

# A Robust Time Scale Based on Maximum Likelihood Estimation

Hamish McPhee, *University of Toulouse, ENSEEIHT-IRIT-TéSA, Toulouse, France*  
Jean-Yves Tournet, *University of Toulouse, ENSEEIHT-IRIT-TéSA, Toulouse, France*  
David Valat, *CNES, Toulouse, France*  
Philippe Paimblanc, *TéSA, Toulouse, France*  
Jérôme Delporte, *CNES, Toulouse, France*  
Yoan Grégoire, *CNES, Toulouse, France*

## BIOGRAPHIES

**Hamish McPhee** first received a Bachelor's degree in Mechanical and Aerospace engineering from the University of Adelaide in 2018. He then moved on to obtain an MSc in Aerospace Engineering from ISAE-Supaero in 2020, in Toulouse, France. He is currently undertaking a PhD in developing an autonomous timescale for a swarm of nanosatellites. His current research tasks are cofinanced by the French space agency CNES and the TéSA laboratory.

**Jean-Yves Tournet** (SM'08, F'19) received the ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT) de Toulouse in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992. He is currently a professor in the university of Toulouse (ENSEEIHT) and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo (MCMC) methods. He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001-2007, 2010-2015, 2019-2021) and the EURASIP SAT committee on Theoretical and Methodological Trends in Signal Processing (2015-2019). He has been serving as an associate editor for the IEEE Transactions on Signal Processing (2008-2011, 2015-2019) and for the EURASIP journal on Signal Processing (2013-2019). In 2019, He becomes president of the French Association GRETSI. He was a member of the board of directors of EURASIP from 2019 to 2021. He is currently the president of EURASIP from January 2022.

**David Valat** received his M.Sc. degree in 1997 in metrology and signal processing. He joined the Paris Observatory, SYRTE department as an RD engineer. For 12 years, he has been involved in time and frequency transfer and time scales activities. In 2011, he joined the French space agency CNES as a time and frequency expert. His main interests are the characterization of space grade oscillators, clocks remote comparison by GNSS/optical links, absolute calibration of GNSS reception chains, and time scales generation.

**Philippe Paimblanc** graduated as an electronics engineer from the ENAC (Ecole Nationale de l'Aviation Civile) in 2002 and received the same year his Master's research degree in signal processing. He performed a PhD at the satellite navigation lab of the ENAC. He is now a research engineer at TéSA laboratory, in Toulouse, France. His research activity is centered on satellite navigation and positioning.

**Jérôme Delporte** graduated as an electronics engineer from the ENAC (Ecole Nationale de l'Aviation Civile) in 1997 and received the same year his Master's research degree in microwave and optical transmissions. He is now a time-frequency senior expert at the French Space Agency CNES. His main interest relate to the timing aspects of GNSS and SBAS (time scales, atomic clocks, time transfer, absolute calibration of GNSS receiver chains, and performance monitoring). He was the chairman of the EFTF Scientific Committee from 2017-2018 and is now a member of the EFTF Executive Committee. He is the current chair of the LNE Time-Frequency Scientific Committee and a member of the French Metrology Committee. He is also a member of the GNSS timing task force of the UN-ICG (United Nations – International Committee on GNSS) that he chaired in 2017, 2019, and 2022.

**Yoan Grégoire** is a radiolocation expert at the French Space Agency (CNES). After being involved in GNSS signal processing and Cospas-Sarsat system performance activities, he is now studying future swarm technologies. In particular, he is involved in the development of localization and synchronization techniques and technologies in a swarm of satellites.

*Preprint submitted to 2023 Precise Time and Time Interval Systems and Applications Meeting*

## ABSTRACT

This paper introduces a new statistical model for clock phases assuming a multivariate Gaussian distribution for the clock phase deviations from a common time scale. This model allows us to derive a maximum likelihood estimator for the clock phases, which is consistent with the current methods of computing a common time scale for a collection of clocks. Detailing a statistical model of the clock phases, which assumes a Gaussian distribution allows us to find the MLE for each clock's phase deviation from a common time scale. For verification, the MLE for the clock phases is shown to be consistent with the result of the existing basic time scale equation. The statistical distribution of the frequency states resulting from this statistical model is Gaussian over a window of past time instants. This property can be used to design a new time scale based on the maximum likelihood estimator of frequency and frequency variances that are alternatives to the exponential filters designed for AT1. With the appropriate number of past frequency samples, this MLE has identical performance to the optimal AT1 algorithm in a nominal context. The statistical distribution of the frequency when the clock suffers a phase jump anomaly is then identified as a Student's t-distribution. The Student's t-distribution models the statistics of datasets contaminated by outliers, leading to the derivation of a different MLE that is robust to those outliers. The time scale using the robust MLE provides estimates of each clock's frequency and frequency variance that are unaffected by phase jump anomalies and improves the long-term frequency stability when each clock in the ensemble experiences phase jump anomalies within some window of time.

## I. INTRODUCTION

Satellites are susceptible to many challenges in maintaining consistent timing due to their remote and harsh environment. Indeed, modification of the payload environment causes variations in the states of onboard clocks that would not be expected in a laboratory. Time transfers between satellites are subject to noise and can be influenced by outliers such as missing data (due to limited windows of access), instantaneous faults in electronics due to cosmic radiation, and hardware degradation due to the impossibility of maintenance. The context of this work revolves around a swarm of nanosatellites whose objectives rely on consistent timing, the "Nanosatellites pour un Observatoire Interferométrique Radio dans l'Espace" (NOIRE) study. The objective of NOIRE is to use a swarm of nanosatellites that form a radio array telescope in lunar orbit where there is no radio pollution (Ceconi et al., 2018). A radio array telescope uses the principle of radio interferometry, which relies heavily on the precise timing and positioning of each satellite to make the desired observations. There exist proposals for Lunar global navigation satellite systems (GNSS) that could aid in the positioning and synchronization of the swarm (Pereira et al., 2022). However, these projects are still young and can benefit from the same methods proposed for time scale generation. The general context can be expanded to satellites in lunar orbit, which will require a robust time-scale algorithm in real-time, that combines data from a large number of satellites, and autonomously compensates for the expected anomalies. Emphasis is placed on the autonomous aspect of the algorithm because access to ground stations or laboratory clocks is considered impossible. The objective of this work is to design a specific robust Maximum Likelihood Estimator (MLE) that accounts for the effects of outliers on the noise statistics of each clock.

For simplicity of mass production for the nanosatellites, it is assumed that the clock technology is homogeneous throughout the swarm. Based on this assumption, fifty simulated clocks have been generated with each matching the typical performance of an Oven Controlled Crystal Oscillator (OCXO). This type of clock is considered an initial design choice for implementation in the NOIRE satellites. The clock data is generated by following the power law representation of clock noise processes as described in (Kasdin and Walter, 1992). Due to the stochastic nature of the clock noise processes, we can consider the phase difference measurements to be random variables. In addition, the clock frequencies are considered random variables. Works that consider Kalman Filters for the determination of a time scale commonly consider the clock noises (and any measurement noise) as Gaussian processes (Barnes et al., 1982; Breakiron, 2002; Coleman and Beard, 2020). The Gaussian assumption has also been recently used in a method of detecting anomalies up to a specified probability of false alarm (Trainotti et al., 2022). It has been shown that the effects of a frequency jump result in datasets that do not strictly follow the Gaussian law (Marszalec et al., 2021). Some other studies have also identified the non-Gaussianity caused by anomalies in general and how that can help in detecting them (Trainotti et al., 2022; Tryon and Jones, 1983).

The methodology of the MLE design consists of two major steps, with a series of assumptions and estimations made during each step. The first step begins with the inputs being a set of clock phase differences between the simulated clocks. The inputs represent phase differences that are obtained through inter-satellite links. For this work, it is assumed that all satellites can always have links with any other satellite. The clock phase differences are then considered to be linear combinations of each clock's phase deviation from a theoretically perfect clock, with random variations introduced by the internal clock noises. To obtain the last observation of our system, we make a prediction of each clock's frequency offset using previously estimated values. This, in turn, allows the prediction of each clock's phase offset from the ideal clock to create a new indirect observation with the weighted sum of phase predictions. The phase prediction error is considered another source of noise in the observation of the offsets from the perfect clock. We then model the joint distribution of phase differences and the weighted sum of phase predictions as a multivariate Gaussian distribution.

