Earthen Embankment Monitoring using LiDAR data by Randomized Consensus of Topological Data Analysis

Austin R.J. Downey[1,2][0000−0002−5524−2416], Jie Wei[3][0000−0003−1781−6412], A Q M Zohuruzzaman[4][0009−0000−2863−1929], Paul T. Schrader[6][0000−0002−6017−0509], Sadik Khan[4][0000−0002−0150−6105], Jason Bakos[5][0000−0002−0821−6258], Weicong Feng[3][0009−0007−9246−8792], Erik Blasch[6][0000−0001−6894−6108], and Erika Ardiles-Cruz[6][0000−0003−3092−4696]

[1] Mechanical Engineering, University of South Carolina, Columbia, South Carolina, USA

[2] Civil and Environmental Engineering, University of South Carolina, Columbia, South Carolina, USA

[3] Computer Science, City College of New York, New Your City, New York, USA

[4] Civil and Environmental Engineering, Jackson State University, Jackson, Mississippi, USA

[5] Computer Science and Engineering, University of South Carolina, Columbia South Carolina, USA

[6] Air Force Research Laboratory, Rome New York, USA

Distribution Statement A. Approved for public release; distribution is unlimited: AFRL – 2024 – 3971

Abstract. Ensuring continuous operation of critical infrastructure necessitates frequent labor-intensive inspections and maintenance. This study leverages computing and sensing techniques within a Dynamic Data Driven Applications Systems (DDDAS) paradigm. Specifically, 3D LiDAR is utilized to capture detailed geometry information. This data is then analyzed using Topological Data Analysis (TDA) techniques to gain insights into the health of the infrastructure. The research focuses on expansive clay terrains over two years during the Summer and Fall seasons at a site in Jackson, Mississippi. The dataset, part of the Slope LiDAR embankment (SLidE) dataset, includes 3D point clouds with 1-6 million points per scan. Due to the high computational demands of TDA, a randomized sampling method is employed to reduce the data size from millions to thousands of points, ensuring efficient analysis without compromising accuracy. Results highlight the deformation of expansive clays, known for their shrink-swell behavior in response to moisture changes, which poses significant geotechnical challenges. These findings are crucial as climate change with increased precipitation, affects the stability of earth-constructed embankments. By capturing dynamic soil behavior through seasonal 3D scanning, the study provides insights into deformation patterns. The approach balances rigorous data representation with manageable computational demands, revealing a 2.35% seasonal variation in slope geometry; potentially correlated with moisture levels.

Keywords: infrastructure· 3D LiDAR · Topological Data Analysis (TDA) · Expansive clay · Moisture dynamics · Shrink-swell behavior DDDAS · Dynamic Data Driven Applications Systems · InfoSymbiotic Systems

1 Introduction

Ensuring the safety and reliability of critical infrastructure is a priority in modern civil engineering, especially for transportation networks. Traditional inspection and maintenance tasks are labor-intensive and pose significant risks to human technicians, making the adoption of non-invasive methodologies increasingly important. Expansive clay terrains present particular geotechnical challenges due to their shrink-swell behavior in response to moisture variations [9]. These soils, which can undergo significant volume changes, pose a persistent threat to the stability of structures such as highway embankments. The dynamic nature of expansive clays, exacerbated by climate change and associated shifts in precipitation patterns, necessitates continuous monitoring to prevent potential failures.

Generally, traditional methods are employed to assess the condition of embankments [11], including collecting soil samples from the subsurface and subjecting them to laboratory tests to determine their physical and mechanical properties, such as grain size distribution, Atterberg limits, shear strength, and compressibility. In-situ tests, like the standard penetration test, cone penetration test, and vane shear test, are also conducted on-site to evaluate soil properties and strength parameters. These destructive techniques are labor-intensive and costly. To overcome the limitations of these destructive techniques, various non-destructive methods have been developed for geotechnical monitoring [6]. Inclinometers, which measure lateral soil movement, are installed vertically in boreholes to detect slope movement. Piezometers are used to monitor pore water pressure within the embankment, as high pore water pressures reduce soil strength and trigger failures. In recent years, technologies such as electrical resistivity imaging, ground-penetrating radar, and remote sensing techniques have been increasingly utilized to inspect and provide data on large areas.

