GeoTerrace-2026-052

Machine Learning-Based Identification of Factors Influencing GNSS Time Series Anomalies: A Case Study of the Dnister Hydropower Complex

O. Haidus, ***I. Brusak** *(Lviv Polytechnic National University)*

SUMMARY

This study extends the semi-automatic Isolation Forest based anomaly detection in GNSS time series towards identification of the driving factors of the detected anomalies. The method combines harmonic trend-seasonal modelling with one-step ARIMA innovations, an Isolation Forest detector with causal features of station motion and the hydrological regime, a catalogue of candidate driving events, and permutation-based statistical verification of the coincidences. The method is applied to six stations of the GeoTerrace and System.NET networks around the Dnister Hydropower Complex in 2020-2024. Non-tidal atmospheric loading is the dominant external driver of the vertical station motion, with a correlation of 0.44-0.50 in the 3–30-day band at all six stations and a 10-14% WRMS reduction after correction with the measured admittance. Significant loading events are traced directly in the daily solutions: the direction of the station displacement matches the modelled deformation in 79% of episodes. With hydrological data and loading features included in the detector feature space, the anomalies concentrate near seismicity: 21 of 22 anomalous days within the seismic catalogue coverage fall within three days of seismic events against a 47% chance level. The results form the basis for factor-based classification of GNSS anomalies in monitoring.

Keywords: GNSS time series; GeoTerrace network; Isolation Forest; Non-tidal loading; Reservoir-triggered seismicity; Dnister Hydropower Complex.

Corresponding Author: ivan.v.brusak@lpnu.ua

Introduction

In previous studies a semi-automatic method for anomaly detection in GNSS time series based on the Isolation Forest algorithm was proposed and demonstrated on 37 stations of the GeoTerrace network (Brusak et al., 2025; 2026). Some of the detected anomalies coincided with known events, such as loading episodes and documented equipment replacements, but the nature of a considerable share of the anomalies remained unexplained. For the Dnister Hydropower Complex area, a relationship between reservoir water level fluctuations and seismic activity, typical for reservoir-triggered seismicity, was established (Savchyn & Pronyshyn, 2020; Zyhar et al., 2023).

The aim of this study is to go beyond anomaly detection by identifying the factors that may drive the observed anomalies. The detected anomalies are compared with a catalogue of potential triggering events, while each association is evaluated against the expected rate of random coincidence and its statistical significance is assessed using permutation-based testing. The Dnister Hydropower Complex is the test area: the reservoir provides a controlled hydrological factor, the microseismic swarm beneath the dam a separate event class, and the ESMGFZ models (Dill and Dobslaw, 2013) an independent prediction of loading deformations.

Method and Theory

The input data are daily N, E, Up coordinate solutions of six GeoTerrace and System.NET stations around the Dnister Hydropower Complex and NDNS GNSS station near the dam (30.07.2020-31.12.2024 time series), the daily reservoir water level, and vertical deformations predicted by the ESMGFZ non-tidal loading models: atmospheric (NTAL), oceanic (NTOL) and hydrological (HYDL) (Dill and Dobslaw, 2013; Mémin et al., 2020; Nistor et al., 2021). Seismicity is described by a combined catalogue: local bulletins of the Carpathian network (Md, 173 events within 30 km) and the EMSC bulletin (ML, with the Dobrovolsky (1979) preparation-zone filter); 24 regional M4.0-5.2 USGS events (up to 350 km) produce no systematic response at NDNS. Each component is modelled in two ways. A structural model with ARMA errors and AIC-based order selection yields a deseasoned series in which multi-day physical responses remain visible (1):

$$x_t = a_0 + a_1 t + \sum_{k\in K}[b_k sin(\omega_k t) + c_k cos(\omega_k t)] + \eta_t, \quad \eta_t \sim ARMA(p, q) \qquad (1)$$

where ω_k are the annual, semi-annual and weekly angular frequencies; an ARIMA model additionally yields one-step innovations robust to trend changes. Anomalies are detected by an Isolation Forest of 800 trees (Liu et al., 2008; Pedregosa et al., 2011), which scores a day x by its mean isolation depth E[h(x)] over the ensemble (2):

$$s(x) = 2^{-\frac{E[h(x)]}{c(\psi)}} \qquad (2)$$

where c(ψ) is the path-length normalisation for the subsample size ψ. The feature space is causal (uses only past values) and combines the innovations, the velocity and acceleration of the deseasoned series, rolling statistics, hydrological regime features and atmospheric loading features; hyperparameters were selected by a label-free stability criterion over a grid of 72 configurations.

The event catalogue uses physically justified thresholds: loading event is considered significant when its expected response $\beta \cdot \Delta L$ exceeds the daily repeatability of the up component $\sigma_{Up} \approx 3.6$ mm (3):

$$|\Delta L_{3d}| \geq \frac{\sigma_{Up}}{\beta} \approx 7\,mm, \quad \beta \approx 0.4-0.5 \qquad (3)$$

while NTAL dominates the sub-monthly non-tidal vertical signal over continental Eurasia (Klos et al., 2021). Anomaly-event matching is performed in ±0...±7-day windows; every explained share is reported together with its background share, and significance is assessed by circular-shift permutation tests. An event-anchored verification is also performed: for each significant NTAL episode the station displacement direction is compared with the modelled one.

