ASTRA: FERRAMENTA COMPUTACIONAL PARA PROCESSAMENTO E ANÁLISE SISMO-IONOSFÉRICA DE DADOS GNSS/TEC POR PCA

Laura Luiz Trigo, Kauê Faria Messias da Silva, Virgínia Klausner.

Universidade do Vale do Paraíba/Instituto de Pesquisa e Desenvolvimento, Avenida Shishima Hifumi, 2911, Urbanova - 12244-000 - São José dos Campos-SP, Brasil, ltrigo@univap.br, kauefaria17@gmail.com, virginia@univap.br.

Resumo

Neste trabalho, foi desenvolvido o ASTRA, uma ferramenta computacional em MATLAB destinada a integrar e automatizar o processamento e a análise de dados GNSS/TEC. O sistema permite organizar séries de TEC, selecionar períodos de análise, controlar dados ausentes, formar matrizes, aplicar a Análise de Componentes Principais (PCA) e realizar a interpolação espacial dos resultados. Como caso de referência, foram analisados dados da rede RAMSAC associados ao terremoto ocorrido em 22 de agosto de 2025, na região sul da Passagem de Drake. Os resultados mostraram que, em condições geomagneticamente calmas, a PCA1 concentrou 95,67% da variância, enquanto no dia do evento sua participação foi reduzida para 35,79%, com maior contribuição das componentes subsequentes. Os mapas das componentes principais apresentaram diferentes padrões de distribuição espacial, com uma estrutura mais localizada observada na PCA3. O ASTRA mostrou-se capaz de integrar as principais etapas do processamento e da análise dos dados GNSS/TEC, estabelecendo uma base para sua aplicação em diferentes períodos, eventos e conjuntos de estações.
Palavras-chave: GNSS/TEC. Análise de Componentes Principais. Ionosfera. Terremotos. Processamento de dados.

Área do Conhecimento: Engenharias.

Introdução

O monitoramento do Conteúdo Eletrônico Total (TEC) por redes de GNSS (*Global Navigation Satellite System*) permite investigar a variabilidade espacial e temporal da ionosfera e identificar perturbações associadas a diferentes fenômenos geofísicos. Eventos sísmicos de grande magnitude podem produzir perturbações atmosféricas capazes de alcançar a ionosfera e alterar temporariamente sua densidade eletrônica, gerando assinaturas detectáveis nas observações de dados GNSS (Heki; Ping, 2005; Pulinets; Ouzounov, 2011). Entretanto, a ionosfera também apresenta variações naturais, solares e geomagnéticas, o que dificulta a identificação de alterações potencialmente relacionadas aos terremotos e torna necessária a utilização de técnicas capazes de organizar e analisar grandes conjuntos de dados observacionais.

Nesse contexto, a Análise de Componentes Principais (PCA) constitui uma abordagem estatística para decompor conjuntos de variáveis correlacionadas em componentes ordenadas conforme a parcela de variância explicada, permitindo destacar os padrões dominantes presentes nas séries ionosféricas (MathWorks, 2026). Essa técnica é particularmente útil no presente estudo devido à quantidade de dados provenientes das diferentes estações GNSS, que possuem séries temporais de TEC e apresentam comportamentos que podem estar relacionados entre si. A aplicação da PCA permite, assim, organizar essas informações e representar a variabilidade dos dados por meio de um número reduzido de componentes, facilitando a identificação dos padrões mais relevantes para a análise das perturbações ionosféricas.

O desenvolvimento do ASTRA surgiu da necessidade de reunir e tornar mais prática a execução das etapas envolvidas na análise dos dados ionosféricos. Inicialmente, o processamento era realizado por meio de diferentes algoritmos para aplicação da PCA aos dados de estações GNSS. Posteriormente, a abordagem foi adaptada aos dados do terremoto de magnitude 7,5 ocorrido em 22 de agosto de 2025, às 02:16 (UTC), na região sul da Passagem de Drake (Estados Unidos, 2025), utilizando estações da rede RAMSAC. Essa experiência evidenciou a necessidade de reunir os algoritmos em um único ambiente, dando origem ao desenvolvimento do ASTRA no MATLAB App Designer. A ferramenta centraliza etapas posteriores à obtenção do TEC, incluindo seleção dos períodos de análise, formação das matrizes, aplicação da PCA e interpolação espacial. Após a criação do executável, o programa pode ser distribuído como aplicação independente para usuários que não possuam o MATLAB, mediante o MATLAB Runtime (MathWorks, 2026). Dessa forma, o objetivo deste

trabalho é apresentar o desenvolvimento e a aplicação do ASTRA como ferramenta computacional para auxiliar a análise de perturbações sismo-ionosféricas a partir de dados GNSS/TEC.

