The GMWM: A New Framework for Inertial Sensor Calibration

Roberto Molinari

Research Center for Statistics (GSEM)
University of Geneva

joint work with
S. Guerrier (UIUC), J. Balamuta (UIUC),
J. Skaloud (EPFL), Y. Stebler (u-blox),
& M.-P. Victoria-Feser (U. Geneva)

February 10, 2016

Introduction

General Framework:

- We propose **a new framework for inertial sensor calibration.**
- It is based on the **Generalized Method of Wavelet Moments (GMWM)** of *Guerrier et al, JASA, 2013* which is a new statistical approach to estimate the parameters of (complex) time series models.
- The GMWM is able to estimate efficiently time series models which are commonly used to describe the errors of inertial sensors.
- This calibration approach provides considerable improvements (terms of navigation performance) compared to existing methods.
- This methodology is **robust** (potentially applicable for FDI purposes) and is able to **automatically select a suitable model**.

IMU Calibration

Errors in Inertial Sensors

- Possible causes:
  - Non-orthogonalities of the sensor axes
  - Environmental conditions (e.g. temperature)
  - Electronics
  - Dynamics
  - ...
- Error types:
  - Deterministic (calibration models, physical models, ...)
  - **Random error components** (typically latent time series models,...)

Correct stochastic sensor error modeling implies:

- Correct stochastic assumptions for inference
- **Better navigation or post-processing performance**

Effect on position of error model

Emulation setting:

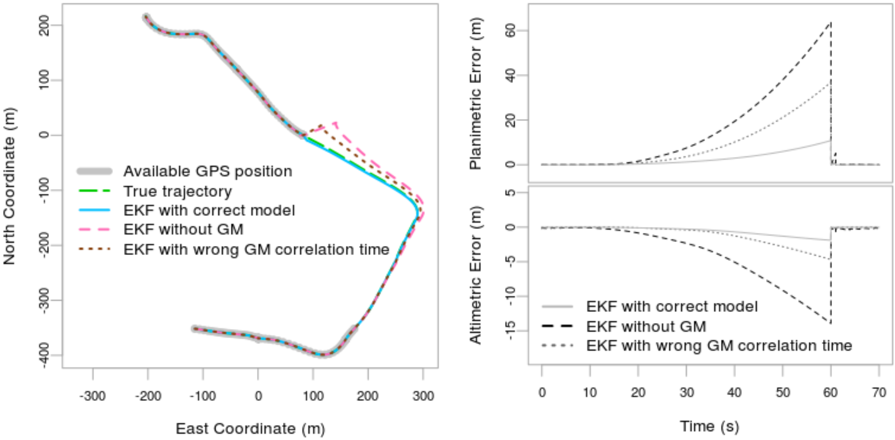
- Suppose the following model for inertial sensors:

$$Y_t = \exp(-\beta\Delta t)Y_{t-1} + \varepsilon_t,\ \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{GM}(1 - \exp(-2\beta\Delta t)))$$
$$X_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2),\ Z_t = X_t + Y_t.$$

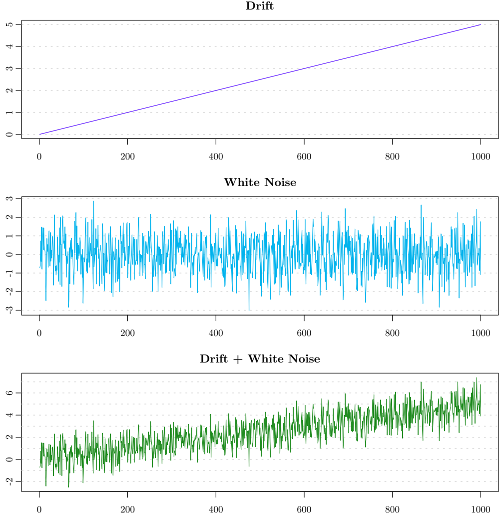
- Consider the following three models:

Sensor	Scenario	β	σ_{GM}	σ_{WN}
Acc.	Correct model	10^{-4}	50.0	70.0
	Wrong β	10^{-2}	50.0	70.0
	Without GM	–	–	70.0
Gyro.	Correct model	10^{-4}	10.0	30.0
	Wrong β	10^{-2}	10.0	30.0
	Without GM	–	–	30.0

