Forecasting of Space Geodesy Data and Investigation of the Relationship with the Solar Activity

Guilherme Oliveira Menezes, Jean-Pierre Raulin, Ray Fernando Hidalgo Ramirez, Leandro Augusto da Silva, Sergio V. D. Pamboukian UPM, Stephen Merkowitz GSFC/NASA

Abstract

This research aims to analyze the parameters that measure the movement of the planet and the natural phenomena that interfere in the variation of the duration of the terrestrial days, in order to predict the future situation in the coming months. The proposed methodology consists in collecting data of the day's variation on the IERS website (International Earth Rotation and Reference Systems Service), sunspot data on the SILSO website (Sunspot index and Long-term Solar Observations) and the Sun's polar magnetic field data on the WSO website (Wilcox Solar Observatory) using Webscrapping technique and store this data in a non-relational database. The predictions of the planet's rotation movement were made using the forecasting of time series through the auto regressions and autoregressive vectors, using the data collected from the VLBI (Very Long Baseline Interferometry). Through statistical analyzes with the graphs, the results show a 22-year LOD cycle and a 22-year PMF cycle, half the 11-year sunspot cycle; and the periodicity of the day's variation and its forecasts resulted an RMSE (root mean squared error) in the order of only 0.2 milliseconds error. Another result was the introduction of the Sun's polar magnetic field (PMF) variable that made possible to create a new forecast curve for the variation of the length of the day (LOD).

Introduction

The Earth is not a rigid body that rotates perfectly. Variations in the length of the day (LOD) that are about 10^{-3} seconds occur each day, and variations in the order of tens of seconds occur over several decades. To explain these variations in the Earth's rotation, it is necessary to understand its relationship with different phenomena such as the behavior of the atmosphere and the oceans, the elasticity of the Earth and a magnetic hydrodynamics of the planet's nucleus (LAMBECK, 1989).

The different phenomena responsible for the variation of the LOD can be seen in Figure 1. It can notice the interference of earthquakes, tides, melting glaciers, movement of tectonic plates, gravitational attractions, interference from the Moon and the Sun, and others.

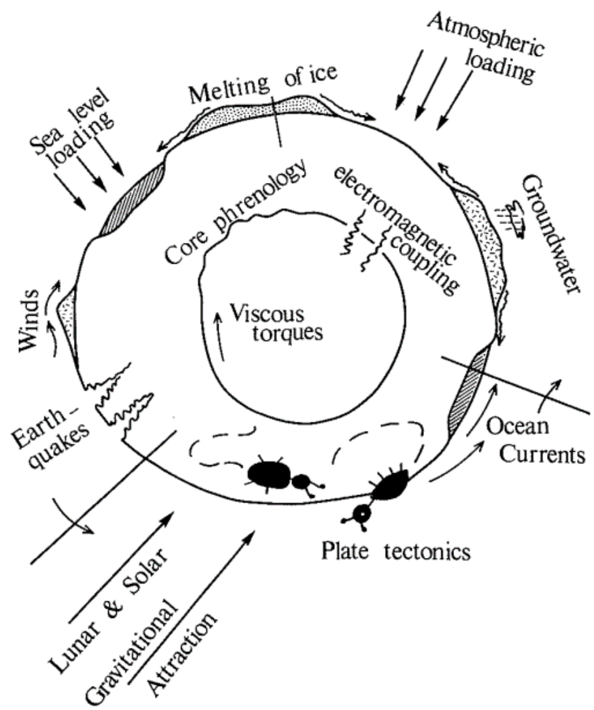

Figure 1 *– The illustration shows different phenomena that cause disturbances in the Earth's rotation, such as the movement of fluids inside the planet, the redistribution of mass, and movements that occur in the atmosphere and in the oceans. (Font: LAMBECK, 1980).*

Carter's study (1986) showed the degree of importance of the angular momentum of the atmosphere in relation to the variation in the length of the LOD day, represented in Figure 2.

