Evaluation of time and frequency data: anticipating failures in frequency patterns that make up an atomic time scale.

B J Jordão[1], D V Magalhães[2], P S Rocha[1] and R J Carvalho[1]

[1] Observatório Nacional, Divisão de Serviços da Hora Legal Brasileira - Laboratório Primário em Tempo e Frequência, Rio de Janeiro, 20921-400, Brazil.
[2] Instituto de Fisica de São Paulo – USP, Departamento de Física e Ciência dos Materiais – Grupo de Óptica, São Paulo, 13560-970, Brazil.

E-mail: bjordan@on.br

Abstract. Anomaly detection in commercial cesium standards is crucial due to their accuracy as these standards are used on atomic time scales and in various technological applications. Anomalies or instability in cesium standards can have a negative impact on systems such as satellite navigation, military communication, network synchronization, and on time scales in many laboratories. The cesium standard is widely used as an atomic time standard due to its stability and precision, requiring continuous monitoring of its parameters to ensure signal integrity and avoid significant deviations. This study uses univariate and multivariate statistical control charts to monitor and detect anomalies in the cesium standards used in the Brazilian atomic time scale. These control charts use Hotelling's T^2 statistical test to detect anomalies and identify their onset before instability occurs in the cesium standard output signal. In addition, an exponentially weighted moving average (EWMA) plot is used to individually evaluate each parameter in the multivariate analysis of the control chart. The study confirms the abnormal behavior of the cesium standard at a specific point and successfully identifies the exact moment of the anomaly using Hotelling's T^2 and EWMA plots, before a significant stability change in the pattern occurs.

1. Introduction

Primary laboratories use atomic clocks to generate an independent atomic time scale [1]. A time scale is a system that unequivocally assigns a temporal coordinate, called a date, to any event [2].

The precision of the cesium standards allowed the development of the current time scale, facilitating the understanding of the relativistic effect [3].

Coordinated Universal Time (UTC), which serves as the international reference time scale, is computed by the BIPM (Bureau International des Poids et Mesures) using data from various atomic clocks maintained in more than 80 institutions [4]. It is essential that these watches are highly stable, accurate and reliable. The commercialization of cesium standards allowed countries to create their own time scales, which serve as a stable reference for calibration and construction of atomic time scales [5]

While it is possible to create a timescale with a single clock, it is not recommended due to the risk of compromise if a failure occurs [6]. It is essential to use several clocks to compose a time scale, whose data are used in a calculation by an algorithm to obtain a weighted average based on the weights of each clock. The result of this calculation is used to correct the accuracy of the signal generated by the scale [7].

In Brazil, the Brazilian Legal Hour Services Division (DISHO) of the National Observatory (ON) is responsible for the atomic time scale. The Brazilian Atomic Time Scale (ETAB1(ONRJ)), used by ON, employs commercial cesium beam and hydrogen maser standards for its realization. Since 1989, UTC(ONRJ) has been compared to UTC(BIPM) using a GPS receiver, and the results are sent to BIPM [8]. Figure 1 shows the block diagram of the ON Brazilian Atomic Time Scale.

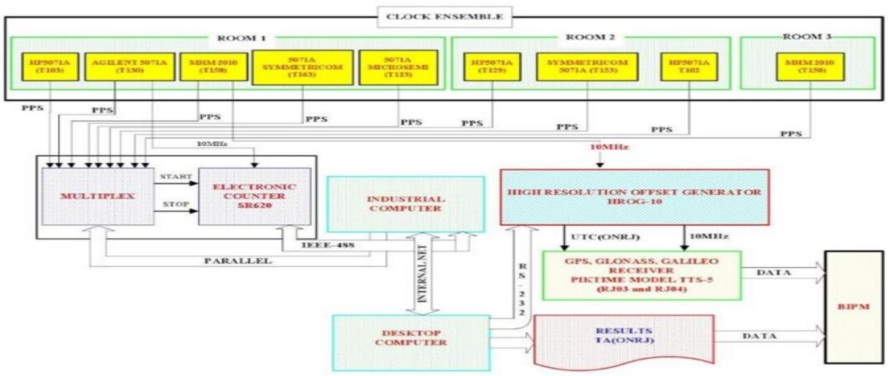

Figure 1. Block diagram (updated) of the Brazilian atomic time scale system of the National Observatory - ON.

Techniques are applied to detect anomalies in cesium patterns in order to keep a country's timescale in full working order. These devices are widely used in defining UTC and are maintained by experts in atomic time scales, ensuring the accuracy and stability of the time reference.

