archive ouverte UNIGE
http://archive-ouverte.unige.ch

Article

Wavelet-Variance-Based Estimation for Composite Stochastic Processes

GUERRIER, Stéphane, *et al*.

Abstract

This article presents a new estimationmethod for the parameters of a times series model.We consider here composite Gaussian processes that are the sum of independent Gaussian processes which, in turn, explain an important aspect of the time series, as is the case in engineering and natural sciences. The proposed estimation method offers an alternative to classical estimation based on the likelihood, that is straightforward to implement and often the only feasible estimation method with complex models. The estimator furnishes results as the optimization of a criterion based on a standardized distance between the sample wavelet variances (WV) estimates and the model-basedWV. Indeed, the WV provides a decomposition of the variance process through different scales, so that they contain the information about different features of the stochastic model. We derive the asymptotic properties of the proposed estimator for inference and perform a simulation study to compare our estimator to the MLE and the LSE with different models. We also set sufficient conditions on composite models for our estimator to be consistent, that are [...]

Reference

GUERRIER, Stéphane, *et al*. Wavelet-Variance-Based Estimation for Composite Stochastic Processes. *Journal of the American Statistical Association*, 2013, vol. 108, no. 503, p. 1021-1030

DOI : 10.1080/01621459.2013.799920

Available at:
http://archive-ouverte.unige.ch/unige:38161

Disclaimer: layout of this document may differ from the published version.

UNIVERSITÉ DE GENÈVE

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

Wavelet-Variance-Based Estimation for Composite Stochastic Processes

Stéphane GUERRIER, Jan SKALOUD, Yannick STEBLER, and Maria-Pia VICTORIA-FESER

This article presents a new estimation method for the parameters of a times series model. We consider here composite Gaussian processes that are the sum of independent Gaussian processes which, in turn, explain an important aspect of the time series, as is the case in engineering and natural sciences. The proposed estimation method offers an alternative to classical estimation based on the likelihood, that is straightforward to implement and often the only feasible estimation method with complex models. The estimator furnishes results as the optimization of a criterion based on a standardized distance between the sample wavelet variances (WV) estimates and the model-based WV. Indeed, the WV provides a decomposition of the variance process through different scales, so that they contain the information about different features of the stochastic model. We derive the asymptotic properties of the proposed estimator for inference and perform a simulation study to compare our estimator to the MLE and the LSE with different models. We also set sufficient conditions on composite models for our estimator to be consistent, that are easy to verify. We use the new estimator to estimate the stochastic error's parameters of the sum of three first order Gauss–Markov processes by means of a sample of over 800,000 issued from gyroscopes that compose inertial navigation systems.

KEY WORDS: Allan variance; Kalman filter; Signal processing; Time series.

1. INTRODUCTION

Let $F_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ be the model associated to the univariate Gaussian time series $\{Y_t; t \in \mathbb{Z}\}$ that is stationary or nonstationary but with stationary backward differences of order d, and let $\{y_t, t = 1, \ldots, T\}$ be the corresponding observed outcome. In this article, we propose a new estimation method for the parameter's vector $\boldsymbol{\theta}$ that is based on the matching between the empirical and model-based wavelet variances (WV). We also set the conditions on $F_{\boldsymbol{\theta}}$ under which our estimator is consistent. In particular, we consider for the empirical WV the Maximal Overlap Discrete Wavelet Transform (MODWT) WV estimator (Greenhall 1991; Percival and Guttorp 1994) based on Haar wavelet filters for which $d = 1$. In this case, we show that for a model made of the sum of independent Gaussian white-noise (WN), drift, quantization noise (QN), random walk (RW), and a finite number of autoregressive models of order 1 (AR(1)), our estimator is consistent (see Corollaries 2 and 3 in the supplementary materials, Section F). This model encompasses most of the stochastic models used in engineering and natural sciences applications.

The processes can be represented without loss of generality, by a state-space model of the form

$$\mathbf{x}_{t+1} = \boldsymbol{\Phi}\mathbf{x}_t + \mathbf{w}_{t+1} + \mathbf{u}_{t+1} \tag{1}$$

with measurements

$$y_{t+1} = \mathbf{h}\mathbf{x}_{t+1} + v_{t+1}, \tag{2}$$

where $\mathbf{x}_t$ is a $q \times 1$ system state vector at time t, $\boldsymbol{\Phi}$ is a $q \times q$ coefficient or state-transition matrix from t to $t+1$, $\mathbf{w}_t$ is a $q \times 1$ multivariate WN vector, that is, $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, $\mathbf{u}_t$ is a $q \times 1$ deterministic input vector, y_t is the one-dimensional output variable, $\mathbf{h}$ is a $1 \times q$ design vector which maps the state vector $\mathbf{x}_t$ into the output y_t, and v_t is a one-dimensional noise such that $v_t \sim \mathcal{N}(0, \sigma^2_{WN})$.

The books by Harvey (1989) and Durbin and Koopman (2001) contain extensive accounts of state-space models and their applications. For linear and/or Gaussian state-space models, the Maximum Likelihood Estimator (MLE) is a natural choice for the estimation of the model's parameters. The Kalman filter, MCMC, or other simulation smoother, together with the specification of a prior to start the latent process, are typically used for computing predictors of the state-variables and one-step-ahead predictors of the observations which are then used in an Expectation-Maximization (EM) algorithm of Dempster, Laird, and Rubing (1977) to compute the MLE (see also, e.g., Shumway and Stoffer 1982). The MLE is based on the density $f_{\boldsymbol{\theta}}(y) = \partial F_{\boldsymbol{\theta}}(y)/\partial y$, with $\boldsymbol{\theta}$ the $p \times 1$ vector of parameters containing the unknown elements of $\boldsymbol{\Phi}, \mathbf{Q}, \mathbf{h}$ and possibly also $\mathbf{u}_{t+1}$ together with σ^2_{WN}. The maximization step can be very complex and finding the MLE is not always a simple task. Moreover, even if new smoothers are regularly proposed (see, e.g., Durbin and Koopman 2002; McCausland, Miller, and Pelletier 2011) that improve the computational efficiency, the task becomes even more challenging when the observed process size is large and the model is, for example, a mixture of three first-order Gauss–Markov (GM) random processes (a reparameterization of AR(1) processes) with different parameters, a model that is used in the case study Section 5. We also found in a simulation exercise in Section 4 that our estimator, unlike the classical MLE or the least squares estimator (LSE), is importantly less biased when the roots of an AR(1) process

Stéphane Guerrier, Research Center for Statistics, HEC Genève, University of Geneva, Geneva, Switzerland (E-mail: *Stephane.Guerrier@unige.ch*). Jan Skaloud (E-mail: *jan.skaloud@epfl.ch*) and Yannick Stebler (E-mail: *yannick.stebler@epfl.ch*) Geodetic Engineering Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland (E-mail *jan.skaloud@epfl.ch*; *yannick.stebler@epfl.ch*). Maria-Pia Victoria-Feser, Research Center for Statistics, HEC Genève, University of Geneva, Geneva, Switzerland (E-mail: *Maria-Pia.VictoriaFeser@unige.ch*). Stéphane Guerrier and Maria-Pia Victoria-Feser are partially supported by a Swiss National Fund Grant (no. 100018-131906). Yannick Stebler was partially supported by the European FP7-GALILEO-2007-GSA-1 grand CLOSE-SEARCH as well as by the project no. 103010.2 of the Swiss Commission for Technology and Innovation (CTI). The authors are grateful to two anonymous reviewers for their comment that helped to improve the content and the presentation of the article.

© 2013 American Statistical Association

lie near the unit circle. However, a formal and more general treatment of the unit root problem is left for future research.

An important useful tool for understanding underlying features of a process is the Power Spectral Density (PSD), which for stationary processes, is given by $S_{F_\theta}(f) = \Delta t \sum_{\tau=-\infty}^{\infty} C_{F_\theta}(\tau) e^{-i2\pi f \tau \Delta t}$ for $|f| \leq f_N = 1/(2\Delta t)$ (f_N is the Nyquist frequency) with $C_{F_\theta}(\tau) = \text{cov}_{F_\theta}[Y_{t+\tau}, Y_t]$ being the Auto-Covariance Function (ACF) of $\{Y_t\}$ such that $\sum_{\tau=-\infty}^{\infty} C_{F_\theta}^2(\tau) < \infty$. The WV are related to the PSD since they decompose the variance process $\text{var}_{F_\theta}[y_t] = \int_{-f_N}^{f_N} S_{F_\theta}(f) df$ as thoroughly described in Percival and Walden (2000) (see also Section 2). The WV can be estimated from the sample $\{y_t, t = 1, \ldots, T\}$ using, for example, the maximal-overlap discrete wavelet transform estimator defined in Serroukh, Walden, and Percival (2000) (see also Percival 1995 and Equation (6) next), and for the matching of these empirical estimates to the WV implied by the model, we propose an optimization criterion based on a standardized distance between both WV estimates, as is done, for example, with the Generalized Method of Moments (GMM) (Hansen 1982). We call the new estimator the Generalized Method of Wavelet Moments (GMWM) estimator. We provide, among others, sufficient conditions on the model F_θ for the GMWM estimator to be consistent and asymptotically normal. Moreover, it can be computed using simulations as is done with indirect inference (Gourieroux, Monfort, and Renault 1993; Smith 1993; Gallant and Tauchen 1996), so that it is very straightforward to implement in practice.