There are still not enough available observations to allow for the estimation of both the mean and variance of the modeled distribution. Hence, we assume fixed variances of the clock phases and derive the MLE of the phase offsets. The result gives the estimator of each clock's deviation from an ideal clock. This estimator is equivalent to the Basic Time Scale Equation (BTSE) that computes the deviation from the common time scale in previously validated time scale algorithms (Thomas et al., 1994; Stein, 2003). The AT1 algorithm is a common solution to compute a time scale under nominal conditions, that uses the BTSE (Weiss and Weissert, 1991). The Gaussian assumption on the clock phase data allows us to derive the MLE for the proposed model. Considering that the Gaussian assumption represents nominal clock operations, we expect some congruence between the AT1 and MLE time scale algorithms. This is represented in the fact that the time scales are computed with the same first step of determining phase offsets. The difference in algorithms is evident in the following step.

In the second step of our estimation strategy, we propose to use a sliding window of past frequency predictions to construct a new dataset. This dataset allows more states to be estimated, e.g., frequency and frequency variances. To perform this estimation, we model the past predictions of each clock's frequency using the previous Gaussian model but with known clock phases and an unknown covariance matrix. This means we can use the frequency predictions from step 1 as inputs for an MLE in the second step. The MLEs of the frequencies and frequency variances are then computed and used to update the predictions of each clock's phase offset and the weights of the clocks. The AT1 algorithm uses a different method for estimating the frequency and frequency variance (Weiss and Weissert, 1993), resulting in different phase predictions and weights to input into the first step.

The types of anomalies that have been identified for clocks operating in space include phase jumps, frequency jumps, periodic components, and missing data (Riley, 2008; Galleani and Tavella, 2012, 2013; Coleman and Beard, 2020). For this paper, the focus is placed on phase jump anomalies because they are observed to cause outliers in the frequency of the clocks. In the realm of robust statistics, methodologies exist that autonomously compensate for the impacts of such outliers without losing performance when the outliers are absent (Maronna et al., 2006). One such method is extending the MLE to a statistical distribution that appropriately accounts for the occurrence of outliers. The resulting robust MLE time scale repeats the previous two-step estimation strategy using a Student's t-distribution to compensate for the impact of phase jumps on the estimates of the frequencies and frequency variances. The derivation of an MLE for the Student's t-distribution requires an Expectation-Maximization algorithm that converges to the desired estimate (Doğru et al., 2018; Hasannasab et al., 2020).

There are several contributions that arise from our new methods of generating a time scale: (i) visualization of the clock states as random variables following a specified statistical distribution, (ii) application of the MLE to detail a justification behind the BTSE, and to propose new estimation methods for frequency and frequency variance, (iii) analysis on the effects of phase jumps on the statistical distribution of the clock data, and comparison to the Student's t-distribution, (iv) derivation of a new robust MLE based on the Student's t-distribution to estimate the clock frequencies and frequency variances, generating a robust time scale for the types of anomalies investigated. This study finally promotes the benefits of using the robust MLE for frequency to help in future work aimed towards detecting anomalies before treating them.

## II. MEASUREMENT MODEL

To reproduce an ideal time scale, we want to have access to the phase difference between each independent clock and a theoretically perfect clock. We can imagine the theoretically perfect clock as an oscillator with constant frequency and no uncertainty in the evolution of the clock phase. The perfectly stable phase state of the perfect clock is denoted as  $h_p(t)$ . Similarly, the predictable phase state of some clock  $i$  is then denoted as  $h_i(t)$ . In reality, clocks suffer deviations from the predictable phase according to a series of unpredictable internal noises. So the actual clock phases are equivalent to the predictable component of each clock plus some random deviation,

$$h'_i(t) = h_i(t) + \varepsilon_i(t), \varepsilon_i(t) \sim \mathcal{N}(0, \sigma_i^2(t)). \quad (1)$$

When observing the phase state of a clock, only a comparison between clock  $i$  and some other reference clock can be measured. In the case of simulated clock data, we generate the individual phase deviations from a perfect clock including the effects of the internal noises

$$x_{i,p}(t) = h'_i(t) - h_p(t), \quad (2)$$

where the true evolution of the phase state from a time instant  $\tau$  seconds earlier is:

$$h'_i(t) = h_i(t - \tau) + \tau f_i(t - \tau) + \varepsilon_i(t). \quad (3)$$

The term  $f_i(t - \tau)$  is the ideally constant frequency for clock  $i$  and we assume any unpredictable deviations from that frequency are contained in  $\varepsilon_i(t)$ . The predictable phase component is then  $h_i(t) = h_i(t - \tau) + \tau f_i(t - \tau)$ . For the propagation of the perfect clock, we obtain the same equation without the random deviations:

$$h_p(t) = h_p(t - \tau) + \tau f_p(t - \tau). \quad (4)$$

The frequency deviation of clock  $i$  from the perfect clock is denoted as  $y_{i,p}(t) = f_i(t) - f_p(t)$ . The phase deviation of clock  $i$  from the perfect clock is then defined as the sum of the ideal propagation component  $\theta_i(t) = h_i(t) - h_p(t)$  and the unpredictable component  $\varepsilon_i(t)$ . The resulting random variable is assumed to follow a Gaussian distribution, leading to the following model

$$x_{i,p}(t) = h'_i(t) - h_p(t) = \theta_i(t) + \varepsilon_i(t), \quad (5)$$

$$x_{i,p}(t) \sim \mathcal{N}(\theta_i(t), \sigma_i^2(t)). \quad (6)$$

To clarify, the term  $\theta_i(t)$  is the phase deviation that clock  $i$  should have from a perfect clock in the case it had a perfectly stable frequency and predictable phase. We identify the  $\theta_i(t)$  values as the parameters we wish to estimate, where any error in the estimates contributes to an error in our estimate of the perfect clock. The result is a phase deviation of each clock with respect to a reference clock that varies from the perfect clock. This reference clock is defined as our ensemble clock or time scale  $h_E(t)$ , yielding,

$$\hat{\theta}_i(t) = h_i(t) - h_E(t) = x_{i,E}(t). \quad (7)$$

In practice the  $x_{i,p}(t)$  values are not observable, instead, we observe only the differences between pairs of clocks  $x_{ij}(t) = x_{i,p}(t) - x_{j,p}(t)$ . The phase differences provide  $N - 1$  non-redundant equations under the assumption of negligible measurement noise. That is, selecting a common reference clock for each of the measurements  $x_{i1}(t)$  allows the reproduction of all other pairs of clocks using linear combinations. For example, the phase difference measurement  $x_{23}(t)$  is not necessary if both  $x_{21}(t)$  and  $x_{31}$  are already available ( $x_{23}(t) = x_{21}(t) - x_{31}(t)$ ). In the context of a swarm of nanosatellites, measurement noise is expected to be influential but for the derivations presented in this work, we maintain the common assumption that measurement noise is negligible in comparison to the clock noises in  $\varepsilon_i(t)$ . It is also assumed that the number of satellites capable of measuring phase differences is constant so there is always access to the  $N - 1$  phase differences. With only  $N - 1$  measurements and  $N$  parameters to estimate, the system is indeterminate. We then use a common method amongst time scale algorithms, defining a constraint on the weighted sum of the predicted  $\theta_i(t)$  values Thomas et al. (1994)

$$\sum_{i=1}^N w_i(t - \tau) \hat{x}_{i,E}(t) = \sum_{i=1}^N w_i(t - \tau) \theta_i(t). \quad (8)$$

This constraint is satisfied for an ideal time scale but depends on the weights and predictions of the phase deviations  $\hat{x}_{i,E}(t)$ . The difficulty in using the predictions is that they can only be made using the previous estimates,

$$\hat{x}_{i,E}(t) = \hat{\theta}_i(t - \tau) + \tau \hat{y}_{i,p}(t - \tau). \quad (9)$$

The above requires appropriate estimates of the previous phase and frequency deviations. For optimal estimation, the error in the previous estimates should approach the value of the unpredictable fluctuations due to the internal clock noises. As we are assuming that the unpredictable fluctuations in both phase and frequency are contained in  $\varepsilon_i(t)$ , we can rewrite the prediction equation as:

$$\hat{x}_{i,E}(t) = \theta_i(t - \tau) + \tau y_{i,p}(t - \tau) + \varepsilon_i(t) = \theta_i(t) + \varepsilon_i(t). \quad (10)$$

The final measurement uses the previously computed weights to take a weighted sum of the phase predictions. The weighted sum is used as a method of constraining the time scale performance. As explained in Stein (2003), the weighted sum of the clock noise should be set to zero to obtain the ideal time scale constraint in (8)

$$\sum_{i=1}^N w_i(t - \tau) \hat{x}_{i,E}(t) = \sum_{i=1}^N w_i(t - \tau) (\theta_i(t) + \varepsilon_i(t)). \quad (11)$$

