Comparisons of Robust Estimators for a Robust Time Scale in a Swarm of Satellites

Hamish McPhee
TéSA Research Laboratory
Toulouse, France
hamish.mcphee@tesa.prd.fr

Lorenzo Ortega
IPSA-TéSA
Toulouse, France
lorenzo.ortega@ipsa.fr

Jean-Yves Tourneret
University of Toulouse, IRIT-ENSEEIHT-TéSA
Toulouse, France
jean-yves.tourneret@toulouse-inp.fr

***Abstract*—This work provides a comparative study of the complexity and performance for a range of different types of robust estimators. The interest of this analysis is to find the preferred robust estimator that can define the system time for a swarm of satellites. The Student's t-distribution is used as a model for the noise corrupting the measurements. The ideal performance of an unbiased estimator for a fixed number of degrees of freedom is known in the form of the Cramér-Rao Bound (CRB). In this article, two examples of a robust M-estimator and an approximation of the Maximum Likelihood Estimator (MLE) resulting from an Expectation-Maximization algorithm are each tested with respect to the performance bounds. Each estimator is also compared with the Gaussian MLE under Gaussian noise, to identify any losses in efficiency under Gaussian conditions. The complexity of the algorithms is also studied by comparing the time until convergence in the iterative update of the robust estimators.**

***Index Terms*—Robust estimation, asymptotic performance, computational complexity**

I. Introduction

When dealing with parameter estimation from observations contaminated with outliers, several choices can be made in defining the estimators. These choices are important in many practical applications, including timing in satellite constellations, a case investigated in this work. For this application, the presence of outliers can cause a reduction in the precision of positioning solutions, which should be avoided. Robust estimators effectively provide a robust system time for the constellation that does not propagate these outliers. This work aims to express the trade-offs between different types of robust estimators in theoretical cases. In practice, readers can then make informed decisions about the most suitable method when estimating the parameters of any contaminated signal model, with a specific use case explored for robust system time.

The heavy tails of a Student's t-distribution assign a higher likelihood to outliers compared to standard normal situations without outliers [1], [2]. Modeling clock anomalies with this distribution in [3] has demonstrated a robust method of defining a time scale. This is relevant in the application of collaborative satellite observations, where the onboard clocks often face rapid fluctuations in their timing measurements caused by environmental factors [4], [5]. Whether combining observations made in a swarm of satellites for scientific reasons or for providing navigation solutions, significant changes in the satellite times can produce errors in the correlation of inter-satellite data [6], [7]. Recent studies [8], [9] on space-based interferometry for radio-astronomy have identified challenges in both localization and timing, which are critical for accurate interferometric image reconstruction. The solution requires the generation of a common ensemble time scale or 'system time' using only onboard timing data, since the satellites could potentially be operating autonomously and remotely. The use of robust estimators to define the system time is still a novel area of study, so the potential options of robust estimators are interesting for this application.

The Student's t-distribution has an asymptotic limit for the estimation error that depends on the number of degrees of freedom. The derived CRB for this distribution [10] establishes a benchmark for the best-case estimation performance. In recent works, the baseline of non-robust estimators has also been characterized using the Misspecified CRB (MCRB) [11], [12], providing insights into the effects of model mismatches on estimation accuracy. In the case that the signal noise is assumed to be Gaussian, while the noise is truly Student's t-distributed, the resulting Gaussian Misspecified MLE (MMLE) has been proven [13] to be asymptotically unbiased and its performance limited by the MCRB derived in [14]. Furthermore, this MCRB equals the Gaussian CRB for the same parameters, which exceeds the CRBs of well-specified models with real symmetric heavy-tailed distributions. It has also recently been shown that knowledge of the true distribution is not required to find estimators that improve upon the MMLE because the semiparametric CRB for the parameters of elliptically symmetric distributions equals the CRB of the parameters of the true distribution [15]. This suggests that a semiparametric robust estimator could perform as well as one based on the exact noise model. While identifying such an estimator is beyond this paper's scope, we argue that a class of estimators may enhance MMLE performance without requiring knowledge of the noise distribution. Here, we explore some of these estimators: i) the approximate MLE of the parameters of the Student's t-distribution determined by the Expectation-Maximization (EM) algorithm [16], [17], ii) an M-estimator tuned specifically for use with a Student's t-distribution [1], and iii) a more general M-estimator that just