While in this article we use wavelets (through their variances) for parametric estimation, wavelets have been used in many different statistical problems mainly for nonparametric estimation. To cite only a few, wavelets have been used in nonparametric density estimation (see, e.g., Doukhan and Leon 1990; Kerkyacharian and Picard 1992; Walter 1992; Donoho et al. 1996), inverse problems (see, e.g., Donoho 1995), nonparametric regression or curve fitting (see, e.g., Donoho and Johnstone 1994 and Donoho et al. 1995) also with correlated errors (see, e.g., Johnstone and Silverman 1997), or with locally stationary noise (see, e.g., Von Sachs and Macgibbon 2000).

This article is organized as follows. In Section 2, we introduce the WV and explain how it is used mainly in the engineering literature. In Section 3, we present the estimator based on the WV and derive its statistical properties. A simulation study is then presented in Section 4 that compares the finite sample performance of the GMWM estimator to the MLE and the LSE. In Section 5, we apply the new methodology to estimate parameters for the stochastic error models in inertial sensors with a real dataset of size 833,685.

2. THE WAVELET VARIANCE

Basically, as pointed out by Percival and Guttorp (1994), the WV can be interpreted as the variance of a process after it has been subject to an approximate bandpass filter. WV can be built using wavelet coefficients issued from a modified Discrete Wavelet Transform (DWT) (Mallat 1999; Percival and Walden 2000) called the Maximal Overlap DWT (MODWT); see Greenhall (1991), Percival and Guttorp (1994). The wavelet coefficients are built using wavelet filters $\{\tilde{h}_{j,l}\}$, $j = 1, \ldots, J$ which for $j = 1$ and for the MODWT satisfy

$$\sum_{l=0}^{L_1-1} \tilde{h}_{1,l} = 0, \quad \sum_{l=0}^{L_1-1} \tilde{h}_{1,l}^2 = \frac{1}{2} \quad \text{and} \quad \sum_{l=-\infty}^{\infty} \tilde{h}_{1,l} \tilde{h}_{1,l+2m} = 0,$$

where $\tilde{h}_{1,l} = 0$ for $l < 0$ and $l \geq L_1$, L_1 is the length of $\tilde{h}_{1,l}$, and m is a nonzero integer. Let also $\widetilde{H}_1(f) = \sum_{l=0}^{L_1-1} \tilde{h}_{1,l} e^{-i2\pi f l}$ be the transfer function of $\tilde{h}_{1,l}$. To obtain the jth-level wavelet filters $\{\tilde{h}_{j,l}\}$ of length $L_j = (2^j - 1)(L_1 - 1) + 1$, one computes the inverse discrete Fourier Transform of

$$\widetilde{H}_j(f) = \widetilde{H}_1(2^{j-1} f) \prod_{l=0}^{j-2} e^{i2\pi 2^l f (L_1-1)} \widetilde{H}_1\left(\frac{1}{2} - 2^l f\right). \quad (3)$$

The MODWT filter is actually a rescaled version of the DWT filter $h_{j,l}$, that is, $\tilde{h}_{j,l} = h_{j,l}/2^{j/2}$. Filtering an infinite sequence $\{Y_t; t \in \mathbb{Z}\}$ using the wavelet filters $\{\tilde{h}_{j,t}\}$ yields the MODWT wavelet coefficients

$$\overline{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} y_{t-l}, \ t \in \mathbb{Z}. \quad (4)$$

We define the WV at dyadic scales $\tau_j = 2^{j-1}$, as the variances of $\{\overline{W}_{j,t}\}$, that is

$$\nu^2(\tau_j) = \text{var}(\overline{W}_{j,t}).$$

Notice that the WV are assumed not to depend on time. The condition for this property to hold is that the integration order d for the series $\{Y_t\}$ has to be stationary such that $d \leq L_1/2$ and $\{\tilde{h}_{j,l}\}$ is based on a Daubechies (Daubechies 1992) wavelet filter; see Percival and Walden (2000), chapter 8. This is because Daubechies wavelet filters of width L_1 contain an embedded backward difference filter of order $L_1/2$. In such a case, the series of wavelet coefficients $\{\overline{W}_{j,t}\}$ is stationary with PSD $S_{W_j}(f) = |\widetilde{H}_j(f)|^2 S_{F_\theta}(f)$, $|\cdot|$ denoting the modulus. Hence (see Serroukh, Walden, and Percival 2000)

$$\nu^2(\tau_j) = \int_{-1/2}^{1/2} S_{W_j}(f) df = \int_{-1/2}^{1/2} |\widetilde{H}_j(f)|^2 S_{F_\theta}(f) df. \quad (5)$$

There is therefore an implicit link between the WV and the parameters of the data generating model F_θ. We exploit this connection when defining an estimator for $\boldsymbol{\theta}$, namely by matching a sample estimate of the WV $\nu^2(\tau_j)$ together with the model-based expression of the WV given by the left side of Equation (5). For WV based on Haar wavelet filters (see Equation (9) next) and for the WN, RW, Drift, QN, AR(1), ARMA(1,1), and ARIMA(0,1,1) models, the integral in Equation (5) is solved and given in the supplementary materials (see Section A), based on the results of Zhang (2008). WV for other models can be computed using the same methodology.

For a finite (observed) process $\{y_t; t = 1, \ldots, T\}$, the MODWT is given by

$$\hat{\nu}^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j}^{T} W_{j,t}^2 \quad (6)$$

with $W_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} y_{t-l}$, $t \in (L_j; T)$ and $M_j = T - L_j + 1$, is a consistent estimator for $\nu^2(\tau_j)$; see Serroukh, Walden, and Percival (2000) who also show that under suitable conditions (see also Theorem 3 in the supplementary materials),

$\sqrt{M_j}(\hat{\nu}^2(\tau_j) - \nu^2(\tau_j))$ is asymptotically normal with mean 0 and variance

$$S_{W_j}(0) = 2\int_{-1/2}^{1/2} S^2_{W_j}(f)df = 2\int_{-1/2}^{1/2} |\overline{H}_j(f)|^4 S^2_{F_\theta}(f)df. \quad (7)$$

Equation (7) can be estimated by means of

$$\widehat{S}_{W_j}(0) = \sum_{\tau=-M_j}^{M_j}\left[\frac{1}{M_j}\sum_{t=L_j}^{T} W_{j,t}W_{j,t+|\tau|}\right]^2 \quad (8)$$

and the asymptotic properties of Equation (8) are given in Percival and Walden (2000), p. 312.

A particular choice for the wavelet filter is the Haar wavelet filter whose first DWT filter ($j = 1$) is

$$\{h_{1,0} = 1/\sqrt{2},\ h_{1,1} = -1/\sqrt{2}\} \quad (9)$$

with length $L_1 = 2$. If the process is stationary with backward differences of order $d > 1$, then other wavelet filters such as Daubechies wavelet filters can be used (Daubechies 1992). When the WV is evaluated with Haar wavelet filters, it is actually equal to half the Allan variance (AV) (Allan 1966). In 1998, the IEEE Standard 1293 (1998) introduced the AV technique as a noise identification method which can be used for determining the characteristics of underlying stochastic processes affecting signals. AV and WV are often used in engineering disciplines and physical sciences as a graphical approach for model building purposes. In the supplementary materials (see Section B), we provide an example of the graphical use of WV. For a more detailed description, see, for example, Percival and Walden (2000).

When a graphical representation is suitable, with Haar WV, the parameters of stochastic processes are sometimes estimated by means of linear regression on preidentified linear regions in a log–log plot of scale τ_j versus WV $\nu^2(\tau_j)$. For example, this approach has been used for over 30 years as a standard routine measure of frequency stability in lasers (Fukuda, Tachikawa, and Kinoshita 2003) or atomic clocks (Allan 1987). More recently, the WV has also been used with optical sensors (Kebabian, Herndon, and Freedman 2005), various types of gas monitoring spectrometers (Werle, Mücke, and Slemr 1993; Bowling et al. 2003; Skrínskı et al. 2009), sonic anemometer-thermometers (Loescher et al. 2005), inertial sensors (El-Sheimy, Hou, and Niu 2008; Guerrier 2009), and radio-astronomical instrumentation (Schieder and Kramer 2001). The WV was also used in Percival and Guttorp (1994) to analyze vertical ocean shear measurements. In Fadel et al. (2004), it was employed to study the variability in heart-beat intervals. In Whitcher (2004), discrete wavelet packet transforms are used to estimate one of the parameters of a seasonal long memory process for the analysis of atmospheric and economic time series. In Gebber, Orer, and Barman (2006), WV is exploited to study nerve activities. WV has also been applied in Earth orientation metrology in Gambis (2002) and with other types of (geo)physical data. However, the linear regression on identified linear regions of the WV plots provides reasonably estimated parameters only for a limited number of processes and is often biased (Stebler et al. 2011). In our simulated example presented in Figure 4, the graphical estimation of first-order GM process parameters mixed with other processes like a WN is not suitable.

In the following section, we propose instead a criterion based on a standardized distance between sample and model-based WV that provides consistent estimators of the model's parameters for a wide range of models.