The design objective is then  $\sum_{i=1}^N w_i(t - \tau) \varepsilon_i(t) = 0$ , indicating that the weight computations should assign low weights to clocks with high variance or high error between predictions and estimates. The measurement model will include the stochastic clock processes from the equation above as variances arising from an assumed distribution. The joint statistical model according to the distribution of the individual clock phases is now presented for the complete set of measurements:

$$\mathbf{z}(t) = \begin{bmatrix} x_{21}(t) \\ x_{31}(t) \\ \vdots \\ x_{N1}(t) \\ \sum_{i=1}^N w_i(t - \tau) \hat{x}_{i,E}(t) \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \theta_2(t) - \theta_1(t) \\ \theta_3(t) - \theta_1(t) \\ \vdots \\ \theta_N(t) - \theta_1(t) \\ \sum_{i=1}^N w_i(t - \tau) \theta_i(t) \end{bmatrix}, \boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t)) \right) = \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}(t)), \boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t))), \quad (12)$$

where the vector  $\boldsymbol{\theta}(t) = [\theta_1(t) \cdots \theta_N(t)]^T$  contains the independent predictable components of the clock phase deviations and  $\boldsymbol{\sigma}^2(t) = [\sigma_1^2(t) \cdots \sigma_N^2(t)]^T$  contains the variances of each clock. Both the mean vector and covariance matrix can be written in the form of matrix products with  $\boldsymbol{\theta}(t)$  and  $\boldsymbol{\sigma}^2(t)$ , respectively. The mean vector is

$$\boldsymbol{\mu}(\boldsymbol{\theta}(t)) = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \\ w_1(t-\tau) & w_2(t-\tau) & w_3(t-\tau) & \cdots & w_N(t-\tau) \end{bmatrix} \begin{bmatrix} \theta_1(t) \\ \theta_2(t) \\ \vdots \\ \theta_N(t) \end{bmatrix} = \mathbf{A}(t)\boldsymbol{\theta}(t). \quad (13)$$

When forming the joint distribution of these random variables it is important to include the correlation between each of the phase difference measurements and the predicted phase deviations. Since clock 1 is considered the common reference for all the phase differences, we can expect that the variance of clock 1 affects all the observations. The measurement noises for each pair of clocks are considered independent. Hence, the covariance matrix of the measurement vector  $\mathbf{z}(t)$  can be written as:

$$\boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t)) = \begin{bmatrix} \sigma_2^2 + \sigma_1^2 & \sigma_1^2 & \cdots & \sigma_1^2 & w_2\sigma_2^2 - w_1\sigma_1^2 \\ \sigma_1^2 & \sigma_3^2 + \sigma_1^2 & \sigma_1^2 & \vdots & \vdots \\ \vdots & \ddots & \ddots & \sigma_1^2 & \vdots \\ \sigma_1^2 & \cdots & \sigma_1^2 & \sigma_N^2 + \sigma_1^2 & w_N\sigma_N^2 - w_1\sigma_1^2 \\ w_2\sigma_2^2 - w_1\sigma_1^2 & \cdots & \cdots & w_N\sigma_N^2 - w_1\sigma_1^2 & \sum_{i=1}^N w_i^2\sigma_i^2 \end{bmatrix} = \mathbf{A}(t)\mathbf{D}(\boldsymbol{\sigma}^2(t))\mathbf{A}^T(t), \quad (14)$$

where we denote  $\mathbf{D}(\boldsymbol{\sigma}^2)$  as the diagonal matrix with the vector  $\boldsymbol{\sigma}^2$  on the diagonal. The above covariance model is an expansion of the model expressed by Levine (2012), which assumes equal variances for the clocks and does not include the  $N$ th row and column. It is known that elliptically symmetric distributions are stable under linear transforms, including the Gaussian distribution (Hasannasab et al., 2020). Hence, the following property is obtained

$$\mathbf{A}^{-1}(t)\mathbf{z}(t) \sim \mathcal{N}(\mathbf{A}^{-1}(t)\boldsymbol{\mu}(\boldsymbol{\theta}(t)), \mathbf{D}(\boldsymbol{\sigma}^2(t))), \quad (15)$$

$$\mathbf{A}^{-1}(t)\mathbf{z}(t) \sim \mathcal{N}(\boldsymbol{\theta}(t), \mathbf{D}(\boldsymbol{\sigma}^2(t))). \quad (16)$$

Pre-multiplying the set of measurements  $\mathbf{z}(t)$  by the inverse of  $\mathbf{A}(t)$  is equivalent to solving the system of  $N$  equations that come from the phase difference measurements and the constraint on the weighted sum of the clock error. This is demonstrated by writing the equations without noise and assuming a perfect estimation for  $\hat{x}_{i,E}(t)$

$$\left. \begin{aligned} x_{21}(t) &= \theta_2(t) - \theta_1(t) \\ x_{31}(t) &= \theta_3(t) - \theta_1(t) \\ &\vdots \\ x_{N1}(t) &= \theta_N(t) - \theta_1(t) \\ \sum_{i=1}^N w_i(t-\tau)\hat{x}_{i,E}(t) &= \sum_{i=1}^N w_i(t-\tau)\theta_i(t) \end{aligned} \right\} \longleftrightarrow \mathbf{z}(t) = \mathbf{A}(t)\boldsymbol{\theta}(t). \quad (17)$$

The explicit solutions for  $\theta_j(t)$  in the above system of equations are

$$\theta_j(t) = \sum_{i=1}^N w_i(t-\tau)(\hat{x}_{i,E}(t) - x_{ij}(t)) \longleftrightarrow \boldsymbol{\theta}(t) = \mathbf{A}^{-1}(t)\mathbf{z}(t). \quad (18)$$

However, if we evaluate the above equations with the clock noises and estimation error, we only obtain estimates of the phase states  $\hat{\theta}_i(t) = x_{i,E}(t)$ , and not the true clock phases. That is, the estimates represent the deviation from some time scale, which is the result of the BTSE as presented in (Weiss and Weissert, 1993; Thomas et al., 1994)

$$x_{j,E}(t) = h_i(t) - h_E(t) = \sum_{i=1}^N w_i(t-\tau)(\hat{x}_{i,E}(t) - x_{ij}(t)) \longleftrightarrow \hat{\boldsymbol{\theta}}(t) = \mathbf{A}^{-1}(t)\mathbf{z}(t). \quad (19)$$

According to (16) and (19), the estimated deviation of each clock from the time scale remains a random variable. In this case, we expect any other calculations that use this estimate to also follow some statistical distribution. For example, if we take the slope between the current and previous phase estimates, we will obtain noisy samples of the clock frequencies. We propose to use a statistical model of these frequency samples to obtain our frequency predictions and weights. In the following section, we derive the new MLE time scale that uses the statistics of the frequencies to determine an estimator for the frequency and the frequency variance. By estimating these parameters, we introduce new methods of finding the frequency predictions used in (9) and determining the weights using the frequency variance estimates.

### III. A NEW TIME SCALE BASED ON A GAUSSIAN MLE

To derive the MLE for the presented assumptions on the statistics of the measurements, we must define the multivariate probability density function (PDF) for the measurements:

$$f(\mathbf{z}(t); \boldsymbol{\mu}(\boldsymbol{\theta}(t)), \boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t))) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t))|}} \exp\left(-\frac{1}{2}(\mathbf{z}(t) - \boldsymbol{\mu}(\boldsymbol{\theta}(t)))^T \boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t))^{-1}(\mathbf{z}(t) - \boldsymbol{\mu}(\boldsymbol{\theta}(t)))\right). \quad (20)$$

This PDF gives the probability of the  $N$  measurements occurring at some instant in time, so we can estimate at most  $N$  parameters using the maximum likelihood estimator. The parameters to be estimated are the  $x_{i,p}(t)$  values, which give us the prediction of the time scale under ideal conditions. To find the  $x_{i,p}(t)$  values that have maximum probability, we can maximize the log-likelihood function  $l(\boldsymbol{\theta}(t); \boldsymbol{\mu}(\boldsymbol{\theta})) =$

$$\log(f) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t))| - \frac{1}{2}(\mathbf{z}(t) - \boldsymbol{\mu}(\boldsymbol{\theta}(t)))^T \boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t))^{-1}(\mathbf{z}(t) - \boldsymbol{\mu}(\boldsymbol{\theta}(t))). \quad (21)$$

Taking the derivative w.r.t.  $\boldsymbol{\theta}(t)$  :

$$\frac{\partial \log(f)}{\partial \boldsymbol{\theta}(t)} = \left( \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta}(t))}{\partial \boldsymbol{\theta}(t)} \right)^T \boldsymbol{\Sigma}(\boldsymbol{\sigma}^2(t))^{-1} [\mathbf{z}(t) - \boldsymbol{\mu}(\boldsymbol{\theta}(t))] = 0, \quad (22)$$

$$\hat{\boldsymbol{\theta}}(t) = \mathbf{A}^{-1} \mathbf{z}(t). \quad (23)$$

The final result is identical to simultaneously solving the  $N$  equations defined in (17). Therefore, the BTSE can be considered the result that maximizes the likelihood for the measurement model specified above. However, the output of the BTSE still follows some statistical distribution (see (16)). Due to this inherited uncertainty in the estimate of the time scale, the prediction of frequency can also be considered a random variable. Using a slope approximation of the frequency, and presenting the slope of the error in the time scale offset as  $\eta_i(t)$

$$y_{i,s}(t) = \frac{x_{i,E}(t) - x_{i,E}(t - \tau)}{\tau} = y_{i,p}(t) + \eta_i(t), \quad (24)$$

$$y_{i,s}(t) \sim \mathcal{N}(y_{i,p}(t), \sigma_{y_i}^2(t)). \quad (25)$$

We assume that the frequency predictions made at consecutive time instants maintain a constant mean and variance. That is, each clock is representative of a clock with a constant frequency that suffers some stochastic deviations at each point in time. Then, we obtain independent samples of the constant frequency using slope approximations across a chosen window of past time steps. The samples are then defined as  $y_{i,s}(t - M\tau), \dots, y_{i,s}(t)$ , where  $M$  is the number of past samples used and the window length is  $M + 1$ .