1.1. General objectives

The stability of cesium standards is essential for the accuracy and reliability of time measurement devices. In the long term, the commercial cesium standard has a stability of approximately 4x10E-12 [9]. If this stability is not guaranteed, time measurements can become inaccurate, affecting areas such as telecommunications and navigation, which depend on accurate synchronization. The Allan Deviation is a widely used statistic to estimate frequency stability, allowing to identify changes in the frequency of an oscillator over time [10]. Significant variations in the stability of cesium standards can result in anomalous behavior in the phase difference between the PPS signals of the standards, as illustrated in figure 2. These variations are due to anomalies in the parameters.

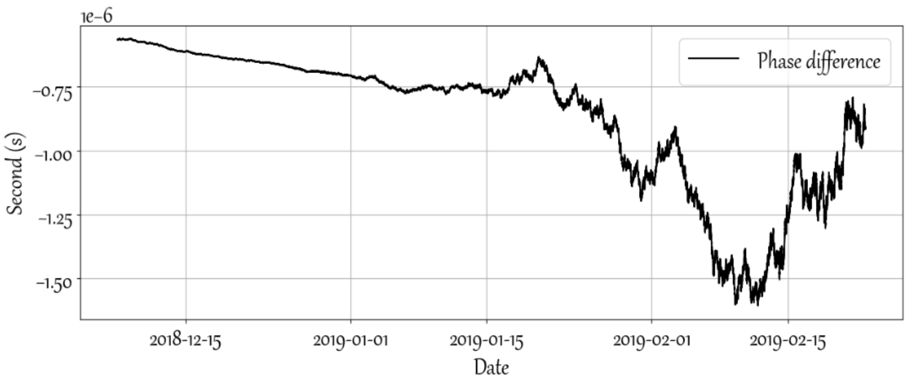

Figure 2. Anomalous behavior of the phase difference between the PPS signal of the reference standard and the PPS signal of the standard under analysis. One can verify the significant variation due to the change in stability.

In this project, the objective is to detect anomalies in the cesium standard parameters using data science techniques (PCA, CEP, EWMA, and Hotelling's T^2) to anticipate corrective actions and preserve the stability of the output signals, ensuring the integrity of the timescale atomic.

1.2. Specific objectives

- Use multivariate PCA techniques to reduce the study variables, preserving information;

Create univariate EWMA and multivariate Hotelling's T^2 control charts to verify statistical control of standard parameters;

- Identifying the onset of anomalies through EWMA or Hotelling's T^2 control charts;
- Estimate the time between the detection of the anomaly and the compromise of the stability of the pattern;
- Observe the behavior of the Allan Deviation before and after the anomaly detection to verify the stability of the frequency.

2. Literature Review

2.1. Related Work

Hewlett-Packard implemented technical improvements in its cesium standards, allowing remote monitoring of its parameters for analysis [11]. This improvement facilitated a study carried out at the United States Naval Observatory (USNO), where differences in the values of certain parameters between different standards were found, however, these differences do not indicate anomalies. Chadsey and Kubik [11] identified seven parameters (E-multiplier, Ion Pump, Osc. Control, Signal Gain, Thermometer, RF amplitude 1 and RF amplitude 2) as important to predict the performance and useful life of the patterns.

2.2. Principal Component Analysis – PCA

Principal Component Analysis (PCA) is a statistical technique that reduces the dimensionality of data, capturing its variability in uncorrelated principal components, facilitating the interpretation and visualization of results [12]. It is widely used in many areas to explore patterns and make informed decisions. It is a powerful tool for exploratory data analysis and decision making in many areas, from scientific studies to practical applications in business and engineering.

2.3. Hotelling's T^2 Multivariate Control Chart

In 1947, Hotelling pioneered control chart techniques to track multiple quality characteristics simultaneously [13].

Hotelling's T^2 control chart is a technique widely used in industry to monitor the variance of a multivariate population by detecting anomalous data points based on measures of variance and covariance. However, it's important to perform additional testing and analysis to confirm issues and determine their source, according to Montgomery [14].

