classical models.

Definition 1 [6] Process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is called the ARCH process of order p if

$$\begin{aligned} E(\varepsilon_t|\mathcal{F}_{t-1}) &= 0, \\ Var(\varepsilon_t|\mathcal{F}_{t-1}) &= \sigma_t^2, \\ \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 &= \omega + A(L)\varepsilon_t^2, \end{aligned}$$

where $\mathcal{F}_{t-1}$ is the σ-field generated by the $\{\varepsilon_{t-j}, j \geq 1\}$ and in order to ensure the non-negativity of the variance it is assumed that $\omega > 0,\ \alpha_i \geq 0$ for $i = 1, \ldots, p$, $A(z) = \sum_{i=1}^{p} \alpha_i z^i$ and L is a backward shift operator.

The very important property of the ARCH process if the fact that under some restrictions it is weak stationary. Namely, if

$$\alpha_1 + \cdots + \alpha_p < 1,$$

then the ARCH(p) process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is weakly (covariance) stationary, [6]. Moreover it can be shown that in this case the autocovariance function (ACVF) of the ARCH process is given by:

$$\gamma(t, t+h) = cov(\varepsilon_t, \varepsilon_{t+h}) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0, \end{cases} \tag{1}$$

where σ^2 is the unconditional variance of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ given by the following relation:

$$\sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i}.$$

The second important property is related to the fact that there is a strong relation between the ARCH models and ARMA systems, namely:

Remark 1 [24] Let $\{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \text{ARCH}(p)$ with conditional variance at time t given by σ_t^2. After introducing the squared innovations process $\{u_t\}$

$$u_t = \varepsilon_t^2 - E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \varepsilon_t^2 - \sigma_t^2$$

we can obtain the AR representation of $\{\varepsilon_t^2\}$ as:

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + u_t.$$

with noise $\{u_t\}$.

Four years after Engle's [6] introduction of the ARCH process, in 1986 Bollerslev in [2] proposed the Generalized ARCH (GARCH) model as a natural solution to the problem with the high ARCH orders. Although, GARCH exhibits almost the same properties as ARCH model, it has one crucial advantage. Namely, the possibility to significantly reduce the number of parameters in the model. In GARCH model the conditional variance is a linear function of past squared innovations and past conditional variances.

Definition 2 [2] The process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is called the GARCH process of order p and q if

$$E(\varepsilon_t|\mathcal{F}_{t-1}) = 0, \tag{2}$$

$$Var(\varepsilon_t|\mathcal{F}_{t-1}) = \sigma_t^2, \tag{3}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 = \omega + A(L)\varepsilon_t^2 + B(L)\sigma_t^2, \tag{4}$$

where $\mathcal{F}_{t-1}$ is the σ-field generated by the $\{\varepsilon_{t-j}, j \geq 1\}$, $\omega > 0$, $\alpha_i \geq 0$ for $i = 1, \ldots, p$ and $\beta_j \geq 0$ for $j = 1, \ldots, q$. Moreover, $B(z) = \sum_{j=1}^{q} \beta_j z^j$.

Formally, rewriting (4) as

$$\sigma_t^2 = \frac{\omega}{1 - B(1)} + \left(\frac{A}{1-B}\right)(L)\varepsilon_t^2 = \phi_0 + \sum_{i=1}^{\infty} \phi_i \varepsilon_{t-i}^2,$$

we see that (at least at a formal level) a GARCH(p, q) process is in fact an ARCH(∞) process. Similar, as in case of ARCH system, we can show that under similarly restricted parameters, the GARCH model is weakly stationary. Moreover, the variance of GARCH(p, q) model is given by:

$$Var(\varepsilon_t) = \sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{m}(\alpha_i + \beta_i)},$$

where $\sum_{i=1}^{m}(\alpha_i + \beta_i) < 1, (m = max(p, q))$ in order to ensure the non-negativity of the variance. Moreover it can be shown that squared GARCH(p, q) model has ARMA representation, namely if $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a GARCH(p, q) process, then the following relation holds:

$$\begin{aligned}
\varepsilon_t^2 &= \sigma_t^2 + u_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + u_t \\
&= \omega + \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 - \sum_{i=1}^{q} \beta_i(\varepsilon_{t-i}^2 - \sigma_{t-i}^2) + u_t \\
&= \omega + \sum_{i=1}^{m=max(p,q)} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 - \sum_{i=1}^{q} \beta_i u_{t-i} + u_t,
\end{aligned}$$

where $\alpha_i = 0\,\forall_{i>p}$ and $\beta_i = 0\,\forall_{i>q}$ and $u_t = \varepsilon_t^2 - \sigma_t^2$. Because $\{u_t\}$ has mean zero and is serially uncorrelated, and hence is a white noise, therefore the given representation of ε_t^2 is in fact the ARMA(p, q).

Although in the classical definition it is assumed that the standardized residuals $\{z_t\}$ of ARCH/GARCH are Gaussian random variables, in practice it turns out that this assumption weakly corresponds to the reality. Especially financial data exhibit leptokurticity, which is a strong evidence against normality. For example Bollerslev [3] in 1987 suggested the conditional distribution of innovations to be Student's-t instead of normal. On the other hand, Nelson [16] in 1991 proposed to use the Generalized Error distribution (GED) to capture the fat tails usually observed in the distribution of financial times series. In this paper we concentrate on the GED distribution.

Definition 3 (*Generalized Error Distribution—GED*, [8]) The random variable X is GED—distributed, i.e. $X \sim G(\mu, \lambda, \kappa)$ or $X \sim G(\mu, \lambda, a)$ if it has the following probability density function:

$$f(t; \mu, \lambda, \kappa) = \frac{\exp\left(-\frac{1}{2}\left|\frac{x-\mu}{\lambda}\right|^{\frac{1}{\kappa}}\right)}{2^{\kappa+1}\lambda\Gamma(\kappa+1)} \tag{5}$$

or equivalently with parametrization $a = \frac{1}{\kappa}$

$$f(t; \mu, \lambda, a) = \frac{a \exp\left(-\frac{1}{2}\left|\frac{t-\mu}{\lambda}\right|^{a}\right)}{2^{1+\frac{1}{a}}\lambda\Gamma\left(\frac{1}{a}\right)}, \tag{6}$$

which is described by the three parameters

- $\mu \in (-\infty, \infty)$—locates the mode of the distribution (location parameter),
- $\lambda \in (0, \infty)$—defines the dispersion of the distribution (scale parameter),
- κ or $a = \frac{1}{\kappa} \in (0, \infty)$—corresponds to the kurtosis of the distribution (shape parameter).

Remark 2 Let us assume that $X \sim G(\mu, \lambda, \kappa)$ with probability density function f given in (5), then we have [8]:

- mean, median and mode of the distribution is given by μ,

$$\frac{\lambda}{\kappa}\left(\exp\left(\frac{\kappa^2}{2}\right)-1\right) = \mu - a\lambda\left(\exp\left(\frac{1}{2a^2}\right)-1\right),$$

- if we denote the variance of GED distribution as σ^2 then we have the following relationship

$$\lambda = \left(\frac{2^{-2\kappa}\Gamma(\kappa)\sigma^2}{\Gamma(3\kappa)}\right)^{\frac{1}{2}} = \left(\frac{2^{-2\frac{1}{a}}\Gamma(\frac{1}{a})\sigma^2}{\Gamma(3\frac{1}{a})}\right)^{\frac{1}{2}} := \tilde{\lambda}\sigma \tag{7}$$

  which leads to the alternative density representation:

$$f(t;\mu,\lambda,a) = \frac{a\exp\left(-\frac{1}{2}\left|\frac{t-\mu}{\tilde{\lambda}\sigma}\right|^a\right)}{2^{1+\frac{1}{a}}\tilde{\lambda}\sigma\Gamma\left(\frac{1}{a}\right)}, \tag{8}$$

- kurtosis of the GED distribution depends only on the shape parameter and is given by

$$Kurt(X) = \frac{\Gamma(5\kappa)\Gamma(\kappa)}{(\Gamma(3\kappa))^2} = \frac{\Gamma\left(5\frac{1}{a}\right)\Gamma\left(\frac{1}{a}\right)}{\left(\Gamma\left(3\frac{1}{a}\right)\right)^2}, \tag{9}$$

- for $\kappa = \frac{1}{2}$ $(a = 2)$ put in (5) we obtain the normal distribution $\mathcal{N}(\mu, \lambda^2)$ since

$$f(t;\mu,\lambda,a) = \frac{\exp\left(-\frac{1}{2}\left|\frac{x-\mu}{\lambda}\right|^2\right)}{2^{\frac{1}{2}+1}\lambda\Gamma(\frac{1}{2}+1)} = \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\lambda}\right)^2\right)}{\sqrt{2^3}\lambda\frac{1}{2}\sqrt{\pi}}$$
$$= \frac{1}{\lambda\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\lambda}\right)^2\right)$$

- for $\kappa = 1$ $(a = 1)$ we obtain the Laplace or Double Exponential distribution with density

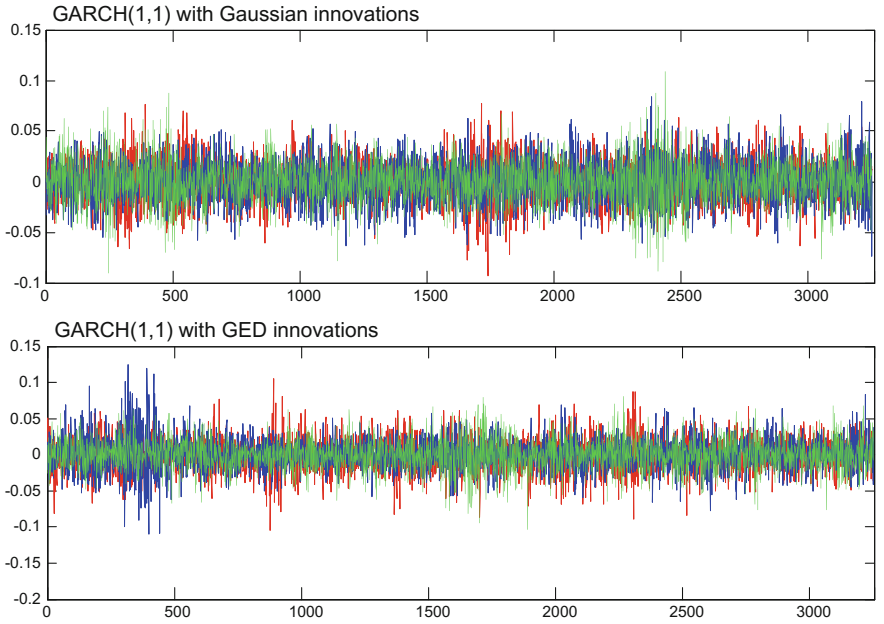


Fig. 1 Comparison of trajectories for GARCH(1, 1) with Gaussian and GED with shape $a = 1.5945$ innovations and parameters $\omega \approx 0$, $\alpha = 0.04991$, $\beta = 0.93224$

$$f(t; \mu, \lambda, a = 1) = \frac{1}{4\lambda} \exp\left(-\frac{1}{2}\left|\frac{x-\mu}{\lambda}\right|\right).$$

In particular, if one wants to consider GED distribution with unit variance and unit shape parameter (so Laplace distribution with unit variance) then

$$\lambda = \left(\frac{2^{-2}\Gamma(1)}{\Gamma(3)}\right)^{\frac{1}{2}} = \sqrt{\frac{\frac{1}{4}}{2}} = \sqrt{\frac{1}{8}} = \frac{1}{2\sqrt{2}}.$$

In Fig. 1 we present GARCH trajectories with various innovations.

3 Testing and Estimation

3.1 Correlation Testing

At the beginning of every time series data analysis, among others, it is always desirable to check if the observed data exhibits relations expressed in the language of correlation function. In this situation the very useful tool is a plot of empirical auto-

correlation function. There are also many statistical tests that can be performed in order to capture autocorrelations. But one should also examine the behavior of the volatility of the data whether is varying in time with tendency to clustering, or in other words, if we are dealing with ARCH effects. It is very important since ignoring existence of these effects may lead to the model misspecification and inappropriate conclusions. That is why in applications we used statistical tests evaluating correlations and ARCH effects.

The first test is an Engle ARCH test (or Engle test, ARCH test) presented by Engle [6]. Note that there is no need to test directly for GARCH effects since if ARCH effects exist, GARCH models can be also considered.

In order to investigate if among data there are statistically significant correlations, we can apply the Ljung-Box test introduced in [14]. We chose this particular test mainly because it has become quite popular in GARCH modeling.

3.1.1 Estimation

In the classical situation, when the variance is constant in time, Ordinary Least Squares is the BLUE estimator (Best Linear Unbiased Estimator) so it is efficient and unbiased. However, under assumption of heteroscedasticity, although the Least Squares (LS) estimator has many advantages like being user-friendly with no optimization requirements, it is relatively less efficient than the estimators based on the likelihood methods. That is why the maximum likelihood methods are the most popular in GARCH parameter estimation. For the Gaussian Maximum Likelihood (Gaussian-ML) estimation we refer the reader to [10]. One should be aware that the assumption of the conditional normality for the $\{\varepsilon_t\}_{t\in\mathbb{Z}} \sim$ GARCH is difficult to justify in many empirical applications and as a matter of fact is not always true. However, since it was often observed that the Gaussian-ML method is robust against the law misspecification we can try to apply it even when the conditional distribution is non-Gaussian. If such situation appears we call it the Quasi or Pseudo Maximum Likelihood method (QML). Namely, the QML estimator is obtained simply by maximizing the likelihood function in the same way as in the ML case although the true probability is not normal. The asymptotic properties (like consistency or asymptotic normality) of QML estimators for ARCH models are established in [23]. In this paper however we consider method based on the GED likelihood function, i.e.

$$L(\theta; \varepsilon_1, \ldots, \varepsilon_n) = \prod_{i=1}^{n} f(\varepsilon_i; \theta) \tag{10}$$

with $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, a, \tilde{\lambda})$ and

$$f(\varepsilon_i; \theta) = \frac{a \exp\left(-\frac{1}{2}\left|\frac{\varepsilon_i}{\tilde{\lambda}\sigma_i}\right|^a\right)}{2^{1+\frac{1}{a}}\tilde{\lambda}\sigma_i \Gamma\left(\frac{1}{a}\right)},$$

where $\sigma_i^2 = \sigma_i^2(\theta) = \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{i-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{i-j}^2$ for GARCH(p, q) and $\sigma_i^2 = \sigma_i^2(\theta) = \omega + \alpha\varepsilon_{i-1}^2 + \beta\sigma_{i-1}^2$ for GARCH$(1, 1)$ process. We can formulate then the following lemma:

Lemma 1 *Assume that innovations* $\{\varepsilon_t\}_{t\in\mathbb{Z}} \sim GARCH(1, 1)$ *are conditionally GED distributed with parameters* μ, $\lambda = \tilde{\lambda}\sigma$ *(see (7)) and* a. *Then the first order conditions for existence of the maximum likelihood estimator of parameter vector* $\theta = (\omega, \alpha, \beta, a, \tilde{\lambda})$ *are given by*

$$\begin{aligned}
\frac{\partial l}{\partial \omega} = 0 &\Leftrightarrow \frac{a}{4}\sum_{i=1}^{n}\left(\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a |\sigma_i|^{-a-2} - \frac{1}{2\sigma_i^2}\right) = 0,\\
\frac{\partial l}{\partial \alpha} = 0 &\Leftrightarrow \frac{a}{4}\sum_{i=1}^{n}\left(\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a |\sigma_i|^{-a-2}\varepsilon_{i-1}^2 - \frac{\varepsilon_{i-1}^2}{2\sigma_i^2}\right) = 0,\\
\frac{\partial l}{\partial \beta} = 0 &\Leftrightarrow \frac{a}{4}\sum_{i=1}^{n}\left(\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a |\sigma_i|^{-a-2}\sigma_{i-1}^2 - \frac{\sigma_{i-1}^2}{2\sigma_i^2}\right) = 0,\\
\frac{\partial l}{\partial a} = 0 &\Leftrightarrow \frac{n}{a} + \frac{n\ln 2}{a^2} + \frac{n\Gamma'(\frac{1}{a})}{a^2\Gamma(\frac{1}{a})} - \frac{1}{2}\sum_{i=1}^{n}\ln\left(\left|\frac{\varepsilon_i}{\tilde{\lambda}|\sigma_i|}\right|\right)\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a |\sigma_i|^a = 0,\\
\frac{\partial l}{\partial \tilde{\lambda}} = 0 &\Leftrightarrow \frac{a}{2}\sum_{i=1}^{n}|\tilde{\lambda}|^{-a-1}|\varepsilon_i|^a|\sigma_i|^{-a} - \frac{n}{\tilde{\lambda}} = 0.
\end{aligned}$$

Proof If we assume GED distribution defined by the probability density function in (8) with location parameter $\mu = 0$, shape parameter a and variance σ_t^2 with $\lambda = \tilde{\lambda}\sigma_t$ then the log-likelihood function l is of the form

$$\begin{aligned}
l(\theta|\varepsilon_1, \ldots, \varepsilon_n) &= \sum_{i=1}^{n}\ln\frac{a\exp\left(-\frac{1}{2}\left|\frac{\varepsilon_i}{\tilde{\lambda}\sigma_i}\right|^a\right)}{2^{1+\frac{1}{a}}\tilde{\lambda}\sigma_i\Gamma\left(\frac{1}{a}\right)}\\
&= \sum_{i=1}^{n}\ln a - \frac{1}{2}\left|\frac{\varepsilon_i}{\tilde{\lambda}\sigma_i}\right|^a - \ln\left(2^{1+\frac{1}{a}}\tilde{\lambda}\sigma_i\Gamma\left(\frac{1}{a}\right)\right)\\
&= \sum_{i=1}^{n} -\frac{1}{2}\left|\frac{\varepsilon_i}{\tilde{\lambda}\sigma_i}\right|^a - \ln\sigma_i\\
&\quad + n\left(\ln a - \left(1+\frac{1}{a}\right)\ln 2 - \ln\left(\Gamma\left(\frac{1}{a}\right)\right) - \ln\tilde{\lambda}\right).
\end{aligned}$$

First we compute the derivatives with respect to ω, α and β. We have

$$\frac{\partial l}{\partial \omega} = \frac{a}{4}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a \left|\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2\right|^{-a/2-1} - \frac{1}{2(\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2)}$$
$$= \frac{a}{4}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a |\sigma_i|^{-a-2} - \frac{1}{2\sigma_i^2},$$

$$\frac{\partial l}{\partial \alpha} = \frac{a}{4}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a \left|\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2\right|^{-a/2-1}\varepsilon_{i-1}^2 - \frac{\varepsilon_{i-1}^2}{2(\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2)}$$
$$= \frac{a}{4}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a |\sigma_i|^{-a-2}\varepsilon_{i-1}^2 - \frac{\varepsilon_{i-1}^2}{2\sigma_i^2}.$$

$$\frac{\partial l}{\partial \beta} = \frac{a}{4}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a \left|\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2\right|^{-a/2-1}\varepsilon_{i-1}^2 - \frac{\sigma_{i-1}^2}{2(\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2)}$$
$$= \frac{a}{4}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a |\sigma_i|^{-a-2}\varepsilon_{i-1}^2 - \frac{\sigma_{i-1}^2}{2\sigma_i^2}.$$

Now we pass to derivatives with respect to a and $\tilde{\lambda}$,

$$\frac{\partial l}{\partial a} = \frac{n}{a} + \frac{n\ln 2}{a^2} + \frac{n\Gamma'(\frac{1}{a})}{a^2\Gamma(\frac{1}{a})} - \frac{1}{2}\sum_{i=1}^{n}\ln\left(\left|\frac{\varepsilon_i}{\tilde{\lambda}|\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2|^{1/2}}\right|\right)\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a \left|\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2\right|^{a/2}$$
$$= \frac{n}{a} + \frac{n\ln 2}{a^2} + \frac{n\Gamma'(\frac{1}{a})}{a^2\Gamma(\frac{1}{a})} - \frac{1}{2}\sum_{i=1}^{n}\ln\left(\left|\frac{\varepsilon_i}{\tilde{\lambda}|\sigma_i|}\right|\right)\left|\frac{\varepsilon_i}{\tilde{\lambda}}\right|^a |\sigma_i|^a,$$

$$\frac{\partial l}{\partial \tilde{\lambda}} = \frac{a}{2}\sum_{i=1}^{n}|\tilde{\lambda}|^{-a-1}|\varepsilon_i|^a|\omega+\alpha\varepsilon_{i-1}^2+\beta\sigma_{i-1}^2|^{-a/2} - \frac{n}{\tilde{\lambda}} = \frac{a}{2}\sum_{i=1}^{n}|\tilde{\lambda}|^{-a-1}|\varepsilon_i|^a|\sigma_i|^{-a} - \frac{n}{\tilde{\lambda}}.$$

□

4 Prediction

Prediction is one of the most important task in volatility modeling. Future value of the financial time series and its conditional volatility can be used among others, in Value at Risk calculations or option pricing as well. Since the squared GARCH model has an ARMA representation, the forecast procedure will be also evaluated

in similar way as in ARMA case. However, in ARCH/GARCH case we allow the conditional variance to vary in time. Of course, we are not particularly interested in forecasting future value of GARCH process itself since, by the definition, we know that $E\varepsilon_t = 0$ for $t = 1, 2, \ldots T$ and for every $T + k$ forecast we obtain $E(\varepsilon_{T+k}) = 0$, $k > 0$, where

$$E(\varepsilon_{T+k}|\mathcal{F}_T) = \begin{cases} E(\varepsilon_{T+1}|\mathcal{F}_T), & k = 1 \\ E(E(\varepsilon_{T+k}|\mathcal{F}_{T+k-1})|\mathcal{F}_T) = E(0|\mathcal{F}_T), & k > 1. \end{cases}$$

But still, we can predict the future value of the squared process $\{\varepsilon_t^2\}_{t\in\mathbb{Z}}$ or equivalently the future conditional variance/volatility of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. Consider the GARCH(1, 1) case with conditional variance given by

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

which has been already estimated in time points $t = 1, \ldots, T$. The time horizon T is the moment when we are making the k-step ahead forecast of future volatility σ_{T+k}^2 with $k > 0$. Let us first denote the conditional expectation as $E\left(\varepsilon_T|\mathcal{F}_T\right) := E_T\varepsilon_T^2$. Our goal is to calculate $E_T\varepsilon_{T+k}^2$. By Ref. [9], we proceed recursively starting with the forecast value at $T + 1$, which is simply

$$E_T\sigma_{T+1}^2 = \omega + \alpha E_T\varepsilon_T^2 + \beta E_T\sigma_T^2 = \omega + \alpha\varepsilon_T^2 + \beta\sigma_T^2.$$

Using the fact that $E_T\varepsilon_{T+1}^2 = E_T\sigma_{T+1}^2$ we can perform the another step, namely the forecast at $T + 2$

$$E_T\sigma_{T+2}^2 = \omega + \alpha E_T\varepsilon_{T+1}^2 + \beta E_T\sigma_{T+1}^2 = \omega + (\alpha + \beta)E_T\sigma_{T+1}^2.$$

After iterating above procedure we obtain the final equation for $T + k$ future variance forecast

$$E_T\sigma_{T+k}^2 = \omega\frac{1 - (\alpha + \beta)^k}{1 - (\alpha + \beta)} + (\alpha + \beta)^{k-1}(\alpha\varepsilon_T^2 + \beta\sigma_T^2).$$

What is more, we observe that in the limit of forecast horizon the conditional variance prediction reduces to the unconditional variance

$$\begin{aligned} \lim_{k\to\infty} E_T\sigma_{T+k}^2 &= \lim_{k\to\infty} \omega\frac{1 - (\alpha + \beta)^k}{1 - (\alpha + \beta)} + (\alpha + \beta)^{k-1}(\alpha\varepsilon_T^2 + \beta\sigma_T^2) \\ &= \omega\frac{1}{1 - (\alpha + \beta)}. \end{aligned}$$

From the above we can also conclude that $\alpha + \beta$ part determines how quickly the variance forecast converges to the unconditional variance. For the empirical methods of forecasting we refer the reader to [5, 17, 18]. In this paper we particularly

are interested in methods based on the bootstrap replications which are more robust in case of the distribution misspecification. We consider $\{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \text{GARCH}(1,1)$ series observed in time points $t = 1, \ldots, T$ with volatility process $\{\sigma_t\}$. Then the $b-$bootstrap replication of $\{\varepsilon_1, \ldots, \varepsilon_T\}$ and $\{\sigma_1, \ldots, \sigma_T\}$ is obtained by the following procedure:

$$\begin{cases} z_1^{*b}, \ldots, z_T^{*b} \sim \hat{F}_{\{\hat{z}_t\}} \\ {\sigma_1^{*b}}^2 = \hat{\sigma}_1^2 \\ \varepsilon_t^{*b} = z_t^{*b}\sigma_t^{*b} & \text{for } t = 1, \ldots, T \\ {\sigma_t^{*b}}^2 = \hat{\omega} + \hat{\alpha}{\varepsilon_{t-1}^{*b}}^2 + \hat{\beta}{\sigma_{t-1}^{*b}}^2 & \text{for } t = 2, \ldots, T \end{cases},$$

where $\hat{F}_{\{\hat{z}_t\}}$ denotes the empirical distribution of estimated standardized residuals $\{\hat{z}_t\}$. Having already this recipe we can proceed to the main algorithm of GARCH(1, 1) k-step ahead forecast:

1. Estimate parameters ω, α, β on the basis of $\{\varepsilon_1, \ldots, \varepsilon_T\} \rightarrow \hat{\omega}, \hat{\alpha}, \hat{\beta}$.
2. Simulate bootstrap replications with respect to $\hat{\omega}, \hat{\alpha}, \hat{\beta} \rightarrow \{\varepsilon_1^{*\,b}, \ldots, \varepsilon_T^{*\,b}\}$, $\{\sigma_1^{*\,b}, \ldots, \sigma_T^{*\,b}\}, b = 1, \ldots, B$.
3. Re-estimate each b-bootstrap replication $\rightarrow \hat{\omega}^{*\,b}, \hat{\alpha}^{*\,b}, \hat{\beta}^{*\,b}, b = 1, \ldots, B$.
4. For every $b-$bootstrap replication simulate k-period future $b-$bootstrap replication with respect to $\hat{\omega}^{*\,b}, \hat{\alpha}^{*\,b}, \hat{\beta}^{*\,b} \rightarrow \{\varepsilon_{T+1}^{*\,b}, \ldots, \varepsilon_{T+k}^{*\,b}\}$, $\{\sigma_{T+1}^{*\,b}, \ldots, \sigma_{T+k}^{*\,b}\}, b = 1, \ldots, B$.
5. Calculate $T + j$ prediction value of the process and volatility as an average of $\{\varepsilon_{T+j}^{*\,1}, \ldots, \varepsilon_{T+j}^{*\,B}\}$ and $\{\sigma_{T+j}^{*\,1}, \ldots, \sigma_{T+j}^{*\,B}\}$ respectively, $j = 1, \ldots, k$.
6. Compute $(1-\gamma)\%$ prediction intervals

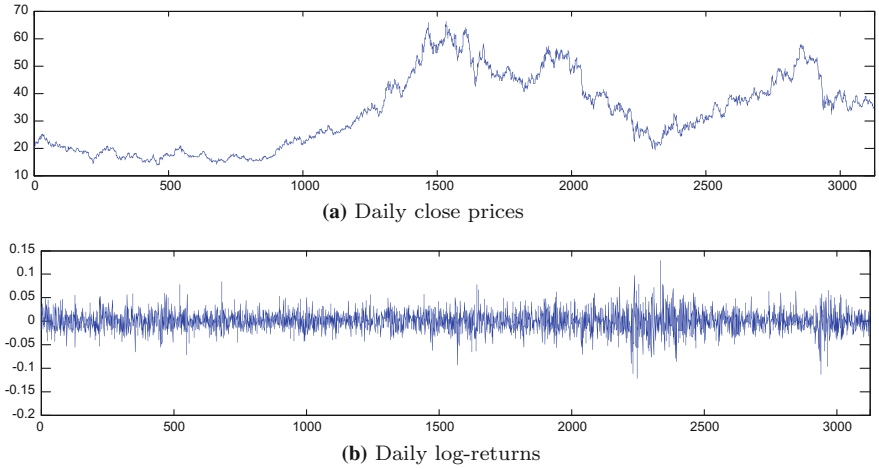
$$[L_{\varepsilon_{T+j}},\ U_{\varepsilon_{T+j}}] = \left[q_{\varepsilon_{T+j}}\left(\frac{1-\alpha}{2}\right),\ q_{\varepsilon_{T+j}}\left(\frac{1+\alpha}{2}\right)\right],$$
$$[L_{\sigma_{T+j},i},\ U_{\sigma_{T+j}}] = \left[q_{\sigma_{T+j}}\left(\frac{1-\alpha}{2}\right),\ q_{\sigma_{T+j}}\left(\frac{1+\alpha}{2}\right)\right],$$

where $q_{\varepsilon_{T+j}}(\gamma)$, $q_{\sigma_{T+j}}(\gamma)$ are empirical γ-quantiles of $\{\varepsilon_{T+j}^{*\,1}, \ldots, \varepsilon_{T+j}^{*\,B}\}$ and $\{\sigma_{T+j}^{*\,1}, \ldots, \sigma_{T+j}^{*\,B}\}$ respectively, $j = 1, \ldots, k$.

We mention, the problem of consistency of bootstrap for dependent and heavy-tailed data is discussed in [13].

5 Applications

In this section we consider real time series from the Polish financial market, i.e. PKN ORLEN (PKN) stocks quoted between 1999-11-26 and 2012-05-11 since the the company has been quoted on the Warsaw Stock Exchange. More precisely, data



(a) Daily close prices

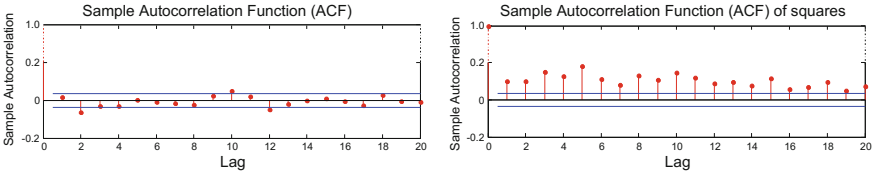
(b) Daily log-returns

Fig. 2 Data taken from the PKN shares with $n = 3125$ observations

consists of $n = 3251$ daily close prices (Fig. 2a). To the analysis we take daily log returns of a given time series—see Fig. 2b. Before we proceed to the more advanced methods, let us first look at the general summary statistics of the calculated log returns R_t of PKN shares. We can see that the average values of our data are very close to zero (mean = 0.00017302, median = 0.00) and its probability distribution function is little bit skewed to the right (skewness = −0.046782) with rather spiky shape (kurtosis = 4.9792).

We can also presume that data cannot be normally distributed since the excess kurtosis is 1.9792 while for normal distribution it should be close zero. Therefore in this case we can not use the classical ARCH/GARCH process introduced by Engle [6] and Bollerslev [2]. On the other hand, if we look at the plot of the daily log-returns (Fig. 2b) we can notice some volatility clustering in particular time periods. Although sample ACF of daily log-returns (Fig. 3a) shows little correlations, the further evidence of volatility clustering can be seen in the sample significant ACF of the squares time series are visible in Fig. 3b. It is worth noting that the similar situation, i.e. exhibition of correlation of squares is observable in ARCH/GARCH autocorrelation function. That is why in order to model the log returns one needs a stochastic process with a nonconstant conditional variance. Fortunately, we know the class of processes that take into account heteroscedasticity. Among others, GARCH models have this property. It is important to investigate whether in our case the GARCH model is a proper choice.

First of all, the Engle test has to be performed in order to check whether log returns R_t exhibits the ARCH effect. We additionally apply the Ljung-Box test as the auxiliary evidence of the autocorrelation presence (but simultaneously keeping in



(a) Log returns **(b)** Squared log returns

Fig. 3 Sample ACF of the PKN daily log returns

Lags	Decision	p-value	Test stat.	critical value
1	1	0	31.986	3.8415
5	1	0	207.55	11.07
10	1	0	259.19	18.307

(a) Engle test

Lags	Decision	p-value	Test stat.	critical value
1	1	0	32.024	3.8415
5	1	0	290.95	11.07
10	1	0	513.55	18.307

(b) Ljung-Box test

Fig. 4 Results of testing autocorrelations for PKN log returns with significance level $\alpha = 0.05$, the Decision = 1 denotes the rejection of H_0 hypothesis (there is no correlation inside the analyzed time series)

mind fact that these tests are mainly applied on the estimated standardized residuals obtained from the fitted model). Results are presented in Fig. 4. It is clearly seen that our data exhibit the ARCH effects in each of the 1–10 lags (all of the p-values are equal to zero and also test statistics exceed the critical values). But this is not enough to determine the GARCH order p, q.

Since we already know that GARCH model is a good choice to describe our data, the next step is to determine the proper distribution of GARCH innovations and p, q orders as well. We perform the maximum likelihood estimation assuming that innovations are GED and Gaussian distributed. We also compare different orders of the GARCH model, i.e. $p = 1, \ldots, 4$ and $q = 0, \ldots, 4$. There is no need to examine higher p, q orders since it may only lead to the unnecessarily complicated models. For each instance we calculate the maximum values for corresponding loglikelihood function (LLF), Akaike Information Criterion (AIC, [1]), Bayesian Information Criterion (BIC, [20]) and Hannan-Quinn Information Criterion (HQIC, [11]). The best model should minimize AIC, BIC, HQIC and maximize LLF. In particular, if we consider the models with equal number of parameters then it is enough to compare only the maximum values of their log-likelihood functions. But, when we deal with models of different orders for p and q then we use the information criterions (AIC, BIC, HQIC) that make adjustment to the log-likelihood function to take into account the number of parameters. The results for selecting ARCH/GARCH orders for PKN daily log-returns are presented in Fig. 5. According to our approach, first we compare maximized LLF for models with the same orders but different distributions. It is easy to notice that for every pair (p, q), $p = 1, 2, 3, 4$ and $q = 0, 1, 2, 3, 4$, the higher values for LLF occur with GED distributed innovations. On the other hand,

LLF

	Gaussian					GED				
p/q	0	1	2	3	4	0	1	2	3	4
1	7547.4	7677.6	7675.2	7673.3	7670.7	11059	11281	11277	11274	11270
2	7566	7675	7675.3	7673.4	7670.8	11087	11277	11278	11275	11271
3	7598	7673.1	7673.4	7673.5	7671.2	11145	11274	11275	11276	11273
4	7611.9	7670.5	7670.8	7670.9	7671.7	11159	11270	11271	11272	11272

AIC

	Gaussian					GED				
p/q	0	1	2	3	4	0	1	2	3	4
1	-15089	-15347	-15340	-15335	-15327	-22111	-22552	-22543	-22534	-22524
2	-15124	-15340	-15339	-15333	-15326	-22163	-22542	-22542	-22534	-22524
3	-15186	-15334	-15333	-15331	-15324	-22277	-22533	-22533	-22534	-22525
4	-15212	-15327	-15326	-15324	-15323	-15296	-15375	-15373	-15371	-15371

BIC

	Gaussian					GED				
p/q	0	1	2	3	4	0	1	2	3	4
1	-15071	-15323	-15310	-15298	-15285	-22086	-22522	-22506	-22492	-22476
2	-15100	-15310	-15302	-15290	-15277	-22133	-22506	-22500	-22485	-22470
3	-15156	-15298	-15291	-15283	-15270	-22241	-22491	-22485	-22479	-22465
4	-15176	-15285	-15277	-15269	-15263	-22262	-22475	-22469	-22464	-22456

HQIC

	Gaussian					GED				
p/q	0	1	2	3	4	0	1	2	3	4
1	-15082	-15339	-15329	-15321	-15312	-22102	-22541	-22530	-22519	-22507
2	-15115	-15329	-15326	-15318	-15308	-22153	-22529	-22527	-22516	-22505
3	-15175	-15321	-15318	-15314	-15305	-22264	-22518	-22516	-22514	-22503
4	-15199	-15312	-15308	-15304	-15302	-22289	-22506	-22504	-22502	-22499

Fig. 5 Information criteria with maximized LLF value

when we search for the minimal value among all models with respect to AIC, BIC and HQIC, then in each case of information criterion we arrive to the GARCH(1, 1) and GED distribution. As a result, for final estimation of PKN daily log-returns we should choose the GARCH(1, 1) process with GED distributed innovations. In the next step we can estimate the parameters of the selected model by using the Maximum Likelihood Estimation (MLE) with loglikelihood function corresponding to the GED distribution (presented in the previous Section). We obtain the following estimates for GARCH parameters ω, α, β and shape parameter a of GED distribution:

$$\begin{aligned}\hat{\omega} &= 8.207805e^{-6}\\ \hat{\alpha} &= 0.04991\\ \hat{\beta} &= 0.93224\\ \hat{a} &= 1.5945\end{aligned}$$

Moreover, in Fig. 6 we present the estimates of innovations of the $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ process and corresponding conditional standard deviations σ_t.

Standardized residuals $\{z_t\}$ have been also evaluated and according to GARCH definition should exhibit no correlations since they are IID random variables. That is

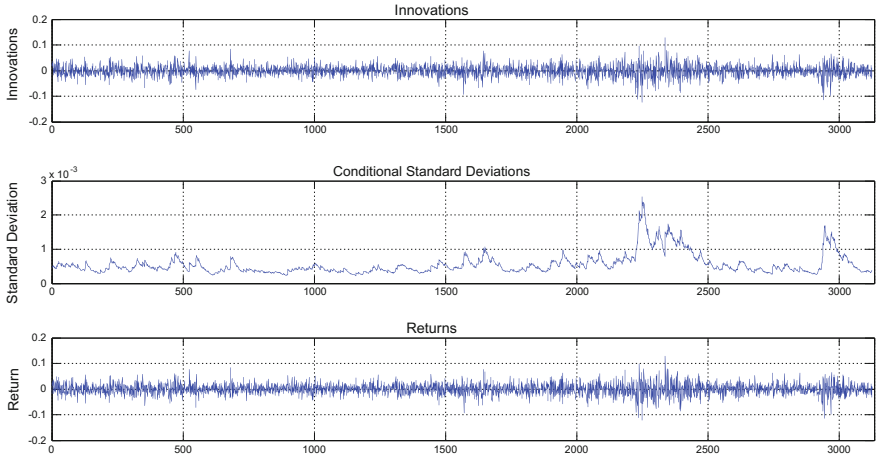
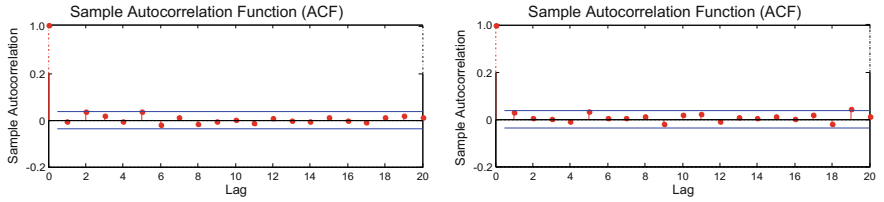


Fig. 6 Estimated innovations and conditional standard deviations together with original PKN daily log returns after Maximum Likelihood estimation

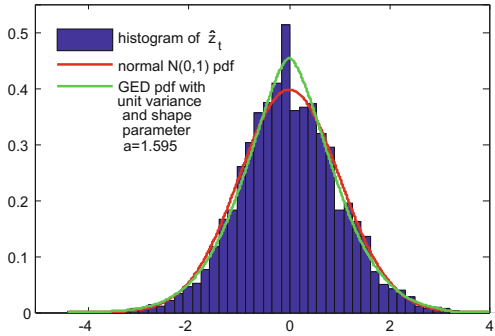


(a) $\{\hat{z}_t\}$ **(b)** $\{\hat{z}_t^2\}$

Fig. 7 Sample ACF of the estimated standardized residuals $\{\hat{z}_t\}$ of PKN daily log returns

why, we plot the ACF of estimated standardized residuals $\hat{z}_t$ (Fig. 7a) and its squares $\hat{z}_t^2$ (Fig. 7b).

However, if the model is correctly specified, then $\{\hat{z}_t\}$ should be not only uncorrelated but also should behave like a GED random variables with zero scale and unit variance in order to satisfy our assumptions. Hence, the histogram has been plotted together with normal $\mathcal{N}(0, 1)$ density plot and GED density with unit variance and estimated shape parameter $a = 1.5945$ (Fig. 8a). We observe that GED with shape parameter a indeed fits better to estimated standardized residuals than Gaussian distribution. Let us check this statement by applying statistical tests. Namely, Jarque-Bera (JB, [12]), Shapiro-Wilk (SW, [21]) and Kolmogorov-Smirnov (KS, [4]) have been performed for evaluating normality of $\hat{z}_t$ (Fig. 8b) and Chi-square (Chi2, [22]) and once again Kolmogorov-Smirnov (KS) for testing GED distribution (Fig. 8c). Results are just as we expected, i.e. all tests reject the null hypothesis stating



(a) Histogram of $\{\hat{z}_t\}$ together with specified theoretical probability density functions

Test	Decision	p-value	Test stat.
JB	1	0.0010	113.7042
KS	1	0.0437	0.0247
SW	1	2.87e-09	0.9946

(b) $H_0 : \hat{z}_t \sim \mathcal{N}(0,1),\ t = 1,\ldots,n$

Test	Decision	p-value	Test stat.
Chi2	0	0.2707	6.3831
KS	0	0.1310	0.0294

(c) $H_0 : \hat{z}_t \sim GED,\ t = 1,\ldots,n$ with unit variance and $a = 1.5945$

Fig. 8 Histogram and results of distribution testing. Decision equal to 1 denotes the rejection of the null hypothesis

normality while in remaining tests there was no evidence to reject the hypothesis evaluating GED distribution. In order to check the testing and estimation procedures for given data we simulate 1000 Monte Carlo (MC) GARCH(1, 1) samples with GED distributed innovations and parameters obtained from the real time series, i.e. $\omega = 8.207805e^{-6}, \alpha = 0.04991, \beta = 0.93224,\ a = 1.5945$. Then, for each MC sample we proceed through the same steps like in case of the original data with PKN daily log-returns, namely:

Step 1 We reestimate each MC sample,

Step 2 We test for existence of correlations among newly estimated MC standardized residuals $\hat{z}_t^{MC}$, $MC = 1,\ldots,1000$ using Ljung-Box and Engle's ARCH test,

Step 3 We test the form of the distribution of $\hat{z}_t^{MC}$, $MC = 1,\ldots,1000$ whether is Gaussian (Jarque-Bera (JB), Shapiro-Wilk (SW), Kolmogorov-Smirnov (KS) test) or GED (Chi-square (Chi2) and Kolmogorov-Smirnov (KS) test).

The results are given respectively for each step:

Ad. Step 1 The estimates of GARCH parameters ω, α, β and also GED distributions shape a can be observed on the boxplots in Fig. 9 while obtained average

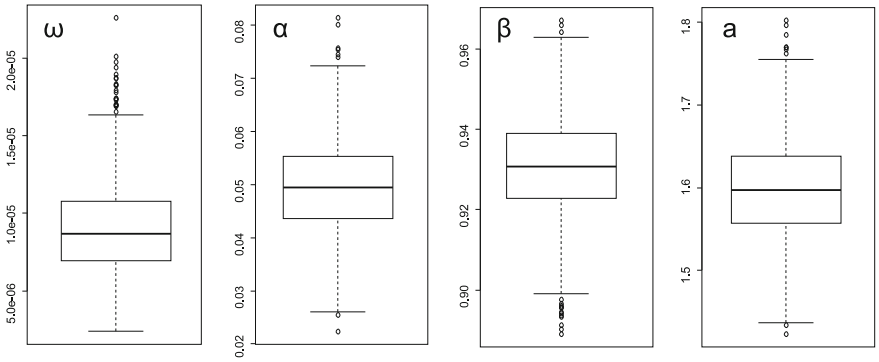


Fig. 9 Boxplots of estimated parameters of ω, α, β, a via maximum likelihood estimation based on 1000 Monte Carlo samples

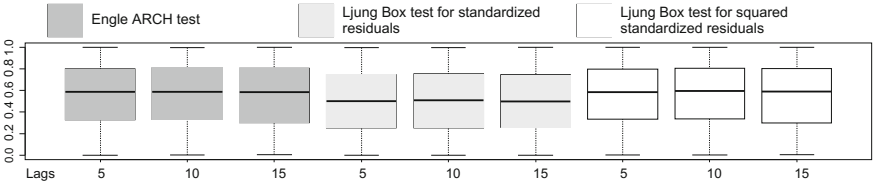
values of estimates are as follows:
$\hat{\omega} = 9.086671e^{-6}$
$\hat{\alpha} = 0.04968$
$\hat{\beta} = 0.9303$
$\hat{a} = 1.5981$

Ad. Step 2 Fig. 10b presents the average results for Engle's ARCH and Ljung-Box test. Additionally, to have overall view, we draw the boxplots of p-values as well (Fig. 10a).

Ad. Step 3 Likewise in Step 2. we plot the boxplots of p-values (Fig. 11a) and give average results of distribution testing (Fig. 11b and c).

To sum up, average maximum likelihood estimates obtained during MC sampling are very similar to those based only on PKN daily log-returns. Moreover, the standardized residuals after MC estimation were uncorrelated since average p-values corresponding to Engle ARCH and Ljung-Box tests exceed or were close to the significance level 0.05, which is needed for accuracy of the model. While evaluating statistical tests we also confirmed the correctness of the chosen GED distribution and its shape parameter. Hence, results for MC estimation have given evidences that assumed model, i.e. GARCH(1, 1) with GED distribution, fits well to our data. Having already that knowledge, we can build model to make a forecast of PKN daily log returns.

All of previous conclusions concerning the choice of the model were made in order to underline the forecast performance. Namely, we intend to predict future values of PKN daily log-returns 100-days ahead using analytical methods, Monte Carlo (MC) or bootstrap as well. The prediction will be based on the 500 MC future samples and also 500 bootstrap future replications. The results we present in Fig. 12.

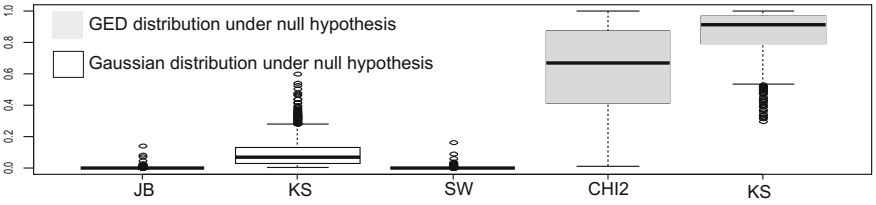


(a) Boxplots of p-values

Test	Lags	Decision	p-value	Test stat.	Critical value
Engle	5	0.0270	0.5582	4.3865	11.0705
	10	0.0370	0.5651	9.0190	18.3070
	15	0.0280	0.5577	13.8948	24.9958
Ljung-Box	5	0.0570	0.4989	5.0457	11.0705
	10	0.0590	0.5033	10.0522	18.3070
	15	0.0590	0.4970	15.1761	24.9958
Ljung-Box for squares	5	0.0310	0.5582	4.3929	11.0705
	10	0.0390	0.5635	9.0578	18.3070
	15	0.0340	0.5560	13.9664	24.9958

(b) Average results

Fig. 10 Average results of Engle ARCH test for $\{\hat{z}_t\}$ and Ljung-Box test for $\{\hat{z}_t\}$ and $\{\hat{z}_t^2\}$ obtained during 1000 MC trials. For each trial, decision equal to 1 denotes the rejection of the null hypothesis.eps



(a) Boxplots of p-values

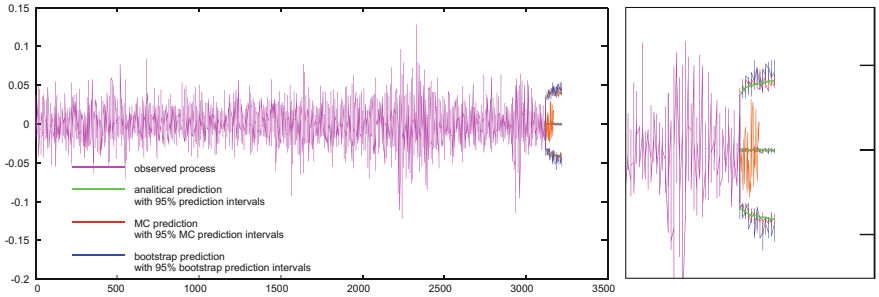
Test	Decision	p-value	Test stat.
JB	0.9980	0.0014	45.4323
KS	0.4390	0.0892	0.0237
SW	0.9970	0.0007	0.9965

(b) $H_0^{MC} : \hat{z}_t^{MC} \sim \mathcal{N}(0,1)$, $MC = 1, \ldots, 1000$, $t = 1, \ldots, n$

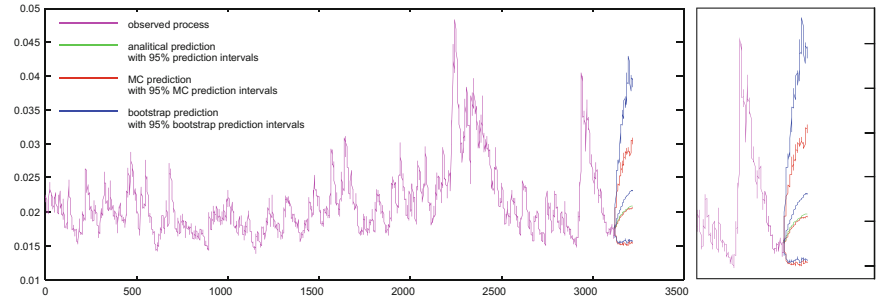
Test	Decision	p-value	Test stat.
Chi2	0.0180	0.6246	5.5337
KS	0.0000	0.8615	0.01028

(c) $H_0^{MC} : \hat{z}_t^{MC} \sim GED$ with unit variance and $a = 1.5945$, $MC = 1, \ldots, 1000$, $t = 1, \ldots, n$

Fig. 11 Average results of distribution testing. Decision equal to 1 denotes the rejection of the null hypothesis H_0^{MC}, $MC = 1, \ldots, 1000$



(a) Prediction of the process



(b) Prediction of the volatility

Fig. 12 100 step ahead prediction of the PKN daily log-returns

6 Conclusions

In this paper we have examined two of the most popular econometric models useful especially in the analysis of data with observable heteroscedastic behavior, namely ARCH and GARCH. We have extended the classical models to the case with non-Gaussian structure, i.e. we have introduced the ARCH/GARCH systems with Generalized Error distribution that is more adequate to real financial time series than classical Gaussian. We have given the main properties of examined systems and proposed testing and estimation procedures. In order to illustrate the theoretical results we have examined real data from Polish financial market in the context of the presented methodology.

References

1. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.
2. Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*(3), 307–327.

3. Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, *69*(3), 542–547.
4. Chakravarti, I. M., Laha R. G., & Roy J. (1967). *Handbook of methods of applied statistics* (Vol. I, pp. 392–394). Wiley
5. Chen, B., Gel, Y., Balakrishna, N., & Abraham, B. (2011). Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes. *Journal of Forecasting*, *30*, 51–71.
6. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrika*, *50*(4), 987–1007.
7. Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, *15*(4), 157–168.
8. Evans, M., Hastings, N., & Peacock B. (2000). *Statistical distributions* (3rd ed., pp. 74–76). Wiley
9. Francq, C., & Zakoian, J.-M. (2010). GARCH models structure. In *Statistical inference and financial applications*. Wiley.
10. Franke, J., Hardle, W. K., & Hafner, C. M. (2011). *Statistics of financial markets: An introduction* (3rd ed.). Springer.
11. Hannan, E. J., & Quin, G. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B*, *41*(2), 190–195.
12. Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review.*, *55*(2), 163–172.
13. Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer series in statistics.
14. Ljung, G. M., & Box, E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.
15. Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.
16. Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.
17. Pascual, L., Romo, J., & Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics and Data Analysis*, *50*(9), 2293–2312.
18. Reeves, J. J. (2005). *Bootstrap Prediction Intervals for ARCH Models*, *21*(2), 237–248.
19. Ritchken, P., & Trevor, R. (1999). Pricing options under generalized GARCH and stochastic volatility processes. *The Journal of Finance*, *54*(1), 377–402.
20. Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.
21. Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611.
22. Snedecor G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed.). Iowa State University Press.
23. Weiss, A. A. (1986). Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory*, *2*(1), 107–131.
24. Zivot, E. (2008). Practical issues in the analysis of univariate GARCH models.

A Residual Based Method for Fitting PAR Models Using Fourier Representation of Periodic Coefficients

Harry Hurd

Abstract In this paper we present a preliminary simulation study of a method for estimating the Fourier coefficients of the periodic parameters of a periodic autoregressive (PAR) sequence. For motivational and comparative purposes, we first examine the estimation of Fourier coefficients of a periodic function added to white noise. The method is based on the numerical minimization of mean squared residuals, and permits the fitting of PAR models when the period T equals the observation size N. For this paper, algorithms and simulations were coded in MATLAB, but an implementation will be available in the **R** package, **perARMA**.

1 Introduction

There exist many natural random processes in which the probability structure has a periodic rhythm, which, in the strict sense means that the probability law is invariant under shifts of length T. To be precise, a process $X_t(\omega) : \Omega \longrightarrow \mathbf{C}$ or $\mathbf{R}$ is called periodically stationary with period T if for every n, collection of times $t_1, t_2, ..., t_n$ in $\mathbf{Z}$ or $\mathbf{R}$, collection of Borel sets $A_1, A_2, ..., A_n$ of $\mathbf{C}$ or $\mathbf{R}$,

$$\begin{aligned} &Pr[X_{t_1+T} \in A_1, X_{t_2+T} \in A_2, ..., X_{t_n+T} \in A_n] \\ &= Pr[X_{t_1} \in A_1, X_{t_2} \in A_2, ..., X_{t_n} \in A_n]. \end{aligned} \tag{1}$$

and there are no smaller values of $T > 0$ for which (1) holds. Synonyms for *periodically stationary* include *periodically non-stationary*, *cyclostationary* (think of cyclically stationary [1]), *processes with periodic structure* [6], and a few others. If $T = 1$, the process is strictly stationary.

When the process is of second order, $X_t \in L_2(\Omega, \mathscr{F}, P)$ with $t \in \mathbf{Z}$, it is called *periodically correlated* [2] (PC), or wide-sense cyclostationary with period T if

H. Hurd (✉)
Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3260, USA
e-mail: hhurd1@nc.rr.com

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_7

$$m(t) = E\{X_t\} = m(t+T)\forall t, \text{ and} \tag{2}$$

$$R(s,t) = E\{X_s\overline{X_t}\} = R(s+T, t+T)\forall s,t \in \mathbf{Z} \tag{3}$$

and there are no smaller values of $T > 0$ for which (2) and (3) hold. If $T = 1$, the process is weakly (or wide-sense) stationary.

A second order stochastic sequence X_t is called PARMA (p, q) with period T if it satisfies, for all $t \in \mathbf{Z}$,

$$\sum_{j=0}^{p} \phi_j(t) X_{t-j} = \sum_{k=1}^{q} \theta_k(t)\xi_{t-k} + \sigma(t)\varepsilon_t \tag{4}$$

where ε_t is a real valued orthogonal process and real parameters, $\phi_j(t) = \phi_j(t+T), \theta_k(t) = \theta_k(t+T)$ and $\sigma(t) = \sigma(t+T)$ for every appropriate j,k. Sometimes we write $\theta_0(t) = \sigma(t)$. Under certain constraints of the parameters, expressed by (8) below, these sequences are PC.

Here we will concentrate on the special case of periodic autoregressive (PAR) sequences, for which

$$\sum_{j=0}^{p} \phi_j(t) X_{t-j} = \sigma(t)\varepsilon_t \quad t \in \mathbf{Z}, \tag{5}$$

where ε_t is an orthogonal process, $\phi_0(t) \equiv 1$, $\phi_j(t) = \phi_j(t+T)$, and $\sigma(t) = \sigma(t+T)$ for every appropriate j. Although Pagano [8] initiated the recent notation and stream of effort on PAR sequences, it is clear that Hannan [3] was aware of them.

Essential information may be obtained by blocking X_t into vectors $\mathbf{X}_n$ of length T as prescribed by Gladyshev [2]; then (5) becomes

$$L\mathbf{X}_n - \sum_{j=1}^{p'} U_j \mathbf{X}_{n-j} = \Gamma\varepsilon_n \tag{6}$$

where L has the form

$$L = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -\phi_1(1) & 1 & 0 & \dots & 0 \\ -\phi_2(2) & -\phi_1(2) & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\phi_{T-1}(T-1) & -\phi_{T-2}(T-1) & -\phi_{T-3}(T-1) & \dots & 1 \end{bmatrix}, \tag{7}$$

$$[U_j]_{nn'} = \phi_{jT+n-n'}(n),$$

and $\varepsilon_n = [\varepsilon_{nT}, \varepsilon_{nT+1}, \dots \varepsilon_{nT+T-1}]'$. The matrix Γ is similarly arranged as L except the diagonal is $\{\sigma(0), \sigma(1), \dots, \sigma(T-1)\}$ and the condition for X_t to be PC is identical to the condition for the vector sequence $\mathbf{X}_n$ to be stationary, namely that

$$\det\left[L - \sum_{j=1}^{p'} U_j \lambda^j\right] \neq 0, \quad |\lambda| \leq 1 \tag{8}$$

The condition (8) was expressed first by Pagano [8] for PAR, and then by Vecchia [9] for general PARMA.

Of course the vector sequence $\mathbf{X}_n$ could also be modeled by a vector AR, (VAR) model, but we note that the number of real autoregressive parameters for a general VAR(p) is on the order of pT^2 because the autoregressive coefficients are $T \times T$ matrices. But for PAR(p) the number is on the order of pT, which can still be sizable when compared to the total length of the series available. See Pagano [8, p. 1316]. For a full PARMA given by (4) the parameter count is seen to be $(p + q + 1)T$. An alternative parameterization of a PARMA system (see Jones and Breslford [6]) can sometimes substantially reduce the number of parameters via representing the periodically varying parameters by Fourier series. In the case of PAR we have

$$\phi_j^{A_j}(t) = a_{j,1} + \sum_{n=1}^{[T/2]} a_{j,2n} \cos(2\pi nt/T) + a_{j,2n+1} \sin(2\pi nt/T) \tag{9}$$

for $t = 0, 1, \ldots, T-1, j = 1, \ldots, p$. The inverse for the $a_{j,n}$ coefficients is given by

$$\begin{aligned} a_{j,1} &= \frac{1}{T} \sum_{t=0}^{T-1} \phi_j(t) \\ a_{j,2n} &= \frac{2}{T} \sum_{t=0}^{T-1} \phi_j(t) \cos(2\pi nt/T) \\ a_{j,2n+1} &= \frac{2}{T} \sum_{t=0}^{T-1} \phi_j(t) \sin(2\pi nt/T) \end{aligned} \tag{10}$$

for $n = 2, \ldots, [T/2], j = 1, \ldots, p$. We also denote

$$A_j = \{a_{j,1}, a_{j,2}, a_{j,3}, \ldots, a_{j,2*[T/2]+1}\}' \tag{11}$$

to be a column vector.

When estimating the natural PAR coefficients $\{\phi_j(t), j = 1, 2, \ldots, p, t = 0, 1, \ldots, T-1$ or their Fourier coefficients, $\{A_j, j = 1, 2, \ldots, p\}$, there is always an issue of the length of the sample N relative to the period T. The two important cases are (1) $N>>T$ and (2) $N = T$. In the case $N>>T$, the usual method for estimating the coefficients is through the Yule-Walker equations and the existence of multiple periods allows the sample covariance to be estimated and used to solve for the unknown coefficients. In the case of $N = T$, although the covariance of the sequence cannot be estimated in the usual way, the Fourier coefficients can still be successfully estimated via *ordinary least squares* (OLS) when the number of coefficients is small relative to N. In this note,

for the purpose of background, we will briefly review the usual method for $N>>T$ and then, for $N = T$, present a simulation study that illustrates the effectiveness of the OLS method. We also include, for the purpose of motivation and comparison, results from application of the method to the estimation of Fourier coefficients of a periodic function added to white noise.

The application of this idea to full PARMA models is of interest but not so straightforward because of the way that the moving average parameters appear in (4). Approaches to this problem are currently under study.

2 Determination of PAR Coefficients by Yule Walker Method

The Yule-Walker method, which is based on minimizing mean square error of a linear predictor, gives an important way for finding the coefficients $\{\phi_j(t) = \phi_j(t+T), j = 1, \dots, p\}$.

