

ACCEPTED MANUSCRIPT • OPEN ACCESS

# Detection and identification of faults in clock ensembles with the generalized likelihood ratio test

To cite this article before publication: Christian Trainotti *et al* 2022 *Metrologia* in press <https://doi.org/10.1088/1681-7575/ac7686>

## Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2022 BIPM & IOP Publishing Ltd.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# Detection and identification of faults in clock ensembles with the generalized likelihood ratio test

Christian Trainotti<sup>1</sup>, Gabriele Giorgi<sup>1</sup> and Christoph Günther<sup>1,2</sup>

<sup>1</sup>Institute of Communications and Navigation, German Aerospace Center (DLR), Münchener Straße 20, 82234 Weßling, Germany

<sup>2</sup>Institute for Communications and Navigation, Technical University of Munich, Theresienstraße 90, 80333 München, Germany

E-mail: christian.trainotti@dlr.de

**Abstract.** In this paper we propose an approach for performing fault detection and identification in clock ensembles based on the generalized likelihood ratio test. We show that by applying a set of purposefully-designed statistical tests, one can successfully detect faults occurring in a clock of the ensemble, and identify which measurement in the ensemble is most likely to have triggered the detection. We first develop the theoretical framework for the characterization of the detectors and their performance, and validate the derivations via Monte Carlo simulations. Then, we apply the statistical tests to an ensemble of cesium clocks, aiming at detecting and identifying three types of non-nominal behaviors. The faulty conditions are obtained by injecting a pattern of phase steps, a phase and frequency drift, and an oscillatory phase component.

*Keywords:* GLRT, clock ensemble, integrity, fault detection, timing

## 1. Introduction

Clock ensembles play a fundamental role in providing a stable and robust time scale for high-integrity applications. Compared to single clock approaches, clock ensembles are characterized by improved stability and robustness. However, clock faults can still impact the stability and availability of the generated time scale. Thus a clock ensemble must be equipped with a dedicated fault monitoring system performing three steps: fault detection, fault identification, and fault adaptation. Detecting a fault means raising an alarm when a fault occurrence is observed, while the identification step aims at locating where in the system the fault is occurring. Finally, fault adaptation consists in taking corrective actions to reduce the impact of the fault on the system, for instance by repairing or temporarily removing the faulty unit. Since in this paper we only tackle the first two steps, we will talk of fault detection and identification (FDI).

A FDI algorithm is implemented in a detector. The detector observes a particular quantity (the observable) and scans for deviations from the expected nominal behavior, which are caused by the occurrence of a fault. Different faults act differently on the system, thus we must choose the right observable (or a set thereof) to properly identify different types of expected errors. In this work we use three observables, namely Kalman filter residuals, phase measurements, and dynamic Allan variance (DAVAR) of differential clock measurements. For each of these we devise a dedicated detector based on the generalized likelihood ratio test (GLRT), which accounts for the statistical distribution of the observable in both nominal and faulty conditions. We derive a model-based test and a self-consistency test. In the model-based test the detector compares the current observation to the value predicted by a clock model: any deviation between the two can originate from a fault, but also from a mismodelling of the expected behavior. We apply the model-based test on the entire set of clock measurements (*overall model test*) to perform fault detection, and then singularly on each measurement to perform fault identification (*w-test*) [1, 2]. In the case of homogeneous ensembles, which only contain clocks of the same type, we construct the *self-consistency test*, that does not require to define clock noise parameters. The availability of multiple clocks of the same type allows the estimation of the nominal clock behavior directly from the observations. Instead of detecting deviations between observed and modelled behaviors, we now detect deviations between one measurement and the average observed behavior of all measurements.

After deriving the different types of statistical tests, we describe the expected performance of each in terms of probability of false alarm and probability

of missed detection. Monte Carlo simulations are then used to validate the theoretical derivation. Finally, we perform an experimental assessment of all the tests devised by injecting different types of fault patterns in a laboratory clock ensemble, and we verify whether the tests can successfully detect and identify the faulty signal.

### 1.1. State of the art

Table 1 shows a selection of literature on fault detection in clocks and clock ensembles. While most of the studies focus on detecting faults on single clock signals, some works specifically target clock ensembles. In [3] the authors propose the usage of a Kalman filter for fault detection in a clock ensemble: by tracking the evolution of the Kalman filter residuals it is possible to detect discrepancies between model and observation. This method is further investigated in [4] and applied in [5, 6]. The use of the GLRT for detecting clock failures is described in [7], and the method is further developed in [8, 9, 10]. In [11] the method is applied to the detection of anomalies on a GPS satellite clock. The DAVAR method, detailed in [12, 13, 14], makes use of the overlapping Allan variance (OAVAR) computed on sliding windows of given length, and compares the temporal changes in the data to predefined DAVAR values. The detector is triggered if the deviations exceed a statistically significant threshold. The application of this method to clocks onboard satellites of global navigation satellite systems (GNSSs) is shown in [15, 16]. In [17] the author describes three algorithms for detection of jumps in frequency measurements: the BLKAVG algorithm computes the average of frequency points on non-overlapping moving windows, detecting a jump when the difference in averages among two windows is larger than a given threshold; the SEQAVG algorithm employs an averaging window only once a potential jump is detected, to confirm or reject the detection; finally, the CUSUM algorithm computes the cumulative sum of the differences between each frequency data point and the average value on the entire data set. The slope of the cumulative sum reflects the average behavior of the measurement and can be used to detect frequency jumps. Other techniques for fault detection in clock signals include trend analysis and filtering of the frequency signal [18, 19, 20], the optimal stopping method [21], and interferometric analysis [22]. Fault detection in clock ensembles is further discussed in [23, 24, 25] using Kalman filters. Instead, the methods in [26, 27, 28] use least-squares fitting, standard deviation computed on sliding windows, and infinite impulse response filters.

In this work we integrate and extend some of these approaches for FDI. Expanding from our previous works [30, 31] we provide a theoretical framework

Table 1: Selected literature on integrity for clocks and clock ensembles.

Method	Single clock	Clock ensemble
Kalman filter residuals	[4, 5, 6, 29]	[3, 23, 24, 25]
Generalized likelihood ratio test on clock measurements	[7, 8, 9, 10, 11]	<i>this work</i>
Dynamic Allan variance	[12, 13, 14, 15, 16]	<i>this work</i>
Trend analysis—Smoothing—Recursive filters	[18, 19, 20]	[27, 28]
Optimal stopping method	[21]	
Average, least-squares, standard deviation on sliding windows	[17]	[26]
Interferometry	[22]	

for the design of fault detectors in clock ensembles. The techniques shown here can be directly applied to real-time dependable timing system, greatly improving the robustness of the generated time scale. The paper shows how to take advantage of the GLRT for the observation of different quantities, so that the integrity layer of a system can be easily expanded with further detectors using the same approach. Furthermore, the GLRT allows us to predict the performance of the detectors in terms of probability of false alarm ( $P_{fa}$ ), probability of missed detection ( $P_{md}$ ), and minimum detectable bias (MDB). This enables tuning the detectors to fulfill the desired integrity requirements, namely which magnitude of fault is detectable and with which probability. By taking as example a precise timing facility running a clock ensemble to provide timing services, these methods would benefit both the provider and its users. Firstly, the provider can guarantee a given integrity level, since the detection performances can be determined beforehand; then, in case of a faulty clock, warnings can be promptly provided to the users, while the fault is located in the system and corrective actions are put in place. An additional advantage of these methods is represented by the self-consistency test, which detects whether one of the clocks diverges from the ensemble's average behavior. Since no a-priori clock noise model is assumed, this test is not triggered by any influence affecting all clocks in the same way. The evaluation of clocks on board GNSS satellites is an example of such a benefit: the estimation of the clocks' states is affected by the residuals of the orbit determination process, which determine an increase of the observed OAVAR for sampling intervals around half of the orbital period [32]. A model-based test would in this case trigger a detection since the observed behavior differs from the clock model, even if the misbehavior is not due to the clock itself. Instead, the self-consistency test would not trigger, since the additive bias is present on all clocks.

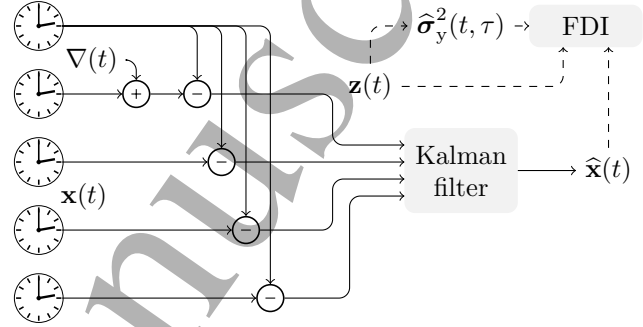


Figure 1: Functional representation of the clock ensemble under study, with measurement topology and processing units: the clocks, whose state vector is  $\mathbf{x}(t)$ , are measured by a device generating clock measurements  $\mathbf{z}(t)$ . A Kalman filter computes the estimates  $\hat{\mathbf{x}}(t)$  of the clock states. The fault detection and identification algorithm (FDI) uses the clock measurements  $\mathbf{z}(t)$ , the states estimate  $\hat{\mathbf{x}}(t)$ , and the dynamic Allan variance (DAVAR)  $\hat{\sigma}_y^2(t, \tau)$  of the clock measurements to observe and evaluate the state of the system. For assessing the FDI algorithm, we inject a fault  $\nabla(t)$  to the signal of the second clock. The computer executing the algorithm is not depicted.

## 2. Clock ensemble

The system under study consists of a set of clocks, a measurement device, a Kalman filter and a number of detectors. While we keep the theoretical development as general as possible concerning number and type of clocks, the experimental verification is performed with an ensemble of 5 cesium frequency references. Figure 1 shows a schematic of the ensemble. The measurement device generates differential clock measurements, which are phase and/or frequency differences between two clocks. In our setup the first clock acts as reference against which we measure the remaining units, but different measurement topologies can be used. The clock measurements are fed to the Kalman filter, which first predicts the states of the clocks according to a specified clock model, and

then updates the predicted states with the available observations. The Kalman filter is executed on a laboratory computer connected to the measurement device. This computer also runs the FDI algorithm. Furthermore, a dedicated device can inject a desired fault pattern into the signal of a clock, so that we can run experiments for assessing the performance of the detectors. The injected fault is indicated by  $\nabla(t)$  in figure 1.

### 2.1. Clock model and ensemble model

We use a 2-state clock model with constant frequency drift and Markov processes, as described in [33, 34, 35]. The state vector of each clock contains one state representing the phase deviation and one state for the frequency deviation. When required for more accurate modelling, the state vector of the single clock is expanded with one or more additional states, related to additive Markov processes [34]. We combine the states of each clock to obtain the state vector  $\mathbf{x}$  of the entire ensemble. In time discrete notation we describe the dynamics of the system and its observation with

$$\mathbf{x}_k = \Phi(\tau_0) \mathbf{x}_{k-1} + \mathbf{d}(\tau_0) + \mathbf{n}_{k-1}, \quad (1)$$

$$\mathbf{z}_k = \mathbf{H} \mathbf{x}_k + \mathbf{v}_k. \quad (2)$$

where the subscript  $\bullet_k$  indicates the quantities at time  $t_k$ ,  $\tau_0 = t_k - t_{k-1}$  is the constant time discretization,  $\Phi$  the state propagation matrix,  $\mathbf{d}$  the drift vector,  $\mathbf{z}$  the vector of clock measurements, and  $\mathbf{H}$  the measurement matrix. The process noise  $\mathbf{n}_k$  and the measurement noise  $\mathbf{v}_k$  are assumed Gaussian:

$$\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(\tau_0)), \quad (3)$$

$$\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad (4)$$

where  $\mathbf{Q}$  is the process noise covariance matrix and  $\mathbf{R}$  is the covariance matrix of the measurement noise. The matrices  $\Phi$  and  $\mathbf{Q}$  are block diagonal, where each block is the corresponding matrix for the single clock  $i$ ,  ${}_i\Phi(\tau_0)$  and  ${}_i\mathbf{Q}(\tau_0)$ , computed with the model parameters of the respective clock  ${}_i\sigma_1$  and  ${}_i\sigma_2$ :

$${}_i\Phi(\tau_0) = \begin{bmatrix} 1 & \tau_0 \\ 0 & 1 \end{bmatrix}, \quad (5)$$

$${}_i\mathbf{Q}(\tau_0) = \begin{bmatrix} {}_i\sigma_1^2 \tau_0 & {}_i\sigma_2^2 \frac{\tau_0^3}{3} & {}_i\sigma_2^2 \frac{\tau_0^2}{2} \\ {}_i\sigma_2^2 \frac{\tau_0^2}{2} & {}_i\sigma_2^2 \tau_0 & 0 \end{bmatrix}. \quad (6)$$

If Markov processes need to be added to model specific behaviors, we expand the matrices accordingly [34]. Denoting the constant frequency drift of the  $i$ -th clock as  ${}_i d$ , we generate the drift vector by stacking the vectors  ${}_i\mathbf{d}(\tau_0)$  of the single clocks:

$${}_i\mathbf{d}(\tau_0) = {}_i d \begin{bmatrix} \frac{\tau_0^2}{2} \\ \tau_0 \end{bmatrix}. \quad (7)$$

The measurement matrix  $\mathbf{H}$  describes how the clocks are measured with respect to each other. It depends

on the measurement topology of the ensemble and on whether differential phase or frequency measurements are available. In our case we measure the phase of each clock with respect to the first clock. Thus we obtain  $M = N - 1$  measurements, where  $N$  is the number of clocks in the ensemble. Without Markov processes the measurement matrix is

$$\mathbf{H} = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & & & & & & & & \\ -1 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}. \quad (8)$$

Additionally, we define a second measurement matrix  $\bar{\mathbf{H}}$ , which maps the clock units to the measurements. It can be obtained from  $\mathbf{H}$  by eliminating the null columns. For instance, the matrix corresponding to the aforementioned topology with 4 clocks and 3 differential measurements is

$$\bar{\mathbf{H}} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}. \quad (9)$$

This matrix represents the measurement topology without dependence on the number of states used to model the system.

### 2.2. Kalman filter

We run a clock ensembling algorithm based on a Kalman filter, as described in [34, 36]. At each time step, the filter provides an estimate of the clock states by executing the following steps:

$$\hat{\mathbf{x}}_k^- = \Phi(\tau_0) \hat{\mathbf{x}}_{k-1} + \mathbf{d}(\tau_0), \quad (10)$$

$$\mathbf{P}_k^- = \Phi(\tau_0) \mathbf{P}_{k-1}' \Phi^T(\tau_0) + \mathbf{Q}(\tau_0), \quad (11)$$

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H} \mathbf{P}_k^- \mathbf{H}^T + \mathbf{R})^{-1}, \quad (12)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H} \hat{\mathbf{x}}_k^-), \quad (13)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{P}_k^-, \quad (14)$$

$$\mathbf{P}_{r,k} = \mathbf{S} \mathbf{P}_k \mathbf{S}^T, \quad (15)$$

$$\mathbf{P}_k' = \frac{\mathbf{P}_{r,k} + \mathbf{P}_{r,k}^T}{2}. \quad (16)$$

$\hat{\mathbf{x}}^-$  is the predicted estimate of the state vector  $\mathbf{x}$ , with associated error covariance matrix  $\mathbf{P}^-$ ,  $\mathbf{K}$  is the Kalman gain, used to generate the updated state estimate  $\hat{\mathbf{x}}$ , with associated error covariance  $\mathbf{P}$ . To avoid the unbounded growth of the components in  $\mathbf{P}$ , we employ the covariance  $x$ -reduction method (15) described in [36]. The reduced covariance matrix  $\mathbf{P}_r$  is obtained via the diagonal matrix  $\mathbf{S}$ , an identity matrix whose entries corresponding to the phase components in the state vector are set to zero. Finally, (16) ensures that the resulting reduced covariance matrix is diagonal. In (14),  $\mathbf{I}$  denotes an identity matrix, whose size is the number of states in  $\mathbf{x}$ .

### 3. Faults, observables and hypotheses

In this paper we consider two different classes of faults. The first kind consists of abrupt changes in the observed phase or frequency signals, such as jumps and steps, which present a very fast dynamics. The second type of faults shows a slower behavior and includes phase or frequency drifts as well as oscillatory components in the signals. A drift can arise due to the aging of the clocks, while the oscillation can generate from periodic effects, such as daily temperature fluctuations or orbital-related dynamics. While abrupt faults are promptly reflected into the Kalman filter residuals, drifts cause a step-to-step phase and frequency deviation which might be too small to be detected by testing the filter residuals. Thus, to detect this type of faults we need a quantity which “accumulates” the error, such as the phase measurements and their DAVAR. Clearly, the required accumulation of a sufficient number of samples introduces a delay between the fault onset and the time of first detection.

In the following sections we introduce the theoretical descriptions of the three observables: Kalman filter residuals, phase measurements and DAVAR. For each observable we define a *null hypothesis*  $\mathcal{H}_0$ , which describes its statistical distribution when the system operates under nominal conditions. An *alternative hypothesis*  $\mathcal{H}_A$  describes instead the observable’s distribution when a fault is occurring in the system. With these hypotheses we then derive a statistical test decision based on the GLRT.

#### 3.1. Kalman filter residuals

The Kalman filter residuals track the deviations between the observed clock measurements  $\mathbf{z}$  and the measurements predicted by the model. Recalling (13) and (14), we express the residuals and their covariance as:

$$\rho_k^{(\text{kf})} = \mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^- \quad (17)$$

$$\Omega_k^{(\text{kf})} = \mathbf{R} + \mathbf{H}\mathbf{P}_k^-\mathbf{H}^\top. \quad (18)$$

In nominal conditions we expect the residuals to remain small, while their magnitude increases if the observed measurements differ from the modelled behavior. We therefore define the two hypotheses:

$$\begin{cases} \mathcal{H}_0^{(\text{kf})} : \rho_k^{(\text{kf})} \sim \mathcal{N}(\mathbf{0}, \Omega_k^{(\text{kf})}) \\ \mathcal{H}_A^{(\text{kf})} : \rho_k^{(\text{kf})} \sim \mathcal{N}(\mathbf{C}\nabla_k, \Omega_k^{(\text{kf})}) \end{cases} \quad (19)$$

Here we introduce the vector of faults  $\nabla$  affecting the system, and the matrix  $\mathbf{C}$  describing how the faults act on the observables. The size of vector  $\nabla$  depends on the shape of matrix  $\mathbf{C}$ , whose free design allows the selection of a given fault mode to be tested. For example, setting  $\mathbf{C} = \mathbf{I}_M$ , an identity matrix of size

$M$ , enables us to perform fault detection by assuming a potential fault on all elements of the observable vector, with  $\nabla$  becoming a vector of  $M$  entries. When assigning  $\mathbf{C} = \mathbf{c}_i$  instead, with  $\mathbf{c}_i$  a  $M$ -elements vector with 1 as the  $i$ -th entry and zero elsewhere, we can scan all single measurements  $i = 1 \dots M$  to locate the faulty clock. In this case  $\nabla$  is a scalar.