Effect on position of error model



An easy latent time series model



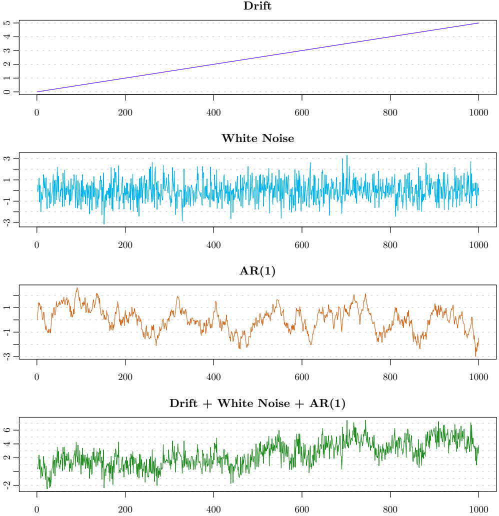
Remarks:

- Simple linear regression model:

$$y_t = \omega t + \varepsilon_t$$
$$\varepsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

- MLE is perfectly fine.
- **What if we add an AR(1) process?**

Adding an autoregressive process



Remarks:

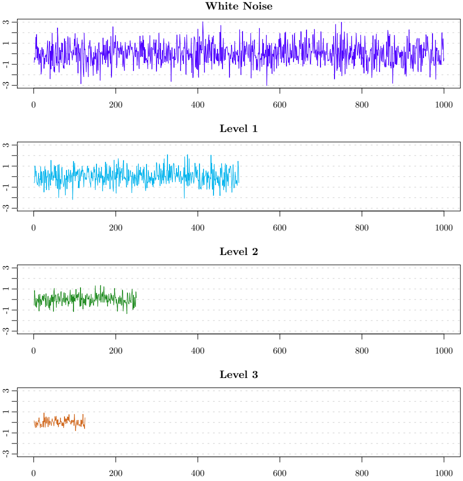
- Not a linear regression model but a **state space model**.
- Computing the likelihood is not an easy task (Kalman filter).
- **MLE (in fact EM-KF) fails**.

Estimation of Latent Time Series Models

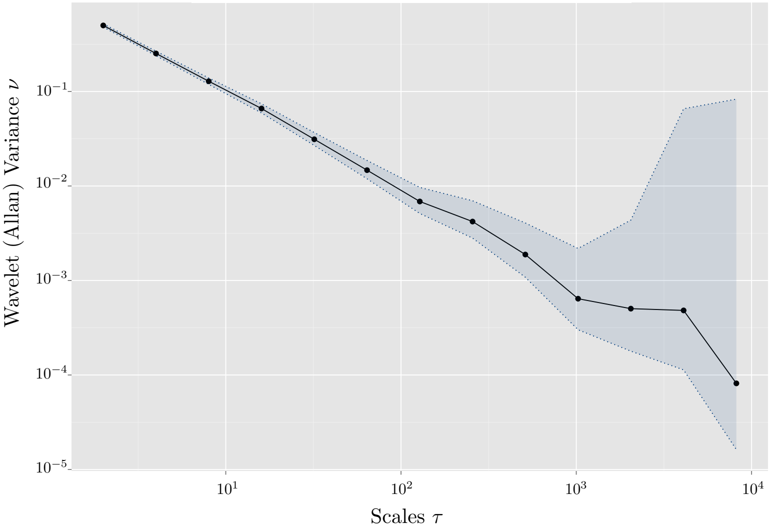
Existing methods:

- Transforming into a "non-latent" model (e.g. ARMA)
  - Does not work in general.
  - Tends to diverge when one latent time series is "close" to unit root.
  - Difficult to "inverse".
- MLE of an associated state-space (possibly using EM algorithm)
  - Computationally intensive.
  - Systematically diverges with "complex" models.
  - A lot of work is needed for a new model (see *Stebler et al, Meas Sci Tech, 2012*).
- "Graphical" method
  - Limited to a few possible models.
  - Not consistent in general (see *Guerrier et al, 2013b*).
  - "Inefficient" (see *Guerrier et al, 2013b*).

