A Robust Time Scale for Space Applications Using the Student's t-distribution

Hamish McPhee[1,2], Jean-Yves Tourneret[2], David Valat[3], Jérôme Delporte[3], Yoan Grégoire[3], Philippe Paimblanc[1]

[1] TéSA, Toulouse, France, [2] University of Toulouse, Toulouse, France, [3] CNES, Toulouse, France

Abstract. In this article, the principles of robust estimation are applied to the standard basic time scale equation to obtain a new method of assigning weights to clocks. Specifically, the Student's t-distribution is introduced as a new statistical model for an ensemble of clocks that are experiencing phase jumps, frequency jumps or anomalies in their measurement links. The proposed robust time scale is designed to mitigate the effects of these anomalies without necessarily identifying them, but through applying a method of robust estimation for the parameters of a Student's t-distribution. The proposed time scale algorithm using the Student's t-distribution (ATST) is shown to achieve comparable robustness to phase jumps, frequency jumps, and anomalies in the measurements with respect to the AT1 oracle time scale. The AT1 oracle is a special realization of the AT1 time scale which corrects all anomalies by having prior knowledge of their occurrences. The similar performance of ATST and AT1 oracle suggests that the ATST algorithm is efficient for obtaining robustness with no prior knowledge or detection of the occurrences of anomalies.

Keywords: Robust time scale, phase jumps, frequency jumps, inter-satellite links, measurement anomalies

1. Introduction

A spatial reference frame is an important piece of information to understand the position of an object. Similarly, a time reference frame is necessary to establish a meaningful time stamp of an event. This type of reference frame is called a time scale, which is defined using the timing information of an ensemble of clocks. The resulting time scale provides a "virtual" or "paper" reference clock that achieves better stability than any individual clock in the ensemble. A robust time scale is required in applications where no manual corrections can be made, although the need for corrections is necessary. Such applications can be remote, technologically limited, and deployed in harsh environments. For example, a constellation of nanosatellites is simultaneously: i) limited by size, weight, power, and cost of clock technologies, and ii) more prone to anomalies due to the environment and time transfer methods. Specifically, malfunctioning clocks and erroneous measurements are challenges that must be faced when generating a robust time scale for space applications. Existing time scale algorithms consider the assignment

of particular weights to each clock in an ensemble depending on their stability. For an abnormally behaving clock, the associated stability is expected to be reduced. Hence, the weights computed in time scale algorithms should be adapted to appropriately de-weight anomalous clocks and measurements.

Many strategies of assigning weights to the clocks of an ensemble have been investigated in the current literature. The primary solutions apply weights to the differences between predictions of clock states and measurements of the clock time differences. This is true for the AT1 algorithm [1–3] and is implicit in the composite clock computed with a Kalman Filter (KF) [4–8]. Both of these time scale algorithms perform somewhat optimally in the nominal case with different motivations in choosing clock weights. Particular methodologies are proposed for the detection and correction of different types of clock anomalies [2, 9–15]. A common limitation in the design of such detection algorithms is the choice of a detection threshold that potentially constrains the magnitude and types of anomalies that can be detected depending on the test statistic being used. The contribution of this article autonomously mitigates anomalies by adjusting clock weights according to a single set of observations and without a limitation linked to a fixed threshold.

The AT1 algorithm is presented in this work as the state-of-the-art solution for a time scale under normal operating conditions. Two separate methodologies have been defined for robustifying the AT1 algorithm, each method detects and compensates for either clock phase jumps or frequency jumps [2, 9]. It is expected that anomalies occurring in the measurement links cannot be detected and corrected with the same methods. A swarm of satellites will be heavily influenced by measurement noise that is possibly contaminated by outliers. Hence, this work aims to introduce a new time scale algorithm that better mitigates such measurement-based anomalies. To the best knowledge of the authors, there are no existing investigations into generating a time scale robust to both link anomalies and jump anomalies.

Anomaly detection is a significant field of research for clock models. This paper studies an alternative to detection by mitigating clock anomalies using a robust estimation algorithm. In a general sense, robust signifies a method that maintains good performance in the absence of anomalies and does not lose significant performance in the presence of a small number of anomalies, regardless if they are actively detected or not. To confirm the validity of a robust time scale, it should perform as well as the AT1 algorithm in the nominal case and the AT1 algorithm with perfect detection of anomalies, herein referred to as AT1 oracle because it effectively knows the occurrence of all anomalies before the moment they occur. With this knowledge, AT1 oracle reduces the weights of any anomalous clock to zero at the exact time of the anomaly, any time there is an anomaly and hence, prevents the time scale from being affected by individual clock faults.

To design a robust time scale, the generated time scale should maintain good frequency stability when abnormalities occur from either source (clocks or measurement links). The proposed methodology to obtain a robust time scale is based on principles of

robust statistics [16, 17]. More precisely, we assume that the combination of normal and corrupted clock data is modeled by a Student's t-distribution, which implicitly assumes that the data has some probability of containing outliers. This distribution comes from a family of heavy-tailed probability distributions that assigns non-zero probabilities to the occurrence of outliers [16]. Therefore, the Maximum Likelihood Estimators (MLEs) for the defining parameters of the Student's t-distribution are robust by taking outliers into account. The method of mitigating the impact of the outliers in an MLE is comparable to the method of assigning weights to the clock measurements and building the time scale. Hence, a robust time scale is proposed that takes advantage of this similarity.

This paper is organized as follows. Section 2 briefly introduces and summarizes the Basic Time Scale Equation (BTSE). Section 3 presents the main contribution of the paper. This contribution is a new time scale algorithm that uses a robust estimation procedure to obtain weights for the clock phases. The methodology aligns with the BTSE and provides a robust solution that does not require active detection of anomalies yet remains robust in the presence of phase jumps, frequency jumps and link anomalies. Finally, section 4 introduces a specific scenario that is simulated to replicate the clock ensemble and anomalies expected in a swarm of nanosatellites. Comparisons are made between the performances of the newly suggested algorithm, AT1 without any anomaly detection, and the perfectly robust AT1 oracle algorithm.

2. Basic Time Scale Equation

The output of a time scale algorithm is the difference in time between each clock contained in the ensemble and the mutual time scale $x_{i,E}(t) = h_i(t) - h_E(t)$, where $h_i(t)$ indicates the absolute time of clock i, which is unobservable. Each clock can then be synchronized because the time scale $h_E(t)$ is a common reference time. To compute $x_{i,E}(t)$, the algorithm requires predictions of each of the clock phases $\hat{x}_{j,E}(t)$, measurements of the time differences between all of the clocks in the ensemble $x_{ji}(t)$, and weights assigned to each clock in the ensemble $w_j(t-\tau)$, where τ is the time between consecutive measurements. A time scale is then realized for an ensemble of N clocks using the BTSE [18–21]:

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

where the predictions, measurements, and weights can be defined differently for each time scale algorithm. By default, the weights are computed at the previous epoch for AT1 because the equation for the weights needs a previous computation of $x_{i,E}(t)$ to estimate the prediction error at each step in time. Including anomaly detection in the AT1 algorithm means that the weights that are computed after the first realization of each $x_{i,E}(t)$ are used to recompute the BTSE and obtain a corrected $x_{i,E}(t)$ if an anomaly is detected. This means the weights would be instead denoted as $w_j(t)$ in the BTSE. The AT1 algorithm combined with perfect detection of any anomaly will be referred to

as AT1 oracle to represent the case that any type or magnitude of anomaly is always detected and the weights of the affected clocks are set to zero before the recomputation of the BTSE.

The BTSE can also be explained as the solution to a system of N equations that includes a constraint on the weighted average of the clock noise

$$\sum_{j=1}^{N} w_j(t-\tau)\left(\hat{x}_{j,E}(t) - h_j(t)\right) = 0, \tag{2}$$

and relevant phase difference measurements ($N-1$ equations for the independent $x_{ji}(t)$) [18, 19]. The constraint (2) cannot be ensured because there is no direct access to $h_j(t)$. However, the weights should compensate for this by minimizing the prediction errors. This leads to an alternative expression for the BTSE:

$$x_{i,E}(t) = h_i(t) + \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - h_j(t)\right]. \tag{3}$$

The aim is to estimate the instantaneous phase deviation of clock i from a theoretically perfect time scale, denoted as $h_i(t)$. Errors between the predictions and observations are expected to be responsible for the difference between our obtainable time scale and the theoretically perfect time scale. Intuitively, the prediction errors $e_j(t) = \hat{x}_{j,E}(t) - h_j(t)$ can represent an unpredictable component, i.e., caused by stochastic processes only. As indicated above, the weights should ideally reduce these stochastic components of the clock phase to zero. Similarly, supplementary time scale equations could be used to determine clock frequency and drift with independent weights that set the corresponding stochastic components to zero [19]. Each of the terms in the BTSE will be detailed independently for two different types of time scale algorithms below. Depending on the algorithm, the methods of computing clock weights, and the definition of the statistics will differ.