3. GMWM ESTIMATOR

We propose to estimate the model's parameters using an estimator which combines on the one hand the WV and on the other hand the Generalized Least Squares (GLS) principle, using the relationship given in Equation (5). More precisely, we propose to find $\hat{\boldsymbol{\theta}}$ such that the WV implied by the model, say $\boldsymbol{\phi}(\boldsymbol{\theta})$, matches the empirical WV, say $\hat{\boldsymbol{\phi}}$, and solves the following GLS optimization problem

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}(\boldsymbol{\theta}))^T \Omega (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}(\boldsymbol{\theta})) \quad (10)$$

in which Ω, a positive-definite weighting matrix, is chosen in a suitable manner (see next). Equation (10) defines the GMWM estimator. $\boldsymbol{\phi}(\cdot) = [\phi_j(\cdot)]_{j=1,\dots,J}$ is a binding function between $\boldsymbol{\theta}$ and $\boldsymbol{\nu}^2 = [\nu^2(\tau_j)]_{j=1,\dots,J}$ such that $\boldsymbol{\phi}(\boldsymbol{\theta}) = \boldsymbol{\nu}^2$, and $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\nu}}^2$ and $\boldsymbol{\phi}(\hat{\boldsymbol{\theta}})$ are two estimators.

In Theorem 3 given in the supplementary materials (see Section C), we use the results of Giraitis and Taqqu (1997) on limit theorems for bivariate Appell polynomials to set the conditions for $\hat{\boldsymbol{\phi}}$ to be asymptotically multivariate normal. In particular, $S_{F_\theta}(f) \leq C|f|^{-\alpha}$, $f \in [-1/2, 1/2]$, $\alpha < 1/2$. This result generalizes the results of Serroukh, Walden, and Percival (2000) to the multivariate case, while Serroukh and Walden (2000a, 2000b) generalize the results of Serroukh, Walden, and Percival (2000) to wavelet covariances.

In Theorem 1 next, we state the conditions on $\hat{\boldsymbol{\Omega}}$, $\hat{\boldsymbol{\phi}}$, and $\boldsymbol{\phi}(\boldsymbol{\theta})$ (hence on F_θ) under which $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$. For that, we follow the methodology used for GMM (see, e.g., Harris and Mátyás 1999) and use the results of Komunjer (2012). Let $\boldsymbol{\phi}(\boldsymbol{\theta}) = (\boldsymbol{\phi}_1(\boldsymbol{\theta}), \boldsymbol{\phi}_2(\boldsymbol{\theta}))$ with $\dim(\boldsymbol{\phi}_1(\boldsymbol{\theta})) = p$ (and $\dim(\boldsymbol{\phi}_2(\boldsymbol{\theta})) = J - p$), and let also $\boldsymbol{g}(\boldsymbol{\theta}) = \mathbb{E}[\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}(\boldsymbol{\theta})] = (\boldsymbol{g}_1(\boldsymbol{\theta}), \boldsymbol{g}_2(\boldsymbol{\theta}))$ (i.e., the same split as with $\boldsymbol{\phi}(\boldsymbol{\theta})$). Consider the following conditions:

- (C.1) If Ω is estimated by $\hat{\Omega}$, then $\hat{\Omega} \xrightarrow{P} \Omega$, that is, the weighting matrix converges to a population weighting matrix.
- (C.2) $\hat{\boldsymbol{\phi}}$ is a consistent estimator of the WV $\boldsymbol{\nu}^2 = [\nu^2(\tau_j)]_{j=1,\dots,J}$.
- (C.3) Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with $\boldsymbol{\Theta}$ being an open subset of $\mathbb{R}^p$. Then, there exists a bijective, twice continuous differentiable function $\boldsymbol{k} : \mathbb{R}^p \to \boldsymbol{\Theta}$, such that $\boldsymbol{\gamma} = \boldsymbol{k}^{-1}(\boldsymbol{\theta}) \in \mathbb{R}^p$, $\boldsymbol{k}(\boldsymbol{\gamma}) = [k_j(\gamma_j)]_{j=1,\dots,p}$, $|(\partial/\partial\boldsymbol{\gamma})\boldsymbol{k}(\boldsymbol{\gamma})| = \prod_{j=1}^p (\partial/\partial\gamma_j)k_j(\gamma_j) > 0$ (i.e., the determinant of the Jacobian matrix is nonnegative) and if $p = 2$, $(\partial/\partial\gamma_j)k_j(\gamma_j) > 0, \forall j$ (i.e., the Jacobian matrix is positive-definite definite).
- (C.4) $\boldsymbol{g}_1(\boldsymbol{\theta})$ is twice continuously differentiable.
- (C.5) $|(\partial/\partial\boldsymbol{\theta})\boldsymbol{g}_1(\boldsymbol{\theta})|$ is nonnegative.
- (C.6) $||\boldsymbol{g}_1(\boldsymbol{\theta})|| \to \infty$ whenever $||\boldsymbol{\theta}|| \to \infty$.
- (C.7) For every $c \in \mathbb{R}^p$, the equation $\boldsymbol{g}_1(\boldsymbol{\theta}) = c$ has countably many (possibly zero) solutions in $\mathbb{R}^p$.
- (C.8) For $p = 2$, $(\partial/\partial\boldsymbol{\theta})\boldsymbol{g}_1(\boldsymbol{\theta})$ is positive definite, and for $p > 2$, the set of points $\boldsymbol{\theta}_s$ for which $\operatorname{rank}[(\partial/\partial\boldsymbol{\theta})\boldsymbol{g}_1(\boldsymbol{\theta})] < p - 1$ is bounded.

Theorem 1. For Ω satisfying (C.1), under (C.2), for $\boldsymbol{g}(\boldsymbol{\theta})$ or $-\boldsymbol{g}(\boldsymbol{\theta})$ satisfying (C.4)–(C.8), and if there exists a function $\boldsymbol{f}$ satisfying (C.3), then $\hat{\boldsymbol{\theta}}$ defined in Equation (10) is consistent.

The proof is given in the supplementary materials (see Section D). Condition (C.3) defines a reparameterization from $\boldsymbol{\theta} \in \Theta$ to $\boldsymbol{\gamma} \in \mathbb{R}^p$. Hence, in practice, for a given model $F_{\boldsymbol{\theta}}$ one has first to check that a function $\boldsymbol{k} = [k_j]_{j=1,\ldots,p}$ exists that satisfies (C.3). For variance parameters $0 < \sigma^2$, the k_j's are the logarithm, and for other restricted parameters with restrictions of the form $a < \theta_j < b$, one can use the logit function on $(\theta_j - a)/(b-a)$, that is, $k_j(\gamma_j) = (b-a)\exp(\gamma_j)((1+\exp(\gamma_j)) + a$.

Condition (C.3) is convenient but actually too restrictive. The following corollary actually extends the conditions of Theorem 1.

Corollary 1. If the bijective, twice continuous differentiable function $\boldsymbol{k}$ does not satisfy the conditions on the Jacobian matrix in (C.3), then $\hat{\boldsymbol{\theta}}$ defined in (10) is consistent under conditions (C.1), (C.2), and (C.4)–(C.8) in which $(\partial/\partial\boldsymbol{\theta})\boldsymbol{g}_1(\boldsymbol{\theta})$ is replaced by $(\partial/\partial\boldsymbol{\theta})\boldsymbol{g}_1(\boldsymbol{\theta})(\partial/\partial\boldsymbol{\gamma})\boldsymbol{k}(\boldsymbol{\gamma})$.

The proof follows directly from the proof of Theorem 1.

In Theorem 2 below, we provide the consistency conditions of the GMWM when $F_{\boldsymbol{\theta}}$ is made of the sum of L independent processes. Let their respective WV be $\phi_j^{(l)}(\boldsymbol{\theta}^{(l)})$ and partition $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = [\boldsymbol{\theta}^{(l)}]_{l=1,\ldots,L}$, we need the following set of conditions:

(C.9) The functions $\phi_j^{(l)}(\boldsymbol{\theta}^{(l)})$ are linearly independent with all linear combinations of the other $\phi_j^{(l')}(\boldsymbol{\theta}^{(l')})$, $l \neq l'$, $l, l' = 1, \ldots, L$, and for all τ_j, that is, one cannot find $a_m, b_m \in \mathbb{R}$, $m = 1, \ldots, L$ such that $\phi_j^{(l)}(\boldsymbol{\theta}^{(l)}) = \sum_{\substack{m=1 \\ m \neq l}}^{L} a_m \phi_j^{(m)}(\boldsymbol{\theta}^{(m)}) + b_m$.

(C.10) For all j, $\phi_j^{(l)}(\boldsymbol{\theta}^{(l)}) = \phi_j^{(l)}(\boldsymbol{\theta}_\star^{(l)})$ if and only if $\boldsymbol{\theta}^{(l)} = \boldsymbol{\theta}_\star^{(l)}$ for all l.

(C.11) Each element of $\frac{\partial}{\partial\theta^{(l,m)}}\phi_j^{(l)}(\boldsymbol{\theta}^{(l)})$, where $\theta^{(l)} = [\theta^{(l,m)}]$, $m = 1, \ldots, \dim(\boldsymbol{\theta}^{(l)})$ depends on τ_j for at least $L-1$ of the $l = 1, \ldots, L$.