The use of control charts involves the application of phases I and II to establish control limits and monitor the process. In Phase I, data is analyzed to verify process control, while in Phase II, they are used to monitor the process, comparing sample statistics to control limits. The Hotelling statistic is a measure that combines the estimated mean vector $\overline{\overline{X}}$ and the estimated covariance matrix (S) of the $\overline{\overline{X}}$ vector under control of the process mean vector [15] according to equation (1).

$$T^2=n\left(\overline{X}-\overline{\overline{X}}\right)^t S^{-1}\left(\overline{X}-\overline{\overline{X}}\right) \tag{1}$$

For Phase I, calculations of UCL and LCL control limits are given by

$$UCL= \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \tag{2}$$

$$LCL=0 \tag{3}$$

Where $F_{\alpha,p,mn-m-p+1}$ is the percentile of the F distribution with p and (mn-p+1) degrees of freedom.

For Phase II, the calculations of the new UCL and LCL limits are established only to monitor future observations [16] and the control limits are given by

$$UCL= \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \tag{4}$$

$$LCL=0 \tag{5}$$

2.4. Exponentially Weighted Moving Average Control Chart - EWMA

The exponentially weighted moving average (EWMA) is a process monitoring technique that uses exponential weights to detect small changes in a single variable [17]. Compared to the Shewhart control chart, EWMA has similar performance but is easier to configure and operate. It can be applied to individual observations or subgroups of size $n > 1$ [14] and is defined by equation (6).

$$z_i= \lambda x_i+(1-\lambda)z_{i-1} \tag{6}$$

Where $0 < \lambda \leq 1$ is a constant and, as the first value, which is used in the first sample, we use

$$z_0= \mu_0$$

Sometimes the mean of the data is used as the initial value of the EWMA, so that $z_0 = \overline{x}$

The z_i of the EWMA is a weighted average of all previous sample means, which can be evidenced by substituting z_{i-1} into the right hand side of equation (6) as shown below

$$\begin{aligned} z_i &= \lambda x_i+(1-\lambda)[\lambda x_{i-1}+(1-\lambda)z_{i-2}] \\ &= \lambda x_i+\lambda(1-\lambda)x_{i-1}+(1-\lambda)^2 z_{i-2} \end{aligned}$$

Substituting recursively for z_{i-j}, j=2,3,…,t, we have

$$z_i= \lambda \sum_{j=0}^{i-1} (1-\lambda)^j \ x_{i-j}+ (1-\lambda)^i \ z_0 \tag{7}$$

As the sample mean age increases, the weights $\lambda(1-\lambda)^j$ progressively decrease. The sum of all weights $\lambda(1-\lambda)^j$ equals 1, as each weight represents a fraction of the total sample

$$\lambda\sum_{j=0}^{i-1}(1-\lambda)^j = \lambda\left[\frac{1-(1-\lambda)^i}{1-(1-\lambda)}\right]=1-(1-\lambda)^i$$

The variance of z_i, when x_i are independent random variables with variance σ^2, according to equation (8):

$$\sigma_{zi}^2= \sigma^2\left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2i}\right] \tag{8}$$

The EWMA control chart is constructed by plotting z_i against sample number (or time). The center line and control limits for this plot are given by equations (9) and (10).

$$\text{UCL}= \mu_0+L\sigma\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]} \tag{9}$$

Center line = μ_0

$$\text{UCL}= \mu_0-L\sigma\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]} \tag{10}$$

where L is the width of the control limits (“Limits of L-sigma”) and σ is the standard deviation when the process is in control.

Hotelling's T^2 control chart and EWMA control chart are tools used to monitor statistical processes. These tools were used in Phase I of this research to monitor the cesium standard parameters. Both control charts were successfully used in Phase I to monitor parameters. These tools were useful to detect anomalies and to verify that the processes are within the established limits.

3. Material end method

30 files were used (58460.txt - 12/08/2018 to 58490.txt - 01/07/2019) containing records collected from minute to minute. 22 files (58460.txt - 12/08/2018 to 58481.txt - 12/29/2018) were used in phase I and 9 files (58482.txt - 12/30/2018 to 58490.txt - 01/07/ 2019), in phase II, respectively. A python algorithm was developed to concatenate the files used in phase I and phase II and transform them into a dataframes. Missing data were replaced using interpolation techniques. The two dataframes contained the standard deviation of the parameters. Each standard deviation value was calculated according to equation (11).

$$s_j = \sqrt{\frac{\sum_{i=j}^{(n-1)+j}(x_i - \bar{x})^2}{n-1}} \tag{11}$$

where n is the window size for calculation, in this case, $n = 60$. The j is the displacement one step ahead ($j = 0,1,2,3,\ldots,m$), where m is the total value of register in dataframe.in this case.