2.1. Measurement noise

Time scales are typically generated after pre-processing any source of measurement noise. This means that the phase difference measurements that are substituted into the BTSE are assumed to have negligible noise. In reality the measurement noise will be more significant in a swarm of nanosatellites. Due to other constraints, the satellites may not necessarily perform the required pre-processing to reduce measurement noise to a negligible level. In the context of using inter-satellite links and pseudo-range solutions to obtain the clock phase differences, the measurement noise is expected to vary as a function of the relative inter-satellite distances. Hence, certain pairs of satellites could provide better-quality measurements. This is an important factor to consider when discussing the BTSE because the result will no longer provide a common reference. In

presence of additive measurement noise, the BTSE can be rewritten as follows:

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - (x_{ji}(t) + n_{ji}(t))\right], \tag{4}$$

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - x_{ji}(t)\right] - \sum_{j=1}^{N} w_j(t-\tau) n_{ji}(t), \tag{5}$$

with measurement noise $n_{ji}(t)$ referring to the random noise on the specific link between clocks i and j. For each clock i, the resulting estimate of $x_{i,E}(t)$ will depend on the weighted sum of the noise on each link that includes clock i. In addition, anomalies in the measurement process can cause certain links to have significantly greater measurement noise than others. The measurements that are affected by anomalies should then be given lower weights in the BTSE to minimize the impact of corrupted measurements.

3. Atomic Time Scale using the Student's t-distribution

Ideally, a robust time scale can compensate for anomalies of a wide range of magnitudes and types without any degradation in the nominal case. As was discussed above, anomaly detection methods have constraints on the choice of test statistics and detection thresholds depending on the type of anomaly. The objective of this section is to present a new weighting procedure for the calculation of the BTSE that provides a robust time scale not reliant on the detection or identification of specific anomalies. The weights will be based on the MLE for the mean of the Student's t-distribution, generating the new Atomic Time scale using the Student's T-distribution (ATST).

3.1. Measurements

The ATST algorithm will consider all unique pairs of satellites as sources of independent phase difference measurements. Unique measurement noise is present on each satellite link. As was defined in Section 2, the phase difference measurements are

$$z_{ji}(t) = x_{ji}(t) + n_{ji}(t). \tag{6}$$

Anomalies in the measurement links can be represented by some outlier being added to the noise on a specific link

$$\tilde{n}_{ji}(t) = n_{ji}(t) + \Delta n_{ji}(t). \tag{7}$$

Phase difference measurements are grouped into sets of $N-1$ measurements according to the common reference clock i. This is also necessary for the BTSE to compute $x_{i,E}(t)$ in the AT1 algorithm.

3.2. Predictions

Identical to the AT1 prediction step, the proposed ATST algorithm assumes a second-order polynomial to predict the clock phases

$$\hat{x}_{i,E}(t) = x_{i,E}(t-\tau) + \tau y_{i,E}(t-\tau) + \frac{\tau^2}{2} d_{i,E}(t-\tau). \tag{8}$$

The true clock phase will deviate from the predictable clock phase by some random amount according to the internal noises. The possibility of phase jumps and frequency jumps will cause the clocks to sometimes deviate from the predictable clock by an even greater amount than the standard clock models expect. The ATST algorithm assumes that any anomaly in the clocks modifies the instantaneous true phase by some unpredictable bias

$$\tilde{h}_i(t) = h_i(t) + \Delta h_i(t). \tag{9}$$

These deviations are effectively outliers in the clock phases at a given instant in time. The outliers due to clock anomalies can be observed in the prediction errors that are defined as the difference between the predicted and true phases

$$\hat{x}_{i,E}(t) - \tilde{h}_i(t) = \tilde{e}_i(t) = e_i(t) + \Delta h_i(t), \tag{10}$$

where $e_i(t)$ is the prediction error in the case no anomaly occurred at time t. We make this assumption at all time instants regardless if any anomaly occurs. Since the phase state is also affected by frequency jumps, we expect that only modifying the assumption on the phase prediction error is sufficient to achieve a robust result.

3.3. Residuals

The BTSE residuals can be expressed in terms of what is already considered to be affected by anomalies, e.g., if there is an anomaly on clock j and/or on link ji:

$$r_{ji}(t) = \hat{x}_{j,E}(t) - z_{ji}(t), \tag{11}$$

$$r_{ji}(t) = \hat{x}_{j,E}(t) - (\tilde{h}_j(t) - h_i(t) + \tilde{n}_{ji}(t)), \tag{12}$$

$$r_{ji}(t) = h_i(t) + e_j(t) + \Delta h_j(t) - n_{ji}(t) - \Delta n_{ji}(t). \tag{13}$$

The above form of the residuals shows a collection of terms that represent the unpredictability of each clock j and each link between clock j and the fixed clock i. The residuals provide N assumed independent observations of the absolute phase of clock i. It is assumed that the stochastic prediction errors $e_j(t)$ and measurement noise $n_{ji}(t)$ combined with the outliers due to internal anomalies $\Delta h_j(t)$ and link anomalies $\Delta n_{ji}(t)$ imply a Student's t-distribution for $r_{ji}(t)$. Additional assumptions allow simplification of the parameters for the defining distribution. The measurement noise is assumed to be zero mean with a uniform variance across all links, each link is independent of the other links and link anomalies are random occurrences that could appear on any link. The

internal clock noises are also considered independent, with random chances of suffering from an anomaly. Consequently, the BTSE residuals follow the Student's t-distribution:

$$r_{ji}(t) \sim T\left(h_i(t), \sigma_i^2(t), \nu_i(t)\right). \tag{14}$$

The parameters of the above distribution depend only on the reference clock i because the phase of clock i is kept fixed for the N residual samples, corresponding to the phase difference measurements made with respect to clock i. That is, for each possible reference clock i there is a unique distribution defined by: the mean $x_{i,E}(t)$, which provides the phase offset of clock i from the designed time scale, the scale parameter $\sigma_i^2(t)$ relating to the dispersion of every other clock compared to clock i, and the shape parameter $\nu_i(t)$ describing how many anomalies are present that can affect the estimation of $x_{i,E}(t)$. The shape factor $\nu_i(t)$ indicates how many of the residuals contain outliers, and the magnitudes of those outliers.

A certain value of $\nu_i(t)$ corresponds to only one out of N residuals being affected by an anomaly. If that anomaly increases to a significantly larger value, then $\nu_i(t)$ decreases to correspond with longer tails of the distribution. If several measurements had outliers, then $\nu_i(t)$ would also decrease to indicate a higher likelihood of those anomalies. This allows us to deal with different mixtures of corrupted measurements by assigning appropriate weights as long as the statistical assumption remains correct. Making a robust estimate of the mean mitigates the anomalies modeled by the t-distribution and provide a robust realization of the time scale. It is true that some measurements are reused for different sets of residuals so there can be some level of correlation between the dispersion and shape parameters for each of the possible reference clocks. Nevertheless, the mean is the main parameter of interest and always remains unique to each reference clock.

Note that the N measurements include $r_{ii}(t) = h_i(t) + \hat{x}_{i,E}(t) - h_i(t) = \hat{x}_{i,E}(t)$, which is simply the sum of the absolute phase of clock i and the prediction error for clock i. This is the same form as $r_{ji}(t) = h_i(t) + \hat{x}_{j,E}(t) - h_j(t)$, representing the observations made by the other BTSE residuals which deviate from $h_i(t)$ by their corresponding prediction errors. In other words, if the anomaly is on clock i, the residual $r_{ii}(t)$ will appear as an outlier compared to the other measurements that have all jumped due to the common observation of clock i, which now deviates from the predicted value. The number of clocks N must be sufficiently large for the statistical model to be efficient, in this work $N = 50$ clocks is assumed to coincide with the available number of clocks envisioned for a swarm of satellites. The assumption of the Student's t-distribution is verified for this number of clocks in Appendix A.1, which fits both a Gaussian and Student's t-distribution to the residuals in the presence of phase jumps, frequency jumps, and link anomalies.

3.4. Estimation of Parameters for the Student's t-distribution

To generate the ATST, only the equations in the Expectation Maximization (EM) algorithm are necessary to obtain a weighted average of BTSE residuals. More precisely, the t-distributed BTSE residuals for a fixed clock i, i.e., $r_{ji}(t) \sim T\left(\mu_i(t), \sigma_i^2(t), \nu_i(t)\right)$, have a mean that is equivalent to the offset of clock i from the generated ATST time scale, i.e., $x_{i,E}(t)$. The parameters depend on the reference clock used for the measurements but we use a generalized notation for the parameters and omit the dependence on time to avoid complexity in the relevant subscripts. The reader is invited to consult Appendix A for a more detailed explanation of the Student's t-distribution and the associated MLE to estimate the parameters μ, σ^2, ν. As each of the parameters of interest in the Student's t-distribution depends on the other two, the MLEs of μ, σ^2, ν cannot be computed directly. However, it is well known that the Student's t-distribution can be represented by an infinite mixture of Gaussian distributions [22]:

$$r_{ji} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{v_j}\right),\ v_j \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \tag{15}$$

where $\mathcal{G}(a, b)$ denotes a gamma distribution with parameters a and b. The joint PDF of $\mathbf{r} = [r_{1,i}, \cdots, r_{N,i}]^T$ and $\mathbf{v} = [v_1, \cdots, v_N]^T$ provides the so-called complete likelihood function. Taking the logarithm of the complete likelihood leads to:

$$l_c(\mathbf{r}, \mathbf{v}) = \frac{n\nu}{2} - n \log \Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu+1}{2} - 1\right) \sum_{j=1}^{N} \log v_j - \frac{N}{2} \log(2\pi) \tag{16}$$

$$-\frac{N}{2} \log\left(\sigma^2\right) - \frac{1}{2} \sum_{j=1}^{N} v_j \left[\nu + \frac{(r_{ji} - \mu)^2}{\sigma^2}\right]. \tag{17}$$

Marginalizing the complete likelihood with respect to $\mathbf{v}$ yields the likelihood as described in (A.2). This representation allows an iterative EM algorithm to be derived [22]. After an initialization of the unknown parameters, the EM alternates between Expectation (E) and Maximization (M) steps:

- **Initialization**: The location and scale parameters are initialized with the Gaussian MLEs and the number of degrees of freedom is chosen to be small because that will help minimize the number of iterations in case there is an anomaly without necessarily increasing demand in the nominal case:

$$\hat{\mu}_0 = \frac{1}{N} \sum_{i=1}^{N} r_{ji}, \tag{18}$$

$$\hat{\sigma}_0^2 = \frac{1}{N-1} \sum_{i=1}^{N} (r_{ji} - \hat{\mu}_0)^2, \tag{19}$$

$$\hat{\nu}_0 = 3. \tag{20}$$

- **E Step**: At iteration k, given $\hat{\boldsymbol{\theta}}_{k-1} = \left(\hat{\mu}_{k-1}, \hat{\sigma}^2_{k-1}, \hat{\nu}_{k-1}\right)^T$, the E step computes the expectation of $l_c(\mathbf{z}, \mathbf{v})$ with respect to the variables v_i, which requires the following computations

$$u_{j,k} = E[v_j | r_{ji}, \hat{\theta}_{k-1}] = \frac{\hat{\nu}_{k-1} + 1}{\hat{\nu}_{k-1} + \frac{(r_{ji} - \hat{\mu}_{k-1})^2}{\hat{\sigma}^2_{k-1}}}, \tag{21}$$

$$w_{j,k} = E[\log(v_j) | r_{ji}, \hat{\theta}_{k-1}] = \psi\left(\frac{\hat{\nu}_{k-1} + 1}{2}\right) - \log\left(\frac{1}{2}\left(\hat{\nu}_{k-1} + \frac{(r_{ji} - \hat{\mu}_{k-1})^2}{\hat{\sigma}^2_{k-1}}\right)\right), \tag{22}$$

  and leads to the objective function Q

$$Q(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_k) = \frac{N\nu}{2} - N \log\Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu+1}{2} - 1\right)\sum_{i=1}^{N} w_{j,k} - \frac{N}{2}\log(2\pi) \tag{23}$$

$$-\frac{N}{2}\log\left(\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N} u_{j,k}\left[\nu + \frac{(r_{ji} - \mu)^2}{\sigma^2}\right]. \tag{24}$$

- **M Step**: At iteration k, the M Step maximizes the Q function with respect to the parameters μ, σ^2, ν, which yields

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} u_{j,k} r_{ji}}{\sum_{i=1}^{N} u_{j,k}}, \tag{25}$$

$$\hat{\sigma}^2_k = \frac{\sum_{i=1}^{N} u_{j,k}(r_{ji} - \hat{\mu}_{k-1})^2}{N-1}, \tag{26}$$

$$\phi\left(\frac{\hat{\nu}_k}{2}\right) + \sum_{i=1}^{N}\left[u_{j,k} - w_{j,k} - 1\right] = 0. \tag{27}$$

The solution to (27) is obtained through the application of Newton's method, which converges to a solution after few iterations. Consider the solution to (27) is the root of the function $f(\nu_k)$, then Newton's method allows us to iteratively approximate that root by the following with each iteration n

$$\nu_{k,n+1} = \nu_{k,n} - \frac{f(\nu_{k,n})}{f'(\nu_{k,n})}, \tag{28}$$

where the initial guess can be the previous estimate of the number of degrees of freedom, $\nu_{k,0} = \hat{\nu}_{k-1}(t)$. The above is repeated until reaching a maximum number of iterations or a chosen difference between consecutive approximations is reached.

Note that the initial value of the number of degrees of freedom is chosen to reduce the number of iterations in the case we have outliers. This initialization does not affect the results when there are no anomalies because the location and scale estimates will already converge to the t-MLE using the Gaussian MLE. Note also that the estimator is declared to have converged when a certain stopping rule has been reached. The stopping rule is usually a minimum difference between estimates on consecutive iterations, say

$\varepsilon < 0.01$, where ε is computed using all the parameters [23]. The resulting EM algorithm is described in Algorithm 1:

Algorithm 1 Expectation Maximization for a Robust MLE

function tEM($[r_{1,i}(t), \cdots, r_{N,i}(t)]$)

Init.: $\hat{\mu}_0(t) = \frac{1}{N}\sum_{j=1}^{N} r_{ji}(t)$, $\hat{\sigma}_0^2(t) = \frac{1}{N-1}\sum_{j=1}^{N}\left(r_{ji}(t) - \hat{\mu}_0(t)\right)^2$,$\hat{\nu}_0(t) = 3$,

while $\varepsilon > 0.01$ **do**

$u_{j,k} = \frac{\hat{\nu}_{k-1}+1}{\hat{\nu}_{k-1}+\frac{(r_{ji}-\hat{\mu}_{k-1})^2}{\hat{\sigma}_{k-1}^2}}$, $\hat{\mu}_k = \frac{\sum_{j=1}^{N} u_{j,k} r_{ji}}{\sum_{j=1}^{N} u_{j,k}}$, $\hat{\sigma}_k^2 = \frac{\sum_{j=1}^{N} u_{j,k}(r_{ji}-\hat{\mu}_{k-1})^2}{N-1}$,

ν_k estimated as the solution of the following equation

$$\phi\left(\frac{\nu_k}{2}\right) - \phi\left(\frac{\hat{\nu}_{k-1}+1}{2}\right) + \sum_{j=1}^{N}\left[u_{j,k} - \log(u_{j,k}) - 1\right] = 0.$$

$k = k + 1$

end while

return $\hat{\mu}_k$, $\hat{\sigma}_k^2$, $\hat{\nu}_k$,$u_{j,k}$

end function

The following comment is appropriate: outliers in the data introduce a bias on the estimate of the mean of the Gaussian distribution because the outlying measurements have equal weight to every other measurement. The estimation of the mean of the Student's t-distribution is robust to anomalies because it assigns lower weights to less probable measurements, i.e., outliers. For this reason, we are interested in using the weights determined by the MLE for the Student's t-distribution to produce a new robust time scale. Next, the above EM algorithm is related to the key components of the BTSE to demonstrate how the weighted average provided in the M step is equivalent to the weighted average in the BTSE if the random variables are appropriately chosen.

3.5. Weights

Once the EM algorithm described above has converged, the normalized values of $u_{j,k}$ used in the estimate of the mean act as the final weights on each of the BTSE residuals. The terms $u_{j,k}$ are designed to mitigate outliers because a greater difference between the data and the estimated mean results in a smaller weight. The robust estimates of the scale parameter and number of degrees of freedom are also included in the weight computation. Both these parameters do not change depending on clock j for the residuals r_{ji} and hence do not affect the weights of individual clocks more than a scaling factor. The estimate $\hat{\mu}$ is obtained according to a weighted average of the BTSE residuals with the normalized weights $w_j = u_{j,k} \Big/ \sum_{i=1}^{N} u_{i,k}$. Hence, the robust MLE of the location parameter is equivalent to a BTSE with specifically designed weights. As in the BTSE, each $x_{i,E}(t)$ is computed using a different clock i as the common reference

for the measurements, and hence, the residuals. The ATST time scale is then obtained

$$x_{i,\mathrm{ATST}}(t) = \mathrm{tEM}\left([r_{1,i}(t), \cdots, r_{N,i}(t)]\right) = \frac{\sum_{j=1}^{N} u_{j,k} r_{ji}(t)}{\sum_{j=1}^{N} u_{j,k}}, \tag{29}$$

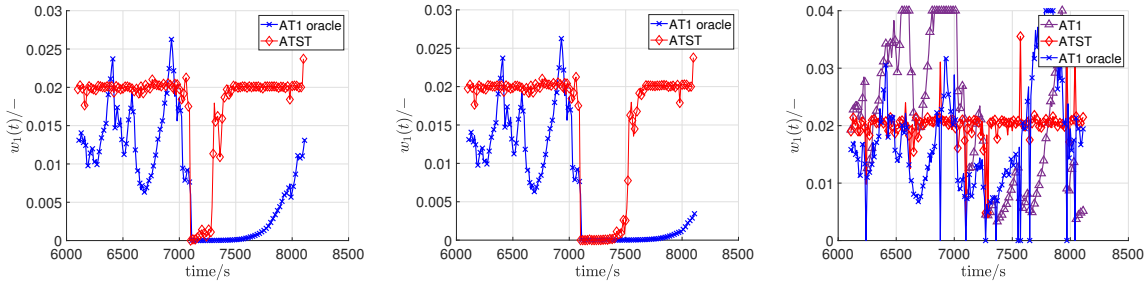
where $\mathrm{tEM}(r_{ji}(t))$ is the output from the EM algorithm and differs depending on which set of residuals are used at the input. At each iteration of the EM algorithm, the weights are further refined thanks to using the adapted estimates $\hat{\mu}_{k-1}$ and $\hat{\sigma}^2_{k-1}$. The iterative nature of the EM algorithm also inherently involves a re-computation of the BTSE at each iteration. This is similar to the strategy for AT1 oracle which compensates for phase jumps and frequency jumps by recomputing the BTSE using the weights computed at time t. The ATST algorithm is essentially an extension of this procedure that re-adapts the weights before each computation. However, instead of a detection threshold or a priori knowledge on the occurrence of anomalies, ATST uses a convergence threshold. Intuitively, the convergence is obtained quickly (often immediately) if no anomalies are present. More iterations are needed when the initial estimate significantly deviates from the first re-computation.

