THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut National Polytechnique de Toulouse (INP Toulouse)*

Présentée et soutenue le *15/11/2024* **par :**

Hamish MCPHEE

Autonomous and Robust Time Scale Algorithm for a Swarm of Nanosatellites

---

JURY

Audrey Giremus	Professeure des Universités	Rapporteure
François Vernotte	Professeur des Universités	Rapporteur
Frédéric Pascal	Professeur des Université	Examinateur
Esa Ollila	Professeur des Université	Examinateur
David Valat	Ingenieur CNES	Examinateur
Corinne Mailhes	Professeure des Universités	Examinatrice
Jean-Yves Tourneret	Professeur des Universités	Directeur de Thèse

---

École doctorale et spécialité :
MITT : Domaine STIC : Réseaux, Télécoms, Systèmes et Architecture

Unité de Recherche :
Institut de Recherche en Informatique de Toulouse (IRIT)

Directeur de Thèse :
Jean-Yves TOURNERET

Rapporteurs :
Audrey Giremus et *François Vernotte*

Résumé

Cette thèse s'intéresse à la construction d'une échelle de temps autonome et robuste aux erreurs d'horloge pour des essaims de satellites. Prévu pour une utilisation dans un essaim de nanosatellites, cette nouvelle échelle de temps appelée ATST (Autonomous Time Scale using the Student's T-distribution) peut traiter les anomalies dues aux imperfections des horloges et aux liens inter-satellites manquants dans un environnement hostile. Plus précisément, les types d'anomalies traités incluent les sauts de phase, les sauts de fréquence, un bruit de mesure élevé dans certains liens et d'éventuelles données manquantes. En calculant la moyenne pondérée des résidus issus de l'équation de l'échelle de temps de base (BTSE), la contribution des satellites avec des mesures anormales est réduite pour la génération de l'échelle de temps. Les poids attribués à chaque horloge sont basés sur l'hypothèse que les résidus de l'ensemble suivent une loi de Student, ce qui permet d'utiliser des méthodes d'estimation robustes à la présence d'éléments aberrants.

La performance de l'algorithme ATST est équivalente à celle de l'algorithme AT1 oracle, qui est une version de l'échelle de temps AT1 avec la capacité de détecter parfaitement toutes les anomalies dans des données simulées. Bien que l'algorithme n'ait pas de méthode de détection explicite, l'algorithme ATST affiche toujours un niveau de robustesse comparable à celui d'un détecteur parfait. Cependant, l'algorithme ATST est conçu pour un essaim avec de nombreuses horloges de types homogènes et est limité par une complexité numérique élevée. De plus, les anomalies sont toutes traitées de la même manière sans distinction entre les différents types d'anomalies. Malgré ces limitations identifiées, ce nouvel algorithme ATST représente une contribution prometteuse dans le domaine des échelles de temps grâce à la robustesse atteinte.

Une méthode de traitement des horloges ajoutées ou retirées de l'ensemble des horloges disponibles est également proposée dans cette thèse en conjonction avec la méthode ATST. La méthode obtenue préserve la continuité de phase et de fréquence de l'échelle de temps en attribuant un poids nul aux horloges concernées lorsque le nombre total d'horloges est modifié. Un estimateur des moindres carrés (Least Squares, LS) est présenté pour montrer comment les mesures des liens inter-satellites peuvent être traitées en amont pour réduire le bruit de mesure et en même temps remplacer les mesures manquantes. L'estimateur LS peut être utilisé avec une méthode de détection qui élimine les mesures anormales. Il peut alors remplacer les mesures supprimées par les estimations correspondantes.

Cette thèse étudie également les performances de l'estimateur du maximum de vraisemblance (MLE) pour les paramètres des lois de probabilités à queues lourdes, plus précisément pour la loi de Student et pour un mélange de lois gaussiennes. Les améliorations obtenues en supposant que ces lois sont effectivement à queues lourdes par rapport à l'hypothèse de la loi gaussienne sont démontrées avec les bornes de Cramér-Rao mal-spécifiées (MCRB). Les expressions obtenues des MCRB pour une loi de Student et un mélange de lois gaussiennes confirment que les lois à queues lourdes sont meilleures pour l'estimation de la moyenne en présence de valeurs aberrantes. Elles permettent également de montrer que l'estimation des paramètres des lois à queues lourdes nécessite au moins 25 horloges pour obtenir une erreur d'estimation proche de la MCRB correspondante, c'est-à-dire que l'estimateur atteigne son efficacité asymptotique.

Des propositions de pistes de recherche futures incluent le traitement des limitations de l'algorithme ATST concernant les types et le nombre d'horloges. Une nouvelle définition des pondérations des résidus issue d'une méthode d'apprentissage statistique utilisant des données d'apprentissage est envisageable grâce à l'utilisation des résidus de l'échelle de temps de base BTSE. Une autre piste de recherche est le traitement des anomalies transitoires qui pose actuellement problème pour l'algorithme ATST. Un traitement de ce type d'erreurs pourrait être envisagé avec un algorithme d'apprentissage statistique ou avec un estimateur robuste de la fréquence des horloges sur une fenêtre de données passées.

Mots clés: Estimation robuste, échelles de temps, détection des anomalies, bornes de Cramér-Rao mal-spécifiées.

Abstract

A new robust time scale algorithm, Autonomous Time scale using the Student's T-distribution (ATST), has been proposed and validated using simulated clock data. Designed for use in a nanosatellite swarm, ATST addresses phase jumps, frequency jumps, anomalous measurement noise, and missing data by making a weighted average of the residuals contained in the Basic Time Scale Equation (BTSE). The weights come from an estimator that assumes the BTSE residuals are modeled by a Student's t-distribution.

Despite not detecting anomalies explicitly, the ATST algorithm performs similarly to a version of the AT1 time scale that detects anomalies perfectly in simulated data. However, ATST is best for homogeneous clock types, requires a high number of clocks, adds computational complexity, and makes no distinction of anomaly types. Despite these identified limitations, the robustness achieved is a promising contribution.

The implementation of ATST includes a method that maintains phase and frequency continuity when clocks are removed or reintroduced into the ensemble by resetting appropriate clock weights to zero. A Least Squares (LS) estimator is also presented to pre-process inter-satellite measurements, reducing noise and estimating missing data. The LS estimator is also compatible with anomaly detection which removes anomalous inter-satellite measurements because it can replace the removed measurements with their estimates.

The thesis also explores optimal estimation of parameters of two heavy-tailed distributions: the Student's t and Bimodal Gaussian mixture. The Misspecified Cramér Rao Bound (MCRB) confirms that assuming heavy-tailed distributions handles outliers better compared to assuming a Gaussian distribution. We also observe that at least 25 clocks are required for asymptotic efficiency when estimating the mean of the clock residuals. The methodology also aids in analyzing other anomaly types fitting different distributions.

Future research proposals include addressing ATST's limitations with diverse clock types, mitigating performance loss with fewer clocks, and exploring robust time scale generation using machine learning to weight BTSE residuals. Transient anomalies can be targeted using machine learning or even a similar method of robust estimation of clock frequencies over a window of past data. This is interesting to research and compare to the ATST algorithm that is instead proposed for instantaneous anomalies.

Key words: Robust estimation, time scales, anomaly detection, signal processing, time and frequency transfer, misspecified bounds.

Remerciements

Il est important de reconnaître tous les membres qui ont rendu possible cette thèse grâce à leur soutien précieux. J'aimerais remercier tout particulièrement l'équipe d'encadrement, tant du côté académique qu'industriel, pour m'avoir énormément appris sur un sujet qui était nouveau et difficile à comprendre au départ. Je remercie mon directeur de thèse, Jean-Yves Tourneret, pour son aide précieuse et utile tout au long de la thèse, ainsi que pour la liberté qu'il m'a accordée dans la recherche des pistes pertinentes qui m'intéressaient avant tout. Merci également à David Valat, Jérôme Delporte, et Yoan Grégoire du CNES pour leur encadrement et leurs divers conseils lors de nos réunions régulières et à d'autres moments nécessaires. L'équipe du CNES a été central pour l'introduction au domaine du temps et de la fréquence, en particulier la compréhension des échelles de temps existantes. Je vous remercie spécialement pour vos efforts lors de nos réunions pour m'écouter et me comprendre en anglais, et pour votre patience et vos encouragements dans le développement de mes compétences en français. À tous mes encadrants, votre confiance en moi a été essentielle pour chaque succès pendant la thèse, et j'espère que ce manuscrit vous rendra fiers de nos travaux ensemble.

Je remercie le laboratoire TéSA d'avoir rendu possible cette thèse en fournissant un co-financement avec le CNES pour un projet très intéressant que j'ai eu plusieurs opportunités de présenter au sein du laboratoire et ailleurs. Merci particulièrement à la directrice de TéSA, Corinne Mailhes, pour son dévouement à offrir à tous les doctorants des conditions de travail agréables et stimulantes. Votre soutien direct pour notre avenir est également grandement apprécié.

Je remercie mes rapporteurs de thèse, Audrey Giremus et François Vernotte, pour le temps qu'ils ont consacré à la relecture de ce manuscrit ainsi que pour leurs remarques et conseils avisés. Je remercie également Frédéric Pascal et Esa Ollila d'avoir accepté d'être membres de mon jury de thèse.

Merci aussi aux membres de la communauté scientifique qui ont pu m'apprendre de nouvelles approches pour aborder mes problèmes grâce à leur expertise, leurs avis et leurs conseils au cours des échanges que j'ai pu avoir avec eux. Je pense en particulier à Lorenzo Ortega, Philippe Paimblanc, Stefano Fortunati, Olivier Besson, Frédéric Pascal, Sébastien Trilles, Daniel Medina, Manuele Dassié et plusieurs autres que j'ai eu le plaisir de rencontrer lors de conférences et d'ateliers divers.

I would like to thank my family for the support provided throughout my whole life to get to where I am now. I am grateful for your belief in my abilities and the encouragement for me to follow the path that I find the most stimulating. A special thank you to my parents that have worked hard to open up the possibilities for me to excel, without the life you made for me I don't think I could have had the dedication to finish the monumental task that is a PhD.

Muchísimas gracias to Andrea for cheering me on in the most difficult times and encouraging me to always do my best. Thank you also for listening to me talk about clocks and anomalies and weights and everything that I've learnt during these past three years. Last but not least, thank you to Leia who has coauthored some of this thesis by sitting on my keyboard and meowing some very useful suggestions.

Contents

Chapter 1

Context

The motivation of this thesis originates from a preliminary study by the French Space Agency (CNES): Nanosatellites pour une Observatoire Interférometrie Radio dans l'Espace (NOIRE) refers to a proposed mission to conduct low frequency radio interferometry using a swarm of nanosatellites [1]. The study of such a mission provides context for general applications that use satellites as distributed systems that make collaborative observations. That is, the particularity of a swarm of nanosatellites is the interesting aspect of the mission that can be applied to many other future scientific endeavors. A swarm in this context is considered as a collection of similar satellites (nano-sized refers to a mass from 1-10 kg) that maintain proximity with each other by having similar orbits. This differs from constellations that try to separate the satellites to increase coverage. In addition, regular inter-satellite communications are pivotal in the function of the swarm, whether it concerns positioning, timing, or correlation of measurements.

Although the broad topic of a swarm of nanosatellites is the core application, the requirements for the application of NOIRE are used as an example of specifications and potential difficulties faced in a swarm. The NOIRE mission is not the only mission that aims to perform radio astronomy observations in space, with other scientific agencies pursuing similar strategies. As radio interference is prevalent in the Earth environment in the frequency bands of interest, Lunar orbits have been chosen for the NOIRE mission.

Radio interferometry requires synchronization of the local clocks in an array of devices separated by some distance. Each device has an individual clock that does not necessarily count time in a synchronized manner with the other devices. For this reason, the design of reference times, otherwise known as time scales is pivotal in obtaining a stable and reliable method of synchronizing clocks in distributed systems, whether that be radio telescopes on Earth or a swarm of satellites in Lunar orbit. In short, a time scale takes the best qualities of an ensemble of clocks to act as a virtual reference clock that is continuous in phase and frequency as well as being more stable than any individual clock in the ensemble.

This chapter will first discuss the principles of radio interferometry and introduce the

potential issues that arise by applying this method of scientific observations in space. Certain requirements for radio interferometry and nanosatellites are introduced to give context for the objectives of this thesis. Since the NOIRE mission is limited by nanosatellite technologies, some key challenges and limitations faced by clocks onboard nanosatellites are also discussed. Such challenges will indicate the clocks of interest and new methods that must be researched in this thesis, i.e., robust methods of generating a time scale using only the clocks onboard nanosatellites. The requirement to depend solely on the onboard clocks makes the resulting time scale autonomous. This autonomy also suggests that compensation of anomalies should be automatic to ensure continuity of phase and frequency in the reference clock. The final contribution should be an autonomous and robust time scale that has better performance than individual satellite clocks, which supports correction of the onboard clocks by steering to the time scale.

By developing a novel solution to the challenge presented by NOIRE, several additional fields can obtain new outlooks on current methodologies. Potential contributions can arise in the fields of time-scale algorithms, robust estimation, autonomous navigation, and applied machine learning. Before such contributions can be properly identified and addressed, the state of the art shall be explored so that the requirements can be appropriately defined.

1.1 Space-based Radio Interferometry

The NOIRE mission was established as a result of collaboration between a series of predecessor studies supported by the European Space Agency (ESA) and from the Dutch teams at ASTRON, in which the common idea has been to deploy a swarm of nanosatellites that cooperate as nodes of a space-based radio interferometer [1]. Some of the earliest of these predecessor studies are Distributed Aperture Array for Radio Astronomy In Space (DARIS) and the Orbiting Low Frequency Array (OLFAR). Studies for the OLFAR mission identified a desired observation target of low-frequency electromagnetic radiation ($\leq$30 MHz) and investigated the potential orbit locations for at least 10 identical and scalable satellites [2, 3, 4]. Additional studies on a space-based radio observatory exist from institutes outside Europe such as the Solar Radio Imaging Array (SIRA) [5], the AERO-VISTA mission using a pair of cubesats [6], and the swarm of nanosatellites for Space Ultra-Low Frequency Radio Observatory (SULFRO) [7]. The principles behind radio interferometry in general should be properly studied to understand the potential challenges that will be faced by these missions in attempting space-based radio interferometry.

Radio interferometry is well-established on Earth in the form of Very Long Baseline Interferometry (VLBI), which uses an array of distributed radio telescopes to synthetically increase the aperture and observe long wavelengths of electromagnetic radiation. The existing Very Long Baseline Array (VLBA) was conceived to observe radio frequencies in the order of 327 MHz to 43 GHz [8]. New scientific discoveries can be achieved by observing the novel low-frequency radio waves that are below 100 MHz and go as low

as 100 kHz. Such low frequencies mean larger wavelengths, and consequently more widely spread VLBI arrays and a need for reduced radio interference. As is demonstrated in Figure 1.1, the distance between adjacent nodes relates directly to a difference in time of arrival (TOA) of the target wavefront. In fact, the difference in TOA (ΔTOA) is the principle observation made in VLBI that allows science based on radio astronomy to be conducted [9]. As a result, the precision of the different TOA is important in the reconstruction of the interferometric measurements.

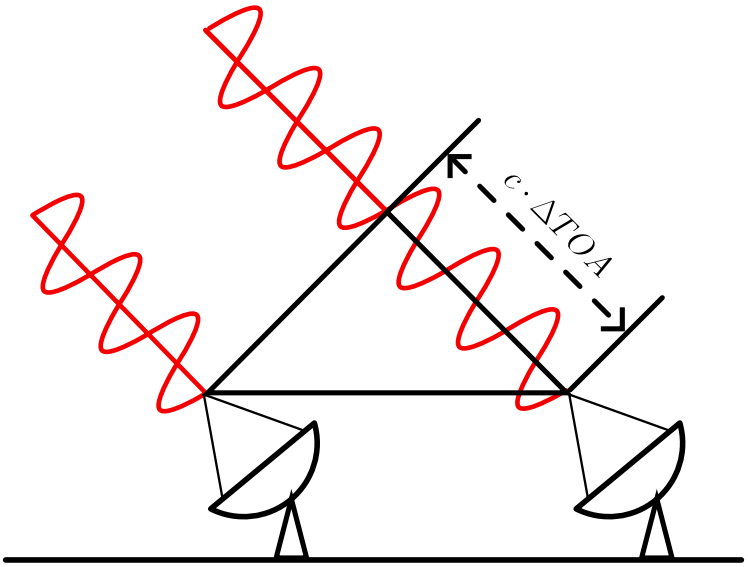


Figure 1.1: Illustration of the wavefront arriving to physically separated sensors with some difference in time of arrival ΔTOA.

Figure 1.2 demonstrates the same principle of VLBI applied to satellites. A key difference in the two different applications is that satellites are not necessarily fixed relative to some common reference or even with respect to each other. Conversely, the dishes fixed to the ground on Earth have well known positions and distances between each other. The resulting interferometry performed by the swarm of satellites will rely on both relative positioning and timing between the satellites. With a high number of satellites, a reasonable method of positioning would be based on the Differential Time of Arrival (DTOA) method used in positioning based on Global Navigation Satellites Systems (GNSS). This reveals another parameter in the swarm that relies on time precision, and hence another motivating factor to provide a reliable time reference.

The precision of TOA measurements are affected by unique and stochastic behaviors of the clocks responsible for the time stamps of the measurements. The clock technologies used in the VLBA are Hydrogen Masers, chosen for their good long-term stability [10], meaning that random fluctuations in the clock frequency over large time intervals are minimized.

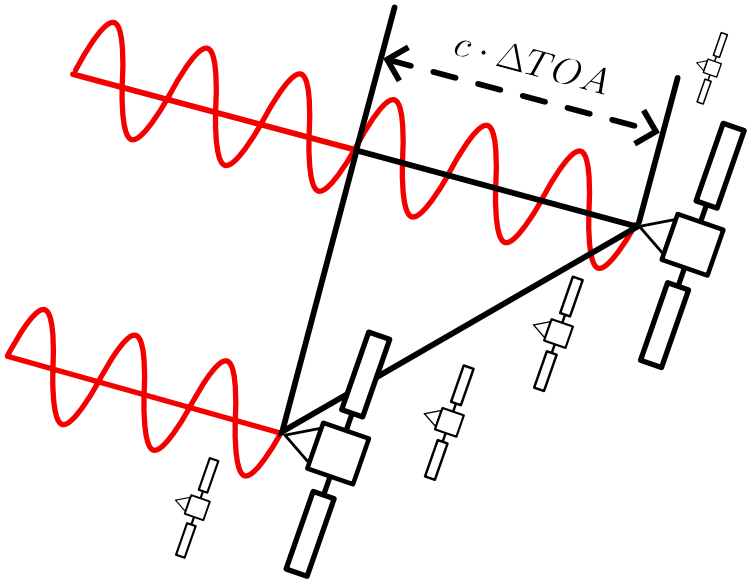


Figure 1.2: Illustration of the wavefront arriving to physically separated sensors with some difference in time of arrival ΔTOA.

As a result, the TOA will be unaffected by additional errors over long observation times due to inconsistencies between clocks in each radio observation device and the end product of radio sky images are not deteriorated.

Unfortunately, Hydrogen maser clocks cannot operate onboard nanosatellites due to size, weight, and power constraints. As a result, only certain clock technologies can be considered for space-based interferometry. These clocks are not necessarily sufficient to achieve the desired stability for correlation of coherent interferometric observations. The next section will introduce the different performance metrics for clocks and how the space-based technologies differ compared to those used in ground stations. If the clock technologies alone are not stable enough, they only need to be synchronized with respect to a time scale that achieves the desired stability. An envisaged time scale is a common reference that is synthetically generated using the timing information from every clock in the swarm of satellites.

Time scale algorithms are inherently designed to provide a reference time that is more stable than any individual clock used in the generation. The noises in each independent clock are essentially averaged down and the resulting time scale can expect an improvement in stability that is a function of the total number of clocks. Since a swarm of satellites will consist of many independent on-board clocks, the time scale is expected to make a significant gain in stability compared to the individual clocks. That is, neglecting any anomalies that may occur on the clocks and any accessibility issues with sharing timing

information throughout the swarm. Realistically, the space environment is not a place that allows us to neglect potential anomalies on either the clocks or the inter-satellite links that allow comparisons of clocks on different satellites.

1.2 Clock technologies

A clock is able to display the time thanks to an internal oscillator with a defined period of oscillations. If that period is stable, the time that passes can be measured by counting the peaks in the output voltage of the oscillator. The actual signal provided by an oscillator is a voltage $V(t)$ affected by stochastic noises [11]:

$$V(t) = V_0(t)(1 + \alpha(t)) \cos\left(2\pi\nu_0 t + \phi(t)\right), \tag{1.1}$$

where the nominal voltage $V_0(t)$ fluctuates according to the amplitude noise $\alpha(t)$, and the phase fluctuations are defined by $\phi(t)$. The nominal frequency ν_0 corresponds to the frequency that would result in perfectly uniform periodic oscillations, which would provide an ideal oscillator for timing purposes. However, deviations from the nominal frequency are always present due to internal clock noises that occur in a range of different types. The instantaneous frequency of the oscillator can be represented in terms of the phase fluctuations [12]:

$$\nu(t) = \nu_0 + \frac{1}{2\pi}\frac{d\phi}{dt}, \tag{1.2}$$

where this instantaneous frequency is typically presented as a fractional frequency $y(t)$ that describes a normalized deviation from the nominal frequency

$$y(t) = \frac{\nu(t) - \nu_0}{\nu_0} = \frac{1}{2\pi\nu_0}\frac{d\phi}{dt} = \frac{dx}{dt}. \tag{1.3}$$

The fractional frequency also functions as the rate of change of the time deviation $x(t)$ with respect to the nominal clock that oscillates with constant frequency. That is, the time difference is also a difference in the absolute phase of clock h_i and a perfect clock h_p

$$x_{i,p}(t) = h_i(t) - h_p(t) = \frac{\phi(t)}{2\pi\nu_0}, \tag{1.4}$$

where $\phi(t)$ has already been defined as the source of random phase noise in the timing of oscillator signal and $h_p(t)$ represents the absolute phase of an oscillator that is perfectly stable.

1.2.1 Standardized Time

The unit of the second defined in the International System of Units (SI) is based on a fixed value of the nominal frequency. The standard definition of a second was declared at

the 13th Conférence Générale des Poids et Mesures (CGPM) to be "the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the Caesium 133 atom". In 1997, it was clarified this is for an atom at rest at a thermodynamic temperature of 0 K [13]. This was further altered to the current definition at the 26th CGPM to "The second, symbol s, is the SI unit of time. It is defined by taking the fixed numerical value of the Caesium frequency $\Delta\nu_{Cs}$, the unperturbed ground-state hyperfine transition frequency of the caesium 133 atom, to be 9 192 631 770 when expressed in the unit Hz, which is equal to s^{-1}" [14]. The key modification being the focus on a fixed numerical value of the defining constant, $\Delta\nu_{Cs}$ (to aid reproducibility) [15].

Since the standard unit of the second is directly related to this specific atomic element, oscillators must then know their relative frequency offsets [16]. Examples of elements used for oscillators besides Caesium are Rubidium (Rb), Quartz (Qz), and Hydrogen microwave lasers (H masers). There is promising work towards a globalized weighted average to define the SI second using the potentially less uncertain optical clock technologies that are still being developed [16]. Even if an oscillator has a different nominal frequency to that which defines the second, the difference in phase due to the frequency offset can be compensated. However, stochastic variations in the phase and frequency of an oscillator will result in time deviations that cannot be predicted and cannot be measured directly. The next section explains the types of stochastic noises that can cause phase fluctuations in different types of oscillators.

1.2.2 Clock noises

A summary of the general types of noise experienced by an oscillator are listed below:

- 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), $\alpha = -2$.

Each noise is linked to a term h_α, where α specifies the type of noise. See Appendix A for a summary of the equations defined in [11, 17] that relate h_α to typical stability metrics such as, power noise spectra or two-sample variances. There are several types of two-sample variances that are useful to independently visualize each of the noise types, the generalized formula for two-sample variance is:

$$\sigma_y^2(\tau) = E\left[\frac{1}{2}(\bar{y}_2 - \bar{y}_1)^2\right], \tag{1.5}$$

where $\bar{y}_2$ and $\bar{y}_1$ are fractional frequencies averaged over an interval of time equal to τ seconds. The most basic and widely used two-sample variance is the Allan variance (AVAR) [18]. The estimate of the AVAR is obtained using the following equation

$$\sigma_y^2(\tau) = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\bar{y}_{k+1} - \bar{y}_k)^2 , \tag{1.6}$$

where M total samples of the fractional frequency are available and the subscript $k+1$ refers to fractional frequency measured τ seconds after the fractional frequency with sunscript k. Due to the effective averaging in the AVAR computation, the variance over large sampling intervals τ requires a long duration of time to collect sufficient samples and obtain an accurate approximation of the long term stability. Besides increasing the total number of samples, the confidence of the AVAR estimate can be improved by taking overlapping samples, i.e., use every possible pair of samples separated by the desired sampling interval. As demonstrated in Figure 1.3, the dependence of overlapping samples is neglected to increase the total number of samples that can be used for a fixed total number of samples.

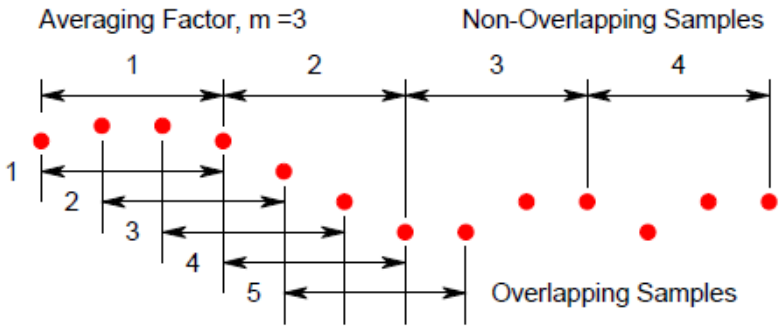


Figure 1.3: The difference between non-overlapping and overlapping samples for estimation of a two-sample variance [12].

The formula for computing the Overlapping AVAR (OAVAR) is then [12]:

$$\sigma_y^2(\tau) = \frac{1}{2m^2(M-2m+1)} \sum_{j=1}^{M-2m+1} \sum_{k=j}^{j+m-1} (\bar{y}_{k+m} - \bar{y}_k)^2 , \tag{1.7}$$

where the increment of k is now equal to the minimum measurement interval τ_0 and the sampling interval is some integer multiple of the minimum measurement interval $\tau = m\tau_0$. Generally, the stability of an oscillator is visualized in terms of deviation instead of variance, i.e., Allan Deviation (ADEV or OADEV for the overlapped version). Even after improving the confidence in the estimate by using overlapping samples, the estimate will

retain a significant level of uncertainty that should be included in the stability analysis. The confidence interval for estimating the OADEV is defined in [17] with specific equations for each different type of noise listed above. The equations for the confidence intervals are detailed in Appendix B. Throughout this work, the OADEV is used to observe the stability of clocks and time scales, where the uncertainty is displayed in the form of error bars. For generality purposes, the discussions of the stability will refer to the stability as ADEV although the overlapping form is used in the associated figures.

The ADEV is a useful metric to assess the types of noise on an oscillator as well as assess the absolute stability of a given clock. Regardless of the order of magnitude of the clock noises, the ADEV will follow a certain slope on the log-log plot over intervals of sampling time that correspond to specific types of noise. Figure 1.4 shows the regions and their corresponding slopes.

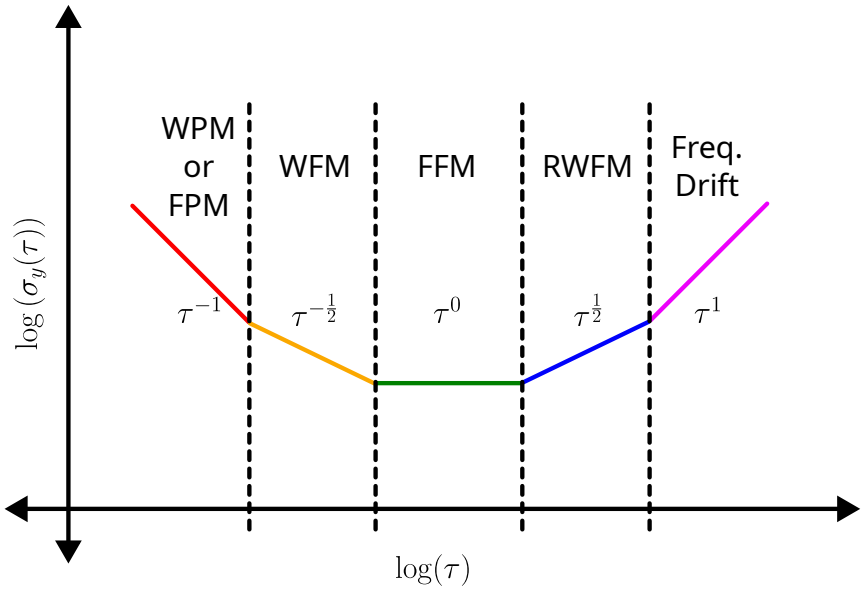


Figure 1.4: The relation between the slope of the ADEV and the types of noises experienced in oscillators [11, 12].

Depending on the type of oscillator some of the above noises are dominant over a larger range of sampling intervals so the ADEV does not necessarily display all of the noise types. Alternative two-sample variances are special cases of 1.5 for special cases of noise analysis. For example, the Hadamard variance (HVAR) helps visualize higher order frequency noises and removes the effect of frequency drift in the estimated variance.

1.3 Satellite-based Clock technologies

An in depth study has been presented in [6] to demonstrate the required clock stability and other characteristics such as jitter noise and temperature sensitivity for their mission. The AERO-VISTA mission [6] is intended to make observations of radio emissions at a frequency of 15 MHz and lower using interferometry, which aligns well with the range desired for NOIRE. The study has identified two key types of clocks for use in the pair of cubesats performing interferometry: the Chip Scale Atomic Clock (CSAC) and the Oven Controlled Crystal Oscillator (OCXO). The choices of these clocks are based on power and size compatibility with a small satellite payload as well as the clock stability requirements. The characteristic noise models for potential clocks to be used in the swarm are defined using the specifications of OCXO and CSAC clocks [19, 20], which then allow each type of clock to be simulated according to the methodology described in C.

1.3.1 Requirements

The clock technologies can potential be used by themselves if they meet the requirements for the mission, i.e., without the aid of a time scale as a reliable reference. However, it is clear that the anomalies that can occur due to the harsh environment will result in reduction in the performance of the individual clocks. This section will discuss the desired requirements for the NOIRE mission and can be extended to other examples of satellite swarms or space based interferometry.

Time Synchronization

The time synchronization requirement refers to the maximum allowable error between each of the clocks in the swarm and the common reference. Even without worrying about space based applications, the relative times of each clock in an ensemble will diverge over time if no steering is applied to re-synchronize the clocks. Since this thesis is focused on the design of the time scale algorithm, it is assumed that the virtual clock generated by the algorithm is used as the synchronization reference. That is, the time scale is generated first, then the onboard clocks are gradually steered such that their phase and frequency match that of the time scale and hence benefit from the improved performance of the time scale. For that reason, the synchronization accuracy of the designed time scale with respect to a perfect reference clock is analyzed to observe the synchronization performance of steered clocks. When moving from the simulated scenarios to real data, the perfect reference is replaced by some other reference that may be available and is significantly reliable and stable.

The International Telecommunications Union (ITU) presents recommendations for clock systems in coherent networks [21]. Although the scientific objectives are needed to specify an exact requirement for the synchronization accuracy, the general limits that are considered in standardized recommendations can be a good first target for assessing the performance of the proposed time scale. The metrics for assessing synchronization performance

are Maximum Time Interval Error (MTIE) and Time Deviation (TDEV), which are related to the offsets of individual clock systems from the Coordinated Universal Time (UTC). In the context of an autonomous time scale, the role of UTC is replaced by the new time scale. This justifies the similarity between the requirements for coherent synchronization of clocks in the swarm and the requirements suggested by the ITU. The MTIE refers to the largest change in clock time with respect to a specific reference over a given time interval τ [22], i.e.,

$$\mathrm{MTIE}(\tau) = \max_{t_0>0} \left\{ \max_{t_0-\tau<t<t_0} \{x_{E,p}(t)\} - \min_{t_0-\tau<t<t_0} \{x_{E,p}(t)\} \right\}, \tag{1.8}$$

for the offset of the designed time scale E from the perfect reference p. The above expression is the theoretical formula for MTIE but in reality, over long periods of time it is complicated to be certain about the maximum value. For that reason the confidence interval of the MTIE is limited to 95% [12]. The TDEV is an extension of the Modified ADEV (MDEV), denoted as ${}^{M}\sigma_y(\tau)$ which is designed to split the effects of white and flicker phase noises. This is important for observing the phase stability instead of frequency stability. The reader is invited to consult Appendix A for details on the computation of the MDEV, which is then converted to TDEV according to the following [11]:

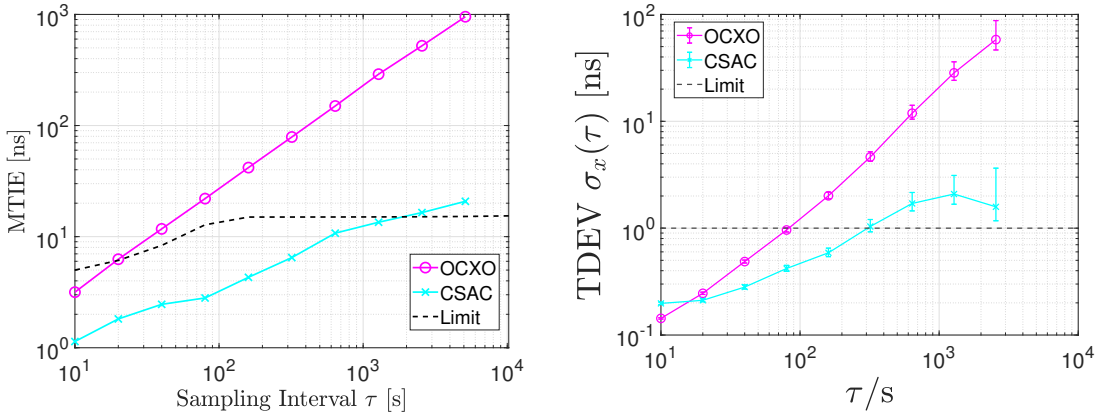
$$\sigma_x(\tau) = \frac{\tau}{\sqrt{3}} {}^{M}\sigma_y(\tau). \tag{1.9}$$

The constraint that is placed on locked clocks is presented in the following table for different intervals. Locked clocks refers to clocks that are steered to the time scale, so there is some feedback control to make sure the error remains small.

MTIE limit (ns)	TDEV limit (ns)	Sampling Interval (ns)
4	1	$0.1 < \tau \leq 1$
$0.11\tau + 3.89$	1	$1 < \tau \leq 100$
$3.75 \times 10^{-5}\tau + 15$	1	$100 < \tau \leq 3 \times 10^4$
$3.75 \times 10^{-5}\tau + 15$	$3.33 \times 10^{-5}\tau$	$3 \times 10^4 < \tau \leq 3 \times 10^5$
$3.75 \times 10^{-5}\tau + 15$	10	$3 \times 10^5 < \tau \leq 4 \times 10^5$
30	10	$\tau > 4^5$

Based on the above requirements, the speed at which the clocks are steered could be designed. The upper limits for the synchronization metrics are illustrated in Figure 1.5 below

alongside the achievable performance for the OCXO and CSAC clocks. It is clear that the suggested upper limits are not respected by the clock technologies by themselves. The true restrictions on the required accuracy could be relaxed depending on the swarm requirements, opening the range of acceptable sampling intervals. Otherwise, through steering to the designed time scale, the MTIE and TDEV will be reduced, therefore attaining an improved range of appropriate sampling intervals for even these requirements.



(a) Maximum Time Interval Error.

(b) Time deviation.

Figure 1.5: The achievable synchronization performance for the simulated OCXO and CSAC clocks alongside the recommended upper limit for telecommunications applications.

Noticeably, the CSAC clock performs much better in terms of synchronization and can even satisfy the MTIE requirements over long intervals. The occurrence of anomalies on the clock will cause the performance of the individual clocks to be perturbed, however, if the time scale algorithm appropriately mitigates the anomalies then the steered clocks can not only achieve improved performance but avoid the negative effects of the anomalies. A specific sub-criteria for the time scale to ensure it has properly mitigated the effects of anomalies is the continuity.

Phase and Frequency Continuity

The requirement of phase and frequency continuity specifically refers to the maximum allowable instantaneous change in either of these parameters for the time scale. That is, a discontinuous clock (or time scale) would result in a significant increase in the short term MTIE and a deterioration of the frequency stability, so the continuity requirement can also be linked to the MTIE and the ADEV. Anomalies can cause the phase or frequency of an individual clock to change instantaneously to another value, these types of anomalies are hereon referred to as jumps. Since an autonomous time scale is built using the clocks in the

swarm, the jumps can propagate into the time scale if it is not sufficiently robust. Missing data also interferes with the continuity of the time scale specifically due to the reduced performance of the ensemble time with fewer clocks.

Without continuity in the individual clocks or the time scale, the coherence of the distributed measurements is at risk of being negatively impacted because more or less time is thought to have passed since the previous correlation. The nature of the jump anomalies will be presented in Section 1.3.2. Such anomalies break the requirement of continuity by default, so the time scale must guarantee continuity in the presence of anomalous and missing data to provide a reference that does not risk errors in the correlation. Continuity can be confirmed by observing the phase and frequency evolution of the time scale over a window that contains both nominal and anomalous clock data. The time scale should maintain a similar magnitude of instantaneous changes before and after the occurrence of the anomaly. Depending on the type of clocks used to generate the time scale this magnitude could vary, hence the requirement of continuity is simply confirmed by the trend of the clock phase and frequency.

Frequency Stability

The requirements of clock stability are derived based on a maximum coherence loss of 15% for the scientific measurements, this is equivalent to a limit on the phase difference between two satellites due to frequency instability. Coherence loss is quantified for an integration time τ by $C(\tau)$ [23]

$$C(\tau) = \left| \frac{1}{\tau} \int_0^\tau \exp\left(i\phi_j(t) - i\phi_k(t)\right) dt \right|, \tag{1.10}$$

where $C(\tau) = 1$ means no loss in coherence. The worst-case loss for the AERO-VISTA mission occurs when measuring 15 MHz over 100 s, corresponding to the longest integration time and highest frequency. The NOIRE mission does not necessarily have the same worse case coherence loss because the scientific objectives still need to be confirmed pending parallel studies, only an upper limit of 100 MHz is stated [1]. This maximum observable radio frequency "without significant loss" can be related directly to the ADEV over a sampling interval corresponding to the integration time [24, 25]

$$f_{obs} \leq \frac{1}{2\pi\tau\sigma_y(\tau)}. \tag{1.11}$$

The corresponding constraint on the ADEV for any sampling interval τ would then be

$$\tau\sigma_y(\tau) \leq \frac{1}{2\pi f_{obs}}, \tag{1.12}$$

$$\tau\sigma_y(\tau) \leq 1.59 \times 10^{-9}. \tag{1.13}$$

Substituting any value of τ would coincide with that interval being the time over which transient events are able to be observed in the scientific observations. Therefore according

to the final application of the NOIRE mission, the ADEV can be optimized for the intended sampling interval. Nevertheless, there is nothing to lose ensuring that the condition (1.13) is satisfied for a wide range of sampling intervals.

A typical metric used to discussed clock performance in a general manner uses the concept of seconds lost (in synchronization) over a certain interval of time. The value of $\tau\sigma_y(\tau)$ provides exactly that metric for an interval of τ seconds. That is, the constraint that allows observation up to 100 MHz means the timing is permitted to lose at most 1.59 ns over any measurement interval. To understand if the OCXO and CSAC clocks can achieve this required stability, simulated ADEV curves for those types of clocks are displayed alongside the limiting curve for ADEV in Figure 1.6. The ADEV and OCXO clocks is estimated using typical values of the characteristic noise experience by each type of clock. More details on the simulation of these types of clocks are found in Appendix C.

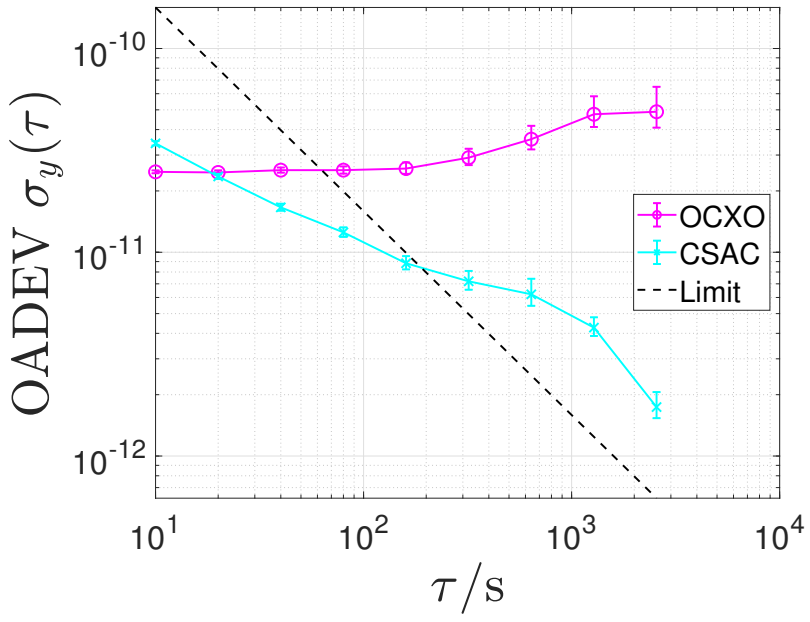


Figure 1.6: The limitation on the clock ADEV for a range of different sampling intervals. The intervals at which the individual clocks ADEVs are lower than the limit are the possible intervals that can be used for observing the desired radio frequency.

The clocks can both satisfy the requirement of maximum ADEV (1.13) for some integration times. However, each type of clock has a maximum integration time where the coherence requirement is broken. An optimally designed time scale will have an improved frequency stability (reduced by a factor $1/\sqrt{N}$) compared to each individual clock, meaning onboard satellites clocks can be steered to the time scale and obtain the required perfor-

mance for the integration times where the stability was not sufficient. If anomalous clocks are used to generate the time scale and the anomalies are not properly compensated, the benefit of steering to a stable and continuous time scale is lost. This motivates the design of a robust time scale to ensure the requirements are still satisfied by the time scale in presence of anomalies while extending the range of possible sampling intervals that achieve the required ADEV.

When analyzing the effects of anomalies on the frequency stability of the time scale, the occurrence rate of the anomalies is important. For example, if an anomaly occurs once over a period of one day but the sampling frequency provides a measurement every 10 seconds, the anomaly is averaged out in the estimation of the ADEV although the continuity of the time scale can still be broken. In the above figure and all following analyses, the total duration of simulated data is taken as a 3 hour interval. Observing the effects of a single anomaly on the different clock types in this duration is equivalent to assuming the satellites will face one anomaly of that magnitude every 3 hours.

Before defining existing time scale algorithms in the next chapter, the types of anomalies are presented, where measurement anomalies are a unique issue to deal with when relying on a time scale instead of only onboard clocks.

1.3.2 Anomalies in space

The expected anomalies to be faced by clocks operating in space are: phase jumps, frequency jumps, anomalous measurements, missing data, changes in variance, periodic deviations, and clock drift. Although it is useful to consider the sources of these anomalies and attempt to reduce their occurrence, this thesis assumes that there are no specially designed protections included in the design of the nanosatellites. With this assumption, the assessment of robustness to the anomalies allows a decision to be made on how much time and money should be invested into preventative measures in the NOIRE mission.

