GeoTerrace-2026-061

Automated Processing of Daily GNSS Coordinate Time Series: Data Correction, Velocity Estimation and Grouping of Station Motion

***A. Paikush, I. Savchyn** *(Lviv Polytechnic National University)*

SUMMARY

The number of permanent GNSS stations, and the volume of daily solutions they produce, has grown to a point where station-by-station manual analysis is no longer practicable. This work presents an automated pipeline that takes daily coordinate series from the Nevada Geodetic Laboratory through five stages: acquisition and sorting by tectonic plate, noise filtering, data correction, velocity estimation, and unsupervised grouping of station motion. A linear velocity with annual and semi-annual harmonics is fitted by least squares, and the velocities are cross-checked against an independent plate-motion model used for verification only, which never filters or removes observations. The correction stage proves decisive, because undetected or wrongly imposed offsets propagate directly into velocity. Segmentation with a prescribed number of breakpoints is replaced by a penalised criterion; the surviving offsets are estimated jointly with the trend and seasonal terms and judged against the empirical level noise of each series rather than a white-noise standard error. On synthetic series with realistic white-plus-flicker noise the procedure leaves 0.18 spurious offsets per series instead of 30, detects offsets from 8 mm, and narrows the velocity error range from 7.58 to 1.24 mm/yr. On a 50-station Eurasian test set the median horizontal velocity falls from 8.27 to 3.58 mm/yr, while undisturbed stations stay unchanged.

Keywords: GNSS time series; offset detection; change point detection; least squares; station velocity; plate motion model; unsupervised clustering.

Corresponding Author: andrii.paikush.asp.2025@lpnu.ua

Introduction

Continuously operating GNSS stations are among the densest sources of geodetic information available. The Nevada Geodetic Laboratory (NGL) updates daily coordinates for roughly ten thousand stations every week and distributes them openly (Blewitt et al., 2018). A single tectonic plate accounts for tens of millions of epochs, so the classical workflow, in which an analyst inspects each series by hand, is no longer realistic.

Automation moves the difficulty rather than removing it. The dominant obstacle is the offset: an abrupt change of station position caused by an equipment change, a monument disturbance, or an earthquake. An offset left uncorrected biases the estimated velocity; one invented when none exists biases it just as effectively. The community benchmark found that manual solutions almost always outperform automated ones, that the smallest reliably detectable offsets are about 5 mm for the best manual and 8 mm for the best automated approaches, and that velocities below roughly 1 mm/yr are not robust (Gazeaux et al., 2013). Recent work still reports true-positive rates near 61 per cent against false positives near 30 per cent (Khazraei & Amiri-Simkooei, 2021).

A second difficulty is statistical. GNSS series are not white-noise processes; their spectrum is well described by white plus flicker noise (Williams et al., 2004), so low-frequency wander makes the level of a series drift by amounts a white-noise standard error declares highly significant, and any offset test built on that assumption reports discontinuities that are not there. Grouping stations by their motion then exposes regional structure (Takahashi & Hashimoto, 2022; Tretyak & Brusak, 2022; Savchyn, Brusak, & Tretyak, 2023; Savchyn, & Bilashuk, 2024). This work builds a pipeline covering the path from raw NGL files to grouped station motion, with emphasis on the correction stage, where the largest and least visible errors arise.

Data

The input is the NGL GlobalStationList, which provides daily topocentric coordinates in .tenv3 format: station code, epoch as a decimal year, the east, north, and up components in metres, their formal uncertainties, and the station coordinates. Global IGS20 series should reproduce the plate-motion prediction; plate-fixed series should show velocities close to zero. The Eurasian plate was chosen for development because it offers both many stations and a wide range of data quality. Its directory holds 8,698 .tenv3 files; requiring at least three years of observations leaves 6,941 stations, or 79.8 per cent, comprising about 29.6 million daily epochs. Series length ranges from 3.0 to 32.3 years, median 11.9, at a median density of 338 epochs per year. Quality is far from uniform, from campaign occupations of about ten epochs per year to a nearly complete daily record, and this heterogeneity matters more than length alone. A 50-station subset, weighted towards the most disturbed series, was used below.