This variation of movements of the planet Earth will be analyzed by the scientists, because they are changing the parameters of orientation of the Earth also known as EOP (Earth Orientation Parameters), parameters that are obtained through the technique of VLBI (Very Long Baseline Interferometry) (MALKIN , 2008).

Some techniques in the fields of statistics and artificial intelligence have been developed to improve the accuracy of the prediction of these parameters. However, there is still no fully satisfactory method to predict all these parameters (MODIRI et al, 2020).

The VLBI network uses the technique of interferometry in continental proportions. This technique consists of having at least two antennas or more placed far apart in order to capture radio waves from space sources. It is an alternative technique to the single antenna telescope (“single-dish”) which is limited to the capture radius in proportion to the diameter of the equipment (CARTER & ROBERTSON, 1986).

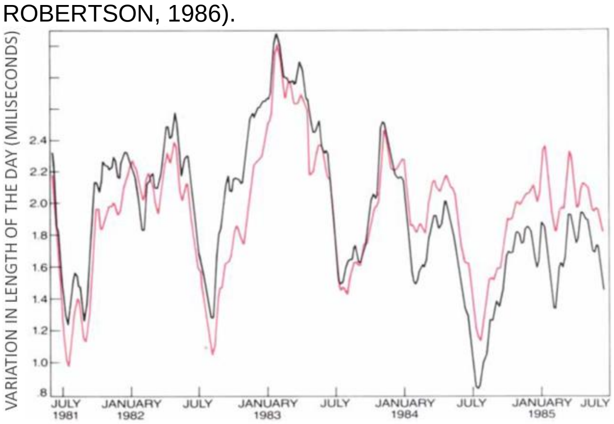

Figure 2 *- It can be noted that there is a correspondence between the values of the duration of the day and the angular momentum of the atmosphere, represented by the black and red lines, respectively (Font: CARTER & ROBERTSON, 1986).*

After capturing the spectrum of waves from space sources, the data is recorded on magnetic tapes that are then correlated, and in this way a result equivalent to that of observing an antenna with a diameter equal to the separation of the antennas is achieved. That is, if the antennas are separated by 100 meters, this corresponds to an observation of an antenna with a diameter of 100 meters (MORISON, 2008).

Studies and technologies have advanced rapidly, and distances of the order of thousands of kilometers have become necessary. Due to this need, the VLBI (*Very Long Baseline Interferometry*) technique arises, to meet the growing demand for greater precision in radio wave observations (KARTTUNEN, 2017).

In practice, the VLBI is an interferometer of intercontinental proportions, where radio telescopes are used in different countries and continents acting together to observe the same star in the universe, thus achieving better resolutions, as if simulating an antenna with thousands of kilometers in diameter.

Among the benefits of this technique, in addition to the significant improvements in spatial resolution, the most profound discoveries of the universe are made, one can study the origins of the formation of the universe, observe quasars, discover new radio galaxies, pulsars, among other diverse applications. One of the most famous was the recent signal captured from a black hole, which appeared in several media at the beginning of 2019.

In addition, VLBI can also be applied to the study of Earth dynamics, or geodesy. The technique is capable of measuring the movements of tectonic plates with millimeter precision, accurately analyzing the different types of movements that the Earth performs, and can even measure the variation in the duration of terrestrial days with a precision of milliseconds, that is, not all days are 86400 seconds, there is a daily variation of milliseconds (THOMPSON et al., 2017).

In analyzing the data collected by VLBI, one can, for example, notice that the Earth accelerates in some years, but reduces its speed in others. So far, it is not known exactly why this phenomenon occurs and this is one of the motivations for this research.

Through the VLBI technique it is also possible to map quasars in the celestial sphere, thus creating an International Celestial Reference System (ICRF). Quasars are so far away that almost no movement is detectable.

Considering all the movements of the planet Earth (rotation, translation, precession, nutation and others) it is possible to transform these ICRF points into an International Terrestrial Reference System (ITRF). For this transformation, it is necessary to know and calculate all the parameters of Earth orientation (EOP).