As clock data is either presented as phase or frequency differences, the affects of the different types of anomalies are expressed in terms of the nominal phase and frequency of clock i plus some contaminating term $a(t)$ that changes depending on the type of anomaly.

$$\tilde{x}_{i,p}(t) = x_{i,p}(t) + a(t), \tag{1.14}$$

and the fractional frequency being the rate of change of the phase offset leads to the contaminating term also being converted to a rate of change

$$\tilde{y}_i(t) = y_i(t) + \dot{a}(t), \tag{1.15}$$

where $\dot{a}(t) = \frac{d}{dt}\left[a(t)\right]$. The clock anomalies can now be expressed in terms of their characteristic contaminating function $a(t)$. The anomalies can affect the clocks in one way, however, a different response could occur in a time scale algorithm if the anomaly is not properly treated. The response of a time scale to the individual anomalies is not necessarily

identical to the effect on the individual clocks due to the methods of assigning weights and estimating frequencies. In fact, the design of a robust time scale algorithm should provide a time reference that is completely unaffected by the expected anomalies. To properly identify what the time scale should avoid, the affects of expected anomalies are visualized.

It is difficult to identify the extent of the anomalies that will be faced by the satellites used for the NOIRE mission. The anomalies known to occur in existing clock technologies are therefore used as an initial assumption of the expected problems to be faced. Regardless of the realistic magnitudes of anomalies that are expected, the inclusion of a sufficient number of satellites in the swarm can still effectively reduce the impact. In the unlikely case that several satellites suffer an anomaly at the same time, the impact will be more substantial than singular anomalous satellites. It is assumed that the more likely scenario is that each satellite experiences an anomaly at some random point in time. If the independent anomalies occur relatively frequently, the impact of even small magnitude jumps can be considerable.

Phase Jumps

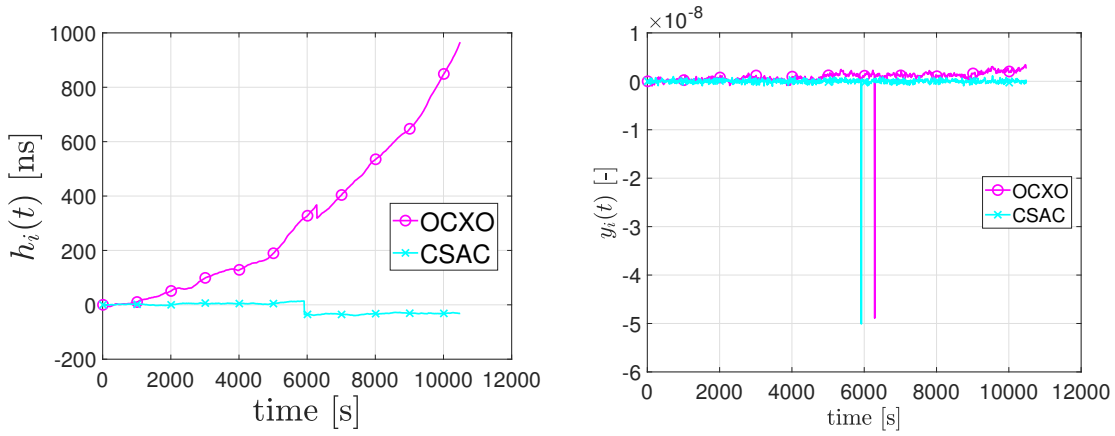
A phase jump anomaly is something already experienced in clocks based on Earth that can be observed and removed from the clock data manually by timing experts. Specialized algorithms already exist to automatically detect these types of jumps and will be discussed in more detail in the next chapter. A phase jump causes the time offset of the affected clock to increase or decrease by a magnitude of Δx indefinitely. The contamination function then has the form

$$a_x(t) = \Delta x H(t - t_a), \tag{1.16}$$

where $H(t - t_a)$ is the Heaviside step function which has a value of zero for $t < t_a$ and one for $t \geq t_a$, and t_a is the time epoch at which the anomaly occurs. Ideally, the anomaly is detected at the instant it occurs to mitigate it's affects in real-time and allow the radio image correlation. The time derivative of the contamination function shows that the clock frequency is also affected by the phase jump, in the form of an outlier at time t_a

$$\dot{a}_x(t) = \Delta x \delta(t - t_a), \tag{1.17}$$

where $\delta(t - t_a)$ is a Dirac distribution. An example of the magnitude of a phase jump observed in monitoring Global Positioning System clocks is around 70 ns [26] but can also be as low as a few nanoseconds [27]. The causes of phase jumps can be considered a characteristic behavior of oscillators or a result of the method of comparing clock phases [28]. The latter case is also considered in this work but as a separate anomaly referred to as a measurement or link anomaly. The presence of a single phase jump results in a discontinuity in both the phase and the frequency of an oscillator so that it is no longer capable of meeting the requirements for the mission. An example of this is shown in Figure 1.7,



(a) Phase evolution.

(b) Fractional frequency.

Figure 1.7: The impact of a simulated 50 ns phase jump on the phase and frequency of the simulated clocks.

Frequency Jumps

Frequency jumps are equivalent to the phase jumps, where the frequency of the clock is abruptly changed at a certain time epoch t_a. Similarly to the phase jump anomalies, there are existing detection methods for this type of anomaly but with a limitation on the detection delay. See the next chapter for more discussion on the existing anomaly detection algorithms. The form of the contaminating function with respect to the frequency state is

$$\dot{a}_y(t) = \Delta y H(t - t_a), \tag{1.18}$$

where the Heaviside step is used to indicate the change in the frequency. In practice, the frequency jump does not necessarily occur instantaneously, so can take some time before it is possible to identify a frequency jump. A temporary frequency jump is also possible where the step would then reduce back to zero at the end of the period with the frequency shift. This type of temporary frequency jump has been experimentally shown to occur during exposure of a Temperature Compensated Crystal Oscillator (TCXO) to radiation or a magnetic field, conditions that are relevant in the space environment [29]. Temporary frequency jumps can also be viewed as a transient phase anomalies, which exist on OCXO clocks due to large changes in temperature [30, 31].

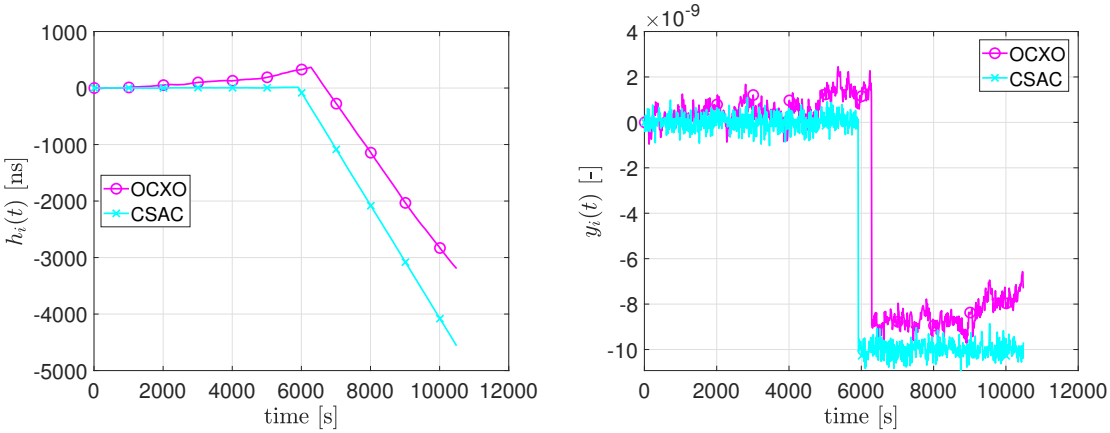
When an oscillator suffers from a frequency jump, the phase is contaminated according to a change in slope that is proportional to magnitude of the anomaly. For an indefinite frequency jump:

$$a_y(t) = \Delta y R(t - t_a), \tag{1.19}$$

where $R(t - t_a)$ is the unit ramp function that equals zero for $t \leq t_a$ and $R(t - t_a) = t - t_a$ for $t > t_a$. Depending on the size of the frequency jump it may not cause a significant

change in the clock phase for several time epochs after the moment the jump occurred. Generally, the magnitude of frequency jumps are smaller than those of phase jumps. This is why the detection of a frequency jump is typically constrained by some time delay.

Figure 1.8 indicates the changes in the phase and frequency of OCXO and CSAC clocks that suffer from a frequency jump of 10 ns/s over a period. The magnitude of the simulated anomaly is larger than the magnitudes experienced in tests to make the change in the clocks more clear in the figure.



(a) Phase evolution.

(b) Fractional frequency.

Figure 1.8: The impact of a simulated 10 ns/s frequency jump on the phase and frequency of the simulated clocks.

Environmental Sensitivities

Temperature coefficients are used to indicate the magnitude of frequency error per degree of temperature change. CSAC clocks have a temperature coefficient around $\pm 5 \times 10^{-12}/°\text{C}$ and OCXO clocks having a coefficient of $\pm 1.33 \times 10^{-11}/°\text{C}$ according to experimental results in [31]. These coefficients are defined for the temperature range from -10°C to 70°C and result in phase errors over several hours in the order of a few microseconds for OCXO and around 1 μs for the tested CSACs. Tests for TCXO clocks exposed to radiation or a magnetic field resulted in jumps in fractional frequency of 0.4 ns/s and 0.05 ns/s, respectively. This gives a rough idea of the expected magnitude of frequency variations that are often transient in nature, and can occur periodically in relation to orbital periods or temporary radiation events.

Periodic fluctuations

Spacecraft in orbit are exposed to periodic temperature and radiation fluctuations due to the periodicity of their orbits. A spacecraft in eclipse will surely experience a temperature drop, and can expect some level of shielding from solar radiation. As a result, the temporary frequency jumps can effectively occur in the form of transient frequency fluctuations over some defined period. Depending on the amplitude of the fluctuations, the ADEV at half the period of the fluctuation is expected to increase with a local maximum due to the increase instability at that time interval. This type of anomaly is therefore specifically disadvantageous for remaining capable of collecting scientific data over a broad range of different time intervals. The contaminating function can be defined for a certain amplitude A and period P

$$a_p(t) = A \sin\left(\frac{2\pi}{P}t\right), \tag{1.20}$$

the amplitude of the anomaly in the frequency is then scaled proportionally to the frequency of the fluctuations

$$\dot{a}_p(t) = 2\pi\frac{A}{P}\cos\left(\frac{2\pi}{P}t\right). \tag{1.21}$$

Measurement Anomalies

The presence of measurement anomalies is an important constraint on making clock comparisons between satellites, which is necessary for generating a time scale. This type of anomaly does not affect the onboard clocks but the time scale that we aim to create by using information from all of the clocks in the swarm. The definition of measurement anomalies considers significant and variable measurement noise on the observed differences between the satellite clocks. To define this type of anomalies, the method of measuring clock differences should be explained.

Assuming a method of two way transfers is possible between neighbouring satellites, the fundamental principle of pseudorange measurements are useful for determining the relative positions of the satellites in the swarm [32]. Generally, the pseudorange considered in GNSS applications includes delays due to atmospheric and ionospheric effects, which can be removed through the use of two way transfers. In lunar orbit, the atmospheric and ionospheric delays are not an issue but any similar effects but any similar effects can be absorbed into the measurement noise or considered an anomaly. A simplified form of the equation for the pseudorange between satellite A and satellite B is:

$$\rho_{AB}(t) = D_{AB}(t) + cx_{AB}(t) + n_{AB}(t) = c\tau_{AB}(t), \tag{1.22}$$

where the pseudorange $\rho_{AB}(t)$ is written simply as the sum of the geometric distance $D_{AB}(t)$, the position error due to the satellite clock differences cx_{AB}, and measurement noise present on the inter-satellite communication link $n_{AB}(t)$. The pseudorange is obtained by measuring the travel time of a signal sent from one satellite to the other, denoted as

$\tau_{AB}(t)$. This explains why the difference in the clock time appears in the pseudorange equation, as a source of error in the inter-satellite distance in addition to the measurement noise. Again, by taking advantage of a two way signal transfer between the satellites, either the distance or clock offset can be obtained from the measured pseudoranges. That is, the pseudorange in the opposite direction must be simultaneously measured with $\rho_{A,B}(t)$ to ensure negligible variations in the measured distance and clock offset:

$$\rho_{BA}(t) = D_{BA}(t) + cx_{BA}(t) + n_{BA}(t) = c\tau_{BA}(t), \tag{1.23}$$

where $D_{BA}(t) = D_{AB}$, $x_{BA} = -x_{AB}$, and the noise variance is equal for both directions but have independent values because the two way transfer is not instantaneous or perfectly simultaneous [33]. Taking the difference between the two pseudoranges allows the time difference between the two clocks to be observed with a residual noise related to the noise distribution for the link between satellites A and B

$$x_{AB}(t) = \frac{\tau_{AB}(t) - \tau_{BA}(t)}{2} + \frac{n_{AB}(t) - n_{BA}(t)}{2c}. \tag{1.24}$$

If the noise in the link between satellites A and B is contaminated by an additional interference or if the inter-satellite separation increases rapidly, the measured clock difference is then anomalous. The determination of the travel time for the signal of interest can also be erroneous due to ambiguities in the sub-millisecond measurements of the signal code for the pseudorange, but several procedures have been implemented to solve for this ambiguity [34, 35]. Ambiguities in the carrier wave compared to the signal code can also result in anomalous measurements of the signal transit time when attempting to improve estimates using carrier phase [36, 37]. This can understandably result in behavior similar to a phase jump. For example, the time difference when satellite A is affected by a phase jump is

$$\tilde{x}_{AB}(t) = x_{A,p}(t) + a_x(t) - x_{B,p}(t) = x_{AB}(t) + a_x(t), \tag{1.25}$$

which can be equivalent to the measured phase difference in the case there is a change in the noise variance or a bias in the travel time measurement due to characteristic anomalies, i.e., the contamination function is now a representation of the error in τ_{AB} due to the measurement process:

$$a_x(t) = \frac{\hat{\tau}_{AB}(t) - \tau_{AB}(t)}{2} + \frac{\hat{\tau}_{BA}(t) - \tau_{BA}(t)}{2}, \tag{1.26}$$

where $\hat{\tau}_{AB}(t)$ is an estimate of the signal travel time and differs from the actual signal travel time according to the signal noise ratio (SNR) and other uncertainties in the observation/estimation procedure. There are many proposals for joint estimation of the clock bias and the inter-satellite distances with more detailed procedures than demonstrated above. The state-of-the-art typically takes into account the motion of the satellite between consecutive two-way communications and also estimates clock frequency differences (clock

skew) [33, 38, 39]. Several of the estimators use a specialized method of Maximum Likelihood Estimation (MLE) or Least Squares (LS) estimation, which allows verification of the optimality of the estimators.

Although the estimators of clock bias are proven to be asymptotically efficient, i.e., proven to achieve minimum error when the signal noise ratio is sufficiently high, these estimators are not necessarily robust to the anomalies discussed above. For the design of the time scale algorithms, it is assumed that anomalies and measurement noise will not necessarily be compensated before obtaining the independent phase difference measurements between any pair of satellite clocks. The contributions presented in this thesis address estimators that are robust to the expected anomalies, integration of the robust estimators in the basic definition of a time scale, and the performance of those robust estimators in the asymptotic regime. As a result, the necessity to remove anomalous measurements is also analyzed by observing the results of using a basic LS estimator on contaminated measurements compared to permitting them to be processed in the robust time scale algorithm.

Missing Measurements

Missing measurement anomalies refer to the case that the distance between satellites becomes too great to establish a communication link or if there are temporary issues in the transmission of signals. This can mean that certain inter-satellite links are either unavailable or so noisy that they are not worth keeping in the determination of the time scale. It can be shown that when a certain satellite becomes completely isolated from the rest of the swarm, the time scale requires special treatment of the remaining clock measurements and the isolated satellite needs to be reintroduced appropriately. There is an identified gap in the literature on a specific methodology to follow when generating a time scale with a missing clock.

It is sure that timing laboratories around the world deal with this problem manually on a regular basis, but an algorithm that can handle this automatically would be ideal for the satellite swarm. The next chapter explains the current time scale algorithms and how they deal with the anomalies presented. The design of a new robust time scale is then based on the same core principles of a general time scale algorithm but introduces a new method of anomaly mitigation. The assessment of the anomaly mitigation in the next chapter does not include the case of missing measurements due to the lack of specialized algorithms to deal with this type of anomaly. A contribution towards a robust time scale algorithm is presented later in Chapter 3 that aims to mitigate phase jumps, frequency jumps, and measurement anomalies. An additional section is included in that chapter to present a contribution towards dealing with missing measurements.

Chapter 2

Time Scales

The context of this thesis has demonstrated a need for a robust time scale. This chapter is dedicated to explaining the current state-of-the-art time scale algorithms. There are several designs of time scales that are defined in different manners for different applications. Nevertheless, there is a common core amongst the designs that has been identified in the research conducted during this thesis. This common requirement for all time scales is referred to as the Basic Time Scale Equation (BTSE) and will be shown where it applies for each corresponding time scale algorithm. By providing several points of view on the BTSE in various time scale algorithms, a new understanding of the state-of-the-art solutions is presented, which is pivotal to the primary contribution of the thesis presented in the next chapter.

Without dedicated anomaly detection and mitigation strategies, the time scale will also be corrupted by the same anomalies that cause the individual clocks to fail to achieve the minimum required stability. Specialized methodologies exist for detection and compensation of phase jumps and frequency jumps in the widely used AT1 time scale [40, 41]. Otherwise, the implementation of a Kalman Filter to estimate clock states has allowed a range of general anomaly detection methods to be applied [42, 43, 44, 45]. Such anomaly detection algorithms are presented in this chapter to appropriately communicate the possible methods that are currently available. By first presenting the current state-of-the-art, the novelty of the contribution presented in the next chapter will be more apparent.

2.1 Coordinated Universal Time

The Coordinated Universal Time (UTC) is a common reference well-known around the world as a basis for time zones or even precise timing in computers. The UTC reference time is linked to the traditional realization of time that relies on the rotation of the Earth. As a result, UTC has discontinuities due to inconsistencies in the rotation of the Earth, which must be corrected with leap seconds [46]. The Temps Atomique International (TAI)

is a continuous reference time that follows the Standard International (SI) definition of the second and acts as the base for UTC before adding leap seconds. The TAI is kept in accordance with the definition of the second by having the frequency of the free atomic time scale (EAL for Echelle Atomique Libre) steered to a Primary Frequency Standard (PFS).

The EAL is the primary realization of a time scale that applies a weighted average to the timing information obtained amongst an ensemble of clocks. This weighted average is effectively the application of what is referred to as the BTSE. Several existing time scale algorithms apply the weighted average with different definitions of the weights, and intermediate steps such as frequency estimation and predictions. Since the focus of this thesis is the development of a new and robust time scale algorithm, the steps of going from EAL to TAI then to UTC are considered outside of the scope. Depending on the application of the time scale, the frequency may need steering to a frequency standard and may benefit from being linked to the global reference UTC, but these steps are well-defined so can be easily implemented on any algorithm that is used to obtain a realization of EAL.

Whatever the chosen algorithm to obtain EAL, each laboratory makes their own realization of UTC, denoted as UTC(k) for laboratory k. The local UTC(k) realizations are made according to the time scale algorithm applied to the measured clock comparisons between physical clocks in the laboratories and appropriately translating from the equivalent EAL. The report BIPM Circular T is published each month and presents the phase deviations between UTC(k) and the global UTC along with uncertainties. This is important for the dissemination of the time reference system used on Earth, so that relevant applications can benefit from the precisions of the best clocks in the world. However, the NOIRE mission will not necessarily need synchronization with respect to UTC. The swarm of nanosatellites will be making their observations autonomously and just need the time stamps with respect to a stable reference. The best method of using a stable and common time scale for all the satellites without needing regular timing information sent from Earth is to compute an independent time scale using only the onboard clocks. This will help reduce the requirements for ground-based communication with the swarm, hence, the cost of the mission is also reduced by keeping a minimum complexity of receiver infrastructure.

A summary of the components that make up the realization of the UTC is illustrated in Figure 2.1. The specificities of steering and including leap seconds are considered outside the scope of this thesis, where the main objective is to provide a novel method of computing the weighted average in the Basic Time Scale Equation.

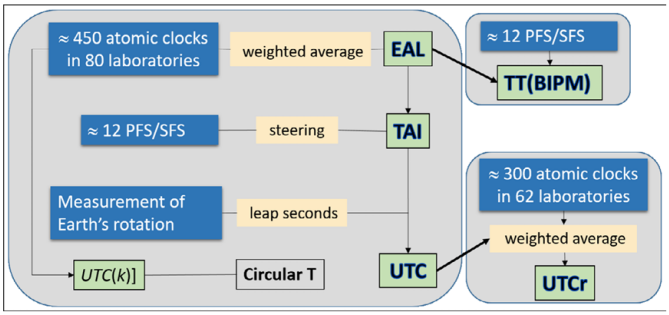


Figure 2.1: Diagram of the different stages that are implemented to realize the UTC time scale as described by the Bureau International des Poids et Mesures (BIPM) [46].

2.2 Basic Time Scale Equation

A time scale in general takes a weighted average of a network of clock data to provide a common ensemble time, being more stable than any individual clock [46, 47, 48, 49, 50]. 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 $x_{j,i}(t) = h_j(t) - h_i(t)$, 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 [49, 51, 52, 46]:

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - x_{j,i}(t)\right], \tag{2.1}$$

where the predictions, measurements, and weights can be defined differently for each algorithm. 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 [51]. By default, the weights are computed using information from the previous epoch for the AT1 time scale algorithm [48]. More about this algorithm is explained in later sections. The need to be robust to anomalies means that the weights should be updated at the current time epoch to compensate the effects in near real-time. That is, the BTSE can be recomputed at the current time with updated weights that obtain information about the performance of each clock at the current time $w_j(t)$.

The BTSE can also be explained as the solution to a system of N equations that includes a weighted average of the clock phase noise (one equation) and relevant phase difference

measurements ($N-1$ equations) [49]. That is, the following system of equations:

$$\sum_{j=1}^{N} w_j(t-\tau)h_j(t) = \sum_{j=1}^{N} w_j(t-\tau)\hat{x}_{j,E}(t), \tag{2.2}$$

$$x_{1,2}(t) = h_1(t) - h_2(t), \tag{2.3}$$

$$\vdots \tag{2.4}$$

$$x_{1,N}(t) = h_1(t) - h_N(t). \tag{2.5}$$

The system of equations above should be solved for the absolute time of each clock $h_i(t)$, yielding solutions that are equivalent to the BTSE defined above. The solutions are not perfect because the predictions in (2.2) are not perfect observations of the true phase due to the stochastic evolution of the clock states.

It can be shown that the error between the solutions to the BTSE and the true absolute phase is identical for all clocks since

$$x_{i,E}(t) = h_i(t) + \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - h_j(t)\right]. \tag{2.6}$$

Based on the definition of the output being a phase deviation from the time scale, the absolute time of the time scale is

$$h_E(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[h_j(t) - \hat{x}_{j,E}(t)\right], \tag{2.7}$$

which is the weighted average of the prediction errors. Intuitively, the prediction errors $e_j(t) = \hat{x}_{j,E}(t) - h_j(t)$ represent the unpredictable component of the clock phase, i.e., caused by stochastic processes only. For a perfectly stable time scale, either the predictions should be as close as possible to the true phase or the weights should be appropriately adjusted for those clocks that are relatively more unpredictable. Since anomalies are by definition unpredictable they clearly require reduced weights otherwise the absolute time of the time scale will change according to the magnitude of the anomaly. Assuming anomalies are properly mitigated or not present in the clock data, the time scale produces a common reference time with a stability improved by a factor of $\sqrt{N}$ in terms of Allan deviation.

The design of the new time scale algorithm in the next chapter includes study of non-uniform measurement noise between inter-satellite communication links in the form of measurement anomalies and another case where the noise is negligible in comparison to the clock stochastic processes but the clocks have internal anomalies. These cases are generally considered nominal operations but with different magnitudes of measurement noise. As discussed in Section 1.3.2, link noise can also be affected by anomalies in the space environment causing certain measurements to have an elevated variance or bias that

is non-negligible. The next chapter explains a contribution in the form of a new time scale algorithm that is robust to such measurement based outliers in the weight computations and another contribution that presents preprocessing opportunities for the measurements in the swarm to mitigate measurement anomalies before computing the time scale.

A step by step list is provided below to detail the required steps to be able to recursively compute the BTSE at each time instant, separated by the sampling interval τ. Assuming that the previous offset of each clock from the time scale is known, the steps to compute the BTSE include

1. **Predictions**: The offset from the time scale at the current time is predicted using the previously known offset.
2. **Measurements**: The absolute phases of each clock are indirectly observed by measuring the clock time differences.
3. **Weight computation**: Depending on the algorithm implemented, the weights are either computed using past values in order to realize the BTSE or if anomalies are detected, the weights are modified.
4. **BTSE**: The offset of each clock from the common time scale is calculated with the weighted average of the BTSE residuals. These BTSE residuals refer to the difference between individual clock phase predictions and the time difference measurements.
5. **Frequency Estimation**: The frequency of the time scale is estimated according to the outputs of the BTSE to aid in the next prediction of the clock phases with respect to the time scale. Higher order clock states such as drift or biases could also be estimated at this step.

Each of the terms in the BTSE is detailed independently in the next sections for two different types of time scale algorithms below. Depending on the algorithm, the methods for predicting clock states, computing clock weights, and the detection and compensation of anomalies will differ. The Kalman Filter methodology is presented first to allow links to the KF formulation in the later section on the AT2 time scale for compensation of frequency jumps.

2.3 Kalman Filter Composite Clock

This section explains how the principle clock states are included in the traditional Kalman Filter (KF) formula to obtain a common reference. This method of obtaining a time scale is most famously used for the GPS composite clock [53], but other timing laboratories implement the KF in a similar manner [54, 55]. At first glance, the equations related to the KF do not seem coherent with the BTSE but it can be demonstrated that the KF algorithm is equivalent to a batch least squares [56] of the clock residuals, i.e., a type of

weighted average. The KF algorithm consists of two key steps: the prediction step, which is made before the measurements are obtained, and the update step, which is equivalent to applying weights as described in the BTSE to the difference between the measurements and the predictions.

2.3.1 Measurements

The only method of measuring true clock phases is through making time comparisons between pairs of clocks and obtaining the phase differences $x_{j,i}(t)$. Since the KF algorithm deals simultaneously with phase, frequency, and drift estimation, the measurements are presented as a linear equation of the state vector:

$$\mathbf{Hx}(t) = \begin{bmatrix} x_{1,p}(t) - x_{2,p}(t) \\ x_{1,p}(t) - x_{3,p}(t) \\ \vdots \\ x_{1,p}(t) - x_{N,p}(t) \end{bmatrix} = \begin{bmatrix} x_{1,2}(t) \\ x_{1,3}(t) \\ \vdots \\ x_{1,N}(t) \end{bmatrix}, \tag{2.8}$$

where the observation matrix $\mathbf{H}$ has dimensions $(N-1) \times 3N$ and contains $+1$ and -1 in the appropriate positions to obtain the measured phase differences:

$$\mathbf{H} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & \dots & \cdots & 0 \\ 1 & 0 & -1 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \end{bmatrix} \tag{2.9}$$

An advantage of the KF algorithm is that it implements the measurement noise into the computation of the timescale. However, the required assumptions must be compatible with reality, otherwise, the KF algorithm will poorly compensate for unexpected issues. We define the set of $N-1$ noisy measurements in the vector:

$$\mathbf{z}(t) = \mathbf{Hx}(t) + \mathbf{n}(t), \tag{2.10}$$

where the vector $\mathbf{n} = [n_{1,2}(t), \cdots, n_{1,N}(t)]$ includes the unique noise associated with each of the links between satellite 1 and the other satellites. The fact that unique inter-satellite links will have different measurement noises is typically not addressed in KF-based time scales, instead it is just assumed that the measurement noise is uniform and follows a Gaussian distribution:

$$\mathbf{n}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{2.11}$$

Not only is the assumption of the distribution inherent in the definition of the KF but the matrix $\mathbf{R}$ must be known or estimated. The variance associated with white phase noise in the clocks is suggested as a good first approximation for the measurement noise matrix. However, this variance is not guaranteed to be fixed and will change significantly in the presence of anomalies. As a result, the KF algorithm must have robust approaches introduced to process potential anomalies in measurements and clock states.

2.3.2 Predictions

To predict the states of a clock, we must assume a certain model of the clock dynamics. A state equation specifies the dynamic propagation for each of the states of interest. These states are: the clock phase with respect to a perfect reference time $x_{i,p}(t) = h_i(t) - h_p(t)$, the clock fractional frequency $y_{i,p}(t)$, which is related to the rate of change of the clock phase, and the clock frequency drift $d_{i,p}(t)$, which is simply the rate of change of the fractional frequency. Due to the nature of atomic clocks, the dynamic evolution of these states is not predictable. That is, there are stochastic noises associated with each clock state. Assuming that clock drift is not negligible, the phase state equation is:

$$x_{i,p}(t) = x_{i,p}(t-\tau) + \tau y_{i,p}(t-\tau) + \frac{\tau^2}{2} d_{i,p}(t-\tau) + \varepsilon(t), \tag{2.12}$$

where $\varepsilon(t)$ is the stochastic component associated with the characteristic phase noise of a clock. We assume a linear evolution of the frequency due to the drift and assume that the drift is constant except for another random noise component

$$y_{i,p}(t) = y_{i,p}(t-\tau) + \tau d_{i,p}(t-\tau) + \eta(t), \tag{2.13}$$

$$d_{i,p}(t) = d_{i,p}(t-\tau) + \xi(t). \tag{2.14}$$

The state vector $\mathbf{x}(t) = [x_{1,p}(t), \cdots, x_{N,p}(t), y_{1,p}(t), \cdots, y_{N,p}(t), d_{1,p}(t), \cdots, d_{N,p}(t)]^T$ contains the phases of each clock, the frequencies of each clock, and the drifts of each clock. The combination of the dynamic equations for each state (2.12), (2.13), and (2.14) provides the overall state equation:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{b}(t), \tag{2.15}$$

where the stochastic components are contained in the process noise vector:

$$\mathbf{b}(t) = [\varepsilon_1(t), \cdots, \varepsilon_N(t), \eta_1(t), \cdots, \eta_N(t), \xi_1(t), \cdots, \xi_N(t)]^T, \tag{2.16}$$

and $\mathbf{A}$ is the matrix that simultaneously specifies the dynamic models in (2.12) to (2.14)

$$\mathbf{A} = \begin{bmatrix} 1 & \tau & \frac{\tau^2}{2} \\ 0 & 1 & \tau \\ 0 & 0 & 1 \end{bmatrix} \otimes \mathbf{I}_N = \mathbf{\Phi} \otimes \mathbf{I}_N, \tag{2.17}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{I}_N$ is the $N \times N$ identity matrix.

$$\mathbf{b}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}). \tag{2.18}$$

The KF-based time scales use a specific model of $\mathbf{Q}$ to represent the variances and covariances of the unpredictable components of clock evolution [57]

$$\mathbf{Q} = \begin{bmatrix} s_1\tau + s_2\frac{\tau^3}{3} + s_3\frac{\tau^5}{20} & s_2\frac{\tau^2}{2} + s_3\frac{\tau^4}{8} & s_3\frac{\tau^3}{6} \\ s_2\frac{\tau^2}{2} + s_3\frac{\tau^4}{8} & s_2\tau + s_3\frac{\tau^3}{3} & s_3\frac{\tau^2}{2} \\ s_3\frac{\tau^3}{6} & s_3\frac{\tau^2}{2} & s_3\tau \end{bmatrix}, \tag{2.19}$$

where s_1, s_2, and s_3 are diffusion coefficients corresponding to the intensity of the white frequency noise, random walk frequency noise, and random walk drift noise, respectively. The white phase noise is contained in the measurement noise of the KF model and the flicker frequency noise can be modeled by a linear combination of Markov processes that are treated as additional states in the KF model [58]. The inclusion of the Markov states is not necessary to obtain a functioning time scale and does not provide further insight into the core problem of clock anomalies. For these reasons the Markov modeling is omitted from the scope of this thesis.

Regardless of how detailed the prediction model is, there will always be some error with respect to the true clock states. The KF estimates the covariance matrix of the estimation error $\mathbf{P}$ with an initial prediction step

$$\mathbf{P}(t|t-\tau) = \mathbf{A}\mathbf{P}(t-\tau|t-\tau)\mathbf{A}^T + \mathbf{Q}. \tag{2.20}$$

Information from the previous time instant is used in the prediction step of the KF algorithm. Parameters computed at time t using past information are specified with $t|t-\tau$. The prediction equation also uses this notation to refer to the a priori KF estimate

$$\hat{\mathbf{x}}(t|t-\tau) = \mathbf{A}\hat{\mathbf{x}}(t-\tau|t-\tau). \tag{2.21}$$

We now have predictions of the clock states and an estimate of the variance of those predictions. The expected distribution of the predictions is Gaussian

$$\hat{\mathbf{x}}(t|t-\tau) \sim \mathcal{N}\left(\mathbf{x}, \mathbf{P}(t|t-\tau)\right). \tag{2.22}$$

Other models for the above process noise and state equations are established for clocks affected by jump-type anomalies [59]. However, these models are more suited to simulations with known magnitudes of anomalies. If the extended models are included in the KF algorithm, the jump magnitudes would then be additional states to estimate. Consequently, the state vector would need to be modified to include N more parameters to estimate for each type of anomaly. This is not necessarily impossible for one type of anomaly, as shown in [50]. Nevertheless, trying to estimate several different types of anomalies on several clock states would result in a loss in the estimation performance and is restricted by the available measurements.

2.3.3 Weights

The update step of the KF is the moment that residuals are weighted with the predetermined Kalman gain matrix to obtain an improved estimate of the clock states. The Kalman gain matrix $\mathbf{K}$ is derived according to a minimization of the state mean square error, leading to [60]:

$$\mathbf{K} = \mathbf{P}(t|t-\tau)\mathbf{H}^T \left(\mathbf{H}\mathbf{P}(t|t-\tau)\mathbf{H}^T + \mathbf{R}\right)^{-1}. \tag{2.23}$$

The resulting matrix is $N \times (N-1)$, where $N-1$ is the number of measurements used to estimate the independent phase values. In the best-case scenario: we can approximate the measurement noise covariance $\mathbf{R}$ perfectly, there are no missing measurements so $\mathbf{H}$ is well-defined by (2.9), and the state estimation covariance $\mathbf{P}$ is representative of the actual clock estimation error. However, in the case of unpredictable anomalies, none of these approximations are necessarily true. In addition, the Kalman gain matrix is calculated using only past information, there is no link between the current measurements and the current weights. For these reasons, we cannot rely on $\mathbf{K}$ alone to mitigate the effects of anomalies. Detection algorithms are a strong field of research for the robustification of KF time scales. The KF update equation uses the gain matrix to simultaneously estimate phase, frequency, and drift:

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t|t-\tau) + \mathbf{K}\left(\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}(t|t-\tau)\right), \tag{2.24}$$

and the resulting estimation covariance is also updated:

$$\mathbf{P}(t) = \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\mathbf{P}(t|t-\tau)\left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)^T + \mathbf{K}\mathbf{R}\mathbf{K}^T. \tag{2.25}$$

The distribution of the updated estimates is then

$$\hat{\mathbf{x}}(t) \sim \mathcal{N}(\mathbf{x}(t), \mathbf{P}(t)). \tag{2.26}$$

An important modification to the KF algorithm for generating a preferred timescale is the method of "covariance x-reduction" [61]. This means the components of the matrix $\mathbf{P}(t)$ associated with phase covariance are set to zero after the update step. If this is not done, the phase covariance will grow without limit because the number of measurements is less than number of states being estimated.

We can rewrite the KF update (2.24) in terms of individual clock phases, frequencies, and drifts, using a notation compatible with the BTSE notation, $x_{i,\mathrm{KF}}(t)$ is the ith element

of $\hat{\mathbf{x}}(t)$, and $\hat{x}_{i,\mathrm{KF}}(t)$ is the ith element of $\hat{\mathbf{x}}(t|t-\tau)$:

$$\begin{bmatrix} x_{1,\mathrm{KF}}(t) \\ \vdots \\ x_{N,\mathrm{KF}}(t) \\ y_{1,\mathrm{KF}}(t) \\ \vdots \\ y_{N,\mathrm{KF}}(t) \\ d_{1,\mathrm{KF}}(t) \\ \vdots \\ d_{N,\mathrm{KF}}(t) \end{bmatrix} = \begin{bmatrix} \hat{x}_{1,\mathrm{KF}}(t) \\ \vdots \\ \hat{x}_{N,\mathrm{KF}}(t) \\ \hat{y}_{1,\mathrm{KF}}(t) \\ \vdots \\ \hat{y}_{N,\mathrm{KF}}(t) \\ \hat{d}_{1,\mathrm{KF}}(t) \\ \vdots \\ \hat{d}_{N,\mathrm{KF}}(t) \end{bmatrix} + \begin{bmatrix} K_{1,1} & \cdots & K_{1,M} \\ \vdots & & \vdots \\ K_{N,1} & \cdots & K_{N,M} \\ K_{N+1,1} & \cdots & K_{N+1,M} \\ \vdots & & \vdots \\ K_{2N,1} & \cdots & K_{2N,M} \\ K_{2N+1,1} & \cdots & K_{2N+1,M} \\ \vdots & & \vdots \\ K_{3N,1} & \cdots & K_{3N,M} \end{bmatrix} \begin{bmatrix} x_{1,2}(t) - (\hat{x}_{1,\mathrm{KF}}(t) - \hat{x}_{2,\mathrm{KF}}(t)) \\ \vdots \\ x_{1,N}(t) - (\hat{x}_{1,\mathrm{KF}}(t) - \hat{x}_{N,\mathrm{KF}}(t)) \end{bmatrix}. \tag{2.27}$$

The above expression includes $M = N - 1$ measurements of the phase differences between clock 1 and all other clocks. This is sufficient to obtain a solution and is sufficient to explain the implicit weights. The elements of the matrix $\mathbf{K}$ depend on past information and predictions so should be a function of $(t - \tau)$ but the notation is omitted for brevity. The above equation is identical to simultaneously computing the BTSE N times for each clock state. In this case, the reference clock for the measurements is fixed but the weights are different for each realization of the BTSE. Each clock has N weights implicitly assigned via the Kalman gain matrix. These implicit weights correspond to a linear combination of the elements in $\mathbf{K}$. The equations defining the implicit weights vary depending on whether or not we are estimating the states for the clock common to all measurements or another clock. Equation (2.27) leads to:

$$x_{i,\mathrm{KF}}(t) = \hat{x}_{i,\mathrm{KF}}(t) + \sum_{j=2}^{N} K_{i,(j-1)} \left(x_{1j}(t) - (\hat{x}_{1,\mathrm{KF}}(t) - \hat{x}_{j,\mathrm{KF}}(t)) \right), \tag{2.28}$$

$$y_{i,\mathrm{KF}}(t) = \hat{y}_{i,\mathrm{KF}}(t) + \sum_{j=2}^{N} K_{(i+N),(j-1)} (x_{1j}(t) - (\hat{x}_{1,\mathrm{KF}}(t) - \hat{x}_{j,\mathrm{KF}}(t))), \tag{2.29}$$

$$d_{i,\mathrm{KF}}(t) = \hat{d}_{i,\mathrm{KF}}(t) + \sum_{j=2}^{N} K_{(i+2N),(j-1)} (x_{1j}(t) - (\hat{x}_{1,\mathrm{KF}}(t) - \hat{x}_{j,\mathrm{KF}}(t))). \tag{2.30}$$

The resulting implicit weights can be written in terms of the elements of the Kalman gain matrix, but the exact expressions will vary depending on the measurements used and which

clock is being estimated. For example, when estimating the phase of the reference clock, the KF's equivalent BTSE is given by the following (with the dependence on time omitted for brevity of the equations):

$$x_{1,KF} = \hat{x}_{1,KF} + \sum_{j=2}^{N} K_{1,(j-1)} \left(x_{1j} - (\hat{x}_{1,KF} - \hat{x}_{j,KF}) \right), \tag{2.31}$$

$$x_{1,KF} = \hat{x}_{1,KF} - \hat{x}_{1,KF} \left(\sum_{j=2}^{N} K_{1,(j-1)} \right) + x_{1,p} \left(\sum_{j=2}^{N} K_{1,(j-1)} \right) + \sum_{j=2}^{N} K_{1,(j-1)} \left(\hat{x}_{j,KF} - x_{j,p} \right), \tag{2.32}$$

$$x_{1,KF} = x_{1,p} + (\hat{x}_{1,KF} - x_{1,p}) \left(1 - \sum_{j=2}^{N} K_{1,(j-1)} \right) + \sum_{j=2}^{N} K_{1,(j-1)} \left(\hat{x}_{j,KF} - x_{j,p} \right), \tag{2.33}$$

$$x_{1,KF} = x_1 + \sum_{j=1}^{N} w_j (\hat{x}_{j,KF} - x_j), \tag{2.34}$$

where the corresponding implicit weights are a function of the Kalman gain which was computed using past values

$$w_j(t-\tau) = 1 - \sum_{i=2}^{N} K_{1,(i-1)}, \; j = 1, \tag{2.35}$$

$$w_j(t-\tau) = K_{1,j-1}, \qquad j > 1. \tag{2.36}$$

Since the sum of the weights is equal to 1, we can simplify the KF with implicit weights to be in an identical form to the BTSE. Substituting the above implicit weights into (2.1) will provide the realization of the KF timescale with respect to clock 1, denoted as $x_{1,\mathrm{KF}}(t)$, and leading to:

$$x_{1,KF}(t) = \sum_{j=1}^{N} w_j(t-\tau)(\hat{x}_{j,KF}(t) - x_{j,1}(t)). \tag{2.37}$$

A similar simplification is possible for the BTSE used to compute $x_{i,\mathrm{KF}}(t)$, where $i \neq 1$:

$$w_j(t-\tau) = -\sum_{l=2}^{N} K_{j,(l-1)}, \qquad j = 1, \tag{2.38}$$

$$w_j(t-\tau) = 1 + K_{j,j-1}, \qquad j = i, \tag{2.39}$$

$$w_j(t-\tau) = K_{i,j-1}, \qquad j \neq 1, j \neq i. \tag{2.40}$$

If a different set of measurements are used, e.g., $N-1$ phase differences with respect to clock 2 and hence $i = 2$, the implicit weights are adjusted appropriately. For each of the

different values of i, the KF update equation (2.24) computes the BTSE simultaneously, each time with different implicit weights to obtain the same common timescale from the specified measurements.

With only $N-1$ linearly independent measurements, the gain matrix will automatically assign the required weights for each state estimate. Additional measurements would introduce additional elements in the matrix $\mathbf{K}$ that appropriately modify the implicit weights. Similarly to the phase, the KF algorithm automatically performs the prediction and estimation of the frequency and drift states while applying optimized weights. The Supplementary Time Scale Equations (STSE) for frequency and drift have implicit weights different to the phase weights and defined by the additional rows of the Kalman gain matrix $K_{N+1}, \cdots, K_{3N}$.

The implicit weights for all of the clock states are not robust to an anomaly occurring at time $t = t_a$ because the Kalman gain matrix is computed only using a priori information from $t = (t_a - \tau)$. Consequently, any anomaly in the clock states will not have its effect reduced by the implicit weights and the timescale will become unstable. For this reason, the KF algorithm requires additional robustness through specific detection methods or an adaptive Kalman gain matrix.

2.3.4 Anomaly detection for KF time scales

Creating a robust KF algorithm consists of detecting anomalies and modifying weights such that the detected anomalies do not contribute to the estimates of the clock states. To maintain some level of brevity, the specific robust modifications to the KF timescale are not introduced in this work but some compatible detection methods are discussed below.

Likelihood ratio test

The Likelihood Ratio Test (LRT) focuses on using a test statistic and a threshold above which this test statistic indicates the presence of an anomaly [62]. The choices of test statistics and thresholds can be made based on performance requirements, such as the probability of a false alarm (PFA) and the probability of non-detection (PND). The PFA refers to how often the test will incorrectly detect an anomaly. An example of a false alarm in time scale applications is measurement noise increasing instantaneously for a specific link, e.g., $x_{1,2}$, and causing the detection of a phase jump on clock 1 or 2, which did not actually have any anomaly. If this occurs on too many links at the same time, the effective number of clocks contributing to the time scale could decrease because their weights are reduced more than necessary. The PND is related to the probability of failing to detect anomalies. For example, phase jumps and frequency jumps of small magnitude may not cause the test to exceed the threshold but still result in instability of the resulting time scale.

