ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 11, nº 2 (2025)

https://doi.org/10.21680/2447-3359.2025v11n2ID39777

The use of machine learning for detecting descontinuities in RBMC station coordinate time series

O uso de machine learning para detecção de descontinuidades nas séries temporais das coordenadas da RBMC

Alberto Luis da Silva[1]; Julio Cesar de Oliveira[2]; William Rodrigo Dal Poz[3]

[1] Federal University of Viçosa/Department of Civil Engineering, Viçosa/MG, Brazil. Email: alberto.silva@ufv.br
ORCID**:** https://orcid.org/0009-0008-4172-1265

[2] Federal University of Viçosa/Department of Civil Engineering, Viçosa/MG, Brazil. Email: oliveirajc@ufv.br
ORCID**:** https://orcid.org/0000-0003-0894-5597

[3] Federal University of Viçosa/Department of Civil Engineering, Viçosa/MG, Brazil. Email: william.dalpoz@ufv.br
ORCID**:** https://orcid.org/0000-0001-9532-3643

Abstract: The GNSS observations collected by RBMC allow estimating, with millimeter precision, the daily coordinates of its stations. As a result, it is possible to generate robust time series, capable of describing geodynamic displacements, such as the movements of lithospheric plates, as well as the effects of terrestrial, oceanic and atmospheric tides. However, discontinuities in these series indicate changes in the reference coordinates of the stations, and need to be considered when significant. Events such as antenna changes and earthquakes are the main sources of discontinuities in time series. In this work, we evaluated the capacity of machine learning algorithms in the automatic identification of discontinuities related to antenna changes. Of the five methods evaluated, Random Forest presented the best performance, with an F1-Score of 0.78 and an accuracy rate of 77.5% for discontinuities equal to or greater than 1 cm. The results demonstrate the potential of machine learning methods in classifying patterns in time series of RBMC coordinates. However, performance depends on adequate data processing and the representativeness of the modeled events.

Keywords: RBMC; GNSS; machine learning.

Resumo: As observações GNSS coletadas pela RBMC permitem estimar, com precisão milimétrica, as coordenadas diárias de suas estações. Como resultado, é possível gerar séries temporais robustas, capazes de descrever deslocamentos geodinâmicos, como os movimentos das placas litosféricas, bem como os efeitos das marés terrestres, oceânicas e atmosféricas. Entretanto, descontinuidades nessas séries podem indicar alterações nas coordenadas de referência das estações, e precisam ser consideradas quando significativas. Eventos como trocas de antenas e terremotos, são as principais fontes causadoras de descontinuidades nas séries temporais. Neste trabalho, avaliou-se a capacidade de algoritmos de *machine learning* na identificação automática de descontinuidades relacionadas a trocas de antenas. Dos cinco métodos avaliados, o *Random Forest* apresentou o melhor desempenho, com um F1-Score de 0,78 e uma taxa de acerto de 77,5% para descontinuidades iguais ou superiores a 1 cm. Os resultados demonstram o potencial dos métodos de *machine learning* na classificação de padrões em séries temporais de coordenadas da RBMC. Contudo, o desempenho depende de um tratamento adequado dos dados e da representatividade dos eventos modelados.

Palavras-chave: RBMC; GNSS; machine learning.

Received: 07/04/2025; Accepted: 13/08/2025; Publication: 26/08/2025.

1. Introduction

Since the establishment of its first station in 1995, in Brasília, the Brazilian Network for Continuous Monitoring of GNSS Systems (RBMC) has been providing observation data from points located on the Earth's surface (COSTA et al., 2012). In total, there are more than 200 stations that serve as support for works requiring a primary reference structure. The continuous operation of the network and systematic GNSS data processing enable the estimation of daily station coordinates with millimeter-level accuracy, thus generating time series suitable for monitoring the geodetic reference frame in Brazil. Nevertheless, natural phenomena such as earthquakes, as well as equipment maintenance activities, including antenna replacement, may introduce discontinuities (offsets) into these time series (DAWIDOWICZ et al., 2023; LE et al., 2024; LAHTINEN et al., 2022).

The identification of discontinuities in coordinate time series results in the segmentation of the original series into multiple solutions, as discussed by Altamimi et al. (2023). The BRAZ station had its series segmented into eight solutions, each with distinct coordinates in ITRF2020 (ALTAMIMI et al., 2022). Tools such as FODITS, available in the Bernese software (DACH et al., 2015), enable the analysis of discontinuities. However, most of these tools require prior information about the events, which makes the process limited and semi-automatic.

The objective of this study is to evaluate the ability to identify discontinuities caused by antenna replacements using machine learning techniques. Time series of coordinates from 198 RBMC stations, estimated by the Nevada Geodetic Laboratory (NGL) at the University of Nevada, Reno (BLEWITT et al., 2018), were employed. Daily relative solutions estimated by IBGE were also used. Among the machine learning algorithms already applied in GNSS studies (SIEMURI et al., 2022), five were tested: Random Forest, Linear Support Vector Machine, K-Nearest Neighbor, Decision Trees, and Naive Bayes. Considering that the time series present standard deviations that can reach up to 0.7 cm, we defined discontinuities equal to or greater than 1 cm as significant, following the criterion adopted by Crocetti et al. (2021).

One of the main challenges in identifying discontinuities caused by antenna replacements is the absence of data or the presence of noisy data. In addition, the offsets do not occur uniformly, as they depend on factors such as the antenna model (DAWIDOWICZ et al., 2023), station location, data quality, and the type of centering device and orientation antenna. Section 2 of this paper presents some characteristics of the RBMC time series, the discontinuities caused by earthquakes and maintenance activities, as well as the criteria adopted for selecting the analyzed series. The methodology employed for sample construction, data pre-processing, and classification evaluation criteria are described in Section 3. Section 4 presents the classification performance indices obtained for each machine learning algorithm, the detection capability for different offset amplitudes, and their application to RBMC time series produced by IBGE. Final conclusions and recommendations are discussed in Section 5.

2. RBMC station coordinate time series

The RBMC station data are continuously processed by several analysis centers, including those associated with SIRGAS, IGS, IBGE itself, and geodetic institutes such as NGL. The latter provides daily coordinates determined with the GipsyX software, in addition to a file named *steps.txt* (http://geodesy.unr.edu/NGLStationPages/steps.txt), which contains information on events considered potential sources of discontinuities, such as earthquakes and antenna replacements (BLEWITT et al., 2018). Information about the equipment can also be found in the *logfiles* of each station (IBGE, 2024).

