

Statistical Anomaly Detection Using LC-Curves: An Application to Geophysical Time Series

Fuad Aleskerov[1,2] and Aleksei Kosychev[1(✉)]

[1] HSE University, Myasnitskaya St. 20, Moscow 101001, Russia
amkosychev@edu.hse.ru
[2] Institute of Control Science of Russian Academy of Science, Profsoyuznaya St. 65, Moscow 117997, Russia

Abstract. We adapt the Life Cycle (LC) curve methodology to the problem of statistical anomaly detection in time series. LC-curves transform a raw non-negative time series into a cumulative share profile that highlights when, within a fixed observation window, most of the signal intensity is concentrated. Deviations of this profile from the uniform-accumulation baseline serve as simple anomaly indicators that require no distributional assumptions. As an application, we apply the method to 5-min GPS displacement time series from the Nevada Geodetic Laboratory, comparing 196 earthquakes with $M \geq 6$ against 104 seismically quiet intervals, and evaluate performance with Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI). The method achieves POD ≈ 0.82 (CSI $= 0.62$, FAR $= 0.29$) at the lowest tested threshold. Pre-event anomalies are statistically distinguishable from calm-period anomalies, demonstrating that LC-curves can effectively convert high-frequency time series into actionable anomaly indicators.

Keywords: LC-curves · non-stationary time series · anomaly detection · non-parametric methods · event-based verification metrics

1 Introduction

Short-term earthquake prediction remains a difficult and controversial problem, despite decades of work on seismic and geodetic precursors. GPS station networks now provide small-scale measurements of ground motion at high temporal resolution, offering the possibility to detect subtle deformation anomalies in the hours to days before large events. However, it is still unclear to what extent such signals can be distinguished from normal variability in a systematic, statistically robust way, and how often similar anomalies appear during seismically quiet periods.

In this study, we address this question by adapting the Life Cycle (LC) curve methodology — originally developed for analyzing non-stationary customer behavior — to the problem of pre-earthquake ground-deformation anomalies. LC-curves method transforms a raw time series into a cumulative profile that highlights when, within a fixed window, most of the signal is accumulated. Applied to 5-min horizontal GPS displacement

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16766, pp. 384–392, 2027.
https://doi.org/10.1007/978-3-032-30539-8_30

series from the Nevada Geodetic Laboratory, this representation allows us to compress pre-event windows into a small number of anomaly indicators.

We build a global dataset that treats earthquakes and calm intervals symmetrically. For 196 earthquakes with $M \geq 6$ and sufficient geodetic coverage, as well as for 104 selected seismically quiet intervals, we compute LC-curve-based anomaly measures from nearby GPS stations. Each event (or calm case) is then represented by a single maximum anomaly value, enabling event-based evaluation with standard verification metrics: Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI). This framework allows us to quantify both apparent precursors and false alarms under consistent assumptions.

2 Literature Review

Research on earthquake prediction spans more than a century, yet classical seismological approaches have long struggled to deliver reliable short-term predictions. As Geller [8] argued, precursor signatures such as seismic quiescence, geomagnetic changes, or hydrological anomalies often lack reproducibility across events and regions. Later studies refined the understanding of crustal deformation [6, 15], while comprehensive reviews summarized decades of precursor research [10]. Together, these works highlight both the diversity of potential precursors and the difficulty of converting them into operational forecasts.

The emergence of artificial intelligence marked a major shift to data-driven modeling. Early reviews [3, 4] showed that machine learning methods (SVM, neural networks, fuzzy–neural hybrids) outperformed classical statistical techniques, especially for nonlinear and noisy datasets. Subsequent regional studies [7, 11] reinforced the value of ML for forecasting ground-motion parameters and short-term seismic behavior in local networks.

Deep learning further transformed earthquake data analysis. The EQTransformer model [12] demonstrated high accuracy in waveform detection and P/S phase picking, enabling the construction of high-resolution catalogs with up to ten times more events than traditional methods. Other works [5] proposed that such dense catalogs reveal previously hidden triggering patterns and fault interactions.

Multi-sensor approaches extend these capabilities. GPS and InSAR displacement time series [9] and satellite-derived thermal and atmospheric indicators [17] have shown measurable anomalies days to weeks before major events. Meanwhile, IoT-enhanced systems [6, 14] integrate ML with dense, real-time sensor networks, improving latency and scalability for early-warning applications.

