Chapter 5: Extensions of the GMWM

Jan Skaloud* & Stéphane Guerrier†

*École Polytechnique Fédérale de Lausanne; †University of Geneva

This document was prepared with the help of Dr. Davide Cucci, Dr. Roberto Molinari, Dr. Samuel Oros, Dr. Mucyo Karemera, Gaetan Bakalli, Cesare Miglioli, Lionel Voirol, Haotian Xu & Yuming Zhang

Material available online: https://gmwm.netlify.com

EPFL - Winter 2020

Extension to Multi-signal

4924 IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 68, NO. 12, DECEMBER 2019

A Multisignal Wavelet Variance-Based Framework for Inertial Sensor Stochastic Error Modeling

Ahmed Radi, Gaetan Bakalli, Stéphane Guerrier, Naser El-Sheimy, Abu B. Sesay, and Roberto Molinari

***Abstract*—The calibration of low-cost inertial sensors has become increasingly important over the last couple of decades, especially when dealing with sensor stochastic errors. This procedure is commonly performed on a single error measurement from an inertial sensor taken over a certain amount of time, although it is extremely frequent for different replicates to be taken for the same sensor, thereby delivering important information which is often left unused. In order to address the latter problem, this paper presents a general wavelet variance-based framework for multisignal inertial sensor calibration, which can improve the modeling and model selection procedures of sensor stochastic errors using all replicates from a calibration procedure and allows to understand the properties, such as stationarity, of these stochastic errors. The applications using microelectromechanical system inertial measurement units confirm the importance of this new framework, and a new graphical user interface makes these tools available to the general user. The latter is developed based on an R package called mgmwm and allows the user to select a type of sensor for which different replicates are available and to easily make use of the approaches presented in this paper in order to carry out the appropriate calibration procedure.**

***Index Terms*—Allan variance (AV), generalized method of wavelet moments (GMWMs), graphical user interface (GUI), inertial sensors calibration, microelectromechanical system (MEMS) inertial measurement unit (IMU), model selection, stochastic error analysis, wavelet variance (WV).**

I. Introduction

THE stochastic error modeling for low-cost microelectromechanical system (MEMS)-based inertial sensors is usually considered to be a nontrivial task due to the complicated model structure that, in many cases, could contain different latent random processes. Various stochastic modeling techniques have been regularly used based on a considerable amount of research going from time-domain to frequency-domain approaches to characterize the random behavior of inertial sensor error signals.

The most common approach to study the properties of the errors is represented by the use of the autocorrelation function (ACF) which, for instance, was used to investigate the correlation time of consumer-grade inertial sensor errors in [1] and [2]. However, this method was shown to be limited when dealing with high dynamic range or higher order random processes [3]. Moreover, Petkov and Slavov [4] and Evans *et al.* [5] used the power spectral density (PSD), which is mathematically the Fourier transform of the ACF, for low-cost inertial sensor error modeling. Although this approach works well for high-frequency noise processes, such as white noise (WN), a major drawback consists in the lack of accuracy in identifying low-frequency noise parameters which are represented in the low-frequency part of the PSD log–log plot where information is still conveyed but with increased uncertainty.

Aside from studying the characteristics of an error signal, many methods have been used to estimate the parameters of the error models identified via the above-mentioned techniques. The main statistically sound approach for this purpose is the maximum likelihood (ML), which has been widely used for the task of inertial sensor stochastic calibration (most recently in [6]). In order to implement this method, the most common approach is through the use of the expectation–maximization algorithm employed within a Kalman filter [7] to allow the ML to deliver parameter estimation of the state-space models characterizing these errors. However, such a procedure tends to diverge when the models are even slightly complex and when the size of the error signals is considerable (as is the case for inertial sensor calibration, see [8]). As an alternative to the ML, the linear-fitting approach that makes use of the Allan variance (AV) has been widely employed to study the stochastic characteristics of inertial errors and retrieve their model parameters (see [3], [9]–[11]). Although computationally more feasible than the ML approach, this method was shown to be statistically inconsistent for most models characterizing small and low-cost MEMS-based inertial measurement units (IMUs) (see [12] and the study in [13]).

Considering the above-mentioned limitations, the generalized method of wavelet moments (GMWMs) was introduced in [14] to provide a general statistical method for the purpose of inertial sensor stochastic calibration. This method allows to overcome the limitations of the previously existing techniques by delivering a statistically sound and computationally

Manuscript received August 16, 2018; revised January 17, 2019; accepted January 24, 2019. Date of publication March 15, 2019; date of current version November 8, 2019. The Associate Editor coordinating the review process was George Xiao. *(Corresponding author: Ahmed Radi.)*

A. Radi and N. El-Sheimy are with the Department of Geomatics Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: ahmed.elboraee@ucalgary.ca; elsheimy@ucalgary.ca).

G. Bakalli is with the Research Center for Statistics, University of Geneva, 1211 Geneva, Switzerland (e-mail: gaetan.bakalli@unige.ch).

S. Guerrier is with the Geneva School Economics and Management, Faculty of Sciences, University of Geneva, 1211 Geneva, Switzerland (e-mail: stephane.guerrier@unige.ch).

A. B. Sesay is with the Department of Electrical Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: sesay@ucalgary.ca).

R. Molinari is with the Department of Statistics, Pennsylvania State University, State College, PA 16802 USA (e-mail: roberto.molinari@hotmail.it).

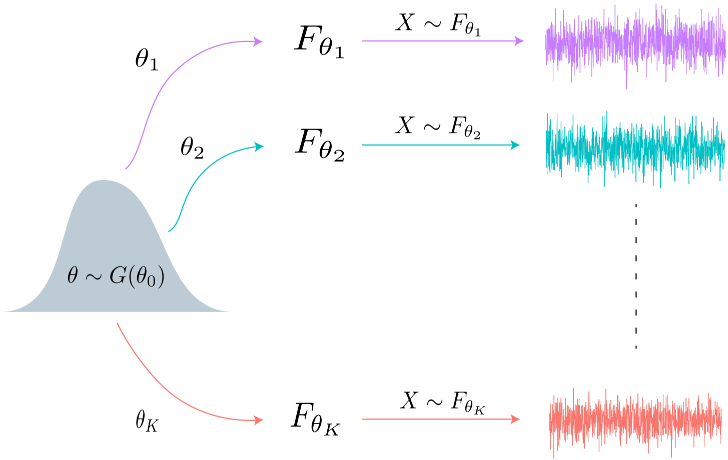
Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2019.2899535

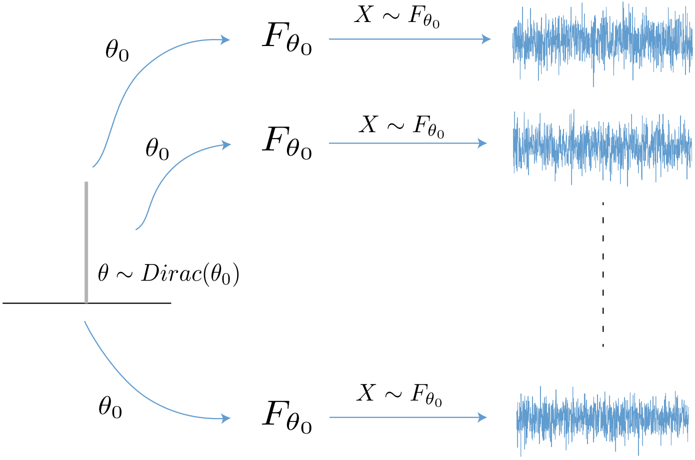
Dr. Ahmed Radi
U. Calgary

Gaetan Bakalli
U. Geneva

Near Stationarity Setting



Stationarity Setting



GMWM with multiple replicates

MuGMWM:

Considering $k = 1, \dots, K$, the number of replicates recorded from the same IMU in static condition, we define the **Multisignal GMWM** estimator as the solution of the following minimization problem:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \, \frac{1}{K} \sum_{k=1}^{K} \|\hat{\boldsymbol{\nu}}_k - \boldsymbol{\nu}(\boldsymbol{\theta})\|^2_{\boldsymbol{\Omega}}. \tag{3.1}$$

Let $\hat{\nu}_{jk}$ be the j^{th} element of the vectors $\hat{\boldsymbol{\nu}}_k$, the empirical WV computed on the k^{th} replicate of size $j = 1, \dots, J_k$. Similarly, let $\nu_j(\boldsymbol{\theta})$ be the j^{th} elements of the model-based WV $\nu(\boldsymbol{\theta})$.

Definition: Near Stationarity

Near stationarity

We define a nearly stationary time series, as one which exhibits the following properties:

1. Same model, but with **different parameter values** for each sequences.
2. The vector of parameter $\boldsymbol{\theta}$ has a **probability distribution $\boldsymbol{G}(\boldsymbol{\theta}_0)$**, where $\mathbb{E}[G] = \boldsymbol{\theta}_0$.
3. The distribution $\boldsymbol{G}(\boldsymbol{\theta}_0)$ can be interpreted as **the internal sensor model**, which may account for unobserved factors (e.g. temperature).

Remark:

Near stationarity is a new statistical concept which **contradicts the almost always assumed strong stationarity**. Further research is needed to understand this framework.

Two Estimators: Average GMWM vs MuGMWM

Replicate Error Signals

Let us assume we have K independent replicates of the observed stochastic error of an inertial sensor. Can we use all this information?

Average GMWM vs MuGMWM

We define $\hat{\boldsymbol{\theta}}^{\circ}$ as the Average GMWM (AGMWM) defined as:

$$\hat{\boldsymbol{\theta}}^{\circ} = \frac{1}{K} \sum_{k=1}^{K} \tilde{\boldsymbol{\theta}}_k \, ,$$

with $\tilde{\boldsymbol{\theta}}_k = \text{argmin}_{\boldsymbol{\theta}_k \in \boldsymbol{\Theta}} \|\hat{\boldsymbol{\nu}}_k - \boldsymbol{\nu}(\boldsymbol{\theta}_k)\|^2_{\boldsymbol{\Omega}_k}$. Remember that we define the Multisignal GMWM (MuGMWM) as

$$\hat{\boldsymbol{\theta}}^{\star} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \, \frac{1}{K} \sum_{k=1}^{K} \|\hat{\boldsymbol{\nu}}_k - \boldsymbol{\nu}(\boldsymbol{\theta})\|^2_{\boldsymbol{\Omega}_k}.$$

Two Estimators: Average GMWM vs MuGMWM

Properties

It turns out that the MuGMWM appears far more appropriate than the AGMWM for two main reasons:

- The MuGMWM is more efficient than the AGMWM, i.e.

$$\frac{\operatorname{tr}\left(\min_{\Omega_i} \operatorname{var}\left[\hat{\boldsymbol{\theta}}^{\circ}\right]\right)}{\operatorname{tr}\left(\min_{\Omega_i} \operatorname{var}\left[\hat{\boldsymbol{\theta}}^{\star}\right]\right)} \xmapsto{\mathcal{P}} c > 1.$$

- Jensen inequality implies that
  - If $\boldsymbol{\nu}(\boldsymbol{\theta})$ is linear, i.e for stochastic processes WN, DR and QN, or if $G(\boldsymbol{\theta}_0)$ is a Dirac function, then

$$\hat{\boldsymbol{\theta}}^{\star} - \hat{\boldsymbol{\theta}}^{\circ} \xmapsto{\mathcal{P}} 0.$$

  - If $\boldsymbol{\nu}(\boldsymbol{\theta})$ is not linear, i.e for stochastic processes RW and AR1, then

$$\hat{\boldsymbol{\theta}}^{\star} - \hat{\boldsymbol{\theta}}^{\circ} \xmapsto{\mathcal{P}} \delta \neq 0.$$

Testing for Near Stationarity

This frameworks allows to use the MuGMWM objective function as a test statistics, which is defined below:

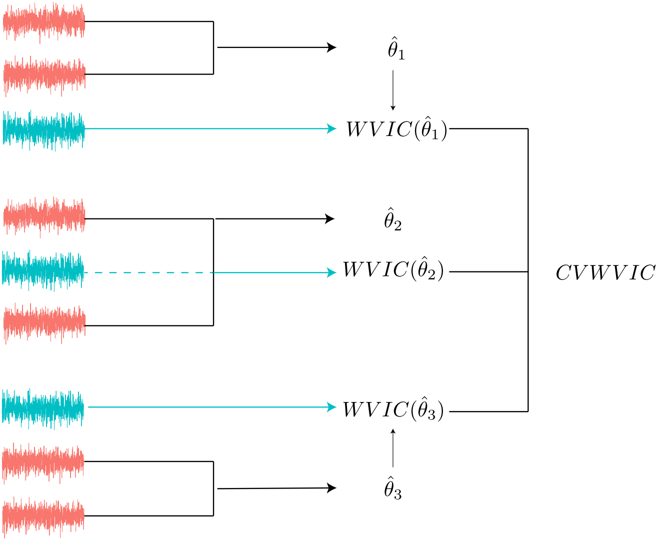
$$g(\hat{\boldsymbol{\theta}}) = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\|^2_{\hat{\boldsymbol{\Omega}}}.$$

Using this statistic we would therefore like to test the following null and alternative hypotheses

$$\mathcal{H}_0 : \boldsymbol{\theta}_k = \boldsymbol{\theta}_0, \; \forall\, k,$$
$$\mathcal{H}_a : \mathcal{H}_0 \text{ is false.}$$

Under $\mathcal{H}_0$, we fall in the case where the distribution G is a Dirac function and $\boldsymbol{\theta}_k = \boldsymbol{\theta}_0, \forall k$.