Leveraging novel computing and sensing technologies, such as 3D LiDAR and Topological Data Analysis (TDA), offers promising solutions by providing detailed insights into infrastructure health with minimal human intervention. TDA is ideal for identifying global and local data structures, making it suitable for analyzing intricate deformation patterns of expansive clays. However, TDA's high computational demands, with an $\mathcal{O}(n^3)$ time complexity [5], require innovative data processing approaches. Others have explored accelerating TDA by clearing birth columns when reducing the boundary matrix under specific conditions [3]. This work seeks a robust data-processing method that limits preconditions.

This paper presents a novel application of TDA to the monitoring of earthen embankments made from expansive clay by employing a randomized sampling strategy to reduce computational load while preserving analytical accuracy. To do so, this work leverages the open-source Slope LiDAR embankment (SLiDE) dataset [8] introduced in prior works by the team [15]. The SLiDE dataset contains 3D LiDAR scans of an embankment along the Terry Road Exit from I-20 in Jackson, Mississippi, over multiple seasons across several years. The contribution of this work is twofold. First, it proposes a computationally efficient TDA approach based on randomized consensus. Second, it demonstrates the effectiveness of TDA in detecting the shrink-swell behavior of expansive clays.

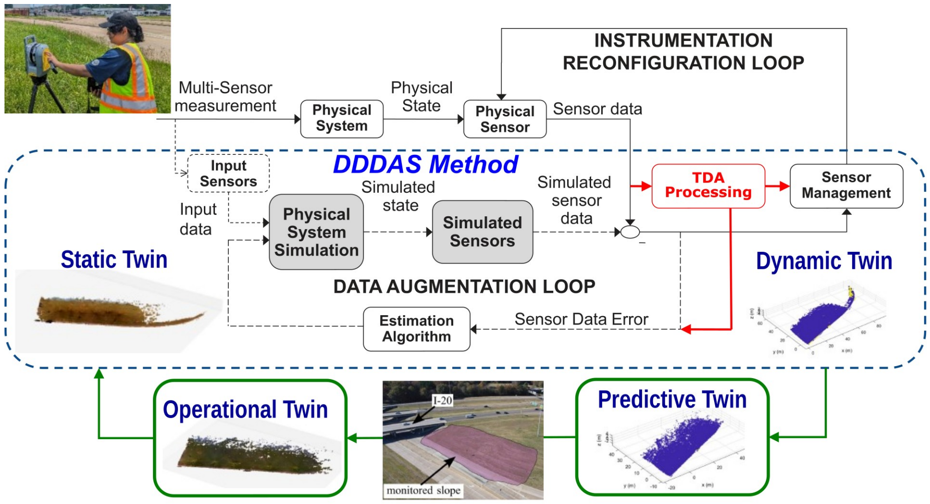

Fig. 1. A proposed Dynamic Data Driven Applications Systems (DDDAS) framework to enable dynamically optimized 3D LiDAR sensing and TDA-based data processing for enhanced and efficient monitoring of earthen embankments

1.1 DDDAS for Intelligent Slope Monitoring

The Dynamic Data Driven Applications Systems (DDDAS) paradigm [1] offers the potential to dynamically optimize 3D LiDAR sensing and TDA-based data processing for enhanced and efficient monitoring of earthen embankments. DDDAS offers a transformative approach to infrastructure monitoring, particularly for geotechnical applications such as slope stability analysis. DDDAS integrates real-time data acquisition with computational models, creating a feedback loop that dynamically updates simulations based on incoming data and, conversely, guides data collection efforts based on the evolving state of the system. This approach enhances the accuracy, responsiveness, and efficiency of monitoring and analysis processes. For example, Parida et al. [10] developed a DDDAS method for structural analysis of performance-based earthquake engineering. The method involves data fusion estimation, soil modeling, and machine learning to assist in bridge structural integrity

A proposed DDDAS-based process is shown in Figure 1 where data collected over seasons can be analyzed and saved. The historical information can drive the allocation (when and where) of limited instrumentation (i.e. a set number of LidAR scanning systems). Combining the historical data with the real-time collections in the DDDAS paradigm affords effective physical interpretation of changes in soil moisture and knowledge of any short-or-long term changes in the embankment to support transportation agencies in making mitigating actions. The integration of DDDAS with this monitoring setup involves several key components:

1. Real-Time Data Integration: Continuous LiDAR scanning provides up-to-date geometric data, which is immediately integrated into computational models which ensures that the models reflect the current state of the slope.
2. Adaptive Modeling: The computational models adapt based on the incoming LiDAR data. For example, if the models detect significant changes in slope geometry indicative of potential instability, they can trigger more frequent or detailed data collection in the affected areas.
3. Enhanced Decision-Making: The insights gained from the models, which now incorporate the latest data, enable more informed decision-making. This can include directing maintenance efforts to areas showing early signs of deformation or instability, thereby preventing more severe issues.
4. Resource Optimization: By focusing data collection and analysis efforts on critical areas identified by the models, DDDAS optimizes the use of resources, reducing unnecessary data processing and focusing attention where needed.