The ATST weighting procedure differs from the standard BTSE because each set of residuals obtains a unique set of weights. That is, the weights are assigned to each unique phase difference measurement and hence, the weights computed for estimating $x_{1,\mathrm{ATST}}(t)$ are not necessarily the same as the weights computed for $x_{2,\mathrm{ATST}}(t)$. Despite this variation in the weights used for each realization of the ATST, it can be demonstrated that each estimate still results in a common time scale. By varying the weights according to the residuals for each unique link j, i, the algorithm can observe increased deviations due to anomalies in both the measurements and the clocks. Such errors result in reduced weights to simultaneously deal with phase jumps, frequency jumps, and link anomalies. The robustness of the ATST time scale to these types of anomalies is demonstrated in the next section alongside the AT1 oracle.

Additionally, the weights computed according to the MLE defined above have naturally ensured that there is no dominant clock by assigning very similar weights at each point in time. This can be observed in Figure 1, where the weight for a single clock in an ensemble of fifty clocks is concentrated around 1/50 except for time epochs with anomalies. By keeping the weights of nominal clocks equally distributed around 1/50, the time scale will not be susceptible to sudden changes in clocks that have a dominant weight at a certain time instant then face an anomaly at the next.

Figure 1 shows only the weights of the three clocks that are affected by anomalies as shown in the histograms of Figure A2. We can observe a reduction in weights at the corresponding time of the anomalies, except for the link anomalies. The weight of a single clock should not necessarily be reduced because a single link has an anomaly, even if that clock is part of the affected link. Instead, the ATST weight corresponding to the affected measurement link is reduced. The weights associated with phase jumps and link anomalies are only reduced for a short period, whereas the frequency jump results in a weight that gradually rejoins the ensemble. The weights of the less stable

clocks are reduced while the weights of the other clocks are increased somewhat equally to compensate.



(a) Phase jump. (b) Frequency jump. (c) Link anomaly.

Figure 1: Weights for clock 1 as defined by the ATST and AT1 oracle time scale when clock 1 suffers from each of the different types of anomalies. When considering the link anomalies, the weight defined by AT1 differs from AT1 oracle which sets the weight of clock 1 to zero whenever an anomaly is on a link including clock 1.

The resulting ATST time scale should also have a frequency estimation procedure. The ATST time scale is not concerned with robust estimation of the frequency, although it would also be possible with the Student's t-distribution [24]. Since the estimation of $x_{i,\mathrm{ATST}}(t)$ is already robust, we expect the first approximation of frequency to also be robust. This frequency is defined as:

$$y_{i,s}(t) = \frac{x_{i,\mathrm{ATST}}(t) - x_{i,\mathrm{ATST}}(t-\tau)}{\tau}. \tag{30}$$

As a result, the same exponential filter as the AT1 algorithm can be applied, i.e.,

$$y_{i,\mathrm{ATST}}(t) = \frac{y_{i,\mathrm{ATST}}(t-\tau) + m_i y_{i,s}(t)}{1+m_i}. \tag{31}$$

The time constant $m_i = 100$ s is chosen based on the types of clocks involved, so the ATST algorithm maintains some level of compatibility with mixtures of different types of clocks. By using the same methodology of frequency updates as the AT1 algorithm, we eliminate one aspect that could affect the relative performance of ATST and AT1. The final comparison will be solely based on the different methods of assigning weights to the BTSE residuals. At the beginning of the algorithm, the frequency estimate can be initialized at zero for a first prediction of the phase. It is assumed that the drift is negligible, so it is initialized at a value of zero and remains at zero throughout the duration of the simulation. With the complete definitions of the estimations used, the ATST algorithm is presented in Algorithm 2:

Algorithm 2 ATST Time Scale Generation

Init.: $\hat{x}_{i,\mathrm{ATST}}(0) = 0$, $y_{i,\mathrm{ATST}}(0) = 0$, $d_{i,\mathrm{ATST}}(0) = 0$,
for $11\tau \leq t \leq t_f$ **do**
  $\hat{x}_{i,\mathrm{ATST}}(t) = x_{i,\mathrm{ATST}}(t-\tau) + \tau y_{i,\mathrm{ATST}}(t-\tau) + \frac{\tau^2}{2} d_{i,\mathrm{ATST}}(t-\tau)$,
  $\hat{y}_{i,\mathrm{ATST}}(t) = y_{i,\mathrm{ATST}}(t-\tau) + \tau d_{i,\mathrm{ATST}}(t-\tau)$,
  $d_{i,ATST}(t) = d_{i,ATST}(t-\tau)$,
  for $1 \leq i \leq N$ **do**
    $r_{ji}(t) = \hat{x}_{j,\mathrm{ATST}}(t) - z_{ji}(t)$, for $j = 1, \cdots, N$
    $x_{i,\mathrm{ATST}}(t) = \mathrm{tEM}\left([r_{1,i}(t), \cdots, r_{N,i}(t)]\right)$,
    $i = i + 1$,
  end for
  $y_{i,s}(t) = \frac{x_{i,\mathrm{ATST}}(t) - x_{i,\mathrm{ATST}}(t-\tau)}{\tau}$,
  $y_{i,ATST}(t) = \frac{\hat{y}_{i,\mathrm{ATST}}(t) + m_i y_{i,s}(t)}{1+m_i}$,
  $t = t + \tau$.
end for

4. Robustness of ATST Time Scale

To verify that the ATST time scale is robust to the anomalies of interest, we will assess three primary criteria: frequency stability, fractional frequency evolution, and phase evolution. The frequency stability of the new time scale should be unaffected and maintain similar performance in comparison to the AT1 oracle. If the ATST matches the performance of AT1 oracle, then it can be concluded that it is sufficiently robust to the tested anomalies. There should be no observable jumps or outliers in the frequency or phase of the time scale at any point in time. Measurement noise is expected to place a higher constraint on the frequency stability of the time scale, but anomalies in the links should not add to that constraint. Before presenting the results, the setup of the simulated data is explained and justified for the application of a swarm of nanosatellites.

4.1. Simulated data

Simulated clocks are used for the realization of the time scale to more easily mimic the case of a swarm of fifty nanosatellites. The characteristic noise models for potential clocks to be used in the swarm are defined using the specifications of Oven Controlled Crystal Oscillators (OCXO). The simulation methodology and how it can be extended to other types of clocks is explained in Appendix B. The noise variance levels of each simulated OCXO clock were varied randomly to obtain unique stochastic behaviours. The size and power consumption of the individual clocks is important for space-based timing technologies, especially for those to be used in nanosatellites. As a result, the OCXO is preferred for large-scale nanosatellite constellations because it achieves these requirements at a lower price than Cesium or Rubidium atomic clocks that could potentially fit in nanosatellite applications. However, clocks installed in

nanosatellites are more susceptible to anomalies due to limitations on the available protection. OCXO clocks are sensitive to their environment and have a more restrictive stability performance. Therefore, an ensemble of similar OCXO clocks is a good example of an application that needs a robust time scale algorithm.

A specific interval between measurements is fixed based on the mission of the swarm. For this work, the interval chosen is 10 s to be compatible with scientific observations of short-duration events. This results in many measurements being made over a day, hence increasing the likelihood of experiencing anomalies. The proportion of anomalies is then assumed to be one phase jump per clock, one frequency jump per clock, and one anomaly per inter-satellite link at random times over a 6 hour simulation period. To investigate each type of anomaly independently, three separate simulations are made, each with only one type of anomaly occurring in the swarm. A final simulation is also conducted to analyze the effects of a mixture of the investigated anomalies occurring randomly in the swarm.

4.2. Metrics

The phase evolution of the time scales is a visualization of their equivalent virtual clocks. To obtain the phase of the time scales we simply take the difference between the output of the BTSE and the simulated clock data

$$h_i(t) - x_{i,E}(t) = h_i(t) - (h_i(t) - h_E(t)) = h_E(t). \tag{32}$$

The fractional frequency of the time scales is obtained by taking the first difference of the phase

$$y_E(t) = \frac{h_E(t) - h_E(t-\tau)}{\tau}. \tag{33}$$

For clarity purposes, the frequency offsets are removed from the phase of the time scales by subtracting the averages of their frequencies over the whole simulation period $\bar{y}_E$

$$h_E(t) := h_E(t) - t\bar{y}_E. \tag{34}$$

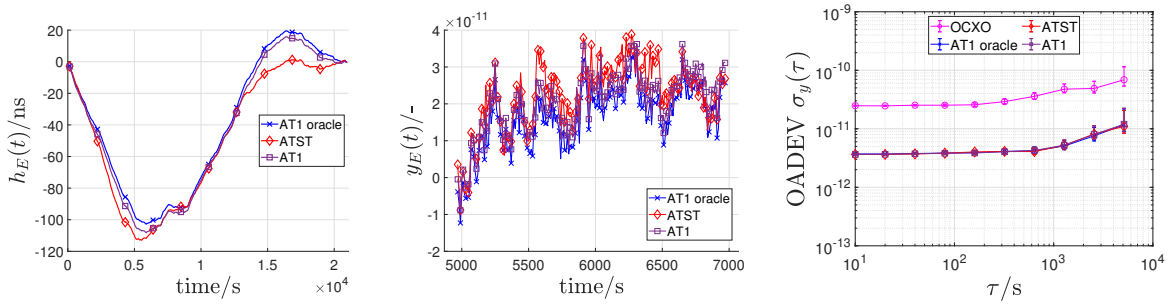
Finally, the OADEV is computed according to the standard formula for a given set of phase samples [25, 26]

$$\sigma_y^2(\tau) = \frac{1}{2(M-1)} \sum_{t=0}^{M-2} \left(\frac{h_E(t+2\tau) - 2h_E(t+\tau) + h_E(t)}{\tau} \right)^2, \tag{35}$$

where $M-2$ realizations of the time scale phase are available, restricting the maximum sampling interval to $\tau = \frac{M-2}{2}$. Confidence intervals are included on the computed values of OADEV to indicate the increase in uncertainty of the OADEV as the number of samples decreases for higher sampling intervals [27].