WVIC with Multiple Signal Replicates



CV-WVIC

Notation

- We partition the K replicates in two parts: the first part of size k_1 (with $1 \leq k_1 < K$), is used for **training purpose**, where $M = \binom{K}{k_1}$ denotes the size of every possible combination $m_i^{(k_1)}$ for $m = 1, \ldots, M$ and for $i = 1, \ldots, k_1$ whereas, the second part of size $k_2 = K - k_1$, for which $m_l^{(k_2)}$, for $l = 1, \ldots, k_2$, defines the complement of $m_i^{(k_1)}$, is used for **validation purpose**.
- Let $\mathcal{M}_\jmath$, $\jmath = 1, \ldots, \mathcal{J}$, denote the $\jmath^{th}$ model out of the set of $\mathcal{J}$ candidate models, with $\boldsymbol{\theta}_\jmath \subset \boldsymbol{\Theta}_\jmath \subseteq \mathbb{R}^p$ being the unknown parameter vector of $\mathcal{M}_\jmath$ and its associated parameter space.

CV-WVIC: Estimation and Validation Step

1. We use $m_i^{k_1}$, one specific combination of the error signals replicates of size k_1 to estimate the parameter vector $\boldsymbol{\theta}_{\jmath}^{(m_i)}$ for the $\jmath^{th}$ model in the following manner

$$\hat{\boldsymbol{\theta}}_{\jmath}^{(m_i)} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\jmath}}{\text{argmin}} \sum_{k \in m_i^{k_1}}^{k_1} ||\hat{\boldsymbol{\nu}}_k - \boldsymbol{\nu}(\boldsymbol{\theta}_{\jmath})||_{\boldsymbol{\Omega}}^2. \tag{3.2}$$

2. We then compute the MuGMWM objective function, evaluated at the parameter estimated in Eq. (3.2), to deliver the WVIC computed on an independent sample $\boldsymbol{X}_t^{(k)}$ with $k \in c_{k_2}^{(m)}$:

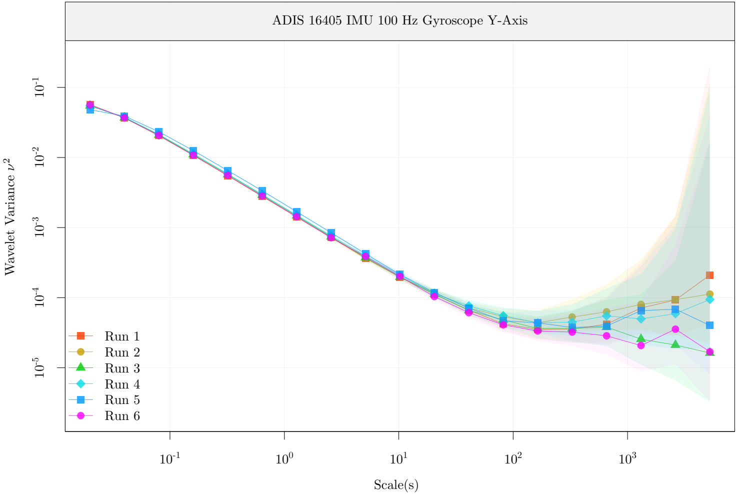
$$\widehat{\mathcal{C}}_{\jmath}^{(m_l)} = \sum_{k \in m_l^{k_2}}^{k_2} \|\hat{\boldsymbol{\nu}}_k - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}}_{\jmath}^{(m_i)})\|_{\boldsymbol{\Omega}}^2. \tag{3.3}$$

CV-WVIC

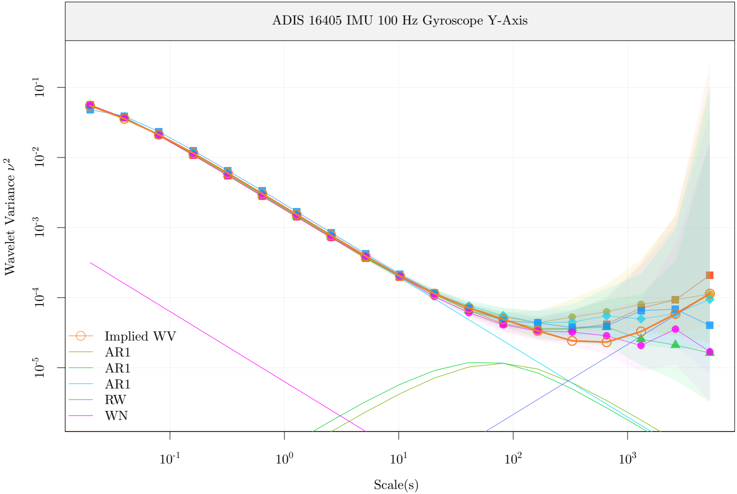
Repeating Eq. (3.2) and (3.3) for $m = 1, \ldots, M$, and computing the average $\widehat{\mathcal{C}}_\jmath = \frac{1}{M}\sum_{m=1}^{M} \widehat{\mathcal{C}}_\jmath^{(m_l)}$, for each model in the $\mathcal{J}$, will allow us to select the model which minimizes $\widehat{\mathcal{C}}_\jmath$, i.e.

$$\hat{\jmath} = \underset{\jmath=1,\ldots,\mathcal{J}}{\text{argmin}}\, \widehat{\mathcal{C}}_\jmath. \tag{3.4}$$

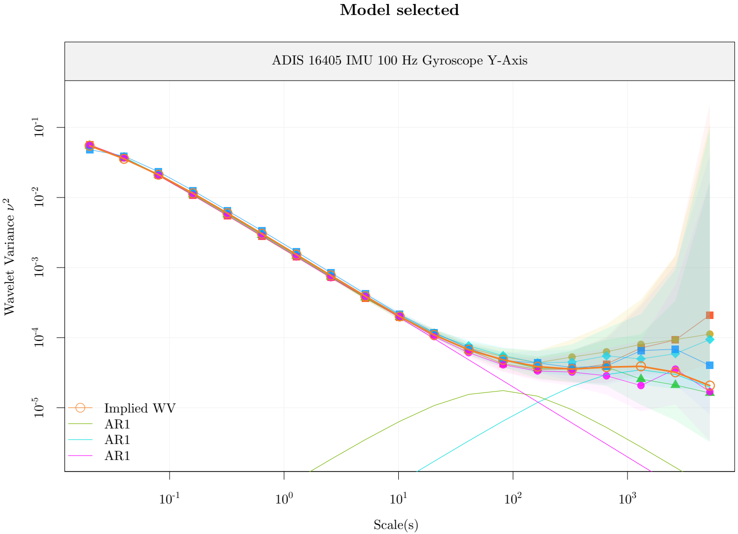
Case Study: The ADIS 16405 IMU



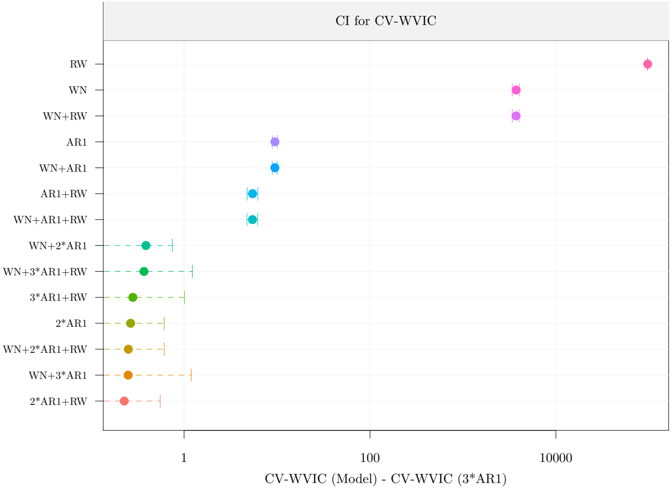
Case Study: The ADIS 16405 IMU



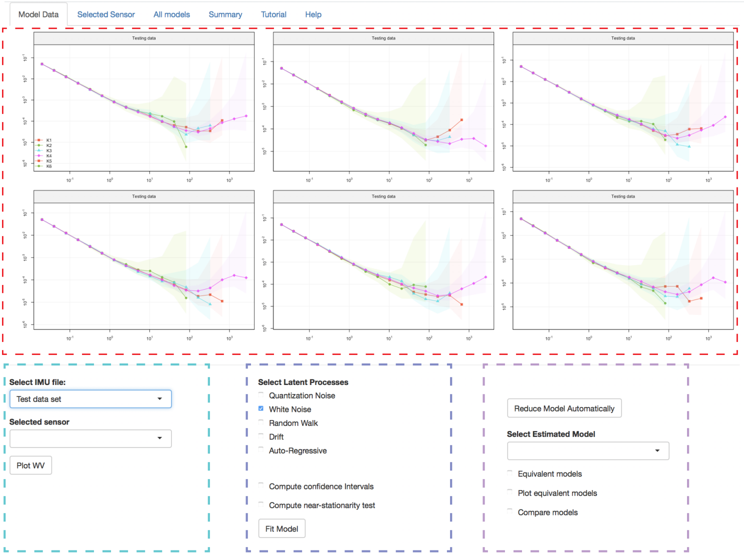
Case Study: The ADIS 16405 IMU



Case Study: The ADIS 16405 IMU



Web-based Platform: MuGMWM



Extension to Multivariate Processes

IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 67, NO. 19, OCTOBER 1, 2019 5143

Multivariate Signal Modeling With Applications to Inertial Sensor Calibration

Haotian Xu, Stéphane Guerrier, Roberto Carlo Molinari, and Mucyo Karemera

Abstract—**The common approach to inertial sensor calibration has been to model the stochastic error signals of individual sensors independently, whether as components of a single inertial measurement unit (IMU) in different directions or arrayed in the same direction for redundancy. For this purpose, research in this domain has been focused on the proposal of various methods to improve the estimation of these models both from a computational and a statistical point of view. However, the separate calibration of the individual sensors is unable to take into account the dependence between each of them which can have an important impact on the precision of the navigation systems. In this paper, we develop a new approach to simultaneously model the individual signals and the dependence between them by studying the quantity called Wavelet Cross-Covariance and using it to extend the application of the Generalized Method of Wavelet Moments. This new method can be used in other settings for time series modeling, especially in cases where the dependence among signals may be hard to detect. Moreover, in the field of inertial sensor calibration, this approach can deliver important contributions among which the possibility to test dependence between sensors, integrate their dependence within the navigation filter and construct an optimal virtual sensor that can be used to simplify and improve navigation accuracy. The advantages of this method and its usefulness for inertial sensor calibration are highlighted through a simulation study and an applied example with a small array of XSens MTi-G IMUs.**

Index Terms—**Generalized method of wavelet moments, wavelet covariance, multivariate time series, signal processing, virtual sensors, IMU, navigation.**

I. Introduction

THE modelling of multivariate time series is an important as well as challenging task in many applied domains of data analysis. Indeed, it is extremely common to measure phenomena over time which not only manifest autocorrelation with their past but also show a form of dependence between them. This setting can be observed, among others, in a variety of longitudinal studies [1], [2], in economic and financial research [3]–[6] as well as in different fields of engineering [7], [8]. In the latter case, an increased attention has been given to the modelling of multivariate signals for the task of inertial sensor calibration where the error signals of the individual accelerometers and gyroscopes that compose the Inertial Measurement Units (IMU) are characterized by error signals that are dependent on each other [8]. To date, these error signals have been modeled independently from each other and this task alone has already been challenging to deal with. In fact, the individual error signals of these instruments are characterized by deterministic and stochastic components where the first can be dealt with through physical models (which are employed for calibration and compensation prior to use) while the latter are often described through complex stochastic models that are made by the sum of different underlying processes. The estimation of these stochastic models has always been complicated and many methods used for this purpose have either suffered from statistical limits and/or are computationally or numerically unstable. Indeed, in most latent variable modelling scenarios, the Maximum-Likelihood Estimator (MLE) cannot be directly applied since the latent stochastic components are unobservable and their marginal distribution is usually unknown. A feasible likelihood-based method is the Expectation-Maximization (EM) algorithm [9] whose implementation is nevertheless limited due to the complexity and, more importantly, the possible non-convexity of its associated optimization problems therefore severely limiting the use of its statistical properties in practice [10]. To avoid the problems of the likelihood-based methods, the linear regression based on the Allan Variance (AVLR) is widely used for estimation and prediction especially in the field of inertial sensor calibration [11]. Although, the AVLR is a computationally feasible method for estimating many complex models, its statistical properties are not optimal as discussed in [12] where the inconsistency of this method was shown for the majority of the latent models used for the purpose of sensor calibration.