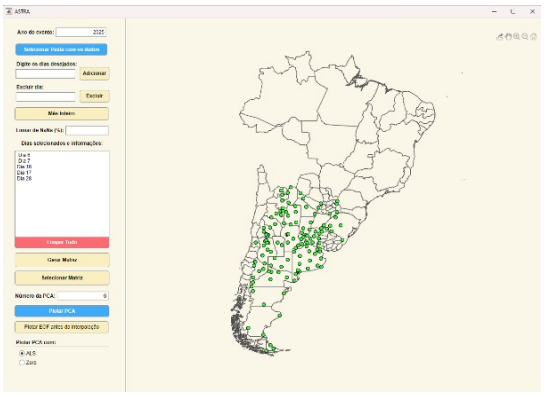
Metodologia

Foram utilizados dados de TEC provenientes de estações GNSS da rede RAMSAC, referentes ao período associado ao terremoto de magnitude 7,5 ocorrido em 22 de agosto de 2025, na região sul da Passagem de Drake (Estados Unidos, 2025). Inicialmente, os dados observacionais GNSS foram obtidos em formato RINEX e posteriormente processados pelo software GPS-TEC, desenvolvido por Seemala (2020), responsável pela obtenção das séries temporais de TEC e pela aplicação das correções necessárias aos dados, incluindo a correção dos efeitos instrumentais. Para as estações da rede RAMSAC, os arquivos RINEX foram obtidos por meio de uma rotina desenvolvida em Python, enquanto os dados de outras redes podem ser obtidos diretamente dos respectivos servidores. Após o processamento no GPS-TEC, as séries de TEC foram organizadas em arquivos .txt com resolução temporal de um minuto.

Os arquivos de TEC gerados pelo GPS-TEC foram posteriormente analisados no UTECDA (*Univap Total Electron Content Data Analysis*), utilizado para a visualização e verificação preliminar das séries temporais das estações GNSS. Nessa etapa, foram avaliados os gráficos de TEC e de desvio padrão (STD), construídos a partir dos dias geomagneticamente calmos selecionados para o período de análise com base nos valores do índice geomagnético Kp, obtidos no site do Kyoto World Data Center (WDC) (WORLD DATA CENTER FOR GEOMAGNETISM, KYOTO, 2026). Essa verificação permitiu observar a disponibilidade e a qualidade das séries temporais e identificar estações com ausência significativa de dados. As estações consideradas válidas após essa verificação foram então utilizadas como entrada para o ASTRA.

A partir dos arquivos previamente verificados no UTECDA, inicia-se o processamento no ASTRA. O usuário seleciona, por meio da interface gráfica, o diretório contendo os arquivos .txt das estações. A partir dos nomes dos arquivos, o programa identifica automaticamente o ano de análise e as estações disponíveis, utilizando a sigla da estação presente no nome do arquivo, e apresenta sua distribuição espacial na interface, permitindo ao usuário verificar visualmente a cobertura da rede antes de gerar a matriz. Cada estação é representada no mapa por um marcador, inicialmente identificado pela cor verde. A representação espacial permite ainda realizar a seleção manual das estações que deverão ser removidas da análise. Para isso, o usuário pode selecionar diretamente o marcador correspondente à estação no mapa, fazendo com que sua identificação seja registrada pelo sistema e que sua respectiva coluna seja posteriormente excluída da matriz de dados. Dessa forma, a remoção manual das estações, que anteriormente exigia a alteração de vetores de identificadores diretamente no código-fonte, passa a ser realizada pela própria interface gráfica.

Figura 1 - Interface do ASTRA com a distribuição espacial das estações GNSS identificadas a partir dos arquivos selecionados.