Theorem 2. Let $F_{\boldsymbol{\theta}}$ be the stochastic process associated to $\{Y_t, t \in \mathbb{Z}\}$ made of the sum of L independent processes with respective WV $\phi_j^{(l)}(\boldsymbol{\theta}^{(l)})$, $\hat{\boldsymbol{\theta}}$ be the GMWM estimator of $\boldsymbol{\theta}$, then for $\boldsymbol{\Omega}$ satisfying (C.1), $\hat{\boldsymbol{\phi}}$ satisfying (C.2), and $\phi_j^{(l)}(\boldsymbol{\theta}^{(l)})$ satisfying conditions (C.9)–(C.11), $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$.

The proof is given in the supplementary materials (see Section E). Condition (C.10) can be verified for each subprocess WV using the conditions of Theorem 1. We do that in the supplementary materials (see Section F) for some of the processes presented in Table 3, and for the composite processes treated in this article, we show the consistency using the conditions in Theorem 2.

Obviously, the number of scales J should be $J \geq p$ but at the same time, as discussed in Section 4, a too large J introduces variability in the estimator. The study of a possible optimal value for J is out of the scope of the present article. However, we may notice that in Guerrier et al. (2013), the authors investigate through simulation studies different choices for the number of scales J for mixtures of WN, QN, drift, RW, and AR(1) models, and found out that from a given number of scales, the variability of the GMWM does not really improve. This suggests that one could choose a number $J^\star$ such that $J^\star = \min(J, a \cdot p)$, with a a bounded (positive) integer.

Then, under the conditions of Theorem 1 (or possibly Corollary 1) or Theorem 2 as well as Theorem 3, we have that $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal, with asymptotic covariance matrix $\mathbf{V}_{\hat{\boldsymbol{\theta}}}$ given by

$$\mathbf{V}_{\hat{\boldsymbol{\theta}}} = \mathbf{B}\mathbf{V}_{\hat{\boldsymbol{\phi}}}\mathbf{B}^T, \tag{11}$$

where

$$\mathbf{B} = (\mathbf{D}^T\Omega\mathbf{D})^{-1}\mathbf{D}^T\Omega \tag{12}$$

and where $\mathbf{D} = \partial\boldsymbol{\phi}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T$, and $\mathbf{V}_{\hat{\boldsymbol{\phi}}}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\phi}}$ with elements given in (A-3). When $\boldsymbol{\Omega} = \mathbf{I}$, then $\mathbf{V}_{\hat{\boldsymbol{\theta}}} = (\mathbf{D}^T\mathbf{D})^{-T}\mathbf{D}^T\mathbf{V}_{\hat{\boldsymbol{\phi}}}\mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}$. The most efficient estimator is obtained by choosing $\Omega = \mathbf{V}_{\hat{\boldsymbol{\phi}}}^{-1}$, leading then to $\mathbf{V}_{\hat{\boldsymbol{\theta}}} = (\mathbf{D}^T\mathbf{V}_{\hat{\boldsymbol{\phi}}}^{-1}\mathbf{D})^{-1}$. In practice, the matrix $\mathbf{D}$ is computed at $\hat{\boldsymbol{\theta}}$.

The estimation of the WV covariance σ_{kl}^2 of $\mathbf{V}_{\hat{\boldsymbol{\phi}}}$ given in (A-3) is in general not straightforward. In the supplementary materials (see Section G), we show that under the assumption of a Gaussian process for Y_t, a suitable estimator is given by

$$\hat{\sigma}_{kl}^2 = \frac{1}{2}\sum_{\tau=-M_{kl}}^{M_{kl}}\left[\frac{1}{M_{kl}}\sum_{t=\max(L_j,L_k)}^{T} W_{k,t}W_{l,t+\tau}\right]^2 + \frac{1}{2}\sum_{\tau=-M_{kl}}^{M_{kl}}\left[\frac{1}{M_{kl}}\sum_{t=\max(L_j,L_k)}^{T} W_{k,t-\tau}W_{l,t}\right]^2, \tag{13}$$

where $M_{kl} = \min(M_l, M_k)$ (see Equation (6)). Alternatively, when the process is not Gaussian or when the sample size is very large as is the case with the dataset analyzed in Section 4 so that the computation of Equation (13) is infeasible, one can use a parametric bootstrap to estimate $\mathrm{cov}(\hat{\nu}^2(\tau_k), \hat{\nu}^2(\tau_l))$. Q samples of size T are simulated from $F_{\hat{\boldsymbol{\theta}}}$ on which Q WV $\hat{\nu}_q^2(\tau_k)$ and $\hat{\nu}_q^2(\tau_l)$, $q = 1, \ldots, Q$, are computed and σ_{kl}^2 is estimated by their empirical covariance.

When $J > p$, that is, the number of WV is greater than the dimension of the parameter vector $\boldsymbol{\theta}$, one can assess the goodness of fit of the model $F_{\boldsymbol{\theta}}$ to the data by testing the hypotheses $H_0 : E[\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}(\boldsymbol{\theta})] = 0$, $H_1 : E[\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}(\boldsymbol{\theta})] \neq 0$ using the χ^2-test statistic

$$T(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}(\hat{\boldsymbol{\theta}}))^T\mathbf{V}_{\hat{\boldsymbol{\phi}}}^{-1}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}(\hat{\boldsymbol{\theta}})) \tag{14}$$

which is asymptotically χ_{J-p}^2 under H_0 (see Hansen 1982). It will be used to assess the fit of the postulated model in the case study in Section 5. The investigation of the finite sample properties of Equation (14) is left for future research.

As a possible extension of the GMWM when analytical expressions for $\boldsymbol{\phi}(\boldsymbol{\theta})$ in Equation (10) are too complicated to compute, one can resort to simulations to compute $\boldsymbol{\phi}(\boldsymbol{\theta})$ and hence place the GMWM in the framework of indirect inference (Gourieroux, Monfort, and Renault 1993; Smith 1993; Gallant and Tauchen 1996). Basically, given a sample of observations $\{y_t, t = 1, \ldots, T\}$ and a hypothetical model $F_{\boldsymbol{\theta}}$, $\hat{\phi}_j$

is defined as the WV $\hat{\nu}^2(\tau_j)$ estimated from the sample using Equation (6). Let also $\hat{\boldsymbol{\phi}}^\star_j(\boldsymbol{\theta}) = \frac{1}{R}\sum_{r=1}^R \hat{\boldsymbol{\phi}}^\star_{j,r}(\boldsymbol{\theta})$ with $\hat{\boldsymbol{\phi}}^\star_{j,r}(\boldsymbol{\theta})$ being the WV estimates $\hat{\nu}^2_\star(\tau_j)$ computed on simulated series $\{y_t^{\star(r)}(\boldsymbol{\theta}), t = 1, \ldots, T\}, r = 1, \ldots, R$. Then, $\hat{\boldsymbol{\phi}} = [\hat{\phi}_j]_{j=1,\ldots,J}$ and $\hat{\boldsymbol{\phi}}^\star(\boldsymbol{\theta}) = [\hat{\phi}^\star_j(\boldsymbol{\theta})]_{j=1,\ldots,J}$ are used in Equation (10) to obtain an estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, whose properties are described in, for example, Gourieroux, Monfort, and Renault (1993). In particular, for R sufficiently large, $\mathbf{V}_{\hat{\theta}} \approx \mathbf{B}\mathbf{V}_{\hat{\boldsymbol{\phi}}}\mathbf{B}^T$. In that case, $\mathbf{B}$ can be computed numerically.

A step-by-step algorithm for computing the GMWM estimator (10) given a dataset of size T and a choice for $F_{\boldsymbol{\theta}}$, is given by the following steps:

- Step 1. Compute the empirical WV $\hat{\boldsymbol{\phi}}$ for a suitable number of scales $J \geq \dim(\boldsymbol{\theta})$ using Equation (6).
- Step 2. Construct the matrix $\boldsymbol{\Omega}$. We recommend using for $\boldsymbol{\Omega}$ the inverse of a diagonal matrix whose elements are $\widehat{S}_{W_j}(0)$ given in Equation (8) for the considered scales. This choice ensures that condition (C.1) is satisfied.
- Step 3. Compute the model-based WV $\boldsymbol{\phi}(\boldsymbol{\theta})$.
- Step 4. Minimize the quadratic form defined in Equation (10) in $\boldsymbol{\theta}$ to obtain the GMWM estimator $\hat{\boldsymbol{\theta}}$.
- Step 5. Compute confidence intervals for $\hat{\boldsymbol{\theta}}$ by parametric bootstrap or using Equation (11) with $\boldsymbol{\theta}$ replaced by $\hat{\boldsymbol{\theta}}$.

Hence, the method is rather easy to implement in practice. Assuming that we have a time series, say y_t of length T and a model $F_{\boldsymbol{\theta}}$ made of the sum of processes satisfying Corollary 3, the estimation of $\boldsymbol{\theta}$ starts with the computation (only once) of the empirical WV using Equation (6) and their estimated variances using Equation (8) (that are used in $\boldsymbol{\Omega}$). The model-based WV is then analytically built using the sum of the model-based WV (in Table 3) of the independent subprocesses in $F_{\boldsymbol{\theta}}$. Given all these ingredients, the quadratic form defined in Equation (10) is then optimized. This optimization problem is very simple. Confidence intervals for $\hat{\boldsymbol{\theta}}$ are, however, generally computed by parametric bootstrap. This task is quite easy to implement, but of course implies simulating data and repeating several times Steps 1–4. This can take in practice about a minute with massive datasets such as the one considered in the case study section.