Given the complexity of these models and the difficulties in estimating them, the signals have thus far been dealt with individually without taking into account the forms of dependence between them. As mentioned earlier, this approach is non-optimal since a multivariate approach to signal analysis would be more appropriate for different reasons. Firstly, it is reasonable to assume that there exists a dependence between individual sensors along the three axes of an IMU and, when

Manuscript received May 3, 2019; revised July 19, 2019; accepted August 4, 2019. Date of publication August 19, 2019; date of current version September 5, 2019. The work of M. Karemera was supported by a Swiss National Science Foundation Grant (SNSF Professorships, 176843). The work of S. Guerrier was supported in part by a Innosuisse grant (IP-ENG, 37308.1) and in part by a Swiss National Science Foundation Grant (SNSF Professorships, 176843). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Remy Boyer. *(Corresponding author: Roberto Carlo Molinari.)*

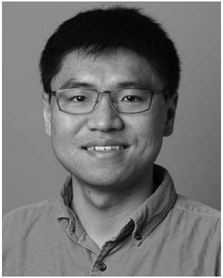
H. Xu and M. Karemera are with the Geneva School of Economics and Management, University of Geneva, Geneva 1205, Switzerland (e-mail: haotian.xu@unige.ch; mucyo.karemera@unige.ch).

S. Guerrier is with the Geneva School of Economics and Management and the Faculty of Science, University of Geneva, Geneva 1205, Switzerland (e-mail: stephane.guerrier@unige.ch).

R. C. Molinari is with the Department of Statistics, Pennsylvania State University, State College, PA 16801 USA (e-mail: rum415@psu.edu).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TSP.2019.2935902



Haotian Xu
U. Geneva

Multivariate process

Definition 5.1 (Common class of multivariate process).

- **WN**: $\boldsymbol{X}_{1,t} \overset{\text{i.i.d.}}{\sim} \mathcal{F}(\mathbf{0}, \boldsymbol{\Sigma})$.
- **RW**: $\boldsymbol{X}_{2,t} = \boldsymbol{X}_{2,t-1} + \boldsymbol{\iota}_t, \;\; \boldsymbol{\iota}_t \overset{\text{i.i.d.}}{\sim} \mathcal{F}(\mathbf{0}, \boldsymbol{\Lambda})$.
- **QN**: $\boldsymbol{X}_{3,t}$. For this process we do not consider a multivariate model and therefore each univariate process is characterized by the parameter $Q_i^2 \in \{x \in \mathbb{R} | x > 0\}$.
- **Dr**: $\boldsymbol{X}_{4,t}$. Define each univariate process as $X_{4,t}^{(i)} = \omega_i t, \;\; \omega_i \in \{x \in \mathbb{R} | x > 0\}$.
- **AR1**: $\boldsymbol{X}_{k,t}$, $\forall \; k = 5, \dots, K$, where $K \in \mathbb{N}$, and defined as

$$\boldsymbol{X}_{k,t} = \boldsymbol{\Phi} \boldsymbol{X}_{k,t-1} + \boldsymbol{\varepsilon}_t,$$

  where $\boldsymbol{\Phi}$ is a diagonal matrix with diagonal elements $\phi_k^{(i)}$, $0 < |\phi_k^{(i)}| < 1$, for $i = 1, \dots, D_k$ and $\boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{F}(\mathbf{0}, \boldsymbol{Z}_k)$.

The covariance matrices $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{Z}_k$ defined above are positive definite $D_k \times D_k$ matrices with $k = 1, 2, 5, ..., K$ respectively, with their respective elements being $[\sigma^{(i,i')}]_{1 \leq i \leq i' \leq D_1}$, $[\lambda^{(i,i')}]_{1 \leq i \leq i' \leq D_2}$ and $[z_k^{(i,i')}]_{1 \leq i \leq i' \leq D_k}$.

Multivariate process

Definition 5.2 (Multivariate process).

Let $(\boldsymbol{X}_t)_{t\in\mathbb{Z}} \subset \mathbb{R}^I$, where $I \in \mathbb{N}$. Consider a multivariate time series composed by K latent processes belonging to the class in Definition 5.1 as

$$\boldsymbol{X}_t \equiv \sum_{k=1}^{K} \boldsymbol{S}_k \boldsymbol{X}_{k,t},$$

$\boldsymbol{X}_t \equiv [X_t^{(i)}]_{i=1,\dots,I}$ is an I-dimensional vector and $\boldsymbol{X}_{k,t} \equiv [X_{k,t}^{(j)}]_{j=1,\dots,D_k}$ is a D_k-dimensional vector ($D_k \leq I$) with k being the index for the type of univariate/multivariate model underlying some or all of the latent processes and $\boldsymbol{S}_k$ is an $I \times D_k$ matrix with either 1 or 0 as its elements that we define as

$$(\boldsymbol{S}_k)_{i,j} = \begin{cases} 1 & \text{if } X_{k,t}^{(j)} \text{ appears in } X_t^{(i)} \\ 0 & \text{if otherwise.} \end{cases}$$

Multivariate process (continue)

Condition of the multivariate process

- $\boldsymbol{X}_{k,t}$ is independent of $\boldsymbol{X}_{k',t}$, $\forall\ k \neq k'$.

Remark 1.

This condition implies that each model that composes a latent model (univariate or multivariate) is independent from the others (e.g. a **WN** process is independent from a **AR1** process and two different **AR1** processes are independent from each other). The reason for stating this condition is simply to provide proofs on the identifiability of these models for the MGMWM method.

Example:

- $\boldsymbol{X}_t = \boldsymbol{X}_{1,t} + \boldsymbol{X}_{2,t} + \boldsymbol{X}_{3,t} + \boldsymbol{X}_{4,t}$.
- $\boldsymbol{X}_t = \boldsymbol{X}_{1,t} + \boldsymbol{X}_{3,t} + \sum_{k=5}^{K} \boldsymbol{X}_{k,t}$.

Wavelet cross-covariance(WCCV)

The wavelet coefficients for $i = 1, \ldots, I$ and level $j = 1, \ldots, J$ is:

$$W_{j,t}^{(i)} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}^{(i)}.$$

Definition 5.3 (Wavelet cross-covariance).

We can define the wavelet cross-covariance at the j-th level as

$$\gamma_{j,h}^{(i,i')}(\boldsymbol{\theta}^{(i,i')}) = \text{cov}\left(W_{j,t}^{(i)},\, W_{j,t+h}^{(i')}\right),\;\; 1 \leq i, i' \leq I,$$

where h represents the lag in time between observations and $\theta^{(i,i')}$ represents the parameter vector defining the dependence within/between signals *(see Definition (5.1))*.

Remark 2.

We limit ourselves to studying the WCCV at lag $h = 0$ since this lag is sufficient to identify and estimate the models considered for this work.

WCCV estimator

Similar to $\hat{\nu}_j^2$, consider a WCCV estimator:

$$\hat{\gamma}_j^{(i,i')} = \hat{\gamma}_{j,0}^{(i,i')} = \frac{1}{M_j} \sum_{t=1}^{M_j} W_{j,t}^{(i)} W_{j,t}^{(i')}.$$

Remark 3.

When $i = i'$, $\gamma_j^{(i,i')}$ and $\hat{\gamma}_j^{(i,i')}$ reduce to ν_j^2 and $\hat{\nu}_j^2$ respectively.

Based on the above notations, the WCCV vector and the vector of estimated WCCV are defined as:

$$\boldsymbol{\nu}(\boldsymbol{\theta}_0) \equiv \left[\gamma_j^{(i,i')}\right]_{\substack{j=1,\dots,J \\ 1\le i\le i'\le I}}, \quad \hat{\boldsymbol{\nu}} \equiv \left[\hat{\gamma}_j^{(i,i')}\right]_{\substack{j=1,\dots,J \\ 1\le i\le i'\le I}},$$

and the parameter vector of the overall model is:

$$\boldsymbol{\theta} \equiv \left[\left[\boldsymbol{\theta}^{(1)}\right]^{\mathsf{T}}, \left[\boldsymbol{\theta}^{(1,2)}\right]^{\mathsf{T}} \dots, \left[\boldsymbol{\theta}^{(I-1,I)}\right]^{\mathsf{T}}, \left[\boldsymbol{\theta}^{(I)}\right]^{\mathsf{T}}\right]^{\mathsf{T}}.$$

Properties of the WCCV estimators

Conditions

- $\boldsymbol{\Delta}_t = \left(\Delta_t^{(1)}, \Delta_t^{(2)}, \dots, \Delta_t^{(I)}\right)^\intercal = \boldsymbol{G}(\mathcal{F}_t)$, where $\mathcal{F}_t = \sigma(\dots, \epsilon_{t-1}, \epsilon_t)$ with ϵ_t be i.i.d. random elements, and $\boldsymbol{G}(\cdot) = \left(g^{(1)}(\cdot), g^{(2)}(\cdot), \dots, g^{(I)}(\cdot)\right)^\intercal$ is a $\mathbb{R}^I$-valued measurable function,
- $\max_{i=1,\dots,I} \|\Delta_t^{(i)}\|_4 \leq \infty$,
- The functional dependence measure $\max_{i=1,\dots,I} \sum_{t=0}^{\infty} \|\Delta_t^{(i)} - \Delta_t^{(i)\,*}\|_4 \leq \infty$, where $\Delta_t^{(i)\,*} = g^{(i)}(\mathcal{F}_t^*)$ and $\mathcal{F}_t^* = \sigma(\dots, \epsilon_{-1}, \epsilon_0^*, \epsilon_1, \dots, \epsilon_t)$ Wu 2005.

Asymptotic result:

Under these conditions, we have

$$\sqrt{T}\,(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\theta_0)) \overset{\mathcal{D}}{\rightarrow} \mathcal{N}(\mathbf{0}, \boldsymbol{V}),$$

where $\boldsymbol{V} = \text{var}(\hat{\boldsymbol{\nu}})$, see Theorem 1 in Xu et al. 2019.

MGMWM

GMWM consider a univariate process $(X_t^{(i)})_{t\in\mathbb{Z}}$ and, assuming that this process is generated from a parametric model with its parameter vector $\boldsymbol{\theta}^{(i)}$ which can **uniquely identify** the WV vector $\gamma^{(i)}(\boldsymbol{\theta}^{(i)})$, the GMWM aims at estimating this parameter vector through the following minimization problem:

$$\hat{\boldsymbol{\theta}}^{(i)} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}^{(i)}}{\operatorname{argmin}} \left(\hat{\gamma}^{(i)} - \gamma^{(i)}(\boldsymbol{\theta})\right)^{\mathsf{T}} \boldsymbol{\Omega}^{(i)} \left(\hat{\gamma}^{(i)} - \gamma^{(i)}(\boldsymbol{\theta})\right).$$

The idea of the GMWM, which is based on the Generalized Method of Moments (GMM) framework, is to inverse the mapping from the WV vector back to the parameter vector $\boldsymbol{\theta}$. With this in mind, this idea can be extended by using the vector of WCCV which includes moments that contain information on the subset of parameters in $\boldsymbol{\theta}$ that contain information on the latent dependence between signals i and i'. For this reason, it is possible to build the MGMWM as an extension of the GMWM as follows:

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

Identifiability

The condition of model identifiability is essential for any method that aims at estimating its corresponding parameters and is consequently necessary to obtain its asymptotic properties for inference. In brief, model identifiability through the MGMWM can be defined as

$$\boldsymbol{\nu}(\boldsymbol{\theta}_0) = \boldsymbol{\nu}(\boldsymbol{\theta}_1) \text{ if and only if } \boldsymbol{\theta}_0 = \boldsymbol{\theta}_1,$$

where $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \in \boldsymbol{\Theta} \subset \mathbb{R}^p$. This basically means that there is a one-to-one correspondence between the WCCV and the parameter vector $\boldsymbol{\theta}$. This property, although essential, is often assumed in practice since it can be considerably challenging to prove.