Using at least  $2N$  samples of frequency (obtainable using only the  $N$  frequency approximations from the current and previous time steps), we can obtain the MLE for both the frequency and the frequency variance of each clock. Under the assumption of a Gaussian distribution, the parameter MLEs for this set of observations are simply the sample mean and sample variance Maronna et al. (2006):

$$\hat{y}_{i,p}(t) = \frac{1}{M+1} \sum_{m=0}^M y_{i,s}(t - m\tau), \quad (26)$$

$$\hat{\sigma}_{y_i}^2(t) = \frac{1}{M+1} \sum_{m=0}^M (y_{i,s}(t - m\tau) - \hat{y}_{i,p}(t))^2. \quad (27)$$

To avoid the need for saving all past samples and increasing memory demand, the estimation is performed on a sliding window of samples with a fixed size. The sliding window size must be large enough to ensure MLE convergence to the optimum estimate. It is noted here that we draw a specific parallel to the nominally optimal AT1 time scale. The algorithm to compute AT1 defines two time constants,  $m_i$  and  $N_\tau$  for the exponential filters in the following relations

$$\hat{y}_{i,p}(t) = \frac{y_{i,s}(t) + m_i \hat{y}_{i,p}(t - \tau)}{1 + m_i}, \quad (28)$$

$$\epsilon_i^2(t) = \frac{\hat{\epsilon}_i^2(t) + N_\tau \epsilon_i^2(t - \tau)}{1 + N_\tau}. \quad (29)$$

Where  $\hat{\epsilon}_i = |x_{i,E}(t) - \hat{x}_{i,E}(t)|$  is the clock prediction error. The exponential filters essentially place lower weights on the more recent predictions to avoid rapid changes that could cause instabilities. The magnitude of the time constants relates to the amount of time between two samples that will vary according to some desired type of stochastic noise. The time constant for the prediction error ( $N_\tau$ ) is set to the period at which the white frequency noise is expected to be constant (Weiss and Weissert, 1991). Similarly, the time constant for the frequency update ( $m_i$ ) is selected to be the period at which the random walk frequency noise first becomes dominant (Weiss and Weissert, 1991). If flicker frequency noise is more suitable than random walk frequency noise,  $m_i$  is the time interval at which the white noise intercepts with flicker noise on the ADEV curve. Both time constants can be set to be equal for the simulated OCXO clocks,  $N_\tau = m_i = 100$  (see Appendix A). We then take this value of the time constants as the constant width of the sliding window for the MLE algorithm. As the MLE is expected to improve estimation accuracy with more samples, we will also assess higher values of  $M$  and correspondingly alter the AT1 time constants for comparison. With the window length defined, the MLEs for the frequency and frequency variance are computed as alternatives to the outputs of the exponential filters in the AT1 algorithm.

The key difference between the AT1 time scale and the Gaussian MLE time scale is that the weights are computed with respect to an estimated frequency variance instead of a prediction error. The weights from the previous iteration of either algorithm cause a bias in the estimation of the variance, which is corrected with the factor  $K_i$  derived by Tavella et al. (1991):

$$\hat{\epsilon}_i^2(t) = |x_{i,E}(t) - \hat{x}_{i,E}(t)|^2 K_i(t), \quad (30)$$

$$\hat{\sigma}_{y_i}^2(t) = \frac{K_i(t)}{M + 1} \sum_{n=0}^M (y_{i,s}(t - n\tau) - \hat{y}_{i,p}(t))^2, \quad (31)$$

$$K_i(t) = \frac{1}{1 - w_i(t - \tau)}. \quad (32)$$

Finally, these corrected estimates of prediction error and frequency variance are inverted and normalized to obtain the weights to be used in the next iteration of the BTSE.

$$w_{i,AT1}(t) = \frac{\frac{1}{\hat{\epsilon}_i^2(t)}}{\sum_{i=1}^N \frac{1}{\hat{\epsilon}_i^2(t)}}, \quad (33)$$

$$w_{i,MLE}(t) = \frac{\frac{1}{\hat{\sigma}_{y_i}^2(t)}}{\sum_{i=1}^N \frac{1}{\hat{\sigma}_{y_i}^2(t)}}. \quad (34)$$

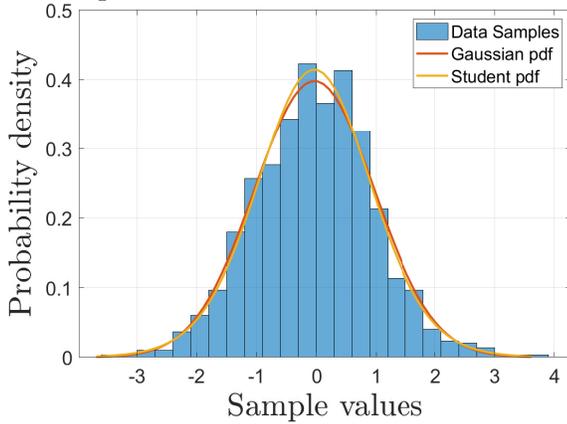
With the equations above, the Gaussian MLE time scale is fully defined. The algorithm follows familiar steps to the existing AT1 algorithm but with an alternate method for predicting the frequency and frequency variance, which leads to changes in the phase predictions and weights, respectively. The resulting Gaussian MLE and AT1 time scales for equal window lengths and time constants are presented in Section VI, together with the MLE derived for a more outlier inclusive distribution defined in the next section. More precisely, we assume a different distribution of the frequency approximations over the sliding window. The Student's t-distribution is selected to represent the effect of phase jump anomalies and is verified in the following section.

#### IV. ANOMALOUS CLOCK STATISTICS AND STUDENT MLE

Anomalies occur at unpredictable times and result in an unpredictable change in the instantaneous clock states. We expect the presence of phase jumps in the ensemble to cause a greater occurrence of outlying samples. A phase jump at some time can be related to a temporary frequency jump (outlier in frequency) at the same time. The histograms presented in this section will show the distribution of the frequency of a specific clock over a window of time samples, with and without outliers. In

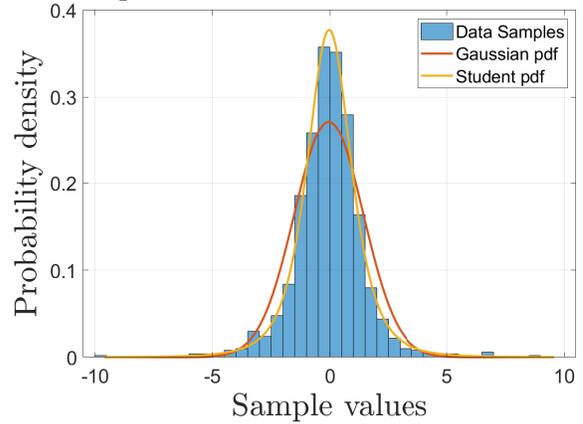
addition, the PDFs corresponding to both a Gaussian distribution and a Student's t-distribution are computed and superimposed to see if they fit the data. The t-distribution is chosen as it belongs to the same family of distributions as the Gaussian but differs according to a shape parameter called the degrees of freedom  $\nu$ . As is illustrated in Figure 1a, for a high value of  $\nu$ , the t-distribution is approximately equivalent to the Gaussian distribution. Conversely, Figure 1b demonstrates that outliers have a higher probability with fewer degrees of freedom (see the scale of the samples on the x-axis).