Figure 2.2 illustrates how the distribution of anomalies can overlap with the distribution

of nominal data and cause such errors in the detection of anomalies. The demonstration in Figure 2.2 can be extended to consider anomalies that occur according to two identical distributions symmetrically mirrored about the mean of the nominal data, as would be the case for clock jumps that have similar probability to occur in either direction.

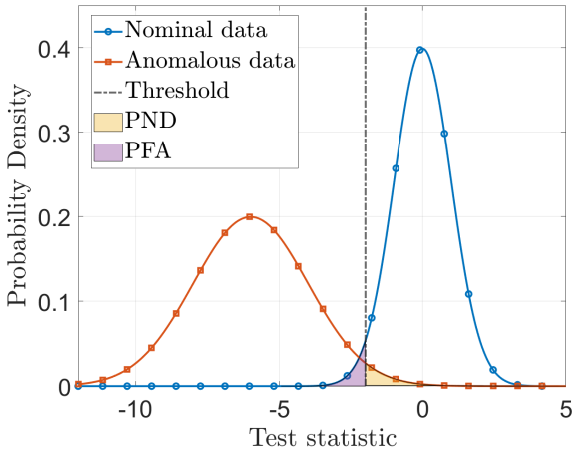


Figure 2.2: A basic illustration of the PFA and PND for a certain distribution of anomalies. The example threshold is shown at -2 and would then be at +2 for symmetrically mirrored anomalies. No threshold can guarantee a perfect separation between the clean and contaminated data when the distributions overlap. Since the anomalies could potentially provide realistic observations, there is a chance of mistaking the anomalies for nominal observations, i.e., the PND. Also there are certain values of the nominal data that may be discarded as anomalies, indicated by the PFA.

There is always a trade-off between PFA and PND because a low PFA implies a high PND and vice versa. The preferred optimization of this trade-off can depend on the correction procedure that is followed after the detection. A false alarm could result in removing a clock that is positively contributing to the time scale, an over-correction that can cause discontinuity and lowered stability in the time scale. For this reason, we would aim to minimize PFA. However, a certain proportion of the undetected anomalies are guaranteed to cause instability in the resulting time scale. By reducing PFA we increase the PND and vice-versa. Therefore, anomaly detection and identification will always retain some risk in the design.

The LRT designs the required threshold for a fixed PFA [62]. This test has been used for the detection of phase jumps [45], frequency jumps [42, 43, 45], and other non-jump-type anomalies such as changes in drift or variance [63, 45]. In each of the references, certain test statistics are better suited to certain types of anomalies, i.e., no individual test is sufficient to optimally identify the occurrence of all types of anomalies. The KF measurement resid-

uals, windows of past data for an estimate of the Dynamic AVAR (DAVAR) [45], average frequencies over a specified window [40, 43], and variance of the KF frequency residual [42] are all examples of test statistics used in the state-of-the-art anomaly detection for KF time scales.

There are also added constraints in the time between an anomaly occurring and its detection. For example, the identification of frequency jumps or clock drifts requires several observations after the time of the anomaly occurrence. The delay between the appearance and detection of the anomaly can differ depending on the magnitude of the frequency jumps or drifts. The detection delay for these anomalies means that the time scale should be retroactively fixed after the anomaly has already made an impact.

2.3.5 Robust Estimation for known anomalies

Besides anomaly detection, robust estimators can be applied to the KF to estimate the states with properly adapted weights for measurements affected by anomalies [64, 65]. This includes implementation of the known robust M-estimator, which similarly requires the tuning of a threshold to ensure performance in nominal operations is not reduced [66]. The challenge of obtaining an optimum threshold for detection or tuning a robust M-estimator is not a problem with another type of robust estimation based on heavy-tailed probability distributions. This type of robust estimation refers to modeling the combined probability of anomalies and normal observations and using knowledge of the probability distribution or an appropriate score function to estimate the impact of the anomalies. Extensions to robust estimation for a KF are not actively applied to the computation of KF based time scales. Similarly, other time scale algorithms rather rely on the detection aspect of robustness and not the estimation which can function more autonomously because there is no need to detect then adapt the time scale.

Taking advantage of estimation can also refer to estimation of deterministic parameters that characterize certain anomalies. For example, environmental conditions can cause a periodic transient effect on the clock phases. By introducing additional states in the KF model, the same KF algorithm can be used to jointly estimate the phase, frequency, drift, and harmonic coefficients that cause the periodic variation for each clock [50]. Similarly, one could define other types of additional states in the KF model to represent other types of anomalies.

2.4 AT1

The other algorithm to be assessed is a state-of-the-art and optimal solution to the BTSE, called the AT1 algorithm. The AT1 algorithm from NIST is widely used for simple representation of a real-time computed time-scale [49]. The algorithm includes all past information through use of exponential filters to provide ideal transient behavior in the predicted clock frequencies and computed clock weights depending on optimal time constants for different

types of clocks [48]. The AT1 algorithm has had several updates that address phase jumps and frequency jumps in the clocks and integrates the principle of the Kalman Filter (KF) to help with frequency estimation in the AT2 algorithm [40, 41].

2.4.1 Measurements

With only $N-1$ independent clock phase comparisons, i.e., $x_{j,i}(t)$ for a fixed satellite i, the BTSE can be computed N times to obtain the N time differences between each clock and the generated time scale $x_{i,E}(t)$. In most cases, the AT1 algorithm is used to compute time scales in timing laboratories where the measurement noise can be averaged down to a negligible level with respect to the clock noise. Nevertheless, the application to a swarm of satellites should not neglect measurement noise, so the noise on independent satellite links is defined. The statistical model of phase difference measurements $z_{ji}(t)$ with additive measurement noise is:

$$z_{ji}(t) \sim \mathcal{N}(h_j(t) - h_i(t), \sigma_{ji}^2(t)), \tag{2.41}$$

where the phase difference measurements for the links between clocks j and i are denoted as $z_{ji}(t) = x_{ji}(t) + n_{ji}(t)$ and the subscript ji refers to the specific link between clocks j and i. The above is equivalent to an assumption of independent measurements that experience the same noise level. Realistically, each link would have a different level of noise variance depending on the inter-satellite distances or anomalies in the links. The impact of anomalies in the measurement process will be specifically investigated by adding instantaneous changes to the measurement noise at random time epochs. Adding anomalies on the measurement links is a manner of expressing a varied measurement noise throughout a swarm. That is, when the measurement noise is generally negligible but it is significantly high for some specific links then those links with large noise could be suffering from an anomaly.

2.4.2 Predictions

The predictions of clock phases are equivalently defined for the AT1 time scale algorithm and other existing algorithms such as the KF, e.g., using a second-order polynomial, we obtain:

$$\hat{x}_{i,E}(t) = x_{i,E}(t-\tau) + \tau y_{i,E}(t-\tau) + \frac{\tau^2}{2} d_{i,E}(t-\tau), \tag{2.42}$$

where the above can be reduced to a linear model if we neglect drift. Similarly to (2.42), the AT1 time scale replicates a prediction equation for both the frequency and drift parameters:

$$\hat{y}_{i,E}(t) = y_{i,E}(t-\tau) + \tau d_{i,E}(t-\tau), \tag{2.43}$$

$$\hat{d}_{i,E}(t) = d_{i,E}(t-\tau). \tag{2.44}$$

The fractional frequency $y_{i,E}(t)$ is a dimensionless parameter that denotes the frequency difference between clock i and the generated time scale as a ratio to the nominal frequency.

The fractional frequency is hereby referred to as frequency for brevity. The drift parameter $d_{i,E}(t)$ is only necessary if the impact of drift is significant on the clocks over the interval τ between predictions. The clocks used in the simulations of Section 3.3 have negligible drift so $d_{i,E}(t) = 0$. The term $y_{i,E}(t)$ is expressed with respect to the time scale to show that it is an estimated parameter, just like $x_{i,E}(t)$. In general, the frequency is simply the rate of change of the phase, i.e., $y(t)$ would be the rate of change of $h(t)$. However, when referring to $y_{i,E}(t)$, we do not simply find the rate of change of $x_{i,E}(t)$ but follow a different estimation process. The final frequency estimates will be explained in Section 2.4.4 because they are obtained with a certain weighting process.

Certain stochastic noises exist in clock technologies that are unpredictable and result in predictions that vary randomly from the true clock states. For consistency in the methods presented in this paper, a statistical model is introduced involving the initial definition of AT1 predictions assuming Gaussian distributions for $h_i(t), y_i(t), d_i(t)$:

$$h_i(t) \sim \mathcal{N}(\hat{x}_{i,E}(t), \epsilon^2_{x_i}(t)), \tag{2.45}$$

$$y_i(t) \sim \mathcal{N}(\hat{y}_{i,E}(t), \epsilon^2_{y_i}(t)), \tag{2.46}$$

$$d_i(t) \sim \mathcal{N}(\hat{d}_{i,E}(t), \epsilon^2_{d_i}(t)). \tag{2.47}$$

The principle of these distributions is that a predictable clock would follow the prediction equations perfectly. However, the true clock states deviate from the predictions by some random amount at each time epoch. The distribution of possible values that the true clock states can have is initially assumed to be Gaussian. In addition, we are assuming no correlation between the clock states, whereas a multivariate solution such as a KF algorithm would be capable of implementing correlated noise. The variances for each of the associated distributions are denoted as $\epsilon^2_{x_i}(t)$, $\epsilon^2_{y_i}(t)$, and $\epsilon^2_{d_i}(t)$. They are not necessarily linked to a specific noise type in the stochastic clock model but rather a collection of all the internal clock noises that contribute to the prediction errors. The equations for the phase prediction variances are presented in Section 2.4.4 because the weights are based on these prediction variances.

2.4.3 BTSE Residuals

The BTSE residuals are presented as the difference between the predictions and the measurements $r_{ji}(t) = \hat{x}_{j,E}(t) - z_{ji}(t)$, which are the exact terms being weighted in the BTSE. Based on the above definitions of the measurements and predictions, the statistical model for the BTSE residuals can be written as

$$r_{ji}(t) \sim \mathcal{N}(x_{i,E}(t), \epsilon^2_{x_j}(t) + \epsilon^2_{x_i}(t) + \sigma^2_{ji}(t)). \tag{2.48}$$

Each of the above residuals differs from a predictable clock (represented by the output of the phase predictions) with variance described by $\epsilon^2_{x_j} + \epsilon^2_{x_i} + \sigma^2_{ji}$. Since all of the residuals used to compute $x_{i,E}(t)$ will have the same term $\epsilon^2_{x_i}(t)$ in the variance, the logical choice of weights

would be the inverse of $\epsilon_{x_j}^2(t) + \sigma_{ji}^2(t)$. The AT1 algorithm neglects the measurement noise variance σ_{ji}^2 and aims to estimate the error of each clock from the prediction by estimating each ϵ_{x_j}. Despite neglecting measurement noise, any added variance due to measurement noise would still impact the variance estimate and, as a result, the weights.

2.4.4 Weights

The BTSE realizations $x_{i,E}(t)$ are computed with weights defined from the previous time epoch for the basic AT1 algorithm. The resulting values should provide a time scale with improved frequency stability compared to the individual clocks. This is done by assigning lower weights to less predictable clocks. As mentioned above, the variance of the BTSE residuals would be a good metric for placing small weights on the clocks that vary the most from a predictable clock. For weight computation, the unique variance of the BTSE residuals $r_{ji}(t)$ is given by only $\epsilon_{x_j}^2$. This is called a unique variance because the variance $\epsilon_{x_i}^2$ is common to each observation and should not affect the weights. The unique phase variances of each clock can be expressed with the fundamental definition of variance for a zero-mean dataset

$$\epsilon_{x_j}^2(t) = E\left[\left(\hat{x}_{j,E}(t) - h_j(t)\right)^2\right], \tag{2.49}$$

where $E[\cdot]$ denotes the expected value. The above variance cannot be computed exactly because we cannot access $h_j(t)$ as a direct measurement. The next best approximation is the estimate of $x_{j,E}(t)$ using the previously computed weights. The error between the phase predictions and the result of the BTSE at time t provides an approximation of the prediction standard deviation

$$\hat{\epsilon}_j(t) = \left|\hat{x}_{j,E}(t) - x_{j,E}(t)\right| C_j(t-\tau), \tag{2.50}$$

where the variance is now denoted $\epsilon_j^2(t)$ to signify that it is not a realization of the actual phase variance. There is a bias induced on the variance according to the previous weights that is compensated by multiplying by the term $C_j(t-\tau) = \frac{1}{\sqrt{1-w_j(t-\tau)}}$ [67]. The estimated error is then squared and subjected to an exponential filter [68, 41, 48]

$$\epsilon_i^2(t) = \frac{\hat{\epsilon}_i^2(t) + N_\tau \epsilon_i^2(t-\tau)}{1 + N_\tau}. \tag{2.51}$$

The time constant N_τ for the prediction errors is chosen to be the sampling interval at which the white frequency noise of the clocks becomes dominated by flicker frequency noise [40], i.e., the weights are not permitted to change rapidly outside the time interval at which the frequency stability is best for the corresponding clocks. Finally, the AT1 weights are calculated with the inverse of the filtered prediction error, then normalized so the sum of all weights equals 1, i.e.,

$$w_i(t) = \frac{\frac{1}{\epsilon_i^2(t)}}{\sum_{i=1}^{N} \frac{1}{\epsilon_i^2(t)}}. \tag{2.52}$$

A limitation is placed on the maximum allowable value of the weights so the time scale does not become too reliant on any single clock [41]. According to (2.50), this algorithm requires the BTSE to be computed before the weights at time t can be determined. Hence, anomaly detection is needed to identify whether the current clock data is impacted by anomalies and if corrections are needed.

The weighting steps of the AT1 algorithm apply weights to not only clock phases but also clock frequencies. We consider the rate of change of the weighted phase estimate over the time interval τ to be our first approximation of the frequency:

$$y_{i,s}(t) = \frac{x_{i,E}(t) - x_{i,E}(t-\tau)}{\tau}, \tag{2.53}$$

where any corrections of phase jumps should be applied before making the first approximation of frequency. The AT1 algorithm proposes an exponential filter on the frequency approximation. This assigns a certain weight (related to a time constant m_i) to a prediction of the frequency and ensures a gradual change to the new intermediate frequency. The time constant m_i can be specifically tuned to provide the best result for each type of clock in the ensemble [48]. The same value as N_τ can be used for the frequency time constant for clocks that have significant flicker frequency noise [48].

$$y_{i,\mathrm{AT1}}(t) = \frac{y_{i,s}(t) + m_i \hat{y}_{i,AT1}(t)}{1 + m_i}. \tag{2.54}$$

The frequency estimation is the only difference between the AT1 and AT2 algorithms. The AT2 algorithm instead estimates frequency using a KF. A univariate KF is defined for each of the N clocks, with the predictions from (2.43) varying according to the random walk Frequency Modulation (FM) and the "measurements" from (2.53) varying due to the white FM. This is shown to be equivalent to another exponential filter with different definitions of time constant depending on the variances of predictions and measurements [48, 40]. Due to this equivalence with the exponential filter, the original AT1 frequency estimation is retained in this work to simplify the choice of time constants. Normally, the AT2 algorithm is used as a method of detecting frequency jumps according to a certain threshold. The next section will discuss the methods of choosing thresholds to detect anomalies. It will also establish the AT1 oracle time scale as a baseline for assessing the robustness of the new time scale.

2.4.5 AT1 Phase jump detection

In the description of the time step detection of the AT1 algorithm, new weights should be used to recompute the ensemble average, i.e., the BTSE when the prediction error exceeds a given threshold [40, 41]. Each of the references have a different smoothing function to adjust the weights gradually for prediction errors that are still close to the detection

threshold.

$$wct_i = 1, k_i \leq 3, \tag{2.55}$$

$$wct_i = 1 - (3 - k_i)^2, 3 < k_i < 4, \tag{2.56}$$

$$wct_i = 4 - k_i, 3 < k_i < 4, \tag{2.57}$$

$$wct_i = 0, k_i \geq 4, \tag{2.58}$$

where (2.56) is defined in [40], (2.57) is defined in [41], and the test parameter is given by:

$$k_i = \frac{|x_{i,E}(t) - \hat{x}_{i,E}(t)|}{\epsilon_i}. \tag{2.59}$$

Weight control terms are used to rescale the weights before recomputing the BTSE with the newly modified weights. This somewhat mimics the idea of robust estimation that adapts the weights according to an outlier score, gradually reduction to zero. Hence, there is motivation to introduce robust estimation to the AT1 algorithm to provide a more rigorous definition of the weight adjustment for anomalous clocks and no need to choose a specific threshold for detection.

2.4.6 AT2 Frequency jump detection

The measurement variance $R_i(t)$ is given by the estimated Allan variance over the sampling interval that corresponds to white frequency noise [40]. The other variance term is computed according to the KF,

$$\hat{P}_i(t|t-\tau) = \hat{P}_i(t-\tau|t-\tau) + Q_i(t), \tag{2.60}$$

where the process noise $Q_i(t)$ is the Allan variance associated with random walk frequency noise. Following the formulation of the standard KF algorithm expressed in the previous section, we obtain the update equations

$$y_{i,\mathrm{AT2}}(t) = \frac{R_i(t)\hat{y}_{i,AT2}(t) + \hat{P}_i(t|t-\tau)y_{i,s}(t)}{R_i(t) + \hat{P}_i(t|t-\tau)}, \tag{2.61}$$

$$\hat{P}_i(t) = \frac{R_i(t)\hat{P}_i(t|t-\tau)}{R_i(t) + \hat{P}_i(t|t-\tau)}. \tag{2.62}$$

In the one-dimensional case, the KF update equation is equivalent to an exponential filter. This time, the time constants are replaced by estimates of the measurement noise $R_i(t)$ and the prediction error variance $\hat{P}_i(t)$, assuming that clock frequencies are uncorrelated. The estimates of the variance parameters can be reliably obtained for laboratory clocks. An ensemble of clocks in space would not necessarily have such reliable models of measurement noise or prediction error due to the increased probability of anomalies. Hence, there are

additional procedures put in place to detect such anomalies and recompute the BTSE [48, 40, 41]. For example, the test statistic used for detecting time steps in AT1 is the phase prediction error

$$\kappa_{x_i} = \left| x_{i,E}(t) - \hat{x}_{i,E} \right|, \tag{2.63}$$

and the threshold is linked to the standard deviation of the phase predictions. The hypothesis for the case with no anomalies is that the test statistic follows a Gaussian distribution, it is then simple to choose a threshold of three standard deviations. The test statistic tells us if there is a phase jump anomaly if

$$\kappa_{x_i} > 3\epsilon_i. \tag{2.64}$$

The proposed test statistic for a frequency jump is also reasonable, being the difference between the average frequency over some past interval of length L, $y_{i,\text{avg}}$, and the frequency estimate at the beginning of that interval $y_{i,E}(t - L\tau)$

$$\kappa_{y_i} = \left| y_{i,\text{avg}}(t) - y_{i,E}(t - L\tau) \right|. \tag{2.65}$$

Conversely, the process of deciding the threshold for frequency jumps is more convoluted. Estimates of the white FM noise variance (σ_α^2) and the random walk FM noise variance (σ_β^2) are required to determine another test variance (σ_L^2) that provides the threshold. The choice of threshold is then four standard deviations (using this test variance) "because it was empirically found to be appropriate" [40]. The details of the threshold for frequency jump detection are described in [40]. As with the other tests, the frequency jumps are detected if the following condition is satisfied:

$$\kappa_{y_i} > 4\sigma_L. \tag{2.66}$$

It is suggested to use several different length values L to be sure of a frequency jump since it occurs over a continuous time period. The estimates of the white and random walk noises can already be obtained in the AT2 frequency update step

$$\sigma_{\alpha_i}^2 = \tau_0 \tau_{x_i} R_i, \tag{2.67}$$

$$\sigma_{\beta_i}^2 = \frac{3n}{2n^2 + 1} Q_i, \tag{2.68}$$

where τ_0 is the smallest time interval between measurements, τ_{x_i} is the specific measurement interval for clock i if there are periods of missing data, and $n = \frac{\tau_{y_i}}{\tau_0}$. We assume there are no periods of missing data in this work so we take $\tau_{x_i} = \tau_0$ for all clocks. Contrarily, we still use different τ_{y_i} to avoid frequency samples contaminated by phase jumps. Next, the test variance is computed using an interval of past frequency values with the following length:

$$L_{\max} = \frac{1}{2} \left(\sqrt{1 + 4 \left(\frac{\sigma_\alpha}{\tau_0 \sigma_\beta} \right)^2} - 1 \right) \tag{2.69}$$

2.4.7 Oracle detection

The AT1/AT2 jump detection methods are not as rigorous as those defined with the likelihood ratio test. As a result, it is hard to assess the performance for different magnitudes of phase jumps, frequency jumps, and anomalies in measurement links. In addition, the delay required to detect and identify a frequency jump is not compatible with the swarm application, which must be robust and autonomous without needing to go back in time and correct a frequency jump in post-processing.

An example of perfect detection can be used instead of comparing a new algorithm to the pre-defined anomaly detection for AT1/AT2, which is not straightforward to implement in the same manner as the original designers. We can use a priori knowledge of simulated anomalies to generate the AT1 oracle time scale. The AT1 oracle time scale is an example of achievable performance if we compensate the phase jumps, frequency jumps, and link anomalies at the exact time they occur because it is known in advance that they will occur at that time. This allows the time scale to be corrected almost perfectly by setting the relevant weights to zero at the exact time of the anomaly. That is, the BTSE becomes:

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t) \left[\hat{x}_{j,E}(t) - x_{ji}(t)\right], \tag{2.70}$$

after computing $w_j(t)$ and knowing that an anomaly occurs at time t. The weights remain high for reliable clocks, keeping them in the ensemble, whereas clocks with low weights have their impact on the ensemble reduced. For phase jump anomalies, the weights are expected to be small due to a larger-than-normal prediction error. Frequency jumps or drifts that may initially have small impacts on the measured clock phases may not necessarily result in computing a low weight. This is where the AT1 oracle algorithm forces the weights to be zero at the exact time of the anomaly to remove the contribution of the anomalous clock or measurement. The method of iteratively applying the weights is similar to the method suggested to compute the EAL time scale in [46], driving weights to some low value (or zero) upon detecting abnormal behavior. The weights of clocks with any measurement links that have an anomaly are forced to zero at the time of the anomaly.

The primary contribution of this thesis aims to design a robust procedure that assigns independent weights to each measurement based on the observations obtained at each instant of time. This is similar to the robust estimation discussed in Section 2.3.5 for implementation in the KF-based algorithm but compatible with the AT1 formulation. The choice to adapt robust estimation to an "AT1-like" time scale algorithm is based on the ability to more explicitly understand the functions of the weights and to provide a direct relation to the statistical model, allowing later studies and adaptations based on other potential models. The next chapter presents the primary contribution of this thesis, being a novel method of deweighting the clock phase measurements according to a robust estimation algorithm.

Chapter 3

A New Robust Time Scale Algorithm

The Autonomous Time scale using the Student's T-distribution (ATST) is a new time scale algorithm that uses robust estimation to simultaneously mitigate the effects of phase jumps, frequency jumps, and anomalous phase difference measurements. More precisely, it is assumed 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 [66]. Therefore, the Maximum Likelihood Estimators (MLEs) for the defining parameters of the Student's t-distribution are robust to anomalies that cause such outliers. The method of mitigating the impact of the outliers in the MLE is comparable to the method of assigning weights to individual clocks in the BTSE to build the time scale. Hence, the new ATST time scale is designed by taking advantage of this similarity.

This chapter presents robust estimation as a method of mitigating anomalies, which acts as an alternative to anomaly detection. Besides dealing with missing measurements, the detection procedures discussed in Chapter 2 for removing measurement anomalies are not necessary in the ATST algorithm. This provides a motivation to use the ATST algorithm because it includes an implicit autonomous detection of anomalies, although, at the cost of other restrictions that will be discussed in this chapter. To confirm the validity of a robust time scale, it should maintain good performance in the absence of anomalies and should not lose performance in the presence of anomalies. The ATST time scale algorithm is therefore assessed by comparing it to the AT1 algorithm in the nominal case and the AT1 algorithm with perfect detection of anomalies otherwise. AT1 with perfect detection is herein referred to as AT1 oracle because it effectively knows the occurrence of all anomalies before the moment they occur.

3.1 Anomalous Clocks and 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 (see Appendix E.1), generating the new autonomous time scale using the Student's T-distribution referred to as ATST. To understand how the ATST attains robustness, the Student's t-distribution should be explained. In general notation, a random variable X that is distributed according to a univariate t-distribution is denoted as $X \sim T(\mu, \sigma^2, \nu)$. 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 $\text{var}(X) = \sigma^2 \frac{\nu}{\nu-2}$ 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 3.1 illustrates this connection between the shape parameter and the normality of the data.

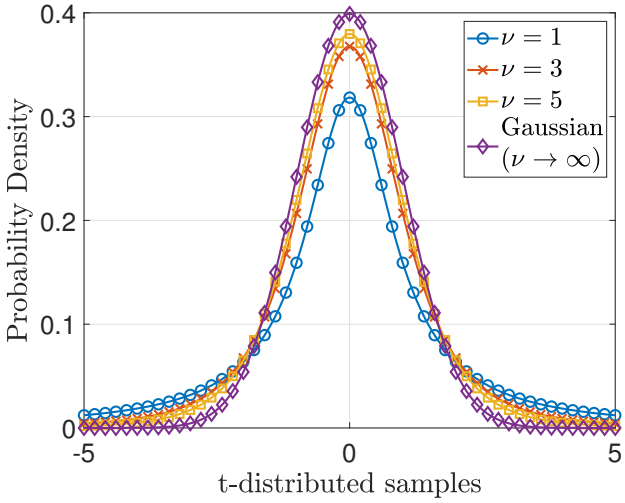


Figure 3.1: Examples of the Student's t-distribution probability density functions for different numbers of degrees of freedom ν. The distribution converges to a Gaussian distribution for infinite value of ν. Lower values of ν correspond to heavier probabilities in the tails, hence a greater proportion 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 given probability of outliers occurring. The PDF defines the required equations to estimate the parameters of the distribution as described in Appendix E.1. These equations show that the degrees of freedom also contribute to robust estimates of the other parameters, i.e., the presence of anomalies is accounted for in the estimates of the mean and variance. The MLE of the mean is a weighted average of the data according to the Expectation Maximization (EM) algorithm specified in Appendix E.1, where the weights are lower for less probable observations according to the PDF of the Student's t-distribution. To summarize, the proposed ATST time scale is generated by assuming the clock data is modeled by a Student's t-distribution and using the robust estimate of the mean of that data as a realization of the robust time scale.

3.1.1 Clock Measurements

The proposed ATST algorithm considers all unique pairs of satellites as sources of independent phase difference measurements. A unique measurement noise is present on each satellite link. As was defined in section 2.2, the phase difference measurements are

$$z_{j,i}(t) = x_{j,i}(t) + n_{j,i}(t). \tag{3.1}$$

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

$$z_{j,i}(t) = x_{j,i}(t) + n_{j,i}(t) + \Delta n_{j,i}(t) = x_{j,i}(t) + \tilde{n}_{j,i}(t). \tag{3.2}$$

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.

3.1.2 Clock Predictions

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

$$\hat{x}_{i,E}(t) = x_{i,E}(t-\tau) + \tau y_{i,E}(t-\tau) + \frac{\tau^2}{2} d_{i,E}(t-\tau). \tag{3.3}$$

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

$$\tilde{h}_i(t) = h_i(t) + \Delta h_i(t). \tag{3.4}$$

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

$$\tilde{e}_i(t) = \tilde{h}_i(t) - \hat{x}_{i,E}(t) = e_i(t) + \Delta h_i(t), \tag{3.5}$$

where we recall here that the prediction error is $e_i(t) = h_i(t) - \hat{x}_{i,E}(t)$. 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.

3.1.3 BTSE Residuals

The BTSE residuals are expressed in terms of the clock phase measurements affected by anomalies, e.g., corrupted clock j provides the phase $\tilde{h}_j(t)$ and/or an anomaly on link j, i introduces a contaminated noise $\tilde{n}_{j,i}(t)$: recall the definition of $z_{j,i}(t)$ as a function of $h_j(t)$ and $h_i(t)$

$$r_{j,i}(t) = \hat{x}_{j,E}(t) - z_{j,i}(t), \tag{3.6}$$

$$r_{j,i}(t) = \hat{x}_{j,E}(t) - (\tilde{h}_j(t) - h_i(t) + \tilde{n}_{j,i}(t)), \tag{3.7}$$

$$r_{j,i}(t) = h_i(t) + e_j(t) - \Delta h_j(t) - n_{j,i}(t) - \Delta n_{j,i}(t). \tag{3.8}$$

The objective of a time scale is to provide a common reference for any clock, $x_{i,E}(t) = h_i(t) - h_E(t)$. Without loss of generality, the noise due to stochastic processes on each clock $j = 1, \cdots, N$ and the measurement links j, i for fixed i are assumed to be noisy observations of the absolute phase of the time scale $h_E(t)$. The residuals can then be written in terms of the time scale that we aim to estimate

$$r_{j,i}(t) = h_i(t) - h_E(t) + b_{j,i}(t). \tag{3.9}$$

It is assumed that the effects of outliers due to internal clock malfunctions and link anomalies imply a Student's t-distribution for $b_{j,i}(t)$ and hence for the residuals $r_{j,i}(t)$. If the phase of clock i is affected by an anomaly, it is equivalent to estimate the actual deviation as $x_{i,E}(t) = \tilde{h}_i(t) - h_E(t)$, which allows steering of the corrupted clock to the common time scale. Additional assumptions allow simplification of the parameters for the defining distribution. The instantaneous measurement noise throughout the ensemble 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 can be modeled by a heavy-tailed distribution such as the Student's t-distribution:

$$r_{j,i}(t) \sim T\left(x_{i,E}(t), \sigma_i^2(t), \nu_i(t)\right). \tag{3.10}$$

The parameters of the above distribution depend only on the reference clock i because the phase of clock i is constant throughout the N BTSE residual samples. 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 $\sigma_i^2(t)$ is related to the dispersion of every other clock compared to clock i, and the shape parameter ν_i describes the impact of anomalies on the estimation of $x_{i,E}(t)$. The shape factor $\nu_i(t)$ is related to the heaviness of the tails so can be linked to the number and magnitude of 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 heavier 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 Student's 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.

3.1.4 Verifying the Student's t-distribution

The assumption of the Student's t-distribution is verified for this number of clocks in Figure 3.2, 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 3.2 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 $r_{jj}(t)$ causes the distribution to still be affected by an outlier. The following comments can be made:

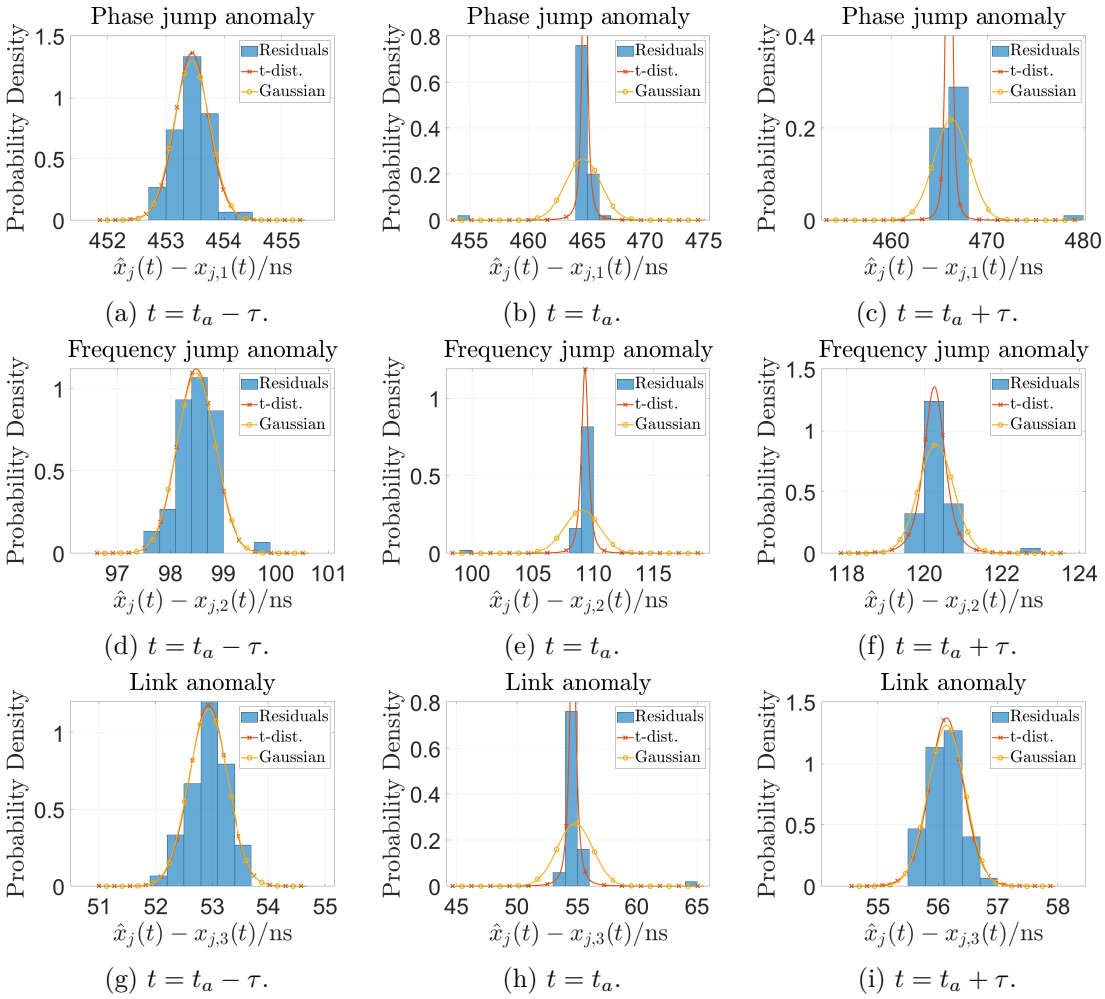
- As expected, the distribution of the BTSE residuals at the time immediately before an anomaly is well approximated by both the Gaussian and Student's t-distributions. Hence, we expect a good performance for the MLE of the t-distribution in the nominal case.

- 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. As can be observed at $t = t_a$, the histogram of the residuals is better fitted by the Student's t-distribution than by the Gaussian distribution.

- 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 one clock, resulting in a different rate of change for the phase of the affected clock. Additionally, 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. Again, the Student's t-distribution provides a good fit for the distribution of the residuals just after the time of the anomaly.

If the total number of clocks is too low, the number of samples for the BTSE residuals is too low to appropriately fit a statistical distribution. Space-based applications often function with redundant technologies to contribute to robustness, e.g., many clocks on many satellites. Hence, the limitation of having few observations is not an immediate concern but is relevant to dealing with missing data, which is addressed in Section 3.4. Simultaneously occurring anomalies would also change the distributions of the data, but it is expected that this simply changes the shape parameter of the Student's t-distribution so it remains a valid model.

The outliers 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. 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.



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

3.2 Weights of ATST Time Scale

The derivation of the MLE for the parameters of the Student's t-distribution is detailed in Appendix E.1. To generate the ATST, only the equations in the EM algorithm are necessary to obtain a weighted average of a certain dataset. More precisely, the t-distributed BTSE residuals for a fixed clock i, i.e., $r_{ji}(t) \sim T\left(\mu(t), \sigma^2(t), \nu(t)\right)$, are used to define the initial values of the estimates as follows:

$$\hat{\mu}_0(t) = \tfrac{1}{N}\sum_{j=1}^{N} r_{ji}(t), \quad \hat{\sigma}_0^2(t) = \tfrac{1}{N-1}\sum_{j=1}^{N}\left(r_{ji}(t) - \hat{\mu}_0(t)\right)^2, \quad \hat{\nu}_0(t) = 3. \tag{3.11}$$

Note that the initial value of the number of degrees of freedom $\nu_0 = 3$ is chosen to always initially assume that outliers are possibly present in the clock measurements. This initialization does not affect the results when there are no anomalies because the location and scale estimates using the Gaussian MLE are close to the Student's t-MLE in this context. Note also that the Student's t-MLE 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 based on the consecutive error between estimates of each parameter [69]. The resulting EM algorithm is described in Algorithm 1:

Algorithm 1 Expectation Maximization for a Robust MLE

function tEM($r_{j,i}(t)$)

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

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

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

ν_k estimated as the solution of the following equation for ν_k

$$N\left[\phi\left(\frac{\nu_k}{2}\right) - \phi\left(\frac{\hat{\nu}_{k-1}+1}{2}\right)\right] + \sum_{j=1}^{N}\left[u_{j,k} - \log(u_{j,k}) - 1\right] = 0. \tag{3.12}$$

$$\varepsilon = \left\|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}\right\| \tag{3.13}$$

k=k+1

end while

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

end function

where the subscript k denotes the iteration number of the EM algorithm executed with the observations obtained at time t. Note that another iterative algorithm (using Newton's method as specified in Appendix(E.22)) is needed to estimate the number of degrees of freedom satisfying (3.12) [69]. Once the EM algorithm has converged, the values of $u_{j,k}$ act as 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. These terms are also able to return to the Gaussian assumption in the case of nominal data because a large number of degrees of freedom results in the weights being approximately $1/N$ for each clock. However, this causes problems if a mixture of different clock types are present, meaning that the weights are not assigned based on each clock's unique stability. Hence, the new algorithm is first evaluated with an ensemble of homogeneous OCXO clocks to prove robustness before integrating different clock behaviors.

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 the residual $r_{j,i}$ and hence affect the weights of each individual clock equally. The MLE of μ, denoted as $\hat{\mu}$ in Algorithm 1 is obtained according to a weighted average of the BTSE residuals with the normalized weights $w_{j,ATST}(t) = u_{j,k}(t) \Big/ \sum_{i=1}^{N} u_{i,k}(t)$ as defined in the above algorithm. Hence, the robust MLE of the location parameter is equivalent to a BTSE with specifically designed weights. As detailed for the BTSE in Chapter 2, 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 as:

$$x_{i,\mathrm{ATST}}(t) = \mathrm{tEM}\left(r_{j,i}(t)\right) = \frac{\sum_{j=1}^{N} u_{j,k} r_{ji}(t)}{\sum_{j=1}^{N} u_{j,k}}, \tag{3.14}$$

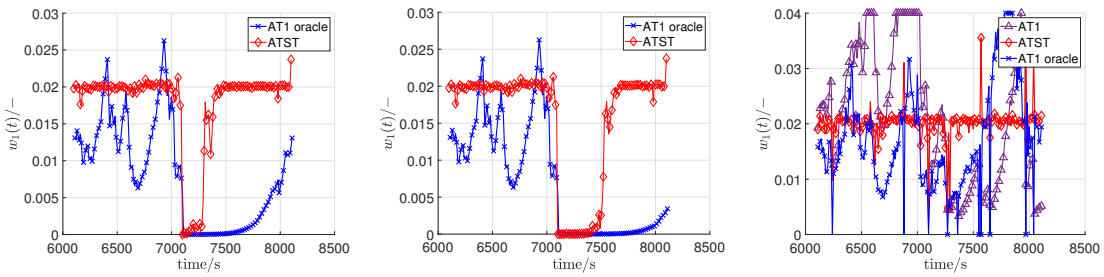
where $\mathrm{tEM}(r_{ji}(t))$ is the output from the EM algorithm and differs depending on which set of residuals is used at the input. At each iteration of the EM algorithm, the weights are further refined thanks to using the adapted estimates $\hat{\mu}_{k-1}$ and $\hat{\sigma}^2_{k-1}$. The iterative nature of the EM algorithm also inherently involves a recomputation of the BTSE at each iteration. This is also how AT1 oracle compensates phase jumps and frequency jumps by recomputing the BTSE using the weights computed at a given time instant, where the weights are set to zero whenever an anomaly is known. The ATST algorithm is essentially a recursive application of this procedure that re-adapts the weights before each computation in each iteration of the EM algorithm. However, instead of a detection threshold or a priori knowledge on the occurrence of anomalies, ATST uses a threshold for the convergence of the EM estimator to confirm that the weights are appropriately attributed to anomalous measurements. 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 recomputation.

The ATST weighting procedure differs from the standard BTSE because each set of residuals yields a unique set of weights, i.e., the weights are assigned based on specific inter-satellite links. The weights computed for estimating $x_{1,\mathrm{ATST}}(t)$ are not necessarily the same as the weights computed for $x_{2,\mathrm{ATST}}(t)$ because they each use a different set of inter-satellite links being $z_{j,1}(t)$ and $z_{j,2}(t)$, respectively. Despite this variation in the weights used for each realization of the BTSE, it can be shown that each estimate still results in a common time scale. By varying the weights according to the unique residuals, the algorithm takes into account the higher deviations of both measurements and clocks affected by anomalies. 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 will naturally ensure there is no dominant clock by assigning very similar weights at each point in time. This can be observed in Figure 3.3, where the weight for a single clock in an ensemble

of fifty clocks is concentrated around 1/50 except for time epochs with anomalies. Unless there is a significant number of anomalous clocks at the same time epoch, the weights of the nominal clocks will not increase drastically. This autonomous constraint on the maximum clock weight is advantageous because the AT1 algorithm normally requires monitoring to ensure no individual clock dominates the time scale. By keeping the weights of clocks unaffected by anomalies close to equally distributed around 1/50, the time scale will not be susceptible to sudden changes in clocks that are more stable at a certain time instant then face an anomaly at the next. However, this is not preferable for a time scale with a mixture of different clock types, e.g., the weights will not necessarily take into account the stability of a Cesium clock over long time intervals when compared to a crystal oscillator at a single time epoch.

Figure 3.3 displays the weights of the three clocks that are affected by different anomalies, and whose histograms are shown in Figure 3.2. 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 3.3: Weights of the ATST and AT1 time scale in the presence of the different types of anomalies. In the case of phase jumps and frequency jumps the weights are defined in the same way for AT1 and AT1 oracle so only one of them is plotted. When considering link anomalies, AT1 oracle manually sets the weight of clock 1 to zero whenever an anomaly is on a link including clock 1.

Since the estimation of the ith phase $x_{i,\mathrm{ATST}}(t)$ is already designed to be robust to phase jumps and frequency jumps, we expect the distribution of past frequencies to be unaffected by those anomalies. As a result, the frequency can be estimated using an average over a

window of past frequency values or use the same exponential filter as the AT1 algorithm

$$y_{i,\mathrm{ATST}}(t) = \frac{\hat{y}_{i,\mathrm{ATST}}(t) + m_i y_{i,s}(t)}{1 + m_i}. \tag{3.15}$$

The time constant m_i 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 must be initialized to obtain a first prediction of the phase. The clock phases are assumed to be measured with respect to a stable reference, obtaining $x_{i,r}(t)$ for the first 10 time steps. The algorithm would then start at the beginning of the eleventh time step assuming that the frequency over the first ten time steps is constant. It is also assumed that the drift is negligible, so it is initialized at a value of zero. With the complete definitions of the estimations used, the ATST algorithm is presented in Algorithm 2:

Algorithm 2 ATST Time Scale Generation

Inputs: $z_{j,i}(t)$, τ, $x_{i,E}(t-\tau)$
Init.: $x_{i,\mathrm{ATST}}(10\tau) = 0$, $y_{i,\mathrm{ATST}}(10\tau) = \frac{x_{i,r}(10\tau) - x_{i,r}(0)}{10\tau}$, $d_{i,\mathrm{ATST}}(10\tau) = 0$,
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**
    $r_{j,i}(t) = \hat{x}_{i,\mathrm{ATST}}(t) - z_{j,i}(t)$,
    $x_{i,\mathrm{ATST}}(t) = \mathrm{tEM}\left(r_{j,i}(t)\right)$,
    $i = i + 1$,
  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

3.3 Robustness of ATST time scale

To verify that the ATST time scale is robust to the anomalies of interest, we will assess the primary criteria against the requirements introduced in Chapter 1. That is, synchronization with respect to MTIE and TDEV, frequency stability in the form of the OADEV, and fractional frequency and phase evolution to confirm the continuity. The time and frequency stabilities of the new time scale should be unaffected by anomalous clocks 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.