This work supported by the TéSA Research Laboratory

aims at eliminating extreme values. The performance metrics used to compare the estimators are the efficiency in nominal operations, Mean Square Error (MSE) in the presence of outliers, and computational complexity. Section II introduces the robust estimators and their implementation. Section III discusses their use in defining a time scale, and Section IV evaluates their performance under nominal and Student's t noise conditions, before concluding with a preferred estimator and future directions.

II. Robust Estimators

As each estimator to be analyzed uses some form of iterative procedure, the initialization is kept constant for each estimator. To ensure that the estimator is initialized with a consistent estimate, the Gaussian MLE is used to initialize the estimate of the location and scale parameters. The Gaussian MLE initializing the iterative algorithms is defined by:

$$\hat{\mu}_0 = \frac{1}{N}\sum_{j=1}^{N} z_j, \;\; \hat{\sigma}_0^2 = \frac{1}{N}\sum_{j=1}^{N}(z_j - \hat{\mu}_0)^2, \tag{1}$$

where $z_1, \cdots, z_N$ are N independent and identically distributed random variables.

A. EM-Student's t-distribution

In theory, the MLE derived based on the probability distribution function (PDF) of the Student's t-distribution should provide an asymptotically efficient and optimal estimate. The PDF of the Student's t-distribution is:

$$p_T(z_j;\mu,\sigma^2,\nu) = \frac{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}}{\sqrt{\nu\pi\sigma^2}}\left(1+\frac{1}{\nu}\left(\frac{z_j-\mu}{\sigma}\right)^2\right)^{-\left(\frac{\nu+1}{2}\right)}. \tag{2}$$

Due to the form of the PDF, the MLE does not have a closed-form when the number of degrees of freedom is unknown. In the case of anomalous measurements, the shape parameter is not necessarily known. Hence, the estimation of the location parameter requires joint estimation of the scale matrix and the number of degrees of freedom. To converge to the MLE, an EM algorithm developed in [17] has been implemented for generation of a robust time scale in [3]. The EM iteratively estimates the latent variables u_j and w_j in Algorithm 1 that allow estimation of the parameters of the Student's t-distribution. Equation (3) is solved using another iterative Newton's method to converge to a solution for the estimated number of degrees of freedom $\hat{\nu}_k$. This estimation of the shape parameter of the underlying distribution is useful in remaining efficient in the nominal case, but at the expense of higher computational cost. For the EM algorithm above, an additional initialization is necessary, i.e., $\nu_0 = 100$ for the number of degrees of freedom. By initializing the number of degrees of freedom close to the value for Gaussian noise, the convergence speed of the EM algorithm in nominal conditions is improved. However, the initialization of the number of degrees of freedom must not be too large, otherwise the EM algorithm risks converging to the Gaussian MLE and not being robust. Alternatively, the number of degrees of freedom can be initialized at a low value $\nu_0 = 2.01$ (without entering the domain with undefined variance $\nu \leq 2$) to improve the convergence speed in the presence of outliers and guarantee robustness. The stopping rule $\epsilon \leq 1e-5$ for each iterative procedure is kept constant, focusing only on the estimate of the location parameter because that is the goal when generating a common time scale.