#### 3.2. Phase measurements

We describe the time-varying distribution of the phase measurements of the clocks with the model equations (1):

$$\mathbf{x}_k \sim \mathcal{N}[\Phi(t_k - t_0)\mathbf{x}(t_0) + \mathbf{d}(t_k - t_0), \mathbf{Q}(t_k - t_0)] \quad (20)$$

where  $\mathbf{x}(t_0)$  contains the initial phase and frequency states of the ensemble. Then, from (2) we have

$$\mathbf{z}_k \sim \mathcal{N}(\zeta_k, \Omega_k^{(z)}), \quad (21)$$

where

$$\zeta_k = \mathbf{H}[\Phi(t_k - t_0)\mathbf{x}(t_0) + \mathbf{d}(t_k - t_0)], \quad (22)$$

$$\Omega_k^{(z)} = \mathbf{H}\mathbf{Q}(t_k - t_0)\mathbf{H}^\top + \mathbf{R}. \quad (23)$$

We define the phase residuals as the difference between observed and expected phase measurements,

$$\rho_k^{(z)} = \mathbf{z}_k - \zeta_k, \quad (24)$$

so that the corresponding hypotheses can be defined:

$$\begin{cases} \mathcal{H}_0^{(z)} : \rho_k^{(z)} \sim \mathcal{N}(\mathbf{0}, \Omega_k^{(z)}) \\ \mathcal{H}_A^{(z)} : \rho_k^{(z)} \sim \mathcal{N}(\mathbf{C}\nabla_k, \Omega_k^{(z)}) \end{cases} \quad (25)$$

These hypotheses have the same form of those derived for the Kalman filter residuals (19). While the latter track the evolution of the phase measurements with respect to the previous time step, the phase measurements residuals describe the overall phase evolution from the initial time instant.

#### 3.3. Dynamic Allan variance

The third observable is the DAVAR of differential phase measurements computed on a sliding window of given length. In [12] the DAVAR value is centered in the middle of the sliding window, whereas in this paper the evaluation time corresponds to the right limit of the window. In this way the DAVAR can be used in a real-time scenario employing the latest available data samples. Thus, the DAVAR at time  $t_k$  uses the measurements in the window  $[t_k - W, t_k]$ , where  $W$  is the length of the sliding window. We use the notation  $_{ij}\hat{\sigma}_{k,\tau,W}^2$  for the DAVAR of the measurement between clocks  $i$  and  $j$ , at time instant  $k$ , for sampling interval  $\tau$  and window length  $W$ . The expected value  $_{ij}\bar{\sigma}_\tau^2$  and the variance  $\text{Var}(_{ij}\hat{\sigma}_{\tau,W}^2)$  of the DAVAR are assumed

constant in time and can be expressed using the clock noise parameters:

$$ij\bar{\sigma}_\tau^2 = 3\frac{ijR}{\tau^2} + \frac{ij\sigma_1^2}{\tau} + \frac{ij\sigma_2^2\tau}{3} + ij\frac{\tau^2}{2}, \quad (26)$$

$$\text{Var}(ij\hat{\sigma}_{\tau,w}^2) = f(\tau, W, ijR, ij\sigma_1^2, ij\sigma_2^2, ij\frac{\tau^2}{2}). \quad (27)$$

Here, the noise parameters refer to the combined contribution of the two clocks included in each measurement. The function in (27) is explicitly derived in [37] for different noise components. Since an expression of the variance also considering Markov processes is not available yet, we use a simplified clock model without additional processes, whose parameters are reported in section 4.4, expression (73).

The DAVAR at time  $t_k$  distributes according to a  $\chi^2$ -distribution [38], namely (in the following we drop the  $ij$  subscript):

$$p_{\tau,w} \frac{\hat{\sigma}_{k,\tau,w}^2}{\bar{\sigma}_\tau^2} \sim \chi^2(p_{\tau,w}), \quad (28)$$

where

$$p_{\tau,w} = 2 \frac{\bar{\sigma}_\tau^4}{\text{Var}(\hat{\sigma}_{\tau,w}^2)} \quad (29)$$

are the degrees of freedom. By combining (28) and (29) we define the normalized DAVAR

$$\xi_{k,\tau,w} = \frac{2\hat{\sigma}_{k,\tau,w}^2 \bar{\sigma}_\tau^2}{\text{Var}(\hat{\sigma}_{\tau,w}^2)}. \quad (30)$$

We express the distribution of  $\xi$  using the gamma distribution

$$\xi_{k,\tau,w} \sim \chi^2(p_{\tau,w}) \equiv \Gamma\left(\frac{p_{\tau,w}}{2}, 2\right), \quad (31)$$

since it allows us to also express the faulty case. If a fault occurs in the system, we expect a change of the measured DAVAR, therefore we modify the theoretical value by the quantity  $\nabla_{k,\tau,w}$ :

$$\begin{aligned} \xi_{k,\tau,w} &= \frac{2\hat{\sigma}_{k,\tau,w}^2 (\bar{\sigma}_\tau^2 + \nabla_{k,\tau,w})}{\text{Var}(\hat{\sigma}_{\tau,w}^2)} \\ &= \frac{2\hat{\sigma}_{k,\tau,w}^2 \bar{\sigma}_\tau^2}{\text{Var}(\hat{\sigma}_{\tau,w}^2)} \frac{\bar{\sigma}_\tau^2 + \nabla_{k,\tau,w}}{\bar{\sigma}_\tau^2} \end{aligned} \quad (32)$$

$$\sim \Gamma\left(\frac{p_{\tau,w}}{2}, 2 \frac{\bar{\sigma}_\tau^2 + \nabla_{k,\tau,w}}{\bar{\sigma}_\tau^2}\right). \quad (33)$$

The quantity  $\xi$  refers to a single measurement. To design an overall model test, we need to express the distribution of a vector of observables  $\boldsymbol{\xi}$  containing the observation of all measurements. However, expanding expression (33) to a multivariate gamma distribution is not trivial. In this work we focus on the scalar case, leaving the extension to the multivariate case for future work. It is worth noting that we can derive an overall model test by approximating the gamma distribution (33) to a Gaussian distribution. This holds for a number of degrees of freedom  $p_{\tau,w}$

sufficiently large, which occurs when the sampling interval  $\tau$  is small compared to the window length  $W$ . Under this approximation we can follow the same procedure previously derived for tests on the phase measurements: compute the DAVAR residuals as the difference between observed and expected behaviors, and derive the null and alternative hypotheses, which would be in the same form of (19) and (25). However, in this paper we only focus on the w-test derived from the exact distribution (33) of  $\xi$ , which is valid for any sampling interval. The corresponding hypotheses read

$$\begin{cases} \mathcal{H}_0^{(\xi)} : \xi_{k,\tau,w} \sim \Gamma\left(\frac{p_{\tau,w}}{2}, 2\right) \\ \mathcal{H}_A^{(\xi)} : \xi_{k,\tau,w} \sim \Gamma\left(\frac{p_{\tau,w}}{2}, 2 \frac{\bar{\sigma}_\tau^2 + \nabla_{k,\tau,w}}{\bar{\sigma}_\tau^2}\right) \end{cases}. \quad (34)$$

#### 4. Fault detection and identification with the GLRT

In this section we derive the statistical tests based on the observables' distributions. In the GLRT framework we devise a model-based test, which can be in the form of an overall model test or a w-test, and a self-consistency test. The derivation of the detectors starts from the definition of the hypotheses on the observed distributions. In section 3 we showed that the hypotheses are of the same form for normally distributed observables (Kalman filter residuals and phase measurements), thus we need only one derivation of a test expression which can then be applied to all of them. Additionally, we need a separate derivation of the w-test for gamma distributed observables (normalized DAVAR). Table 2 gives an overview of the observables and the tests derived in this work.

Assume we have a vector of observables with given distribution  $\mathbf{y} \sim f(\mathbf{y}|\alpha)$ , where the parameter  $\alpha$  takes values in the set  $\mathcal{G}$ . If we consider the hypotheses

$$\begin{cases} \mathcal{H}_0 : \alpha \in \mathcal{G}_0 \\ \mathcal{H}_A : \alpha \in \mathcal{G} \end{cases}, \quad (35)$$

where  $\mathcal{G}_0 \subset \mathcal{G}$ , the GLRT for this problem is then defined as

$$\text{Reject } \mathcal{H}_0 \text{ if } \frac{\max_{\alpha \in \mathcal{G}_0} f(\mathbf{y}|\alpha)}{\max_{\alpha \in \mathcal{G}} f(\mathbf{y}|\alpha)} < a. \quad (36)$$

The threshold  $a$  defines the critical region  $\mathcal{K}$ , i.e. the set of values of  $\mathbf{y}$  leading to a rejection of  $\mathcal{H}_0$ .

Since the measurements are affected by noise, the test can deliver a wrong result. We talk of false alarm (or type I error) if  $\mathcal{H}_0$  is rejected when in fact  $\mathcal{H}_0$  is true. The probability of false alarm  $P_{fa}$  of the test (or size of the test) is

$$P_{fa} = \int_{\mathcal{K}} f(\mathbf{y}|\mathcal{H}_0) d\mathbf{y}. \quad (37)$$

Table 2: Overview of the observables used and of the tests derived in this work.

Quantity	Notation			Model-based test		Self-consistency test
	Observable	Residual	Distribution	Overall model test	w-test	
Kalman filter residuals	-	$\rho_k^{(kf)}$	$\mathcal{N}$	✓	✓	$\chi^{[a]}$
Phase measurements	$\mathbf{z}_k$	$\rho_k^{(z)}$	$\mathcal{N}$	✓	✓	✓
Normalized DAVAR	$\xi_{k,\tau,w}$	-	$\Gamma$	$\chi^{[b]}$	✓	$\chi^{[b]}$

[a] self-consistency test is not applicable, since Kalman filter residuals already include model information.

[b] not derived in this work.

A missed detection (type II error) occurs instead when  $\mathcal{H}_0$  is wrongly accepted, when in fact  $\mathcal{H}_A$  is true. The probability of missed detection  $P_{\text{md}}$  is

$$P_{\text{md}} = 1 - \int_{\mathcal{K}} f(\mathbf{y}|\mathcal{H}_A) d\mathbf{y} . \quad (38)$$

#### 4.1. Model-based GLRT for normally distributed observables

The Kalman filter residuals and the phase measurements residuals follow a normal distribution. Thus, we generalize the hypotheses (19) and (25) by removing all observable-specific annotations:

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{\rho} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}) \\ \mathcal{H}_A : \boldsymbol{\rho} \sim \mathcal{N}(\mathbf{C}\boldsymbol{\nabla}, \boldsymbol{\Omega}) \end{cases} . \quad (39)$$

This allows us to derive a general expression of the test, which is then used for the different observables by using the respective matrices. By following the derivation shown in Appendix A, we obtain the test

$$\text{Reject } \mathcal{H}_0 \text{ if } T = \hat{\boldsymbol{\nabla}}^\top \mathbf{Q}_{\hat{\boldsymbol{\nabla}}\hat{\boldsymbol{\nabla}}}^{-1} \hat{\boldsymbol{\nabla}} > k , \quad (40)$$

with

$$\hat{\boldsymbol{\nabla}} = (\mathbf{C}^\top \boldsymbol{\Omega}^{-1} \mathbf{C})^{-1} \mathbf{C}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\rho} , \quad (41)$$

$$\mathbf{Q}_{\hat{\boldsymbol{\nabla}}\hat{\boldsymbol{\nabla}}} = (\mathbf{C}^\top \boldsymbol{\Omega}^{-1} \mathbf{C})^{-1} . \quad (42)$$

If  $q$  is the number of entries in  $\hat{\boldsymbol{\nabla}}$ , the test  $T$  is  $\chi^2$ -distributed with  $q$  degrees of freedom. The distribution is central under the null hypothesis and non-central under the alternative hypothesis:

$$\begin{cases} \mathcal{H}_0 : T \sim \chi^2(q, 0) \\ \mathcal{H}_A : T \sim \chi^2(q, \lambda) \end{cases} , \quad (43)$$

with non-centrality parameter

$$\lambda = \boldsymbol{\nabla}^\top \mathbf{Q}_{\hat{\boldsymbol{\nabla}}\hat{\boldsymbol{\nabla}}}^{-1} \boldsymbol{\nabla} . \quad (44)$$

We design two versions of this test by varying the shape of matrix  $\mathbf{C}$ . In the limit case of  $q = M$  we obtain the overall model test, while setting  $q = 1$  yields the w-test.

**4.1.1. Overall model test:** in the overall model test we have  $q = M$  and  $\mathbf{C} = \mathbf{I}$ , leading to  $\hat{\boldsymbol{\nabla}} = \boldsymbol{\rho}$ . The estimate of fault vector  $\hat{\boldsymbol{\nabla}}$  completely absorbs the residuals  $\boldsymbol{\rho}$ , and the test reduces to

$$\text{Reject } \mathcal{H}_0 \text{ if } T = \boldsymbol{\rho}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\rho} > k_M . \quad (45)$$

The test is  $\chi^2$ -distributed with  $M$  degrees of freedom, and we compute the threshold  $k_M$  from the expression

$$P_{\text{fa}} = \int_{k_M}^{+\infty} \chi^2(\eta|q = M, \lambda = 0) d\eta . \quad (46)$$

The probability of missed detection for a given value of non-centrality parameter  $\lambda$  is

$$P_{\text{md}} = \int_0^{k_M} \chi^2(\eta|q = M, \lambda) d\eta . \quad (47)$$

We employ the overall model test for the detection step, in which all the entries of the observation vector are tested as a whole.

**4.1.2. w-test:** the w-test is derived by setting  $q = 1$ . The matrix  $\mathbf{C}$  becomes a column vector  $\mathbf{c}_i$ , where the  $i$ -th entry is equal to one, and zero elsewhere. By iteratively applying the test to all measurements in the ensemble (varying  $i$  from 1 to  $M$ ), we can sweep all the measurements and locate the one which is most probably faulty. This process is also known as data-snooping. The test becomes

$$\text{Reject } \mathcal{H}_0 \text{ if } T = \frac{\hat{\boldsymbol{\nabla}}^2}{\sigma_{\hat{\boldsymbol{\nabla}}}^2} > k_1 , \quad (48)$$

where  $\hat{\boldsymbol{\nabla}} = \frac{\mathbf{c}_i^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\rho}}{\mathbf{c}_i^\top \boldsymbol{\Omega}^{-1} \mathbf{c}_i}$  and  $\sigma_{\hat{\boldsymbol{\nabla}}}^{-2} = \mathbf{c}_i^\top \boldsymbol{\Omega}^{-1} \mathbf{c}_i$  are now scalar. The test is  $\chi^2$ -distributed with 1 degree of freedom, central under the null hypothesis, non-central under the alternative hypothesis. We compute the threshold  $k_1$  from the expression

$$P_{\text{fa}} = \int_{k_1}^{+\infty} \chi^2(\eta|q = 1, \lambda = 0) d\eta , \quad (49)$$

while the non-centrality parameter  $\lambda = \boldsymbol{\nabla}^\top \mathbf{c}_i^\top \boldsymbol{\Omega}^{-1} \mathbf{c}_i$  is related to the  $P_{\text{md}}$ :

$$P_{\text{md}} = \int_0^{k_1} \chi^2(\eta|q = 1, \lambda) d\eta . \quad (50)$$



The MDB represents the minimum magnitude of fault that can be detected by the w-test with a given value of  $P_{\text{md}}$ :

$$|\nabla_{\text{MDB}}| = \sqrt{\frac{\lambda}{\mathbf{c}_i^T \boldsymbol{\Omega}^{-1} \mathbf{c}_i}} \quad (51)$$

#### 4.2. Self-consistency test for normally distributed observables

The model-based test requires to know a-priori how the clocks behave in their environment to determine the parameters of the clock model. This is problematic if the clocks cannot be directly controlled or accessed, for instance for clocks onboard satellites: in this cases the application of a clock model not reproducing the actual nominal conditions of the system can lead to mismodelling, missed detection, and in general poor performance of the integrity monitoring techniques. This motivates the self-consistency test, in which we only set the structure of the ensemble and the distribution function of the measurements, but the model parameters are left free and are estimated from the observations [1]. In the case of homogeneous ensembles, where the clocks are of the same type, the test estimates the expected value and variance of the measurements, as well as assessing the occurrence of a fault in a measurement. While the model-based tests are computed from the residuals, defined as the difference between the observation and the modelled behavior, the self-consistency test is computed directly from the observations. We develop the self-consistency test for the phase measurements (21), while the derivation for gamma distributed observables is left for future work.