The first dataframe was used to build the multivariate control chart with Hotelling's T^2 and to build the univariate EWMA control charts for the four parameters of the cesium standard, considered Phase I. To determine this Phase, past data were used where the cesium standard presented parameters “under control”. In Hotelling's T^2 plot, the dataframe dataset was submitted to PCA to decompose the matrix of variances and covariances of the data into principal components. The second dataframe was used to build the future data graphs in phase II.

3.1. Selection of cesium standard parameters
The parameters selected for the analysis were Osc. Control, RF-Amplitude 1, RF-Amplitude 2, and E-multiplier. These parameters were chosen based on their temporal evaluation, as it was noticed that they exhibited significant variations in value practically at the same moment, as shown in figure 3. These variations were not observed in the other parameters or occurred at much later times.

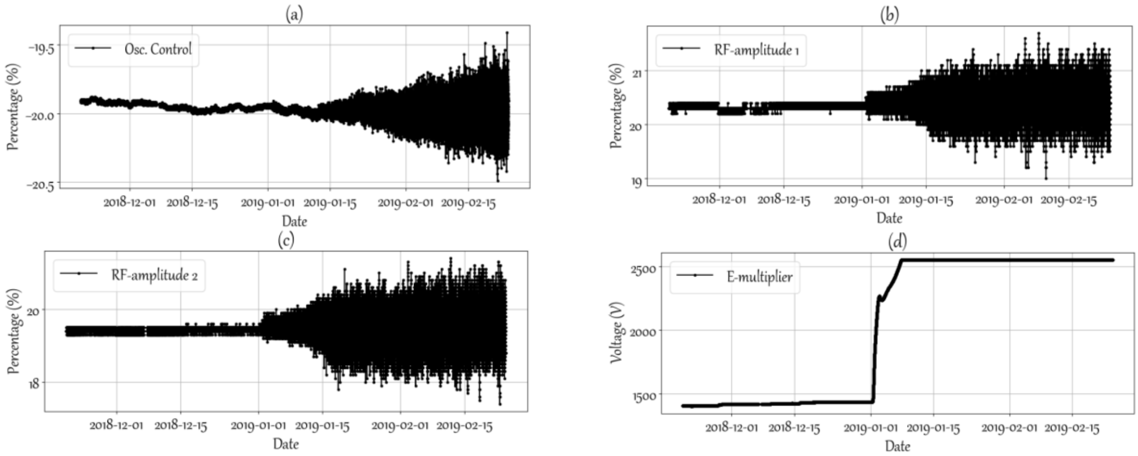

Figure 3. Parameters chosen to carry out the study. The parameters Osc_Control (a), RF_amplitude_1 (b), RF_amplitude_2 (c) and E_multiplier (d) before January 2019 showed normal behavior. After January, he began to show abnormal behavior.

3.2. *Building the control charts (Phase I)*
Data were split into two dataframes after treatment and removal of missing values. In the first dataframe, multivariate control charts were constructed with Hotelling's T^2 and univariate EWMA control charts. The PCA applied to this dataframe data showed that three principal components explain 91.4% of the variability, as shown in figure 4.

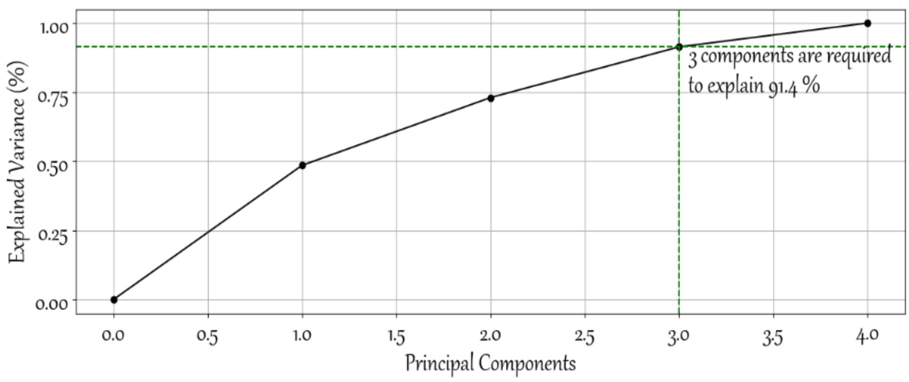

Figure 4. Graph the number of principal components needed to explain the variance of the data.