3.3.1 Simulated data

The proposed robust time scale can be appropriately tested by using simulated clock data to obtain a high number of clocks corresponding to the clocks onboard the nanosatellites in the swarm. As was mentioned earlier, a homogeneous ensemble of OCXO clocks is used because of the nature of the ATST weights under nominal conditions. The simulation methodology and how it can be extended to other types of clocks is explained in Appendix C. The noise variance levels of each simulated clock were varied randomly by multiplying the variance by a normally distributed random variable to obtain unique but relatively similar stochastic behaviors.

For this work, the interval between measurements is fixed at 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 in the measurements. 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. For OCXO clocks, the time constants for the error and frequency estimates in the AT1 algorithm are established from the respective OADEV.

3.3.2 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

$$h_i(t) - x_{i,E}(t) = h_i(t) - (h_i(t) - h_E(t)) = h_E(t). \tag{3.16}$$

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

$$y_E(t) = \frac{h_E(t) - h_E(t-\tau)}{\tau}. \tag{3.17}$$

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 $\bar{y}_E$

$$h_E(t) := h_E(t) - t\bar{y}_E. \tag{3.18}$$

Finally, the OADEV is computed according to the standard formula for a given set of phase samples [11, 12]

$$\sigma_y^2(\tau) = \frac{1}{2(M-1)} \sum_{t=0}^{M-2} \left(\frac{h_E(t+2\tau) - 2h_E(t+\tau) + h_E(t)}{\tau} \right)^2, \tag{3.19}$$

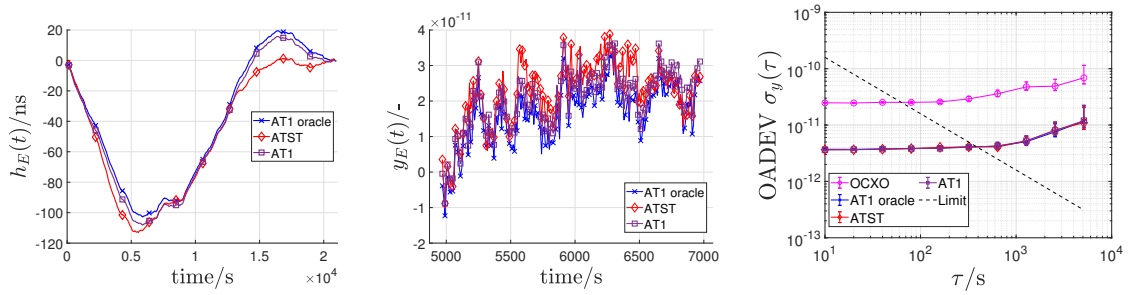
where $M-2$ realizations of the time scale phase are available, restricting the maximum sampling interval to $\tau = \frac{M-2}{2}$. Confidence intervals are included on the computed values of OADEV to indicate the increase in uncertainty of the OADEV as the number of samples decreases for higher sampling intervals [17].

The ATST time scale is first validated under nominal conditions before evaluating the responses to the anomalies of interest. Figure 3.4 displays the phase, frequency, and frequency stability of the ATST algorithm alongside the AT1 algorithm without any anomalies. It is confirmed that when all the clocks are nominal throughout the simulation period, the ATST time scale obtains very similar performance to the AT1 algorithm. This is true because the weights are approximately $1/N$ and each of the clocks have similar performances.

The criteria used to assess the synchronization of the time scale are displayed next in Figure 3.5, where the ATST algorithm is shown to match the performance of the AT1 time scale under nominal conditions. The validity of the new time scale is confirmed by demonstrating the equivalence of both time and frequency stability compared to an existing, optimal time scale such as AT1. These figures also demonstrate how the time scale expands the range of functional sampling intervals in which the requirements stated in Chapter 1 are satisfied.

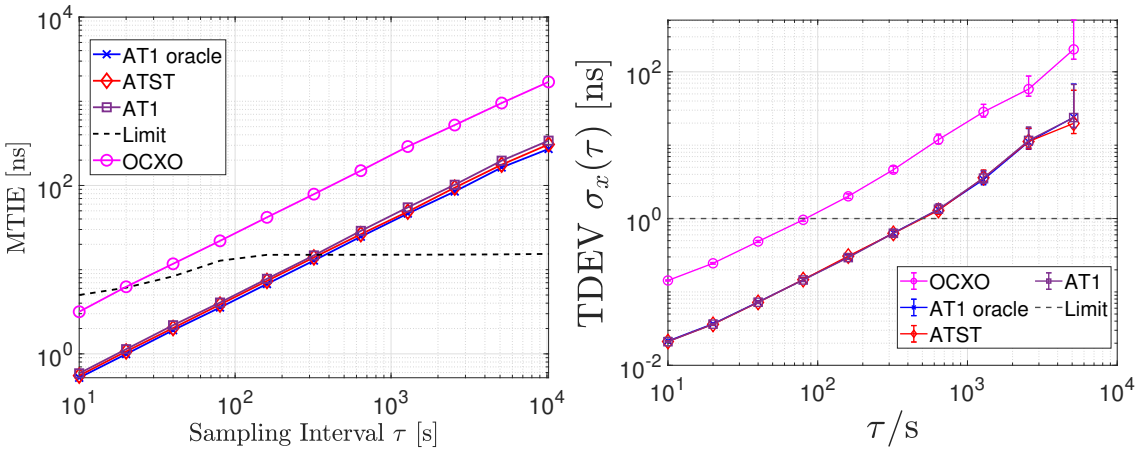
3.3.3 Measurement Noise

The effect of Gaussian measurement noise on the time scales is presented before assessing the effects of link anomalies. Figure 3.6 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



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

Figure 3.4: Nominal stability performance 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.



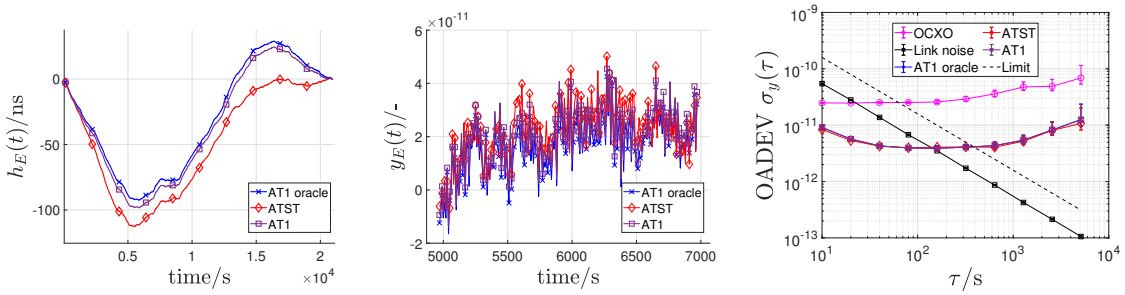
(a) Max time error. (b) Time dev.

Figure 3.5: Nominal timing performances of the ATST and AT1 oracle time scales using the above-mentioned ensemble of fifty OCXO clocks.

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 3.6c, the short-term OADEV of the time scales is affected according to the OADEV of the link noise. Since the measurement noise is included 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.

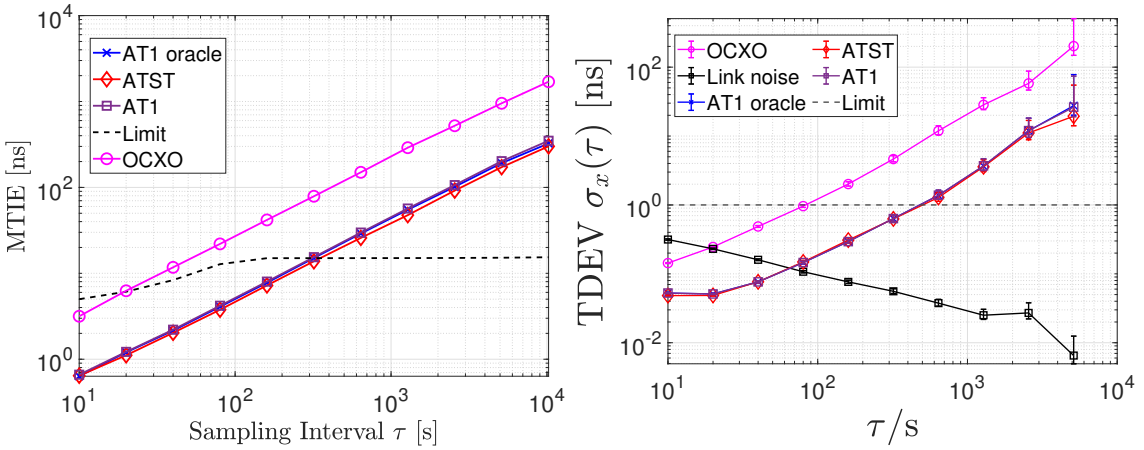
Figure 3.6: Continuity and stability performance of the time scales using the above-mentioned ensemble of fifty OCXO clocks and a uniform noise on every satellite link with variance $\sigma_n^2 = 10^{-19}$.

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. The effect of measurement noise on the MTIE and TDEV is also limited to the short term, with each of the time scales compensating the noise by the same amount.

3.3.4 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 throughout the measurement process, so each unique link is simulated with an anomaly at a random point in time.

Figure 3.8 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



(a) Max time error. (b) Time dev.

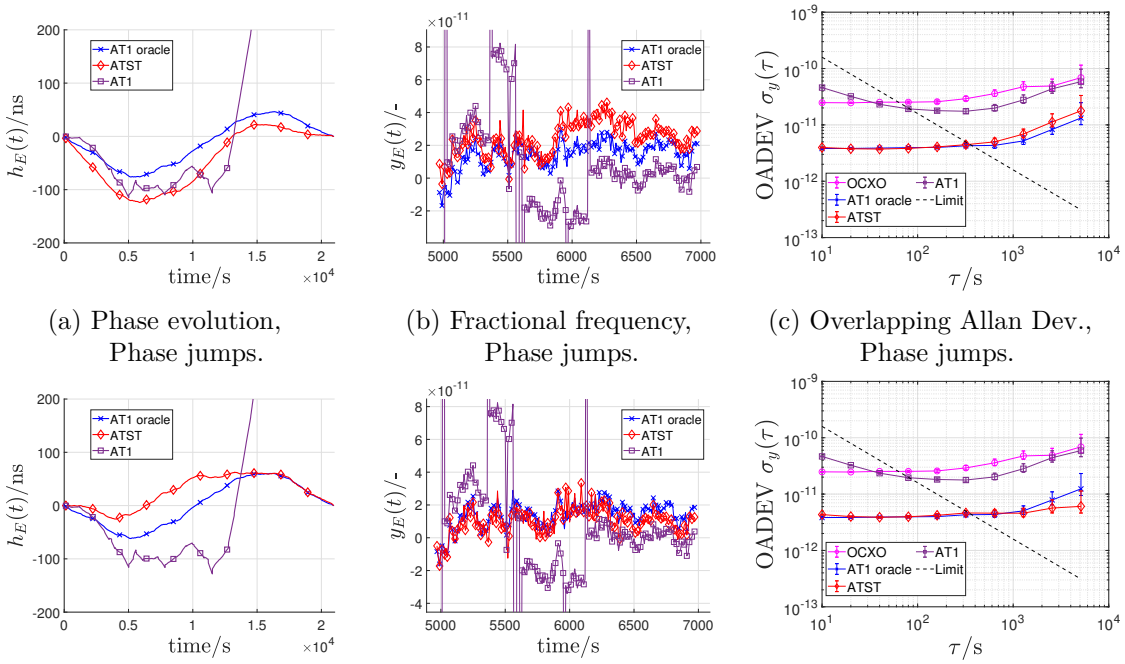
Figure 3.7: Timing accuracy assessed for the ATST and AT1 oracle time scales with uniform measurement noise on every satellite link with variance $\sigma_n^2 = 10^{-19}$.

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 section are presented so that the impact is visible on the ensemble of 50 clocks. For conciseness of this chapter, other simulations with smaller or larger anomaly magnitudes are available in Appendix D, where the same level of robustness is confirmed.

Figures 3.8b and 3.8e 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 an order of magnitude similar to the phase jumps are introduced on every unique link to generate the link anomalies. This magnitude of the link anomalies in Figure 3.11 is chosen to cause the resulting contaminated measurement noise to exceed the original measurement noise with a variance of 10^{-19}. Alternatively, the measurement noise can be reduced to a negligible level so that link anomalies with a magnitude of 10 ns are impactful. The results with such lower intensity anomalies are shown in Appendix D. Appendix D also indicates the responses of the time scales with link anomalies of larger magnitudes to prove that there is not an upper limit on the magnitude of anomalies that can be mitigated.

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 3.8c. 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 3.8b, where the effect of a phase jump on the frequency of AT1 is identical to that seen in Figure 3.8e 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.

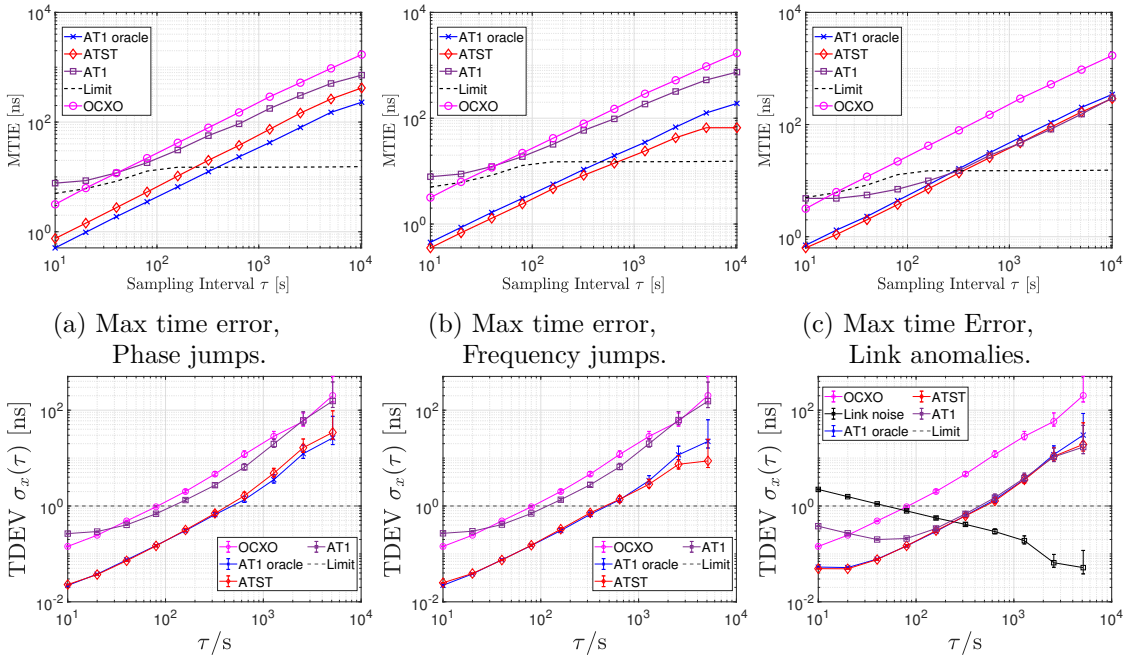
(e) Fractional frequency, Frequency jumps.

(f) Overlapping Allan Dev., Frequency jumps.

Figure 3.8: 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.

The following figure presents the metrics that classify the synchronization performance for the time scale. Since the MTIE and TDEV have remained relatively unchanged in Figure 3.9 compared to the nominal case, the errors in synchronization of any clocks steered to the time scales are verified to be robust to the presence of anomalies. Again, the ATST time scale is able to achieve very similar (or close) performance to the AT1 oracle time scale, and the time scales are able to achieve the required performance for more values of sampling intervals. The difference in the MTIE performance in the case of phase jumps is approximately constant for all sampling intervals so is assumed to be linked to a frequency offset between the two time scales that can be later compensated.

Figure 3.10 demonstrates that the ATST weights differ to the AT1 weights while still



(a) Max time error, Phase jumps.

(b) Max time error, Frequency jumps.

(c) Max time Error, Link anomalies.

(d) Time dev., Phase jumps.

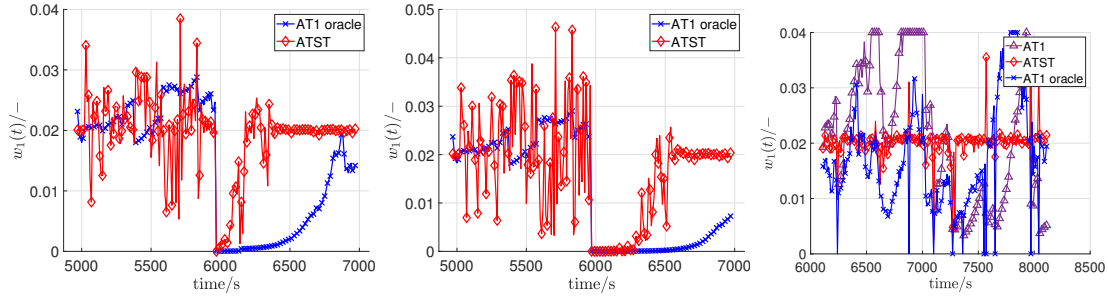
(e) Time dev., Frequency jumps.

(f) Time dev., Link anomalies.

Figure 3.9: Maximum Time Interval Error (top row) and Time Deviation (bottom row) for the new time scale compared to the AT1 oracle with phase jumps (left column), frequency jumps (middle column), and anomalies on the measurement links (right column).

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.

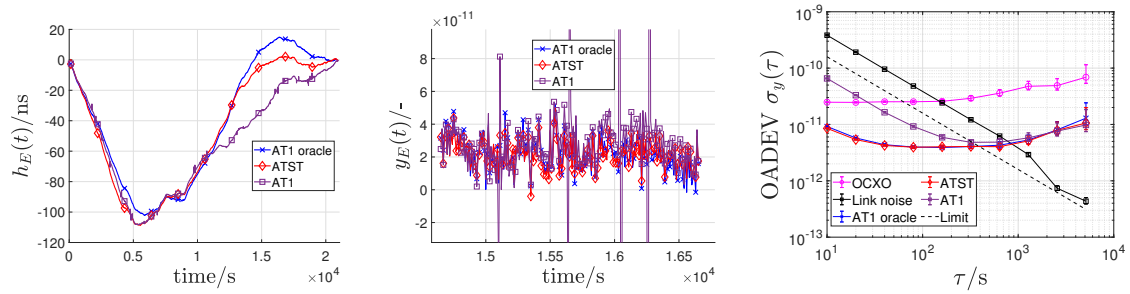
Figures 3.11a, 3.11b, and 3.11c present the robustness of the ATST time scale to link anomalies. The metrics linked to synchronization are detailed above in Figures 3.9e and 3.9f, where the anomalies are appropriately mitigated. This is a novel contribution because the measurement anomalies have previously been neglected in the design of time scales. The AT1 oracle time scale has a priori knowledge of the link anomalies and sets $w_i(t_a) = w_j(t_a) = 0$ for a link anomaly on measurement $z_{j,i}(t_a)$. The result is a time scale that can remove all link anomalies perfectly. In practice, this is not compatible with real time processing. However, 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



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

Figure 3.10: 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 3.3, the weights are seen to change more frequently due to anomalies occurring in other clocks in the ensemble.

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 a similar phase evolution, it can be concluded that ATST is robust to all of these anomalies, as designed in Section 3.1.

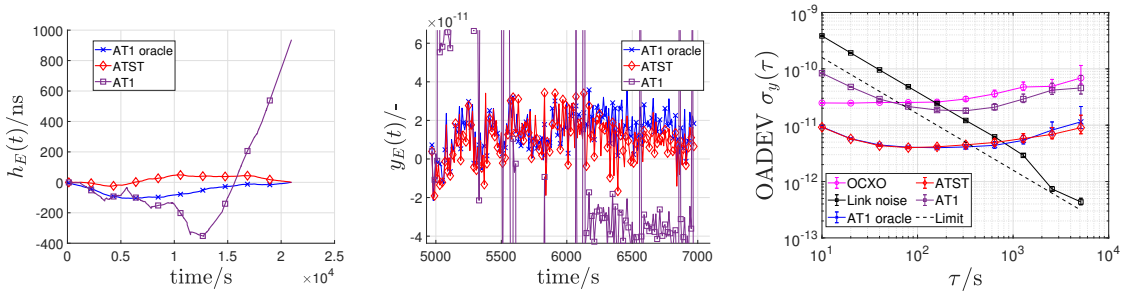


(a) Phase evolution, Link anomalies. (b) Fractional frequency, Link anomalies. (c) Overlapping Allan Dev., Link anomalies.

Figure 3.11: 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 3.12, 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.

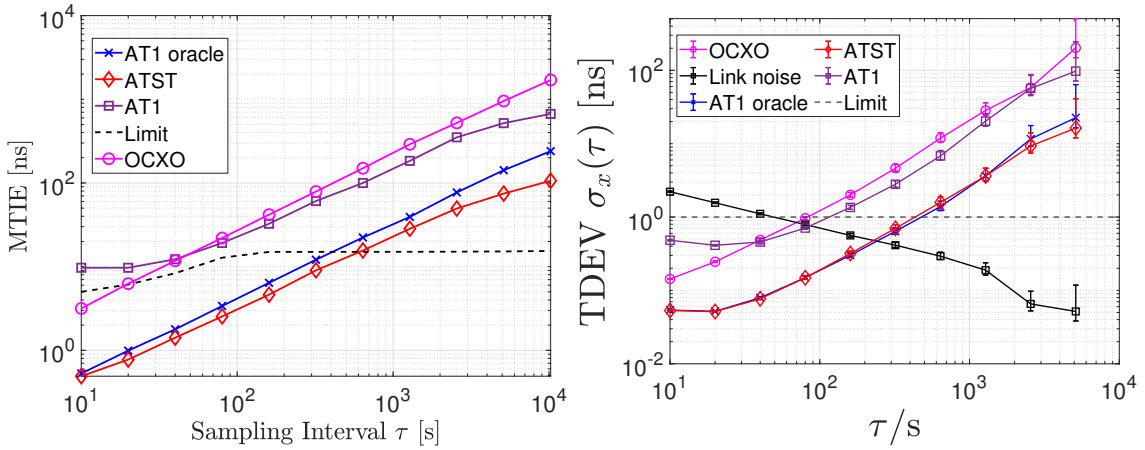
(c) Overlapping Allan Dev., All Anomalie.s

Figure 3.12: 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 ATST algorithm has also mitigated the effects of the combined anomalies on the synchronization performance, as detailed in Figure 3.13. Again the newly proposed robust time scale is able to match the performance of the AT1 oracle algorithm within some margin of error and the required limit is only exceeded at longer time intervals. Compared to the case with phase jump anomalies where AT1 oracle achieved a lower MTIE, the ATST algorithm is now shown to obtain slightly better MTIE in this case. This suggests some level of uncertainty in the estimate of MTIE, which should be better explored for future analyses.

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. As explained in section 3.1, the BTSE residuals can only be modeled by the Student's t-distribution if enough samples (clocks) are available.

Further analysis on the missing data problem is presented in Section 3.4 and an idea of the required minimum number of clocks is inferred from the results in Chapter 4 on optimum estimation performance. Additionally, ATST computes the weights and the time scale in an iterative EM algorithm, which introduces additional computational complex-



(a) Max time error. (b) Time dev.

Figure 3.13: Timing accuracy compared for the ATST and AT1 oracle time scales with each of the above anomalies occurring in the swarm.

ity when compared to AT1 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. A proposal to combine the ATST with the AT1 algorithm and gain the ability to profit from the performances of different clock types is explained in Chapter 5.

3.4 Missing data

A majority of time scale algorithms neglect measurement noise in their definitions. This is because measurement noise is usually dealt with in pre-processing. Any remaining noise after initial treatment is usually a simple white phase noise that only slightly affects the short-term stability of the resulting time scale. For a swarm of nanosatellites, the presence of measurement noise cannot be neglected. More importantly, the potential errors in the measurement process can act as another source of anomalies and cause certain pairs of satellites to have even more significant measurement noise. The measured clock bias data will hence be corrupted by non-uniform measurement noise with outliers. In addition, any averaging of the noise over a long period of time is assumed to be incompatible with the scientific mission of the swarm because the period of time between collecting clock data and making scientific observations is too short. A contribution of this thesis is to consider all possible measurements that can be made in the swarm at a single time instant, and assess the ability to reduce the measurement noise. The amount of possible measurements

is expected to change as the relative positions of the satellites vary. Therefore, a certain portion of the measurements thought to be available could be lost at a given time instant. The impact of losing different amounts of measurements is also assessed in the processing of measurement noise.

Two cases of missing measurements are considered: (i) missing inter-satellite links, and (ii) missing satellites. The first case refers to the link between a specific pair of satellites being lost, whereas the second scenario would mean a certain satellite has lost contact with all other satellites in the swarm. The reason for losing measurements can be because a satellite has drifted too far from its neighbors temporarily, a malfunction has occurred at some point in the communication chain, or measurements can be deliberately removed because they are identified as anomalous. There are plenty of methods that could be implemented to detect anomalies in a general data set but there has not been any publication on detecting anomalies in the clock bias measurements made in satellite constellations. Certain detection methods are aimed at anomalies in the clocks themselves but do not consider the possibility of a corrupt measurement being confused for a phase jump in the measured clocks. Basing anomaly detection on the measured clock biases helps eliminate specific bad links in the swarm that could corrupt the estimates even after applying phase jump detection algorithms like those proposed in earlier chapters.

3.4.1 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:

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - (x_{ji}(t) + n_{ji}(t))\right], \tag{3.20}$$

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - x_{ji}(t)\right] - \sum_{j=1}^{N} w_j(t-\tau) n_{ji}(t), \tag{3.21}$$

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.

The stochastic processes of an oscillator are not necessarily the only noise that contributes to the stability of the time scale. When clock comparisons are made between satellites, a non-negligible noise is expected to be introduced in the measurements. This modifies the system of equations with an additive noise of $n_{ji}(t)$ on each unique link between satellites j and i. As a result, the BTSE no longer provides a unique reference for every solution:

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - (x_{j,i}(t) + n_{j,i}(t))\right], \tag{3.22}$$

which can be separated into the classical BTSE and an additional weighted average of the measurement noise:

$$x_{i,E}(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[\hat{x}_{j,E}(t) - x_{j,i}(t)\right] + \sum_{j=1}^{N} w_j(t-\tau)n_{j,i}(t). \tag{3.23}$$

The resulting expression for the absolute time of the time scale now depends on the reference clock i that is used for observing the inter-satellite time differences:

$$\tilde{h}_E(t) = \sum_{j=1}^{N} w_j(t-\tau)\left[h_j(t) - \hat{x}_{j,E}(t)\right] + \sum_{j=1}^{N} w_j(t-\tau)n_{j,i}(t) = h_E(t) + \sum_{j=1}^{N} w_j(t-\tau)n_{j,i}(t). \tag{3.24}$$

Depending on the number of clocks and amplitude of the measurement noise, the weighted average may be sufficient to eliminate the noise or reduce it to a level that is negligible compared to the variance of the time scale. If the noise is not fully mitigated by the weighted average, then the generated time scale will not be a common reference amongst all the satellites because the noise is unique to each clocks.

The measurements with and without measurement noise can be defined in two vectors:

$$\mathbf{z}(t) = [z_{1,2}(t), \cdots, z_{1,N}(t), z_{2,3}(t), \cdots, z_{(N-1),N}(t)]^T, \tag{3.25}$$

$$\mathbf{x}(t) = [x_{1,2}(t), \cdots, x_{1,N}(t)]^T, \tag{3.26}$$

where the measurements with noise $\mathbf{z}(t)$ include all possible unique links between satellites and their associated noise; however, the measurements without noise in $\mathbf{x}(t)$ are simply the $N-1$ unique and non-redundant measurements required to compute the BTSE. Anomalous clocks or measurements will result in a larger deviation from the prediction than if there was no problem. When such a deviation exceeds a certain threshold, the corresponding clock or measurement can be classified as an anomaly by dedicated detection algorithms [41, 45].

Appropriately reducing the weights corresponding to anomalous clocks and corrupted measurement links allows the time scale to be robust to problems faced in the swarm. However, this method requires anomaly detection and quantification to adjust weights correctly. This work instead considers the case of missing measurements as a generalization for dealing with all types of anomalies. That is, contaminated measurements can be removed from the data set if they are detected, becoming missing measurements. As a result, the proposed methodology to deal with missing communication links can be extended to removing measurements classified as anomalies.

3.4.2 Noise and anomaly mitigation with Missing measurements

Assuming that neighboring satellites can communicate all unique clock comparisons between them, a total of $N(N-1)/2$ unique observations are available in a swarm of N satellites. With negligible noise, $N-1$ unique and independent measurements are required to generate a time scale using the BTSE. As shown in (3.21), noise in the measurements introduces uncertainty in the realization of a time scale. Since each time difference measurement can be written as a linear combination of the other measurements, the noise can be reduced with an LS estimator for the $N-1$ required measurements. That is, the set of all possible clock comparisons $\mathbf{z}(t)$ can be written as a set of linear equations of the required measurements $\mathbf{x}(t)$. An example is written for the case of $N=4$ satellites but can be easily expanded to greater numbers of satellites

$$\mathbf{z}(t) = \begin{bmatrix} z_{1,2}(t) \\ z_{1,3}(t) \\ z_{1,4}(t) \\ z_{2,3}(t) \\ z_{2,4}(t) \\ z_{3,4}(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_{1,2}(t) \\ x_{1,3}(t) \\ x_{1,4}(t) \end{bmatrix} + \begin{bmatrix} n_{1,2}(t) \\ n_{1,3}(t) \\ n_{1,4}(t) \\ n_{2,3}(t) \\ n_{2,4}(t) \\ n_{3,4}(t) \end{bmatrix}, \tag{3.27}$$

or otherwise,

$$\mathbf{z}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{n}(t). \tag{3.28}$$

The matrix $\mathbf{A}(t)$ depends on time because its number of rows depends on the number of available measurements. For example, the link between satellites 1 and 2 may not be physically possible to obtain or it could be neglected due to a detected anomaly. As a

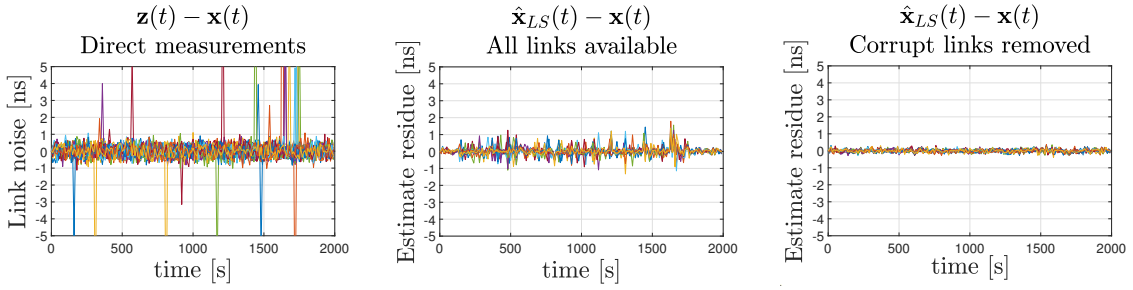
result, the first row would not appear in $\mathbf{A}(t)$. The LS estimator is defined as:

$$\hat{\mathbf{x}}_{LS}(t) = \min_{\mathbf{x}(t)} \left\{ \|\mathbf{z}(t) - \mathbf{A}(t)\mathbf{x}(t)\|^2 \right\}, \tag{3.29}$$

$$\hat{\mathbf{x}}_{LS}(t) = \left(\mathbf{A}(t)^T \mathbf{A}(t)\right)^{-1} \mathbf{A}(t)^T \mathbf{z}(t) \tag{3.30}$$

The number of measurements in $\mathbf{z}(t)$ will reduce for each unavailable or removed measurement, hence, the number of rows in matrix $\mathbf{A}(t)$ will also decrease. While the rank of this matrix remains equal to $N-1$, the LS estimator can still estimate the required measurements contained in $\mathbf{x}$ with residual error proportional to the number of available links.

To visualize the noise reduction achievable by the LS estimator, a situation with $N = 25$ satellites is simulated to represent a swarm of nanosatellites. All inter-satellite clock comparisons are simulated with Gaussian zero-mean additive measurement noise with a standard deviation of 0.3 ns. Anomalies of 10 ns are introduced at randomly chosen times on each unique link, some links may have anomalies at the same instant but each link only has one anomaly. Figure 3.14a indicates the intensity of the measurement noise on the $N-1$ required measurements made with respect to satellite 1. These measurements are sufficient to be substituted directly into (3.21) but will result in an unstable reference.



(a) Direct measurement noise with a single anomaly occurring on each link at a random point in time.

(b) Remaining error on the LS estimator when using all possible unique clock comparison measurements.

(c) Remaining error on the LS estimator when removing all measurements marked as anomalies to estimate.

Figure 3.14: Least squares estimator noise reduction for the 24 inter-satellite links measuring the time differences $x_{1,2}, \cdots, x_{1,25}$ by using redundant measurements in the satellite swarm.

The LS estimator provides estimates with a residue lower than the noise shown in Figure 3.14b. The anomalous measurements are not removed so that the LS estimator can be assessed in the case of non-detection. The anomalies are seen to be reduced because there is a sufficient number of additional measurements available simultaneously without anomalies.

With one out of 300 possible measurements affected by an anomaly at a single time instant, the spike due to that anomaly is noticeably filtered. However, when one of the required measurements, e.g., $x_{1,2}(t)$ has no anomaly, an anomaly occurring on one of the redundant measurements, e.g., $x_{3,4}(t)$ can cause the estimate to have a significant deviation compared to the direct measurement at that time instant. Depending on the magnitude of measurement noise and the magnitude of the anomaly, either the direct measurement or the LS estimator could be preferred.

The LS estimator still functions with certain measurements removed. Therefore, measurements classified as anomalies can be removed and the LS estimator can use the additional measurements to estimate those removed measurements. Figure 3.14c illustrates the benefit of removing the anomalous measurements (assuming they can be perfectly detected) by showing that the error is reduced in comparison to the direct measurements seen in Figure 3.14a. The proposed LS estimator is then useful in preprocessing of phase difference measurements that includes anomaly detection and sharing of redundant phase difference measurements.

If an anomaly affects the clock states, e.g., a phase or frequency jump, the number of measurements removed results in a reduction of the rank of $\mathbf{A}(t)$. In this case, the weights and the BTSE should be modified at the instant the measurements are removed. This is similar to the case of dealing with completely isolated clocks, which is explained in the next section.

3.4.3 Removal of Missing clocks

A swarm has access to N satellites in the best-case scenario. Considering N_m satellites can lose connection from the swarm at time t_m, how do we maintain continuity in the time scale when only using measurements from the remaining $N_a = N - N_m$ clocks? The change in the time scale due to having a reduced number of clocks is defined as:

$$C_{N_m}(t_m) = x_{i,E}(t_m)|_N - x_{i,E}(t_m)|_{N_a}, \tag{3.31}$$

where the time scale computed with the remaining N_a clocks is

$$x_{i,E}(t_m)|_{N_a} = \sum_{j=1}^{N_a} u_j(t_m - \tau) r_{j,i}(t_m), \tag{3.32}$$

and the time scale if all clocks were still available, is

$$x_{i,E}(t_m)|_N = \sum_{j=1}^{N_a} w_j(t_m - \tau) r_{j,i}(t_m) \tag{3.33}$$

$$+ \sum_{j=N_a+1}^{N} w_j(t_m - \tau) r_{j,i}(t_m), \tag{3.34}$$

which is split into two sums for the remaining and missing clocks. Neglecting the fact that $r_{j,i}(t_m)$ is not available for the missing clocks, we can compute the change in the time scale

$$C_{N_m}(t_m) = \sum_{j=1}^{N_a} \left(w_j(t_m - \tau) - u_j(t_m - \tau)\right) r_{j,i}(t_m) \tag{3.35}$$

$$+ \sum_{j=N_a+1}^{N} w_j(t_m - \tau) r_{j,i}(t_m). \tag{3.36}$$

The weights that were computed at the previous time instant are used to generate the AT1 time scale, with weights calculated for the full system denoted as $w_j(t-\tau)$, and weights computed for the reduced system denoted as $u_j(t-\tau)$. Since the weights of the missing clocks were computed using the information from the previous time epoch, they do not consider that those clocks are missing at $t = t_m$. Since it is simple to detect which clocks are missing at $t = t_m$, the weights of those clocks can be set to zero. If the weights of the missing clocks are set to zero then the remaining weights must be renormalized to maintain the uniqueness of the time scale for each clock. This means that the weights should be modified such that

$$w_j(t_m - \tau) = w_j(t_m - \tau) \Big/ \sum_{i=1}^{N_a} w_i(t_m - \tau)\ , \qquad j \leq N_a, \tag{3.37}$$

$$w_j(t_m - \tau) = 0, \qquad j > N_a. \tag{3.38}$$

If the above values are substituted into (3.36) then the expression for the change in the time scale due to losing N_m clocks becomes

$$C_{N_m}(t_m) = \sum_{j=1}^{N_a} \left(w_j(t_m - \tau) - u_j(t_m - \tau)\right) r_{j,i}(t_m). \tag{3.39}$$

This can be set to zero, ensuring phase continuity of the time scale by setting $u_j(t_m - \tau) = w_j(t_m - \tau)$ for the weights used to compute the reduced time scale at $t = t_m$. This proves why resetting the weights of missing clocks to zero allows continuity in the time scale when the number of available clocks changes. The above derivation can similarly justify continuity for an increasing number of clocks in the ensemble. However, if the weights always stay at zero for new (or returning) clocks then the time scale will never benefit from the increased number of clocks available.

3.4.4 Reintroduction of Missing clocks

When inter-satellite links are reintroduced into the swarm at $t = t_r$, the time scale must again ensure continuity to avoid instantaneous jumps by ensuring the returning clocks have zero weights when computing $x_{i,E}(t_r)$. Since the weights of the missing clocks were

already reset in the past, $w_j(t_r - \tau) = 0$ is guaranteed for returning clocks j and $x_{j,E}(t_r)$ is computed using the non-returning clocks. At the time instant after the clocks have returned ($t = t_r + \tau$), the new clock measurements were still effectively "missing" at the previous epoch but there is some information provided by $x_{j,E}(t_r)$.

To profit from the new measurements, the weights must eventually be non-zero, but they must not change rapidly because that will cause a jump in the time scale. Hence, a gradual reintroduction of the weights for the new clocks is used after the first reset at $t = t_r$. The AT1 time scale algorithm is compatible with this requirement thanks to an exponential filter that gradually increases weights over time according to an optimally chosen time constant. This time constant defines the amount of time before the reintroduced clocks are capable of contributing to the time scale again. The weights computed by the ATST algorithm do not explicitly control the rate of change of the weights so the algorithm may not function as well as AT1.

Figure 3.15 demonstrates how the weights are reset for the AT1 and ATST time scale. The weights after the moment of reintroduction at $t = 8000$ s are shown to be gradually increased for both algorithms. However, the exponential filter that controls the reintroduction for AT1 is not present in the ATST algorithm. It is possible that the exponential filter on the frequency estimates affects the temporal evolution of the weights for the ATST algorithm because the weights are determined with respect to prediction errors, which in turn depend on the frequency estimate. This is important for the reintroduction to be gradual and maintain continuity in the time scale. For a greater number of clocks being introduced at the same time, the period of reintroduction may not be sufficiently long for the ATST weights. That being said, a much higher number of clocks going missing and being reintroduced at the exact same time should not exceed the number tested in the below figures.

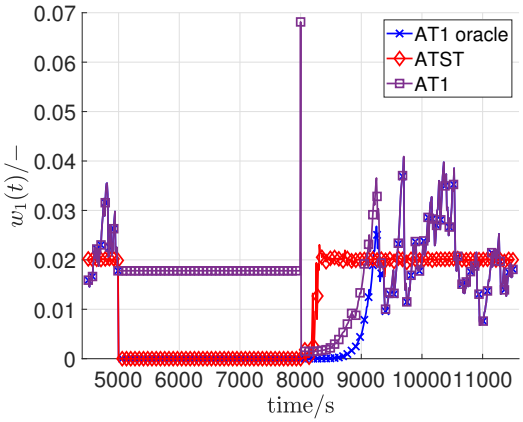
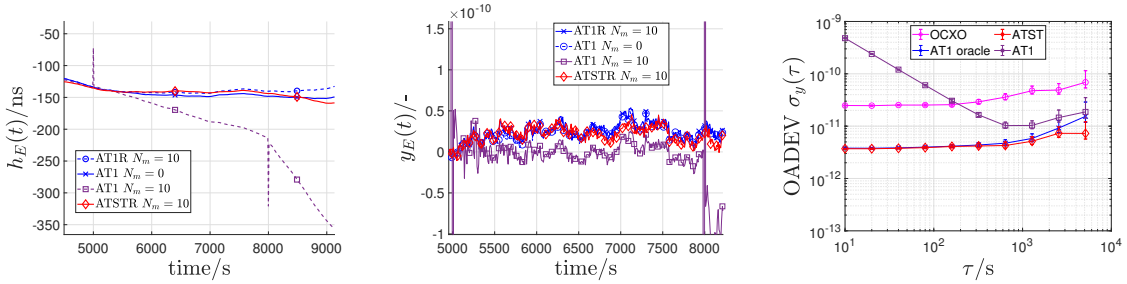


Figure 3.15: Weights for the different time scale algorithms with a reset occurring at the moments of clock removal and reintroduction.

Figure 3.16 illustrates the resulting time scale with and without the appropriate weights being reset at the required times. AT1R and ATSTR are used to refer to the AT1 and ATST time scales using the proposed weight reset methodology. The time scale is computed with 10 out of 50 total simulated clocks being removed at once (at $t = 5000$ s). All of the isolated clocks are reintroduced simultaneously after 50 minutes. This is a potential occurrence in the application of a swarm of satellites where the availability of inter-satellite links is intermittent. The results prove that continuity is still maintained with something as simple as forcing the weights to zero at the appropriate times. The ATST algorithm maintains continuity at the point of reintroduction because the weights do not change too rapidly although they still increase faster than the weights in the AT1 algorithm.



(a) Phase evolution, Missing Clocks.

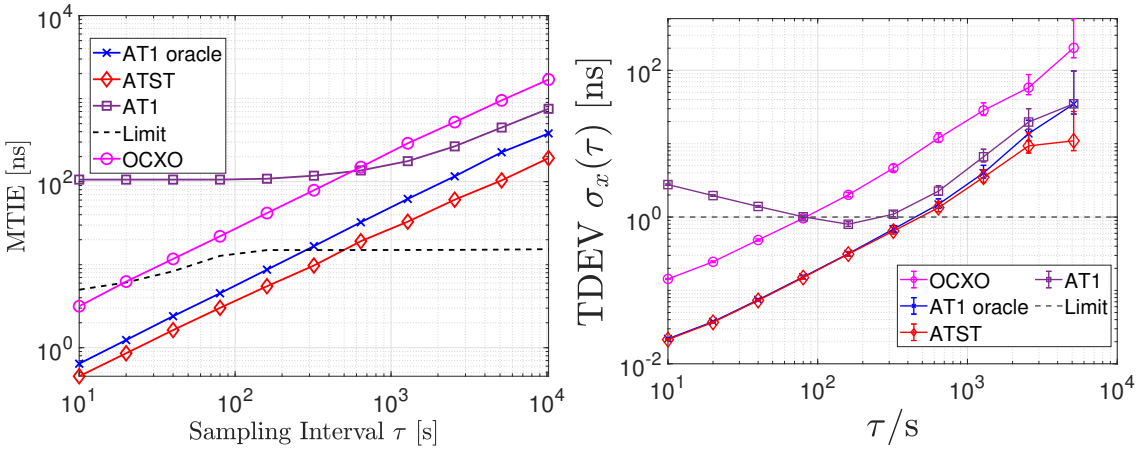
(b) Fractional frequency, Missing Clocks.