In an homogeneous ensemble, we expect all clock measurements to exhibit the same mean value and variance, thus we rewrite expression (21) as

$$\mathbf{z}_k \sim \mathcal{N}(\zeta_k \mathbf{u}, v_k^2 \boldsymbol{\Psi}) , \quad (52)$$

where  $\mathbf{u}$  is a  $M$ -sized vector of ones,  $\zeta_k$  the magnitude of the expected value,  $\boldsymbol{\Psi}$  dictates the shape of the covariance matrix, and  $v_k^2$  controls the magnitude of the covariance matrix. The test directly estimates the scalars  $\zeta_k$  and  $v_k^2$  from the observations  $\mathbf{z}_k$ . We set the matrix  $\boldsymbol{\Psi}$  according to the measurement topology of the ensemble:

$$\boldsymbol{\Psi} = \left| \overline{\mathbf{H}} \overline{\mathbf{H}}^T \right|^{-1} \overline{\mathbf{H}} \overline{\mathbf{H}}^T = \frac{1}{M} \overline{\mathbf{H}} \overline{\mathbf{H}}^T . \quad (53)$$

By dropping the subscript for time dependence, the hypotheses read

$$\begin{cases} \mathcal{H}_0^{(s)} : \mathbf{z} \sim \mathcal{N}(\zeta \mathbf{u}, v^2 \boldsymbol{\Psi}) \\ \mathcal{H}_A^{(s)} : \mathbf{z} \sim \mathcal{N}(\zeta \mathbf{u} + \mathbf{C} \nabla, v^2 \boldsymbol{\Psi}) \end{cases} . \quad (54)$$

The derivation in Appendix B leads to the self-consistency test

$$\text{Reject } \mathcal{H}_0^{(s)} \text{ if } T = \frac{\hat{\nabla}^T \mathbf{R}_{\hat{\nabla} \hat{\nabla}}^{-1} \hat{\nabla}}{q \hat{v}_A^2} > k_s , \quad (55)$$

with

$$\mathbf{R}_{\hat{\nabla} \hat{\nabla}} = (\mathbf{C}^T \boldsymbol{\Psi}^{-1} \mathbf{P}_u^\perp \mathbf{C})^{-1} , \quad (56)$$

$$\hat{\nabla} = \mathbf{R}_{\hat{\nabla} \hat{\nabla}} \mathbf{C}^T \boldsymbol{\Psi}^{-1} \mathbf{P}_u^\perp \mathbf{z} , \quad (57)$$

$$\hat{v}_A^2 = (M - q - 1)^{-1} \mathbf{z}^T \boldsymbol{\Psi}^{-1} \mathbf{P}_C^\perp \mathbf{P}_u^\perp \mathbf{z} , \quad (58)$$

where  $q$  is the number of columns in  $\mathbf{C}$ , and we use the following idempotent matrices:

$$\mathbf{P}_u^\perp = \mathbf{I} - \mathbf{u} (\mathbf{u}^T \boldsymbol{\Psi}^{-1} \mathbf{u})^{-1} \mathbf{u}^T \boldsymbol{\Psi}^{-1} , \quad (59)$$

$$\mathbf{P}_C^\perp = \mathbf{I} - \mathbf{C} (\mathbf{C}^T \boldsymbol{\Psi}^{-1} \mathbf{C})^{-1} \mathbf{C}^T \boldsymbol{\Psi}^{-1} , \quad (60)$$

$$\mathbf{P}_u^\perp = \mathbf{I} - \mathbf{u} (\mathbf{u}^T \boldsymbol{\Psi}^{-1} \mathbf{P}_C^\perp \mathbf{u})^{-1} \mathbf{u}^T \boldsymbol{\Psi}^{-1} \mathbf{P}_C^\perp . \quad (61)$$

The test (55) is  $F$ -distributed

$$\begin{cases} \mathcal{H}_0^{(s)} : T \sim F(q, M - 1 - q, 0) \\ \mathcal{H}_A^{(s)} : T \sim F(q, M - 1 - q, \lambda) \end{cases} , \quad (62)$$

with non-centrality parameter

$$\lambda = \nabla^T \mathbf{R}_{\hat{\nabla} \hat{\nabla}}^{-1} \nabla v^{-2} . \quad (63)$$

The test threshold  $k_s$  can be found by inverting the expression for the  $P_{\text{fa}}$ :

$$P_{\text{fa}} = \int_{k_s}^{+\infty} F(\eta | q, M - 1 - q, \lambda) d\eta . \quad (64)$$

In the following we apply the self-consistency test in a manner similar to the w-test, thus we set  $\mathbf{C} = \mathbf{c}_i$ ,  $q = 1$ , and we sweep the different measurements by changing the position  $i$  of the non-zero entry in the vector  $\mathbf{c}_i$ . The MDB for the self-consistency test is

$$|\nabla_{\text{MDB}}^{(s)}| = \sqrt{\frac{\lambda}{R_{\hat{\nabla} \hat{\nabla}}^{-1}} v^2} , \quad (65)$$

where in this case  $R_{\hat{\nabla} \hat{\nabla}}$  is a scalar.

#### 4.3. w-test for gamma-distributed observables

As previously shown, the normalized DAVAR  $\xi_{k,\tau,W}$  follows a gamma distribution, for which a dedicated test is here developed. In this work we focus on the scalar version, leading to the w-test. Dropping all subscripts, the hypotheses are

$$\begin{cases} \mathcal{H}_0^{(r)} : \xi \sim \Gamma\left(\frac{p}{2}, 2\right) \\ \mathcal{H}_A^{(r)} : \xi \sim \Gamma\left(\frac{p}{2}, 2 \frac{\bar{\sigma}^2 + \nabla}{\bar{\sigma}^2}\right) \end{cases} . \quad (66)$$

The derivation in Appendix C leads to

$$T = \frac{\hat{\sigma}^2}{\bar{\sigma}^2} , \quad (67)$$

which distributes according to a gamma distribution

$$\begin{cases} \mathcal{H}_0^{(r)} : T \sim \Gamma\left(\frac{p}{2}, \frac{2}{p}\right) \\ \mathcal{H}_A^{(r)} : T \sim \Gamma\left(\frac{p}{2}, \frac{2}{p} \frac{\bar{\sigma}^2 + \nabla}{\bar{\sigma}^2}\right) \end{cases} \quad (68)$$

Thus, the w-test on the DAVAR becomes

$$\text{Reject } \mathcal{H}_0^{(r)} \text{ if } T < k_l \text{ or } T > k_u, \quad (69)$$

where  $k_l$  is the lower threshold and  $k_u$  is the upper threshold, which can be computed from the expression of the  $P_{fa}$ :

$$P_{fa} = 1 - \int_{k_l}^{k_u} \Gamma\left(\eta \left| \frac{p}{2}, \frac{2}{p} \right.\right) d\eta. \quad (70)$$

The  $P_{md}$  can be found from:

$$P_{md} = \int_{k_l}^{k_u} \Gamma\left(\eta \left| \frac{p}{2}, \frac{2}{p} \frac{\bar{\sigma}^2}{\bar{\sigma}^2 + \nabla_{MDB}} \right.\right) d\eta. \quad (71)$$

The test (69) is double-sided. One can argue that a fault in the system can only cause a degradation of the observed stability and thus an increase in DAVAR, making the lower side of the test pointless. However, this test detects differences between observation and model, therefore we can use the lower side to detect mismodelling. If we select a model predicting an DAVAR higher than the real value, the test will fall below the lower threshold, and the triggering of the test can be accordingly marked.

#### 4.4. Monte-Carlo analysis

We run a set of Monte Carlo simulations of an ensemble comprising 5 cesium references to verify the theoretical distributions of the devised tests, and to compare the observed values of  $P_{fa}$  and  $P_{md}$  to the expected ones. This is only possible in simulation, since we need a large number of runs to evaluate the statistical behavior. Two simulations are carried out: one for the nominal case, and one for the faulty case, where a fault is injected in the first measurement. The selected model parameters of the cesium standards are

$$\begin{aligned} {}_i\sigma_1^2 &= 9 \cdot 10^{-24} \text{ s}, \quad {}_i\sigma_2^2 = 0, \quad {}_id = 0, \\ {}_iU_M &= 10^{-23} \text{ s}, \quad {}_iR_M = 0.57 \text{ s}^{-1}, \end{aligned} \quad (72)$$

where  ${}_iU_M$  and  ${}_iR_M$  parametrize the Markov process. As discussed in section 3.3, we also need a model without Markov processes to obtain a prediction model for the DAVAR, thus we choose

$${}_i\sigma_1^2 = 4.5 \cdot 10^{-23} \text{ s}, \quad {}_i\sigma_2^2 = 0, \quad {}_id = 0, \quad {}_iU_M = 0. \quad (73)$$

In both cases we assume a value for the measurement noise variance of  $\mathbf{R} = \mathbf{I}_M \cdot 10^{-25} \text{ s}^2$ . These values are obtained by a laboratory characterization of the cesium frequency references used in this work. Long term measurements of the clocks are taken and their DAVAR

computed, which is then fitted to derive the model parameters. We set the probability of false alarm to

$$P_{fa} = 10^{-3}, \quad (74)$$

leading to the threshold value for the overall model test  $k_M = 18.5$ , for the w-test  $k_l = 10.8$ , and for the self-consistency test  $k_s = 998.5$ . The thresholds  $k_l$  and  $k_u$  for the DAVAR test depend on the length of sliding window  $W$  and on the sampling interval. A representation is given in the following experimental results. In the different simulations, the  $P_{md}$  can be computed using the threshold value and the magnitude of the injected drift.

**4.4.1. Kalman filter residuals test:** we run the Kalman filter in the nominal case and in the faulty case, where we inject a fault of magnitude  $\nabla = 10^{-11} \text{ s}$  on the first residual, at all time steps. Thus we have  $\nabla = \nabla \mathbf{c}_1$ , where  $\mathbf{c}_1 = [1 \ 0 \ \dots \ 0]^T$  is a column vector of  $M$  entries. We then check the distributions of the overall model test and of the w-test under the two hypotheses. In the faulty case we expect a time dependent non-centrality parameter

$$\lambda_k = \nabla^2 \mathbf{c}_1^T (\mathbf{\Omega}_k^{(kf)})^{-1} \mathbf{c}_1. \quad (75)$$

The matrix  $\mathbf{\Omega}_k^{(kf)}$  depends on time, but it converges after some iterations of the Kalman filter, as long as the measurements do not show abrupt changes. For  $t_k = 100 \text{ s}$  we obtain  $\lambda_{100} = 5.2$ . From expressions (47) and (50) we expect a  $P_{md}$  for the overall model test of  $P_{md}^{(om)}(\lambda_{100}) = 0.94$ , while for the w-test of  $P_{md}^{(wt)}(\lambda_{100}) = 0.84$ .

Figure 2 shows the distributions of the overall model test and of the w-test in the nominal case  $\mathcal{H}_0^{(kf)}$  and faulty case  $\mathcal{H}_A^{(kf)}$ , as well as the theoretical expected distributions. By counting the number of triggerings in the nominal case based on  $10^5$  simulations, we compute  $\hat{P}_{fa} = 9.6 \cdot 10^{-4}$ , which is in good agreement with the expected value of  $P_{fa} = 10^{-3}$ . If instead we count the times the test was not triggered in the faulty case, we obtain  $\hat{P}_{md}^{(om)} = 0.94$  and  $\hat{P}_{md}^{(wt)} = 0.84$  for the w-test. These values are in agreement with the theoretical derivation of the tests on the Kalman filter residuals.

**4.4.2. Phase measurements test:** in this case we inject a fault with linearly increasing magnitude with slope  $\frac{d\nabla}{dt} = 2.5 \cdot 10^{-12} \text{ s/s}$  in the first entry of the clock measurements vector  $\mathbf{z}$ . While the distribution of test in the null hypothesis is constant in time, in the faulty case the distribution is time dependent, since the injected fault is growing. Thus we verify the distribution for  $t_k = 500 \text{ s}$ , leading to  $\lambda_{500} = 57.2$  for the overall model test and w-test, and  $\lambda_{500}^{(s)} = 53.6$  for the self-consistency test. Figure 3 shows the theoretical and observed distributions of the three tests in the

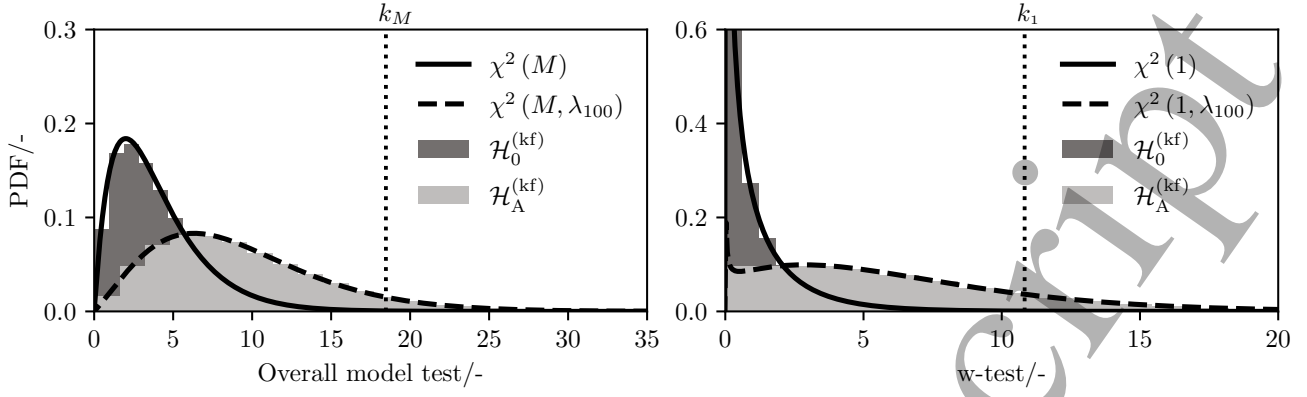


Figure 2: Distribution at  $t_k = 100$  s of the overall model test and of the w-test applied to the Kalman filter residuals in the nominal case  $\mathcal{H}_0^{(kf)}$  and in the faulty case  $\mathcal{H}_A^{(kf)}$  ( $\nabla = 10^{-11}$  s added on the first residual at every time step). The histograms show the distributions of the test from Monte Carlo simulations, while the lines show the theoretical distributions.

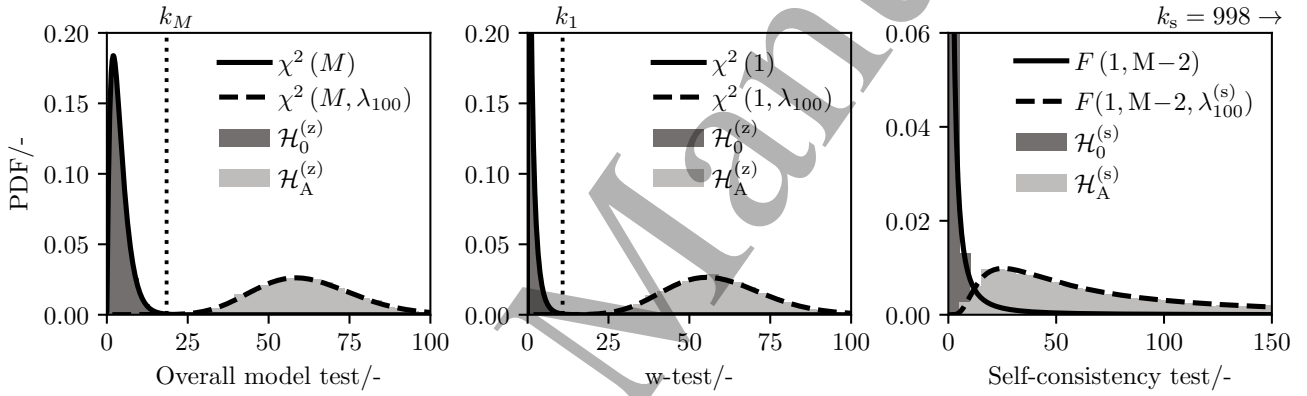


Figure 3: Left: distribution of the overall model test applied to the phase residuals. Center: distribution of the w-test applied to the phase residuals. Right: distribution of the self-consistency test applied to the phase measurements (threshold out of scale). All distributions are provided in the nominal and in the faulty case for  $t_k = 500$  s, where a drift  $\nabla$  with slope  $\frac{d\nabla}{dt} = 2.5 \cdot 10^{-12}$  s/s is added on the first measurement. The histograms show the distributions of the test from Monte Carlo simulations, while the lines show the theoretical distributions.

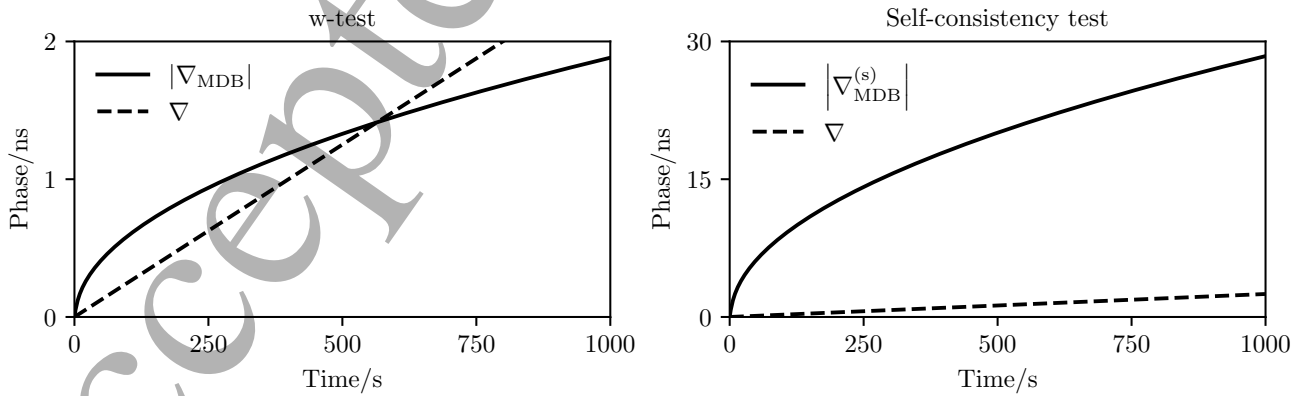


Figure 4: Evolution of the minimum detectable bias (MDB) for the w-test ( $|\nabla_{MDB}|$ ) and self-consistency test ( $|\nabla_{MDB}^{(s)}|$ ) on the phase measurements, as function of time. The MDB grows in time since the magnitude of the observables' covariance matrix increases. The dashed lines represent the magnitude of the injected fault  $\nabla$ . In this particular case, the self-consistency test (right plot) would not detect the slowly increasing drift.