Example of Student's t-distribution  $\nu = 300$



(a) Examples of the probability density for 1000 samples generated by the Student's t-distribution parameterized as  $t(0, 1, 300)$ .

Example of Student's t-distribution  $\nu = 3$

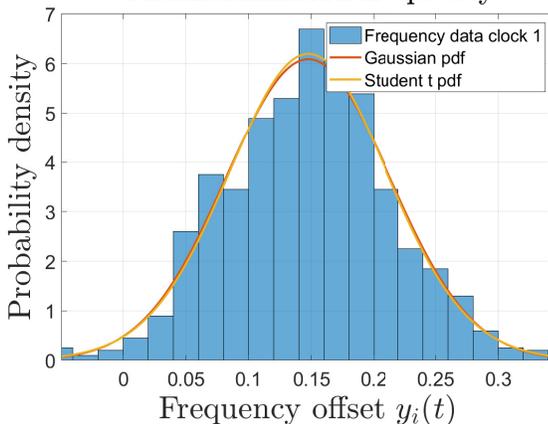


(b) Examples of the probability density for 1000 samples generated by the Student's t-distribution parameterized as  $t(0, 1, 3)$ .

**Figure 1:** Histograms of synthetically generated data.

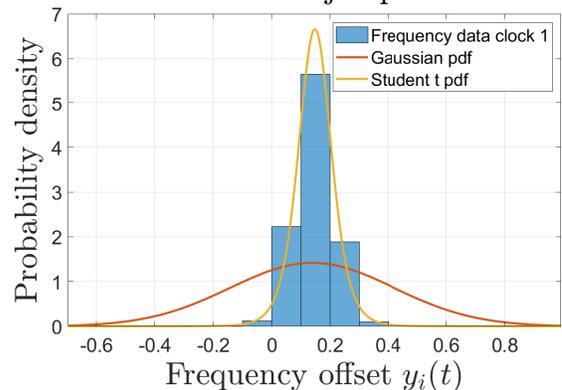
To verify the Student's t-distribution as a reasonable assumption for clock data in presence of anomalies, we include significant magnitudes of phase jumps on the simulated clock data. These magnitudes are  $\Delta x = 10$  ns multiplied by a random scalar value to keep the same order of magnitude but have an unpredictable amplitude. A greater number of jumps occurring on a single clock further modifies the statistics of the data for that clock. In Figure 2a we present the baseline for a simulated clock's frequency statistics in nominal operations. As expected, including a phase jump on a clock introduces an outlier in the frequency state of that clock. As seen in Figure 2b, the effect of one clock having one phase jump over a window of 1000 samples causes a noticeable change in the frequency statistics. The Gaussian PDF that is computed to fit this distribution is skewed by the outlier, resulting in an inflated variance and a shift in the mean. This indicates that robust estimation of the frequency based on the t-distribution should neglect the impact of the phase jump.

Uncontaminated frequency



(a) Baseline frequency statistics for a single clock under nominal operating conditions.

Contaminated frequency with Phase jump



(b) Frequency distribution for a window of 1000 samples, with one phase jump at the end of the window.

**Figure 2:** Histograms of clock frequency data.

The objective of a robust MLE is to provide the best estimate of the parameters of the distribution that includes outliers, where the outliers do not have any negative impact. Assuming that the frequency states of the clocks have a t-distribution, the following distribution of the frequency can be obtained:

$$y_{i,s}(t) \sim t(y_{i,p}(t), \sigma_{y_i}^2(t), \nu_i). \quad (35)$$

Since we cannot guarantee knowledge of which clock is experiencing the anomaly, it is not possible to make this assumption on only the faulty clock. Considering the potential for anomaly detection algorithms (Galleani and Tavella, 2012; Trainotti et al., 2022), this could be an eventual application. However, for the remainder of this work, we make the same assumption for all clocks and place a phase jump on each clock at a different time. This is equivalent to stating that each clock has some chance of suffering a phase jump within our total simulation period. This assumption is not baseless for the context of this project, where the NOIRE swarm has a collection of homogeneous clocks operating in the same hostile environment. Then, the joint assumption on all clocks allows us to introduce the multivariate PDF for  $\mathbf{y}_s(t) \sim t(\boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}(t), \nu(t))$  (Doğru et al., 2018):

$$\begin{aligned} f(\mathbf{y}_s(t); \boldsymbol{\mu}_y(t), \boldsymbol{\Sigma}(t), \nu(t)) \\ = \prod_{m=0}^M \frac{\Gamma\left(\frac{\nu(t)+M}{2}\right) |\boldsymbol{\Sigma}(t)|^{-\frac{1}{2}}}{(\pi\nu(t))^{M/2} \Gamma\left(\frac{\nu(t)}{2}\right)} \left[1 + \frac{1}{\nu(t)} (\mathbf{y}_s(t-m\tau) - \boldsymbol{\mu}_y(t))^T \boldsymbol{\Sigma}(t)^{-1} (\mathbf{y}_s(t-m\tau) - \boldsymbol{\mu}_y(t))\right]^{-\frac{(\nu(t)+M)}{2}}, \end{aligned} \quad (36)$$

where  $\Gamma(a)$  is the gamma function, the frequency slope approximations are contained in the random vector  $\mathbf{y}_s(t) = [y_{1,s}(t) \cdots y_{N,s}(t)]^T$ , and the ideal frequency offsets for each clock are  $\boldsymbol{\mu}_y(t) = [y_{1,p}(t) \cdots y_{N,p}(t)]^T$ . The exact MLEs for the parameters in the above PDF are difficult to derive. However, by introducing latent variables that are related to the target parameters, an approximation of the MLE can be obtained (Doğru et al., 2018). An Expectation-Maximization (EM) algorithm is used to generate a sequence of parameters that converge to a local maximum of the Student's t-likelihood. For iteration  $k$  of the EM algorithm, the parameters are updated according to the following equations:

$$\hat{u}_m^{(k)} = \frac{\hat{\nu}^{(k-1)} + N}{\hat{\nu}^{(k-1)} + (\mathbf{y}_s(t-m\tau) - \hat{\boldsymbol{\mu}}_y^{(k-1)})^T \hat{\boldsymbol{\Sigma}}_{(k-1)}^{-1} (\mathbf{y}_s(t-m\tau) - \hat{\boldsymbol{\mu}}_y^{(k-1)})} \quad (37)$$

$$\hat{\boldsymbol{\mu}}_y^{(k)}(t) = \frac{\sum_{n=0}^M \hat{u}_n^{(k)} \mathbf{y}_s(t-m\tau)}{\sum_{n=0}^M \hat{u}_n^{(k)}}, \quad (38)$$

$$\hat{\boldsymbol{\Sigma}}_{(k)}(t) = \frac{1}{M+1} \sum_{m=0}^M \hat{u}_m^{(k)} (\mathbf{y}_s(t-m\tau) - \hat{\boldsymbol{\mu}}_y^{(k)}(t)) (\mathbf{y}_s(t-m\tau) - \hat{\boldsymbol{\mu}}_y^{(k)}(t))^T, \quad (39)$$

$$\hat{\nu}^{(k)}(t) : \text{solve } \psi\left(\frac{\hat{\nu}^{(k)}(t)}{2}\right) - \psi\left(\frac{\hat{\nu}^{(k-1)}(t) + N}{2}\right) + \sum_{m=0}^M (u_m^{(k)} - \log(u_m^{(k)}) - 1) = 0, \quad (40)$$

where the terms  $\hat{u}_m^{(k)}$  are weighting factors that assign less importance to outliers, and  $\psi(a)$  is the digamma function. The correction factor  $K_i(t)$  is applied after the EM algorithm to remove biases due to past weights. A term depending on  $\nu$  is also included to convert the estimated scale matrix  $\hat{\boldsymbol{\Sigma}}$  to an equivalent covariance matrix  $\mathbf{C}(t) = \frac{\hat{\nu}^{(k)}(t)}{\hat{\nu}^{(k)}(t)-2} \hat{\boldsymbol{\Sigma}}(t)$  for the t-distribution. The new correction term can then be applied as follows:

$$K_i(t) = \frac{1}{1 - w_i(t-\tau)} \frac{\hat{\nu}^{(k)}(t)}{\hat{\nu}^{(k)}(t) - 2}, \quad (41)$$

$$\hat{\sigma}_i^2(t) = K_i(t) \hat{\boldsymbol{\Sigma}}_{(i,i)}(t), \quad (42)$$

where  $\hat{\boldsymbol{\Sigma}}_{(i,i)}(t)$  is the  $i$ th diagonal element of the matrix  $\hat{\boldsymbol{\Sigma}}(t)$ . The implemented EM algorithm uses the ‘‘classical EM’’ code from Hasannasab et al. (2020). It requires us to initialize the parameters, which can be done using the Gaussian MLE results before computing  $\hat{u}_m$  and consequently updating the parameter estimates until reaching the stopping rule  $e < 10^{-5}$  with

$$e = \frac{\sqrt{\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_{k-1}\|^2 + \|\boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_{k-1}\|^2}}{\sqrt{\|\boldsymbol{\mu}_{k-1}\|^2 + \|\boldsymbol{\Sigma}_{k-1}\|^2}} + \frac{\sqrt{(\log(\nu_k) - \log(\nu_{k-1}))^2}}{|\log(\nu_{k-1})|}. \quad (43)$$

It is acceptable to consider that each clock in the ensemble has the same number of degrees of freedom because each clock has a chance to experience the same type of anomaly with the same order of magnitude. The time scale generated with the results of the robust MLE is referred to as the Student MLE time scale. It is constructed in the same manner as the Gaussian MLE time scale but obtains the frequency estimates and computes weights using (38) and (39), respectively.