Method and Theory

The processing is organised as five sequential blocks (Figure 1), with an independent verification branch attached to the velocity block.

Block 1, data acquisition. Stations are read from the GlobalStationList and downloaded in parallel for a chosen reference frame and file type, then written into directories named after the plate NGL assigns to each station; series shorter than three years are excluded.

Block 2, noise filtering. Each component is smoothed with a Savitzky-Golay filter, which preserves the curvature of the trend while attenuating high-frequency scatter.

Block 3, data correction. Gross errors, epochs deviating from a rolling median by more than a multiple of the median absolute deviation, are deleted. Offsets are corrected rather than deleted, in two stages. First, a preliminary trend and seasonal model is removed and candidate breakpoints are proposed in the residual by binary segmentation of the sum-of-squares criterion, with a penalty rather than a prescribed number of breakpoints as the stopping rule (Truong et al., 2020): requesting a fixed count returns exactly that count whether or not a discontinuity exists, so a clean series must be free to return none. Second, the candidates enter a single least-squares solution with the trend and seasonal terms, the design matrix gaining one Heaviside column each; the weakest offset is discarded and the system re-solved until all survivors clear the threshold. Estimating offsets jointly with the harmonics prevents an imperfectly modelled annual signal from being decomposed into a chain of spurious steps.

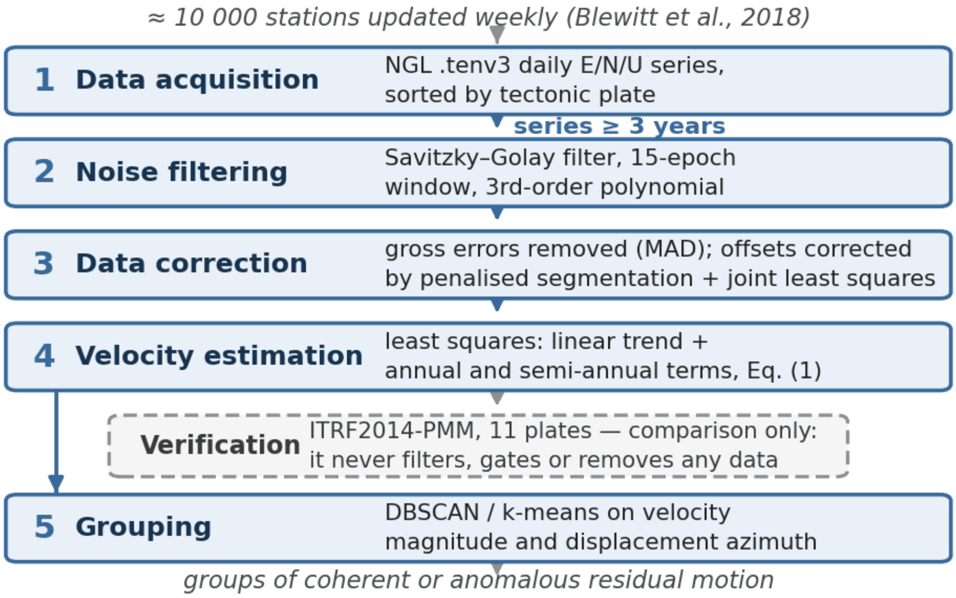

Figure 1. *Block diagram of the automated processing pipeline. The plate-motion model enters only as an independent check on the estimated velocities*