3.2.1. T^2 Hotelling Charter

Using the dataframe with the 3 principal components, the graph in figure 5 was plotted with 95.0% confidence limits (T^2 = 7.81) and 99.0% confidence limits (T^2 = 11.3).

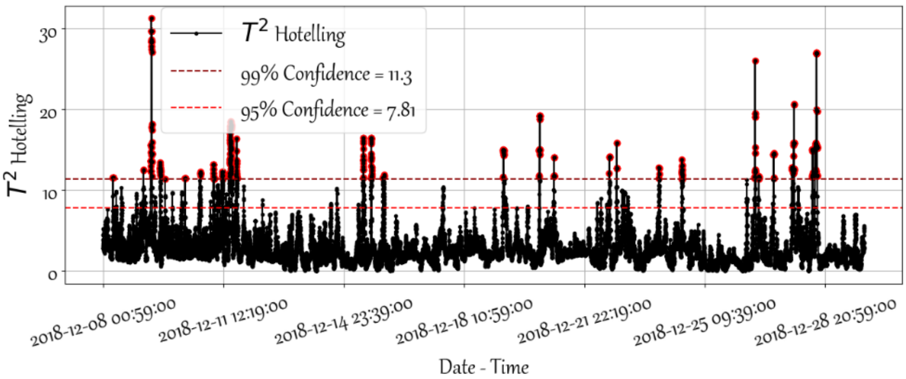

Figure 5. Hotelling's T^2 control chart for Phase I. Obtaining 95% and 99% confidence limits.

Although some points were outside the control limits above the 99% confidence level, the statistical control remained normal, which was expected based on the parameter data.

3.2.2. EWMA Charter

For each parameter of the cesium standard, an EWMA control chart was constructed, figure 6. For all cesium standard parameters, the following values were assigned to the function parameters: Lambda = 0.05, L = 2.7 and Sigma = 1.

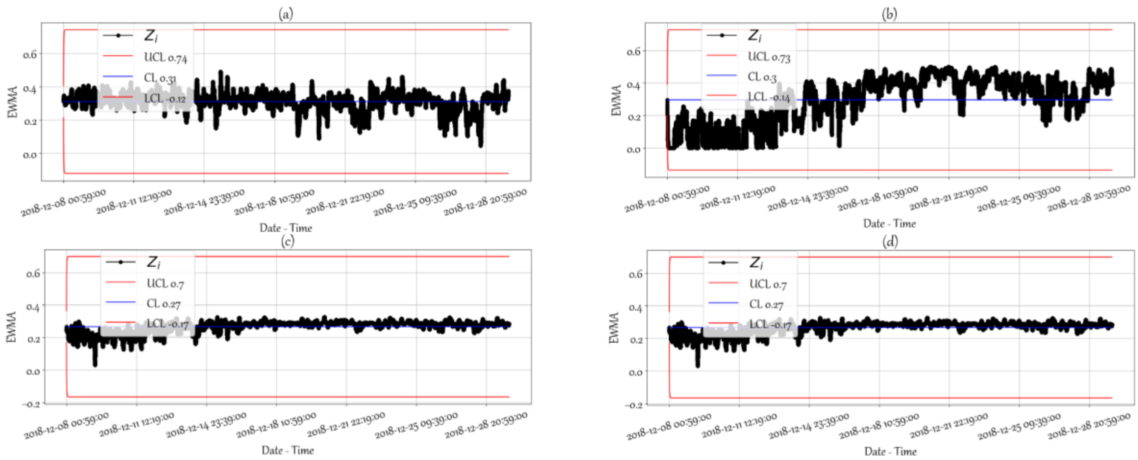

Figure 6. Control charts for Osc parameters. Control (a), RF-amplitude 1 (b), RF-amplitude 2 (c) and E-multiplier (d) using Exponential Weighted Moving Average - EWMA with its control limits.

Table 1 shows the calculated control limits when parameter data were submitted to control charts - EWMA.

Table 1. Lower control limits (LCL), center line (CL) and upper control limits (UCL) using Exponential Weighted Moving Average - EWMA for each parameter of the cesium standard.

	Lower control limits (LCL)	Center line (CL)	Upper control limits (UCL)
Osc. Control (a)	-0.12	0.31	0.74
RF-amplitude 1 (b)	-0.14	0.30	0.73
RF-amplitude 2 (c)	-0.17	0.27	0.70
E-multiplier (d)	-0.42	0.02	0.45

The Hotelling and EWMA T^2 control charts are tools used to monitor statistical processes and were used in Phase I to monitor the parameters of the cesium standard when it behaved normally.