Overall, the literature shows a clear progression from precursor-centric approaches with limited reliability toward ML- and DL-based methods that extract predictive structure directly from seismic, geodetic, and satellite data, increasingly combined with multi-sensor fusion. Despite remaining challenges (data scarcity for large earthquakes, interpretability, regional variability), current advances provide strong evidence that high-resolution, multimodal data combined with modern analytical approaches can reveal meaningful short-term predictive patterns previously inaccessible to traditional methods.

3 Life Cycle (LC) Curves

The LC-curves approach provides a way to analyze the inherently non-stationary behavior of a system by examining how a chosen indicator evolves over a given time horizon. Unlike traditional econometric techniques for non-stationary time series where the temporal dimension typically enters the model as a parameter, the LC-curves method focuses directly on the cumulative dynamics of the process. This framework makes it possible to treat both stationary and non-stationary time series without requiring any assumptions about stationarity.

Conceptually, an LC-curve represents the life cycle of an object (e.g., a customer in the original formulation). The foundational models behind this methodology were first introduced in [1] and further developed in [2].

Earlier applications of the methodology to earthquake data were first presented in a plenary talk at the eighth International Conference on Dynamics of Disasters (DOD 2025), Nanjing, China, 2025 (forthcoming [19]), and at the 13th International Conference on Geological and Civil Engineering (ICGCE 2026), Fukuoka, Japan, 2026.

3.1 Construction of LC-Curves

Let n be the number of time moments, and for a particular customer non-negative time series is defined as $\{Vi\}$, $Vi \geq 0$, $i = 1, \ldots, n$. To construct an LC-curve at time t the sum of the values of the time series Vi is calculated starting from time t, till time n. Then the share of this sum to the sum of all values of Vi from 1 to n is calculated. Namely,

$$LC(t) = \frac{\sum_{i=t}^{n} v_i}{\sum_{i=1}^{n} v_i}$$

Where Vi is the value of the time series at time i.

Notice that in the first moment of time $t = 1$. $LC(1) = 1$. At the last time moment $t = n$, $LC(n)$ is not always close to 0, for instance, if Vi is equal to 0 for all i, except n. then $LC(n) = 1$. Nevertheless, for the most of the cases, $LC(n)$ is close to 0 as the number of periods of time n increases.

3.2 Examples

Let us consider two examples taken directly from the GPS dataset. We use the vertical displacement record at station CONZ, located 40 km from the epicenter of the M = 6.2 earthquake (event usp000hkg9). On a seismically quiet day at this station the input series consists of small fluctuations, and the corresponding LC-curve almost coincides with the uniform-accumulation bisector (Fig. 1).

In the second example (Fig. 2) we use the same station and component on the day of the earthquake. The input series is dominated by a sharp burst of displacement at the moment the seismic wave reaches the station. The LC-curve responds with a pronounced step and stays well below the bisector for the remainder of the window.