For some fixed t, the linear predictor of X_t, based on $\{X_{t-p}, \dots, X_{t-1}\}$, that minimizes the MS error is given by the orthogonal projection of X_t onto $\mathscr{M}(t-1;p) = \mathrm{sp}\{X_s, s \in \{t-p, \dots, t-1\}\}$. We denote

$$\widehat{X}_{t,t-1;p} = (X_t | \mathscr{M}(t-1;p)), \quad \text{and set} \quad \widehat{X}_{t,t-1;0} = 0.$$

Specializing to *real* sequences we then need to determine the coefficients $\alpha_{j,p}^{(t)}$ in

$$\widehat{X}_{t,t-1;p} = \sum_{j=1}^{p} \alpha_{j,p}^{(t)} X_{t-j}.$$

The normal equations arising from the orthogonal projection are

$$E\{[X_t - \widehat{X}_{t,t-1;p}]X_s\} = 0, \quad s = t-p, \dots, t-1$$

or in matrix form

$$\begin{bmatrix} R(t,t-1) \\ \vdots \\ R(t,t-p) \end{bmatrix} = \begin{bmatrix} R(t-1,t-1) & \cdots & R(t-p,t-1) \\ R(t-1,t-2) & \cdots & R(t-p,t-2) \\ \vdots & \vdots & \vdots \\ R(t,t-p) & \cdots & R(t-p,t-p) \end{bmatrix} \begin{bmatrix} \alpha_{1,p}^{(t)} \\ \alpha_{2,p}^{(t)} \\ \vdots \\ \alpha_{p,p}^{(t)} \end{bmatrix}$$

and in a shorter notation

$$\mathbf{r}_{t,t-1:t-p} = \mathbf{R}_{t-1,p} \boldsymbol{\alpha}_p^{(t)}. \tag{12}$$

Any $\boldsymbol{\alpha}_p^{(t)} = [\alpha_{1,p}^{(t)} \alpha_{2,p}^{(t)} \ldots \alpha_{p,p}^{(t)}]'$ that solves (12) (the normal equations) implements the projection. If $\mathbf{R}_{t-1,p}$ is invertible, the solution is unique but if not, then any pseudo-inverse still yields a predictor that minimizes MS error. Other results using this notation may be found in [4, 5].

Since for PC-T processes the covariances are invariant (see (3)) under shifts of length T, then the prediction coefficients will be periodic in t with period T. So for a sample of length KT, there are multiple occurrences (order of K) of products $X_{t_1+kT} X_{t_2+kT}$ from which we estimate the covariances $\mathbf{r}_{t,t-1:t-p}$ and $\mathbf{R}_{t-1,p}$ appearing in (12). Specifically,

$$\widehat{R}_K(t_1, t_2) = \frac{1}{K} \sum_{k=0}^{K-1} [X_{t_1+kT} - \widehat{m}_{t_1,N}][X_{t_2+kT} - \widehat{m}_{t_2,N}] \tag{13}$$

is the estimator for some entry $R(t_1, t_2)$.

Then the estimator $\widehat{\boldsymbol{\alpha}}_p^{(t)}$ is obtained by solving $\widehat{\mathbf{r}}_{t,t-1:t-p} = \widehat{\mathbf{R}}_{t-1,p} \widehat{\boldsymbol{\alpha}}_p^{(t)}$ and estimates for the Fourier coefficients A_j are obtained via (11). But to estimate $R(t_1, t_2)$ in this manner requires K to be of nontrivial size, some say at least 40. Here we seek a methodology to estimate the A_j when the number of periods K available in the sample is small, say $K = 1$.

3 OLS Fit for Periodic Function with Additive Noise

In order to develop some intuition for the PAR estimation problem, we first examine the simpler case of estimating the Fourier coefficients of a periodic function added to white noise. Given a trajectory of observations $\{X_0, X_2, \ldots, X_{N-1}\}$, we wish to minimize

$$Q(\mathbf{A}) = \sum_{t=1}^{N} \left[X_t - \left\{ a_1 + \sum_{n=1}^{[T/2]} a_{2n} \cos(2\pi nt/T) + a_{2n+1} \sin(2\pi nt/T) \right\} \right]^2 \tag{14}$$

where $\{a_1, a_2, \ldots\} = \mathbf{A}$. Although there is a closed form solution due to the mutual orthogonality of the set of sines and cosines, we do the minimization numerically to prepare for the application to PAR, for which there is no closed form solution.

To see the idea in very simple example, suppose we wish to fit just the $\cos(2\pi t/T)$ term to X_t using ordinary least squares (OLS). The OLS estimate for a_2 is well known to be $\hat{a}_2 = \frac{2}{T} \sum_{t=0}^{T-1} X_t \cos(2\pi t/T)$ and more generally

$$\hat{a}_{2n} = \frac{2}{T} \sum_{t=0}^{T-1} X_t \cos(2\pi nt/T) \quad \hat{a}_{2n+1} = \frac{2}{T} \sum_{t=0}^{T-1} X_t \sin(2\pi nt/T). \tag{15}$$

If $X_t = \zeta_t + f_t$ for $t \in \{0, 1, \ldots, N-1\}$, where ζ_t is Gaussian white noise with zero mean and variance σ^2_{noise}, and $f_t = A\cos(2\pi t/T)$, with $T = N$, then it is easy to see that $\hat{a}_{2n}$ and $\hat{a}_{2n+1}$ are Gaussian and that

$$E\{\hat{a}_2\} = A, \qquad E\{\hat{a}_3\} = 0, \qquad \text{Var}\ \{\hat{a}_2\} = \text{Var}\ \{\hat{a}_3\} = \frac{\sigma^2_{noise}}{T} \tag{16}$$

and

$$\begin{aligned} E\{\hat{a}_2\hat{a}_3\} &= \frac{4}{T^2}\sum_{t=0}^{T-1}\sum_{t'=0}^{T-1}\cos(2\pi t/T)\sin(2\pi t'/T)E\{Z_t Z_{t'}\} \\ &= \frac{4}{T^2}\sum_{t=0}^{T-1}\cos(2\pi t/T)\sin(2\pi t/T) = 0. \end{aligned} \tag{17}$$

Simulated series of X_t for $T = N = 1024$, $\sigma_{noise} = 1$ and $A = 0, 0.2, 0.4, 1.0$ were produced; note the coefficients of the Fourier series in (14) are therefore $a_1 = 0, a_2 = A$ and $a_j = 0, j \geq 3$.

These simulated series were processed by a MATLAB script that implements the minimization with respect to $\mathbf{A}$ of $Q(\mathbf{A})$, where the single trajectory $\{X_0, X_2, \ldots, X_{N-1}\}$ is treated as fixed. Figure 1 (top panel) shows the signal $f_t = A\cos(2\pi t/T)$ with $A = 1$ in red, the sum X_t in blue and the estimated signal $\hat{f}_t = \hat{A}\cos(2\pi t/T)$ in green. Although both are present, the difference between the red and green curves is nearly imperceptible on the scale used.

The middle panel of Fig. 1 is the residual Z_t of the OLS fit and the bottom panel is the sample Fourier transform (computed via FFT) of the residual, showing no clear residual periodic component.

Some questions that we can address by simulation: (1) Sample distribution of parameter estimates; (2) Variance of estimates as function of $N = T$; (3) Variance of estimates as function of number of frequencies searched.

Figures 2 and 3 are the sample histograms of $\hat{a}_2$ and $\hat{a}_3$ when the true values are $a_2 = 0.2$ and $a_3 = 0.0$; in each, $\sigma_{noise} = 1$, $N = T = 4096$. These histograms were produced by $NSAMP = 500$ replicates of the simulation-estimation process. In both of these histograms the p-value of a Lilliefors test for normality were both ≥ 0.5, indicating no evidence for rejection of normality. The Lilliefors test is a Kolmogorov-Smirnov type of test for normality in which the null is normal with parameters estimated from the data [7]; thus a large p_L indicates the normality of the sample distribution cannot be rejected.

The sample variances were 0.023 and 0.022, whereas the values computed via (16) were 0.0221.

For each parameter, the empirical dependence of $\hat{\sigma}$ on $N = T$ can be seen by the least squares fit of a straight line, $y = mx + b$, to the pairs $(N, \hat{\sigma})$, where both N and $\hat{\sigma}$ are transformed to a log scale, so the expected $T^{-1/2}$ dependence becomes $m = -1/2$.

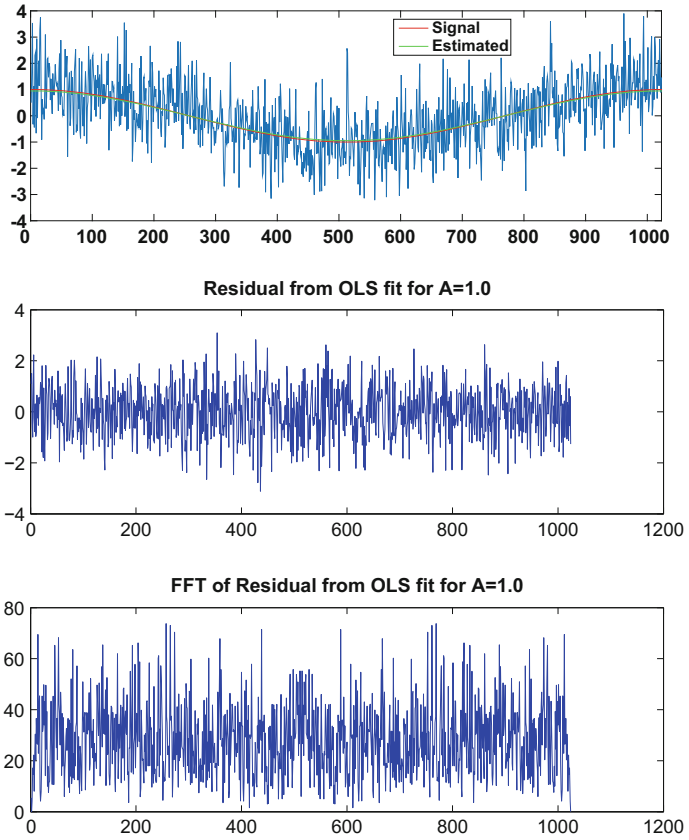


Fig. 1 *Top panel red* is true signal $f_t = A\cos(2\pi t/T)$ with $A = 1$, *blue* is signal plus noise, *green* is estimated signal where $\hat{A} = 0.947$ is determined by minimizing $Q(\mathbf{A})$ in (14). *Middle panel* is residual series $Z_t = X_t - \hat{A}\cos(2\pi t/T)$; *bottom* is FFT of Z_t

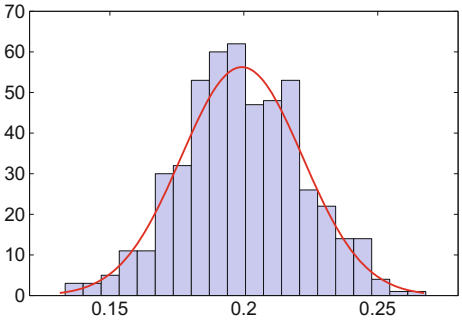


Fig. 2 Estimates $\hat{a}_2$ with $a_2 = 0.2$, $N = T = 4096$, $\sigma_{noise} = 1$, $\hat{\mu}_2 = 0.199$, $\hat{\sigma}_2 = 0.023$

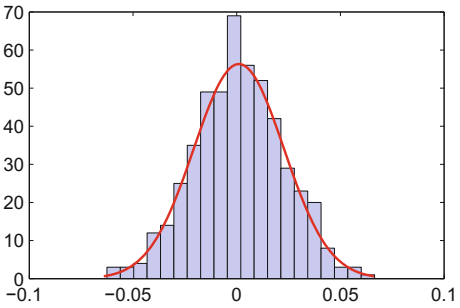


Fig. 3 Estimates $\hat{a}_3$ with $a_2 = 0$, $N = T = 4096$, $\sigma_{noise} = 1$, $\hat{\mu}_3 = 0.001$ $\hat{\sigma}_3 = 0.022$

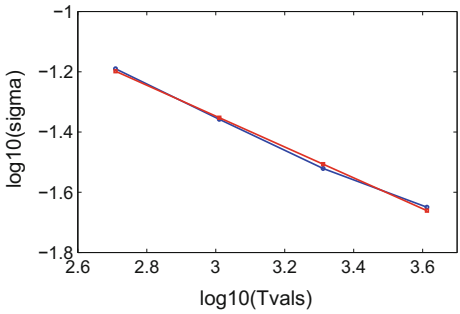


Fig. 4 Estimating a_2 where $\sigma_{noise} = 1$, $a_2 = A = 0.4$, slope $= -0.512$

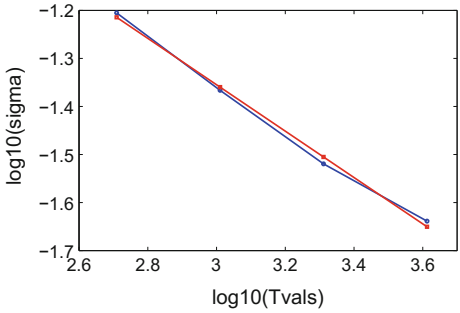


Fig. 5 Estimating a_3 where $\sigma_{noise} = 1$, $a_2 = A = 0.4$, slope $= -0.483$

Figures 4 and 5 illustrate this fitting for parameters $\{a_2, a_3\}$, where for each parameter the values $A = 0.4$, $N = \{512, 1024, 2048, 4096\}$ are used. The resulting slope estimates are $m = \{-0.512, -0.483\}$, where blue lines connect the observed data and the red lines are the least squares fit to the pairs $(N, \hat{\sigma})$.

In order to show the variability of parameter estimates when parameter values are zero, we set parameters a_1 through a_{17} to be active whereas only $a_2 = A = 0.4$ was nonzero. Figure 6 shows the boxplots, based on 500 replicates, of all 17 parameters estimated. Box vertical boundaries are 25th and 75th percentiles and red line is median. The ability to perceive non-nullity of parameters is visually clear. In the next section we use the t-test for testing for this non-nullity.

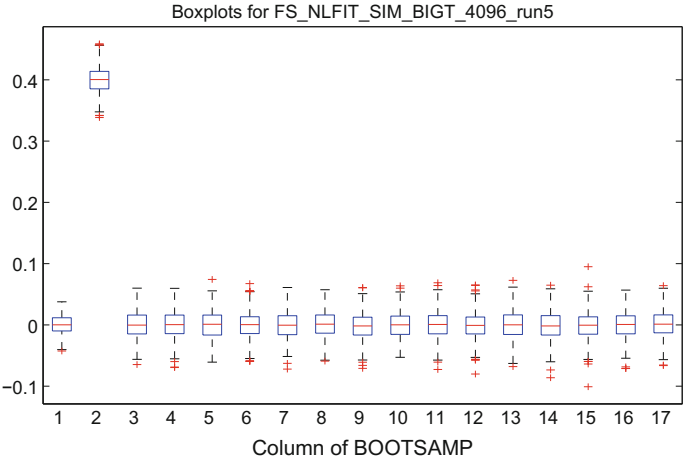


Fig. 6 Active parameters are $\{a_1, a_2, \ldots, a_{17}\}$ with $A = 0.4$, $N = 4096$. Boxplots of parameter estimates are based on 500 replicates

4 OLS fit of a Fourier Series Parametrization of a PAR Model

For a PAR model as in (5), we designate the following estimating procedure as `parmsef`. First we minimize the objective function

$$Q(\mathbf{A}) = \frac{1}{N-p} \sum_{t=p+1}^{N} \left[X_t - \sum_{j=1}^{p} \phi_j^{\mathbf{A}}(t) X_{t-j} \right]^2 \tag{18}$$

where

$$\phi_j^{\mathbf{A}}(t) = a_{j,1} + \sum_{n=1}^{[T/2]} a_{j,2} \cos(2\pi nt/T) + a_{j,2n+1} \sin(2\pi nt/T)$$

and we set $N = T$.

The OLS estimate of $\mathbf{A}$ is the value of $\mathbf{A}$ that minimizes $Q(\mathbf{A})$ for $\mathbf{A} \in \mathbf{S}_1$, the parameter search space, defined as $\mathbf{S}_1 = \text{sp}\{e_j \in \mathbf{R}^N : j \in I_{\mathbf{A}}\}$ where $I_{\mathbf{A}}$ is the set of indexes identifying the active $\mathbf{A}$ parameters. If $\mathbf{A}^*$ minimizes the mean square residuals Q, then denoting $\hat{Z}_t$ as the residual sequence $\hat{Z}_t = X_t - \sum_{j=1}^{p} \phi_j^{A_j^*}(t) X_{t-j}$, we then determine the OLS estimate of $\sigma(t)$ by minimization of

$$Q_\sigma(\mathbf{B}) = \frac{1}{N-p} \sum_{t=p+1}^{N} \left[\hat{Z}_t^2 - \sigma_B^2(t) \right]^2, \tag{19}$$

using

$$\sigma_B(t) = \sum_{n_j \in J_B} b_{2n_j} \cos(2\pi n_j t/T) + b_{2n_j+1} \sin(2\pi n_j t/T). \tag{20}$$

The minimization is with respect to the active parameters b_{n_j} in the collection $B = \{b_0, b_1, b_2, b_3, \ldots, b_{2*[T/2]+1}\}$.

To exercise this program we simulated a PAR(2) by specifying the coefficients $\phi_1(t)$ and $\phi_2(t)$, shown in Fig. 7 to be

$$\phi_1(t) = 1.1 + 0.6 \cos 2\pi t/4096 \tag{21}$$

$$\phi_2(t) = -0.3450 - 0.33 \cos 2\pi t/4096 - 0.045 \cos 4\pi t/4096 \tag{22}$$

and $\sigma_B(t) = 1$.

Figure 8 is a simulated series of $N = T = 4096$ samples using $\phi_1(t)$ and $\phi_2(t)$ given in (22) and illustrated in Fig. 7. Note the higher amplitudes and lower frequency fluctuations at the beginning and end of the series in comparison to the middle section.

In the first experiment with the `parmsef` algorithm we set the seven autoregressive parameters shown in Table 1 to be active where the true values, the sample mean and standard deviations and the Lilliefors p-value, p_L, are also given in the table.

The sample distributions for all seven of the estimated a_{jk} parameters are found to be consistent with the normal; six of these are shown in Fig. 9a–f. Additionally, but not shown here, the sample distribution for the first few b parameters from (20)

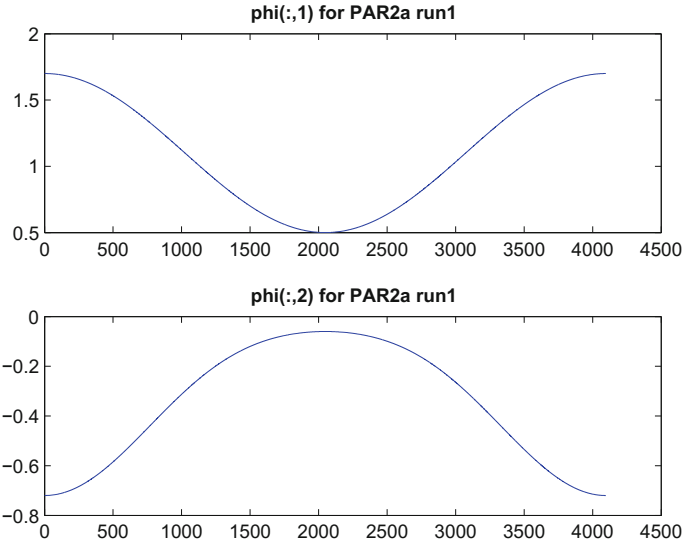


Fig. 7 $\phi_1(t)$ and $\phi_2(t)$ given in (22) for PAR2a run1, $N = 4096$

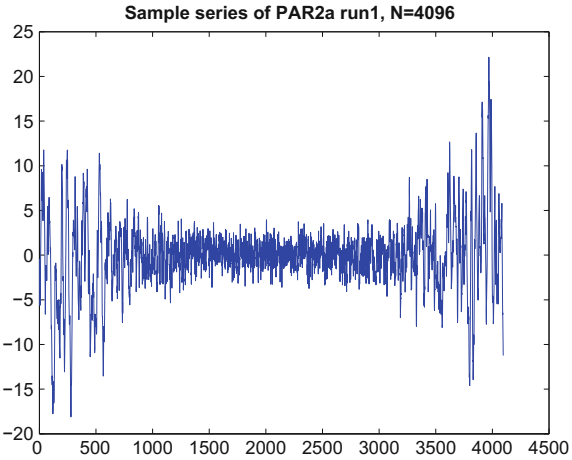


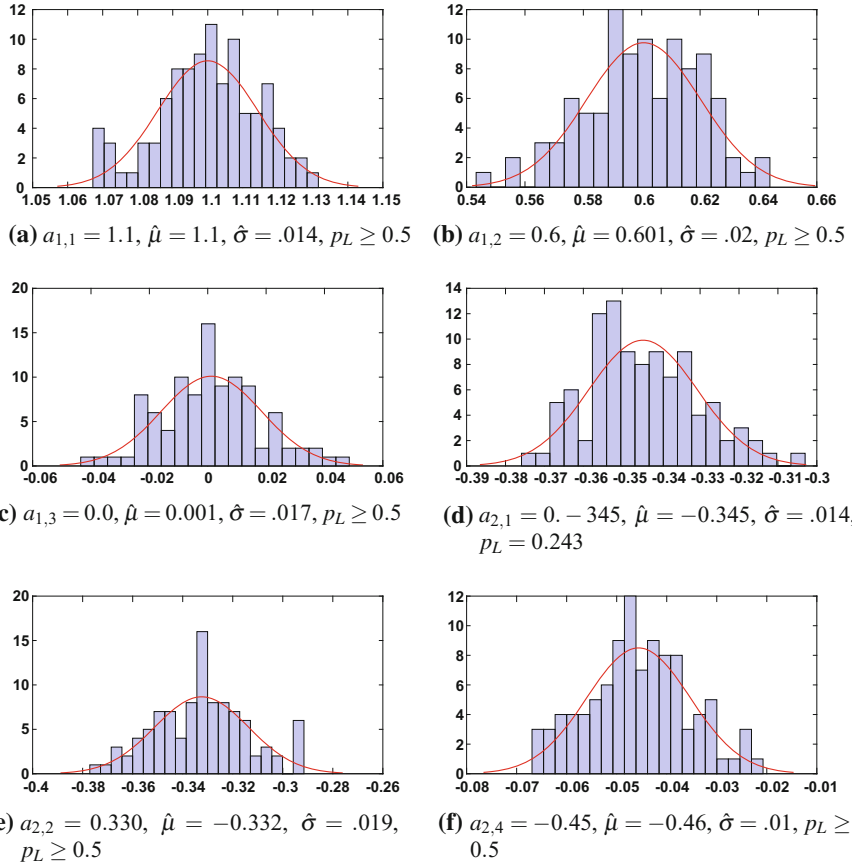
Fig. 8 Sample series using $\phi_1(t)$ and $\phi_2(t)$ given in (22) and illustrated in this figure with $N = 4096$

Table 1 Sample μ and σ with $N = T = 4096$ for selected Fourier coefficients estimated by `parmsef`. Estimates are based on $NSAMP = 100$ replicates

Parameter	Value	$\hat{\mu}$	$\hat{\sigma}$	p_L	t score	p-value for $\mu = 0$
$a_{1,1}$	1.1	1.1	0.014	≥0.5	785.7	0.0
$a_{1,2}$	0.6	0.601	0.02	≥0.5	300.5	0.0
$a_{1,3}$	0.0	0.001	0.017	≥0.5	0.588	0.721
$a_{2,1}$	−0.345	−0.345	0.014	0.243	−246.4	0.0
$a_{2,2}$	−0.330	−0.332	0.019	≥0.5	−174.7	0.0
$a_{2,3}$	0.0	−0.001	0.019	≥0.5	0.526	0.701
$a_{2,4}$	−0.45	−0.46	0.010	≥0.5	−460	0.0

are consistent with normal and with sample variances similar to those of the a parameters. Finally, for the estimates of each parameter, we include the t-score and the p-value of the t-test for $\mu = 0$ based on $NSAMP = 100$ replicates. Although these tests correctly differentiate the null from the nonnull parameters, we note that in the usual time series analysis, there is only one sample available on which one can base a test.

As in Figs. 4 and 5, for each estimated parameter the dependence of $\hat{\sigma}$ on the series length $T = N$ can be seen by fitting a straight line to the $(N, \hat{\sigma})$ as we did in Sect. 3 for the OLS fit to a periodic function with additive noise. Values of $\hat{\sigma}$ were determined for $N = T = (512, 1024, 2048, 4096)$, and this fitting is illustrated in Fig. 10a, b for parameters a_{11} and a_{14}, producing values $m = -0.524, -0.508$ in the two cases; the observed data are in blue and and the red line is the result of the OLS straight line fit. The empirical dependence on N is slightly steeper than the expected $m = -1/2$.



(a) $a_{1,1} = 1.1$, $\hat{\mu} = 1.1$, $\hat{\sigma} = .014$, $p_L \geq 0.5$

(b) $a_{1,2} = 0.6$, $\hat{\mu} = 0.601$, $\hat{\sigma} = .02$, $p_L \geq 0.5$

(c) $a_{1,3} = 0.0$, $\hat{\mu} = 0.001$, $\hat{\sigma} = .017$, $p_L \geq 0.5$

(d) $a_{2,1} = 0. - 345$, $\hat{\mu} = -0.345$, $\hat{\sigma} = .014$, $p_L = 0.243$

(e) $a_{2,2} = 0.330$, $\hat{\mu} = -0.332$, $\hat{\sigma} = .019$, $p_L \geq 0.5$

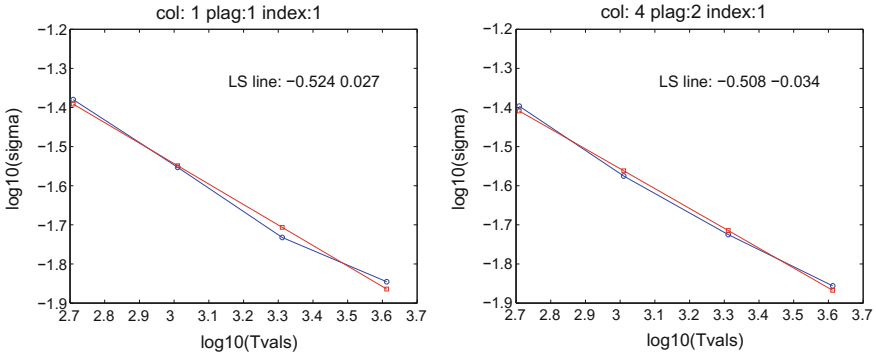
(f) $a_{2,4} = -0.45$, $\hat{\mu} = -0.46$, $\hat{\sigma} = .01$, $p_L \geq 0.5$

Fig. 9 Sample distributions of parameter estimates with $N = T = 4096$ for selected Fourier coefficients estimated by `parmsef`. Estimates are based on $NSAMP = 100$ replicates

As a check that the fit has successfully explained the correlation structure in the simulated series, the empirical ACF and PACF were computed for the residual $\hat{Z}_t$ of the fit, resulting in the plots of Fig. 11. Both ACF and PACF are consistent with uncorrelated noise.

Finally, to again see the effect of more coefficients with null values we made a run in which parameters $\{a_{11}, \ldots, a_{19}, a_{21}, \ldots, a_{28}\}$ (a total of 17) were estimated, although only 5 had nonzero true values. Figure 12 illustrates the ability to visually perceive the non zero values among the 17 from a sample of 100 simulations.

For each parameter, Table 2 presents true values and estimated means and standard deviations; in addition, the p-value of the Lilliefors test, p_L, t-scores and p-values for t-tests for $\mu = 0$ are given. As in Table 1, the t-tests correctly differentiate the null from the nonnull parameters, but the more important issue, not addressed



(a) $a_{1,1}, \hat{m} = -.524$ **(b)** $a_{1,4}, \hat{m} = -.508$

Fig. 10 OLS fit (*red*) of $y = mx + b$ to $\hat{\sigma}$ (*blue*) as function of $N = T = (512, 1024, 2048, 4096)$ using log10 scales

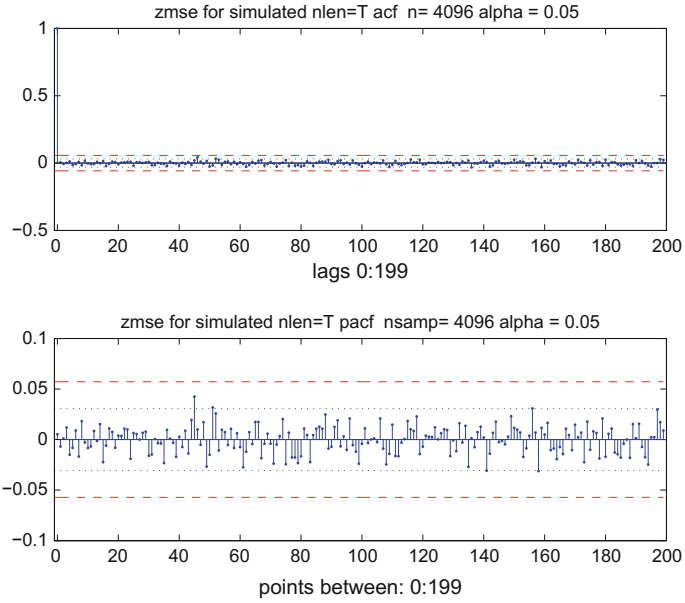


Fig. 11 PAR2a run1 $N = 4096$, ACF, PACF of PARMSEF residuals from one realization

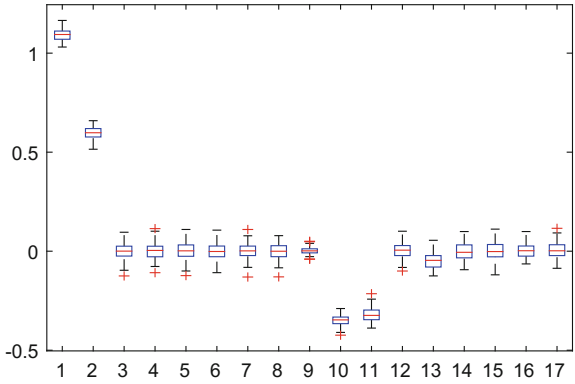


Fig. 12 Boxplots for PAR2a run1, $N = T = 1024$. Active parameters are $\{a_{11}, \ldots, a_{19}, a_{21}, \ldots, a_{28}\}$ (17 total), all null except $a_{11} = 1.1, a_{12} = 0.6, a_{21} = -0.345, a_{22} = -0.33, a_{24} = -0.045$. Boxplots of parameter estimates are based on 100 replicates

Table 2 Sample μ and σ for 17 selected Fourier coefficients estimated by `parmsef`. *NSAMP* = 100, $N = T = 1024$

Parameter	Value	$\hat{\mu}$	$\hat{\sigma}$	p_L	t score	pv for $\mu = 0$
a_{11}	1.10	1.09	0.03	≥0.50	382.90	0.00
a_{12}	0.60	0.60	0.03	≥0.50	189.79	0.00
a_{13}	0.00	0.00	0.04	≥0.50	−0.77	0.44
a_{14}	0.00	0.00	0.04	≥0.50	0.15	0.89
a_{15}	0.00	0.00	0.04	≥0.50	0.016	0.99
a_{16}	0.00	0.00	0.04	0.389	−0.83	0.41
a_{17}	0.00	0.00	0.04	0.422	0.33	0.74
a_{18}	0.00	0.00	0.04	0.440	−0.11	0.91
a_{19}	0.00	0.00	0.02	≥0.50	1.43	0.16
a_{21}	−0.34	−0.35	0.03	≥0.50	−128.00	0.00
a_{22}	−0.33	−0.32	0.03	0.211	−93.63	0.00
a_{23}	0.00	0.00	0.04	≥0.50	0.85	0.40
a_{24}	−0.04	−0.05	0.04	0.265	−11.44	0.00
a_{25}	0.00	0.00	0.04	0.278	−0.46	0.65
a_{26}	0.00	0.00	0.04	≥0.50	0.22	0.83
a_{27}	0.00	0.00	0.03	0.492	−0.03	0.97
a_{28}	0.00	0.00	0.04	0.214	0.98	0.33

here, is the ability of these tests to detect non null parameters from only one sample. Methods for accomplishing this may be based on (1) computed parameter variances (2) estimates of parameter variances based on bootstrapping or simulation.

5 Conclusions

We demonstrated the use of an OLS minimization to estimate the Fourier coefficients of the periodic parameters in a periodic autoregressive model. This method is shown to be effective even when the sample size N is small relative to the period T, say $N = T$. Simulations show that the empirical distributions of parameter estimates are typically normal and standard errors diminish as $N^{-1/2}$ as expected. Topics for future research include (1) improvement of computational methods (2) direct (parametric) computation of estimator standard errors to facilitate the identification of important Fourier coefficients (3) use of simulation or bootstrapping to characterize empirical distributions of parameter estimates (4) extension to PARMA.

Acknowledgements The author would like to acknowledge the efforts of Dr. Wioletta Wójtowicz for assistance in the simulations described here.

References

1. Dehay, D., & Hurd, H. (1993). Representation and estimation for periodically and almost periodically correlated random processes. In W. A. Gardner (Ed.), *Cyclostationarity in communications and signal processing*. IEEE Press.
2. Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Mathematics*, *2*, 385–388.
3. Hannan, E. J. (1955). A test for singularities in Sydney rainfall. *Australian Journal of Physics*, *8*, 289–297.
4. Hurd, H. L. (2004–2005). Periodically correlated sequences of less than full rank. *Journal of Statistical Planning and Inference*, *129*, 279–303.
5. Hurd, H. L., & Miamee, A.G. (2007). *Periodically Correlated Sequences: Spectral Theory and Practice*, Wiley, Hoboken, NJ.
6. Jones, R., & Brelsford, W. (1967). Time series with periodic structure. *Biometrika*, *54*, 403–408.
7. Lilliefors, H. (1967). On the Kolmogorov—Smirnov test for normality with mean and variance unknown, *Journal of American Statistical Association*, *62*, 399402.
8. Pagano, M. (1978). On periodic and multiple autoregressions. *Annals of Statistics*, *6*, 1310–1317.
9. Vecchia, A. V. (1985). Periodic autoregressive moving average (PARMA) modeling with applications to water resources. *Water Resources Bulletin*, *21*, 721–730.

Vectorial Periodically Correlated Random Processes and Their Covariance Invariant Analysis

I. Javorskyj, I. Matsko, R. Yuzefovych and O. Dzeryn

Abstract The first and the second order moment functions of the vectorial periodically correlated random processes—the mathematical models of the vectorial physical phenomena—are analyzed. The properties of the linear and the quadratic invariants of the covariance tensor-function are described. The representations of the covariance tensor-function and its invariants in the form of Fourier series are considered. Fourier coefficient properties of these series are analyzed. Obtained relations are specified for the amplitude and phase-modulated signals. The examples of using the vector PCRP methods for the rolling bearing vibration analysis are given.

Keywords Vectorial periodically correlated random processes · Covariance tensor-function · Linear invariants · Quadratic invariants

I. Javorskyj · I. Matsko (✉) · R. Yuzefovych · O. Dzeryn
Karpenko Physico-Mechanical Institute of National Academy of Sciences of Ukraine,
Naukova Str. 5, Lviv 79601, Ukraine
e-mail: matskoivan@gmail.com

I. Javorskyj
e-mail: javor@utp.edu.pl

R. Yuzefovych
e-mail: roman.yuzefovych@gmail.com

O. Dzeryn
e-mail: dzeryn_oksana@ukr.net

I. Javorskyj
Institute of Telecommunications, University of Science and Technology,
Al. Prof. Kaliskiego 7, 85796 Bydgoszcz, Poland

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_8

1 Introduction

Vectorial periodically correlated random processes (PCRP) are the adequate mathematical model for the analysis of stochastic recurrence structure of vectorial physical quantities, for example, density of electrical and magnetic fields, wind and sea flow speeds, displacement, speed and acceleration of vibration. There is a number of methods for vectorial processes analysis [1–4]: (a) component-wise one, where a vector on the plane is considered as a pair of numbers, which corresponds to its projections in the Cartesian coordinates; (b) a method of rotary components based on the substitution of the vector by a complex number and representation of time series in the form of superposition of round oscillations with left and right polarization; (c) complex-value one, where vector is considered as a complex number which real and imaging parts coincide with the Cartesian coordinates vector; (d) vector-algebraic one, where probability characteristics of the random processes are introduced by means of operations, accepted in Euclid algebra space [1, 4].

In the case of the last method application main characteristics are: vector of mean function, covariance function, which is defined as expected value of the tensor product of the random vector processes taken at moments of time t and $t+u$, and spectral density, that is the Fourier transformation of covariance function on the time lag u. Covariance and spectral tensors are corresponding tensor-functions. Each of the invariants allows one to develop one or another property of vectorial random processes independently of the coordinate system, where its components are measured.

Covariance and spectral properties of wind and sea flow speeds were investigated, using the characteristics of vectorial stationary random processes in monographs [1, 4]. Vectorial PCRP methods for investigation of stochastic recurrence structure of those quantities were proposed in [2]. It is shown in [5, 6] that using the vectorial algebraic PCRP-approach to vibration signals analysis allows one to improve the efficiency of mechanical system damage detection as well as separation of their sources. The importance of a detail analysis of probabilistic characteristics properties of the given class of random processes proceeds from the above mentioned.

The analysis of the first and the second order moment functions of vectorial PCRP, linear and the quadratic invariant properties of covariance tensor-functions and also the harmonic analysis of these quantities are the aim of this paper.

2 The First and the Second Order Moment Functions

Definition 2.1 Vectorial random process $\vec{\xi}(t) = \vec{\imath}\xi_1(t) + \vec{\jmath}\xi_2(t)$, where $\vec{\imath}$, $\vec{\jmath}$ are basis unit vectors, is called periodically correlated if its mean function $\vec{m}_{\underset{\xi}{\rightarrow}}(t) = E\vec{\xi}(t)$ is a periodical vector

$$\vec{m}_{\underset{\xi}{\rightarrow}}(t) = \vec{m}_{\underset{\xi}{\rightarrow}}(t+T), \tag{1}$$

and covariance function $\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u) = E\,\overset{\circ}{\vec{\xi}}(t) \otimes \overset{\circ}{\vec{\xi}}(t+u)$, where $\overset{\circ}{\vec{\xi}}(t) = \vec{\xi}(t) - \vec{m}_{\underset{\xi}{\rightarrow}}(t)$, and $\otimes$ is a sign of tensor product,[1] is called the time-periodical tensor-function

$$\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(\boldsymbol{t},\boldsymbol{u}) = \boldsymbol{b}_{\underset{\xi}{\rightarrow}}(\boldsymbol{t}+\boldsymbol{T},\boldsymbol{u}). \tag{2}$$

For fixed t and u the quantity $\overrightarrow{b}_{\underset{\xi}{\rightarrow}}(t,u)$ is a dyadic tensor. It follows from the matrix representation of a tensor

$$\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u) = \begin{bmatrix} b_{\xi_1}(t,u) & b_{\xi_2\xi_1}(t,u) \\ b_{\xi_1\xi_2}(t,u) & b_{\xi_2}(t,u) \end{bmatrix} \tag{3}$$

and condition (2) that auto-covariance functions of vector components $\overrightarrow{\xi}(t)$ $b_{\xi_1}(t,u)$ and $b_{\xi_2}(t,u)$, and also their cross-covariance functions $b_{\xi_1\xi_2}(t,u)$ and $b_{\xi_2\xi_1}(t,u)$ are periodic functions of time. Covariance tensor function $\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u)$ describes correlations between directed fluctuations of vector $\vec{\xi}(t)$ at the moments of time t and $t+u$, and the variance $\boldsymbol{d}_{\underset{\xi}{\rightarrow}}(t,0) = \boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,0)$ that is a tensor function of time t, gives numerical estimate of such variations intensity and their orientation in a chosen coordinate system.

Random processes $\xi_1(t)$ and $\xi_2(t)$ are periodically correlated and can be represented in the form [7–12]:

$$\xi_1(t) = \sum_{k\in Z} \xi_k^{(1)}(t) e^{ik\omega_0 t}, \tag{4}$$

[1]The second-rank tensor, which components are defined by the elements of the matrix $\begin{pmatrix} a_x b_x & a_x b_y \\ a_y b_x & a_y b_y \end{pmatrix}$ is called the tensor product $\vec{a}\otimes\vec{b}$ of the two vectors $\vec{a} = a_x\vec{\imath} + a_y\vec{\jmath}$ and $\vec{b} = b_x\vec{\imath} + b_y\vec{\jmath}$ (of the first-rank tensor).

$$\xi_2(t) = \sum_{k \in Z} \xi_k^{(2)}(t) e^{ik\omega_0 t}. \tag{5}$$

It follows from (4) and (5) that

$$b_{\xi_p}(t,u) = \sum_{n \in Z} e^{in\omega_0 t} \sum_{k \in Z} R_{k-n,k}^{(p)}(u) e^{ik\omega_0 u}, \tag{6}$$

$$b_{\xi_p \xi_q}(t,u) = \sum_{n \in Z} e^{in\omega_0 t} \sum_{k \in Z} R_{k-n,k}^{(p,q)}(u) e^{ik\omega_0 u}, \tag{7}$$

where $R_{kl}^{(p)}(u) = E\overline{\xi}_k^{(p)}(t)\overset{\circ}{\xi}_l^{(p)}(t+u)$, $R_{kl}^{(p,q)}(u) = E\overline{\xi}_k^{(p)}(t)\overset{\circ}{\xi}_l^{(p)}(t+u)$. Conditions of auto- and cross-covariance functions periodicity for random processes $\xi_1(t)$ and $\xi_2(t)$, obtained from (6) and (7), are independent. If stationary components $\xi_k^{(1)}(t)$ and $\xi_k^{(2)}(t)$ do not correlate the cross-covariance functions $b_{\xi_p \xi_q}(t,u)$ degenerates into a quantity independent of time t—cross-covariance functions of the jointly stationary random process:

$$b_{\xi_p \xi_q}(t,u) = \sum_{k \in Z} R_{kk}^{(p,q)}(u) e^{ik\omega_0 u}.$$

Definition 2.2 Vectorial random processes with periodic in time mean function $\vec{m}_{\vec{\xi}}(t)$ and spur of matrix (3) will be called the vector diagonal PCRP.

2.1 The Vector of a Mean Function

Since $\vec{m}_{\vec{\xi}}(t) = \vec{\imath} m_{\xi_1}(t) + \vec{\jmath} m_{\xi_2}(t)$, expression (1) means, that mean functions of each vector component $\xi_1(t)$ and $\xi_2(t)$ are also periodic in time: $m_{\xi_1}(t) = m_{\xi_1}(t+T)$, $m_{\xi_2}(t) = m_{\xi_2}(t+T)$.

It follows from (4) and (5) that

$$m_{\xi_1}(t) = \sum_{l \in Z} m_l^{(1)} e^{il\omega_0 t} = m_0^{(1)} + \sum_{l \in Z} \left[\tilde{m}_l^{(1)} \cos l\omega_0 t + \tilde{\tilde{m}}_l^{(1)} \sin l\omega_0 t \right],$$
$$m_{\xi_2}(t) = \sum_{l \in Z} m_l^{(2)} e^{il\omega_0 t} = m_0^{(2)} + \sum_{l \in Z} \left[\tilde{m}_l^{(2)} \cos l\omega_0 t + \tilde{\tilde{m}}_l^{(2)} \sin l\omega_0 t \right],$$

where

$$m_l^{(k)} = E\xi_l^{(k)}(t),\ m_l^{(k)} = \frac{1}{2}\left[\tilde{m}_l^{(k)} - i\tilde{\tilde{m}}_l^{(k)}\right],\ \forall\, l \neq 0\, k = \overline{1,2}. \tag{8}$$

Thus, we have the following:

Proposition 2.1 *The vector of the mean function is represented by the sum of vectors*

$$\vec{m}_{\vec{\xi}}(t) = \vec{m}_0 + \sum_{l \in N} \vec{m}_l(x_l, y_l),$$

where $\vec{m}_0 = m_0^{(1)}\vec{\imath} + m_0^{(2)}\vec{\jmath}$ *is some constant vector while the coordinates of the vectors* $\vec{m}_l(x_l, y_l)$ *change harmonically in time, that is*

$$x_l(t) = A_l \sin(l\omega_0 t + \varphi_l),\ y_l(t) = B_l \cos(l\omega_0 t + \psi_l), \tag{9}$$

where $A_l = \sqrt{\left(\tilde{m}_l^{(1)}\right)^2 + \left(\tilde{\tilde{m}}_l^{(1)}\right)^2}$, $B_l = \sqrt{\left(\tilde{m}_l^{(2)}\right)^2 + \left(\tilde{\tilde{m}}_l^{(2)}\right)^2}$, $\varphi_l = arctg\left(\tilde{m}_l^{(1)} / \tilde{\tilde{m}}_l^{(1)}\right)$, $\psi_l = arctg\left(\tilde{m}_l^{(2)} / \tilde{\tilde{m}}_l^{(2)}\right)$, and $\tilde{m}_l^{(k)}$, $\tilde{\tilde{m}}_l^{(k)}$ *are defined by expression (8).*

Consider separately harmonic components of vector $\vec{m}_{\vec{\xi}}(t)$. Formulae (9) are parametric equations of some lines at the *Oxy* plane that are called the Lissajous figures [13]. Represent these equations in the form

$$\frac{x_l}{A_l} = \sin(l\omega_0 t + \varphi_l),\ \frac{y_l}{B_l} = \sin(l\omega_0 t + \psi_l)$$

and assume that $\varphi_l - \psi_l = 2k\pi$, so $\varphi_l = 2k\pi + \psi_l$. Then

$$\frac{x_l}{A_l} = \sin(l\omega_0 t + \psi_l),\ \frac{y_l}{B_l} = \sin(l\omega_0 t + \psi_l).$$

Subtracting these equations we obtain

$$y_l = \frac{B_l}{A_l} x_l.$$

Since $|x_l| \leq A_l$, $|y_l| \leq B_l$, this formula is an equation of the interval with beginning at the point with coordinates $(-A_l; -B_l)$ and ending at point $(A_l; B_l)$.

If $\varphi_l - \psi_l = k\pi$, where k is an odd number, we obtain

$$y_l = -\frac{B_l}{A_l} x_l,$$

that is an equation of the interval with beginning at $(-A_l; B_l)$ and ending at $(A_l; -B_l)$.

Suppose $\varphi_l - \psi_l = k\frac{\pi}{2}$. Then

$$\frac{x_l}{A_l} = \cos(l\omega_0 t + \psi_l),\ \frac{y_l}{B_l} = \sin(l\omega_0 t + \psi_l).$$

After squaring and summation we obtain the equation of an ellipse, which axes are defined by the squares of harmonic components amplitudes:

$$\frac{x_l^2}{A_l^2} + \frac{y_l^2}{B_l^2} = 1. \tag{10}$$

If $A_l = B_l$, then (10) is an equation of a circle.

Equation

$$\frac{x_l^2}{A_l^2} + \frac{y_l^2}{B_l^2} = 1 - \cos[2l\omega_0 t + (\varphi_l + \psi_l)]\cos(\varphi_l - \psi_l)$$

in a general case describes some central closed curve which characteristic sizes are defined by amplitudes A_l and B_l and by phases φ_l and ψ_l.

Taking into account that the curve according to the mean function $\vec{m}_{\underrightarrow{\xi}}(t)$ is the superposition of the curves, the parametric equation of each of them has the form (9), we have:

Proposition 2.2 *The parametric equations of the closed curve for the vector mean function* $\vec{m}_{\underrightarrow{\xi}}(t)$ *on the Oxy plane is represented in the form*

$$\begin{cases} x(t) = m_0^{(1)} + \sum\limits_{l \in \mathbb{N}} A_l \cos(l\omega_0 t + \varphi_l), \\ y(t) = m_0^{(2)} + \sum\limits_{l \in \mathbb{N}} B_l \cos(l\omega_0 t + \psi_l). \end{cases}$$

The concrete form of the mean function curve is defined by amplitudes A_l, B_l and phases φ_l, ψ_l of the harmonic components. This form reflects *ipso facto* properties of random vector $\overrightarrow{\xi}(t)$ deterministic part from the geometric point of view.

2.2 The Covariance Tensor-Function

Analogically to the matrices for dyadic vector we can define the operation of transposition of covariance tensor-function (3), which is reduced to changing the rows by columns with saving their numbers. Thus we have

$$\boldsymbol{b}^{T}_{\vec{\xi}}(t,u)=\begin{bmatrix} b_{\xi_1}(t,u) & b_{\xi_2\xi_1}(t,u) \\ b_{\xi_1\xi_2}(t,u) & b_{\xi_2}(t,u) \end{bmatrix}.$$

Since

$$b_{\xi_p}(t,-u)=b_{\xi_p}(t-u,u),\ b_{\xi_p\xi_q}(t,-u)=b_{\xi_q\xi_p}(t-u,u), \tag{11}$$

then

$$\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,-u)=\boldsymbol{b}^{T}_{\underset{\xi}{\rightarrow}}(t-u,u).$$

Taking into account that sum $\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u)+\boldsymbol{b}^{T}_{\underset{\xi}{\rightarrow}}(t,u)$ is symmetrical matrix and difference $\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u)-\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u)$ is asymmetrical matrix, we have:

Proposition 2.3 *The covariance tensor-function* $\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u)$ *in the only way can be represented in the form of the sum of symmetrical* $\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)$ *and asymmetrical* $\boldsymbol{b}^{A}_{\underset{\xi}{\rightarrow}}(t,u)$ *parts*:

$$\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u)=\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)+\boldsymbol{b}^{A}_{\underset{\xi}{\rightarrow}}(t,u),$$

where

$$\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)=\begin{bmatrix} b_{\xi_1}(t,u) & \frac{1}{2}\left[b_{\xi_1\xi_2}(t,u)+b_{\xi_2\xi_1}(t,u)\right] \\ \frac{1}{2}\left[b_{\xi_1\xi_2}(t,u)+b_{\xi_2\xi_1}(t,u)\right] & b_{\xi_2}(t,u) \end{bmatrix}, \tag{12}$$

$$\boldsymbol{b}^{A}_{\underset{\xi}{\rightarrow}}(t,u)=\begin{bmatrix} 0 & \frac{1}{2}\left[b_{\xi_1\xi_2}(t,u)-b_{\xi_2\xi_1}(t,u)\right] \\ \frac{1}{2}\left[b_{\xi_2\xi_1}(t,u)-b_{\xi_1\xi_2}(t,u)\right] & 0 \end{bmatrix}. \tag{13}$$

Corollary 2.1 *The transpositions of the symmetrical and asymmetrical parts of the covariance tensor-function* $\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u)$ *satisfy the relations:*

$$\left[\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)\right]^{T}=\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u),\ \left[\boldsymbol{b}^{A}_{\underset{\xi}{\rightarrow}}(t,u)\right]^{T}=-\boldsymbol{b}^{A}_{\underset{\xi}{\rightarrow}}(t,u), \tag{14}$$

$$\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,-u)=\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t-u,u),\ \boldsymbol{b}^{A}_{\underset{\xi}{\rightarrow}}(t,-u)=\boldsymbol{b}^{A}_{\underset{\xi}{\rightarrow}}(t-u,u). \tag{15}$$

We obtain relations (14) after transpositions of (12) and (13). Relations (15) follow from Eq. (11).

All the components of tensor function $\boldsymbol{b}^{A}_{\underset{\xi}{\rightarrow}}(t,u)$ *at* $u=0$ are equal to zero, that's why the variance tensor for vectorial PCRP is symmetrical.

3 The Linear Invariants of the Covariance Tensor-Function

Tensor components in general case depend on the choice of the coordinate system, however there are functions of tensor components that are invariant to the coordinate system choice. Such functions are called tensor invariants [3].

Definition 3.1 The expected value of the scalar product of vector $\vec{\xi}(t)=\xi_1(t)\vec{\imath}+\xi_2(t)\vec{\jmath}$ and vector $\vec{\xi}(t+u)=\xi_1(t+u)\vec{\imath}+\xi_2(t+u)\vec{\jmath}$ is called the linear invariant $I_1(t,u)$, i.e.

$$I_1(t,u)=E\left\{\overset{\circ}{\vec{\xi}}(t)\cdot\overset{\circ}{\vec{\xi}}(t+u)\right\}=E\left|\overset{\circ}{\vec{\xi}}(t)\right|\left|\overset{\circ}{\vec{\xi}}(t+u)\right|\cos\left(\angle\overset{\circ}{\vec{\xi}}(t)\,\overset{\circ}{\vec{\xi}}(t+u)\right), \tag{16}$$

where $\angle\overset{\circ}{\vec{\xi}}(t)\,\overset{\circ}{\vec{\xi}}(t+u)$ is an angle between $\overset{\circ}{\vec{\xi}}(t)$ and $\overset{\circ}{\vec{\xi}}(t+u)$.

The invariant $I_1(t,u)$ can be considered as a measure of vectors $\overset{\circ}{\vec{\xi}}(t)$ and $\overset{\circ}{\vec{\xi}}(t+u)$ collinearity.

Proposition 3.1 *The linear invariant* $I_1(t,u)$ *is defined by the sum of main diagonal elements of the symmetrical part* $\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)$ *of the covariance tensor function* $\boldsymbol{b}_{\underset{\xi}{\rightarrow}}(t,u)$.

Proof The scalar product of the vectors $\overrightarrow{\xi}(t)$ and $\overrightarrow{\xi}(t+u)$ is given by

$$\begin{aligned}\overset{\circ}{\vec{\xi}}(t)\,\overset{\circ}{\vec{\xi}}(t+u)&=\left[\overset{\circ}{\underset{1}{\xi}}(t)\overrightarrow{i}+\overset{\circ}{\underset{2}{\xi}}(t)\overrightarrow{j}\right]\left[\overset{\circ}{\underset{1}{\xi}}(t+u)\overrightarrow{i}+\overset{\circ}{\underset{2}{\xi}}(t+u)\overrightarrow{j}\right]\\&=\overset{\circ}{\underset{1}{\xi}}(t)\,\overset{\circ}{\underset{1}{\xi}}(t+u)+\overset{\circ}{\underset{2}{\xi}}(t)\,\overset{\circ}{\underset{2}{\xi}}(t+u).\end{aligned}$$

It follows from the above that

$$I_1(t,u) = b_{\xi_1}(t,u) + b_{\xi_2}(t,u). \tag{17}$$

Corollary 3.1 *The time changes of the least square value of vector* $\overset{\circ}{\vec{\xi}}(t)$ *are defined by the time changes of the invariant* $I_1(t,u)$ *at* $u=0$, *i.e.*

$$E\left|\overset{\circ}{\vec{\xi}}(t)\right|^2 = I_1(t,0) = b_{\xi_1}(t,0) + b_{\xi_2}(t,0).$$

The last equation follows from (16) and (17) at $u=0$.

Taking into account the property of symmetry of PCRP covariance (11) we obtain:

$$I_1(t,-u) = I_1(t-u,u). \tag{18}$$

Definition 3.2 The expected value of the skew product of vectors $\overset{\circ}{\vec{\xi}}(t)$ and $\overset{\circ}{\vec{\xi}}(t+u)$ is called the linear invariant $D(t,u)$, i.e.

$$D(t,u) = E\left\{\overset{\circ}{\vec{\xi}}(t) \times \overset{\circ}{\vec{\xi}}(t+u)\right\} = E\left|\overset{\circ}{\vec{\xi}}(t)\right|\left|\overset{\circ}{\vec{\xi}}(t+u)\right| \sin\left(\angle\, \overset{\circ}{\vec{\xi}}(t)\, \overset{\circ}{\vec{\xi}}(t+u)\right). \tag{19}$$

Skew product of vectors is an analogue to the vector product (defined in a three-dimensional space), but contrary to the last one it is not a vector but a scalar.

Note that in the case, when the angle between vectors $\overset{\circ}{\vec{\xi}}(t)$ and $\overset{\circ}{\vec{\xi}}(t+u)$ is equal to zero, the invariant $D(t,u)$ is equal to zero too. Therefore this quantity is often called a rotation indicator.

Proposition 3.2 *The linear invariant* $D(t,u)$ *defines the properties of asymmetrical part* $\boldsymbol{b}^A_{\vec{\xi}}(t,u)$ *of covariance tensor-function* $\boldsymbol{b}_{\vec{\xi}}(t,u)$.

Proof For skew product of vectors $\overset{\circ}{\vec{\xi}}(t)$ and $\overset{\circ}{\vec{\xi}}(t+u)$ we have

$$\begin{aligned}\overset{\circ}{\vec{\xi}}(t) \times \overset{\circ}{\vec{\xi}}(t+u) &= \left[\overset{\circ}{\underset{1}{\xi}}(t)\vec{\imath} + \overset{\circ}{\underset{2}{\xi}}(t)\vec{\jmath}\right] \times \left[\overset{\circ}{\underset{1}{\xi}}(t+u)\vec{\imath} + \overset{\circ}{\underset{2}{\xi}}(t+u)\vec{\jmath}\right] \\ &= \overset{\circ}{\underset{1}{\xi}}(t)\, \overset{\circ}{\underset{2}{\xi}}(t+u) - \overset{\circ}{\underset{2}{\xi}}(t)\, \overset{\circ}{\underset{1}{\xi}}(t+u).\end{aligned}$$

Hence

$$D(t,u) = E\left\{\overset{\circ}{\vec{\xi}}(t) \times \overset{\circ}{\vec{\xi}}(t+u)\right\} = b_{\xi_1\xi_2}(t,u) - b_{\xi_2\xi_1}(t,u).$$

On the other hand the asymmetric part $\boldsymbol{b}_{\vec{\xi}}^{A}(t,u)$ can be represented in the form

$$\boldsymbol{b}_{\vec{\xi}}^{A}(t,u)=\frac{1}{2}D(t,u)\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$$

Thus, the quantity $D(t,u)$ is the invariant of asymmetrical part of covariance tensor-function and fully defines its properties.

Since $b_{\xi_1\xi_2}(t,-u)=b_{\xi_2\xi_1}(t-u,u)$ then

$$D(t,-u)=-D(t-u,u). \tag{20}$$

It follows from this equation that $D(t,0)=0$. The invariants (16) and (19) for the stationary random processes are time-independent, i.e. they are functions of time lag u only, that is $I_1(t,u)=I_1(u)$ and $D(t,u)=D(u)$. It follows from (18) and (20) that in this case $I_1(u)$ is an even function and $D(u)$ is an odd function: $I_1(-u)=I_1(u),\ D(-u)=D(u)$.

Taking into account expressions (16) and (19), we can interpret the invariants $I_1(t,u)$ and $D(t,u)$ as the covariance functions of collinear $\overset{\circ}{\vec{\xi}}{}^{||}(t)$ and orthogonal $\overset{\circ}{\vec{\xi}}{}^{\perp}(t)$ components of the vectorial process $\vec{\xi}(t)$:

$$I_1(t,u)=E\left\{\overset{\circ}{\vec{\xi}}(t)\cdot\overset{\circ}{\vec{\xi}}{}^{||}(t)\right\}=E\left\{\overset{\circ}{\vec{\xi}}{}^{||}(t)\cdot\overset{\circ}{\vec{\xi}}(t+u)\right\}, \tag{21}$$

$$D(t,u)=E\left\{\overset{\circ}{\vec{\xi}}(t)\cdot\overset{\circ}{\vec{\xi}}{}^{\perp}(t+u)\right\}=E\left\{\overset{\circ}{\vec{\xi}}{}^{\perp}(t)\cdot\overset{\circ}{\vec{\xi}}(t+u)\right\}. \tag{22}$$

In the general case functions (21) and (22) change their sign, so when interpreting them it is advisable to consider their absolute values as coefficients of intensity of the respective changes, and the signs as coefficients of predominant direction of these changes.

Absolute value of $|I_1(t,u)|$ characterizes the intensity of collinear changes of random vector $\overset{\circ}{\vec{\xi}}(t)$ and the sign of $I_1(t,u)$ characterizes their direction: if the sign is positive the changes of vector in interval $[t,\ t+u]$ are mainly unidirectional, and if the sign is negative—multidirectional.

A skew product sign depends on the relative positions of $\overset{\circ}{\vec{\xi}}(t)$ and $\overset{\circ}{\vec{\xi}}(t+u)$, because the angle is measured from $\overset{\circ}{\vec{\xi}}(t)$ to $\overset{\circ}{\vec{\xi}}(t+u)$. If the sign is positive then for fixed moments of time t and $t+u$ vector $\overset{\circ}{\vec{\xi}}(t+u)$ is located to the left from $\overset{\circ}{\vec{\xi}}(t)$, and if the sign is negative—to the right.