The ATST time scale should be validated under nominal conditions before evaluating the responses to the anomalies of interest. Figure 2 displays the phase,

frequency, and frequency stability of the ATST algorithm alongside the AT1 algorithm without any anomalies.



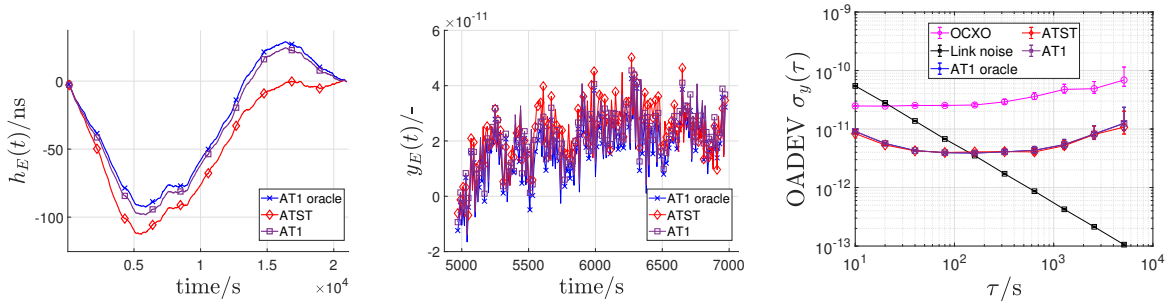
(a) Phase evolution. (b) Fractional frequency. (c) Overlapping Allan Dev.

Figure 2: Nominal performances of the ATST and AT1 oracle time scales using the above-mentioned ensemble of fifty OCXO clocks. The OADEV (c) compares the time scales to a singular OCXO clock to show the improvement in stability with respect to a single clock. Note that the ATST performance is very close to that of "AT1 oracle". Error bars on the OADEV are included to indicate the 68 % confidence interval for the simulation period used.

4.3. Measurement Noise

The effect of Gaussian measurement noise on the time scales should be presented before assessing the effects of link anomalies. Figure 3 displays the phase, frequency, and frequency stability of the ATST algorithm alongside the AT1 algorithm and measurement noise, referred to below as link noise. The magnitude of the noise variance is chosen such that it exceeds the Allan deviation of the individual clocks in the ensemble at the measurement interval of 10 s. This noise is assumed to be a white phase modulation noise. This is chosen to illustrate the impact of measurement noise on the two different time scales.

As shown in Figure 3c, the short-term OADEV of the time scales is affected according to the OADEV of the link noise. Since the measurement noise is injected in the time scale equation, there is a magnitude of noise variance at which the short-term stability of the time scale becomes worse than that of the individual clocks in the ensemble. At this point, the objective of the time scale to provide a virtual clock better than any individual clocks is not achieved. Anomalies in the measurement noise will effectively increase the variance of the link noise and hence, the short-term stability. As a result, a negligible level of measurement noise can become significant enough to deteriorate the time scale if anomalies in the measurements are not properly treated.



(a) Phase evolution. (b) Fractional frequency. (c) Overlapping Allan Dev.

Figure 3: Performances of the time scales using the above-mentioned ensemble of fifty OCXO clocks and a uniform noise on every satellite link with variance $\sigma_n^2 = 10^{-19}$.

Each time scale algorithm is comparable in nominal frequency stability with and without measurement noise. The measurement noise is shown to increase the short-term Allan deviation equivalently for both time scales. This means neither time scale is necessarily preferred for applications with Gaussian measurement noise.

4.4. Results in presence of anomalies

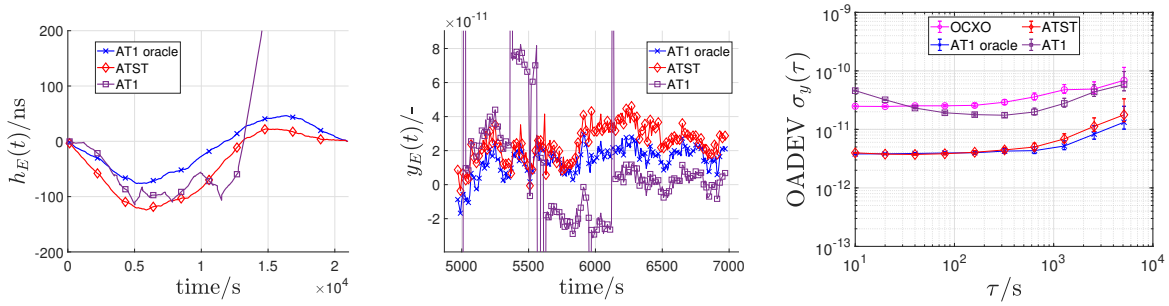
Anomalies in both the clocks and the measurements are analyzed in this section for both ATST and AT1 oracle. Note that AT1 has no specialized method of dealing with measurement anomalies. The AT1 oracle algorithm simply mitigates the anomalies by setting the weights of the clocks in the affected measurement links to zero, assuming it knows exactly which links are affected by anomalies. It is likely that several links are affected by anomalies a throughout the measurement process, so each unique link is simulated with an anomaly at a random point in time.

Figure 4 displays the phase, frequency, and ADEV of ATST, AT1, and AT1 oracle for the jump type anomalies in absence of measurement noise. The anomalies are investigated separately by introducing a single phase jump on each clock or a single frequency jump on each clock in two different simulations. The magnitudes are randomly distributed values with zero mean and a standard deviation of 100 ns, and 100 ns/s for the phase jump and the frequency jump, respectively. The resulting maximum values faced are then approximately equal to the 3σ values, i.e., ± 300 ns for phase jumps and ± 300 ns/s for frequency jumps. This is to illustrate a scenario that has a more significant impact on the time scale created with fifty clocks. The magnitudes tested in this article are presented so that the impact is visible on the ensemble of 50 clocks. For conciseness of this article, other simulations with smaller or larger anomaly magnitudes are available in the companion technical report [28].

Figures 4b and 4e demonstrate spikes in the frequency of AT1 without anomaly compensation. Nevertheless, both AT1 oracle and ATST are able to compensate the investigated phase jump and frequency jump anomalies. Outliers with similar magnitudes to the phase jumps are introduced on every unique link to generate the

link anomalies. This magnitude of the link anomalies in Figure 6 is chosen to cause the resulting contaminated measurement noise to exceed the original measurement noise with a variance of 10^{-19}.

Phase jumps cause intermediate frequency values to be outliers at the time of the anomaly and as a result, introduce a loss in frequency stability for AT1 without any treatment of anomalies. This is seen in the OADEV of the OCXO clock in Figure 4c. Phase jumps can also cause frequency jumps in the non-robust time scales because the frequency approximations are significantly affected by the phase jump, which consequently affect the following prediction of phase if the frequency time constant is not high enough. This is visible in Figure 4b, where the effect of a phase jump on the frequency of AT1 is identical to that seen in Figure 4e for a frequency jump anomaly at the same point in time. Link anomalies introduce unexpected measurements in the clock phase differences, also producing outliers in the frequency of the non-robust time scale. As a result, the short-term OADEV is increased for the AT1 time scale without anomaly mitigation.



(a) Phase evolution, Phase jumps.

(b) Fractional frequency, Phase jumps.

(c) Overlapping Allan Dev., Phase jumps.

(d) Phase evolution, Frequency jumps.

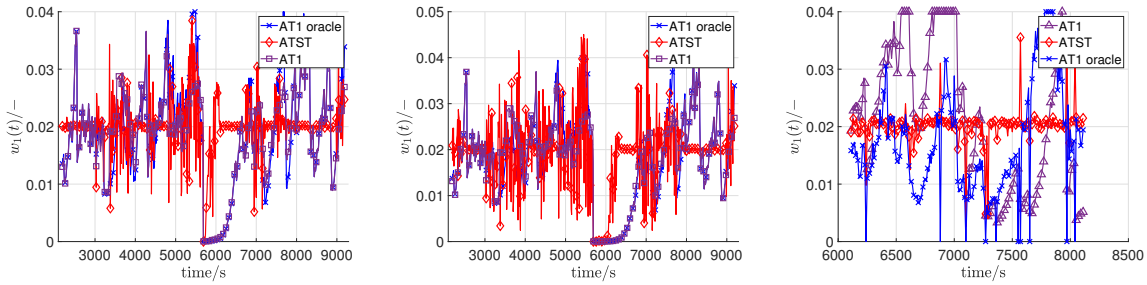
(e) Fractional frequency, Frequency jumps.

(f) Overlapping Allan Dev., Frequency jumps.

Figure 4: Time scale performance with phase jumps (top row) and frequency jumps (bottom row) each with order of magnitude randomly assigned in the interval -300 to 300 in the units of ns or ns/s, respectively.