Algorithm 1 EM for Student's t-distribution

while $\epsilon > 1e-5$ **do**

k = k+1

$u_{j,k} = \frac{\hat{\nu}_{k-1}+1}{\hat{\nu}_{k-1}+\frac{(z_i-\hat{\mu}_{k-1})^2}{\hat{\sigma}^2_{k-1}}}$

$w_{j,k} = \psi\left(\frac{\hat{\nu}_{k-1}+1}{2}\right) - \log\left(\frac{1}{2}\left(\hat{\nu}_{k-1}+\frac{(z_i-\hat{\mu}_{k-1})^2}{\hat{\sigma}^2_{k-1}}\right)\right)$

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

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

$$N\phi\left(\frac{\hat{\nu}_k}{2}\right) + \sum_{i=1}^{N}\left[u_{i,k} - w_{i,k} - 1\right] = 0 \tag{3}$$

$\epsilon = \hat{\mu}_k - \hat{\mu}_{k-1}$

end while

B. M-estimators

The classical least squares estimator minimizes the sum of the squared residuals and is optimal in the nominal case of Gaussian noise.

$$\hat{\mu}_{\text{LS}} = \arg\min_{\mu}\sum_{i=1}^{N}(z_i-\mu)^2 = \frac{1}{N}\sum_{i=1}^{N} z_i. \tag{4}$$

A robust M-estimator can be implemented by using an Iteratively Reweighted Least Squares (IRLS) procedure. The IRLS minimizes a different loss function $\rho(z_i - \mu)$ that reduces the impact of outlying values

$$\hat{\mu}_M = \arg\min_{\mu}\sum_{i=1}^{N}\rho(z_i-\mu). \tag{5}$$

The sample median is a robust estimate of the location that is defined by the loss function $\rho(x) = |x|$ [1]. Other known loss functions provide a decreasing weight for outliers, with the Huber function and Tukey's bisquare function being the most commonly implemented [1]. The bisquare loss function is defined as follows:

$$\rho_b(x) = \begin{cases} 1-\left(1-\left(\frac{x}{b}\right)^2\right)^3, & \text{for } |x| \leq b, \\ 1, & \text{for } |x| > b, \end{cases} \tag{6}$$

where the threshold b is chosen to fix a certain level of performance in the nominal case. The value $b = 4.685$ provides 95% efficiency in the nominal case and is used in this work as a benchmark robust estimator [1]. The above loss function results in the following definition of weights, which are calculated using the derivative of the loss function [1]

$$W_b(x) = \begin{cases} \rho'(x)/x = \left(1 - \left(\frac{x}{b}\right)^2\right)^2, & \text{for } |x| \le b, \\ \rho''(0) = 0, & \text{for } |x| > b. \end{cases} \quad (7)$$

Multiplicative constants that do not affect the minimization of (5) are neglected in the weighting function. The bisquare loss function provides a redescending M-estimator because the derivative of the loss function (otherwise referred to as the influence function) tends to zero at infinity. This results in increased robustness to large outliers. A redescending influence function is also defined in [1] for the Student's t-distribution with shape parameter ν, which is usually unknown

$$\psi_\nu(x) = \rho'_\nu(x) = \frac{x}{x^2 + \nu}. \quad (8)$$

The IRLS algorithm that solves (5) with the above loss function requires the computation of the weighting function

$$W_\nu(x) = \begin{cases} \frac{1}{x^2+\nu}, & \text{for } x \neq 0, \\ \frac{1}{\nu}, & \text{for } x = 0. \end{cases} \quad (9)$$

Similarly to the M-estimators based on the bisquare loss function, the number of degrees of freedom ν is fixed to a value that provides robustness to outliers but does not lose too much efficiency under nominal noise conditions. For a loss in performance of no greater than 5% when the noise is Gaussian, the number of degrees of freedom can be fixed to $\nu = 12$ (empirically found). This is different from the EM algorithm in Section II-A that actively updates the number of degrees of freedom to be efficient in both nominal and anomalous cases. The resulting IRLS algorithm is summarized below