Looking differently at a time series using the Wavelet Variance



The Wavelet Variance



Initial idea:

Match the WV:

- Exploit the relationship that exists between the model $F_{\boldsymbol{\theta}}$ and the WV $\boldsymbol{\nu}(\boldsymbol{\theta})$ (i.e. **mapping $\boldsymbol{\theta} \mapsto \boldsymbol{\nu}(\boldsymbol{\theta})$**).
- "Inverse" this mapping by minimizing some discrepancy between empirical WV ($\hat{\boldsymbol{\nu}}$) and the theoretical WV $\boldsymbol{\nu}(\boldsymbol{\theta})$.
- This should provide an approximation of the point $\boldsymbol{\theta}(\hat{\boldsymbol{\nu}})$.

Wavelet Variance

Empirical WV:

- The WV (ν_{τ_j}) is the **variance of wavelet coefficients** for the scale τ_j.
- Wavelet coefficients ($\overline{W}_{j,t}$) are weighted averages computed on the series Y_t.
- The weights are called wavelet filters h_j: e.g. the Haar wavelet filter.
- The wavelet filters give non-zero weights to observations at a given lag (windows size of length L_j). Hence, there are as many WV as there are scales.
- The wavelet filters can be computed on consecutive windows, or on overlapping windows (to get $\widetilde{W}_{j,t}$ using $\tilde{h}_j$). Overlapping windows lead to more efficient estimators (such as the MODWT).

Wavelet Variance

Definition:

The Wavelet Variance (WV) is the variance of the wavelet coefficients, i.e.

$$\nu_{\tau_j} = \mathrm{var}\left(\widetilde{W}_{j,t}\right), \quad \text{where} \quad \widetilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} y_{t-l}, \; t \in \mathbb{Z}$$

and where $\tilde{h}_{j,l}$ are wavelet filters based on MODWT.

Remark:

The Allan variance is a special case of the WV, in fact $\sigma_{\bar{y}}(\tau) = 2\nu_\tau$ where ν_τ is based on Haar wavelet.

Estimation of the WV

MODWT estimator:

A consistent estimator for ν_{τ_j} is given by the MODWT estimator defined in *Percival, Biometrika, 1995*

$$\hat{\nu}(\tau_j) = \frac{1}{M(T_j)} \sum_{t \in T_j} \widetilde{W}_{j,t}^2$$

Theorem: Asymptotic Normality

Serroukh et al, JASA, 2000 show that under suitable conditions

$$\sqrt{M(T_j)}\left(\hat{\nu}(\tau_j) - \nu_{\tau_j}\right) \xmapsto[T \to \infty]{\mathcal{D}} \mathcal{N}\left(0, S_{W_j}(0)\right)$$

A more general theorem

Theorem: Multivariate Extension

We extended this result to the multivariate case and demonstrated that under some regularity conditions

$$\sqrt{T}\left(\hat{\boldsymbol{\nu}} - \mathbb{E}[\hat{\boldsymbol{\nu}}]\right) \xmapsto[T\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \mathbf{V}\right)$$

where $\mathbf{V} = [\sigma^2_{kl}]_{k,l=1,\dots,J}$.

Remark:

- This theorem generalizes *Serroukh et al, JASA, 2000* result and enables to compute the (asymptotic) covariance between the WV (or the AV) at two different scales.
- We proposed an estimator for σ^2_{kl}.

Theoretical WV

WV implied by $F_{\boldsymbol{\theta}}$:

Given a model $F_{\boldsymbol{\theta}}$ one can compute the theoretical WV as:

$$\nu_{\tau_j} = f\left(\boldsymbol{\theta}\right) = \int_{-1/2}^{1/2} |\widetilde{H}_j(f)|^2 S_{F_{\boldsymbol{\theta}}}(f) df$$

Example:

Consider an AR(1):

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \stackrel{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \;\; t = 1, ..., T$$