Figure 5 demonstrates that the ATST weights differ to the AT1 weights while still managing to compensate the anomalies. In response to a frequency jump, the ATST weight remains low for some period before it is allowed to increase, which is likely related

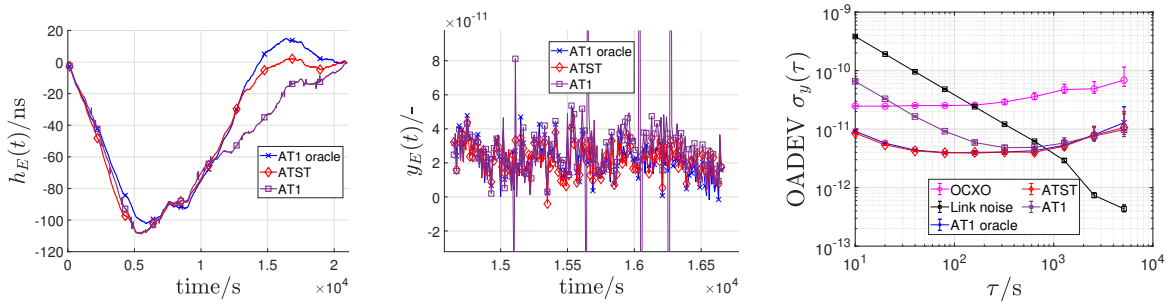
to the exponential filter on the frequency estimates. The weights are increased much more quickly in response to the other anomalies. This could be considered advantageous because the effective number of clocks is not reduced for a long period of time as is the case for the AT1 algorithm.



(a) Phase jumps. (b) Frequency jumps. (c) Link anomalies.

Figure 5: Weights for clock 1 as defined by the ATST and AT1 oracle time scale when every clock in the ensemble suffers from an anomaly at some point int time. Compared to Figure 1, the weights are seen to change more frequently due to anomalies occurring in other clocks in the ensemble.

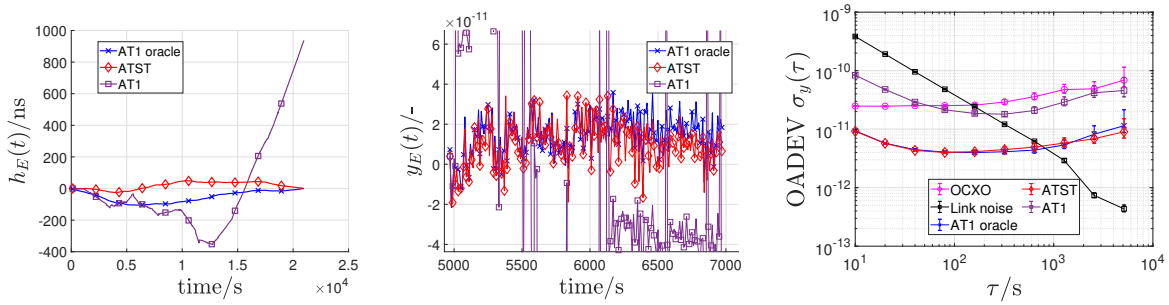
Figures 6a, 6b, and 6c present the robustness of the ATST time scale to link anomalies. This is a novel contribution because the measurement anomalies have previously been neglected in the design of time scales. The AT1 oracle time scale has a priori knowledge of the link anomalies and sets weights $w_j(t_a)$ and $w_i(t_a)$ to zero for a link anomaly on measurement $z_{ji}(t_a)$, removing all link anomalies perfectly. The AT1 oracle time scale provides the performance limit for the best method to deal with link anomalies, serving as a good basis for comparing the proposed ATST time scale. Since ATST remains just as stable as AT1 oracle in both the long and short terms, experiences no jumps in frequency, and retains continuity in phase evolution, it can be concluded that ATST is robust to all of these anomalies, as designed in section 3.



(a) Phase evolution, Link anomalies. (b) Fractional frequency, Link anomalies. (c) Overlapping Allan Dev., Link anomalies.

Figure 6: Time scale performance with link anomalies with randomized amplitudes for each jump with a maximum order of magnitude of 300 ns. The base measurement noise without anomalies is normally distributed with zero mean and a variance of $\sigma_n^2 = 10^{-19}$.

One advantage of using the ATST algorithm is that it does not require the establishment of some method to detect and identify link anomalies. The ATST time scale is robust to phase jumps, frequency jumps, and link anomalies, without differentiating between the anomalies. Assuming the Student's t-distribution as a model of the BTSE residuals results in an estimate of the mean that automatically considers outliers in the data when they are present. When there is no anomaly, the assumption of t-distributed residuals naturally simplifies to a Gaussian assumption. The basis of using the Student's t-distribution to assign weights encompasses all of these types of anomalies in the same assumption. This is supported in the results shown in Figure 7, where phase jumps, frequency jumps and link anomalies are all present throughout the simulation period. The performance in presence of all anomalies is proven to be robust because the ATST time scale maintains similar performance to the AT1 oracle. Robustness to all anomalies without need to differentiate between them is beneficial in the context of a swarm of satellites because it is not simple to identify the types of anomalies or their sources.



(a) Phase evolution, All Anomalies.

(b) Fractional Frequency, All Anomalies.

(c) Overlapping Allan Dev., All Anomalie.s

Figure 7: Time scale performance with a combination of all of the above anomalies with the same orders of magnitude but at different times throughout the simulation period.

The weaknesses of the ATST algorithm include identification of anomalies, minimum sample size, computational complexity, and heterogeneity of clock weights. The ATST algorithm does not differentiate between different types of anomalies so is more suitable for autonomous applications where anomalies are unavoidable in both clocks and measurements but identifying their source is not a priority. The BTSE residuals can only be modeled by the Student's t-distribution if enough samples (clocks) are available. Generally, time scales aim to use a high number of clocks to gain the best improvement in frequency stability, so sample size should not be a problem but raises questions for functionality with missing data. The efficiency for the EM algorithm with different sample sizes indicates a minimum number of samples (clocks) that are required to converge to the optimal estimator performance [29]. If observations are obtained exactly from a Student's t-distribution, a minimum of around 25 clocks should obtain a mean square error close to the optimal value. The difference between the actual

distribution of clock residuals and the Student's t-distribution is also not taken into account so this number could vary depending on the type and magnitude of anomalies experienced because the assumption of the Student's t-distribution is not a good enough fit for the data. Additionally, ATST computes the weights and the time scale in an iterative EM algorithm, which introduces computational complexity and may impact the ability to apply the time scale in real time applications. Lastly, by assigning the weights approximately equal for all clocks, the potential to benefit from different types of clocks is lowered. That does not necessarily mean the ATST cannot be adapted for mixtures of different clock types, but the current version is optimized for homogeneous clocks.

5. Conclusion

Critical space-based timing applications require a robust time scale. Specifically, a focus on nanosatellites places constraints on the clock technologies and any associated protections. For that reason, this paper has designed a new robust time scale algorithm that was tested for a homogeneous swarm of OCXO clocks. Assuming a Student's t-distribution for the model of the contaminated data has allowed a novel definition of the clock weights leading to a new time scale referred to as ATST.

The proposed ATST algorithm is capable of compensating for phase jumps, frequency jumps, and link anomalies by using the weights associated with the MLE for the mean of the t-distribution. By assuming perfect detection of anomalies, the AT1 oracle time scale was generated as a best case scenario to compare to ATST. All investigated types of anomalies are considered the same in the ATST algorithm, i.e., a source of undesired errors that can be modeled by a specific statistical distribution designed for outliers. The resulting time scale is therefore robust to each of the types of anomalies, even providing equivalent performance to the AT1 oracle time scale. This is expected to be useful in applications with no need to identify anomalies but to mitigate their effects autonomously. The new robust time scale is constrained to ensembles of homogeneous clocks for now, but a path forward for including diversity in clock types is possible and worth investigating.

Further work should be conducted on appropriately accounting for the benefits of different types of clocks in the ATST algorithm. Besides the extension to other types of clocks, certain other types of anomalies can be expected for space-based applications. For instance, the periodic temperature changes caused by orbits could introduce periodic changes in clock behavior. The Student's t-distribution may not be the ideal statistical model to use for this type of anomaly so other models should be explored. Another anomaly to explore is the impact of losing links between certain satellites or losing a satellite completely for the proposed time scale. Missing data is expected to be an important issue in the application of a satellite swarm so a robust time scale should also be designed to manage missing measurements of different scales. Finally, the application of the robust time scale to position determination and scientific observations should be

investigated.

Appendix A. Maximum Likelihood Estimator - Student's t-distribution

To understand how the ATST attains robustness, the Student's t-distribution should be explained. The parameters of the Student's t-distribution are the location parameter μ, which is also the mean value, the scale parameter σ^2, and the shape parameter given by the number of degrees of freedom ν. The parameters σ^2 and ν are related to the variance of the distribution by

$$\mathrm{var}(X) = \sigma^2 \frac{\nu}{\nu - 2}, \tag{A.1}$$

for $\nu > 2$. The degrees of freedom parameter is directly related to the level of abnormality of the distribution. Indeed, as $\nu \to \infty$ the t-distribution approaches a Gaussian distribution. Conversely, low values of ν coincide with a probability distribution heavily impacted by outliers. Figure A1 illustrates this connection between the shape parameter and the normality of the data.

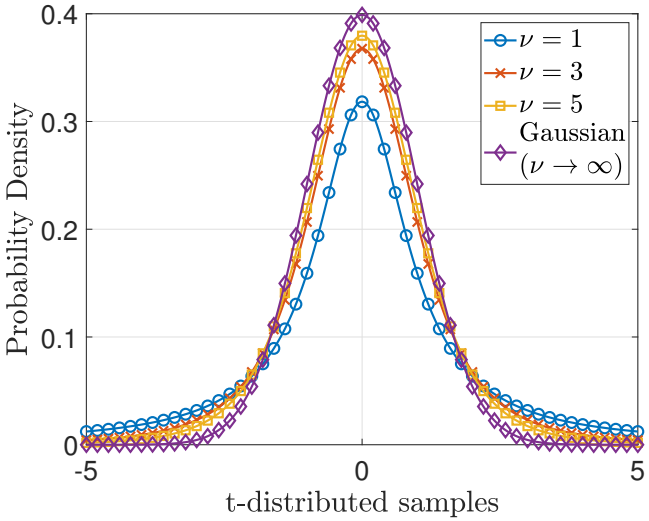


Figure A1: Examples of the Student's t-distribution probability density functions for different numbers of degrees of freedom ν. The distribution converges to a Gaussian distribution for infinite value of ν. Lower values of ν correspond to heavier probabilities in the tails, hence a greater proportion of contaminated data.