Algorithm 2 IRLS for M-estimator

$\nu = 12$
while $\epsilon > 1e-5$ **do**
$k = k + 1$
$\hat{\sigma}_k^2 = \text{MAD}(\mathbf{z} - \hat{\mu}_{k-1})$
$w_{j,k} = W\left(\frac{z_j - \hat{\mu}_{k-1}}{\hat{\sigma}_k}\right)$
$\hat{\mu}_k = \frac{\sum_{i=1}^N w_{i,k} z_i}{\sum_{i=1}^N w_{i,k}}$
$\epsilon = \hat{\mu}_k - \hat{\mu}_{k-1}$
end while

where $\text{MAD}(\mathbf{z} - \hat{\mu}_{k-1})$ is the Median Absolute Deviation for samples $\mathbf{z} = [z_1, \cdots, z_N]$ at the current snapshot, providing a robust estimate of the scale parameter. To simplify, the two different M-estimators will be referred to as the M_b-estimator and M_ν-estimator for the bisquare weighting function and the Student weighting function, respectively.

III. Time Scale Estimation

A time scale is constructed with a weighted average of clock prediction errors to act as a common reference or system time for a distributed network of clocks. The clock data is only observable by satellite i measuring time differences with all other satellites $j \neq i$, obtaining $x_{i,j}(t) = x_i(t) - x_j(t) + n_{i,j}(t)$ with measurement noise $n_{i,j}(t)$ for each inter-satellite link. The prediction errors are observed by comparing these measurements with the predicted values using information from a previous epoch $\hat{x}_j(t|t-\tau)$. Outliers in the clock data arise from two different sources: i) the onboard clocks change values due to environmental changes such as radiation or temperature and ii) the exchange of clock timing information is corrupted by noise on the inter-satellite communication links. As a result, either the onboard time deviates from the predicted value or the clock measurements provide incorrect information about the true time. The advantage of modeling the prediction errors with the Student's t-distribution is that both these sources of outliers are addressed simultaneously [3]. In addition, the time offset of satellite i with respect to the designed system time $x_{i,E}(t)$ is a parameter of the Student's t-distribution that models the measurements and prediction errors:

$$x_{i,j}(t) - \hat{x}_j(t|t-\tau) \sim T(x_{i,E}(t), \sigma^2(t), \nu(t)). \quad (10)$$

The same model can be assumed for each of the N satellites in the swarm by changing the reference satellite i for the clock measurements. The location parameter then corresponds to the time offset from the system time for the reference satellite. The above model is equivalent to a series of observations z_j following a Student's t-distribution where $\mu = x_{i,E}(t)$ is the location parameter and allows access to the system time. Omitting the dependence of time to represent a single snapshot, the observations of clock prediction error are $z_j \sim T(\mu, \sigma^2, \nu)$.

Each robust estimator provides an estimate $\hat{\mu}$, which should mitigate the impact of outliers in the noise model. In practice, the robust estimators provide a time scale that is not impacted by the independent sources of anomalies, allowing a reliable correlation of collaborative images or localization information. The MSE for the robust estimators should be better than the limit described by the MCRB. The reliability of the time scale is linked to the error of the estimate of the mean, a larger MSE would indicate that the prediction errors and link noises with outliers are contributing more to the time scale than the optimal case of only using nominal timing information.

IV. Estimator Analysis

For reproducibility of the results, the noise is generated with Student-t and Gaussian distributions, both with zero-mean and scale parameter $\sigma^2 = 1$. The number of degrees of freedom for the Student-t distribution is set to $\nu = 3$. The initial analysis evaluates the efficiency of robust estimators in the absence of outliers. Fig. 1 presents the Asymptotic Relative Efficiency (ARE), calculated empirically as the ratio of the MSE of the Gaussian MLE to the MSE of the robust estimators under nominal conditions, $\text{ARE} = \text{MSE}_{\hat{\mu}_G}/\text{MSE}_{\hat{\mu}}$. For brevity, the MLE for the parameters of the Student's t-distribution approximated by the EM algorithm is referred to as the EM estimator throughout the rest of this article. It is capable of adapting the appropriate parameter estimates, achieving near 100% ARE when the number of degrees of freedom is appropriately initialized. In contrast, the M_ν-estimator shows a tradeoff: lower efficiency with fewer degrees of freedom