Identifiability results

Under certain technical conditions (see Xu et al. 2019 for details), we have the identifiability for the following two general models

- $\boldsymbol{X}_t = \boldsymbol{X}_{1,t} + \boldsymbol{X}_{2,t} + \boldsymbol{X}_{3,t} + \boldsymbol{X}_{4,t}$.
- $\boldsymbol{X}_t = \boldsymbol{X}_{1,t} + \boldsymbol{X}_{3,t} + \sum_{k=5}^{K} \boldsymbol{X}_{k,t}$.

Asymptotic properties of the MGMWM

Under certain technical conditions, we have:

- **Consistency**: $\hat{\boldsymbol{\theta}} \overset{p}{\to} \boldsymbol{\theta}_0$, see Theorem 2 in Xu et al. 2019.
- **Asymptotic normality**: $\sqrt{T}\,\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \overset{\mathcal{D}}{\to} \mathcal{N}(\mathbf{0}, \boldsymbol{\Xi})$, where $\boldsymbol{\Xi}$ is the asymptotic covariance of $\hat{\boldsymbol{\theta}}$, see Theorem 3 in Xu et al. 2019.

Algorithm

1. Compute empirical WCCV $\hat{\boldsymbol{\nu}}$ for a suitable number of scales $J < \log_2(T)$
2. Compute starting value $\tilde{\boldsymbol{\theta}}$: **for** $i \leftarrow 1$ **to** I **do**

```
    for j ← i to I do
        if i == j then
            Compute Ω^(i) using the approach considered in Guerrier et al. 2013;
            Compute θ̃^(i) (based on Ω^(i)) using the algorithm described in
              Balamuta et al. 2017
        else
            θ̃^(i,j) ⟵ 0
        end
    end
end
```

3. Compute $\hat{\boldsymbol{\theta}}$:
   1. Estimate $\boldsymbol{\Omega}$ by $\boldsymbol{\Omega}_T$ (for example taking the inverse of the covariance matrix estimator Andrews 1991).
   2. Compute $\hat{\theta}$ using $\boldsymbol{\Omega}_T$ and the starting value $\tilde{\boldsymbol{\theta}}$ by the algorithm described in Balamuta et al. 2017

Simulation: Combining Low-Cost Sensors

We consider 3 calibration signals which are correlated. The true model is:

$$X_t = \mathsf{RW}_t(\boldsymbol{\Lambda}) + \mathsf{WN}_t(\boldsymbol{\Sigma}).$$

The covariance matrix $\boldsymbol{\Sigma}$ of the multivariate WN process is given by

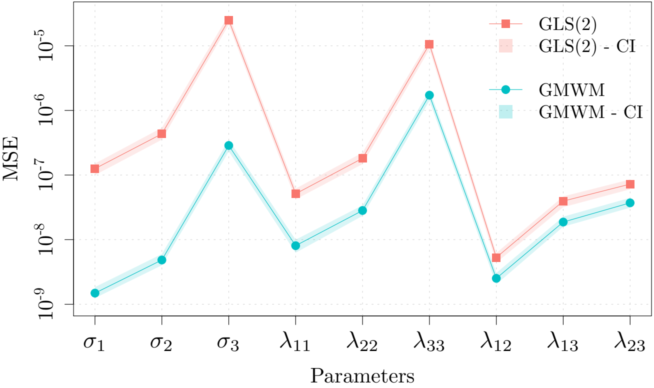
$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} = 10^{-3} \begin{bmatrix} 0.1010 & 0 & 0 \\ 0 & 0.0712 & 0 \\ 0 & 0 & 0.0490 \end{bmatrix},$$

while the covariance matrix $\boldsymbol{\Lambda}$ of the multivariate RW process is given by

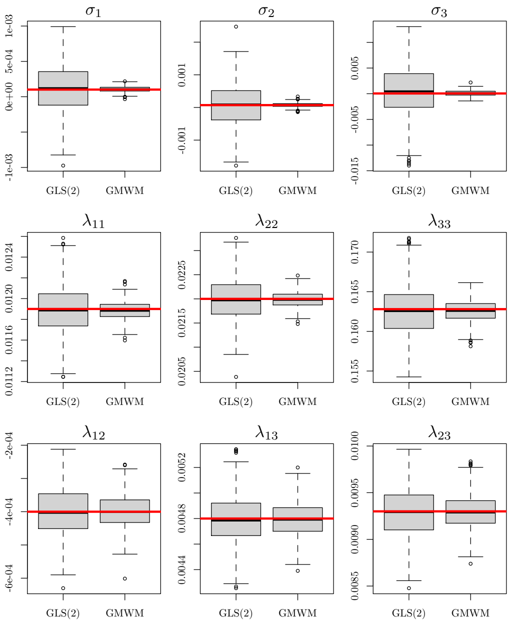
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{12} & \lambda_{22} & \lambda_{23} \\ \lambda_{13} & \lambda_{23} & \lambda_{33} \end{bmatrix} = \begin{bmatrix} 0.0119 & -0.0004 & 0.0048 \\ -0.0004 & 0.0220 & 0.0093 \\ 0.0048 & 0.0093 & 0.1628 \end{bmatrix}.$$

The MGMWM estimators of $\boldsymbol{\Lambda}(\lambda_{11}, \lambda_{22}, \lambda_{33}, \lambda_{12}, \lambda_{13}, \lambda_{23})$ and $\boldsymbol{\Sigma}(\sigma_1, \sigma_2, \sigma_3)$ are obtained and compared with the estimators proposed recently by Vaccaro and Zaki 2017. It is shown that our estimator is outperformed in terms of finite sample bias and efficiency. In Vaccaro and Zaki 2017, they proposed a GLS estimator based on the sample AV.

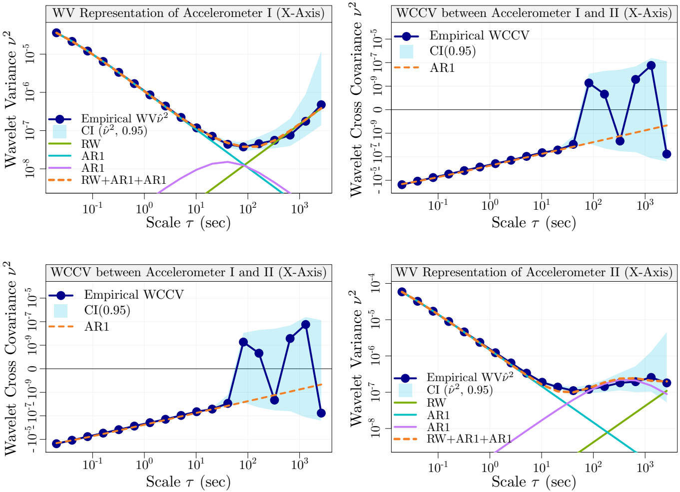
Simulation: Combining Low-Cost Sensors



Solid lines represent the sample MSE of GLS(2) and GMWM estimators for a multivariate WN + RW process based on 500 replications and sample size $T = 10^5$. Lower and upper dotted lines represent respectively the bootstrap 2.5% and 97.5% confidence intervals.



σ_1
σ_2
σ_3
λ_{11}
λ_{22}
λ_{33}
λ_{12}
λ_{13}
λ_{23}
GLS(2)
GMWM
1e-03
5e-04
0e+00
-1e-03
0.001
-0.001
0.005
-0.005
-0.015
0.0124
0.0120
0.0116
0.0112
0.0225
0.0215
0.0205
0.170
0.165
0.160
0.155
-2e-04
-4e-04
-6e-04
0.0052
0.0048
0.0044
0.0100
0.0095
0.0090
0.0085



Given the empirical WV (XSens MTi-G IMU) and the characteristics of the theoretical WV of the latent variables we considered, we propose to use the sum of RW + AR1 + AR1 to model both these two processes. Given the empirical WCCV, we allow the second AR1 component to be correlated.

Application to Virtual Sensor

An Optimal Virtual Inertial Sensor Framework using Wavelet Cross Covariance

Yuming Zhang
Department of Statistics
Pennsylvania State University
University Park, United States
yfz5097@psu.edu

Haotian Xu
Research Center for Statistics
University of Geneva
Geneva, Switzerland
haotian.xu@unige.ch

Ahmed Radi
Department of Geomatics Engineering
University of Calgary
Calgary, Canada
ahmed.elboraee@ucalgary.ca

Roberto Molinari
Department of Statistics
Pennsylvania State University
University Park, United States
roberto.molinari@hotmail.it

Stéphane Guerrier
Department of Statistics
and Institute for CyberScience
Pennsylvania State University
University Park, United States
stephane@psu.edu

Mucyo Karemera
Department of Statistics
Pennsylvania State University
University Park, United States
mucyo.karemera@gmail.com

Naser El-Sheimy
Department of Geomatics Engineering
University of Calgary
Calgary, Canada
elsheimy@ucalgary.ca

***Abstract*—The practice of inertial sensor calibration has commonly been carried out by taking into account the deterministic and stochastic components of the error measurements issued from a calibration session. Once the deterministic components have been taken into account through physical models, the remaining stochastic component has always been dealt with for each sensor separately. The latter process involves estimating complex probabilistic models for each sensor which has been proven to be extremely complicated over the past years, although recent proposals have allowed to overcome most of the limitations that have characterized this task. However, the separate stochastic calibration of the individual sensors composing an inertial measurement unit may not be wise in many cases since there can be considerable degrees of dependence between the sensors, especially between the gyroscopes. For this reason, there has been growing attention towards this issue in order to consider the influence of the stochastic behaviour of the sensors on each other, with few proposals that address this problem. Among these proposals there has been the idea of integrating the information coming from the different gyroscopes so as to build a virtual gyroscope. In this paper we build on this idea and, using a recently proposed method for multivariate signal modelling, we deliver a general and flexible framework that allows to consider many different modelling options which provide the basis to construct a virtual sensor that optimally combines the information from the individual sensors and considerably improves navigation accuracy.**

***Index Terms*—Stochastic Error Modeling, Inertial Sensors Calibration, Inertial Measurement Unit, Wavelet Variance, Wavelet Cross-Covariance, Generalized Method of Wavelet Moments**

I. Introduction

The utilization of inertial sensors in the design of Inertial Navigation Systems (INS) to provide short term position, velocity and attitude information is constantly growing for navigation purposes for different applications. Generally speaking, inertial sensors suffer from numerous errors that are accumulated through time and should be accurately modeled and compensated to achieve acceptable navigation results and to avoid the rapid degradation of the navigation solution. Hence, developing accurate and efficient models is highly needed to decrease the effect of inertial sensor errors. The modeling process of such inertial sensor errors is considered to be one of the most challenging tasks in the design of any navigation system, especially for low-cost ones. Indeed, inertial sensors are typically corrupted by two categories of errors: deterministic and stochastic. A considerable amount of literature has contributed to the calibration of deterministic errors which can be found, for example, in [1] (and the references therein). However, the calibration of deterministic errors is beyond the scope of this study which deals with approaches used to calibrate stochastic errors that are still limited due to the complexity of this task. In fact, these stochastic errors are often very complex in nature and can only be represented through composite processes (i.e. the sum of multiple independent latent processes) which has lead to a considerable amount of research in order to improve calibration in this setting.

Yuming Zhang
U. Geneva

Redundant Sensor Architecture

Goal

Combine all information coming from an array of sensors to provide false detection and isolation and to improve the overall navigation performance.