Antenna replacements at stations, even when substituted with identical models, may introduce discontinuities due to the fact that phase center offset and variation corrections are established in laboratory conditions rather than at the station site, and are therefore influenced by local phenomena such as multipath (DAWIDOWICZ et al., 2023). Antenna orientation and changes in antenna height, when not properly defined, are additional maintenance-related factors that may also generate discontinuities (HUANG et al., 2025).

Figure 1 shows the time series of daily coordinates for the POAL station. The blue dashed lines represent earthquakes reported by Blewitt et al. (2018), while the red lines indicate the antenna replacements.

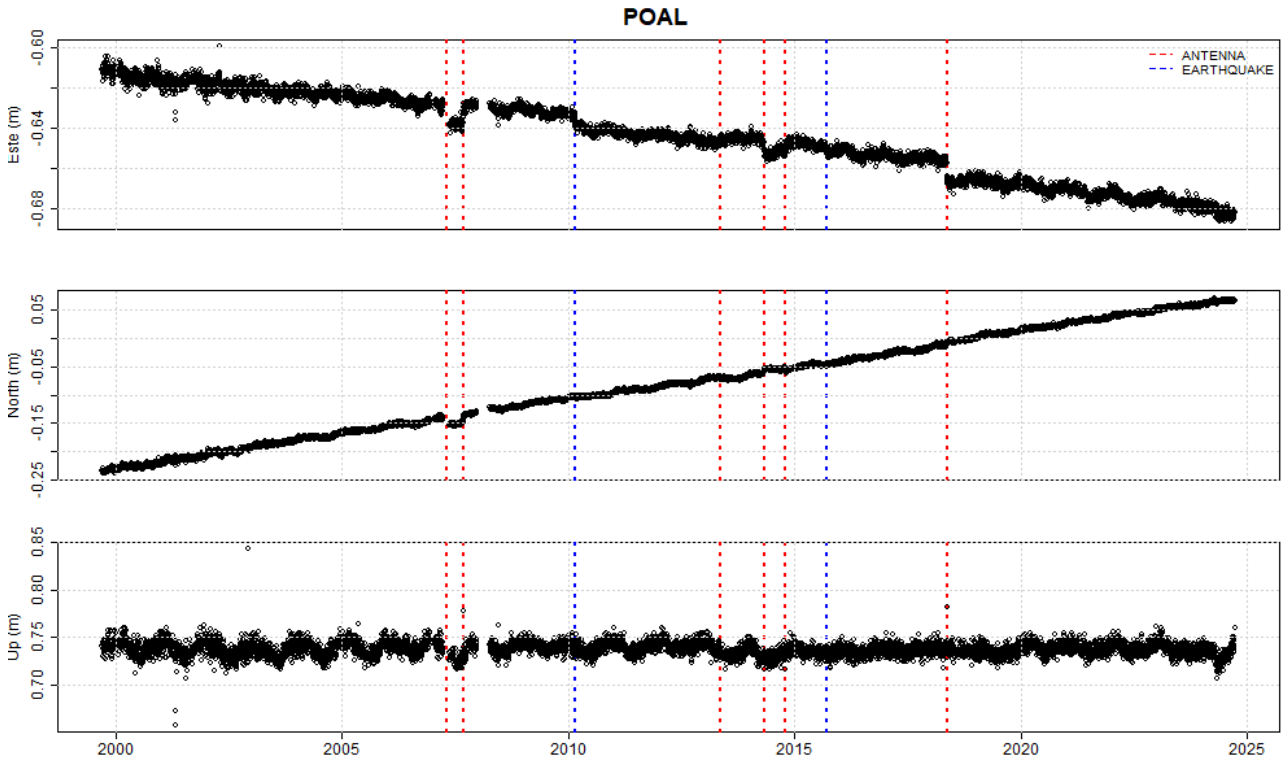


Figure 1 – Time series of the POAL station (Porto Alegre, RS) determined from NGL solutions. Discontinuities can be observed resulting from multiple antenna replacements and from the earthquake that occurred in January 2010 in Concepción, Chile. Source: Authors (2024).

3. Methodology

For training a machine learning algorithm, it is essential that the data are representative of the target classes and available in sufficient quantity to enable effective classification (SIEMERS et al., 2022; RAJPUT et al., 2023). In the case of discontinuities in time series caused by antenna replacements, observations are not always available, as many of these replacements occur due to equipment failures. In other cases, the data are available but noisy, i.e., of low quality, which is precisely the reason why the equipment is being replaced (Figure 2). In both situations, it is not advisable to use these periods in model training, since the discontinuity may not be present in the time window or may fail to exhibit a representative pattern (GAO et al., 2024).

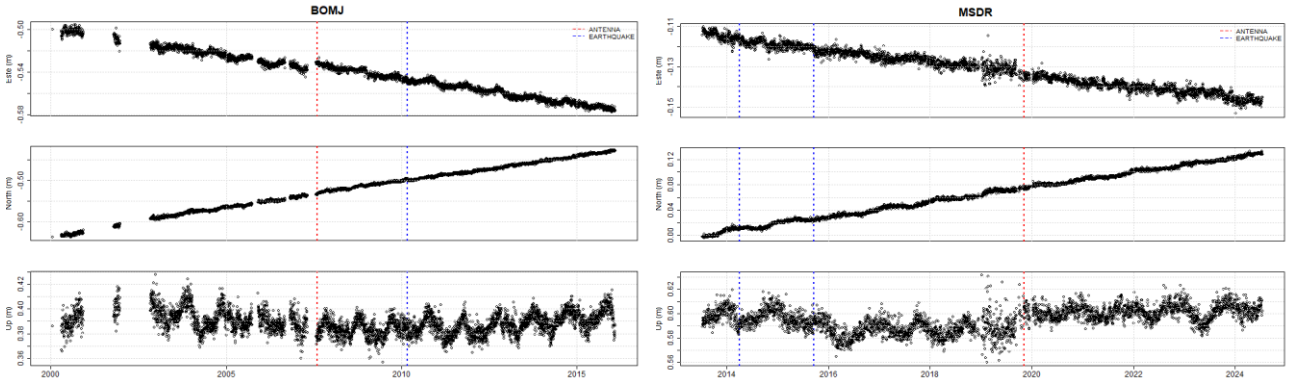


Figure 2 – Discontinuities resulting from antenna replacements, but difficult to model due to the lack of data prior to the replacement (BOMJ – August 1st, 2007) and the presence of highly noisy data (MSDR – November 4th, 2019). Source: Authors (2024).