2 Methodology

This section introduces the SLidE Dataset and the TDA processing technique.

2.1 SLidE Dataset

The SLidE Dataset is publicly available [8] and was initially reported by Zohuruzzaman et al. [15]. SLidE contains multiple LiDAR scans of an earthen embankment along the I-20E exit toward Terry Road in Jackson, Mississippi (32°16'-48.92"N, 90°12'44.03"W) [7]. This location was chosen due to its geological composition predominantly consisting of Yazoo clay, a high-plasticity soil known for its challenging shrink-swell behavior. The dataset spans multiple years, with LiDAR scans conducted at different time intervals between Summer 2021 and Fall 2023, enabling the analysis of seasonal variations in the embankment's stability and deformation patterns. The 15 ft. high slope, depicted in Figure 2, has a grade ranging from 3.5:1 to 4:1 (V:H). The slope experienced shallow landslides in the past, which were remediated using steel H-piles. It now consists of reinforced and as-built sections, primarily composed of Yazoo clay.

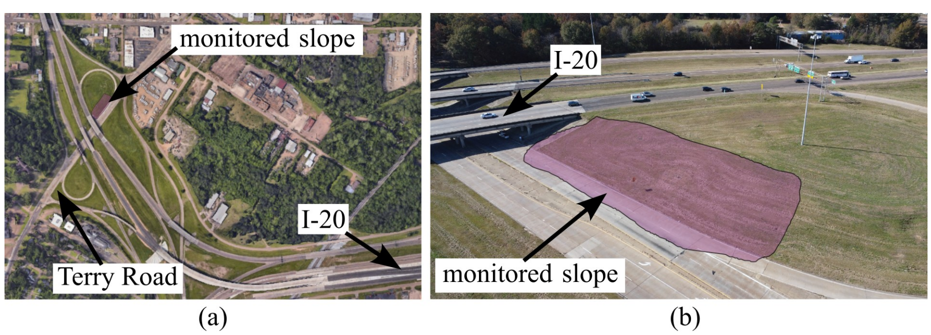

Fig. 2. The earthen embankment monitored in the SLidE Dataset that is located on Terry Road near the I-20 East exit in Jackson Mississippi, showing: (a) satellite image of the reference slope (base image credit Google Earth), (b) an aerial view of the slope taken from the North-East side of the slope.

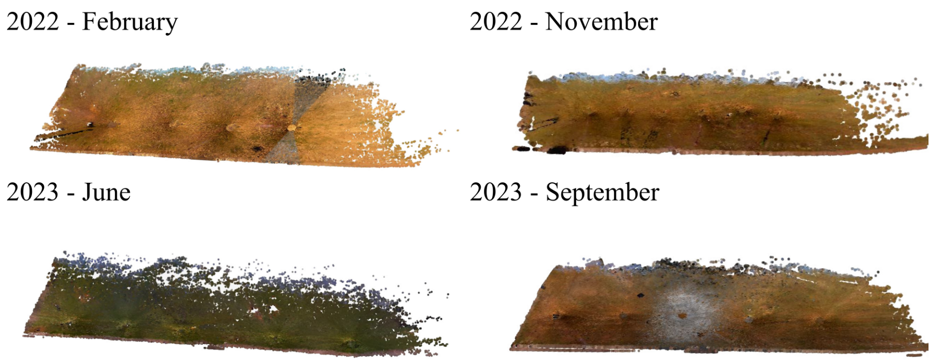

Fig. 3. LiDAR point cloud surface topography of the monitored embankment from scans conducted between 2022 and 2023.

The dataset includes dense point cloud data collected from multiple LiDAR scans over several years, creating a 3D surface and topography of the slope. Scans were conducted at five to six stations using Terrestrial LiDAR equipment (Trimble X7), generating around 20 million points. The data was digitally processed using Trimble RealWorks software to create a unified point cloud with approximately 12 million points, which includes various infrastructural features like the adjacent highway and bridge. Additional processing with Autodesk ReCap removed unrelated elements like road signs, foliage, and personnel. The dataset is available in the compressed 'LAZ' format, a standard for 3D point cloud data. It was converted from e57 format to LAZ using the e57tolas tool from the LASTools software collection, preserving the original point-source IDs. Some point clouds from the public repository are shown in 3. For simplicity, this work leverages the four most recent LiDAR scans in the SLidE Dataset, consisting of scans taken in February and November 2022 along with scans from June and September 2023.