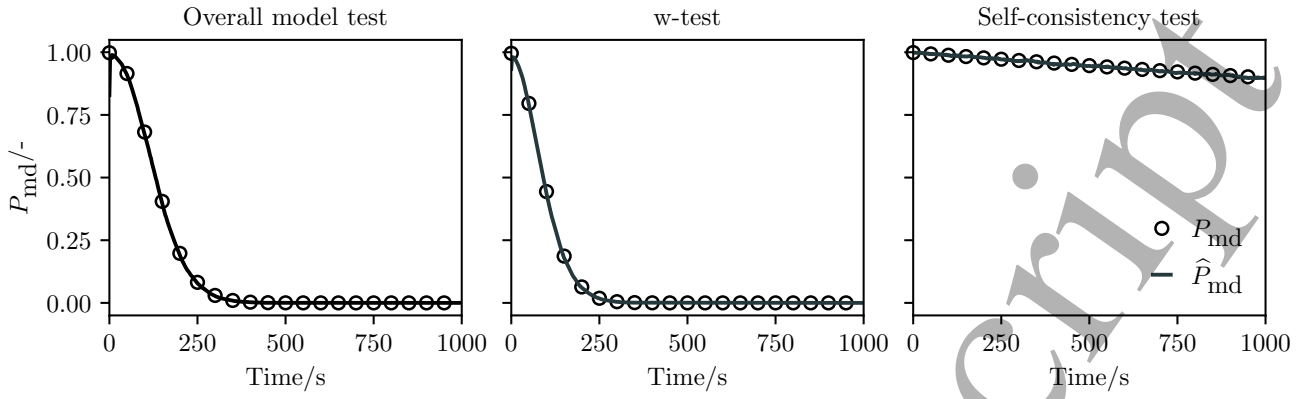


Figure 5: Expected ( $P_{md}$ ) and observed ( $\hat{P}_{md}$ ) probability of missed detection for the tests applied to phase measurements. The theoretical values are found using the corresponding integrals defining the  $P_{md}$ , while we compute the observed values by counting the number of triggerings of the given test during the Monte Carlo simulation. As the injected fault increases faster than then the MDB (see figure 4), the  $P_{md}$  decreases over time.

nominal and faulty cases for  $t_k = 500$  s, while in figure 4 the magnitude of the injected fault is compared to the MDB for the w-test and self-consistency test. Since the entries of the matrix  $\Omega_k^{(z)}$  used to compute the non-centrality parameter are increasing, the MDB grows in time. A slowly increasing drift can be detected only when its magnitude becomes larger than the MDB. In figure 5 we plot a comparison of the expected and observed probabilities of missed detection for the overall model test, w-test and self-consistency test as function of time. The theoretical values are computed with (47) for the overall model test, (50) for the w-test and (71) for the self-consistency test. At the beginning, the injected fault is smaller than the MDB and remains undetected. The detection occurs later when the fault magnitude exceeds the MDB, which is reflected in the decrease in  $P_{md}$ .

**4.4.3. DAVAR test:** to verify the distributions of the tests on the DAVAR, we simulated an ensemble in the nominal case using the clock parameters (73), while in the faulty case we increase the first parameters of the second clock by a factor 3, so that  $2\sigma_1^2 = 1.35 \cdot 10^{-22}$  s. In this way the first entry in the measurement vector shows a fault, whose magnitude  $\nabla(\tau)$  is computed by subtracting the theoretical DAVAR values of the first and second clocks. We employ a sliding window of 1000 s, and we compute the test for different values of sampling interval  $\tau$ . The upper plots in figure 6 show the distribution of the normalized DAVAR (66) for an increasing sampling interval  $\tau$ , while the lower plots show the distribution of the corresponding w-test developed using the gamma distribution (68). Figure 7 shows the observed and expected  $P_{md}$  for the DAVAR w-test at increasing values of  $\tau$ , while figure 8 compares the magnitude of the injected fault  $\nabla(\tau)$  to the value

of the MDB ( $\nabla_{MDB}$ ). The error can be captured at shorter sampling intervals, while in the long term the statistical uncertainty in the determination of the DAVAR increases, thus weakening the power of detection.

## 5. Hardware setup and measurements

In our laboratory we run an ensemble of 5 cesium frequency references (Symmetricron 5071A), structured as in figure 1. With a SpectraDynamics high resolution offset generator (HROG) we inject a fault of desired magnitude in the signal of the second cesium. Finally, a computer collects the measurements generated by a K+K FXE80 counter, and runs the ensembling and FDI algorithms in Matlab. We run three scenarios with different faults: in the first we inject a series of phase steps; the second scenario involves a phase and frequency drift; finally, a periodic phase component is injected in the third case. Here we set the integrity requirements as  $P_{fa} = 10^{-3}$ ,  $P_{md} = 10^{-6}$ .

### 5.1. Scenario I: injection of phase steps

Figure 9 shows the phase measurements in the first scenario along with the injected phase steps. The phase pattern includes steps of different size and sign. For this scenario we only compute the tests on the Kalman filter residuals and on the phase measurements, but not on the DAVAR.

**5.1.1. Tests on the Kalman filter residuals:** applying the overall model test on the Kalman filter residuals generates the results shown in figure 10. All the injected steps cause a spike in the test value, whose level depends on the size of the respective step. Only

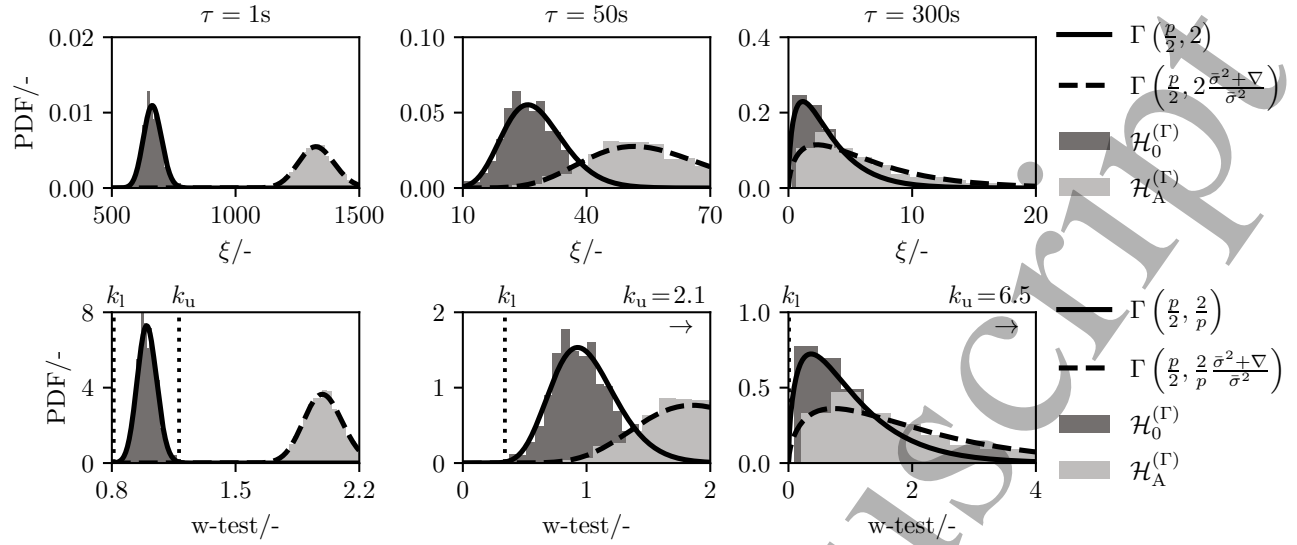


Figure 6: The upper plots show the distribution of the normalized DAVAR  $\xi$  for different sampling intervals, for a sliding window set to  $W = 10^3\text{s}$ . The lower plots show the distributions of the corresponding w-test and the test thresholds. The histograms show the distributions resulting from Monte Carlo simulations, while the lines show the theoretical distributions. Please note the different scaling of the axes.

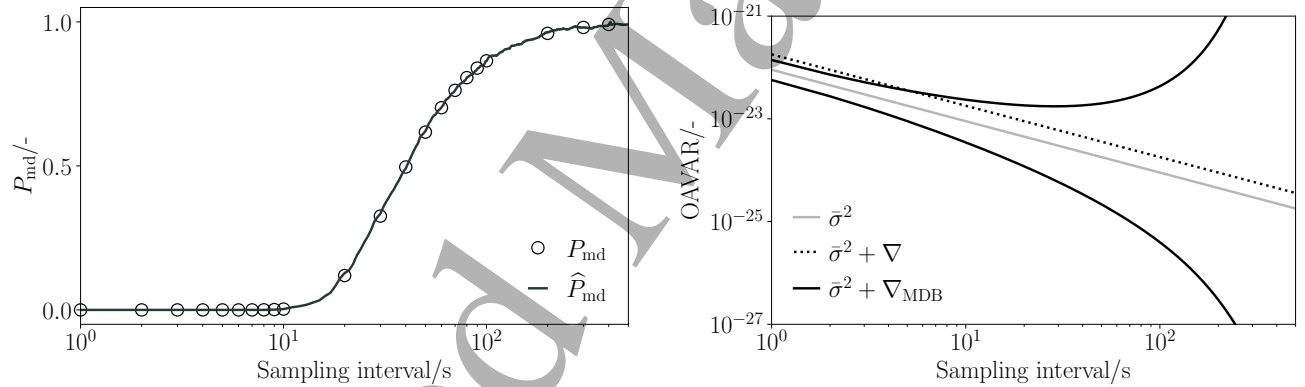


Figure 7: Expected ( $P_{md}$ ) and observed ( $\hat{P}_{md}$ ) probability of missed detection for the w-test on the DAVAR, for different sampling intervals and fixed length of the sliding window  $W = 10^3\text{s}$ .

Figure 8: The solid gray line represents the nominal OAVAR  $\sigma^2(\tau)$ , the dotted line shows the faulty OAVAR, which is shifted from the nominal by the injected fault  $\nabla$ , and the solid black lines represent the minimum detectable bias  $\nabla_{MDB}$  for the DAVAR test, using  $W = 10^3\text{s}$ .

the smallest steps around  $t_k = 270\text{s}$  do not trigger the test, but this is expected as their size is comparable with the MDB ( $|\nabla_{MDB}^{(kf)}| = 52.1\text{ps}$ ). For locating the faulty measurement once the overall model test is triggered, we perform the w-test shown in figure 10b. Although the test on the first channel (using  $\mathbf{c}_1$ ) shows the highest values, the tests on the other channels trigger as well. This is due to the correlation between the differential measurements using a common clock, also encoded by the non-diagonal structure of the matrix  $\mathbf{\Omega}_k$ . Thus, we cannot unambiguously identify the faulty measurement yet. To perform identification we apply the following strategy:

- (i) find which vector  $\mathbf{c}_j$  generates the highest test value;
- (ii) create a reduced observables vector  $\boldsymbol{\rho}_R$  by removing the entry  $j$  corresponding to  $\mathbf{c}_j$ ;
- (iii) perform the overall model test on the reduced observables  $\boldsymbol{\rho}_R$ :
  - (a) if the test is triggered,  $\boldsymbol{\rho}_R$  still contains the faulty measurement. Restart from step (i) to remove one more entry from the reduced vector  $\boldsymbol{\rho}_R$ ;
  - (b) if the test is *not* triggered, the faulty measure-

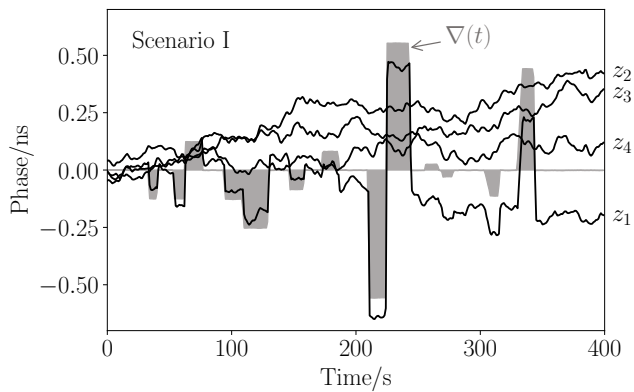


Figure 9: Phase measurements  $\mathbf{z}(t)$  in the first scenario. The shaded area shows the injected pattern of phase steps  $\nabla(t)$ .

ment was correctly removed: measurement  $j$  is considered as faulty;

- (iv) repeat the process until a faulty measurement is identified. If all entries are removed from the observables vector  $\boldsymbol{\rho}$ , the algorithm cannot identify a faulty measurement, and the identification process fails.

The result of this process represents the measurement which is most probably faulty, shown in figure 10c. Here, we use the symbol  $\emptyset$  when the fault identification process is unsuccessful. As expected, the first measurement is identified as faulty in all the detected steps. Figure 10d shows the overall model test computed on the reduced vector of observables. The reduced test is never triggered, meaning that the algorithm successfully identifies the faulty measurement. Removing entries in the vector of observables requires recomputing the test threshold with a smaller number of degrees of freedom, leading to the dips shown in the threshold value.

**5.1.2. Tests on the phase measurements:** although this scenario is not designed to assess the phase measurements tests, we show in figure 11 the performance of the overall model test. As expected, only the largest steps trigger the test, since the magnitude of the covariance matrix (and therefore the magnitude of the MDB) increases over time.

## 5.2. Scenario II: injection of a drift

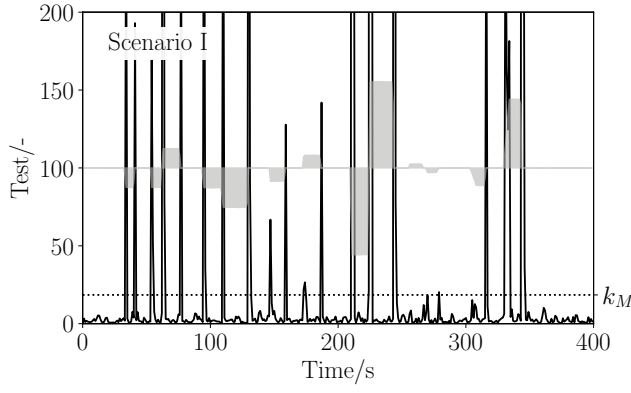
In the second experiment we inject a phase and frequency drift, shown in figure 12. It consists of three parts: until  $t_k = 10^5$  s no drift is added, while the frequency drift increases linearly for  $10^5$  s  $< t_k < 2 \cdot 10^5$  s, reaching a maximum value of 1 mHz, which is then kept constant until the end of the experiment. Corre-

spondingly, the phase drift shows first a quadratic increase and finally a linear growth. Figure 13 shows the OAVAR computed on the entire measurement length and the OAVAR of the clock models: the dotted points describe the model without Markov process, while the crossed points correspond to the model with Markov process.

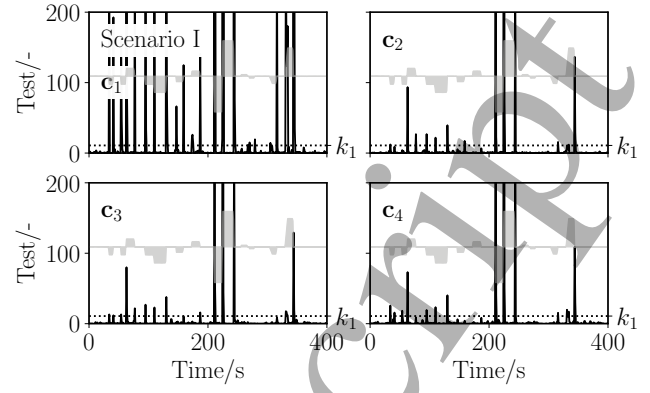
**5.2.1. Tests on the phase measurements:** figure 14 shows the results of the test applied to the phase measurement residuals. The injected drift is successfully detected about 4320 s after its onset at  $t_k = 10^5$  s. The overall model test in figure 14a shows a small increase at the beginning of the experiment. This is due to the nature of this test, which is more sensitive to small phase deviations during the initialization of the detection process. The w-test (figure 14b) correctly locates the drift on the first measurement, as seen in figure 14c. The inset plots show that the w-tests computed using  $\mathbf{c}_2$ ,  $\mathbf{c}_3$ , and  $\mathbf{c}_4$  grow more slowly than with  $\mathbf{c}_1$ , and are triggered on average 5000 s later. Figure 14d shows the overall model test on the reduced vector of phase residuals. The self-consistency test results in the values in figure 14e. This test identifies the drift with a delay of 21945 s using  $\mathbf{c}_1$ . Although the detection with this test occurs with a significant delay, one should consider that the injected drift is relatively small (100 ps/s), thus a longer time is needed for the fault to exceed the detection threshold.

**5.2.2. Test on the DAVAR:** we expect the test on the DAVAR to detect the fault at long sampling intervals, since the drift increases the measured DAVAR in the region above  $\tau = 10^2$  s (see figure 13). However, the successful detection of the drift depends on the length of the sliding window and on the current time step. The drift can be seen in the DAVAR only if the sliding window is long enough, in this case at least  $W > 10^3$  s. Figure 15 shows the DAVAR and the w-test (67) for three windows of length  $W = 10^5$  s starting at different time instants: before the drift onset, during the quadratic phase increase, and during the linear phase drift. The shaded area in the lower test plot represents the area where the null-hypothesis is accepted, delimited by the lower threshold  $k_l$  and the upper threshold  $k_u$ . The DAVAR computed on the window during the quadratic drift clearly shows that the fault can be detected in the region above  $\tau = 200$  s. However, the fault remains undetected in the last data segment, since the DAVAR is invariant to linear phase drifts. In the lower plot of figure 15 we note that the test is triggered also for short sampling intervals. This triggering is due to mismodelling: in figure 13 a mismatch can be noted between the measured DAVAR and the values predicted by both clock models. In

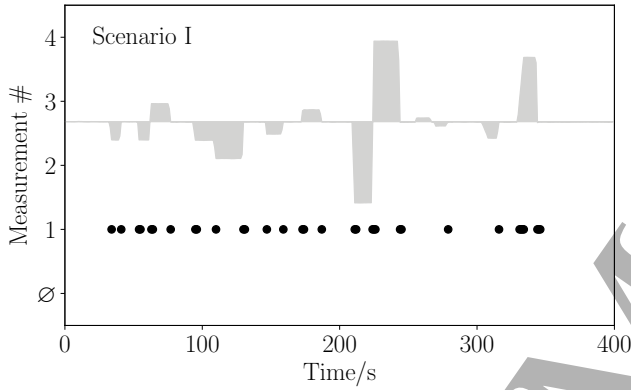




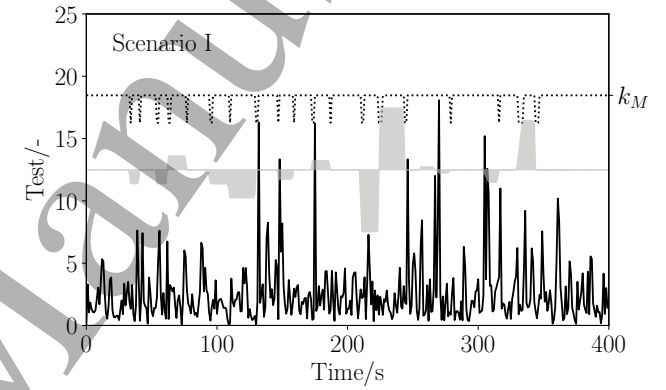
(a) Kalman filter residuals overall model test. The dotted line is the test threshold, the shaded curve shows the injected steps.



