Article scientifique   Article   2023   Submitted version   Open Access

- -

On Performance Evaluation of Inertial Navigation Systems: The Case of Stochastic Calibration

- -

Cucci, Davide A.; Voirol, Lionel; Khaghani, Mehran; Guerrier, Stéphane

This publication URL: https://archive-ouverte.unige.ch/unige:169711
Publication DOI: 10.1109/TIM.2023.3267360

On Performance Evaluation of Inertial Navigation Systems: the Case of Stochastic Calibration

On Performance Evaluation of Inertial Navigation Systems: the Case of Stochastic Calibration

Davide A. Cucci, Lionel Voirol, Mehran Khaghani, Stéphane Guerrier

Abstract—In this work we address the problem of rigorously evaluating the performances of a inertial navigation system under design in presence of multiple alternative choices. We introduce a framework based on Monte-Carlo simulations in which a standard extended Kalman filter is coupled with realistic and user-configurable noise generation mechanisms and attempts to recover a reference trajectory from noisy measurements. The evaluation of several statistical metrics of the solution, aggregated over hundreds of realizations, gives a reasonable estimate of the expected performances of the system in real-world conditions and allow the user to operate the choice between alternative setups. To show the generality of our approach, we consider an example application to the problem of stochastic calibration. Two competing stochastic modeling techniques, namely, the widely popular Allan variance linear regression, and the emerging generalised method of wavelet moments are rigorously compared in terms of the framework defined metrics and in multiple scenarios. We find that the latter provides substantial advantages and should be preferred, at least for certain classes of inertial sensors. Our framework allows to consider a wide range of problems related to the quantification of navigation system performances such as, for example, the robustness of an INS with respect to outliers or other modeling imperfections. While real world experiments are essential to assess to performance of new methods they tend to be costly and are typically unable to lead to a sufficient number of replicates to evaluate, for example, the correctness of estimated uncertainty. Therefore, our method can bridge the gap between these experiments and pure statistical consideration as done, for example, in the stochastic calibration literature.

Index Terms—Kalman Filter, Monte-Carlo Simulations, Allan variance, Generalized Method of Wavelet Moments.

I. INTRODUCTION

NAVIGATION refers to the determination of the position and the orientation of a body with respect to a well-defined coordinate reference system. Inertial Navigation System (INS) is found at the core of many modern navigation systems: inertial sensors keep track of the acceleration and the angular velocity (rotation rate) of the body, which can be integrated over time to estimate its position, velocity, and attitude [1]. INS is autonomous (largely unaffected by the surrounding environment), widely available and affordable, high-rate in output (up to a few kHz), and accurate in short-time. However, it suffer from accumulation of random errors in time, eventually leading to unreliable position and orientation estimates. This limitation is magnified with low-grade inertial sensors, as those typically employed in smartphones or drones, where position estimates can become practically unusable in as low as few dozens of seconds.

To control this inevitable position and orientation drift in accuracy, inertial measurements are typically fused with aiding information coming from other sensors. In outdoor applications, such as in planes, cars, ships, and drones, the aiding data typically comes from Global Navigation Satellite Systems (GNSS). These systems provide absolute Position, Velocity and Timing (PVT) data at lower rates if compared to INS (up to 20 Hz), but immune to error accumulation. Therefore, GNSS are complementary to INS and the combination INS-GNSS is widely used in practice. In indoor applications, or whenever the system can not depend on the reliable and continuous reception of uncorrupted GNSS signals, such aiding information can come for example from cameras, in visual-inertial systems [2], or ultra wide band beacons [3]. The fusion of data from INS and other aiding systems, such as GNSS, is typically performed via variants of Kalman filter such as the commonly used Extended Kalman Filter (EKF) [4].

One of the most critical steps in the design of a navigation system is the quantification of its performances: when several types of heterogeneous sensor, such as inertial and GNSS, are fused together, many factors influence the statistics of the estimation error and their combined impact tends to be difficult to predict. Once the system is assembled, the experimental validation is complex or impossible because of, for example, the difficulty of acquiring a sufficiently accurate ground truth or because the platform guidance system (such as the autopilot in drones) depends on the correct functioning of the navigation system, potentially leading to undesirable accidents. Furthermore, the validation of performance indicators such as the consistency of position and orientation uncertainty estimates requires the execution of hundreds or thousands of trajectories, which is unfeasible in practice.

In this work we propose a framework for quantifying navigation systems performances that relies on Monte-Carlo simulations. We consider a classical EKF for INS/GNSS navigation that the user can configure according to different, competing setups, such as sensor selection or modeling choices. Our framework allows to perform thousands of Monte-Carlo simulations using a user-specified trajectory that is representative of the chosen application scenario. In each run, we generate

D. A. Cucci is Senior Research Associate, Geneva School of Economics and Management, University of Geneva, 1205, Switzerland.

L. Voirol is Ph.D. candidate, Geneva School of Economics and Management, University of Geneva, 1205, Switzerland.

M. Khaghani is Researcher, Geneva School of Economics and Management, University of Geneva, 1205, Switzerland.

S. Guerrier is Assistant Professor, Faculty of Science & Geneva School of Economics and Management, University of Geneva, 1205, Switzerland. (E-mail: Stephane.Guerrier@unige.ch).

noise-free sensor readings from the user-supplied trajectory and corrupt those with realistic noise samples, either taken from static data (typically available and widely used in the case of inertial sensors) or sampled from an user-specified noise model. The results of all simulations are aggregated to compute several type of metrics and allowing the user to compare the results quantitatively. While this approach can not entirely substitute the experimental evaluation of the final system, the use of realistic noise samples and user-defined noise generation mechanisms allow users to nearly reproduce real-world operation scenarios.

To show the benefits of our approach, we consider an example application focusing on the problem of establishing a suitable stochastic model for inertial sensors. Indeed, inertial sensors, like any other sensor, have errors, both deterministic and stochastic. Deterministic errors, such as the stable parts of scale factors and axis non-orthogonality can be pre-calibrated and removed from the measurements directly (see e.g., [5] and the references therein). The additive stochastic part of the error, for example composed of white noise, turn-on biases and other time-correlated processes, can only be taken into account "on-flight" during navigation, provided that a suitable stochastic model for the sensor errors has been established beforehand. This process is often referred to as "stochastic calibration" and it is an important step in navigation system design and implementation. Indeed, an accurate stochastic calibration allows to, among others, to maximize the estimation accuracy by enabling the correct estimation and removal of the maximum possible portion of the stochastic errors from sensor measurements and to identify misbehaving sensors or measurements via fault detection and exclusion/isolation mechanisms.

Stochastic calibration of inertial sensors has been widely studied in the last decades and several methods are available, ranging from the Allan Variance (AV) Linear Regression method [6, 7], which we refer to as AVLR, to the more recent Generalized Method of Wavelet Moments (GMWM) [8, 9], not to mention methods based on the analysis of the power spectral density of the sensor errors [10, 11], correlation of filtered sensor outputs [12], and maximum-likelihood estimation [13]. In this context, each technique allows to consider a class of possible stochastic models and selects one within this class that appears to be the most suitable according to some specific goodness-of-fit metric. This means that different techniques typically yield different stochastic models, and that many models can be practically equivalent given the considered metric. Furthermore, it is difficult to relate these goodness-of-fit metrics used in stochastic modeling to the actual navigation performances that users will obtain in their applications. Following our proposed approach, it is possible to compare different stochastic modeling techniques according to the criteria that are highly significant from the user's perspective, such as the obtained position and orientation estimation error statistics or the consistency of the derived confidence intervals, as obtained by the navigation filter in simulated, but realistic, scenarios. By means of extensive simulation analyses based on real-world trajectories, and further backed-up by comparison against data from a real sensor, we show that in general the models obtained with GMWM perform better in navigation. However, the difference are significant when low-cost Micro-ElectroMEchanical System (MEMS) Inertial Measurement Units (IMUs) are considered, while less important if the stochastic nature of the error processes is the one typical of tactical or navigation grade sensors.

The rest of the paper is organized as follows. In Sec. II we summarize the main contributions of the paper. A brief review on stochastic calibration is given in Sec. III, specifically in relation with the two methods considered in this work, namely, the AVLR and the GMWM, discussed in Sec. III-A and III-B, respectively. In Sec. IV we introduce the proposed framework to evaluate the performances of navigation systems and we discuss its components in details. Such framework is applied in Sec. V and VI to compare the impact of the stochastic calibration technique, being it the AVLR or the GMWM, on the performances of the navigation filter, both based on simulated sensor and real-world ones. Finally, Sec. VII concludes.

II. MAIN CONTRIBUTIONS

The main contributions of this work are summarized below.

1) We introduce a framework based on Monte-Carlo simulations and realistic, user-configurable noise generation mechanisms that allow to closely replicate real-world navigation conditions. This framework can be used to assess and quantify the performances of an INS in different conditions, such as, for example, nominal, in presence of GNSS outages, when sensor measurements are contaminated with outliers or if the stochastic model for any sensor has been misspecified. The performance metrics evaluated by the framework also include measures of the correctness and of the accuracy of the uncertainty for the navigation states, as estimated by the INS, which is often difficult to verify in practice.

2) To illustrate the generality of the proposed approach we consider an example application targeting an open research question: we study the impact of different statistical procedures to determine a suitable stochastic calibration for an inertial sensors, namely, the AVLR and the GMWM. We propose to compare different techniques not in terms of the usual metrics used in system identification, such as information criteria or the likelihood, but in terms of the actual navigation performances that can be expected once the determined model is employed in an INS. Due to its generality and flexibility, our framework provides a suitable environment to evaluate the performance of new stochastic calibration methods compared to existing ones.

3) A comprehensive simulation study suggests that the AVLR, which is widely used in practice, in certain cases may lead to stochastic models characterized by worse performances in navigation, e.g., in terms of position and orientation error or accuracy of the estimated uncertainty. In contrast, this appears not to be the case if the GMWM is employed and our study sheds new light on what are the conditions under which the differences in navigation performances are substantial.

4) Our framework is implemented in an open-source `R` package, making it available widely and allowing any user to easily perform a wide range of assessments of his own INS and reproduce our results.

III. STOCHASTIC CALIBRATION OF INERTIAL SENSORS

The measurements of inertial sensors are intrinsically affected by random errors that originate from the internal physics of the device and whose nature depends on the actual technology employed [14]. For instance, in a typical gyroscope, the error terms include white noise, correlated random noise, bias instability, and angle random walk. The sources of these errors differ from one gyroscope to another. Most of this error sources are correlated in time. Stochastic calibration refers to the modeling and characterization of such error sources for a specific device prior to its usage [5]. Indeed, any navigation filter performing information fusion between inertial, GNSS and possibly other sensors needs correct stochastic models of the measurement errors. The quality of those models impacts directly the accuracy of the navigation solution and of its estimated uncertainty.