The 95% and 99% confidence limits of the Hotelling T^2 control chart and the UCL, CL and LCL control limits of the EWMA control charts obtained in Phase I will be used in Phase II to observe the statistical control of the parameters of the cesium standard for future data.

Hotelling's T^2 control chart was used to monitor the variability of data from the PCA. The EWMA control chart was used to monitor the individual stability of each parameter.

These tools were useful to detect anomalies in the process and to verify if the processes are within the established limits.

4. Results

4.1. Anomaly Detection - Hotelling's T^2 Control Chart (Phase II)

In Phase II of the control chart analysis, a second dataframe consisting of 12,960 records from 9 files was used. These files contained future data ranging from 30/12/2018 to 07/01/2019. The analysis also incorporated the standard deviation of the data.

During the analysis, an anomaly was detected at 11:17 AM on January 1, 2019, as indicated in figure 7 Following this point, there was a noticeable escalation in process instability, with a significant increase in the values of each subsequent data point.

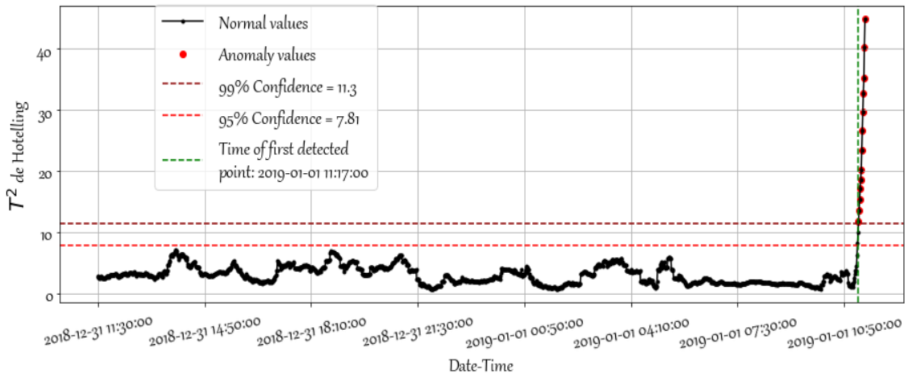

Figure 7. Hotelling's T^2 Control Chart identifying anomaly data. The first abnormal data point was detected at 11:17 AM on January 1, 2019.

The anomaly ellipse was utilized for data visualization, as depicted in figure 8. This ellipse was constructed using the second and third principal components.

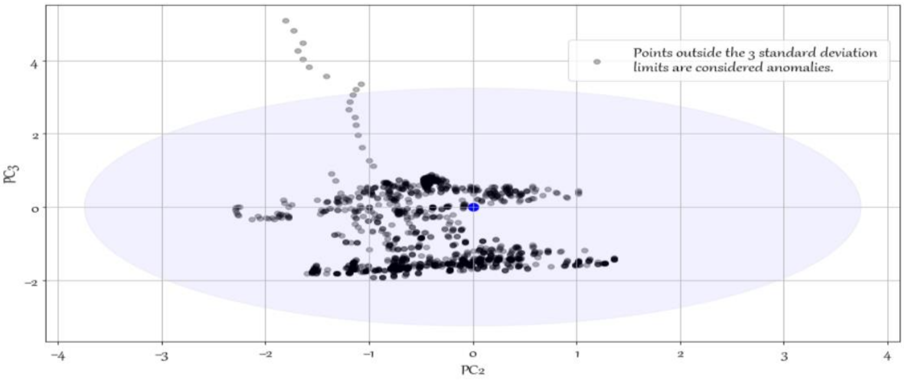

Figure 8. The ellipse plot shows that some data points are outside the three standard deviation limits. These data points are considered anomalies.

In the same time window as figure 7, the data for the parameters Osc. control, RF-amplitude 1, RF-amplitude 2, and E-multiplier were subjected to an Exponentially Weighted Moving Average (EWMA) control chart, as shown in figures 9a, 9b, 9c, and 9d. This was done to observe which parameter changed first and significantly influenced the Hotelling's T^2 control chart.

The purpose of this analysis was to determine which specific parameter exhibited an early and notable deviation from the expected values, thus impacting the overall performance of the Hotelling's T^2 control chart.