Currently, the human being's routine is very dependent on technology, and it seems that a millisecond more or less in the duration of the day will not interfere in people's daily lives. However, there would be an impact on communication systems if you lose access to the GPS (*Global Positioning System*) system. This millisecond over the duration of the days causes the Earth to rotate a little more, thus generating a small angle between the displacement of the previous day and the displacement of the current day (Figure 3). This small angle represents a small distance on the Earth's surface. However, at altitudes corresponding to the orbits of geostationary satellites, the accumulated delay in some days corresponds to a large distance, which can result in the loss of communication with the satellite and consequently affect the GPS. Thus, frequent monitoring is necessary to make the correct corrections of the satellite's position.

The study and application of the VLBI (Very Long Baseline Interferometry) technique to measure variations in the duration of the day allows us to better understand the behavior of the Earth's movements, which will bring benefits both in the new discoveries for the scientific and practical environments. This better understanding of this phenomenon will help in the monitoring of satellites and the navigation of spaceships, since small angular errors on Earth end up generating large errors when it is very distant, which is the case with satellites, which requires corrections in the positioning of the antennas.

Objective

The general objective of this work is to predict future values of LOD data using autoregressive techniques for time series and autoregressive vectors to compare the performances (BROWNLEE, 2020; HAYKIN, 2009; HYNDMAN & ATHANASOPOULOS, 2018). For this, it is set as specific objectives to investigate the periodicity of variability in the duration of days using the data obtained through the technique of VLBI and the movements of the Earth, make exploratory analyzes of such data, seek patterns and work on comparisons with solar activity to investigate what might be related with this effect.

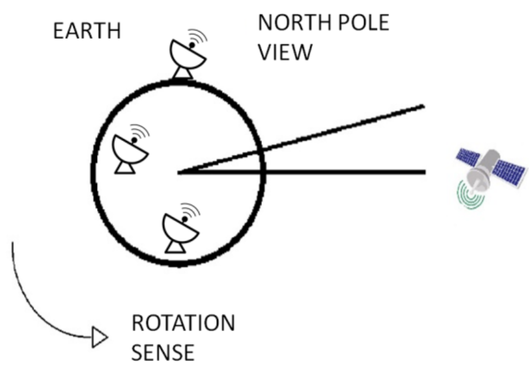

***Figure 3**. This is the view from the north pole of planet Earth. The arrow represents the direction of the Earth's rotation, while the lines represent a small displacement on Earth of milliseconds more, this can generate a large displacement in space (the figure is not scaled).*

Method

The purpose of this work is to predict the LOD curve and investigate the influence of natural phenomena on the Earth's rotation parameters, specifically the duration of the day (LOD).

The LOD can be estimated by the SLR (Satellite Laser Ranging), GRACE (Gravity Recovery And Climate Experiment), and the VLBI (Very Long Baseline Interferometry) technique (BOURDA, 2008).

This investigation to understand the behavior of the planet's movements and the variation of time will be done through the VLBI technique. Through autoregressive algorithm in time series analyzes, search for patterns and comparisons with other factors, such as solar activity and magnetic field, and other factors that explain how the parameters of the planet Earth are being influenced. Especially the length of the day (LOD). The daily updated data will be collected using the webscrapping technique, stored in a database and analyzed. The methodological scheme is illustrated in Figure 4.

Before forecasting a time series using an Auto regressive model, it is necessary to check if the serie has a stationary behavior. For that, it is necessary to perform a unit root test, such as the Dickey Fuller test. When failing the test, the differencing technique can be applied to transform the data stationary, after this stage the series must pass a new unit root test, and repeat this cycle over and over until it is approved.