(b) Kalman filter residuals w-test, for different vectors  $c_i$ . The dotted line is the test threshold, the shaded curve shows the injected steps.



(c) Most probable faulty measurement according to the Kalman filter residuals test. The shaded curve shows the injected steps. The symbol  $\emptyset$  is used for unsuccessful identification (a fault is detected but not identified).



(d) Overall model test on the Kalman filter reduced residuals vector, where the entry of the faulty measurement is removed. The dotted line is the test threshold, the shaded curve shows the injected steps. When a entry is removed, we recompute the threshold with a smaller value of  $q$ , leading to the dips shown in the plot.

Figure 10: Scenario I - tests on the Kalman filter residuals in the presence of phase steps. Figure 10a shows the overall model test, figure 10b the w-test selecting the single measurements, figure 10c shows which measurement is most probably deemed as faulty, while figure 10d shows the reduced test, when the faulty measurement is excluded from the observation vector.

summary, we expect the DAVAR test to trigger due to the drift for  $\tau > 200$  s,  $W > 10^3$  s, and  $10^5$  s  $< t_k < 2 \cdot 10^5$  s. The rejections due to mismodelling should occur for all window lengths, all times, and  $\tau < 10$  s.

The value of the DAVAR test is function of time step  $t_k$ , sampling interval  $\tau$ , length of the sliding window  $W$ , and measurement channel. To show these dependencies, we introduce the compact visualization shown in figure 16. Here we plot the results of the w-test on the DAVAR for the first channel, for one value of sliding window length  $W$ . This plot is obtained by reworking the three-dimensional surface of the DAVAR as function of time and sampling interval, corresponding to the waterfall plots in [13].

Since we are mainly interested in knowing whether the test value lies within the thresholds (accept  $\mathcal{H}_0^{(r)}$ ) or outside (reject  $\mathcal{H}_0^{(r)}$ ), we can forgo plotting the test value on the z-axis, and we use instead a color code for showing test acceptance or rejection. Thus, we reduce the 3D surface to a plot on the  $(t_k - \tau)$  plane: the black pixels show the pairs  $(t_k, \tau)$  where the test is triggered, while we leave a white pixel when accepting  $\mathcal{H}_0^{(r)}$ . More information can be added to the visualization: for example, by cutting the three-dimensional DAVAR curve at a given time and draw a projection onto the  $(T - \tau)$  plane, we can visualize the test value as function of the sampling interval, along with the test thresholds (upper plot).

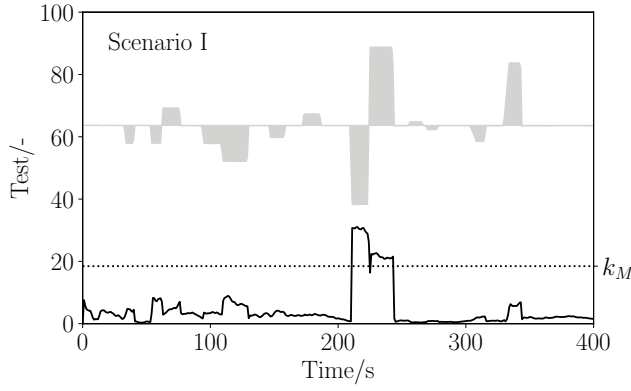


Figure 11: Scenario I - Phase measurement residuals overall model test in the presence of phase steps. The dotted line is the test threshold, the shaded curve shows the injected steps.

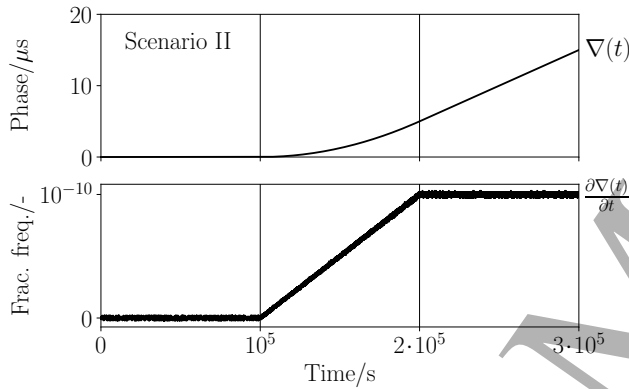


Figure 12: Scenario II - injected fault in terms of phase and fractional frequency differences. Three parts can be identified: initially no fault is added, then the injected phase grows quadratically, finally reaching a linear behavior.

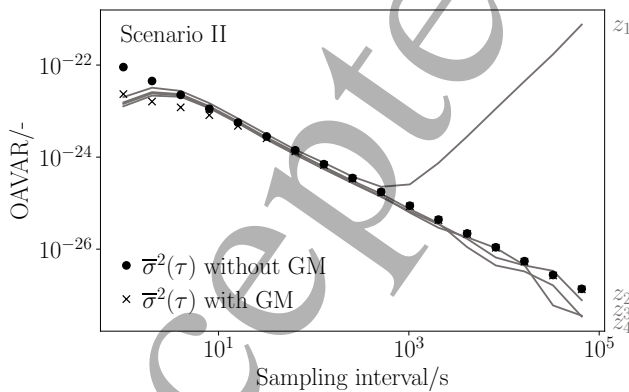


Figure 13: Measured OAVAR  $\hat{\sigma}^2$  of the four channels computed on the entire batch of experimental data, and theoretical expected OAVAR  $\bar{\sigma}^2$  for the two clock models, with and without Gauss-Markov processes (GM). The drift injected to the first measurement  $z_1$  is visible at long sampling intervals.

Conversely, if we cut the surface at a given sampling interval, we project onto the  $(T - t_k)$  plane on the right side the evolution of the test value as function of time (the shaded area represents the injected drift). The hatched area covers points where the test value is not available. The first value is available at  $t_k = W$ , since we must wait to collect enough points for computing the DAVAR on the window  $[0, W]$ . In this example we see that the test detects the mismodelling at short values of  $\tau$ , and the drift at longer sampling intervals. The detection delay depends on the window length and on the sampling interval. In figure 17a the w-test on the first measurement computed for  $W = 10^4$  s is plotted as function of time: the different lines show the test computed for different sampling intervals. Higher values of  $\tau$  result in higher test values and the detection tends to occur earlier. This is visualized also in figure 17b, where we plot the detection delay as function of the sampling interval for the first measurement. The minima occur for both window lengths at  $\tau = 2 \cdot 10^3$  s, with detection delays of about 3000 s.

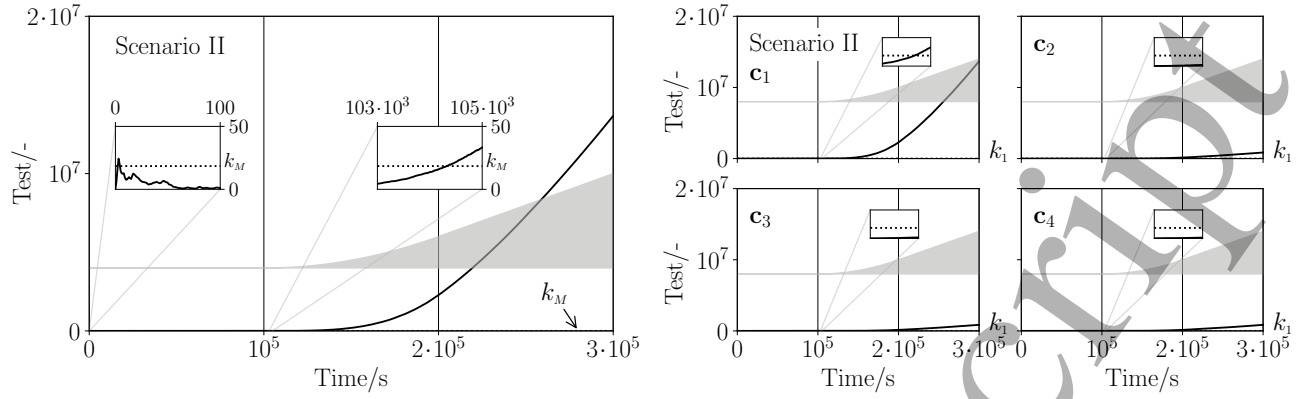
Figure 18 shows the result of the w-test on the DAVAR (68) based on the gamma distribution, for the 4 measurement channels  $z_i$ , and for different lengths of sliding window. We note how shorter sliding windows provide results at earlier time instants, but only for shorter sampling intervals. Conversely, longer windows yield values of DAVAR at longer sampling intervals, but only at later times. As expected, the test detects the drift only on the first measurement  $z_1$ , when using sufficiently long sliding windows (area “D”). The drift is detected only during the quadratic part, while the mismodelling is detected at short  $\tau$  at any time, for any window length, and on all the measurements (area “M”).

### 5.3. Scenario III: injection of a periodic component

In the third scenario we inject an oscillating phase component with a period of 90 minutes and an increasing amplitude. The amplitude increases similarly to the second scenario: for  $t_k < 10^5$  s the amplitude is zero, for  $10^5 \text{ s} < t_k < 2 \cdot 10^5$  s it increases linearly to reach a final amplitude of 1.38 ns (corresponding to  $5^\circ$  at a nominal frequency of 10 MHz). Figure 19 shows the injected fault.

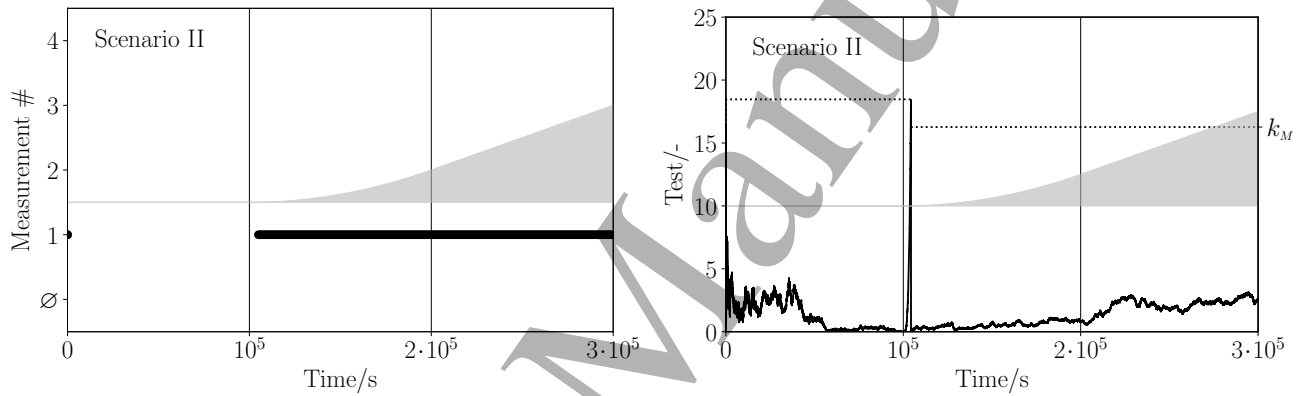
**5.3.1. Tests on the phase measurements:** figure 20 collects the results for the tests on the phase measurements. This test does not detect any fault, since the injected phase oscillation is relatively small compared to the process and measurement noises, and it lies within the bounds of the expected statistical variance. The self-consistency test on the phase measurements of figure 20b shows some





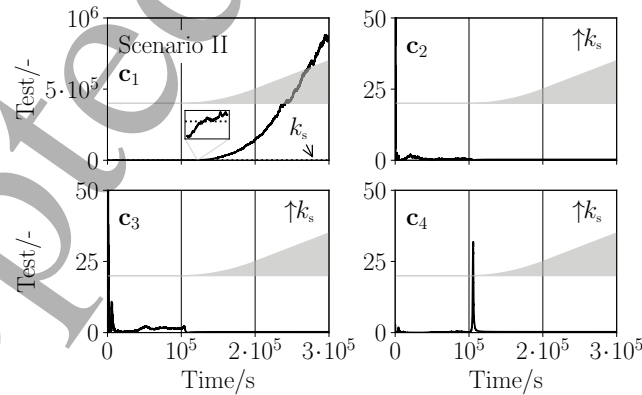
(a) Phase measurement residuals overall model test. The dotted line is the test threshold, the shaded curve represents the injected drift.

(b) Phase measurement residuals w-test, for different vectors  $\mathbf{c}_i$ . The dotted line is the test threshold, the shaded curve represents the injected drift. All the inset plots show the range  $1.03 \cdot 10^5$  s to  $1.05 \cdot 10^5$  s.



(c) Most probable faulty measurement according to the phase measurement residuals test. The shaded curve represents the injected drift. The symbol  $\emptyset$  is used for unsuccessful identification.

(d) Overall model test on the phase measurement reduced residuals vector, where the entry of the faulty measurement is removed. The dotted line is the test threshold, the shaded curve represents the injected drift.



(e) Self-consistency test on the phase measurements. Note the different scales of the y-axis: the test threshold  $k_s = 998$  is out of scale, except for the plot for  $\mathbf{c}_1$ . The shaded curve represents the injected drift.

Figure 14: Scenario II - tests on the phase measurements in the presence of phase drift. Figure 14a shows the overall model test, figure 14b the w-test on single measurements, figure 14c shows which measurement is most probably deemed as faulty, while figure 14d shows the reduced test, when the faulty measurement is removed. Finally, figure 14e shows the self-consistency test on the phase measurements.

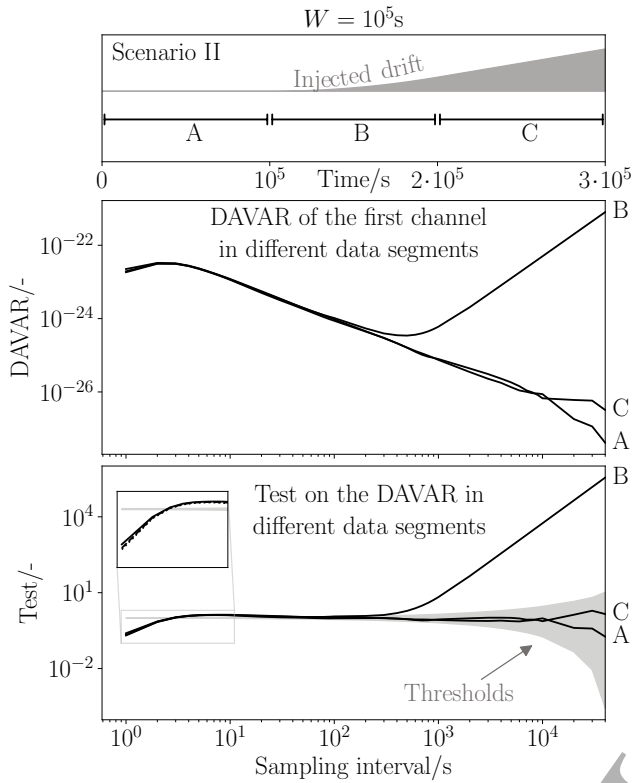


Figure 15: Scenario II - Measured DAVAR  $\hat{\sigma}_{k,\tau}^2$  of the faulty channel and corresponding test, computed on three data segments (A, B, and C) starting at different times. The upper plot shows the alignment of the data segments with the injected drift (shaded area): the first window is computed before the drift onset, the second during the part with quadratic drift, while the third in the part with linear drift. The central plot shows the DAVAR computed in the three segments. As expected, only the quadratic drift causes an increase in the observed values. The lower plot shows the test computed on the three data segments. The shaded area represents the interval between the lower and upper test thresholds, i.e. the region where  $\mathcal{H}_0^{(T)}$  is accepted.

triggerings, which are not related to the fault injection. The number of activation falls within the expected probability of false alarm.

**5.3.2. Test on the DAVAR:** in figure 21 we plot the DAVAR computed on data segments starting at three different time instants. The injection of the oscillation produces two “bumps” in the DAVAR for sampling intervals around the period length, seen in both the second and third part of the experiment. Figure 22 shows the w-test based on the gamma distribution applied to the four measurements, for increasing lengths of the sliding window. The test successfully detects the oscillation (area “D”), but only

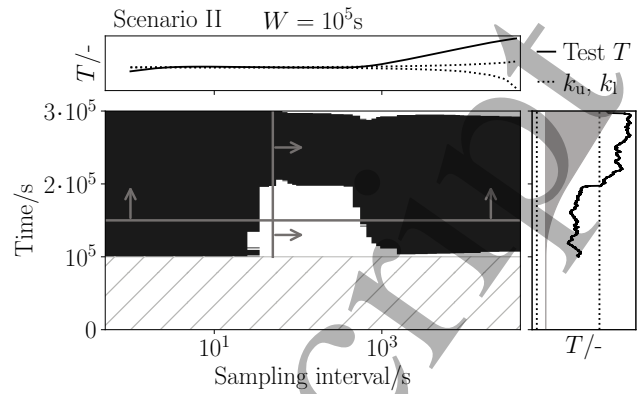


Figure 16: Result of applying the DAVAR test to the first channel in the second scenario, for a sliding window of length  $W = 10^5$  s. The black shading represents the points in time and sampling interval where the test is triggered, while the hatched area covers missing data. The upper and right plots show the projections at a selected time instant and at a selected sampling interval. The dotted lines show the test thresholds.

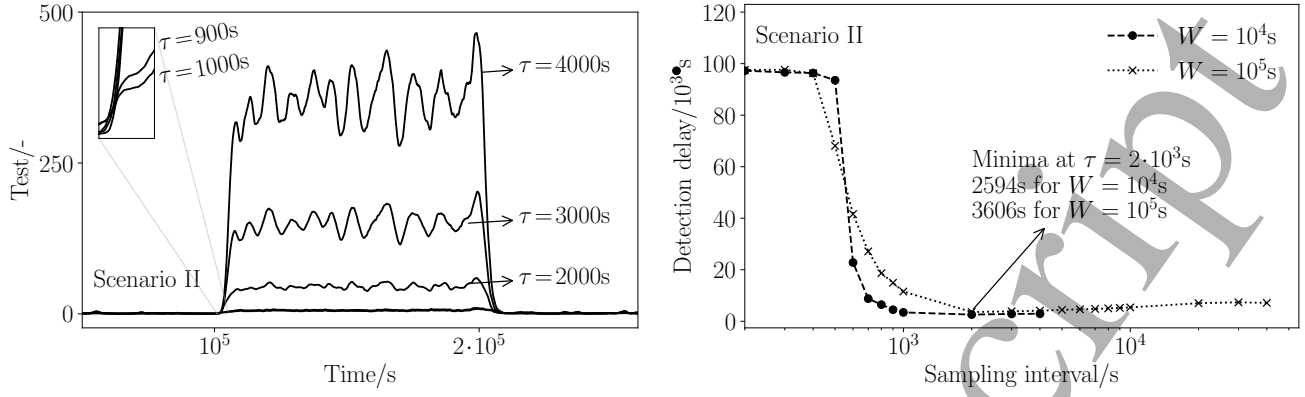
for longer windows, since for  $W = 10^3$  s we cannot use sampling intervals long enough to detect the bumps. Again, we can see the detection of the mismodelling at short values of  $\tau$  (area “M”).