If a simulation-based approach (indirect inference) is preferred, then Step 3 is not necessary and Step 4 is replaced by the following steps:

- Step 4a. Choose a sufficiently large value for R ($R = 100$ is usually a sensible choice)
- Step 4b. Minimize the quadratic form defined in Equation (10) in which $\boldsymbol{\phi}(\boldsymbol{\theta})$ is replaced by $\boldsymbol{\phi}^\star(\boldsymbol{\theta})$, each element being computed using Equation (6) in which the observations are replaced by a simulated sample from $F_{\boldsymbol{\theta}}$ (for a given current value of $\boldsymbol{\theta}$) of size $R \cdot T$.

Step 5 can be then used for inference, for sufficiently large R (see, e.g., Genton and Ronchetti 2003).

4. SIMULATION STUDY

This section is dedicated to the finite sample performance evaluation of the GMWM estimator compared to the MLE and the LSE. For the GMWM estimator, one has to make a choice about $\boldsymbol{\Omega}$ in Equation (10). The choice is not straightforward in practice, as it is the case for the moment-based estimator GMM or Indirect-Inference-based estimators. Asymptotically, it is well known that the optimal choice for $\boldsymbol{\Omega}$ is any matrix proportional to $\mathbf{V}^{-1}_{\hat{\boldsymbol{\phi}}}$. This does not mean that this choice is better in finite samples, especially because the asymptotic covariance matrix needs to be estimated. A simple choice is the identity matrix, and in our case an intermediate solution is the inverse of the diagonal matrix made of the variance estimates of the MODWT using Equation (8). Indeed, in our simulations, we found that choosing $\boldsymbol{\Omega} = \mathbf{V}^{-1}_{\hat{\boldsymbol{\phi}}}$ may lead in some cases to numerical instability, probably due to the estimation of the covariances using Equation (13). Anyway, asymptotically these three choices for $\boldsymbol{\Omega}$ lead to a consistent GMWM (provided the other conditions are also satisfied). When optimizing Equation (10), we use a quasi-Newton optimization method. For the MLE, we use the EM algorithm together with the Kalman smoother (EM-KF) as proposed by Shumway and Stoffer (1982) (see also Holmes 2010).

An important engineering application is sensor calibration in which the behavior of the errors affecting signals has to be well understood and modeled. In particular, gyroscopes and accelerometers which are part of inertial navigation systems used for space, aeronautical, ground, and underwater applications are subject to random errors that largely affect the quality of the positioning and navigation solution over time. A widely accepted model for the error behavior of an inertial sensor is given by the following state-space model

$$\begin{aligned} x_{t+1} &= e^{-\beta\Delta t}x_t + w_{t+1} + u_{t+1} \\ y_{t+1} &= x_{t+1} + v_{t+1,} \end{aligned} \tag{15}$$

where Δt is the time interval between two consecutive measurements, $u_t = \omega\Delta t$, $w_t \sim \mathcal{N}(0, q)$ with $q = \sigma^2_{\mathrm{GM}}(1 - e^{-2\beta\Delta t})$, and $v_t \sim \mathcal{N}(0, \sigma^2_{\mathrm{WN}})$. The observed series is hence made up of the sum of a first-order GM process with correlation time β^{-1} and variance σ^2_{GM}, a drift with slope ω, and a WN with variance σ^2_{WN}. Even with such a relatively simple model, the estimation task is nontrivial.

For our simulation study, we consider simulated processes $\{y_t,\ t = 1, \ldots T\}$ from model (15) as well as an AR(1) with autocorrelation parameter near the boundary value of 1. From model (15), we actually generate three types of samples which correspond to three different (sub)models. In Model 1, we set $\sigma^2_{\mathrm{WN}} = \omega = 0$, hence we have only a first-order GM, in Model 2, we set $\omega = 0$, hence we remove the drift only from the complete model (15) (Model 3). We generate 100 processes of size $T = 6000$ with $\Delta t = 1$ and $\boldsymbol{\theta} = (\sigma^2_{\mathrm{WN}}, \sigma^2_{\mathrm{GM}}, \beta, \omega) = (4, 16, 0.05, 0.005)$ at the complete model. For the submodels, the parameters are constrained accordingly and they are not estimated.

For both the GMWM (simulation based or not) and the EM-KF, the initial values for the optimizations were set to $\boldsymbol{\theta}_{(\mathrm{start})} = (1.0, 1.0, 1.0, 0.0)$, which is relatively far away from the true simulation values. We found that the choice for the starting

Table 1. RMSE and relative RMSE (R-RMSE) of the GMWM and EM-KF estimators for 100 simulated processes of size $T = 6000$ from model (15) (Model 3) with $\Delta t = 1$ and $\boldsymbol{\theta} = (\sigma^2_{\mathrm{WN}}, \sigma^2_{\mathrm{GM}}, \beta, \omega) = (4, 16, 0.05, 0.005)$ and from submodels with $\sigma^2_{\mathrm{WN}} = \omega = 0$ (Model 1) and with $\omega = 0$ (Model 2)

		GMWM		EM-KF	
		RMSE	R-RMSE	RMSE	R-RMSE
Model 1	σ^2_{GM}	2.00	0.13	1.26	0.08
	β	$8.14 \cdot 10^{-3}$	0.16	$4.09 \cdot 10^{-3}$	0.08
Model 2	σ^2_{GM}	1.92	0.12	1.55	0.10
	β	$1.05 \cdot 10^{-2}$	0.21	$4.94 \cdot 10^{-3}$	0.10
	σ^2_{WN}	0.13	0.03	0.14	0.04
Model 3	σ^2_{GM}	0.96	0.06	74.57	4.66
	β	$4.63 \cdot 10^{-3}$	0.09	0.04	0.85
	σ^2_{WN}	0.11	0.03	0.16	0.04
	ω	$2.79 \cdot 10^{-4}$	0.06	0.12	23.58

values is not a serious issue for the computation of the GMWM. The root mean-squared errors (RMSE) as well as the relative RMSE (relative to the true parameter value) are presented in Table 1 for Models 1–3, and for the GMWM and the EM-KF. We also tried a simulation-based version of the GMWM with $R = 100$ and found out that the RMSE are of the same order as for the GMWM with analytical WV (results not presented here). The sample WV were computed for $J = 12 (< \log(6000)/\log(2) \approx 12.55)$ scales for all models. The results show that for the smaller models 1 and 2, the RMSE is smaller for the EM-KF than for the GMWM estimator, while the RMSE of the EM-KF explodes for the complete model 3. This last feature is actually well known with models with a drift component. For example, Hinrichsen and Holmes (2009) consider a multivariate state-space model composed of a drift with unknown rate and a random walk process to model the growth of an ecological population with observations that are subject to measurement error. They actually use a two-step estimation procedure by which the drift is first estimated by linear regression and then removed from the state equation leading to a simpler model that is well estimated by means of the EM-KF (see also Stebler et al. 2011).

When the EM-KF behaves well (models 1 and 2), it has a better performance in terms of RMSE than the GMWM estimator. However, one can further improve the efficiency of the latter by decreasing the number of scales J at which the WV are estimated. Indeed, in this example, $J = 12$ scales are used to estimate two or three parameters, and if more scales are added (supposing a larger T), this only introduces more variability in the GMWM estimator. An optimal choice (in terms of GMWM estimator's efficiency) of the scales and their number is beyond the scope of the present article. The number of scales is obviously a function of the number of parameters p, but their choice (among the possible ones) depends on the model that is considered. For example, with model (15) without the drift element (i.e., Model 2), according to the WV graph in Figure 4, if the last three to four scales are ignored, then the WV are still able to capture information about the other model components. Actually, removing the last four scales improves the efficiency of the GMWM estimator in the simulation study for Model 2 (results not shown here).

An important feature of the GMWM estimator is that it can be used for models for which the MLE (or other classical estimators) is seriously biased. For example, maximum likelihood estimation of ARMA processes, and/or least squares estimation (based on the Yule–Walker equations) of AR processes, when the roots lie near the unit circle, provide estimators that can be seriously biased (see, e.g., Andrews 1993). With the GMWM estimator with Haar coefficients for the WV, we do not have this boundary problem, since, for example, with an AR(1) process, with autocorrelation parameter ρ, when the latter tends to the value of 1, we get a random walk which is a nonstationary process. However, the wavelet coefficients $W_{j,t}$ actually are backward differences of order at least 1, so that the problem of nonstationarity disappears. Note also that taking first differences of a stationary AR(1) process is known to lead to the problem of overdifferencing, since the resulting series is not invertible (it becomes an ARMA(1,1)) which has more parameters than the original one. Consequently, its parameters will be difficult to estimate, and it will tend to degrade the quality of forecasts, so differencing is not always a suitable method to overcome the estimation problem.

As an illustration, we simulated 1000 AR(1) processes of size $T = 5000$ with ρ ranging from 0.8 to 1 (and with residual variance $\sigma^2 = 4$) and estimated the LSE based on the Yule–Walker equations together with the GMWM ($J = 12$). Figure 1 presents the relative RMSE of the LSE and GMWM estimator for both parameters. The GMWM estimator of ρ is less precise than the LSE, except when ρ is close to 1. For σ^2, the RMSE of both estimators are quite similar for ρ smaller than 0.9 but for higher values the bias and variance of the LSE explodes showing that there is a boundary effect that affects the LSE. The MLE provides RMSE that are sensibly smaller than the LSE (not presented here) but for values of ρ larger than 0.95 the MLE is difficult if not impossible to compute.