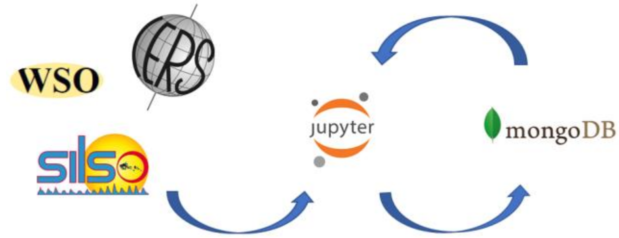

***Figure 4**: Methodological scheme of this work*.

Forecasts were made using the AR (Auto regressive) model. They are based on previous observations of the same variable, this means that the algorithm will try to create an equation that corresponds to the series studied based on the information, this means that the variables on which the forecast will be based is the variable itself in previous moments.

Auto Regressive model with n auto regressive terms (lag), according to Equation 1:

$$y'_t = \Theta_0 + \Theta_1 y'_{t-1} + \Theta_2 y'_{t-2} + \ldots + \Theta_n y'_{t-n} + \epsilon_t \quad (1)$$

The predicted value y_t is represented by the sum of a constant Θ_0 and the combination linear of predecessor values $y'_{t-1} \ldots y'_{t-n}$ with the inclinations $\Theta_1 \ldots \Theta_n$, added to white noise ϵ_t (HYNDMAN & ATHANASOPOULOS, 2018)

In the AR model, the forecast variable is influenced by the predictors of the series itself. A VAR (Vector Autoregressions) model is a generalisation of the univariate autoregressive model for forecasting a vector of time series. In this framework, all variables are treated symmetrically. They are all modelled as if they all influence each other equally.

The right hand side of each equation includes a constant and lags of all of the variables. Considering a two variable VAR with one lag, a two-dimensional VAR (1) can be represented by the following Equations 2 and 3:

$$y_{1,t} = c_1 + \phi_{11,1} y_{1,t-1} + \phi_{12,1} y_{2,t-1} + e_{1,t} \quad (2)$$

$$y_{2,t} = c_2 + \phi_{21,1} y_{1,t-1} + \phi_{22,1} y_{2,t-1} + e_{2,t} \quad (3)$$

Where $e_{1,t}$ e $e_{2,t}$ are the white noise. The coefficient $\phi ii,\ell$ captures the influence of the ℓth lag of variable yi on itself, while the coefficient $\phi ij,\ell$ captures the influence of the ℓth lag of variable yj on yi (HYNDMAN & ATHANASOPOULOS, 2018).

Results

Some exploratory analyzes were performed on the data to get knowledge about the distribution and time series of day length variations (LOD). With the work developed, Figure 5 was found, a cycle of approximately 22 years was noted in the LOD, exactly half of the 11-year solar activity cycle, obtained through SILSO sunspot data (Data were collected on April 28, 2020).

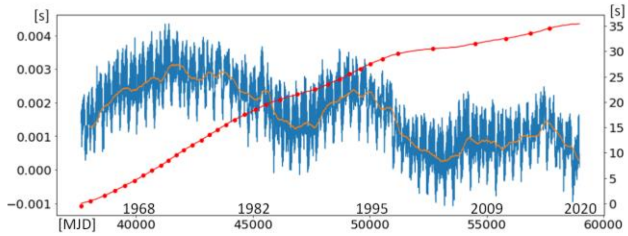

***Figure 5**: The graph shows the variation in the duration of days over the years, the blue line is the actual LOD measurement, the orange line is the annual moving average and the red line is the accumulated value. When a sum of the accumulated value reaches the value of one second (red dots - Leap Seconds) clocks from around the world need to be corrected.*

The graph in Figure 5 shows an acceleration trend in the rotation speed of the Earth, as the observed values are below zero, indicating days with less than 86400 seconds. This phenomenon was very observed in the year 2020, when there were new records of short days in the months of July and August.