The nature of the shape parameter allows the estimates of the mean and variance to simplify to the normal case if the data is not contaminated with outliers. This is of interest because a robust time scale should not sacrifice performance in the nominal case to mitigate anomalies. The probability density function (PDF) of the Student's t-distribution has a specific form that assigns a larger probability of outliers occurring compared to other statistical models. An MLE is defined based on knowledge

of a likelihood function that fits the observations being made. In this work, the Student's t-distribution is used to model the distribution of clock data contaminated with anomalies. The likelihood of a sample distributed according to the univariate Student's t-distribution is defined as follows [16]:

$$L(\mathbf{z};\mu,\sigma^2,\nu)=\prod_{i=1}^{N}p(z_i;\mu,\sigma^2,\nu)=\prod_{i=1}^{N}\frac{1}{\sqrt{\nu\pi\sigma^2}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left[1+\frac{1}{\nu}\left(\frac{z_i-\mu}{\sigma}\right)^2\right]^{-\frac{(\nu+1)}{2}}, \tag{A.2}$$

with $\mathbf{z}=(z_1,...,z_N)^T$. The parameters μ, σ^2, and ν are the mean, scale, and shape parameters, respectively. The MLE for each of the parameters μ, σ^2, and ν aims to identify the values of those parameters that maximize the likelihood (A.2) for a given sample $\mathbf{z}$:

$$\left[\hat{\mu},\hat{\sigma}^2,\hat{\nu}\right]^T=\arg\max_{\mu,\sigma^2,\nu}\left\{L(\boldsymbol{z};\mu,\sigma^2,\nu)\right\}. \tag{A.3}$$

To simplify the derivations, it is usual to derive the expression of the MLE by minimizing the negative log-likelihood [30]

$$\left[\hat{\mu},\hat{\sigma}^2,\hat{\nu}\right]^T=\arg\min_{\mu,\sigma^2,\nu}\left\{-\log L(\boldsymbol{z};\mu,\sigma^2,\nu)\right\}. \tag{A.4}$$

In the case of the univariate Student's t-distribution, the MLEs of the unknown parameters are the solutions to the following equations

$$\frac{\partial\log(L)}{\partial\mu}=\frac{\nu+1}{\sigma^4}\sum_{i=1}^{N}\frac{z_i-\mu}{\nu+\left(\frac{z_i-\mu}{\sigma}\right)^2}=0,\quad \frac{\partial\log(L)}{\partial\sigma^2}=\frac{1}{\sigma^2}+\frac{\nu+1}{\sigma^4}\sum_{i=1}^{N}\frac{(z_i-\mu)^2}{\nu+\left(\frac{z_i-\mu}{\sigma}\right)^2}=0, \tag{A.5}$$

$$\frac{\partial\log(L)}{\partial\nu}=\phi\left(\frac{\nu}{2}\right)-\phi\left(\frac{\nu+1}{2}\right)+\sum_{i=1}^{N}\left[\frac{\nu+1}{\nu+\left(\frac{z_i-\mu}{\sigma}\right)^2}-\log\left(\frac{\nu+1}{\nu+\left(\frac{z_i-\mu}{\sigma}\right)^2}\right)-1\right]=0, \tag{A.6}$$

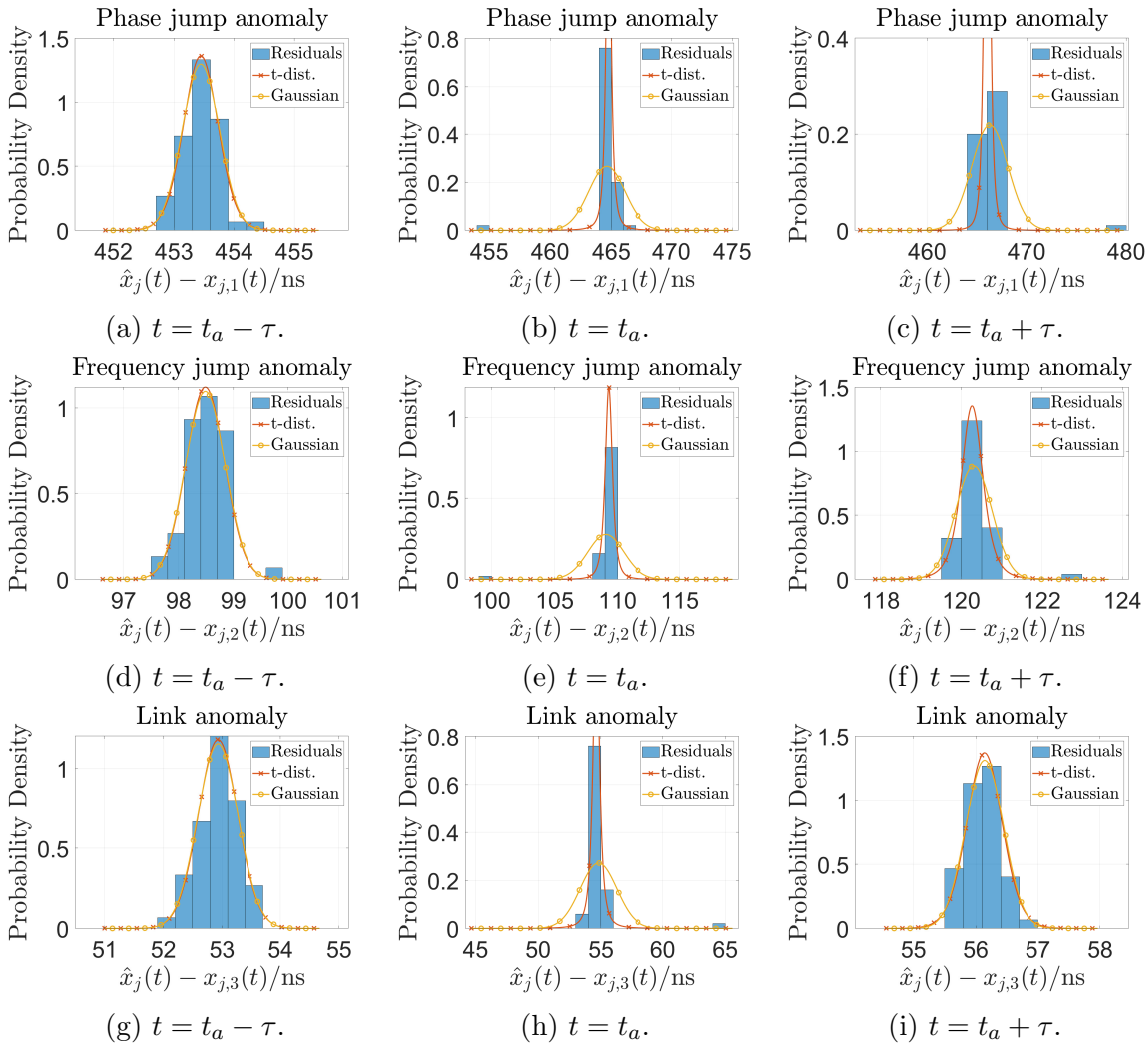
where $\phi(x)=\psi(x)-\log(x),\quad x>0$, and the digamma function $\psi(x)$ is given by:

$$\psi(x)=\frac{d}{dx}\log\left[\Gamma(x)\right]=\frac{\Gamma'(x)}{\Gamma(x)}. \tag{A.7}$$

Appendix A.1. Statistics of Anomalous Clocks

Figure A2, which shows the histograms of the BTSE residuals with phase jumps, frequency jumps, and outliers on the measurement links. The figure is divided into columns by time epochs: before, during, and after the occurrence of an anomaly t_a. The rows of Figure A2 are separated by type of anomaly. The distributions are plotted using the corrupted clock as the common reference for all the measurements. This shows that even with an identical anomaly on all phase difference measurements, the inclusion

of $r_{ii}(t)$ causes the distribution to still be affected by an outlier. As expected, the BTSE residuals at the time immediately before an anomaly are well approximated by both a Gaussian distribution and the t-distribution. Hence, we expect a good performance for the MLE of the t-distribution in the nominal case. At the epoch of the anomalies, there are outliers in the residuals, corresponding to the magnitude of the jumps. For these histograms, fifty OCXO clocks are simulated with a jump on a single clock at $t = t_a$ with a magnitude of 10^{-8}. The units of the jumps are compatible with the relevant states, i.e., 10 ns for phase and measurement jumps, and 10 ns/s for a jump in the fractional frequency y_i.



(a) $t = t_a - \tau$. (b) $t = t_a$. (c) $t = t_a + \tau$.

(d) $t = t_a - \tau$. (e) $t = t_a$. (f) $t = t_a + \tau$.

(g) $t = t_a - \tau$. (h) $t = t_a$. (i) $t = t_a + \tau$.

Figure A2: Student's t PDFs are shown to model the statistics of the residuals with and without an outlier on one of the measurement links.

The outlier in the measurement only occurs at $t = t_a$, so we see that the distribution returns to a Gaussian distribution at the time after the link anomaly. We observe the phase jump resulting from the frequency jump is also 10 ns, but this outlier does not disappear at the following time epoch because the frequency has changed for the

anomalous clock. Lastly, the phase jump outlier switches sides at the following epoch due to the prediction error over that interval ($\tau = 10$ s) being significantly different than from the now-shifted phase. Nevertheless, the t-distribution provides a good fit for each of the resulting distributions.