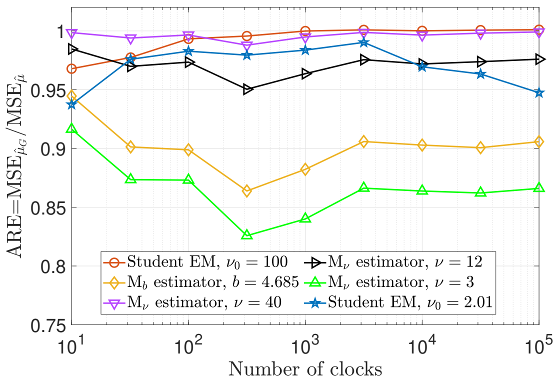

Fig. 1. Asymptotic Relative Efficiency (ARE) for the investigated robust estimators in the absence of anomalies. The closer the ARE to 1, the better the associated estimator performs under nominal conditions.

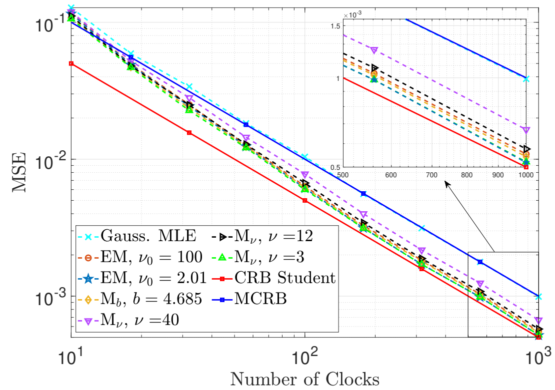

Fig. 2. Asymptotic Mean Square Error (MSE) for the different robust estimators in the presence of anomalies compared to the performance limits.

but improved ARE as ν increases, at the cost of reduced robustness. Proper tuning of ν can control efficiency loss. Note that same tradeoff applies to the M$_b$-estimator since the EM estimator is autonomously efficient. Note that this estimator is preferred for remote space operations where manual tuning can be impractical. However, other factors must be considered before making the final choice of estimator.

To assess the asymptotic performance of the proposed estimators with contaminated data, Fig. 2 shows the MSE for the location estimate as a function of the number of clocks. The number of Monte Carlo runs is set to 2000. Moreover, we show the theoretical performance corresponding to the CRB of the Student's t-distribution (red squares) and the MCRB (light-blue squares). Fig. 2 confirms that the MSE of the location for the EM estimator (orange circles) asymptotically approaches the CRB (red squares). The M$_\nu$-estimator with the correct degrees of freedom $\nu = 3$ (green triangles) effectively mitigates outliers and converges to the CRB. However, as shown in Fig. 1 the M-estimator with correctly specified numbers of degrees of freedom does not remain efficient under nominal conditions. In contrast, the EM estimator dynamically estimates the number of degrees of freedom, maintaining both efficiency in the nominal case and robustness to outliers.

Fig. 2 also shows that the MSE of the Gaussian MLE (blue crosses) converges to the MCRB. The other robust estimators are not optimal but still outperform the Gaussian MLE. Setting the number of degrees of freedom to $\nu = 12$ offers a reasonable balance, providing an acceptable level of efficiency in the nominal case (see Fig. 1) while maintaining similar performance as the EM estimator initialized with $\nu_0 = 100$. The M$_b$-estimator achieves slightly better performance in the robust case (yellow diamonds compared to black triangles and orange circles in Fig. 2). This confirms that it is not necessary to assume the exact noise model to obtain a robust estimator. However, the ARE remains worse than the appropriately tuned M$_\nu$-estimator so is less preferable. The case of fixing the number of degrees of freedom to a high value for better nominal performance shows significant losses (purple triangles) for the M$_\nu$-estimator compared to other robust methods.