## 6. Conclusions

In this paper we devised and tested FDI in clock ensembles based on the GLRT. The detectors observe three quantities: Kalman filter residuals, phase measurements, and DAVAR of phase measurements. We designed a model-based test, which performs fault detection with the overall model test and identification with the w-test. Furthermore, we introduced a self-consistency test to identify the faulty unit by observing the phase measurements without a-priori assuming a clock model.

The framework of the GLRT offers us a structured way to develop new detectors, and adapt them to the observation of further quantities, provided that they show similar statistics. Since in the GLRT we not only define a nominal hypothesis, but also a model for the faulty case, we can evaluate the probability of missed detection as function of the size of the fault. This is an advantage of these tests compared to other methods where only the nominal behavior is defined. Thanks to this, we can estimate the performance of a detector beforehand, given the integrity requirements of the application.

We performed three experiments to test the capability of the detectors in different scenarios. Considering the results in section 5.1, the test on the



(a) Scenario II - DAVAR w-test for  $W = 10^4$  s on the first measurement as function of time. The different lines show the test for different sampling intervals.

(b) Scenario II - Detection delay of the DAVAR w-test on the first measurement as function of the sampling interval, for different lengths of the sliding window  $W$ .

Figure 17: Scenario II - Analysis of the detection delay for the DAVAR w-test. We define the detection delay as the difference between the start of drift injection at  $t_k = 10^5$  s and the time of first detection.

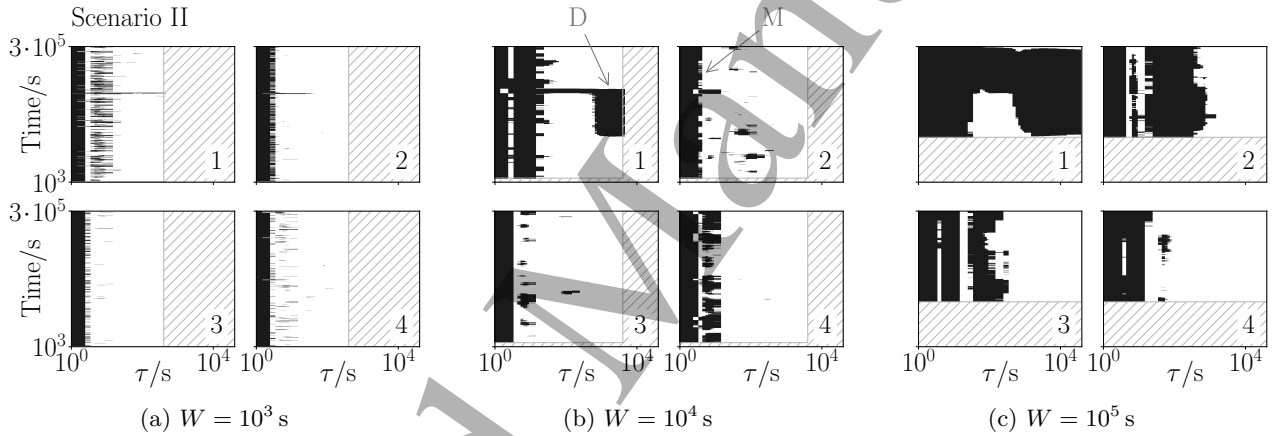


Figure 18: Scenario II - DAVAR w-test based on the gamma distribution, for different lengths of sliding window  $W$ , for the measurement channels 1 to 4. The black areas indicate fault detection, white means no detection. The hatched areas cover missing data: shorter windows provide values only at shorter sampling times, but longer windows can be evaluated only at later times. The area “D” shows drift detection, while the area “M” corresponds to triggering due to mismodelling.

Kalman filter residuals is the first choice for detecting faults such as phase steps, since this test detects without delay the occurrence of abrupt faults. Phase jumps can also be detected using the tests on the phase measurements, however the main disadvantage of this detector lies in the fact that the MDB increases in time with a square root law. Thus, the detection performance decreases in time, until the phase measurements are zeroed and the process is restarted. Nevertheless, as shown in section 5.2, the test on the phase measurements successfully detects and identifies the drift affecting the clock. Obviously, the detection occurs with a certain delay after the error onset; this requires further analyses to understand how the delay changes as function of the drift magnitude.

Both in the second and third scenarios we evaluate the test on the DAVAR, which successfully detects the injected drift and oscillation. However, the  $\chi^2$ -distribution of the DAVAR complicates the derivation of the GLRT. In this paper we derived a scalar version in section 4.3: although this test performs well in detecting drifts, it is applied to the single measurements and thus it does not take advantage of the structure of the clock ensemble. In future work we want to extend the statistical description of the DAVAR of the clock ensemble using a multivariate gamma distribution, and thus develop an exact overall model test on the DAVAR. The approximation to a Gaussian distribution allows us to develop an overall model test and a self-consistency

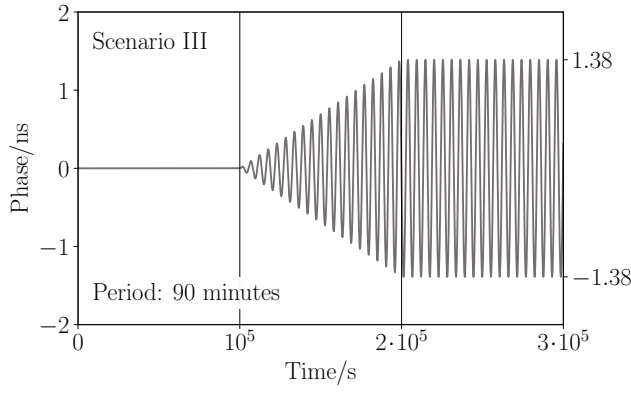
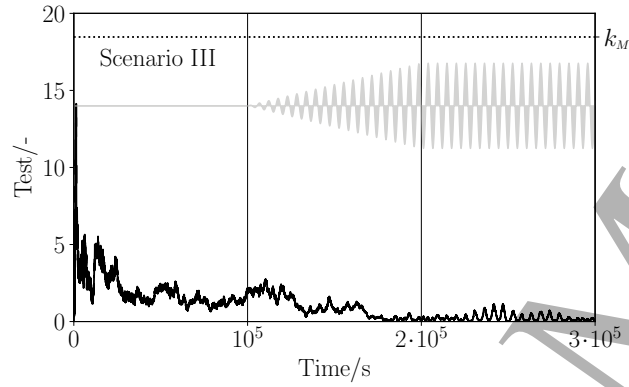
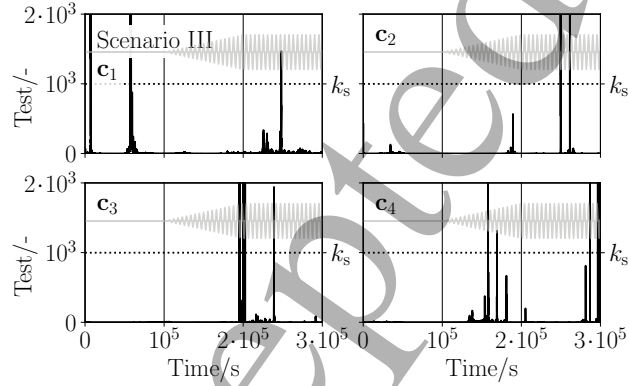


Figure 19: Scenario III - Injected fault phase difference  $\nabla(t)$ . We can see three parts: initially no fault is added, then the amplitude grows quadratically, finally reaching an oscillation of constant amplitude.



(a) Phase measurement residuals overall model test. The dotted line is the test threshold, the shaded curve shows the injected oscillation.



(b) Self-consistency test on the phase measurements. The dotted line is the test threshold, the shaded curve shows the injected oscillation.

Figure 20: Scenario III - tests on the phase measurements in the presence of an oscillation in phase. Figure 20a shows the overall model test, and figure 20b shows the self-consistency test on the phase measurements. As expected, neither test can successfully detect the injected oscillation.

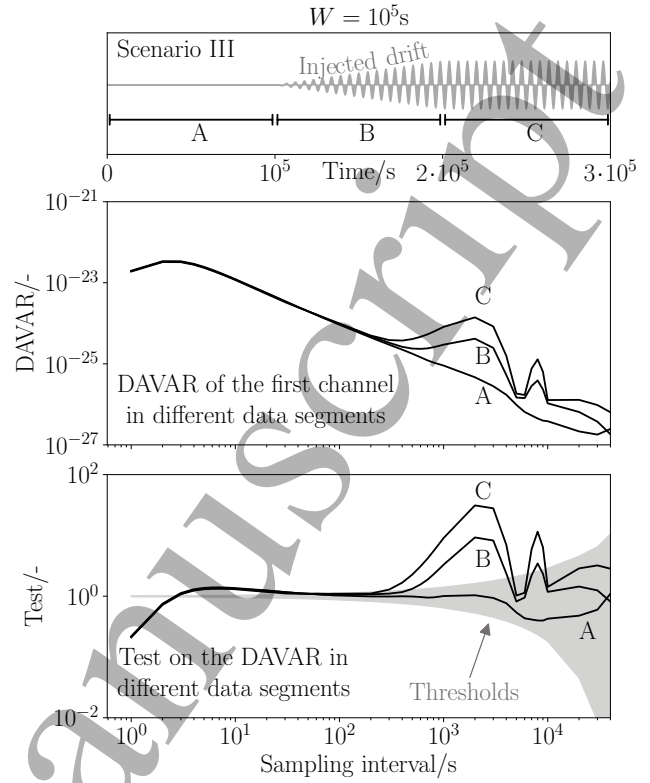


Figure 21: Scenario III - Measured DAVAR  $\hat{\sigma}^2$  of the faulty channel and corresponding test, computed on three data segments (A, B, and C) starting at different times. The upper plot shows the alignment of the segments with the injected drift (shaded area): the first window is computed before the drift onset, the second during the part with increasing amplitude, while the third in the part with constant amplitude. The central plot shows the DAVAR in the three batches, and the expected effects of the injected oscillation. The lower plot shows the test computed on the three data segments. The shaded area represents the interval between the lower and upper test thresholds.

test on the DAVAR. Although the approximation works well for short sampling intervals, this assumption degrades the test results at long term, which is exactly the region of interest when detecting drifts. Thus, longer sliding windows must be employed, which in turn delays the detection.

The detection of the mismatching in the DAVAR test at short sampling intervals leads us to an interesting discussion point. On the one hand, the model-based tests clearly require a precise clock model to provide efficient fault detection without triggering too many false alarms. On the other hand, we can consider this as an opportunity to warn the user that wrong clock models may be in use. Our examples showed exactly this case: since the DAVAR tests

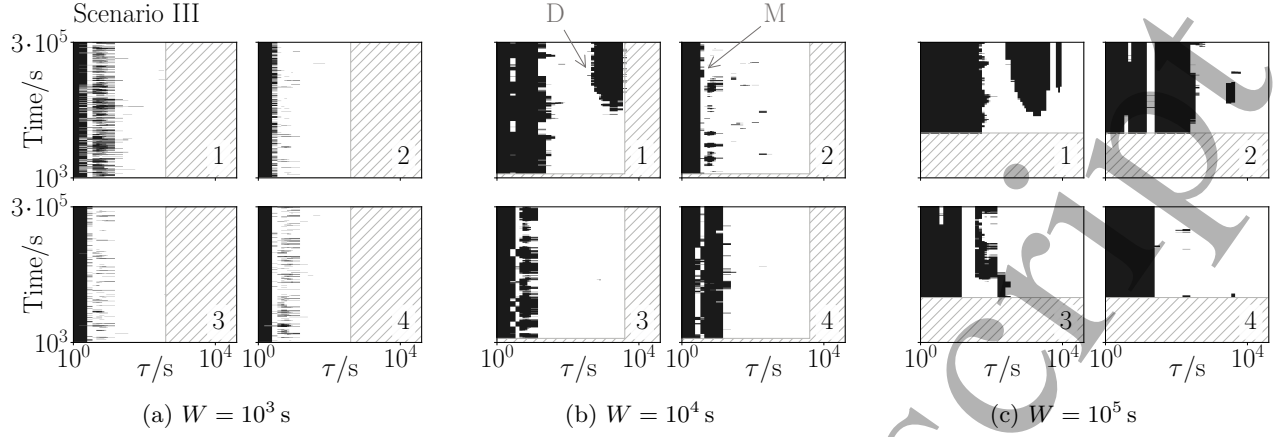


Figure 22: Scenario III - DAVAR w-test based on the gamma distribution, for different lengths of sliding window  $W$ , for the measurement channels 1 to 4. The black areas indicate fault detection, white means no detection. The hatched areas cover missing data. The area “D” shows detection of the injected oscillation, while the area “M” corresponds to triggering due to mismodelling.

trigger for short sampling intervals at all the times, on all the channels, and in different scenarios, we conclude that our clock model is imprecise in that region, which is confirmed by observing the difference between modelled and observed DAVAR in figure 13. Clearly, the detector can not autonomously discern between faulty behavior or wrong model, but an experienced user shall be able to determine which one is the case.

An help in this direction comes from the self-consistency test. Since the test compares different measurements between each other, this method is not influenced by mismodelling. Therefore, by comparing the results of w-test and self-consistency test, we are able to determine which triggerings are due to faulty clock behaviors and which ones are due to mismodelling or other unmodelled effects. These effects include for instance daily temperature fluctuations, orbital effects, and in general everything equally affecting all the measurements and not predicted by the clock model. However, the self-consistency test works only on homogeneous clock ensembles, containing only clocks of the same type. Furthermore, the estimation of the expected value and of the magnitude of the covariance matrix requires at least three measurements. Finally, the derivation of the self-consistency test for  $\chi^2$ -distributed observables is not straightforward and requires further analyses.

The different tests developed in this work complement each other, and could be bundled together in an operational scenario to form an ensemble of powerful fault detectors.

## Acknowledgments

This work was partially supported by the DLR Galileo

Competence Center in the framework of the project “RPTF - Robust Precise Time Facility”.

## Appendix A. Derivation of the GLRT for normally distributed observables

The ratio in (36) becomes

$$\text{Reject } \mathcal{H}_0 \text{ if } \frac{\max_{\nabla=0} f(\rho|\nabla)}{\max_{\nabla \neq 0} f(\rho|\nabla)} = \frac{f(\rho|0)}{f(\rho|\hat{\nabla})} < a, \quad (\text{A.1})$$

where the probability density function (PDF) is

$$f(\rho|\nabla) = (2\pi)^{-\frac{M}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2}(\rho - C\nabla)^\top \Omega^{-1}(\rho - C\nabla)}, \quad (\text{A.2})$$

and

$$\hat{\nabla} = \underset{\nabla \neq 0}{\operatorname{argmax}} f(\rho|\nabla). \quad (\text{A.3})$$

To find  $\hat{\nabla}$  we take the derivative of the logarithm of (A.2) and set it to zero. This operation does not influence the position of the maximum. Thus, since  $\Omega^{-1}$  is symmetric, it results

$$\begin{aligned} 0 &= -\frac{1}{2} \frac{\partial}{\partial \nabla} [(\rho - C\nabla)^\top \Omega^{-1}(\rho - C\nabla)]|_{\hat{\nabla}} \\ &= C^\top \Omega^{-1} (\rho - C\hat{\nabla}), \end{aligned} \quad (\text{A.4})$$

from which

$$\begin{aligned} \hat{\nabla} &= (C^\top \Omega^{-1} C)^{-1} C^\top \Omega^{-1} \rho \\ &= Q_{\hat{\nabla}\hat{\nabla}} C^\top \Omega^{-1} \rho, \end{aligned} \quad (\text{A.5})$$

where we defined the matrix  $Q_{\hat{\nabla}\hat{\nabla}} = (C^\top \Omega^{-1} C)^{-1}$ . By substituting (A.5) in (A.1), the ratio becomes

$$\exp\left(\frac{1}{2} \hat{\nabla}^\top Q_{\hat{\nabla}\hat{\nabla}}^{-1} \hat{\nabla} - \rho^\top \Omega^{-1} C\hat{\nabla}\right). \quad (\text{A.6})$$

From (A.5) we have

$$\mathbf{Q}_{\hat{\mathbf{V}}\hat{\mathbf{V}}}^{-1} \hat{\mathbf{V}} = \mathbf{C}^\top \mathbf{\Omega}^{-1} \boldsymbol{\rho} = (\boldsymbol{\rho}^\top \mathbf{\Omega}^{-1} \mathbf{C})^\top, \quad (\text{A.7})$$

which we can substitute in the second term of (A.6):

$$\boldsymbol{\rho}^\top \mathbf{\Omega}^{-1} \mathbf{C} \hat{\mathbf{V}} = \hat{\mathbf{V}}^\top \mathbf{Q}_{\hat{\mathbf{V}}\hat{\mathbf{V}}}^{-1} \hat{\mathbf{V}}. \quad (\text{A.8})$$

Thus, the test (A.1) becomes

$$\text{Reject } \mathcal{H}_0 \text{ if } \exp\left(-\frac{1}{2} \hat{\mathbf{V}}^\top \mathbf{Q}_{\hat{\mathbf{V}}\hat{\mathbf{V}}}^{-1} \hat{\mathbf{V}}\right) < a, \quad (\text{A.9})$$

which can be equivalently rewritten by taking the logarithm of both sides, leading to

$$\text{Reject } \mathcal{H}_0 \text{ if } T = \hat{\mathbf{V}}^\top \mathbf{Q}_{\hat{\mathbf{V}}\hat{\mathbf{V}}}^{-1} \hat{\mathbf{V}} > 2 \log(a) = k. \quad (\text{A.10})$$