The theoretical (haar) WV of such process is given by

$$\nu_{\tau_j} = \frac{\left(\frac{\tau_j}{2} - 3\phi_1 - \frac{\tau_j \phi_1^2}{2} + 4\phi_1^{\frac{\tau_j}{2}+1} - \phi_1^{\tau_j+1}\right)\sigma_\varepsilon^2}{\frac{\tau_j^2}{2}\left(1-\phi_1\right)^2\left(1-\phi_1^2\right)}$$

WV of latent time series models

A very useful property:

Suppose we have

$$Y_t = X_t^{(1)} + ... + X_t^{(k)}$$

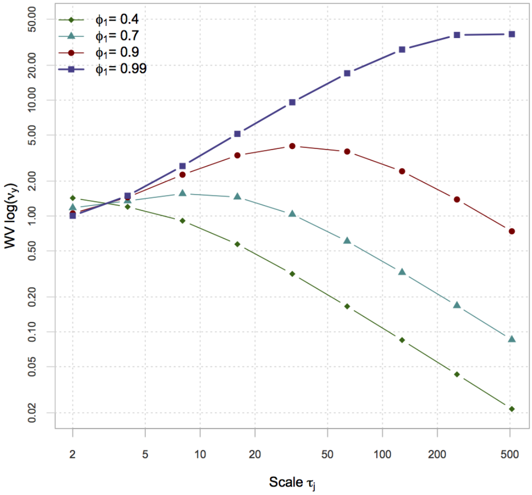
then the PSD of Y_t is

$$S_{Y_t} = S_{X_t^{(1)}} + ... + S_{X_t^{(k)}}$$

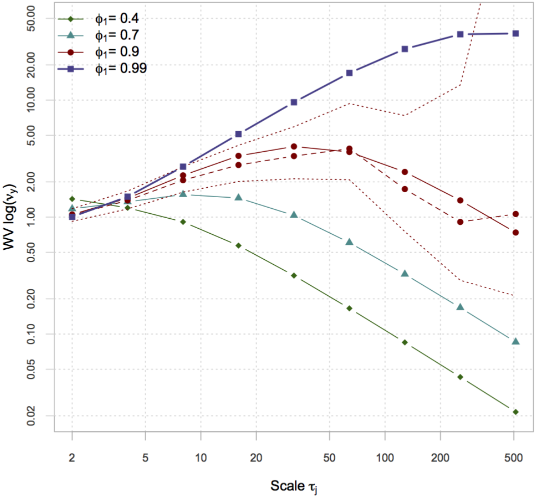
so the WV of Y_t is given by

$$\nu_{Y_t,\tau_j} = \int_{-1/2}^{1/2} |\widetilde{H}_j(f)|^2 \left(\sum_{i=1}^{k} S_{X_t^{(i)}} \right) df = \sum_{i=1}^{k} \nu_{X_t^{(i)},\tau_j}$$

Principle of the GMWM



Principle of the GMWM



The GMWM estimator

Definition:

The GMWM estimator is the solution of the following optimization problem

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \; \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)^T \boldsymbol{\Omega} \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)$$

in which $\boldsymbol{\Omega}$, a positive definite weighting matrix, is chosen in a suitable manner such that the above quadratic form is convex.

Theorem: Consistency

$\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$ (under some regularity conditions) for a large class of (latent) models.

Asymptotic distribution of $\hat{\boldsymbol{\theta}}$

Theorem: Asymptotic Normality

We showed that (under some regularity conditions)

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \boldsymbol{\Sigma}\right)$$

where $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{V}\mathbf{B}^T$, $\mathbf{B} = \left(\mathbf{D}^T\boldsymbol{\Omega}\mathbf{D}\right)^{-1}\mathbf{D}^T\boldsymbol{\Omega}$, $\mathbf{D} = \partial\boldsymbol{\nu}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T$ and $\mathbf{V}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\nu}}$.

Choosing $\boldsymbol{\Omega}$:

The most efficient estimator is (asymptotically) obtained by choosing $\boldsymbol{\Omega} = \mathbf{V}^{-1}$, leading then to $\boldsymbol{\Sigma} = (\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}$.

A small Example...