In general, it is difficult to model the physical processes behind inertial measurement error, and anyways such model would be specific to a single device or family. Therefore, stochastic modeling methods are typically employed instead. Typically, a mixture of simple stochastic processes is selected such that it provides a suitable approximation of the stochastic properties of the measurement error. For example, one common choice is the sum of a white noise and a random walk processes. The selected model is then integrated in the navigation filter of choice (e.g., an EKF), after its characterizing parameters have been estimated. In the remainder of the section we discuss two of the most common method for the calibration of inertial sections, namely the AV based and GMWM methods, whose performances are compared later in Sec. V and VI.

A. The Allan Variance Linear Regression Method

For inertial sensors, the most widespread modeling technique among practitioners and device manufacturers is the AVLR [6], conceived for the characterization of phase and frequency instability of precision oscillators, and suggested for the stochastic characterization of interferometric fiber optic gyroscopes in [11]. This is mainly a graphical technique based on the manual inspection of the AV plot. The AV plot is derived from a sufficiently long error time series: several hours of sensor data are acquired while the device is not moving, and thus no signal is observed other than the measurement error and constant quantities such as the Earth rotation rate and the gravity. A stochastic model for the measurement error, along with its characterizing parameters, can be determined based on the assumption that the different stochastic processes composing the total error appear in seemingly different regions of the AV plot as distinctive patters, such as linear regions, as illustrated in Fig. 1. Practitioners typically model the error as a sum of simple stochastic processes, e.g., quantization noise, white noise, random walk, etc., based on which patterns can

be qualitatively identified in the AV plot. After the model has been selected, its parameters, can be determined from those, for example by means of linear regression. This procedure is commonly referred as the Allan Variance Linear Regression method and it is summarized in the IEEE standard [11]: "*A first approximation [...] can be estimated by sketching in the asymptotes to the charted data analysis, and computing approximate model coefficients*". A comprehensive discussion of this approach is presented in [7].

The stochastic modeling procedures based on the AV suffer from several limitations. First, it relies on the assumption that only one underlying process is completely determining the shape of the AV at a given region of the plot. This perfect separation of the processes is actually not true as all underlying stochastic processes have an impact on the entire AV. Consequently, the AV slope method leads to inconsistent estimated parameters characterized in practice by large biases even in the case of simple stochastic models (e.g., the sum of a white noise and a random walk [15]). The second limitation is that no pragmatic rule is available to the users to solve the model selection problem, i.e., decide which stochastic processes compose the total measurement error. Indeed, in realistic scenarios where multiple underlying stochastic processes are present, it is difficult to approximate the observed AV with simple stochastic processes having a linear representation in the standard AV log-log graph. More often, the empirical Allan variance has a complex shape and the user needs to resort to their intuition and experience to select a suitable model and the relevant scales on which to perform the linear regression. Finally, the AV technique is only able to estimate the parameters of models having a linear representation in the AV log-log plot. In order to obtain reasonable point estimates, it is also needed that the parameters of the underlying stochastic processes are such that each process *dominates* in a distinct region of of the AV, which is often not the case in practice. The presence of correlated noise, e.g., first order Gauss-Markov processes, that are not linear in the AV log-log plot, can render the use of AV slope method even more challenging.

B. Moment Matching Techniques

Many methods have been proposed which aim minimizing the distance between an empirical quantity (e.g., the AV) and its model based counterpart (e.g., the AV implied by a given stochastic model) which can be expressed as an function of its parameters. An example of such methods developed in the context of inertial sensor stochastic calibration is the the Generalized Method of Wavelet Moments (GMWM), proposed in [8, 16]. This method uses the Wavelet Variance (WV) instead of the AV. This choice is due to the statistical properties of the WV which have been studied further than the ones of the AV. However, these two quantities are very similar and the interpretation of the AV presented in Sec. III-A also applies to the WV. Other moment-matching methods such as the Autonomous Regression Method for Allan Variance [17] have been proposed for the estimation of inertial sensors stochastic models. However, [9] demonstrated that these methods can be seen as special cases of the standard GMWM approach.
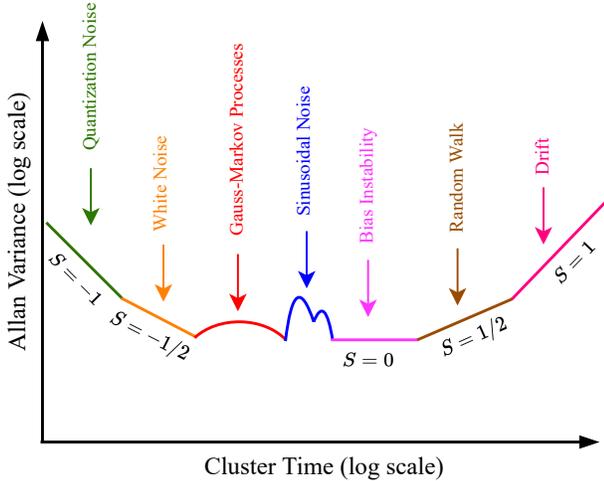
Fig. 1: Allan variance analysis noise terms results, after [14].

As previously mentioned, the underlying idea of the GMWM approach is to match the empirical and model-based WVs. Indeed, the theoretical WV for several stochastic processes, such as AutoRegressive–Moving-Average (ARMA), Quantization Noise (QN), White Noise (WN), etc., is known analytically as a function of the model parameters due to the results of [18]. Moreover, the wavelet variance of the sum of several stochastic processes is simply the sum of all the processes wavelet variances. Thus, the model parameters say $\theta \in \Theta$ can be obtained as follows:

$$\hat{\theta} = \underset{\theta \in \Theta}{\mathrm{argmin}} \left[\hat{\nu} - \nu(\theta) \right]^T \Omega \left[\hat{\nu} - \nu(\theta) \right], \tag{1}$$

where θ is the vector of model parameters, $\hat{\nu}$ is the empirical wavelet variance of the time series, $\nu(\theta)$ is the theoretical wavelet variance implied by the parameter vector θ and Ω is a weight vectors that depends on the uncertainty of $\hat{\nu}$. The optimization problem in (1) is typically solved iteratively using gradient-based or Gauss-Newton methods. If the wavelet variance of a given model is linear in the model parameters, which is the case for arbitrary sums of quantization noise, white noise, random walk and drift, the optimization problem can be solved in closed form [9]. Additional information on this method can be found for example in [8, 16].

IV. Assessment Framework for Stochastic Model Impact on Navigation

As previously mentioned, stochastic calibration is required in order to characterize the inertial sensor measurements uncertainty, so that the chosen sensor fusion algorithm can properly integrate them with other sensor measurements and estimate a reliable navigation solution. While many techniques exist for such purpose, such as those reviewed in Section III, it remains difficult to quantify the actual impact of a given choice of stochastic models on the navigation performances of the system in a given application scenario. For instance, the stochastic calibration procedure may yield an accurate model of the long term correlations of the measurement error,

whereas a much simpler model could perform similarly for the user application where, for instance, a drone is flown only for a few minutes at a time. As another example, many alternative models may exist for a given sensor that perform comparably with respect to the metric employed in stochastic calibration (i.e., similar value for the objective function defined in (1) at the solution), making the choice difficult and potentially subjective for the users.

In this section, we propose an approach to quantify the impact on navigation performances of different, competing setups, such as sensor selection or modeling choices. The proposed approach is depicted in Fig. 2. To illustrate the generality of our framework, we will later consider the problem of comparing the stochastic calibrations obtained from different statistical procedures, namely, the AV slope method and the GMWM. More precisely, we consider a user-specified trajectory which is typical of the application scenario, available as samples of the body position and orientation. This trajectory will be used as a ground truth. The available samples are differentiated to obtain higher order kinematic properties, such as velocities and accelerations, which are used to calculate *noise-free* readings, as they would be measured by perfect sensors mounted on the body. The *noise-free* readings are then corrupted with noise samples taken contiguously from random portions of static acquisition data (e.g., as collected during standard stochastic calibration procedures), and thus consisting of samples of only the actual sensor noise. A sensor fusion algorithm is configured with the stochastic model under investigation, for example determined with the AV slope method. The sensor fusion algorithm processes the noisy readings and estimates the final navigation solution. This procedure is repeated in a Monte-Carlo fashion. The set of obtained navigation solutions is aggregated to compute a suite of statistics relevant to assess the expected performances of the system in the given scenario, for example in terms of mean position and orientation error, or consistency of the estimated uncertainty. In order to evaluate the impact of different choices of sensor stochastic models, the user needs just to configure appropriately the sensor fusion algorithm and compare the different statistics obtained after the Monte-Carlo simulations. As previously mentioned, the proposed approach has been implemented in the form of an open-source R package[1]. This software allows to perform easily all the steps outlined in Fig. 2 and to compute many relevant statistics to compare competing stochastic models. We use this software package to perform all the investigations presented in the rest of this work and therefore our results can be easily replicated. In the remainder of the section, we discuss the different components of the framework in further detail.

A. Noise-free Measurements Generation

We assume in this section that a suitable trajectory is available. This trajectory can be chosen to resemble the user's use case, for example in terms of dynamics of the body motion and duration. The latter is available as a sequence of positions, r_t, and orientations, $R_{b,t}^n$, of the body frame b at the target

---

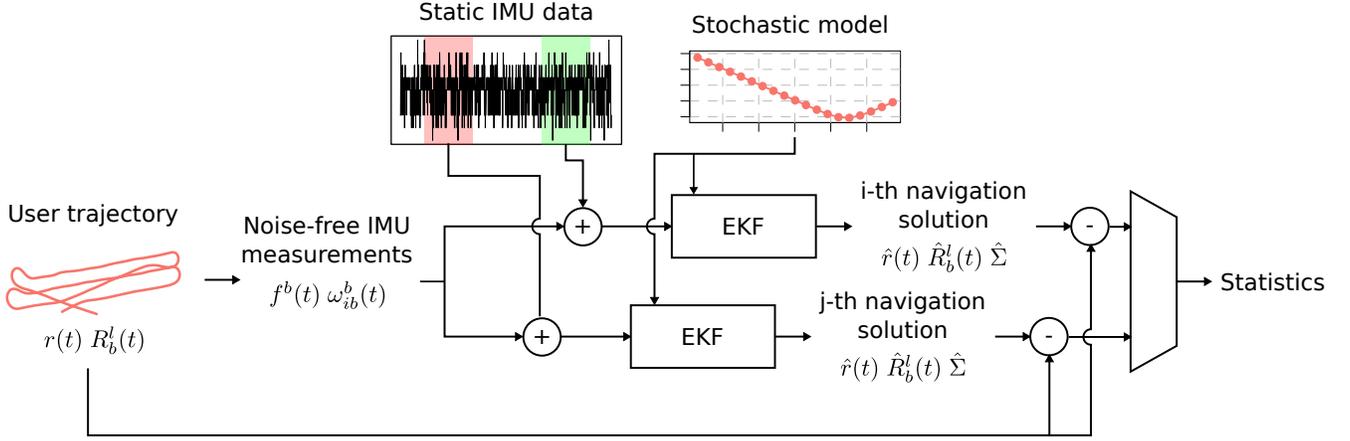[1]Navigation "R" package: https://github.com/SMAC-Group/navigation

Fig. 2: The proposed framework for the evaluation of inertial navigation systems.

inertial sensor rate. These are expressed with respect to a fixed, Cartesian, non-rotating navigation frame n. The trajectory is assumed to be the ground truth in the following, therefore neither its accuracy or how it was originally determined are relevant for the following.