Fonte: Os autores (2026).

Após a identificação das estações, os arquivos são então lidos individualmente e submetidos a uma etapa de normalização, na qual a linha de cabeçalho é desconsiderada, os valores decimais são convertidos para o formato numérico utilizado pelo MATLAB e os valores inválidos representados por -999.99 são substituídos por NaN (*Not a Number*). As três primeiras colunas dos arquivos, correspondentes às informações de média, média acrescida do desvio padrão e média subtraída do

desvio padrão, são removidas da estrutura utilizada no processamento, permanecendo as séries de VTEC necessárias às etapas seguintes.

Antes da matriz ser concatenada, o usuário pode, opcionalmente, definir um limiar de dados ausentes para controlar quais estações serão mantidas no processamento. O limiar estabelece a porcentagem máxima de valores NaN permitida em cada série temporal. Quando um valor é informado, as estações que ultrapassam o limite definido são excluídas automaticamente da matriz, para evitar que séries com elevada quantidade de dados ausentes sejam utilizadas nas etapas posteriores do processamento. Os valores apresentados na tabela abaixo são exemplos de limiares que podem ser utilizados, como 0%, 5%, 10%, 20% ou 50%. A tabela também apresenta o critério correspondente a cada valor.

Tabela 1 - Critérios de exclusão de estações em função do percentual de dados ausentes.

Limiar	Estilo
0%	Remove estações com qualquer valor ausente
5%	Remove estações com mais de 5% de valores ausentes
10%	Remove estações com mais de 10% de valores ausentes
20%	Remove estações com mais de 20% de valores ausentes
50%	Remove estações com mais de 50% de valores ausentes

Fonte: Os autores (2026).

Após a aplicação do critério de dados ausentes, os arquivos das estações mantidas são concatenados para formar a matriz utilizada nas etapas posteriores do processamento. Nesse procedimento, as séries temporais correspondentes aos diferentes dias do período selecionado são organizadas sequencialmente, de modo que os dados de um dia são posicionados imediatamente após os dados do dia anterior, formando uma única série temporal contínua para cada estação. Assim, considerando um período de 31 dias e resolução temporal de um minuto, cada estação contribui com $31 \times 1.440 = 44.640$ observações, enquanto cada estação corresponde a uma coluna da matriz. Dessa forma, a matriz final reúne, em uma única estrutura, todas as observações temporais das estações selecionadas, sendo posteriormente utilizada no cálculo do dTEC e na aplicação da PCA.

Posteriormente à criação da matriz de dTEC, o ASTRA permite definir o tratamento dos valores ausentes antes da aplicação da PCA. O usuário pode optar pela substituição dos valores NaN por zero ou pelo método *Alternating Least Squares* (ALS). A utilização de métodos baseados em ALS é uma alternativa para a análise de dados incompletos por PCA, conforme discutido por Gómez-Sánchez et al. (2024). A substituição por zero é mantida como uma opção adicional da ferramenta para situações em que a quantidade ou distribuição dos dados ausentes inviabiliza a estimativa pelo ALS. Após o tratamento, a PCA é aplicada à matriz resultante.

Durante a etapa de interpolação espacial, o ASTRA realiza uma extensão do domínio por simetrização espacial. Para isso, as posições das estações e os respectivos valores da componente principal são replicados por reflexão em relação aos limites do conjunto original, formando cópias ao redor do domínio de análise. Essa extensão é utilizada exclusivamente como recurso numérico para reduzir os efeitos de borda durante a interpolação pelo método *Natural Neighbor* (Gary, 2024). Dessa forma, as réplicas não correspondem a novas estações ou a novas observações da ionosfera, mas constituem uma extensão espacial dos dados originais utilizada para auxiliar a interpolação e a representação gráfica dos resultados da PCA.

Resultados

A Figura 2 apresenta a interface principal do ASTRA desenvolvida no MATLAB App Designer. O sistema reúne, em um único ambiente, controles para configurar o processamento, incluindo seleção do ano e do diretório dos dados, definição dos dias de interesse, exclusão de períodos, escolha do limite aceitável de dados ausentes e seleção do método de tratamento dos valores NaN. A interface também permite selecionar a componente principal a ser analisada e visualizar espacialmente as estações GNSS utilizadas. Dessa forma, etapas que anteriormente dependiam da execução individual de diferentes rotinas passaram a ser acessadas por meio de um fluxo único e configurável.