## V. ALGORITHMS

The time scale algorithms introduced in this paper are compared with the AT1 time scale as a reference. Before analyzing the resulting characteristics of the time scales, the similarities and differences between the algorithms are presented in the following algorithm summaries. The initialization, prediction, and time scale generation steps are identical for AT1, Gaussian MLE, and Student MLE, although the numerical output of these steps differs because of the reliance on the later steps.

---

### Algorithm 1 AT1 algorithm

---

**Input:**  $x_{ij}(t)$ ,  $N = 50$ ,  $t_f = 10486$  s,  $x_{i,p}(0)$ ,  $x_{i,p}(10)$

**Initialization:**  $m_i = 100$  s,  $N_\tau = 100$  s,  $w_i(10) = \frac{1}{N}$ ,  $\tau = 1$  s,  $\hat{y}_{i,p}(t - \tau) = \frac{x_{i,p}(10) - x_{i,p}(0)}{10\tau}$ ,  $x_{i,E}(10) = 0$   
**for**  $t = 11, \dots, t = t_f$  **do**

**Prediction:**

$$\hat{x}_{i,E}(t) = x_{i,E}(t - \tau) + \tau \hat{y}_{i,p}(t - \tau),$$

**Time scale generation:**

$$x_{i,E}(t) = \sum_{j=1}^N w_j(t - \tau) (\hat{x}_{j,E}(t) - x_{ji}(t)),$$

**Frequency update:**

$$y_{i,s}(t) = \frac{x_{i,E}(t) - x_{i,E}(t - \tau)}{\tau}, \hat{y}_{i,p}(t) = \frac{y_{i,s}(t) + m_i \hat{y}_{i,p}(t - \tau)}{1 + m_i},$$

**Weight computation:**

$$K_i(t) = \frac{1}{1 - w_i(t - \tau)}, \hat{\epsilon}_i^2(t) = |x_{i,E}(t) - \hat{x}_{i,E}(t)|^2 K_i(t), \epsilon_i^2(t) = \frac{\hat{\epsilon}_i^2(t) + N_\tau \epsilon_i^2(t - \tau)}{1 + N_\tau}, w_i(t) = \frac{\frac{1}{\epsilon_i^2(t)}}{\sum_{i=1}^N \frac{1}{\epsilon_i^2(t)}}$$

**end for**

**Output:**  $x_{i,E}(t)$

---

The key design differences between algorithms are in the frequency update and weight computation steps. The AT1 algorithm uses the concept of exponential filters to reduce the influence of more recent estimates. In the initialization step, appropriate time constants are chosen so that only variations over a certain interval can contribute to the frequency and weight estimates. The idea of the MLE algorithms is to make a similar choice in time constant by choosing a number of past frequency samples  $M$  that provides a certain statistical distribution of data, i.e., a certain level of random variations affects the frequency.

---

### Algorithm 2 Gaussian MLE algorithm

---

**Input:**  $x_{ij}(t)$ ,  $N = 50$ ,  $t_f = 10486$  s,  $x_{i,p}(0)$ ,  $x_{i,p}(10)$

**Initialization:**  $M = 100$  s,  $w_i(10) = \frac{1}{N}$ ,  $\tau = 1$  s,  $\hat{y}_{i,p}(t - \tau) = \frac{x_{i,p}(10) - x_{i,p}(0)}{10\tau}$ ,  $x_{i,E}(10) = 0$   
**for**  $t = 11, \dots, t = t_f$  **do**

**Predict:**

$$\hat{x}_{i,E}(t) = x_{i,E}(t - \tau) + \tau \hat{y}_{i,p}(t - \tau),$$

**Time scale generation:**

$$x_{i,E}(t) = \sum_{j=1}^N w_j(t - \tau) (\hat{x}_{j,E}(t) - x_{ji}(t)),$$

**Frequency update:**

$$y_{i,s}(t) = \frac{x_{i,E}(t) - x_{i,E}(t - \tau)}{\tau}, \hat{y}_{i,p}(t) = \frac{1}{M+1} \sum_{n=0}^M y_{i,s}(t - n\tau),$$

**Weight computation:**

$$K_i(t) = \frac{1}{1 - w_i(t - \tau)}, \hat{\sigma}_{y_i}^2(t) = \frac{K_i(t)}{M+1} \sum_{n=0}^M (y_{i,s}(t - n\tau) - \hat{y}_{i,p}(t))^2, w_i(t) = \frac{\frac{1}{\hat{\sigma}_{y_i}^2(t)}}{\sum_{i=1}^N \frac{1}{\hat{\sigma}_{y_i}^2(t)}}$$

**end for**

**Output:**  $x_{i,E}(t)$

---

The weights of the MLE algorithms are defined as the inverse of the estimated frequency variances. The estimate of the frequency variance concerns the uncertainty in the slope approximation of the frequency, which differs from the phase prediction error  $\epsilon_i(t)$  used in the AT1 algorithm. Nevertheless, both algorithms observe the instability in the change of phase using the new estimates. The AT1 algorithm applies another exponential filter to avoid large contributions from the most recent prediction error. The MLE algorithms use the sliding window to ensure that each new sample is only 1 out of  $M$  samples that contribute to the variance estimates.

---

**Algorithm 3** Student MLE algorithm

---

**Input:**  $x_{ij}(t)$ ,  $N = 50$ ,  $t_f = 10486$  s,  $x_{i,p}(0)$ ,  $x_{i,p}(10)$

**Initialization:**  $M = 100$  s,  $w_i(10) = \frac{1}{N}$ ,  $\tau = 1$  s,  $\hat{y}_{i,p}(t - \tau) = \frac{x_{i,p}(10) - x_{i,p}(0)}{10\tau}$ ,  $x_{i,E}(10) = 0$   
**for**  $t = 11, \dots, t = t_f$  **do**

**Predict:**

$$\hat{x}_{i,E}(t) = x_{i,E}(t - \tau) + \tau \hat{y}_{i,p}(t - \tau),$$

**Time scale generation:**

$$x_{i,E}(t) = \sum_{j=1}^N w_j(t - \tau) (\hat{x}_{j,E}(t) - x_{j,i}(t)),$$

**Frequency update:**

$$y_{i,s}(t) = \frac{x_{i,E}(t) - x_{i,E}(t - \tau)}{\tau}, \hat{y}_{i,p}(t) = \frac{\sum_{m=0}^M \hat{u}_m^{(k)} y_{i,s}(t - m\tau)}{\sum_{m=0}^M \hat{u}_m^{(k)}},$$

$$\hat{u}_m = \frac{\hat{\nu} + N}{\hat{\nu} + s_m}, s_m = \sum_{i=1}^N \frac{(y_{i,s}(t - m\tau) - y_{i,E}(t))^2}{\hat{\sigma}_i^2(t)}$$

**Weight computation:**

$$K_i(t) = \frac{1}{1 - w_i(t - \tau)} \frac{\nu}{\nu - 2}, \hat{\sigma}_{y_i}^2(t) = \frac{K_i(t)}{M+1} \sum_{m=0}^M \hat{u}_m (y_{i,s}(t - m\tau) - \hat{y}_{i,p}(t))^2, w_i(t) = \frac{\frac{1}{\hat{\sigma}_{y_i}^2(t)}}{\sum_{i=1}^N \frac{1}{\hat{\sigma}_{y_i}^2(t)}}$$