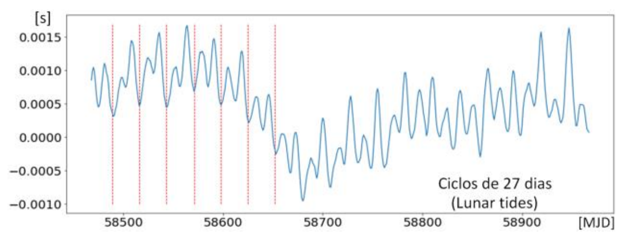

***Figure 6**: When the series was analyzed in detail, 27-day cycles were found in the behaviour of the serie, which are related to the lunar tide cycle detected by Wahr in 1988.*

In order to facilitate the study of the LOD series, the trend was removed and a comparison was made with sunspots, considering that the cycle found is twice that of a cycle of solar activity.

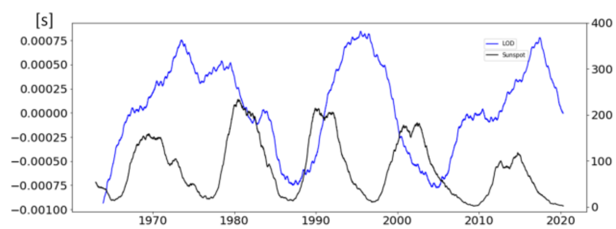

***Figure 7:** The blue line represents the LOD variation in seconds and the black line the sunspots variation number.*

This relationship of LOD with sunspots leads us to better investigate their interaction with the Sun, and when is analyzed the data from the Polar Magnetic Field of the Sun (PMF-avg), a 22-year cycle is noted, which when plotted on a graph shows the peaks and valleys of the two series together over the years, as shown in Figure 8.

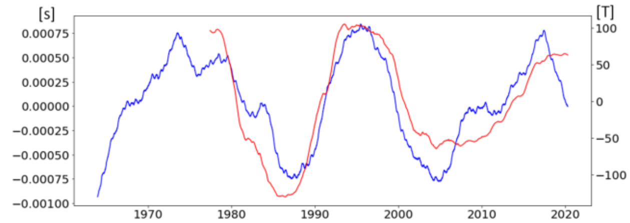

***Figure 8**: The blue series represents the annual moving average of the LOD and the red series the annual moving average of the Polar Magnetic Field Average (PMF-avg).*

When analyzing this similarity in the behavior of the two series over the years, it is intuitive imagine that they will continue with this behavior in the coming years, and therefore, when the LOD forecast is made, it is expected that the PMF will have a similar behavior.

In order to improve the best model found using AR, which uses only LOD data, new simulations were realized using VAR, a technique that allows introducing the PMF vector in LOD prediction, to see if this will help the prediction or not.

To do the AR, the data were separated for algorithm training and validation tests, and as there is no standardization in the LOD forecast researches, a forecast of 500 was chosen to be able to compare with other researches and explore the long-term forecasts more.

To choose the best Lag, a partial autocorrelation (PACF) graph, represented by Figure 9, was made:

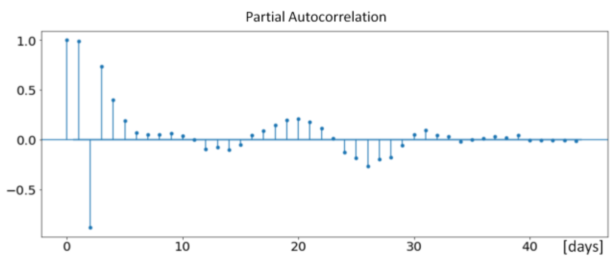

***Figure 9**: The Partial Autocorrelation graph shows us the degree of correlation of the days that precede the predicted day have.*

Simulations were made for lags with PACF above 0.25 and 0.05, with lags that go up to 100, 500 and 5000, as shown in Figure 10.

Experiments were realized with different lags in search of the smallest RMSE. Lags were simulated every 50 to 50, and later lags were simulated every 500 to 500. The best model found in this first stage of simulations was for a lag of 7500, whose RMSE error was $2{,}126\text{x}10^{-4}$ seconds. In a second stage, values between 7 and 8 thousand were simulated, varying from 100 to 100, the 7500 model continued to be the best. In a third step, lag values between 7400 and 7600 were simulated, always reducing by half the value in relation to the best rmse value found, thus the lag values were converging to the lag of 7468, whose RMSE error was $2{,}112\text{x}10^{-4}$ seconds, as shown in Figure 11.