Figura 2 - Interface principal do ASTRA para configuração, processamento e análise dos dados GNSS/TEC.



Fonte: Os autores (2026).

Como caso de referência utilizado no desenvolvimento e na avaliação do ASTRA, foram considerados os resultados obtidos para o setor argentino durante o período associado ao evento sísmico. Nos dias geomagneticamente calmos, a primeira componente principal concentrou 95,67% da variância total, enquanto as componentes PCA2, PCA3 e PCA4 representaram, respectivamente, 4,12%, 0,12% e 0,05%. No dia do evento sísmico, a PCA1 representou 35,79% da variância, enquanto a PCA2 apresentou 31,26%, a PCA3 13,27% e a PCA4 10,89%. Assim, as quatro primeiras componentes concentraram 91,21% da variância no dia do evento, com maior participação relativa dos modos secundários em comparação à condição geomagneticamente calma.

Tabela 2 - Variância explicada pelas quatro primeiras componentes principais no setor argentino.

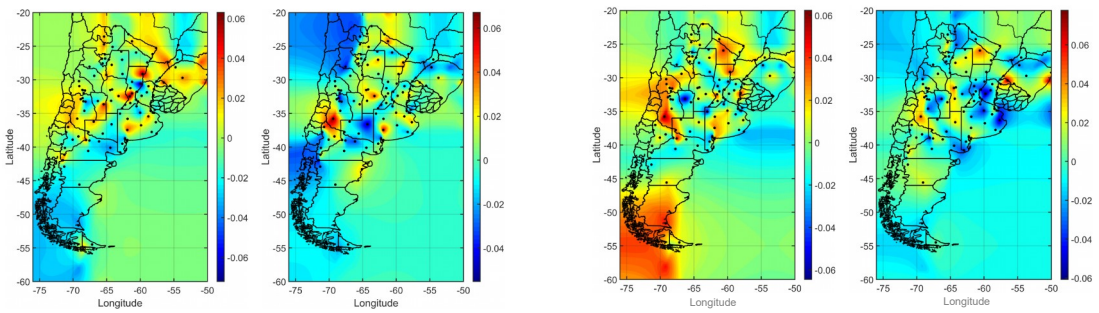
Componentes	**Dias calmos (%)**	**Dia do evento (%)**
PCA1	95,67	35,79
PCA2	4,12	31,26
PCA3	0,12	13,27
PCA4	0,05	10,89
Total acumulado	99,96	91,21

Fonte: Os autores (2026).

A comparação entre o dia do evento sísmico e os dias geomagneticamente calmos foi realizada com o objetivo de estabelecer uma condição de referência para o comportamento natural da ionosfera. Os dias calmos, por apresentarem menor influência de perturbações geomagnéticas intensas, permitem caracterizar de forma mais representativa a variabilidade ionosférica esperada na ausência de condições geomagnéticas significativamente perturbadas. Dessa forma, a análise do dia do terremoto em relação a essa linha de base de referência permite identificar alterações na distribuição espacial das componentes principais que se diferenciem do comportamento observado em condições geomagneticamente mais estáveis.

A Figura 3 apresenta a distribuição espacial das quatro primeiras componentes principais obtidas para o setor argentino no dia do terremoto de 22 de agosto de 2025. A PCA1 apresenta distribuição mais abrangente, enquanto as componentes de ordem superior apresentam estruturas mais regionalizadas, destacando-se a PCA3 pela concentração de um padrão na região do epicentro.

Figura 3 - Distribuição espacial das quatro primeiras componentes principais para o setor argentino no dia do terremoto de 22 de agosto de 2025.



Fonte: Os autores (2026).

Discussão

A concentração de 95,67% da variância na PCA1 durante os dias geomagneticamente calmos indica que, nessas condições, grande parte da variabilidade ionosférica do setor argentino pôde ser explicada, majoritariamente, por um único modo dominante. No dia do terremoto, a redução da participação da PCA1 e o aumento relativo das componentes seguintes mostram uma redistribuição da variabilidade entre diferentes modos espaciais. Essa comparação não permite atribuir isoladamente ao terremoto a origem das diferenças observadas, uma vez que a ionosfera está sujeita simultaneamente a variações solares, geomagnéticas e atmosféricas. Assim, os dias geomagneticamente calmos são utilizados como linha de base como referência para a comparação com o comportamento observado durante o período do evento.