The minimum amplitude of a jump characterized as a discontinuity depends on the natural variability of the time series, which results from the accuracy of the coordinates, the effects of terrestrial, oceanic, and atmospheric tides, as well as local geodynamic movements (SAVCHUK et al., 2023). In the RBMC station time series estimated by NGL, the average moving standard deviation of the East, North, and Up components, considering a three-week time window, was 1 mm, 1 mm, and 5 mm, respectively, with maximum values reaching up to 7 mm in the Up component. Since antenna replacement may generate discontinuities in all components, but with a higher probability of occurrence in the vertical component, values lower than the average standard deviation of the Up component should not be considered significant for classification purposes. In this study, a threshold of 1 cm was adopted for significant discontinuities, following the same criterion used by Crocetti et al. (2021). Discontinuities were computed as the difference between the median of the last ten days before the replacement and the median of the first ten days after the replacement.

Of the 233 antenna replacements recorded in the 198 time series analyzed from RBMC stations, only 33 stations exhibited discontinuities greater than 1 cm. Considering that 80% of the series are used for model development, only 26 stations would be selected for training. However, due to the particularities of each time series, the use of only 13% of the data from RBMC stations proves insufficient for establishing patterns that allow the classification of discontinuities across the entire network with acceptable accuracy. To address this low occurrence of events, several procedures were adopted in this study with the aim of achieving a more representative and efficient classification.

3.1. Selection of data for training

Considering the low number of known, significant, and suitable discontinuities for training a machine learning model, and at the same time the large amount of data available, we selected time series segments that exhibited data continuity, captured the seasonal variations of each station, and were free of gaps that could lead to the identification of false discontinuities, as well as free of antenna replacements and potentially disruptive earthquakes. Discontinuities were then intentionally introduced in order to evaluate the efficiency of the methods for detecting and handling such events (Section 3.3). We selected only periods with at least 95% of the data available, no more than one consecutive week without coordinates, and a minimum duration of four years, which ensures the capture of seasonal variations and allows for a more reliable estimation of velocities (TUNINI et al., 2024). Based on these criteria, from the 198 available solutions, we identified time series from 108 RBMC stations, of which 87 were used for training and 21 for testing (Figure 3).

Figure 3 – RBMC stations randomly selected for training (80%) and testing (20%) the machine learning models. Source: Authors (2024).

3.2. Removal of outliers and noisy segments

The data used for training the algorithm were preprocessed to remove undesired observations, such as those resulting from low-quality measurements or inadequate processing. First, a reference precision was determined for each time series component by calculating the median of the moving standard deviation estimated over a three-week time window, as presented in Crocetti et al. (2021). Next, the moving median of the time series, estimated for the same time window, was subtracted from the series values. Observations with differences greater than three times the reference standard deviation were considered outliers and therefore removed. After outlier removal, the identification of noisy segments was carried out. For this purpose, the reference values were estimated again. Observations were classified as noisy when the moving median of the standard deviation, estimated over a three-week window, exceeded twice the reference precision of the respective component (CROCETTI et al., 2021).

3.3. Introducing discontinuities into real data series

To enable the machine learning algorithm to be trained with samples containing well-defined jumps, discontinuities were simulated in real time series data. Thus, the only discontinuities present in the selected series were those intentionally introduced. For this purpose, a 1 cm jump was inserted into each time series by applying values of 0, +1, or −1 to the East, North, and Up components, where 0 indicates no jump, +1 indicates a positive jump, and −1 indicates a negative jump. Combinations among these three components were generated randomly, with the exception of the [0, 0, 0] combination, which was discarded as it does not represent any discontinuity. It should be noted that jumps caused by antenna

replacements can affect any of the coordinate components, although they occur more frequently and with greater intensity in the vertical component (WANNINGER, 2009). The epoch for inserting the jump was defined through a scan starting from the midpoint of the time series, searching for a continuous data interval corresponding to twice the sample window used in model training (21 days in this study). The scan was performed in both temporal directions, and the chosen period was the one whose midpoint was closest to the center of the time series (Figure 4).

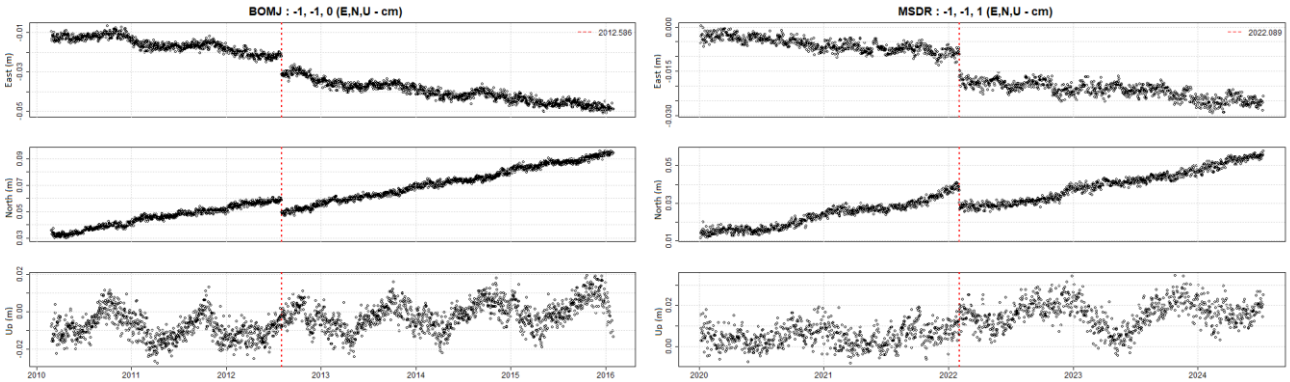


Figure 4 – Randomly inserted discontinuities in the time series of BOMJ (Bom Jesus da Lapa, BA) on August 1st, 2012 (−1 cm in East and −1 cm in North) and MSDR (Dracena, MS) on March 30th, 2022 (−1 cm in East, −1 cm in North, and +1 cm in Up). Source: Authors (2024).

3.4. Constructing sample for training

This step consisted of transforming a time series with n epochs into a set of $n - m$ samples. Each sample contains m epochs, extracted using a moving window that spans the entire series. Each of these samples represents a pattern which, when associated with information about the presence or absence of a jump and its position (if any), enables algorithm training. Since the coordinates are composed of three components (East, North, and Up), all of which may be affected by discontinuities, the algorithm was trained using the combined set of samples from the three components simultaneously. This resulted in a study matrix with dimensions $(n - m) \times 3m$. According to Crocetti et al. (2021), sample sizes that perform best in detecting discontinuities in coordinate time series are those composed of 21 or 28 epochs. In this study, we opted for the configuration with 21 epochs, which, considering the three components, yields 63 elements per sample, representing a 21-day window containing data from the East, North, and Up components. This configuration allows not only the identification of the presence of a jump but also its position within the window. Algorithm training was conducted using a vector, referred to as the target vector, in which values of 0 indicate the absence of a discontinuity in the sample, and values from 1 to 20 indicate the occurrence of a discontinuity and its relative position within the sampling window. It is worth noting that if the jump occurs precisely in the first epoch of the sample, the target vector takes the value 0, since in this case the sample will contain only post-jump data (Figure 5).