It follows from the mentioned above that invariants $I_1(t,u)$ and $D(t,u)$ of covariance tensor (3) allow one to analyze collinear and orthogonal variability of random processes $\overset{\circ}{\vec{\xi}}(t)$ independently of the chosen coordinate system.

4 The Quadratic Invariant Properties

Form a quadratic form on the base of the symmetrical part (12) of the covariance tensor-function (3):

$$f(x,y) = ax^2 + 2bxy + cy^2, \tag{23}$$

where

$$a = b_{\xi_1}(t,u),\ b = \frac{1}{2}\left[b_{\xi_1\xi_2}(t,u) + b_{\xi_2\xi_1}(t,u)\right],\ c = b_{\xi_2}(t,u).$$

Definition 4.1 The quantity

$$I_2(t,u) = \begin{vmatrix} a & b \\ b & c \end{vmatrix} = b_{\xi_1}(t,u)b_{\xi_2}(t,u) - \frac{1}{4}\left[b_{\xi_1\xi_2}(t,u) + b_{\xi_2\xi_1}(t,u)\right]^2$$

which is matrix (12) determinant, is called discriminant of quadratic form (23) or quadratic invariant of the symmetrical part $\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)$ of the covariance tensor-function.

The quadratic invariant $I_2(t,u)$ is a determining feature for quadratic form (23) classification [14]. Suppose $a = b_{\xi_1}(t,u) \neq 0$.

Then

$$f(x,y) = \frac{1}{a}\left(a^2x^2 + 2abxy + acy^2\right) = \frac{1}{a}\left[(ax+by)^2 + I_2(t,u)y^2\right]. \tag{24}$$

If $I_2(t,u) > 0$ and $a \geq 0$, formula (24) is positive definite $f(x,y) \geq 0$, and if $a < 0$—negative definite $f(x,y) \leq 0$. The surface $z = f(x,y)$ (Fig. 1a) is cup-shaped and its isolines $f(x,y) = k$ are elipses.

If $I_2(t,u) = 0, f(x,y)$ equals to zero on the line $ax + by = 0$, and outside the line it has the same sign as a. The surface $z = f(x,y)$ in this case is gutter-shaped and its isolines are straight lines (Fig. 1b).

For negative $I_2(t,u) < 0$ expression (24) is a subtraction of squares and it can be represented in the form

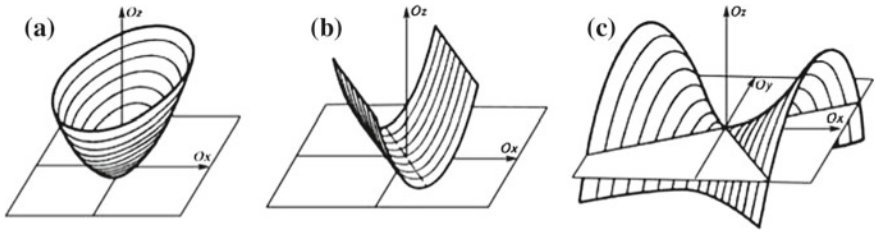


Fig. 1 Typical graphic dependences of the quadratic form $z=f(x,y)$

$$f(x,y)=\frac{1}{a}\left[ax+\left(b+\sqrt{-I_2(t,u)}\right)y\right]\left[ax+\left(b-\sqrt{-I_2(t,u)}\right)y\right]. \qquad (25)$$

This function equals zero on the lines

$$ax+\left(b+\sqrt{-I_2(t,u)}\right)y=0,\ ax+\left(b-\sqrt{-I_2(t,u)}\right)y=0.$$

These lines cross through the coordinate system origin and divide the plane Oxy into four sections. When passing to the opposite section one of the multipliers changes its sign and the other does not change it, that's why the quadratic form changes its sign too (Fig. 1c).

If $c=b_{\xi_2}(t,u)\neq 0$ expression (23) can be rewritten as

$$f(x,y)=\frac{1}{c}\left(cax^2+2bcxy+c^2y^2\right)=\frac{1}{c}\left[(bx+cy)^2+I_2(t,u)x^2\right].$$

The properties of this function are similar to the properties of (24).

If $a=c=0$, then $f(x,y)=2bxy$ and

$$I_2(t,u)=-\frac{1}{4}\left[b_{\xi_1\xi_2}(t,u)+b_{\xi_2\xi_1}(t,u)\right]^2<0.$$

In this case function $z=f(x,y)$ is alternating.

Proposition 4.1 *The quadratic invariant of variance tensor-function* $I_2(t,0)$ *is positive definite* $I_2(t,0)\geq 0$.

Proof Since for zero time lag $b_{\xi_1\xi_2}(t,0)=b_{\xi_2\xi_1}(t,0)$ then

$$I_2(t,0)=b_{\xi_1}(t,0)b_{\xi_2}(t,0)-b^2_{\xi_1\xi_2}(t,0).$$

Taking into account the known inequality

$$b^2_{\xi_1\xi_2}(t,0)\leq b_{\xi_1}(t,0)b_{\xi_2}(t,0)$$

we conclude that $I_2(t,0)\geq 0$.

Corollary 4.1 *The quadratic form (24) of the variance tensor-function is positive definite* $[f(x,y)]_{u=0} \geq 0$.

The quadratic form is a cup-shaped, if the strong inequality $b^2_{\xi_1\xi_2}(t,0) < b_{\xi_1}(t,0)b_{\xi_2}(t,0)$ is satisfied, and it has a qutter-shaped form, if $b^2_{\xi_1\xi_2}(t,0) = b_{\xi_1}(t,0)b_{\xi_2}(t,0)$. In the first case the isolines are ellipses with respective directions of their axes, and in the second case—the straight lines.

In the general case auto-covariance functions $b_{\xi_1}(t,u)$ and $b_{\xi_2}(t,u)$ for some t and u have the values with different signs, so condition $I_2(t,u) \geq 0$ is not satisfied. It means that depending on the variations characters of auto- and cross-covariance functions in time t and lag u it is possible to obtain the given above types of quadratic forms. Note, that in the case of alternating quadratic forms the isolines $f(x,y) = c$ are hyperbolas.

Write quadratic form $f(x,y)$ in the matrix form:

$$f(x,y) = X^T \boldsymbol{b}^C_{\underset{\xi}{\rightarrow}}(t,u)X,$$

where $X^T = [x\ y]$. Matrix $\boldsymbol{b}^C_{\underset{\xi}{\rightarrow}}(t,u)$ is real and symmetric, that's why such orthogonal transformation

$$X = \tilde{\boldsymbol{b}}^C_{\underset{\xi}{\rightarrow}}(t,u)S, \tag{26}$$

exists, where $S^T = [s\ \nu]$, that reduces this form to a canonical one, that contains squares of variables s and ν only, i.e.

$$f(s,\nu) = \lambda_1 s^2 + \lambda_2 \nu^2 = S^T \Lambda S,$$

where Λ is the diagonal matrix:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Numbers λ_1 and λ_2 are eigen-values of matrix $\boldsymbol{b}^C_{\underset{\xi}{\rightarrow}}(t,u)$, namely

$$\boldsymbol{b}^C_{\underset{\xi}{\rightarrow}}(t,u)S = \lambda S,$$

or

$$\left[\boldsymbol{b}^C_{\underset{\xi}{\rightarrow}}(t,u) - \lambda E\right]S = 0,$$

where E is unity matrix.

Write the last matrix equation in the form of the system of equations:

$$\left[b_{\xi_1}(t,u)-\lambda\right]s+\frac{1}{2}\left[b_{\xi_1\xi_2}(t,u)+b_{\xi_2\xi_1}(t,u)\right]\nu=0,$$
$$\frac{1}{2}\left[b_{\xi_1\xi_2}(t,u)+b_{\xi_2\xi_1}(t,u)\right]s+\left[b_{\xi_2}(t,u)-\lambda\right]\nu=0.$$

This system has nonzero solutions in case its determinant equals zero:

$$\begin{vmatrix} b_{\xi_1}(t,u)-\lambda & \frac{1}{2}\left[b_{\xi_1\xi_2}(t,u)+b_{\xi_2\xi_1}(t,u)\right] \\ \frac{1}{2}\left[b_{\xi_1\xi_2}(t,u)+b_{\xi_2\xi_1}(t,u)\right] & b_{\xi_2}(t,u)-\lambda \end{vmatrix}$$
$$=\lambda^2-\left[b_{\xi_1}(t,u)+b_{\xi_2}(t,u)\right]\lambda+b_{\xi_1}(t,u)b_{\xi_2}(t,u)-\frac{1}{4}\left[b_{\xi_1\xi_2}(t,u)+b_{\xi_2\xi_1}(t,u)\right]^2=0.$$

Its solutions are:

$$\lambda_{1,2}(t,u)=\frac{1}{2}\left[b_{\xi_1}(t,u)+b_{\xi_2}(t,u)\pm\sqrt{\begin{bmatrix} b_{\xi_1}(t,u)- \\ -b_{\xi_2}(t,u)\end{bmatrix}^2+\begin{bmatrix} b_{\xi_1\xi_2}(t,u)+ \\ +b_{\xi_2\xi_1}(t,u)\end{bmatrix}^2}\right] \quad (27)$$
$$=\frac{1}{2}\left[I_1(t,u)\pm\sqrt{I_1^2(t,u)-4I_2(t,u)}\right].$$

It is clear, that

$$\lambda_1(t,u)+\lambda_2(t,u)=I_1(t,u),\ \lambda_1(t,u)\lambda_2(t,u)=I_2(t,u).$$

Orthogonal transformation (22) is a transition to a new system of coordinates $Os\nu$, obtained by rotating the old system Oxy by angle α (Fig. 2).

It can be represented as

$$x=s\cos\alpha-\nu\sin\alpha,$$
$$y=s\sin\alpha+\nu\cos\alpha.$$

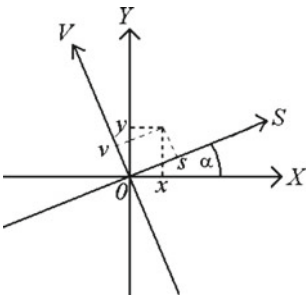


Fig. 2 Coordinate points on the plane in the old and new coordinate systems

Substituting these ratios into (20) we obtain

$$f(s,\nu) = s^2\left(a\cos^2\alpha + b\sin 2\alpha + c\sin^2\alpha\right) + \nu^2\left(a\sin 2\alpha - b\sin 2\alpha + c\cos^2\alpha\right) + s\nu\left[2b\left(\cos^2\alpha - \sin^2\alpha\right) + 2(c-a)\cos\alpha\sin\alpha\right].$$

Component depending on product $s\nu$ is absent if

$$2b\left(\cos^2\alpha - \sin^2\alpha\right) + 2(c-a)\sin\alpha\cos\alpha = 0.$$

From this we obtain

$$tg2\alpha = \frac{2b}{a-c}.$$

Taking into account the foregoing we can formulate now the following proposition.

Proposition 4.2 *The symmetrical part of the covariance tensor-function* $\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)$ *in its own basis Osv which is obtained by rotating the old system Oxy by angle*

$$\alpha = \frac{1}{2}arctg\frac{b_{\xi_2\xi_1}(t,u) + b_{\xi_2\xi_1}(t,u)}{b_{\xi_1}(t,u) - b_{\xi_2}(t,u)},$$

is represented by the diagonal matrix

$$\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u) = \begin{bmatrix} \lambda_1(t,u) & 0 \\ 0 & \lambda_2(t,u) \end{bmatrix},$$

where $\lambda_1(t,u)$ *and* $\lambda_2(t,u)$ *are its eigen-values, defined by expression (27).*

Central quadratic curve

$$\lambda_1(t,u)s^2 + \lambda_2(t,u)\nu^2 = 1. \tag{28}$$

can be put in accordance with the symmetrical part of tensor $\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)$ in its own basis. Invariants $\lambda_1(t,u)$ and $\lambda_2(t,u)$, that are eigen-values of symmetrical tensor $\boldsymbol{b}^{C}_{\underset{\xi}{\rightarrow}}(t,u)$, define typical sizes of this curve. They are the extreme values of covariance functions in the orthogonal direction. The inverse quantities define squares of the major and minor axes of the central quadratic curve. If $\lambda_1(t,u) > 0$ and $\lambda_2(t,u) > 0$, the tensor curve is an ellipse, and if $\lambda_1(t,u) = \lambda_2(t,u)$—circle. If $\lambda_1(t,u) > 0$ and $\lambda_2(t,u) < 0$, then (28) is an equation of hyperbola.

Proposition 4.3 *The quadratic curve (28) is an ellipse or a circle if* $I_2(t,u) > 0$, *and it is a hyperbola if* $I_2(t,u) < 0$. *If* $I_2(t,u) = 0$, *curve (28) degenerates into straight lines.*

Taking into account (17) and (21) we have

$$I_1^2(t,u) - 4I_2(t,u) = \left[b_{\xi_1}(t,u) - b_{\xi_2}(t,u)\right]^2 + \left[b_{\xi_1\xi_2}(t,u) + b_{\xi_2\xi_1}(t,u)\right]^2 \geq 0.$$

Thus, $\sqrt{I_1^2(t,u) - 4I_2(t,u)}$ is a real function $\forall\, t, u \in R$. From (27) it follows that for $I_2(t,u) > 0 \lambda_1(t,u) > 0$ and $\lambda_2(t,u) > 0$. If $I_2(t,u) < 0$ then $\lambda_1(t,u) > 0$ and $\lambda_2(t,u) < 0$. In case $I_2(t,u) = 0$, $\lambda_2(t,u) = 0$.

Corollary 4.2 *The quadratic curve (28) accordant with the variance tensor-function for all moment time t, for which $I_2(t,0) \neq 0$, is an ellipse or a circle.*

Proposition 4.4 *The values of covariance function $\boldsymbol{b}\overset{C}{\underset{\xi}{\rightarrow}}(t,u)$ of the vector projections onto the direction that forms angle β with own basis are calculated by formula:*

$$b_\beta(t,u) = \lambda_1(t,u)\cos^2\beta + \lambda_2(t,u)\sin^2\beta. \tag{29}$$

Proof The vector projections onto chosen direction are given by

$$\eta(t) = \eta_1(t)\cos\,\beta + \eta_2(t)\sin\,\beta,$$

where $\eta_1(t)$ and $\eta_2(t)$ are the vector components in own basis. The random processes $\eta_1(t)$ and $\eta_2(t)$ are non-correlated. Then

$$b_\beta(t,u) = E\overset{\circ}{\underset{1}{\eta}}(t)\,\overset{\circ}{\eta}(t+u) = b_{\eta_1}(t,u)\cos^2\beta + b_{\eta_2}(t,u)sin^2\beta.$$

Taking into account that $b_{\eta_1}(t,u) = \lambda_1(t,u)$ and $b_{\eta_2}(t,u) = \lambda_2(t,u)$, we obtain formula (29).

Note, that analyzing real oscillations for the sake of suitable physical interpretation it is worth considering linear invariants I_1 and D, that determine properties of vectorial processes without preliminary finding the covariance function tensor own basis. Using invariants I_1 and I_2 we can determine this basis, define the form of the tensor central quadratic curve, its typical sizes and orientation of its own basis. Based on these investigations the typical dependences of the given characteristics on time t and lag u are obtained, and the conclusions about the properties of objects, that are the sources of analyzed vector oscillations, can be drawn.

5 The Harmonic Analysis of the Covariance Tensor-Functions and Their Invariants

Assume that

$$\int_0^T \left|b_{\xi_p}(t,u)\right| dt < \infty,\ p = \overline{1,2}, \tag{30}$$

$$\int_0^T \left| b_{\xi_p \xi_q}(t,u) \right| dt < \infty,\ p,q = 1,2. \tag{31}$$

The auto-covariance $b_{\xi_p}(t,u)$ and cross-covariance function can be represented as Fourier series

$$b_{\xi_p}(t,u) = \sum_{k \in Z} B_k^{(\xi_p)}(u) e^{ik\omega_0 t},\ b_{\xi_p \xi_q}(t,u) = \sum_{k \in Z} B_k^{(\xi_p \xi_q)}(u) e^{ik\omega_0 t}. \tag{32}$$

Proposition 5.1 *If inequalities (30) and (31) are satisfied, the covariance tensor-function* $b_{\vec{\xi}}(t,u)$ *is represented by Fourier series*

$$\boldsymbol{b}_{\vec{\xi}}(t,u) = \sum_{k \in Z} \boldsymbol{b}_k^{(\vec{\xi})}(u) e^{ik\omega_0 t} = \vec{\boldsymbol{b}}_0^{\vec{\xi}}(u) + \sum_{k \in N} \left[\boldsymbol{C}_k^{(\vec{\xi})}(u) \cos\, k\omega_0 t + \boldsymbol{S}_k^{(\vec{\xi})}(u) \sin\, k\omega_0 t \right],$$

where

$$\boldsymbol{b}_0^{(\vec{\xi})}(u) = \begin{bmatrix} B_0^{(\xi_1)}(u) & B_0^{(\xi_1 \xi_2)}(u) \\ B_0^{(\xi_2 \xi_1)}(u) & B_0^{(\xi_2)}(u) \end{bmatrix},\ \boldsymbol{b}_k^{(\vec{\xi})}(u) = \begin{bmatrix} B_k^{(\xi_1)}(u) & B_k^{(\xi_1 \xi_2)}(u) \\ B_k^{(\xi_2 \xi_1)}(u) & B_k^{(\xi_2)}(u) \end{bmatrix}$$

$$\boldsymbol{C}_k^{(\vec{\xi})}(u) = \begin{bmatrix} C_k^{(\xi_1)}(u) & C_k^{(\xi_1 \xi_2)}(u) \\ C_k^{(\xi_2 \xi_1)}(u) & C_k^{(\xi_2)}(u) \end{bmatrix},\ \boldsymbol{S}_k^{(\vec{\xi})}(u) = \begin{bmatrix} S_k^{(\xi_1)}(u) & S_k^{(\xi_1 \xi_2)}(u) \\ S_k^{(\xi_2 \xi_1)}(u) & S_k^{(\xi_2)}(u) \end{bmatrix},$$

and $\boldsymbol{b}_k^{(\vec{\xi})}(u) = \frac{1}{2} \left[\boldsymbol{C}_k^{(\vec{\xi})}(u) - i\boldsymbol{S}_k^{(\vec{\xi})}(u) \right],\ k \neq 0$. *The quantities are called the covariance Fourier tensor-component* [2]. *Taking into account the equalities*

$$B_k^{(\xi_p)}(-u) = B_k^{(\xi_p)}(u) e^{-ik\omega_0 u},\ B_k^{(\xi_p \xi_q)}(-u) = B_k^{(\xi_q \xi_p)}(u) e^{-ik\omega_0 u},$$

for Fourier tensor-components we obtain

$$\boldsymbol{b}_k^{(\vec{\xi})}(-u) = \left[\boldsymbol{b}_k^{(\vec{\xi})}(u) \right]^T e^{-ik\omega_0 u}, \tag{33}$$

$$\boldsymbol{C}_k^{(\vec{\xi})}(-u) = \left[\boldsymbol{C}_k^{(\vec{\xi})}(u) \right]^T \cos\, k\omega_0 u - \left[\boldsymbol{S}_k^{(\vec{\xi})}(u) \right]^T \sin\, k\omega_0 u, \tag{34}$$

$$\boldsymbol{S}_k^{(\vec{\xi})}(-u) = \left[\boldsymbol{C}_k^{(\vec{\xi})}(u) \right]^T \sin\, k\omega_0 u + \left[\boldsymbol{S}_k^{(\vec{\xi})}(u) \right]^T \cos\, k\omega_0 u. \tag{35}$$

Corollary 5.1 *If the inequalities (30) and (31) are satisfied, the symmetrical and asymmetrical parts of the covariance tensor-function are represented by Fourier series*

$$\boldsymbol{b}^{A}_{\vec{\xi}}(t,u)=\sum_{k\in Z}\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{k}(u)e^{ik\omega_0 t}=\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{0}(u)+\sum_{k\in N}\left[\tilde{\boldsymbol{C}}^{(\vec{\xi})}_{k}(u)\cos\,k\omega_0 t+\tilde{\boldsymbol{S}}^{(\vec{\xi})}_{k}(u)sink\omega_0 t\right],$$

$$\boldsymbol{b}^{C}_{\vec{\xi}}(t,u)=\sum_{k\in Z}\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{k}(u)e^{ik\omega_0 t}=\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{0}(u)+\sum_{k\in N}\left[\tilde{\boldsymbol{C}}^{(\vec{\xi})}_{k}(u)\cos\,k\omega_0 t+\tilde{\boldsymbol{S}}^{(\vec{\xi})}_{k}(u)sink\omega_0 t\right],$$

where

$$\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{0}(u)=\begin{bmatrix} B^{(\xi_1)}_{0}(u) & \frac{1}{2}\left[B^{(\xi_1\xi_2)}_{0}(u)+B^{(\xi_2\xi_1)}_{0}(u)\right]\\ \frac{1}{2}\left[B^{(\xi_1\xi_2)}_{0}(u)+B^{(\xi_2\xi_1)}_{0}(u)\right] & B^{(\xi_2)}_{0}(u)\end{bmatrix},$$

$$\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{0}(u)=\begin{bmatrix} 0 & \frac{1}{2}\left[B^{(\xi_1\xi_2)}_{0}(u)-B^{(\xi_2\xi_1)}_{0}(u)\right]\\ \frac{1}{2}\left[B^{(\xi_1\xi_2)}_{0}(u)-B^{(\xi_2\xi_1)}_{0}(u)\right] & 0\end{bmatrix},$$

$$\tilde{\boldsymbol{C}}^{(\vec{\xi})}_{k}(u)=\begin{bmatrix} C^{(\xi_1)}_{k}(u) & \frac{1}{2}\left[C^{(\xi_1\xi_2)}_{k}(u)+C^{(\xi_2\xi_1)}_{k}(u)\right]\\ \frac{1}{2}\left[C^{(\xi_1\xi_2)}_{k}(u)+C^{(\xi_2\xi_1)}_{k}(u)\right] & C^{(\xi_2)}_{k}(u)\end{bmatrix},$$

$$\tilde{\boldsymbol{C}}^{(\vec{\xi})}_{k}(u)=\begin{bmatrix} 0 & \frac{1}{2}\left[C^{(\xi_1\xi_2)}_{k}(u)-C^{(\xi_2\xi_1)}_{k}(u)\right]\\ \frac{1}{2}\left[C^{(\xi_2\xi_1)}_{k}(u)-C^{(\xi_1\xi_2)}_{k}(u)\right] & 0\end{bmatrix},$$

$$\tilde{\boldsymbol{S}}^{(\vec{\xi})}_{k}(u)=\begin{bmatrix} S^{(\xi_1)}_{k}(u) & \frac{1}{2}\left[S^{(\xi_1\xi_2)}_{k}(u)+S^{(\xi_2\xi_1)}_{k}(u)\right]\\ \frac{1}{2}\left[S^{(\xi_1\xi_2)}_{k}(u)+S^{(\xi_2\xi_1)}_{k}(u)\right] & S^{(\xi_2)}_{k}(u)\end{bmatrix},$$

$$\tilde{\boldsymbol{S}}^{(\vec{\xi})}_{k}(u)=\begin{bmatrix} 0 & \frac{1}{2}\left[S^{(\xi_1\xi_2)}_{k}(u)-S^{(\xi_2\xi_1)}_{k}(u)\right]\\ \frac{1}{2}\left[S^{(\xi_1\xi_2)}_{k}(u)-S^{(\xi_2\xi_1)}_{k}(u)\right] & 0\end{bmatrix},$$

$$\text{and}\,\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{k}(u)=\frac{1}{2}\left[\tilde{\boldsymbol{C}}^{(\vec{\xi})}_{k}(u)-\mathrm{i}\tilde{\boldsymbol{S}}^{(\vec{\xi})}_{k}(u)\right],\ \tilde{\boldsymbol{b}}^{(\vec{\xi})}_{k}(u)=\frac{1}{2}\left[\tilde{\boldsymbol{C}}^{(\vec{\xi})}_{k}(u)-\mathrm{i}\tilde{\boldsymbol{S}}^{(\vec{\xi})}_{k}(u)\right],\ k\neq 0. \tag{36}$$

Conditions (33)–(35) for symmetric and asymmetric parts of tensor-components are written as

$$\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{k}(-u)=\tilde{\boldsymbol{b}}^{(\vec{\xi})}_{k}(u)e^{-\mathrm{i}k\omega_0 u},\ \tilde{\boldsymbol{C}}^{(\vec{\xi})}_{k}(-u)=\tilde{\boldsymbol{C}}^{(\vec{\xi})}_{k}(u)\cos\,k\omega_0 u-\tilde{\boldsymbol{S}}^{(\vec{\xi})}_{k}(u)\sin\,k\omega_0 u, \tag{37}$$

$$\tilde{\boldsymbol{S}}_k^{(\vec{\xi})}(-u) = \tilde{\boldsymbol{C}}_k^{(\vec{\xi})}(u)\sin k\omega_0 u + \tilde{\boldsymbol{S}}_k^{(\vec{\xi})}(u)\cos k\omega_0 u, \tag{38}$$

$$\tilde{\boldsymbol{b}}_k^{(\vec{\xi})}(u) = -\tilde{\boldsymbol{b}}_k^{(\vec{\xi})}(u)e^{-\mathrm{i}k\omega_0 u},\ \tilde{\boldsymbol{C}}_k^{(\vec{\xi})}(-u) = -\tilde{\boldsymbol{C}}_k^{(\vec{\xi})}(u)\cos k\omega_0 u + \tilde{\boldsymbol{S}}_k^{(\vec{\xi})}(u)\sin k\omega_0 u, \tag{39}$$

$$\tilde{\boldsymbol{S}}_k^{(\vec{\xi})}(-u) = -\tilde{\boldsymbol{C}}_k^{(\vec{\xi})}(u)\sin k\omega_0 u - \tilde{\boldsymbol{S}}_k^{(\vec{\xi})}(u)\cos k\omega_0 u. \tag{40}$$

It can be easily seen, that linear invariants of the Fourier tensor-components of symmetric part

$$B_k^{(I_1)}(u) = B_k^{(\xi_1)}(u) + B_k^{(\xi_2)}(u),$$
$$C_k^{(I_1)}(u) = C_k^{(\xi_1)}(u) + C_k^{(\xi_2)}(u),\ S_k^{(I_1)}(u) = S_k^{(\xi_1)}(u) + S_k^{(\xi_2)}(u),$$

are respective Fourier coefficients of linear invariant $I_1(t,u)$, i.e.

$$I_1(t,u) = \sum_{k\in Z} B_k^{(I_1)}(u)e^{-\mathrm{i}k\omega_0 t} = B_0^{(I_1)}(u) + \sum_{k\in N}\left[C_k^{(I_1)}(u)\cos k\omega_0 t + S_k^{(I_1)}(u)\sin k\omega_0 t\right].$$

They define the amplitudes and phases of harmonic components of the expected value of scalar product for vectors $\overset{\circ}{\vec{\xi}}(t)$ and $\overset{\circ}{\vec{\xi}}(t+u)$—covariance function of complementary components of vector $\vec{\xi}(t)$, and if $u=0$—the amplitudes and phases of harmonic components of vector $\vec{\xi}(t)$ fluctuations intensity.

It follows from (37) and (38) that

$$B_k^{(I_1)}(-u) = B_k^{(I_1)}(u)e^{-\mathrm{i}k\omega_0 u},\ C_k^{(I_1)}(-u) = C_k^{(I_1)}(u)\cos k\omega_0 u - S_k^{(I_1)}(u)\sin k\omega_0 u,$$
$$S_k^{(I_1)}(-u) = C_k^{(I_1)}(u)\sin k\omega_0 u + S_k^{(I_1)}(u)\cos k\omega_0 u.$$

Similarly, invariants

$$B_k^{(D)}(u) = B_k^{(\xi_1\xi_2)}(u) - B_k^{(\xi_2\xi_1)}(u),$$
$$C_k^{(D)}(u) = C_k^{(\xi_1\xi_2)}(u) - C_k^{(\xi_2\xi_1)}(u),\ S_k^{(D)}(u) = S_k^{(\xi_1\xi_2)}(u) - S_k^{(\xi_2\xi_1)}(u)$$

define harmonic parameters of the expected value of the skew product for vectors $\overset{\circ}{\vec{\xi}}(t)$ and $\overset{\circ}{\vec{\xi}}(t+u)$, i.e. correlations between orthogonal components of vectorial process $\overset{\circ}{\vec{\xi}}(t)$:

$$D(t,u) = \sum_{k\in Z} B_k^{(D)}(u)e^{\mathrm{i}k\omega_0 t} = B_0^{(D)}(u) + \sum_{k\in N}\left[C_k^{(D)}(u)\cos k\omega_0 t + S_k^{(D)}(u)\sin k\omega_0 t\right].$$

Taking into account (39) and (40) we have:

$$B_k^{(D)}(-u) = -B_k^{(D)}(u)e^{-ik\omega_0 u},\ C_k^{(D)}(-u) = -C_k^{(D)}(u)\cos k\omega_0 u + S_k^{(D)}(u)\sin k\omega_0 u,$$
$$S_k^{(D)}(-u) = -C_k^{(D)}(u)\sin k\omega_0 u - S_k^{(D)}(u)\cos k\omega_0 u.$$

Zero coefficient $B_0^{(I_1)}(u)$ is time average of the invariant $I_1(t,u)$. It is an even function of lag u: $B_0^{(I_1)}(-u) = B_0^{(I_1)}(u)$. Zero coefficient $B_0^{(D)}(u)$ is time average of the invariant $D(t,u)$. It is odd function: $B_0^{(D)}(-u) = -B_0^{(D)}(u)$. These quantities can be considered as linear invariants of the stationary approximation of the vector PCRP [2].

Note, that Fourier coefficients of the symmetrical part of covariance tensor-function are tensor-functions of lag u and we can also analyze their properties on the base of their quadratic invariants. For example, for zero tensor-component (35) we have

$$I_2^0(u) = B_0^{(\xi_1)}(u)B_0^{(\xi_2)}(u) - \frac{1}{4}\left[B_0^{(\xi_1\xi_2)}(u) + B_0^{(\xi_1\xi_2)}(u)\right]^2. \tag{41}$$

This quantity is a quadratic form discriminant

$$g(x,y) = B_0^{(\xi_1)}(u)x^2 + \left[B_0^{(\xi_1\xi_2)}(u) + B_0^{(\xi_2\xi_1)}(u)\right]xy + B_0^{(\xi_2)}(u)y^2. \tag{42}$$

Proposition 5.2 *The quadratic invariant (41) at* $u = 0$ *is positive definite* $I_2^{(0)}(0) \geq 0$.

Proof For $u = 0$ we have

$$I_2^{(0)}(0) = B_0^{(\xi_1)}(0)B_0^{(\xi_2)}(0) - \left[B_0^{(\xi_1\xi_2)}(0)\right]^2. \tag{43}$$

Consider the inequality

$$\left(\alpha \overset{\circ}{\underset{1}{\xi}}(t) + \overset{\circ}{\underset{2}{\xi}}(t)\right)^2 \geq 0.$$

For its expected value we have

$$E\left(\alpha \overset{\circ}{\underset{1}{\xi}}(t) + \overset{\circ}{\underset{2}{\xi}}(t)\right)^2 = \alpha^2 b_{\xi_1}(t,0) + 2\alpha b_{\xi_1\xi_2}(t,0) + b_{\xi_2}(t,0) \geq 0.$$

Taking into account Fourier series (32) after time-averaging we obtain

$$\alpha^2 B_{\xi_1}(0) + 2\alpha B_{\xi_1\xi_2}(0) + B_{\xi_2}(0) \geq 0.$$

Hence it follows that

$$\left[B_0^{(\xi_1\xi_2)}(0)\right]^2 \le B_{\xi_1}(0)B_{\xi_2}(0)$$

and accordingly $I_2^{(0)}(0) \ge 0$.

Corollary 5.2 *The quadratic form at* $u=0$ *is positive definite* $[g(x,y)]_{u=0} \ge 0$.

Formula (42) can be written as

$$g(x,y) = \frac{1}{B_0^{(\xi_1)}(u)}\left[\left[B_0^{(\xi_1)}(u)x + \frac{1}{2}\left[B_0^{(\xi_1\xi_2)}(u) + B_0^{(\xi_2\xi_1)}(u)\right]y\right]^2 + I_2^{(0)}(u)y^2\right],$$

when $B_0^{(\xi_1)}(u) \ne 0$ and in the form

$$g(x,y) = \frac{1}{B_0^{(\xi_2)}(u)}\left[\left[\frac{1}{2}\left[B_0^{(\xi_1\xi_2)}(u) + B_0^{(\xi_2\xi_1)}(u)\right]x + B_0^{(\xi_2)}(u)y\right]^2 + I_2^{(0)}(u)x^2\right].$$

when $B_0^{(\xi_2)}(u) \ne 0$. It follows from these equalities that $[g(x,y)]_{u=0} \ge 0$.

The matrix (36) is real and symmetric. Taking into consideration the results obtained in Sect. 4 now we formulate the following corollaries:

Corollary 5.3 *The symmetrical part of the tensor-component function* $\tilde{\boldsymbol{b}}_0^{(\vec{\xi})}(u)$ *in its own basis Osv, obtained by rotation of the old system Oxy by angle*

$$\alpha = \frac{1}{2}arctg\frac{B_0^{(\xi_1\xi_2)}(u) + B_0^{(\xi_2\xi_1)}(u)}{B_0^{(\xi_1)}(u) - B_0^{(\xi_2)}(u)}$$

is represented by the diagonal matrix

$$\tilde{\boldsymbol{b}}_0^{(\vec{\xi})}(u) = \begin{bmatrix} \lambda_1^{(0)}(u) & 0 \\ 0 & \lambda_2^{(0)}(u) \end{bmatrix},$$

where $\lambda_1^{(0)}(u)$ *and* $\lambda_2^{(0)}(u)$ *are its eigen-values, defined by expression*

$$\lambda_{1,2}^{(0)}(u) = \frac{1}{2}\left[I_1^{(0)}(u) \pm \sqrt{\left[I_1^{(0)}(u)\right]^2 - 4I_2^{(0)}(u)}\right] \tag{44}$$

and $I_1^{(0)}(u) = B_0^{(\xi_1)}(u) + B_0^{(\xi_2)}(u)$.

Corollary 5.4 *The quadratic curve*

$$\lambda_1^{(0)}(u)s^2+\lambda_2^{(0)}(u)v^2=1 \tag{45}$$

is an ellipse or a circle if $I_2^{(0)}(u)>0$ *and it is a hyperbola if* $I_2^{(0)}(u)<0$. *If* $I_2^{(0)}(u)=0$, *curve (45) degenerates into a straight line*.

Since

$$\left[I_1^{(0)}(u)\right]^2-4I_2^{(0)}(u)=\left[B_0^{(\xi_1)}(u)-B_0^{(\xi_2)}(u)\right]^2+\left[B_0^{(\xi_1\xi_2)}(u)+B_0^{(\xi_2\xi_1)}(u)\right]^2\geq 0,$$

then it follows from (44) that $\lambda_1(u)>0$ and $\lambda_2(u)>0$ if $I_2^{(0)}(u)>0$, and $\lambda_1(u)>0$ but $\lambda_2(u)<0$ if $I_2^{(0)}(u)<0$. For $I_2^{(0)}(u)=0$ we obtain $\lambda_1^{(0)}(u)=\frac{1}{2}I_1^{(0)}(u)$ and $\lambda_2^{(0)}(u)=0$.

Corollary 5.5 *Quadratic curve (45) at* $u=0$ *is an ellipse or a circle if* $I_2(0)\neq 0$.

Note. *It follows from (43) that* $I_2(0)=0$ *only in the case in which the variance of one of the vector components* $\xi_1(t)$ *or* $\xi_2(t)$ *is equal to zero*.

Taking into account relationships (24), (30) and (31) we can obtain formulae for Fourier coefficients of nonlinear invariant $I_2(t,u)$.

Since

$$b_{\xi_p}(t,u)b_{\xi_q}(t,u)=\sum_{k\in Z}e^{\mathrm{i}k\omega_0 t}\sum_{l\in Z}B_l^{\left(\xi_p\right)}(u)\bar{B}_{l-k}^{\left(\xi_q\right)}(u),$$

$$b_{\xi_p\xi_q}(t,u)b_{\xi_q\xi_p}(t,u)=\sum_{k\in Z}e^{\mathrm{i}k\omega_0 t}\sum_{l\in Z}B_l^{\left(\xi_p\xi_q\right)}(u)\bar{B}_{l-k}^{\left(\xi_q\xi_p\right)}(u),$$

so

$$I_2(t,u)=\sum_{k\in Z}B_k^{(I_2)}(u)e^{\mathrm{i}k\omega_0 t}, \tag{46}$$

where

$$B_k^{(I_2)}(u)=\sum_{l\in Z}\left[B_l^{(\xi_1)}(u)\bar{B}_{l-k}^{(\xi_2)}(u)-\frac{1}{4}\left[\begin{array}{c}B_l^{(\xi_1\xi_2)}(u)\bar{B}_{l-k}^{(\xi_1\xi_2)}(u)+\\+B_l^{(\xi_2\xi_1)}(u)\bar{B}_{l-k}^{(\xi_2\xi_1)}(u)-\\-2B_l^{(\xi_1\xi_2)}(u)\bar{B}_{l-k}^{(\xi_2\xi_1)}(u)\end{array}\right]\right]. \tag{47}$$

Thus we can formulate the following proposition:

Proposition 5.2 *The quadratic invariant* $I_2(t,u)$ *is represented by Fourier series (46), the coefficients of which are defined by formula (47).*

Note, that the expression defining a quadratic invariant is nonlinear, that's why the formula for $B_0^{(I_2)}(u)$ *differs from expression (41) for* $I_2^{(0)}(u)$ *that defines the quadratic invariant of the symmetrical part of tensor-component function* $\boldsymbol{b}_0^{(\vec{\xi})}(u)$.

Corollary 5.6 *Zero Fourier coefficient* $B_0^{(I_2)}(u)$ *at* $u=0$ *is positive definite* $B_0^{(I_2)}(u) \geq 0$.

We get the inequality $B_0^{(I_2)}(u) \geq 0$ after time-averaging of inequality $I_2(t,0) \geq 0$.

It is difficult to obtain the formulae for Fourier coefficients of invariants $\lambda_1(t,u)$ and $\lambda_2(t,u)$ in common case because they are defined by a square root of the invariants $I_1(t,u)$ and $I_2(t,u)$. But using the equalities $\lambda_1(t,u)+\lambda_2(t,u)=I_1(t,u)$ and $\lambda_1(t,u)\lambda_2(t,u)=I_2(t,u)$ we easily obtain the nonlinear equations determined these coefficients.

Taking into account Fourier series

$$\lambda_1(t,u)=\sum_{k\in Z} B_k^{(\lambda_1)}(u)e^{ik\omega_0 t},\ \lambda_2(t,u)=\sum_{k\in Z} B_k^{(\lambda_2)}(u)e^{ik\omega_0 t} \tag{48}$$

we have proposition:

Proposition 5.3 *The invariants* $\lambda_1(t,u)$ *and* $\lambda_2(t,u)$ *are represented by Fourier series (48) and their Fourier coefficients are defined by equations*

$$B_k^{(\lambda_1)}(u)+B_k^{(\lambda_2)}(u)=B_k^{(I_1)}(u), \tag{49}$$

$$\sum_{l\in Z} B_{l+k}^{(\lambda_1)}(u)\bar{B}_l^{(\lambda_2)}(u)=B_k^{(I_2)}(u). \tag{50}$$

Zero coefficients $B_0^{(\lambda_1)}(u)$ and $B_0^{(\lambda_2)}(u)$ determine the time average of the parameters of the quadratic curve (28). We should note that the non-linearities of Eqs. (49) and (50) cause that these quantities do not coincide with the invariants of zero tensor-component function $\tilde{B}_0^{(\vec{\xi})}(u)$, defined by expression (44).

6 Invariant Covariance Analysis of Modulated Signals

Covariance tensor-components structure can be specified using the harmonic representation of vector components (4) and (5). So, in the general case we have:

$$B_k^{(I)}(u)=\sum_{l\in Z}\left[R_{l-k,l}^{(\xi_1)}(u)+R_{l-k,l}^{(\xi_2)}(u)\right]e^{\mathrm{i}k\omega_0 u}, \tag{51}$$

$$B_k^{(D)}(u)=\sum_{l\in Z}\left[R_{l-k,l}^{(\xi_1\xi_2)}(u)-R_{l-k,l}^{(\xi_2\xi_1)}(u)\right]e^{\mathrm{i}k\omega_0 u}. \tag{52}$$

Equations (51) and (52) can be significantly simplified for a quadrature model, if only the first jointly stationary components in relationships (4) and (5) are non-zero

$$\xi_{\pm 1}^{(1)}(t) = \frac{1}{2}[\mu_c(t) \mp \mathrm{i}\mu_s(t)],\ \xi_{\pm 1}^{(2)}(t) = \frac{1}{2}[\nu_c(t) \mp \mathrm{i}\nu_s(t)].$$

So vector $\vec{\xi}(t)$ components are written as:

$$\xi_1(t) = \mu_c(t)\cos\omega_0 t + \mu_s(t)\sin\omega_0 t,\ \xi_2(t) = \nu_c(t)\cos\omega_0 t + \nu_s(t)\sin\omega_0 t.$$

On the base of these expressions we obtain

$$b_{\xi_1}(t,u) = B_0^{(\xi_1)}(u) + C_2^{(\xi_1)}(u)\cos 2\omega_0 t + S_2^{(\xi_1)}(u)\sin 2\omega_0 t,$$
$$b_{\xi_2}(t,u) = B_0^{(\xi_2)}(u) + C_2^{(\xi_2)}(u)\cos 2\omega_0 t + S_2^{(\xi_2)}(u)\sin 2\omega_0 t,$$
$$b_{\xi_p\xi_q}(t,u) = B_0^{(\xi_p\xi_q)}(u) + C_2^{(\xi_p\xi_q)}(u)\cos 2\omega_0 t + S_2^{(\xi_p\xi_q)}(u)\sin 2\omega_0 t,$$

and

$$B_0^{(\xi_1)}(u) = \frac{1}{2}\left[R_{\mu_c}(u) + R_{\mu_s}(u)\right]\cos\omega_0 + R^{-}_{\mu_c\mu_s}(u)\sin\omega_0 u, \tag{53}$$

$$C_2^{(\xi_1)}(u) = \frac{1}{2}\left[R_{\mu_c}(u) - R_{\mu_s}(u)\right]\cos\omega_0 + R^{+}_{\mu_c\mu_s}(u)\sin\omega_0 u, \tag{54}$$

$$S_2^{(\xi_1)}(u) = R^{+}_{\mu_c\mu_s}(u)\cos\omega_0 u + \frac{1}{2}\left[R_{\mu_s}(u) - R_{\mu_c}(u)\right]\sin\omega_0 u, \tag{55}$$

$$B_0^{(\xi_2)}(u) = \frac{1}{2}[R_{\nu_c}(u) + R_{\nu_s}(u)]\cos\omega_0 + R^{-}_{\nu_c\nu_s}(u)\sin\omega_0 u, \tag{56}$$

$$C_2^{(\xi_2)}(u) = \frac{1}{2}[R_{\nu_c}(u) - R_{\nu_s}(u)]\cos\omega_0 + R^{+}_{\nu_c\nu_s}(u)\sin\omega_0 u, \tag{57}$$

$$S_2^{(\xi_2)}(u) = R^{+}_{\nu_c\nu_s}(u)\cos\omega_0 u + \frac{1}{2}[R_{\nu_s}(u) - R_{\nu_c}(u)]\sin\omega_0 u, \tag{58}$$

$$B_0^{(\xi_1\xi_2)}(u) = \frac{1}{2}\left[R_{\mu_c\nu_c}(u) + R_{\mu_s\nu_s}(u)\right]\cos\omega_0 + \frac{1}{2}\left[R_{\mu_c\nu_s}(u) - R_{\mu_s\nu_c}(u)\right]\sin\omega_0 u, \tag{59}$$

$$C_2^{(\xi_1\xi_2)}(u) = \frac{1}{2}\left[R_{\mu_c\nu_c}(u) - R_{\mu_s\nu_s}(u)\right]\cos\omega_0 + \frac{1}{2}\left[R_{\mu_c\nu_s}(u) + R_{\mu_s\nu_c}(u)\right]\sin\omega_0 u, \tag{60}$$

$$S_2^{(\xi_1\xi_2)}(u) = \frac{1}{2}\left[R_{\mu_c\nu_s}(u) + R_{\mu_s\nu_c}(u)\right]\cos\omega_0 + \frac{1}{2}\left[R_{\mu_s\nu_s}(u) - R_{\mu_c\nu_c}(u)\right]\sin\omega_0 u, \tag{61}$$

where $R_{\mu_{c,s}}(u)$, $R_{\nu_{c,s}}(u)$, $R_{\mu_{c,s}\nu_{c,s}}(u)$ are auto- and cross-covariance functions of random processes $\mu_c(t)$, $\mu_s(t)$, $\nu_c(t)$ and $\nu_s(t)$, and $R^{\pm}_{\mu_c\mu_s}(u)$, $R^{\pm}_{\nu_c\nu_s}(u)$—odd and even parts of the cross-covariance functions. Formulae for covariance components $B_0^{(\xi_2\xi_1)}(u)$, $C_2^{(\xi_2\xi_1)}(u)$ and $S_2^{(\xi_2\xi_1)}(u)$ are similar to (59)–(61) with the difference in order of correlations between modulating processes $\mu_c(t)$, $\mu_s(t)$ and $\nu_c(t)$, $\nu_s(t)$.

On the base of ratios (53)–(61) we obtain expressions for invariant $I_1(t,u)$ covariance components:

$$B_0^{(I)}(u)=\frac{1}{2}\left[\begin{array}{l}\left[R_{\mu_c}(u)+R_{\nu_c}(u)\right]+\\ +\left[R_{\mu_s}(u)+R_{\nu_s}(u)\right]\end{array}\right]\cos\omega_0 u+\left[R^{-}_{\mu_c\mu_s}(u)+R^{-}_{\nu_c\nu_s}(u)\right]\sin\omega_0 u,$$

$$C_2^{(I)}(u)=\frac{1}{2}\left[\begin{array}{l}\left[R_{\mu_c}(u)-R_{\mu_s}(u)\right]+\\ +\left[R_{\nu_c}(u)-R_{\nu_s}(u)\right]\end{array}\right]\cos\omega_0 u++\left[R^{+}_{\mu_c\mu_s}(u)+R^{+}_{\nu_c\nu_s}(u)\right]\sin\omega_0 u,$$

$$S_2^{(I)}(u)=\left[R^{+}_{\mu_c\mu_s}(u)+R^{+}_{\nu_c\nu_s}(u)\right]\cos\omega_0 u+\frac{1}{2}\left[\begin{array}{l}\left[R_{\mu_c}(u)-R_{\mu_s}(u)\right]+\\ +\left[R_{\nu_c}(u)-R_{\nu_s}(u)\right]\end{array}\right]\sin\omega_0 u.$$

For the zeros lags $u=0$ we have:

$$B_0^{(I)}(0)=\frac{1}{2}\left[R_{\mu_c}(0)+R_{\nu_c}(0)+R_{\mu_s}(0)+R_{\nu_s}(0)\right],$$

$$C_2^{(I)}(0)=\frac{1}{2}\left[R_{\mu_c}(0)-R_{\mu_s}(0)+R_{\nu_c}(0)-R_{\nu_s}(0)\right],\ S_2^{(I)}(0)=R^{+}_{\mu_c\mu_s}(0)+R^{+}_{\nu_c\nu_s}(0).$$

As it follows from the obtained expressions the value of the zero covariance component $B_0^{(I)}(u)$ and the absolute value of the second covariance component $\left|B_2^{(I)}(u)\right|$ increase comparatively with the values of the quantities that characterize the properties of covariance functions of the individual vector components, i.e. comparatively with the last time-averaged value of invariant $I_1(t,u)$ and depth of its time variation increase. It allows one to estimate the state of the system that generate vectorial signal $\vec{\xi}(t)$ with a higher efficiency.

Covariance components of invariant $D(t,u)$ have the form:

$$B_0^{(D)}(u)=\left[R^{-}_{\mu_c\nu_c}(u)+R^{-}_{\mu_s\nu_s}(u)\right]\cos\omega_0 u+\left[R^{+}_{\mu_c\nu_s}(u)+R^{-}_{\mu_s\nu_s}(u)\right]\sin\omega_0 u,$$

$$C_2^{(D)}(u)=\left[R^{-}_{\mu_c\nu_c}(u)+R^{-}_{\nu_s\mu_s}(u)\right]\cos\omega_0 u+\left[R^{-}_{\mu_s\nu_c}(u)+R^{-}_{\mu_c\nu_s}(u)\right]\sin\omega_0 u,$$

$$S_2^{(D)}(u)=\left[R^{-}_{\mu_s\nu_c}(u)+R^{-}_{\mu_c\nu_s}(u)\right]\cos\omega_0 u+\left[R^{-}_{\mu_c\nu_c}(u)+R^{-}_{\nu_s\mu_s}(u)\right]\sin\omega_0 u,$$

where signs "+" and "–" determine the even and odd correlations respectively. The feature of the second covariance components is that they are defined only by the odd components of cross-covariance functions of processes that modulate harmonics of orthogonal components of vector $\vec{\xi}(t)$. This fact, for example, can be

used in vibration diagnosis for separation of the moving and fixed defects, since the character of odd correlations for both types of defects, typical of vibration signals, is different.

7 The Examples of Using the Covariance Invariants for Vibration Analysis

Vibration oscillations measured at different points on mechanism are different. For example, in Fig. 3 the graphs of the zeroth covariance component estimators for vibration acceleration measured at different points on bearing housing are shown. In the first as well as in the second case the graphs have the form of damping oscillation but these oscillations significantly differ by their power and parameters of their individual components. It means that analyzing oscillations measured only in the one direction we may significantly impair the opportunity of damage detection. So it is reasonable to provide measurement of oscillations at several points simultaneously and to use for the analysis such their common characteristics which do not depend on coordinate system where measurements were provided.

Let us bring some results of the analysis of rolling bearing vibrations as a vectorial PCRP. In order to investigate changes of their invariant characteristics as

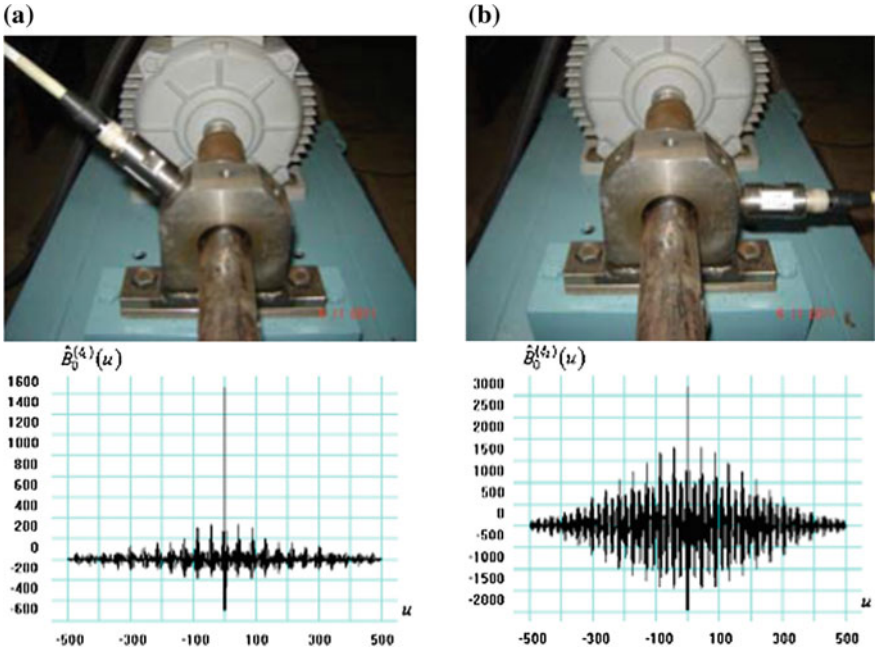


Fig. 3 Estimators of the zeroth covariance components for vibration accelerations measured at angle 135° (**a**) and 0° (**b**)

bearing state degradation the experiment on the test bench was provided. The signals of vibration acceleration were measured in two perpendicular directions with using the vibration diagnosis system Vector created in Karpenko Physico-Mechanical Institute of NAS of Ukraine. In order to increase bearing damaging the outer race of the bearing was skewed.

We should note that invariant characteristics for some type of damages can be built on the base of the mean function of vectorial PCRP. Among them, for example, can be parameters of the curve in the rectangular coordinate system Oxy defined by functions which describe components of vector $\vec{m}_{\underrightarrow{\xi}}(t)$:

$$x_l(t) = A_l \sin(l\omega_0 t + \varphi_l),\ y_l(t) = B_l \sin(l\omega_0 t + \psi_l).$$

Curves built on the base of the first harmonics of differently directed vector components are presented in Fig. 4. These harmonics parameters were calculated using the component method [15, 16] at the beginning of the experiment. As it can be seen from Fig. 4 these curves have a form of ellipses which parameters and orientation do not change under coordinate system rotation (coordinate axes were chosen on direction of mounted sensors (Fig. 4a)).

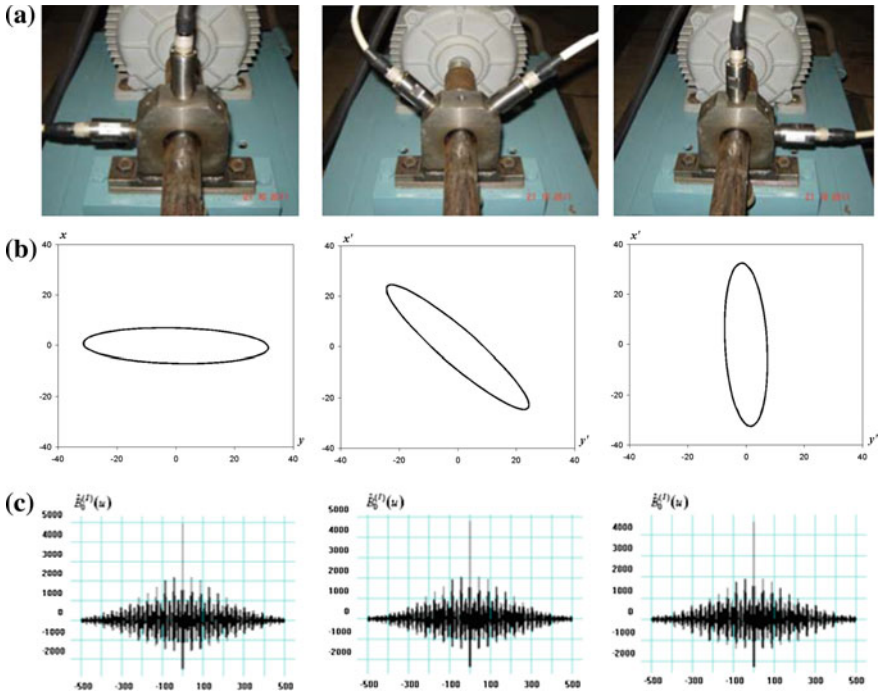


Fig. 4 Estimators of the vibration signals invariants: **a** position of sensors on bearing housing; **b** the first harmonic of the mean function vector; **c** the zeroth tensor-component $\hat{B}_0^{(l)}(u)$

Invariant characteristics were computed by formulas which define them (Sect. 4), i.e. on the base of estimators of auto- and cross-covariance functions of vector $\vec{\xi}(t)$. These functions were estimated by the component method [15, 16].

The graphs of the zeroth tensor-component $\hat{B}_0^{(I)}(u)$ estimators for different positions of the sensors are shown in Fig. 4c. They prove independence of this quantity of the chosen coordinate system.

The second order curves built for the fixed time lag u at the beginning of the experiment have a form of concentric circles which radii slightly differ.

Diagram that illustrates the correlations changes with respect to the direction has a similar form. Such changes are defined by function $b_\beta(t,u)$ (formula (29)).

When the damage appear the second order curves degrade into ellipses. The bigger axe of ellipses increases as damage grows and has its own orientation (Fig. 5a). The diagram of changes of correlation with respect to direction has a similar orientation (Fig. 5b). As it was found after bearing demounting and its cutting this orientation showed the place of the outer race damage (Fig. 6).

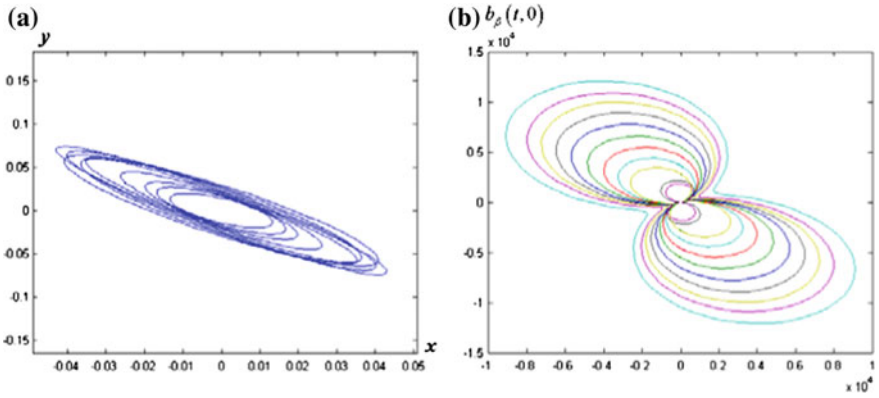


Fig. 5 The second order curves (**a**) and changes of correlations with respect to direction (**b**) at $u = 0$

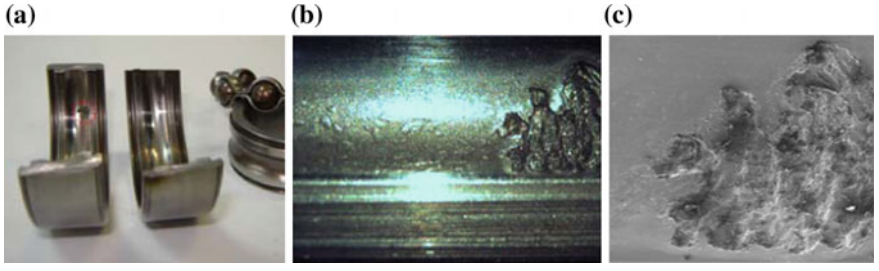


Fig. 6 Damage on the bearing outer race: **a** bearing elements; **b** damage zoomed 10 times; **c** damage zoomed 20 times

Based on the results of carried out experiment we can state that the analysis of vibrations using the vectorial PCRP methods open new opportunities for improvement of the damages detection efficiency and also their localization and typification.

8 Conclusions

Methods of vectorial PCRP are suitable for a common analysis of vector components time variation for which recurrence and stochasticity are typical. The feature of this approach is the possibility to analyze the properties of random vector invariant quantities independently of the coordinate system, where signals were measured in. This feature is very important for the solution of vibration diagnosis issues. Use of the linear invariants of covariance tensor function, defined by the expected values of scalar and skew products of random vectors respectively, allows one to improve the efficiency of diagnosis, separate moving and fixed defects. Using quadratic invariants we can investigate the spatial properties of vibration, compute changes of correlations in arbitrary direction and localize a defect. The properties of the central quadratic curves, which axes are defined by eigen-values of symmetric part of covariance tensor-function, are effective features for defects localization too. These eigen-vales determine the maximal values of covariance function by the axis direction in the own basis.

References

1. Belyshev, A. P., Klevantsov, Y., & Rozhkov, V. A. (1983). *Probability analysis of sea currents*. Leningrad: Gidrometeoizdat. (in Russian).
2. Ya, D., Rozhkow, V., & Javors'kyj, I. (1987). *The methods of probabilistic analysis of oceanological rhythms*. Leningrad: Gidrometeoizdat. (in Russian).
3. Kochin, V. E. (1961). *Vectorial computation and basics of tensor calculus*. Moscow: Izdatelstvo AN SSSR. (in Russian).
4. Rozhkov, V. A. (1974). *The methods of probability analysis of oceanologic processes*. Leningrad: Gidrometeoizdat. (in Russian).
5. Javors'kyj, I., Matsko, I., Yuzefovych, R., et al. (2014). Vectorial diagnosis of rolling bearing with growing defect on the outer race. *Vibratsiyi v technitsi i technolohiyah, 2*(76), 101–110. (in Ukrainian).
6. Javors'kyj, I., Yuzefovych, R., Kravets, I., et al. (2013). Information-measuring system for multi-dimentional vibration diagnosis. *Problemy mashynobuduvannya, 16*(3), 45–50. (in Ukrainian).
7. Ya, D. (1969). On periodically correlated random processes and systems with periodic parameters. *Otbor i Peredacha Informatsiji, 30,* 16–24. (in Russian).
8. Ya, D. (1975). On representation of periodically correlated random processes by stationary components. *Otbor i Peredacha Informatsiyi, 45,* 7–20. (in Russian).
9. Ya, D., & Javors'kyj, I. (1982). *Rhythms of sea waves and underwater acoustic signals*. Kyiv: Naukova Dumka. (in Russian).