**end for**

**Output:**  $x_{i,E}(t)$

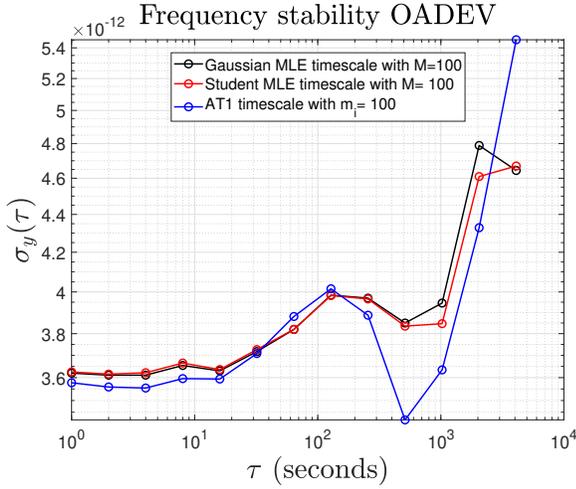
---

## VI. RESULTS

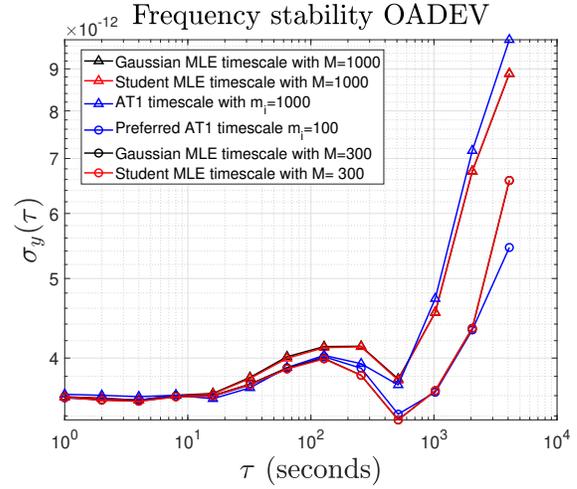
The time scales are assessed in nominal conditions with the simulated clocks as detailed in Appendix A. Figure 3a shows the frequency stability performance of each time scale with the best choice of time constants and MLE window size chosen to match those time constants. The metric used to assess the frequency stability is the Overlapping Allan Deviation (OADEV) and is plotted against the averaging time  $\tau$ . The time scales have a similar trend in OADEV with a similar order of magnitude. However, the OADEV of the AT1 time scale with sampling intervals of 500 seconds is preferred over the MLE time scales. To assess the similarity in the choice of window size and time constants, the same time scales are again computed with different initialization of those parameters. Figure 3b shows that when  $M = 300$ , the MLE time scales approach the same OADEV as the AT1 time scale with the preferred time constants. Furthermore, increasing both the window size and time constants to 1000 seconds also results in equal performance (seen with triangle data points), although this performance has higher OADEV.

This gives insight into the selection of the window size for the MLE time scales. As expected, the number of samples included in the window should be sufficient enough to ensure that the EM algorithm converges to accurate estimates. However, the number of samples should not be too large because the dominant noise of the clocks could change to a non-preferred type, as is the case with the AT1 time constants. For the next analysis on performance with phase jump anomalies, the time constants for AT1 are kept as  $m_i = N\tau = 100$ , as shown above, and the window size is  $M = 300$ . The robustness of the time scale algorithms is next to be assessed. By including a phase jump at a random time on each clock in the ensemble, we appropriately represent some common effects on all clocks due to the space environment. A phase jump on a clock causes an outlier in that clock's frequency. We expect this outlier to be automatically accounted for with the robust estimator, i.e., the Student MLE. Figure 4a shows that this is the case for the frequency estimation, where the Gaussian MLE and AT1 estimates of frequency are clearly impacted by the outlier. Similarly, the weight estimate from the Student MLE is shown to neglect the outlier in Figure 4b. This is also expected for a robust estimator, that considers the outliers as actual samples occurring with some designated low probability.

A weight that is blind to an outlier is typically useless in the time scale design. Nevertheless, the robust estimator has achieved its task of eliminating the outlier from its estimates. As detailed in (37), (38), and (39) there are different weights assigned to the past samples of the frequency to determine the robust frequency and variance estimates, without a priori knowledge of the anomaly and without losing performance in the nominal case. The end results are the approximated values of  $y_{i,E}(t)$  and  $w_i(t)$  that would be obtained if clock  $i$  had no outlier. The performances of the time scales are presented in Figure 5. Despite the weight being poorly assigned, the improved frequency estimate has resulted in some gain in the long-term frequency stability for the Student MLE time scale.

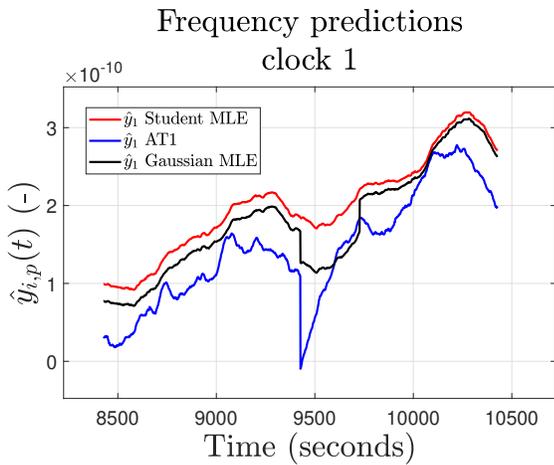


(a) Performance of the designed MLE time scales compared to AT1,  $M = m_i = N_\tau = 100$

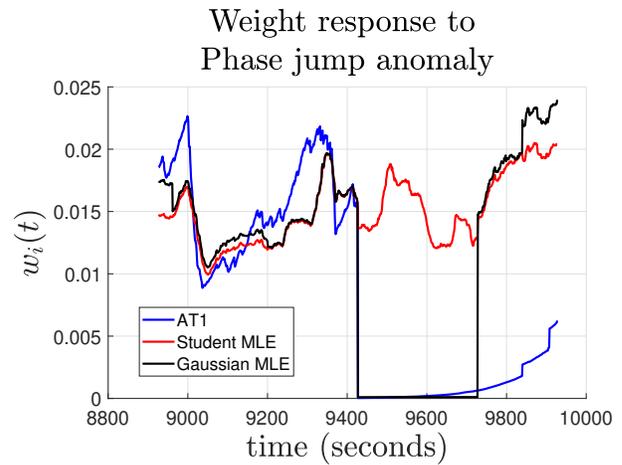


(b) Change in the MLE time scale performances depending on the window length and similarity to the AT1 time constant.

**Figure 3:** Frequency stability for different window sizes (note Gaussian MLE curves are superimposed with Student MLE curves on the right).



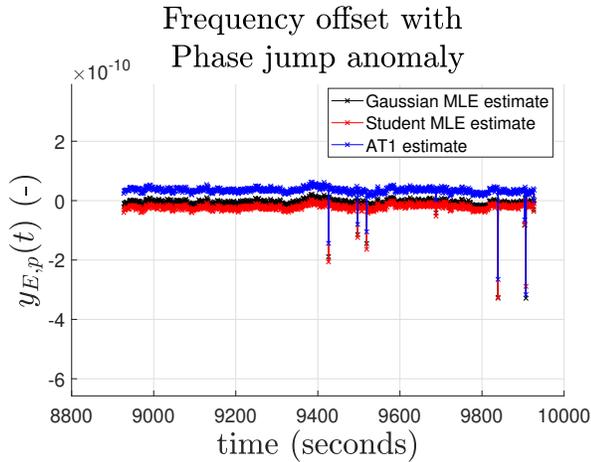
(a) Example of robust frequency estimation for clock 1 with phase jump.



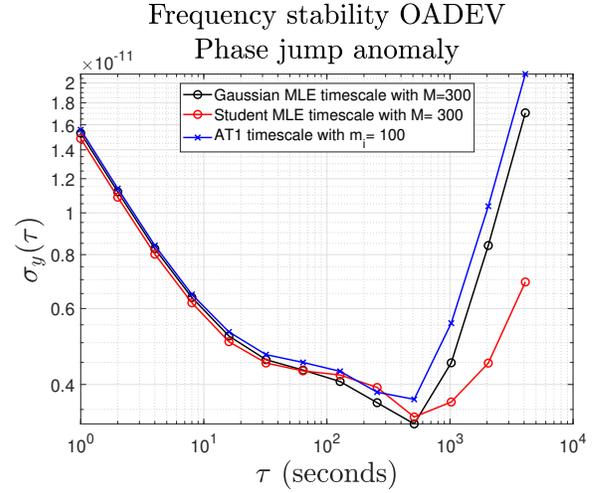
(b) Weight computations from the robust and non-robust variance estimates.

**Figure 4:** Robust estimates.

Using the robust MLE solution is a passive way of lowering the impact of the phase jump anomalies, where there is no design choice required in terms of thresholds, probabilities of false alarms, or empirically chosen weight smoothing functions. For future work, the results from the robust MLE could be included in an active anomaly detection procedure. One potential method could include choosing a threshold for the difference between the robust estimates and the non-robust estimates. This would then be comparable to the active phase jump detection methods suggested for AT1 in Levine (2012) and Weiss and Weissert (1993) (which were not used in our realization of AT1). The same weight control terms that were suggested for use in AT1 could even be applied using a new threshold based on the robust estimate of frequency variance.