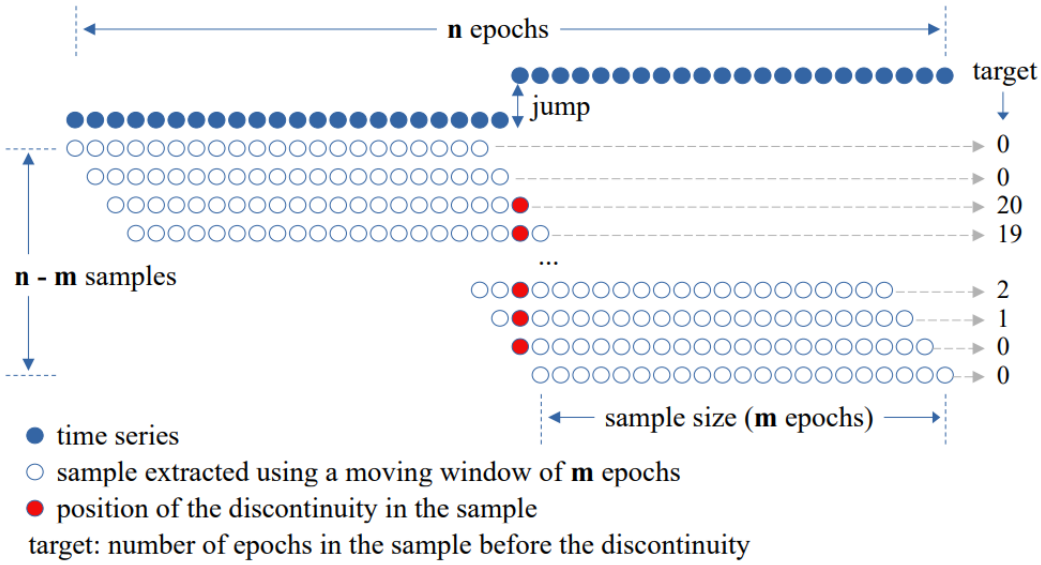


Figure 5 – Schematic representation of the study matrix built from a sliding window and the corresponding target vector. Source: Authors (2024).

In addition to the construction of the study matrix, another strategy recommended by Crocetti et al. (2021) and adopted in this work was the individual normalization of the values of each sample. This procedure aims to standardize the data scale, thereby facilitating the algorithm's learning process. Furthermore, three additional columns were inserted into the study matrix, corresponding to the original range of each sample for the East, North, and Up components, respectively. It is expected that, in the presence of a discontinuity, the range of the affected component will exhibit a proportional change in relation to the magnitude of the jump, when compared to samples without discontinuities in that same component.

3.5. Performance evaluation of the classification

Once the model was trained, it was necessary to evaluate its performance against the data reserved for testing. For this purpose, classification was performed according to the sample construction procedure, and the results were compared with the previously known labels using the confusion matrix (ÖZBEY et al., 2024; HEYDARIAN et al., 2022). Samples with discontinuities correctly classified were defined as True Positives (TP), while samples without discontinuities correctly identified were True Negatives (TN). Samples incorrectly classified as containing discontinuities correspond to False Positives (FP), and those that actually contained discontinuities but were classified as without jumps are False Negatives (FN). Additionally, there are cases in which discontinuities were correctly identified but with a temporal offset relative to the actual epoch of occurrence. In these cases, the sample was also classified as a True Positive (TP*), but not considered in the quality indicators (CROCETTI et al., 2021). Based on the confusion matrix values, the main performance metrics of the classification were determined: Precision (Pr), indicating the proportion of samples classified with discontinuities that actually occurred; Recall (Re), representing the ability of the model to correctly identify existing discontinuities; and the F1-score, which is the harmonic mean between precision and recall and was adopted in this study as the main indicator to select the algorithm that best fits the analyzed samples. Equations 1, 2, and 3 present the mathematical expressions used to compute these indicators (DE DIEGO et al., 2022). A flowchart summarizing the adopted methodology is presented in Figure 6.