10. Hurd, H. L. (1989). Representation of strongly harmonizable periodically correlated processes and their covariances. *Journal of Multivariate Analysis, 29*(1), 53–67.
11. Hurd, H. L., & Miamme, A. (2007). *Periodically correlated random sequences, spectral theory and practice*. New York: Wiley.
12. Ogura, H. (1971). Spectral representation of periodic nonstationary random processes. *IEEE Transactions Information Theory*, 20, IT-17, 143–149.
13. Yavorskii, B., & Detlaf, A. (1981). *Physics reference*. Moskow: Nauka. (in Russian).
14. Gantmacher, F. R. (1959). *The theory of matrix*. New York: Chelsea Publication Company.
15. Javors'kyj, I., Dehay, D., & Kravets, I. (2014). Component statistical analysis of second order hidden periodicities. *Digital Signal Processing, 26,* 50–70.
16. Javors'kyj, I., Isayev, I., Majevski, I., & Yuzefovych, R. (2010). Component covariance analysis for periodically correlated random processes. *Signal Processing, 90,* 1083–1102.

Periodically Correlated Sequences with Rational Spectra and PARMA Systems

Andrzej Makagon

Abstract An acronym PARMA is used in different configurations, we talk about PARMA systems, PARMA sequences, or PARMA models. This paper is a result of the author search to understand this complex world of PARMAs.

1 Introduction

Periodically correlated sequences (PC) are sequences that are obtained by listing elements of a multi-variate stationary sequence in a linear order. There is no surprise therefore that both theories are strictly related. VARMA models are representations of multi-variate stationary sequences in a form of vector difference equations (VDE). PARMA models result from nonhomogeneous (periodic) VDE representations and under mild conditions yield periodically correlated sequences. PARMA models form a subset of VARMA models. Since only stationary sequences with rational densities admit VARMA models, it is natural that the study of PARMA models should involve an analysis of periodically correlated sequences with rational densities.

Sequences with rational densities play an important role in the theory of multi-variate stationary sequences. These are the only multi-variate stationary sequences for which the theoretical prediction problem has an explicit solution (under small additional assumptions), that is the only multi-variate stationary sequences for which it is possible to explicitly compute the coefficients of the innovation representation of the sequence from its density. This solution, however, is not fully satisfactory since it depends on infinitely many parameters (i.e. innovation coefficients). A VARMA model is an ingenious concept of reducing the number of parameters. The idea is to represent a sequence as a solution of a vector difference equations that involves only finitely many terms. In spectral domain this translates to a problem of writing a transfer function of the sequence (i.e. a square factor of its density) as a "quotient" of two polynomial matrices. The main purpose of this paper is to transfer these and other

A. Makagon (✉)
Department of Mathematics, Hampton University, Hampton, VA 23668, USA
e-mail: andrzej.makagon@hamptonu.edu

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_9

relevant results known for multi-variate stationary sequences to those periodically correlated sequences given by PARMA models.

A paper is organized as follows. In Sect. 2 we provide notation and definitions needed in the sequel. Section 3 contains an extensive review of the theory of periodically correlated sequences and multi-variate stationary sequences that sometimes goes beyond the needs of this paper. A focus is on the relation between PC and T-variate stationary sequences and on sequences with rational densities. In Sect. 4 we discuss a relationship between PC sequences with rational densities and PARMA systems. The next Sect. 5 contains few remarks about PARMA models. In that section we limit our attention to full rank sequences; a general case of any rank sequences seems to be still open even for multi-variate stationary sequences.

Most of the facts about multi-variate stationary sequences come from [11], while for the theory of VARMA systems from [4, 5]. Up-to-date review of PARMA models and related topics, with emphasis on statistics, can be found in [2]. Sources of information about periodically correlated sequences will be cited as needed. To our best knowledge, periodically correlated sequences with rational densities have not been studied before.

2 Preliminaries

Sets and Matrices. In what follows C denotes the set of complex numbers, $D_{<r} = \{z \in C : |z| < r\}$ is an open disk of radius r, and $D_r = \{z \in C : |z| = r\}$ is a circle of radius r, $r > 0$. The interval $[0, 2\pi)$ will be understood as a group with addition modulo 2π and with standard Lebesgue measure structure. The abbreviation *a.e.* will mean *almost everywhere* with respect to the Lebesgue measure on $[0, 2\pi)$. A function on $[0, 2\pi)$ will be interpreted as a function of the unit circle D_1 and will be written as $f(e^{it})$ or $f(z)$, $z = e^{it}$, $t \in [0, 2\pi)$. This notation is convenient in analysis of sequences with rational densities and we use it all over the paper. The letter T will be always a fixed positive integer and $\mathcal{Z}$ will denote the set of integers. The symbols $\mathrm{q}(m)$ and $\langle m \rangle$ stand for the quotient and the nonnegative remainder in division of m by T, so that $m = \mathrm{q}(m)T + \langle m \rangle$. The entries of matrices and vectors in this paper are indexed from 0 instead from 1. The (i, j) entry of a matrix A will be denoted by $A^{i,j}$. For any matrix A, A^* is the complex conjugate of A, i.e. $(A^*)^{i,j} = \overline{A^{j,i}}$. If B is a square $n \times n$ matrix then the *adjugate* B^A of B is an $n \times n$ matrix whose (i, j) entry is given by $(B^A)^{ij} = (-1)^{i+j} \det(B_{[j,i]})$, where $B_{[j,i]}$ is obtained from B by deleting j-th row and i-th column. A square matrix B is invertible iff $\det(B) \neq 0$ and if it is then $B^{-1} = (\det(B))^{-1} B^A$. An $n \times n$ diagonal matrix A will be denoted $diag[A^{0,0}, A^{1,1}, \ldots A^{n-1,n-1}]$. A *minor of degree* k, $k \leq \min(m, n)$, of an $m \times n$ matrix A is the determinant of a square $k \times k$ sub-matrix of A obtained by deleting $m - k$ rows and $n - k$ columns from A. A rank of a matrix A is the maximum number of linearly independent columns or rows of A; $\mathrm{rank}(A) = r$ iff there is a minor of order r which is nonzero and all minors of bigger order are zero. A *principal minor* of order k of a square $n \times n$ matrix A is a minor

obtained by deleting last $n-k$ columns and the last $n-k$ rows. A square matrix A is non-negative ($A \geq 0$) if for every $a \in C^T$, $aAa^* \geq 0$; $A \geq 0$ iff all principal minors are non-negative. A $1 \times n$ matrix will be called a *vector*.

A *rational* $m \times n$ matrix $R(z)$ is a matrix whose entries $R(z)^{j,k}$ are rational functions of a complex variable $z \in C$; that is, each entry is a ratio of two polynomials. We will be *always* assuming that the entries of a rational matrix are written in the simplest form. Note that since minors of a rational matrix are rational functions, if a minor is non-zero for one z than it is nonzero for all $z \in C$ except finitely many points. Therefore the rank of a rational matrix is constant except for finitely many z's. The *poles* of a rational matrix are numbers z such that $R(z)$ does not exist and the zeros of $R(z)$ are complex numbers z at which the matrix $R(z)$ drops its rank (e.g. [5], or [7], Sect. 3.2.). A *polynomial* $m \times n$ matrix is a matrix whose entries are polynomials of a complex variable z. A matrix function $R(e^{it})$ on $[0, 2\pi)$ is called rational (or polynomial) if there is a rational (polynomial) matrix $R(z)$ on C such that $R(e^{it})$ equals $R(z)$ a.e. on D_1. If $P(z)$ is a polynomial matrix and $p(z) = \det P(z)$ is not identically zero, then $p(z) \neq 0$ for all $z \in C$ except finitely many, and $P(z)^{-1} = p(z)^{-1}P(z)^A$ is a rational matrix. The most important fact from the theory of rational matrices will be for us the following theorem proved by Rozanov.

Theorem 2.1 ([11], Theorem 10.1) *Each a.e. nonegative rational square matrix function* $F(e^{it})$ *on* $[0, 2\pi)$ *of rank r (i.e. rank of* $F(z)$ *is r) can be represented in the form* $F(e^{it}) = G(e^{it})G(e^{it})^*$ *a.e. where* $G(z)$ *is rational, analytic on* $D_{<1}$*, and the rank of* $G(z)$ *is r for* all $z \in D_{<1}$.

In terms of zeros and poles the last statement means that all zeros and poles of $G(z)$ are outside of the open unit disk $D_{<1}$.

A square polynomial matrix is $U(z)$ is called *unimodular* if $\det(U(z))$ is constant (i.e. does not depend on z). An $m \times m$ matrix $L(z)$ is called a *left (common) divisor* of $m \times n$ polynomial matrices $A(z)$ and $B(z)$ if there are $m \times n$ polynomial matrices $A_1(z)$ and $B_1(z)$ such that $A(z) = L(z)A_1(z)$ and $B(z) = L(z)B_1(z)$. Note that if $\det A(z)$ is not identically zero, then also $\det L(z)$ is not. A left divisor $L(z)$ is called a *greatest left divisor* of $A(z)$ and $B(z)$ if for any other left divisor $L_1(z)$ of $A(z)$ and $B(z)$ there is a polynomial $m \times m$ matrix $T(z)$ such that $L(z) = L_1(z)T(z)$. Polynomial matrices $A(z)$ and $B(z)$ are called *left coprime* if the only left divisors of $A(z)$ and $B(z)$ are the unimodular once. For those and other interesting facts from the theory of polynomial and rational matrices we refer the reader to [5] or [7], Sect. 3.2.

Hilbert Spaces. A Hilbert space will be denoted $\mathcal{H}$ (or $\mathcal{K}$), and (x, y) will denote the inner product of $x, y \in \mathcal{H}$ (or $\mathcal{K}$). All Hilbert spaces are assumed to be complex and separable. If $\mathcal{H}$ is a Hilbert space and M is a closed subspace of $\mathcal{H}$, then $(x|M)$ denotes the orthogonal projection of $x \in \mathcal{H}$ onto M. If S is any subset of $\mathcal{H}$ then $\overline{\text{sp}}\{S\}$ denote the smallest closed linear subspace of $\mathcal{H}$ containing S. A linear mapping (operator) from $\mathcal{H}$ onto $\mathcal{K}$ is called unitary if $(Ux, Uy)_{\mathcal{K}} = (x, y)_{\mathcal{H}}$ for every $x, y \in \mathcal{H}$. A sequence (ξ_n), $n \in \mathcal{Z}$, of elements of $\mathcal{H}$ is *orthonormal* if $(\xi_n, \xi_m) = 1$ if $m = n$, and zero otherwise. Important for us will be the space C^T of all row vectors of the length T with entries in C and standard Euclidean inner product, the Hilbert space L^2 of all measurable complex functions (in fact equivalence classes of functions) on $[0, 2\pi)$

which are square integrable w.r.t the Lebesgue measure dt on $[0, 2\pi)$, and the Hilbert space $L^2(C^T)$ of all C^T-valued functions f on $[0, 2\pi)$ such that their entries are in L^2. The inner product in $L^2(C^T)$ is $(f, g) = \int_0^{2\pi} f(e^{it})g(e^{it})^* dt$. The symbol L^2_+ will denote the subspace of L^2 consisting functions f whose Fourier coefficients with negative indices are zero, i.e. such that $f(e^{it}) = \sum_{k=0}^{\infty} f_k e^{itk}$; $L^2_+(C^T)$ denotes the subspace of functions in $L^2(C^T)$ such that their entries are in L^2_+. The *standard orthonormal basis* for C^T is $e_k, k = 0 \dots, T-1$, where e_k is the row vector that has 1 at the k-th place and zero otherwise. The *standard orthonormal basis* for $L^2(C^T)$ is the family of functions $\zeta_n(e^{it})$, $n \in \mathcal{Z}$, defined by $\zeta_{nT+r}(e^{it}) = (1/\sqrt{2\pi})e^{-int}e_r$, $n \in \mathcal{Z}$, $r = 0, \dots, T-1$. If $G(e^{it})$ is a matrix function then we say that G is square integrable if all entries of G are in L^2.

Stochastic Sequences. We adopt a Hilbert space approach. A (univariate) *stochastic sequence* $(x(n))$ in a Hilbert space $\mathcal{H}$ is a sequence of elements of $\mathcal{H}$ indexed by the set of all integers $\mathcal{Z}$. The *correlation function* of $(x(n))$ is the function on $\mathcal{Z}^2$ defined by $R_x(m, n) = (x(m), x(n))$. If $(x(n))$ is a stochastic sequence then we denote $M_x = \overline{\text{sp}}\{x(m) : m \in \mathcal{Z}\}$. Two stochastic sequences $(x(n))$ in $\mathcal{H}$ and $(y(n))$ in possibly different Hilbert space $\mathcal{K}$ are said to be *equivalent* if $R_x(m, n) = R_y(m, n)$ for every $m, n \in \mathcal{Z}$ or, equivalently, if there is a unitary mapping $\boldsymbol{\Phi}$ from M_x onto M_y such that $\boldsymbol{\Phi}(x(n)) = y(n), n \in \mathcal{Z}$. The concept of an equivalence of stochastic sequences makes the Hilbert space $\mathcal{H}$ appearing in the definition of a stochastic sequence irrelevant and we will stop writing it unless will be necessary. If T is a positive integer, then a *T-variate stochastic sequence in $\mathcal{H}$* is a family of T stochastic sequences $(x^k(n))$, $k = 0, \dots, T-1$, in $\mathcal{H}$. It is convenient to write it as a sequence of column vectors $\boldsymbol{x}(n) = [x^k(n)]$ with entries in $\mathcal{H}$ and $x^0(n)$ being at the top. The *correlation function* of a T-variate stochastic sequence $\boldsymbol{x}(n) = [x^k(n)]$, $n \in \mathcal{Z}$, is $T \times T$ matrix valued function $R_{\boldsymbol{x}}$ on $\mathcal{Z}^2$ defined as $R_{\boldsymbol{x}}(m, n)^{j,k} = (x^j(n), x^k(m))$. A T-variate sequence $\boldsymbol{\xi}_n = [\xi^j_n]$, $n \in \mathcal{Z}$, is called a *T-variate orthonormal sequence* if $R_{\boldsymbol{x}}(n, n) = I$, the identity matrix, and $R_{\boldsymbol{x}}(m, n) = 0$ for all $m \neq n$.

If $(\boldsymbol{x}(n))$ is a T-variate stochastic sequence then we define $M_{\boldsymbol{x}}(n) = \overline{\text{sp}}\{x^k(m) : k = 0, \dots, T-1, m \leq n\}$, $M_{\boldsymbol{x}} = M_{\boldsymbol{x}}(+\infty)$, and $N_{\boldsymbol{x}}(n) = M_{\boldsymbol{x}}(n) \ominus M_{\boldsymbol{x}}(n-1)$. A sequence $(\boldsymbol{x}(n)$ is called *regular* if $\bigcap_n M_{\boldsymbol{x}}(n) = \{0\}$. One of the main goals of prediction theory is to find an orthogonal projection $\hat{\boldsymbol{x}}(n) = (\boldsymbol{x}(n)|M_{\boldsymbol{x}}(n-1) := [(x^k(n)|M_{\boldsymbol{x}}(n-1)]$. In probabilistic terms $\hat{\boldsymbol{x}}(n)$ represents the best linear estimate (predictor) of a random vector $\boldsymbol{x}(n)$ from the past. All the above notation and definitions are valid for univariate sequences, that is when $T = 1$.

3 PCs and T-variate Stationary Sequences

A T-variate stochastic sequence $(\boldsymbol{x}(n))$ is called *stationary* if for every $n \in \mathcal{Z}$, $R_{\boldsymbol{x}}(n + r, r)$ is constant in $r \in \mathcal{Z}$. If it is so, then the function $K_{\boldsymbol{x}}(n) := R_{\boldsymbol{x}}(n, 0)$ is called the *(matrix) correlation function of a stationary sequence* $(\boldsymbol{x}(n))$. If $(\boldsymbol{x}(n))$ is T-variate stationary then the mapping $Ux^k(n) = x^k(n + 1)$, $n \in \mathcal{Z}$, $k = 0, \dots, T-1$, extends

linearly to a unitary operator on $M_{\boldsymbol{x}}$ which is called the *shift* of $(\boldsymbol{x}(n))$. A T-variate stochastic sequence $(\boldsymbol{x}(n))$ is completely described by its shift and the vector $\boldsymbol{x}(0)$; namely $\boldsymbol{x}(n) = U^n\boldsymbol{x}(0) := [U^n x^k(0)]$. If we write $U^n = \int_0^{2\pi} e^{-iun}E(du)$, $n \in \mathcal{Z}$, (e.g. [3]) then we obtain that $K_{\boldsymbol{x}}(n) = \int_0^{2\pi} e^{-iun}F(du)$ where F is a $T \times T$ nonnegative matrix measure on $[0, 2\pi)$ defined by $F^{j,k}(\Delta) = (E(\Delta)x^k(0), x^j(0))$. The measure F is called the *spectrum* of a T-variate stationary sequence $(\boldsymbol{x}(n))$.

An important example is a sequence $(H(n))$ of $T \times T$ matrix valued functions with rows in $L^2(C^T)$ defined as $H(n)(e^{it}) = e^{-int}H(e^{it})$, $n \in \mathcal{Z}$, $t \in [0, 2\pi)$. The sequence $(H(n))$ can be viewed as a T-variate stationary sequence in $\mathcal{H} = L^2(C^T)$. The k-th coordinate $H^k(n)$ of $H(n)$ is the k-th row $H^{k\cdot}(e^{it})$ of the matrix $H(e^{it})$ multiplied by e^{-int}, the shift of $(H(n))$ is multiplication by e^{-it}, the correlation function of $(H(n))$ is $K_H(n) = \int_0^{2\pi} e^{-int}H(e^{it})H(e^{it})^* dt$, so the spectrum F of $(H(n))$ is $F(dt) = H(e^{it})H(e^{it})^* dt$.

A (univariate) sequence $(x(n))$ is called *periodically correlated with period T* (we will abbreviate it T-PC), if for every $n \in \mathcal{Z}$ the sequence $R_x(n + r, r) = (x(n + r), x(r))$ is T-periodic in $r \in \mathcal{Z}$. The discrete Fourier transform of $R_x(n + r, r)$ with respect r will be denoted $a_j(n)$; more precisely

$$a_j(n) := \sum_{r=0}^{T-1} e^{-2\pi ijr/T}R_x(n + r, r), \qquad j = 0, \dots, T - 1. \tag{1}$$

If $(x(n))$ is T-PC then the mapping $Wx(n) = x(n + T)$, $n \in \mathcal{Z}$, extends linearly to a unitary operator in M_x which is called the T-*shift* of a T-PC sequence $(x(n))$. To describe a T-PC sequence we need $x(0)$ and two unitary operators; namely, a sequence $(x(n))$ is T-PC iff $(x(n))$ is of the form

$$x(n) = U^n \left[(1/T) \sum_{j=0}^{T-1} e^{-2\pi ijn/T} V^j x(0) \right], \qquad n \in \mathcal{Z}, \tag{2}$$

where U and V are unitary operators in some Hilbert space $\mathcal{K} \supseteq M_x$, $V^T = I$, and U, V satisfy a *canonical commutation relation* $VU = e^{-2\pi i/T}UV$. If we write operators U and V in terms of their spectral resolutions as follows $U^n = \int_0^{2\pi} e^{-iun}E(du)$, $n \in \mathcal{Z}$, and $V^j = \sum_{k=0}^{T-1} e^{2\pi ikj/T}P_k$, $j = 0, \dots, T - 1$, then the formula (2) takes the form $x(n) = U^n P_{\langle n \rangle}x(0)$ and we obtain that $R_x(n + r, r) = (U^n x(0), P_{\langle r \rangle}x(0))$, $n, r \in \mathcal{Z}$. Consequently $a_j(n) = \int_0^{2\pi} e^{-iun}\gamma^j(du)$, where $\gamma^j(\Delta) = (E(\Delta)x(0), V^j x(0))$, $j = 0, \dots, T - 1$ (see [9] for details). The vector measure $\gamma = (\gamma^0, \dots, \gamma^{T-1})$ is called the *spectrum* of a T-PC sequence $(x(n))$. The existence of measures γ^j can be proved in a different way (e.g. [6], Sect. 6.2).

If the spectrum F or γ is absolutely continuous with respect to Lebesgue measure on $[0, 2\pi)$, then we call the respective sequence *absolutely continuous*, and abbreviate it *a.c.*. The Radon-Nikodym derivative of an a.c. spectrum with respect to the Lebesgue measure on $[0, 2\pi)$ will be called a *density* of the sequence. As indicated

in Sect. 2 we will look at a density as a function on the unit circle rather than on $[0, 2\pi)$. Here are the precise definitions:

1. A *density of a T-variate stationary a.c. sequence* $(\boldsymbol{x}(n))$ is a $T \times T$ matrix function F on the unit circle D_1 with integrable entries such that

$$K_{\boldsymbol{x}}(n) = \int_0^{2\pi} e^{-itn} F\left(e^{it}\right) dt, \quad n \in \mathcal{Z}. \tag{3}$$

2. A *density of an a.c. T-PC sequence* $(x(n))$ is a vector function $g = (g^0, \dots, g^{T-1})$ on the unit circle D_1 with integrable entries such that

$$a_j(n) = \int_0^{2\pi} e^{-itn} g^j\left(e^{it}\right) dt, \quad j = 0, \dots, T-1, \ n \in \mathcal{Z}. \tag{4}$$

A *square factor of a density* F of a T-variate stationary sequence is a $T \times T$ matrix function H on the unit circle D_1 with entries in L^2 such that $F\left(e^{it}\right) = H\left(e^{it}\right) H\left(e^{it}\right)^*$ a.e. A *square factor of a density* g of a T-PC sequence is defined as a vector function h on the unit circle D_1 with values in C^T and with entries in L^2 such that for every $j = 0, \dots, T-1$, $g^j(e^{it}) = h(e^{it})h(e^{i(t+2\pi j/T)})^*$ a.e. If g is a density of a T-PC sequence then g admits at least one square factor (see [8] or Lemma 3.1). Note that our definition of a square factor is slightly different from the one given in [8], where it was defined as $g^j(e^{it}) = (1/T)h(e^{it})h(e^{i(t+2\pi j/T)})^*$.

Remark 1 In the case of stationary sequences H is often called a *transfer function* (the nomenclature seems to come from signal processing). Suppose that $(\boldsymbol{x}(n))$ is T-variate stationary sequence with a density F and H is a square factor of F. We can think about a filter whose input is a T-variate "white noise" on $[0, 2\pi)$, that is, in our approach, a Hilbert space valued vector measure $\boldsymbol{z}(\Delta) = [z^k(\Delta)]$ on $[0, 2\pi)$ such that $(z^k(\Delta_1), z^j(\Delta_2)) = \ell(\Delta_1 \cap \Delta_2)$ if $k = j$ (where ℓ is the Lebesgue measure) and zero otherwise. The output is a T-variate sequence $\mathbf{y}(n) = \int_0^{2\pi} e^{-int} H\left(e^{it}\right) \boldsymbol{z}(dt)$ which is stationary and its spectral density is equal to F, that is $(\mathbf{y}(n))$ unitary equivalent to the sequence $(\boldsymbol{x}(n))$. A square factor h of a density g of a T-PC sequence $(x(n))$ has a similar interpretation. Now the input of a filter is a univariate "white noise" $z(\Delta)$. The output of the filter is

$$y(n) = \int_0^{2\pi} e^{-int} \left[\frac{1}{T} \sum_{k=0}^{T-1} e^{-i2\pi nk/T} h\left(e^{i(t+2\pi k/T)}\right) \right] z(dt). \tag{5}$$

Corollary 3.1 shows that $(y(n))$ is a T-PC sequence equivalent to $(x(n))$. □

There is a natural one-to-one correspondence between T-variate stationary sequences and T-PC sequences. Namely, if $\boldsymbol{x}(n) = [x^k(n)]$, $n \in \mathcal{Z}$, is a T-variate stationary sequence then by arranging $x^k(n)$'s in one sequence we obtain a T-PC sequence $x(n) = x^{\langle n \rangle}(\mathrm{q}(n))$, $n \in \mathcal{Z}$, and conversely, if $x(n)$ is T-PC and we define

$x^k(n) = x(nT + k), k = 0, \dots, T-1, n \in \mathcal{Z}$, then $\boldsymbol{x}(n) = [x^k(n)], n \in \mathcal{Z}$, is a T-variate stationary. Given a T-PC sequence $(x(n))$, the T-variate stationary sequence $(\boldsymbol{x}(n)) = [x(nT + k)]$ defined above will be called the *block sequence corresponding to* $(x(n))$; given a T-variate stationary sequence $(\boldsymbol{x}(n))$, the sequence $x(n) = x^{\langle n \rangle}(\mathrm{q}(n))$, $n \in \mathcal{Z}$, will be called the T-PC sequence *corresponding to* $(\boldsymbol{x}(n))$. Since $M_{\boldsymbol{x}}(n) = M_x(nT + T - 1)$, both sequences are simultaneously regular or not. The following lemma describes the relation between the spectra and square factors of $(x(n))$ and $(\boldsymbol{x}(n))$.

Lemma 3.1 *Let $(x(n))$ be T-PC and $(\boldsymbol{x}(n))$ be the corresponding T-variate stationary block sequence.*

1. *Suppose that $(\boldsymbol{x}(n))$ is a.c. and $H\left(e^{it}\right)$ is a square factor of its density $F\left(e^{it}\right)$. Let $H^{k\cdot}$ denote the k-th row of H, and let*

$$h(e^{it}) := \sum_{k=0}^{T-1} e^{ikt} H^{k\cdot}(e^{iTt}), \quad t \in [0, 2\pi).$$

   Then $(x(n))$ is a.c. and h is a square factor of a density $g = (g^0, \dots, g^{T-1})$ of $(x(n))$, that is $g^j(e^{it}) = h(e^{it})h(e^{i(t+2\pi j/T)})^*$ a.e., $j = 0 \dots, T-1$.
2. *Suppose that $(x(n))$ is a.c. and $h(e^{it})$ is a square factor of a density g of $(x(n))$. Define*

$$f_k(e^{it}) = (1/T) \sum_{j=0}^{T-1} e^{-ik(t+2\pi j/T)} h(e^{i(t+2\pi j/T)}), \quad k = 0, \dots, T-1.$$

   Then f_k is $2\pi/T$-periodic, and hence there is a function $h_k(e^{it})$ such that $f_k(e^{it}) = h_k(e^{iTt})$. Let $H(e^{it})$ be the $T \times T$ matrix function which k-th row $H^{k\cdot}$ is equal h_k. Then $(\boldsymbol{x}(n))$ is a.c. and H is a square factor of its density F, that is $F(e^{it}) = H(e^{it})H(e^{it})^*$ a.e.

The lemma is a consequence of Lemmas 3.2 and 3.3 from [8] (applied for μ to be the Lebsgue measure). For a convenience of the reader, and because the proof for a.c. sequences is much easier than in a general case, we will give a full proof in Appendix. The proof is not a replacement nor a simplification of the proof given in [8] since it works only for a.c. sequences.

We will express the above relations in a matrix form. To simplify the notation we denote $e^{it} = z$ and $e^{i2\pi/T} = d$. Note that $d^k = d^{\langle k \rangle}$ and that $\sum_{k=0}^{T-1} d^{jk} = 0$ unless $j = 0$ modulo T when it is equal to T. Let $H = [H^{j,k}]$ and $h = (h^0, \dots, h^{T-1})$ be the square factors defined in Lemma 3.1. Given h we define the *companion matrix functions* $H_d(z)$ by $H_d^{j,k}(z) = h^k(zd^j)$. Note that h is just a first row of H_d. Further let $D_z = diag[1, z, z^2 \dots, z^{T-1}]$, and let D be the matrix with entries $D^{j,k} = d^{jk}$. Easy computation shows $D^{-1} = (1/T)[d^{-jk}]$. Under these notations the part 1 of Lemma 3.1 says that $h(z) = (1, z, \dots, z^{T-1})H(z^T)$, while part 2 says that

$$H^{k\cdot}(z^T) = (1/T)z^{-k}\sum_{j=0}^{T-1} d^{-jk}h(zd^j) = (D_z^{-1}D^{-1}H_d(z))^{k\cdot}, \tag{6}$$

which also shows that $H_d(z) = DD_zH(z^T)$. Summing up we have that

$$h(z) = (1, z, \dots, z^{T-1})H(z^T) \tag{7}$$

$$H(z^T) = D_z^{-1}D^{-1}H_d(z), \quad \text{where } H_d^{j,k}(z) = h^k(zd^j) \tag{8}$$

The operations (7) and (8) are inverse to each other, that is if we start with H, construct h as in (7), and then use (8), we will end up with the same H; indeed

$$H(z) \xrightarrow{(7)} h(z) = (1, z, \dots, z^{T-1})H(z^T) \xrightarrow{(8)} D_z^{-1}D^{-1}H_d(z) = H(z^T).$$

Therefore the correspondence $h \leftrightarrow H$ described in Lemma 3.1 is one-to-one and onto.

The substitution $z = e^{it}$ is not just symbolic convenience. In many cases it defines a concrete function of complex variable; for example if $R(e^{it})$ is a rational function of t then $R(z)$ is a rational function of complex variable that coincides with $R(e^{it})$ on the unit circle D_1.

From Lemma 3.1 we obtain the following functional model for a.c. T-PC sequences, which is a special a.c. case of [8], Theorem 3.2 (and also [9], Theorem 3.3).

Corollary 3.1 *Let $(x(n))$ be an a.c. T-PC sequence with density g and let h be a square factor of g. Let $(y(n))$ be a sequence in $L^2(C^T)$ defined as*

$$y(n)(e^{it}) = (1/T)\sum_{j=0}^{T-1} e^{-in(t+2\pi j/T)}h(e^{i(t+2\pi j/T)}), \quad t \in [0, 2\pi).$$

Then $(y(n))$ and $(x(n))$ are unitarily equivalent.

Proof From (6) it follows that, in terms of $z = e^{it}$, $d = e^{i2\pi/T}$ and $H(z)$, $y(n)(z) = (1/T)\sum_{j=0}^{T-1} z^{-n}d^{-jn}h(zd^j) = z^{-\mathrm{q}(n)T}H^{\langle n\rangle\cdot}(z^T)$. Hence $y(n)(e^{i\cdot})$, $n \in \mathcal{Z}$, is a T-PC sequence in $L^2(C^T)$ that corresponds to a T-variate stationary sequence $(\mathbf{y}(n))$ in $L^2(C^T)$ defined by $y^k(n)(e^{it}) = e^{-i\mathrm{q}(n)Tt}H^{\langle n\rangle\cdot}(e^{itT})$. Since for any integrable 2π periodic function ϕ, $\int_0^{2\pi}\phi(e^{iTt})dt = \int_0^{2\pi}\phi(e^{it})dt$, $(\mathbf{y}(n))$ has the same covariance as the T-variate block sequence corresponding to $(x(n))$, and hence $(y(n))$ and $(x(n))$ are unitarily equivalent. □

An immediate consequence of Lemma 3.1 is that the corresponding sequences $(\boldsymbol{x}(n))$ and $(x(n))$ are simultaneously a.c. or not. Since their densities are respectively $F(z) = H(z)H(z)^*$ and $g(z) = h(z)H_d(z)^* = e_0H_d(z)H_d(z)^*$, $z = e^{it}$, where $e_0 =$

$(1, 0, \dots, 0)$, the relations (7) and (8) yield the following relations between the density g of a T-PC sequence $(x(n))$ and the density F of the corresponding T-variate stationary block sequence:

$$F(z^T) = (1/T)D_z^{-1}D^{-1}G(z)DD_z, \tag{9}$$

where $G(z) = H_d(z)H_d(z)^*$, that is $G(z)^{j,k} = h(zd^j)h(zd^k)^* = g^{\langle k-j\rangle}(zd^j)$; and

$$g(z) = e_0 DD_z F(z^T) D_z^* D^*. \tag{10}$$

Remember that in the above formulas $z = e^{it}$ and $d = e^{i2\pi/T}$. Therefore we have obtained the following corollary.

Corollary 3.2 *Let $(x(n))$ be T-PC and $(\boldsymbol{x}(n))$ be the corresponding T-variate stationary block sequence. Suppose that both are a.c., and let g and F be their densities, respectively. Then*

$$F^{j,k}(z^T) = (1/T^2)z^{k-j}\sum_{p=0}^{T-1}\sum_{q=0}^{T-1} d^{kq-jp} g^{\langle q-p\rangle}(zd^p),$$

$$g^k(z) = \sum_{p=0}^{T-1}\sum_{q=0}^{T-1} z^{p-q} d^{-qk} F^{p,q}(z^T),$$

where $z = e^{it}$ and $d = e^{i2\pi/T}$.

For not a.c. sequences, the relation between spectral measures of $(x(n))$ and $(\boldsymbol{x}(n))$ can be found in [8].

Definition 3.1 A T-PC sequence $(x(n))$ is said to have a *rational density* if $(x(n))$ is a.c. and there is a rational vector function $g(z)$ of complex variable such that $g(e^{it})$ is a density of $(x(n))$. A T-variate stationary sequence $(\boldsymbol{x}(n))$ is said to have a *rational density* if $(\boldsymbol{x}(n))$ is a.c. and there is a rational $T \times T$ matrix function $F(z)$ of complex variable such that $F(e^{it})$ is a density of $(\boldsymbol{x}(n))$.

The formulas (9) and (10) show, in particular, that F is rational iff g is rational. We summarize our discussion in the following theorem.

Theorem 3.1 *Let $(x(n))$ be an a.c. T-PC sequence and $(\boldsymbol{x}(n))$ be the corresponding T-variate stationary block sequence. Then $(x(n))$ has a rational density iff $(\boldsymbol{x}(n))$ has a rational density. Moreover, if $F(z)$ is a rational density of $(\boldsymbol{x}(n))$ and $H(z)$ is a rational square factor of $F(z)$, then $h(z)$ defined by (7) is a rational square factor of a density $g(z)$ of $(x(n))$; and vice verse, if $g(z)$ is a rational density of $(x(n))$ and $h(z)$ is a rational square factor of $g(z)$, then $H(z)$ defined by (8) is a rational square factor of a density $F(z)$ of $(\boldsymbol{x}(n))$. Everywhere above $z = e^{it}$, $t \in [0, 2\pi)$.*

Sequences with rational density may have square factors which are not rational.

A T-PC sequence $(x(n))$ is called a *moving average (MA)* if there exist an orthonormal system (ξ_n) in some Hilbert space $\mathcal{H} \supseteq M_x$ and a set of scalars $(c_k(n))$, $n, k \in \mathcal{Z}$, such that each $(c_k(n))$, $k \in \mathcal{Z}$, is T-periodic in n,

$$x(n) = \sum_{k=-\infty}^{\infty} c_k(n)\xi_{n-k}, \quad n \in \mathcal{Z}. \tag{11}$$

and $U^T x(n) = x(n+T)$, $n \in \mathcal{Z}$, where U is the shift of (ξ_n) defined by $U\xi_n = \xi_{n+1}$. Note that we allow some ξ_k's to be outside M_x; and we allow some $c_0(n)$'s to be zero. If two sequences $(x(n))$ and $(y(n))$ are unitarily equivalent and one has an MA representation, then the other also does. A sequence may have many different MA representations. We recognize an MA representation of $(x(n))$ (if it exists) by listing its coefficients $(c_k(n))$. A T-PC sequence has an MA representation iff it is a.c.. An MA representation $(c_k(n))$ of a T-PC sequence $(x(n))$ in called an *innovation representation of* $(x(n))$ if for every $n \in \mathcal{Z}$

$$M_x(n) = \overline{\mathrm{sp}}\{c_0(m)\xi_m \,:\, m \le n\} := M_{c\xi}(n). \tag{12}$$

For (12) to be true for every n it is enough that it is true for $n = 0, \dots, T-1$. An MA representation $(c_k(n))$ of $(x(n))$ is an innovation representation iff $c_k(n) = 0$ for all $k < 0$ and $n \in \mathcal{Z}$, and $c_0(n)\xi_n$ is a one-step prediction error at n, that is $x(n) - (x(n)|M_x(n-1)) = c_0(n)\xi_n$, $n \in \mathcal{Z}$. A T-PC sequence $(x(n))$ has an innovation representation iff it is regular. If it does then the number of nonzero elements in the set $\{c_0(m) \,:\, m = 0, \dots, T-1\}$, is called the *rank* of $(x(n))$.

A *moving average (MA)* representation of a T-variate stationary stochastic sequence $(\boldsymbol{x}(n))$ is a representation of $(\boldsymbol{x}(n))$ in the form

$$\boldsymbol{x}(n) = \sum_{k=-\infty}^{\infty} C_k \boldsymbol{\xi}_{n-k}, \quad n \in \mathcal{Z}, \tag{13}$$

where C_k's are $T \times T$ matrices, and $\boldsymbol{\xi}_n = [\xi_n^j]$ is a T-variate orthonormal sequence in some space $\mathcal{H} \supseteq M_{\boldsymbol{x}}$ such that $W\boldsymbol{x}^k(n) = \boldsymbol{x}^k(n+1)$, $n \in \mathcal{Z}$, $k = 0, \dots, T-1$, where W is the shift of $(\boldsymbol{\xi}_n)$ defined as $W\xi_n^k = \xi_{n+1}^k$. An MA representation of $(\boldsymbol{x}(n))$ (if exist) will be recognized by listing its matrix coefficients (C_k). A T-variate stationary sequence $(\boldsymbol{x}(n))$ has an MA representation iff $(\boldsymbol{x}(n))$ is a.c.. An MA representation (C_k) of a T-variate stationary sequence $(\boldsymbol{x}(n))$ is called an *innovation representation of* $(\boldsymbol{x}(n))$ iff for every $n \in \mathcal{Z}$,

$$M_{\boldsymbol{x}}(n) = \overline{\mathrm{sp}}\{aC_0\boldsymbol{\xi}_m \,:\, m \le n, a \in C^T\}, \tag{14}$$

that is iff $C_k = 0$ for all $k < 0$ and for each n, $\boldsymbol{x}(n) - (\boldsymbol{x}(n)|M_{\boldsymbol{x}}(n-1)) = C_0(n)\boldsymbol{\xi}_n$, $n \in \mathcal{Z}$. Here $(\boldsymbol{x}(n)|M_{\boldsymbol{x}}(n-1)) = [(\boldsymbol{x}^k(n)|M_{\boldsymbol{x}}(n-1))]$. For (14) to be true it is enough that it is true for $n = 0$. A T-variate stationary sequence has an innovation representation

iff it is regular. If it does then then dimension of the space $N_{\boldsymbol{x}}(0) = \overline{\text{sp}}\{aC_0\boldsymbol{\xi}_0 : a \in C^T\}$ (which is equal to the rank of matrix C_0) is called the *rank* of the sequence. Note that since $N_{\boldsymbol{x}}(0) = N_x(0) \oplus \ldots N_x(T-1)$, the rank of a regular T-PC sequence $(x(n))$ is equal to the rank of the corresponding T-variate stationary block sequence $(\boldsymbol{x}(n))$ and is equal to the (matrix) rank the matrix

$$\Sigma := (\boldsymbol{x}(n) - (\boldsymbol{x}(n)|M_{\boldsymbol{x}}(n-1))(\boldsymbol{x}(n) - (\boldsymbol{x}(n)|M_{\boldsymbol{x}}(n-1))^*.$$

If (C_k) is an innovation representation of $(\boldsymbol{x}(n))$ then $\Sigma = C_0C_0^*$.

There is a one-to-one correspondence between MA representations $(c_k(n))$ of a T-PC sequence $(x(n))$ and MA representations (C_k) of its corresponding T-variate stationary block sequence $(\boldsymbol{x}(n))$ given by

$$C_k^{i,j} = c_{kT+i-j}(i) \quad \text{or} \quad c_k(n) = C_{-\text{q}(n-k)}^{\langle n\rangle,\langle n-k\rangle}. \tag{15}$$

Recall that $\text{q}(m)$ and $\langle m\rangle$ stand for the quotient and the remainder in division of m by T, respectively. To see (15) note that after substituting $x^j(n) = x(nT+j)$ and $\xi_{n-p}^k = \xi_{(n-p)T+k}$ into (13) we obtain that

$$\sum_{p=-\infty}^{\infty}\sum_{k=0}^{T-1} C_p^{j,k}\xi_{(n-p)T+k} = x(nT+j) = \sum_{r=-\infty}^{\infty} c_r(j)\xi_{nT+j-r},$$

because $c_r(nT+j) = c_r(j)$. Multiplying both sides by $\xi_{(n-p)T+k}$ we obtain that

$$C_p^{j,k} = \sum_{r=-\infty}^{\infty} c_r(j)(\xi_{nT+j-r}, \xi_{(n-p)T+k}) = c_{pT-k+j}(j)$$

On the other hand multiplying both sides by ξ_{nT+j-r} and noting that $(n-p)T+k = nT+j-r$ iff $-p = \text{q}(j-r)$ and $k = \langle j-r\rangle$, we obtain that $c_r(j) = C_{-\text{q}(j-r)}^{j,\langle j-r\rangle}$.

We will refer to the two MA representations described in (15) as *corresponding* to each other. The relation (15) is visualized in a matrix form below:

$$\left[\begin{array}{c|ccc|ccc|c}
\ldots & c_T(0) & c_{T-1}(0) & \ldots & c_1(0) & c_0(0) & c_{-1}(0) & \ldots\ c_{-T+1}(0) & \ldots \\
\ldots & c_{T+1}(1) & c_T(1) & \ldots & c_2(1) & c_1(1) & c_0(1) & \ldots\ c_{-T+2}(1) & \ldots \\
\ldots & \ldots & \ldots & \mathbf{C_1} & \ldots & \ldots & \ldots & \mathbf{C_0}\ \ldots & \ldots \\
\ldots & c_{2T-1}(T-1) & c_{2T-2}(T-1) & \ldots & c_T(T-1) & c_{T-1}(T-1) & c_{T-2}(T-1) & \ldots\ c_0(T-1) & \ldots
\end{array}\right]$$

If $(c_k(n))$ in (15) is an innovation representation of $(x(n))$ then (C_k) is an innovation representation of $(\boldsymbol{x}(n))$. The converse is not true even if C_0 is lower triangular, a counterexample is given in [10]. However, if additionally to being lower triangular C_0 is invertible (i.e. $(\boldsymbol{x}(n))$ is of full rank), then $(c_k(n))$ is an innovation representation of $(x(n))$.

Lemma 3.2 *Let $(x(n))$ be a regular T-PC sequence and $(\boldsymbol{x}(n))$ be the corresponding T-variate stationary block sequence. Let (C_k) be an innovation representation of*

$(\boldsymbol{x}(n))$ and $(c_k(n))$ be the corresponding MA representation of $(x(n))$ defined in (15). If C_0 is lower triangular and invertible, then $(c_k(n))$ is an innovation representation of $(x(n))$.

Proof Recall that $x^k(0) = x(k)$, $\boldsymbol{\xi}_0^k = \xi_k$, $k = 0, \dots, T-1$, $M_{\boldsymbol{x}}(0) = M_x(T-1)$, and $M_{\boldsymbol{x}}(-1) = M_x(-1)$. Also note that since C_0 is invertible and C_0 is lower triangular, then all $C_0^{k,k}$ are different than zero. By assumption $C_0\boldsymbol{\xi}_0$ is equal to the orthogonal projection of $\boldsymbol{x}(0)$ onto $N_{\boldsymbol{x}}(0) = M_{\boldsymbol{x}}(0) \ominus M_{\boldsymbol{x}}(-1)$. Since C_0 is lower triangular, we have that $x(k) - (x(k)|M_{\boldsymbol{x}}(-1)) = C_0^{k,0}\xi_0 + \cdots C_0^{k,k}\xi_k$, $k = 0, \dots, T-1$. Suppose first that $k = 0$. Then from $M_{\boldsymbol{x}}(-1) = M_x(-1)$ it follows that $x(0) - (x(0)|M_x(-1)) = x(0) - (x(0)|M_{\boldsymbol{x}}(-1)) = C_0^{0,0}\xi_0$. Since $C_0^{0,0} \neq 1$, we conclude that $N_x(0) = M_x(0) - M_x(-1) = \overline{\mathrm{sp}}\{\xi_0\}$. Assume that we have already shown that $x(j) - (x(j)|M_x(j-1)) = C_0^{j,j}\xi_j$ for $j = 0, \dots, k-1$, $0 < k < T-1$. Then $x(k) - (x(k)|M_{\boldsymbol{x}}(-1)) = C_0^{k,0}\xi_0 + \cdots + C_0^{k-1,k-1}\xi_{k-1} + C_0^{k,k}\xi_k$, and hence

$$x(k) - C_0^{k,k}\xi_k = (x(k)|M_{\boldsymbol{x}}(-1)) + C_0^{k,0}\xi_0 + \cdots C_0^{k-1,k-1}\xi_{k-1} = (x(k)|M_x(k-1)),$$

i.e. $x(k) - (x(k)|M_x(k-1)) = C_0^{k,k}\xi_k$. Note that the proof will not work if $C_0^{0,0} = 0$ but both $C_0^{1,0}$ and $C_0^{1,1}$ are different than zero. □

There is an obvious one-to-one correspondence between MA representations (C_n) of a T-variate stationary sequence $(\boldsymbol{x}(n))$ and square factors H of a density F of $(\boldsymbol{x}(n))$ given by

$$\boldsymbol{x}(n) = \sum_{k=-\infty}^{\infty} C_k\boldsymbol{\xi}_{n-k} \longleftrightarrow H\left(e^{it}\right) = (1/\sqrt{2\pi}) \sum_{k=-\infty}^{\infty} C_k e^{ikt}. \tag{16}$$

This relation is easily seen if we choose $\boldsymbol{\xi}_n^j = (1/\sqrt{2\pi})e^{-int}e_j$. Square factors of F that correspond to innovation representations of $(\boldsymbol{x}(n))$ are called *maximal factors* (Rozanov [11]). The corresponding notion for PC sequences was introduced in [10] under the name an *i-factor*. Maximal factors can be characterized in terms of subspaces of L_+^2 spanned by their coordinates (see for example [11] for T-variate stationary case, and [10] for the PC case). Finding a maximal factor is equivalent to finding coefficients of an innovation representation of a sequence. The latter constitutes a solution to so called *prediction problem*. So far the prediction problem has been solved only for full rank stationary T-variate sequences having rational densities [4, 5, 11]. Theorem 3.1 and Lemma 3.2 allow us to obtain the solution for full rank PC sequences with rational densities. A procedure is following: given a full rank T-PC sequence $(x(n))$ with rational density g compute the rational density F of the corresponding T-variate block sequence $(\boldsymbol{x}(n))$ using formula (9), find a maximal rational square factor G of F using a construction given in [5] or [11], multiply G by a proper unitary matrix Q so that the "zero" term of Fourier series of $H(e^{it}) = G(e^{it})Q$ is a lower triangular matrix, and then use Lemma 3.2 to recover innovation coefficients of $(x(n))$.

4 PARMA Systems

A VARMA system of dimension $T \geq 1$ is a system of vector difference equations (VDE)

$$\sum_{j=0}^{l} A_j \boldsymbol{x}(n-j) = \sum_{j=0}^{r} B_j \boldsymbol{\xi}_{n-j}, \qquad n \in \mathcal{Z}, \tag{17}$$

where A_j's, and B_j's are complex $T \times T$ matrices, A_0 is invertible, A_l, B_0, B_r are nonzero, and $\boldsymbol{\xi}_n$, $n \in \mathcal{Z}$, is a given T-variate orthonormal sequence in some Hilbert space $\mathcal{H}$. A *proper stationary solution* to a VARMA system (17) is a T-variate stationary sequence $(\boldsymbol{x}(n))$ which satisfies the system and such that $x^j(n) \in M_{\xi}$ and $Wx^j(n) = x^j(n+1)$, $n \in \mathcal{Z}$, $j = 0, \ldots, T-1$, where W denotes the shift of $(\boldsymbol{\xi}_n)$. In many publications and books the last requirement is replaced by some additional assumptions about the coefficients of the system or by a requirement that the solution has an MA representation (see for example in [5]). Without any additional assumptions the system (17) may have multiple or not regular stationary solutions ([5], p. 13). A PARMA system is an infinite system of difference equations

$$\sum_{j=0}^{l} a_j(n) x(n-j) = \sum_{j=0}^{r} b_j(n) \xi_{n-j}, \qquad n \in \mathcal{Z}, \tag{18}$$

where $l, r \geq 0$, $a_j(n)$ and $b_j(n)$ are T-periodic (in n) sequences of complex numbers, $a_0(n) = 1$ for every $n \in Z$, and none of the sequences $(b_0(n))$, $(a_l(n))$, and $(b_r(n))$ is identically zero. Let U be the shift of (ξ_n), that is $U\xi_n = \xi_{n+1}$, $n \in Z$. We will be interested only in T-PC solutions $(x(n))$ to the system whose T-shift coincides with U^T, that is such that $x(n) \in M_{\xi}$ and $U^T x(n) = x(n+T)$, $n \in \mathcal{Z}$. We label them *proper PC solutions*. The assumption $U^T x(n) = x(n+T)$ allows us to avoid having multiple or not regular PC solutions. If we arrange the coefficients $a_j(n)$ into an $T \times (L+1)T$ matrix $[A_L \ \ldots \ A_1 \ A_0]$ where L is such that the matrix contains all nonzero $a_j(n)$'s as shown below

$$\left[\begin{array}{c|cccc|cccc} \ldots & a_T(0) & \ldots & a_2(0) & a_1(0) & a_0(0) & 0 & \ldots & 0 \\ \ldots & a_{T+1}(1) & \ldots & a_3(1) & a_2(1) & a_1(1) & a_0(1) & \ldots & 0 \\ \ldots & \ldots & \mathbf{A(1)} & \ldots & \ldots & \ldots & \mathbf{A(0)} & \ldots & \ldots \\ \ldots\ldots & a_{2T-1}(T-1) & \ldots & a_{T+1}(T-1) & a_T(T-1) & a_{T-1}(T-1) & a_{T-1}(T-1) & \ldots & a_0(T-1) \end{array}\right]$$

and do the same for the $b_j(n)$'s creating $T \times (R+1)T$ matrix $[B_R \ \ldots \ B_0]$, then, using matrices A_j and B_j, the system (18) can be written as a VARMA system

$$\sum_{j=0}^{L} A_j \boldsymbol{x}(n-j) = \sum_{j=0}^{R} B_j \boldsymbol{\xi}_{n-j}, \qquad n \in \mathcal{Z}, \tag{19}$$

where A_0 and B_0 are lower triangular, and $(\boldsymbol{x}(n))$ and $(\boldsymbol{\xi}_n)$ are T-variate block sequences corresponding to $(x(n))$ and (ξ_n) respectively, that is $\boldsymbol{x}(n) = [x^k(n)]$ with

$x^k(n) = x(nT + k)$ and $\boldsymbol{\xi}_n = [\xi_n^k]$ with $\xi_n^k = \xi_{nT+k}$. The system (17), and hence the system (18), is completely described by a pair of polynomial matrices $(A(z), B(z))$ defined as

$$A(z) = \sum_{k=0}^{L} A(k)z^k \quad \text{and} \quad B(z) = \sum_{k=0}^{R} B(k)z^k. \tag{20}$$

In the sequel we will identify both (17) and (18) by giving the corresponding pair $(A(z), B(z))$. The only difference between PARMA and general VARMA systems is that in PARMA systems A_0 and B_0 are lower triangular, and $A_0^{i,i} = 1$ for each $i = 0, \dots, T-1$, so PARMA systems form a subset of the family of VARMA systems. Note that since $a(z) = \det(A(z))$ is a polynomial and by assumption $a(0) = \det(A_0) \neq 0$, $a(z) \neq 0$ for all $z \in C$ except finitely many points, and consequently $A(z)^{-1}$ exists for all $z \in C$ except finitely many points.

Theorem 4.1 *A PARMA system $(A(z), B(z))$ has a proper PC solution iff the rational matrix function $A(z)^{-1}B(z)$ has no poles of modulus 1. If a proper PC solution $(x(n))$ exists, then it is unique, absolutely continuous, and its density $g = (g^0, \dots, g^{T-1})$ is given by*

$$g^j(e^{it}) = h(e^{it})h(e^{i(t+2\pi j/T)})^*, \quad a.e. \tag{21}$$

where $h(z) = (1, z, \dots, z^{T-1})H(z^T)$ and $H(z) = (1/\sqrt{2\pi})A(z)^{-1}B(z)$, $z = e^{it}$.

The theorem remains true when in the above formulation we replace *PARMA* by *VARMA*, *PC* by *T-variate stationary*, $(x(n))$ by $(\boldsymbol{x}(n))$, g by F, and the formula (21) by $F(e^{it}) = H(e^{it})H(e^{it})^*$, where $H(z) = (1/\sqrt{2\pi})A(z)^{-1}B(z)$.

Proof From the preceding discussion it follows that a PARMA system (18) has a proper T-PC solution $(x(n))$ iff the associated VARMA system (19) has a proper T-variate stationary solutions $(\boldsymbol{x}(n))$, and if it does then $(\boldsymbol{x}(n))$ is the T-variate stationary block sequence corresponding to $(x(n))$. Because $(\boldsymbol{x}(n))$ is to be proper, it is enough to find $\boldsymbol{x}(0)$ since then $\boldsymbol{x}(n) = W^n\boldsymbol{x}(0)$. Therefore it is enough to solve (19) just for $n = 0$, that is solve the equation $\sum_{j=0}^{L} A_j\boldsymbol{x}(-j) = \sum_{j=0}^{R} B_j\boldsymbol{\xi}_{-j}$. Substituting $\boldsymbol{x}(-j) = W^{-j}\boldsymbol{x}(0)$ and $\boldsymbol{\xi}_j = W^{-j}\boldsymbol{\xi}_0$, we can write the above equation as

$$\sum_{j=0}^{L} A_j W^{-j}\boldsymbol{x}(0) = \sum_{j=0}^{R} B_j W^{-j}\boldsymbol{\xi}_0, \tag{22}$$

or symbolically, using polynomials $A(z)$ and $B(z)$ defined in (20), as $A(W^{-1})\boldsymbol{x}(0) = B(W^{-1})\boldsymbol{\xi}(0)$. To solve (22) let us consider a T-variate orthonormal system $\boldsymbol{\zeta}_n = [\zeta_n^k]$ in $L^2(C^T)$ defined as $\zeta_n^k = (1/\sqrt{2\pi})e^{-int}e_k$, where (e_k) is the standard basis in C^T, and define the unitary operator $\boldsymbol{\Phi} : M_\xi \to L^2(C^T)$ by $\boldsymbol{\Phi}(\xi_n^k) = \zeta_n^k$, $n \in \mathcal{Z}$, $k = 0, \dots, T-1$. Note that the shift of $(\boldsymbol{\zeta}_n)$ is the operator of multiplication by e^{-it} and that $\boldsymbol{\zeta}_0 = [\zeta_0^k] = (1/\sqrt{2\pi})I$, where I is the $T \times T$ identity matrix. The mapping $\boldsymbol{\Phi}$ transfers the equation (22) into matrix equation

$$A(e^{it})H(e^{it}) = (1/\sqrt{2\pi})B(e^{it}), \tag{23}$$

where $H(e^{it})$ is a $T \times T$ matrix function with rows $H^{k\cdot}(e^{it}) = \boldsymbol{\Phi}(x^k(0))$. Summing up, (22) has a solution $\boldsymbol{x}(0) \in M_\xi$ iff there is a $T \times T$ matrix function H with rows in $L^2(C^T)$ that satisfies (23). Since $A(e^{it})^{-1}$ exists a.e., the only candidate for H is $H(e^{it}) = (1/\sqrt{2\pi})A^{-1}(e^{it})B(e^{it})$. Hence (22) has a solution iff all entries of $A^{-1}(e^{it})B(e^{it})$ belong to L^2. Since $(e^{it} - c)^{-1}$ is square integrable always except when $|c| = 1$, we conclude the system (19) has a proper T-variate stationary solution iff $A^{-1}(z)B(z)$ has no poles of modulus 1, assuming as always that all entries of the rational matrix $A^{-1}(z)B(z)$, are written in the simplest forms. If this condition is satisfied then the solution $(\boldsymbol{x}(n))$ to (19) is given by $x^k(n) = \boldsymbol{\Phi}^{-1}(e^{-int}H^{k\cdot}(e^{it}))$, where $H(e^{it}) = (1/\sqrt{2\pi})A^{-1}(e^{it})B(e^{it})$. The uniqueness follows from the fact that, because of a.e. invertibility of $A(e^{it})$, $H(e^{it})$ defined above is the only matrix function satisfying (23). The covariance of $(\boldsymbol{x}(n))$ is

$$K_{\boldsymbol{x}}^{j,k} = (\boldsymbol{\Phi}^{-1}(e^{-int}H^{j\cdot}(e^{it})), \boldsymbol{\Phi}^{-1}(H^{k\cdot}(e^{it}))) = \int_0^{2\pi} e^{-int}H^{j\cdot}(e^{it})H^{k\cdot}(e^{it})^* dt,$$

and hence $(\boldsymbol{x}(n))$ is a.c. and its density is $F(e^{it}) = H(e^{it})H(e^{it})^*$. A proper T-PC solution the original PARMA system $(A(z), B(z))$ is therefore the T-PC sequence that corresponds to $(\boldsymbol{x}(n))$, that is $x(n) = x^{\langle n\rangle}(\mathrm{q}(n))$, $n \in \mathcal{Z}$. From Theorem 3.1 we conclude that $(x(n))$ is a.c. and that $h(z) = (1, z, \ldots, z^{T-1})H(z^T)$, $z = e^{it}$, is a square factor of the density g of $(x(n))$, which proves the formula (21). □

The first part (existence) of Theorem 4.1 is well known but difficult to attribute to a particular name. In fact more is known. From a description of all solutions to (19) given for example in [4], p. 11, it follows that if additionally $\det A(z) \neq 0$ for all $|z| = 1$, then the system has only one bounded solution which therefore must be a proper T-PC solution. Regarding computing a density of a proper T-PC solution, two other different procedures were given in [12, 13]. Our formula seems similar to [13].

The formula (21) shows that a density of a proper PC solution to any PARMA system is a rational function. From the next theorem it follows that the opposite is also true.

Theorem 4.2 *Let $(x(n))$ be a T-PC sequence. Then the following conditions are equivalent:*

1. *$(x(n))$ is a proper T-PC solution to some PARMA system.*
2. *$(x(n))$ has a rational density.*
3. *there exists a PARMA system $(A(z), B(z))$ such that:*
   a. *polynomial matrices $A(z)$ and $B(z)$ are left co-prime,*
   b. *$A(z)$ has no zeros in an open disk $D_{<r}$ of a radius $r > 1$, and $B(z)$ has no zeros in the open disk $D_{<1}$ of radius 1,*
   c. *$(x(n))$ is the only T-PC solution to the system $(A(z), B(z))$.*

The theorem remains valid for T-variate stationary sequences, that is when in the above formulation we replace *T-PC* by *T-variate stationary*, $(x(n))$ by $(\boldsymbol{x}(n))$, and the word *PARMA* by *VARMA*.

Proof Let $(\boldsymbol{x}(n))$ denote the T-variate stationary block sequence corresponding to $(x(n))$ and F be its spectrum.

(1. $\Rightarrow$ 2.) From the proof of Theorem 4.1 it follows that if $(x(n))$ is a proper PC solution to some PARMA system (18), then the corresponding block sequence $(\boldsymbol{x}(n))$ satisfies the associated VARMA system (19) and that $(\boldsymbol{x}(n))$ is unitary equivalent to the T-variate stationary sequence $(H(n))$ in $\mathcal{H} = L^2(C^T)$ defined by $H(n) = e^{-int}H(e^{it})$, where $H(e^{it}) = (1/\sqrt{2\pi})A^{-1}(e^{it})B(e^{it})$, $n \in \mathcal{Z}$. Note that $H(z) = (1/\sqrt{2\pi})A(z)^{-1}B(z)$ is rational. The correlation function of $(H(n))$ is $K_H(n) = \int_0^{2\pi} e^{-int}H(e^{it})H(e^{it})^* dt$. Hence $(\boldsymbol{x}(n))$ is a.c. and its spectral density is $F(e^{it}) = H(e^{it})H(e^{it})^*$ a.e., that is $H(z)$ is a rational square factor of $F(z)$. From Lemma 3.1 part 1. we conclude that the spectrum of $(x(n))$ is a.c. and $h(z) = \sum_{k=0}^{T-1} z^k H^{k\cdot}(z^T)$ is a square factor of the density g of $(x(n))$. Since $H(z)$ is a rational matrix, $h(z)$ is also, and consequently all $g^j(z) = h(z)h(d^j z)^*$, $d = e^{2\pi/T}$, are rational.