(c) Overlapping Allan Dev., Missing Clocks.

Figure 3.16: Performance of the proposed weight reset method for the AT1 and ATST time scale algorithms when 10 clocks become isolated simultaneously and are reintroduced simultaneously after 3000 seconds.

Although continuous, the overall stability of the time scale is still reduced compared to the case with no isolated clocks. This is unavoidable because the optimal reduction in ADEV for a time scale is proportional to the number of clocks. The ADEV is estimated using simulated samples over 20000 seconds with a data rate of 10 seconds. This means that the outage period of 3000 seconds for the missing clocks has a marginal effect on the overall ADEV. If the ADEV were instead estimated over only the duration of isolation, the time scale stability would be degraded to the achievable level for the reduced number of available clocks.

The continuity can be further confirmed for the tested time scales by observing the MTIE and TDEV in Figure 3.17. The AT1 time scale is presented as an example of a time scale that does not apply the required reweighting at the time a clock goes missing or is introduced. The loss in the time accuracy due to not treating the missing clocks is apparent in the short term MTIE, whereas the proposed method successfully mitigates this loss.



(a) Max time error.

(b) Time dev.

Figure 3.17: Timing accuracy assessed for the ATST and AT1 oracle time scales with $N_m = 10$ and the proposed reweighting methodology to ensure continuity.

3.5 Summary

The main contribution of this thesis is presented in this chapter together with two other minor contributions. The ATST time scale provides a new and robust method of adjusting the weights associated to clocks in the BTSE. The performance has been shown to be similar to an oracle version of the AT1 time scale algorithm that has perfect detection of phase jumps, frequency jumps and anomalies in inter-satellite measurement links. It has also been shown that the ATST algorithm is compatible with removing and reintroducing up to 10 clocks simultaneously. It is recommended to use the proposed ATST time scale for swarm applications with many satellites (at least 25) and where anomalies are frequent and the time scale must remain unaffected.

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 heavy-tailed statistical distribution. The resulting time scale is robust to each of the types of anomalies but cannot automatically identify which types of anomalies are present. The weights defined by the ATST algorithm are stable for nominal clocks, being fixed at approximately $1/N$ and only changing from that value when an anomaly occurs. The ATST is recommended for ensembles of homogeneous clocks to take advantage of the equal weights under nominal conditions, but a path forward for including diversity in clock types is possible and worth investigating. The first idea is to attempt a combination of the AT1 and ATST algorithms, aiming to match the detection performance of AT1 oracle by integrating the weights from the ATST algorithm. This idea is expanded upon in Chapter 5 with some preliminary results.

The proposed LS estimator provides estimates of time difference measurements with reduced error residuals thanks to the many redundant measurements that can be shared within the satellite swarm. Using an LS estimator to filter the measurement noise is the last contribution of this chapter that provides a simple method to reduce the problem of measurement noise in the swarm while also accounting for missing communication links and link anomalies. Indeed, the LS estimator is capable of mitigating jumps on measurement links given enough nominal redundant measurements. The main advantage of the LS estimator is the compatibility with a good detection algorithm to remove any inter-satellite measurement links that are corrupted by anomalies and obtain estimates of the removed measurements. Future work should be conducted to analyze the detection capabilities of the ATST algorithm or to find other novel methods that can reliably detect such link anomalies.

Chapter 4

Estimation Performance Limits

The proposed ATST time scale has introduced a new statistical model for the clock residuals allowing some robustness to the potential presence of outliers affecting the clock measurements. The Cramér-Rao Bound (CRB) is a useful metric to assess the estimation performance for parameters defining a precise statistical model. In the case of the clock residuals following a Student's t-distribution, the corresponding CRB can demonstrate the minimum possible Mean Square Error (MSE) attainable for an estimator of the location parameter. The CRB for certain statistical distributions presents a performance limit for time scales defined by assuming the clock residuals follow those distributions. In this chapter, the bounds are analyzed for the Gaussian distribution, the Student's t-distribution, and another typical model for anomalous data being a specific Gaussian mixture distribution.

Besides the CRB, the effect of incorrectly assuming the statistical model can be assessed with the Misspecified Cramér-Rao Bound (MCRB). For example, the assumption that the clock residuals follow a Student's t-distribution provides a model that encompasses the effects of outliers. Assuming that the data follows a Gaussian distribution instead is expected to deteriorate the estimation performance when outliers exist. The MCRB for the case of assuming a Gaussian distribution is derived when the true data follows either a Student's t-distribution or a two-variance Bimodal Gaussian Mixture (BGM). The results show just how much the MSE is restricted by neglecting the presence of anomalies in the assumed model. The benefit of using the ATST time scale over assigning equal weights to each clock is quantified and validated by evaluating the case where the data is actually distributed by a Student's t-distribution. The other statistical model that includes a probability of outliers then demonstrates if assuming that distribution could also be beneficial to generate a robust time scale. By having a quantifiable improvement, it is simply to make a choice about which assumption would be preferred for each possible true distribution of the clock residuals.

This chapter derives the MCRB in a generalized manner so it can be followed for other specific statistical distributions with known Probability Density Functions (PDFs), allowing

future comparison of different types of distributions that could potentially model the clock residuals. The results of the derivations should coincide with the recent generalized results from [70], i.e., the Gaussian assumption has an MCRB for the location parameter that is identical to the true variance of the data. The effects of misspecification are also established for certain parameters of a BGM or a general N-mode Gaussian Mixture [71]. Instead of misspecifying parameters, this work analyzes misspecification of the assumed type of distribution. Although this simplifies to a special case of parameter misspecification, the general form of the corresponding MCRB is not available in the literature.

4.1 Statistical models

The true models that are considered to represent contaminated data are types of heavy-tailed probability distributions. The Student's t-distribution has been defined in Chapter 3 with it's parameters estimated using the EM algorithm detailed in Appendix E.1. The BGM distribution consists of the sum of two Gaussian distributions with different variances but equal means. The parameters of the BGM can similarly be estimated using an EM algorithm so the assumption of this distribution for the clock residuals can generate a time scale in a similar manner to the ATST.

The BTSE residuals $r_{ji}(t)$ as introduced in Chapter 2 allow N time observations of clock i, which is common to each of the residual measurements. As was stated in the definition of the ATST, the residuals with fixed clock i are assumed to follow a Student's t-distribution with parameters $\boldsymbol{\eta}_t = [\mu_t, \sigma_t^2, \nu]^T$:

$$r_{ji}(t) \sim T(\mu_T, \sigma_T^2, \nu), \tag{4.1}$$

where μ_T is the location parameter, σ_T^2 is the scale parameter and ν is the number of degrees of freedom (assumed to be constant). These parameters depend on the fixed reference clock i and each sample j is considered some random deviation from the time of clock i with respect to the time scale. The parameters are written in a simplified form without the dependence on clock i to generalize the derivation of the bounds. The Student's t-distribution provides a good model of a random variable that has a non-zero probability of outliers occurring [72]. The joint PDF of a vector $\mathbf{r}(t) = (r_{1,i}(t), \cdots, r_{N,i}(t))^T$ of N i.i.d. random variables distributed according to a univariate Student's t-distribution is:

$$p_T(\mathbf{r}; \boldsymbol{\eta}_t) = \prod_{j=1}^{N} \frac{1}{\sigma_T \sqrt{\pi\nu}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{1}{\nu}\left(\frac{r_{ji} - \mu_T}{\sigma_T}\right)^2\right)^{-\left(\frac{\nu+1}{2}\right)}, \tag{4.2}$$

where the dependence on time for the residual samples is omitted for brevity. Another statistical model for the residuals $r_{ji}(t)$ is a BGM that mixes two Gaussian distributions with equal means but different variances, scaled by a factor of $k \geq 1$, with $k = 1$ simplifying to the classical Gaussian distribution:

$$r_{ji}(t) \sim GM(\mu_{\mathrm{GM}}, \sigma_{\mathrm{GM}}^2, \varepsilon, k). \tag{4.3}$$

If the true distribution is described by the BGM above, the true parameter vector is $\boldsymbol{\eta}_{\mathrm{GM}} = [\mu_{\mathrm{GM}}, \sigma^2_{\mathrm{GM}}, \varepsilon, k]^T$, where $0 \leq \varepsilon \leq 1$ refers to the probability of a sample being anomalous. The BGM consists of the nominal mode, which follows a Gaussian distribution denoted as $g_1(\mathbf{r}; \mu_{\mathrm{GM}}, \sigma^2_{\mathrm{GM}})$, and an anomalous mode that generates outliers with probability ε according to another Gaussian distribution denoted as $g_2(\mathbf{r}; \mu_{\mathrm{GM}}, k\sigma^2_{\mathrm{GM}})$. The PDF of $\mathbf{r}(t)$ assuming its components are independent is:

$$p_{\mathrm{GM}}(\mathbf{r}; \boldsymbol{\eta}_{\mathrm{GM}}) = \prod_{j=1}^{N} \left[\frac{(1-\varepsilon)}{\sqrt{2\pi\sigma^2_{\mathrm{GM}}}} \exp\left(-\frac{1}{2} \frac{(r_{ji} - \mu_{\mathrm{GM}})^2}{\sigma^2_{\mathrm{GM}}} \right) + \frac{\varepsilon}{\sqrt{2\pi k\sigma^2_{\mathrm{GM}}}} \exp\left(-\frac{1}{2} \frac{(r_{ji} - \mu_{\mathrm{GM}})^2}{k\sigma^2_{\mathrm{GM}}} \right) \right], \tag{4.4}$$

where the time dependency is again omitted for brevity. In both cases, the estimation of the location parameter μ_t or μ_{GM} acts as an estimator of the time difference of clock i from the generated time scale. Since each estimation procedure results from an EM algorithm, the other parameters must be jointly estimated. If the presence of the anomalies is neglected then the estimator of the location parameter is instead defined as the sample average. That is, the clock bias measurements and the internal clock noises are misspecified and assumed to be Gaussian with misspecified parameter vector $\boldsymbol{\theta} = [\mu_G, \sigma^2_G]^T$:

$$r_{ji}(t) \sim \mathcal{N}(\mu_G, \sigma^2_G). \tag{4.5}$$

The Gaussian assumption would then be an incorrect model for the random variable, providing the following misspecified PDF with different location and scale parameters

$$p_G(\mathbf{r}(t); \mu_G, \sigma^2_G) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2_G}} \exp\left[-\frac{1}{2} \left(\frac{r_{ji}(t) - \mu_G}{\sigma_G} \right)^2 \right]. \tag{4.6}$$

The reference time computed for the constellation is expected to lose stability in the presence of anomalies if the Gaussian assumption is made. By deriving the MCRB for this case, the benefits of correctly fitting a heavy-tailed distribution to the anomalous data are revealed. The assumptions of the Student's t-distribution or the BGM distribution presented earlier could also be considered misspecified if the true distribution is really modeled by the other distribution. This implies that other MCRBs could be relevant for the investigated heavy-tailed distributions.

4.2 Cramér Rao Bounds

The case where the assumed statistical model is correct implies that the minimum possible MSE is specified by the CRB. That is, for estimation of the parameter μ, the MSE is

bounded as follows:

$$E\left[(\hat{\mu} - E[z])^2\right] \geq \mathrm{CRB}_{\mu}, \tag{4.7}$$

where $E[z]$ refers to the mathematical expectation of the data with respect to the true distribution, i.e., the true mean of the data. Knowledge of the CRB is convenient for assessing the performance of an MLE because it is known that the MSE of the correctly specified MLE converges to the CRB in the asymptotic regime. The asymptotic regime refers to the case where the observed information has a sufficient number of samples or a reasonably high signal-noise ratio. The CRB for the estimation of the mean of a Gaussian distribution is well-defined [62]:

$$\mathrm{CRB}_{\mu_G} = \frac{\sigma_G^2}{N}. \tag{4.8}$$

In the application of time scale generation, the number of samples N is equivalent to the number of clocks. The CRB is then reminiscent of the expected gain in stability of a time scale being proportional to the number of clocks. Depending on the true model of the data, the form of the CRB can change depending on additional parameters. The CRB for the univariate Student's t-distribution is [73]:

$$\mathrm{CRB}_{\mu_T} = \left(\frac{\nu+3}{\nu+1}\right)\frac{\sigma_T^2}{N}, \tag{4.9}$$

and for the BGM with scaled variance, the CRB is not available in a simple closed-form expression, however, the CRB can be lower bounded by a value that can be approximated by the following expression [71]:

$$\mathrm{CRB}_{\mu_{\mathrm{GM}}} = \left(\frac{k}{k-\varepsilon(k-1)}\right)\frac{\sigma_{\mathrm{GM}}^2}{N}. \tag{4.10}$$

The CRBs for the two heavy-tailed distributions are compared impartially to the CRB of the Gaussian distribution by including additional scaling factors that normalize the variance of the true distribution. The normalized variances of each of the distributions are then considered the same and equal to $\sigma^2 = \sigma_G^2$, leading to:

$$\mathrm{CRB}_{\mu_T} = \left(\frac{\nu+3}{\nu+1}\right)\left(\frac{\nu-2}{\nu}\right)\frac{\sigma^2}{N}, \tag{4.11}$$

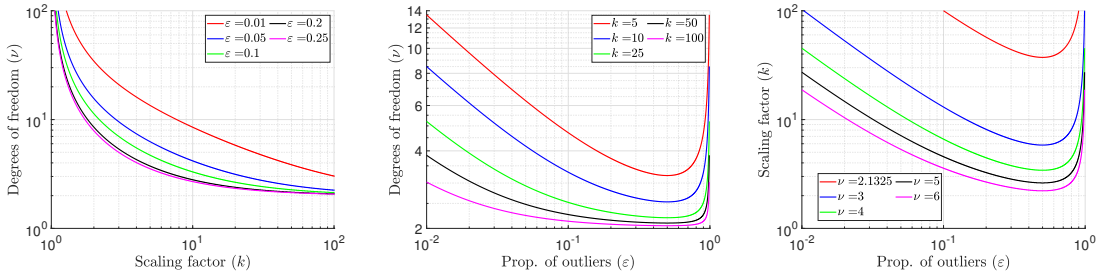
$$\mathrm{CRB}_{\mu_{\mathrm{GM}}} = \left(\frac{k}{k-\varepsilon(k-1)}\right)\left(\frac{1}{\varepsilon(k-1)+1}\right)\frac{\sigma^2}{N}. \tag{4.12}$$

The parameters that describe the outliers in the data for the Student's t-distribution (ν) and the BGM distribution (ε, k) provide some scaling factor with respect to the Gaussian CRB. These heavy-tailed distributions can obtain an equivalent lower bound on the estimation performance by equating the scaling factors of the normalized variance and finding

a relation between ν and ε, k. With the variances of the two distributions normalized, the link between the bounds for the t-distribution and the BGM is obtained by using the following relation:

$$\left(\frac{\nu+3}{\nu+1}\right)(\varepsilon(k-1)+1) = \left(\frac{\nu}{\nu-2}\right)\frac{k}{k-\varepsilon(k-1)}. \tag{4.13}$$

For any two fixed parameters out of ν, ε, and k, the complimentary third parameter that ensures an equal CRB can be found. Figure 4.1 provides curves that satisfy (4.13) for certain fixed values of each of the parameters. An example of a combination that will be investigated is $\varepsilon = 0.1$, $k = 100$ for the BGM, which coincides with a number of degrees of freedom of $\nu = 2.1325$ for the Student's t-distribution. The figure below provides a joint visualization of the dependencies between the parameters used to define outliers in the two heavy-tailed distributions.



(a) Fixed fraction of outliers ε. (b) Fixed scaling factor k. (c) Fixed degrees of freedom ν.

Figure 4.1: Useful curves for finding the parameters for the two heavy-tailed distributions that result in an equivalent CRB.

It makes sense that low numbers of degrees of freedom coincide with a large scaling factor in Figure 4.1a because they both relate to the intensity of outliers. Figure 4.1b shows that the proportion of outliers has a turning point at $\varepsilon = 0.5$ to represent the point that the contaminating Gaussian distribution becomes dominant in the BGM. As ε approaches one, the data becomes Gaussian again but with a different variance. This is observed in Figure 4.1b where the degrees of freedom diverge to infinity for $\varepsilon \approx 1$, corresponding to a Gaussian distribution. Similar behavior is shown for the scaling factor in Figure 4.1c but this is rather due to the constraint that the BGM and Student's t-distributions have the same CRB. The main outcome from these figures is the ability to infer the parameters of the heavy-tailed distributions that provide an equivalent CRB. Estimators that assume the incorrect heavy-tailed distribution for the data could then obtain estimates satisfy the above constraint and hence reach an MSE close to the correctly specified CRB.

4.3 Misspecified Cramér Rao Bounds

Table 4.1 indicates the possible combinations of true and assumed distributions between those presented above and shows the relevant bounds. A general result has already been obtained for the case of assuming the Gaussian distribution while the true distribution could be any elliptically symmetric distribution [70]. This result is verified and a closed-form expression for specific examples is provided by deriving the MCRB from first principles with the true distribution specified by either the Student's t-distribution or the BGM distribution. For precision in the notations, the MCRB is presented with subscripts to

Assumed \ True	Gaussian (p_G)	BGM (p_{GM})	Student's t (p_T)
Gaussian (p_G)	CRB_{μ} ✓	$\mathrm{MCRB}_{\mu}(p_{\mathrm{GM}}\|\|p_G)$ ✓*	$\mathrm{MCRB}_{\mu}(p_T\|\|p_G)$ ✓*
BGM (p_{GM})	$\mathrm{MCRB}_{\mu}(p_G\|\|p_{\mathrm{GM}})$?	$\mathrm{CRB}_{\mu_{\mathrm{GM}}}$ ✓	$\mathrm{MCRB}_{\mu}(p_T\|\|p_{\mathrm{GM}})$?
Student's t (p_T)	$\mathrm{MCRB}_{\mu}(p_G\|\|p_T)$?	$\mathrm{MCRB}_{\mu}(p_{\mathrm{GM}}\|\|p_T)$?	CRB_{μ_T} ✓

Table 4.1: Combinations of assumed and true distributions used to compute the MCRBs of interest. Check marks (✓) signify which combinations of assumed and true distributions have derived closed-form bounds. The question marks symbolize the combinations that have not yet been defined and potentially do not have a closed form, and * means "derived in this work".

refer to the parameters being estimated. For example $\mathbf{MCRB}_{\boldsymbol{\theta}}$ refers to the matrix that contains the bounds (independent and joint) for each parameter in $\boldsymbol{\theta}$. As an example, in the Gaussian case, one has:

$$\mathbf{MCRB}_{\boldsymbol{\theta}}(p||q) = \begin{bmatrix} \mathrm{MCRB}_{\mu}(p||q) & \mathrm{MCRB}_{\mu,\sigma^2}(p||q) \\ \mathrm{MCRB}_{\sigma^2,\mu}(p||q) & \mathrm{MCRB}_{\sigma^2}(p||q) \end{bmatrix}. \tag{4.14}$$

The MCRB will also specify the true (p) and assumed (q) distributions in parentheses because several combinations are considered in this work. Specifically, $\mathbf{MCRB}_{\boldsymbol{\theta}}(p_t||p_G)$ refers to the MCRB when assuming the Gaussian distribution with PDF p_G but with the true model following a Student's t-distribution with PDF p_t. Correspondingly, we can write $\mathbf{MCRB}_{\boldsymbol{\theta}}(p_{\mathrm{GM}}||p_G)$ for the other bound of interest.

The starting point to derive the MCRB is the definition of the Kullback Leibler Divergence (KLD), a statistical similarity measure between the true and assumed models defined as follows [74]:

$$D_{\mathrm{KL}}(p(\mathbf{z};\boldsymbol{\eta})||q(\mathbf{z};\boldsymbol{\theta})) = E_p\left[\log\left(\frac{p(\mathbf{z};\boldsymbol{\eta})}{q(\mathbf{z};\boldsymbol{\theta})}\right)\right], \tag{4.15}$$

where the subscript of E_p indicates the expectation with respect to the true PDF p, which remains the general notation for declaring any true distribution. To derive the MCRB, the pseudo-true parameters must be derived, which are defined as the parameters that minimize the KLD between the true and assumed models

$$\tilde{\boldsymbol{\theta}}_p = \arg\min_{\boldsymbol{\theta}} \{D_{\mathrm{KL}}\} = \arg\min_{\boldsymbol{\theta}} \left\{ E_p \left[\log \left(\frac{p(\mathbf{z};\boldsymbol{\eta})}{q(\mathbf{z};\boldsymbol{\theta})} \right) \right] \right\}. \tag{4.16}$$

The subscript p is included to indicate that the pseudo-true parameters depend on the true distribution. Conveniently, the expression for the KLD can be simplified to remove the components that do not depend on $\boldsymbol{\theta}$:

$$\tilde{\boldsymbol{\theta}}_p = \arg\min_{\boldsymbol{\theta}} \{-E_p \left[\log \left(q(\mathbf{z};\boldsymbol{\theta}) \right) \right]\}. \tag{4.17}$$

We obtain the following pseudo-true values for the two true distributions that model observations with outliers:

$$\tilde{\mu}_{p_T} = \mu_T, \qquad \tilde{\mu}_{p_{\mathrm{GM}}} = \mu, \tag{4.18}$$

$$\tilde{\sigma}^2_{p_T} = \sigma^2_T \frac{\nu}{\nu-2}, \qquad \tilde{\sigma}^2_{p_{\mathrm{GM}}} = \sigma^2 \left((k-1)\varepsilon + 1 \right), \tag{4.19}$$

where the full derivations are detailed in Appendix F.1. The pseudo-true parameters are denoted as $\tilde{\boldsymbol{\theta}}_p = [\tilde{\mu}_p, \tilde{\sigma}^2_p]^T$, the subscript p refers to the true distribution and can be replaced by p_t or p_{GM} for the Student's t-distribution or the BGM, respectively. It is noteworthy that the parameters that minimize the KLD between the misspecified Gaussian distribution and the distributions that model anomalous data are the mean and variance of the corresponding true distributions. The variance of the t-distribution, and as a result, the pseudo-true scale parameter is only defined for numbers of degrees of freedom $\nu > 2$. This restriction must also exist for the bounds, so we assume the number of degrees of freedom is always greater than two. The next steps to derive the MCRB involve the computation of two matrices [75]:

$$\mathbf{A}(\tilde{\boldsymbol{\theta}}_p) = E_p \left[\left(\frac{\partial^2 \log(q(z_j;\boldsymbol{\theta}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}, \tag{4.20}$$

$$\mathbf{B}(\tilde{\boldsymbol{\theta}}_p) = E_p \left[\left(\left(\frac{\partial \log(q(z_j;\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \log(q(z_j;\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^T} \right) \right] \right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \tag{4.21}$$

Note that the above matrices are only concerned with the marginal PDF of a single sample z_j. Under the assumption that each sample is i.i.d. the MCRB is defined as:

$$\mathbf{MCRB}_{\boldsymbol{\theta}}(p||p_G) = \frac{1}{N} \mathbf{A}(\tilde{\boldsymbol{\theta}}_p)^{-1} \mathbf{B}(\tilde{\boldsymbol{\theta}}_p) \mathbf{A}(\tilde{\boldsymbol{\theta}}_p)^{-1}. \tag{4.22}$$

The derivatives in (4.21) are provided in Appendix F.2. The results are then used to compute the MCRB for each of the two true distributions in Appendix F.3. For the Student's t-distribution, the following result is obtained:

$$\mathbf{MCRB}_{\boldsymbol{\theta}}(p_T||p_G) = \begin{bmatrix} \frac{\tilde{\sigma}^2_{p_T}}{N} & 0 \\ 0 & \left(\frac{\nu-1}{\nu-4}\right)\frac{2\tilde{\sigma}^4_{p_T}}{N} \end{bmatrix}, \tag{4.23}$$

where the bound associated with σ^2 is undefined for $\nu \leq 4$. The pseudo-true scale parameter is simply the sample variance of the true distribution. Therefore the bound for the estimation of the location parameter is equivalent to the CRB defined for the Gaussian distribution. As expected, the misspecified bound for σ^2 also simplifies to the Gaussian CRB for $\nu \to \infty$ because the true distribution then simplifies to a Gaussian. Similar results are visible when the true model is the BGM:

$$\mathbf{MCRB}_{\boldsymbol{\theta}}(p_{\mathrm{GM}}||p_G) = \begin{bmatrix} \frac{\tilde{\sigma}^2_{p_{\mathrm{GM}}}}{N} & 0 \\ 0 & \frac{Q(\phi)}{2(\phi+1)^2}\frac{2\tilde{\sigma}^4_{p_{\mathrm{GM}}}}{N} \end{bmatrix}, \tag{4.24}$$

where $\phi = \varepsilon(k-1)$ and $Q(\phi) = -\phi^2 + (3k+1)\phi + 2$. The predefined domains of ε and k suggest that $0 \leq \phi \leq k-1$, which means that there are no values of ε or k that result in an undefined bound. The value of $\frac{Q(\phi)}{2(\phi+1)^2}$ is always greater than one for the defined range of values for ϕ so the bound is never an undefined value and exceeds the equivalent Gaussian CRB for $0 < \phi < k-1$.

4.4 Analysis of Bounds

The MCRB provides the minimum MSE that can be attained in the asymptotic regime by the MLE that assumes a misspecified model for the observed data [76, 77, 78]. Similarly to the definition of the CRB, for an unbiased estimator:

$$E_p\left[(\hat{\mu}_q - E_p[z])^2\right] = \mathrm{MSE}_{\hat{\mu}_q}(p||q) \geq \mathrm{MCRB}_{\mu}(p||q), \tag{4.25}$$

where the misspecified estimator of the location parameter is denoted as $\hat{\mu}_q$ and is made assuming that the distribution of the data follows a statistical model with PDF q and the true distribution is with PDF p and location parameter μ. Despite the combinations in Table 4.1 marked with question marks having unknown bounds, it is still possible to assess the performance of assuming those distributions in a misspecified case with toy examples. That is, the estimators that assume either a Student's t distribution or the BGM distribution can be tested on data that is truly generated by specific statistical models. As

a result, the performance in presence of both nominal (Gaussian) and contaminated (heavy-tailed) data is tested and the compatibility between the two heavy-tailed distributions can be analyzed.

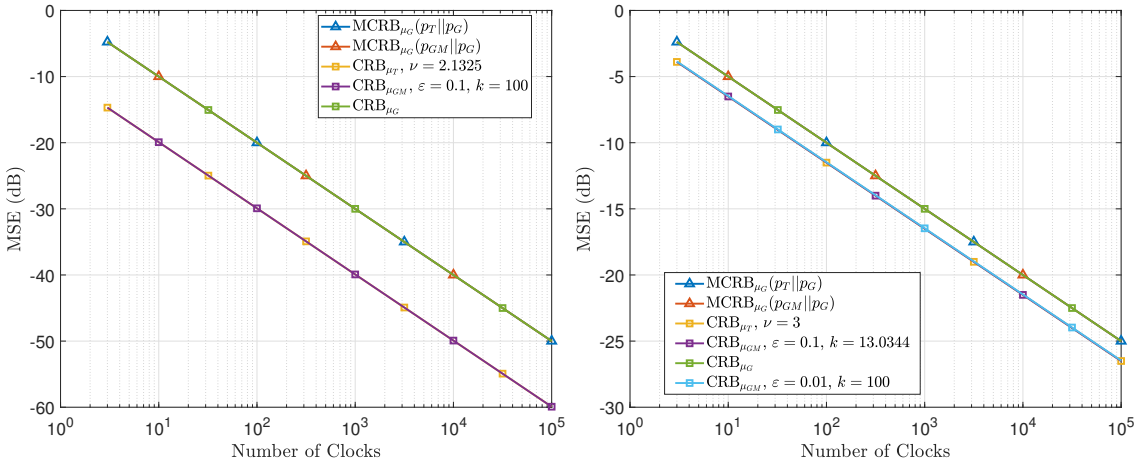
Without the need for any estimators to be implemented, the derived MCRBs and CRBs can be compared to show the gain in MSE that is achievable when the presence of outliers is correctly specified. Figure 4.2a demonstrates the equivalence of the MCRBs derived in this work (blue and red triangles) and the Gaussian CRB (green squares). The significance of this result is that assuming a Gaussian distribution without knowing the true distribution is actually heavy-tailed results in an estimator of the mean with the MSE bounded by the actual variance of the data. Since the true distribution is still symmetric and unbiased, the Gaussian MLE does not experience any increase in the bound due to the misspecification of the type of distribution.

Figure 4.2a also displays the effect of misspecification on the achievable MSE of a given estimator. The well-specified estimator is expected to reach the associated CRB (see CRB_{μ_t} and $\text{CRB}_{\mu_{GM}}$), resulting in a lower MSE compared to the MCRB that bounds the misspecified estimator. This confirms that misspecification on the type of distribution results in higher MSE, which is not desirable in any case. The bounds are displayed such that the variance of the data is normalized and the x-axis displays the total number of samples (clocks) available. This ensures that the resulting bounds are appropriately compared for a common dataset with an arbitrary statistical model but equal variance for any assumed distribution.

The improvement due to assuming the correct distribution depends on the values of the parameters that describe the intensity of the outliers. As demonstrated in Figure 4.2b, the CRB (yellow squares) approaches the MCRB (blue and red triangles) as the contamination becomes weaker, i.e., as the number of degrees of freedom increases for the Student's t-distribution, and if the proportion or magnitude of outliers decreases for the BGM. Figure 4.2b also shows several combinations of ε and k that result in a CRB equal to the CRB for the Student's t-distribution with $\nu = 3$ (yellow, green and purple squares).

Due to the lower bounds for the two heavy-tailed distributions, it is hypothesized that assuming the Student's t-distribution although the true model follows the BGM (and vice versa) does not necessarily cause the error of the misspecified estimator to deviate from the CRB. Therefore, the associated MCRBs are assumed to not deviate much from the CRB. This conclusion provides insight into $\text{MCRB}(p_T||p_{\text{GM}})$ and $\text{MCRB}(p_{\text{GM}}||p_T)$, which are not straightforward to derive.

Instead of using real clock data, the analysis is conducted with toy examples where the true model of the data is known exactly. The experiments have used random number generators such that N i.i.d. samples are generated according to a chosen statistical distribution. To validate the derived MCRBs, the performance of specific estimators is evaluated alongside the relevant bounds in the next sections. The Gaussian MLE is considered a misspecified MLE (MMLE) when the true distribution is not Gaussian. The MLEs for the Student's t-distribution and the BGM are not closed-form and require iterative Expecta-



(a) Heavy-tailed distributions parameterized for strong effects from anomalies.

(b) Heavy-tailed distributions parameterized for weaker effects from anomalies.

Figure 4.2: Both of the derived MCRBs are shown to be superimposed with the Gaussian CRB. The CRBs associated with the heavy-tailed distributions are also overlapping because they are parametrized so that their respective CRBs coincide. The CRBs being lower than the MCRBs confirms that it is preferred to assume the correct model of the data.

tion Maximization (EM) algorithms. As explained in Chapter 3, the derivation of the EM algorithm for the Student's t distribution is detailed in Appendix E.1, the EM algorithm for the BGM is derived in Appendix E.2.

4.4.1 True Distribution: Gaussian

To justify the assumptions of the heavy-tailed distributions, it is preferred that the MSEs of the associated estimators converge to the CRB for a Gaussian distribution

$$E_{p_G}\left[(\hat{\mu}_{t,\mathrm{GM}} - E_{p_G}[z])^2\right] \rightarrow \mathrm{CRB}_{\mu_G} = \frac{\sigma_G^2}{N}, \tag{4.26}$$

where the arrow signifies asymptotic convergence. Although not rigorous, verifying (4.26) in simulations can suggest that $\mathrm{MCRB}_{\mu}(p_G||p_T)$ and $\mathrm{MCRB}_{\mu}(p_G||p_{GM})$ are both equivalent to the Gaussian CRB. The convergence is illustrated in figure 4.3. This is a minor proof that assuming these specific heavy-tailed distributions does not introduce a restriction on the MSE of the estimator if the data is truly modeled by a Gaussian distribution.

This result is not surprising for the estimators assuming heavy-tailed distributions because the EM algorithms initialize their estimators with the Gaussian MLE. In the case the data truly has a Gaussian distribution, the convergence criteria for the EM algorithms is already satisfied with the initial estimates so the resulting performance is equivalent to

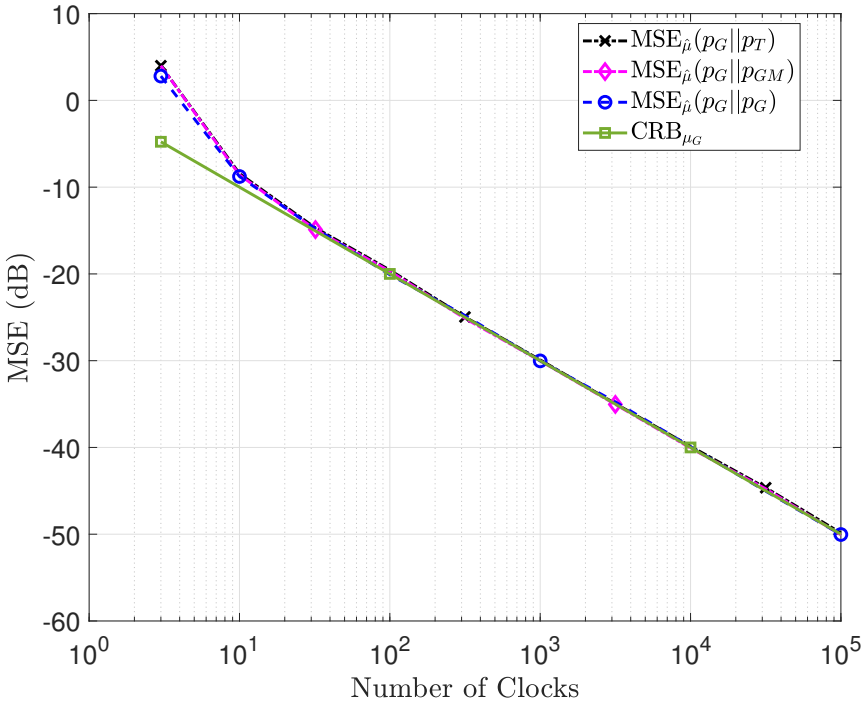


Figure 4.3: Convergence of investigated estimators when the true distribution is Gaussian.

the Gaussian MLE. The above figure also demonstrates that the number of clocks needed to converge to the optimum estimation error is around 15 clocks if the distribution is Gaussian. This corresponds to nominal clock performance and setting the weights of each clock equal to $1/N$.

4.4.2 True Distribution: Student's t

The correctly specified estimator is the one defined in the EM algorithm for the Student's t-distribution, this estimator is expected to converge to the associated CRB:

$$E_{p_t(\mu_T,\sigma_T^2,\nu)}\left[(\hat{\mu}_T - E_{p_T}[z])^2\right] \rightarrow \text{CRB}_{\mu_T}(\sigma_T^2,\nu), \tag{4.27}$$

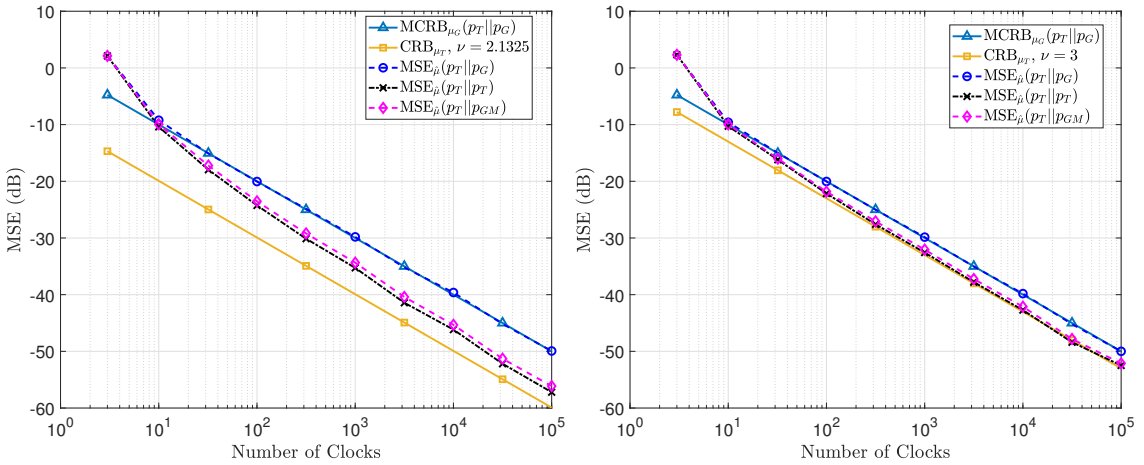
the MCRB corresponding to assuming a BGM can be analyzed by using simulations and assessing the convergence. Assuming a BGM results in joint estimation of the parameters μ_{GM}, σ_{GM}^2, ε, and k where the combination of the estimated values $\hat{\varepsilon}$ and $\hat{k}$ are assumed to provide an approximation of the true CRB, $\text{CRB}_{\mu_{GM}}(\hat{\sigma}_{GM}^2,\hat{\varepsilon},\hat{k}) \approx \text{CRB}_{\mu_t}(\sigma_T^2,\nu)$. That is, $\hat{\varepsilon}$ and $\hat{k}$ satisfy (4.13) within some margin of error. This approximation of the true bound can relate to an approximation of the true distribution so that the convergence of

the estimator assuming a BGM can be analyzed with respect to the true CRB:

$$E_{p_{GM}(\hat{\mu}_{GM},\hat{\sigma}^2_{GM},\hat{\varepsilon},\hat{k})}\left[(\hat{\mu}_{GM} - E_{p_{GM}}[z])^2\right] \rightarrow \text{CRB}_{\mu_{GM}}(\hat{\sigma}^2_{GM},\hat{\varepsilon},\hat{k}) \approx \text{CRB}_{\mu_T}(\sigma^2_T,\nu). \quad (4.28)$$

The condition of $\hat{\varepsilon}$ and $\hat{k}$ satisfying (4.13) is not guaranteed so any error in those estimates would contribute to some level of misspecification. Analyzing the convergence of the misspecified estimator allows a visualization of the significance of any errors in the estimates that cause losses with respect to the true CRB.

Figure 4.4a shows that the estimator assuming the Gaussian distribution (blue circles) converges to $\text{MCRB}_{\mu_G}(p_T||p_G)$ (blue triangles), verifying the derived MCRB and that assuming the Gaussian distribution is limited by that same bound. However, the MSE of the correctly specified MLE (black crosses) for data generated by a Student's t-distribution is lower than the MCRB so there is an improvement in estimation accuracy by correctly specifying the model. As predicted based on the equivalence between the CRB of the two heavy-tailed distributions, assuming the BGM (magenta diamonds) obtains a similar MSE to the correctly specified MLE when the data has a Student's t-distribution.



(a) Number of degrees of freedom $\nu = 2.1325$. (b) Number of degrees of freedom $\nu = 3$.

Figure 4.4: MSE performance of the correctly specified and misspecified estimators, when the data is generated by a Student's t-distribution. The errors for each of the estimators are presented as $\text{MSE}_{\hat{\mu}}(p_1||p_2)$, where p_1 is the true distribution and p_2 is the assumed distribution.

Figure 4.4b reveals that for $\nu = 3$, the MSE of the correctly specified estimator (black crosses) converges rapidly to the associated CRB (yellow squares). Indeed, the minimum number of clocks required to converge to the optimum estimation error is large but there is a noticeable improvement in MSE with respect to the misspecified estimator ofr as few clocks as 35. As demonstrated in Figure 4.4a, the MSE of the MLE (black crosses)

requires significantly more samples before converging to the CRB for the t-distribution (yellow squares) with a low number of degrees of freedom. A first assumption on the cause of this delayed convergence is an increased variance and hence uncertainty of the data due to more severe or frequent outliers. Additionally, the EM algorithm is potentially not sufficient to converge to the MLE for each of the estimated parameters as the number of degrees of freedom approaches the limit of $\nu = 2$. There are modified algorithms that improve upon the EM used to estimate te parameters of the Student's t-distribution, some especially designed to improve the estimation of the degrees of freedom [69]. Future work can be conducted to assess the impact of the different EM algorithms on the convergence of the MSE.

4.4.3 True Distribution: Bimodal Gaussian Mixture

The estimator that correctly specifies the true distribution is now denoted as $\hat{\mu}_{GM}$, which is expected to converge in MSE to the appropriate CRB.

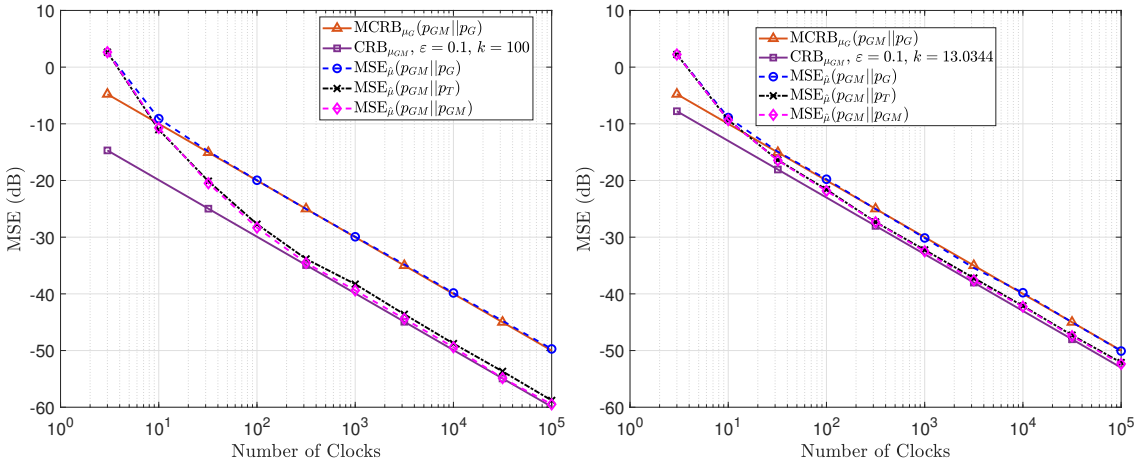
$$E_{p_{\mathrm{GM}}}\left[(\hat{\mu}_{\mathrm{GM}} - E_{p_{\mathrm{GM}}}[z])^2\right] \to \mathrm{CRB}_{\mu_{\mathrm{GM}}}. \tag{4.29}$$

The estimator defined by the Student's t-distribution jointly estimates μ_t, σ_t^2, and ν, where the estimator for the number of degrees of freedom can provide an approximation of the true CRB, $\mathrm{CRB}_{\mu_t}(\hat{\sigma}_t^2, \hat{\nu}) \approx \mathrm{CRB}_{\mu_{GM}}(\sigma_{GM}^2, \varepsilon, k)$. That is, $\hat{\nu}$ satisfies (4.13) for the true values of ε and k, within some margin of error. Let this approximation of the true bound relate to an approximation of the true distribution so that the convergence of the estimator assuming the Student's t-distribution can be analyzed:

$$E_{p_T(\hat{\mu}_T, \hat{\sigma}_T^2, \hat{\nu})}\left[(\hat{\mu}_T - E_{p_T}[z])^2\right] \to \mathrm{CRB}_{\mu_T}(\hat{\sigma}_T^2, \hat{\nu}) \approx \mathrm{CRB}_{\mu_{GM}}(\sigma_{GM}^2, \varepsilon, k). \tag{4.30}$$

Again, it is not guaranteed that $\hat{\nu}$ is perfect at approximating the true BGM distribution so any error in the corresponding estimator would cause misspecification. The importance of the misspecification can be properly assessed through simulation of the misspecified estimator.