$$Pr = \frac{TP}{TP+FP} \quad (1)$$

$$Re = \frac{TP}{TP+FN} \quad (2)$$

$$F1score = \frac{2xPrxRe}{Pr+Re} = \frac{2TP}{2TP+FP+FN} \quad (3)$$

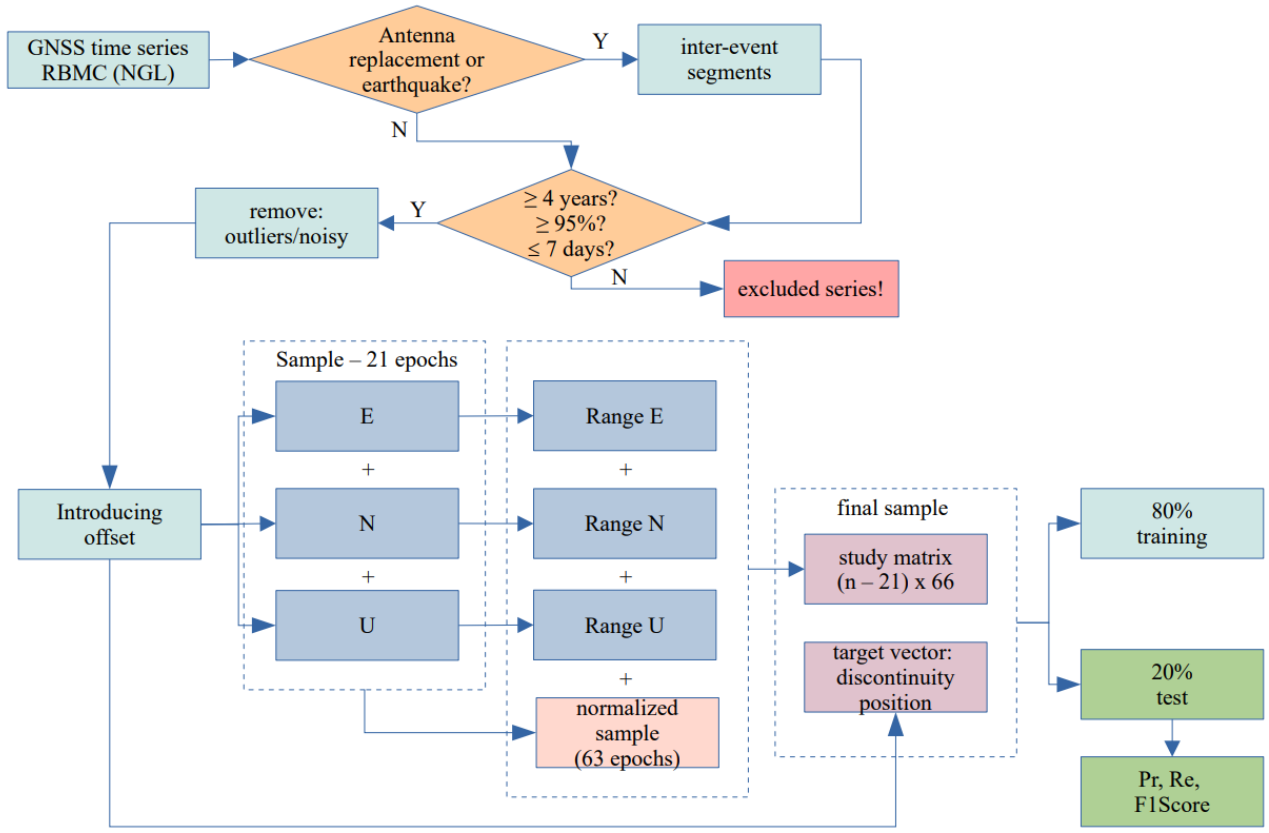


Figure 6 – Flowchart illustrating the methodology applied for data selection, outlier and noise detection, sample construction, and classification quality analysis. Source: Authors (2024).

4. Results

Machine learning classification performance is influenced by multiple factors, including the algorithm's logic, data quality and behavior, the methodology adopted for sample construction, preprocessing and filtering strategies, and the representativeness of classes within the training set. In the context of detecting discontinuities in coordinate time series from a continental-scale network such as RBMC, additional challenges arise. These include the pronounced seasonal behavior of the vertical component at certain stations, the differing accuracies among coordinate components, the wide geographic distribution of the stations, and the lack of spatial correlation between events. Moreover, the absence of observations in the days immediately preceding a discontinuity hampers the characterization of its pattern, thereby increasing the complexity of the classification task.

4.1. Classification algorithms

We analyzed the performance of five machine learning algorithms applied to multiclass classification: Random Forest – RF (ALI et al., 2012), Support Vector Machine – SVM (LORENA and CARVALHO, 2007), k-Nearest Neighbors – KNN (SUN et al., 2018), Decision Tree – DT (ROKACH and MAIMON, 2005), and Naive Bayes – NB (WEBB, 2011). As shown in Figure 7, and in agreement with Crocetti et al. (2021), the Random Forest algorithm achieved the best performance, with an F1-Score of 0.78, and was therefore adopted in this study. It is noteworthy that 77.5% of the existing discontinuities were correctly identified (TP + TP*) by Random Forest, with 39.6% being classified exactly at the epoch of their occurrence (TP). It was also observed that no false positives (FP) occurred, meaning that no discontinuity was classified without a corresponding record, resulting in a precision (Pr) of 1 for Random Forest. On the other hand, since 22.5% of the discontinuities were not identified (FN), the recall (Re) was 0.64. It is important to highlight that, for discontinuity detection, the Re value is more relevant than Pr, since it reflects the algorithm's ability to detect all events of interest.

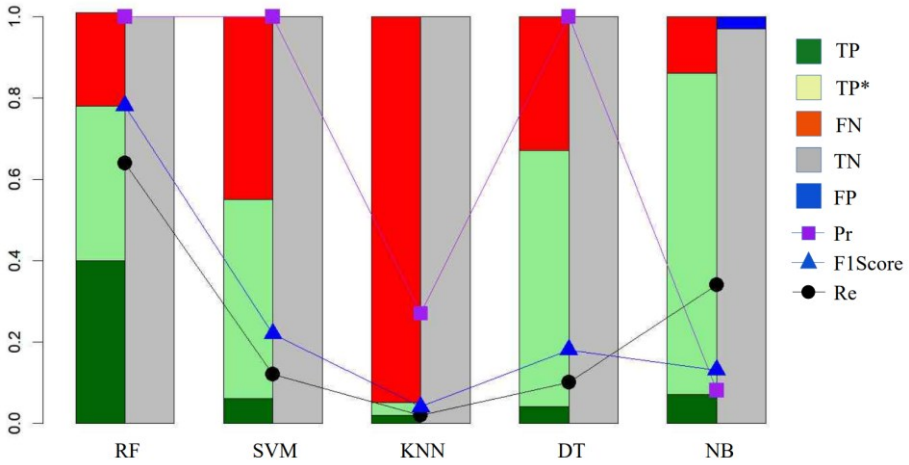


Figure 7 – Quality indicators of the algorithms tested in the classification of discontinuities in the RBMC time series. Source: Authors (2024).

4.2. Detection capability regarding jump size

It is expected that larger jumps in a time series improve the performance of classification algorithms, as discontinuities become more distinguishable from the natural variability of the samples. In this study, when the Random Forest algorithm, previously trained with 1 cm discontinuities, was applied to samples containing jumps of 1 cm, 2 cm, and 3 cm, performance metrics improved with increasing jump magnitude. The F1-score increased from 0.78 to 0.86, driven primarily by a substantial rise in True Positives (TP), from 39.6% to 67.3% (Figure 8). Notably, the proportion of False Negatives (FN) remained nearly constant at approximately 20%. This behavior suggests that, despite the larger magnitude of the discontinuities, the trained algorithm was unable to fully capture the patterns associated with the jumps, indicating a potential limitation in the representativeness of the training set.