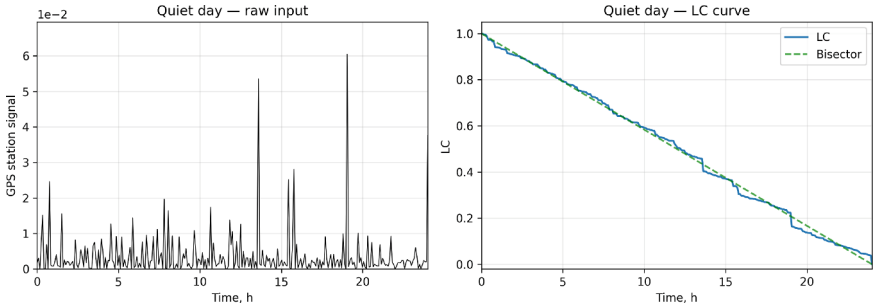

Fig. 1. A GPS displacement series at station CONZ on a seismically quiet day

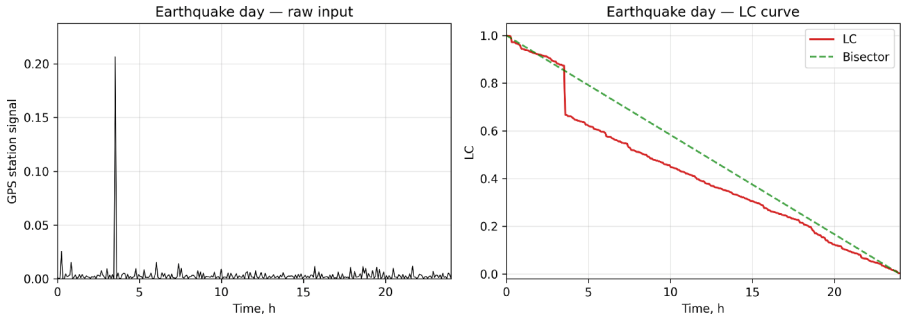

Fig. 2. GPS displacement series at station CONZ before earthquake.

4 Data Description

This study relies on several datasets that together enable a systematic analysis of GPS anomalies preceding major seismic events. We begin with a global catalogue of earthquakes with magnitude $M \geq 6$, containing 2750 events in the 2004–2025 period (U.S. Geological Survey [16]), and compute the distance between each event and every GPS station available from the Nevada Geodetic Laboratory (NGL) [13]. We then restrict the dataset to earthquakes that have at least one GPS station within 50 km radius from the epicenter, resulting in a total of 590 events.

To ensure that the selected earthquakes are not influenced by earlier local seismicity, we incorporate a second global catalogue that includes all earthquakes with $M \geq 2$ covering the same time period. For each of the 590 earthquakes, we analyze the surrounding seismic activity in a pre-event window defined by 14 days prior to the mainshock and within a 30 km radius. Events that are not preceded by any smaller earthquakes in this window form a filtered subset of 329 earthquakes with $M \geq 6$.

We complement these seismic datasets with a global catalogue of 22 287 GPS stations. For each of the 329 selected earthquakes, we identify all stations located within 50 km radius, forming 1919 earthquake–station pairs. From these pairs, we successfully download 1161 unique GPS time series from NGL. Each downloaded file contains 5-min observations of the north–south, east–west, and vertical displacement components

recorded throughout a year. Out of the 1161 files, 978 contain a 4-day interval preceding the earthquake and can therefore be processed by the model. The processed files correspond to 196 unique earthquakes.

To evaluate false alarms, we additionally analyze GPS displacements during seismically quiet periods. For each of the 176 stations associated with the 196 earthquakes, a calm window is defined as an interval between two successive $M \geq 2$ events located within 50 km of the station, drawn from the same global catalogue used in the earthquake-screening step. Only calm windows of at least 124 days are retained — this margin ensures that the 4-day analysis segment lies well inside a seismically quiet interval — and for every station we select the longest such window in which a continuous 4-day block of 5-min observations is available. This procedure yields 104 fully complete calm-period datasets, each suitable for LC-curve computation in the same 4-day format used for the earthquake cases.

5 Evaluation Metrics

To assess the performance of the model in detecting earthquakes, we employ three widely used verification metrics Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI). These metrics are standard in event-based evaluation, particularly in applications involving rare or spatially sparse phenomena [18].

In all metric definitions, we use the standard contingency-table notation, where TP (true positives) denotes correctly detected events, FP (false positives) represents cases where an event was predicted but did not occur, and FN (false negatives) corresponds to events that occurred but were not predicted.

POD measures the model's ability to correctly identify positive events.

$$POD = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$$

A POD value of 1 indicates perfect detection.

FAR describes the fraction of predicted events that did not actually occur.

$$FAR = \mathrm{FP}/(\mathrm{TP} + \mathrm{FP})$$

Lower FAR values indicate fewer false alarms.

CSI provides an overall measure of event-detection performance by jointly penalizing both missed events and false alarms.

$$CSI = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP} + \mathrm{FN})$$

Unlike many other metrics, CSI does not include true negatives, making it particularly suitable for sparse-event domains.

6 LC-Curve Model Implementation

In this section we describe the LC-curve model implementation for earthquake cases and calm-day scenarios to verify the true nature of anomalies in earth displacements.