(2. $\Rightarrow$ 3.) Suppose now that $(x(n))$ is a.c. and its density g is rational. Then the corresponding T-variate stationary block sequence $(\boldsymbol{x}(n)$ is a.c. and, by Theorem 3.1, its density $F(z)$ is rational. The Rozanov's Theorem 2.1 implies that $F(z)$ has a factorization $F(z) = G(z)G(z)^*$ where G is rational, analytic in the the open disk $D_{<1}$ and $G(z)$ has no zeros in $D_{<1}$. Since entries of $F(e^{it})$ are integrable, the function $G(z)$ has no poles of modulus 1. Therefore $G(z)$ is analytic in some open disk $D_{<r}$, $r > 1$. The matrix $G(0)$ is different than 0 since otherwise $z = 0$ would be a zero of $G(z)$. If we factor the least common multiple, say $q(z)$, of all denominators of entries of $G(z)$, then we can write $G(z) = P(z)/q(z)$ where $P(z)$ is $T \times T$ polynomial matrix with no zeros in $D_{<1}$ and $q(z)$ is a scalar polynomial with all zeros outside $D_{\leq 1}$, and hence outside of a certain disk $D_{<r}$ of radius $r > 0$. In particular $z = 0$ is not a zero of $q(z)$, so $q(0) \neq 0$. Let $A_0(z) = (q(z)/q(0))I$ and $B_0(z) = (1/q(0))P(z)$. Then both are analytic polynomial matrices, the constant term of $A_0(z)$ is the identity matrix I, $A_0(z)^{-1} = (q(0)/q(z))I$ and $A_0(z)^{-1}B_0(z) = (q(0)/q(z))(1/q(0))P(z) = G(z)$. Factoring out the left greatest common divisor $L(z)$ of $A_0(z)$ and $B_0(z)$ we obtain that $A_0(z) = L(z)C(z)$ and $B_0(z) = L(z)D(z)$. Since $\det A_0(z) \neq 0$ on some $D_{<r}$, $r > 1$, both $\det L(z) \neq 0$ and $\det C(z) \neq 0$ on $D_{<r}$, $r > 1$. Hence $L(z)^{-1}$ exists and we conclude that $C(z)^{-1}D(z) = A_0(z)^{-1}B_0(z) = G(z)$. Let S be an invertible matrix such that $SC(0)$ is lower triangular, and let Q be a unitary matrix such that $SD(0)Q$ is lower triangular. Define $A(z) = SC(z)$ and $B(z) = (\sqrt{2\pi})SD(z)Q$. Since $A(0)$ and $B(0)$ are lower triangular, $(A(z), B(z))$ is a PARMA system. Polynomial matrices $A(z)$ and $B(z)$ satisfy conditions a. and b. of 3., and $(1/\sqrt{2\pi})A(z)^{-1}B(z) = C(z)^{-1}D(z)Q$, $z \in C$. Define $H(z) = (\sqrt{2\pi})A(z)^{-1}B(z)$. Then $H(z) = G(z)Q$, and hence $H(z)H(z)^* = G(z)G(z)^* = F(z)$. Hence $(\boldsymbol{x}(n))$ is a proper T-variate stationary solution to the system $(A(z), B(z))$. Since $(\boldsymbol{x}(n))$ is a block sequence corresponding to $(x(n))$, the sequence $(x(n))$ is a proper T-PC solution to the system $(A(z), B(z))$. Since $\det A(z) = (\det S)(\det A_0(z))/(\det L(z)) \neq 0$ on the circle $|z| = 1$,

the system $(A(z), B(z))$ has only one bounded solution, and hence only one T-PC solution.

(3. $\Rightarrow$ 1.) In view of Theorem 4.1, the condition b. itself implies that that system $(A(z), B(z))$ has a proper T-PC solution. Moreover it implies that $\det A(z) \neq 0$ on the circle $|z| = 1$, and hence that the system has only one bounded solution. Therefore the sequence $(x(n))$ must be a proper T-PC solution to $(A(z), B(z))$. □

An immediate consequence of the above theorem is regularity of every T-PC sequence with rational density.

Corollary 4.1 *Every T-PC (or T-variate stationary) sequence with rational density is regular. Consequently, if a PARMA (or VARMA) system has a proper PC (or proper T-variate stationary) solution, then this solution is regular.*

Proof Suppose that $(x(n))$ is a T-PC sequence with rational density g, and let $(\boldsymbol{x}(n))$ be the corresponding T-variate stationary block sequence. From Theorem 4.2, part 3., we conclude that there is a PARMA system $(A(z), B(z))$ such that $(\boldsymbol{x}(n))$ is a proper T-variate stationary solution of the system $(A(z), B(z))$ and the rational function $H(z) = (\sqrt{2\pi})A(z)^{-1}B(z)$ is analytic in some open disk $D_{<r}$ of radius $r > 1$. Moreover $H(e^{it})$ is a square factor of the density of $(\boldsymbol{x}(n))$. Being analytic, H has an expansion $H(z) = \sum_{k=0}^{\infty} C_k z^k$, $|z| < r, r > 1$. Therefore the corresponding MA representation of $(\boldsymbol{x}(n))$ is one-sided and hence $(\boldsymbol{x}(n))$, and also $(x(n))$, are regular. □

5 PARMA Models

We failed to find a precise definition of a PARMA (or VARMA) model, so we have assumed the following.

Definition 5.1 A PARMA (or VARMA) system $(A(z), B(z))$ is called a *PARMA (or VARMA) model* if the polynomial matrices $A(z)$ and $B(z)$ satisfy the conditions a. and b. of part 3 of Theorem 4.2

Theorems 4.1 and 4.2 show that every PARMA (or VARMA) model has a unique T-PC (or T-variate stationary) solution and this solution has rational density, and vice verse, that every T-PC (or T-variate stationary) sequence with rational density admits a PARMA (respectively VARMA) model. The sole reason that we added the condition a. saying that $A(z)$ and $B(z)$ are left co-prime is to reduce the set of allowed models. Theorem 4.2 remains valid if we remove this condition from part 3.

Why do we like to have a model? Because then the sequence $(b_0(n)\xi_n)$ in (18) or $(B_0\boldsymbol{\xi}_n)$ in (17) are innovation sequences for $(x(n))$ and $(\boldsymbol{x}(n))$, that is $\overline{\text{sp}}\{b_0(n)\xi_n\} = N_x(n)$ and $\overline{\text{sp}}\{aB_0\boldsymbol{\xi}_n : a \in C^T\} = N_{\boldsymbol{x}}(n)$, respectively. At least we believe so, since so far we can prove it only under some additional assumptions: a *miniphase assumption* about $B(z)$, or a *full density rank* assumption about the sequence. If $(b_0(n)\xi_n)$ (or $(B_0\boldsymbol{\xi}_n)$) are innovations then the one step prediction of an element $x(n)$ (or $\boldsymbol{x}(n)$) based

on the immediate past is obtained by simply setting $\xi_n = 0$ in the model equation (18) (or $\boldsymbol{\xi}_n = 0$ in (17), respectively).

The *miniphase assumption* is the assumption that $\det B(z)$ is not identically zero ([5], p. 25). In the case of VARMA systems, the miniphase assumption implies that a proper T-variate stationary solution $(\boldsymbol{x}(n))$ of the system is a full rank sequence and, hence, its density $F(e^{it})$ is a.e. invertible. It is easy to see that inverse also holds true, that is if a T-variate stationary sequence $(\boldsymbol{x}(n))$ has a rational density F with $\det F(e^{it}) \neq 0$ a.e., and if $(A(z), B(z))$ is a VARMA model for $(\boldsymbol{x}(n))$, then $B(z)$ must satisfy the miniphase assumption and $(\boldsymbol{x}(n))$ is full rank. Below we discuss a consequence of the *miniphase assumption* for PARMA models. Note that until this moment we have not assumed anything about the rank of a sequence or the matrix rank of a polynomial matrix $B(z)$.

Definition 5.2 For any a.c. T-PC sequence $(x(n))$ with density g let $G(e^{it})$ be the $T \times T$ matrix function defined by $G^{j,k}(e^{it}) = g^{\langle k-j\rangle}(e^{i(t+2\pi j/T)})$, $t \in [0, 2\pi)$, $k,j = 0, \dots, T-1$. If $\det G(e^{it}) \neq 0$ a.e. then we say that $(x(n))$ is of a *full density rank*.

Note that if h is a square factor of g, then $G(z) = H_d(z)H_d(z)^*$, where $z = e^{it}$ and $H_d(z)$ is the matrix whose k-th row is equal $h(zd^k)$ as in (8). Also recall that the *rank r of a T-PC sequence* $(x(n))$ is the number of nonzero elements in a sequence $x(k) - (x(k)|M_x(k-1)), k = 0, \dots, T-1$. A T-PC sequence $(x(n))$ is said to be of *full rank* if $r = T$.

Theorem 5.1 *Suppose that a T-PC sequence $(x(n))$ has a rational density g and that $(A(z), B(z))$ is a PARMA model (18) for $(x(n))$. Assume additionally that* $\det B(z)$ *is not identically zero (a miniphase assumption). Then $(x(n))$ has full density rank and the sequence (ξ_n) in (18) is an innovation sequence for $(x(n))$. Consequently $x(n)$ is of full rank.*

Proof Let $(\boldsymbol{x}(n))$ be a T-stationary block sequence corresponding to $(x(n))$ and F be its density. Let $(A(z), B(z))$ be a PARMA model for $(x(n))$. Then $H(z) = (1/\sqrt{2\pi})$ $A(z)^{-1}B(z)$ is a square factor of F. Since $A(z)$ has no zeros in an open disk $D_{<r}$, $r > 1$, $\det A(z) \neq 0$ for all $|z| < r$. By assumption $b(z) = \det B(z)$ is not identically zero and hence, because $b(z)$ is a polynomial, $b(z) \neq 0$ everywhere except finitely many z's. Since $B(z)$ has no zeros in the open disk $D_{<1}$, $\det B(z) \neq 0$ everywhere on $D_{<1}$. Summing up, $H(z)$ is analytic on $D_{<1}$ and $\det H(z) = (1/\sqrt{2\pi})(\det B(z))/(\det A(z)) \neq 0$ on $D_{<1}$. From [11], p. 76, we conclude that $H(z)$ is a maximal factor of F, which means that if we write $H(z)$ as a power series $H(z) = (1/\sqrt{2\pi}) \sum_{k=0}^{\infty} C_k z^k$, $|z| < r$, then

$$\boldsymbol{x}(n) = \sum_{k=0}^{\infty} C_k \boldsymbol{\xi}_{n-k}, \tag{24}$$

is an innovation representation of $(\boldsymbol{x}(n))$. Note that, because $B(z) = \sqrt{2\pi}A(z)H(z)$ and all three are analytic in $D_{<r}$, $B(0) = \sqrt{2\pi}A(0)C_0$. Also $A(0)$ is invertible because otherwise $z = 0$ would be a zero of $A(z)$. This and the fact that by definition both

$B(0)$ and $A(0)$ are lower triangular imply that $C_0 = (1/\sqrt{2\pi})A(0)^{-1}B(0)$ is lower triangular and invertible. From Lemma 3.2, we conclude that the MA representation of $(x(n))$ generated by (24)

$$x(n) = \sum_{k=0}^{\infty} c_k(n)\xi_{n-k}, \quad c_k(n) = C_{-q(n-k)}^{\langle n\rangle,\langle n-k\rangle},$$

is an innovation representation of $(x(n))$, that is $(c_0(n)\xi_n)$ is an innovation for $(x(n))$. Since by (15), $c_0(mT + j) = c_0(j) = C_0^{j,j} \neq 0$ for all $j = 0, \ldots, T-1$, and $m \in \mathcal{Z}$, (ξ_n) is also an innovation for $(x(n))$. This implies that the rank of $(x(n))$ is T. Moreover, since $\det H(e^{it}) \neq 0$ a.e., the matrix function $H_d(e^{it})$ appearing in (8) is also invertible a.e., and hence $G(e^{it}) = H_d(e^{it})H_d(e^{it})^*$ is also invertible a.e., that is $(x(n))$ is a full density rank. □

The next theorem is sort of inverse and shows that if $(x(n))$ is a T-PC sequence with full rank rational density, then any PARMA model for $(x(n))$ must satisfy the miniphase assumption.

Theorem 5.2 *Suppose that $(x(n))$ is T-PC with a rational density g of full density rank. Let $(A(z), B(z))$ be a PARMA model for $(x(n))$. Then $\det B(z)$ is not identically zero, and (ξ_n) in (18) is an innovation sequence for $(x(n))$. Moreover $(x(n))$ is of full rank.*

Proof Full density rank means that $\det G(e^{it}) = |\det H_d(e^{it}))|^2 \neq 0$ a.e. From the relation (8) it follows that also $|\det H(e^{itT})| = (1/T)|\det H_d(e^{it})| \neq 0$ a.e. Hence $H(z) = (1/\sqrt{2\pi})A(z)^{-1}B(z)$ is invertible except finitely many z's. Consequently $(A(z), B(z))$ satisfies a miniphase assumption and the rest follows from Theorem 5.1. □

We do not know whether Theorems 5.1 or 5.2 are true for sequences of not full density rank r or without a miniphase assumption.

An immediate consequence of the two theorems is that if a T-PC sequence $(x(n))$ has a rational density, then $(x(n))$ is of full rank iff it has a full density rank (we already know that every T-PC sequence $(x(n))$ with a rational density is regular).

In some publications a VARMA (as well as a PARMA) model is defined as a system $(A(z), B(z))$ which additionally to the condition b. of part 3 of Theorem 4.2, satisfies the so called *invertibility assumption* which says that $\det B(z) \neq 0$ for all $|z| \leq 1$ (e.g. [1], p. 409). The invertibility assumption (together with b.) immediately gives that a T-variate stationary solution $(\boldsymbol{x}(n))$ to the system is proper, full rank, and $(\boldsymbol{\xi}_n)$ an innovation sequence for $(\boldsymbol{x}(n))$. The invertibility assumption is much stronger that a miniphase assumption which allows $\det B(z) = 0$ for finitely many z's of modulus 1, and significantly reduces a number of T-variate stationary sequences with rational densities that can be modeled in that way. For example a pair $A(z) = 1$ and $B(z) = 1 - z$ is a model for a univariate stationary sequence $x(n) = \xi_n - \xi_{n-1}$, $n \in \mathcal{Z}$, but $B(z)$ does not satisfy an invertibility assumption.

Note is that a construction of the rational matrix $G(z)$ in Rozanov's Theorem 2.1 is explicit, as well as all constructions presented in this paper are explicit. Therefore

given a PARMA system we can explicitly compute a density of its T-PC solution via Theorem 4.1, while a construction given in the proof of Theorem 4.2 allows us to construct a PARMA model for any T-PC sequence with rational density given its density g. A needed procedure for finding a left greatest common divisor of polynomial matrices $A_0(z)$ and $B_0(z)$ can be found in [5], p. 38, or in [7], Sect. 1.15.2. Moreover, if g is if full density rank then the proofs of Theorems 5.1 and 5.2 show us how to find coefficients of an innovation representation of $(x(n))$.

Remark 2 For each T-PC sequence (or T-variate stationary sequence) with rational density one can find many different PARMA (respectively VARMA) models, even if we assume the miniphase assumption. This is because there are many pairs $(A_k(z), B_k(z))$ satisfying the conditions of Definition 5.1 and such that that $H_k(z)H_k(z)^*$, where $H_k(z) = (1/\sqrt{2\pi})A_k(z)^{-1}B_k(z)$, coincide a.e. on the unit circle $|z| = 1$. Although each model serves its prediction purpose, this lack of uniqueness is not convenient in model identification for it would be nice to have a unique, preferably minimal, set of model coefficient to be estimated. Because of this Hannan introduced the notion of *identifiability*. The idea is to impose some constraints on allowable models such that each sequence with rational density would have one and only one model satisfying these constraints. Identifiability problem for VARMA models is discussed in [5]. We do not address this question in our paper. □

Appendix: Proof of Lemma 3.1

Let $(x(n))$ be T-PC, $(\boldsymbol{x}(n) = [x^k(n)])$ be the corresponding T-variate stationary block sequence, and γ and F be respectively their spectral measures.

1. First we prove that: *if* $(\boldsymbol{x}(n))$ *is a.c.,* $H\left(e^{it}\right)$ *is a square factor of its density* $F\left(e^{it}\right)$ *of* $(\boldsymbol{x}(n))$, $H^{k\cdot}$ *denotes the k-th row of* H, *and we define* $h(e^{it}) = \sum_{k=0}^{T-1} e^{ikt}H^{k\cdot}(e^{iTt})$, *then* $(x(n))$ *is a.c. and h is a square factor of a density* $g = (g^0, \dots, g^{T-1})$ *of* $(x(n))$. Write $H^{k\cdot}$ as a Fourier series $H^{k\cdot}(e^{it}) = \sum_{n=-\infty}^{\infty} H_n^k e^{int}$. Then

$$h(e^{it}) = \left[\sum_{n=-\infty}^{\infty}\sum_{k=0}^{T-1} H_n^k e^{i(nT+k)t}\right] = \sum_{p=-\infty}^{\infty} h_p e^{ipt}, \quad \text{where } h_p := H^{\langle p\rangle}_{\mathrm{q}(p)}.$$

Denoting $d = e^{2\pi/T}$, we obtain that

$$\begin{aligned} b_j(n) &:= \int_0^{2\pi} e^{-int}h(e^{it})h(e^{i(t+2\pi j/T)})^* dt \\ &= \sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty} h_p h_q^* d^{-jq}\int_0^{2\pi} e^{i(p-q-n)t}dt = 2\pi\sum_{q=-\infty}^{\infty} h_{q+n}h_q^* d^{-jq}. \end{aligned}$$

Because $\sum_{j=0}^{T-1} d^{j(r-q)}$ is nonzero only if $q = mT + r$ for some $m \in \mathcal{Z}$, the inverse discrete Fourier transform of $b_j(n)$ is

$$(1/T)\sum_{j=0}^{T-1} d^{jr} b_j(r) = (2\pi/T)\sum_{q=-\infty}^{\infty} h_{q+n}h_q^* \sum_{j=0}^{T-1} d^{j(r-q)}$$
$$= 2\pi \sum_{m=-\infty}^{\infty} h_{mT+r+n}h_{mT+r}^* = 2\pi \sum_{m=-\infty}^{\infty} H^{\langle r+n\rangle}_{m+\mathrm{q}(r+n)}(H^r_m)^*,$$

$r = 0, \dots, T-1$. Recall that the spectrum of $(x(n))$ is a vector measure $\gamma = (\gamma^0, \dots, \gamma^{T-1})$ such that $a_j(n) = \int_0^{2\pi} e^{-int}\gamma^j(dt)$, and that $R_x(n+r, r) = (1/T)\sum_{j=0}^{T-1} e^{2\pi ijr/T} a_j(r)$. Since for $r = 0, \dots, T-1$,

$$R_x(n+r, r) = K_{\boldsymbol{x}}^{\langle n+r\rangle, r}(\mathrm{q}(n+r)) = \int_0^{2\pi} e^{-i\mathrm{q}(n+r)t} H^{\langle n+r\rangle\cdot}(e^{it}) H^{r\cdot}(e^{it})^* dt$$
$$= \sum_{p=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} H_p^{\langle n+r\rangle} H_m^r \int_0^{2\pi} e^{i(p-m-\mathrm{q}(n+r))t} dt$$
$$= 2\pi \sum_{m=-\infty}^{\infty} H^{\langle n+r\rangle}_{m+\mathrm{q}(n+r)} H^r_m, \tag{25}$$

comparing this with the inverse discrete Fourier transform of $(b_j(n))$, we conclude that $b_j(n) = a_j(n)$, $n \in \mathcal{Z}$, $j = 0, \dots, T-1$, and hence γ is a.c. and the density of γ^j is equal $g^j(e^{it}) = h(e^{it})h(e^{i(t+2\pi j/T)})^*$ a.e.

2. We will show opposite, that is *assuming that* $(x(n))$ *is a.c., from a factor* $h(e^{it})$ *of a density* g *of* $(x(n))$ *we will construct a square factor* H *of the density of* $F(e^{it})$ *of* $(\boldsymbol{x}(n))$, *showing at the same time that* $(\boldsymbol{x}(n))$ *is a.c.*. Write $h(e^{it}) = \sum_{p=-\infty}^{\infty} h_p e^{ipt}$, so that $e^{-ikt}h(e^{it}) = \sum_{q=-\infty}^{\infty} h_{q+k}e^{iqt}$. We want to construct a function whose Fourier coefficients are equal h_{q+k} if $q = mT$, and other are zero. It is easy to see that the function $f_k(e^{iu})) = \frac{1}{T}\sum_{j=0}^{T-1} e^{-ik(u+2\pi j/T)}h(e^{i(u+2\pi j/T)})$ has this property. Indeed, denoting as previously $d = e^{2\pi i/T}$ we obtain that

$$f_k(e^{it}) = \sum_{p=-\infty}^{\infty}\left(\frac{1}{T}\sum_{j=0}^{T-1} d^{(p-k)j}\right) h_p e^{i(p-k)t} = \sum_{m=-\infty}^{\infty} h_{mT+k}e^{imTt}.$$

Clearly each function f_k is a function of tT, hence there is $h_k(e^{it})$ such that $h_k(e^{iTt}) = f_k(e^{it})$. The Fourier series of h_k is $h_k(e^{it}) = \sum_{m=-\infty}^{\infty} h_{mT+k}e^{imt}$. Let $H(e^{it})$ be the $T \times T$ matrix function which k-th row $H^{k\cdot}$ is equal h_k, i.e.

$$H^{k\cdot}(e^{it}) = \sum_{n=-\infty}^{\infty} H_n^k e^{int}, \quad \text{where } H_n^k = h_{nT+k},$$

and let $(\mathbf{y}(n))$ be a T-variate stationary sequence with the density $H(e^{it})H(e^{it})^*$. Repeating computation (25) we conclude that the covariance of the T-PC sequence $(y(n))$ corresponding to $(\mathbf{y}(n))$ equals

$$R_y(n+r,r) = 2\pi \sum_{m=-\infty}^{\infty} H^{\langle n+r\rangle}_{m+\mathrm{q}(n+r)}(H^r_m)^* = 2\pi \sum_{m=-\infty}^{\infty} h_{Tm+n+r}h^*_{mT+r}.$$

On the other hand, as it was computed in part 1.,

$$\begin{aligned} R_x(n+r,r) &= (1/T)\sum_{j=0}^{T-1} d^{jr} \int_0^{2\pi} e^{-int}h(e^{it})h(e^{i(t+2\pi j/T)})^* dt \\ &= 2\pi \sum_{m=-\infty}^{\infty} h_{mT+r+n}h^*_{mT+r}. \end{aligned}$$

Comparing this with the previous formula we see that $R_x = R_y$, and hence they have the same spectrum. We therefore conclude that $(\boldsymbol{x}(n))$ is a.c. and its density $F(e^{it}) = H(e^{it})H(e^{it})^*$ a.e.

References

1. Brockwell, P. J., & Davis, R. A. (1987). Time series: Theory and methods. In *Springer series in statistics*. Springer.
2. Dudek, A., Hurd, H., & Wójtowicz, W. (2016). PARMA methods based on Fourier representation of periodic coefficients. *Wiley interdisciplinary reviews/Computational statistics*, *8*(3), 130–149.
3. Dunford, N., & Schwartz, J. T. (1988). *Linear operators part II: Spectral theory*. Wiley.
4. Hannan, E. J. (1970). *Multiple time series*. Wiley.
5. Hannan, E. J., & Deistler, M. (2012). *The statistical theory of linear systems*. Society for Industrial and Applied Mathematics (SIAM).
6. Hurd, H. L., & Miamee, A. (2007). *Periodically correlated random sequences; spectral theory and practice*. Wiley.
7. Kaczorek, T. (2007). *Polynomial and rational matrices*. Springer.
8. Makagon, A., & Miamee, A. G. (2013). Spectral representation of periodically correlated sequences. *Probability and Mathematical Statistics*, *33*(1), 175–188.
9. Makagon, A., & Miamee, A. G.: Structure of PC sequences and the 3rd prediction problem. In F. Chaari, J. Leskow, A. Neapolitano & A. Sanchez-Ramirea (Eds.), *Cyclostationarity: Theory and methods* (pp. 53–72). Lecture notes in mechanical engineering. Springer.
10. Makagon, A. (2015). Innovation and factorization of the density of a regular PC sequence. *Probability and Mathematical Statistics*, *35*(1), 15–30.
11. Rozanov, Y. A. (1967). *Stationary random processes*. Holden-Day series in time series analysis. Holden-Day.
12. Sakai, H. (1991). On the spectral density matrix of a periodic ARMA process. *Journal of Time Series Analysis*, *12*, 73–82.
13. Wylomanska, A. (2008). Spectral measures of PARMA sequences. *Journal of Time Series Analysis*, *29*(1), 1–13.

Fault Detection in Belt Conveyor Drive Unit via Multiple Source Data

Piotr Kruczek, Jakub Sokołowski, Jakub Obuchowski, Mateusz Sawicki, Agnieszka Wyłomańska and Radosław Zimroz

Abstract The fault detection for the belt conveyor drive unit gives a possibility to avoid unexpected breaks in production and to plan the repairs for the most convenient time. Trouble-free operation of the belt conveyor system in an underground mine is one of the most essential requirement in case of production plan realization. In order to maintain the system, several features (energy consumption, output, temperature, vibrations, belt speed etc.) might be monitored. For instance, belt conveyors can be equipped with sensors, which measure the temperature of their transmission and the drive unit electricity consumption. The paper aims to examine the possibility of multivariate data analysis application for fault detection. The authors focused on the extraction of additional information from the relation between the transmission temperature and electrical energy consumption (electric current). Namely, in the first step, the electric current signal is smoothed in order to reduce its high-frequency volatility. Then the relationship between smoothed current signal and the temperature is examined. It was observed that the specifically smoothed current might explain the variation of the temperature. Variability of fitted model parameters may indicate a change in the drive unit condition. The findings

P. Kruczek (✉) · J. Sokołowski · J. Obuchowski · M. Sawicki · R. Zimroz
KGHM CUPRUM Ltd, CBR, Sikorskiego 2-8, Wrocław 53-659, Poland
e-mail: pkruczek@cuprum.wroc.pl

J. Sokołowski
e-mail: jsokolowski@cuprum.wroc.pl

J. Obuchowski
e-mail: jobuchowski@cuprum.wroc.pl

M. Sawicki
e-mail: msawicki@cuprum.wroc.pl

R. Zimroz
e-mail: rzimroz@cuprum.wroc.pl

A. Wyłomańska
Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center,
Wroclaw University of Science and Technology, Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
e-mail: agnieszka.wylomanska@pwr.edu.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_10

are illustrated by analysis of signals acquired on the machine and it covers period before damage occurred till the repair. Finally, model outputs for the three stages have been analyzed: before the damage, during its development and after repair. The obtained results prove that the relationship between current and temperature depends on the condition of the drive system. The outcome confirms that there is a possibility to apply context-based, multiple sources data analysis for fault detection in belt conveyor drive unit.

Keywords Belt conveyor · Fault detection system · Drive unite diagnostics · Multiple source data analysis · Gearbox temperature dynamics

1 Introduction

Trouble-free operation of mechanical systems is of high importance in many industries. Automotive, chemical, aerospace and mining industries are prominent examples of branches, where condition of mechanical systems plays an important role [1–9]. Any unexpected break might result in problems with fulfillment of the production schedule. Today's technological opportunities give possibility to monitor a lot of informative features related to operational condition of machinery systems. This raises a need to develop continuous-time fault detection systems that would help to maintain complex systems that consists of many machines. Surely, mining networks of belt conveyors are illustrative examples of such systems. Belt conveyors, especially these that are connected serially, are exposed to unexpected downtimes due to fault of their elements. One of the most crucial components of the belt conveyor is the conveyor belt. There are many systems that quantify condition of the belt. Each type of conveyor belt is related to specific approaches which can be beneficial in terms of fault detection. For instance, belts with steel cord might be diagnosed using systems that measure magnetic disruptions [10–14]. Condition of fabric conveyor belts might be assessed using X-ray imaging or other non-destructive methods [15, 16]. Both of these approaches require advanced methods of signal processing [17, 18]. Another crucial part of belt conveyor network is the drivetrain, which usually consists of electric motors, shafts, gearboxes and pulleys. In this case, monitoring systems are often equipped with sensors that measure vibrations, temperatures, weight of the transported material, electricity consumed by motors, etc. [19–25]. Fault detection systems require not only to monitor the acquired data, but the main purpose is to provide decisions whether the considered part of the conveyor is faulty or not. In this paper we propose a data processing method that integrates recordings from different sources in order to indicate technical condition of the considered machine—a part of the belt conveyor system. Namely, it is presented that the electric current of engines and temperature measured on gearboxes might provide a reliable diagnostic information that might

help to minimize stall time of the belt conveyor network in the underground mine. The methodology presented in this paper is based on the concept presented in [26, 27], where the Authors analyze not the diagnostic signals themselves, but their relations to another, auxiliary parameters.

The paper is structured as follows. In Sect. 2 comprehensive data description is provided. Data pre-processing required for further analysis is presented in Sect. 3. Methodology that helps to assess condition of the belt conveyor drive unit is contained in Sect. 4. Application of the proposed methodology to real data from a commercial monitoring system is presented in Sect. 5. The last section contains conclusion.

2 Analyzed Data Description

In this paper temperature and current records acquired by commercial, multichannel low frequency data logger are analyzed. The measurement systems operate continuously in an underground copper ore mine. Data is collected on the four gearboxes and electric motors installed on the same belt conveyor. Scale of the belt conveyor system causes that the monitoring system measures a lot of signals. In order to avoid overwhelming sizes of the data files, variables are quantized into finite sets denoted $\mathrm{T} = \{\tau_1, \tau_2, \ldots, \tau_{n_t}\}: \tau_j < \tau_{j+1}$ for temperature, and $I = \{i_1, i_2 \ldots, i_{n_i}\}: i_j < i_{j+1}$ for current. Namely, the system saves a particular observation if it differs from previous one more than a given threshold. Therefore, when the phenomena is relatively stable, the records are not registered. Each record possesses information of time (year, month, day, hour, minute and second) and related physical quantity. Unfortunately, this memory saving solution comes with difficulties in signal processing. During specific time interval each channel might possess different amount of values recorded in different time points. Therefore, integration of data from different channels might be difficult. Thus, signal interpolation is strongly recommended and can solve this problem. Moreover, in analyzed observation one can notice some outliers. Let τ_j be a temperature recorded at the k-th time point, and assume that it corresponds to j-th value from the predefined set of quantized temperature data, namely $x_k = \tau_j$. Due to nature of temperature and continuous work characteristic of a system, x_{k+1} should belong to $\{\tau_{j-1}, \tau_{j+1}\}$. However, some records do not fulfill this pattern. They might take values from a set $\{-30\ °\mathrm{C},\ 129\ °\mathrm{C}\}$ for a few seconds or minutes and afterwards return to usual values of temperature. Such observations are recognized as errors and should be removed from the data sample, since temperatures of transmission in this mine are not lower than 20 °C nor higher than 90 °C (even for overheated faulty gearbox). In the Fig. 1 the exemplary conveyor belt gearbox and the temperature sensor are presented.

Fig. 1 Conveyor belt gearbox and the temperature sensor

3 Data Pre-processing

As it was mentioned in previous section, analysis of relation between raw temperature and current signals are difficult since sampling is time-varying and it is different for each channel. Firstly, error values should be erased from data and both temperature and current have to be interpolated at identical time axis for each channel in order to make analysis simpler. It was decided to interpolate data at equally spaced time points with one minute time intervals. Due to physical properties of analyzed quantities it was decided to use linear interpolation of temperatures and stepwise for electric current. Linear interpolation of temperature is motivated by its slow variation in time. Jumps and periods with constant value of the electric current (engine is switched off) lead to stepwise interpolation. The procedure of such data pre-processing is presented in Fig. 2. The linear interpolation is curve fitting by the first order polynomial. Namely, the interpolated values between two records are calculated from fitted linear function. The stepwise method is even simpler. All of the interpolated variables are equal to the last recorded quantity.

3.1 Data Smoothing

We assume that electric current consumed by the conveyor belt engine has an impact on the gearbox temperature. Moreover, it is expected that the temperature depends rather on past current values measured through some period than on its instantaneous value. In order to process the real data the smoothing algorithms can be used. They provide somehow averaged time series corresponding to a given time horizon. There are many well-known methods that can be applied. In this article we present two of them. The first one is moving average (MA) and the second one is exponential smoothing.

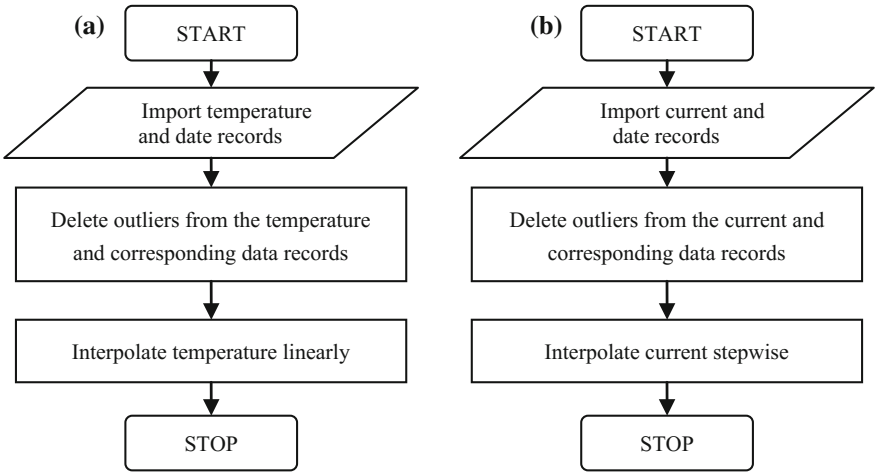


Fig. 2 Data pre-processing steps for temperature (**A**) and electric current (**B**)

3.1.1 Moving Average

This method is based on averaging the data on the given time period with the series of weights. The formula for signal X_t, smoothing length k (horizon) and weights w_i, where $i = 1, \ldots, k$ is given:

$$Y_t = \sum_{i=1}^{k} w_i X_{t-i}. \tag{1}$$

In our data we use the equal weights for all $i = 1, \ldots, k$, i.e. $w_i = \frac{1}{k}$. Then all past k electric current recorded values have the same influence on the averaged temperature Y_t. The discussion on appropriate choice of the time horizon k is contained in Sect. 5.

3.1.2 Exponential Smoothing

In the second method impact of the historical electric current data decreases with respect to time (according to the smoothing parameter m), what might be more relevant for real data than equally weighted moving average. The formula for given smoothing parameter $m \in (0,\ 1)$ and signal X_t is presented below:

$$\begin{aligned} Y_t &= mX_t + (1-m)Y_{t-1}, t > 1 \\ Y_1 &= X_1. \end{aligned} \tag{2}$$

Moreover the explicit form of smoothed signal can be expressed by following equation:

$$Y_t = \sum_{j=0}^{t-2} m(1-m)^j X_{t-j} + (1-m)^{t-1} X_1. \tag{3}$$

In this method the smoothing parameter m has to be specified. Smaller value of m results in slow decrease of weights, thus past values has relatively large influence on current Y_t. Higher value of m stands for high influence of latest values on the smoothed signal. Thus, the output of the smoothing procedure might still contain some high-frequency variations. The parameter m equals to 1 results in $Y_t = X_t$ for each t. From the above equation we can derive the sum $S_{m,l}$ of weights applied to l latest recordings:

$$S_{m,l} = \sum_{j=0}^{l} m(1-m)^j = \sum_{j=1}^{l+1} m(1-m)^{j-1} = 1-(1-m)^{l+1}. \tag{4}$$

From above equation one can estimate the influence of the l latest observations to the currently calculated Y_t for given m. Moreover, the formula for smoothing parameter m depending on given sum $S_{m,l}$ can be derived:

$$m = 1 - \sqrt[l+1]{1-S_{m,l}}. \tag{5}$$

Therefore, one can require that latest l observations are responsible for $p\%$ of currently calculated Y_t. Alternatively, one can assume a smoothing parameter and calculate the influence of l latest weights to Y_t. For example, for sampling frequency of data equal to 1 min and smoothing parameter $m = \frac{1}{120}$, the sum of weights related to 120 latest recordings (2 h) is $S_{1/120,\,120} \approx 0.6367$. Therefore, 120 latest recordings account for about 64% of currently calculated Y_t.

In further analysis we exploit temperature signals interpolated linearly at equally spaced one-minute intervals and stepwise-interpolated current signals smoothed using both mentioned methods, i.e. moving average with equal weights and exponential smoothing.

4 Diagnostic Methodology

In this section we propose the methodology for fault detection in belt conveyor drive unit. It is relies on the change in the relation between electric current and temperature acquired on the electric motor and the gearbox, respectively. The analysed data is sampled on one minute basis. In the first step the pre-processed signals are separated into 24 h groups starting at 6 a.m. This is motivated by the fact that the working day in the mine starts at 6 a.m. and it consists of 4 complete shifts

of equal duration. Due to the work cycle, the belt conveyors usually do not work from Saturday, 6 a.m. till Monday, 6 a.m. Moreover, signals obtained on Mondays (Mon, 6 a.m.–Tue, 6 a.m.) are significantly different from the rest of the week, since on Mondays the machines start working after a long brake. Therefore, signals from 4 working days (Tuesday 6 a.m.–Saturday 6 a.m.) are easily comparable.

From the diagnostic point of view, the most informative parts of the monitored signals are related to periods with high load (the load of conveyor is often high, if only the conveyor operates) or operation without load (increasing energy consumption during idle state might indicate increasing resistance). The monitoring system installed in this particular underground mine covers the most important conveyors among the network. These conveyors operate usually under high load, thus there might be insufficient amount of data related to idle operation to perform diagnostics. In order to select data related to high load, the proper thresholding for smoothed electric current might be used. Namely, we investigate relation between current and temperature for which the averaged current is higher than a given threshold. Moreover, in the analysis different values of thresholds are being tested, that the best one can be chosen.

Clearly, the data records can be separated into two groups with respect to temperature trend. The authors decided to consider only data for which the temperature does not decrease, since it corresponds to heating dynamics.

The last stage is focused on the proper model selection. Authors observed that for such pre-processed data the linear regression is suitable. Namely, the following model might be applied:

$$T(t) = aES(t) + b, \tag{6}$$

where T is temperature, t is time point, ES smoothed electric current and a, b are the model parameters to be estimated. In order to test the goodness of fit the R^2 measure is applied. Linear regression can explain the relation between variables. Changes of the model parameters during machine operation might indicate damage, similar to the methodology presented in [26] for wind turbines condition monitoring. To wit, faulty machine would reach higher temperature than the healthy one, for a given load. Moreover the model is simple, does not require time-consuming calculations and it is easy to interpret. In order to diagnose a machine the regression line (namely slope and intercept) can be analysed. The results of the application of proposed method for real signals are presented in the next section.

5 Real Data Application Results

The analysed data consist of temperature and electric current recorded on 4 gearboxes from one belt conveyor. Signals are pre-processed as it is described in Sect. 2. Furthermore, one of the machine (no. 54) was damaged and repaired during the analysed period. The main goal is to show capability of damage detection with

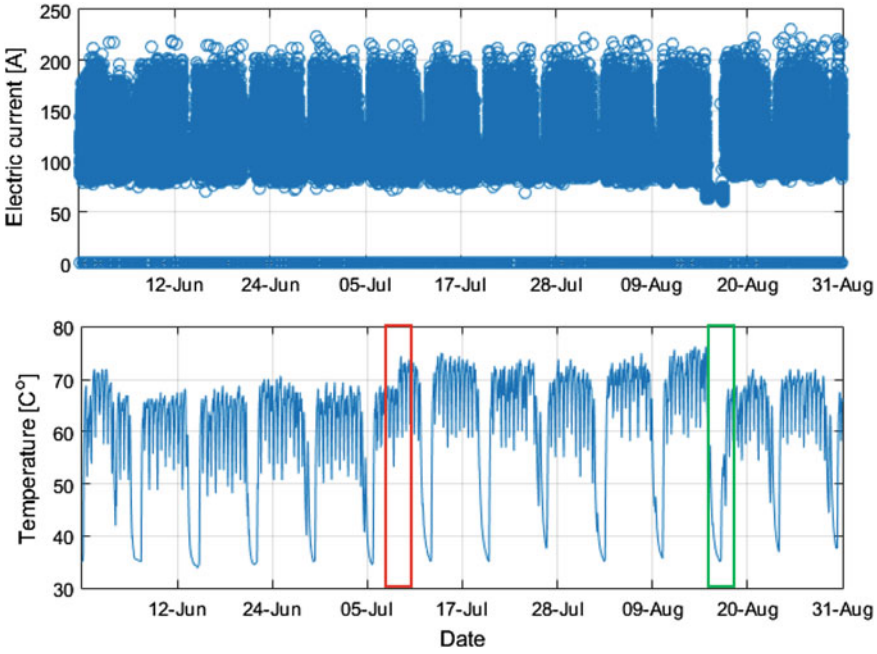


Fig. 3 Current and Temperature signals recorded on damaged gearbox. *Red box* marks the period, when the temperature increased significantly. The *green box* shows the period when the machine was repaired

proposed methodology. Thus, the model was applied to faulty gearbox. The comparison of the result with transmission in good condition is also provided.

First of all, the signal of electric current and temperature recorded on gearbox 54 are presented in the Fig. 3.

Clearly, before the fault component was replaced, the anomaly temperature is observed. After the repair the temperature returned to normal level. Additionally, one can be interested in developing an automatic procedure for detection of such anomaly. As it was mentioned in previous sections the model parameters are estimated on the smoothed electric current instead of the raw signal. The comparison of two introduced averaging methods is depicted in Fig. 4. Furthermore, the corresponding values of temperature are also plotted. The exponential smoothed parameter was set to $\frac{1}{120}$ and MA parameter is equal to 180 (180 min, half of one shift). One can observe specific relation between current and temperature. The increase of smoothed electric current implies increase of the temperature.

In order to compare the smoothing methods the linear regression was applied. Then the R^2 measure provides information, which averaging procedure is more suitable for the real signals. The whole data can be divided into three subgroups, depending on the stage of damage development. First one, before the damage appeared, corresponds to relatively low temperature (usually below 70 °C). When

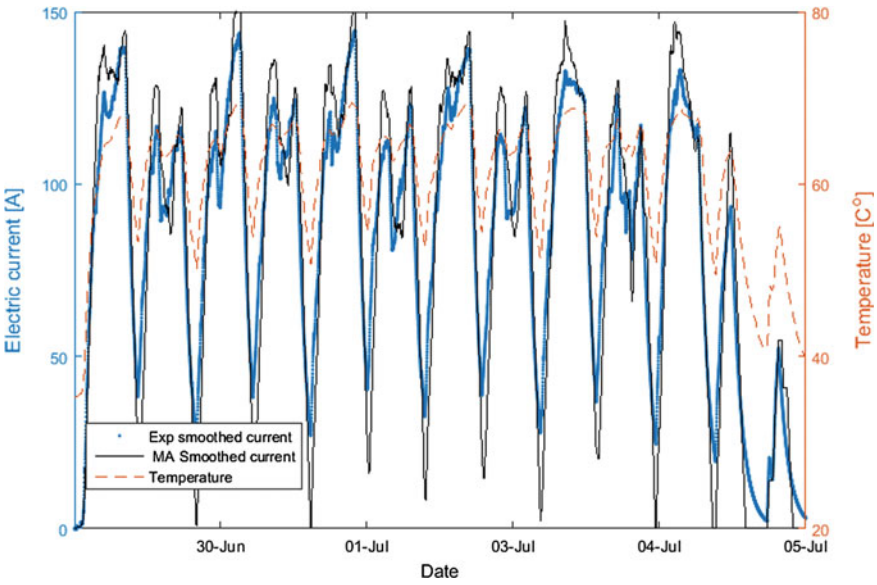


Fig. 4 Smoothed electric current and temperature comparison

the temperature rises, the second subgroup can be noticed—it stands for damage development. Finally, the third subgroup consists of signals acquired after repair, which corresponds to a good condition. In Fig. 5 there are presented the linear regression results for each day during the experiment. The plots on the top are related to the day before the growth of the temperature observed in the Fig. 3. The middle plots represent the day after the anomaly increase. The temperature-current relation is presented on the bottom panel. Furthermore, two electric current averaging methods are also compared. One could expect that the exponentially smoothing is more adequate, because the weights decrease with respect to time. Indeed, the goodness of fit for this smoothing method is much better. All R^2 are higher than 0.7 and for good condition they are even higher than 0.9. It means that temperature data, during heating periods, can be well explained by exponentially smoothed electric current. On the other hand, one can easily observe that moving average method is not good enough. The R^2 are much smaller (except one day) and the linear regression should not be applied for such data.

Once the smoothing method is chosen, the performance of the proposed diagnostic method can be examined. Apparently, the differences between healthy and faulty machine can be observed. Plots for days before temperature growth and after repair looks similarly (Fig. 6). On the contrary, the middle plot is almost shifted in parallel, with the higher value of intercept b and similar slope a.

Figure 5 presents only exemplary results from the analysed time horizon. In Fig. 6 there are presented the outputs of the fitted models for each analysed day. As it could be expected, the difference between healthy and faulty condition states is clear. The parallel shift of regression lines is noticeable, the ones from the damage

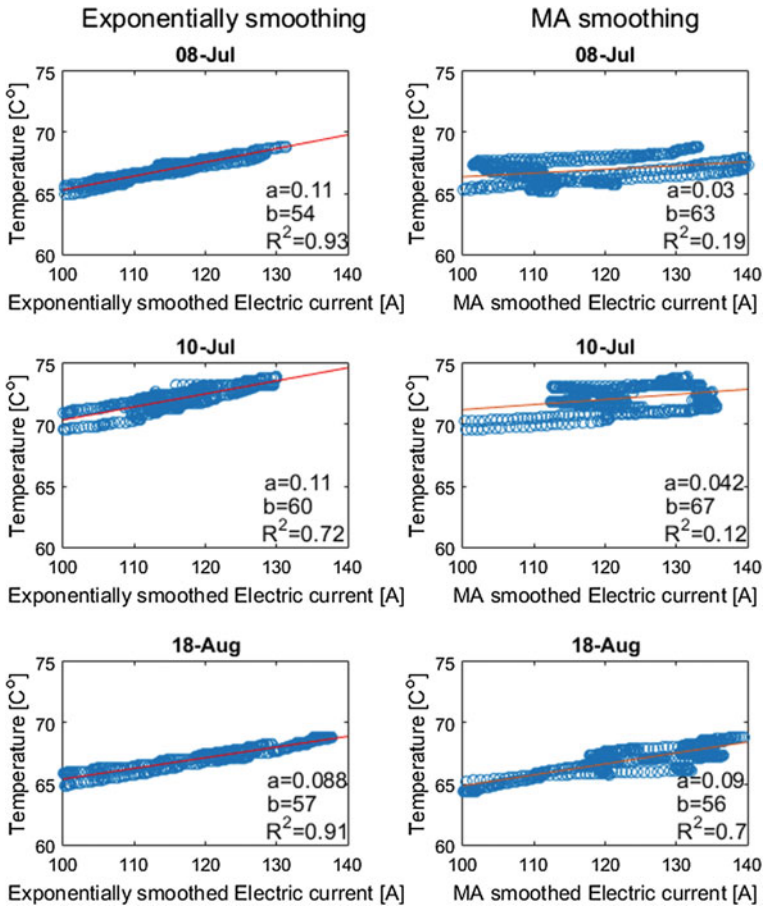


Fig. 5 Linear regression applied to temperature versus current. A day before damage (*top panels*), during damage (*middle panels*) and after repair (*bottom panels*). The results are presented for exponentially smoothed (*left side*) and MA smoothed (*right side*) data

development period are much above the other ones. Thus, the linear models fitted to conveyor belt data can be considered as an indicator of machine's condition.

The presented result are obtained for the exponential smoothing parameter equal to $\frac{1}{120}$. The lower bound for smoothed electric current was set to 100 A—it indicates that the load is applied to the system. Moreover, only the data acquired during temperature increase is analyzed and signals during weekends and one day after them are not considered. Thus, the data from different days are comparable and it can be concluded that the changes in fitted model parameters are related to change of machine condition.

The comparison of the outputs of the model fitted to different gearboxes is presented in Fig. 7. Condition of each gearbox is represented by value of the fitted

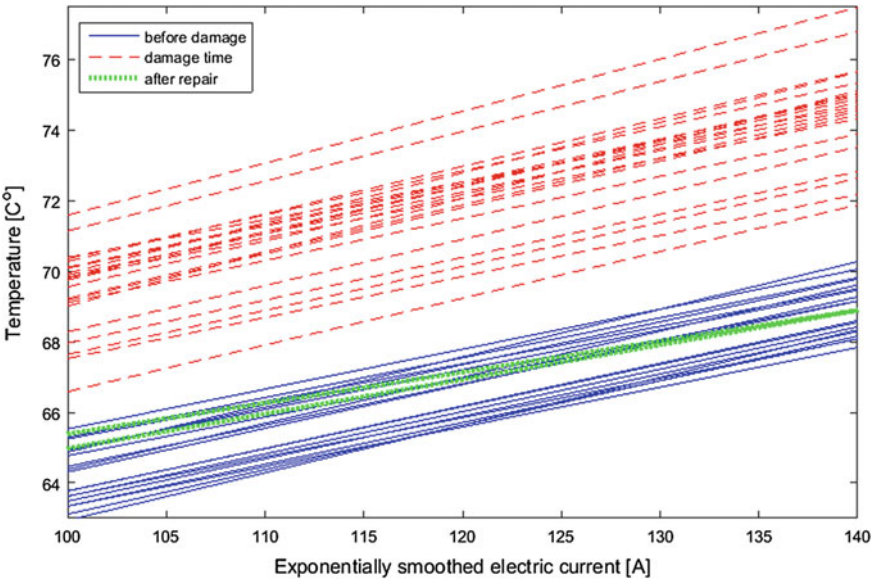


Fig. 6 Regression lines for days divided into 3 groups with respect to machine condition

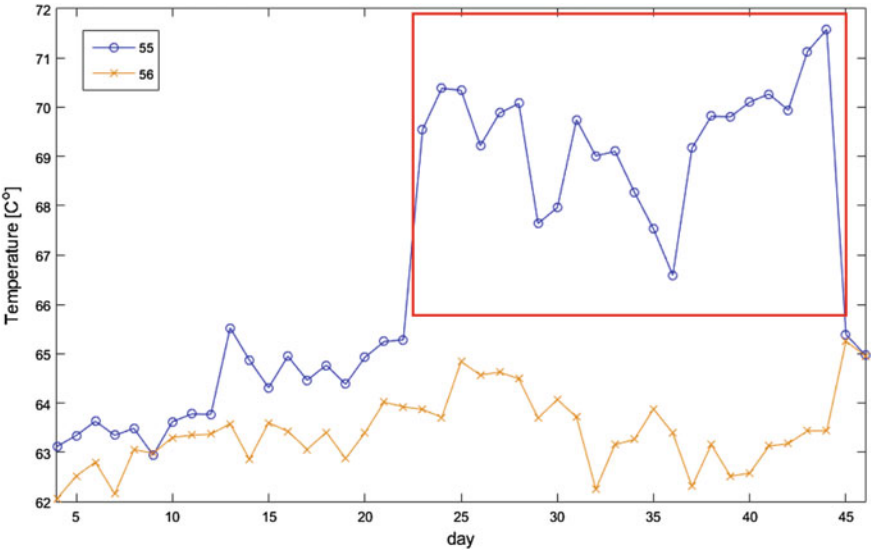


Fig. 7 Values of the regression lines at smoothed electric current equal to 100 A, two different gearboxes. The damaged one is no. 55. *Red box* marks the period, when machine was in faulty condition

regression line at electric current equal to 100 A (diagnostic feature). The damage of gearbox no. 55 can be clearly noticed therein. After the repair, the value of the proposed diagnostic feature returns to the level before damage development. Clearly, only on the gearbox no. 55 the diagnostic feature changes significantly. Furthermore, the proposed feature does not increase significantly during the entire considered period. Thus, the proposed model that combines smoothed interpolated data of temperature and electric current is able to detect abnormalities in technical condition of the gearbox.

6 Conclusions

In this paper the temperature and electric current signals from belt conveyor were analyzed. One of the gearboxes revealed a damage during the investigated time horizon. The main goal was to propose a damage detection method based on available quantities, i.e. electric current from engines and temperatures measured on gearboxes. The data was pre-processed and two different smoothing methods were investigated. Selection of appropriate parts of the data was motivated by their usability to indicate temporal influence of electric current to heating dynamics. The regression model was proved as suitable connection between two investigated quantities. Exponential smoothing method turned out to be the most suitable in terms of coefficient of determination. Upon the fitted model parameters, appropriate diagnostic feature was proposed. It was shown that there is a possibility to indicate temporal condition of the machine using the proposed feature. Due to simplicity of this parameter, the proposed method might be easily applied in the underground mine from which the data originate. Although, there is a need to analyze whether such idea of combining temperature and load might be applied to other parts of the belt conveyor system, e.g. pulleys or rollers. It has to be mentioned that proposed method needs a training period when the machine is healthy, thus it cannot be applied for one-off measurements. It is suitable for continuous monitoring of the machine condition and is able to detect abnormal cases.

References

1. Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483–1510.
2. Czech, P., Wojnar, G., Burdzik, R., Konieczny, Ł., & Warczek, J. (2014). 1268. Application of the discrete wavelet transform and probabilistic neural networks in IC engine fault diagnostics. *Journal of Vibroengineering, 16*(4).
3. Konieczny, Ł., Burdzik, R., & Łazarz, B. (2013). 1113. Application of the vibration test in the evaluation of the technical condition of shock absorbers built into the vehicle. *Journal of Vibroengineering, 15*(4).

4. Randall, R. B. (2011). *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. Wiley.
5. Yin, S., Li, X., Gao, H., & Kaynak, O. (2015). Data-based techniques focused on modern industry: an overview. *IEEE Transactions on Industrial Electronics, 62*(1), 657–667.
6. Trutt, F. C., Sottile, J., & Kohler, J. L. (2002). Online condition monitoring of induction motors. *IEEE Transactions on Industry Applications, 38*(6), 1627–1632.
7. Blazej, R., Jurdziak, L., Kirjanow, A., & Kozlowski, T. (2015). Evaluation of the quality of steel cord belt splices based on belt condition examination using magnetic techniques. *Diagnostyka, 16*(3).
8. Kołowrocki, K., & Soszyńska-Budny, J. (2015). Statistical identification of complex technical system operation process. *Diagnostyka, 16*(4).
9. Figlus, T. (2015). 1592. The application of a continuous wavelet transform for diagnosing damage to the timing chain tensioner in a motorcycle engine. *Journal of Vibroengineering, 17*(3).
10. Mao, Q., Ma, H., Zhang, X., & Zhang, D. (2011). Research on magnetic signal extracting and filtering of coal mine wire rope belt conveyer defects. In *2011 3rd International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (Vol. 3, pp. 18–22). IEEE.
11. Błażej, R., Kirjanów, B., & Kozłowski, T. (2014). A high resolution system for automatic diagnosing the condition of the core of conveyor belts with steel cords. *Diagnostyka, 15*(4), 41–45.
12. Langebrake, F., Klein, J., & Gronau, O. (1998). Non-destructive testing of steel-cord conveyor belts. *Bulk Solids Handling(Switzerland), 18*(4), 565–570.
13. Harrison, A. (1985). A magnetic transducer for testing steel-cord deterioration in high-tensile strength conveyor belts. *NDT International, 18*(3), 133–138.
14. Alport, M. J., Basson, J. F., & Padayachee, T. (2007). Digital magnetic imaging of steel cord conveyor belts. In *14th International Materials Handling Conference* (pp. 1–2).
15. Fourie, J. H., Alport, M. J., Basson, J. F., & Padayachee, T. (2005). Condition monitoring of fabric-reinforced conveyor belting using digital x-ray imaging. *Bulk Solids Handling, 25*(5), 290.
16. Harrison, A. (1996). 15 years of conveyor belt nondestructive evaluation. *Bulk Solids Handling(Switzerland), 16*(1), 13–19.
17. Xian-Guo, L., Chang-Yun, M., Wen, W., & Yan, Z. (2011). Research on fault detection algorithm of steel cord conveyor belt based on gabor filtering and image fusion. *Journal of Convergence Information Technology, 6*(11).
18. Obuchowski, J., Wylomańska, A., & Zimroz, R. (2014). Recent developments in vibration based diagnostics of gear and bearings used in belt conveyors. In *Applied Mechanics and Materials* (Vol. 683, pp. 171–176). Trans Tech Publications.
19. Zimroz, R., Hardygóra, M., & Blazej, R. (2015). Maintenance of belt conveyor systems in poland–An overview. In *Proceedings of the 12th International Symposium Continuous Surface Mining-Aachen 2014* (pp. 21–30). Springer.
20. Stefaniak, P. K., Wyłomańska, A., Zimroz, R., Bartelmus, W., & Hardygóra, M. (2016). Diagnostic features modeling for decision boundaries calculation for maintenance of gearboxes used in belt conveyor system. In *Advances in Condition Monitoring of Machinery in Non-Stationary Operations* (pp. 251–263). Springer.
21. Kacprzak, M., Kulinowski, P., & Wedrychowicz, D. (2011). Computerized information system used for management of mining belt conveyors operation. *Eksploatacja I Niezawodnosc-Maintenance and Reliability, 2,* 81–93.
22. Stefaniak, P. K., Wyłomańska, A., Obuchowski, J., & Zimroz, R. (2015). Procedures for decision thresholds finding in maintenance management of belt conveyor system–statistical modeling of diagnostic data. In *Proceedings of the 12th International Symposium Continuous Surface Mining-Aachen 2014* (pp. 391–402). Springer.
23. Peruń, G., & Opasiak, T. (2016). Assessment of technical state of the belt conveyor rollers with use vibroacoustics methods—preliminary studies. *Diagnostyka, 17*(1), 75–81.

24. Hardygóra, M., Bartelmus, W., Zimroz, R., Król, R., & Błażej, R. (2009). Maintenance, diagnostics and safety of belt conveyors in the operations. *Transport and Logistics (Belgrade), 6,* 351–354.
25. Zimroz, R., Krol, R., Hardygora, M., Gorniak-Zimroz, J., Bartelmus, W., Gładysiewicz, L., & Biernat, S. (2011). A maintenance strategy for drive units used in belt conveyors network. In *22nd World Mining Congress & Expo, Istanbul, September* (pp. 11–16).
26. Bartelmus, W., & Zimroz, R. (2009). A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing, 23*(5), 1528–1534.
27. Zimroz, R., Bartelmus, W., Barszcz, T., & Urbanek, J. (2014). Diagnostics of bearings in presence of strong operating conditions non-stationarity—A procedure of load-dependent features processing with application to wind turbine bearings. *Mechanical Systems and Signal Processing, 46*(1), 16–27.

Application of Independent Component Analysis in Temperature Data Analysis for Gearbox Fault Detection

Jacek Wodecki, Pawel Stefaniak, Mateusz Sawicki and Radoslaw Zimroz

Abstract In the real multisource signal analysis one of the main problems is the fact that true information is divided partially among the individual signals and/or measured signal is a mixture of different sources. This comes from the fact that input channels are typically related, and carry information about different processes occurring during the measurement. Those processes can be thought of as independent sources of vaguely understood "information". In many cases separation and extraction of those sources can be crucial. In this paper we present the usage of Independent Component Analysis as a tool for information extraction from real-life multichannel temperature data measured on heavy duty gearboxes used in mining industry. Original signals, due to operational factors reveal cyclic variability and detection of damage was difficult. Thanks to proposed procedure, from four channels acquired from 4 gearboxes driving belt conveyor we have extracted one of 4 components, that is related to change of condition of a single gearbox. For new signal visibility of change is clear and simple automatic detection rule can be applied.