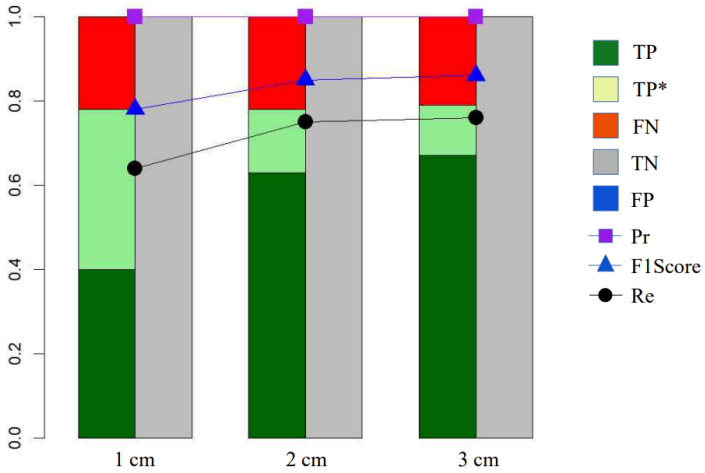
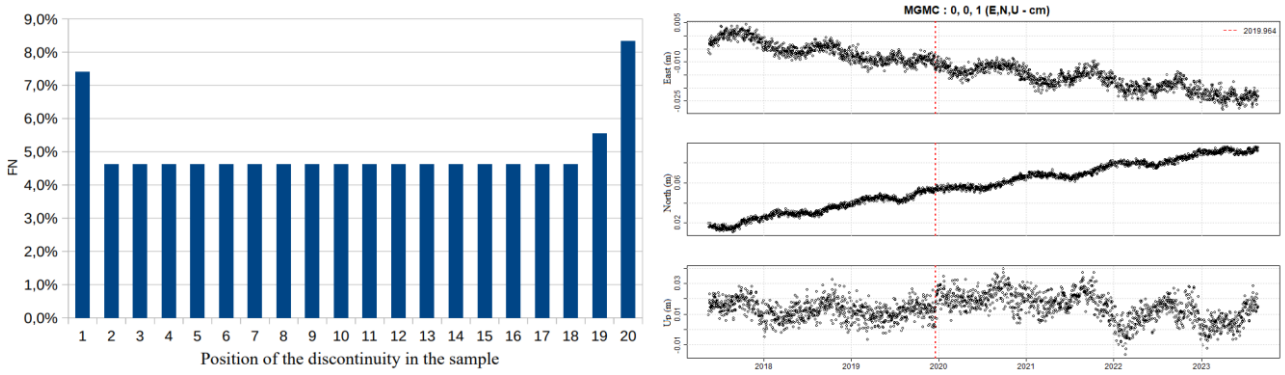


Figure 8 – Performance indicators of the Random Forest algorithm tested in the classification of different discontinuity magnitudes. Source: Authors (2024).

4.3. Discontinuities not detected in the samples

To investigate why approximately 20% of the discontinuities present in the samples were not identified by the Random Forest algorithm, an analysis focused on the position of the jumps within the time windows was performed. It was found that discontinuities located at the ends of the samples were more difficult to detect, due to the insufficient number of values before or after the jump, which compromises the identification of the trained pattern (Figure 9a). However, samples with jumps positioned near the center were also observed to be misclassified by the algorithm. In these cases, the number of samples containing a jump remains constant, suggesting that the discontinuities are not sufficiently clear in the series. An illustrative example is shown in Figure 9b for the MGMC station (Montes Claros, MG), where a 1 cm jump was inserted in the vertical component on December 18th, 2019. In this specific case, a higher dispersion of the Up component data can be observed, which makes it difficult to detect the discontinuity.



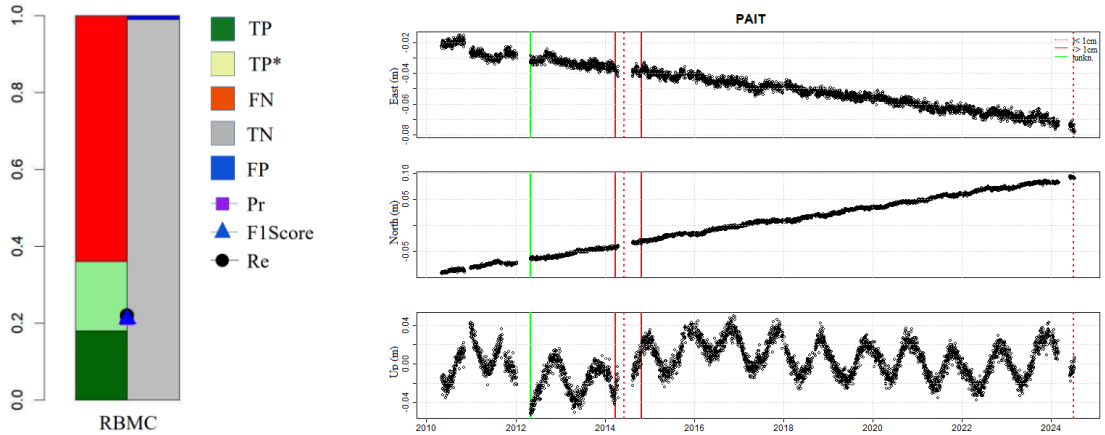
(a) (b)

Figure 9 – (a): Distribution of the position of FN within the samples, indicating greater concentration at the extremities. (b): A 1 cm discontinuity in the vertical component of the MGMC station (Montes Claros, MG) not identified by the classification algorithm, most likely due to data dispersion within the samples. Source: Authors (2024).

4.4. Performance of discontinuity classification in RBMC coordinate time series with real events

To assess the capability of the trained model in identifying real discontinuities in the RBMC time series, 37 stations presenting jumps larger than 1 cm due to antenna replacements were selected, totaling 47 events. As expected, the performance indicators obtained were lower than those reported in Section 4.1, with an F1-Score of 0.21 and approximately

36% of discontinuities correctly identified in the samples (TP + TP*). However, discontinuities detected by the model but not previously documented as events exceeding 1 cm were also observed. Although classified as False Positives (FP), such occurrences should not be automatically interpreted as algorithmic failures, since they may correspond to real but undocumented events in the series. It is noteworthy, however, that when present in the training data, these FP may compromise the model's learning effectiveness. Figure 10a presents the performance indicators of the classification against real discontinuities. Figure 10b illustrates the time series of the PAIT station (Itaituba, PA), highlighting known discontinuities larger than 1 cm (solid red lines), smaller than 1 cm (dashed red line), and an additional discontinuity identified by the classification whose origin remains unknown (solid green line).

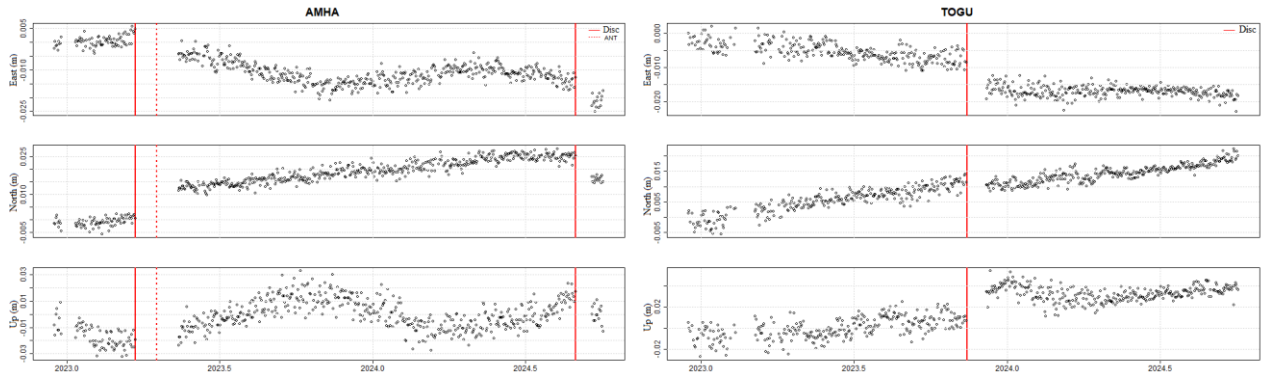


(a) (b)

Figure 10 – (a): Performance indicators of the machine learning classification considering real discontinuities in the RBMC time series; (b) discontinuity equal to or greater than 1 cm (solid red line) identified at the PAIT station (Itaituba-PA) between April 19th, 2012 and May 1st, 2012, smaller than 1 cm (dashed red line), and without an associated antenna replacement record (solid green line). Source: Authors (2024).