From the position and orientation samples we derive higher order kinematic properties of the body frame as follows:

$$v_t^n = \frac{r_{t+1} - r_t}{\Delta t}, \quad a_t^n = \frac{v_{t+1}^n - v_t^n}{\Delta t},$$
$$\omega_{nb,t}^b = \frac{\log(R_{b,t}^{n\,T} R_{b,t+1}^n)}{\Delta t}, \tag{2}$$

where v_t^n and a_t^n are the body frame velocity and acceleration at time t expressed in n, respectively. Moreover, $\omega_{nb,t}^b$ is the body frame angular velocity of b with respect to n, expressed in b, $\log(\cdot)$ is the logarithmic map in the 3D rotation group SO(3) and Δt is the trajectory sampling time.

Next, *noise-free* inertial sensor readings are computed. While gyroscope measurements are already given by $\omega_{nb,t}^b$ in (2), the specific force accelerometer reading is given by:

$$f_t^b = R_{b,t}^{n\,T}(a_t^n + g^n),$$

where g^n is the gravity acceleration expressed in n. Here we ignore effects such as the Earth rotation, apparent Coriolis forces, coning and sculling [19], or a position dependant gravity vector. While these effect are important in real world navigation systems, at least above certain accuracy requirements, they are deterministic and, once they have been accounted for in the sensor fusion algorithm, they play limited or no role when looking at the performances of inertial stochastic models. Thus, they are ignored in the following. For simplicity, position and orientation derivatives are evaluated in (2) by means of first and second order forward finite difference scheme. If desired, more sophisticated differentiators can be employed, see for example [20]. However, the impact of this choice is limited, as long as the chosen differentiation scheme matches the forward integration mechanization employed in the sensor fusion algorithm detailed in Sec. IV-C.

B. Noisy measurement generation

We assume that the user has collected data from the inertial sensor under evaluation in static conditions for stochastic calibration purposes. Since the device is static, the collected data consists of actual noise samples that can be used to corrupt *noise-free* measurements.

Under the assumption that inertial sensors performances are relatively independent of environmental conditions, the noise samples collected in static conditions have similar, if not the same, stochastic properties of the noise during real-world operation. Thus, we add contiguous chunks of the available static data to the *noise-free* inertial measurements to obtain the noisy ones. A different start time for each chunk is randomly selected for each Monte-Carlo simulation. However, the assumed independence between IMU stochastic models and environmental conditions does not hold entirely, at least for low-cost IMUs. Indeed, the stochastic properties of the measurement error and/or the deterministic calibration (e.g., scale factors) may depend on temperature (see e.g., [21]) and motion dynamics [22]. The effect of the device temperatures can readily be investigated with the proposed framework simply collecting static data in a temperature chamber and varying the temperature according to the user application. However, the dependency on motion dynamics is more difficult both to model and then to compensate for in a sensor fusion system, and it is not considered in the following. Alternatively, reference stochastic model can be assumed for the inertial sensors and sampled to obtain noise data. This is useful if static data is not available, or for example, if one wants to quickly evaluate different sensors available on the market.

An a priori stochastic model is also used to generate noise samples for the GNSS sensor. Indeed, differently from inertial sensors, it is well known that the noise properties of GNSS measurements are very much dependent on environmental conditions, such as receiver surroundings, satellite constellation, ionosphere and troposphere conditions, and so on, so that static data is probably not representative of real world operations. Since this work focus on inertial sensors, the GNSS

measurement error is assumed to be white and uncorrelated. More realistic GNSS error samples could be obtained using a GNSS simulator and then used to corrupt *noise-free* position and velocity measurements.

C. Sensor fusion algorithm

Let N denote the number of Monte-Carlo simulations. For simulation i, with $i \in \{1, ..., N\}$, a sensor fusion algorithm is employed to fuse together noisy inertial and GNSS observations and estimate the navigation solution, $^{(i)}\hat{r}_t$ and $^{(i)}\hat{R}_{b_t}^n$, as well as their estimated covariance $^{(i)}\Sigma_t$.

For the sensor fusion algorithm we chose the EKF as this method is at the core of most of the real time navigation systems and its behaviour is well known and understood among practitioners. The choice of the sensor fusion algorithm does not affect the principle of operation of the approach proposed in Fig. 2 and any other method, such as unscented Kalman filters, particle filters, or even smoothing based methods such as dynamic networks [23], could replace the EKF and directly allow to compare the performances of different state estimation algorithms.

We consider a standard error-state formulation of the EKF as presented for example in [1]. In such EKF, the inertial sensor process noise models can be configured as an arbitrary sum of the following stochastic processes:

1) random constant,
2) White Noise (WN),
3) AutoRegressive process of order one (AR1), which is equivalent to a first-order Gauss-Markov process,
4) Random Walk (RW),
5) drift.

Multiple AR1s (or Gauss-Markov processes) can be considered, as it will be demonstrated in Section VI-A2. By combining the previously mentioned processes, a wide class of models can be constructed which includes standard models for inertial sensor errors Other Gaussian processes, such as AutoRegressive Moving-Average (ARMA) models of order higher than one can be considered, see for example [24]. The proposed framework could be easily extended to include such processes, however, this choice does not appear to be widespread in practice and is left for future research.

The implemented EKF employs a first order Euler method to integrate the process model. This method is simplistic and higher order integration methods are preferable in real-world applications. However, here the goal is to evaluate the impact of the stochastic models for the sensor errors, not the quality of the filter mechanization. Thus, the latter has been chosen to match the *noise-free* measurement generation mechanism presented in Section IV-A, so that it produces exactly the original position and orientation samples when integrating *noise-free* measurements. This choice prevents spurious integration noise to impact further analyses.

D. Navigation performance statistics

The navigation solutions are then aggregated to compute statistics that are useful to evaluate the performances of the navigation system. The details are given in the following:

1) Position and orientation error: The original trajectory provided by the user gives the ground truth, and each estimated solution can be compared to the reference in terms of position and orientation error:

$$
\begin{aligned}
^{(i)}\Delta r_t &= {}^{(i)}\hat{r}_t - r_t, \\
^{(i)}\Delta R_t &= \log\left[\left(^{(i)}\hat{R}_{b,t}^n\right)^T R_{b,t}^n\right]^\vee,
\end{aligned}
\tag{3}
$$

where the $(\cdot)^\vee$ operator gives the three elements composing the skew-symmetric matrix returned by $\log(\cdot)$, thus being a 3D representation of the difference between the estimated and reference orientations. In the following, for simplicity we will average both position and orientation error over the three axes.

The error samples computed with (3) can then be aggregated with respect to the different navigation solutions, with respect to time, or both, by means of any sample based operation, such as the sample mean, sample covariance, and so on. Given the statistical properties of such sample based estimators, and provided that N is sufficiently large (e.g., several hundreds), those quantities will approach the true value of the underlying quantities.

E. Normalized Estimation Error Squared

The Normalized Estimation Error Squared (NEES) is a commonly employed metric to evaluate whether the error in position and orientation is consistent with their estimated covariance. An in-depth discussion can be found in [4, Chapter 3.7.4].

The NEES is defined as follows:

$$
^{(i)}\text{NEES}_t = \left[\begin{array}{c} ^{(i)}\Delta r_t \\ ^{(i)}\Delta R_t \end{array}\right]^T {}^{(i)}\Sigma_t^{-1} \left[\begin{array}{c} ^{(i)}\Delta r_t \\ ^{(i)}\Delta R_t \end{array}\right],
\tag{4}
$$

In a Monte-Carlo setup with N simulations, the average NEES at time t, which we denote as $\overline{\text{NEES}}_t$, follows (approximately) a chi-square distribution with $6N$ degrees of freedom (6 being the dimension of the stacked $^{(i)}\Delta r_t$ and $^{(i)}\Delta R_t$ vectors). Thus, a two-sided confidence interval for $\overline{\text{NEES}}_t$ with level $1 - \alpha$ can be expressed as $[\epsilon_1, \epsilon_2]$ where

$$
\epsilon_1 = \chi_{6N}^2\left(\frac{\alpha}{2}\right) \quad \text{and} \quad \epsilon_2 = \chi_{6N}^2\left(1 - \frac{\alpha}{2}\right),
\tag{5}
$$

where $\chi_n^2(\alpha)$ is the α-th quantile of a chi-squared distribution with n degrees of freedom. A one-sided confidence interval can be constructed similarly.

The evolution of $\overline{\text{NEES}}_t$ with respect to the bounds $[\epsilon_1, \epsilon_2]$ is an important metric used in practice to assess whether the sensor fusion algorithm is *overconfident*, meaning that the actual error is typically larger than the estimated uncertainty of the position and orientation estimates, or *underconfident*, the other way around, or none of the two.

F. Coverage

The consistency and the efficiency of the estimated position and orientation, and of their estimated covariance, can be evaluated in an alternative way, as commonly done in the

statistical literature. We first define the one-sided confidence interval of level $1 - \alpha$ for $^{(i)}\text{NEES}_t$ as follows:

$$[\epsilon'_1, \epsilon'_2] = \left[0, \chi^2_N \left(1 - \alpha\right)\right]. \tag{6}$$

Next, we introduce the following binary variable:

$$^{(i)}c_t = \begin{cases} 1 & \text{if } ^{(i)}\text{NEES}_t \in [\epsilon'_1, \epsilon'_2], \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

If the estimated covariance $^{(i)}\hat{\Sigma}_t$ is sufficiently well estimated $^{(i)}c_t$ follows approximately a Bernoulli distribution with parameter $p = 1 - \alpha$, allowing to assess the simulation error. The *coverage* of the confidence intervals defined in (6) is given by the average of $^{(i)}c_t$ over the N Monte-Carlo runs, $\bar{c}_t$.