Appendix B. Simulated clocks

The simulation of OCXO clocks has been conducted based on the phase noise of a typical OCXO clock. The resulting simulator uses characteristics of the Allan Variance (AVAR) for the typical clock and generates the phase of new clocks such that the AVAR remains similar. Specifically, the slopes of the AVAR over certain sampling intervals are linked to the different types of noises experienced in oscillators by some constant term h_α, where α specifies the type of noise [25, 27]. A summary of the general types of noise experienced by an oscillator are:

- White Phase Modulation (WPM), $\alpha = 2$,
- Flicker Phase Modulation (FPM), $\alpha = 1$,
- White Frequency Modulation (WFM), $\alpha = 0$,
- Flicker Frequency Modulation (FFM), $\alpha = -1$,
- Random Walk Frequency Modulation (RWFM), $\alpha = -2$.

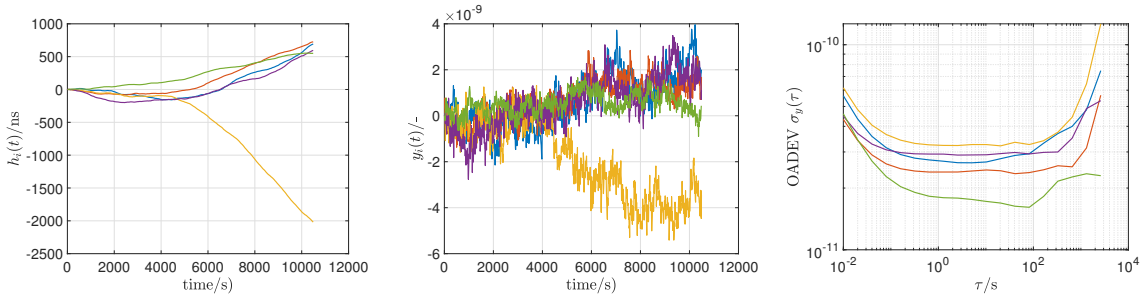
Some of these noises are more dominant than the others so the AVAR does not always display all of them. Nevertheless, all of the above values of α are used in the clock simulator. The power spectral density for clock phase can be written as a sum of each of the different noise types

$$S_x(f) = \frac{S_y(f)}{(2\pi f)^2} = \frac{1}{(2\pi f)^2} \sum_{\alpha=-2}^{2} h_\alpha f^\alpha = \sum_{\beta=-4}^{0} g_\beta f^\beta, \tag{B.1}$$

where $g_\beta = \frac{h_\alpha}{(2\pi)^2}$ and $\beta = \alpha - 2$. The variance associated with each clock noise is then computed

$$Q^d(\beta) = \frac{g_\beta}{2(2\pi)^\beta \left(\tau_0^{\beta+1}\right)}. \tag{B.2}$$

The clock noises are then generated independently according to the discrete generation method highlighted in [31]. Finally, the noise terms are summed to produce a simulation of an oscillator that has the same noise characteristics as the original source of the h_α coefficients. Slight variability has been added to the variance $Q^d(\beta)$ for each simulated clock to ensure that they do not have identical frequency stability. The simulator can similarly use the typical OADEV of any other type of clock to replicate several simulated clocks with similar performance. Figure B1 shows the phase, frequency, and OADEV for a small collection of simulated OCXO clocks. It is clear that they each have their unique characteristic noise but maintain the typical performance for an OCXO clock.



(a) Phase evolution. (b) Fractional frequency. (c) Overlapping Allan Dev.

Figure B1: Examples of 5 simulated OCXO clocks and the diversity of the clock behaviours.

References

[1] M. Weiss and T. Weissert, "AT2, A New Time Scale Algorithm: AT1 Plus Frequency Variance," *Metrologia*, vol. 28, pp. 65–74, 01 1991.

[2] M. A. Weiss and T. P. Weissert, "Sifting through nine years of NIST clock data with TA2," in *Proc. 7th European Frequency and Time Forum*, (Neuchatel, Switzerland), pp. 199–210, 03 1993.

[3] B. Rueff, J. Davis, and E. Pardo, "The development of a clock algorithm at NPL," in *Proc. 14th European frequency and time forum*, (Torino, Italy), pp. 398–402, 3 2000.

[4] C. A. Greenhall, "Forming stable timescales from the Jones–Tryon Kalman filter," *Metrologia*, vol. 40, p. S335, 06 2003.

[5] M. J. Coleman and R. L. Beard, "Autonomous clock ensemble algorithm for GNSS applications," *NAVIGATION*, vol. 67, pp. 333–346, 06 2020.

[6] L. A. Breakiron, "A Kalman Filter for Atomic Clocks and Timescales," tech. rep., NAVAL OBSERVATORY WASHINGTON DC, 01 2002.

[7] S. Stein and R. Filler, "Kalman filter analysis for real time applications of clocks and oscillators," in *Proceedings of the 42nd Annual Frequency Control Symposium, 1988.*, pp. 447–452, 1988.

[8] J. Brown, Kenneth R., "The Theory of the GPS Composite Clock," in *Proc. of the 4th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1991)*, (Albuquerque, NM), pp. 223–242, 09 1991.

[9] J. Levine, "Invited Review Article: The statistical modeling of atomic clocks and the design of time scales," *Review of Scientific Instruments*, vol. 83, p. 021101, 02 2012.

[10] L. Galleani and P. Tavella, "Detection of atomic clock frequency jumps with the kalman filter," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 3, pp. 504–509, 2012.

[11] L. Galleani and P. Tavella, "Characterization of atomic clock anomalies in the dynamic allan variance domain," in *2013 Joint European Frequency and Time Forum International Frequency Control Symposium (EFTF/IFC)*, pp. 645–648, 2013.

[12] L. Galleani and P. Tavella, "Robust detection of fast and slow frequency jumps of atomic clocks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 2, pp. 475–485, 2017.

[13] C. Zucca and P. Tavella, "A mathematical model for the atomic clock error in case of jumps," *Metrologia*, vol. 52, pp. 514–521, 2015.

[14] C. Zucca, P. Tavella, and G. Peskir, "Detecting atomic clock frequency trends using an optimal stopping method," *Metrologia*, vol. 53, pp. 89–95, 2016.

[15] C. Trainotti, G. Giorgi, and C. Günther, "Detection and identification of faults in clock ensembles with the generalized likelihood ratio test," *Metrologia*, vol. 59, p. 045010, 07 2022.

[16] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics, Wiley, 1 ed., 03 2006.

[17] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 1 ed., 10 2018.

[18] P. Tavella and C. Thomas, "Comparative study of time scale algorithms," *Metrologia*, vol. 28, pp. 57–63, 01 1991.

[19] S. Stein, "Time scales demystified," in *Proc. IEEE International Frequency Control Symposium and PDA Exhibition Jointly with the 17th European Frequency and Time Forum*, (Tampa, FL, USA), pp. 223–227, 2003.

[20] G. Panfilo and E. F. Arias, "Studies and possible improvements on the EAL algorithm," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 57, pp. 154–160, 01 2010.

[21] G. Panfilo and F. Arias, "The Coordinated Universal Time (UTC)," *Metrologia*, vol. 56, p. 042001, 08 2019.

[22] F. Doğru, Y. M. Bulut, and O. Arslan, "Doubly reweighted estimators for the parameters of the multivariate t-distribution," *Communications in Statistics - Theory and Methods*, vol. 47, pp. 4751–4771, 03 2018.

[23] M. Hasannasab, J. Hertrich, F. Laus, and G. Steidl, "Alternatives to the EM algorithm for ML estimation of location, scatter matrix, and degree of freedom of the student t distribution," *Numerical Algorithms*, vol. 87, pp. 77–118, 09 2020.

[24] H. McPhee, J.-Y. Tourneret, D. Valat, P. Paimblanc, J. Delporte, and Y. Grégoire, "A Robust Time Scale Based on Maximum Likelihood Estimation," in *Proc. of the 54th Annual Precise Time and Time Interval Systems and Applications Meeting*, (Long Beach, California), pp. 61–75, 02 2023.

[25] E. Rubiola and F. Vernotte, "The companion of the Enrico's chart for phase noise and two-sample variances," 01 2022.

[26] W. Riley and D. Howe, "Handbook of frequency stability analysis," 07 2008.

[27] "IEEE standard definitions of physical quantities for fundamental frequency and time metrology–random instabilities," *IEEE Std 1139-2022 (Revision of IEEE Std 1139-2008)*, pp. 1–60, 2022.

[28] H. McPhee, J.-Y. Tourneret, D. Valat, J. Delporte, Y. Grégoire, and P. Paimblanc, "Technical Report associated with the paper entitled "Robust Time Scale for Space Applications using the Student's t-distribution" submitted to *Metrologia*," 2024. Available at: `http://perso.tesa.prd.fr/jyt/publis_fichiers/Tech_report_ATST_Anomaly_mag.pdf`.

[29] R. Piché, "Cramér-Rao Lower Bound for Linear Filtering with t-Distributed Measurement Noise," in *Proc. 19th International Conference on Information Fusion*, (Heidelberg, Germany), 2016.

[30] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall signal processing series, Englewood Cliffs, N.J: Prentice-Hall PTR, 1993.

[31] N. Kasdin and T. Walter, "Discrete simulation of power law noise (for oscillator stability evaluation)," in *Proceedings of the 1992 IEEE Frequency Control Symposium*, pp. 274–283, 1992.