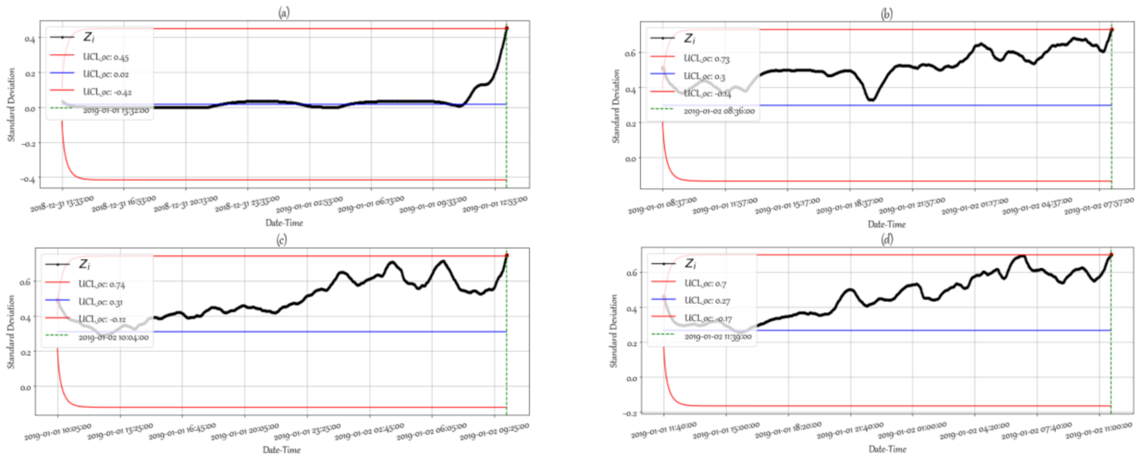

Figure 9. In the EWMA control chart, the moments of the first anomaly detections in the parameters were identified. The first parameter, E-multiplier, exhibited an anomaly at 1:32 PM on January 1, 2019 (a). Next, the RF-amplitude 1 parameter had an anomaly detection at 8:36 AM on January 2, 2019 (b). On the same day, at 10:04 AM, an anomaly was detected in the Osc. control parameter (c). Finally, at 11:39 AM on January 2, 2019, the first anomaly detection in the RF-amplitude 2 parameter occurred (d).

4.2. Verifying the stability of the cesium standard after detecting the Hotelling's T^2 control chart.
When analyzing the phase difference graph between PPS signals, it was observed that the control charts detected anomalies at specific moments. Interestingly, these detections occurred prior to significant variations in the phase difference values between the PPS signals. This finding is evident in the graph displayed in figure 10.

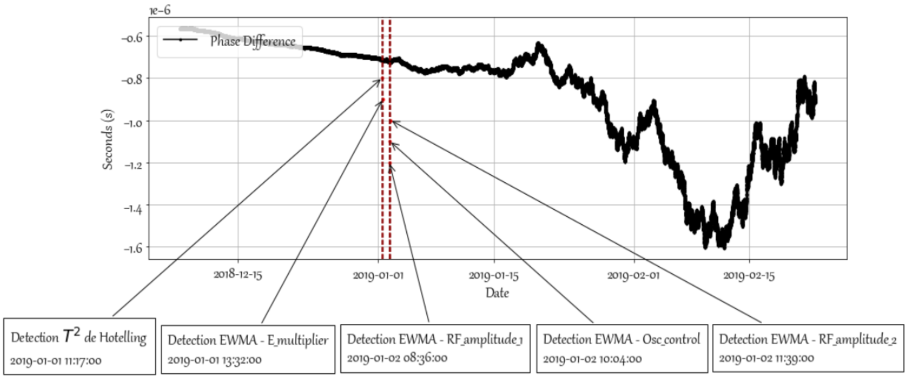

Figure 10. Graph of the phase difference between the PPS signal of the reference standard and the PPS signal of the standard under analysis. Time instances in which each parameter exhibited anomalies in its data.

After identifying the moment of the first detection using the Hotelling's T^2 control chart, the Allan deviation values were examined before and after the detection. The Allan deviation values remained below σ=2e-11 in the days preceding and following the detection, which occurred at 11:17 AM on January 01, 2019. However, starting from the 18th day, specifically on January 19, 2019, the Allan deviation values began to increase significantly, as shown in the red area of figure 11.