The coverage, $\bar{c}_t$, measures how often, in practice, the confidence intervals constructed from $^{(i)}\hat{\Sigma}_t$ and centered at $^{(i)}\hat{r}$ and $^{(i)}\hat{R}^n_{b,t}$ include, or *cover*, the true position and orientation. In other words, such confidence intervals are trustworthy if $\bar{c}_t$ is close to $1 - \alpha$. This measure provides an alternative approach to assess the quality of the computed navigation solution.

V. Case Study: Allan Variance Linear Regression Method versus GMWM

In this section we apply the proposed framework to compare different statistical estimators leading to different stochastic models for the inertial sensors in terms of navigation performances. More precisely, we compare the AV slope method and the GMWM, both briefly presented in Sec. III. This example is relevant since it allows to investigate whether, and when, a more sophisticated statistical method such as the GMWM should be used instead of the commonly employed AV slope method. Indeed, we consider an experimental setup composed of two steps:

1) a calibration step, in which the stochastic models are determined from available noise data (for example collected in during static acquisitions) using the two previously mentioned estimators,
2) Monte-Carlo navigation simulations, where the framework depicted in Fig. 2 is employed to quantify the navigation performances of the estimated models.

In our experiment, the estimation of the parameters of the stochastic models taking place in the first step is performed based several hours of data collected from a static device. Since the device is static, the collected measurements will be composed by noise samples only, and the selected stochastic modeling techniques are employed to determine the two sets of stochastic models (each set being composed of the gyroscope and accelerometer model) that we seek to compare. Next, we evaluate the determined stochastic models in a realistic scenario using the proposed Monte-Carlo simulation framework. We consider a trajectory of a real fixed-wing Unpiloted Aerial Vehicle (UAV) performing an aerial mapping mission lasting approximately 40 minutes. By performing hundreds of simulations, each time employing different noise realisations, we obtain reliable statistics of the navigation performances that can be obtained employing each of the two sets of models. From these statistics, which model should be preferred for implementation in the final application can be assessed.

As discussed in details in Sec. VI, we consider two scenarios. In the first scenario, we consider synthetic inertial sensors for which the true noise model is assumed to be known. The chosen models aims to mimic the observed AV/WV shape of real-world low-to-mid grade inertial sensors. These models are sampled to generate both synthetic noise data for the stochastic calibration and for the generation of noisy inertial readings for the Monte-Carlo simulations. In the second scenario, a real sensor is considered and multiple static acquisitions are performed to collect both the data for the stochastic calibration and the realistic noise samples to corrupt inertial readings during navigation.

A. A Flexible Set of Stochastic Models

We define a general set of models that is equal or can well approximate the vast majority of models used in inertial sensors. Such set is defined as the sum of M AR1s, which we write as first-order Gauss-Markov processes because of the practical meaning of the parameters (correlation time and variance of the innovation):

$$\begin{aligned} e(t) &= \sum_i^N x_i(t), \\ \dot{x}_i(t) &= -\frac{1}{\tau_{c,i}} x_i(t) + \xi_i(t), \quad \tau_c \geq 0, \end{aligned} \tag{8}$$

where $e(t)$ is the measurement error affecting the inertial sensor at time t, $\xi_i(t)$ is a (continuous time) WN with power spectral density q_i and $\tau_{c,i}$ is the i-th process correlation time.

This class of models is very general and includes most of the processes typically considered in modeling inertial sensors, where some well-known special cases are given in the following:

1) **White noise** (WN), typically referred to as *angular random walk* for gyroscopes and *velocity random walk* for accelerometers, is obtained when $\tau_{c,i} \to 0$,
2) **Random walk** (RW), or *rate random walk*, in the case of gyroscopes, when $\tau_{c,i} \to \infty$ (any long-term correlation observed in practice can be modeled with a sufficiently high $\tau_{c,i}$),
3) **Bias instability**, or flicker noise, it is defined in terms of its power spectral density:

$$S_\Omega = \begin{cases} \frac{B^2}{2\pi f} & f < f_0, \\ 0 & f \geq f_0. \end{cases} \tag{9}$$

Since no state-space model can be derived for this process, it can not be employed directly in state estimation algorithms, e.g., in an EKF. As suggested in [25, Section 4.3], or in [11], it is "sometimes approximated by a Markov model or a multiple stage ARMA model", such as the one in Eq. (8).
4) **Turn-on bias**, or random constant, the constant part of the sensor error that changes at every power cycle can be modeled setting $q_i = 0$ and $\tau_{c,i} \to \infty$.

Two other commonly employed processes do not fall in the proposed model family. Those are the quantization noise and the drift, or *rate ramp* in gyroscopes. Since both of these can
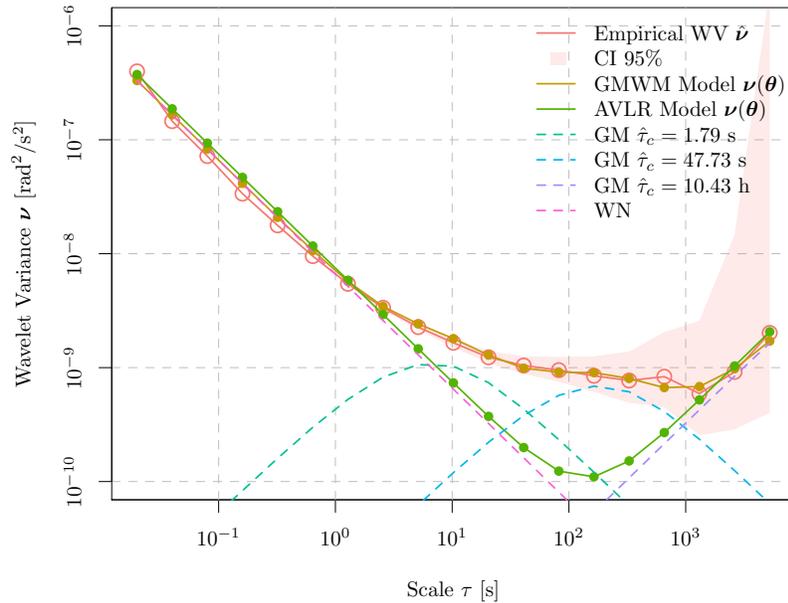
Fig. 3: Empirical WV of 2 h static acquisition for the KVH1750 accelerometer (in red) and theoretical WV as implied by the model obtained with AVLR (green) and the GMWM (brown). The decomposition of the GMWM model in 4 first-order Gauss-Markov processes, one of which degenerates into a white noise, is also shown for the GMWM model.

be estimated with both the AVLR and the GMWM, they are not considered in this study.

In stochastic calibration, is more customary to consider the equivalent, discrete-time formulation of (8):

$$
e_t = \sum_i^N x_{i,t}, \tag{10}
$$
$$
x_{i,t} = \phi_i x_{i,t-1} + \xi_{i,t}, \quad \xi_{i,t} \sim \mathcal{N}(0, \sigma_i^2),
$$

where the well known relation between the correlation time and the innovation power spectral densities, and the discrete time $\phi_i \in [0,1]$ and $\sigma_i^2 > 0$ are given below:

$$
\phi_i = \exp\left(-\frac{1}{\tau_{c,i} f}\right),
$$
$$
\sigma_i^2 = \frac{q_i \tau_{c,i}}{2}\left[1 - \exp\left(-\frac{2}{\tau_{c,i} f}\right)\right], \tag{11}
$$

where f is the sampling frequency. For a given Gauss-Markov process i, the two different parameterizations (ϕ_i, σ_i^2) and $(\tau_{c,i}, q_i)$ are equivalent given f and they will be used interchangeably depending on which one is more intuitive in the given context.

The GMWM provides estimates having suitable statistical properties for the set presented in (10), when the number of Gauss-Markov processes, M, is arbitrary but known, and also to determine the optimal M for a given calibration data. On the contrary, the AVLR method would typically estimate model where $M = 2$ Guass-Markov processes belonging to the WN (or angular or velocity random walk) and RW (rate random

walk) special cases. The AVLR is also capable of estimating bias instability in terms of B and f, as in (9). However, it is difficult to relate those to a suitable state-space model that can be employed in an EKF.

Given the specific features of the two stochastic calibration methods presented before, their models may be different. To illustrate this difference, we considered approximately two hours static data collected using the Novatel KVH 1750 IMU. In Fig 3 we compare the empirical WV (which can be interpreted similarly to the AV) of the X axis accelerometer with the models estimated with AVLR (in green) and GMWM (in brown). It can be observed that the AVLR model does not provide a close mach to the the empirical WV between scales 10^1 s and 10^3. Indeed, the AVLR method does not allow to estimate state-space stochastic models that would well approximate the flat region in the WV evident at those scales. On the contrary, the GMWM estimates appear to provide a better model fit considering a model composed of $N = 4$ Gauss-Markov processes, the first of which degenerates into a white noise ($\tau_{c,1} = 0$). The theoretical WV of the resulting model matches the observed one to an excellent degree. In particular, the GMWM allows to obtain a model for the long-term correlation of the error, for example, between scales 10^1 s and 10^3 s. No such model is provided by the manufacturer [26] or can be estimated reliably with the available graphical methods based on the AV.

In the next section, we will investigate in which cases the extra modeling power offered by GMWM would lead to superior performances in navigation.

VI. Simulation Results

In this section we compare the performance of stochastic calibration based the AVLR and the GMWM, respectively in two different scenarios. In the first scenario, we assume that the stochastic model behind noise generation in inertial sensors is known and belongs to the set of models introduced in Sec. V-A. Many different parameter values are considered. In the second scenario we consider a real sensor, the Xsens MTI-g IMU [27], for which data from a static acquisition is available. This sensor is characterized by a significant bias-instability and the WV of the static acquisitions match the ones of some of the models considered in the first scenario.

A. Scenario 1: Known Stochastic Models

We consider a simulation scenario in which the true stochastic model behind inertial sensor noise is known. In the following, we first consider the class of models introduced (10), with $M = 2$, in Sec. VI-A1. Next, in Sec. VI-A2 we consider a more complex model family where $M = 3$. In both cases, the first Gauss-Markov process (or AR1) always degenerates to a WN, e.g., $\tau_{c,1} \to 0$.