First, we take the nearest GPS stations to the earthquake. Then, from the next point in the station's time series (i) we subtract the value of the previous one (i − 1). To get the correct results when using the LC-curves method, we need to get rid of the negative values of the earth's surface displacements. To do this, we extract the minimum values for each of the displacement directions from the data obtained. The resulting values are then converted by raising to the fourth power. Lower powers and the absolute value were also tested. The fourth power gave the strongest separation between bursts and background noise and was therefore retained.

As a scalar deviation metric for each LC-curve we use its maximum Euclidean distance to the uniform-accumulation bisector. This single value is referred to throughout as the anomaly value and is the quantity to which all detection thresholds in the following sections are applied. In fact, much more extended metrics can be used as well.

6.1 LC-Curve Analysis of Pre-Seismic Displacements

The number of GPS stations available per earthquake varies substantially. According to the processed dataset, 87 earthquakes are associated with only one GPS station, while the remaining events involve multiple stations, in some cases extending to more than ten stations per event. This variability reflects differences in regional network density but does not affect the workflow, because we retain only the station exhibiting the largest LC-curve deviation as the observation for each earthquake. This procedure results in a single anomaly metric per event, enabling consistent event-level evaluation.

To quantify the model's ability to detect earthquake-related anomalies, we compute true-positive (TP) and false-negative (FN) counts for a sequence of anomaly thresholds ranging from 0.09 to 0.15. The resulting detection statistics are summarized in Table 1.

Table 1. True positives and false negatives for all earthquakes at different thresholds.

Threshold	TP	FN
0.09	161	35
0.10	145	51
0.11	125	71
0.12	109	87
0.13	103	93
0.14	88	108
0.15	82	114

6.2 Model Performance for Non-Earthquake Scenarios

To evaluate the baseline behavior of the LC-curves model in the absence of seismic activity, we apply the same anomaly-detection procedure to long, seismically quiet intervals.

Following the multi-stage filtering and completeness checks described in Sect. 4, we obtain 104 stations that meet the criteria for uninterrupted observations.

For each station, we compute the LC-curve deviation for the selected calm-day interval and extract the maximum anomaly value. We evaluate the number of calm-day anomalies that exceed a series of thresholds from 0.09 to 0.15. For each threshold, we count the number of false positives (FP) and true negatives (TN). The results are summarized in Table 2.

Table 2. False positives and true negatives for 104 calm-period intervals.

Threshold	FP	TN
0.09	65	39
0.10	50	54
0.11	39	65
0.12	32	72
0.13	22	82
0.14	19	85
0.15	14	90

At the lowest thresholds (0.09–0.10), the model registers 65 and 50 false positives, reflecting that a small fraction of calm-day intervals can occasionally exhibit moderate deviations. As the threshold increases, FP counts decrease steadily from 39 at 0.11 to 14 at 0.15, while TN counts increase correspondingly. This monotonic transition indicates that higher anomaly thresholds effectively suppress non-earthquake detections.

7 Discussion

In this section, we evaluate the model's ability to discriminate between earthquake-related anomalies and background variability by comparing all earthquake cases against all non-earthquake cases. We apply the same set of event-based metrics POD, FAR, and CSI across a sequence of detection thresholds. The resulting metrics are summarized in Table 3.

Table 3. Detection metrics (POD, FAR, CSI) for 196 earthquakes vs. 104 calm-period intervals.

Threshold	POD	FAR	CSI
0.09	0.82	0.29	0.62
0.10	0.74	0.26	0.59
0.11	0.64	0.24	0.53

(*continued*)

Table 3. (*continued*)

Threshold	POD	FAR	CSI
0.12	0.56	0.23	0.48
0.13	0.53	0.18	0.47
0.14	0.45	0.18	0.41
0.15	0.42	0.15	0.39

Across the entire threshold range, we observe a monotonic decline in POD, indicating that stricter detection requirements inevitably lead to an increase in missed events. At the same time, the FAR steadily decreases, reflecting a reduction in false alarms as the model becomes more conservative. The CSI also declines smoothly with increasing threshold. Since CSI penalizes both types of errors (misses and false alarms), the gradual reduction of CSI indicates an overall deterioration of balanced performance as the model is shifted into a low-sensitivity regime. Importantly, the maximum CSI ≈ 0.62 is observed at the lowest threshold, which corresponds to the highest POD of 0.82 but also the highest FAR of 0.29.