Keywords ICA · Gearbox · Belt conveyor · Multichannel data · Temperature

J. Wodecki (✉) · P. Stefaniak · M. Sawicki · R. Zimroz
KGHM Cuprum Ltd., CBR Sikorskiego 2-8, 53-659 Wrocław, Poland
e-mail: jwodecki@cuprum.wroc.pl

P. Stefaniak
e-mail: pkstefaniak@cuprum.wroc.pl

M. Sawicki
e-mail: msawicki@cuprum.wroc.pl

R. Zimroz
e-mail: rzimroz@cuprum.wroc.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_11

1 Introduction

Information extraction is a very important topic in many branches of the field of signal processing. It is not easy for individual signals, but multivariate (multi-channel, multidimensional) signals can provide some interesting opportunities. For example, we can assume that information is hidden and distributed among many input channels and/or measured signal is a mixture of different sources. Blind Source Separation (BSS) methods come in handy in that case [1–6]. The idea is the separation of a set of source signals from a set of mixed signals, without the aid of information (or with very little information) about the source signals or the mixing process. In that case we consider the input channels to be mixtures of some underlying sources, processes or events. This problem is in general highly underdetermined, but useful solutions can be derived under a surprising variety of conditions. One of very useful methods is called Independent Component Analysis, and its approach can be particularly effective if one requires not the whole signal, but merely its most salient features [7–11]. It is a computational method for separating a multivariate signal into additive subcomponents. This is done by assuming that the subcomponents are non-Gaussian signals and that they are statistically independent from each other. This paper presents the usage of ICA to extract feature carrying information about non-typical failure-related signal changes over time.

2 Methodology

In this section we present the methodology which we use to real temperature signals from the set of heavy duty gearboxes of belt conveyor driving station used in mining industry. Functional scheme of the procedure is presented in Fig. 1.

Temperature data acquired from SCADA systems are frequently used for condition monitoring purposes especially for mechanical system with relatively hard access as wind turbines farm, off shore or mining machines operating underground [12–21]. However, in our case, due to several factors influencing variability of temperature, raw data are difficult to process and decision making process related to damage detection is problematic.

In previous work we have tried to simplify structure of the signal by segmentation, data driven or model based decomposition etc. In this paper we provide the alternative to those approaches using ICA. It is well known that segmentation is commonly used as pre-processing stage in many applications (speech, seismic, physics, radiation etc. [20–27]) where one can extract information related to "event", cycle, or specific signal properties. In our case such segment might be related to week or shift related cycles.

However, it is believed that "optimal" procedure should be quick and exploit information hidden in all channels simultaneously. It was proved in condition

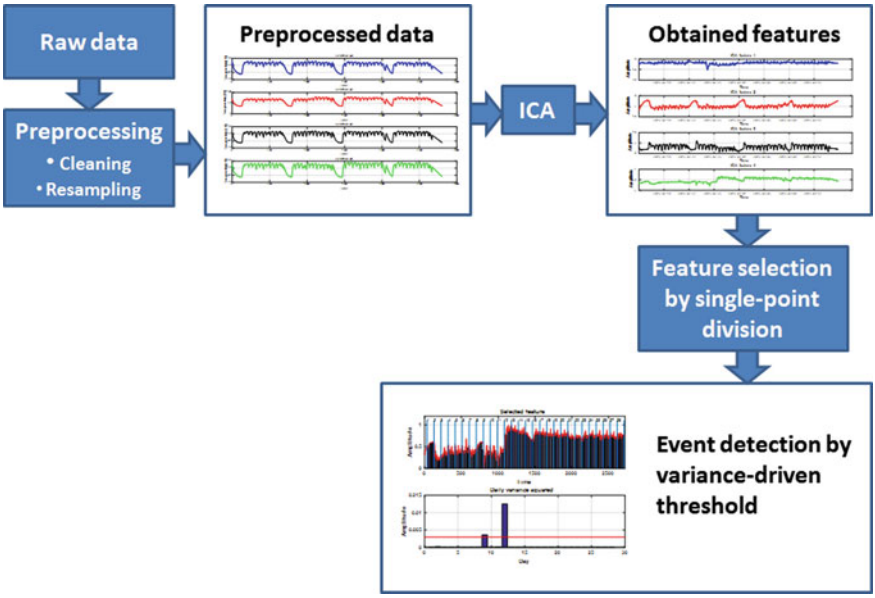


Fig. 1 Functional scheme of the algorithm

monitoring, especially for nonstationary context, that multidimensional analysis (for several channels or for several variables) might be very powerful in the context of removing redundancy and environmental influence (for example loading conditions) [28–32].

It is worth mentioning here that validation of signals acquired in harsh underground mining conditions is often a critical stage. In our case we have tried to minimize all indirect stages of processing, however some basic "data cleaning" and resampling was required.

Presented method consists of four main steps: preprocessing of the raw data, feature extraction with Independent Component Analysis, selection of desired informative feature, and event detection according to selected feature.

2.1 Preprocessing

Firstly, outliers have been removed. It turned out that there were several samples of negative values, which had to be measurement errors. They were replaced with the previous value for simplicity. Secondly, non-uniform sampling had to be dealt with.

Raw data are sampled with time-varying period depending on dynamics of the signal (i.e. temperature variability). If temperature is nearly constant, sampling frequency becomes small, if temperature is changing sampling of signal becomes

more frequent. It helps to monitor behavior of the process and in the same time saves storage required to store measurements.

Data vectors have been resampled with linear interpolation, which is a way of curve fitting with first-order polynomial. Although it is the simplest way of interpolation, it is very appropriate for this application, because we know that newly created sample between two already existing is going to take value between them. Otherwise, change-sensitive acquisition system would have register it in the first place. It also matches the character of the signal with its other property: slow variations of value.

2.2 Independent Component Analysis (ICA)

Independent component analysis was originally designed to solve so called "cocktail party problem" [7–11]. Consider two people speaking at the same time, being captured by two microphones positioned in different places, that capture two different records and producing two time signals. Let's denote those signals by $x_i(t)$ and original speech signals by $s_i(t)$. This situation can then be expressed as a linear combination:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix}$$

where a_{ij} are coefficients related to distances from microphones to speakers. Those coefficients are not known, so the problem cannot be dealt with by solving the equation with traditional methods. ICA on the other hand is designed to estimate those coefficients based on statistical independence of original sources. In our problem we treat measured signals as mixtures $x_i(t)$, and we hope to obtain distinct features from the ICA.

If we assume that the set of measured signals $\boldsymbol{X} = \{x_1(t), x_2(t), \ldots, x_n(t)\}$ is the linear combination of independent sources $\boldsymbol{S} = \{s_1(t), s_2(t), \ldots, s_m(t)\}$ where $m \leq n$, then matrix form of the ICA problem is

$$X = AS$$

where $\boldsymbol{A}$ is the coefficient matrix consisting of a_{ij} elements, and we don't know either $\boldsymbol{A}$ or $\boldsymbol{S}$ matrices. We regard the noise to be one of the sources. ICA method attempts to estimate a separating matrix $\boldsymbol{W}^T = \boldsymbol{A}^{-1}$ to be able to obtain the sources S:

$$S = A^{-1}X = W^T X$$

From the Central Limit Theorem we know that the distribution of a sum of independent random variables tends to the Gaussian distribution. Hence, since sum

of sources is expected to be more Gaussian than the sources, maximizing the non-Gaussianity of $\boldsymbol{W}^T\boldsymbol{X}$ will result in obtaining independent components. There are many measures of non-Gaussianity. In this application, the one proposed by Hyvarinen and Oja [8] based on maximum-entropy principle is used, where negentropy is defined as follows:

$$J(y) \propto \left[E\{G(y)\} - E\{G(y_{gauss})\}\right]^2,$$

where y_{gauss} has ideal Gaussian distribution, y_{gauss} and y are centered and have unity variance, $E\{\}$ is an averaging operator, and G is nonlinear function. This application of ICA algorithm uses approximation where $y = \boldsymbol{W}^T\boldsymbol{X}$ and $G(y) = tanh(y)$. Unfortunately the ICA method cannot identify uniquely neither correct ordering of the source signals, nor their proper scaling (including sign). The sign issue however can be sometimes dealt with, if our understanding of type and character of signals permits us to modify the unmixing matrix $\boldsymbol{W}^T$.

Selected properties of ICA:

- ICA can separate only linear combinations of sources;
- Order of input signals is irrelevant;
- ICA separates sources basing on maximization of non-Gaussianity, so if there were two or more perfectly Gaussian sources, they cannot be separated;
- Even if sources are not perfectly independent, ICA finds such space, in which they are maximally independent.

In the case analyzed here physical meaning of sources is obviously different than for original cocktail party context. We assume that we have mixture of four sources (processes) that describe weekly variation and showing weekend drops of value; change of state in the process connected to failure of a machine; oscillating character of signal during the days and shifts; and finally an additive white Gaussian noise which is present in every physical measurement as well as in every real-life process.

2.3 Feature Selection

Afterwards, algorithm selects feature that carries viable information about the failure-related changes in the process. For every ICA feature we search for such point in time, that splits the data into two parts of highest difference in mean. Then we check for which of features the mean difference is greatest for found point, and this is the feature that we are interested in. After that, cross-correlation of selected feature with all input signals, provides us with information which signal the misbehavior is coming from.

2.4 Event Detection

As a last step, selected feature is segmented into individual days basing on knowledge of our resampling parameters. After that, daily variance is calculated. To emphasize differences in values even more, variance is squared. At the end, threshold for variance is defined basing on variance mean, to enable automatic event detection and localization.

3 Application to Real Data

Described method was used to analyze temperature data measured on set of four heavy-duty gearboxes used in belt conveyor driving station in underground mine.

3.1 Description of Real Data

The data acquisition system used for performing measurements is a commercial, multichannel low frequency data logger that is operating continuously. The system stores the value of a variable only if it changes by some predefined value $\pm\Delta T$. This allows to reduce amount of data being recorded, which is convenient for very low frequency processes. On the other hand it can lead to some difficulties from the signal processing point of view. Firstly, measurement channels will have a different amount of samples for the same time period. Data is not distributed uniformly over time, although there is common clock for all of them, so they are in practice synchronized. It causes problems even to try to simply compare the data. Because of this situation such signal cannot be considered a time series, and it has to be preprocessed before any further actions [33].

Although signals originate from different physical sources, gearboxes drive the same conveyor within the same station, so their behavior is heavily connected. Basing on this assumption we can consider the signals describing one process in a multidimensional manner (Fig. 2). It was proven by Cempel and other authors that multidimensional analysis for condition monitoring, especially under nonstationary operations is very interesting approach [28–31]. Signals were checked for stationarity with the result of nonstationary character, due to changes in mean value.

3.2 Results

At first, signals had to be preprocessed according to methodology described in Sect. 2.1, and the output sampling period was set to 15 min. To better visualize

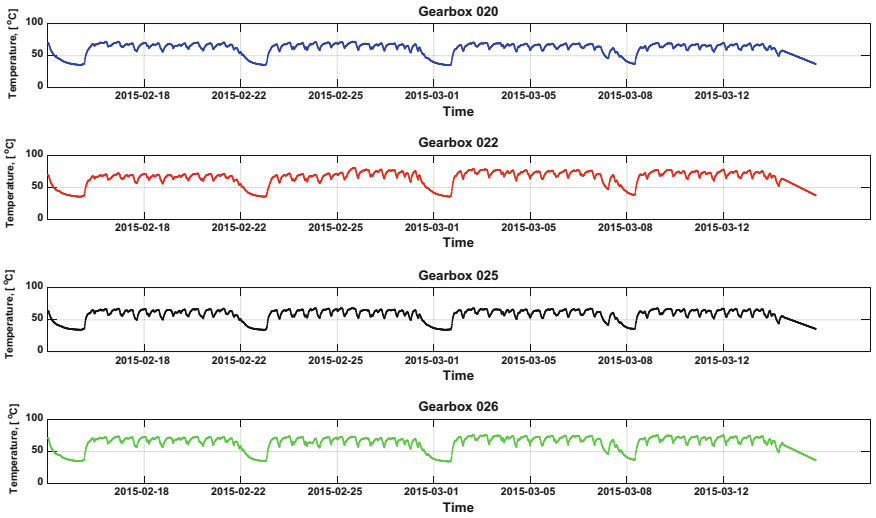


Fig. 2 Preprocessed input signals

relative changes in signal behavior, four data vectors are presented together in Fig. 3.

First conclusion when comparing Figs. 2 and 3 is that looking of each signal separately is non-effective from diagnostic point of view. It is hard to notice change in the signal, but even simple presentation all signals together shows that there is a change in the middle of week 2.

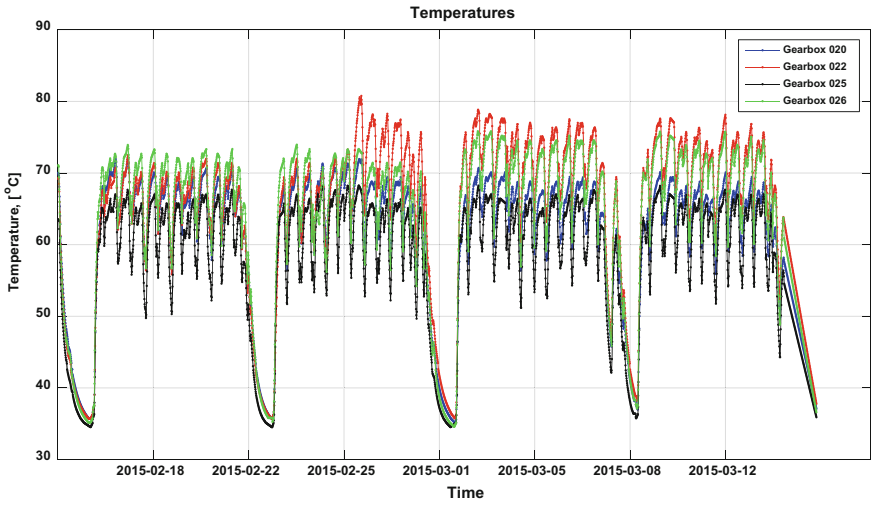


Fig. 3 Raw signals presented together

However, we are looking for an automatic technique that will allow to extract process related to change of condition only. So, ICA approach was used. In our case ICA returns four features with estimated unmixing matrix W^T:

$$W^T = \begin{bmatrix} 0.469 & 0.006 & -0.504 & 0.076 \\ -0.104 & -0.151 & -0.586 & 0.725 \\ -0.304 & 0.335 & -0.035 & -0.019 \\ -0.343 & -0.163 & 0.417 & 0.165 \end{bmatrix}$$

For reasons described in Sect. 2.2 arrangement of output features will be different every time. As we can see in Fig. 4 features 1 and 4 carry information about common factors of signals behavior, feature 2 presents highest frequency details, and feature 3 informs about trend of behavior change, so in this case third feature is selected for further analysis. Besides, cross-correlation provides the information that selected feature is connected with behavior of signal coming from gearbox 022.

Selected feature along with daily variances is presented in Fig. 5. One can see that two events can cause an alarm. Event occurring during 12th day can be clearly

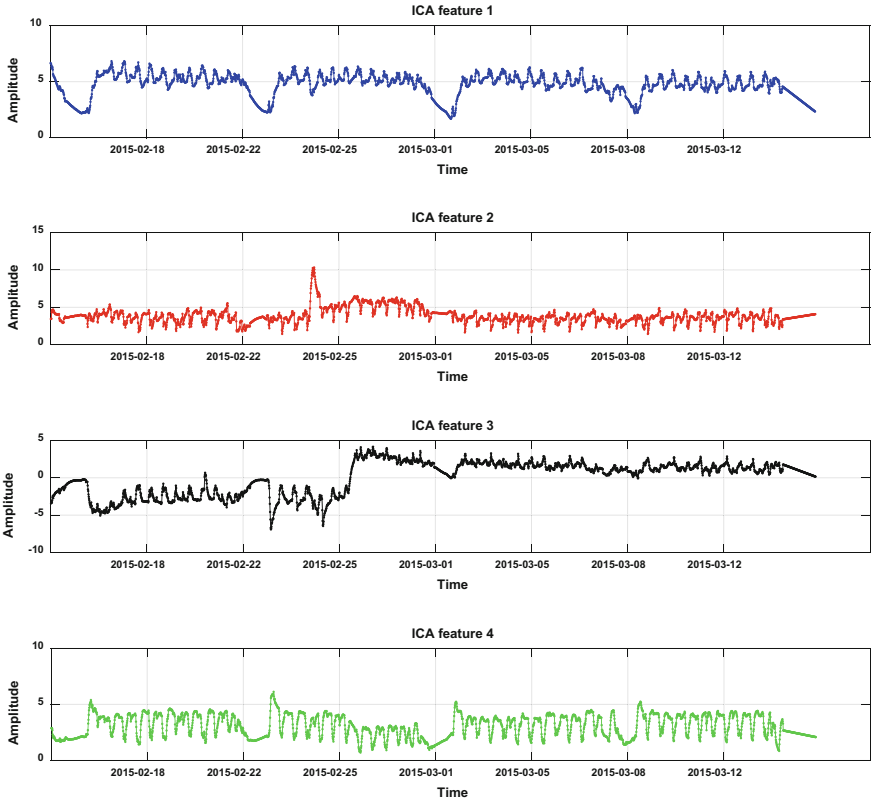


Fig. 4 Four features obtained after ICA

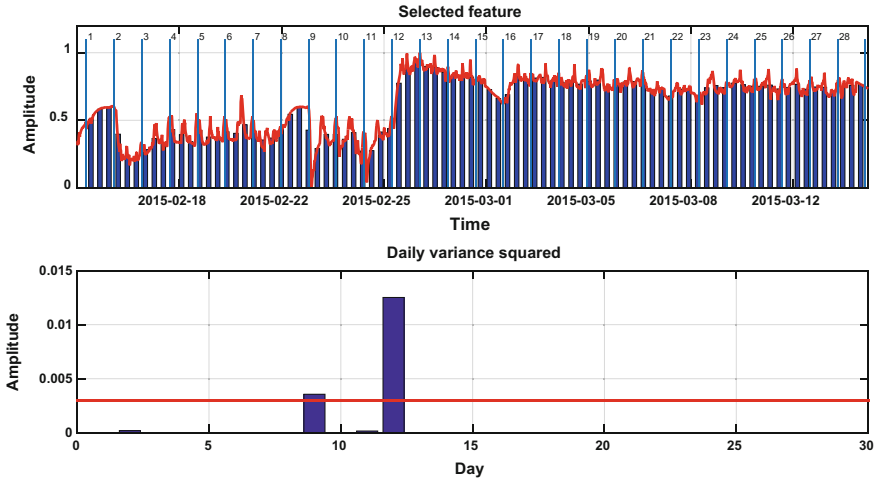


Fig. 5 Selected feature with consecutive days indicated. Daily variance squared has been proposed as event selector

visible in Fig. 3 for signal from gearbox 022. On the other hand, event occurring during 9th day is not easily visible. Figure 6 presents this situation. One can see that during 24th of February, temperature of gearboxes 026 and 022 dropped unusually compared to behavior from previous work cycles.

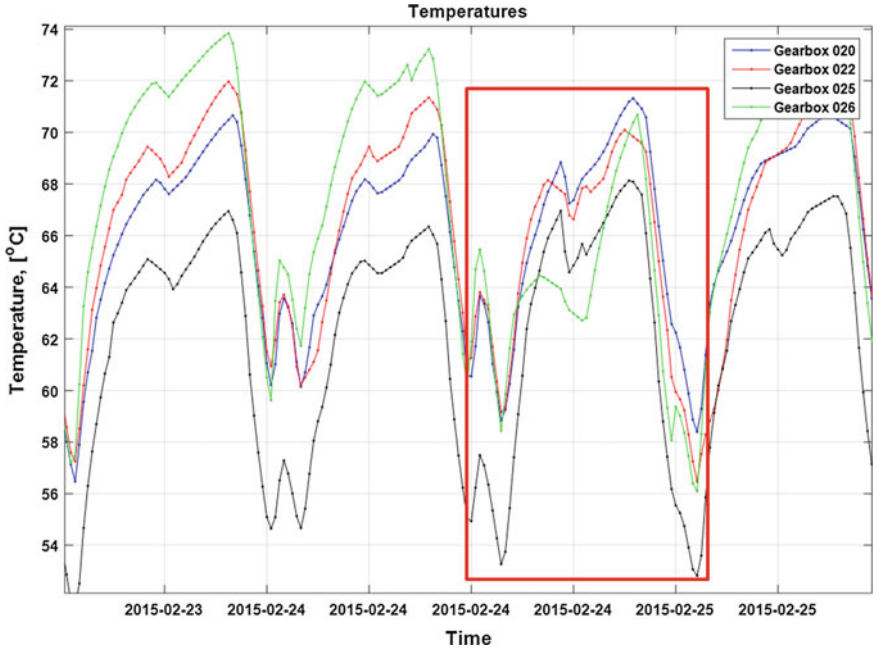


Fig. 6 Event from day 9

4 Conclusions

In this paper we have presented application of Independent Component Analysis for signal feature extraction applied to real temperature signal from set of heavy duty gearboxes used in mining industry. The methodology is based on the analysis of multichannel time series. In order to extract information about the damage we analyze the features obtained by applying the ICA to four-channel input signal. Extracted features allow to detect unusual behavior of gearboxes, identify misbehaving device and provide behavioral feature for further analysis and interpretation. By using ICA we are able to separate different sources influencing shape of measured temperature signal, or in other words, we are able to remove from signal contribution related to operational factors.

Acknowledgements This work is supported by the Framework Programme for Research and Innovation Horizon 2020 under grant agreement n. 636834 (DISIRE—Integrated Process Control based on Distributed In Situ Sensors into Raw Material and Energy Feedstock).

References

1. Roan, M. J., Erling, J. G., & Sibul, L. H. (2002). A new, non-linear, adaptive, blind source separation approach to gear tooth failure detection and analysis. *Mechanical Systems and Signal Processing*, *16*(5), 719–740.
2. Lin, J., & Zhang, A. (2005). Fault feature separation using wavelet-ICA filter. *NDT and E International*, *38*(6), 421–427.
3. Wang, J., Gao, R. X., & Yan, R. (2014). Integration of EEMD and ICA for wind turbine gearbox diagnosis. *Wind Energy*, *17*(5), 757–773.
4. Wang, H., Li, R., Tang, G., Yuan, H., Zhao, Q., & Cao, X. (2014). A compound fault diagnosis for rolling bearings method based on blind source separation and ensemble empirical mode decomposition. *PLoS ONE*, *9*(10), Article number e109166.
5. Antoni, J. (2015). Blind separation of vibration components: Principles and demonstrations. *Mechanical Systems and Signal Processing*, *19*(6), 1166–1180. ISSN 0888-3270.
6. Cichocki, A., Zdunek, R., & Amari, S.-I. (2006). New algorithms for non-negative matrix factorization in applications to blind source separation. In *2006 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2006 Proceedings* (Vol. 5).
7. Hyvarnen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, (9), 1483–1492.
8. Hyvarnen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, (13), 411–430.
9. Li, L., & Qu, L. (2002). Machine diagnosis with independent component analysis and envelope analysis. In IEEE ICIT'02 (pp. 1360–1364).
10. He, Q., Feng Z., & Kong, F. (2007). Detection of signal transients using independent component analysis and its application in gearbox condition monitoring. *Mechanical Systems and Signal Processing*, *21*(5), 2056–2071.
11. Zuo, M. J., Lin, J., & Fan, X. (2005). Feature separation using ICA for a one-dimensional time series and its application in fault detection. *Journal of Sound and Vibration*, *287*(3), 614–624.
12. Yang, W., Little, C., Tavner, P. J., & Court, R. (2014). Data-driven technique for interpreting wind turbine condition monitoring signals. *IET Renewable Power Generation, 8*(2), 151–159.

13. Guo, P., Infield, D., & Yang, X. (2012). Wind turbine generator condition-monitoring using temperature trend analysis. *IEEE Transactions on Sustainable Energy*, *3*(1), 124–133. Art. no. 5970135. doi:10.1109/TSTE.2011.2163430.
14. Astolfi, D., Castellani, F., & Terzi, L. (2014). Fault prevention and diagnosis through SCADA temperature data analysis of an onshore wind farm. *Diagnostyka*, *15*(2), 71–78.
15. Yang, W., Court, R., & Jiang, J. (2013). Wind turbine condition monitoring by the approach of SCADA data analysis. *Renewable Energy*, (53), 365–376.
16. Wilkinson, M., Darnell, B., Van Delft, T., & Harman, K. (2014). Comparison of methods for wind turbine condition monitoring with SCADA data. *IET Renewable Power Generation*, *8* (4), 390–397.
17. Zhang, X., & Fan, T.-N. (2012). The research of distribute temperature monitoring system early warning fire in coal belt conveyor. *Advanced Materials Research, 548,* 890–892.
18. Nembhard, A. D., Sinha, J. K., Pinkerton, A. J., & Elbhbah, K. (2013). Fault diagnosis of rotating machines using vibration and bearing temperature measurements. *Diagnostyka*, *14* (3), 45–51.
19. Bongers, D. R., & Gurgenci, H. (2008). Fault detection and identification for longwall machinery using SCADA data, complex system maintenance handbook. In *Springer series in reliability engineering* (pp. 611–641).
20. Zimroz, R., Wodecki, J., Krol, R., Andrzejewski, M., Sliwinski, P., & Stefaniak, P. K. (2014). Self-propelled mining machine monitoring system—Data validation, processing and analysis. In C. Drebenstedt & R. Singhal (Eds.), *Mine Planning and Equipment Selection: Proceedings of the 22nd MPES Conference* (pp. 1285–1294). Dresden.
21. Wylomanska, A., & Zimroz, R. (2015). The analysis of stochastic signal from LHD mining machine. *Springer Proceedings in Mathematics and Statistics, 122,* 469–478.
22. Zimroz, R., Madziarz, M., Żak, G., Wyłomańska, A., & Obuchowski, J. (2015). Seismic signal segmentation procedure using time-frequency decomposition and statistical modelling. *Journal of Vibroengineering*, *17*(6), 3111–3121.
23. Polak, M., Obuchowski, J., Madziarz, M., Wyłomańska, A., & Zimroz, R. (2016). Time-varying group delay as a basis for clustering and segmentation of seismic signals. *Journal of Vibroengineering, 18*(1), 267–275.
24. Gajda, J., Sikora, G., & Wyłomańska, A. (2013). Regime variance testing—A quantile approach. *Acta Physica Polonica B*, *44*(5), 1015–1035.
25. Makowski, R., & Hossa, R. (2014). Automatic speech signal segmentation based on the innovation adaptive filter. *Journal of Applied Mathematics and Computer Science, 24*(2), 259–270.
26. Przylibski, T., Wyłomanska, A., Zimroz, R., & Fijałkowska-Lichwa, L. (2015). Application of spectral decomposition of 222RN activity concentration signal series measured in Niedzwiedzia Cave to identification of mechanism responsible for different time-period variations. *Applied Radiation and Isotopes*, *104*, 74–86.
27. Crossman, J. A., et al. (2003). Automotive signal fault diagnostics—Part I: Signal fault analysis, signal segmentation, feature extraction and quasi-optimal feature selection. *IEEE Transactions on Vehicular Technology, 52*(4), 1063–1075.
28. Cempel, C., & Tabaszewski, M. (2007). Multidimensional condition monitoring of machines in non-stationary operation. *Mechanical Systems and Signal Processing, 21*(3), 1233–1241.
29. Zimroz, R., & Bartkowiak, A. (2013). Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing, 38*(1), 237–247.
30. Bartkowiak, A., & Zimroz, R. (2014). Dimensionality reduction via variables selection - Linear and nonlinear approaches with application to vibration-based condition monitoring of planetary gearbox. *Applied Acoustics, 77,* 169–177.
31. Zimroz, R., Bartelmus, B., Barszcz, T., & Urbanek, J. (2014). Diagnostics of bearings in presence of strong operating conditions non-stationarity—A procedure of load-dependent features processing with application to wind turbine bearings. *Mechanical Systems and Signal Processing, 46*(1), 16–27.

32. Wodecki, J., Stefaniak, P., Obuchowski, J., Wyłomańska, A., & Zimroz, R. (2015). Combination of ICA and time-frequency representations of multichannel vibration data for gearbox fault detection. *Vibroengineering PROCEDIA*, *6*, 133–138.
33. Sawicki, M., Zimroz, R., Wyłomańska, A., Obuchowski, J., Stefaniak, P., & Żak, G. (2015). An automatic procedure for multidimensional temperature signal analysis of a SCADA system with application to belt conveyor components. *Procedia Earth and Planetary Science*, *15*, 781–790. ISSN 1878-5220.

Time-Frequency Identification of Nonlinear Phenomena for Mechanical Systems with Backlash

Rafał Burdzik, Łukasz Konieczny, Jakub Młyńczak, Piotr Kruczek, Jakub Obuchowski and Maciej Zawisza

Abstract The paper presents result of non-linearity analysis in the steering gear with various backlash. Due to mechanism of steering gear clearance of gear teeth usually generate non-linearity of movement transfer. Thus, it can be considered as main source of non-linearity in vibration signal. The main goal of the article is to observe how the gear backlash level influences the vibration signal and nonlinear phenomena. The signal was recorded on automobile steering system model with a special modified steering gear. The experiment consists of vibration measurements in 3 orthogonal directions for steering gear without backlash, 0.5 mm of backlash and with 0.9 mm of backlash. The advanced non-stationary signal processing methods for evaluation of steering gear backlash were applied. The presented methodology allows complex observation of the vibration signal, especially when non-linear processes occur in the system. Therefore, energy and dynamics of

R. Burdzik (✉) · Ł. Konieczny · J. Młyńczak
Faculty of Transport, Silesian University of Technology, Krasińskiego 8,
40-019 Katowice, Poland
e-mail: rafal.burdzik@polsl.pl

Ł. Konieczny
e-mail: lukasz.konieczny@polsl.pl

J. Młyńczak
e-mail: jakub.mlynczak@polsl.pl

P. Kruczek · J. Obuchowski
KGHM CUPRUM Ltd., CBR, Sikorskiego 2-8, 53-659 Wrocław, Poland
e-mail: pkruczek@cuprum.wroc.pl

J. Obuchowski
e-mail: jobuchowski@cuprum.wroc.pl

M. Zawisza
Institute of Machine Design Fundamentals, Warsaw University of Technology,
Narbutta 84, 02-524 Warsaw, Poland
e-mail: mzawisza@simr.pw.edu.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_12

the signal can be observed. Furthermore, it is possible to analyse amplitudes of time-frequency representation for different gear backlashes, which enables evaluation of non-linearity of the signal.

Keywords Steering gear · Backlash · Non-linear vibration

1 Introduction

Drivers and vehicle operators are continuously exposed to vibrational stimulus. Vehicle steering process is one of the most important due to its role in controlling the vehicle and informing the driver of the vehicle's dynamic state. The steering system can be considered as a multiple-element mechanism that converts rotational movement of steering wheel into horizontal movement of driving wheels. It consists of shafts, joints, pinion, and rack. It has to be assumed that the driver must select the steering wheel angle to keep deviation low. However, there is no definite relationship between the turning angle of the steering wheel performed by the driver and the change in driving direction, because their correlation is not linear.

In the typical steering gear there is a reduction of movement between the steering wheel and road wheels. Thus, the effort required to turn the steering wheel is decreased. Recently, the predominant proportion of rack-and-pinion steering systems used as steering gears might be observed in automotive engineering. This applies for hydraulic or electrically servo-assisted steering systems and for non-servo-assisted steering systems. These steering gears generally have a displaceable steering housing with a rack and pinion mounted therein in the longitudinal direction. A steering pinion is rotatable arranged in a pinion housing and engaged in compatible gearing of the rack. The steering pinion in turn is connected to a steering column and a steering-wheel in order to operate the steering system [1, 2].

The paper presents novel approach for nonlinear phenomena observation by using time-frequency representation (TFR) of vibration signals measured on steering gear housing.

2 Steering System as Vibration Transmitter

In the steering system forces and vibration are transferred from road into driver and the other way around. On the other hand, front wheels shocks with high intensity and small amplitude are transmitted through the steering gear to the driver's hands. Therefore, vibration measurements on steering gear can provide a valuable information. Comprehensive investigation might contain analysis of damping of vibration transferred from road roughness into the steering system, through

evaluation of technical condition of steering gear or vibration gear properties identification.

The analysis of perception of rotational hand–arm vibration transferred from steering wheel is one of the most common issues. Indeed, steering system is the vibration source which affects the driver. It is particularly important due to the sensitivity of the hand tactile receptors and lack of intermediate structures which can attenuate vibration (e.g. shoes or clothing). During the driving process, power spectral densities of steering wheel vibrations can reach frequencies of up to 350 Hz [3–5]. Vibrational modes of the steering wheel and column often produce large resonances in the frequency range from 20 to 50 Hz [6]. Steering wheel vibration might cause discomfort and annoyance of the driver. Interestingly, it has been already proved that translational hand–arm vibration might produce similar levels of perceptible discomfort to translational whole-body vibration when its acceleration level is 5–7 times larger [7]. The European Parliament and the Council of the European Union prepared Directive on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (vibration)—Directive 2002/44/EC. The Directive defines 'hand-arm vibration as mechanical vibration that, when transmitted to the human hand-arm system, entails risks to the health and safety of workers, in particular vascular, bone or joint, neurological or muscular disorders'.

The human subjective response to hand-arm vibration caused by the automotive steering wheel depends on the amplitude modulation of the waveform, the frequency bandwidth contained in the steering wheel vibration and the repetition rate of transient events caused by the road surface irregularities [8–11]. In [12] influences of vibration level, direction of vibration, grip force and hand–arm posture on the absorption of translational energy by the hand–arm system are presented. Authors concluded that energy absorption depends mainly on the frequency and direction of vibration. Moreover, sensation curves developed for translational vibration have contributed to the definition of the frequency weighting which is currently used in both International Organization for Standardization and British Standards Institution [13–17]. Frequency weighting was primarily defined in order to measure and report hand–arm exposures for the purpose of quantifying health effects and it is based on simplification and an extrapolation of sensation data [18, 19]. It can be also used in the automotive industry for evaluating steering wheel vibrational discomfort.

The properties and spectral content of steering wheel vibration depends on numerous factors including the direction of vibration [20], the nature of the road surface irregularities [21], the dynamic characteristics of the tires and the design of the vehicle main suspension and steering mechanism. Thus, steering gear vibration properties should also be considered. Moreover, it is very important element responsible movement transfer and rotation and, in consequence, vibration transfer.

3 Problem Definition

Due to construction of steering system the most dangerous situation occurs when the connections between elements are broken or some clearance occurs. There is lack of well-known methods for evaluation of connections state between elements of steering system [22–25]. The gear is one of the most important part of the steering system, therefore in this paper we analyze signals acquired on the steering gear. Furthermore, the steering gear is the most common source of backlash in steering system. Thus, we would like to apply advanced non-stationary signal processing methods for evaluation of steering gear backlash level [26–28].

4 Test Rig and Experiment Description

In this paper we provide the analysis of non-linear steering gear vibration process for various gear backlash. Intuitively, once the investigated object and exciter mechanism do not change in time, vibration response of the object should be almost constant. Thus, technical state of the steering system model was identified and all adjunctive parameters were constant during the experiment. Additionally, kinematic inductor used as vibration exciter was generating the same programmed induction cycle. The only varying parameter was gear backlash. Due to the mechanism of steering gear function clearance of gear teeth generate non-linearity of movement transfer. Hence, it can be considered as main source of non-linearity in the vibration signal [29, 30].

Backlash can be treated as clearance or lost motion in a mechanism. It is quantified by the maximum distance or angle through which any part of analyzed system may be moved in one direction. Thus, gear backlash is defined as the amount of clearance between mated gear teeth. Backlash of an ideal system, in theory, should equal to zero. Nevertheless, due to lubrication and some manufacturing errors or even thermal expansion some backlash is used.

The investigated object is a model of automobile steering system with specifically modified steering gear. There is a possibility to adjust the gear backlash size. The test rig contains model of steering system with modified steering gear and kinematic inductors. The studied object is depicted in Fig. 1. Vibration measurements were performed for system without backlash and steering gear with backlash equal to 0.5 and 0.9 mm.

Fig. 1 Test rig and modified steering gear

5 Methodology

In this section the proposed methodology for identification of nonlinear phenomena in system with backlash is described. The algorithms are based on the time frequency representation of the signals. Data acquired during the experiment consists of signal from three different axes and three different backlash levels. Furthermore, the measurements are not equal in phase. Thus, in the first step the signals with different backlash should be synchronized. This procedure is performed using cross-correlation, which for parameter m and signal x is given by formula:

$$R_{x,y}(m) = E\left(x_{n+m} y_n^*\right). \tag{1}$$

We are looking for such parameter m which maximizes the cross-correlation value. In our case we compare the signals with backlash with the reference signal without backlash. The proposed shift for the signal backlash 0.5 $(sh05)$ and 0.9 $(sh09)$ are:

$$\begin{aligned} sh05 &= \max_{\mathrm{m}}\{R_{x0,x05}(m)\}, \\ sh09 &= \max_{\mathrm{m}}\{R_{x0,x09}(m)\}, \end{aligned} \tag{2}$$

where $x0, x05, x09$ are signal without backlash, with backlash 0.5 and 0.9, respectively. Moreover, the length of analyzed signals is not equal, thus they have to be truncated. Once the signals are synchronized and have the same length we can proceed the analysis. Firstly, the data is decomposed into time-frequency representation. Then the absolute value of short-time Fourier transforms are calculated:

$$\mathrm{S(t,f)} = |STFT(t,f)| = \left| \sum_{k=0}^{N-1} X_k w(t-k) e^{-\frac{2i\pi fk}{N}} \right|, \quad t \in T, f \in F, \tag{3}$$

where t is a time point, f is a frequency, x_t input signal of length N and $w(t-k)$ is a shifted window function. Then the time-frequency representation of the signal can be compared. In order to perform the comparison, the new map is calculated. For each point from the spectrograms with backlash ($S_{05}(\cdot,\cdot)$, $S_{09}(\cdot,\cdot)$) the new value is calculated:

$$\begin{aligned} SD_{05}(t,f) &= \frac{S_{05}(t,f) - S_0(t,f)}{S_0(t,f)}, \\ SD_{09}(t,f) &= \frac{S_{09}(t,f) - S_0(t,f)}{S_0(t,f)}, \end{aligned} \tag{4}$$

where $S_0(\cdot,\cdot)$ is the spectrogram from the signal without the backlash. Such maps provide an information how the backlash influences the energy in given frequency bands. Finally, we are able to integrate this map with respect to time or frequency. As a result one dimensional function are obtained:

$$\begin{aligned} ITSD(t) &= \frac{1}{n}\sum_f SD(t,f), \\ IFSD(f) &= \frac{1}{n}\sum_t SD(t,f). \end{aligned}$$

These functions inform the reader for which frequency bands and time points the impact of backlash to the result is the greatest. It enables to identify nonlinear phenomena in the system with backlash. It is expected that values of *ITSD* and *IFSD* are influenced by system's backlash level.

6 Real Data Application

This section contains the results of proposed algorithm application to the real data acquired during the experiment. Firstly, let us present recorded signals with three backlash levels (0, 0.5, 0.9). In Fig. 2 waveforms for different backlash level from the x axis are presented. Signals are normalized, hence their amplitudes are between [−1, 1]. Moreover, the signal lengths vary with respect to backlash. One can observe that signals are inconsistent in phase. Thus, in the first step we would like to synchronize them.

Such procedure can be performed from z axis waveforms. The cycles are easily observed in these plots (Fig. 3). It gives a possibility to equalize signals in phase.

The synchronization is performed with cross correlation (Fig. 4). We are looking for the shift lag for which the highest value of cross correlation is obtained. Due to the cyclic behavior of the signal cycles in cross correlation are detected. One can

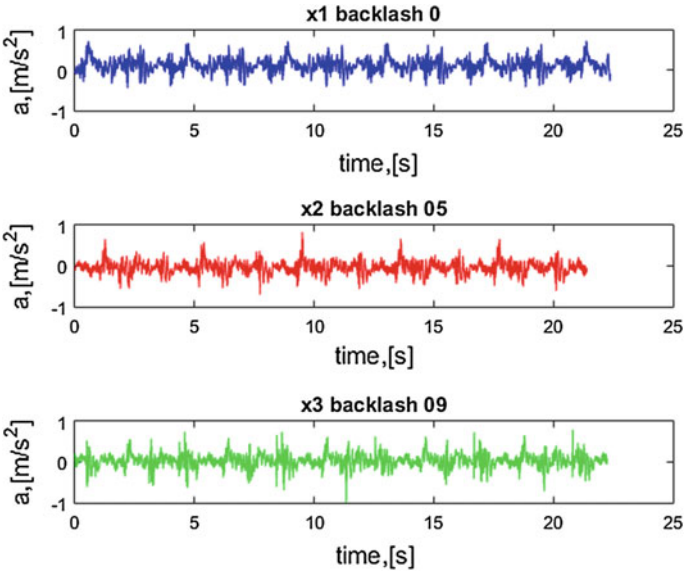


Fig. 2 Waveforms for different backlash level for x axis

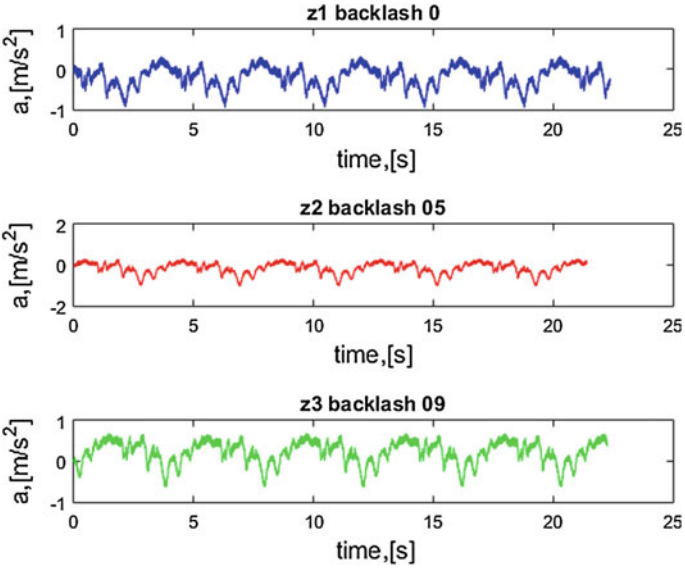


Fig. 3 Waveforms for different backlash level for z axis

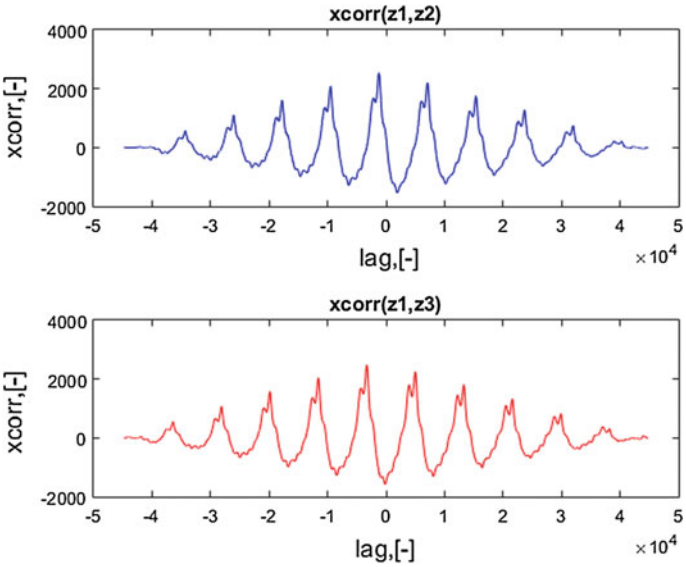


Fig. 4 Cross correlation for z axis signals. Backlash 0 is the reference signal. Backlash signal 05 has to shifted 0.4864 s to the *left*. Backlash signal 09 has to shifted 1.3176 s to the *left*

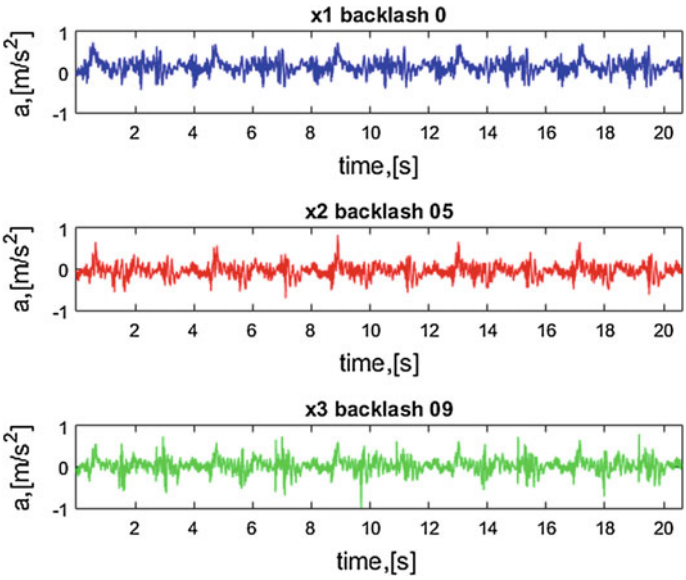


Fig. 5 Waveforms for different backlash level for synchronized x axis

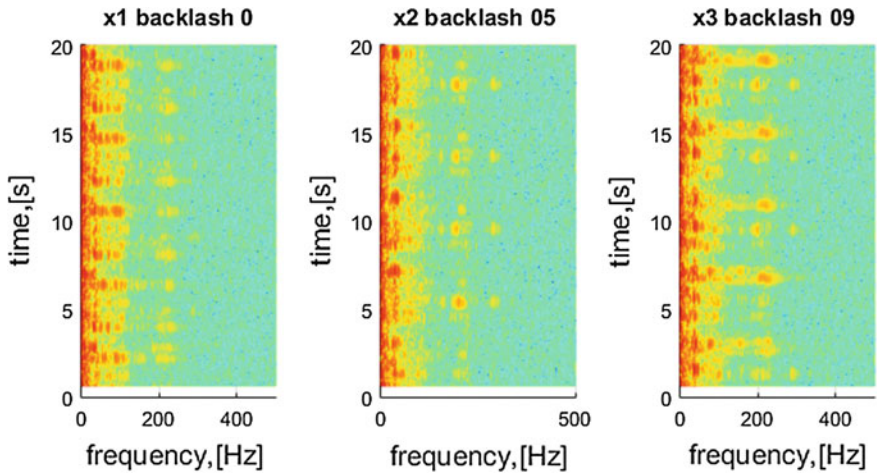


Fig. 6 Spectrograms for different backlash for x axis. Spectrogram parameters are: 2500 length Hamming window with 96% overlapping and 4096 FFT points. The energy amplitude is in the interval [−84, 50]

observe that both backlash signals 05 and 09 should be slightly shifted, 0.4864 and 1.3176 s to the left, respectively.

The output from cross correlation gives a possibility to analyze signals consistent in phase. The sensor is three-axial, it means that the measurements for all axis are recorded simultaneously. Thus, the same shift can be applied for each axis. In Fig. 5 the synchronized time plots for x axis are presented. One can observe that the phase is almost equal among the signals, thus such signals can be compared.

In order to detect for which carrier frequencies the excitations reveal, let us present signals in time-frequency domain. In Fig. 6 the spectrogram for signals from x axis are presented. The highest energy concentrates for low energy frequency bands [0 Hz, 50 Hz]. On the other hand, for frequencies [50 Hz, 300 Hz] the differences with respect to the backlash can be observed. Once the backlash is obtained in the system the periodic impulses reveal. Moreover, the higher backlash results in higher energy of such impulses.

In the next step the differences between the reference value and the spectrogram for signals with backlash are computed (Fig. 7). Clearly, nonlinearity in the energy pattern changes can be observed. In case of the system with backlash equal to 0.5 some impulses reveal in frequency band [200−300 Hz]. Furthermore, the spectrogram of system with backlash 0.9 contain the cyclic excitation with higher amplitude.

Finally, the quantification can be obtained from the integrated $SD(t,f)$ differences map. The results for integration with respect to time (Fig. 8) and frequency (Fig. 9) are attached. Intuitively, bigger differences are obtained for higher backlash

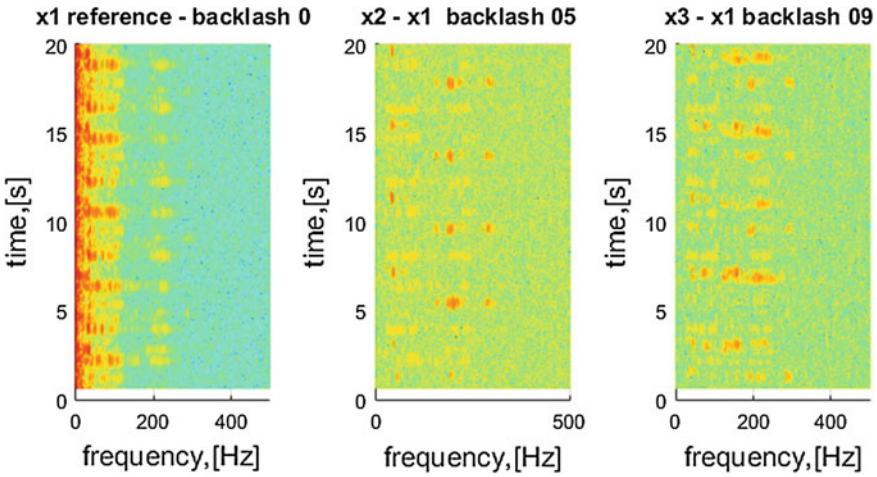


Fig. 7 Differences between spectrogram and reference spectrogram for backlash 0. *Left panel* reference spectrogram for system without backlash, *middle panel* SD_{05} map and *right panel* SD_{09} map

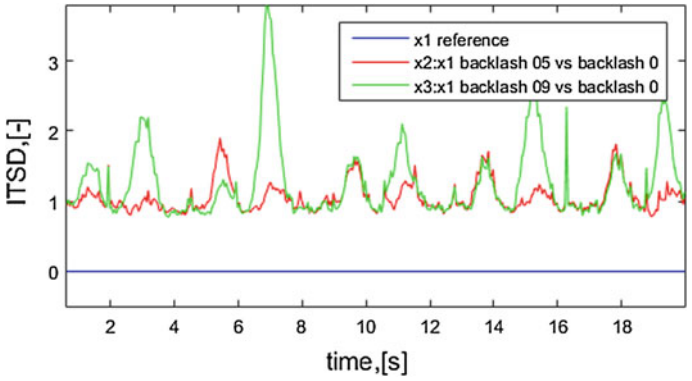


Fig. 8 The percentage difference between the reference signal without backlash and signals with backlash in time domain

level. One can observe that peaks are detected for some time points and the level of noise is similar to system with both backlash levels.

Moreover, according to Fig. 9 signals are not similar only for frequencies lower than 300 Hz. Furthermore, the influence of the nonlinear components is the biggest for frequencies around 200 Hz. Indeed, the amplitudes from integrated maps have the highest values, thus system with backlash differs significantly from the reference one. Results ensure that the backlash influence the vibration signals and give a possibility to distinguish signals with different backlash.

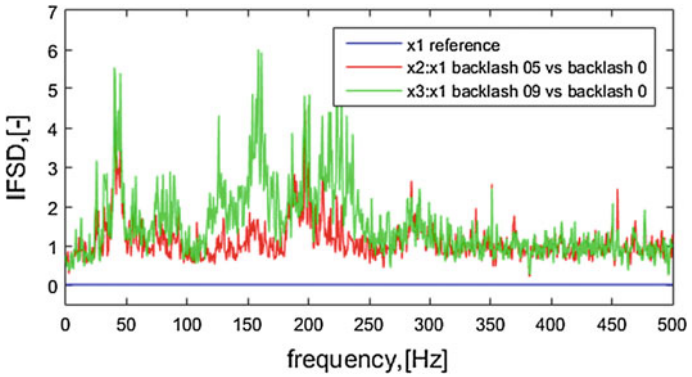


Fig. 9 The percentage difference between the reference signal without backlash and signals with backlash for frequency domain

7 Conclusions

Presented set of signals allows complex observation of the vibration pattern, especially when we consider non-linear processes that occur in the system. According to the results, the increase of backlash influences vibration of steering gear housing. Furthermore, it is not sufficient to compare the waveforms RMS (Root Mean Square) [1]. Analysis of frequency or time-frequency representation of the signal allows to identify high energy components which reveal for system with backlash. It was observed that nonlinear phenomena cause low energy changes in the signal, but after application of the proposed signal processing method it is clearly visible especially for the frequency range 100−300 Hz.

Presented methodology contains the synchronization process performed with cross correlation. In result it allows observation of non-linearity increase based on differences between spectrogram and reference spectrogram for backlash 0. Such time-frequency maps provide an information of backlash influence on energy in given frequency bands. The maps can be integrated with respect to time or frequency as one dimensional function, which are easier to analyze. It can be observed for which frequency bands or time points the influence of backlash on vibration level is the highest.

References

1. Wandor, M., & Burdzik, R. (2015). Research on vibration of steering gear of automotive vehicle. *Vibroengineering Procedia, 6,* 264–267.
2. Pak, C. H., Lee, U. S., Hong, S. C., Song, S. K., Kim, J. H., & Kim, K. S. (1991). A study on the tangential vibration of the steering wheel of passenger car. *SAE Paper, 912565,* 961–968.
3. Gescheider, G. A., Bolanowski, S. J., & Verrillo, R. T. (2004). Some characteristics of tactile channels. *Behavioural Brain Research, 148,* 35–40.

4. Fujikawa, K. (1998). Analysis of steering column vibration. *Motion and Control, 4,* 37–41.
5. Ajovalasit, M., & Giacomin, J. (2009). Non-linear dependency of the subjective perceived intensity of steering wheel rotational vibration. *International Journal of Industrial Ergonomics, 39,* 58–67.
6. Pottinger, M. G., Marshall, K. D., Lawther, J. M., & Thrasher, D. B. (1986). A review of tire/pavement interaction induced noise and vibration. In M. G. Pottinger & T. J. Yager (Eds.), *ASTM STP929 the tire pavement interface* (pp. 183–287). Philadelphia: ASTM.
7. Peruzzetto, P., (1998). Assessing the relative importance of hand vibration with respect to whole-body vibration. The United Kingdom and French Joint Meeting on Human Responses to Vibration, I.N.R.S., Vandoeuvre, France, 26–28 September, 1988.
8. Ajovalasit, M., & Giacomin, J. (2005). Human subjective response to steering wheel vibration caused by diesel engine idle. *Proceedings of the Institution of Mechanical Engineers, Part D, 219,* 499–510.
9. Woo, Y., & Giacomin, J. (2006). The role of the scale and the frequency bandwidth of steering wheel vibration on road surface recognition. In *8th International Symposium on Advance Vehicle Control, AVEC'06*, Taipei, Taiwan.
10. Giacomin, J., Shayaa, M. S., Dormegnie, E., & Richard, L. (2004). Frequency weighting for the evaluation of steering wheel rotational vibration. *International Journal of Industrial Ergonomics, 33,* 527–541.
11. Morioka, M., & Griffin, M. J. (2006). Magnitude-dependence of equivalent comfort contours for fore-and-aft, lateral and vertical hand-transmitted vibration. *Journal of Sound and Vibration, 295,* 633–648.
12. Bustrom, L., & Lundstrom, R. (1994). Absorption of vibration energy in the hand and arm. *Ergonomics, 37*(5), 879–890.
13. The European Parliament and the Council of the European Union Directive on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (vibration)—Directive 2002/44/EC.
14. British Standards Institution. Measurement and evaluation of human exposure to vibration transmitted to the hand. British Standard. London: BSI, 1987:6842.
15. International Organization for Standardization. (1986). *Mechanical vibration—Guidelines for the measurement and the assessment of human exposure to hand-transmitted vibration* (p. 5349). Geneva: ISO.
16. International Organization for Standardization. (2001). Mechanical vibration—measurement and evaluation of human exposure to hand-transmitted vibration—Part 1: General requirements (pp. 5349–1(E)). Geneva: ISO.
17. International Organization for Standardization (2002). Mechanical vibration—measurement and evaluation of human exposure to hand-transmitted vibration—Part 2: Practical guidance for measurement at the workplace (5349–2(E)). ISO.
18. Griffin, M. J. (1990). *Handbook of human vibration*. London: Academic Press.
19. Griffin, M. J. (1997). Measurement, evaluation, and assessment of occupational exposures to hand-transmitted vibration. *Occupation and Environmental Medicine, 54,* 73–89.
20. Burdzik, R. (2013). Research on structure and directional distribution of vibration generated by engine in the location where vibrations penetrate the human organism. *Diagnostyka, 14*(2), 57–61.
21. Burdzik, R. (2014). Identification of structure and directional distribution of vibration transferred to car-body from road roughness. *Journal of Vibroengineering, 16*(1), 324–333.
22. Korbiel, T., & Blaut, J. (2014). Evaluating the effectiveness of mechanical transmission based on the assessment of energy dissipation. *Diagnostyka, 15*(4), 15–20.
23. Jedlinski, L., Caban, J., Krzywonos, L., et al. (2015). Application of vibration signal in the diagnosis of IC engine valve clearance. *Journal of Vibroengineering, 17*(1), 175–187.
24. Budzik, G., & Dziubek, T. (2015). Methodology of measurement aeronautical bevel gears utilizing an optical 3D scanner using white and blue light. *Diagnostyka, 16*(1), 51–56.

25. Popiołek, K., & Pawlik, P. (2016). Diagnosing the technical condition of planetary gearbox using the artificial neural network based on analysis of non-stationary signals. *Diagnostyka, 17*(2), 57–64.
26. Zimroz, R., & Bartkowiak, A. (2013). Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing, 38*(1), 237–247.
27. Obuchowski, J., Zimroz, R., & Wylomanska, A. (2015). Identification of cyclic components in presence of non-Gaussian noise application to crusher bearings damage detection. *Journal of Vibroengineering, 17*(3), 1242–1252.
28. Wyłomańska, A., Obuchowski, J., Zimroz, R., & Hurd, H. (2014). Periodic autoregressive modeling of vibration time series from planetary gearbox used in bucket wheel excavator. In *Cyclostationarity. Theory and Methods Lecture Notes in Mechanical Engineering* (pp. 171−186).
29. Dąbrowski, Z., & Zawisza, M. (2015). The choice of vibroacoustic signal measures, in mechanical fault diagnosis of diesel engines. *Solid State Phenomena, 236,* 220–227.
30. Grzeczka, G., Listewnik, K., Kłaczyński, M., & Cioch, W. (2015). Examination of the vibroacoustic characteristics of 6 kW proton exchange membrane fuel cell. *Journal of Vibroengineering, 17*(7), 4025–4034.

Vibration Signal Processing for Identification of Relation Between Liquid Volume and Damping Properties in Hydraulic Dampers

Łukasz Konieczny, Rafał Burdzik, Jakub Młyńczak, Jakub Obuchowski, Piotr Kruczek and Dariusz Laskowski

Abstract The paper presents analysis of vibration transferred on the car body under different damping properties of suspension. Vehicle must be considered as a multi-technical system in which non-linear phenomena occur. The springing elements in modern cars are coil springs with non-linear characteristics, and the damping elements are telescopic shock absorbers with asymmetrical, strongly non-linear damping characteristics. The scope of the research contains relation between liquid volume and damping properties in hydraulic dampers and analysis of vibration of vehicle forced to vibration by an exciter machine. For comprehensive analysis of the vibration signals their representation in time, frequency and joint time-frequency domains are computed. The paper investigates several different liquid volumes, i.e. from 0% to 100%. The methodology is based on advanced signal analysis designed to determine the influence of fluid volume on the damper characteristics.