The EM-based estimator is robust to the anomalies while maintaining nominal performance and achieves nearly optimal performance when the number of degrees of freedom is properly initialized. Thus, for a balance of nominal efficiency and robustness, the EM algorithm performs best, but only marginally. Nevertheless, using an EM algorithm comes with a cost in computational complexity before converging to the best estimate.

Fig. 3 shows the average time before convergence for each of the iterative algorithms when the noise is Student's t distributed. The simulations were conducted on an Intel(R) Core(TM) i9-10980XE CPU, using MATLAB R2022a. The M$_\nu$-estimator takes the least amount of time because the difficult-to-estimate shape parameter is fixed. Fig. 3 shows that the M$_b$-estimator achieves a similar convergence time to the case with the exact number of degrees of freedom. However, this estimator is still slower than the M$_\nu$-estimator with the tuned number of degrees of freedom while obtaining the same MSE performance. Fig. 3 also demonstrates the difference in initializing the shape parameter at either a small or high value for the EM estimator. There is an improvement when the initial number of degrees of freedom is assumed to be low.

Fig. 4 shows a larger increase in computation time when initializing at low values of the number of degrees of freedom in nominal data. This justifies the choice of initializing the EM with the Gaussian estimates to remain somewhat competitive with the computation time of the M-estimators. Additionally, Fig. 4 shows that both types of M-estimator converge faster than the EM solution. This includes the M$_b$-estimator that does not specify the exact noise distribution but remains robust and fast. Nevertheless, the M$_\nu$-estimator can provide further improvement in computation time while remaining robust and having a better ARE than the M$_b$-estimator.

The computation time for each estimator can have different values depending on the level of precision chosen in the stopping rule. It is assumed that for the same defined level of precision in each of the algorithms, the relative perfor-

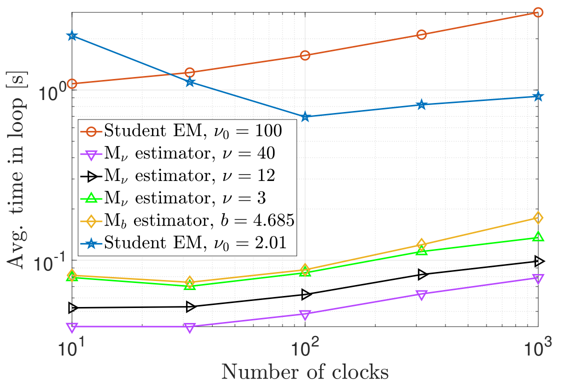

Fig. 3. Average time spent in the iterative loop for each of the robust estimators (Student-t noise).

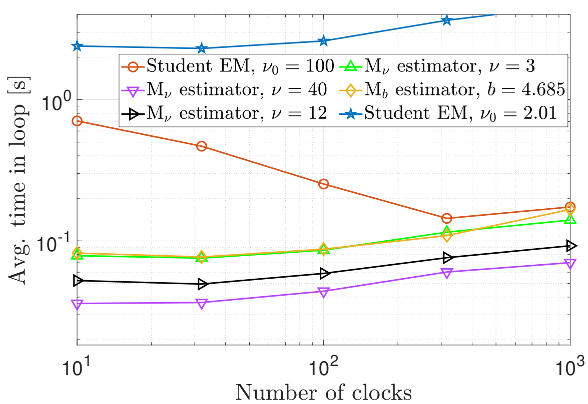

Fig. 4. Average time spent in the iterative loop for each of the robust estimators (Gaussian noise).

mance should have a similar magnitude. The figures presented above highlight the advantages and trade-offs for each of the investigated estimators. The M_ν-estimator provides the fastest convergence but requires a specific tuning to achieve the desired ARE. The EM estimator reliably and automatically obtains good ARE as well as good robustness to outliers at the cost of increased computation time.