1) Model 1: 1 WN + 1 AR1: The stochastic model for the gyroscope is assumed to be the sum of a white noise and a Gauss-Markov process: 24 different instances from this model class, each one being characterized by different parameters, have been chosen as follows:

- the standard deviation of the white noise is fixed at $\sigma_\xi = 10^{-6}$ rad s^{-1}.
- For the first 12 instances (large set from now on) the variance of the AR1 process is large compared to the white noise, and it is small for the remaining 12 (small set from now on). The variance of the AR1 process is given by $\sigma_\xi^2/(1-\phi^2)$ and it is kept constant to 5×10^{-8} rad^2 s^{-2} for the large set and to 5×10^{-9} rad^2 s^{-2} for the small set.
- The j-th instances of both the large and the small sets, with $j \in \{1, 2, ..., 12\}$, have the same value of ϕ, and ϕ_j spans $[0, 1]$. In other words, we consider processes with increasing correlation time τ_C, in the first order Gauss-Markov parameterization.

Moreover, we consider a fixed model for the accelerometer noise, being a white noise with standard deviation $\sigma_\xi = 5 \times 10^{-5}$ m s^{-2}. The resulting stochastic processes are realistic for MEMS IMUs. The theoretical WVs of the considered models are depicted in Fig. 4, the large set being on the left and the small set on the right.

For each of the model instance we sample a time-series, mimicking the static acquisition process that would provide the stochastic calibration data for a real inertial sensor. Next, we use the generated noise time-series to estimate a stochastic model using the GMWM and the AVLR. The estimated model is used in an EKF to process noisy GNSS and inertial readings, corrupted by noise sampled the same way as for the synthetic static acquisition data. The performances of the navigation solution are measured according to the metrics presented in

the Sections IV-D, IV-E, and IV-F. This procedure is repeated 500 times for each model instance in a Monte-Carlo fashion.

From Fig. 4 it is possible to see that the WVs of some of the model instances (e.g., the one in violet) have large parts that are linear in the scales τ and that can be well approximated by the sum of a WN and a RW using the AVLR. However, this does not hold for most of the others (e.g., the ones highlighted in red, green and light blue): for those cases, the model obtained by means of the AVLR is different from the true one, in terms of its WV. In contrast, by means of the GMWM it is possible to recover a good approximation of the true model for all the model instances.

If the model employed in the EKF is substantially different from the true one, used to generate noisy inertial readings, the quality of the navigation solution may degrade. In the following, we quantify such degradation in terms of position and orientation error within GNSS coverage, i.e., when a position fix is available, and during a GNSS outage period of 60 s. The position and orientation error for each model instance are presented in Fig. 5 and 6. These quantities corresponds to $^{(i)}\Delta r_t$ and $^{(i)}\Delta R_t$, as defined in Sec. IV, averaged over the Monte-Carlo simulations and the three axes. The results are expressed in relative units, in terms of the performances of the model estimated by means of the AVLR versus the one estimated with the GMWM. For example, a value of 110% implies that the model estimated with the AVLR achieves, on average, a position (or orientation) error 10% higher than the model estimated with the GMWM. In these figures, and the ones that will follow, the x axis is the time relative to the beginning of the GNSS outage, at second 0, while on the y axis we have the time constant τ_C of the auto-regressive process contained in that specific model instance.

From Fig. 5 and 6 we can observe the following.

- When the WV of the true model can be well approximated by the sum of a WN and a RW, i.e., when $\phi \to 1$ or, equivalently, τ_C is high, both the GMWM and the AVLR achieve similar performances both in terms of position and orientation error. This result is somewhat expected as in this case both the AVLR and the GMWM are able to provide suitable approximations of the underlying data generating process.
- When the WV of the true model presents a peak or anyways it is not linear in the scales τ (see again Fig. 4, e.g., the model highlighted in green), only GMWM can estimate an accurate model and the performances degrade by up to 50% in position and 80% in orientation with AVLR method as this method, unlike the GMWM, is unable to provide an accurate approximation of the underlying data generating process.
- The higher the variance of the AR1 process is (large set compared to small set, on the left and on the right in Fig. 5 and 6, respectively), the higher the degradation of the performances will be. This is intuitive since the AR1 process is the one that can not be estimated with the AVLR technique.
- The degradation in terms of position error is mostly visible during the GNSS outage period only. Indeed, the GNSS position fix corrects for any error accumulated
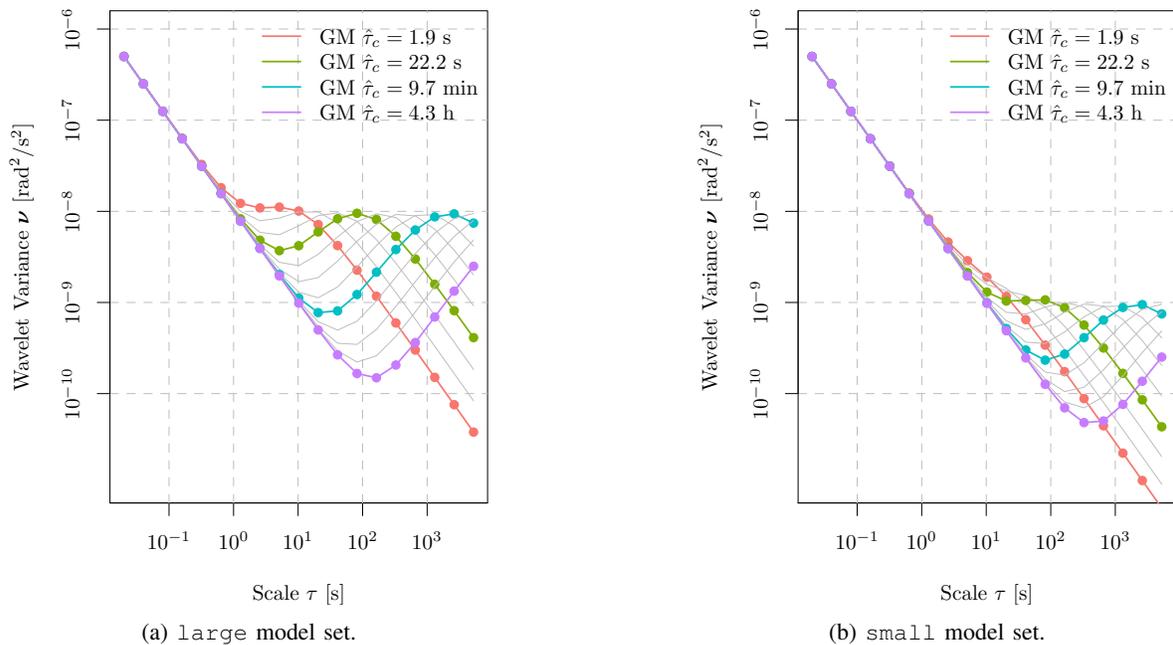
(a) large model set.

(b) small model set.

Fig. 4: Wavelet variances of the considered noise models (grey lines). Four out of twelve instances are highlighted with colors in each figure.

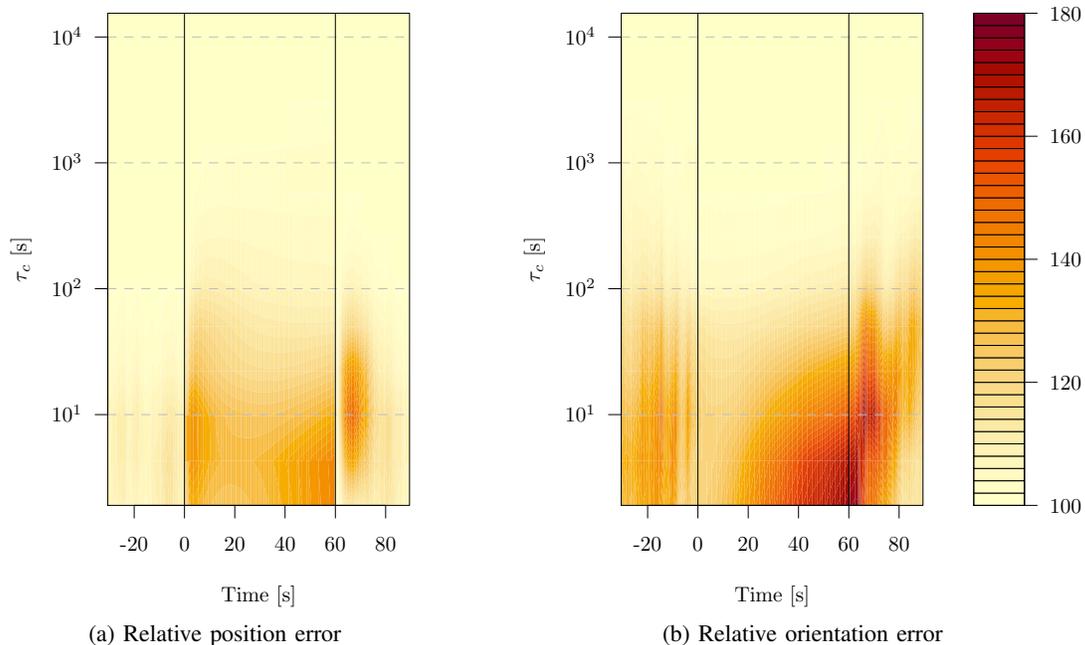(a) Relative position error

(b) Relative orientation error

Fig. 5: Relative position and orientation error of the AVLR model with respect to the GMWM one for the large model set, before, during and after a GNSS outage of 60 seconds starting at $t = 0$ s.

because of poorly modeled inertial readings. If the AR1 has large variance compared to the white noise, a peak in the position error is visible also soon after the GNSS position fixes have become again available. A possible explanation for this effect is due to a wrong model for inertial sensor which will build a wrong covariance matrix in the EKF, requiring more time to correct after recovery of GNSS measurements. Additional information of this effect are shown when discussing the estimated

uncertainty of position and orientation.

- An inaccurate stochastic model for the inertial sensors will have more impact on the orientation error than on the position error. In addition, this impact can be observed even when the GNSS position fix is available. This is expected because the orientation estimates are known to be more dependent on the quality of inertial sensor measurements and models.