Results

Non-tidal atmospheric loading is the dominant external driver of the station motion. In the 3-30 day band the deseasoned Up component correlates with NTAL at 0.44-0.50 at all six stations (out-of-sample; significance confirmed with autocorrelation-corrected tests, $p<0.001$). The joint regression explains $R^2 \approx 0.19$-0.26 of the variance, almost entirely due to NTAL. Correcting the series with the measured admittance reduces the WRMS in this band by 10-14% (Figure 1). The admittance $\beta \approx 0.4$-0.5 is systematic across the network and is consistent with partial absorption of the common-mode signal by the network relative solution.

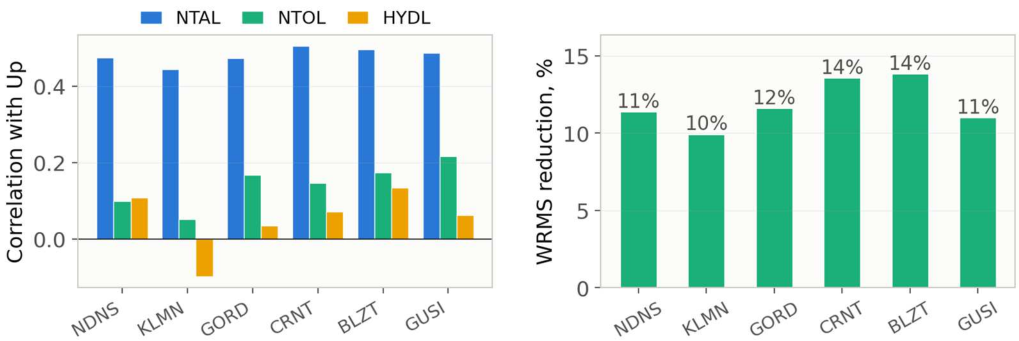

Figure 1. *Correlation of the deseasoned Up with the ESMGFZ loading models in the 3–30-day band (left) and the WRMS reduction after admittance correction (right)*

The event-anchored verification confirms that significant NTAL events are traced directly in the daily solutions: out of 34 episodes with $|\Delta\text{NTAL}| \geq 7$ mm, in 27 (79%) the direction of the station displacement matches the direction of the modelled deformation (binomial test against 50%, $p = 0.0004$; Figure 2). For rapid reservoir water level changes no direct deformation response is found (sign agreement at chance level), which points to an indirect, seismogenic character of the water influence.

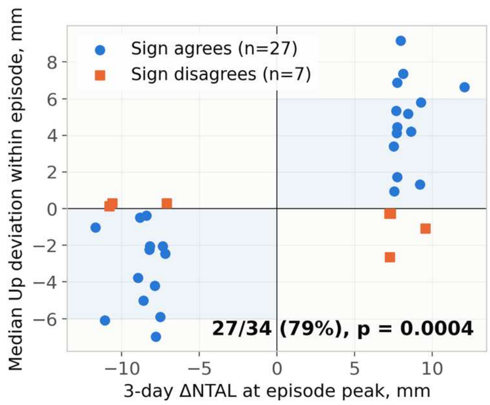

Figure 2. *Event-anchored verification - station displacement vs modelled deformation direction for 34 significant NTAL episodes*

The key detection-level result is obtained when hydrological and NTAL features are included in the detector feature space: the anomalies concentrate near seismicity (Table 1). At the 2% operating point all major anomaly episodes (26.11-15.12.2022, 02-07.02.2023, 07-10.10.2023) fall into periods of seismic activity, and the December 2022 episode coincides with the independently published network-wide vertical group anomaly of the GeoTerrace network (Haidus et al., 2024). Neither the GNSS nor the water features alone produce such concentration - it is the joint 'crustal motion + hydrological regime' feature space that is informative, in agreement with the reservoir-triggered seismicity mechanism (Zyhar et al., 2023). The anomalies outside the catalogue coverage (November 2023) emphasise the need for an updated 2024 microseismic catalogue (Figures 3 and 4).

Table 1. *Share of anomalous days within ±3 days of seismic events and circular-shift permutation p-values, for two operating points of the detector*

Detector feature space	*share (c = 2%)*	*p*	*share (c = 5%)*	*p*
GNSS only	42–52%	> 0.1	42–52%	> 0.1
GNSS + hydro level	19/19 (100%)	0.006	30/37 (81%)	0.008
GNSS + hydro level + NTAL	21/22 (95%)	0.008	34/43 (79%)	0.007

For the detector built on GNSS features alone the explained share of anomalies stays at the background level across the whole hyperparameter grid (Table 1): a considerable part of purely GNSS anomalies is driven by factors outside the catalogue - local and hardware effects and the documented NDNS vertical trend change in 2023-2024. Further work will incorporate station log files and the complete 2024 microseismic catalogue for factor-based classification.