## Appendix B. Derivation of the self-consistency test for normally distributed observables

The hypotheses (54) can be joined as

$$\mathbf{z} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\eta}, v^2 \boldsymbol{\Psi}), \quad (\text{B.1})$$

with

$$\begin{cases} \mathcal{H}_0^{(s)} : \mathbf{A} = \mathbf{u}, \boldsymbol{\eta} = \zeta \\ \mathcal{H}_A^{(s)} : \mathbf{A} = [\mathbf{u} \quad \mathbf{C}], \boldsymbol{\eta} = \begin{bmatrix} \zeta \\ \hat{\mathbf{V}} \end{bmatrix} \end{cases}, \quad (\text{B.2})$$

and PDF

$$f(\mathbf{z}|\boldsymbol{\eta}, v^2) = \frac{(2\pi v^2)^{-\frac{M}{2}}}{\sqrt{|\boldsymbol{\Psi}|}} e^{-\frac{1}{2v^2} (\mathbf{z} - \mathbf{A}\boldsymbol{\eta})^\top \boldsymbol{\Psi}^{-1} (\mathbf{z} - \mathbf{A}\boldsymbol{\eta})}. \quad (\text{B.3})$$

The maximum likelihood estimates are obtained by finding the maxima of the PDF:

$$\begin{aligned} \hat{\boldsymbol{\eta}} &= \arg\max_{\boldsymbol{\eta}} f(\mathbf{z}|\boldsymbol{\eta}, v^2) \\ &= (\mathbf{A}^\top \boldsymbol{\Psi}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \boldsymbol{\Psi}^{-1} \mathbf{z}, \end{aligned} \quad (\text{B.4})$$

and

$$\begin{aligned} \hat{v}^2 &= \arg\max_{v^2} p(\boldsymbol{\rho}|\boldsymbol{\eta}, v^2) \\ &= M^{-1} (\mathbf{z} - \mathbf{A}\hat{\boldsymbol{\eta}})^\top \boldsymbol{\Psi}^{-1} (\mathbf{z} - \mathbf{A}\hat{\boldsymbol{\eta}}) \\ &= M^{-1} \hat{\mathbf{e}}^\top \boldsymbol{\Psi}^{-1} \hat{\mathbf{e}}, \end{aligned} \quad (\text{B.5})$$

where we introduced the estimation error vector

$$\begin{aligned} \hat{\mathbf{e}} &= \mathbf{z} - \mathbf{A}\hat{\boldsymbol{\eta}} \\ &= [\mathbf{I}_M - \mathbf{A}(\mathbf{A}^\top \boldsymbol{\Psi}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \boldsymbol{\Psi}^{-1}] \mathbf{z} \\ &= (\mathbf{I}_M - \mathbf{P}_A) \mathbf{z}, \end{aligned} \quad (\text{B.6})$$

and  $\mathbf{P}_A = \mathbf{A}(\mathbf{A}^\top \boldsymbol{\Psi}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \boldsymbol{\Psi}^{-1}$  is the idempotent projection matrix associated with  $\mathbf{A}$ . Under the two hypotheses we have

$$\hat{\mathbf{e}}_0 = \mathbf{z} - \mathbf{u}\hat{\zeta}_0, \quad (\text{B.7})$$

$$\hat{\mathbf{e}}_A = \mathbf{z} - \mathbf{u}\hat{\zeta}_A - \mathbf{C}\hat{\mathbf{V}}. \quad (\text{B.8})$$

The estimate (B.4) is unbiased

$$\mathcal{E}(\boldsymbol{\eta}) = (\mathbf{A}^\top \boldsymbol{\Psi}^{-1} \mathbf{A})^{-1} \mathbf{A}^\top \boldsymbol{\Psi}^{-1} \mathcal{E}(\mathbf{z}) = \boldsymbol{\eta} \quad (\text{B.9})$$

since  $\mathcal{E}(\mathbf{z}) = \mathbf{A}\boldsymbol{\eta}$ , where  $\mathcal{E}(\bullet)$  is the expectation operator. However, (B.5) is biased:

$$\begin{aligned} \mathcal{E}(v^2) &= M^{-1} \mathcal{E}(\hat{\mathbf{e}}^\top \boldsymbol{\Psi}^{-1} \hat{\mathbf{e}}) \\ &= M^{-1} \text{tr}[\mathcal{E}(\hat{\mathbf{e}}\hat{\mathbf{e}}^\top) \boldsymbol{\Psi}^{-1}]. \end{aligned} \quad (\text{B.10})$$

We observe that the covariance of  $\hat{\mathbf{e}}$  is

$$\mathbf{Q}_{\hat{\mathbf{e}}\hat{\mathbf{e}}} = \mathcal{E}(\hat{\mathbf{e}}\hat{\mathbf{e}}^\top) - \mathcal{E}(\hat{\mathbf{e}}) \mathcal{E}(\hat{\mathbf{e}})^\top = \mathcal{E}(\hat{\mathbf{e}}\hat{\mathbf{e}}^\top). \quad (\text{B.11})$$

since

$$\mathcal{E}(\hat{\mathbf{e}}) = \mathcal{E}(\mathbf{z}) - \mathbf{A}\mathcal{E}(\boldsymbol{\eta}) = \mathbf{0}. \quad (\text{B.12})$$

By introducing (B.6) and exploiting the properties of idempotent matrices we obtain

$$\begin{aligned} \mathbf{Q}_{\hat{\mathbf{e}}\hat{\mathbf{e}}} &= (\mathbf{I}_M - \mathbf{P}_A) \mathcal{E}(\mathbf{z}\mathbf{z}^\top) (\mathbf{I}_M - \mathbf{P}_A)^\top \\ &= v^2 (\mathbf{I}_M - \mathbf{P}_A) \boldsymbol{\Psi}. \end{aligned} \quad (\text{B.13})$$

Substitution of (B.11) and (B.13) in (B.10) yields

$$\begin{aligned} \mathcal{E}(v^2) &= v^2 M^{-1} \text{tr}(\mathbf{I}_M - \mathbf{P}_A) \\ &= v^2 M^{-1} (M - n), \end{aligned} \quad (\text{B.14})$$

where  $n$  is the number of column in  $\mathbf{A}$ :  $n = 1$  under the null hypothesis and  $n = q + 1$  under the alternative hypothesis. The estimate (B.14) is biased, so we can use a corrected estimator:

$$\hat{v}^2 = \hat{v}^2 \frac{M}{M - n} \quad (\text{B.15})$$

The estimates (B.4) and (B.5) can be now rewritten for the null hypothesis  $\mathcal{H}_0$

$$\begin{aligned} \hat{\zeta}_0 &= (\mathbf{u}^\top \boldsymbol{\Psi}^{-1} \mathbf{u})^{-1} \mathbf{u}^\top \boldsymbol{\Psi}^{-1} \mathbf{z} \\ \hat{v}_0^2 &= (M - 1)^{-1} \hat{\mathbf{e}}_0^\top \boldsymbol{\Psi}^{-1} \hat{\mathbf{e}}_0, \end{aligned} \quad (\text{B.16})$$

and for the alternative hypothesis  $\mathcal{H}_A$

$$\begin{aligned} \begin{bmatrix} \hat{\zeta}_A \\ \hat{\mathbf{V}} \end{bmatrix} &= \begin{bmatrix} \mathbf{u}^\top \boldsymbol{\Psi}^{-1} \mathbf{u} & \mathbf{u}^\top \boldsymbol{\Psi}^{-1} \mathbf{C} \\ \mathbf{C}^\top \boldsymbol{\Psi}^{-1} \mathbf{u} & \mathbf{C}^\top \boldsymbol{\Psi}^{-1} \mathbf{C} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u}^\top \boldsymbol{\Psi}^{-1} \\ \mathbf{C}^\top \boldsymbol{\Psi}^{-1} \end{bmatrix} \mathbf{z} \\ \hat{v}_A^2 &= (M - q - 1)^{-1} \hat{\mathbf{e}}_A^\top \boldsymbol{\Psi}^{-1} \hat{\mathbf{e}}_A \end{aligned} \quad (\text{B.17})$$

Now that the the maximum likelihood estimates are available, we can build the test. The GLRT becomes

$$\begin{aligned} \frac{f(\mathbf{z}|\hat{\zeta}_0, \hat{v}_0^2)}{f(\mathbf{z}|\hat{\zeta}_A, \hat{\mathbf{V}}, \hat{v}_A^2)} &= \frac{\hat{v}_A^M \exp\left(-\frac{1}{2\hat{v}_0^2} \hat{\mathbf{e}}_0^\top \boldsymbol{\Psi}^{-1} \hat{\mathbf{e}}_0\right)}{\hat{v}_0^M \exp\left(-\frac{1}{2\hat{v}_A^2} \hat{\mathbf{e}}_A^\top \boldsymbol{\Psi}^{-1} \hat{\mathbf{e}}_A\right)} \\ &= \left(\frac{\hat{v}_A^2}{\hat{v}_0^2}\right)^{\frac{M}{2}} e^{-\frac{q}{2}}. \end{aligned} \quad (\text{B.18})$$

The test becomes

$$\text{Reject } \mathcal{H}_0^{(s)} \text{ if } \frac{\hat{v}_0^2}{\hat{v}_A^2} > \left(\alpha e^{\frac{q}{2}}\right)^{-\frac{2}{M}} = \bar{\alpha}, \quad (\text{B.19})$$

or equivalently:

$$\text{Reject } \mathcal{H}_0^{(s)} \text{ if } \frac{M - q - 1}{M - 1} \frac{\hat{\mathbf{e}}_0^\top \boldsymbol{\Psi}^{-1} \hat{\mathbf{e}}_0}{\hat{\mathbf{e}}_A^\top \boldsymbol{\Psi}^{-1} \hat{\mathbf{e}}_A} > \bar{\alpha}. \quad (\text{B.20})$$

Please note that the term  $M - q - 1$  is always positive since the maximum value of  $q$  is  $M - 2$ . We



want now to rewrite the test as function of  $\hat{\mathbf{V}}$ . By multiplying (B.17) by a matrix  $\mathbf{B}$ ,

$$\mathbf{B} \begin{bmatrix} \mathbf{u}^\top \Psi^{-1} \mathbf{u} & \mathbf{u}^\top \Psi^{-1} \mathbf{C} \\ \mathbf{C}^\top \Psi^{-1} \mathbf{u} & \mathbf{C}^\top \Psi^{-1} \mathbf{C} \end{bmatrix} \begin{bmatrix} \hat{\zeta}_A \\ \hat{\mathbf{V}} \end{bmatrix} = \mathbf{B} \begin{bmatrix} \mathbf{u}^\top \Psi^{-1} \\ \mathbf{C}^\top \Psi^{-1} \end{bmatrix} \mathbf{z}, \quad (\text{B.21})$$

with

$$\mathbf{B} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}^\top \Psi^{-1} \mathbf{u} (\mathbf{u}^\top \Psi^{-1} \mathbf{u})^{-1} & \mathbf{I} \end{bmatrix}, \quad (\text{B.22})$$

we obtain

$$\begin{bmatrix} \mathbf{u}^\top \Psi^{-1} \mathbf{u} & \mathbf{u}^\top \Psi^{-1} \mathbf{C} \\ \mathbf{0} & \mathbf{C}^\top \Psi^{-1} \mathbf{P}_u^\perp \mathbf{C} \end{bmatrix} \begin{bmatrix} \hat{\zeta}_A \\ \hat{\mathbf{V}} \end{bmatrix} = \begin{bmatrix} \mathbf{u}^\top \Psi^{-1} \\ \mathbf{C}^\top \Psi^{-1} \mathbf{P}_u^\perp \end{bmatrix} \mathbf{z}, \quad (\text{B.23})$$

where

$$\mathbf{P}_u^\perp = \mathbf{I} - \mathbf{P}_u = \mathbf{I} - \mathbf{u} (\mathbf{u}^\top \Psi^{-1} \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1}. \quad (\text{B.24})$$

From the second equation in (B.23) we obtain an estimate of the fault magnitude

$$\begin{aligned} \hat{\mathbf{V}} &= (\mathbf{C}^\top \Psi^{-1} \mathbf{P}_u^\perp \mathbf{C})^{-1} \mathbf{C}^\top \Psi^{-1} \mathbf{P}_u^\perp \mathbf{z} \\ &= \mathbf{R}_{\hat{\mathbf{V}}\hat{\mathbf{V}}} \mathbf{C}^\top \Psi^{-1} \mathbf{P}_u^\perp \mathbf{z}, \end{aligned} \quad (\text{B.25})$$

where we defined the matrix

$$\mathbf{R}_{\hat{\mathbf{V}}\hat{\mathbf{V}}} = (\mathbf{C}^\top \Psi^{-1} \mathbf{P}_u^\perp \mathbf{C})^{-1}. \quad (\text{B.26})$$

We can rewrite the estimation error vectors under the two hypotheses

$$\begin{cases} \mathcal{H}_0^{(s)} : \hat{\mathbf{e}}_0 = \mathbf{P}_u^\perp \mathbf{z} \\ \mathcal{H}_A^{(s)} : \hat{\mathbf{e}}_A = \mathbf{P}_C^\perp \mathbf{P}_u^\perp \mathbf{z} \end{cases}, \quad (\text{B.27})$$

where we used following idempotent matrices

$$\mathbf{P}_C^\perp = \mathbf{I} - \mathbf{C} (\mathbf{C}^\top \Psi^{-1} \mathbf{C})^{-1} \mathbf{C}^\top \Psi^{-1}, \quad (\text{B.28})$$

$$\mathbf{P}_u^\perp = \mathbf{I} - \mathbf{u} (\mathbf{u}^\top \Psi^{-1} \mathbf{P}_C^\perp \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1} \mathbf{P}_C^\perp. \quad (\text{B.29})$$

From (B.16) we have

$$(\mathbf{u}^\top \Psi^{-1} \mathbf{u}) \hat{\zeta}_0 = \mathbf{u}^\top \Psi^{-1} \mathbf{z}, \quad (\text{B.30})$$

which can be inserted in the first equation of (B.23):

$$\hat{\zeta}_A - \hat{\zeta}_0 = -(\mathbf{u}^\top \Psi^{-1} \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1} \mathbf{C} \hat{\mathbf{V}}. \quad (\text{B.31})$$

With this, we can express the difference of the residuals:

$$\begin{aligned} \hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A &= \mathbf{u} (\hat{\zeta}_A - \hat{\zeta}_0) + \mathbf{C} \hat{\mathbf{V}} \\ &= [\mathbf{I} - \mathbf{u} (\mathbf{u}^\top \Psi^{-1} \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1}] \mathbf{C} \hat{\mathbf{V}} \\ &= \mathbf{P}_u^\perp \mathbf{C} \hat{\mathbf{V}}. \end{aligned} \quad (\text{B.32})$$

The ratio in (B.20) can be reworked as

$$\frac{\hat{\mathbf{e}}_0^\top \Psi^{-1} \hat{\mathbf{e}}_0}{\hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} = \frac{\hat{\mathbf{e}}_0^\top \Psi^{-1} \hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A}{\hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} + 1. \quad (\text{B.33})$$

The numerator is equal to

$$(\hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A)^\top \Psi^{-1} (\hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A) + 2 (\hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A)^\top \Psi^{-1} \hat{\mathbf{e}}_A, \quad (\text{B.34})$$

but the term  $(\hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A)^\top \Psi^{-1} \hat{\mathbf{e}}_A$  is zero. In fact, by using the property of symmetry and idempotence of  $\mathbf{P}_u^\perp$  and  $\mathbf{P}_C^\perp \mathbf{P}_u^\perp$  we have

$$\begin{aligned} (\hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A)^\top \Psi^{-1} \hat{\mathbf{e}}_A &= \mathbf{z}^\top (\mathbf{P}_u^\perp - \mathbf{P}_C^\perp \mathbf{P}_u^\perp) \Psi^{-1} \mathbf{P}_C^\perp \mathbf{P}_u^\perp \mathbf{z} \\ &= \mathbf{z}^\top (\mathbf{P}_u^\perp - \mathbf{I}) \mathbf{P}_C^\perp \mathbf{P}_u^\perp \Psi^{-1} \mathbf{z} \\ &= \mathbf{z}^\top \mathbf{P}_C^\perp \mathbf{P}_u^\perp \mathbf{P}_u \Psi^{-1} \mathbf{z}. \end{aligned} \quad (\text{B.35})$$

We can show that the product  $\mathbf{P}_u^\perp \mathbf{P}_u$  is zero

$$\begin{aligned} \mathbf{P}_u^\perp \mathbf{P}_u &= (\mathbf{I} - \mathbf{u} (\mathbf{u}^\top \Psi^{-1} \mathbf{P}_C^\perp \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1} \mathbf{P}_C^\perp) \mathbf{u} \\ &\quad \cdot (\mathbf{u}^\top \Psi^{-1} \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1} \\ &= (\mathbf{u} - \mathbf{u} (\mathbf{u}^\top \Psi^{-1} \mathbf{P}_C^\perp \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1} \mathbf{P}_C^\perp \mathbf{u}) \\ &\quad \cdot (\mathbf{u}^\top \Psi^{-1} \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1} \\ &= (\mathbf{u} - \mathbf{u}) (\mathbf{u}^\top \Psi^{-1} \mathbf{u})^{-1} \mathbf{u}^\top \Psi^{-1} \\ &= \mathbf{0}. \end{aligned} \quad (\text{B.36})$$

Thus, the ratio becomes

$$\frac{\hat{\mathbf{e}}_0^\top \Psi^{-1} \hat{\mathbf{e}}_0}{\hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} = \frac{(\hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A)^\top \Psi^{-1} (\hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_A)}{\hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} + 1. \quad (\text{B.37})$$

By inserting (B.32) we obtain

$$\begin{aligned} \frac{\hat{\mathbf{e}}_0^\top \Psi^{-1} \hat{\mathbf{e}}_0}{\hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} &= \frac{\hat{\mathbf{V}}^\top \mathbf{C}^\top \Psi^{-1} \mathbf{P}_u^\perp \mathbf{C} \hat{\mathbf{V}}}{\hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} + 1 \\ &= \frac{\hat{\mathbf{V}}^\top \mathbf{R}_{\hat{\mathbf{V}}\hat{\mathbf{V}}}^{-1} \hat{\mathbf{V}}}{\hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} + 1. \end{aligned} \quad (\text{B.38})$$

Thus the test (B.20) can be rewritten as

$$\text{Reject } \mathcal{H}_0^{(s)} \text{ if } \frac{\hat{\mathbf{V}}^\top \mathbf{R}_{\hat{\mathbf{V}}\hat{\mathbf{V}}}^{-1} \hat{\mathbf{V}}}{\hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} > \beta, \quad (\text{B.39})$$

with

$$\beta = \frac{M-1}{M-q-1} \bar{\alpha} - 1. \quad (\text{B.40})$$

The numerator in (B.39) distributes according to a  $\chi^2$ -distribution with  $q$  degrees of freedom, central under  $\mathcal{H}_0^{(s)}$  and non-central under  $\mathcal{H}_A^{(s)}$ . The non-centrality parameter is  $\lambda = \mathbf{V}^\top \mathbf{R}_{\hat{\mathbf{V}}\hat{\mathbf{V}}}^{-1} \mathbf{V}$ . The denominator is also  $\chi^2$ -distributed, central under both hypotheses and with  $M-q-1$  degrees of freedom.