Figure 4.5a demonstrates that the MMLE assuming the Gaussian distribution (blue circles) follows the other MCRB (red triangles), verifying the derivation. The correctly specified estimator (magenta diamonds) has an MSE reaching the CRB (purple squares) despite using the EM algorithm. This does not necessarily mean the CRB is always reached, different magnitudes for the parameters ε and k should be tested in future work. The same equivalence between the two heavy-tailed distributions is observed in Figure 4.5a, where the MMLE assuming the t-distribution (black crosses) obtains a very similar MSE to the correctly specified MLE. This combination of results implies that there is not necessarily a preferred assumption to make between the two investigated outlier-compatible distributions in terms of MSE. Nevertheless, a preferred assumption could be decided based on the



(a) BGM distribution with proportion of outliers $\varepsilon = 0.1$, and variance scaling factor $k = 100$.

(b) BGM distribution with proportion of outliers $\varepsilon = 0.1$, and variance scaling factor $k = 13$.

Figure 4.5: MSE performance of the correctly specified and misspecified estimators, when the data is generated by a BGM distribution. The errors for each of the estimators are presented as $\text{MSE}_{\hat{\mu}}(p_1||p_2)$, where p_1 is the true distribution and p_2 is the assumed distribution.

simplicity of estimation in either of the associated EM algorithms. This further analysis is saved for future work.

Both of the estimators defined by the MLE for a heavy-tailed distribution are capable of reaching the CRB with the true distribution being a BGM, although a slight offset is present in the misspecified case. This means that the delay in convergence noticed in Figure 4.4a is most probably linked to the properties of the true distribution of the data and not necessarily the EM algorithm. In fact, the parameters $\varepsilon = 0.1$ and $k = 100$ were shown to provide a CRB equivalent to the CRB of a Student's t-distribution with $\nu = 2.1325$. Hence, if the cause of the delay was solely due to the EM algorithm then the MSE of the estimators would be expected to converge just as slowly to the CRB in Figure 4.5a. These toy examples show that heavy-tailed distributions can make good assumptions for other heavy-tailed distributions but some have more restricting performance. Hence, the BTSE residuals in the swarm can be assumed to follow either distribution with a minimum of 35 clocks required in either case to obtain an improved and robust estimation performance in comparison to the misspecified Gaussian assumption.

4.4.4 Misspecified Estimation of Scale Parameter

The misspecified estimation performance for the scale parameter is basically provided by the MSE of the sample variance estimator that assumes a Gaussian distribution. The per-

formance of this estimator relative to the true scale factor of the heavy-tailed distributions is also analyzed for verification of the MCRB that was derived for the case of joint estimation of the location and scale parameters. The estimator is expected to converge in MSE to the MCRB plus the squared bias of the estimator due to misspecification. For the case of assuming a Gaussian distribution when the true distribution is some other heavy-tailed, real, elliptically symmetric distribution p:

$$E_p[(\hat{\sigma}_G - \sigma_p)^2] \to \text{MCRB}_{\sigma_G^2} + (\Delta\sigma^2)^2, \tag{4.31}$$

where $\Delta\sigma^2 = \tilde{\sigma}_p^2 - \sigma_p^2$ is the difference between the pseudo-true and true scale parameters, which is known in closed form using the results of the pseudo-true derivations. The same addition of the bias is necessary for the previously verified bound but it was shown in the derivation of the pseudo-true parameters that the misspecified estimator of the location parameter is not biased so $\Delta\mu = 0$.

Figure 4.6 shows the MSE of the misspecified estimator for each of the three true distributions analyzed above. This figure is intended to validate the derived MCRB and show that there is a bias in the misspecified estimator. The MSE of the biased estimator depends on the true distribution of the data according to the difference shown in Figure 4.6 between the Student's t-distribution (red crosses) and the BGM (black diamonds) with the same parametrization that results in equal CRBs for the location parameter.

The context for the misspecified estimator of the scale parameter is not strongly linked to the time scale application so the results will only be used as justification of the derived bound, which is a contribution that can be applied to more general scenarios. Further work should be conducted to appropriately compare the correctly specified estimators of the scale parameter to the misspecified bounds. That is, the EM algorithms' ability to improve the MSE for the scale estimator should be tested. Besides assessment of the EM algorithms estimation accuracy of the scale parameter, the importance of estimating the scale can be investigated for other applications. The NOIRE mission specifically, may find use of a robust estimator of the scale of the true distribution to better process measurements made for scientific observations, positioning solutions, studies of the relativistic effects of satellites in orbit, and even can be interesting in defining a Lunar Coordinated Time scale.

4.5 Bounds for Time Scale analysis

The use of CRB and MCRB are not typical analysis tools for time scale performance, but the link between the BTSE and robust estimation has made it clear that time scale algorithms can still benefit from classical statistical tools. The estimation of the location parameter is related to the realization of a time scale $x_{i,E}(t)$, as was defined for ATST. The realization of the time scale is then considered an estimate of the true location, which is assumed to be the offset of clock i from a perfectly stable time scale. Then the MSE can

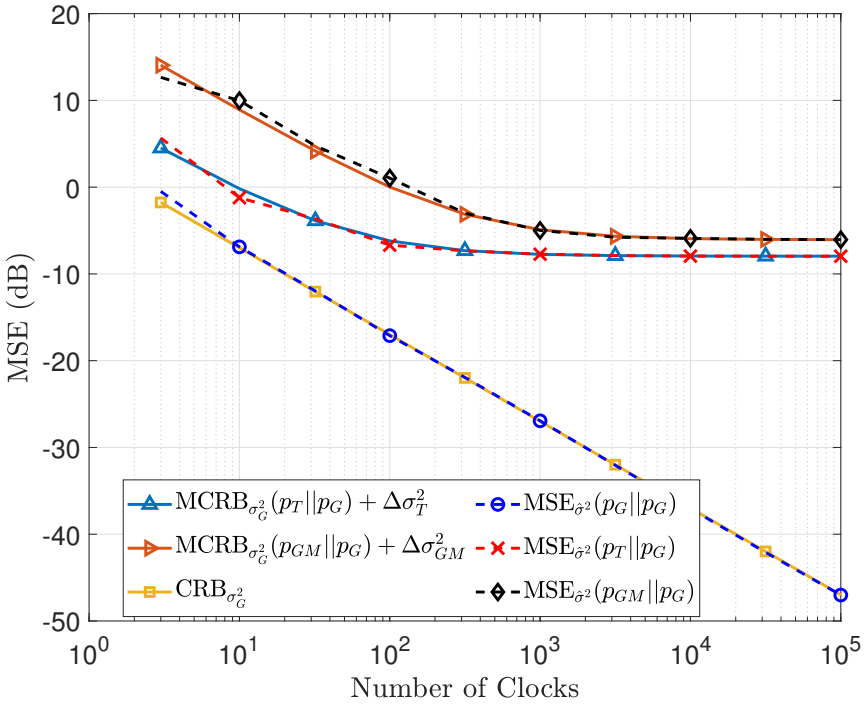


Figure 4.6: Convergence of the correctly specified and misspecified estimators of the scale parameter when the assumed distribution is Gaussian. The estimators have MSE plotted with the label explaining the true distribution: $\hat{\sigma}_G^2 - \hat{\sigma}_p^2$ is the error of the estimator that assumes a Gaussian distribution while the true distribution of the data is p.

be related to the phase of the time scale itself:

$$E_{p_i(t)}\left[(x_{i,E}(t) - x_{i,p}(t))^2\right] = E\left[x_{E,p}^2(t)\right] \geq \mathrm{CRB}_{\mu_i}, \tag{4.32}$$

where $p_i(t)$ refers to the specific probability distribution that models the BTSE residuals $r_{ji}(t)$. In the context of time scales, the convergence of the time scale MSE to the bound is only achieved for ensembles with a sufficient number of clocks. The bound is defined based on the true distribution of the data, which may vary according to the presence of anomalies, changes in measurement noise, or simply due to the different stochastic behaviors present in the clocks at different time intervals. Assuming the CRB remains constant at every time instant and the MSE of the designated estimator (or BTSE equivalently) converges to the bound at times $t-\tau$ and t then the resulting time scale could be considered perfectly stable over the time interval τ.

In reality, the distribution defining the BTSE residuals varies at each time epoch. The primary constraint is the natural evolution of the dispersion of the residuals due to internal clock noises, which will always result in some evolution of the CRB over time.

However, unnatural variations in the data such as jumps and outliers can be mitigated by appropriately choosing a heavy-tailed distribution. To observe the benefit of the heavy-tailed distributions in this context, it is better to observe the bounds without the normalized variances.

We assume that the only change in the dispersion of the data (and hence the bound) over some time interval is due to an anomaly. That is, the scale parameter σ^2 remains constant although the true variance of the data will increase due to an anomaly. The choice of the distribution should minimize the difference between the bounds for data with and without anomalies. The bounds before and after the occurrence of an anomaly can be described by the true distribution at the time epoch before an anomaly $t_a - \tau$ being Gaussian and the true distribution at t_a being a Student's t distribution where the specification of the true and assumed distributions for the MCRBs are now stated in a superscript to make room for the time dependencies. The approximation for the MCRB listed above is valid according to the convergence of the estimator performance shown in the previous section. The change in the CRB is smaller over the time interval τ when assuming the Student's t distribution compared to assuming the Gaussian distribution since $\frac{\nu}{\nu-2}\frac{\sigma^2}{N} > \frac{\nu+3}{\nu+1}\frac{\sigma^2}{N}$ for all $\nu > 2$.

According to the equivalence of the time scale phase to the bounds shown in (4.32), the change in continuity of the time scale can be related to the change in the bounds if the number of clocks is sufficient to obtain convergence. This means the MTIE at the measurement interval is given by the change in the bounds due to the anomaly changing the true distribution. Therefore, the loss in accuracy of the time scale due to assuming a Gaussian distribution when the data is truly given by a Student's t-distribution is always greater than the loss assuming the correct distribution. The jump that exists when assuming the true distribution can be considered small enough that it is negligible for a relatively high number of degrees of freedom, i.e., for a low number of anomalies occurring at the same time. Additionally, the time interval error when assuming the correct distribution could potentially remain negligible compared to the change in the time scale due to the natural evolution over that interval.

Assumed \ True	Before anomaly Gaussian	At time of anomaly Student's t
Gaussian	$\mathrm{CRB}_\mu(t_a - \tau) = \frac{\sigma^2}{N}$	$\mathrm{MCRB}_\mu^{(p_t \| p_G)}(t_a) = \left(\frac{\nu}{\nu-2}\right)\frac{\sigma^2}{N}$
Student's t	$\mathrm{MCRB}_\mu^{(p_G \| p_t)}(t_a - \tau) \approx \frac{\sigma^2}{N}$	$\mathrm{CRB}_\mu(t_a) = \left(\frac{\nu+3}{\nu+1}\right)\frac{\sigma^2}{N}$

Table 4.2: Bounds for a Gaussian distribution before the anomaly and a Student's t distribution at the time of anomaly.

A similar analysis can be conducted where the model of the clock data changes from a Gaussian distribution to a BGM distribution when the anomalies occur:

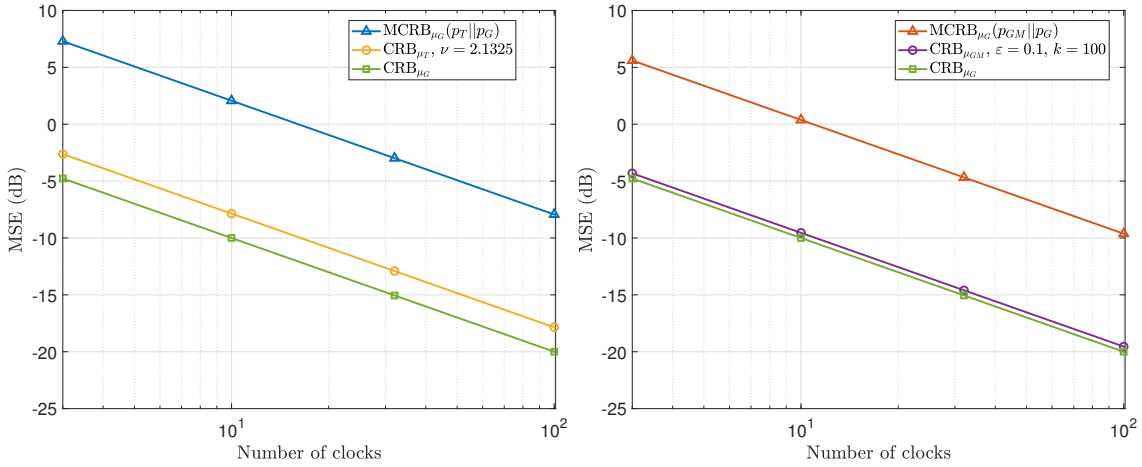
Assumed \ True	Before anomaly Gaussian	At time of anomaly BGM
Gaussian	$\mathrm{CRB}_{\mu}(t_a-\tau)=\frac{\sigma^2}{N}$	$\mathrm{MCRB}_{\mu}^{(p_{\mathrm{GM}}\|\|p_G)}(t_a)=(\varepsilon(k-1)+1)\,\frac{\sigma^2}{N}$
BGM	$\mathrm{MCRB}_{\mu}^{(p_G\|\|p_{\mathrm{GM}})}(t_a-\tau)\approx\frac{\sigma^2}{N}$	$\mathrm{CRB}_{\mu}(t_a)=\left(\frac{k}{k-\varepsilon(k-1)}\right)\frac{\sigma^2}{N}$

Table 4.3: Bounds for a Gaussian distribution before the anomaly and a BGM distribution at the time of anomaly.

It can be shown that the change in the bound is also smaller for the correctly specified estimator when the true distribution is a BGM. That is $\left(\frac{k}{k-\varepsilon(k-1)}\right)\frac{\sigma^2}{N}<(\varepsilon(k-1)+1)\,\frac{\sigma^2}{N}$ for the defined ranges of ε, and k. Depending on the values of the parameters ν, ε, and k that appropriately model the actual anomalies, the change in the bound from $t_a-\tau$ to t_a could be smaller for one of the two distributions. In the general case that the exact distribution is not known, the estimated parameters for ν, ε, and k would be the ones that define how much of a change in the time scale is experienced. Some examples of different parameterizations are shown in Figure 4.7 to indicate how the bounds can be roughly linked to the stability analysis of the time scale.

The two heavy-tailed distributions have been parametrized with the combination of ν, ε, and k shown above that results in equivalent CRBs for the two distributions with normalized variance. The variances have not been normalized in the figure above to appropriately show the effect of anomalies on the minimum attainable error, hence an increase in variance from time $t_a-\tau$ to t_a is expected. For the chosen parameters ν, ε, and k, the increase of the bounds will be greater for those associated with anomalies that result in the residuals following a Student's t-distribution because $\frac{\nu+3}{\nu+1}>\left(\frac{k}{k-\varepsilon(k-1)}\right)$ and $\left(\frac{\nu}{\nu-2}\right)>(\varepsilon(k-1)+1)$. This is illustrated in Figure 4.7 where the CRB for the Student's t-distribution (yellow circles) and $\mathrm{MCRB}_{\mu_G}^{(p_T\|\|p_G)}(t_a)$ (blue triangles) are higher than the respective bounds corresponding to the anomalies resulting in a BGM as the true distribution. This simply indicates that a Student's t-distribution results in a greater variance of the data for the parameters chosen to define the two heavy-tailed distributions.

The true magnitude of the increase of estimation error and hence, loss in stability will depend on how the anomalies contaminating the clock data actually modify the statistics of the clock residuals. In any case, by correctly assuming the distribution of the contaminated clock data, the losses are shown to be significantly reduced. An algorithm such as ATST is theoretically supported by the analysis of the change in the bounds as an improved



(a) Bounds when the true distribution at the time before anomaly is Gaussian and Student's t when the anomaly occurs at t_a.

(b) Bounds when the true distribution at the time before anomaly is Gaussian and BGM when the anomaly occurs at t_a.

Figure 4.7: Examples of the change in the estimation performance limits depending on the assumption made on the statistical model at time instances with and without a anomalies. Estimators applied to toy examples are included to verify the bounds at the different points in time and to represent the MCRBs that have not been explicitly derived.

method of estimating the clock phase compared to assuming clock data follows a Gaussian distribution. The assumption of the Gaussian distribution is equivalent to assuming the data is not contaminated by anomalies and the weights of each clock being assigned equally. While other existing time scale algorithms such as AT1 cannot be similarly compared by this analysis of bounds, any future algorithms that are designed with similar statistical modeling can be assessed quite easily. Therefore, the presented bound analysis is a useful contribution for future work on statistical models of clocks affected by different types of anomalies.

Chapter 5

Future Work

Although the new robust time scale algorithm was shown to be robust to phase jumps, frequency jumps, and anomalies in the simulated inter-satellite measurements, the performance with real data must still be verified and there are still limitations that should be overcome. The first issue to address is the functionality with a mixture of different types of clocks, where the stability over different time intervals is taken into account when computing the time scale. It is suggested to address this problem by introducing the exponential filter on the prediction errors from the AT1 algorithm into the ATST algorithm. This hopefully results in a robust time scale that can match the AT1 oracle even with a mixture of different clock types.

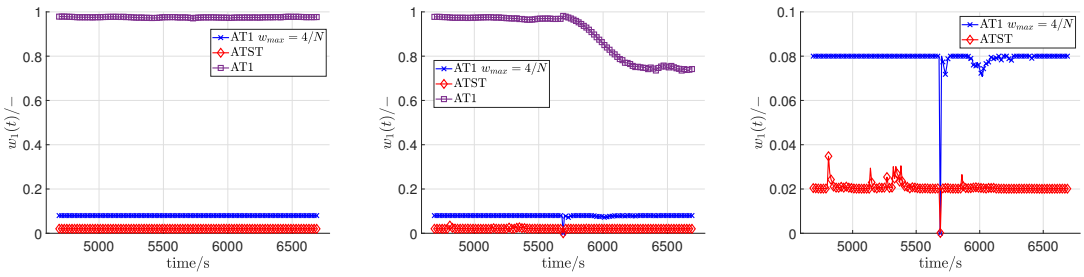
There is still room for improvement in generating a time scale that is robust to other types of anomalies such as those that occur over longer time intervals as opposed to the instantaneous jump type anomalies already addressed in this thesis. After properly understanding the value of the BTSE residuals in classifying anomalies, other methods of computing the clock weights could be proposed for an alternative algorithm. The advantages of machine learning in detection and classification of anomalies is at the forefront of research in many domains. The next proposal for future work explains how machine learning can assign weights to each clock based on information about the predictability of each clock over several time intervals. This is possibly useful in optimizing time scale stability over specific sampling intervals and addressing anomalies that occur over time by correctly choosing the time intervals.

Other than the new method of computing weights, transient anomalies could potentially be compensated by using robust estimation or anomaly detection based on historical data of clock frequencies. This is an extension to the work conducted on robust frequency estimation during the thesis and would require further analysis on the appropriate statistical distributions used to model the effects of such anomalies over the chosen window of frequency data.

5.1 Combining ATST with AT1

A simple test of the difference between the ATST weights and AT1 weights for a mixture of different clock types allows confirmation of the principle behind the weighting in the ATST algorithm. Since there is no filtering of the weights in the ATST algorithm, the stability of each of the clocks over extended sampling intervals is not taken into account. This is the advantage offered by the AT1 algorithm that filters the weights according to previous values with a time constant coinciding with the most stable sampling interval of the specific clock type. An advantage of instead following the ATST weight methodology is that any individual clock could potentially be affected by an anomaly at any instant in time. As the weights are assigned according to only the current measurements, an anomalous clock that showed great stability in the past will not necessarily have a dominant weight due to past behavior.

A property of the ATST weights is that clocks that function nominally receive a weight that is approximately equal to $1/N$, meaning that the case with no anomalies is equivalent to taking an average amongst all the clocks in the ensemble. Figure 5.1 confirms that this is the case even if a clock with orders of magnitude greater stability is included in the ensemble, e.g., an H maser clock. The weights computed using the AT1 algorithm will assign an appropriate weight to the H maser clock so that the time scale benefits from the improved stability of the H maser. However, it is important to put a constraint on the maximum weight to avoid the H maser becoming too dominant.



(a) Weights computed under nominal conditions.

(b) Weight response to phase jump.

(c) Phase jump, zoom on limited weights.

Figure 5.1: The weights computed by the AT1 and ATST algorithms for a single H maser clock that appears in an ensemble with $N = 50$ clocks. It is demonstrated that the ATST algorithm maintains a balanced weight assignment even though the H maser has significantly better stability. It is also shown that the ATST algorithm can still detect jumps in the H maser by reducing the weight at the appropriate time.

If the H maser experiences a phase jump, frequency jump or an anomaly in the respective measurements, the high weight from the previous time epoch will bias the newly

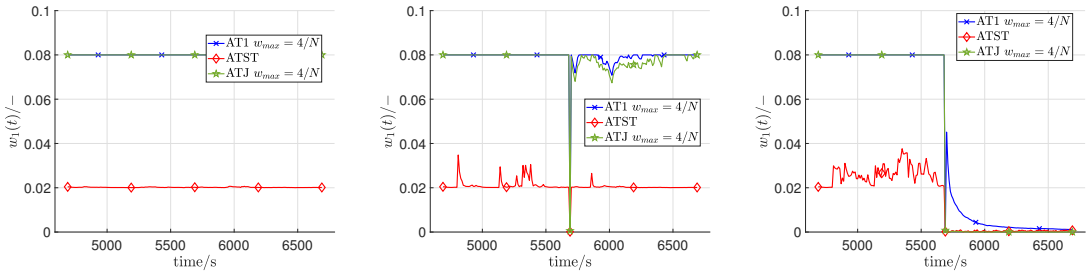
computed weight in the AT1 algorithm, meaning the anomaly may still impact the time scale. Another contribution of the ATST algorithm is proposed such that the weights computed using the Student's t-distribution are used to deweight the anomalous clocks in the AT1 algorithm. Since the ATST weight corresponding to the H maser is not biased by previous stability, and since the ATST algorithm has been shown to detect anomalies as good as an oracle detector (for certain anomalies), the advantages of both ATST and AT1 can be fused by appropriately combining the weighting procedures.

The weight control for the AT1 algorithm in presence of a phase jump adjusts the weights computed as the inverse of the estimated prediction error according to specific deweighting functions. Instead of using the weight control terms that are suggested for the AT1 algorithm, the weights that are computed in the ATST algorithm can be used to adjust the weights. That is, a joint algorithm, denoted as ATJ is proposed that still uses the exponentially filtered prediction error to compute the weights but scaled by a factor given by the weights computed by the robust estimator for the Student's t-distribution:

$$w_{i,ATJ}(t) = \frac{\frac{w_{i,\mathrm{ATST}}(t)}{\epsilon_i^2(t)}}{\sum_{j=1}^{N} \frac{w_{j,\mathrm{ATST}}(t)}{\epsilon_j^2(t)}}. \tag{5.1}$$

This is similar to the weight control term that was derived for phase jumps in AT1 but should also work for frequency jumps and link anomalies, as shown in Section 3.3.4. When the weights computed by the ATST algorithm are low due to phase jumps, frequency jumps or link anomalies, the weights of the joint algorithm will also be low and hence, the anomalies will not impact the resulting time scale. For nominal clocks, the weights computed by ATST are all approximately $1/N$, so the joint algorithm will function basically identically to the AT1 algorithm. This proposed joint algorithm is expected to take advantage of the different clock types like in the AT1 algorithm but can remain robust to the anomalies that the ATST algorithm can compensate. Since the ATST algorithm was shown to deal with both phase jumps and frequency jumps, as well as measurement anomalies this method is preferred for robustifying AT1 compared the separate and complicated methods using AT2.

The resulting weights for the joint algorithm are shown in Figure 5.2 with responses to both phase jumps and frequency jumps. The weights are equivalent to the AT1 oracle but without "cheating" with a priori knowledge of the anomalies occurring. All anomalies have been deweighted by the weights computed in the ATST algorithm while the AT1 algorithms ability to include the benefits of different clock types is not lost in the periods of nominal operations. The need for a maximum weight constraint is still required for the joint ATJ algorithm but this limitation is easy to implement in practice.



(a) No anomaly. (b) Phase jump. (c) Frequency jump.

Figure 5.2: The weights computed by the AT1 and ATST algorithms compared to the joint ATJ algorithm. The weights are appropriately reduced due to the reduction of the ATST weights to match the AT1 oracle anomaly detection. The AT1 weight increases rapidly for the H maser after the frequency jump but this is not present with the joint algorithm.

A result that requires further analysis is the jump in the weight that occurs for the AT1 algorithm immediately after the frequency jump. This could be because of the good performance of the H maser dominating the weight computations. In that case, it is advantageous to use the joint algorithm because it does not experience the same spike in the weights. Further study is required to determine the optimal weight limitation. The proposed joint algorithm is presented in the future work chapter because it requires further tests for realistic proportions of mixed clock types. This can be interesting in future studies that use real clock data when comparing the ATST and AT1 algorithms, the AT1ST time scale should achieve the best of both algorithms.

5.2 Machine Learning Weights

A novel detection method can be investigated that helps identify anomalous measurements using a machine learning methodology such as the Isolation Forest (IF) [79], allowing the removal of those measurements. An example of this detection method can generate an IF using the available measurements at the current time. The IF consists of a collection of isolation trees that divide the input data according to split branches until each observation is completely isolated from the others. In theory, an anomalous data point will become isolated with few separations. After construction of the IF, outlier scores are assigned to each measurement according to the following expression:

$$s(\mathbf{x}) = 2^{\frac{-E[h(\mathbf{x})]}{c(n)}}, \tag{5.2}$$

where the average path length to isolate an individual sample $\mathbf{x}$ is given by $E[h(\mathbf{x})]$ and $c(n)$ is the average path length for all n samples. The measurements $\mathbf{x}$ can be multivariate and contain a wider range of information about the errors in each of the clocks. As was

detailed in Chapter 2, the BTSE aims to minimize the weighted average of the prediction errors so the scores that are assigned to the prediction errors for each clock are a good candidate for weights for another new time scale algorithm. The IF score approaches a value of one for observations that are more likely to be an anomaly, whereas a score close to zero or less than 0.5 corresponds to a normal data point. For this reason, the scores can be used to assign weights to each clock by simply taking the normalized inverse scores as follows:

$$w_i(t) = \frac{\frac{1}{s(\mathbf{x}_i)}}{\sum_{j=1}^{N} \frac{1}{s(\mathbf{x}_j)}}. \tag{5.3}$$

If the scores are greater than a predetermined threshold, the associated measurements are considered outliers and should be discarded, i.e., by setting the weights to zero. The chosen threshold is subject to the anomaly detection trade-off on PND vs. PFA but is necessary if this method of using machine learning is to be used to determine the weights because there is no guarantee of having outlier scores that result in weights close to zero.

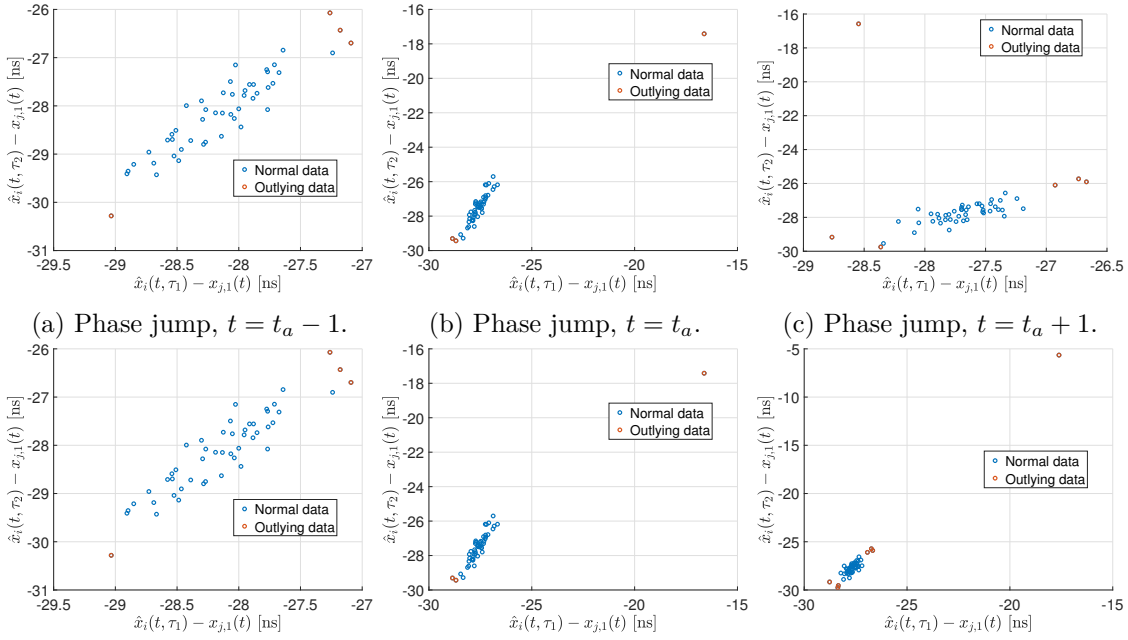
To base the scores on the performance of clocks over various time intervals, the BTSE residuals can be computed with predictions over several time intervals to get additional measurements of the predictability of each clock. The clock phase predictions can then be written as a function of the sampling interval τ as follows:

$$\hat{x}_{i,E}(t,\tau) = x_{i,E}(t-\tau) + \tau y_{i,E}(t-\tau). \tag{5.4}$$

For two different sampling intervals τ_1, and τ_2, a different scale of prediction error will be present in the corresponding BTSE residuals. This allows the resulting scores to be computed based on the stability over the different time intervals. Neglecting measurement noise means only one reference clock is required for the BTSE residuals. In this case, the data that will be used to generate the IF would then take the following form

$$\mathbf{Z} = \begin{bmatrix} \hat{x}_1(t,\tau_1) - x_{1,1}(t) & \hat{x}_1(t,\tau_2) - x_{1,1}(t) \\ \hat{x}_2(t,\tau_1) - x_{2,1}(t) & \hat{x}_2(t,\tau_2) - x_{2,1}(t) \\ \vdots & \vdots \\ \hat{x}_N(t,\tau_1) - x_{N,1}(t) & \hat{x}_N(t,\tau_2) - x_{N,1}(t) \end{bmatrix} = \begin{bmatrix} x_{1,p}(t) + e_1(t,\tau_1) & x_{1,p}(t) + e_1(t,\tau_2) \\ x_{1,p}(t) + e_2(t,\tau_1) & x_{1,p}(t) + e_2(t,\tau_2) \\ \vdots & \vdots \\ x_{1,p}(t) + e_N(t,\tau_1) & x_{1,p}(t) + e_N(t,\tau_2) \end{bmatrix} \tag{5.5}$$

where $\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1, & \cdots, & \mathbf{z}_N \end{bmatrix}^T$ is an $N \times 2$ matrix with each column corresponding to a different sampling interval and the prediction error over different intervals is denoted as $e_i(t,\tau)$. The above allows an illustration of the two dimensional problem in Figure 5.3 where simulated clock data with and without anomalies is sorted according to a generated IF that was automatically tuned using the default MATLAB function "iforest()".



(a) Phase jump, $t = t_a - 1$. (b) Phase jump, $t = t_a$. (c) Phase jump, $t = t_a + 1$.

(d) Frequency jump, $t = t_a - 1$. (e) Frequency jump, $t = t_a$. (f) Frequency jump, $t = t_a + 1$.

Figure 5.3: Scatter plots of the predictor variables that are obtained by making predictions over two different sampling intervals. The values with an outlier score greater than three times the median absolute deviation are highlighted to indicate the samples that should have reduced weights.

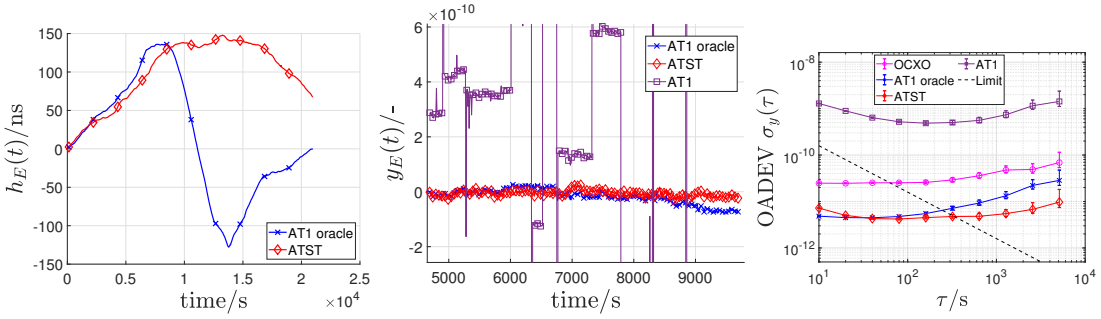
The above figures show that the scores can allow detection of outlying values in the presence of phase jumps and frequency jumps. By introducing more sampling intervals, the detection of more transient types of anomalies could be facilitated. Other machine learning methodologies such as Support Vector Machines (SVM) and Local Outlier Factor (LOF) similarly assign outlier scores to the data that could in turn produce more effective weights than the IF or ATST solutions. For this reason, it is interesting to investigate different methods of machine learning and find the best possible information to feed into the algorithms to obtain robustness to all kinds of anomalies. Another potential set of data can be residuals from the STSE for the clock frequency (similar to the residuals from the BTSE), taking advantage of both the phase and frequency differences between clocks to assign weights.

5.3 Transient Anomalies

The anomalies presented in Chapter 1 have not all been addressed in the core contributions of the thesis. Phase jumps, frequency jumps, and anomalies in the inter-satellite

measurements are important issues that have been appropriately addressed in the proposed solutions. For a more generalized robust time scale algorithm, the mitigation of the more transient clock anomalies should be further explored. For example, clock frequency drift, periodic oscillations, and temporary frequency jumps are all anomalies that have a strong impact on space applications but not necessarily mitigated by the robust estimation used to generate the ATST time scale.

Modeling the statistics for a window of past frequency data is one path to explore that may allow robust estimation or anomaly detection to be applied to these types of anomalies. The idea of applying machine learning to obtain weights that are based on different time intervals can also be compatible with the mitigation of anomalies that occur over specific time intervals. Without any special modifications to deal with transient anomalies, the ATST algorithm is compared to the AT1 oracle in Figure 5.4 in response to a temporary frequency jump of 10 ns/s for a period of 3000 s. The same collection of 50 simulated OCXO clocks is used with the anomalies added once on each clock with known times of occurrence and duration. Only a temporary frequency jump is tested because this type of anomaly is most similar to the jumps and missing data anomalies that were already investigated. The AT1 oracle sets the weights to zero at the beginning and end of the temporary jumps to simulate perfect detection.



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

Figure 5.4: Examples of the time scale performance in presence of temporary frequency jumps with a period of 3000 s and a magnitude of 10 ns/s.

The ATST time scale is shown to be better at mitigating effects on long term stability whereas the AT1 oracle time scale performs better at short term and indeed suffers losses in the long term. The losses for the AT1 oracle in long term may be linked to the number of frequency jumps that occur in the simulation period, i.e., 100 jumps in the ensemble representing the start and end of each temporary anomaly. As a result of the large number of times the clocks have their weights set to zero, the effective number of clocks in the ensemble may be reduced for extended periods of time and hence the stability cannot match an ensemble with all the clocks. A time scale algorithm using the machine learning based

weights or the joint AT1ST time scale could potentially show improvements to the ATST algorithm for this type of temporary frequency jump as well as other types of transient anomalies. Besides such promising methods, the continuation of some previous work on robust frequency estimation could help with the problem of transient anomalies.

5.4 Robust Frequency Estimation

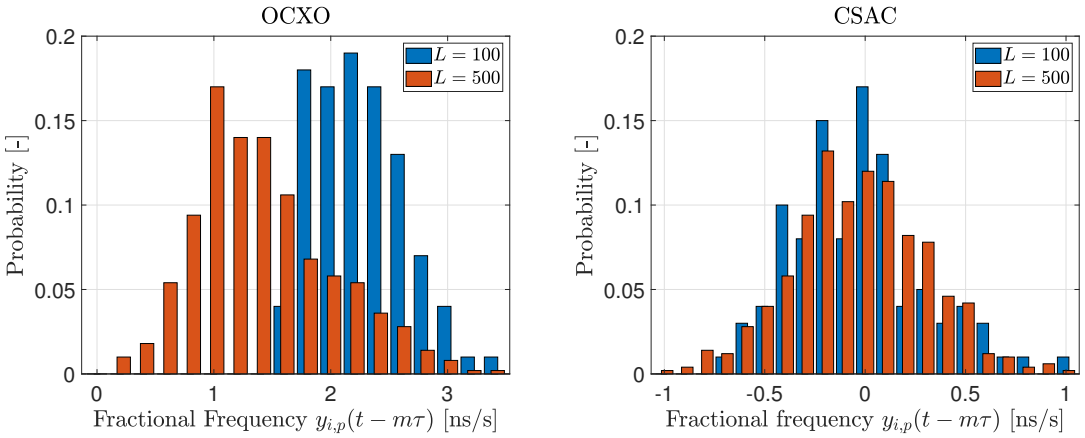
The results of using a time scale algorithm that uses robust estimation of the frequency based on the Student's t-distribution were published in [80]. That work illustrates that the robust frequency estimate is robust only to the effects of phase jumps and the weights computed with the robust estimate of frequency variance were not useful in mitigating the impact in the BTSE. This is not as good as the ATST solution that provides robust estimates by appropriately assigning weights for several types of anomalies.

The robust frequency estimator is instead suggested as an aid in detecting anomalies, because the robust estimate is a good baseline for determining thresholds. The ATST algorithm was designed for robustness against instantaneous anomalies but could potentially fail when faced with transient anomalies that occur over extended periods of time. For that reason, robust frequency estimation using a history of past frequency data might be explored to help deal with such anomalies.

An ideal clock frequency is constant over time, implying good stability and a predictable phase. As a result, the frequency estimate can be made using frequency data over a window of past time instances. This is similar to the existing solution of the ALGOS time scale algorithm [49]. Instead of taking an average frequency over the window, the robust frequency estimation would take a weighted average of the frequencies over the window by assuming that they follow an appropriate statistical distribution for the expected anomalies.

The end of this section explains how to build a robust time scale based on frequency estimation, as proposed in [80]. Each sample of frequency for clock i in the window of $L+1$ samples can be collected as $[y_{i,s}(t-L\tau), \cdots, y_{i,s}(t-\tau), y_{i,s}(t)]$. The choice of the window length varies depending on the distribution of the frequency samples. The length must be sufficient for the estimator to converge to its optimum performance but the distribution must also remain well modeled by the assumed distribution. As the types of noise vary over different sampling intervals, the resulting distribution from a large window of past samples may result in more uncertainty in the resulting estimate. Figure 5.5 displays the different distributions that can model the true frequency of a single simulated OCXO or CSAC. The actual frequency of the simulated clocks is used to demonstrate the expected distribution of frequencies without any dependence on the time scale algorithm or frequency estimation chosen.

The histograms displayed in Figure 5.5 are created using the frequency of the simulated clocks as described in Appendix C. The OCXO clock has a clear change in the distribution when increasing the window length. The distribution of the OCXO frequency is close to



(a) Histogram of frequency data over windows of different length for a simulated OCXO clock.

(b) Histogram of frequency data over windows of different length for a simulated CSAC clock

Figure 5.5: Frequency samples of a single simulated clock spread over a window of past values separated by sampling interval $\tau = 10$ s. Each sample corresponds to an integer multiple of τ where $m = 0, \cdots, L$. Two different numbers of samples L are used to show how the distribution may change due to different noise types occurring over different time intervals.

symmetric and approximately Gaussian, allowing an estimate of the mean frequency. The mean value of the OCXO clock distributions changes as a function of the window length because these types of clocks experience random walk frequency modulations and frequency drift over long time intervals, as observed in the Allan deviation. Conversely, the histogram for the CSAC clock shows two symmetric distributions with similar mean values for the two window lengths. This is explained by the absence of random walk frequency noise in these types of clocks.

Both types of clocks have a frequency distribution over the window that can be well modeled by a Gaussian distribution for the window lengths tested. Further work is required to identify the best type of distribution to assume for the transient anomalies. It would also be interesting to analyze the MCRB with respect to the actual distribution to appropriately demonstrate the gain in correctly modeling the frequency affected by transient anomalies.

Conclusion

A time scale algorithm robust to phase jumps, frequency jumps, anomalies in measurements, and missing clock data has been proposed and verified using simulated clock data. The new robust time scale is intended to be used in a swarm of nanosatellites operating in a remote region of space, so the onboard clocks are expected to suffer from these types of anomalies and more. The proposed time scale algorithm is the Autonomous Time scale using the Student's T-distribution (ATST), which generates a common reference time by taking a weighted average of the Basic Time Scale Equation (BTSE) residuals. These residuals refer to the terms that are weighted in the existing AT1 time scale and implicitly in the Kalman Filter based time scale. By assuming the BTSE residuals are modeled by a Student's t-distribution when anomalies occur, we are able to obtain weights that mitigate the contribution of any of the aforementioned anomalies in the generation of the time scale.

The performance of the ATST time scale is compared to a realization of the AT1 time scale with perfect detection of the investigated anomalies. Since the data is simulated, the algorithm is told exactly when an anomaly has occurred to automatically reduce the weights of the associated clock and obtain the AT1 oracle time scale. For the types of anomalies tested, the ATST algorithm was capable of matching the performance of AT1 oracle, which is promising for an algorithm that does not explicitly detect the anomalies. However, the limitations of the ATST are also presented: the algorithm is only suitable for swarms with homogeneous clock types, the total number of clocks must be sufficiently high to converge to an optimum estimator, the algorithm introduces additional computational complexity, and the lack of ability to differentiate between the different types of anomalies.

Another contribution that is implemented in the design of ATST, and is applicable to other time scales is the derivation of a procedure that deals with removing and reintroducing clocks to an ensemble. The simple proof has shown that continuity in the phase of the time scale can be maintained by setting the weights of removed and returned clocks to zero and renormalizing the weights of the other clocks so the sum is still equal to one. In addition to this method of maintaining continuity when the total number of available clocks changes, a Least Squares (LS) estimator that pre-processes the inter-satellite measurements is introduced to make use of redundant measurements in the swarm for reduction of measurement noise and anomalies, and estimation of the missing measurements when possible. In the presence of anomalies, the LS estimator is compatible with anomaly detection

methodologies because anomalous measurement links can be removed while the estimator provides a replacement using the redundant information shared within the swarm.

The final contribution of the thesis analyzes the optimal estimation performance of the Maximum Likelihood Estimator (MLE) for the parameters of a Student's t-distribution. More specifically, the loss in performance by instead assuming the BTSE residuals follow a Gaussian distribution is demonstrated by deriving the Misspecified Cramér Rao Bound (MCRB). Thanks to this derivation, it has been confirmed that the sample mean is limited in mean square error by the variance of the true distribution, which is inflated in the presence of outliers so the estimation performance is reduced. By assuming a heavy-tailed statistical distribution like the Student's t-distribution or a Bimodal Gaussian Mixture, the outliers are taken into account by the associated MLEs and it is shown that a lower bound is obtained. This analysis shows the benefits of using Student's t residuals for construcitng a robust time scale. It also allows a conclusion on the minimum number of clocks necessary for the ATST to be efficient, being 25 clocks but this number could increase for more extreme anomalies. Lastly, the bounds are linked to the concept of time scale continuity by showing the change in the bound due to the presence of anomalies changing the true distribution. This new methodology can be useful for analyzing other types of anomalies that fit other types of statistical distributions.

Besides the contributions directly presented in this thesis, discussions about potential future research have also been provided. Some of these future studies directly address the limitations identified for the ATST algorithm to allow introduction of different types of clocks and potentially mitigate the loss in performance for applications with fewer numbers of clocks in the ensemble. Other proposals explore a different path to generating a robust time scale but take advantage of the knowledge gained on how the BTSE residuals can be weighted according to any existing weighting algorithms. Specifically machine learning is interesting because of outlier scores that are assigned in unsupervised learning models, which can directly provide weights to mitigate anomalies that can be observed over many different time intervals. Another application of robust estimation to frequency estimation could similarly produce meaningful results for transient anomalies so several options are available for exploring a different robust time scale algorithm.