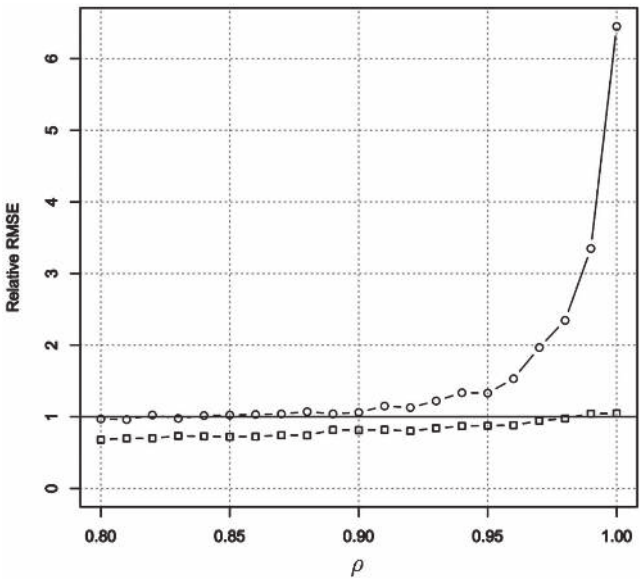

Figure 1. Relative RMSE based on 1000 simulations from AR(1) processes with ρ ranging from 0.8 to 1 and with residual variance $\sigma^2 = 4$, of the LSE relative to the GMWM estimator of ρ ("□") and of σ^2 ("o").

5. CASE STUDY: INERTIAL SENSORS

An inertial navigation system exploits observation from usually an orthogonal triad of gyroscopes and accelerometers measuring angular rates and specific forces, respectively. Once initialized, the navigation is a dead-reckoning process (i.e., the solution at one epoch is computed from current observations as well as from the solution of the previous epoch) in which gyroscope signals (after suppressing Earth rotation) are integrated once to yield orientation, and accelerometers (after accounting for gravity and apparent forces) are integrated twice to get the velocity, and finally the position in three-dimensional space. The advantage of inertial navigation is in the autonomy (i.e., no external infrastructure is needed) and the large sensor bandwidth (unlike satellite positioning system), while its weakness lies in the time-dependent error behavior (due to the integration process). Indeed, the sensor signals are corrupted by random errors, making the resulting positioning and attitude (or orientation) error increase rapidly with time. To bound the error growth, inertial sensors are combined with other sensors (e.g., GPS, odometer, and altimeter) through optimal (e.g., Kalman) filtering. Such combined navigation systems are not only the base for manned and unmanned vehicle (e.g., spacecrafts, aircrafts, cars, and robots) trajectory and orientation determination, but also for pedestrian and indoor localization devices. The gyroscope and accelerometer error behavior is modeled through state augmentation in the system where the augmented states account for sensor errors usually modeled as stochastic processes. This requires that the parameters of the stochastic processes in the state-space model are carefully set up prior to filtering. The quality of this setup (or estimation) strongly influences the quality of the predicted trajectory. Traditionally, the Allan variance with ad hoc graphical estimation (see, e.g., IEEE Standard 1293 1998), or the KF-Self-Tuning approach (see Niu et al. 2007; Waegli 2009; Stebler et al. 2012), or even the MLE on state-space models (with simple models) (Stebler et al. 2011) are used to both determine the error's model structure and to estimate their parameters. The GMWM estimator offers an alternative approach that is especially suitable when the stochastic error process is rather complex.

In this section, we consider the angular rate signal issued from a real Micro-Electro-Mechanical System (MEMS) gyroscope where the spectral structure of errors is often complex. Such sensors have a great potential in navigation due to ergonomic (smaller and lighter equipment) and economical considerations (El-Sheimy and Niu 2007). Therefore, MEMS sensors are increasingly used in a very wide range of applications such as 3D input devices, robotic, virtual reality, vehicle stability control and so forth, and research focuses on modeling and compensating for their large and variable measurement errors. To validate the use of the GMWM for modeling the MEMS' errors, we first analyze the signal produced in static conditions at a frequency of 100 Hz (833, 685 measurements). We provide GMWM estimates on a rather complex model that cannot be estimated by an alternative standard method. To show the impact of the importance of model structure and estimation precision of its parameters in positioning, we then present an emulation study in which the trajectory of a helicopter performing airborne laser scanning is used and for which a nearly perfect (i.e., reference) trajectory can be computed. The emulated measurements along the real trajectory are corrupted with an error signal observed in static condition, and the three methods (AV, the KF-Self-Tuning Approach, and GMWM) are used to model the error signal. The predicted trajectories based on the corrupted measurements and estimated error model, together with GPS measurements, are then compared to determine the positioning error.

After removing the mean in the signal made of 833, 685 measurements produced in static conditions at a frequency of 100 Hz, the time-varying part of the sensor errors is directly available and is presented in the time domain in Figure 3 (upper panel), together with the Haar WV of this process with 95% confidence intervals (lower panel). The WV computed on the original signal give an indication of the underlying stochastic processes that are summed up to build up a composite process. A possible model is a mixture of three different first-order GM processes $x_{t+1} = e^{-\beta\Delta t}x_t + w_t, \quad w_t \sim \mathcal{N}(0, q)$, with $q = \sigma^2_{\text{GM}}(1 - e^{-2\beta\Delta t})$ and $\Delta t = 0.01$ s, the sampling interval. In such a case, the parameters to be estimated are $\boldsymbol{\theta} = \{\sigma_1^2, \beta_1, \sigma_2^2, \beta_2, \sigma_3^2, \beta_3\}$ and the GMWM is identifiable since a first order-Markov random process is a one-to-one reparameterization of an AR(1) process whose sum allows identifiable GMWM (see Corollaries 2 and 3). The GMWM estimates of the parameter set $\boldsymbol{\theta}$ and its corresponding 95% confidence intervals using the WV covariances estimated by means of a parametric bootstrap (400 samples) are given in Table 2. The suitability of the estimated model for the data at hand can be judged graphically by a matching of the empirical WV and the parametric WV using the estimates in Table 2, as it is done in Figure 3 (lower panel). It can be seen that the estimated process matches the observed one very well across WV scales, yielding an indication of a suitable fit. Moreover, one can also compute the goodness of fit test statistic in Equation (14) for competing models and test to what extent the chosen one yields a suitable fit. For that, we also considered a single GM process and the sum of two GM processes. The test statistics and corresponding p-values are 384363.9 (on 17 df) and p-value $< 10^{-5}$ for the single GM process, 84787.01 (on 15 df) and p-value $< 10^{-5}$ for a sum of two GM processes, and 16.81637 (on 13 df) and p-value $= 0.208$ for a sum of three GM processes, respectively, clearly indicating that the last model is suitable. The corresponding estimated WV are drawn in Figure 3 (lower panel).

It should be noted that the sum of three first-order GM random processes can be reparameterized as an ARMA(3,2) process (see, e.g., Granger and Morris 1976; Terasvirta 1977), so that one could, in principle, estimate the latter instead of the

Table 2. Estimated parameters with associated 95% confidence intervals for the mixture of three first-order GM random processes with the Gyroscope signal data

	Estimates	IC(·,0.95)
β_1	$2.1720 \cdot 10^2$	$(2.1465 \cdot 10^2 \,;\, 2.1906 \cdot 10^2)$
σ_1	$7.4521 \cdot 10^{-3}$	$(7.4477 \cdot 10^{-3} \,;\, 7.4688 \cdot 10^{-3})$
β_2	$6.0693 \cdot 10^{-1}$	$(2.7890 \cdot 10^{-1} \,;\, 7.8655 \cdot 10^{-1})$
σ_2	$2.9691 \cdot 10^{-4}$	$(2.9494 \cdot 10^{-4} \,;\, 2.9870 \cdot 10^{-4})$
β_3	$3.5563 \cdot 10^{-3}$	$(1.5333 \cdot 10^{-3} \,;\, 3.9086 \cdot 10^{-3})$
σ_3	$5.5127 \cdot 10^{-4}$	$(5.4038 \cdot 10^{-4} \,;\, 5.6223 \cdot 10^{-4})$

former. However, one should then be able to invert the estimated ARMA since very often, and in particular in the example at hand, a sum of first-order GM random processes is the one that we believe explains the real underlying process, which for gyroscopes reflect the short- and long-term correlated noise due to quantization and mechanical thermal noise phenomena present in MEMS sensors (see Kittel 1958; Drexler 1992). To recover the sum of GM processes' parameters from an estimated ARMA process, and also their standard errors, several conditions need to be satisfied. First, the roots of the processes must lie outside the unit circle. Second, the Jacobian matrix of the transformation between the two parameterizations must be invertible to apply the delta method (Rao 1973; Benichou and Gail 1989). With the inertial sensor data, the estimated processes have roots that are near the unit circle, and moreover, the Jacobian matrix of the transformation evaluated at the estimated parameters is not invertible. In that case at least, estimating an ARMA process and converting the estimated model and inference to the sum of first-order GM processes is infeasible.