Approaches

(1) Navigation filter

- Construct a state space model by taking all measurements from both gyros and accelerometers (or any external aiding).
- Usually computationally challenging and the gain is often not considerable.

(2) **Virtual sensor**

- Construct by taking a (typically linear) combination of all sensors.
- More computationally efficient.

Proposed Method

Goal

Construct an optimal (based on WV) virtual inertial sensor by finding the optimal weights on each sensor depending on the navigation requirements of applications.

Advantages

- Consider navigation requirements of short term and long term applications.
- Consider dependence among sensors.
- Consider flexible multivariate time series models.
- Computationally efficient and numerically stable.

Remark

Our following discussion will mainly focus on applications in gyros, but the proposed method can also be used on an array of accelerometers.

Optimal Virtual Gyro Construction

Virtual Gyro

Let's consider the composite stochastic error $\mathbf{Y}_t$ of an array of p gyros. Then the composite stochastic error of the virtual gyro V_t is defined as

$$V_t = \mathbf{c}^T \mathbf{Y}_t,$$

where $\mathbf{c}$ is the coefficient vector defined as

$$\mathbf{c} = [c_1, \ \ldots, \ c_p]^T \in \mathcal{C} = \{\mathbf{c} \in \mathbb{R}^p : ||\mathbf{c}||_1 = 1\}.$$

Optimal Virtual Gyro Construction

Optimal Weights on Gyros

The coefficient vector $\mathbf{c}$ is found by

$$\hat{\mathbf{c}} = \underset{\mathbf{c} \in \mathcal{C}}{\text{argmin}} \sum_{j=1}^{J} w_j \nu_j(\mathbf{c}),$$

where

- $\nu_j(\mathbf{c})$ is the WV of virtual gyro based on the $\mathbf{c}$,
- J is the total number of scales related to T (i.e. $J \equiv \lfloor \log_2(T) \rfloor - 1$),
- $\mathbf{w} = [w_1, \ \ldots, \ w_J]^T \in {\mathbb{R}^+}^J$ is a constant vector selected by users based on their application requirements.

Illustration Examples

Settings

In order to illustrate the proposed method, we consider the following examples:

- Example 1: Two **Independent** Gyros
  Gyro 1: WN + RW, Gyro 2: WN + RW
- Example 2: Three **Independent** Gyros
  Gyro 1: WN + RW, Gyro 2: WN + RW, Gyro 3: WN + RW
- Example 3: Three **Correlated** Gyros
  Gyro 1: WN + RW, Gyro 2: WN + RW, Gyro 3: WN + RW

Remark

Orange indicates the process with relatively largest variance.
Red indicates the processes that are correlated.

Example 1: Two Independent Gyros

Setting

Suppose we consider an array of two univariate processes with 10 million observations corresponding to $J = 22$.

- **Gyro 1:** **WN** + RW
- **Gyro 2:** WN + **RW**
- Two gyros are **independent**

Application Requirements

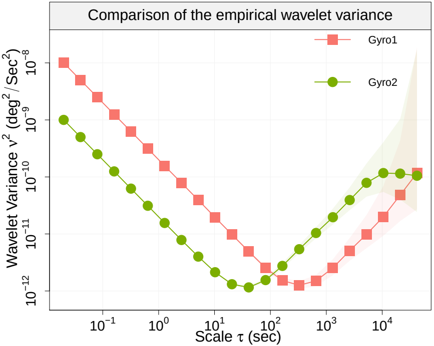
- **Short Term:** $w_j^{(1)} = \begin{cases} \frac{1}{3} & \text{if } j = 1, \dots, 3 \\ 0 & \text{otherwise} \end{cases}$
- **Long Term:** $w_j^{(2)} = \begin{cases} \frac{1}{3} & \text{if } j = 20, \dots, 22 \\ 0 & \text{otherwise} \end{cases}$

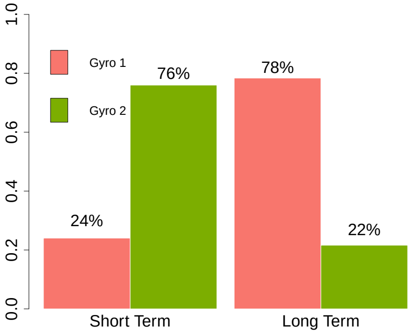
Setting in Example 1

Suppose we consider an array of two univariate processes with 10 million observations corresponding to $J = 22$.

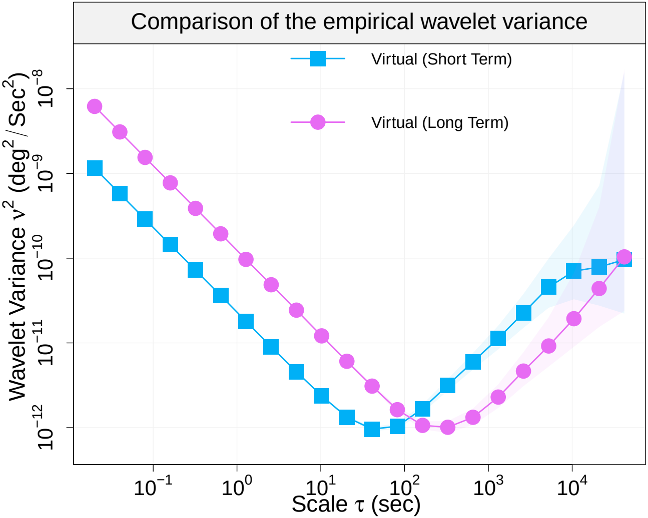
- **Gyro 1:** $X_t^{(1)} = X_t^{(1.1)} + X_t^{(1.2)}$, where
  $X_t^{(1.1)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-8})$,
  $X_t^{(1.2)} = X_{t-1}^{(1.2)} + \epsilon_t^{(1.2)}$, $\epsilon_t^{(1.2)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-16})$.
- **Gyro 2:** $X_t^{(2)} = X_t^{(2.1)} + X_t^{(2.2)}$, where
  $X_t^{(2.1)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-9})$,
  $X_t^{(2.2)} = X_{t-1}^{(2.2)} + \epsilon_t^{(2.2)}$, $\epsilon_t^{(2.2)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-15})$.
- Two gyros are **independent**.

Example 1: Two Independent Gyros





Example 1: Two Independent Gyros



Example 2: Three Independent Gyros

Setting

Now suppose we consider an array of three univariate processes with 10 million observations corresponding to $J = 22$.

- **Gyro 1: WN** + RW
- **Gyro 2:** WN + **RW**
- **Gyro 3: WN** + **RW**
- Three gyros are **independent**

Application Requirements

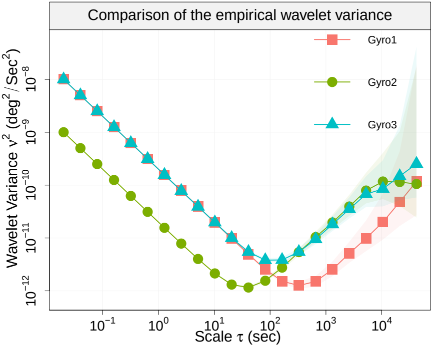
- **Short Term:** $w_j^{(1)} = \begin{cases} \frac{1}{3} & \text{if } j = 1, \dots, 3 \\ 0 & \text{otherwise} \end{cases}$
- **Long Term:** $w_j^{(2)} = \begin{cases} \frac{1}{3} & \text{if } j = 20, \dots, 22 \\ 0 & \text{otherwise} \end{cases}$

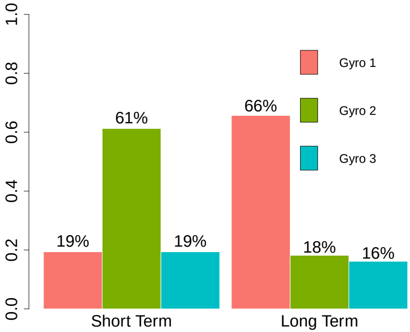
Setting in Example 2

Now suppose we consider an array of three univariate processes with 10 million observations corresponding to $J = 22$.

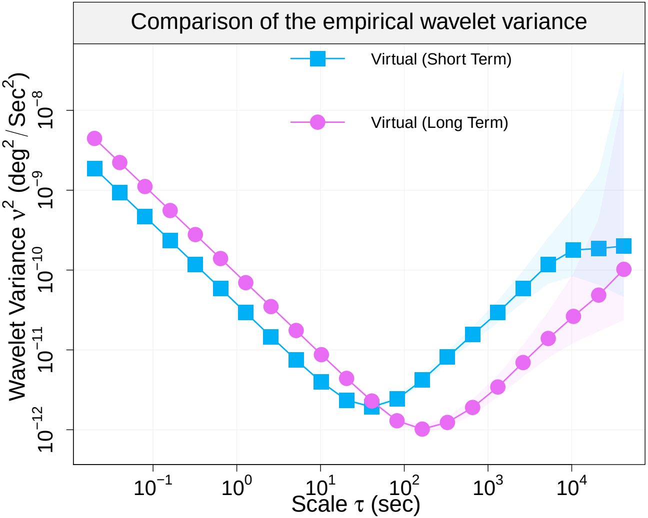
- **Gyro 1:** $X_t^{(1)} = X_t^{(1.1)} + X_t^{(1.2)}$, where
  $X_t^{(1.1)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-8})$,
  $X_t^{(1.2)} = X_{t-1}^{(1.2)} + \epsilon_t^{(1.2)}$, $\epsilon_t^{(1.2)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-16})$.
- **Gyro 2:** $X_t^{(2)} = X_t^{(2.1)} + X_t^{(2.2)}$, where
  $X_t^{(2.1)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-9})$,
  $X_t^{(2.2)} = X_{t-1}^{(2.2)} + \epsilon_t^{(2.2)}$, $\epsilon_t^{(2.2)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-15})$.
- **Gyro 3:** $X_t^{(3)} = X_t^{(3.1)} + X_t^{(3.2)}$, where
  $X_t^{(3.1)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-8})$,
  $X_t^{(3.2)} = X_{t-1}^{(3.2)} + \epsilon^{(3.2)}$, $\epsilon_t^{(3.2)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-15})$.
- Three gyros are **independent**.

Example 2: Three Independent Gyros





Example 2: Three Independent Gyros



Example 3: Three Correlated Gyros

Setting

Suppose we still consider the same three gyros as Example 2, but now there exists dependence between gyro 2 and gyro 3.

- **Gyro 1:** **WN** + RW
- **Gyro 2:** WN + **RW**
- **Gyro 3:** **WN** + **RW**
- Correlation is **0.9**.

Application Requirements

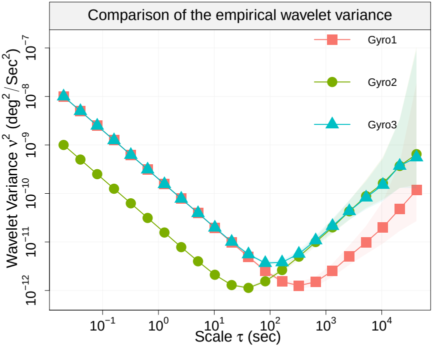
- **Short Term:** $w_j^{(1)} = \begin{cases} \frac{1}{3} & \text{if } j = 1, \dots, 3 \\ 0 & \text{otherwise} \end{cases}$
- **Long Term:** $w_j^{(2)} = \begin{cases} \frac{1}{3} & \text{if } j = 20, \dots, 22 \\ 0 & \text{otherwise} \end{cases}$

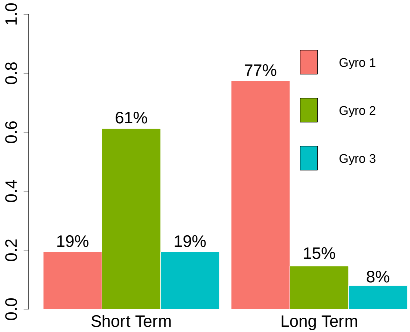
Setting in Example 3

Suppose we still consider the same three gyros as Example 2, but now there exists dependence between gyro 2 and gyro 3.