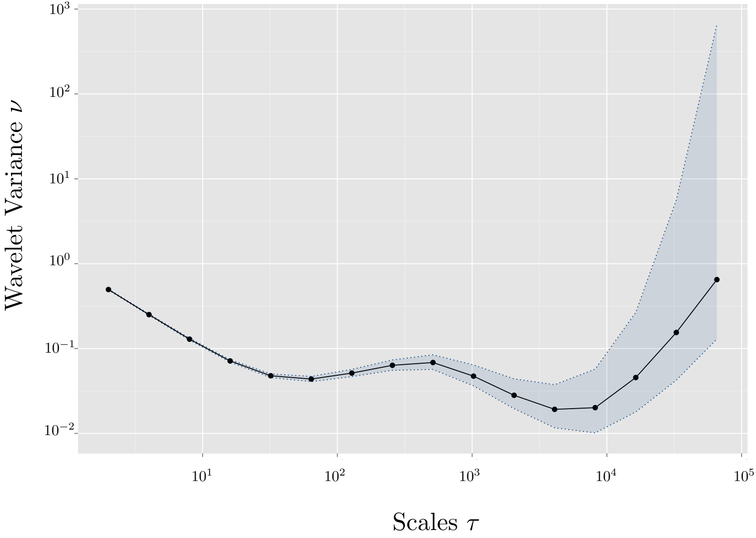
A simulated example:

Let $(y_t) \ : t = 1, ..., 10^5$ be a simulated signal composed of a:

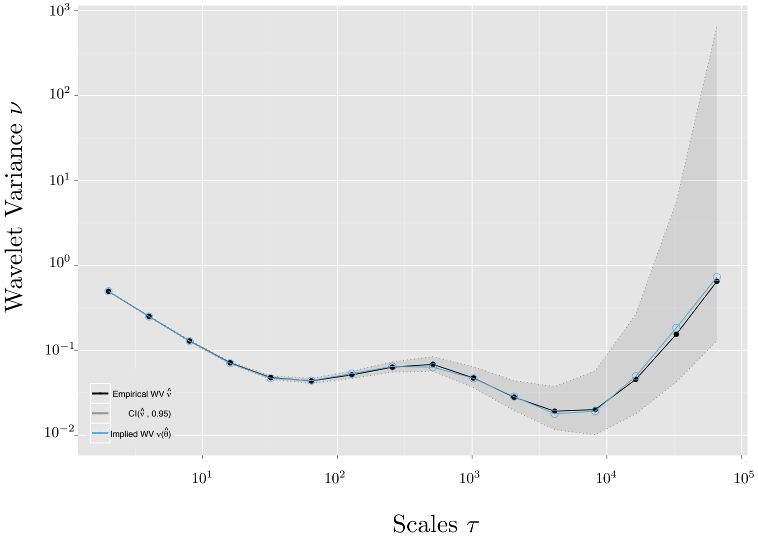
- First-order Gauss-Markov:
  $Y_t = \exp(-\beta \Delta t) Y_{t-1} + \varepsilon_t, \ \ \varepsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2_{GM}\left(1 - \exp(-2\beta \Delta t)\right)\right)$
- Gaussian White Noise: $X_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2_{WN}\right)$
- Rate ramp (drift): $R_t = \omega t$

The observed process is therefore $Z_t = Y_t + X_t + R_t$ and we have that $Z_t \sim F_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta} = (\sigma_{WN}, \sigma_{GM}, \beta, \omega)$

A small Example...



GMWM estimation results



GMWM estimation results

Estimated parameters:

	θ_0	$\hat{\theta}$	$\text{IC}(\theta_0, 0.95)$
σ^2_{WN}	1.00	1.00	$(0.99; 1.01)$
σ^2_{GM}	0.60	0.58	$(0.55; 0.61)$
β	10^{-2}	$1.07 \cdot 10^{-2}$	$(0.99 \cdot 10^{-2}; 1.12 \cdot 10^{-2})$
ω	$5 \cdot 10^{-5}$	$4.87 \cdot 10^{-5}$	$(4.67 \cdot 10^{-5}; 5.07 \cdot 10^{-5})$

Selecting the best model(s)

Goodness-of-Fit (GoF) test

The GMWM properties allow to perform a test for over-identifying restrictions (also known as "J-test" or "Goodness-of-Fit" (GoF)) based on the objective function $g(\boldsymbol{\theta}) \equiv (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}))^T \boldsymbol{\Omega} (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}))$.