To illustrate the impact of the importance of model structure and estimation precision in device positioning, we also performed an emulation study that was presented in Stebler et al. (2012). Using extremely reliable equipments, the trajectory of a helicopter performing airborne laser scanning can be precisely determined. From this trajectory, we computed the theoretical perfect inertial measurements (i.e., accelerations and angular rotations) that should be observed along this trajectory. These perfect measurements were then corrupted with a real static error signal acquired such as the one analyzed above. Then, these "pseudo" measurements were used together with the real GPS measurements to compute a trajectory. At one point, we introduced relatively short artificial gaps in GPS observations of 1 min duration (note that, in practice, such gaps are very common) and compared the deviations of the trajectories obtained with different inertial error models estimated from the data. The two first estimated models are based on the AV and on the KF-Self-Tuning approach. A detailed discussion on these methods can be found for example in Stebler et al. (2012). The GMWM was also used as an alternative estimator and model building approach. The trajectories are depicted in Figure 2 (adapted from Stebler et al. 2012) along with the "true" trajectory. It can be seen that the GMWM-based model (black-dashed line) limits significantly the error growth during the GPS-signal outage compared with the other two benchmark methods which diverged from the "true" trajectory by several thousand of meters! The poor performances of the standard methods explains the recent explosion of the research conducted to determine the stochastic modeling of MEMS-type inertial sensors.

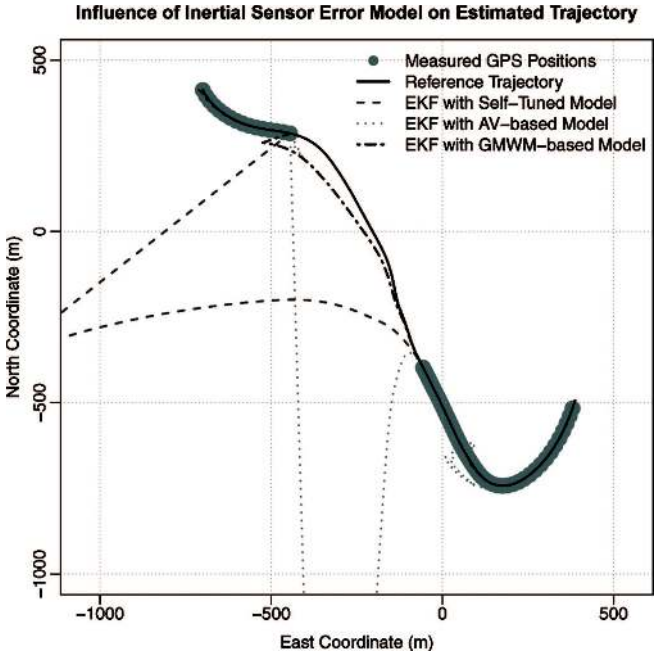

Figure 2. Comparison between a reference trajectory (black dotted line) issued from a mapping flight in which a GPS outage was introduced, with estimated error model based on the AV (light-gray line), the KF-Self-Tunning (dark gray line), and the GMWM (black dashed line).

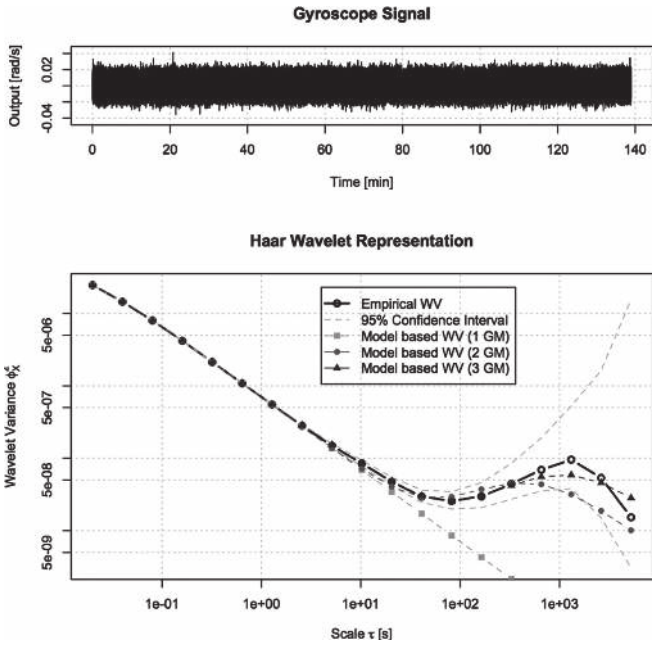

Figure 3. Gyroscope-observed error process (top panel) and graphical comparison (log–log scale) between the Haar WV (line "o") computed from the observed signal and the analytical signal using the estimated parameters of, respectively, the sum of three GM processes (line "▲"), the sum of two GM processes (line ●), and one GM process (line■).

6. CONCLUDING REMARKS

In this article, we present a new estimator for the parameters of composite stochastic processes which is shown to be consistent for the class made of the sum of independent white-noise, drift, quantization noise, random walk, and $k < \infty$ AR(1). As demonstrated in Stebler et al. (2012), the GMWM has many advantages over existing alternative methods for applications in engineering or natural sciences. The enlargement of the class of models for which the GMWM is consistent involves verifying the conditions provided in Theorem 2 and more generally in Theorem 1 (with Corollary 1) and is left for future research. Also, one could use in principle another consistent estimator for

4C/Art

the Haar WV, such as the one proposed in Nason, von Sachs, and Kroisandt (2000).

As an anonymous referee pointed out, one could wonder why to use wavelets instead of making inference directly on the spectral density for which a similar estimation procedure could be used. We believe that our approach is more suitable for the following reasons. First, inference on the PSD would make the optimization of a least-squares-type measure (between the empirical and model-based PSD) more difficult to solve when the PSD has large variability over very narrow frequency bands. As shown in Percival and Walden (2000), the wavelet coefficients at scale τ_j are associated with frequencies in the interval $[1/2^{j+1}, 1/2^j]$ and relation (5) can be approximated by

$$\nu_Y^2(\tau_j) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S_Y(f) df. \quad (16)$$

This means that the WV summarizes the information in the PSD using just one value per octave frequency band. Thus, it is particularly useful when the PSD is relatively featureless within each octave band. In the case of the widely used pure power law processes ($S_Y(f) \propto |f|^\alpha$), for example from Equation (16) one gets $\nu_Y^2(\tau_j) \propto \tau_j^{-\alpha-1}$, meaning that no information is lost in using the PSD summary given by the WV. Second, the computation of empirical WV is more straightforward than nonparametric PSD; for example, the periodogram is an inconsistent estimator of $S_Y(f)$ and can be badly biased even for large samples sizes (because of frequency leakage effects) so that more sophisticated PSD estimators and/or smoothing techniques such as prewhitening or tapering should be used. Third, the PSD of two important processes in sensor error models, namely the drift and the random walk, cannot be distinguished (both have slope of -2 in a log–log representation of the PSD). Finally, the MODWT on which the WV computation is based requires a number of multiplications of order $T \log_2 T$, which is the same order as the widely used fast Fourier transform algorithm, so that the use of WV does not increase the computational burden.

SUPPLEMENTARY MATERIALS

NOTES

A C++ program for computing the GMWM for the models treated in this article is available upon request from the authors.

[Received March 2012. Revised March 2013.]

REFERENCES

Allan, D. (1966), "Statistics of Atomic Frequency Standards," in *Proceedings of the IEEE*, 54, 221–230. [3]

——— (1987), "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 34, 647–654. [3]

Andrews, D. W. (1993), "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61, 139–165. [6]

Benichou, J., and Gail, M. H. (1989), "A Delta Method for Implicitly Defined Random Variables," *The American Statistician*, 43, 41–44. [8]

Bowling, D. R., Sargent, S. D., Tanner, B. D., and Ehleringer, J. R. (2003), "Tunable Diode Laser Absorption Spectroscopy for Stable Isotope Studies of Ecosystem–Atmosphere CO2 Exchange," *Agricultural and Forest Meteorology*, 118, 1–19. [3]

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia: SIAM. [2,3]

Dempster, A. P., Laird, N. M., and Rubing, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society,* Series B, 39, 1–38. [1]

Donoho, D. L. (1995), Nonlinear Solution of Linear Inverse Problems by Wavelet-Vaguelette Decomposition, *Applied and Computational Harmonic Analysis*, 2, 101–126. [2]

Donoho, D. L., and Johnstone, I. M. (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, 81, 425–455. [2]

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995), "Wavelet Shrinkage: Asymptopia?," *Journal of the Royal Statistical Society,* Series B, 57, 301–369. [2]

——— (1996), "Density Estimation by Wavelet Thresholding," *The Annals of Statistics*, 24, 508–539. [2]

Doukhan, P., and Leon, J. (1990), "Déviation Quadratique D'Estimateur De Densité par Projection Orthogonale," *Comptes Rendus de l'Académie des Sciences, Paris (A)*, 310, 424–430. [2]

Drexler, K. E. (1992), *Nanosystems—Molecular Machinery, Manufacturing, and Computation*, New York: Wiley. [8]

Durbin, J., and Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford Statistical Series 24. Oxford, UK: Oxford University Press. [1]

——— (2002), "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89, 603–615. [1]

El-Sheimy, N., and Niu, X. (2007), "The Promise of MEMS to the Navigation Community," *Inside GNSS*, 2, 46–56. [7]

El-Sheimy, N., Hou, H., and Niu, X. (2008), "Analysis and Modeling of Inertial Sensors Using Allan Variance," *IEEE Transactions on Instrumentation and Measurement*, 57, 140–149. [3]