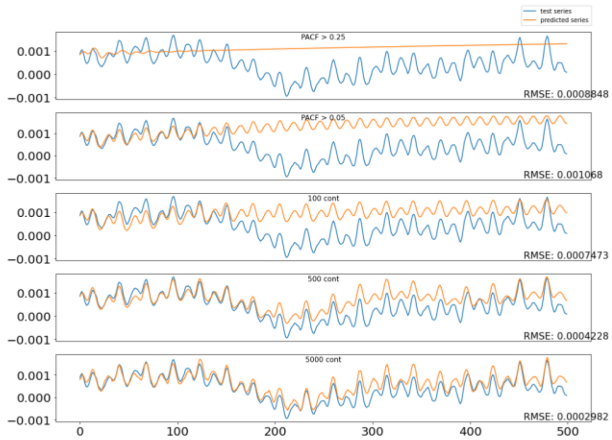

Figure 10*: Simulations to find the best model. The orange line shows the forecast found and the blue line represents the last 500 days of the original series, used to calculate the RMSE.*

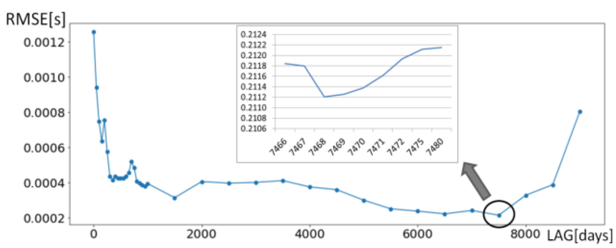

Figure 11: *Each point on the graph represents the error of a simulation performed for different lags for the 500-days forecast. The best model found was AR (7468) with RMSE error of 2,112x10^{-4} seconds.*

When compared to other studies that used other techniques to predict LOD, it is noted that the result was very competitive in the 500-days forecasts. For the other projection horizons, the optimal lag would need to be further investigated, but some of these simulations are shown for comparison purposes in the table in Figure 13.

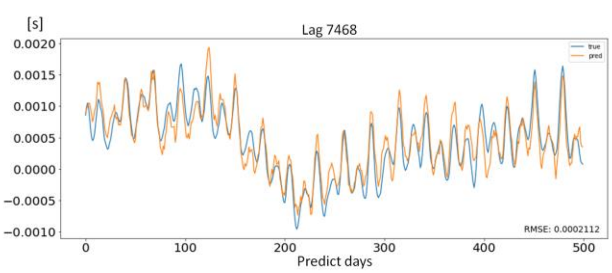

Figure 12: *This is the best prediction found. Lag 7468 with RMSE of 2,112x10^{-4} seconds. Orange line is the forecast, while the blue line is the real data.*

After the intense search for the best AR lag for 500 days, the Polar Magnetic Field (PMF) vector was introduced to see if there is an improvement in the LOD forecast. Other tests were realized considering only the trend withdrawal from the series.

Author	Techniques	Performance Measure	Performance [ms]	Projection horizon [days]
Liao 2012	BP ANN	RMSE	0.4	400
Lei Yu 2015	Gaussian processes	RMSE	0.25 0.29	30 360
Lei Yu 2015	ELM Kalman filter	RMSE	0.32 0.22	500 500
Modiri 2020	Archimedean+ SSA	MAE	0.09	10
This research	AR (7468)	RMSE	**0.21**	500

Figure 13: *Comparison with other researches that made the LOD forecasts.*

As can be seen in Figure 14, for the different lags from 1 to 100 were tested: the AR analyzing the original LOD series; the two-dimensional VAR using the PMF+LOD vectors; the AR diff, analyzing the series after applying differencing; and AR, VAR without the trend. In this way it is possible to see that these procedures cause improvements in the forecasts for these lags.