(a) Time scale frequencies.



(b) Time scale stabilities.

**Figure 5:** Frequency performance when each clock experiences a phase jump.

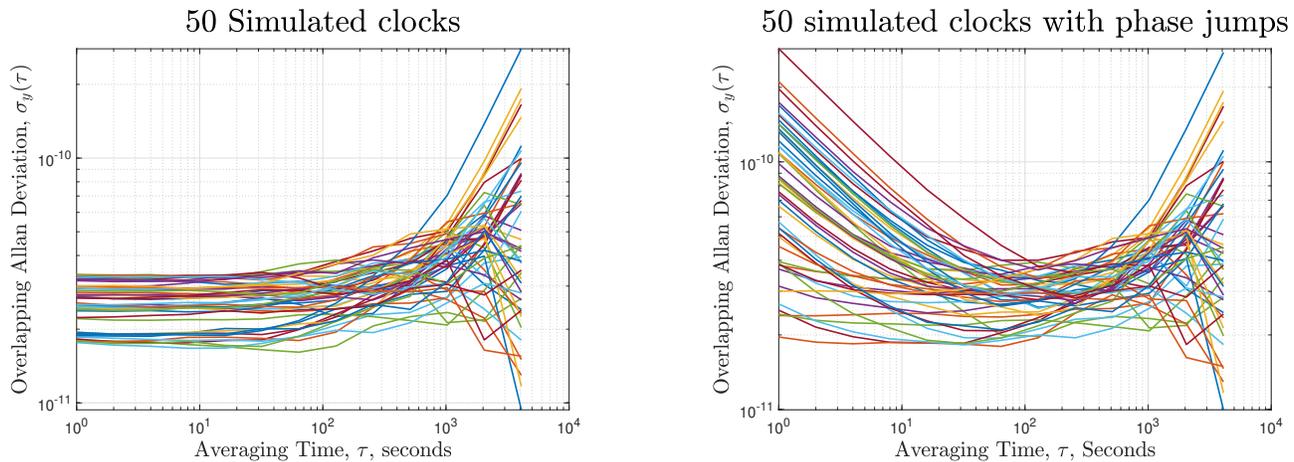
## VII. CONCLUSION

Statistical models of clock phase and frequency have been used to derive the MLE of clock phases, frequencies, and frequency variances. It has been proven that the result of the BTSE maximizes the likelihood of the specified model of phase measurements. Consequently, the new MLE-based time scales use identical prediction and update equations to the estimates of phase. The key differences in the design of the MLE-based time scales and the AT1 time scale are the frequency updates and weight computations. Indeed, another MLE solution is obtained using a sliding window of past frequency approximations to replace the exponential filters of the AT1 algorithm. The sliding window size must be large enough to ensure a good convergence of the EM algorithm, but not so large that the frequency variance is defined by some undesirable clock noise process.

Nanosatellites operating in a harsh environment will be prone to a variety of anomalies. It is important to mitigate the presence of these anomalies in the generation of a time scale amongst a swarm of nanosatellites. This paper focuses on phase jump anomalies that are equivalent to outliers in frequency data, which can then be modeled by the Student's t-distribution. The time scales based on the Gaussian MLE and Student MLE are confirmed to match the performance of the AT1 time scale under nominal conditions. Finally, when each clock in the ensemble is affected by a phase jump anomaly at some random point in time, the Student MLE time scale is shown to robustly estimate the frequency and frequency variance of the anomalous clock, and obtain slight improvements in the long-term frequency stability. The next steps of this work include exploring the statistics resulting from different clock anomalies and using the robust MLE methodology to aid in detecting anomalies.

### A. SIMULATED CLOCKS

The outputs of our clock simulation model are detailed in Figure 6. The clock generator code was created by inputting characteristics from the expected Allan deviation curves for typical space-qualified OCXO clocks. The clock variances were randomly modified to ensure that each clock has unique behaviors from the other clocks generated. Finally, phase jumps with the order of magnitude of 10 ns are used to produce the clocks in the presence of anomalies.



**Figure 6:** Overlapping Allan Deviation for the simulated clocks, without anomalies (left), and with phase jumps on each clock (right).

## ACKNOWLEDGEMENTS

The authors of this paper would like to thank both T SA laboratory and CNES for providing funding to complete this study. In particular, the expertise offered by CNES has been very valuable in understanding the implementation of time scales and the challenges to overcome.

## REFERENCES

- Barnes, J. A., Jones, R. H., Tryon, P. V., and Allan, D. W. (1982). Stochastic models for atomic clocks. In *Proc. of the 14th Ann. Precise Time and Time Interval (PTTI) Appl. Planning Meeting*, pages 295–306, Greenbelt, MD, USA.
- Breakiron, L. A. (2002). A Kalman Filter for Atomic Clocks and Timescales. In *Proc. 33rd Annual Precise Time and Time Interval Systems and Applications Meeting*, pages 431 – 444, Long Beach, CA, USA.
- Cecconi, B., Dekkali, M., Briand, C., Segret, B., Girard, J., Laurens, A., Lamy, A., Valat, D., Delpech, M., Bruno, M., G lard, P., Bucher, M., Nenon, Q., Griessmeier, J.-M., Boonstra, A.-J., and Bentum, M. (2018). NOIRE study report: Towards a low frequency radio interferometer in space. In *Proc. Aerospace Conf.*, pages 1–19, Big Sky, MT, USA.
- Coleman, M. J. and Beard, R. L. (2020). Autonomous clock ensemble algorithm for GNSS applications. *Navigation*, 67(2):333–346.
- Dođru, F. Z., Bulut, Y. M., and Arslan, O. (2018). Doubly reweighted estimators for the parameters of the multivariate t-distribution. *Communications in Statistics-Theory and Methods*, 47(19):4751–4771.
- Galleani, L. and Tavella, P. (2012). Detection of atomic clock frequency jumps with the Kalman filter. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, 59(3):504–509.
- Galleani, L. and Tavella, P. (2013). Characterization of atomic clock anomalies in the dynamic allan variance domain. In *Proc. Joint European Frequency and Time Forum International Frequency Control Symposium (EFTF/IFC)*, pages 645–648, Prague, Czech Republic.
- Hasannasab, M., Hertrich, J., Laus, F., and Steidl, G. (2020). Alternatives to the EM algorithm for ML estimation of location, scatter matrix, and degree of freedom of the student t distribution. *Numerical Algorithms*, 87(1):77–118.
- Kasdin, N. and Walter, T. (1992). Discrete simulation of power law noise (for oscillator stability evaluation). In *Proc. Int. Frequency Control Symposium*, Hershey, PA, USA.
- Levine, J. (2012). Invited Review Article: The statistical modeling of atomic clocks and the design of time scales. *Review of Scientific Instruments*, 83(2):021101.
- Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics. Wiley, 1st edition.

- Marszalec, M., Lusawa, M., and Osuch, T. (2021). Efficient frequency jumps detection algorithm for atomic clock comparisons. *Metrology and Measurement Systems*, vol. 28(1):107–121.
- Pereira, F., Reed, P. M., and Selva, D. (2022). Multi-objective design of a lunar gnss. *NAVIGATION: Journal of the Institute of Navigation*, 69(1).
- Riley, W. J. (2008). Algorithms for frequency jump detection. *Metrologia*, 45(6):S154.
- Stein, S. (2003). Time scales demystified. In *Proc. Int. Frequency Control Symposium*, pages 223–227, Boulder, CO.
- Tavella, P., Azoubib, J., and Thomas, C. (1991). Study of the clock-ensemble correlation in ALGOS using real data. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, 34(78):435–441.
- Thomas, C., Wolf, P., and Tavella, P. (1994). *Time Scales*. BIPM.
- Trainotti, C., Giorgi, G., and Günther, C. (2022). Detection and identification of faults in clock ensembles with the generalized likelihood ratio test. *Metrologia*, 59(4):045010.
- Tryon, P. and Jones, R. (1983). Estimation of Parameters in Models for Cesium Beam Atomic Clocks. *Journal of Research of the National Bureau of Standards*, 88(1).
- Weiss, M. and Weissert, T. (1991). AT2, A New Time Scale Algorithm: AT1 Plus Frequency Variance. *Metrologia*, 28(2):65–74.
- Weiss, M. A. and Weissert, T. P. (1993). Sifting through nine years of NIST clock data with TA2. In *Proc. 7th European Frequency and Time Forum*, pages 199–210, Neuchatel, Switzerland.