The hypotheses tested are

$$H_0 : g(\boldsymbol{\theta}) = 0$$
$$H_1 : g(\boldsymbol{\theta}) \neq 0$$

Distribution of $g(\hat{\boldsymbol{\theta}})$

Under H_0, $g(\hat{\boldsymbol{\theta}}) \sim \chi^2_{J-p}$

Selecting the best model(s)

Wavelet Information Criterion (WIC)

To jointly evaluate the descriptive and predictive capacity of a model, *Guerrier et al, 2015* proposed the Wavelet Information Criterion (WIC):

$$WIC = \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right)^T \boldsymbol{\Omega} \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right) + 2\,\mathrm{tr}\left[\widehat{\mathrm{cov}}\left[\hat{\boldsymbol{\nu}}, \boldsymbol{\Omega}\boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right]^T\right].$$

The model(s) which minimize(s) this criterion can be considered as the best.

Extensions and Developments

The GMWM has many extensions:

- A **robust version** of the GMWM has been developed to reduce the impact of outliers on the estimation procedure and provides a *basis for Fault Detection and Isolation*.
- A **multivariate version** of the GMWM is being developed to take into account the *dependence between gyros and accelerometers* from the same IMU to improve estimation and hence navigation precision.
- A **non-stationary version** of the GMWM is being developed to address the problems linked to the estimation of IMU error signals in dynamic conditions.

Software

R package

- Supports the Classical and Robust forms of GMWM.
- Ability to **automatically select models**.
- **Computationally efficient.** Indeed, it has the ability to process large time series ($n < 10$ million) in under 2 minutes.
- Computational backend written using the **Armadillo C++ Linear Algebra Library**.
- The computational code is platform independent (e.g. compatible with MATLAB's C++ API).

Conclusions

The GMWM approach...

- is a statistically rigorous approach for sensor calibration
- is able to model signals of complex spectral structure (where classical methods fail!)
- is consistent (unlike AV) and asymptotically *normally distributed*
- can be computed (in some situations) using a computationally efficient algorithm (unlike EM) even with very large sample sizes (e.g. few hours of static IMU data)
- can be used to model any type stochastic process that can be simulated (and fulfill some regularity conditions)
- **considerably increases navigation accuracy**.

Any questions before the “fun” part?

Main References on the GMWM

Methodology-related:

- Guerrier, S., Skaloud, J., Stebler, Y. and Victoria-Feser, M.P. Wavelet Variance based Estimation for Composite Stochastic Processes. Journal of the American Statistical Association, 2013
- Guerrier, S., Molinari, R. and Victoria-Feser,M.-P. Estimation of Time Series Models via Robust Wavelet Variance. Austrian Journal of Statistics, 2014
- Guerrier, S., Molinari, R., Stebler, Y., Theoretical Limitations of Allan variance-based Regression for Time Series Models Estimation. Submitted working paper, 2016
- Guerrier, S., Molinari, R. and Victoria-Feser,M.-P. Robust Inference for Time Series Models: a Wavelet-based Framework. Submitted working paper, 2015

Engineering-related:

- Stebler, Y., Guerrier, S., Skaloud, J. and Victoria-Feser, M.-P., The Generalized Method of Wavelet Moments for Inertial Navigation Filter Design, IEEE Transactions on Aerospace and Electronic Systems, 2014
- Stebler, Y., Guerrier S. and Skaloud, J., Study and Modeling of MEMS-based Inertial Sensors Operating in Dynamic Conditions, IEEE Transactions on Instrumentation and Measurement, 2015
- Stebler, Y., Guerrier, S., Skaloud, J., and Victoria-Feser, M.P.: A Framework for Inertial Sensor Calibration Using Complex Stochastic Error Models. Position Location and Navigation Symposium (PLANS), IEEE/ION (2012)
- Stebler, Y., Guerrier, S., Skaloud, J., and Victoria-Feser, M.P.: Constrained expectation-maximization algorithm for stochastic inertial error modeling: study of feasibility. Measurement Science and Technology 22(8) (2011)