If (B.39) is rewritten as

$$\text{Reject } \mathcal{H}_0^{(s)} \text{ if } \frac{q^{-1} \hat{\mathbf{V}}^\top \mathbf{R}_{\hat{\mathbf{V}}\hat{\mathbf{V}}}^{-1} \hat{\mathbf{V}}}{(M-q-1)^{-1} \hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A} > k_s, \quad (\text{B.41})$$

with

$$k_s = \frac{M-q-1}{q} \beta, \quad (\text{B.42})$$

we can express the distribution of the ratio (B.41) with a  $F$ -distribution, which corresponds to the ratio of two independent  $\chi^2$ -distributed variables:

$$F(n_1, n_2, \lambda) = \frac{n_1^{-1} \cdot \chi^2(n_1, \lambda)}{n_2^{-1} \cdot \chi^2(n_2, 0)}. \quad (\text{B.43})$$

From (B.17) we note that the denominator in (B.41) equals to  $\hat{v}_A^2$ , so that the self-consistency test finally becomes

$$\text{Reject } \mathcal{H}_0^{(s)} \text{ if } T = \frac{\hat{\nabla}^\top \mathbf{R}_{\hat{\nabla}}^{-1} \hat{\nabla}}{q \cdot \hat{v}_A^2} > k_s, \quad (\text{B.44})$$

which distributes according to

$$T \sim F(q, M - q - 1, \lambda). \quad (\text{B.45})$$

The terms in (B.44) can be computed with (B.25) and

$$\begin{aligned} \hat{v}_A^2 &= (M - q - 1)^{-1} \hat{\mathbf{e}}_A^\top \Psi^{-1} \hat{\mathbf{e}}_A \\ &= (M - q - 1)^{-1} \mathbf{z}^\top \Psi^{-1} \mathbf{P}_C^\perp \mathbf{P}_u^\perp \mathbf{z}. \end{aligned} \quad (\text{B.46})$$

### Appendix C. Derivation of the w-test for gamma-distributed observables

To obtain an expression for the w-test for gamma-distributed observables in section 4.3, we need to find the maxima in the GLRT:

$$\text{Reject } \mathcal{H}_0^{(r)} \text{ if } \bar{T} = \frac{\max_{\nabla=0} f(\xi|\nabla)}{\max_{\nabla \neq 0} f(\xi|\nabla)} < a, \quad (\text{C.1})$$

where

$$\begin{aligned} f(\xi|\nabla) &= \frac{1}{2} \left( \frac{\hat{\sigma}^2 \bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} \right)^{\frac{p}{2}-1} \frac{1}{\Gamma(\frac{p}{2})} \left( \frac{\bar{\sigma}^2}{\bar{\sigma}^2 + \nabla} \right)^{\frac{p}{2}} \\ &\quad \cdot \exp \left( -\frac{\hat{\sigma}^2 \bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} \frac{\bar{\sigma}^2}{\bar{\sigma}^2 + \nabla} \right). \end{aligned} \quad (\text{C.2})$$

The numerator is trivial:

$$\begin{aligned} \max_{\nabla=0} f(\xi|\nabla) &= f(\xi|0) \\ &= \frac{1}{2} \left( \frac{\hat{\sigma}^2 \bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} \right)^{\frac{p}{2}-1} \frac{1}{\Gamma(\frac{p}{2})} \exp \left( -\frac{\hat{\sigma}^2 \bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} \right). \end{aligned} \quad (\text{C.3})$$

For the denominator we have

$$\max_{\nabla \neq 0} f(\xi|\nabla) = f(\xi|\hat{\nabla}), \quad (\text{C.4})$$

with

$$\hat{\nabla} = \underset{\nabla}{\text{argmax}} f(\xi|\nabla). \quad (\text{C.5})$$

To find  $\hat{\nabla}$  we take the logarithm of the PDF (C.2), since this does not influence the location of the maximum:

$$\begin{aligned} L_\Gamma(\xi|\nabla) &= \log[f(\xi|\nabla)] \\ &= \left( \frac{p}{2} - 1 \right) \log \left( \frac{\hat{\sigma}^2 \bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} \right) - \log \left[ \Gamma \left( \frac{p}{2} \right) \right] + \\ &\quad + \log \left( \frac{\bar{\sigma}^2}{\bar{\sigma}^2 + \nabla} \right) \frac{p}{2} - \frac{\hat{\sigma}^2 \bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} \frac{\bar{\sigma}^2}{\bar{\sigma}^2 + \nabla}. \end{aligned} \quad (\text{C.6})$$

By taking the derivative we obtain

$$\frac{\partial}{\partial \nabla} L_\Gamma = -\frac{p}{2} \frac{1}{\bar{\sigma}^2 + \nabla} + \frac{\hat{\sigma}^2 \bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} \frac{1}{(\bar{\sigma}^2 + \nabla)^2}. \quad (\text{C.7})$$

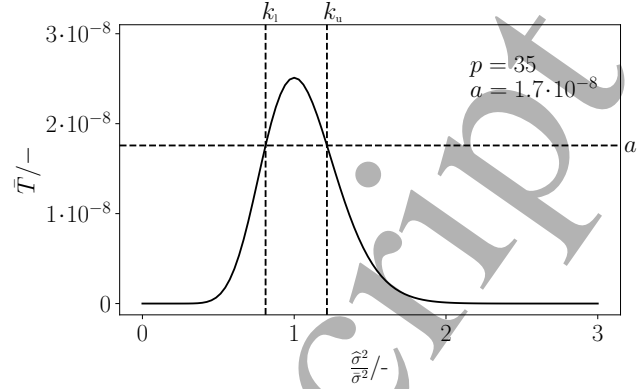


Figure C1: Qualitative example of the function  $\bar{T} \left( \frac{\hat{\sigma}^2}{\bar{\sigma}^2} \right)$ , for some arbitrary values of degrees of freedom  $p$  and threshold  $a$ . Setting  $\bar{T} < a$  is equivalent to  $\frac{\hat{\sigma}^2}{\bar{\sigma}^2} < k_l \vee \frac{\hat{\sigma}^2}{\bar{\sigma}^2} > k_u$ .

If we impose  $\frac{\partial}{\partial \nabla} L_\Gamma|_{\hat{\nabla}} = 0$ , we can compute  $\hat{\nabla}$ :

$$\hat{\nabla} = \frac{2}{p} \frac{\hat{\sigma}^2 \bar{\sigma}^4}{\text{Var}(\hat{\sigma}^2)} = \hat{\sigma}^2 - \bar{\sigma}^2, \quad (\text{C.8})$$

where we substituted the expression for the degrees of freedom

$$p = \frac{2\bar{\sigma}^4}{\text{Var}(\hat{\sigma}^2)}. \quad (\text{C.9})$$

This point is indeed a maximum, since the second derivative computed for  $\nabla = \hat{\nabla}$  is negative:

$$\frac{\partial^2}{\partial \nabla^2} L_\Gamma|_{\hat{\nabla}} = -\frac{\bar{\sigma}^4}{\hat{\sigma}^4 \text{Var}(\hat{\sigma}^2)} < 0. \quad (\text{C.10})$$

Thus, the second maximum becomes

$$\begin{aligned} f(\xi|\hat{\nabla}) &= \frac{1}{2\Gamma(\frac{p}{2})} \left( \frac{\hat{\sigma}^2 \bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} \right)^{\frac{p}{2}-1} \\ &\quad \cdot \exp \left( -\frac{\bar{\sigma}^4}{\text{Var}(\hat{\sigma}^2)} \right), \end{aligned} \quad (\text{C.11})$$

so that the ratio (C.1) yields

$$\begin{aligned} \bar{T} &= \left( \frac{\hat{\sigma}^2}{\bar{\sigma}^2} \right)^{\frac{p}{2}} \exp \left[ -\frac{\bar{\sigma}^2}{\text{Var}(\hat{\sigma}^2)} (\hat{\sigma}^2 - \bar{\sigma}^2) \right] \\ &= \left( \frac{\hat{\sigma}^2}{\bar{\sigma}^2} \right)^{\frac{p}{2}} \exp \left[ -\frac{p}{2} \left( \frac{\hat{\sigma}^2}{\bar{\sigma}^2} - 1 \right) \right]. \end{aligned} \quad (\text{C.12})$$

The test  $\bar{T}$  is a bell shaped function of the ratio  $\frac{\hat{\sigma}^2}{\bar{\sigma}^2}$ , as shown in figure C1, where we plot the function (C.12) for some arbitrary values of degrees of freedom  $p$  and threshold  $a$ . We observe that

$$\bar{T} < a \Leftrightarrow \frac{\hat{\sigma}^2}{\bar{\sigma}^2} < k_l \vee \frac{\hat{\sigma}^2}{\bar{\sigma}^2} > k_u, \quad (\text{C.13})$$

thus the test (C.1) can be equivalently rewritten as

$$\text{Reject } \mathcal{H}_0^{(r)} \text{ if } T < k_l \text{ or } T > k_u, \quad (\text{C.14})$$

where now the test value is  $T = \frac{\hat{\sigma}^2}{\bar{\sigma}^2}$ .



The distribution of the test (C.14) can be derived from (31):

$$\xi = p \frac{\hat{\sigma}^2}{\bar{\sigma}^2} \sim \Gamma\left(\frac{p}{2}, 2 \frac{\bar{\sigma}^2 + \nabla}{\bar{\sigma}^2}\right). \quad (\text{C.15})$$

Thus by applying the scaling property of the gamma distribution we finally obtain

$$T = \frac{1}{p} \left( p \frac{\hat{\sigma}^2}{\bar{\sigma}^2} \right) \sim \Gamma\left(\frac{p}{2}, \frac{2}{p} \frac{\bar{\sigma}^2 + \nabla}{\bar{\sigma}^2}\right). \quad (\text{C.16})$$

## References

- [1] Teunissen P J G 2009 *Testing Theory: an introduction* Mathematical Geodesy and Positioning (Delft Academic Press, VSSD) ISBN 978-9040719752 URL <http://www.vssd.nl/hlf/a030.htm>
- [2] Teunissen P J G 2017 *Batch and Recursive Model Validation* (Cham: Springer International Publishing) pp 687–720 ISBN 978-3-319-42928-1 URL [https://doi.org/10.1007/978-3-319-42928-1\\_24](https://doi.org/10.1007/978-3-319-42928-1_24)
- [3] Jones R H and Tryon P V 1983 *Journal of Research of the National Bureau of Standards* **88** 17–24
- [4] Galleani L and Tavella P 2011 Using the Kalman filter to detect frequency jumps in atomic clocks *2011 Joint Conference of the IEEE International Frequency Control and the European Frequency and Time Forum (FCS) Proceedings* (IEEE) pp 1–3
- [5] Huang X, Gong H and Ou G 2014 *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **61** 747–755
- [6] Khare A, Arora R, Banik A and Mehta S D 2016 *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **63** 326–335
- [7] Nunzi E and Carbone P 2007 *Proc. SPIE* **6603** 237–245
- [8] Nunzi E and Carbone P 2008 *Metrologia* **45** S103–S107
- [9] Nunzi E, Carbone P and Tavella P 2008 Fault detection in atomic clock frequency standards affected by mean and variance changes and by an additive periodic component: the GLRT approach *2008 IEEE Instrumentation and Measurement Technology Conference* (IEEE) pp 1594–1597
- [10] Nunzi E and Saltanocchi G 2009 A simplified detector for atomic clock anomalies detection *2009 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems* (IEEE) pp 15–18
- [11] Nunzi E and Saltanocchi G 2010 Real-time detection of anomalies for atomic clock in space by means of the GLRT *EFTF-2010 24th European Frequency and Time Forum* (IEEE) pp 1–6
- [12] Galleani L and Tavella P 2003 The characterization of clock behavior with the dynamic Allan variance *Proceedings of the 2003 IEEE International Frequency Control Symposium and PDA Exhibition Jointly with the 17th European Frequency and Time Forum, 2003* pp 239–244
- [13] Galleani L and Tavella P 2009 *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* **56** 450–464
- [14] Galleani L and Tavella P 2016 *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **63** 624–635
- [15] Sesia I, Galleani L and Tavella P 2011 *IEEE Transactions on Aerospace and Electronic Systems* **47** 884–895
- [16] Rodríguez-Pérez I, García-Serrano C, Catalán C C, García A M, Tavella P, Galleani L and Amarillo F 2011 *Advances in Space Research* **47** 197–212 ISSN 0273-1177 scientific applications of Galileo and other Global Navigation Satellite Systems - I URL <https://www.sciencedirect.com/science/article/pii/S0273117710004916>
- [17] Riley W J 2008 *Metrologia* **45** S154–S161 URL <https://doi.org/10.1088/0026-1394/45/6/s21>
- [18] Galleani L and Tavella P 2010 An Algorithm for the Detection of the Frequency Jumps in Space Clocks *Proceedings of the 42nd Annual Precise Time and Time Interval Systems and Applications Meeting* (Reston, Virginia) pp 503–508
- [19] Galleani L and Tavella P 2017 *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **64** 475–485
- [20] Formichella V and Tavella P 2018 A recursive clock anomalies detector with double exponential smoothing *2018 European Frequency and Time Forum (EFTF)* (IEEE) pp 245–248
- [21] Zucca C, Tavella P and Peskir G 2016 *Metrologia* **53** S89–S95 URL <https://doi.org/10.1088/0026-1394/53/3/s89>
- [22] Chan Y C, Johnson W A, Karuza S K, Young A M and Camparo J C 2010 *IEEE Transactions on Instrumentation and Measurement* **59** 330–334
- [23] Süß M 2014 *Robust Composite Clock Using Modified Kalman Filtering* Dissertation University of Stuttgart URL <https://elib.dlr.de/122015/>
- [24] Gödel M and Furrner J 2017 Robust Ensemble Time On-board a Satellite *Proceedings of the 48th Annual Precise Time and Time Interval Systems and Applications Meeting ION* (Monterey, California: ION) pp 26–43 30. Jan. – 1. Feb. 2017, Monterey, US. URL <https://doi.org/10.33012/2017.15007>
- [25] Zenzinger A, Bartusch T, Kuehl C, Fischer S and Shrestha A 2012 Failure detection and correction for clock ensemble in space *2012 6th ESA Workshop on Satellite Navigation Technologies (Navitec 2012) & European Workshop on GNSS Signals and Signal Processing* (IEEE) pp 1–8
- [26] Wang Q and Rochat P 2009 An Anomaly Clock Detection Algorithm for a Robust Clock Ensemble *Proceedings of the 41st Annual Precise Time and Time Interval Systems and Applications Meeting* pp 121–130
- [27] Wang Q, Droz F and Rochat P 2015 Robust clock ensemble for time and frequency reference system *2015 Joint Conference of the IEEE International Frequency Control Symposium & the European Frequency and Time Forum* (IEEE) pp 374–378
- [28] Wang Q and Rochat P 2019 Algorithms Development and Verification for Next Generation On-board Clock Monitoring and Control Unit *2019 Joint Conference of the IEEE International Frequency Control Symposium and European Frequency and Time Forum (EFTF/IFC)* (IEEE) pp 1–5
- [29] Vidarsson L, Pullen S, Green G and Enge P 2001 Satellite autonomous integrity monitoring and its role in enhancing GPS user performance *Proceedings of the 14th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2001)* (Salt Lake City, Utah) pp 690–702
- [30] Trainotti C and Giorgi G 2020 Detection and identification of phase and frequency drifts in clock ensembles *Proceedings of the 51st Annual Precise Time and Time Interval Systems and Applications Meeting* (San Diego, CA, USA) URL <https://www.ion.org/ptti/abstracts.cfm?paperID=8062>
- [31] Trainotti C, Giorgi G and Furrner J 2019 Detection and identification of faults in clock ensembles *Proceedings of the 32nd International Technical Meeting of the Satellite Division of the Institute of Navigation 2019* pp 3342–3358 URL <https://www.ion.org/publications/abstract.cfm?articleID=17104>
- [32] Beard R and Senior K 2017 *Clocks* (Cham: Springer International Publishing) pp 121–164 ISBN 978-3-319-42928-1

URL [https://doi.org/10.1007/978-3-319-42928-1\\_5](https://doi.org/10.1007/978-3-319-42928-1_5)

[33] Zucca C and Tavella P 2005 *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **52** 289–296 URL <http://ieeexplore.ieee.org/document/1406554/>

[34] Davis J A, Greenhall C A and Stacey P W 2005 *Metrologia* **42** 1–10

[35] Trainotti C, Schmidt T and Furthner J 2019 Comparison of clock models in view of clock composition, clock steering and measurement fitting *Proceedings of the 50th Annual Precise Time and Time Interval Systems and Applications Meeting (PTTI 2019)* (Reston, Virginia) pp 265–283 URL <https://www.ion.org/publications/abstract.cfm?articleID=16756>

[36] Greenhall C A 2006 *Metrologia* **43** S311–S321 URL <http://iopscience.iop.org/article/10.1088/0026-1394/43/4/S19/meta>

[37] Galleani L 2011 *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* **58** 1550–1558

[38] Howe D A, Allan D U and Barnes J A 1981 Properties of signal sources and measurement methods *Thirty Fifth Annual Frequency Control Symposium* (IEEE) URL <https://ieeexplore.ieee.org/document/1537481>