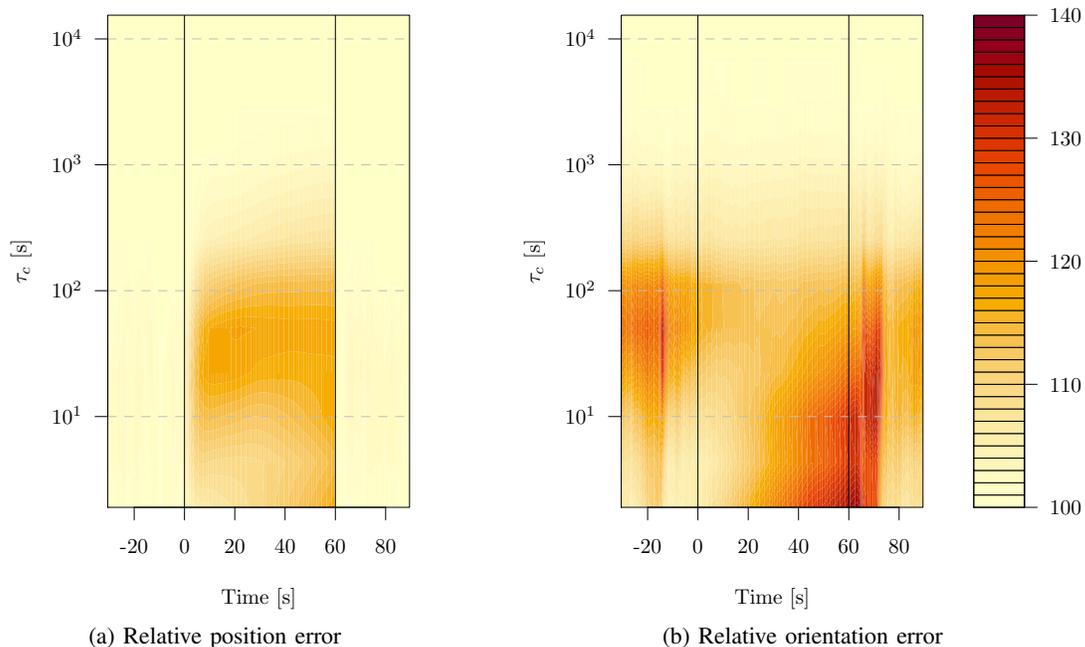Next, we analyze the quality of the confidence intervals

(a) Relative position error

(b) Relative orientation error

Fig. 6: Relative position and orientation error of the AVLR model with respect to the GMWM one for the `small` model set, before, during and after a GNSS outage of 60 seconds starting at $t = 0$ s.
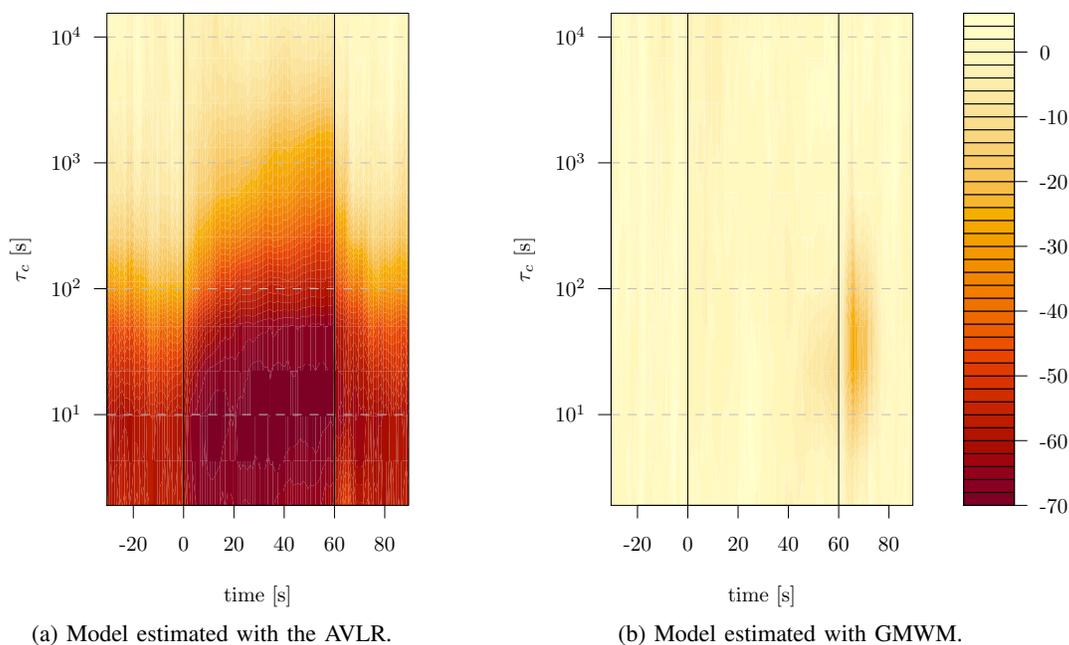

(a) Model estimated with the AVLR.

(b) Model estimated with GMWM.

Fig. 7: Coverage error as a function of the time relative to the GNSS gap (between 0 and 60 s) and the correlation time τ_c for the `large` model set.

for the position and orientation as estimated by the EKF. Indeed, during navigation the filter maintains a covariance matrix over the states, from which confidence intervals can be derived for those. If the stochastic model for the inertial measurement assumed in the EKF does not correspond to the one behind true noise generation, the covariance matrix will not be a good estimate of the probability distribution of the states. To quantify this, we evaluate the coverage as defined in Sec. IV-F with $1-\alpha = 70\%$: if the states uncertainty estimated

by the EKF is correct, we expect that $66\% < \bar{c}_t < 74\%$, the interval accounting for the Monte-Carlo simulation error. Indeed, assuming the filter to provide exact coverage, the random variable $^{(i)}c_t$ follows a Bernoulli distribution with parameter $1 - \alpha$. Therefore, using the central limit theorem an approximate 95% confidence interval for the averaged $^{(i)}c_t$ over N Monte-Carlo simulation is given by

$$(1-\alpha) \pm 1.96\sqrt{\frac{\alpha(1-\alpha)}{N}} \approx [66\%, \ 74\%] \, ,$$

(a) Model estimated with the AVLR.
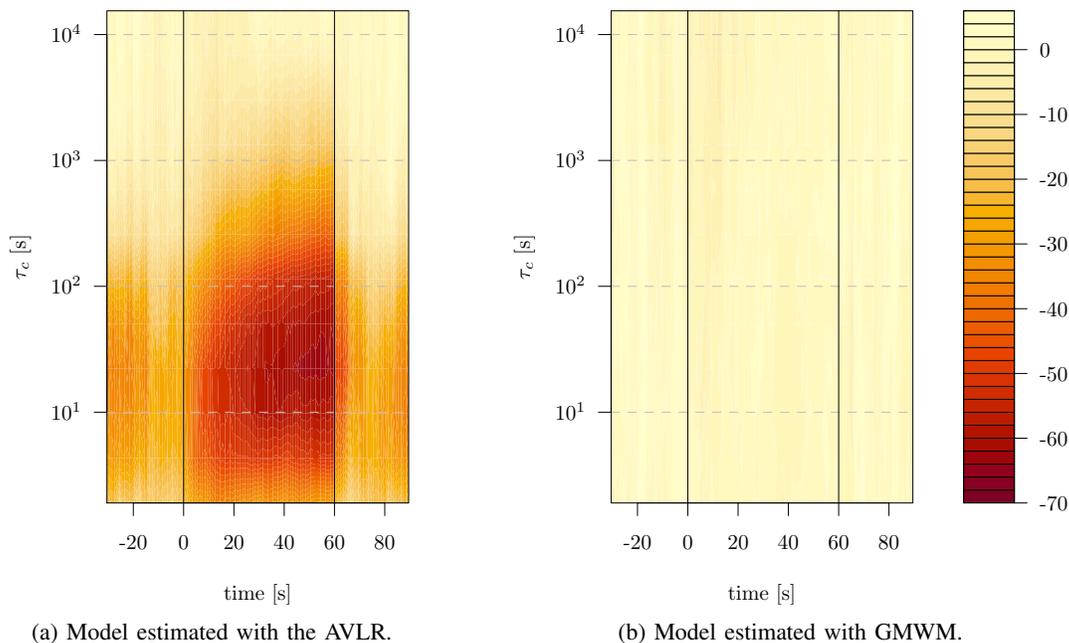
(b) Model estimated with GMWM.

Fig. 8: Coverage error as a function of the time relative to the GNSS gap (between 0 and 60 s) and the correlation time τ_c for the `small` model set.

for $\alpha = 30\%$. The results are depicted in Fig. 7 for the `large` model set and in Fig. 8 for the `small` model set. The results are expressed in terms of coverage error which we define as the difference between the assumed probability (i.e., 70%) and empirical probability that the true states is within the computed 70% confidence interval, $\bar{c}_t$. Consequently, a negative value indicates an over-confident estimate of the position and orientation uncertainty, and vice-versa.

We can observe that the models obtained with the GMWM have the correct coverage (around 70%, or around 0% coverage error) at all times, while the models obtained with the AVLR lead to a large underestimation of the position and orientation uncertainty when $\phi \to 1$ or, equivalently, $\tau_C \to \infty$. This effect is more severe for the `large` model set, and less severe for the `small` one, as happened for the position and orientation relative error. This result suggests that, in practice, if the user is particularly interested in having a reliable estimate of the navigation states uncertainty, the stochastic models obtained with GMWM can provide substantial advantages over the commonly used AVLR, at least for specific structure of the underlying noise generation mechanism.

2) Model 2: 1 WN + 2 AR1s: We consider a more complex case in the following. This is similar to the one discussed before, with the exception that this time the true gyroscope noise model is the sum of two AR1s and a white noise ($M = 3$). The parameters of such auto-regressive processes have been chosen so that a flat region in the high scales of the theoretical WV appears. In this case, 12 different instances have been considered and their theoretical WVs have been plot in Fig. 9a. Such shapes are typical of low-to-medium grade inertial sensors, e.g., MEMS, exhibiting non negligible bias instability. In Fig. 9b we have plotted one of the chosen instances, in green in both figures, and compared with the

empirical WV of a static acquisition of a real sensor, the MTI-g MEMS IMU. It can be observed that the shape of the WV is very similar, suggesting that our experimental setup considers a realistic scenario. In Fig. 12a it is also possible to see that the accelerometer noise model of the MTI-g would present a similar pattern in the WV. The accelerometer noise model will be discussed more in detail in Sec. VI-B.

Similarly to the results obtained in the previous section, it can be observed that a good estimate of the original model using the AVLR can only be obtained for some of the specific parameter values corresponding to correlation times of the two auto-regressive processes that are both either small or high. Instead, the GMWM allows to estimate arbitrary mixtures of auto-regressive processes, regardless of their correlation time and thus recover the underlying noise model with sufficient accuracy.

We apply the same experimental procedure as the one discussed in Section VI-A1. The relative position and orientation error achieved employing the models estimated with the AVLR versus the ones obtained with GMWM are shown in Fig. 10. The coverage error, as defined in the previous section, is also presented in Fig. 11, where it is possible to see that the models obtained with the AVLR lead to a large underestimation of the position and orientation uncertainty in the EKF.