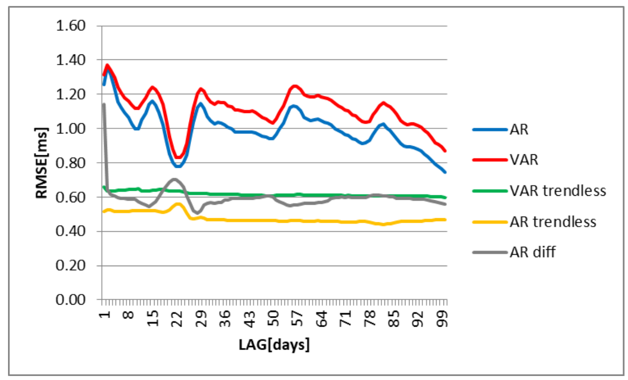

Figure 14: *When using lags between 1 and 100, it is easy to compare which one has better performance. The graph shows the variation of the error of the LOD forecasts in relation to the lag used.*

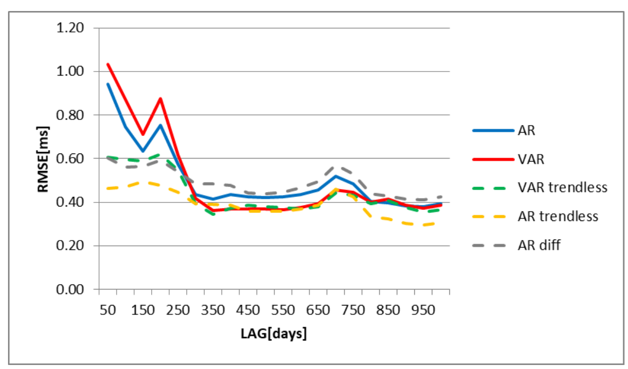

Figure 15: *Continuation of the graph in Figure 14. The AR diff, VAR and AR trendless series are dashed to highlight the improvement of the VAR (red line) in relation to the AR (blue line).*

When increased the lags to hundreds, the results were very close, as can be seen in the graph in Figure 15. The introduction of the PMF vector in the VAR (red line) improved the predictions for Lags from 300 to 800 approximately, when compared to the AR (blue line).

Therefore, the relationship between this new vector (PMF) and LOD needs to be further investigated.

The same tests were expanded to the order of thousands as can be seen in Figure 16.

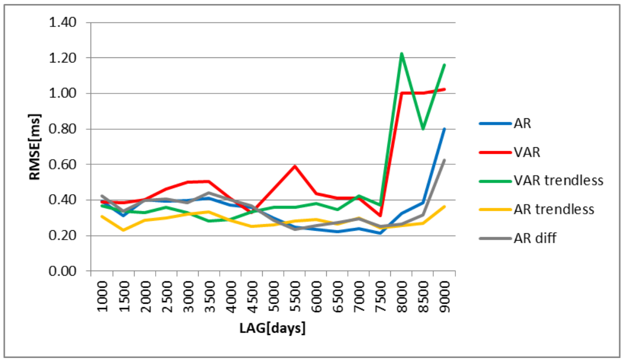

Figure 16: *Continuation of the graph in Figure 15. Tests realized for lags ranging from 500 units to 9000. In all experiments the errors start to increase when approaching the 9000 lag.*

Despite the diverse experiments realized with the LOD series in search of the best forecast, the first experiment remained better than the others.

Conclusions

Given the similarities identified in the behavior of the LOD series with the PMF series, the relationship between these phenomena should be further investigated.

With the exploratory analyzes performed, cycles of 27 days and 22 years were found in the LOD data. In the case of sunspots the cycles of 11 years were identified, and in the data of the polar magnetic field of the Sun cycles of 22 years. The accumulated value of the LOD in the year 2020 for the first time was negative (leap second), the international time will withdraw a second instead of adding.