Fadel, P. J., Orer, H. S., Barman, S. M., Vongpatanasin, W., Victor, R. G., and Gebber, G. L. (2004), "Fractal Properties of Human Muscle Sympathetic Nerve Activity," *American Journal of Physiology—Heart and Circulatory Physiology*, 286, H1076. [3]

Fukuda, K., Tachikawa, M., and Kinoshita, M. (2003), "Allan Variance Measurements of Diode Laser Frequency-Stabilized With a Thin Vapor Cell," *Applied Physics B: Lasers and Optics*, 77, 823–827. [3]

Gallant, A. R., and Tauchen, G. (1996), "Which Moments to Match?," *Econometric Theory*, 12, 657–681. [2,4]

Gambis, D. (2002), "Allan Variance in Earth Rotation Time Series Analysis," *Advances in Space Research*, 30, 207–212. [3]

Gebber, G. L., Orer, H. S., and Barman, S. M. (2006), "Fractal Noises and Motions in Time Series of Presympathetic and Sympathetic Neural Activities," *Journal of Neurophysiology*, 95, 1176–1184. [3]

Genton, M. G., and Ronchetti, E. (2003), "Robust Indirect Inference," *Journal of the American Statistical Association*, 98, 67–76. [5]

Giraitis, L., and Taqqu, M. S. (1997), "Limit Theorems for Bivariate Appell Polynomials. Part i: Central Limit Theorems," *Probability Theory and Related Fields*, 107, 359–381. [3]

Gourieroux, C., Monfort, A., and Renault, E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118. [2,4]

Granger, C. W., and Morris, M. J. (1976), "Time Series Modelling and Interpretation," *Journal of the Royal Statistical Society,* Series A, 139, 246–257. [7]

Greenhall, C. A. (1991), "Recipes for Degrees of Freedom of Frequency Stability Estimators," in *IEEE Transactions on Instrumentation and Measurements* (vol. 40), IEEE, pp. 994–999. [1,2]

Guerrier, S. (September 22–25, 2009), "Improving Accuracy With Multiple Sensors: Study of Redundant MEMS-IMU/GPS Configurations," In *22nd International Meeting of the Satellite Division of The Institute of Navigation*, Savannnah, GA, USA: The Institute of Navigation, pp. 3114–3121. [3]

Guerrier, S., Stebler, Y., Skaloud, J., and Victoria-Feser, M.-P. (2013), "Limits of the Allan Variance and Optimal Tuning of Wavelet Variance Based Estimators," Working Paper, Submitted. [4]

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica: Journal of the Econometric Society*, 50, 1029–1054. [2,4]

Harris, D., and Mátyás, L. (1999), "Introduction to the Generalized Method of Moments Estimation," in *Generalized Method of Moments Estimation*, Themes in Modern Econometrics, László Mátyás, ed., Cambridge University Press, pp. 3–30. [3]

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press. [1]

Hinrichsen, R. A., and Holmes, E. E. (2009), *Spatial Ecology Chapter, Using Multivariate State-Space Models to Study Spatial Structure and Dynamics,* Mathematical and Computational Biology Series, Boca Raton, FL: CRC/Chapman & Hall, pp. 145–164. [6]

Holmes, E. E. (2010), "Derivation of the EM Algorithm for Constrained and Unconstrained Multivariate Autoregressive State-Space (MARSS) Models," Technical Report, Northwest Fisheries Science Center, Seattle, WA: NOAA Fisheries. [5]

IEEE Standard 1293 (1998), "IEEE Standard Specification Format Guide and Test Procedure for Linear, Single-Axis, Non-Gyroscopic Accelerometers," Technical Report, IEEE. [3,7]

Johnstone, I. M., and Silverman, B. W. (1997), "Wavelet Threshold Estimators for Data With Correlated Noise," [2]

Kebabian, P. L., Herndon, S. C., and Freedman, A. (2005), "Detection of Nitrogen Dioxide by Cavity Attenuated Phase Shift Spectroscopy," *Analytical Chemistry*, 77, 724–728. [3]

Kerkyacharian, G., and Picard, D. (1992), "Density Estimation in Besov Spaces," *Statistics and Probability Letters*, 13, 15–24. [2]

Kittel, C. (1958), *Elementary Statistical Physics*, New York: Wiley. [8]

Komunjer, I. (2012), "Global Identification in Nonlinear Models With Moment Restrictions," *Econometric Theory*. Forthcoming. [3]

Loescher, H. W., Ocheltree, T., Tanner, B., Swiatek, E., Dano, B., Wong, J., Zimmerman, G., Campbell, J., Stock, C., and Jacobsen, L. (2005), "Comparison of Temperature and Wind Statistics in Contrasting Environments Among Different Sonic Anemometer-Thermometers," *Agricultural and Forest Meteorology*, 133, 119–139. [3]

Mallat, S. (1999), *A Wavelet Tour of Signal Processing*, New York: Academic Press. [2]

McCausland, W. J., Miller, S., and Pelletier, D. (2011), "Simulation Smoothing for State Space Models: A Computational Efficiency Analysis," *Computational Statistics and Data Analysis*, 55, 199–212. [1]

Nason, G. P., von Sachs, R., and Kroisandt, G. (2000), "Wavelet Processes and Adaptive of the Evolutionary Wavelet Spectrum," *Journal of the Royal Statistical Society,* Series B, 62, 271–292. [9]

Niu, X., Nasser, S., Goodall, C., and El-Sheimy, N. (2007), "A Universal Approach for Processing any MEMS Inertial Sensor Configuration for Land-Vehicle Navigation," *Journal of Navigation*, 60, 233–246. [7]

Percival, D. B. (1995), "On Estimation of the Wavelet Variance," *Biometrika*, 82, 619–631. [2]

Percival, D. B., and Guttorp, P. (1994), "Long-Memory Processes, the Allan Variance and Wavelets," *Wavelets in Geophysics*, 4, 325–344. [1,2,3]

Percival, D. B., and Walden, A. T. (2000), *Wavelet Methods for Time Series Analysis*, Cambridge Series in Statistical and Probabilistic Mathematics. New York: Cambridge University Press. [2,3,9]

Rao, C. R. (1973), *Linear Statistical Inferentcea and Its Applications* (2nd ed.), New York: Wiley. [8]

Schieder, R., and Kramer, C. (2001), "Optimization of Heterodyne Observations Using Allan Variance Measurements," *Astronomy and Astrophysics*, 373, 746–756. [3]

Serroukh, A., and Walden, A. T. (2000a), "Wavelet Scale Analysis of Bivariate Time Series I: Motivation and Estimation," *Journal of Nonparametric Statistics*, 13, 1–36. [3]

——— (2000b), "Wavelet Scale Analysis of Bivariate Time Series II: Statistical Properties for Linear Processes," *Journal of Nonparametric Statistics*, 13, 37–56. [3]

Serroukh, A., Walden, A. T., and Percival, D. B. (2000), "Statistical Properties and Uses of the Wavelet Variance Estimator for the Scale Analysis of Time Series," *Journal of the American Statistical Association*, 95, 184–196. [2,3]

Shumway, R. H., and Stoffer, D. S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3, 253–264. [1,5]

Skrínskı , J., Janecková, R., Grigorová, E., Strizík, M., Kubát, L., Herecová, P., Nevrlı , V., Zelinger, Z., and Civis, S. (2009), "Allan Variance for Optimal Signal Averaging-Monitoring by Diode-Laser and CO2 Laser Photo-Acoustic Spectroscopy," *Journal of Molecular Spectroscopy*, 256, 99–101. [3]

Smith, A. A. (1993), "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84. [2,4]

Stebler, Y., Guerrier, S., Skaloud, J., and Victoria-Feser, M.-P. (2011), "Constrained Expectation-Maximization Algorithm for Stochastic Inertial Error Modeling: Study of Feasibility," *Measurement Science and Technology*, 22. [3,6,7]

——— (2012), "A Framework for Inertial Sensor Calibration Using Complex Stochastic Error Models," in *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*, IEEE, pp 849–861. [7,8]

Terasvirta, T. (1977), "The Invertibility of Sums of Discrete MA and ARMA Processes," *Scandinavian Journal of Statistics*, 4, 165–170. [7]

Von Sachs, R., and Macgibbon, B. (2000), "Non-Parametric Curve Estimation by Wavelet Thresholding With Locally Stationary Errors. [2]

Waegli, A. (2009), *Trajectory Determinationand Analysis in Sports by Satellite and Inertial Navigation*, PhD thesis, Switzerland: Ecole Polytechnique Fédérale de Lausanne, Available at *http://library.epfl.ch/theses/?nr=4288*, *http://library.epfl.ch/theses/?nr=4288*. [7]

Walter, G. G. (1992), "Approximation of Delta Function by Wavelets," *Journal of Approximation Theory*, 71, 329–343. [2]

Werle, P., Mücke, R., and Slemr, F. (1993), "The Limits of Signal Averaging in Atmospheric Trace-Gas Monitoring by Tunable Diode-Laser Absorption Spectroscopy (TDLAS)," *Applied Physics B: Lasers and Optics*, 57, 131–139, 1993. [3]

Whitcher, B. (2004), "Wavelet-Based Estimation for Seasonal Long-Memory Processes," *Technometrics*, 46, 225–238. [3]

Zhang, N. F. (2008), "Allan Variance of Time Series Models for Measurement Data," *Metrologia*, 45, 549–561. [2]