Glossary

ADEV Allan Deviation

AERO-VISTA Auroral Emissions Radio Observer - Vector Interferometry Space Technology using AERO

AIAA American Institute of Aeronautics and Astronautics

ALGOS Algorithm defined by the BIPM with a specialized method of using past frequency data but no specific acronym definition.

ASTRON Netherlands Institute for Radio Astronomy

AT1 Algorithm defined by the NIST with exponential filtering but no specific acronym definition.

AT2 Extension of AT1 algorithm defined by the NIST with a KF but no specific acronym definition.

ATJ The Joint time scale resulting from combining the proposed ATST algorithm and the existing AT1 algorithm.

ATST Autonomous Time scale using the Student's T-distribution

AVAR Allan Variance

BGM Bimodel Gaussian Mixture

BIPM Bureau International de Poids et Mesures

BTSE Basic Time Scale Equation

CGPM Conférence Générale des Poids et Mesures

CNES Centre National d'Études Spatiales

CRB Cramér Rao Bound

CSAC Chip Scale Atomic Clock

DARIS Distributed Aperture Array for Radio Astronomy In Space

DAVAR Dynamic Allan Variance

DTOA Differential Time Of Arrival

EAL Échelle Atomique Libre

EFTF European Frequency and Time Forum

EM Expectation Maximization

ESA European Space Agency

EUSIPCO European Signal Processing Conference

FFM Flicker Frequency Modulation

FM Frequency Modulation

FPM Flicker Phase Modulation

GLONASS Global Navigation Satellite System

GM Gaussian Mixture

GNSS Global Navigation Satellite System

GPS Global Positioning System

HVAR Hadamard Variance

ICASSP International Conference on Acoustics, Speech, and Signal Processing

IEEE Institute of Electrical and Electronics Engineers

IF Isolation Forest

INP Institut National Polytechnique

ION Institute of Navigation

IRIT Institut de Recherche en Informatique de Toulouse

ITU International Telecommunication Union

KF Kalman Filter

KLD Kullback Liebler Divergence

LOF Local Outlier Factor

LRT Likelihood Ratio Test

LS Least Squares

MCRB Misspecified Cramér Rao Bound

MDEV Modified Allan Deviation

MLE Maximum Likelihood Estimator

MMLE Misspecified Maximum Likelihood Estimator

MTIE Maximum Time Interval Error

MVAR Modified Allan Variance

NIST National Institute of Standards and Technology

NOIRE Nanosatellites pour un Observatoire Interférométrique Radio dans l'Espace

NPL National Physical Laboratory

OADEV Overlapping Allan Deviation

OAVAR Overlapping Allan Variance

OCXO Oven Controlled Crystal Oscillators

OLFAR Orbiting Low Frequency Array

PDF Probability Density Function

PFA Probability of False Alarm

PFS Primary Frequency Standard

PM Phase Modulation

PND Probability of Non-Detection

RW Random Walk

RWFM Random Walk Frequency Modulation

SIRA Solar Radio Imaging Array

SNR Signal Noise Ratio

STSE Supplementary Time Scale Equation

SULFRO Space Ulta-Low Frequency Radio Observatory

SVM Support Vector Machine

TéSA Télécommunications Spatiales et Aéronautique

TA2 Algorithm defined by NIST without specific acronym definition

TAI Temps Atomique International

TCXO Temperature Controlled Crystal Oscillator

TDEV Time Deviation

TOA Time Of Arrival

UTC Coordinated Universal Time

VLBA Very Long Baseline Array

VLBI Very Long Baseline Interferometry

WFM White Frequency Modulation

WPM White Phase Modulation

Appendix A

Two-sample Variances and Noises

This appendix provides an extract from the seminal Enrico's chart that provides a useful summary of oscillator noises and two-sample variances [11]. The variances that are displayed in this appendix are the Allan Variance (AVAR), Modified AVAR (MVAR), and the Hadamard Variance (HVAR). Although there are other types of two-sample variances that can be relevant for other applications, the AVAR, MVAR, and HVAR are presented to demonstrate how they can provide knowledge about different characteristic oscillator noises. Only the AVAR is used to assess the frequency stability in the context of this project. The Modified ADEV or MDEV is the square root of the MVAR and is used to compute the TDEV for assessment of the synchronization. The approximation of the MVAR is given by the following expression [12]:

$$ {}^{M}\sigma_y^2(\tau) = \frac{1}{2m^2\tau^2\,(M-3m+2)} \sum_{j=1}^{M-3m+2} \left(\sum_{i=j}^{j+m-1} [x_{i+2m} - 2x_{i+m} + x_i] \right)^2, \qquad \text{(A.1)} $$

where $\tau = m\tau_0$ for minimum sampling period (or measurement interval)τ_0 and M is the total number of frequency samples available.

The HVAR is a tool that could aid in future analysis of the effects of clock drift on the proposed time scale algorithm by comparing to the other variances.

The table below presents power spectral noise as a function of frequency f and two-sample variances as functions of the sampling interval τ. The cutoff frequency f_H is a characteristic of the physical oscillator. The total variance or spectral noise is obtained by taking the sum of the expressions for each noise type.

Table A.1: The expressions for different uncertainty measures of typical noise processes in oscillators.

Noise Type	$S_y(f)$	AVAR $\sigma_y^2(\tau)$	MVAR ${}^M\sigma_y^2(\tau)$	HVAR ${}^H\sigma_y^2(\tau)$
White PM	$h_2 f^2$	$\frac{3f_H}{4\pi^2}\frac{h_2}{\tau^2}$	$\frac{3}{8\pi^2}\frac{h_2}{\tau^3}$	$\frac{5f_H}{6\pi^2}\frac{h_2}{\tau^2}$
Flicker PM	$h_1 f^1$	$\frac{1.731-\log(2)+3\log(2\pi f_H\tau)}{4\pi^2}\frac{h_1}{\tau^2}$	$\frac{(24\log(2)-9\log(3))}{8\pi^2}\frac{h_1}{\tau^2}$	$\approx \frac{5\left[\gamma+\log\left(48^{0.1}\pi f_H\tau\right)\right]}{6\pi^2}\frac{h_1}{\tau^2}$
White FM	$h_0 f^0$	$\frac{1}{2}\frac{h_0}{\tau}$	$\frac{1}{4}\frac{h_0}{\tau}$	$\frac{1}{2}\frac{h_0}{\tau}$
Flicker FM	$h_{-1} f^{-1}$	$2\log(2)h_{-1}$	$\frac{(27\log(3)-32\log(2)}{8}h_{-1}$	$\frac{8\log(2)-3\log(3)}{2}h_{-1}$
Random Walk FM	$h_{-2} f^{-2}$	$\frac{2\pi^2}{3}h_{-2}\tau$	$\frac{11\pi^2}{20}h_{-2}\tau$	$\frac{2\pi^2}{6}h_{-2}\tau$
Integrated Flicker FM	$h_{-3} f^{-3}$	not converging	not converging	$\frac{\pi^2[27\log(3)-32\log(2)]}{6}h_{-3}\tau^2$
Integrated RW FM	$h_{-4} f^{-4}$	not converging	not converging	$\frac{44\pi^2}{60}h_{-4}\tau^3$
Linear drift D_y	N/A	$\frac{1}{2}D_y^2\tau^2$	$\frac{1}{2}D_y^2\tau^2$	0

Appendix B

Confidence Limits

The method of computing the confidence intervals for the estimation of the ADEV is different depending if the overlapping or non-overlapping estimation method is used. Since this thesis always presents an OADEV, this is the only confidence interval presented in this appendix. Although the other forms are available in the same references used to obtain the following equations [12, 17]. Since the estimate of the Allan variance is a sum of the squared two-sample differences, it is assumed to follow a χ^2_k distribution with k degrees of freedom. A test statistic is defined for the chi-squared distribution as

$$\chi^2_{k_e} = \frac{k_e(\tau)\hat{\sigma}^2_y(\tau)}{\sigma^2_y(\tau)}, \tag{B.1}$$

where $k_e(\tau)$ is an equivalent number of degrees of freedom that is defined according to the type of noise experienced by the oscillator at the sampling interval over which the ADEV is estimated. The estimated variance $\hat{\sigma^2_y}(\tau)$ is differs from the true variance $\sigma^2_y(\tau)$. The typical estimation uncertainty presented in published figures of ADEV is the 1σ or 68% confidence interval. That is, the test statistic and equivalent number of degrees of freedom are related to the probability that a random variable x lies within a certain region of values:

$$p_{max} = Pr\left(x \leq \chi^2_{k_e}\right) = 0.84, \tag{B.2}$$

and

$$p_{min} = Pr\left(x \geq \chi^2_{k_e}\right) = 0.16. \tag{B.3}$$

Given the equivalent number of degrees of freedom, the above expressions can provide an upper and lower bound on the test statistic such that there is 68% certainty that the true variance lies within the corresponding interval. The resulting constraint on the true variance is then presented below:

$$\frac{k_e(\tau)}{\chi^2_{k_e}(p_{max})}\hat{\sigma}^2_y(\tau) \leq \sigma^2_y(\tau) \leq \frac{k_e(\tau)}{\chi^2_{k_e}(p_{min})}\hat{\sigma}^2_y(\tau). \tag{B.4}$$

Table B.1 lists empirical formulas for k_e depending on the type of noise that is dominant over the sampling interval τ. In the below table, the number of samples is denoted as N, and the sampling interval τ is represented by a ratio with the minimum sampling interval τ_0, $m = \frac{\tau}{\tau_0}$.

Table B.1: Empirical equations for equivalent number of degrees of freedom used to determine confidence interval of ADEV estimate [17].

Noise Process	**Degrees of freedom** $k_e(\tau)$
White PM	$\frac{(N+1)(N-2m)}{2(N-m)}$
Flicker PM	$\exp\left(\log\left(\frac{N-1}{2m}\right)\log\left(\frac{(2m+1)(N-1)}{4}\right)\right)^{\frac{1}{2}}$
White FM	$\left(\frac{3(N-1)}{2m} - \frac{2(N-2)}{N}\right)\frac{4m^2}{4m^2+5}$
Flicker FM	$\frac{2(N-2)^2}{2.3N-4.9}$ for $m = 1$ $\frac{5N^2}{4m(N+3m)}$ for $m \geq 2$
Random Walk FM	$\left(\frac{(N-1)^2-3m(N-1)+4m^2}{(N-3)^2}\right)\frac{N-2}{m}$

For each sampling interval used to estimate the ADEV, the equivalent number of degrees of freedom is found using the expressions listed above. The known inverse cumulative distribution function for the chi-squared distribution permits us to obtain the test statistic for a given combination of k_e and p, which can then be substituted into (B.4) to obtain the upper and lower limits of the true variance. These limits are displayed in all of the figures of ADEV in this paper in the form of error bars. It is known that the error bars should increase as the sampling interval increases due to having less samples and hence more uncertainty.

Appendix C

Simulation of Satellite Clocks

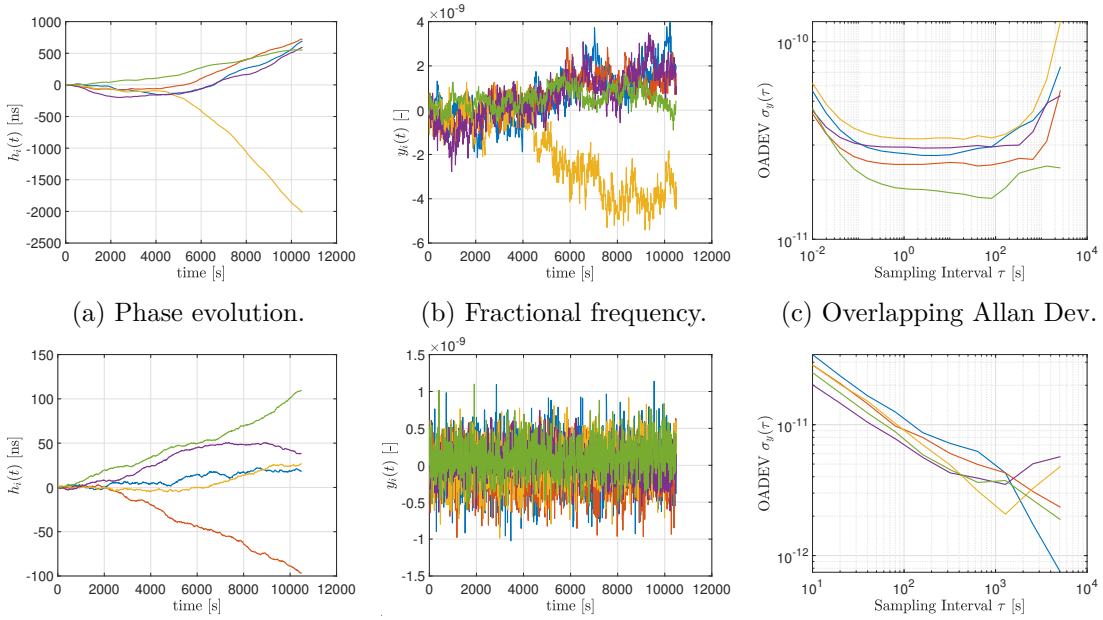
To simulate clock data, we use a noise generator that follows the expected stochastic behavior of the type of clock that we are interested in simulating. The power spectral density for clock phase can be written as a sum of each of the different noise types

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

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

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

The clock noises are then generated independently according to the discrete generation method highlighted in [81]. 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 stabilities. The simulator can similarly use the typical OADEV of any other type of clock to replicate several simulated clocks with similar performance. Figure C.1 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.

(d) Phase evolution. (e) Fractional frequency. (f) Overlapping Allan Dev.

Figure C.1: Examples of 5 simulated OCXO clocks (top) and 5 simulated CSAC clocks (bottom) with diversity in the clock behaviours.

Appendix D

Clock Anomaly Magnitudes

This section addresses alternative magnitudes of anomalies to elaborate on the resolution of the robust time scale proposed in Chapter 3. The proposed ATST time scale is already demonstrated to be as robust as the AT1 oracle time scale (i.e., AT1 with perfect detection of anomalies) in Chapter 3 for the following anomalies:

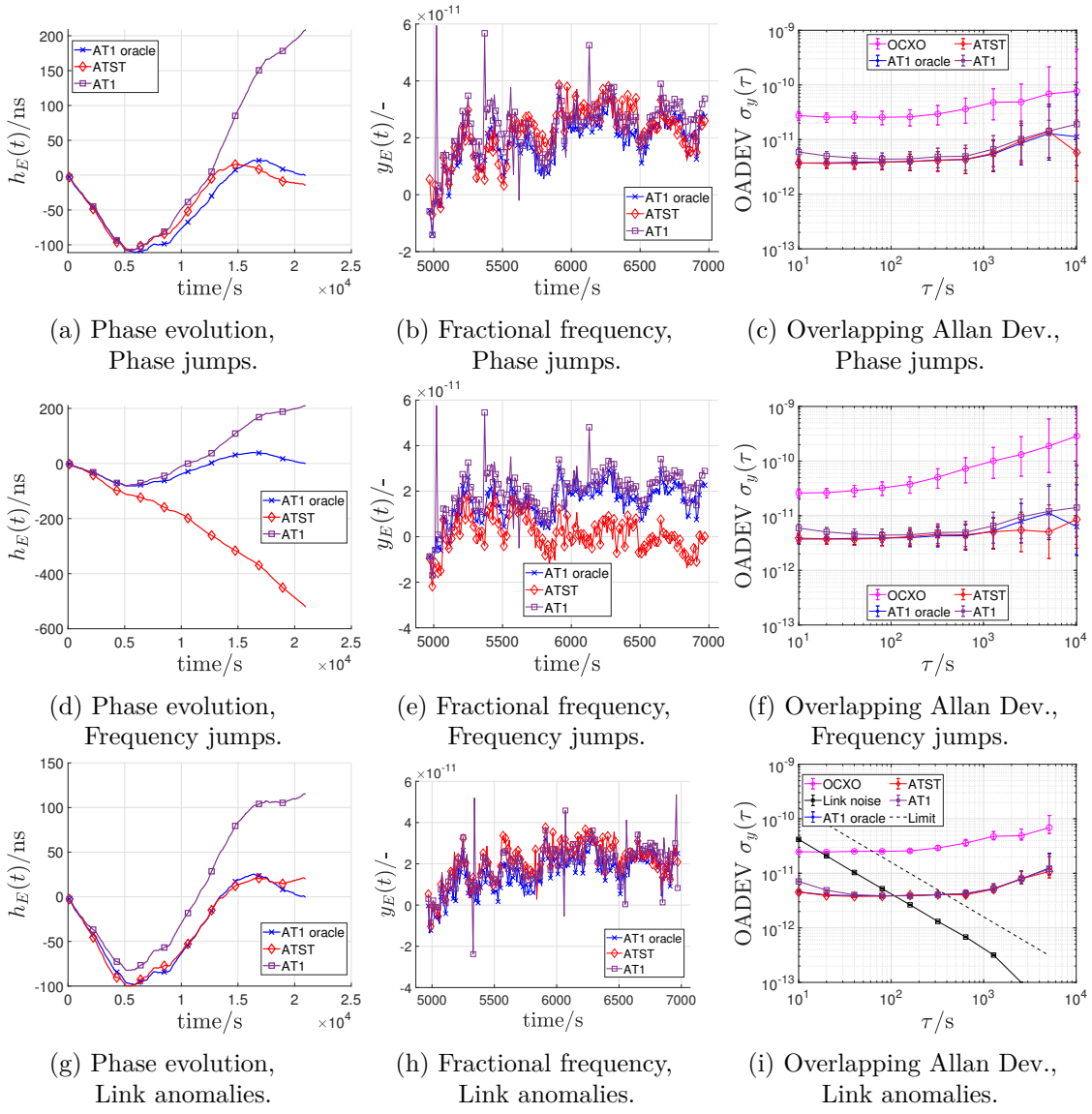
- Phase jumps with magnitude randomly generated by a Gaussian distribution with a standard deviation of $\sigma_{\Delta x} = 100$ ns.

- Frequency jumps with magnitude randomly generated by a Gaussian distribution with a standard deviation of $\sigma_{\Delta y} = 100$ ns/s.

- Anomalies on measurement links j, i between clock j and clock i with magnitude randomly generated by a Gaussian distribution with a standard deviation of $\sigma_{\Delta n_{j,i}} = 100$ ns.

The phase jumps and frequency jumps refer to anomalies in the states of the clocks. The link anomalies are instantaneous changes in measurement noise for certain inter-satellite links in a swarm of satellites. The precise amplitude of the anomalies is randomly assigned according to a standard Normal distribution but the order of magnitude is maintained to describe the anomalies on all clocks.

Figures D.1 and D.2 display the time scale performance with smaller magnitudes of anomalies: $\sigma_{\Delta x} = 10$ ns, $\sigma_{\Delta y} = 10$ ns/s, and $\sigma_{\Delta n_{j,i}} = 10$ ns. The proposed ATST algorithm remains equivalent to the AT1 oracle for the small magnitudes, which means it is not necessarily blind to weaker effects in the assignment of clock weights. This can potentially be a problem for algorithms that rely on a detection threshold that is too large to detect weaker anomalies.



(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.

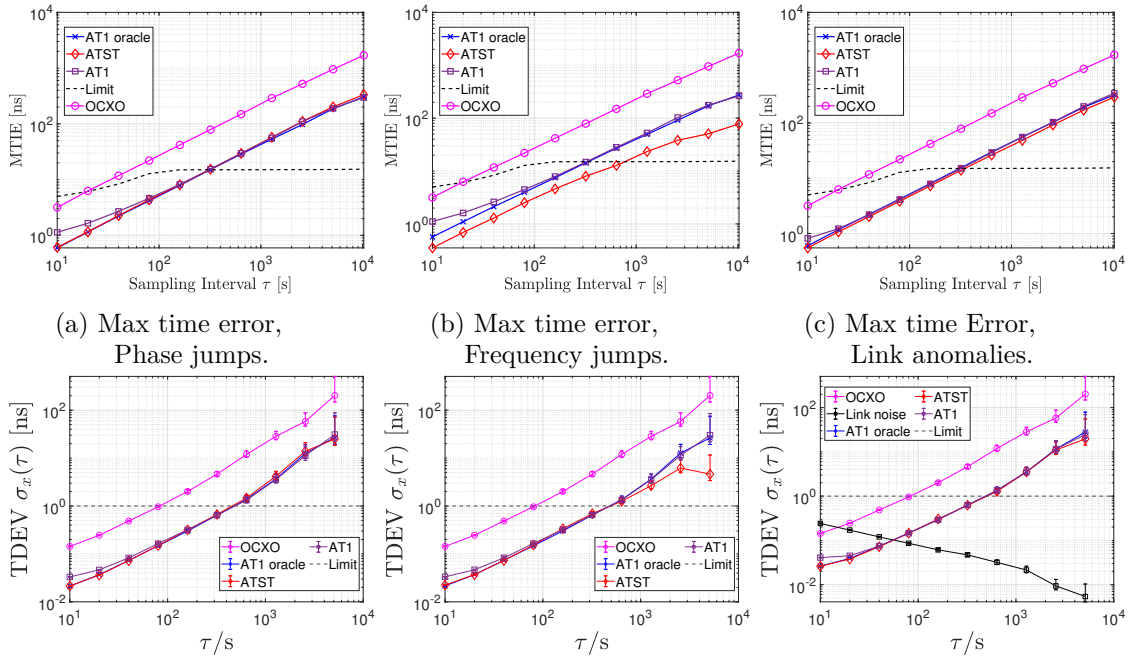
(f) Overlapping Allan Dev., Frequency jumps.

(g) Phase evolution, Link anomalies.

(h) Fractional frequency, Link anomalies.

(i) Overlapping Allan Dev., Link anomalies.

Figure D.1: Time scale performance with phase jumps (top row), frequency jumps (middle row), and link anomalies (bottom row). The base measurement noise without anomalies is normally distributed with a variance of $\sigma_n^2 = 10^{-20}$.



(a) Max time error, Phase jumps.

(b) Max time error, Frequency jumps.

(c) Max time Error, Link anomalies.

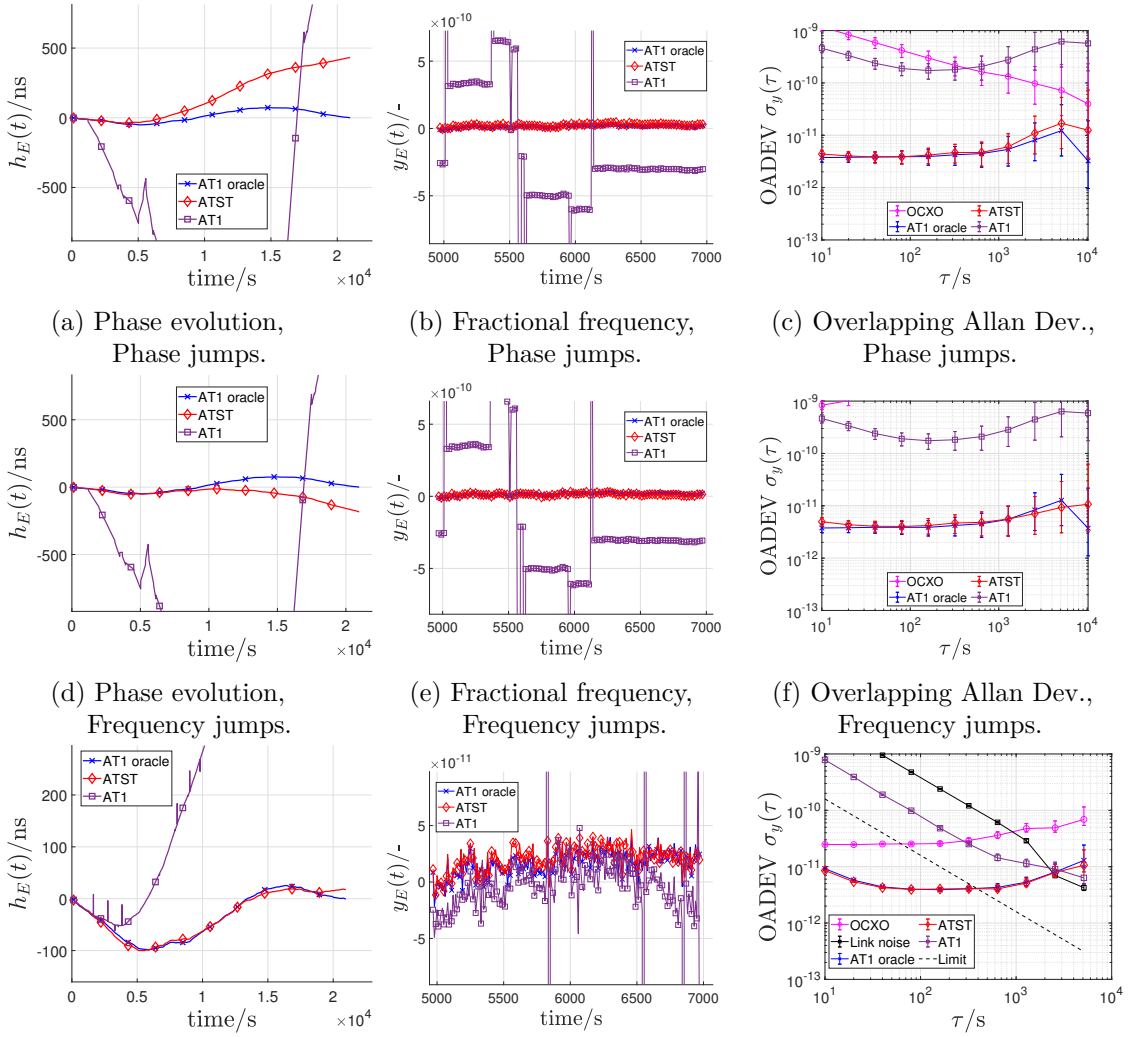
(d) Time dev., Phase jumps.

(e) Time dev., Frequency jumps.

(f) Time dev., Link anomalies.

Figure D.2: Maximum Time Interval Error (top row) and Time Deviation (bottom row) for the new time scale compared to the AT1 oracle with phase jumps of 10 ns (left column), frequency jumps of 10 ns/s (middle column), and anomalies on the measurement links of 10 ns (right column).The base measurement noise without anomalies is normally distributed with a variance of $\sigma_n^2 = 10^{-20}$.

Figures D.3 and D.4 display the time scale performance with larger magnitudes of anomalies: $\sigma_{\Delta x} = 1$ µs, $\sigma_{\Delta y} = 1$ µs/s, and $\sigma_{\Delta n_{ji}} = 1$ µs. The larger scale anomalies may not be as realistic but in the case they do occur it is useful to identify the behavior of the time scale algorithm. The ATST time scale does not match the AT1 oracle perfectly in ADEV but still manages to autonomously mitigate the effects to a reasonable level.



(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.

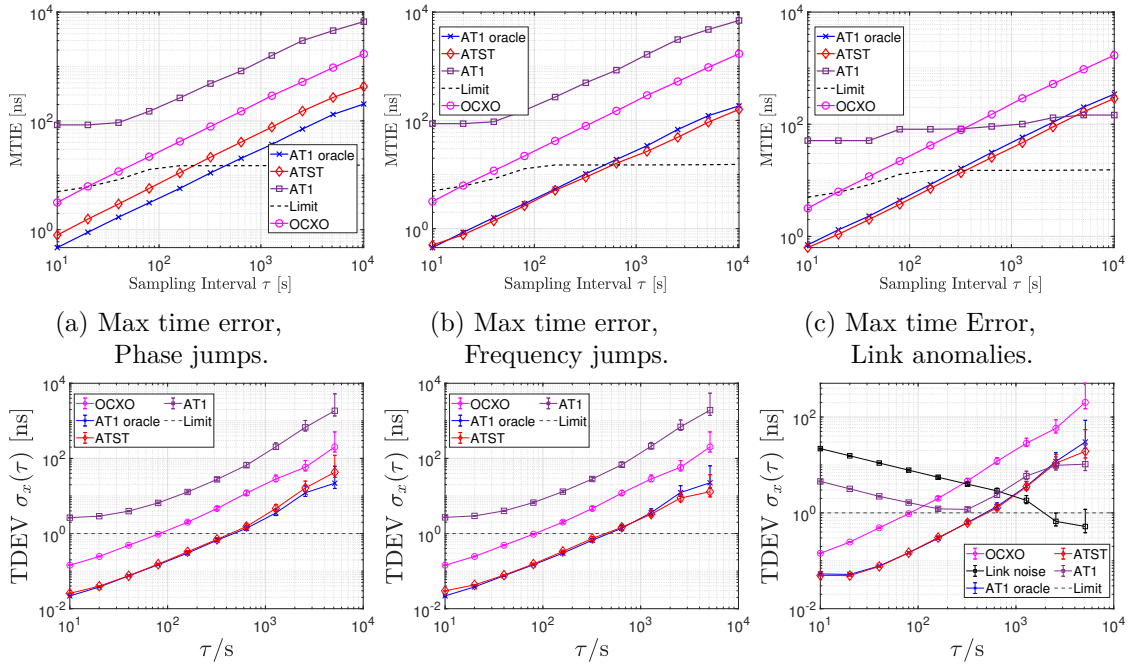
(f) Overlapping Allan Dev., Frequency jumps.

(g) Phase evolution, Link anomalies.

(h) Fractional frequency, Link anomalies.

(i) Overlapping Allan Dev., Link anomalies.

Figure D.3: Time scale performance with phase jumps of 1 µs (top row), frequency jumps (middle row) of 1 µs/s, and link anomalies of 1 µs (bottom row). The base measurement noise without anomalies is normally distributed with a variance of $\sigma_n^2 = 10^{-19}$.



(a) Max time error, Phase jumps.

(b) Max time error, Frequency jumps.

(c) Max time Error, Link anomalies.

(d) Time dev., Phase jumps.

(e) Time dev., Frequency jumps.

(f) Time dev., Link anomalies.

Figure D.4: Maximum Time Interval Error (top row) and Time Deviation (bottom row) for the new time scale compared to the AT1 oracle with phase jumps of 1 µs (left column), frequency jumps of 1 µs/s (middle column), and anomalies on the measurement links of 1 µs (right column).The base measurement noise without anomalies is normally distributed with a variance of $\sigma_n^2 = 10^{-19}$.

Appendix E

Estimator Derivations

E.1 Maximum Likelihood Estimator - Student's t-distribution

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 [66]:

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

with $\mathbf{z}=(z_1,...,z_N)^T$. The parameters μ, σ^2, and ν are the mean, scale, and shape parameters, respectively. The shape parameter is also referred to as the number of degrees of freedom and is related to the quantity of outliers in the data. A random variable z_i that follows the Student's t-distribution is denoted as $z_i \sim T(\mu,\sigma^2,\nu)$. The MLE for each of the parameters μ, σ^2, and ν aims to identify the values of those parameters that maximize the likelihood (E.1) for a given sample $\mathbf{z}$:

$$\left[\hat{\mu},\hat{\sigma}^2,\hat{\nu}\right]^T=\arg\max_{\mu,\sigma^2,\nu}\Big\{L(\boldsymbol{z};\mu,\sigma^2,\nu)\Big\}. \quad \text{(E.2)}$$

To simplify the derivations, it is usual to derive the expression of the MLE by minimizing the negative log-likelihood [62]

$$\left[\hat{\mu},\hat{\sigma}^2,\hat{\nu}\right]^T=\arg\min_{\mu,\sigma^2,\nu}\Big\{-\log L(\boldsymbol{z};\mu,\sigma^2,\nu)\Big\}. \quad \text{(E.3)}$$

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

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

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

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

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

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

As each parameter depends on the other two, the MLEs of μ, σ^2, ν cannot be computed directly. However, it is well known that the Student distribution can be represented by an infinite mixture of Gaussian distributions:

$$z_i | v_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{v_i}\right), \; v_i \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{2}{\nu}\right), \tag{E.8}$$

where $\mathcal{G}(a, b)$ denotes a gamma distribution with parameters a and b. The joint PDF of $\mathbf{z} = [z_1, \cdots, z_N]^T$ and $\mathbf{v} = [v_1, \cdots, v_N]^T$, referred to as complete likelihood, is expressed as:

$$L_c(\mathbf{z}, \mathbf{v}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)} v_i^{\frac{\nu+1}{2}-1} \exp\left(-\frac{v_i}{2} \left(\nu + \left(\frac{z_i - \mu}{\sigma^2} \right)^2 \right) \right), \tag{E.9}$$

Marginalizing the complete likelihood with respect to $\mathbf{v}$ yields the likelihood as described in (E.1). This representation allows an expectation maximization (EM) algorithm to be derived [82]. The EM algorithm is based on the so-called complete log-likelihood, which is the logarithm of the joint distribution of $(\mathbf{z}, \mathbf{v})$:

$$l_c(\mathbf{z}, \mathbf{v}) = \frac{N\nu}{2} \log\left(\frac{N\nu}{2}\right) - N \log\Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu+1}{2} - 1\right) \sum_{i=1}^{N} \log v_i - \frac{N}{2} \log(2\pi) \tag{E.10}$$

$$- \frac{N}{2} \log\left(\sigma^2\right) - \frac{1}{2} \sum_{i=1}^{N} v_i \left[\nu + \frac{(z_i - \mu)^2}{\sigma^2} \right]. \tag{E.11}$$

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:

$$\hat{\mu}_0 = \frac{1}{N}\sum_{i=1}^{N} z_i, \tag{E.12}$$

$$\hat{\sigma}_0^2 = \frac{1}{N-1}\sum_{i=1}^{N}(z_i - \hat{\mu}_0)^2, \tag{E.13}$$

$$\hat{\nu}_0 = 3. \tag{E.14}$$

- **E Step**: At iteration k, given $\hat{\boldsymbol{\theta}}_{k-1} = \left(\hat{\mu}_{k-1}, \hat{\sigma}_{k-1}^2, \hat{\nu}_{k-1}\right)^T$, 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

$$u_{i,k} = E[v_i | z_i, \hat{\theta}_{k-1}] = \frac{\hat{\nu}_{k-1} + 1}{\hat{\nu}_{k-1} + \frac{(z_i - \hat{\mu}_{k-1})^2}{\hat{\sigma}_{k-1}^2}}, \tag{E.15}$$

$$w_{i,k} = E[\log(v_i) | z_i, \hat{\theta}_{k-1}] = \psi\left(\frac{\hat{\nu}_{k-1}+1}{2}\right) - \log\left(\frac{1}{2}\left(\hat{\nu}_{k-1} + \frac{(z_i - \hat{\mu}_{k-1})^2}{\hat{\sigma}_{k-1}^2}\right)\right), \tag{E.16}$$

and leads to the objective function Q

$$Q(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_k) = \frac{N\nu}{2}\log\left(\frac{N\nu}{2}\right) - N\log\Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu+1}{2} - 1\right)\sum_{i=1}^{N} w_{i,k} - \frac{N}{2}\log(2\pi) \tag{E.17}$$

$$-\frac{N}{2}\log\left(\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N} u_{i,k}\left[\nu + \frac{(z_i - \mu)^2}{\sigma^2}\right]. \tag{E.18}$$

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

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} u_{i,k} z_i}{\sum_{i=1}^{N} u_{i,k}}, \tag{E.19}$$

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^{N} u_{i,k}(z_i - \hat{\mu}_k)^2}{N}, \tag{E.20}$$

$$N\phi\left(\frac{\hat{\nu}_k}{2}\right) + \sum_{i=1}^{N}\left[u_{i,k} - w_{i,k} - 1\right] = 0. \tag{E.21}$$

The terms $u_{i,k}$ act as weights in the estimate of the location and scale parameters. These weights take small values for samples that are far from the estimated mean and are approximately equal if the degrees of freedom are large, reducing the estimator to the classical sample mean and variance or MLEs for the parameters of a Gaussian distribution. By assigning lower values to outliers the weighted average is able to mitigate the effects of anomalies.

The solution to (E.21) is obtained through the application of Newton's method, which converges to a solution after few iterations. Consider the solution to (E.21) is the root of the function $f(x)$, then Newton's method allows us to iteratively approximate that root by the following

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{E.22}$$

where the initial guess can be the previous estimate of the number of degrees of freedom, $x_0 = \hat{\nu}_{k-1}(t)$. The above is repeated until reaching a maximum number of iterations or a minimum difference between consecutive approximations is reached. Once the error between consecutive estimates has reached the chosen minimum, the estimators have converged.

E.2 Maximum Likelihood Estimation - Bimodal Gaussian Mixture

Assuming the Bimodal Gaussian Mixture distribution is used to model contaminated data with proportion of anomalous data ε and variance scaling factor k, the likelihood function is:

$$\begin{aligned} L(\mathbf{z};\boldsymbol{\theta}) &= \prod_{i=1}^{N} p(z_i;\boldsymbol{\theta}) \\ L(\mathbf{z};\boldsymbol{\theta}) &= \prod_{i=1}^{N} \frac{1-\varepsilon}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{z_i-\mu}{\sigma}\right)^2\right) + \frac{\varepsilon}{\sqrt{2\pi k\sigma^2}} \exp\left(-\frac{1}{2k}\left(\frac{z_i-\mu}{\sigma}\right)^2\right), \end{aligned} \tag{E.23}$$

with $\mathbf{z} = (z_1, ..., z_N)^T$ and $\boldsymbol{\theta} = [\mu, \sigma^2, \varepsilon, k]^T$. To simplify the expression of the PDF and furthermore simplify the following derivations, the PDF is rewritten such that the exponentials are presented by two sub-functions $g_0(z_i;\boldsymbol{\eta}_0)$ and $g_1(z_i;\boldsymbol{\eta}_1)$

$$L(\mathbf{z};\boldsymbol{\theta}) = \prod_{i=1}^{N} (1-\varepsilon) g_0(z_i;\boldsymbol{\eta}_0) + \varepsilon g_1(z_i;\boldsymbol{\eta}_1), \tag{E.24}$$

where each of the sub-functions g_0 and g_1 is the PDF of an equivalent Gaussian distribution and their respective reduced parameter vectors are $\boldsymbol{\eta}_0 = [\mu, \sigma^2]^T$ and $\boldsymbol{\eta}_1 = [\mu, \sigma^2, k]^T$. The

MLE for the parameters defining the BGM are the respective values that maximize the likelihood (E.1) for a given sample $\mathbf{z}$:

$$\hat{\boldsymbol{\theta}} = \left[\hat{\mu}, \hat{\sigma}^2, \hat{\varepsilon}, \hat{k}\right]^T = \arg\max_{\boldsymbol{\theta}} \{L(\boldsymbol{z}; \boldsymbol{\theta})\}. \tag{E.25}$$

The log-likelihood is maximized instead of the likelihood function, which still results in the appropriate MLE for each parameter

$$\log\left(L(\mathbf{z}; \boldsymbol{\theta})\right) = \sum_{i=1}^{N} \log\left[(1-\varepsilon) g_0(z_i; \boldsymbol{\eta}_0) + \varepsilon g_1(z_i; \boldsymbol{\eta}_1)\right]. \tag{E.26}$$

In the case of the univariate BGM, the MLE of the location parameter is the solution to the following equation

$$\frac{\partial \log\left(L(\mathbf{z}; \boldsymbol{\theta})\right)}{\partial \mu} = \sum_{i=1}^{N} \frac{1}{L(z_i; \boldsymbol{\theta})} \left[(1-\varepsilon) \frac{\partial g_0(z_i; \boldsymbol{\eta}_0)}{\partial \mu} + \varepsilon \frac{\partial g_1(z_i; \boldsymbol{\eta}_1)}{\partial \mu}\right] = 0, \tag{E.27}$$

with the derivatives of the sub-functions

$$\frac{\partial g_0(z_i; \boldsymbol{\eta}_0)}{\partial \mu} = \left(\frac{z_i - \mu}{\sigma^2}\right) g_0(z_i; \boldsymbol{\eta}_0), \tag{E.28}$$

$$\frac{\partial g_1(z_i; \boldsymbol{\eta}_1)}{\partial \mu} = \left(\frac{z_i - \mu}{\sigma^2}\right) \frac{g_1(z_i; \boldsymbol{\eta}_1)}{k}, \tag{E.29}$$

so the equation for the MLE of the mean becomes

$$\frac{\partial \log\left(L(\mathbf{z}; \boldsymbol{\theta})\right)}{\partial \mu} = \sum_{i=1}^{N} \frac{\left(\frac{z_i - \mu}{\sigma^2}\right)}{L(z_i; \boldsymbol{\theta})} \left[(1-\varepsilon) g_0(z_i; \boldsymbol{\eta}_0) + \frac{\varepsilon}{k} g_1(z_i; \boldsymbol{\eta}_1)\right] = 0. \tag{E.30}$$

The above expression does not have a closed form expression for μ that does not depend on the other parameters. It can similarly be shown that the estimators for the other parameters will also depend on the mean. Due to this interdependence of the estimators, the method of Expectation Maximization (EM) is required to iteratively estimate each parameter until converging to an approximation of the MLE.