V. Conclusion

A weighted average helps define a robust time scale in a swarm of satellites. The method of computing the weights depends on the type of robust estimation that is chosen. Hence, several robust estimators have been analyzed to help choose one that remains robust and efficient while not being computationally expensive. Small-scale satellites operating autonomously in distant orbits might have limited computational capacities. An appropriately tuned M_ν-estimator based on the loss function for a Student's t-distribution is shown to provide a low-cost and robust solution. Nevertheless, the more computationally expensive EM algorithms are more autonomous, being able to adapt to nominal conditions with improved efficiency compared to the M_ν-estimator. Using the more commonly implemented M_b-estimator based on the bisquare loss function was also shown to provide a good balance of speed and robust estimation performance, although not as good as the tuned M_ν-estimator. Since the EM estimator is both robust and efficient in the nominal case, it is recommended for use in applications that have sufficient computation budgets. Otherwise, the tuned M_ν-estimator is preferred for computationally limited applications with a permitted maximum loss in nominal efficiency. The performance of the M_b-estimator shows that robust estimators do not necessarily rely on the true definition of the underlying distribution of the anomalous measurements. Future work can investigate other types of robust EM estimators with improved speed, or other fast and robust estimators that do not consider the exact noise model.

References

[1] 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.

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

[3] H. McPhee, J.-Y. Tourneret, D. Valat, J. Delporte, Y. Grégoire, *et al.*, "A robust time scale for space applications using the Student's t-distribution," *Metrologia*, vol. 61, p. 055010, Sep. 2024.

[4] W. Krzewick, "White paper chip-scale atomic clock (CSAC) performance during rapid temperature change," tech. rep., Microsemi: A Microchip company, Mar. 2020.

[5] 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.

[6] P. J. Napier, D. S. Bagri, B. G. Clark, A. E. Rogers, J. D. Romney, *et al.*, "The very long baseline array," *Proceedings of the IEEE*, vol. 82, no. 5, pp. 658–672, 1994.

[7] F. Abel, P. Sundaramoorthy, and R. T. Rajan, "Frequency Augmented Clock Synchronization for Space-based interferometry," in *Proc. 4S Symposium*, (Vilamoura, Portugal), May 2022.

[8] B. Cecconi, M. Dekkali, C. Briand, B. Segret, J. N. Girard, *et al.*, "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.

[9] R. T. Rajan, S. Engelen, M. Bentum, and C. Verhoeven, "Orbiting low frequency array for radio astronomy," in *Proc. IEEE Aerospace Conference*, (Big Sky, MT, USA), pp. 1–11, 2011.

[10] R. Piché, "Cramér-Rao lower bound for linear filtering with t-distributed measurement noise," in *Proc. Int. Conf. Information Fusion*, (Heidelberg, Germany), 2016.

[11] C. D. Richmond and L. L. Horowitz, "Parameter bounds on estimation accuracy under model misspecification," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2263–2278, 2015.

[12] 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 Trans. Signal Process.*, vol. 64, no. 9, pp. 2387–2399, 2016.

[13] S. Fortunati and L. Ortega, "On the efficiency of misspecified Gaussian inference in nonlinear regression: Application to time-delay and doppler estimation," *Signal Process.*, vol. 225, p. 109614, 2024.

[14] H. McPhee, J.-Y. Tourneret, D. Valat, J. Delporte, Y. Grégoire, *et al.*, "Misspecified Cramér-Rao bounds for anomalous clock data in satellite constellations," in *Proc. European Signal Processing Conference (EUSIPCO)*, (Lyon, France), pp. 1222–1226, 2024.

[15] L. Ortega and S. Fortunati, "Robust semiparametric efficient estimator for time delay and doppler estimation," *Submitted to IEEE Signal Processing Letters*, 2025.

[16] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, 1996.

[17] 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.