The threshold is calibrated from the data. Because the noise is coloured, the standard error of a mean understates how far the level of a real series wanders by chance; the threshold is therefore a multiple of the robust scatter of differences between medians of adjacent blocks of epochs, measured on each series and so valid whatever its noise colour. A ceiling of one offset per 120 epochs is imposed as well: a series needing a correction every few weeks is defective rather than offset-affected, and filling it with corrections would fabricate a signal. Block 4, velocity estimation. For each component the following model is fitted by least squares:

$$f(t_i) = y_0 + v \cdot \tau_i + \Sigma_{f=1,2} [\, c_f \sin(2\pi f\, t_i) + d_f \cos(2\pi f\, t_i) \,] + \varepsilon_i \quad (1)$$

Here τ_i is the time from the reference epoch, v the linear velocity, y_0 the offset at that epoch, $f = 1$ and $f = 2$ the annual and semi-annual terms, and ε_i the residual. Residuals are evaluated against the cleaned but unsmoothed series, so the reported scatter describes the data rather than the smoothing window.

Verification. Each horizontal velocity is compared with the ITRF2014 plate motion model, which supplies Euler poles for eleven major plates (Altamimi et al., 2017). The comparison is a diagnostic column only: it never filters, gates or removes an observation, and every station passing the input check is detrended however far it departs. Block 5, grouping. Each station is summarised by the magnitude of its velocity and the azimuth of its horizontal displacement, then grouped with DBSCAN or k-means. Restricting the feature space to one or two interpretable scalars keeps every group explainable in physical terms (Takahashi & Hashimoto, 2022).

Results

The correction block was validated first on synthetic series carrying a known velocity, seasonal signals and white plus flicker noise of realistic amplitude. Across 150 clean series, it introduced 0.18 spurious offsets per series, against the 30 a fixed-count segmentation imposes by construction. Detection probability reaches 0.97 at 8 mm and unity from 10 mm, corresponding to the threshold reported for the best automated solutions in DOGEx (Gazeaux et al., 2013). Over 100 realisations with zero to three offsets, the 5th-to-95th percentile range of the velocity error narrowed from 7.58 to 1.24 mm/yr. Table 1 summarises the effect on the 50-station test set: applied offsets fall from 1,374 to 331 and the median horizontal velocity from 8.27 to 3.58 mm/yr. The two regimes separate cleanly. Series with a genuine discontinuity return to plausible velocities, while undisturbed series are left untouched, their median horizontal velocity staying at 0.83 mm/yr with no station changing by more than 0.79 mm/yr. Figure 2 shows the first case: once a level shift of about one metre is removed, the seasonal signal dominates and the velocity falls from -64.4 to -0.38 mm/yr.

Table 1. *Velocity and offset counts before and after the change to the correction block, for representative stations of the Eurasian plate test set*

Station	**Epochs per year**	**Offsets before**	**Offsets after**	**|vH| before, mm/yr**	**|vH| after, mm/yr**
OBE2	347	30	10	811.5	1.01
KELM	331	30	8	250.3	1.77
1SKU	361	30	2	0.19	0.08
LECO	251	30	0	0.97	0.90
All 50 (median)	**306**	**30**	**5**	**8.27**	**3.58**

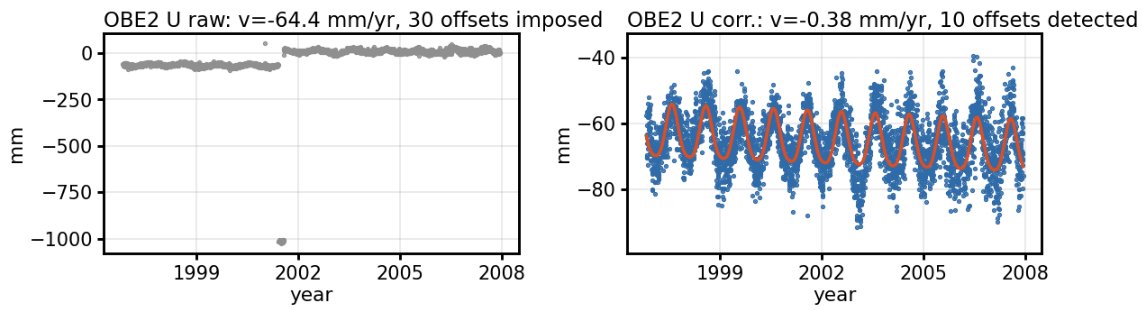

Figure 2. *Station OBE2, up component: raw (left, grey) and corrected (right, blue) with the fitted model in red. Note the change of vertical scale between panels*