- **Gyro 1:** $X_t^{(1)} = X_t^{(1.1)} + X_t^{(1.2)}$, where
  $X_t^{(1.1)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-8})$,
  $X_t^{(1.2)} = X_{t-1}^{(1.2)} + \epsilon_t^{(1.2)}$, $\epsilon_t^{(1.2)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-16})$
- **Gyro 2:** $X_t^{(2)} = X_t^{(2.1)} + X_t^{(2.2)}$, where
  $X_t^{(2.1)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-9})$,
  $X_t^{(2.2)} = X_{t-1}^{(2.2)} + \epsilon^{(2.2)}$, $\epsilon_t^{(2.2)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-15})$
- **Gyro 3:** $X_t^{(3)} = X_t^{(3.1)} + X_t^{(3.2)}$, where
  $X_t^{(3.1)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-8})$,
  $X_t^{(3.2)} = X_{t-1}^{(3.2)} + \epsilon_t^{(3.2)}$, $\epsilon_t^{(3.2)} \overset{\text{iid}}{\sim} N(0, 2 \cdot 10^{-15})$
- **$\text{Corr}(\epsilon^{(2.2)}, \epsilon^{(3.2)}) = 0.9$**

Example 3: Three Correlated Gyros

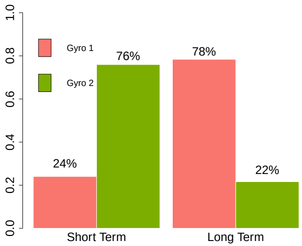


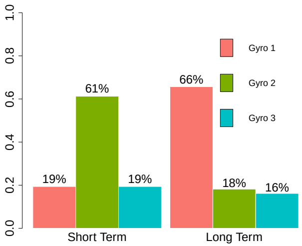


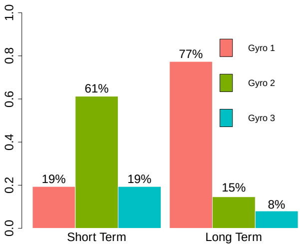
Comparison of Weights for Gyros

Settings

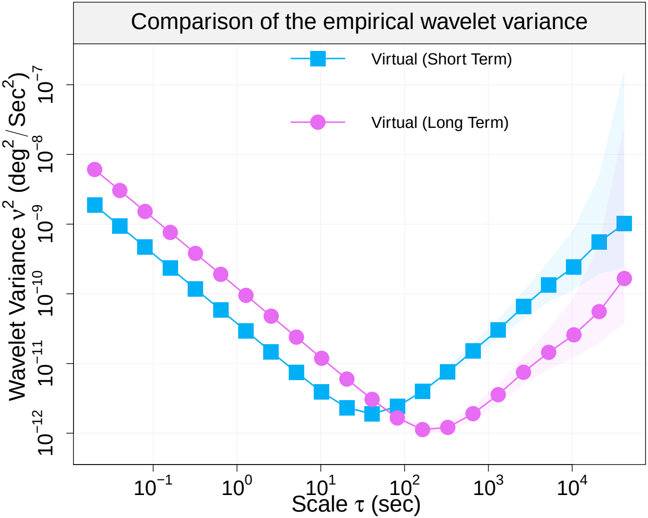
- Example 1: **Gyro 1:** WN + RW, **Gyro 2:** WN + RW
- Example 2: **Gyro 1:** WN + RW, **Gyro 2:** WN + RW, **Gyro 3:** WN + RW
- Example 3: **Gyro 1:** WN + RW, **Gyro 2:** WN + RW, **Gyro 3:** WN + RW







Example 3: Three Correlated Gyros



Remark

Problem

Recall that $\hat{\mathbf{c}}$ is found by $\hat{\mathbf{c}} = \underset{\mathbf{c} \in \mathcal{C}}{\operatorname{argmin}} \sum_{j=1}^{J} w_j \nu_j(\mathbf{c})$, where $\nu_j(\mathbf{c})$ is the WV of virtual gyro. However in general, $\nu_j(\mathbf{c})$ is unknown.

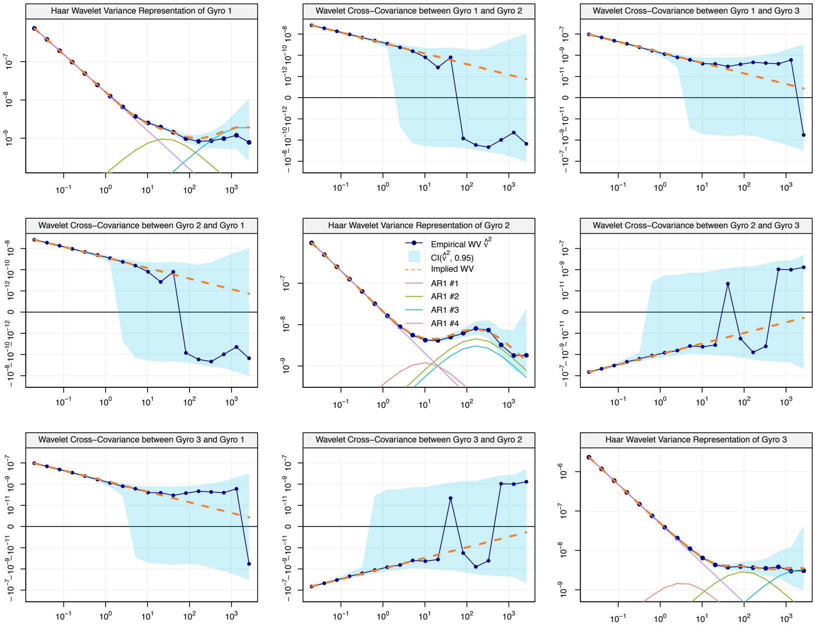
Solution

When solving the optimization problem to find $\hat{\mathbf{c}}$, we replace the WV of virtual gyro, $\nu_j(\mathbf{c})$, with its model implied estimate, $\nu_j(\hat{\boldsymbol{\theta}}, \mathbf{c})$. The **MGMWM** allows us to fit a model and estimate the corresponding parameters, $\hat{\boldsymbol{\theta}}$, by taking into account the dependence among processes, and therefore, allowing us to find $\nu_j(\hat{\boldsymbol{\theta}}, \mathbf{c})$.

Case Study

We analyze the simultaneous error signal measurements on each gyro on the X-axis of three **identical** MTi-G-710 IMUs with sample size approximately **one million** collected over a **3 hours** period at a frequency of **100Hz** corresponding to $\mathbf{J = 18}$.

By looking into the Haar WV representation of each gyro, we consider some model candidates and check how well they fit the data. Taking into account the fact that these three gyros are identical, we propose identical stochastic model structure, **the sum of 4AR1**, for all three gyros.


Haar Wavelet Variance Representation of Gyro 1
Wavelet Cross−Covariance between Gyro 1 and Gyro 2
Wavelet Cross−Covariance between Gyro 1 and Gyro 3
Wavelet Cross−Covariance between Gyro 2 and Gyro 1
Haar Wavelet Variance Representation of Gyro 2
Wavelet Cross−Covariance between Gyro 2 and Gyro 3
Wavelet Cross−Covariance between Gyro 3 and Gyro 1
Wavelet Cross−Covariance between Gyro 3 and Gyro 2
Haar Wavelet Variance Representation of Gyro 3
Empirical WV $\hat{\nu}^2$
CI($\hat{\nu}^2$, 0.95)
Implied WV
AR1 #1
AR1 #2
AR1 #3
AR1 #4

Case Study

Navigation Requirements for Different Applications

- **Short Term:** $w_j^{(1)} = \begin{cases} \frac{1}{6} & \text{if } j = 1, \ldots, 6 \\ 0 & \text{otherwise} \end{cases}$
- **Long Term:** $w_j^{(2)} = \begin{cases} \frac{1}{6} & \text{if } j = 13, \ldots, 18 \\ 0 & \text{otherwise} \end{cases}$

Evaluation Metric

Evaluation Metric

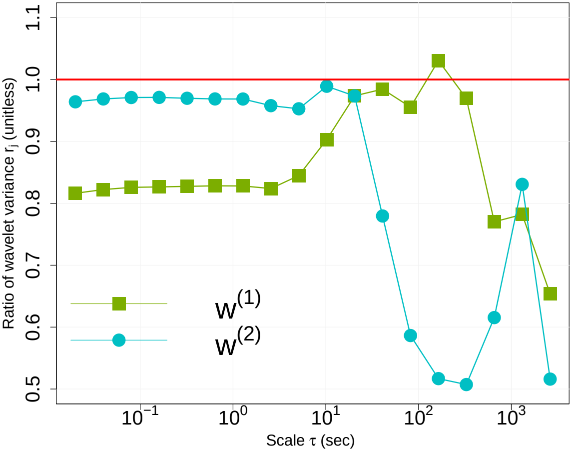
In order to compare the gain we can obtain using the proposed method, we define the evaluation metric

$$r_j \equiv \frac{\hat{\nu}_j(\hat{\mathbf{c}}_{\mathbf{w}})}{\hat{\nu}_j(\bar{\mathbf{c}})},$$

where $\hat{\nu}_j(\hat{\mathbf{c}}_{\mathbf{w}})$ is the empirical WV on scale j of the virtual gyro constructed with the weights $\hat{\mathbf{c}}_{\mathbf{w}}$ obtained based on weights on scales $\mathbf{w}$ with the proposed method, and $\hat{\nu}_j(\bar{\mathbf{c}})$ is the empirical WV on scale j of the virtual gyro constructed with weights $\bar{\mathbf{c}}$ defined as

$$\bar{\mathbf{c}} = \left[\frac{1}{p}, \ldots, \frac{1}{p}\right]^{\mathsf{T}}.$$

Case Study


1.1
1.0
0.9
0.8
0.7
0.6
0.5
Ratio of wavelet variance r_j (unitless)
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Scale τ (sec)
$W^{(1)}$
$W^{(2)}$

Mixed GMWM: Environmental Applications

Dr. Roberto Molinari
Penn State

Mixed GMWM: Environmental Applications

General context

A growing area of interest in ecological research lies in the **effects of changes in climate on biological communities** and the **effects of changes in biological communities on weather and climate**:

- Climate conditions often are measured by ecologists as time series either with meteorological measurements (e.g., temperature, precipitation, relative humidity) recorded at weather stations or with environmental sensors at fixed time intervals.
- The goal is often to **compare the mean values of climate variables** across locations or along an ecological gradient (e.g. among land use categories or biomes).

The Problem of Interest

An experiment whose aim is to compare the **vapor pressure deficit** (a metric that combines temperature and relative humidity data) across different locations with different densities of **Amur Honeysuckle** (*Lonicera maackii*), an invasive understory plant that can be found throughout the Midwest in the USA.

The Problem of Interest

- This experiment was conducted as a component of a broader study of **mosquito abundance in relation to honeysuckle density** which, if appropriately managed, is found to reduce local mosquito abundance.
- It is reasonable to think that this finding is a result of the change in Vapor Pressure Deficit (VPD) since the latter is closely related to several important life history traits in mosquitoes.

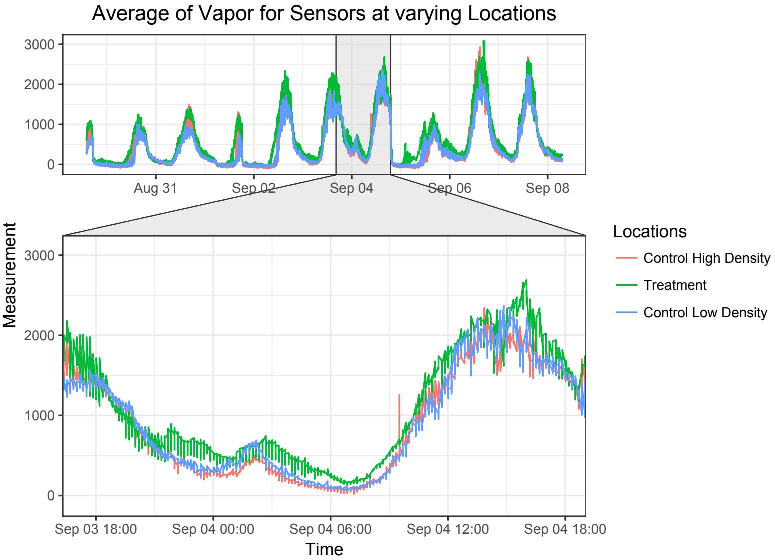
Experimental Setting

The experiment collected VPD measurements over time and was characterized by the following variables:

- **Three locations** were selected among which the first had a *high density* of honeysuckle, the second had a *low density* and the third location had an *absence* of honeysuckle (denote location with i, for $i = 1, 2, 3$).
- **Four sensors** for each location (12 in total) to measure the VPD (denote sensor for a specific location with j, for $j = 1, \dots, 4$).
- **About 1400 measurements in time** for each sensor and location taken at 10 minute intervals over a couple of weeks (denote time with t, for $t = 1, \dots, 1400$).