Taken together, these patterns provide a systematic view of the model's performance and allow us to identify threshold ranges where the balance between missed detections and false alarms is most appropriate for operational use.

8 Conclusion

This study examined whether LC-curves, originally developed to describe customer behavior, can be adapted to detect short-term GPS displacement anomalies preceding large earthquakes. Using LC-based anomaly measures for 196 earthquakes ($M \geq 6$) and 104 seismically quiet intervals, we evaluated detection skill with POD, FAR, and CSI across a range of anomaly thresholds. The results show that LC-curves do capture statistically distinct patterns in pre-seismic GPS time series compared to calm periods, with the highest POD ($\approx$0.82) at the lowest tested threshold of 0.09.

Overall, our findings suggest that LC-curves provide a simple way to compress high-frequency GPS time series into anomaly indicators that carry modest predictive value for large earthquakes.

Acknowledgments. This work has been supported by the grants of the Russian Science Foundation, RSF No. 24-61-00030, https://rscf.ru/project/24-61-00030/.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aleskerov, F., Golubenko, M.: On the Evaluation of a Symmetry of Political Views and Polarization of Society. Working paper WP7/2003/04. HSE University, Moscow (2003)
2. Aleskerov, F., Volsky, V., Yakuba, V.: Application of the symmetry and polarization indices to the analysis of consumption dynamics of customers. In: Proceedings 4th International Conference on Control Problems, Moscow (2009)
3. Azam, F., Sharif, M., Yasmin, M., Mohsin, S.: Artificial intelligence based techniques for earthquake prediction: a review. Sci. Int. **26**(4), 1495–1502 (2014)
4. Banna, M.H.A., et al.: Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges. IEEE Access. **8**, 192880–192923 (2020)
5. Beroza, G.C., et al.: Machine learning and earthquake forecasting — next steps. Nat. Commun. **12**, 4761 (2021)
6. Bilal, M.A., et al.: An early warning system for earthquake prediction from seismic data using batch normalized graph convolutional neural network. Sensors. **22**(17), 6482 (2022)
7. Essam, Y., et al.: Exploring the reliability of different artificial intelligence techniques in predicting earthquake for Malaysia. Soil Dyn. Earthq. Eng. **147**, 106826 (2021)
8. Geller, R.J.: Earthquake prediction: a critical review. Geophys. J. Int. **131**(3), 425–450 (1997)
9. Gitis, V., Derendyaev, A., Petrov, K.: Analyzing the performance of GPS data for earthquake prediction. Remote Sens. **13**(9), 1842 (2021)
10. Huang, F., et al.: Studies on earthquake precursors in China: a review for recent 50 years. Geod. Geodyn. **8**(1), 1–12 (2017)
11. Marhain, S., et al.: Investigating the application of artificial intelligence for earthquake prediction in Terengganu. Nat. Hazards. **108**(2), 977–999 (2021)
12. Mousavi, S.M., et al.: Earthquake transformer — an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nat. Commun. **11**, 3952 (2020)
13. Blewitt, G., Hammond, W.C., Kreemer, C.: Harnessing the GPS data explosion for interdisciplinary science. Eos. **99** (2018)
14. Pwavodi, J., et al.: The role of artificial intelligence and IoT in prediction of earthquakes: review. Artif. Intell. Geosci. **5**, 100075 (2024)
15. Shen, Z., Jackson, D.D., Ge, B.X.: Crustal deformation across and beyond the Los Angeles Basin from geodetic measurements. J. Geophys. Res. **101**, 27957–27980 (1996)
16. Geological Survey, U.S.: Earthquake Hazards Program: Lists, Maps, and Statistics. https://www.usgs.gov/programs/earthquake-hazards
17. Xiong, P., et al.: Towards advancing the earthquake forecasting by machine learning of satellite data. Sci. Total Environ. **771**, 145256 (2021)
18. Wilks, D.S.: Statistical Methods in the Atmospheric Sciences, 3rd edn. Academic (2011)
19. Aleskerov, F., Kosychev, A.: Observations of ground surface displacement as a factor in earthquake onset using the LC-curves method. In: Dynamics of Disasters, Springer Optimization and its Applications, vol. 229. Springer (2026)