About the Forecast, the best autoregressive model found so far was an AR (7468) that presented an RMSE error of $2{,}112\times10^{-4}$ seconds. The idea of doing a forecasting by introducing the vector polar magnetic field of the Sun jointly with the duration of the LOD day, proved to be quite promising because there are still no studies that prove this relationship. For lags from 300 to 800, the introduction of the PMF vector managed to improve these forecasts, this is another aspect that needs to be better studied as it is a variable that has not yet been considered in other forecasting models.

In this way, this research helped to understand a little better and to predict the behavior of the Earth's rotation using a predictive artificial intelligence technique.

Acknowledgments

The authors would like to acknowledge the support of CAPES (Project 88881.310386/2018-01).

References

BOURDA, Géraldine. Length-of-day and Space-geodetic Determination of the Earth's Variable Gravity Field. *Journal of Geodesy*, vol. 82, 4 ed, p.295–305, 2008.

BROWNLEE, Jason. *Introduction to Time Series Forecasting with Python:* How to Prepare Data and Develop Models to Predict the Future. v1.9, 2020.

CARTER, William. E.; ROBERTSON, Douglas. S. Studying the Earth by Very-long-baseline Interferometry. *Scientific American*, v. 255, n. 5 p. 46–54A, 1986.

HAYKIN, Simon. *Neural networks and learning machines*. Upper Saddle River, NJ. Pearson Education, 2009.

HYNDMAN, Rob. J.; ATHANASOPOULOS, George. *Forecasting:* principles and practice, 2 ed, Melbourne, Australia: OTexts, 2018.

KARTTUNEN, Hannu. et al. Observations and Instruments: Radio Telescopes. In: KARTTUNEN, Hannu. et al. *Fundamental Astronomy*. 6th ed. Heidelberg: Springer, 2017. 74-80.

LAMBECK, Kurt. Changes in length-of-day and atmospheric circulation. *Nature*, n. 286, p.104–105, 1980.

LAMBECK, Kurt. The Earth's Variable Rotation: Some Geophysical Causes. In: BABCOCK, A. K.; WILKINS, G.A. (ed). *The Earth's Rotation and Reference Frames for Geodesy and Geodynamics* [S.l.]: Springer Netherlands, 1989. p. 241–284.

LIAO, D.C; WANG, Q. J.; ZHOU, Y. H.; LIAO, X. H.; HUANG, C. L. Long-term Prediction of the Earth Orientation Parameters by the Artificial Neural Network Technique. *Journal of Geodynamics*. 62, 87-92, 2012.

MALKIN, Zinovy. On comparison of the Earth Orientation Parameters Obtained From Different VLBI Networks and Observing Programs. [S.l.]: *Journal of Geodesy*, 83, 547–556, 2008.

MODIRI, Sadegh; BELDA, Santiago; HOSEINI, Mostafa; HEINKELMANN, Robert; FERRÁNDIZ, José. M.; SCHUH, Harald. A New Hybridmethod to Improve The ultra-short-term Prediction of Lod. *Journal of Geodesy*, v. 94, p.23, 2020.

MORISON, Ian. Observing the Universe: Radio Telescopes. In: MORISON, Ian. *Introduction to Astronomy and Cosmology*. Chichester: John Wiley & Sons, 2008. 181-193.

THOMPSON, Richard. A.; MORAN, James. M.; SWENSON, George. W. J. *Interferometry and Synthesis in Radio Astronomy*. 3rd. ed [S.l.]: Springer, 2017.

YU, Lei.; DANNING, Zhao; HONGBING, Cai. Prediction of Length-of-day Using Extreme Learning Machine. *Geodesy and Geodynamics*, v. 6, n.2, p.151-159, 2015.

YU, Lei; DAN-NING, Zhao; YU-PING, Gao; HONG-BING, Cai. Prediction of Length-of-day Variations Based on Gaussian Processes. *Chinese Astronomy and Astrophysics*, v. 39, p. 368-379, 2015.