4.5. Identification of discontinuities in coordinate time series from IBGE

The trained algorithm can be applied to any solution involving the systematic estimation of coordinates, even if they are in different reference frames. The essential requirement is that the time series remain homogeneous and present comparable quality. In this study, we employed the Random Forest model, trained with NGL data, for the detection of discontinuities in daily solutions estimated by IBGE. These solutions, aligned to IGS20 frame, have been available since November 27, 2022, and are determined through relative network processing using the Bernese V5.4 software (DACH et al., 2015), as part of the internal routine of the Geodesy Coordination for data quality control. Similar to what was observed in the NGL series, the algorithm was also able to identify discontinuities associated with antenna replacements in the IBGE solutions, as in the cases of the AMHA (Humaitá, AM) and TOGU (Gurupi, TO) stations, illustrated in Figure 11. For the AMHA station in particular, the first discontinuity identified does not coincide exactly with the antenna replacement date. This discrepancy is due to the absence of data in the days immediately preceding the event, which prevented the algorithm from detecting the jump at the exact time it occurred.



(a) (b)

Figure 11 – Discontinuities identified in IBGE-estimated time series: (a) two discontinuities related to antenna replacements at AMHA (Humaitá, AM), with the first detected on a date not coinciding with the epoch of the event; (b) discontinuity at TOGU (Gurupi, TO) identified at the same time as the antenna replacement. Source: Authors (2024).

5. Conclusion and recommendation

The observations from the RBMC allow for the estimation of daily station coordinates with millimeter-level precision. This level of detail enables the construction of highly refined time series, providing valuable information such as the velocity and direction of lithospheric plate motion. To ensure the temporal consistency of these coordinates and, when necessary, to define multiple solutions, it is essential to account for disruptive events that may affect station positions, such as antenna replacements and earthquakes. In this study, we assess the capability of machine learning algorithms to identify discontinuities in RBMC time series caused by antenna replacements. For this purpose, we used series produced by the Nevada Geodetic Laboratory (NGL) at the University of Nevada, Reno, along with information on the date of earthquakes and antenna replacement events at each station.

The use of real discontinuities generated by antenna replacements in the training of the algorithm was not possible due to the absence of data immediately preceding the event epochs. To address this issue, criteria were established for the selection of time periods that would allow the inclusion of the largest possible number of RBMC stations, while ensuring data representativeness and the presence of seasonal effects observed in the series. In total, 108 stations were selected, each of which was randomly assigned at least one discontinuity of 1 cm. These data were organized into temporal samples generated by a 21-day moving window applied across the entire series for each coordinate component, thus constructing the study matrix and simultaneously defining the target vector, which indicated the presence (or absence) of a discontinuity and its position within the sample.

Among the machine learning algorithms evaluated [Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Naive Bayes (NB)], Random Forest achieved the best performance, with an F1-Score of 0.78. The model correctly identified 77.5% of the discontinuities in the samples (TP + TP*), with 39.6% detected exactly at the time of their occurrence (TP). When the magnitude of the jumps was increased from 1 cm to 3 cm, the percentage of correctly identified TP increased to 67.3% and the F1-Score to 0.86. However, approximately 22% of the discontinuities were not classified (FN), highlighting the need for a more representative dataset for model training. The trained algorithm was also applied to RBMC series containing real discontinuities, achieving an F1-Score of 0.21 and identifying approximately 36% of the jumps (TP + TP*). Although the effectiveness was lower than that obtained with simulated discontinuities, this result was expected, given that part of the data immediately preceding antenna replacement events is either missing or of low quality. The same evaluation was conducted on the series produced by the IBGE as part of the RBMC quality control framework, confirming the strong potential of the model for practical applications, as it was also capable of identifying discontinuities in these solutions.

This study demonstrated the capability of a machine learning model to identify discontinuities in time series associated with antenna replacements at RBMC stations. However, considering that algorithms learn to recognize patterns from data and that discontinuities represent only a small fraction of the overall observations, further research is recommended, focusing on training the model with a broader and more diverse set of stations. To this end, the incorporation of data from other GNSS networks, such as SIRGAS-CON, IGS, NOAA CORS, EUREF, among others, is suggested.

Acknowledgments

The authors thank the Nevada Geodetic Laboratory (NGL) and the Brazilian Institute of Geography and Statistics (IBGE) for providing the coordinate time series of the RBMC stations.

References

ALI, Jehad; KHAN, Rehanullah; AHMAD, Nasir; MAQSOOD, Imran. Random forests and decision trees. *International Journal of Computer Science Issues*, v. 9, n. 5, p. 272–278, set. 2012. Available at: http://www.ijcsi.org/papers/IJCSI-9-5-3-272-278.pdf.

ALTAMIMI, Z.; REBISCHUNG, P.; COLLILIEUX, X.; MÉTIVIER, L.; CHANARD, K. *ITRF2020* [Data set]. IERS ITRS Center hospedado por IGN e IPGP, 2022. Available at: https://doi.org/10.18715/IPGP.2023.LDVIOBNL.

ALTAMIMI, Z.; REBISCHUNG, P.; COLLILIEUX, X. et al. ITRF2020: an augmented reference frame refining the modeling of nonlinear station motions. *Journal of Geodesy*, v. 97, p. 47, 2023. Available at: https://doi.org/10.1007/s00190-023-01738-w.

BLEWITT, G.; HAMMOND, W. C.; KREEMER, C. Harnessing the GPS data explosion for interdisciplinary science. *Eos*, v. 99, 2018. Available at: https://doi.org/10.1029/2018EO104623.

COSTA, S.; LIMA, M.; MOURA JR, N.; DE ABREU, M.; SILVA, A.; FORTES, L.; RAMOS, A. RBMC in real time via NTRIP and its benefits in RTK and DGPS surveys. 2012. Available at: https://doi.org/10.1007/978-3-642-20338-1_115.