B. Scenario 2: Real Sensor Noise

In this scenario we consider a real IMU sensor, the Xsens MTi-g, a MEMs IMU, for which real data is available from static acquisitions in a controlled environment. As anticipated in the previous section, and shown in Fig. 9b, such sensor exhibits non-negligible bias instability which appears as a flat region in the wavelet variance plot in correspondence of high scales τ. The AVLR does not allow to estimate a suitable
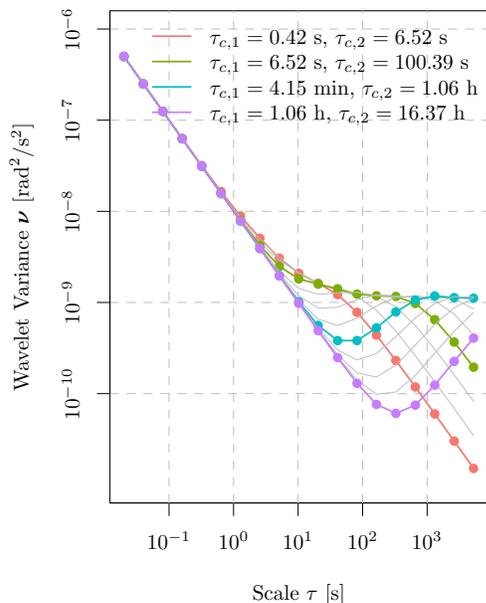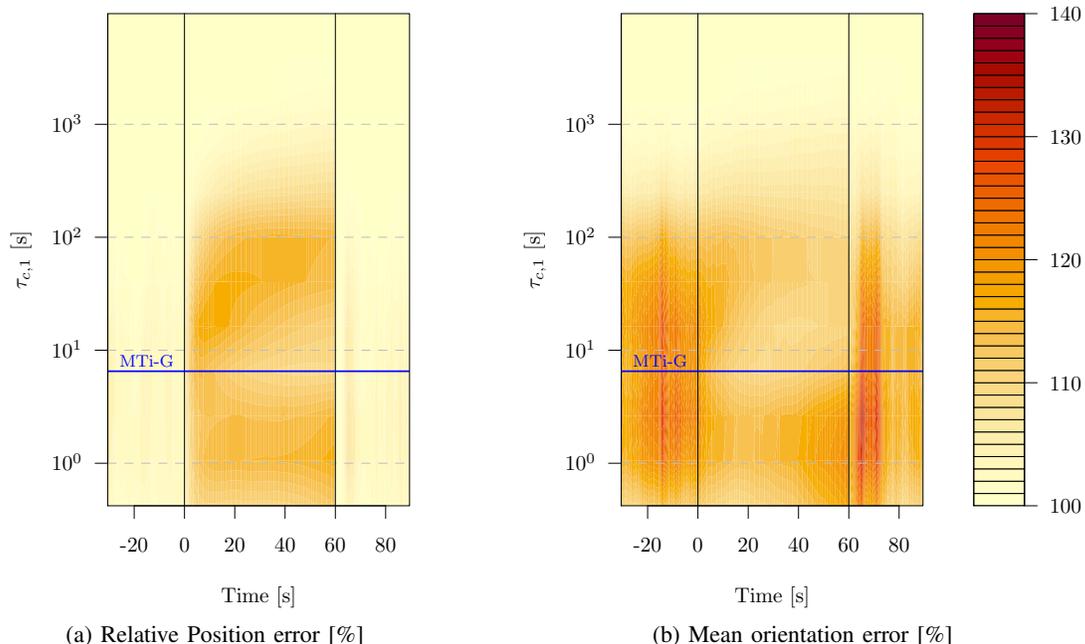
(a) Theoretical WVs of the considered noise models.
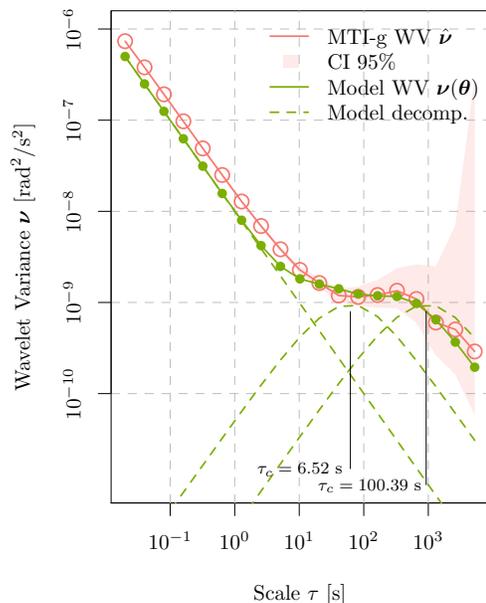
(b) One of the considered model versus the empirical WV of the MTI-g MEMS IMU.

Fig. 9: Comparison of the WV corresponding to the considered parameter values used for the simulation results presented in Sec. VI-A2 (on the left) and compared to the WV of the MTI-g MEMS IMU (on the right).

(a) Relative Position error [%]

(b) Mean orientation error [%]

Fig. 10: Relative position and orientation error of the AVLR model with respect to the GMWM one before, during and after a GNSS outage of 60 second starting at $t = 0$ s and for different time constants $\tau_{c,1}$ of the first auto-regressive process.

state-space model for this data. Instead, a model for this can be obtained approximating the flat region of the empirical WV as the sum of multiple auto-regressive processes that can be effectively estimated by means of the GMWM. The empirical WV of the noise data collected during the static acquisition is presented in Fig. 12, along with the theoretical WV of the model estimated with GMWM and its decomposition. The empirical WV of the real noise data and the theoretical one of

the estimated with the GMWM model match to a remarkable extent once three AR1s are considered for the accelerometer and two for the gyroscope.

We evaluate the performances of the models estimated by means of the AVLR against GMWM in an EKF. This time, instead of simulating noise samples from a stochastic model known a priori, which is not available in the case of a real sensor, we corrupt inertial readings using real noise samples

(a) Model estimated with the Allan variance.
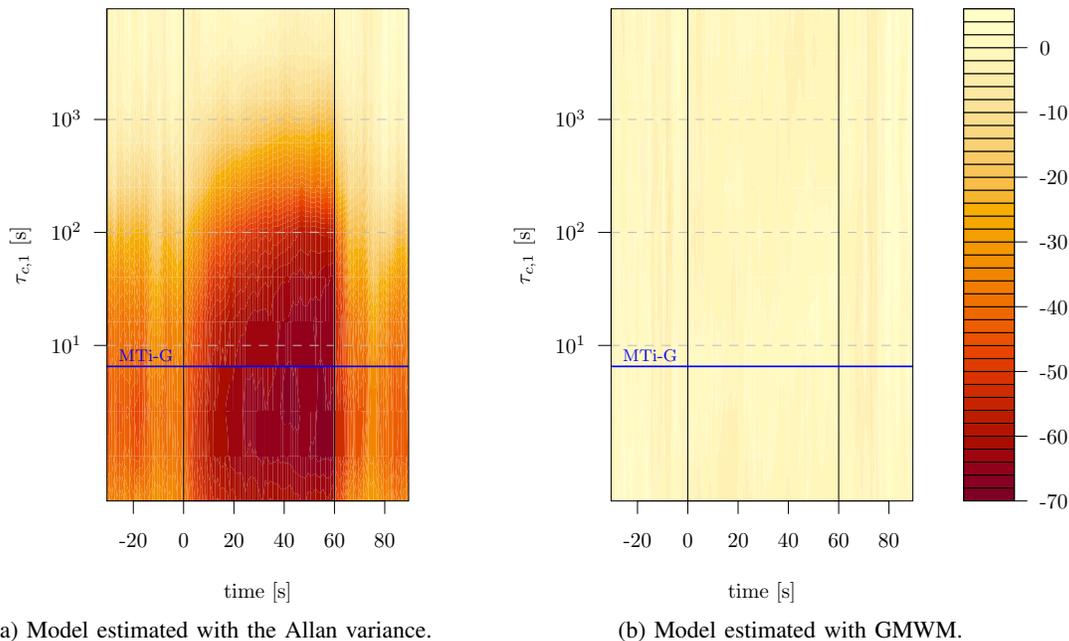
(b) Model estimated with GMWM.

Fig. 11: Coverage error as a function of the time relative to the GNSS gap (between 0 and 60 s) and the correlation time $\tau_{c,1}$ of the first auto-regressive process.
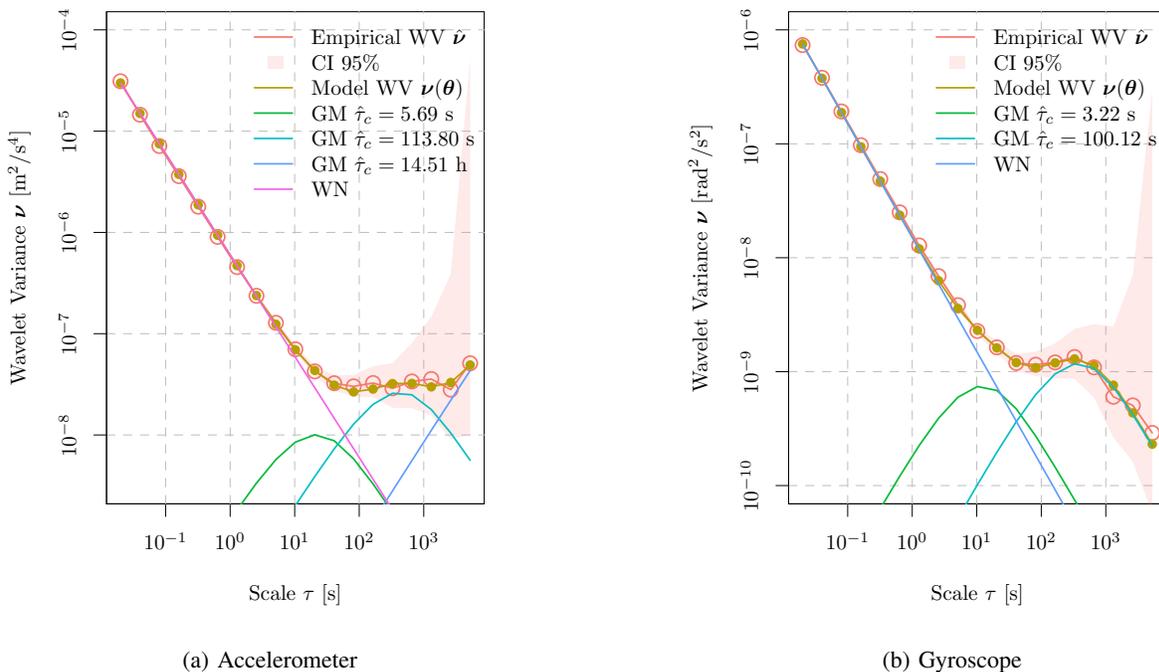

(a) Accelerometer

(b) Gyroscope

Fig. 12: Empirical WV of the 2 h of static data collected from an Xsens MTI-g IMU (in red) and the model estimated with the GMWM (in brown), considering the gyroscope and accelerometer X axis on the left and on the right, respectively. The model decomposition in terms of WN and multiple auss-Markov processes is also shown.

obtained during a static acquisition: several hours of data are available and each time different contiguous sets are chosen in a Monte-Carlo fashion, as discussed in Sec. IV. This approach permits to evaluate the performances of a given stochastic

model when realistic noise samples are provided, more closely to a real-world application scenario.