Ł. Konieczny · R. Burdzik (✉) · J. Młyńczak
Faculty of Transport, Silesian University of Technology,
Krasińskiego 8, 40-019 Katowice, Poland
e-mail: rafal.burdzik@polsl.pl

Ł. Konieczny
e-mail: lukasz.konieczny@polsl.pl

J. Młyńczak
e-mail: jakub.mlynczak@polsl.pl

J. Obuchowski · P. Kruczek
CBR, KGHM CUPRUM Ltd., Sikorskiego 2-8, 53-659 Wroclaw, Poland
e-mail: jobuchowski@cuprum.wroc.pl

P. Kruczek
e-mail: pkruczek@cuprum.wroc.pl

D. Laskowski
Faculty of Electronics, Military University of Technology,
gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland
e-mail: dariusz.laskowski@wat.edu.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_13

Keywords Shock absorber · Damping properties · Non-linear vibration · STFT

1 Introduction

Vibration phenomena occurring in vehicles are very relevant for safety, comfort and reliability issues. It is worth mentioning that vehicle vibration can increase the fuel consumption, accelerate the elements wear and might even cause vibration of roadside buildings. The evaluation of comfort is a very challenging task because of subjective human perception of vibration. The balance between comfort and safe driving is very difficult to achieve. For the driving safety it is extremely important to provide continuous contact of vehicle wheels with the road surface. It requires a relatively high damping coefficient. On the other hand, comfort of the passengers requires minimization of the vibration perception. It can be achieved by the gradual and smooth vibration absorbing [1–4]. The driver and passengers' exposure to whole-body vibration of the vehicle can cause short-term body discomfort, as well as long-term physiological damage. The vibrations of the vehicle body are the main problem in ride comfort [5–7]. Thus, the analysis of complex vibration occurring in vehicle should be conducted.

Vehicle must be considered as a multi-technical system with non-linear characteristics. The maximum response amplitude depends on damping properties of the system. These properties can be described as inverse proportion between damping coefficient and vibration amplitude arise. Within low damping, there is a sudden increase in amplitude as the excitation frequency approaches the natural frequency. While the damping level is higher, the increase to the maximum amplitude is much more gradual. Consequently, one of the major reasons for vibration analysis is to predict when these types of resonances may occur and determine what should be done to avoid them. The magnitude can be reduced if the natural frequency is shifted away from the excitation frequency by changing the stiffness or mass of the system. This is why the choice of a correct analytical tool for observation of the relation between fluid level and damping characteristics is so important. The data recorded during the experiment and analyzed in this paper belongs to a group of non-stationary signals. Thus, time-frequency methods should be applied in order to analyze them. Such signals are described as time-varying because their frequency content varies with respect to time. The signal processing methods (e.g. short-time Fourier transform) allow for information extraction concerning the time-frequency structure of the signal and comprehensive relation identification between fluid level and the damping properties.

The paper presents new signal processing approach for identification of hydraulic damper properties. The instantaneous frequency (IF) provided by Hilbert transform was calculated and tracked in further analysis. An estimator proposed in this study is the average ratio of harmonics of order higher than 1.

2 Experiment Description

Our main goal was to analyze vibration transferred into car-body under different damping properties of suspension. Thus, the series of experiments were conducted. The research can be divided into two main parts. The former one is focused on damping properties of shock absorber and the latter one is dedicated to vibration of vehicle forced to vibration by exciter machine i.e. a wheel stroke shaker. It allows to force vibration with continuous frequency regulation in predefined range, controlled by an inverter.

Shock absorbers are main elements of the vehicle responsible for vibration absorption and converts vibrations into thermal energy. Properties of shock absorbers determine the transfer of vibrations to the car-body, pressure forces of wheels and traction coefficient of tires. Their asymmetrical, strongly non-linear damping characteristics determine rebound and compression areas of wheels pressure forces, and vibration transferred to car-body [8] (Fig. 1).

In our experiment the damping properties of the shock absorber were changed and determined during indicatory tests. As the main component related to damping factor of hydraulic shock absorbers the volume of liquid flowing by the valves was assumed. Moreover, the phenomenon of liquid leakage is one of the most common defects and influences the result of shock absorbers. It is similar to car technical state diagnostic examination. Liquid leakage might be simulated by changing its volume in shock-absorber. This modification directly affects damping factor. The scope of research includes liquid level starting from empty shock absorber and ending with 100% liquid level with 20% gradation steps.

The second part of the study is related to vibration test performed on encased shock absorbers with different damping properties in the vehicle. The car was set in motion and tested with harmonic forced vibration methods. Therefore, it can be considered as a mechanical system generating a vibration signal, which is a non-stationary random process. The concept of the conducted analysis is presented in Fig. 2.

The accelerations of the platform, sprung and unsprung mass were recorded. The signals were acquired in digital form with 2500 Hz sampling frequency. The acceleration sensors were placed on the test platform (Fig. 3), unsprung mass

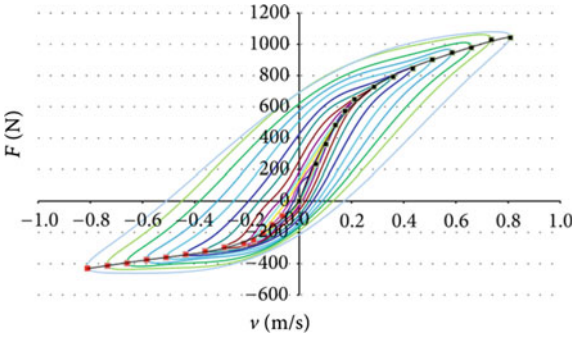


Fig. 1 Working velocity loops of shock absorber and determination of damping characteristics (non-linear function—*black line*)

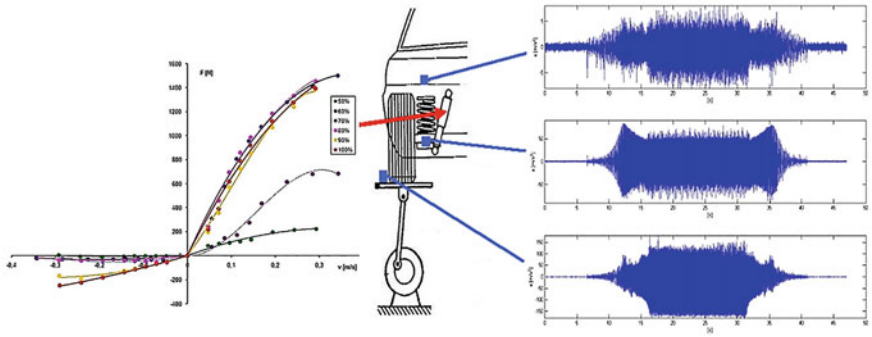


Fig. 2 The illustrated concept of the conducted research

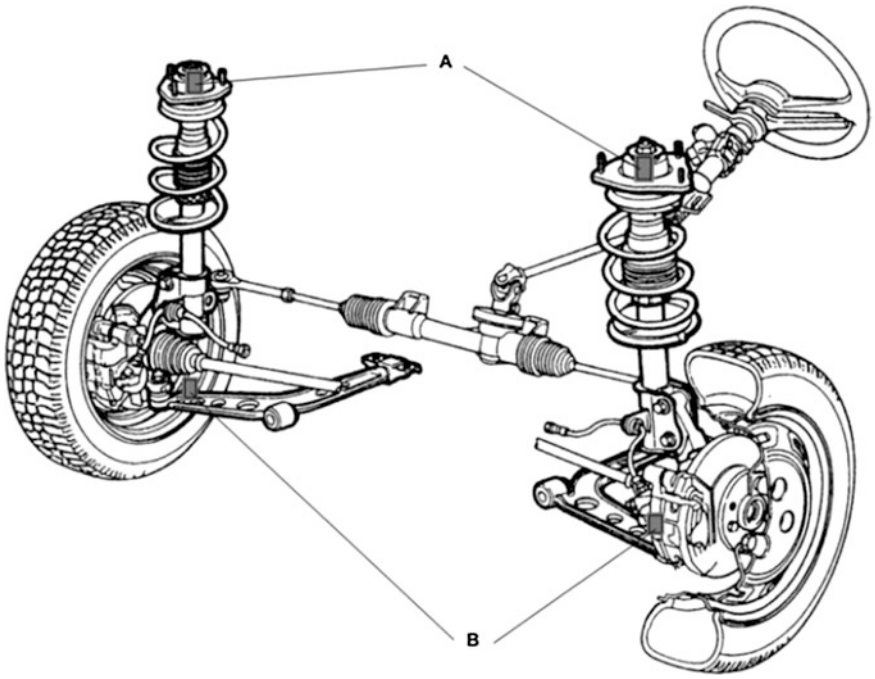


Fig. 3 Points of sensors mounted: *A*—on sprung mass, *B*—on unsprung mass

(on the arm near the wheel) and sprung mass (on the point of upper McPherson strut mount) of vehicle.

For the vibration measurements the accelerometers were used. An accelerometer is a piezo-electric device that is frequently used to monitor vibration. The reason to choose such sensors is related to some problems which encountered in estimating the errors of laser interferometer measurement of resonance frequencies and the quality factor from the resonance curves of vibration amplitudes [9]. The scheme of

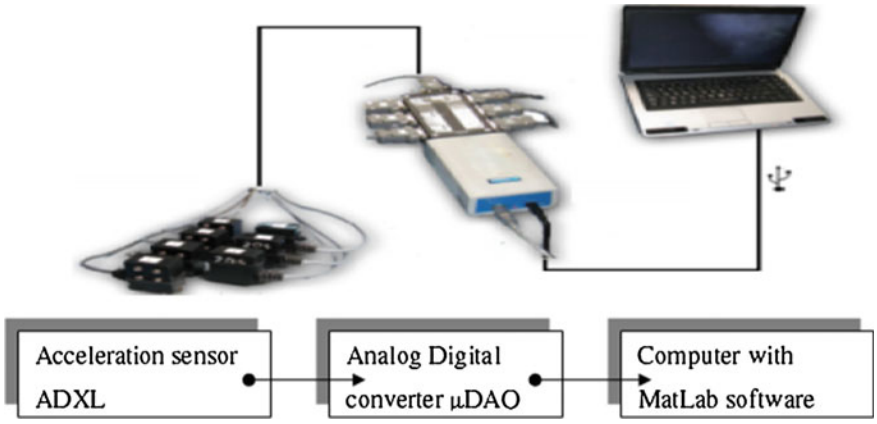
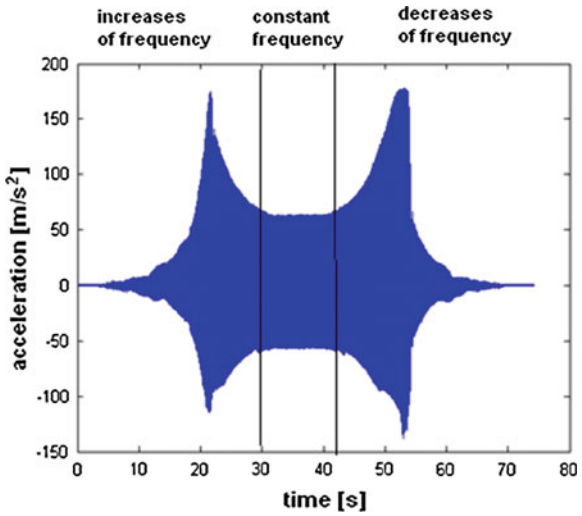


Fig. 4 Scheme of the measurement system

Fig. 5 Example time realization of unsprung mass accelerations



measurement system is depicted in Fig. 4. The acceleration of vibration were measured by ADXL 204 and ADXL 321 sensors. Those are modern parametric sensors built in chip. For the data acquisition the analog-digital card was used. The μDAQ 30 adapted for use with PC computers support the USB interface. The main technical parameters of the μDAQ 30 are 14-bit resolution, max 250 kHz sampling.

The main source of vehicle vibration is excitation of rough road. This source should be considered as a non-stationary random vibration process when traveling at variable speeds [10, 11]. Thus, force excitation process during conducted research consists of three phases: increase of frequency, vibration with constant frequency and decrease of frequency (Fig. 5).

In consequence, the vehicle tested on a vibration exciter can be regarded as a non-linear and non-stationary mechanical system. To investigate a system of this type, the methods of non-stationary random signal analysis can be useful [12–22].

3 Problem Definition

The main problem of our studies is to determine a simulator of fluid leakage in a hydraulic damper. Such problem is of high importance, because fluid is the medium of force transmission and the volume of the fluid determines the damping characteristics (Fig. 1). The vibrations occurring in the vehicle during the test are non-stationary signals. Moreover, the elements of the suspension system (shock absorber, spring elements, bushing, etc.) are subject to considerable dynamic loads. Complex suspension system is a strongly non-linear system. Thus, for the proper observation of the vibration signal a time-frequency transformation has to be applied. The paper presents novel methodology of advanced signal processing application as a multiple tool technique for identification of relation between liquid volume and damping properties in hydraulic dampers.

4 Methodology

In this section we present the methodology allowing for analysis of the relation between input and output in a hydraulic damper system. We propose to analyze this relation using time frequency representation of the vibration signals acquired during the experiment. Due to time-varying frequency characteristics of the input provided to the system the output signal is expected to be nonstationary, thus analysis in time-frequency domain might provide viable information about phenomena occurring in the system. Upon analysis of the entire data acquired during the experiment, the vibration signals are segmented and further processed. Finally, two features are proposed in order to quantify relation between liquid volume and dumping.

In the first step we decompose the signals into time-frequency domain using the short-time Fourier transform defined as [23]:

$$\mathrm{STFT}(t,f) = \sum_{k=0}^{N-1} X_k w(t-k) e^{-\frac{2j\pi fk}{N}}, \quad t \in T, \ f \in F, \tag{1}$$

where t—time point, f—frequency, X_t—input signal of length N and $w(t-k)$—shifted window. Further analysis are performed on the squared amplitude of STFT called spectrogram, namely $S[t,f] = |\mathrm{STFT}[t,f]|^2$. We propose to involve relatively long window, since sampling frequency is much higher than dominating frequency

of input. Moreover, accurate identification of dominating frequency and its harmonics is required for calculation of the proposed features quantifying the relation between fluid level and damping characteristics.

In the next step an algorithm for automatic identification of dominating frequency of the input signal is proposed. Given spectrogram $S[t,f]$ the frequency D dominating in the signal in time t might be specified as:

$$D(t) = \arg\max_{f \in F} S[t,f] \tag{2}$$

where F is the set of frequency bins in which the STFT is calculated. We compare the frequency indicated by $D(t)$ with the instantaneous frequency (IF) provided by the Hilbert transform. In order to obtain IF using the Hilbert transform the considered signal is required to be a monocomponent. Therefore, we apply a narrowband bandpass filter on the acquired signals. Firstly, we compute the discrete Hilbert transform of the signal X_k [24]:

$$H(k) = \sum_{p=0}^{N-1} \left(h(p) \sum_{m=0}^{N-1} X_m e^{-\frac{2j\pi pm}{N}} \right) e^{\frac{2j\pi kp}{N}},\ k = 1, \ldots, N, \tag{3}$$

where $h(p)$ is defined as:

$$h(p) = \begin{cases} 1 & for \quad p = 1, \frac{n}{2} + 1 \\ 2 & for \quad p = 2, \ldots, \frac{n}{2} \\ 0 & for \quad p = \frac{n}{2} + 2, \ldots, n. \end{cases}$$

In other words, $H(k)$ is computed as inverse DFT of an element-wise product of DFT of the signal and $h(p)$.

Next, the analytic signal is designed as

$$A(k) = X(k) + jH(k) \tag{4}$$

where j is the imaginary unit. Finally, the phase of $A(k)$ is unwrapped and normalized:

$$\mathrm{IF}(k) = \frac{f_s}{2\pi} (\mathrm{unwrap}(\arg(A(k))) - \mathrm{unwrap}(\arg(A(k-1)))), \tag{5}$$

where f_s is the sampling frequency, $\arg(z)$ is the argument (phase) of the complex number z and $\mathrm{unwrap}(r)$ is the unwrapping operator that minimizes the incremental angle variation by constraining it to the range $[-\pi, \pi]$.

In this study we analyze the vibration signals on the time interval T_c during which the dominating frequency was nearly constant. Visual inspection of both D and IF as functions of time allows to determine this period, although some automatic algorithms might be designed e.g. to find the longest period with a priori defined maximum deviation of dominating frequency. The next step of the

methodology for quantification of the relation between input and output in a hydraulic damper system consists of designing two features describing the input-output relation in context of damping system properties. The first feature f_1 we define as

$$f_1 = \frac{\sum_{t \in T_c} S_n[D(t), t]}{\sum_{t \in T_c} S_p[D(t), t]}. \tag{5}$$

In other words, the first feature is the ratio of average amplitude of the dominating frequency related to the signal acquired on the vehicle body to average amplitude related to the vibration platform. In general, lower ratio indicates better damping characteristics. One exception of such rule is the case when the wheel loses the contact with the vibration platform. Then the vibration level registered on the vehicle body might be considerably low, but it indicates a dangerous operational mode rather than excellent damping properties. The second feature proposed in this study is an average ratio of harmonics of order higher than 1, i.e. f_2 is defined as follows:

$$f_2 = \frac{\sum_{t \in T_c, k \geq 2} S_n[kD(t), t]}{\sum_{t \in T_c, k \geq 2} S_p[kD(t), t]}. \tag{6}$$

where k is the iterator of dominating frequency harmonics, $k = 1, 2, \ldots$ up to the Nyquist frequency. Therefore, f_2 is focused on harmonics of order higher than 1 which reflects a non-linear character of the damping system. Such features are calculated for each level of liquid volume in the damper that is considered in the experiment. Thus, the relation between the liquid level and damping characteristics might be analyzed using f_1 and f_2.

5 Real Data Application

In this section we apply our algorithm to real data acquired during the experiment. Recall that there are 2 measurement points and several levels of liquid have been investigated, namely 0, 20, 40, 60, 80 and 100% of liquid in the damper. Exemplary waveforms and time-frequency representations of the considered signals are presented in Fig. 6.

One can notice that the vibration signal acquired on the platform in case of empty damper reveals a resonance phenomenon. It appears in 20th and 55th second of the experiment. This phenomenon is manifested as raised amplitude of the signal in time domain. Amplitudes are even higher than these related to maximum frequency of the input. In Figs. 7 and 8 spectrograms of all signals acquired during the experiment are presented. The liquid volume varies from 0% to 100% by 20%; sensors are located on the vibration platform and the vehicle body. It can be noticed

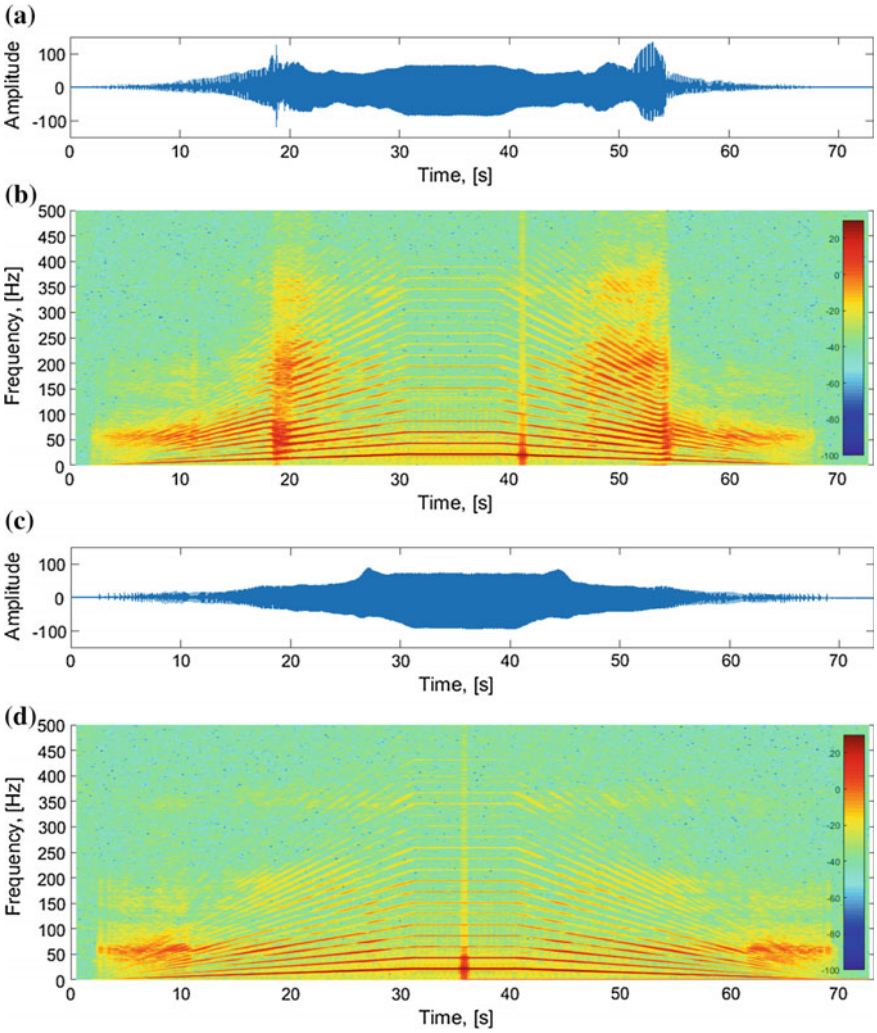


Fig. 6 Waveforms (**a**, **c**) and spectrograms (**b**, **d**) of signals acquired in case of 0% (**a**, **b**) and 100% (**c**, **d**) of liquid in the damper. Sensor located on the vibration platform

that in all cases amplitudes of the excitation signal are comparable. Dominating frequency as a function of time is piecewise linear—it increases up to about 21.5 Hz, then stays on this level for approx. 10 s and finally decreases. The highest amplitude in the excitation signals (acquired on the vibration platform) reveals at the frequency of 21.5 Hz and many harmonics of this component might be noticed. Moreover, there are also sub-harmonics of order 1:2 in almost every signal except the case of sensor located on vehicle body and fluid level of 40%—there are harmonics of order 1:3 therein. The resonance frequency that is manifested mainly

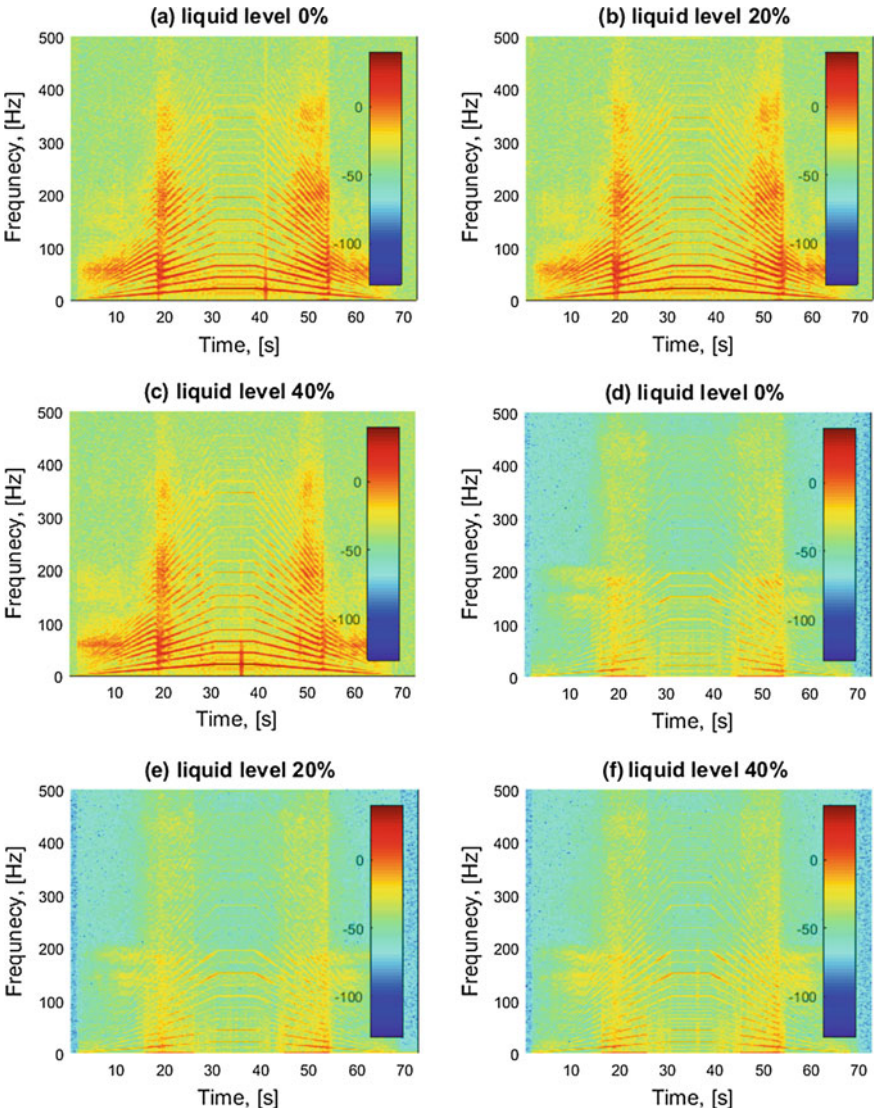


Fig. 7 Spectrograms of signals acquired under liquid volume 0% (**a**, **c**), 20% (**b**, **e**) and 40% (**c**, **f**). Sensor located on the vibration platform (**a**, **b**, **c**) and on the vehicle body (**d**, **e**, **f**)

in Fig. 7a–c is close to 19 Hz. This phenomenon is barely visible in other signals. In the signals acquired on the vehicle body the amplitudes are significantly lower for liquid level 0%, 20% and 40% comparing to amplitudes for 60%, 80% and 100%.

Figure 9 presents results of two different algorithms applied to the signals from the vibration platform acquired under every considered liquid level, from 0% up to 100%. It can be noticed that the Hilbert transform-based method results in some

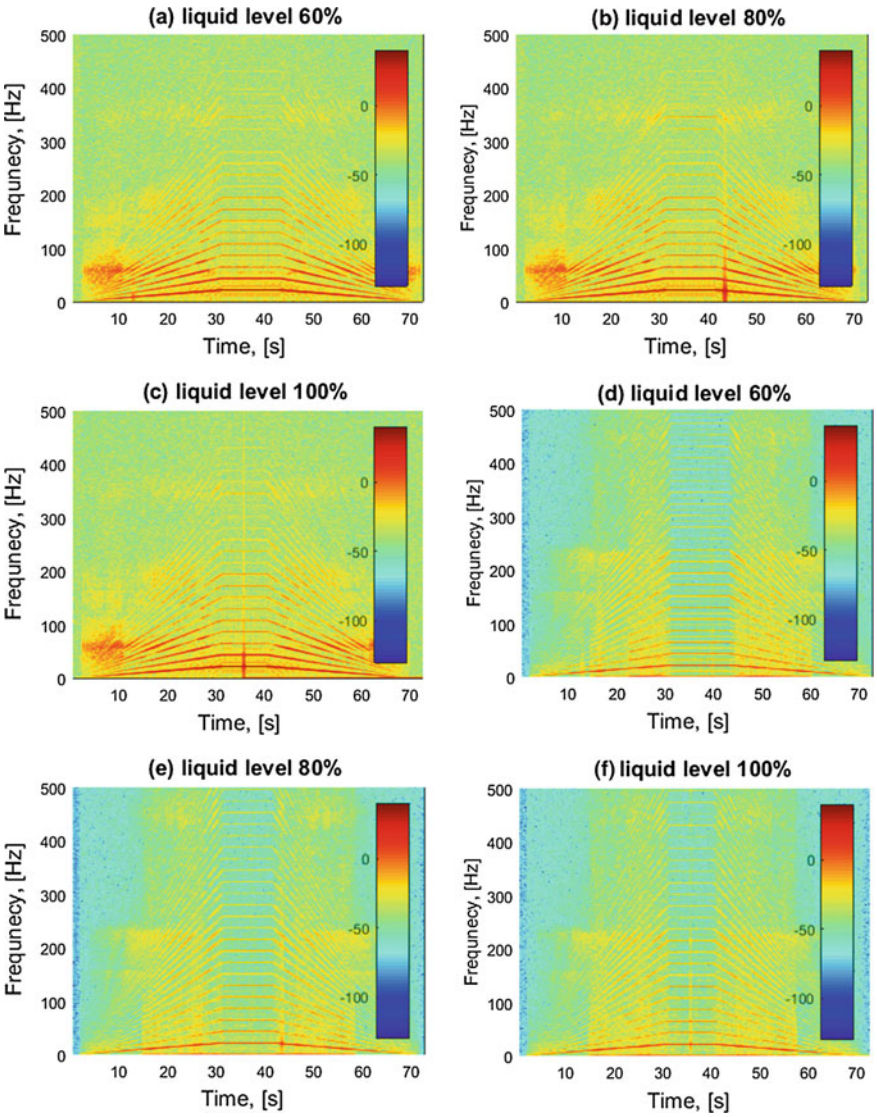


Fig. 8 Spectrograms of signals acquired under liquid volume 60% (**a**, **c**), 80% (**b**, **e**) and 100% (**c**, **f**). Sensor located on the vibration platform (**a**, **b**, **c**) and on the vehicle body (**d**, **e**, **f**)

significant outliers, i.e. instantaneous frequency indicated by this method might be completely different than this obtained by the method proposed in this paper. Although, even the latter one does not indicate properly the frequency set on the vibration platform when it reaches the resonance frequency. Nevertheless, both methods properly indicate the period with nearly constant frequency (21.5 Hz), despite some short-time artifacts (Figs. 7a, c and 8b, c). Duration of these time

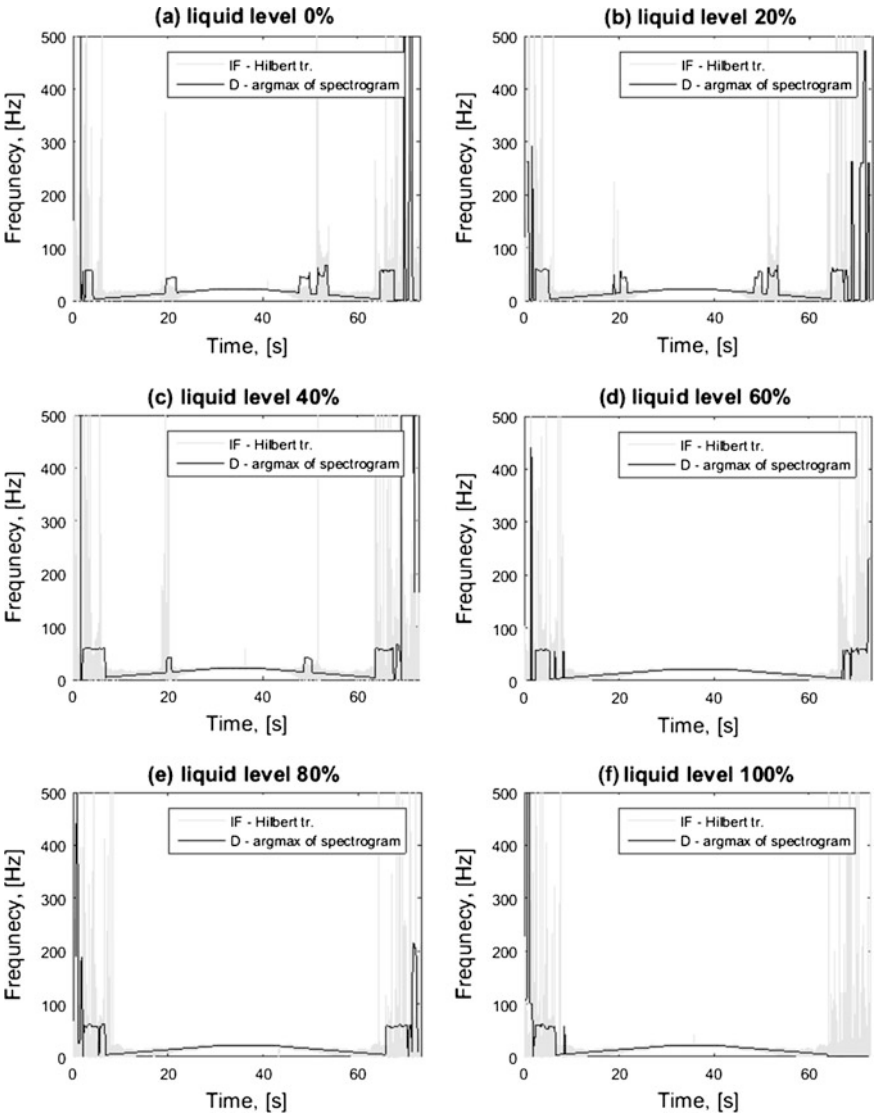


Fig. 9 Comparison of dominating frequency obtained using Hilbert transform and the proposed method involving argmax function on the spectrogram. Sensor located on the vibration platform

intervals is close to 10 s—they begin in 30th second of the experiment and end in 40th. In the next step of the methodology the average spectra have to be calculated. Figure 10 illustrates averaged spectra calculated from the signals acquired during the periods of almost constant dominating frequency. One can easily notice that in case of liquid levels 0%, 20% and 40% the system behaves close to the linear one, since the set of harmonics of the input (vibration platform) is similar to the set of

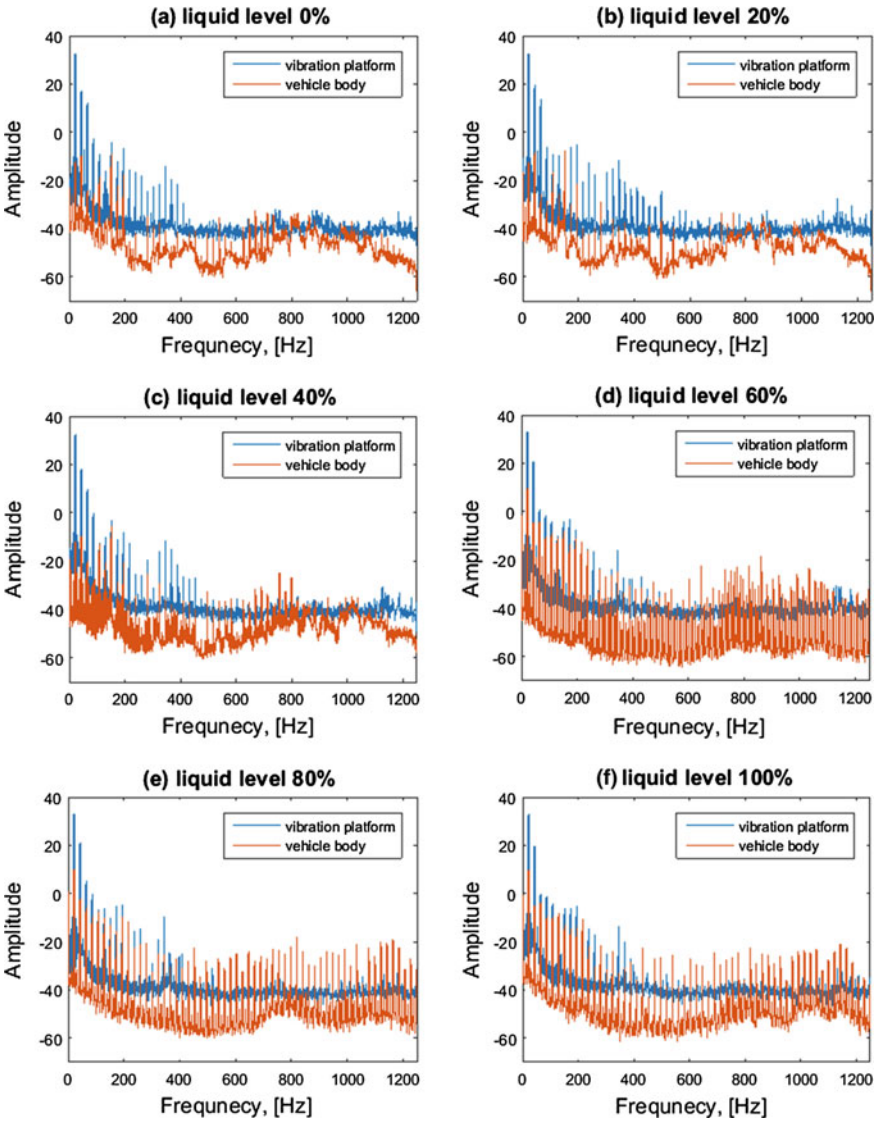


Fig. 10 Averaged spectra calculated during the period of almost constant dominating frequency for each considered liquid level. Sensors located on vibration platform and on the vehicle body

harmonics of the output (vehicle body) signal. On the other hand, the output signals in case of liquid level 60%, 80% and 100% contain of a long train of harmonics, up to the Nyquist frequency. Thus, the system is very non-linear here, since new harmonics are created in the output.

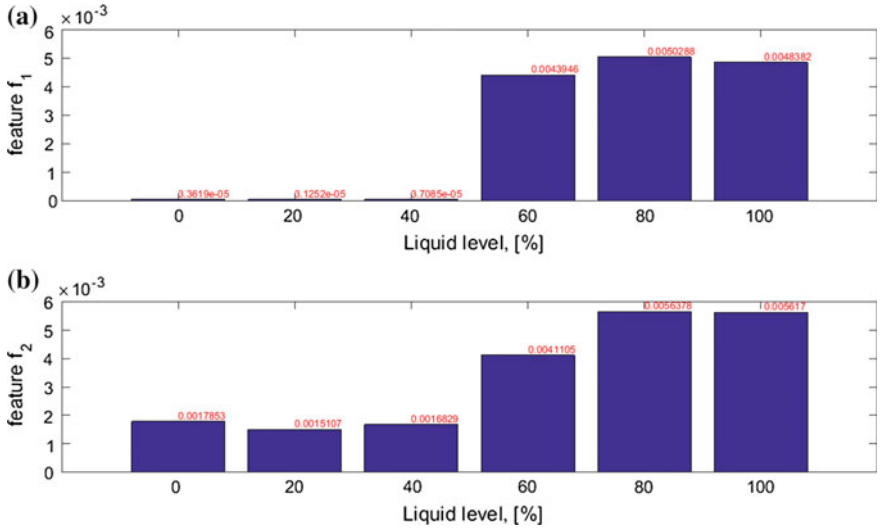


Fig. 11 Values of the features f_1 (**a**) and f_2 (**b**) for each level of liquid

Values of features proposed in this paper are contained in Fig. 11. They prove that the ratio of dominating frequency amplitudes (output to input) is two orders lower in case of liquid level equal to 0%, 20% and 40% while compared to 60%, 80% and 100% (Fig. 11a). The difference between these ratios is lower when higher-order harmonics are considered (Fig. 11b). This indicates that the relative amplitudes of high-order harmonics in the output signal are higher than these in the input. Thus it can be assumed that the higher liquid level result as more non-linear system in such a case.

6 Conclusions

The results of conducted analysis provide interesting conclusions. The proposed methodology allows to determine an estimator that reflects the volume of fluid in the damper. Values of features proposed in this paper are presented in Fig. 11. It proves that the ratio of dominating frequency amplitudes (output to input) is two orders lower in case of liquid level equal to 0%, 20% and 40% while compared to 60%, 80% and 100%. The analysis leads to the conclusion that changes in the volume of fluid in the shock absorber generally have a step change (the fluid volume of less than 60% of the shock absorber in practice stops damping of vibrations). Thus, the higher liquid level, the more non-linear system is. The proposed method is based on advanced transformation of vibration signals and it is effective in the identification of amount of fluid in the shock absorber. Furthermore, studies based on the methodology presented in this article will be continued for the different cases of damage in the vehicle suspension system elements.

The lower level of vibration registered for empty shock absorber is caused by possible loss of wheel-platform contact. Such phenomenon results in decrease of transmission of force into car body by the damper.

References

1. Burdzik, R., & Konieczny, Ł. (2014). Application of vibroacoustic methods for monitoring and control of comfort and safety of passenger cars. *Solid State Phenomena, 210,* 20–25.
2. Burdzik, R., & Konieczny, Ł. (2013). Research on structure, propagation and exposure to general vibration in passenger car for different damping parameters. *Journal of Vibroengineering, 15*(4), 1692–1700.
3. Burdzik, R. (2013). Research on the influence of engine rotational speed to the vibration penetration into the driver via feet—multidimensional analysis. *Journal of Vibroengineering, 15*(4), 2114–2123.
4. Warczek, J. (2014). The model of diagnostic of nonlinear viscoelastomeric damper. *Diagnostyka, 15*(4), 29–34.
5. Griffin, M. J. (2007). Discomfort from feeling vehicle vibration. *Vehicle System Dynamics, 45* (7–8), 679–698.
6. Drozyner, P., Mikolajczak, P., Szuszkiewicz, J., & Jasiulewicz-Kaczmarek, M. (2011). Management standardization versus quality of working life. Ergonomics and health aspects of work with computers. *Lecture Notes in Computer Science, 6779*, 30–39.
7. Konieczny, Ł. (2016). Analysis of simplifications applied in vibration damping modelling for a passive car shock absorber. *Shock and Vibration, 2016*, Article ID 6182847.
8. Konieczny, Ł., Burdzik, R., & Warczek, J. (2013). The uncertainty of determining shock absorber damping characteristic on indicator test stand. *Diagnostyka, 14*(2), 63–66.
9. Atavin, V. G., Kuranov V. V., & Khudyakov, Yu. V. (2003). Determination of the errors of resonance frequency and quality factor measurements from the resonance curves of vibration amplitudes. *Measurement Techniques*, *46*(3), 249–253.
10. He, L., et al. (2008). Non-stationary random vibration analysis of vehicle with fractional damping. In *2008 International Conference on Intelligent Computation Technology and Automation (ICICTA)* (Vol. 2). IEEE.
11. Zhang, L.-J., Lee, C.-M., & Wang, Y. S. (2002). A study on nonstationary random vibration of a vehicle in time and frequency domains. *International Journal of Automotive Technology, 3*(3), 101–109.
12. Boashash, B. (2015). *Time-frequency signal analysis and processing: A comprehensive reference*. Academic Press.
13. Zimroz, R., & Bartkowiak, A. (2013). Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing, 38*(1), 237–247.
14. Jedliński, Ł., Caban, J., Krzywonos, L., Wierzbicki, S., & Brumerčík, F. (2015). Application of vibration signal in the diagnosis of IC engine valve clearance. *Journal of Vibroengineering, 17*(1), 175–187.
15. Dąbrowski, Z., & Zawisza, M. (2015). The choice of vibroacoustic signal measures, in mechanical fault diagnosis of diesel engines. *Solid State Phenomena, 236,* 220–227.
16. Zimroz, R., et al. (2014). Diagnostics of bearings in presence of strong operating conditions non-stationarity—A procedure of load-dependent features processing with application to wind turbine bearings. *Mechanical Systems and Signal Processing, 46*(1), 16–27.
17. Żak, G., et al. (2014). Application of ARMA modelling and alpha-stable distribution for local damage detection in bearings. *Diagnostyka, 15*(3), 3–10.

18. Obuchowski, J., Zimroz, R., & Wylomanska, A. (2015). Identification of cyclic components in presence of non-Gaussian noise application to crusher bearings damage detection. *Journal of Vibroengineering, 17*(3), 1242–1252.
19. Kohut, P., Kurc, K., Szybicki, D., Cioch, W., & Burdzik, R. (2015). Vision-based motion analysis and deflection measurement of a robot's crawler unit. *Journal of Vibroengineering, 17*(8), 4112–4121.
20. Grzeczka, G., Listewnik, K., Kłaczyński, M., & Cioch, W. (2015). Examination of the vibroacoustic characteristics of 6 kW proton exchange membrane fuel cell. *Journal of Vibroengineering, 17*(7), 4025–4034.
21. Rahmi, Guclu. (2005). Fuzzy logic control of seat vibrations of a non-linear full vehicle model. *Nonlinear Dynamics, 40*(1), 21–34.
22. Silvester, B. C. (1966). Vibration reduction in motor cars. *Journal of the Society of Environmental Engineers*, *4*.
23. Boashash, B. (2003). *Time frequency signal analysis and processing: A comprehensive reference*. Oxford: Elsevier.
24. Marple, S. L. (1999). Computing the discrete-time analytic signal via FFT. *IEEE Transactions on Signal Processing, 47*(1999), 2600–2603.

Modified Protrugram Method for Damage Detection in Bearing Operating Under Impulsive Load

Piotr Kruczek and Jakub Obuchowski

Abstract The problem of damage detection in rotating machines is related not only to specific speed or load fluctuations. Sometimes the environment of a gear or bearing makes the diagnosis a challenging problem. In this paper a novel signal processing method based on cyclostationary approach is proposed. It is designed in order to indicate and separate source of the signal that could be related to damage of the machine. The vibration signal related to such source can be considered as a cyclic pulse train. Thus, it is worth to benefit not only from impulsivity of the signal but also from its cyclic nature—taking advantage from these two properties in the same time might be more beneficial. We propose a method that indicates frequency bands where the cyclic pulse train occurs. On the other hand, the method is robust to contamination which might be observed in real-world examples, e.g. non-cyclic high-energy impulses or amplitude modulated sine waves. The motivation for such method is derived from a real-world example—a rolling element bearing operating in a hummer crusher. We compare our method with two other well-known methods for damage detection in rotating machines, namely protrugram and spectral kurtosis.

1 Introduction

Diagnosis of rotating machines is a key part of many industrial mechanical systems. Vibration-based methods play an important role in modern condition monitoring systems [1]. Many signal processing methods that help to understand the vibration signal are available. In general, these methods are often designed to find a series of impulses that arise when force is applied to the faulty area (e.g. inner/outer race, rolling element, gear tooth). The choice of method used to process the vibration signal has to be followed by understanding of machine's operation and the environment

P. Kruczek (✉) · J. Obuchowski
KGHM CUPRUM sp. z o.o. Centrum Badawczo-Rozwojowe,
ul. Gen. Wl. Sikorskiego 2-8, 53-659 Wroclaw, Poland
e-mail: pkruczek@cuprum.wroc.pl

J. Obuchowski
e-mail: jobuchowski@cuprum.wroc.pl

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_14

in which the machine operates. For instance, time-varying speed or/and load have an effect on frequency or/and amplitude modulation of the vibration signal. Thus, it is worth to analyze the signal in domains other than time or frequency i.e. time-frequency domain [2], order domain [3–7], bi-frequency domain [8–12] etc. The time-frequency domain allows for instance to track time-varying speed and load [13–16]. Such information could be beneficial, since distances between impulses related to damage and their amplitudes vary in time, thus it is impossible to notice a specific fault frequency in the signal. Availability of the speed profile allows to analyze the signal in order domain in which the distances between impulses related to damage are approximately constant [17, 18]. Bi-frequency domain illustrates amplitude modulation of the signal with respect to carrier and modulation frequency [9, 19]. Such representation might be used in order to indicate the frequency band where carrier waves are modulated by an impulsive function (a lot of harmonics of the modulation frequency appear therein). This frequency band can also be indicated from the time-frequency or scale-frequency representation assuming that the local damage is revealed by series of wideband excitations, not necessarily equally distributed in time [20–26]. Cyclic impulsive signal can be indicated automatically using a filter bank and an indicator of impulsive pulse train. In [27] the Authors propose a method called "protrugram" that indicates the frequency band for demodulation using kurtosis of envelope spectrum for each narrowband signal obtained from a filter bank. Protrugram is known for its excellent properties: identification of cyclic pulse train, robustness to impulsive noise and high-energy discrete-spectrum components. Although, the main drawback of protrugram occurs for signal components which are modulated by a sine wave or another function with discrete amplitude spectrum [28, 29]. Therefore, kurtosis of envelope of these components could be higher than for cyclic pulse train [30]. Thus, the signal filtered using protrugram could not represent cyclic pulse train even if the considered machine reveals a local damage. This drawback can be avoided using an additional information ("extraction criterion") about construction of the machine [31]. Although, incorporation of such information in the damage detection step could make the diagnostics time-consuming and more complicated. It is more convenient to detect the damage blindly and search for the (eventually) obtained fault frequency in the list of characteristic frequencies. It is worth to notice that vibration signal processing methods based on cyclicity of the pulse train related to damage are often robust to impulsive noise. This is a useful feature in industrial applications where impulsive noise is often related to the environment or to specific operation of the machine [30, 32–34]. In this paper we propose another method that quantifies presence of cyclic pulse train in given frequency band, designs a filter characteristic and, finally, filter the vibration signal. The method is designed to be robust not only to impulsive noise or high-energy discrete-spectrum components—it is also robust to components modulated by a sine wave or other discrete-spectrum function. Such requirements are motivated by analysis of real vibration signal from a copper ore crusher operating in a mineral processing plant. Impulsive noise in vibration signal is related to the crushing process. Series of experiments on different crushers revealed presence of narrowband signal modu-

lated by a sinusoidal signal [33]. Thus, envelope spectrum of the signal demodulated around this frequency band contains a single spike.
The paper is structured as follows. Model of the synthetic signal that follows structure of the real one is described in Sect. 2. Section 3 describes the method that filters the signal in order to extract cyclic pulse train. Results that illustrate effectiveness of the proposed method are presented in Sect. 4. The last section contains conclusions.

2 Model of the Signal

In order to validate the proposed method and illustrate its properties we simulate a signal, which resemble the real one. Length of the signal is 1 s and sampling frequency is 25,000 Hz. The signal consists of the following components: (a) two amplitude modulated sine waves with frequencies 500 and 1000 Hz (4 Hz sine wave as the modulation function), (b) non-cyclic impacts (Gaussian-shaped impulses with center frequency 5000 Hz and random bandwidths between 0 and 12500 Hz), (c) cyclic impulses located in 2 frequency bands—2500–3000 Hz and 8500–9500 Hz, (d) background Gaussian noise. The cyclic impulses were simulated using Matlab function gauspuls with center frequencies $Fc_1 = 3000$ Hz and $Fc_2 = 9000$ Hz and random bandwidth uniformly distributed on intervals [5000–6250 Hz] and [2500–3750 Hz], respectively. Each impulse is replicated in order to obtain cyclicity of the signal of interest (SOI). Cyclic frequency of this pulse train is 30 Hz. The signal is designed in order to follow the real vibration acceleration signal from a hummer crusher. Amplitude modulated sine waves (500 and 1000 Hz) correspond to amplitude modulated harmonics. Non-cyclic impulses might be related to the crushing process—raw material stream enters the crusher and a lot of impulses appear in the vibration signal measured at the bearing housing. These impulses appear at random and their amplitudes are random, as well. Cyclic pulse train corresponds to a local damage. Its cyclic frequency—30 Hz—is related to local damage of the outer race (ball pass frequency outer—BPFO). Two frequency bands where the cyclic impulses occur correspond to two resonance frequency bands. The entire list of the bearing characteristic frequencies might be found in [33]. It is worth to notice that the center frequency of the non-cyclic component is located between two frequency bands where the periodic impulses occur. There are 30 non-cyclic impulses uniformly distributed on the time axis. The time series and a spectrogram of the simulated signal are presented in Fig. 1. We used Kaiser 500-sample length window with 470 overlapping samples and 1024 FFT points.

3 Methodology

The algorithm presented in this paper can be treated as an extension of the protrugram. Recall that the protrugram is a signal processing method that indicates the center frequency and frequency band where kurtosis of the signal's envelope is the

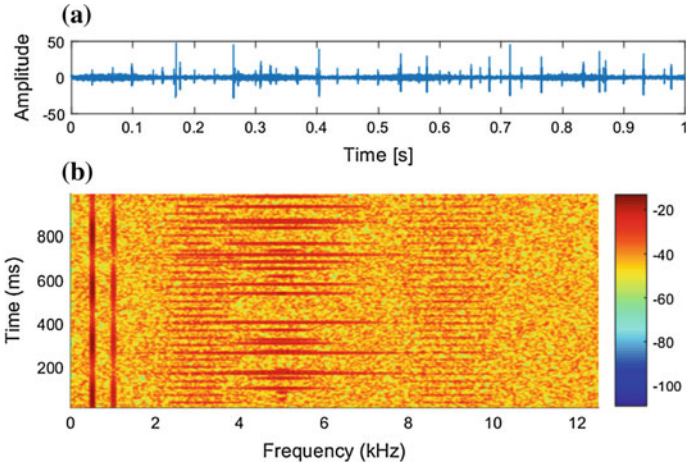


Fig. 1 Original signal (**a**) and spectrogram (**b**)

highest. It is known that for signals with amplitude modulated sinusoidal components the protrugram cannot detect the proper fault frequency (when the modulation function is with discrete spectrum). It is due to the fact that modulation function with discrete spectrum results in high kurtosis of the envelope spectrum. In this chapter we propose solution for the issue mentioned in the Sect. 1. The problem is to detect periodic impulsive excitation (cyclic pulse train). The solution should not be sensitive to non-cyclic impulses and narrowband amplitude modulated components. The problem raises to the fact that a locally damaged bearing can be associated with the periodic pulse train. It can be observed in the envelope spectrum of the signal, as a lot of equally spaced spectral components. Moreover, envelope spectrum of the amplitude modulated components consists of a small number of harmonics. Thus, kurtosis of such envelope spectrum is higher. Presence of a lot of equally spaced spectral components could be indicated using e.g. the empirical autocorrelation function. A desirable feature of the algorithm is a possibility to detect more than one frequency band with cyclic impulses. Furthermore, it would be beneficial to introduce a procedure, in which user does not have to precise fault frequency to detect the damage.

The idea is presented in the flowchart (Fig. 2). First of all, the bandwidth and step size have to be indicated. Then, using a band-pass filter the sub-signal ($SRF_{CF,BW}$) is obtained. The envelope of the SRF is calculated via Hilbert transform. Namely:

$$eSRF_{CF,BW} = |SRF_{CF,BW} + i * \mathcal{H}(SRF_{CF,BW})|, \tag{1}$$

where $\mathcal{H}(\cdot)$ is the Hilbert transform and $|\cdot|$ is the absolute value. Hilbert transform $\mathcal{H}(y(t))$ of $y(t)$ is simply a convolution of $y(t)$ with $h(t) = 1/\pi t$. Afterwards we calculate the mean-normalized amplitude spectrum of the envelope in order to indicate periodicity of the envelope. It is given by Eq. (2).

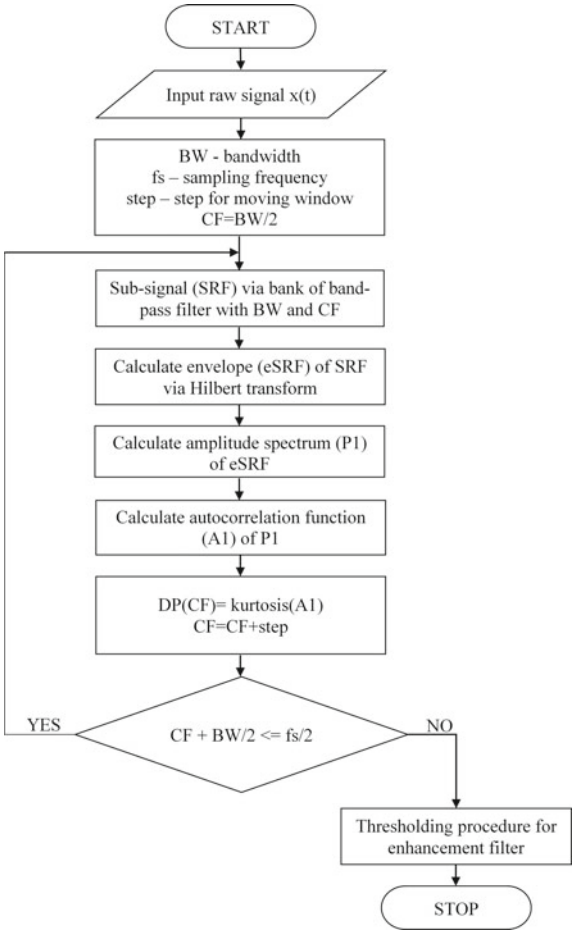


Fig. 2 Flowchart of the double protrugram

$$P1_{CF,BW}(k) = \left|DFT_k(eSRF_{CF,BW} - \mu_e)\right| = \\ = \left|\sum_{n=1}^{N}(eSRF_{CF,BW}(n) - \mu_e) * e^{-iw\pi kn/N}\right|, \tag{2}$$

where μ_e and N are the mean and length of the envelope, respectively. The envelope amplitude spectrum indicate both the cyclic impulses and amplitude modulated components with discrete-spectrum modulation function.. In order to distinguish these two components one can benefit from the property that cyclic impulses are represented by a lot of harmonics in the envelope spectrum. We propose to use autocorrelation function (ACF) in order to indicate whether the envelope spectrum consists of a lot of equally spaced harmonics or not. For a given lag k it is defined by Eq. (3):

$$ACF_{P1}(k) = \frac{1}{N-1} \frac{\sum_{n=1}^{N-k}(P1(n) - \overline{P1})(P1(n+k) - \overline{P1})}{\sigma_0^2}, \tag{3}$$

where N is length of the vector $P1$, $\overline{P1}$ is the mean value of the vector $P1$ and σ_0^2 is sample variance. This function measures the normalized correlation between the signal and its shifted (by k samples) copy. Because of that it is easy to detect the periodicity in the data, since a cyclic signal is similar to its shifted copy. ACF of an envelope spectrum with a lot of equally spaced harmonics should reveal significantly high values at lags corresponding to the spacing between harmonics. Thus, kurtosis of such ACF should be high. On the other hand, ACF of an envelope spectrum that consists of a single spike should not be spiky. Equation (4) presents formula for kurtosis of the autocorrelation function calculated for envelope spectrum $P1$.

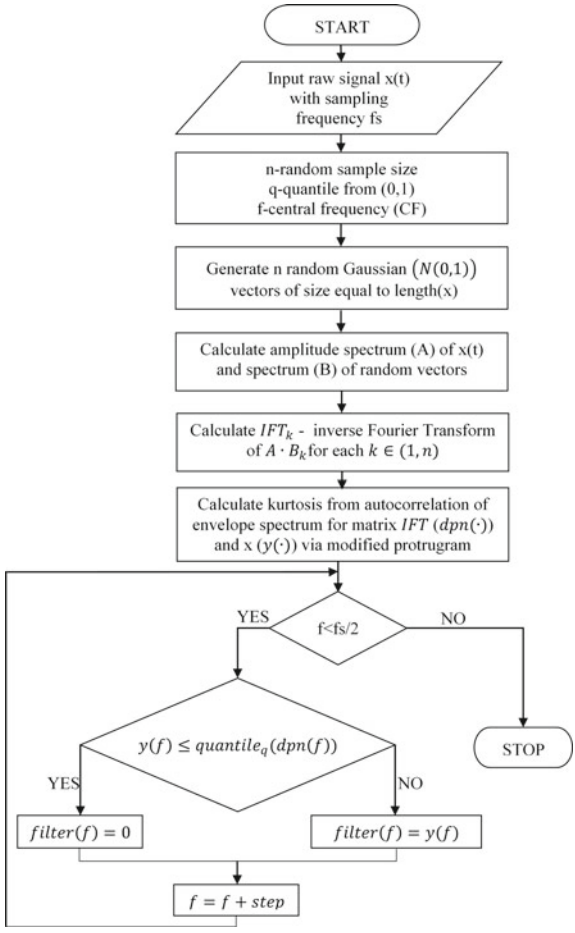


Fig. 3 Flowchart of the thresholding procedure

$$K(X) = \frac{\mu_4}{\sigma^4} = \frac{\frac{1}{N}\sum_{k=1}^{N}\left(ACF_{P1}(k) - \overline{ACF_{P1}}\right)^4}{\left(\frac{1}{N}\sum_{k=1}^{N}\left(ACF_{P1}(k) - \overline{ACF_{P1}}\right)^2\right)^2}, \tag{4}$$

which can be considered as an measure of impulsiveness. It is worth mentioning that for small k it is expected that the autocorrelation function would have high values for the mean-normalized envelope spectrum, despite occurrence of cyclic pulse train in the analyzed signal. According to this fact kurtosis should be calculated from the envelope spectrum with few first observations omitted. In our algorithm we decided to start from lag 10. Lags from 1 to 10 are related to spacing of envelope spectrum harmonic greater than 2500 Hz (under sampling frequency 25,000 Hz). Clearly, such high value cannot stand for a modulation frequency, thus the first lags can be removed. This procedure will provide the kurtosis with respect to frequency.

In order to enhance the filter, a thresholding procedure was applied (Fig. 3). Using the Monte Carlo method and choosing a high quantile (e.g. 0.95) of the generated values we can automatically obtain enhanced amplitude response of the filter. In short, when the kurtosis at a consider frequency is smaller than the quantile, we put 0 as the filter's amplitude response.

4 Results

In this section we compare three methods for damage detection in a rolling element bearing. The comparison is performed for the data described in Sect. 2. Recall, that the signal consists of amplitude modulated sine wave, cyclic and non-cyclic impulses and background noise. It will be shown that for such signal neither spectral kurtosis (SK) method nor protrugram can properly select informative frequency bands. The proposed algorithm outperforms these methods.

It can be expected that spectral kurtosis will detect the non-cyclic inputs as the SOI. It has been already described in [27] that in such signal SK method fails. This procedure is sensitive to impulsiveness, but it is unable to highlight periodicity. In case of the crusher, non-cyclic impulses can come from the crushing process, thus they appear regardless the condition of the bearing. On the other hand, protrugram, as it will be presented, assigns the largest kurtosis to the spectral frequencies related to the sinusoidal amplitude modulated discrete component of the signal. Indeed, in such case the discrete envelope spectrum consists of only one impulse. Obviously, it results in high value of kurtosis.