CROCETTI, L.; SCHARTNER, M.; SOJA, B. Discontinuity detection in GNSS station coordinate time series using machine learning. *Remote Sensing*, v. 13, p. 3906, 2021. Available at: https://doi.org/10.3390/rs13193906.

DACH, R.; LUTZ, S.; WALSER, P.; FRIDEZ, P. *Bernese GNSS Software Version 5.2: user manual*. Bern: Bern Open Publishing, 2015. ISBN 978-3-906813-05-9. DOI: https://doi.org/10.7892/boris.72297.

DAWIDOWICZ, K.; KRZAN, G.; WIELGOSZ, P. Offsets in the EPN station position time series resulting from antenna/radome changes: PCC type-dependent model analyses. *GPS Solutions*, v. 27, p. 9, 2023. Available at: https://doi.org/10.1007/s10291-022-01339-8.

DE DIEGO, I. M.; REDONDO, A. R.; FERNÁNDEZ, R. R. et al. General performance score for classification problems. *Applied Intelligence*, v. 52, p. 12049–12063, 2022. Available at: https://doi.org/10.1007/s10489-021-03041-7.

GAO, W.; LI, Z.; CHEN, Q. et al. Modelling and prediction of GNSS time series using GBDT, LSTM and SVM machine learning approaches. *Journal of Geodesy*, v. 96, art. 71, 2022. Available at: https://doi.org/10.1007/s00190-022-01662-5.

GAO, W.; WANG, C.; FENG, Y. A machine-learning-based missing data interpolation method for GNSS time series. In: YANG, C.; XIE, J. (org.). *China Satellite Navigation Conference (CSNC 2024) Proceedings*. Singapore: Springer, 2024. (Lecture Notes in Electrical Engineering, v. 1092). Available at: https://doi.org/10.1007/978-981-99-6928-9_20.

HEYDARIAN, M.; DOYLE, T. E.; SAMAVI, R. MLCM: multi-label confusion matrix. *IEEE Access*, v. 10, p. 19083–19095, 2022. Available at: https://doi.org/10.1109/ACCESS.2022.3151048.

HUANG, J.; HE, X.; HU, S.; MING, F. Impact of offsets on GNSS time series stochastic noise properties and velocity estimation. *Advances in Space Research*, v. 75, n. 4, p. 3397–3413, 2025. Available at: https://doi.org/10.1016/j.asr.2024.12.016.

IBGE. Brazilian Institute of Geography and Statistics, 2024. Available at: https://geoftp.ibge.gov.br/informacoes_sobre_posicionamento_geodesico/rbmc/relatorio/log_sirgas/. Accessed on Oct. 24, 2024.

JOHNSTON, G.; RIDDELL, A.; HAUSLER, G. The International GNSS Service. In: TEUNISSEN, P. J.; MONTENBRUCK, O. (eds.). *Springer Handbook of Global Navigation Satellite Systems*. Cham: Springer, 2017. p. 739–776. Available at: https://doi.org/10.1007/978-3-319-42928-1_33.

LAHTINEN, S.; JIVALL, L.; HÄKLI, P. et al. Updated GNSS velocity solution in the Nordic and Baltic countries with a semi-automatic offset detection method. *GPS Solutions*, v. 26, p. 9, 2022. Available at: https://doi.org/10.1007/s10291-021-01194-z.

LE, N.; MÄNNEL, B.; BUI, L. K. et al. Identifying neotectonic motions in Germany using discontinuity-corrected GNSS data. *Pure and Applied Geophysics*, v. 181, p. 87–108, 2024. Available at: https://doi.org/10.1007/s00024-023-03390-z.

LORENA, A. C.; DE CARVALHO, A. C. P. L. F. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Available at: https://doi.org/10.22456/2175-2745.5690.

ÖZBEY, V.; ERGINTAV, S.; TARI, E. GNSS time series analysis with machine learning algorithms: a case study for Anatolia. *Remote Sensing*, v. 16, p. 3309, 2024. Available at: https://doi.org/10.3390/rs16173309.

RAJPUT, D.; WANG, W. J.; CHEN, C. C. Evaluation of a decided sample size in machine learning applications. *BMC Bioinformatics*, v. 24, p. 48, 2023. Available at: https://doi.org/10.1186/s12859-023-05156-9.

ROKACH, L.; MAIMON, O. Decision Trees. In: MAIMON, O.; ROKACH, L. (eds.) *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer, 2005. Available at: https://doi.org/10.1007/0-387-25465-X_9.

SAVCHUK, S.; DOSKICH, S.; GOŁDA, P.; RURAK, A. The Seasonal Variations Analysis of Permanent GNSS Station Time Series in the Central-East of Europe. *Remote Sensing*, v. 15, p. 3858, 2023. Available at: https://doi.org/10.3390/rs15153858.

SIEMERS, F.; FELDMANN, C.; BAJORATH, J. Minimal data requirements for accurate compound activity prediction using machine learning methods of different complexity. *Cell Reports Physical Science*, 2022. Available at: https://doi.org/10.1016/j.xcrp.2022.101113.

SIEMURI, A.; SELVAN, K.; KUUSNIEMI, H.; VÄLISUO, P.; ELMUSRATI, M. A systematic review of machine learning techniques for GNSS use cases. *IEEE Transactions on Aerospace and Electronic Systems*, p. 1–42, 2022. Available at: https://doi.org/10.1109/TAES.2022.3219366.

SUN, J. W.; DU, W. X.; SHI, N. C. A Survey of kNN Algorithm. *Information Engineering and Applied Computing*, v. 1, Article ID: 770, 2018. Available at: https://doi.org/10.18063/ieac.v1i1.770.

TUNINI, L.; MAGRIN, A.; ROSSI, G.; ZULIANI, D. Global Navigation Satellite System (GNSS) time series and velocities about a slowly convergent margin processed on high-performance computing (HPC) clusters: products and robustness evaluation. *Earth System Science Data*, v. 16, p. 1083-1106, 2024. Available at: https://doi.org/10.5194/essd-16-1083-2024.

WANNINGER, L. Correction of apparent position shifts caused by GNSS antenna changes. *GPS Solutions*, v. 13, p. 133-139, 2009. Available at: https://doi.org/10.1007/s10291-008-0106-z.

WEBB, G. I. Naïve Bayes. In: SAMMUT, C.; WEBB, G. I. (eds.) *Encyclopedia of Machine Learning*. Boston, MA: Springer, 2011. Available at: https://doi.org/10.1007/978-0-387-30164-8_576.