Along with VPD, an additional variable that was measured was the **temperature** that we denote as $(C_{i,j,t})$.

Experimental Setting



A Possible (Standard) Model

Let $\mathbf{Y} \in \mathbb{R}^N$, with $N = 16800$, be the vector containing the VPD measurements that we individually denote as $(Y_{i,j,t})$. Then we model VPD through the standard linear model:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon,$$

where

- $\mathbf{X}$ is the contrast matrix containing information on location, sensor, temperature and possible interactions,
- β is the parameter vector of interest,
- ϵ is a stochastic error which is assumed to follow the multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.

Regression Model

The Model

The model chosen to describe the average impact of the location (hence of Amur honeysuckle) on VPD is the following:

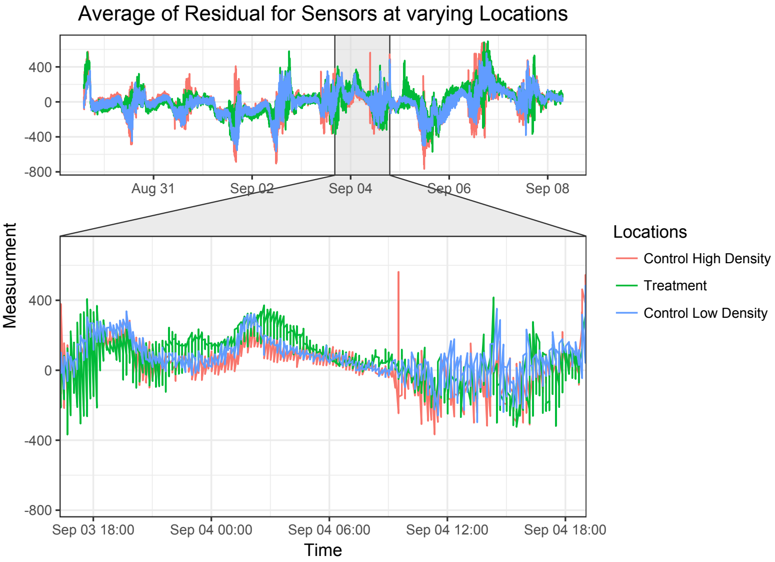
$$Y_{i,j,t} = \beta_0 + \beta_1 \mathbb{1}_{i=2} + \beta_2 \mathbb{1}_{i=3} + \sum_{l=1}^{11} \beta_{l+2} \mathbb{1}_{l=j} + \beta_{14} C_{i,j,t} + \beta_{15} C_{i,j,t}^2 + \epsilon_{i,j,t}$$

where $\beta_3, \dots, \beta_{13}$ are the average impact of the sensors on the VPD (*sum contrast*) while the coefficients of interest are β_1 and mainly β_2 (*treatment contrast*)

Robust Regression

A preliminary analysis of the data suggests using a **robust analysis** since there appear to be significant differences with the classical analysis, indicating a possible presence of outliers.

Regression Residuals



Modeling Residuals

Inference on β

It is important to correctly model the residuals in order to achieve correct and more precise inference regarding the parameter of interest $\boldsymbol{\beta}$. Assuming *iid* residuals appears particularly inappropriate in this case.

The Covariance Matrix $\boldsymbol{\Sigma}$

We can also denote the covariance matrix of the residuals as $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a generic parameter vector characterizing the covariance structure of the residuals. Some examples are:

- $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 \, \mathbb{I}$, where σ^2 is the variance of the observations and $\mathbb{I}$ is the identity matrix.
- $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbb{I} \otimes \Gamma$ where $\otimes$ is the Kronecker product and Γ is a covariance matrix for each "subject" measured over time.

Complex Residual Structure

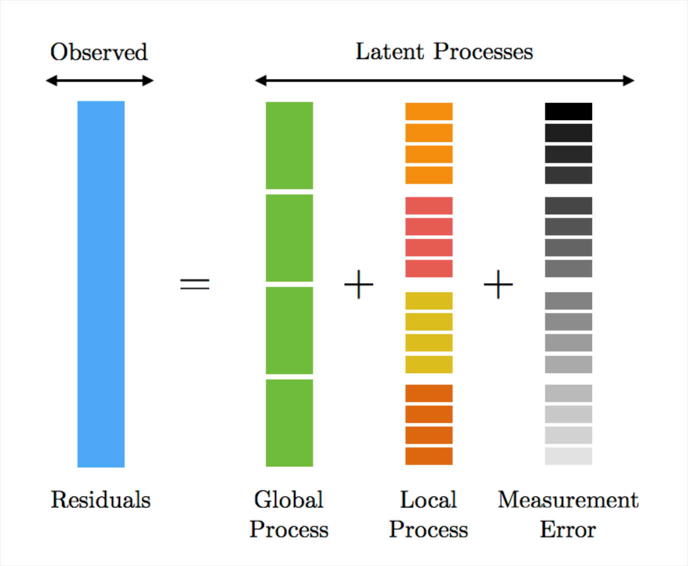
In the case of the experiment of interest, **the structure of $\Sigma(\theta)$ is quite complex** since we need to consider the following sources of dependence through time:

1. process generating the general evolution of VPD due (for example) to unobserved variables.
2. process generating the specific evolution of VPD per location due to the specific "features" of each location
3. process generating the measurement error of each sensor which is autocorrelated.

Overlapping Structure

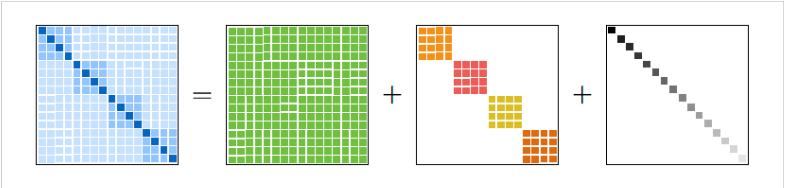
The first process is *shared among all observations*, the second is *shared among observations from the same location* while the third is *specific to individual measurements*.

Latent Processes to Model Residuals?



Complex Residual Structure

Given the setting, the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ of the residuals has the following structure:



To use methodology, we need to:

1. Extend the classical GMWM to a multivariate setting with shared latent processes,
2. Find a suitable time series model for the various latent processes.

Estimation of $\boldsymbol{\theta}$

Maximum Likelihood Estimation (MLE)

The MLE has optimal statistical properties in many cases but can be **numerically unstable** and **computationally inefficient** given the need to inverse $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ (in our case of size 16800×16800).

Generalized Method of Moments (GMM)

The GMM has good statistical properties and is often computationally efficient but can be challenging in large sample sizes to **inverse weighting matrix** and **select appropriate moments**.

Mixed GMWM

We propose to extend the GMWM to this multivariate setting.

Mixed GMWM

Mixed GMWM (MiGMWM)

The MiGMWM is defined as follows:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} \, (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}))^T \boldsymbol{\Omega} (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})),$$

where

- $\hat{\boldsymbol{\nu}} = [\hat{\boldsymbol{\nu}}_m]_{m=1,\dots,M}$ and $\boldsymbol{\nu}(\boldsymbol{\theta}) = [\boldsymbol{\nu}_m(\boldsymbol{\theta})]_{m=1,\dots,M}$
- $\hat{\boldsymbol{\nu}}_m$ and $\boldsymbol{\nu}_m(\boldsymbol{\theta})$ are the estimated and theoretical WV vectors for the m^{th} time series
- $\boldsymbol{\Omega}$ is a positive-definite weighting matrix

Differences with classical GMWM

The MiGMWM **(i)** uses the WV vector for each time series m and **(ii)** the sub-vectors $\boldsymbol{\nu}_m(\boldsymbol{\theta})$ depend on the models underlying the m^{th} time series.

Mixed GMWM

Asymptotic Properties of the MiGMWM

Based on the asymptotic properties of the GMWM, the MiGMWM has the following properties under some basic conditions:

- **Identifiable**: can identify the parameters of a wide range of multivariate latent models

$$\boldsymbol{\nu}(\boldsymbol{\theta}_1) = \boldsymbol{\nu}(\boldsymbol{\theta}_2) \iff \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$$

- **Consistent**: tends in probability to the true parameter $\boldsymbol{\theta}_0$

$$\hat{\boldsymbol{\theta}} \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$$

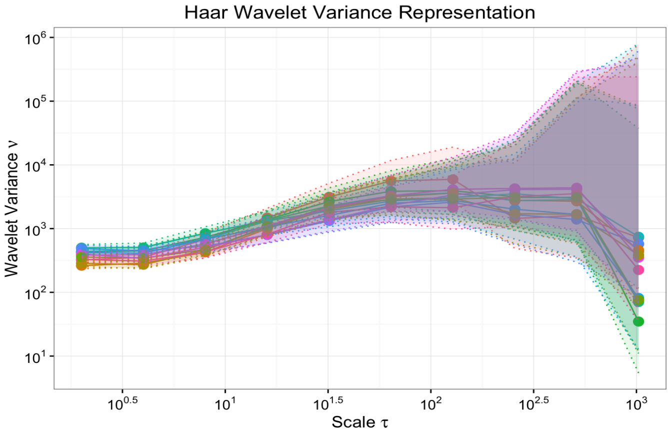
- **Normally Distributed**: tends in distribution towards a multivariate normal distribution

$$\sqrt{L}\boldsymbol{s}^T \boldsymbol{\Lambda}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

WV Analysis

Multivariate Latent Model

The plot shows the estimated WV (and relative confidence intervals) for the 12 time series in the experiment, which seem to follow similar patterns.



Latent Model for Residuals

Proposed Model:

After discussing with experts and analyzing the WV patterns, the models chosen for the different components are as follows:

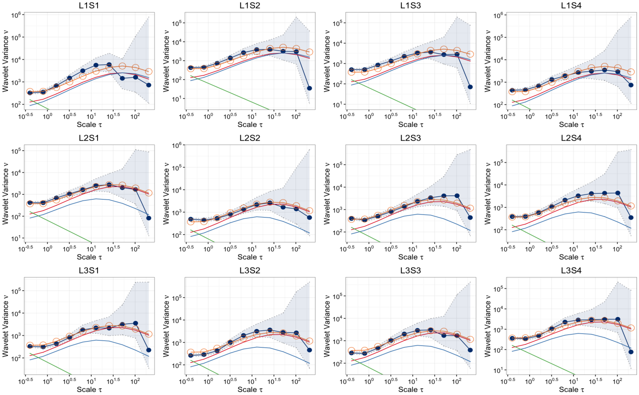
- **Global Process**: second-order autoregressive model (AR(2)) for the general VPD process.
- **Local Process**: first-order autoregressive model (AR(1)) for the VPD in a specific location i.
- **Measurement Error**: a white-noise model (WN) for the measurement error process of a specific sensor j.

This model seems to provide a close approximation to the true data generating process (e.g. null hypothesis of goodness-of-fit test cannot be rejected).

Regression Results

WV implied by $\hat{\theta}$

When using the MiGMWM to estimate this model, we can see that it fits the twelve estimated WVs well.



Interpretation of Estimated Parameters

Explained Variance

Using the estimated parameters it is interesting to compare the amount of variance explained by each process. The measurement error being nearly constant, the table below presents the average per location:

Location	Global Process (AR2)	Local Process (AR1)	Measurement Error (WN)
Control	49.5%	49.3%	1.2%
Low Density	71.2%	27.0%	1.8%
High Density	76.8%	21.4%	1.8%

This covariance structures indicates a very clear form of dependence across location and sensor.

Inference on β

Does location have an impact?

If we didn't consider the multivariate latent model in the residuals:

Parameter	Interpretation	Estimate	p-value	Conclusion
β_0	...	...	...	...
β_1	Low density AH	53.75	≈ 0	Significant impact
β_2	High density AH	87.09	≈ 0	Significant impact
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Distribution under Null Hypothesis of $\beta_i = 0$

If we don't consider the structure in the residuals, the distribution under the null hypothesis could be different and conclusions could change!