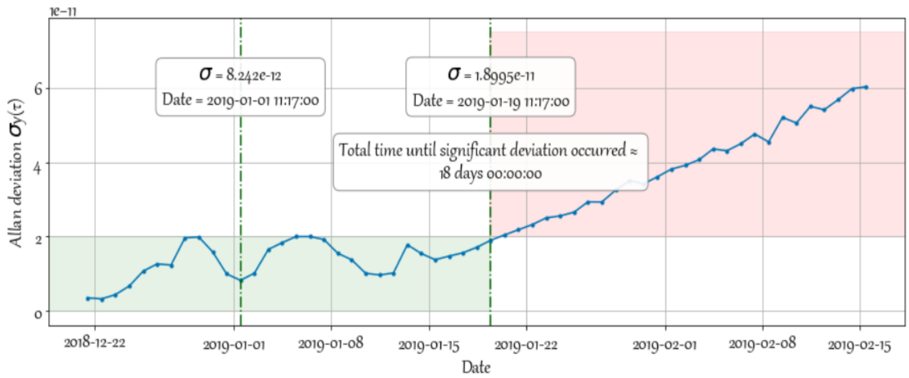

Figure 11. Frequency stability graph displaying only the Allan deviation values for tau = 60 seconds. In the green region, the Allan deviation is under control. In the red region, the Allan deviation is out of control.

5. Conclusion

This work demonstrated the use of multivariate control charts (T^2 Hotelling) and univariate control charts (EWMA) for anomaly detection in atomic cesium standards. The charts were able to identify anomalies before the stability of the standard was affected, emphasizing the importance of precise and efficient monitoring.

The T^2 Hotelling chart detected the first anomaly at 11:17 AM on 01/01/2019. The EWMA charts for individual parameters also detected anomalies, with the E-multiplier, RF-amplitude 1, Osc. control, and RF-amplitude 2 parameters showing anomalies at different times on 01/01/2019 and 02/01/2019.

Using principal component analysis (PCA) and applying the T^2 Hotelling chart after dimensionality reduction proved to be more effective than monitoring parameters individually. However, individual charts improved the anomaly detection performance.

After 18 days from the first anomaly detection using the T^2 Hotelling chart, the cesium standard showed instability, as indicated by a significant increase in Allan deviation values. This provides enough time to take corrective actions and prevent significant issues in the atomic time scale.

References

[1] Carvalho R J and Osório P L 1996 *Estabelecimento da Escala de Tempo Atômico Brasileira*. Anais do 2º Seminário Internacional de Metrologia Elétrica, pp.61-66

[2] Guinot B 1995 *Scale of Time* Metrologia vol 31 pp 431-440

[3] Guinot B 2011 *Time scales in the context of general relativity* Philosophical Transaction of the Royal Society vol 369 n 1953 pp 4131-4142

[4] Tavella P et al 2022 *Towards a consensus on a continuous coordinated universal time.* Metrologia, vol 60, n 1 p 014001

[5] Allan D W et al 1975 *An Accuracy Algorithm for an Atomic Time Scale* Metrologia pp 133-138

[6] Tavella P and Thomas C 1991 *Comparative Study of Time Scale Algorithms* Metrologia vol 28 pp 57-63

[7] Thomas C et al 1994 *Time scale*

[8] Carvalho R J and Osório P L 1996 *Estabelecimento da Escala de Tempo Atômico Brasileira*. Anais do 2º Seminário Internacional de Metrologia Elétrica, pp.61-66

[9] Teles F et al 2002 *Characterization of the Brazilian Cs atomic-frequency standard: evaluation of major shifts* Metrologia vol 39 n 2 p 135

[10] Lombardi M 2008 *The accuracy & stability of quartz watches* Horological Journal vol 150 n 2 p 57

[11] Chadsey H and Kubik A 1997 *Maintenance of HP 5071A primary frequency standards at USNO* Naval Observatory Washington Dc Time Service Div

[12] Guinot B 1995 *Scale of Time* Metrologia vol 31 pp 431-440

[13] Hotelling H 1947 *Multivariate quality control, illustrated by the air testing of sample bombsights* Techniques of Statistical Analysis pp 111-184

[14] Montgomery D C 2012 *Introduction to statistical quality control* 7 ed

[15] Kieling A C et al 2012 *Estudo do Gráfico T^2 de hotelling no controle estatístico do processo multivariado* VII Congresso Nacional de Engenharia Mecânica

[16] Konrath A C et al 2002 *Decomposição da estatística do Gráfico de controle multivariado T^2 de hotelling por meio de um algoritmo computacional*

[17] Prabhu S S and Runger G C 1997 *Designing a multivariate EWMA control chart* Journal of Quality Technology vol 29 n 1 pp 8-15