A interpretação quantitativa apresentada deve ser considerada no contexto do conjunto de dados analisado. Como o estudo de caso contempla um único evento sísmico, os resultados não permitem estabelecer generalizações sobre a resposta da PCA a terremotos de diferentes magnitudes ou sob diferentes condições ionosféricas e geomagnéticas. A realização de análises estatísticas adicionais e a aplicação do ASTRA a um conjunto ampliado de eventos constituem etapas futuras para avaliar a estrutura dos padrões identificados.

Quanto ao tratamento de dados ausentes, a versão atual do ASTRA mantém a substituição por zero como uma alternativa de processamento, além do método ALS. Entretanto, essa abordagem não foi avaliada quantitativamente quanto ao erro de reconstrução neste estudo. Dessa forma, a implementação de métricas como RMSE e MAE, juntamente com a comparação entre ALS, interpolação temporal e outras estratégias de preenchimento, constituem uma etapa necessária para avaliar os efeitos do tratamento dos dados ausentes sobre os resultados da PCA.

Os mapas obtidos para o dia do evento apresentaram diferentes estruturas espaciais entre as componentes analisadas. A PCA1 apresentou uma distribuição mais abrangente, enquanto componentes de ordem superior apresentaram padrões mais regionalizados, destacando-se a PCA3 pela concentração de uma estrutura em uma região específica do domínio analisado. Esses resultados mostram como a decomposição por PCA, associada à interpolação espacial, pode auxiliar na identificação e representação de diferentes padrões presentes nas séries de TEC.

A utilização de observações GNSS em estudos sismo-ionosféricos é relevante porque terremotos podem produzir alterações detectáveis no TEC. Estudos anteriores mostram que redes GNSS podem registrar distúrbios ionosféricos associados a eventos sísmicos (Heki; Ping, 2005; Pulinets; Ouzounov, 2011). Entretanto, a PCA não permite, isoladamente, atribuir essas alterações ao terremoto, sendo utilizada neste trabalho para decompor a variabilidade do TEC e identificar padrões que devem ser interpretados em conjunto com as condições geofísicas do período.

Como análise complementar, os diagramas tempo-distância (TTD) obtidos a partir das séries de STEC para os PRNs 1 e 3 apresentaram alinhamentos temporais coerentes entre diferentes estações, permitindo investigar possíveis padrões de propagação das perturbações observadas. Para o PRN1, foi estimada uma velocidade aparente de aproximadamente 1,95 km/s, enquanto os três alinhamentos identificados para o PRN3 resultaram em velocidades aparentes de aproximadamente 4,09, 2,45 e 1,33 km/s. Esses resultados complementam a análise espacial realizada pela PCA, embora sua interpretação como assinatura exclusivamente associada ao terremoto também dependa da consideração das demais condições ionosféricas e geomagnéticas do período.

Do ponto de vista computacional, a principal contribuição do ASTRA está na integração de etapas que anteriormente dependiam da execução e adaptação de diferentes rotinas. A seleção de períodos, organização das estações, cálculo do dTEC, controle de dados ausentes, consolidação matricial, aplicação da PCA e interpolação espacial passam a compor um fluxo único configurável pelo usuário. Essa integração reduz a intervenção manual e torna o processamento reutilizável para diferentes períodos e conjuntos de estações, sem a necessidade de reconstruir individualmente as etapas computacionais.

Conclusão

O ASTRA foi desenvolvido como uma ferramenta computacional em MATLAB para integrar e automatizar etapas de tratamento e análise de dados GNSS/TEC aplicadas à investigação sismo-ionosférica. A partir do estudo do terremoto de magnitude 7,5 ocorrido em 22 de agosto de 2025, na região sul da Passagem de Drake, o sistema passou a reunir em um único ambiente a seleção dos períodos de análise, organização das estações, cálculo do dTEC, controle de dados ausentes, consolidação matricial, aplicação da PCA e interpolação espacial dos resultados.