Explain Testing Procedure

Parametric Bootstrap Approach:

Suppose we are interested in testing the following hypotheses:

$$\begin{aligned} \mathrm{H}_0 &: \quad \beta_1 = 0, \quad \text{against} \\ \mathrm{H}_1 &: \quad \beta_1 > 0. \end{aligned}$$

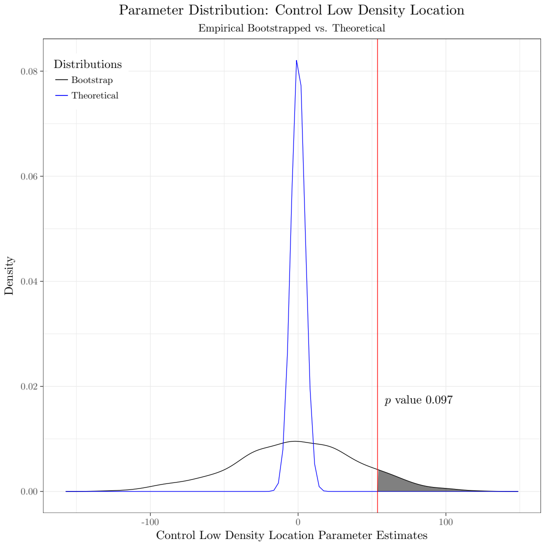
This set of hypotheses is of particular interest for this study as accepting H_1 could indicate a connection between mosquito abundance and honeysuckle density. A p-value for this test can be derived as following:

1. Simulate a sample $\mathbf{Y}^* = \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{H}_0} + \boldsymbol{\epsilon}^*$, where $\hat{\boldsymbol{\beta}}_{\mathrm{H}_0}$ corresponds to $\hat{\boldsymbol{\beta}}$ where β_1 is set to 0 and $\boldsymbol{\epsilon}^*$ is a realization of $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$.
2. Based on $\mathbf{Y}^*$ and $\mathbf{X}$ compute $\hat{\beta}_1$.
3. Repeat Steps 1 and 2, B times to derive the distribution of $\hat{\beta}_1$ under the null which can (e.g.) be used to compute a p-value.

Location with Low Density of Amur Honeysuckle

Significance of β_1:

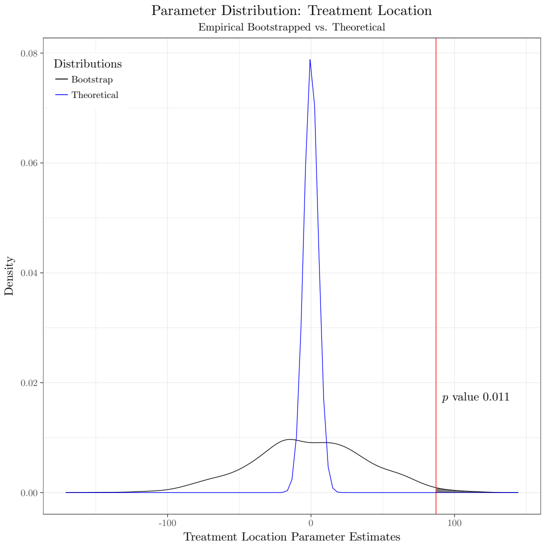
Clear difference of distributions under the null hypothesis <u>with</u> residual structure (black density) and <u>without</u> (blue density). However, the **conclusion is different**: the low presence of Amur honeysuckle does not have a significant positive impact on VPD ($\alpha = 5\%$).



Location with High Density of Amur Honeysuckle

Significance of β_2:

Clear difference of distributions under the null hypothesis with residual structure (black density) and without (blue density). In this case, the **conclusion remains the same**: the high presence of Amur honeysuckle has a significative positive impact on VPD.



Preliminary Conclusions

Experimental Conclusions

- Considering the **multivariate latent structure on the residuals can change conclusions** since low densities of Amur honeysuckle do not seem to have a significant impact.
- Based on these preliminary results, it appears that a **high presence of Amur honeysuckle can have a significant impact** on the presence of mosquitos within a certain area.

Statistical Conclusions

- The MiGMWM allows to estimate the multivariate latent residual structure and **overcomes numerical and computational limitations of other methods**.
- The MiGMWM allows to **consider new complex structures to model multivariate time series** which can have a direct practical interpretation.

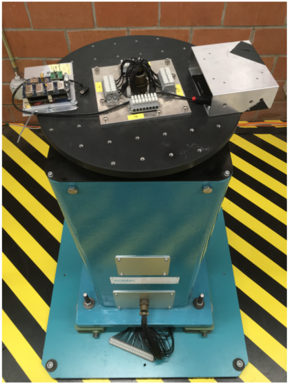
Extension to Covariate Dependency

Dr. Samuel Orso
U. Geneva

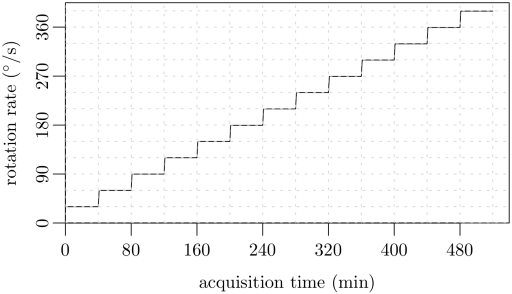
Dr. Philipp Clausen
EPFL

GMWM Example: covariate influence

Rotation table

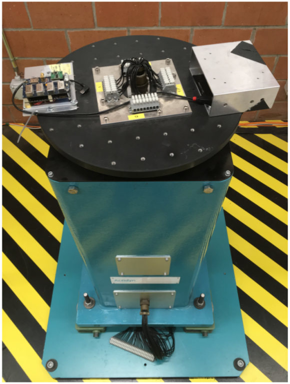


Procedure

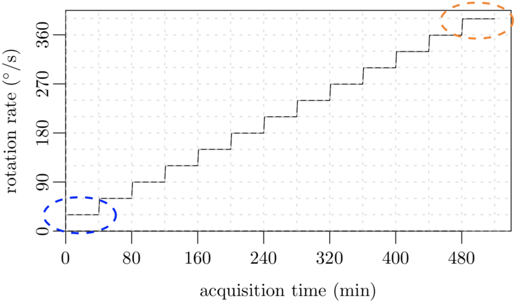


GMWM Example: covariate influence

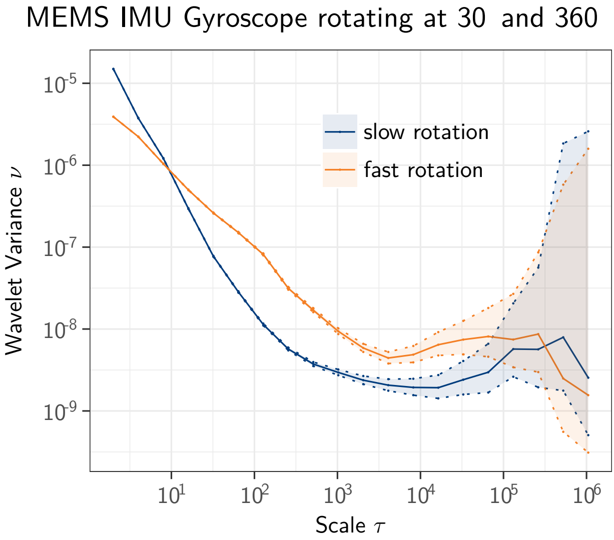
Rotation table



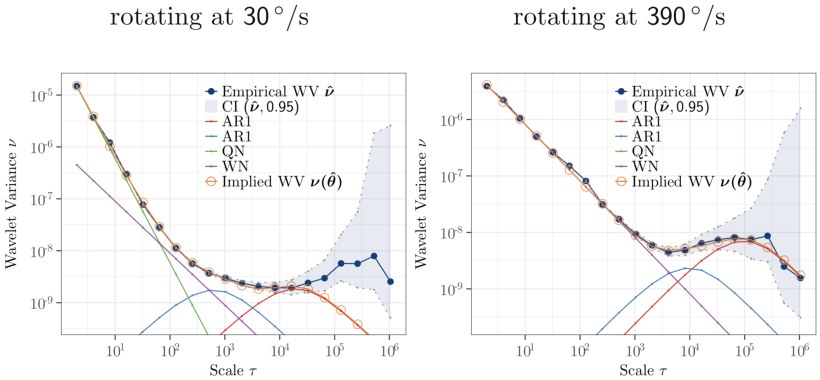
Procedure



GMWM Example: covariate influence



GMWM Example: covariate influence



Process	Units	Datasheet	rotation rate 30 °/s	rotation rate 390 °/s
WN	$^\circ/\sqrt{\text{hr}}$	0.18	0.146	0.439
QN	$(\text{rad}/\text{s})^2$	-	9.756e-06	2.220e-16

Extension to Covariate Dependency

Definition

- External process: $X_t, t \in \mathbb{N}$ (previous example - rotational speed).
- *White Noise* process:

$$V_t|X_t \overset{iid}{\sim} \left(0, \gamma_t^2\right), \quad \gamma_t^2 = g\left(\varsigma_1 + \varsigma_2 X_t\right)$$

- *Auto-Regressive* process of order 1:

$$\begin{aligned} u_t|X_t &= \phi_t \, u_{t-1} + \varepsilon_t, \quad & \phi_t &= h(\varphi_1 + \varphi_2 X_t), \\ \varepsilon_t|X_t &\overset{iid}{\sim} \left(0, \eta_t^2\right), & \eta_t^2 &= k\left(\upsilon_1 + \upsilon_2 X_t\right) \end{aligned}$$

Extended parameter vector

$$\boldsymbol{\theta} = \left[\boldsymbol{\varsigma}^T \;\; \boldsymbol{\varphi}_1^T \;\; \ldots \;\; \boldsymbol{\varphi}_d^T \;\; \boldsymbol{\upsilon}_1^T \;\; \ldots \;\; \boldsymbol{\upsilon}_d^T\right]^T \in \boldsymbol{\Theta}$$

Extension

Dynamic GMWM estimator

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \frac{1}{K} \sum_{k=1}^{K} \left\| \hat{\boldsymbol{\nu}}_k - \boldsymbol{\nu}(\boldsymbol{\theta}, c_k) \right\|^2_{\hat{\boldsymbol{\Omega}}_k},$$

where c_k explains the covariate influence on the WV of bin k, see Clausen et al. 2018 for more details.

And the properties?

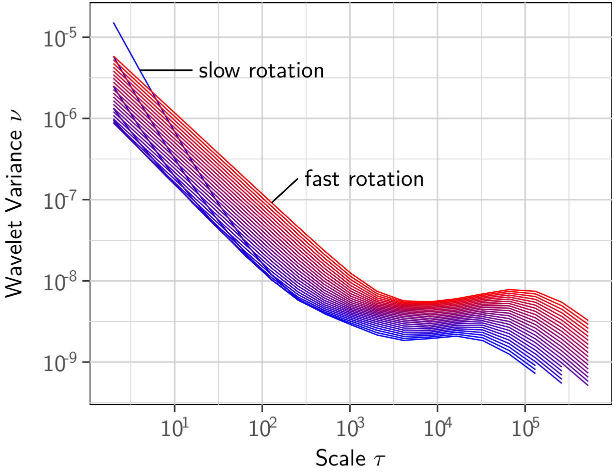
Identifiable ✓

Consistent ✓

Asymptotically Normal ✓

GMWM Example: covariate influence

MEMS IMU Gyroscope stochastic noise as a function of rotational speed



References I

Andrews, Donald WK (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". In: *Econometrica: Journal of the Econometric Society*, pp. 817–858.

Balamuta, James J et al. (2017). "A computationally efficient framework for automatic inertial sensor calibration". In: *IEEE Sensors Journal* 18.4, pp. 1636–1646.

Clausen, Philipp et al. (2018). "Construction of dynamically-dependent stochastic error models". In: *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. IEEE, pp. 1336–1341.

Guerrier, S. et al. (2013). "Wavelet-Variance-Based Estimation for Composite Stochastic Processes". In: *Journal of the American Statistical Association* 108.503.

Vaccaro, Richard J and Ahmed S Zaki (2017). "Reduced-Drift Virtual Gyro from an Array of Low-Cost Gyros". In: *Sensors* 17.2, p. 352.

Wu, Wei Biao (2005). "Nonlinear system theory: Another look at dependence". In: *Proceedings of the National Academy of Sciences* 102.40, pp. 14150–14154.

Xu, Haotian et al. (2019). "Multivariate Signal Modeling With Applications to Inertial Sensor Calibration". In: *IEEE Transactions on Signal Processing* 67.19, pp. 5143–5152.