To define the EM algorithm, latent variables must be introduced that allow a complete likelihood function to be defined. Since there are two modes of the BGM, one linked to normal data and the other linked to anomalous data, a latent variable γ_i is introduced as an outlier label for each sample so can take two possible values with known probabilities. The latent variable is then a Bernoulli random variable

$$\gamma_i \sim \mathcal{B}(1, \varepsilon), \tag{E.31}$$

where the two possible labels are $\gamma_i = 0$ for normal data and $\gamma_i = 1$ for anomalous data. The probability that either label is assigned to the latent variable is linked to the

contamination proportion $L(\gamma_i = 0) = 1-\varepsilon$ and $L(\gamma_i = 1) = \varepsilon$. The conditional probability for the samples z_i given a certain label is linked to the sub-functions expressed above

$$L(z_i|\gamma_i = l) = g_l(z_i; \boldsymbol{\eta}_l), \tag{E.32}$$

i.e., the samples labeled as normal data will effectively be samples of the nominal Gaussian mode $g_0(z_i; \boldsymbol{\eta}_0)$ and the anomalies come from the contaminating Gaussian mode $g_1(z_i; \boldsymbol{\eta}_1)$. The joint PDF of $\mathbf{z} = [z_1, \cdots, z_N]^T$ and $\gamma = [\gamma_1, \cdots, \gamma_N]^T$, referred to as complete likelihood, is expressed as:

$$L_c(\mathbf{z}, \boldsymbol{\gamma}) = \varepsilon^{N_a}(1-\varepsilon)^{N-N_a} \prod_{i=1}^{N} g_0(z_i, \boldsymbol{\eta}_0)^{1-\gamma_i} g_1(z_i, \boldsymbol{\eta}_1)^{\gamma_i}, \tag{E.33}$$

where $N_a = \sum_{i=1}^{N} \gamma_i$ is the number of samples that have been labeled as anomalous. The EM algorithm is based on the so-called complete log-likelihood, which is the logarithm of the joint distribution of $(\mathbf{z}, \boldsymbol{\gamma})$:

$$l_c(\mathbf{z}, \boldsymbol{\gamma}) = N_a \log(\varepsilon) + (N - N_a)\log(1-\varepsilon) \tag{E.34}$$

$$+ \sum_{i=1}^{N} (1-\gamma_i) \log\left(g_0(z_i, \boldsymbol{\eta}_0)\right) + \sum_{i=1}^{N} \gamma_i \log\left(g_1(z_i, \boldsymbol{\eta}_1)\right). \tag{E.35}$$

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:

$$\hat{\mu}_0 = \frac{1}{N} \sum_{i=1}^{N} z_i, \tag{E.36}$$

$$\hat{\sigma}_0^2 = \frac{1}{N-1} \sum_{i=1}^{N} (z_i - \hat{\mu}_0)^2, \tag{E.37}$$

$$\hat{\varepsilon}_0 = 0.01, \tag{E.38}$$

$$\hat{k}_0 = 1.5. \tag{E.39}$$

- **E Step**: At iteration j, given $\hat{\boldsymbol{\theta}}_{j-1} = \left(\hat{\mu}_{j-1}, \hat{\sigma}_{j-1}^2, \hat{\varepsilon}_{k-1}, \hat{k}_{j-1}\right)^T$, the E step computes the expectation of $l_c(\mathbf{z}, \boldsymbol{\gamma})$ with respect to the variables γ_i, and leads to the objective

function Q

$$
\begin{aligned}
Q(\boldsymbol{\theta};\hat{\boldsymbol{\theta}}_j) = E[l_c(\mathbf{z},\boldsymbol{\gamma})|\mathbf{z},\hat{\boldsymbol{\theta}}_{j-1}] &= \log(\varepsilon)\sum_{i=1}^{N} E\left[\gamma_i|\mathbf{z},\hat{\boldsymbol{\theta}}_{j-1}\right] + N\log(1-\varepsilon) \\
&- \log(1-\varepsilon)\sum_{i=1}^{N} E\left[\gamma_i|\mathbf{z},\hat{\boldsymbol{\theta}}_{j-1}\right] + \sum_{i=1}^{N}\log\left(g_0(z_i;\boldsymbol{\eta}_0)\right) \\
&+ \sum_{i=1}^{N} E\left[\gamma_i|\mathbf{z},\hat{\boldsymbol{\theta}}_{j-1}\right]\left(\log\left(\frac{g_1(z_i;\boldsymbol{\eta}_1)}{g_0(z_i;\boldsymbol{\eta}_0)}\right)\right)
\end{aligned} \tag{E.40}
$$

The objective function requires some estimation of the latent variable in the form of it's expected value given the measurements z_i and previous estimates of the parameters of interest $\hat{\boldsymbol{\theta}}_{j-1}$. For a Bernoulli random variable it is known that the expectation is equivalent to the probability of of success, in this case, the likelihood of an outlier

$$
u_{i,j} = E[\gamma_i|z_i,\hat{\boldsymbol{\theta}}_{j-1}] = L(\gamma_i = 1|z_i) = \frac{L(\gamma_i=1)L(z_i,\hat{\boldsymbol{\theta}}_{j-1}|\gamma_i=1)}{L(z_i;\hat{\boldsymbol{\theta}}_{j-1})}, \tag{E.41}
$$

$$
u_{i,j} = \frac{\hat{\varepsilon}_{j-1}g_1(z_i;\hat{\boldsymbol{\eta}}_{1,j-1})}{(1-\hat{\varepsilon}_{j-1})g_0(z_i;\hat{\boldsymbol{\eta}}_{0,j-1}) + \hat{\varepsilon}_{j-1}g_1(z_i;\hat{\boldsymbol{\eta}}_{1,j-1})}, \tag{E.42}
$$

where the estimated reduced parameter vectors are $\hat{\boldsymbol{\eta}}_{0,j-1} = [\hat{\mu}_{j-1},\hat{\sigma}^2_{j-1}]^T$ and $\hat{\boldsymbol{\eta}}_{1,j-1} = [\hat{\mu}_{j-1},\hat{\sigma}^2_{j-1},\hat{k}_{j-1}]^T$. The term computed above provides a sort of normalized probability of an outlier occurring, i.e., $u_{i,j}$ is large if the likelihood of sample i being an outlier is close to the total likelihood of that observation according to the BGM. The objective function is updated to include the expectation of the latent variable and written in terms of the parameters to be estimated

$$
\begin{aligned}
Q(\boldsymbol{\theta};\hat{\boldsymbol{\theta}}_j) &= \log(\varepsilon)\sum_{i=1}^{N} u_{i,j} + N\log(1-\varepsilon) - \log(1-\varepsilon)\sum_{i=1}^{N} u_{i,j} \\
&- \frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{(z_i-\mu)^2}{\sigma^2} - \frac{\log(k)}{2}\sum_{i=1}^{N} u_{i,j} \\
&+ \frac{1}{2}\sum_{i=1}^{N} u_{i,j}\frac{(z_i-\mu)^2}{\sigma^2} - \frac{1}{2k}\sum_{i=1}^{N} u_{i,j}\frac{(z_i-\mu)^2}{\sigma^2}
\end{aligned} \tag{E.43}
$$

- **M Step**: At iteration k, the M Step maximizes the Q function with respect to the

parameters μ, σ^2, ε, k, which yields

$$\hat{\mu}_j = \frac{\sum_{i=1}^{N} w_{i,j} z_i}{\sum_{i=1}^{N} w_{i,j}}, \tag{E.44}$$

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^{N} w_{i,j} (z_i - \hat{\mu}_k)^2}{N-1}, \tag{E.45}$$

$$\hat{\varepsilon}_j = \frac{\sum_{i=1}^{N} u_{i,j}}{N}, \tag{E.46}$$

$$\hat{k}_j = \frac{\sum_{i=1}^{N} u_{i,j} (z_i - \hat{\mu}_j)^2}{\hat{\sigma}_j^2 \sum_{i=1}^{N} u_{i,j}}. \tag{E.47}$$

The weights used to estimate the location and scale parameters are

$$w_{i,j} = 1 - u_{i,j} + \frac{u_{i,j}}{\hat{k}_{j-1}}, \tag{E.48}$$

which can be understood as the normalized probability that a sample is nominal $(1 - u_{i,j)}$ summed with the normalized probability of the sample being an anomaly divided by the scaling factor of the anomalous variance $(\frac{u_{i,j}}{k})$. For anomalous data, $u_{i,j}$ is large as well as k. Therefore the weights $w_{i,j}$ decrease proportionally with the scaling factor of the contaminating variance and are small for anomalous samples. This results in a robust estimate of the mean and nominal variance of the BGM.

The iterative EM steps are repeated until the error between consecutive estimates reduces below a chosen threshold or a maximum number of iterations has occurred. By appropriately initializing the estimates, the number of iterations should remain low. Nevertheless the presence of anomalies complicates the difference between the initial estimates and true parameters so computational limitations should be considered when applying EM algorithms defined for different distributions.

Appendix F

Bound Derivations

F.1 Derivation of pseudo-true parameters.

The pseudo-true parameters $\tilde{\boldsymbol{\theta}} = [\tilde{\mu}, \tilde{\sigma}^2]^T$ are the parameters of the assumed distribution that minimize the KLD from the true distribution

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{D_{\mathrm{KL}}\} = \arg\min_{\boldsymbol{\theta}} \left\{ E_p \left[\log \left(\frac{p(\mathbf{z}; \boldsymbol{\eta})}{q(\mathbf{z}; \boldsymbol{\theta})} \right) \right] \right\}, \tag{F.1}$$

where $p(z; \boldsymbol{\eta})$ represents the true distribution of the data and can be replaced by either $p_t(z; \mu_t, \sigma_t^2, \nu)$ or $p_{\mathrm{GM}}(z; \mu, \sigma^2, k, \varepsilon)$. The choice of the true distribution is based on how anomalies can appear in the observations. The cost function for finding the pseudo-true parameters is simplified to only include the parameters of the Gaussian distribution:

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{-E_p [\log (q(\mathbf{z}; \boldsymbol{\theta}))]\}. \tag{F.2}$$

Substituting the log-likelihood function for the Gaussian distribution leads to:

$$-E_p [\log (q(\mathbf{z}; \boldsymbol{\theta}))] = \frac{N}{2} E_p \left[\log \left(2\pi\sigma_G^2 \right) \right] \tag{F.3}$$

$$+ \frac{1}{2} \sum_{i=1}^{N} E_p \left[\left(\frac{z_i - \mu_G}{\sigma_G} \right)^2 \right]. \tag{F.4}$$

We then expand the quadratic function of z to separate the misspecified parameters μ_G and σ_G^2 from the expectation

$$-E_p [\log (q(\mathbf{z}; \boldsymbol{\theta}))] = \frac{N}{2} \log \left(2\pi\sigma_G^2 \right) + \frac{1}{2\sigma_G^2} \left(\sum_{i=1}^{N} E_p \left[z_i^2 \right] - 2\mu_G \sum_{i=1}^{N} E_p[z_i] + N\mu_G^2 \right). \tag{F.5}$$

The mean and variance of a random variable that follows the Student's t-distribution are known to be:

$$E_{p_t}[z] = \mu_t, \mathrm{var}(z) = E_{p_t}[z_i^2] - E_{p_t}[z_i]^2 = \sigma_t^2 \frac{\nu}{\nu - 2}.$$

Similarly, the moments of the BGM model are known. The expectation of the BGM can be written as a linear combination of the expectations with respect to each of the modes. Using the BGM PDF

$$p_{\mathrm{GM}}(z; \mu, \sigma, k, \varepsilon) = (1 - \varepsilon) g_1(z; \mu, \sigma^2) + \varepsilon g_2(z; \mu, k\sigma^2), \tag{F.6}$$

where $g_1(z; \mu, \sigma^2)$ and $g_2(z; \mu, k\sigma^2)$ are the PDFs of the corresponding Gaussian distributions, the expected values can be evaluated as follows

$$E_{p_{\mathrm{GM}}}[z_i] = (1 - \varepsilon) E_{g_1}[z_i] + \varepsilon E_{g_2}[z_i], \tag{F.7}$$

$$E_{p_{\mathrm{GM}}}[z_i] = \mu, \tag{F.8}$$

$$E_{p_{\mathrm{GM}}}\left[z_i^2\right] = (1 - \varepsilon) E_{g_1}\left[z_i^2\right] + \varepsilon E_{g_2}\left[z_i^2\right], \tag{F.9}$$

$$E_{p_{\mathrm{GM}}}\left[z_i^2\right] = (1 - \varepsilon + \varepsilon k)\sigma^2 + \mu^2 = \mathrm{var}(z) + E_{p_{\mathrm{GM}}}[z]^2. \tag{F.10}$$

The objective function to be optimized becomes:

$$\begin{aligned} -E_p\left[\log\left(q(\mathbf{z}; \boldsymbol{\theta})\right)\right] &= \frac{N}{2} \log\left(2\pi\sigma_G^2\right) \\ &+ \frac{1}{2\sigma_G^2}\left(\sum_{i=1}^{N} \mathrm{var}(z_i) + \sum_{i=1}^{N} E_p[z_i]^2 - 2\mu_G \sum_{i=1}^{N} E_p[z_i] + N\mu_G^2\right), \end{aligned} \tag{F.11}$$

which can be further simplified to make it easier to obtain a generalized result for the pseudo-true parameter μ_{pt}:

$$\begin{aligned} -E_p\left[\log\left(q(\mathbf{z}; \boldsymbol{\theta})\right)\right] &= \frac{N}{2} \log\left(2\pi\sigma_G^2\right) \\ &+ \frac{N}{2\sigma_G^2}\left(\mathrm{var}(z_i) + (E_p[z_i] - \mu_G)^2\right). \end{aligned} \tag{F.12}$$

The value of μ_G that minimizes the above cost function is obtained when $\mu_G = E_p[z]$. Therefore,

$$\tilde{\mu} = E_p[z], \tag{F.13}$$

where the true distribution determines whether the pseudo-true location parameter coincides with the true parameter. As was shown above, random variables following the Student's t-distribution and the BGM both have a simple expression for the expected value. For the pseudo-true scale parameter, the following result is obtained:

$$-\frac{\partial}{\partial \sigma_G^2} E_p\left[\log\left(q(\mathbf{z}; \boldsymbol{\theta})\right)\right] = \frac{N}{2\sigma_G^2} \tag{F.14}$$

$$- \frac{N}{2(\sigma_G^2)^2}\left(\mathrm{var}(z_i) + (E_p[z_i] - \mu_G)^2\right). \tag{F.15}$$

Substituting the pseudo-true parameter for μ_G, one obtains:

$$\frac{N\sigma_G^2}{2} - \frac{N}{2}\text{var}(z) = 0. \tag{F.16}$$

The resulting pseudo-true parameter is the sample variance of the Student's t-distribution:

$$\tilde{\sigma}^2 = \text{var}_{p_t}(z), \tag{F.17}$$

where the subscript p_t specifies the variance of the true distribution of the data. With the above result, we can conclude that the mean and variance of the Student's t-distribution are the pseudo-true parameters that minimize the KLD between a Gaussian distribution and a Student's t-distribution

$$\tilde{\mu}_t = \mu_t, \tag{F.18}$$

$$\tilde{\sigma}_t^2 = \sigma_t^2 \frac{\nu}{\nu - 2}. \tag{F.19}$$

The same is true for the BGM with the variance being a linear combination of the variances of two Gaussian modes because each mode has the same mean.

$$\tilde{\mu}_{\text{GM}} = \mu, \tag{F.20}$$

$$\tilde{\sigma}_{\text{GM}}^2 = \sigma^2 \left((k-1)\varepsilon + 1\right). \tag{F.21}$$

F.2 Deriving Matrices A and B

The log-likelihood function of N i.i.d. Gaussian random variables in $\mathbf{z} = (z_1, \cdots, z_N)^T$ is:

$$\log(q(\mathbf{z}; \boldsymbol{\theta})) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma_G^2) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{z_i - \mu_G}{\sigma_G}\right)^2. \tag{F.22}$$

Instead of using the joint PDF, we take advantage of the linearity of derivatives and expectations to compute **A** and **B** using the marginal PDF for a single sample:

$$\log(q(z_i; \boldsymbol{\theta})) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_G^2) - \frac{1}{2}\left(\frac{z_i - \mu_G}{\sigma_G}\right)^2. \tag{F.23}$$

The equation for the MCRB then takes into account the N i.i.d. random variables. Computing the Hessian with respect to the parameter vector $\boldsymbol{\theta} = [\mu_G, \sigma_G^2]$ provides all the terms

required to compute the required matrices:

$$\mathbf{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} = \left(E_p \left[\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}, \tag{F.24}$$

$$= \left(E_p \begin{bmatrix} \frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G^2} & \frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G \partial \sigma_G^2} \\ \frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \sigma_G^2 \partial \mu_G} & \frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial (\sigma_G^2)^2} \end{bmatrix} \right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \tag{F.25}$$

Straightforward computations lead to:

$$\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G} = \frac{z_i - \mu_G}{\sigma_G^2}, \tag{F.26}$$

$$\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G^2} = -\frac{1}{\sigma_G^2}, \tag{F.27}$$

$$\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G \partial \sigma_G^2} = -\frac{z_i - \mu_G}{(\sigma_G^2)^2}, \tag{F.28}$$

$$\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \sigma_G^2} = -\frac{1}{2\sigma_G^2} + \frac{1}{2}\frac{(z_i - \mu_G)^2}{(\sigma_G^2)^2}, \tag{F.29}$$

$$\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial (\sigma_G^2)^2} = \frac{1}{2(\sigma_G^2)^2} - \frac{1}{(\sigma_G^2)^2}\left(\frac{z_i - \mu_G}{\sigma_G}\right)^2, \tag{F.30}$$

$$\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \sigma_G^2 \partial \mu_G} = -\frac{z_i - \mu_G}{(\sigma_G^2)^2}. \tag{F.31}$$

We follow a similar process to Appendix F.1 to find the expectations of the above expressions. Substituting the pseudo-true values $\mu_G = \tilde{\mu}_p$, $\sigma_G^2 = \tilde{\sigma}_p^2$ yields:

$$A_{1,1} = \left(E_p \left[\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G^2} \right] \right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} = -\left(\mathrm{var}_p(z_i)\right)^{-1}, \tag{F.32}$$

$$A_{1,2} = \left(E_p \left[\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G \partial \sigma_G^2} \right] \right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} = 0, \tag{F.33}$$

$$A_{2,2} = \left(E_p \left[\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial (\sigma_G^2)^2} \right] \right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} = -\frac{1}{2}\left(\mathrm{var}_p(z_i)\right)^{-2}, \tag{F.34}$$

$$A_{2,1} = \left(E_p \left[\frac{\partial^2 \log(q(z_i;\boldsymbol{\theta}))}{\partial \sigma_G^2 \partial \mu_G} \right] \right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} = 0. \tag{F.35}$$

where the term $A_{2,2}$ has been computed using $E_p\left[\left(\frac{z_i-\mu_G}{\sigma_G}\right)^2\right]$ from Appendix F.1:

$$\left(E_p\left[\left(\frac{z_i-\mu_G}{\sigma_G}\right)^2\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} = \frac{\left(\mathrm{var}_p(z_i) + (E_p[z_i] - E_p[z_i])^2\right)}{\mathrm{var}_p(z_i)} = 1. \tag{F.36}$$

This provides all expressions necessary for **A**:

$$\mathbf{A}(\tilde{\boldsymbol{\theta}}_p) = \begin{bmatrix} -\left(\mathrm{var}_p(z_i)\right)^{-1} & 0 \\ 0 & -\frac{1}{2}\left(\mathrm{var}_p(z_i)\right)^{-2} \end{bmatrix}. \tag{F.37}$$

The elements of **B** can be computed using the following derivatives:

$$\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G}\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G} = \left(\frac{z_i-\mu_G}{\sigma_G^2}\right)^2, \tag{F.38}$$

$$\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G}\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \sigma_G^2} = \left(\frac{z_i-\mu_G}{\sigma_G^2}\right)\left(-\frac{1}{2\sigma_G^2} + \frac{1}{2}\frac{(z_i-\mu_G)^2}{(\sigma_G^2)^2}\right), \tag{F.39}$$

$$\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \sigma_G^2}\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \sigma_G^2} = \left(-\frac{1}{2\sigma_G^2} + \frac{1}{2}\frac{(z_i-\mu_G)^2}{(\sigma_G^2)^2}\right)^2, \tag{F.40}$$

hence

$$\mathbf{B} = \left(E_p\left[\left(\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^T}\right)\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} \tag{F.41}$$

$$= \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix}. \tag{F.42}$$

The expectation of $B_{1,1}$ is computed using (F.36):

$$E_p\left[\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G}\frac{\partial \log(q(z_i;\boldsymbol{\theta}))}{\partial \mu_G}\right] = \frac{1}{\sigma_G^2}E_p\left[\left(\frac{z_i-\mu_G}{\sigma_G}\right)^2\right]. \tag{F.43}$$

Substituting the pseudo-true values and using (F.36) leads to:

$$B_{1,1} = \left(\mathrm{var}_p(z_i)\right)^{-1}. \tag{F.44}$$

The terms $B_{1,2}$ and $B_{2,1}$ are:

$$B_{1,2} = B_{2,1} = -\frac{1}{2}\left(E_p\left[\frac{z_i - \mu_G}{(\sigma_G^2)^3}\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} + \left(\frac{1}{2\sigma_G^3}E_p\left[\left(\frac{z_i - \mu_G}{\sigma_G}\right)\left(\frac{z_i - \mu_G}{\sigma_G}\right)^2\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}, \tag{F.45}$$

$$B_{1,2} = \left(\frac{1}{2\sigma_G^3}E_p\left[\left(\frac{z_i - \mu_G}{\sigma_G}\right)^3\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \tag{F.46}$$

The third-order moment as described above is zero for a symmetric distribution around the value μ_G. However, the true distribution p is symmetric around $E_p[z_i]$, so we have to change the expression

$$B_{1,2} = \left(\frac{1}{2(\sigma_G^2)^3}E_p\left[\left((z_i - E_p[z_i]) + (E_p[z_i] - \mu_G)\right)^3\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \tag{F.47}$$

Expanding the cubic function, one has:

$$\begin{aligned} B_{1,2} = &\left(\frac{1}{2(\sigma_G^2)^3}E_p\left[(z - E_p[z_i])^3\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} \\ &+ \left(\frac{3(E_p[z_i] - \mu_G)}{2(\sigma_G^2)^3}E_p\left[(z_i - E_p[z_i])^2\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} \\ &+ \left(\frac{3(E_p[z_i] - \mu_G)^2}{2(\sigma_G^2)^3}E_p\left[(z_i - E_p[z_i])\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} \\ &+ \left(\frac{(E_p[z_i] - \mu_G)^3}{2(\sigma_G^2)^3}\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \end{aligned} \tag{F.48}$$

Substituting the pseudo-true parameter for the mean $\mu_G = E_p[z_i]$ yields:

$$B_{1,2} = B_{2,1} = \left(\frac{1}{2(\sigma_G^2)^3}E_p\left[(z_i - E_p[z_i])^3\right]\right)_{\sigma_G^2=\tilde{\sigma}_p^2}. \tag{F.49}$$

The above expectation is zero because we are now taking the third-order moment of a symmetric distribution around its mean, $E_p[z_i]$. Therefore, for the true symmetric distributions considered, the off-diagonal elements of $\mathbf{B}$ are:

$$B_{1,2} = B_{2,1} = 0. \tag{F.50}$$

Next, we expand the expression for $B_{2,2}$ and substitute the pseudo-true parameters

$$B_{2,2} = E_p\left[\frac{1}{4(\tilde{\sigma}_p^2)^2} - \frac{1}{2\tilde{\sigma}_p^2}\frac{(z_i-\tilde{\mu}_p)^2}{(\tilde{\sigma}_p^2)^2} + \frac{1}{4}\left(\frac{(z_i-\tilde{\mu}_p)^2}{(\tilde{\sigma}_p^2)^2}\right)^2\right], \tag{F.51}$$

then, factoring the constant terms

$$\begin{aligned} B_{2,2} = \frac{1}{4(\tilde{\sigma}_p^2)^2} &- \left(\frac{1}{2(\sigma_G^2)^3}E_p\left[(z_i-\mu_G)^2\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} \\ &+ \left(\frac{1}{4(\sigma_G^2)^4}E_p\left[(z_i-\mu_G)^4\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \end{aligned} \tag{F.52}$$

Again, we use the result of (F.36) and the independence of each z_i to further simplify

$$B_{2,2} = \frac{1}{4(\tilde{\sigma}_p^2)^2} - \frac{1}{2(\tilde{\sigma}_p^2)^2} + \left(\frac{1}{4(\sigma_G^2)^4}E_p\left[(z_i-\mu_G)^4\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \tag{F.53}$$

Expanding the 4th-order polynomial so it is presented in terms of the parameters of a Student's t-distribution:

$$E_p\left[(z_i-\mu_G)^4\right] = E_p\left[((z_i-E_p[z_i]) + (E_p[z_i]-\mu_G))^4\right] \tag{F.54}$$

$$\begin{aligned} E_p\left[(z_i-\mu_G)^4\right] &= E_p\left[(z_i-E_p[z_i])^4\right] \\ &+ 4(E_p[z_i]-\mu_G)E_p\left[(z_i-\mu_t)^3\right] \\ &+ 6(E_p[z_i]-\mu_G)^2E_p\left[(z_i-E_p[z_i])^2\right] \\ &+ 4(E_p[z_i]-\mu_G)^3E_p\left[z_i-E_p[z_i]\right] + (E_p[z_i]-\mu_G)^4. \end{aligned} \tag{F.55}$$

Again, substituting the pseudo-true value of the location parameter simplifies all terms that contain $E_p[z_i]-\mu_G = E_p[z_i]-\tilde{\mu}_p = 0$ for the true distributions investigated. The expression then becomes:

$$B_{2,2} = \left(-\frac{1}{4(\sigma_G^2)^2} + \frac{1}{4(\sigma_G^2)^4}E_p\left[(z_i-E_p[z_i])^4\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}. \tag{F.56}$$

For the final element of the **B** matrix, the fourth-order central moment of the true distribution should be evaluated. The expression will differ depending on the specific distribution that truly describes the data so the **B** matrix is split based on the true distribution. The fourth-order central moment is known for the Student's t-distribution [83]:

$$\left(E_{p_t}\left[(z_i-\mu_t)^4\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} = \frac{3\nu^2}{(\nu-2)(\nu-4)}(\sigma_t^2)^2, \tag{F.57}$$

We substitute the above results into the expression for $B_{2,2}$ to give the result for the misspecified scenario with the two possible true models. The first result being for the Student's t-distribution

$$B_{2,2}(p_t||q) = -\frac{1}{4(\tilde{\sigma}_{p_t}^2)^2} + \frac{\left(\sigma_t^2 \frac{\nu}{\nu-2}\right)^2}{4(\tilde{\sigma}_{p_t}^2)^4} \frac{3(\nu-2)}{(\nu-4)}, \tag{F.58}$$

and simplify,

$$B_{2,2}(p_t||q) = -\frac{1}{4}\left(\sigma_t^2 \frac{\nu}{\nu-2}\right)^{-2} + \frac{3(\nu-2)}{4(\nu-4)}\left(\sigma_t^2 \frac{\nu}{\nu-2}\right)^{-2}. \tag{F.59}$$

Therefore, we obtain the following matrix:

$$\mathbf{B}(p_t||q) = \begin{bmatrix} \left(\sigma_t^2 \frac{\nu}{\nu-2}\right)^{-1} & 0 \\ 0 & \left(\frac{3(\nu-2)}{4(\nu-4)} - \frac{1}{4}\right)\left(\sigma_t^2 \frac{\nu}{\nu-2}\right)^{-2} \end{bmatrix}. \tag{F.60}$$

For conciseness, we can write both $\mathbf{A}$ and $\mathbf{B}$ in terms of the pseudo-true scale parameter, which we know to be the variance of the true distribution, in this case being the Student's t-distribution

$$\mathbf{A} = \begin{bmatrix} -\left(\tilde{\sigma}_p^2\right)^{-1} & 0 \\ 0 & -\frac{1}{2}\left(\tilde{\sigma}_p^2\right)^{-2} \end{bmatrix}, \tag{F.61}$$

$$\mathbf{B}(p_t||q) = \begin{bmatrix} \left(\tilde{\sigma}_p^2\right)^{-1} & 0 \\ 0 & \left(\frac{\nu-1}{2(\nu-4)}\right)\left(\tilde{\sigma}_{p_t}^2\right)^{-2} \end{bmatrix}. \tag{F.62}$$

All terms are the same in both $\mathbf{A}$ and $\mathbf{B}$ for any other type of true distribution when considering the Gaussian model. The only term that varies so far is $B_{2,2}$, which depends on the true distribution. For the BGM, the last term should be derived separately. The expression for the fourth-order central moment can be simplified as a linear combination of the fourth-order moments for each of the modes

$$\begin{aligned} &\left(E_{p_{\mathrm{GM}}}\left[\left(z_i - E_{p_{\mathrm{GM}}}\left[z_i\right]\right)^4\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p} \\ &= (1-\varepsilon)\left(E_{g_1}\left[(z_i-\mu)^4\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} + \varepsilon\left(E_{g_2}\left[(z_i-\mu)^4\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}. \end{aligned} \tag{F.63}$$

The fourth-order moment for a Gaussian distribution is also already established, so for the distributions $g_1(z_i;\mu,\sigma^2)$ and $g_2(z_i;\mu,k\sigma^2)$ are:

$$E_{g_1}\left[(z_i-\mu)^4\right] = 3\sigma^4, \tag{F.64}$$

$$E_{g_2}\left[(z_i-\mu)^4\right] = 3k^2\sigma^4. \tag{F.65}$$

Leading to the resulting central moment for the BGM

$$\begin{aligned}&\left(E_{p_{GM}}\left[\left(z_i - E_{p_{GM}}\left[z_i\right]\right)^4\right]\right)_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_p}\\&= 3(1-\varepsilon)\sigma^4 + 3\varepsilon k^2\sigma^4 = 3\sigma^4((k^2-1)\varepsilon+1).\end{aligned} \tag{F.66}$$

The above is substituted into (F.56) and allows us to evaluate the last term of the corresponding matrix

$$B_{2,2}(p_{GM}||q) = -\frac{1}{4(\tilde{\sigma}^2_{p_{GM}})^2} + \frac{1}{4(\tilde{\sigma}^2_{p_{GM}})^4} 3\sigma^4((k^2-1)\varepsilon+1), \tag{F.67}$$

with significant simplifications after substituting the pseudo-true parameter for dispersion:

$$B_{2,2}(p_{GM}||q) = \frac{-\varepsilon^2(k-1)^2 + \varepsilon(k-1)(3k+1)+2}{4\sigma^4(\varepsilon(k-1)+1)^4}. \tag{F.68}$$

To reduce the size of the following expressions, the term $\phi(\varepsilon,k) = \varepsilon(k-1)$ is defined and the numerator of $B_{2,2}(p_{GM}||q)$ is denoted as the quadratic function of that term, i.e.,

$$Q(\phi(\varepsilon,k)) = -\varepsilon^2(k-1)^2 + \varepsilon(k-1)(3k+1)+2. \tag{F.69}$$

The matrix $\mathbf{B}(p_{GM}||q)$ is then written as:

$$\mathbf{B}(p_{GM}||q) = \begin{bmatrix} \left(\tilde{\sigma}^2_p\right)^{-1} & 0 \\ 0 & \frac{Q(\phi)}{4(\phi+1)^2}\left(\tilde{\sigma}^2_{p_{GM}}\right)^{-2} \end{bmatrix}. \tag{F.70}$$

F.3 MCRB computation

The final computation of the MCRB is a simple matrix multiplication, where the inverse of $\mathbf{A}$ is trivial

$$\mathbf{MCRB}_{\boldsymbol{\theta}}(p||q) = \frac{1}{N}\mathbf{A}^{-1}\mathbf{B}(p||q)\mathbf{A}^{-1}. \tag{F.71}$$

For the same misspecified assumption of a Gaussian distribution for the data, only the term $\mathbf{B}$ varies according to the true distribution. First the misspecified bound where the true data follows a Student's t-distribution is shown in (F.73) and the MCRB for the case where the true distribution is a BGM is then shown in (F.75).

$$\begin{aligned}&\mathbf{MCRB}_{\boldsymbol{\theta}}(p_t||q)\\&= \frac{1}{N}\begin{bmatrix} -\tilde{\sigma}^2_{p_t} & 0 \\ 0 & -2\left(\tilde{\sigma}^2_{p_t}\right)^2 \end{bmatrix}\begin{bmatrix} \left(\tilde{\sigma}^2_{p_t}\right)^{-1} & 0 \\ 0 & \left(\frac{\nu-1}{2(\nu-4)}\right)\left(\tilde{\sigma}^2_{p_t}\right)^{-2} \end{bmatrix}\begin{bmatrix} -\tilde{\sigma}^2_{p_t} & 0 \\ 0 & -2\left(\tilde{\sigma}^2_{p_t}\right)^2 \end{bmatrix},\end{aligned} \tag{F.72}$$

$$= \frac{1}{N}\begin{bmatrix} \tilde{\sigma}^2_{p_t} & 0 \\ 0 & \left(\frac{2(\nu-1)}{(\nu-4)}\right)\left(\tilde{\sigma}^2_{p_t}\right)^2 \end{bmatrix}, \tag{F.73}$$

$$\mathbf{MCRB}_{\boldsymbol{\theta}}(p_{GM}||q)$$

$$= \frac{1}{N}\begin{bmatrix} -\tilde{\sigma}^2_{p_{GM}} & 0 \\ 0 & -2\left(\tilde{\sigma}^2_{p_{GM}}\right)^2 \end{bmatrix}\begin{bmatrix} \left(\tilde{\sigma}^2_{p_{GM}}\right)^{-1} & 0 \\ 0 & \frac{Q(\phi)}{4(\phi+1)^2}\left(\tilde{\sigma}^2_{p_{GM}}\right)^{-2} \end{bmatrix}\begin{bmatrix} -\tilde{\sigma}^2_{p_{GM}} & 0 \\ 0 & -2\left(\tilde{\sigma}^2_{p_{GM}}\right)^2 \end{bmatrix}, \tag{F.74}$$

$$= \frac{1}{N}\begin{bmatrix} \tilde{\sigma}^2_{p_{GM}} & 0 \\ 0 & \frac{Q(\phi)}{(\phi+1)^2}\left(\tilde{\sigma}^2_{p_{GM}}\right)^{-2} \end{bmatrix}. \tag{F.75}$$

Bibliography

[1] B. Cecconi, M. Dekkali, C. Briand, B. Segret, J. N. Girard, A. Laurens, A. Lamy, D. Valat, M. Delpech, M. Bruno, P. Gelard, M. Bucher, Q. Nenon, J.-M. Griesmeier, A.-J. Boonstra, and M. Bentum, "NOIRE study report: Towards a low frequency radio interferometer in space," in *Proc. IEEE Aerospace Conference*, (Big Sky, MT, USA), pp. 1–19, Mar. 2018.

[2] N. Saks, A. Boonstra, R. T. Rajan, M. Bentum, F. Belien, and C. van't Klooster, "DARIS, a fleet of passive formation flying small satellites for low frequency radio astronomy," in *Small Satellite Systems and Services - The 4S Symposium 2010*, pp. 1–15, Jan. 2010.

[3] S. Engelen, C. Verhoeven, and M. Bentum, "OLFAR, a radio telescope based on nano satellites in moon orbit," in *Proc. 24th AIAA/USU Conference on Small Satellites*, (Logan, UT, USA), Utah State University, Oct. 2010.

[4] R. T. Rajan, S. Engelen, M. Bentum, and C. Verhoeven, "Orbiting low frequency array for radio astronomy," in *Proc. IEEE Aerospace Conference*, pp. 1–11, 2011.

[5] R. J. MacDowall, S. D. Bale, L. Demaio, N. Gopalswamy, D. L. Jones, M. L. Kaiser, J. C. Kasper, M. J. Reiner, and K. W. Weiler, "Solar Radio Imaging Array (SIRA): a multispacecraft mission," in *Proc. Enabling Sensor and Platform Technologies for Spaceborne Remote Sensing* (G. J. Komar, J. Wang, and T. Kimura, eds.), vol. 5659, pp. 284 – 292, International Society for Optics and Photonics, SPIE, 2005.

[6] N. Belsten, C. Payne, P. Serra, K. Cahoy, F. D. Lind, M. Knapp, P. J. Erickson, and F. Robey, "Design and verification of a clock system for orbital radio interferometry," in *Proc. 34th AIAA/USU Conference on Small Satellites*, (Logan, UT, USA), AIAA, 2020.

[7] S. Wu, W. Chen, Y. Zhang, W. A. Baan, and T. An, "SULFRO: a Swarm of Nano-/Micro-Satellite at SE L2 for Space Ultra-Low Frequency Radio Observatory," in *Proc. 28th AIAA/USU Conference on Small Satellites*, (Utah, USA), Utah State University, Apr. 2014.

[8] P. J. Napier, D. S. Bagri, B. G. Clark, A. E. Rogers, J. D. Romney, A. R. Thompson, and R. C. Walker, "The very long baseline array," *Proceedings of the IEEE*, vol. 82, no. 5, pp. 658–672, 1994.

[9] K. Kingham, "VLBI Basics: From Station through Correlator," in *Proc. 6th IVS Technical Operations Workshop*, (Westford, MA, USA), International VLBI Service, May 2011.

[10] H. Peters, B. Owings, T. Oakley, and L. Beno, "Hydrogen masers for radio astronomy," in *Proc. 41st Annual Frequency Control Symposium*, pp. 75–81, May 1987.

[11] E. Rubiola and F. Vernotte, "The Companion of Enrico's Chart for Phase Noise and Two-Sample Variances," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 7, pp. 2996–3025, 2023.

[12] W. J. Riley and D. A. Howe, "Handbook of frequency stability Analysis," 2008.

[13] "Resolution 1 of the 13th CGPM (1967) - BIPM." https://www.bipm.org/en/committees/cg/cgpm/13-1967/resolution-1. Accessed: 2022-01-05.

[14] "The International System of Units (SI Brochure)," 2019. https://www.bipm.org/en/publications/si-brochure. Accessed: 2024-04-26.

[15] "Resolution 1 of the 26th cgpm (2018) - BIPM." https://www.bipm.org/en/committees/cg/cgpm/26-2018/resolution-1. Accessed: 2022-01-05.

[16] J. Lodewyck, "On a definition of the SI second with a set of optical clock transitions," *Metrologia*, vol. 56, p. 055009, Sept. 2019.

[17] IEEE, "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.

[18] D. Allan, "Time and frequency (time-domain) characterization, estimation, and prediction of precision clocks and oscillators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 34, no. 6, pp. 647–654, 1987.

[19] Syrlinks, "EWOS™ : Low power & high precision OCXO," 2023.

[20] M. T. Inc., "Chip Scale Atomic Clock CSAC-SA45S," 2023.

[21] International Telecommunication Union Standardization Sector, "Timing characteristics of coherent network primary reference time clocks," Recommendation ITU-T G.8272.2, 2024.

[22] A. Dobrogowski and M. Kasznia, "Time effective methods of calculation of maximum time interval error," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 732–741, 2001.

[23] B. Alachkar, A. Wilkinson, and K. Grainge, "Frequency reference stability and coherence loss in radio astronomy interferometers application to the ska," *Journal of Astronomical Instrumentation*, vol. 07, no. 01, p. 1850001, 2018.

[24] A. E. E. Rogers and J. M. Moran, "Coherence limits for very-long-baseline interferometry," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-30, no. 4, pp. 283–286, 1981.

[25] R. T. Rajan, M. Bentum, and A. Boonstra, "Synchronization for space based ultra low frequency interferometry," in *Proc. IEEE Aerospace Conference*, (Big Sky, MT, USA), pp. 1–8, Mar. 2013.

[26] F. Lahaye, P. Collins, P. Heroux, M. Daniels, and J. Popelar, "Using the Canadian Active Control System (CACS) for Real-Time Monitoring of GPS Receiver External Frequency Standards," in *Proceedings of the 14th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2001)*, (Salt Lake City, UT, USA), pp. 2220 – 2228, Sept. 2001.

[27] Q. Ai, K. Maciuk, P. Lewińska, and L. Borowski, "Characteristics of onefold clocks of gps, galileo, beidou and glonass systems," *Sensors*, vol. 21, Mar. 2021.

[28] M. J. Underhill, "Fundamental phase jumps in oscillators?," in *Proceedings of the 20th European Frequency and Time Forum*, (Braunschweig, Germany), pp. 125–132, 2006.

[29] W. Krzewick, "Sub-microsecond holdover timing capabilities of the miniature atomic clock (MAC) rubidium oscillator," 2023. Presentation at 54th Annual Precise Time and Time Interval Systems and Applications Meeting, not published in proceedings.

[30] A. Novick, M. A. Lombardi, M. Demetrios, and J. Clark, "Measurement of Transient Environmental Effects in GPS-Disciplined Clocks," in *Proceedings of the 54th Annual Precise Time and Time Interval Systems and Applications Meeting*, (Long Beach, California, USA), pp. 10–14, Jan. 2023.

[31] W. Krzewick, "White paper chip-scale atomic clock (csac) performance during rapid temperature change," tech. rep., Microsemi: A Microchip company, Mar. 2020.

[32] P. J. Teunissen and O. Montenbruck, *Springer handbook of global navigation satellite systems*, vol. 10 of *Springer Handbooks*. Springer International Publishing, July 2017.

[33] R. T. Rajan and A.-J. van der Veen, "Joint ranging and synchronization for an anchorless network of mobile nodes," *IEEE Transactions on Signal Processing*, vol. 63, no. 8, pp. 1925–1940, 2015.

[34] B. Peterson, R. Hartnett, and G. Ottman, "GPS Receiver Structures for the Urban Canyon," in *Proc. of the 8th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1995)*, (Palm Springs, CA), pp. 1323–1332, Sept. 1995.

[35] F. Van Diggelen, *A-GPS: Assisted GPS, GNSS, and SBAS*. Artech House Gnss Technology and Applications Library, Artech House, 2009.

[36] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*. Ganga-Jamuna Press, 2011.

[37] F. van Diggelen, "GPS and GPS+GLONASS RTK," in *Proc. of the 10th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1997)*, (Kansas City, MO), pp. 139–144, Sept. 1997.

[38] K.-L. Noh, Q. M. Chaudhari, E. Serpedin, and B. W. Suter, "Novel clock phase offset and skew estimation using two-way timing message exchanges for wireless sensor networks," *IEEE Transactions on Communications*, vol. 55, no. 4, pp. 766–777, 2007.

[39] M. Leng and Y.-C. Wu, "On clock synchronization algorithms for wireless sensor networks under unknown delay," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 182–190, 2010.

[40] 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, Mar. 1993.

[41] 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, Feb. 2012.

[42] 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.

[43] 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.

[44] L. Galleani, "Time-frequency analysis of atomic clock anomalies," in *2018 European Frequency and Time Forum (EFTF)*, pp. 281–284, 2018.

[45] 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, July 2022.

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

[47] G. Petit, "Towards an optimal weighting scheme for TAI computation," *Metrologia*, vol. 40, pp. S252–S256, June 2003.

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

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

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

[51] S. Stein, "Time scales demystified," in *IEEE International Frequency Control Symposium and PDA Exhibition Jointly with the 17th European Frequency and Time Forum, 2003. Proceedings of the 2003*, pp. 223–227, 2003.

[52] 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, Jan. 2010.

[53] 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, Sept. 1991.

[54] C. Trainotti, T. D. Schmidt, and J. Furthner, "Simulating the realization of a mixed clock ensemble," in *Proc. of the Joint Conference of the IEEE International Frequency Control Symposium and European Frequency and Time Forum (EFTF/IFC)*, (Orlando, FL, USA), pp. 1–11, 2019.

[55] P. Roldan, S. Trilles, X. Serena, and A. Tajdine, "Novel Composite Clock Algorithm for the Generation of Galileo Robust Timescale," in *Proc. of the 35the International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+2022)*, (Denver, CO, USA), pp. 2790–2799, Sept. 2022.

[56] O. G. Crespillo, D. Medina, J. Skaloud, and M. Meurer, "Tightly coupled gnss/ins integration based on robust m-estimators," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, (Monterey, CA), pp. 1554–1561, 2018.

[57] C. Zucca and P. Tavella, "The clock model and its relationship with the allan and related variances," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 52, no. 2, pp. 289–296, 2005.

[58] J. A. Davis, C. A. Greenhall, and P. W. Stacey, "A kalman filter clock algorithm for use in the presence of flicker frequency modulation noise," *Metrologia*, vol. 42, pp. 1–10, Jan. 2005.

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

[60] H. Masnadi-Shirazi, A. Masnadi-Shirazi, and M.-A. Dastgheib, "A step by step mathematical derivation and tutorial on kalman filters," 2019.

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

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

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

[64] J.-A. Ting, E. Theodoru, and S. Schaal, "A Kalman Filter for Robust Outlier Detection," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, (San Diego, USA), pp. 1514–1519, Oct. 2007.

[65] A. Bellés, D. Medina, P. Chauchat, S. Labsir, and J. Vilà-Valls, "Robust m-type error-state kalman filters for attitude estimation," in *Proc.31st European Signal Processing Conference (EUSIPCO)*, (Helsinki, Finland), pp. 840–844, 2023.

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

[67] P. Tavella, J. Azoubib, and C. Thomas, "Study of the clock-ensemble correlation in algos using real data," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 34, no. 78, pp. 435–441, 1991.

[68] 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, Mar. 2000.

[69] 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, Sept. 2020.

[70] L. Ortega and S. Fortunati, "Misspecified time-delay and doppler estimation over non gaussian scenarios," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 9346–9350, 2024.

[71] S. Kalyani, "On CRB for parameter estimation in two component gaussian mixtures and the impact of misspecification," *IEEE Transactions on Communications*, vol. 60, no. 12, pp. 3734–3744, 2012.

[72] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.

[73] R. Piché, "Cramér-rao lower bound for linear filtering with t-distributed measurement noise," in *19th International Conference on Information Fusion*, (Heidelberg, Germany), 2016.

[74] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79 – 86, 1951.

[75] A. Mennad, S. Fortunati, M. N. El Korso, A. Younsi, A. M. Zoubir, and A. Renaux, "Slepian-bangs-type formulas and the related misspecified cramér-rao bounds for complex elliptically symmetric distributions," *Signal Processing*, vol. 142, pp. 320–329, 2018.

[76] S. Fortunati, F. Gini, and M. S. Greco, "The misspecified Cramér-Rao bound and its application to scatter matrix estimation in complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2387–2399, 2016.

[77] C. D. Richmond and L. L. Horowitz, "Parameter bounds on estimation accuracy under model misspecification," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2263–2278, 2015.

[78] Q. H. Vuong, *Cramér-Rao bounds for misspecified models.* working paper 652, Div. of the Humanities and Social Sci., Caltech, Pasadena, USA, 1986.

[79] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. Eighth IEEE International Conference on Data Mining*, (Pisa, Italy), pp. 413–422, 2008.

[80] 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, Feb. 2023.

[81] 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.

[82] F. Z. 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, no. 19, pp. 4751–4771, 2018.

[83] J. Kirkby, N. Dang, and D. Nguyen, "Moments of Student's t-distribution: A unified approach," Dec. 2019.