In Fig. 4 we present amplitude responses of filters obtained using double protrugram, protrugram and spectral kurtosis. Predictably, protrugram (Fig. 4b) assigns the largest kurtosis to the frequency corresponding to the amplitude modulated sinusoidal components (500 and 1000 Hz consist of one harmonic only). Furthermore,

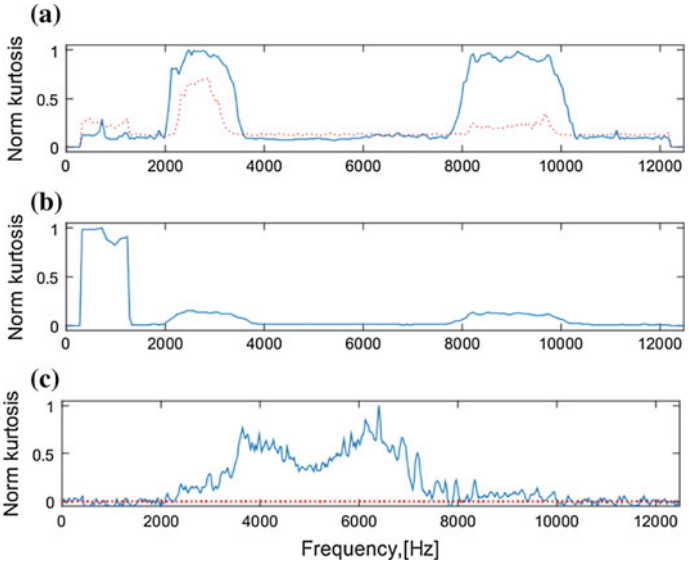


Fig. 4 Filter amplitude responses obtained using: **a** double protrugram, **b** protrugram, **c** spectral kurtosis. *Red line* is a cut off applied to a filter, everything below is assumed to be equal 0

as it was expected, SK (Fig. 4c) method indicates both the non-cyclic impulses and SOI. The proposed algorithm detects properly both of the carrier frequencies of the SOI (Fig. 4a). The thresholding procedure is beneficial, since it leads to automatic enhancement of the amplitude response of the filter.

The comparison of the original and filtered signals in the time domain is presented in Fig. 5. One can clearly notice that the signal filtered using protrugram consists only of amplitude modulated sinusoidal components and in SK case there are non-periodic impulses only. Spectrograms of the original and filtered signals are presented in Fig. 6. In the raw data (Fig. 6a) we can clearly see all of the components (amplitude modulated sine waves, cyclic and non-cyclic impulses). The proposed method leads to the signal with cyclic impulses in frequency bands 2000–3500 Hz and 8000–10,000 Hz and low-energy remains after the thresholding procedure (∼600 and ∼7000 Hz). Finally, envelope spectra are presented in Fig. 7. In case of double protrugram the modulation frequency is 30 Hz and a lot of harmonics are revealed. This stands for an impulsive modulation function, thus a local damage with fault frequency 30 Hz occurs. Envelope spectra of the raw and protrugram-filtered signals indicate modulation frequency of 4 Hz. Only one harmonic is presented, which corresponds to a sinusoidal modulation function. It means that the novel method is able to highlight the SOI in the signal contaminated by amplitude modulated sinusoidal components and non-cyclic impulses with high amplitudes.

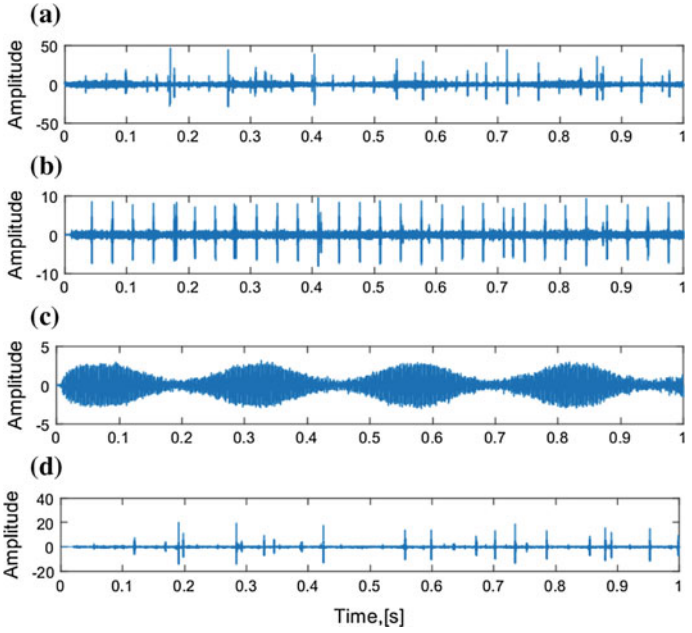


Fig. 5 **a** original signal and signals filtered via: **b** double protrugram, **c** protrugram, **d** spectral kurtosis

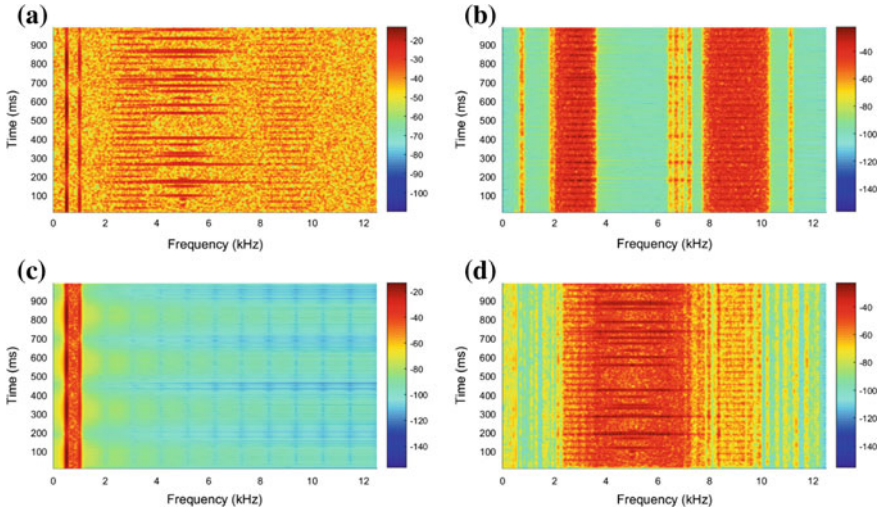


Fig. 6 Spectrograms of the **a** original signal, **b** double protrugram, **c** protrugram, **d** spectral kurtosis

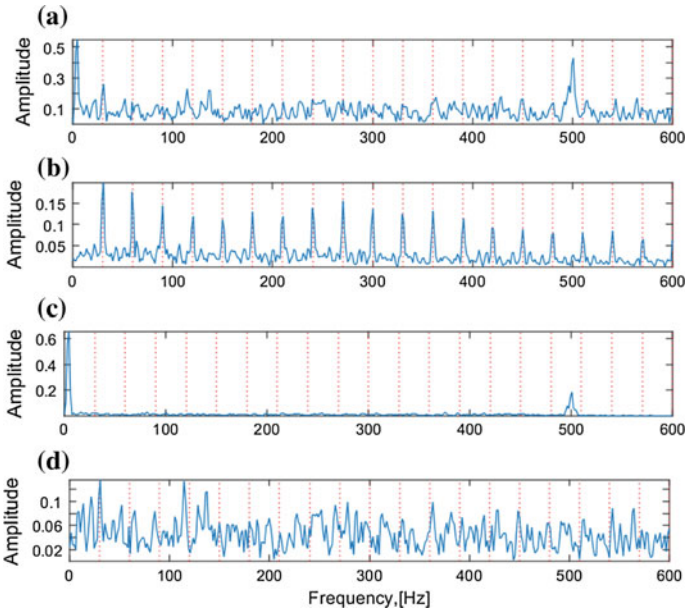


Fig. 7 Envelope spectrum of the **a** original signal and signals filtered via: **b** double protrugram, **c** protrugram, **d** spectral kurtosis

5 Conclusions

The main goal of this paper was to introduce a method for fault detection in bearings operating in machines or environment with sources of additional vibration signals (amplitude modulated sinusoidal and non-cyclic impulsive components). It was already shown [30] that in such situation an extension of protrugram could be beneficial. In this paper we present how to obtain an enhanced filter that separates cyclic impulsive source from other contaminating sources, which can occur in a signal. The algorithm takes advantage of both impulsive and cyclic character of the signal of interest. Moreover, the thresholding procedure is exploited in order to enhance the filter.

The proposed double protrugram algorithm is designed in order to indicate periodic pulse train. Moreover, it is not sensitive to contaminations like amplitude modulated sinusoidal components and non-cyclic wideband impulses. The presented filtering method does not only search for the highest value of an indicator and use a band-pass filter around appropriate frequency. It assesses for every frequency of the spectrum and automatically provides an enhanced amplitude response of the filter. This improvement is obtained via application of the thresholding procedure. This method is highly effective, although it is also time consuming, since it incorporates Monte Carlo simulations. This property can be seen as an major drawback of the method. In the future a faster algorithm for thresholds calculation is of high importance.

References

1. Randall, R. B. (2011). *Vibration-based condition monitoring: Industrial, aerospace and automotive applications*. New York: Wiley.
2. Feng, Z., Liang, M., & Chu, F. (2013). Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mechanical Systems and Signal Processing*, *38*, 165–205.
3. Gade, S., Herlufsen, H., Konstantin-Hansen, H., & Wismer, N. J. (1995). Order tracking analysis. Bruel & Kjaer Technical Review 2.
4. Bonnardot, F., Randall, R. B., & Antoni, J. (2004). Enhanced unsupervised noise cancellation using angular resampling for planetary bearing fault diagnosis. *International Journal of Acoustics and Vibration*, *9*(2), 51–60.
5. Caesarendra, W., Kosasih, B., Tieu, Anh Kiet, & Moodie, C. A. S. (2014). Circular domain features based condition monitoring for low speed slewing bearing. *Mechanical Systems and Signal Processing*, *45*(1), 114–138.
6. Zhao, M., Lin, J., Xu, X., & Lei, Y. (2013). Tacholess envelope order analysis and its application to fault detection of rolling element bearings with varying speeds. *Sensors*, *13*, 10856–10875.
7. Bai, M., Huang, J., Hong, M., & Su, F. (2005). Fault diagnosis of rotating machinery using an intelligent order tracking system. *Journal of Sound and Vibration*, *280*(3–5), 699–718.
8. Antoni, J. (2009). Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, *23*(4), 987–1036.
9. Antoni, J. (2007). Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*, *21*(2), 597–630.
10. Antoni, J. (2007). Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and Vibration*, *304*(3), 497–529.
11. Akkarakaran, S., & Vaidyanathan, P. P. (2000). Bifrequency and bispectrum maps: A new look at multirate systems with stochastic inputs. *IEEE Transactions on Signal Processing*, *48*(3), 723–736.
12. Urbanek, J., Antoni, J., & Barszcz, T. (2012). Detection of signal component modulations using modulation intensity distribution. *Mechanical Systems and Signal Processing*, *28*, 399–413.
13. Combet, F., & Zimroz, R. (2009). A new method for the estimation of the instantaneous speed relative fluctuation in a vibration signal based on the short time scale transform. *Mechanical Systems and Signal Processing*, *23*(4), 1382–1397.
14. Zimroz, R., Millioz, F., & Martin, N. (2010). A procedure of vibration analysis from planetary gearbox under non-stationary cyclic operations by instantaneous frequency estimation in time-frequency domain. In *Conference on Condition Monitoring and Machinery Failure Prevention Technologies (CM and MFPT 2010)* (p. nc).
15. Cardona-Morales, O., Sierra-Alonso, E. F., & Castellanos-Dominguez, G. (2016). Blind extraction of instantaneous frequency for order tracking in rotating machines under non-stationary operating conditions. In *Advances in Condition Monitoring of Machinery in Non-Stationary Operations* (pp. 99–110). Springer International Publishing.
16. Millioz, F., & Martin, N. (2006, July). Time-frequency segmentation for engine speed monitoring. In *Thirteen International Congress on Sound and Vibration, ICSV13*.
17. Villa, L. F., Reñones, A., Peráin, J. R., & De Miguel, L. J., (2011). Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation. *Mechanical Systems and Signal Processing*, *25*(6), 2157–2168.
18. Andre, H., Daher, Z., Antoni, J., & Rémond, D. (2010). Comparison between angular sampling and angular resampling methods applied on the vibration monitoring of a gear meshing in non stationary conditions. In *Proceedings of ISMA2010 including USD2010* (pp. 2727–2736).
19. Urbanek, J., Barszcz, T., & Antoni, J. (2014). Integrated modulation intensity distribution as a practical tool for condition monitoring. *Applied Acoustics*, *77*, 184–194.

20. Antoni, J., & Randall, R. B. (2006). The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, *20*(2), 308–331.
21. Obuchowski, J., Wyłomańska, A., & Zimroz, R. (2014). Selection of informative frequency band in local damage detection in rotating machinery. *Mechanical Systems and Signal Processing*, *48*(1), 138–152.
22. Tse, P. W., & Wang, D. (2013). The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as "Two automatic vibration-based fault diagnostic methods using the novel sparsity measurement—Parts 1 and 2". *Mechanical Systems and Signal Processing*, *40*, 499–519.
23. Peter, W. T., & Wang, D. (2013). The automatic selection of an optimal wavelet filter and its enhancement by the new sparsogram for bearing fault detection: Part 2 of the two related manuscripts that have a joint title as "Two automatic vibration-based fault diagnostic methods using the novel sparsity measurement—Parts 1 and 2". *Mechanical Systems and Signal Processing*, *40*(2), 520–544.
24. Chatterton, S., Borghesani, P., Pennacchi, P., & Vania, A. (2014, August). Optimal frequency band selection for the square envelope spectrum in the diagnostics of rolling element bearings. In *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. V008T11A003–V008T11A003). American Society of Mechanical Engineers.
25. Bo, L., & Peng, C. (2014). Fault diagnosis of rolling element bearing using more robust spectral kurtosis and intrinsic time-scale decomposition. *Journal of Vibration and Control*, 1077546314547727.
26. Żak, G., Wyłomańska, A., & Zimroz, R. (2015). Application of alpha-stable distribution approach for local damage detection in rotating machines. *Journal of Vibroengineering*, *17*(6), 1731.
27. Barszcz, T., & Jabloński, A. (2011). A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram. *Mechanical Systems and Signal Processing*, *25*(1), 431–451.
28. Antoni, J. (2015). The infogram: Entropic evidence of the signature of repetitive transients. *Mechanical Systems and Signal Processing* (In Press).
29. Antoni, J. (2014). Optimization of the envelope spectrum with the entropic uncertainty principle. In *ISMA International Conference on Noise and Vibration Engineering, Leuven, Belgium, 2014* (pp. 379–392).
30. Zimroz, R., Obuchowski, J., & Wyłomańska, A. (2016). Vibration analysis of copper ore crushers used in mineral processing plant-problem of bearings damage detection in presence of heavy impulsive noise. In *Advances in Condition Monitoring of Machinery in Non-Stationary Operations* (pp. 57–70). Springer International Publishing.
31. Barszcz, T., & Jabłoński, A. (2010). Selected methods of finding optimal center frequency for amplitude demodulation of vibration signals. *Diagnostyka*, 25–28.
32. Barszcz, T., Zimroz, R., Urbanek, J., Jabłoński, A., & Bartelmus, W. (2013). Bearings fault detection in gas compressor in presence of high level of non-Gaussian impulsive noise. *Key Engineering Materials*, *569*, 473–480.
33. Obuchowski, J., Zimroz, R., & Wylomanska, A. (2015). Identification of cyclic components in presence of non-Gaussian noise—application to crusher bearings damage detection. *Journal of Vibroengineering*, *17*(3), 1589.
34. Lee, S. K., & White, P. R. (1998). The enhancement of impulsive noise and vibration signals for fault detection in rotating and reciprocating machinery. *Journal of sound and Vibration*, *217*(3), 485–505.

Use of Cyclostationarity Based Condition Indicators for Gear Fault Diagnosis Under Fluctuating Speed Condition

Vikas Sharma and Anand Parey

Abstract The vibration based gear health monitoring is one of the condition monitoring techniques, widely used in industry. Under fluctuating speed conditions, vibration based conventional gear fault diagnosis methods like FFT and condition indicators (CI) like rms and kurtosis, fails to differentiate a faulty gear from a healthy one. Under such conditions, cyclic changes are observed in mean and variance of a vibration signal. CI using such statistical parameters for non-stationary gear vibration signal may mislead the entire fault diagnosis approach. In this chapter, CI based on cyclostationarity has been explained and used for gear fault diagnosis for fluctuating speed conditions. This chapter shows an advantage of using cyclostationarity based CI over conventional CI, for example rms and kurtosis, to diagnose fault. Result shows the effectiveness of the cyclostationarity based CI in differentiating gear health.

Keywords Condition indicators · Cyclostationarity · Fluctuating speed · Gear fault

1 Introduction

The variation in speed regulates the torque at the output of the gearbox. This variation in speed and the sliding friction [1–3] between the gears tooth pair is one of the contributors of vibrations produced in gearbox. In real life, these fluctuations in speed may appear in any fashion. Under such the fluctuating speed conditions the vibration signals generated are non-stationary. But, the present techniques for signal processing and statistical analysis used for gearbox fault diagnosis assume the vibration signal to be stationary.

V. Sharma · A. Parey (✉)
Department of Mechanical Engineering, Indian Institute of Technology Indore, Indore 453552, India
e-mail: anandp@iiti.ac.in

F. Chaari et al. (eds.), *Cyclostationarity: Theory and Methods III*,
Applied Condition Monitoring 6, DOI 10.1007/978-3-319-51445-1_15

According to signal processing concepts, fluctuations in speed result in modulation of vibration signals. Such modulations contribute and hide the existing modulations which were present due to fault. And because of this, conventional CI like rms and kurtosis fails to sense fault under fluctuating speed conditions [4–10]. Thus, vibration based gearbox fault diagnosis methods and conventional CI are inefficient under real life condition, i.e. fluctuating speed conditions.

The cause of modulation in vibration signal and its detection has been a topic of substantial research for over two decades. In order to detect modulation caused due to fault in gears, there are two possible approaches: feature-based (FB) methods [11, 12] and similarity based or likelihood-based (LB) methods [13, 14]. Signals acquired under real-life situation are not accurately stated as non-stationary; rather they can be represented as cyclostationary. Cyclostationary signals are defined as a class of non-stationary signals which are characterized by the periodic variation in their nth order moment with respect to period T. The demonstration of cyclostationarity of a signal is based on cyclic moment. Fault detection on the basis of cyclostationarity is a FB technique.

Gearboxes produce cyclostationary signals which are combination of both first order and second order cyclostationarity. The deterministic part of gearbox vibration signal results in first-order cyclostationarity (CS1) components appeared due to unbalances, eccentricities, etc. whereas; second-order cyclostationarity (CS2) results from phenomenon of random nature which is observed due to impacts, frictions etc. [15]. Cyclostationary analysis detects modulation caused due to faults by evaluating time varying cyclic cumulants and cyclic spectrum. This makes the cyclostationarity based CI to detect fault successfully even in low signal to noise ratio environments. Therefore, early fault detection using CI based cyclostationary analysis could prove to be novel as well as robust approach under fluctuating speed conditions.

This chapter explores the use of cyclostationarity for vibration based gear fault diagnosis under the fluctuating speed conditions. The cyclostationarity based CI (Sect. 3.2) has been explored for detecting gear fault. The gearbox is subjected to fluctuating profile of speed which has been imitated as the working conditions of wind turbines [16, 17]. The chapter is organized as follows. The state-of-art literature review about cyclostationarity is presented in Sect. 2. The fundamentals of cyclostationarity process and statistical properties of indicators are explained in Sect. 3. An experimental illustration is presented in Sect. 4. Discussion of results and the fault highlighting capacity is shown in Sect. 5. Future prospects and conclusions are presented in subsequent Sect. 6.

2 State of the Art Literature Review

In the year 2000, Dalpiaz et al. [18] compared the results based on wavelet transform and cyclostationarity analyses to time-synchronous average and cepstrum analysis. Moreover, the sensitivity to fault severity in terms of the fourth statistical

central moment was assessed by considering two different depths of the crack. Later in 2004, Antoni et al. [19] investigated the relationship between angle and time cyclostationarity by proving that rotating machine signals are intrinsically cyclostationary in angular domain rather than time domain. He also differentiated between degrees of pure cyclostationarity and impure cyclostationarity. In 2005, Zhu et al. [15] presented the usefulness of first order to the third order cyclostationarity in gearbox health monitoring. Both the spectrum of the second-order cyclic cumulant and the cyclic bispectrum of cumulant were found efficient in proving the degradation of the gearbox by the interactions between cyclic frequencies and, rotating and meshing frequency. Later in 2005, Bonnardot et al. [20] designed a time-varying linear-periodic Wiener filter, to discriminate noise without knowing its source, so that second-order cyclostationary source can be extracted for vibration analysis. Afterwards in 2007, Estupinan et al. [21] focused on the fault detection in the roller bearing using the cyclostationary analysis by using a cyclostationarity index and compared it with rms value; his proposed indicator demonstrated better sensitivity with respect to the occurrence of a fault. In 2009, vibration diagnostic methods for planetary gearbox synthesized by Bartelmus [22], also aimed cyclostationary analysis for fault signature extraction. Meanwhile in the same year, for monitoring planetary gearbox used in the mining industry, Zimroz and Bartelmus [23] explored the application of cyclostationary properties of signals to develop a diagnostic feature on the basis of spectral coherence map.

Till early/mid 2009, gear fault diagnostic analysis using cyclostationarity based CIs were restricted to stationary conditions. But later, gear fault diagnosis using cyclostationarity was considered at varying speed conditions. Fu and Li [24] in 2009, evaluated degree of cyclostationarity as a diagnostic feature for gear fault diagnosis under run up situation by converting the time domain transient signal into angular domain stationary signal. In 2011, Feng et al. [25] proposed a ratio of the sums of the cyclic spectral density magnitude at the cyclic frequencies of modulating frequency to 0 Hz along the frequency axis to figure out localized gear fault from amplitude modulated-frequency modulated (AM-FM) signal. By 2015, second-order cyclostationarity analysis was also used to analyze the model's non-linearity of the propagating crack during fatigue damage and expressed that random component which increases with fatigue fault. This magnitude of the cyclic frequency can be used as an effective indicator for fatigue damage [26]. For fluctuating speed conditions, Abboud et al. [27] in 2015, demonstrated analytical and experimental effectiveness of the angle/time strategy over classical cyclostationary approach for the vibration emitted by rolling element bearing fault.

Recently in 2016, cyclostationarity has been widely accepted in the field of early fault diagnosis at fluctuating speed profiles by scientific communities, as evident in two fine reviews [28, 29] where by the problems of statistical function estimation, signal detection and cycle frequency estimation is discussed. A summary of the authors focusing different methods or issues related to cyclostationary for the gear fault diagnosis is presented in Table 1.

Table 1 Summary of the authors focusing on cyclostationary

Methodologies or issues	Authors
Modeling strategies	Gardner [30], Antoni [28], Antoni et al. [19]
Effect of fluctuating speed	Abboud et al. [27], Fu and Li [24], Baudin et al. [31]
Cyclostationary indicators	Estupinan et al. [21], Girondin et al. [32], Napolitano [29], Kebabsa et al. [33], Feng et al. [25], Raad et al. [34], Delvecchio [35], Borghesani et al. [36], Léonard [37]
Application in gear fault diagnosis	Kidar et al. [4], Feng and Chu [38], Zhu et al. [15], Dalpiaz et al. [18], Bonnardot et al. [19], Boungou et al. [26], Eltabach et al. [39], Randall and Antoni [40], Leclere and Hamzaoui [41], Assaad et al. [42]

3 Cyclostationarity for Gear Fault Diagnosis

A non-stationary stochastic process is said to be *n*th order cyclostationarity on the basis of periodicity in *n*th statistical moment with time period T. The fundamental frequency; α of the periodicity is called cyclic frequency of the signal.

3.1 The First-Order and Second-Order Cyclostationarity

Cyclostationarity as stated above is defined on the basis of periodicity of their moments. A possible reason of failure of conventional CI i.e., in terms of detecting fault under fluctuating speed could be the periodicity of the moments. The characterization of signals can be done by using CI based on cyclostationarity.

Let us consider a time domain signal $x(t)$ which can be referred as *n*th order cyclostationary with period T. When the first-order moment, i.e. mean $m_x(t)$ is periodic with period T, then it is said to be a first-order cyclostationary (CS1) signal, which can be expressed as:

$$m_x(t) = E\{x(t)\} = m_x(t+T) \tag{1}$$

where, $E(\bullet)$ is the expected value and implies ensemble average.

Signals which exhibits a time varying auto-correlation function (ACF) can be termed as signal with second-order cyclostationarity (CS2) [28, 31]. Either amplitude modulations or frequency modulations or both appearing in stochastic processes are typical examples of CS2 signals. The ACF of a signal $x(t)$ is a periodic function with period T and is expressed as:

$$R_{xx}(t,\tau) = E\left\{x(t+\overline{\beta}\tau)x(t-\beta\tau)^*\right\} = R_{xx}(t+T,\tau) \tag{2}$$

For a second order stationary signal, autocorrelation function is independent of t, i.e., $R_{xx}(t,\tau) = R_{xx}(\tau)$. When a gearbox is operating, it will show cyclic behavior; this cyclic behavior is due to inherent periodicities embedded in it as kinematic variables viz. loads and speed fluctuates which are periodic to some angles. It implies that the signals are fundamentally cyclostationary in angular domain rather than time-cyclostationary [19]. Therefore to preserve their intrinsic cyclostationarity property, signal can be sampled as a function of an angular variable ϑ and a cyclic period Θ in the angular domain with angle lag as φ. It implies that the signal defined in angular domain can be written as $x(\vartheta)$. Therefore, for angular domain both Eq. (1) and (2) can be rewritten as:

$$m_x(\vartheta) = E\{x(\vartheta)\} = m_x(\vartheta + \Theta) \tag{3}$$

$$R_{xx}(\vartheta,\varphi) = E\left\{x(\vartheta + \overline{\beta}\varphi)x(\vartheta - \beta\varphi)^*\right\} = R_{xx}(\vartheta + \Theta, \varphi) \tag{4}$$

3.2 Cyclostationarity Based Condition Indicators

Cyclostationary approach has been used for fault diagnosis by estimating the cyclostationarity based CI: indicators of nth order cyclostationarity (ICS_{nx}) and the degree of second-order cyclostationarity (DCS^{α}) [34]. In order to define cyclostationarity based CI for signals, cumulant or moments can be used. The use of cumulants is worthy because the nth order cumulant gives significant information which is not carried in lower-order moments.

Assuming the signals to be cycloergodic, ensemble averaging can be replaced by synchronous averaging with the period T for evaluating first order cumulants [30]. Therefore, first order cumulant for period T with N number of averages is calculated as:

$$C_{1x} = \frac{1}{N}\sum_{n=0}^{N-1} x(t - nT) \tag{5}$$

For first order, the cumulant and the moment are same. So, subtracting the first order moment from the signal will derive the residual signal i.e., $r_x(\vartheta) = x(\vartheta) - m(\vartheta)$. With this expression, the characteristics frequencies and their harmonics are subtracted. This residual signal is used to calculate the second order cumulant. The second order cumulant is the autocovariance function of residual signal with zero mean and is expressed as:

$$C_{2x}(\vartheta,\varphi) = R_{xx}(\vartheta,\varphi) = E\{r_x(\vartheta)r_x(\vartheta - \varphi)\} \tag{6}$$

Angle lag will affect the second order cumulant as expressed in Eq. (6), because it measures the similarity between different observations as a function of angle lag

in angular domain and as a function of the time lag in time domain. With considering the angle-lag $\varphi = 0$, the Eq. (6) modified as:

$$C_{2x}(\vartheta, 0) = E\left\{r_x^2(\vartheta)\right\} \tag{7}$$

In angular domain both cumulants $C_{1x}(\vartheta)$ and $C_{2x}(\vartheta, \varphi)$ are considered periodic with period Θ, which can be expanded into a Fourier series as follows:

$$C_{1x}(\vartheta) = \sum_{\alpha \in A} C_{1x}^{\alpha}(\varphi) e^{j2\pi\alpha(\vartheta/\Theta)} \tag{8}$$

$$C_{2x}(\vartheta, \varphi) = \sum_{\alpha \in A} C_{2x}^{\alpha}(\vartheta, \varphi) e^{j2\pi\alpha(\vartheta/\Theta)} \tag{9}$$

And the cyclic cumulants can be written as:

$$C_{1x}^{\alpha} = \lim_{\Theta \to \infty} \frac{1}{\Theta} \int_{\Theta} C_{1x}^{\alpha}(\vartheta) e^{-j2\pi\alpha(\vartheta/\Theta)} d\vartheta \tag{10}$$

$$C_{2x}^{\alpha}(\varphi) = \lim_{\Theta \to \infty} \frac{1}{\Theta} \int_{\Theta} C_{2x}^{\alpha}(\vartheta, \varphi) e^{-j2\pi\alpha(\vartheta/\Theta)} d\vartheta \tag{11}$$

where, A is the set of cyclic orders. As per Raad et al. [34] at zero angle lag ($\varphi = 0$), the cyclic cumulants summarizes all the spectral information of second-order cyclostationary content of the signal. Therefore, they can be conveniently used to define the following indicators of cyclostationarity (ICS_{1x} and ICS_{2x})

$$ICS_{1x} = \frac{\sum_{\alpha \in \overline{A}, \alpha \neq 0} \left|C_{1x}^{\alpha}\right|^2}{\left|C_{2x}^{\alpha}(0)\right|^2} \tag{12}$$

$$ICS_{2x} = \frac{\sum_{\alpha \in \overline{A}, \alpha \neq 0} \left|C_{2x}^{\alpha}(0)\right|^2}{\left|C_{2x}^{\alpha}(0)\right|^2} \tag{13}$$

where, $\overline{A}$ is the set of interested cyclic frequencies which is possibly a subset of A. These indicators are dimensionless and they are normalized by the energy of residual signal.

Using Eq. (12) and (13) the ICS_{1x} and ICS_{2x} can be evaluated. Additionally, to find the information about components of the CS2 present in a signal for a cyclic frequency α by computing another CI the degree of cyclostationarity (DCS) as mentioned below [35]:

$$DCS^{\alpha} = \frac{\int \left|S_{2x}^{\alpha}(v)\right|^2 dv}{\int \left|S_{2x}^{0}(v)\right|^2 dv} = \frac{\int \left|C_{2x}^{\alpha}(\varphi)\right|^2 d\varphi}{\int \left|C_{2x}^{0}(\varphi)\right|^2 d\varphi} \tag{14}$$

where $S_{2x}^{\alpha}(\upsilon)$ is the cyclic spectrum, which implies the measure of the cyclic power distribution of the signal as a function of spectral order.

CIs for the cyclostationarity like ICS_{2x} and DCS^{α} fundamentally have the same physical interpretation as of CS2 signal. Nevertheless ICS_{2x} and DCS^{α} show some different characteristic features:

(1) DCS^{α} can provide information about the fault, if the signal shows some CS2 components; on contrary ICS_{2x} provide only a numerical value which being a summation over a cyclic frequency range $\overline{A}$.
(2) DCS^{α} is slow to evaluate, as it requires computation of cyclic cumulants for all angle lags; while, the evaluation of ICS_{2x} is much faster as it is based on FFT algorithm.

Concluding the characteristics of indicators with respect of gear fault diagnosis it can be stated that,

- For identifying fault periodicities, DCS^{α} can be used.
- To specify the gearbox state by means of simple numerical parameters, ICS_{2x} is useful and can be used as a suitable CI for neural network based intelligent pass/fail algorithms.

In Sect. 5, an attempt has been made to use cyclostationary based CI i.e., ICS_{1x}, ICS_{2x} and DCS^{α} on signals captured from gearbox under fluctuating speed conditions for fault detection.

4 Experimental Investigation

4.1 The Test Rig

The gear test rig uses an assembly of motor-drive-brake type, where gear can easily be mounted/dismounted to mesh. Figure 1 shows the pictorial view of the gear test rig for the present study. The drive to the test rig was provided by an AC drive motor using a 2.237 kW, 3 phase motor via a flexible coupling. A wide range of continuously variable speed from 0 to 3000 rev/min was provided by the control panel of the AC motor. The test rig can withstand a maximum radial load of 24.5 Nm (at the test gear) by magnetic brake type loading arrangement as shown in Fig. 1. The load on test gear and the speed of AC motor was programmed through NV Gate software. The transducer was attached on the bearing case of the test gear shaft and remains outside the gearbox. The layshaft of the gearbox is connected to the magnetic brake.

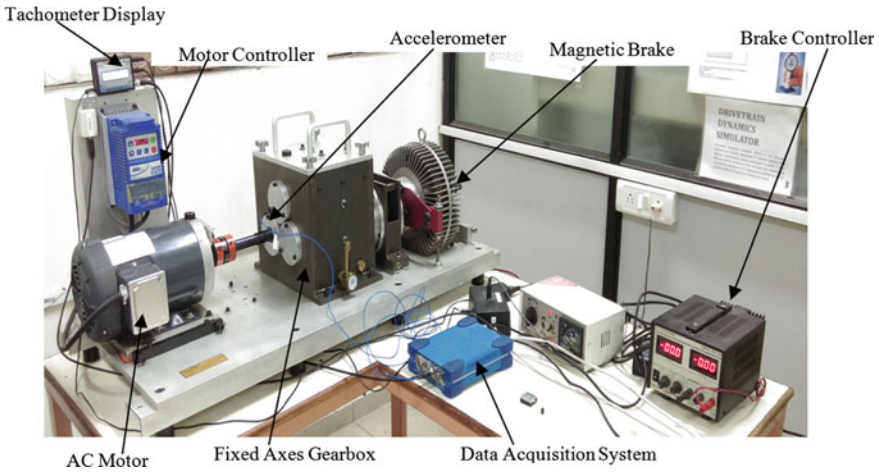


Fig. 1 Experimental test rig

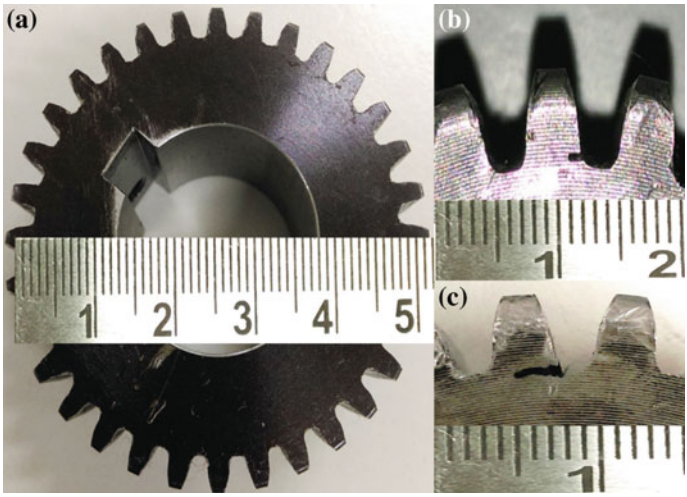


Fig. 2 **a** Healthy gear; **b** initial crack gear; **c** advanced crack gear (crack lengths in mm)

4.2 Test Gears

Spur gears with 20º pressure angle were used in this study with involute profile of medium carbon steel (AISI1045). Different crack on teeth were introduced in the gears by wire electro-discharge machining. Gears with no crack, initial crack and advanced crack as shown in Fig. 2 were tested under fluctuating speed conditions. The details of gears are listed in Table 2.

Table 2 Gear specification

Parameter	Input gear	Output gear
Number of teeth	32	80
Dimeteral pitch (per mm)	0.6299	0.6299
Module (mm)	1.5879	1.5879
Face width (mm)	12.5	12.5

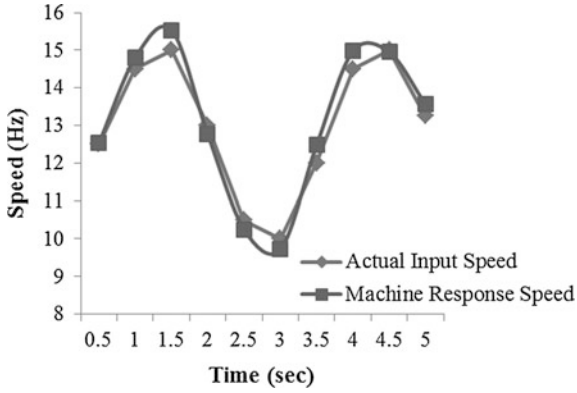


Fig. 3 Sinusoidal fluctuating profile of speed

4.3 Measurement Conditions

The speed was varied sinusoidally as shown in Fig. 3, for 750 rpm ranging from 600 to 900 rpm by keeping the load constant at 25%. Vibration signals were acquired with a sample rate 6.4 kHz for 6.25 s resulting total length of signal i.e., 40000 samples and 10 datasets comprising 5 signals each. These datasets were used to track the error in signal acquisition and processing.

A tachometer signal with 1pulse/rev was used to determine the angular position of the gear shaft and to resample the accelerometer signal in the angular domain with an angular resolution of 1°.

5 Results and Discussion

This section presents the results of experimental analysis of vibration signals captured, when the gearbox are made to work under fluctuating speed. Figure 4 shows the vibration signals for one revolution of different gear health, (a–c) in time domain and (d–f) in angle domain. From Fig. 4, the vibration signal of advance crack is undoubtedly noticeable as it produces a series of high transients while meshing. This implies that the occurrence of the tooth root crack fault possibly

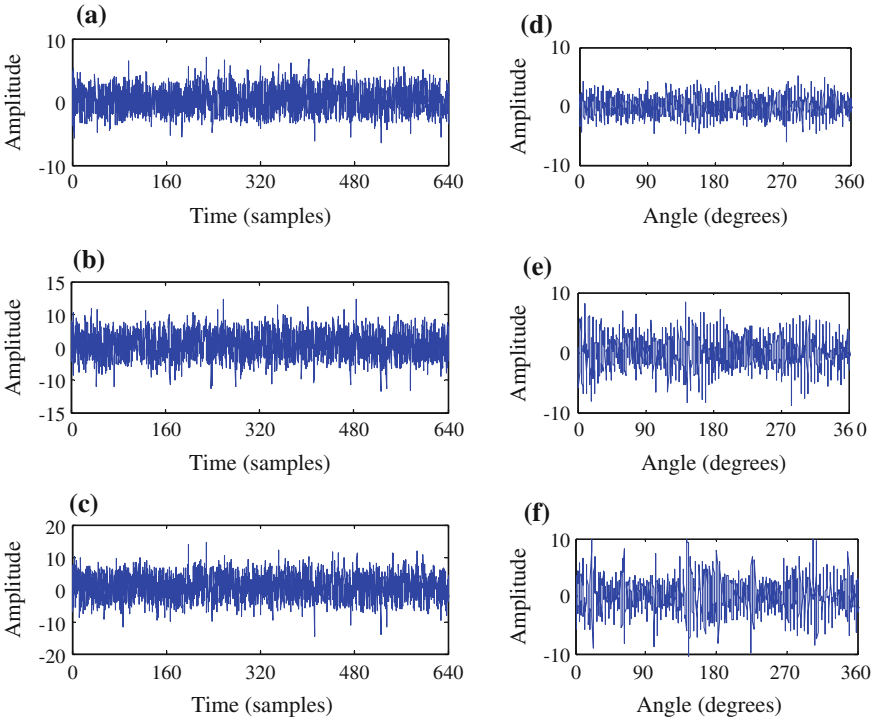


Fig. 4 Gear vibration signal for one revolution in time domain (*left column*) angular domain (*right column*), healthy gear (**a** and **d**), gear with an initial crack tooth (**b** and **e**), gear with advanced crack (**c** and **f**)

match an increase in order of cyclostationarity for the gearbox vibration signals. Comparison of the power spectral densities is shown in Fig. 5. It suggests that there is a considerable increase in cyclostationarity behaviour, because of occurrence of heavy sideband density around the gear-mesh and its harmonics with advancing fault under fluctuating speed conditions.

Both rms and kurtosis are computed for time domain and angular domain vibration signals. Table 3 provides information about the rms and kurtosis value for the faulty gears under both constant speed and fluctuating speed conditions.

It has been observed from Fig. 6 that rms and kurtosis both are sensitive towards fault for constant speed condition. However, in time domain, kurtosis fails to sense the fault under fluctuating speed condition. Similarly, in angular domain, rms fails to respond against fault. Therefore, under fluctuating speed conditions, these CI could be unreliable. The effectiveness of cyclostationarity based CI can now experimentally evaluated by identifying the fault within the gearbox.

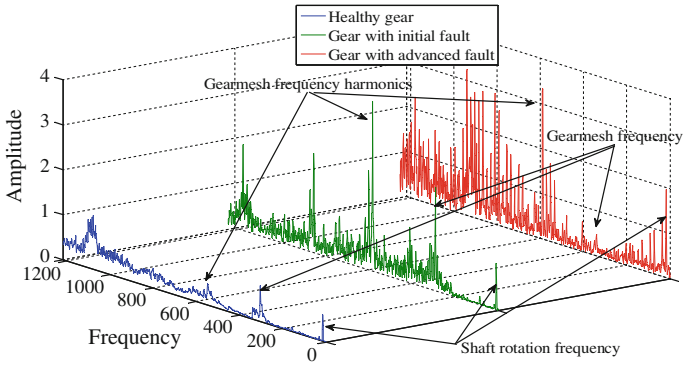


Fig. 5 Power spectral density for captured vibration signals of different gears under sinusoidal fluctuating speed environment

Table 3 CI evaluated for different conditions

	Time domain			Angular domain		
Features	No crack	Initial crack	Advanced crack	No crack	Initial crack	Advanced crack
(*a*) *Constant input speed*						
RMS	1.086	2.245	6.967	1.1547	1.2057	3.574
Kurtosis	3.440	4.060	9.243	3.2715	5.142	7.8514
(*b*) *Sinusoidal fluctuating input speed*						
RMS	0.195	0.351	1.247	1.2245	0.9814	1.8757
Kurtosis	3.299	3.651	2.932	3.5461	4.5801	8.0841

5.1 Application of Cyclostationarity Based CI

Under the effect of fluctuating speed conditions, synchronous averaging fails to detect the fault transients [8, 9, 43–45]. This restricts us to compute the first order cyclic cumulants and ICS_{1x}. However, the central information i.e., the residue more or less remains same, thus residue of the signal can be evaluated. Therefore, on the basis of this assumption, the cyclic cumulants for second order are estimated. Figure 7 illustrates the behaviour of estimated cyclic cumulants $C_{2x}^{\alpha}(0)$, as calculated from Eq. (10) in the cyclic order range of 0–70 and computed for the signals of healthy and faulty conditions of gearbox.

Second-order cyclic cumulants $C_{2x}^{\alpha}(0)$ displays few important components at this cyclic order for healthy gear vibration signal (Fig. 7a), on the other hand for faulty conditions, i.e., initial crack and advanced crack (Fig. 7b, c). It also depicts significant components above the threshold level of cyclic cumulants i.e., 0.075.

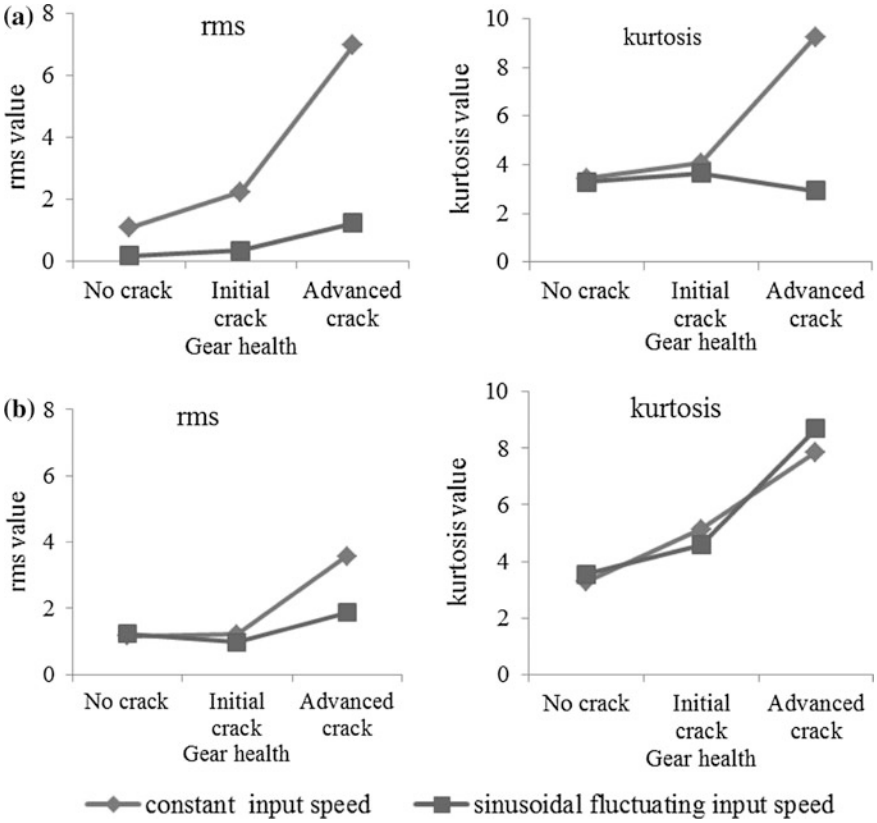


Fig. 6 Performance of rms and kurtosis for **a** time domain signals, **b** angular domain signals under different speed and gear health conditions

The threshold value is considered as 0.75 times of the maximum value of cumulants for healthy gear. Cyclic order 0–70 has been taken as cyclic order the range and order 70 as maximum value so as to generalize the validity of the indicator evaluation. The ICS_{2x} are evaluated separately for every gear, by considering the two sets of cyclic frequencies one for healthy signal and another for faulty signal, i.e., $\varpi_h = \{z \cdot O_h\}_{z=-Z}^{+Z}$ and $\varpi_f = \{q \cdot O_f\}_{q=-Q}^{+Q}$; where, $O_f = 0.5$ is the fundamental faulty order, $O_h = 2$ is the fundamental healthy order. The fundamental faulty order and fundamental healthy order is represented by Z and Q respectively. Order refers to frequency multiple of the reference rotational speed. Thus, healthy order and faulty order corresponds to healthy and fault frequencies. Also ϖ_h is a subset of ϖ_f. Mathematically, $\varpi_f \subseteq \varpi_h$.

Values computed for ICS_{2x} are displayed in Table 4 for 10 datasets under sinusoidal fluctuating speed conditions. First of all, it is seen that ICS_{2x} associated with the healthy gear is found low when the system assumed in good operation and

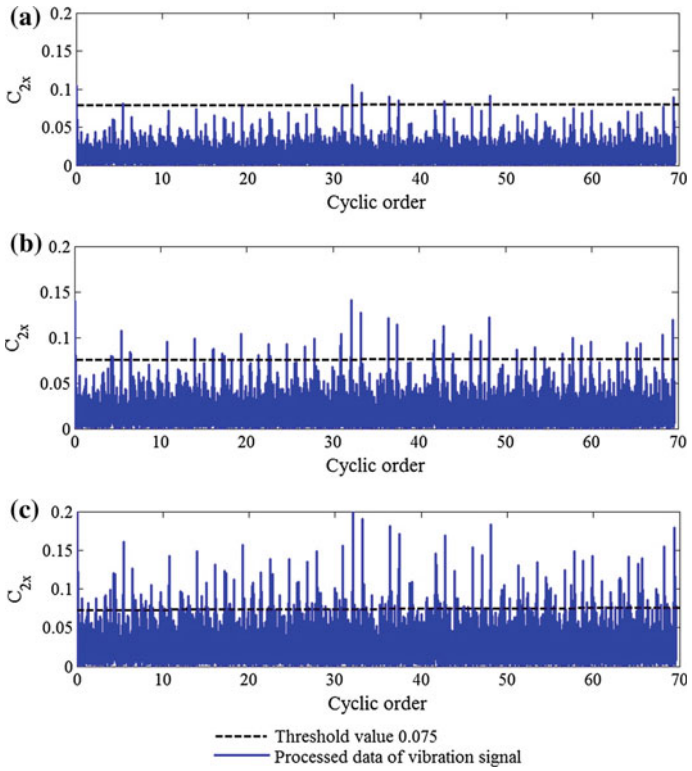


Fig. 7 Estimated cyclic cumulants $C_{2x}^{\alpha}(0)$ plotted in the cyclic order range 0–70 for vibration signals in different condition **a** healthy gear, **b** gear with initial crack tooth, **c** gear with advanced crack

Table 4 Value of ICS_{2x} for sinusoidal fluctuating input speed

Dataset #	No crack	Initial crack	Advanced crack
1	0.5701	10.24	16.101
2	0.3991	9.651	15.932
3	0.2975	10.144	16.814
4	0.0914	8.651	14.905
5	0.1975	10.824	16.984
6	0.3394	9.151	15.152
7	0.4571	9.054	15.870
8	0.3498	9.887	15.932
9	0.2570	9.751	15.778
10	0.3299	9.315	15.852

suggest that no fault development on the gear. On the other hand, the vibration signal with the help of ICS_{2x} exhibits a sudden increase of CS2 cyclic cumulant with the initiation of gear crack. This specifies that even though the speed of gearbox is fluctuating, cyclostationary based CI shows significant impact on the

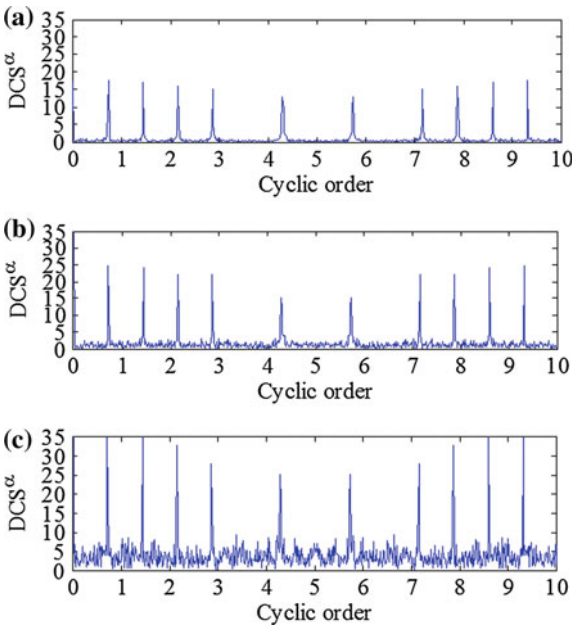


Fig. 8 DCS^{α} plotted in the cyclic order range 0–10 for vibration signals in different condition **a** healthy gear, **b** gear with initial crack tooth, **c** gear with advanced crack

diagnosing capability, even if the fault becomes critical enough to modulate the vibration signals significantly. This phenomenon justifies the second-order cyclostationary properties of this faulty signal. Further, with an increase in crack length, the DCS^{α} in cyclic order range 0–10 also increases showing a characteristic fault periodicity and its harmonics (Fig. 8). It is because of the fact that fault (crack in present study), produces CS2 components, occurring once in each cycle of gear rotation and mesh.

6 Conclusion

This chapter deals with the concepts and application of cyclostationarity based CI. The objective being gearbox fault diagnosis using CI, particularly ICS_{2x} and DCS^{α} under fluctuating speed condition. Unfortunately, conventional CI like rms and kurtosis fails to detect faults reliably under fluctuating speed condition; in both time and angular domain. In present study cyclostationarity based CI viz. ICS_{2x} and DCS^{α} were evaluated for gear fault diagnosis. ICS_{2x} was found sensitive toward both the crack levels for fluctuating speed conditions. DCS^{α} exhibited increase in amplitude of cyclic order for different gear health conditions. Thus, for fluctuating speed conditions, CIs based on cyclostationarity were found effective.

References

1. Jiang, H., Shao, Y., & Mechefske, C. K. (2014). Dynamic characteristics of helical gears under sliding friction with spalling defect. *Engineering Failure Analysis, 39,* 92–107.
2. Hun, C., Smith, W. A., Randall, R. B., & Peng, Z. (2016). Development of a gear vibration indicator and its application in gear wear monitoring. *Mechanical Systems and Signal Processing, 76–77,* 319–336.
3. Li, Z., & Mao, K. (2013). Frictional effects on gear tooth contact analysis. *Advances in Tribology*. doi:10.1155/2013/181048.
4. Kidar, T., Thomas, M., Badaoui, M. El & Guilbault, R. (2013). Diagnosis of gear faults by cyclostionaty. http://surveillance7.sciencesconf.org/conference/surveillance7/27_diagnosis_of_gear_faults_by_cyclostationarity.pdf.
5. Decker, H. J. (2002) Gear crack detection using tooth analysis‖, NASA-TM2002-211491.
6. Decker, H. J. & Lewicki, D. G. (2003). Spiral bevel pinion crack detection in a helicopter gearbox. In *Proceedings of the American Helicopter Society 59th Annual Forum* (pp. 1222–1232). Phoenix, AZ.
7. Zakrajsek J. J., Townsend D. P. & Decker H. J. (1993, January). An analysis of gear fault detection methods as applied to pitting fatigue failure data, Technical Report NASA TM-105950, AVSCOM TR-92-C-035, NASA and the US Army Aviation Systems Command.
8. Ahamed, N., Pandya, Y., & Parey, A. (2014). Spur gear tooth root crack detection using time synchronous averaging under fluctuating speed. *Measurement, 52,* 1–11.
9. Sharma, V., & Parey, A. (2016). Gear crack detection using modified TSA and proposed fault indicators for fluctuating speed conditions. *Measurement, 90,* 560–575. doi:10.1016/j.measurement.2016.04.076.
10. Pan, H. & Yuan, J. (2013). Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8228 LNCS (PART 3) (pp. 425–432).
11. Al-Balushi, K. R., & Samanta, B. (2002). Gear fault diagnosis using energy-based features of acoustic emission signals. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 216*(3), 249–263. doi:10.1177/095965180221600304.
12. Feng, Z., Chen, X., & Liang, M. (2016). Joint envelope and frequency order spectrum analysis based on iterative generalized demodulation for planetary gearbox fault diagnosis under nonstationary conditions. *Mechanical Systems and Signal Processing, 76–77,* 242–264.
13. Yang, H., Mathew, J., & Ma, L. (2003). Vibration feature extraction techniques for fault diagnosis of rotating machinery—a literature survey. In *Asia Pacific Vibration Conference*. Gold Coast, Australia.
14. Gelman, L., Petrunin, I., Jennions, I. K. & Walters M. (2012). Diagnostics of local tooth damage in gears by the wavelet technology. International Journal of Prognostics and Health Management. ISSN: 2153-2648, 2012 005.
15. Zhu, Z. K., Feng, Z. H., & Kong, F. R. (2005). Cyclostationarity analysis for gearbox condition monitoring: Approaches and effectiveness. *Mechanical Systems and Signal Processing, 19,* 467–482.
16. Zhao, M., Lin, J., Wang, X., Lei, Y., & Cao, J. (2013). A tacho-less order tracking technique for large speed variations. *Mechanical Systems and Signal Processing, 40,* 76–90.
17. Li, C., & Liang, M. (2013). Time–frequency signal analysis for gear box fault diagnosis using a generalized synchrosqueezing transform. *Mechanical Systems and Signal Processing, 26,* 205–217.
18. Dalpiaz, G., Rivola, A., & Rubini, R. (2000). Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears. *Mechanical Systems and Signal Processing, 14*(3), 387–412.

19. Antoni, J., Bonnardot, F., Raad, A., & Badaoui M. El. (2004). Cyclostationary modelling of rotating machine vibration signals. *Mechanical Systems and Signal Processing, 18*, 1285–1314.
20. Bonnardot, F., Randall, R. B., & Guillet, F. (2005). Extraction of second-order cyclostationary sources-application to vibration analysis. *Mechanical Systems and Signal Processing, 19,* 1230–1244.
21. Estupinan, E., White, P., & Martin, C. S. (2007). *A cyclostationary analysis applied to detection and diagnosis of faults in helicopter gearboxes*. Berlin, Heidelberg: Springer.
22. Bartelmus, W. (2009). Vibration Diagnostic Method for Planetary Gearboxes Under Varying External Load With Regard to Cyclostationary Analysis, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
23. Zimroz, R., & Bartelmus, W. (2009). Gearbox condition estimation using cyclostationary properties of vibration signal. *Key Engineering Materials, 413–414,* 471–478.
24. Fu, L., & Li, H. (2009). Gear fault diagnosis based on order tracking and degree of cyclostationarity. *IEEE.* doi:10.1109/CISP.2009.5301557.
25. Feng, Z., Zuo, M. J., Hao, R., Chu, F., & Badaoui, M. El. (2011). Gear damage assessment based on cyclic spectral analysis. *IEEE Transactions on Reliability, 60*(1), 21–32.
26. Boungou, D., Guillet, F., Lyonnet, P., & Rosario, T. (2015). Fatigue damage detection using cyclostationarity. *Mechanical Systems and Signal Processing, 58–59,* 128–142.
27. Abboud, D., Antoni, J., Eltabach, M., & Sieg-Zieba, S. (2015). Angle/time cyclostationarity for the analysis of rolling element bearing vibrations. *Measurement, 75,* 29–39.
28. Antoni, J. (2009). Cyclostationarity by examples. *Mechanical Systems and Signal Processing, 23,* 987–1036.
29. Napolitano, A. (2016). Cyclostationarity: New trends and applications. *Signal Processing, 120,* 85–408.
30. Gardner, W. (1986). Measurement of spectral correlation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *34*(5), 1111–1123.
31. Baudin, S., Rémond, D., Antoni, J., & Sauvage, O. (2016). Non-intrusive rattle noise detection in non-stationary conditions by an angle/time cyclostationary approach. *Journal of Sound and Vibration, 366,* 501–513.
32. Girondin, V., Pekpe, K. M.a, Morel, H. & Cassar J. P. (2013). Bearings fault detection in helicopters using frequency readjustment and cyclostationary analysis. *Mechanical Systems and Signal Processing*, *38*, 499–514.
33. Kebabsa, T., Ouelaa, N., Antoni, J., Djamaa, M. C., Khettabi, R., & Djebala, A. (2015). Experimental study of a turbo-alternator in industrial environment using cyclostationarity analysis. *International Journal of Advanced Manufacturing Technology, 81,* 537–552.
34. Raad, A., Antoni, J., & Sidahmed, M. (2008). Indicators of cyclostationarity: Theory and application to gear fault monitoring. *Mechanical Systems and Signal Processing, 22,* 574–587.
35. Delvecchio, S., D'Elia, G., & Dalpiaz, G. (2015). On the use of cyclostationary indicators in IC engine quality control by cold tests. *Mechanical Systems and Signal Processing, 60–61,* 208–228.
36. Borghesani, P., Pennacchi, P., Ricci, R., & Chatterton, S. (2013). Testing second order cyclostationarity in the squared envelope spectrum of non-white vibration signals. *Mechanical Systems and Signal Processing, 40,* 38–55.
37. Léonard, F. (2015). Time domain cyclostationarity signal-processing tools. *Mechanical Systems and Signal Processing, 62–63,* 100–112.
38. Feng, Z. & Chu, F. (2015). Cyclostationary analysis for gearbox and bearing fault diagnosis. *Shock and Vibration*, 542472, http://dx.doi.org/10.1155/2015/542472.
39. Eltabach, M., Sieg-Zieba, S., Vervaeke, T., Padioleau, E. & Berlingen, S. (2009). Monitoring volumetric gear pumps using cyclostationarity of the downstream pressure signals. Cetim.
40. Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—a tutorial. *Mechanical Systems and Signal Processing, 25,* 485–520.

41. Leclere, Q., & Hamzaoui, N. (2014). Using the moving synchronous average to analyze fuzzy cyclostationary signals. *Mechanical Systems and Signal Processing, 44,* 149–159.
42. Assaad, B., Eltabach, M., & Antoni, J. (2014). Vibration based condition monitoring of a multistage epicyclic gearbox in lifting cranes. *Mechanical Systems and Signal Processing, 42,* 351–367.
43. Jafarizadeh, M. A., Hassannejad, R., Ettefagh, M. M., & Chitsaz, S. (2008). Asynchronous input gear damage diagnosis using time averaging and wavelet filtering. *Mechanical Systems and Signal Processing, 22,* 172–201.
44. Bonnardot, F., Badaoui, M. EI, Randall, R. B., Danière, J. & Guillet F. (2005). Use of the acceleration signal of the gearbox in order to perform angular Resampling (with limited speed fluctuation). *Mechanical Systems and Signal Processing*, *19*, 766–785.
45. Combet, F., & Gelman, L. (2007). An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor. *Mechanical Systems and Signal Processing, 21,* 2590–2606.