Two limitations remain. Stations with few or strongly interrupted records keep implausible velocities: the correction block rightly declines to invent offsets in them, making the data problem explicit rather than hiding it. Separately, the gross-error filter still passes clusters of consecutive outliers, because a contaminated rolling median is no longer a reliable reference.

Conclusions

An automated pipeline carries daily GNSS coordinate series from the NGL archive through filtering, correction, velocity estimation and unsupervised grouping, with the plate-motion model entering only as an independent check and never as a filter. On the Eurasian plate, it processes 6,941 stations, 79.8 per cent of the files.

The correction block governs the quality of the result. A penalised criterion instead of a prescribed number of breakpoints, offsets estimated jointly with the trend and seasonal terms, and significance calibrated against the empirical level noise of each series together reduce spurious offsets from 30 per series to 0.18, detect offsets from 8 mm, and narrow the velocity error range to 1.24 mm/yr.

The residual difficulty is no longer the algorithm but the input data. Stations with few or heavily interrupted records still yield unusable velocities, and upcoming work will add a quality criterion based on observation density and the internal consistency of the fit — derived from the solution itself and independent of the plate-motion model.

References

Altamimi, Z., Métivier, L., Rebischung, P., Rouby, H., & Collilieux, X. (2017). ITRF2014 plate motion model. Geophysical Journal International, 209(3), 1906-1912. https://doi.org/10.1093/gji/ggx136

Blewitt, G., Hammond, W. C., & Kreemer, C. (2018). Harnessing the GPS data explosion for interdisciplinary science. Eos, 99. https://doi.org/10.1029/2018EO104623

Gazeaux, J., Williams, S., King, M., Bos, M., Dach, R., Deo, M., Moore, A. W., Ostini, L., Petrie, E., Roggero, M., Teferle, F. N., Olivares, G., & Webb, F. H. (2013). Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment. Journal of Geophysical Research: Solid Earth, 118(5), 2397-2407. https://doi.org/10.1002/jgrb.50152

Khazraei, S. M., & Amiri-Simkooei, A. R. (2021). Improving offset detection algorithm of GNSS position time-series using spline function theory. Geophysical Journal International, 224(1), 257-270. https://doi.org/10.1093/gji/ggaa453

Takahashi, A., & Hashimoto, M. (2022). Cluster analysis of dense GNSS velocity field reveals characteristics associated with regional tectonics in New Zealand. Journal of Geophysical Research: Solid Earth, 127(11), e2022JB024793. https://doi.org/10.1029/2022JB024793

Tretyak, K., & Brusak, I. (2022). Modern deformations of Earth crust of territory of Western Ukraine based on «GEOTERRACE» GNSS network data. Geodynamics, 1(32), 16-25. https://doi.org/10.23939/jgd2022.02.016

Savchyn, I., Brusak, I., & Tretyak, K. (2023). Analysis of recent Antarctic plate kinematics based on GNSS data. Geodesy and Geodynamics. 14(2). 99-110. https://doi.org/10.1016/j.geog.2022.08.004

Savchyn, I., & Bilashuk, A. (2024). Recent movements of the Earth's surface of the Carpathian mountain system according to GNSS data. Geodynamics, 2(37), pp. 5-15. https://doi.org/10.23939/jgd2024.02.005

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. Signal Processing, 167, 107299. https://doi.org/10.1016/j.sigpro.2019.107299

Williams, S. D. P., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R. M., Prawirodirdjo, L., Miller, M., & Johnson, D. J. (2004). Error analysis of continuous GPS position time series. Journal of Geophysical Research, 109, B03412. https://doi.org/10.1029/2003JB002741