The obtained models are compared in terms of the relative position and orientation error when configuring the EFK with
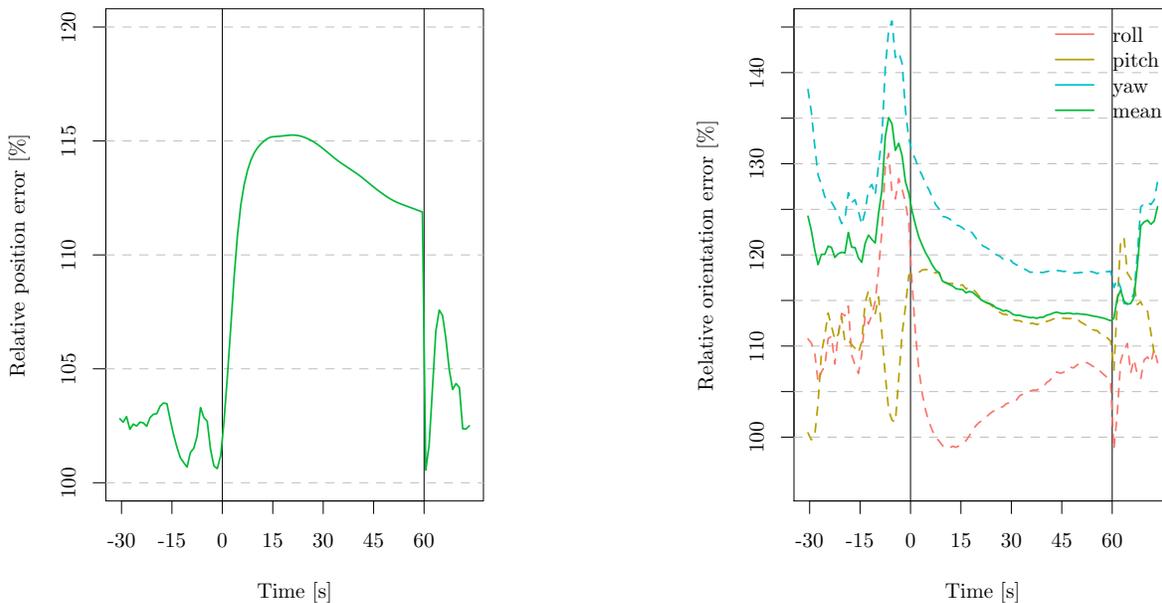
(a) Relative position error of the AVLR with respect to the GMWM stochastic calibration.

(b) Relative orientation of the AVLR with respect to the GMWM stochastic calibration.

Fig. 13: Position and orientation error.

the associated models. In Fig. 13 we show the position and orientation error of the models estimated with the AVLR relative to the ones obtained with the GMWM. As anticipated, the models obtained by means of the AVLR fail to model accurately the stochastic nature of the error signal and this translates in degraded navigation performances, up to 15% in position and 45% in orientation. The decomposition of the orientation error along the three axes is also show to point out that, as expected, the maximum error is found on the Yaw axis and that the AVLR model performs up to 50% worse than GMWM on that axis. These results show the potential benefit of performing stochastic calibration with GMWM on a real device and are in line with what obtained in Sec. VI-A2: this can be seen by comparing Fig. 10, along the blue line.

VII. CONCLUSIONS

In this work we propose a simulation-based framework allowing to assess the performances of INS/GNSS navigation systems in realistic settings. Indeed, our approach allows to consider flexible and user-configurable noise generation mechanisms and permits to quantify INS/GNSS navigation systems performances in different conditions (e.g, GNSS outages, inaccurate stochastic model for any sensor and so on). Using this framework, we study the impact on navigation performances of different statistical procedures (AVLR and GMWM) used to obtain stochastic inertial stochastic models. By comparing a various simulation settings, our results suggest that the commonly AVLR method can lead to to considerably worse performances compared the GMWM in terms of position and orientation error as well as in terms of accuracy of the estimated states, especially for inertial sensors characterized by important bias-instability. These results can have important practical implications, suggesting

that the GMWM can provide significant improvements, in particular when MEMS sensors are considered, where bias-instability is frequent. Moreover, our framework provides a systematic method allowing researchers and practitioners to compare the performance of existing, and potentially new, stochastic calibration methods. In this work we considered stochastic calibration as a proof of concept, but other problems related to the quantification of navigation system performances can be easily investigated with our framework, such as, for example, the robustness of an INS with respect to outliers or other modeling imperfections. Our method can bridge the gap between real world experiments, which are always needed but often impractical, costly and difficult to replicate a large number of time (e.g., to evaluate the correctness of estimated uncertainty) and pure statistical parameter estimation as done, for example, in the stochastic calibration literature.

REFERENCES

[1] D. Titterton, J. L. Weston, and J. Weston, *Strapdown inertial navigation technology*, vol. 17. IET, 2004.

[2] G. Huang, "Visual-inertial navigation: A concise review," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 9572–9582, IEEE, 2019.

[3] Q. Fan, B. Sun, Y. Sun, and X. Zhuang, "Performance enhancement of MEMS-based INS/UWB integration for indoor navigation applications," *IEEE Sensors Journal*, vol. 17, no. 10, pp. 3116–3130, 2017.

[4] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[5] P. Clausen, *Calibration Aspects of INS Navigation*. PhD thesis, EPFL, 2019.

[6] D. W. Allan, "Statistics of atomic frequency standards," *Proceedings of the IEEE*, vol. 54, no. 2, pp. 221–230, 1966.

[7] N. El-Sheimy, H. Hou, and X. Niu, "Analysis and modeling of inertial sensors using Allan variance," *IEEE Transactions on instrumentation and measurement*, vol. 57, no. 1, pp. 140–149, 2007.

[8] S. Guerrier, J. Skaloud, Y. Stebler, and M.-P. Victoria-Feser, "Wavelet-variance-based estimation for composite stochastic processes," *Journal of the American Statistical Association*, vol. 108, no. 503, pp. 1021–1030, 2013.

[9] S. Guerrier, J. Jurado, M. Khaghani, G. Bakalli, M. Karemera, R. Molinari, S. Orso, J. Raquet, C. Schubert, J. Skaloud, *et al.*, "Wavelet-Based Moment-Matching Techniques for Inertial Sensor Calibration," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7542 – 7551, 2020.

[10] R. Allen and D. Chang, "Performance testing of the systron donner quartz gyro," *JPL Engineering Memorandum, EM*, pp. 343–1297, 1993.

[11] M. J. Ma, Z. H. Jin, Y. D. Liu, C. H. Lim, T. S. Lim, V. C. Koo, K. Zhang, J. Q. Bai, B. L. Gu, and A. Abbas, "IEEE Standard Specification Format Guide and Test Procedure for Single-axis Interferometric Optic Gyros," *IEEE Std 952-2020 (Revision of IEEE Std 952-1997)*, 1998.

[12] Y. Yuksel, N. El-Sheimy, and A. Noureldin, "Error modeling and characterization of environmental effects for low cost inertial MEMS units," in *Proceedings of IEEE/ION PLANS 2010*, pp. 598–612, 2010.

[13] J. Nikolic, P. Furgale, A. Melzer, and R. Siegwart, "Maximum likelihood identification of inertial sensor noise model parameters," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 163–176, 2015.

[14] M. El-Diasty and S. Pagiatakis, "Calibration and Stochastic Modelling of Inertial Navigation Sensor Errors," *Positioning (POS) Journal Information*, p. 80, 2010.

[15] S. Guerrier, R. Molinari, and Y. Stebler, "Theoretical limitations of Allan variance-based regression for time series model estimation," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 597–601, 2016.

[16] Y. Stebler, S. Guerrier, J. Skaloud, and M.-P. Victoria-Feser, "Generalized method of wavelet moments for inertial navigation filter design," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 3, pp. 2269–2283, 2014.

[17] J. Jurado, C. M. Schubert Kabban, and J. Raquet, "A regression-based methodology to improve estimation of inertial sensor errors using Allan variance data," *Navigation*, vol. 66, no. 1, pp. 251–263, 2019.

[18] N. F. Zhang, "Allan variance of time series models for measurement data," *Metrologia*, vol. 45, no. 5, p. 549, 2008.

[19] M. B. Ignagni, "Duality of optimal strapdown sculling and coning compensation algorithms," *NAVIGATION, Journal of the Institute of Navigation*, vol. 45, no. 2, pp. 85–96, 1998.

[20] A. Bruton, C. Glennie, and K. Schwarz, "Differentiation for high-precision GPS velocity and acceleration deter-mination," *GPS solutions*, vol. 2, no. 4, pp. 7–21, 1999.

[21] M. El-Diasty, A. El-Rabbany, and S. Pagiatakis, "Temperature variation effects on stochastic characteristics for low-cost MEMS-based inertial sensor error," *Measurement Science and Technology*, vol. 18, no. 11, p. 3321, 2007.

[22] A. Radi, S. Nassar, and N. El-Sheimy, "Stochastic error modeling of smartphone inertial sensors for navigation in varying dynamic conditions," *Gyroscopy and Navigation*, vol. 9, no. 1, pp. 76–95, 2018.

[23] D. A. Cucci, M. Rehak, and J. Skaloud, "Bundle adjustment with raw inertial observations in UAV applications," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 1–12, 2017.

[24] S. Nassar, K.-P. Schwarz, and N. El-Sheimy, "INS and INS/GPS accuracy improvement using autoregressive (AR) modeling of INS sensor errors," in *Proceedings of the 2004 national technical meeting of the institute of navigation*, pp. 936–944, 2004.

[25] Y. Yuksel and H. B. Kaygisiz, "Notes on stochastic errors of low cost MEMS inertial units," *Online. Available: http://www.instk.org/web/static/bibliography/Introduction_to_Sensor_Errors.pdf.*, 2011.

[26] KVH Industries, Inc., *High-Performance Inertial Sensor for a Wide Range of Unmanned Applications*, 2019. Rev. 7.19.

[27] "Xsens 3D motion tracking." www.xsens.com. Accessed: 2022-06-28.