Figure 3. *Reservoir water level (top) and the monthly number of seismic event days (bottom); shading - the period without microseismic catalogue coverage*

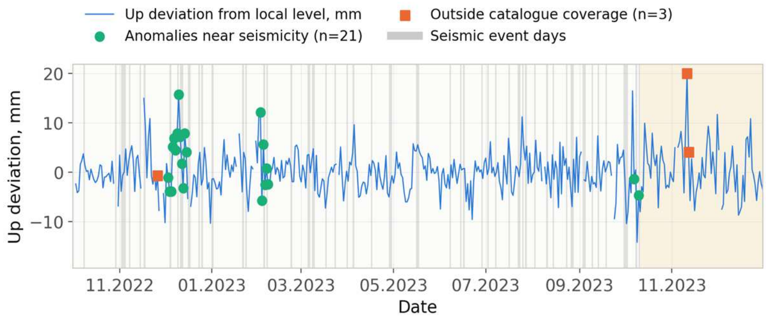

Figure 4. *Deseasoned Up deviation of NDNS with the anomalous days of the 2% operating point: green - within ±3 days of seismicity, orange - outside catalogue coverage; grey - seismic event days*

Conclusions

The method for GNSS anomaly detection with identification of the driving factors is proposed and tested. NTAL is the dominant external driver of the vertical station motion (correlation 0.44-0.50, $p<0.001$; 10-14% WRMS reduction), and its significant events are verified directly in the daily solutions (79% of episodes, $p=0.0004$). Including the hydrological regime and loading into the detector feature space concentrates the anomalies near seismicity (95% within the catalogue coverage, p=0.008), in line with the reservoir-triggered seismicity mechanism, and forms the basis of automated factor-based anomaly classification for near real-time monitoring of the GeoTerrace GNSS network.

Acknowledgements

The research was carried out within the framework of the DB/ GEOS project «Analysis of the Impact of Military Operations on Spatial Deformations of the Dnister Hydropower Complex Using Contact and Remote Methods» (No. 0125U000496).

References

Brusak, I., Haidus, O., & Kuplovskyi, B. (2026). Application of machine learning methods to study the relationships between seismicity and geodynamic features of the Dnister Hydropower Complex. Geofizychnyi Zhurnal, 48(2), 60-76. https://doi.org/10.24028/gj.v48i2.350985

Brusak, I., Maciuk, K., & Haidus, O. (2025). Detection of geodynamic anomalies in GNSS time series using machine learning methods. Geodynamics, 1(38), 37-48. https://doi.org/10.23939/jgd2025.01.037

Dill, R., & Dobslaw, H. (2013). Numerical simulations of global-scale high-resolution hydrological crustal deformations. Journal of Geophysical Research: Solid Earth, 118(9), 5008-5017. https://doi.org/10.1002/jgrb.50353

Dobrovolsky, I.P., Zubkov, S.I., & Miachkin, V.I. (1979). Estimation of the size of earthquake preparation zones. Pure and Applied Geophysics, 117(5), 1025-1044. https://doi.org/10.1007/BF00876083

Haidus, O., Maciuk, K., & Brusak, I. (2024). Semi-automatic anomaly detection in GNSS time series using machine learning algorithms. In International Conference of Young Professionals «GeoTerrace-2024». EAGE. https://doi.org/10.3997/2214-4609.2024510048

Klos, A., Dobslaw, H., Dill, R., & Bogusz, J. (2021). Identifying the sensitivity of GPS to non-tidal loadings at various time resolutions: examining vertical displacements from continental Eurasia. GPS Solutions, 25, 89. https://doi.org/10.1007/s10291-021-01135-w

Liu, F.T., Ting, K.M., & Zhou, Z.-H. (2008). Isolation Forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413-422). IEEE. https://doi.org/10.1109/ICDM.2008.17

Mémin, A., Boy, J.-P., & Santamaría-Gómez, A. (2020). Correcting GPS measurements for non-tidal loading. GPS Solutions, 24, 45. https://doi.org/10.1007/s10291-020-0959-3

Nistor, S., Suba, N. S., Maciuk, K., Kudrys, J., Nastase, E. I., & Muntean, A. (2021). Analysis of noise and velocity in GNSS EPN-repro 2 time series. Remote Sensing, 13(14), 2783. DOI: 10.3390/rs13142783

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830. DOI: 10.5555/1953048.2078195.

Savchyn, I., & Pronyshyn, R. (2020) Differentiation of recent local geodynamic and seismic processes of technogenic-loaded territories based on the example of Dnister Hydro Power Complex. Geodesy and Geodynamics, 11(5), 391—400. https://doi.org/10.1016/j.geog.2020.06.001.

Zyhar, A., Yushchenko, Yu., & Savchyn, I. (2023). A study of the influence of water level fluctuations on the geodynamic situation in the natural and technical geosystem of the Dnister HPP and PSPP cascade. Geodesy, cartography and aerial photography, (97), 24—31. DOI: 10.23939/istcgcap2023.97.024.