Embora os resultados obtidos demonstrem a aplicabilidade do ASTRA, a ferramenta ainda apresenta etapas de desenvolvimento necessárias para ampliar sua utilização. Atualmente, o funcionamento do programa foi testado para os conjuntos de estações utilizados nas análises do Brasil de 2011 e da Argentina de 2025, sendo necessário ampliar o cadastro para contemplar as demais estações disponíveis no UTECDA. Essa expansão permitirá tornar a ferramenta mais escalável e reduzir a dependência de configurações específicas para cada evento ou região. Também pretende-se incorporar métricas quantitativas para avaliar o desempenho dos diferentes métodos de tratamento de dados ausentes, permitindo comparar a reconstrução obtida pelo ALS com outras estratégias de preenchimento e avaliar seus efeitos sobre os resultados da PCA. Dessa forma, a versão atual do ASTRA estabelece uma base computacional funcional para a automatização do processamento GNSS/TEC, que poderá ser ampliada e aperfeiçoada em etapas futuras.

Referências

ESTADOS UNIDOS. United States Geological Survey. *M 7.5 - 2025 Southern Drake Passage Earthquake*, 2025. Earthquake Hazards Program. Disponível em: https://earthquake.usgs.gov/earthquakes/eventpage/us6000r2xq/executive. Acesso em: 22 ago. 2026.

GÓMEZ-SÁNCHEZ, A.; VITALE, R.; RUCKEBUSCH, C.; DE JUAN, A. *Solving the missing value problem in PCA by Orthogonalized-Alternating Least Squares* (O-ALS). Chemometrics and Intelligent Laboratory Systems, v. 250, 105153, 2024. DOI: 10.1016/j.chemolab.2024.105153. Disponível em: https://doi.org/10.1016/j.chemolab.2024.105153. Acesso em: 27 set. 2026.

GARY, W. *An introduction to natural neighbor interpolation*. [S. l.]: G. W. Lucas, 2024. Tinfour Projects. Disponível em: https://gwlucastrig.github.io/TinfourDocs/NaturalNeighborIntro/index.html. Acesso em: 22 ago. 2026.

HEKI, K.; PING, J. Directivity and apparent velocity of the coseismic ionospheric disturbances observed with a dense GPS array. *Earth and Planetary Science Letters*, v. 236, n. 3, p. 845-855, 2005. Disponível em: https://www.sciencedirect.com/science/article/pii/S0012821X05003742. Acesso em: 22 ago. 2026.

WORLD DATA CENTER FOR GEOMAGNETISM, KYOTO. *World Data Center for Geomagnetism, Kyoto*. Kyoto: Kyoto University, Graduate School of Science. Disponível em: https://wdc.kugi.kyoto-u.ac.jp/. Acesso em: 23 ago. 2026.

MATHWORKS. *PCA - principal component analysis*. Natick: MathWorks, 2026. MATLAB & Simulink Help Center. Disponível em: https://www.mathworks.com/help/stats/pca.html. Acesso em: 16 maio 2026.

PULINETS, S.; OUZOUNOV, D. Lithosphere-atmosphere-ionosphere coupling (LAIC) model - an unified concept for earthquake precursors validation. *Journal of Asian Earth Sciences*, v. 41, n. 4, p. 371-382, 2011. Disponível em: https://www.sciencedirect.com/science/article/pii/S1367912010000830. Acesso em: 23 ago. 2026.

SEEMALA, G. K. *GPS-TEC Analysis Application*, 2020. Disponível em: http://seemala.blogspot.com/. Acesso em: 23 ago. 2026.

Agradecimentos

Agradeço à minha orientadora, Prof.ª Dr.ª Virgínia Klausner, pela orientação, confiança e por todo o conhecimento compartilhado ao longo deste trabalho. Também agradeço ao meu colega de iniciação científica, Kauê Faria, por toda a ajuda e colaboração, que foram essenciais para o desenvolvimento deste trabalho. Por fim, ao Instituto de Pesquisa e Desenvolvimento (IP&D) e à Universidade do Vale do Paraíba (UNIVAP) pelo apoio e infraestrutura.