2.2 Topological Data Analysis

TDA is a field with broad scientific impacts originating from studies in applied algebraic and combinatorial topology along with computational geometry. TDA is motivated by the idea that topology and geometry expose qualitative and quantitative global features of data through local behavior and characteristics, which are stable under small perturbations. Its objective is to supply rigorous mathematical and statistical algorithmic methods that yield meaningful analytics when applied to topological and geometric structures of significant complexity by measuring, recording, and tracking linear representations of an underlying data set. Often these data sets are point clouds embedded in Euclidean or some more general metric space where a defined notion of distance exists.

Computational topology drives the field of TDA. The workhorse and most successful of its methodologies is persistent homology (PH). PH is a data compression scheme quantifying critical points of continuous spaces and addressing more general notions of multi-scale characteristics, high-dimensional features, and abstract data structures with the use of discrete metrics. The reader can find foundational details of TDA and PH in Edelsbrunner and Harer [4]. In summary, these characteristics make TDA and PH an attractive choice for digital applications where its implementations has brought numerous results [2,12,13]. The approaches addressing autonomy in infrastructure health introduced in this paper are inspired by Paul Schrader's work and recent TDA-based algorithmic successes detailed in [12]. There he applied TDA and PH to automatic target recognition derived from multimodal sensor data and its fused aggregates.

TDA is a valuable approach to capture the local and global properties innate to 3D point clouds. However, one major problem of TDA is its prohibitively high computational demands: to collect all possible cycles and spheroids from all possible subsets, the current TDA algorithm needs $\mathcal{O}(n^3)$ time complexity, where n is the number of points in the point clouds [5]. In the SLidE Dataset dataset, n ranges between 1 to 6 million, which is computationally infeasible to apply TDA directly on all n points. As done in other image or video processing, the statistical redundancy existing in natural image data should be exploited to release the prohibitive computation involved. By observation, the dense 3D points in our dataset, as depicted in Figure 3, exhibit spatial redundancy. To take advantage of this, the proposed approach uses random sampling to reduce the computation needed for TDA and PH evaluation. For each of the given point clouds, instead of using n points, we randomly draw $m = 5000$ to be fed to the TDA algorithm for analysis, the associated computing cost $\mathcal{O}(m^3)$ is now entirely feasible. There are many different TDA parameters/metrics to capture the global properties of a 3D point cloud. Through empirical investigation, we chose the *Persistence Entropy (PE)* metric for H_0 and H_1 [5]. As a result, each point cloud is represented by a 2D point cloud of ordered pairs (H_0, H_1).

To further eliminate the possible bias in the random sampling, each random sampling of m samples is performed 10 times and the median values of the TDA PE parameters are taken to be the TDA results, which mimics a typical RANSAC (RANdom SAmple Consensus) procedure. The computing procedure RANSAC-TDA for point cloud n is thus summarized in Algorithm 1:

Algorithm 1 Pseudo code for the proposed RANSAC-TDA methodology.

```
procedure RANSAC-TDA(n)
    for K times do
        Randomly sample a subset m_i of 5000 points from point cloud n
        Compute v_i = PE(m_i)
    end for
    return median(v_i)
end procedure
```

3 Results and Discussion

RANSAC-TDA algorithm results are shown in Figure 4, indicating that the June 2023 LiDAR scan has an H1 entropy 2.35% different from the average of the other three scans. These findings align with the author's earlier curvature histogram-based method [15], supporting the hypothesis that changes in H1 entropy are likely due to increased moisture in the slope during summer.

We conducted our RANSAC-TDA algorithm to four 3D point clouds with a casual desktop (CPU: i9-13900, RAM: 64 GB) in Python 3.10 with the giotto-tda package [14]. In our test, the repeated time in the RANSAC-TDA algorithm K is chosen to be 10. On average each run of the RANSAC-TDA for one point cloud takes 69.0 sec. We also inspected the $K=10$ different v_i's, where the 10 different PE metrics are mostly similar, indicating that the spatial redundancy of the point clouds is innate. Thus, random sampling reducing the original 3D point cloud data count by 2-3 orders of magnitudes can still yield a valuable consensus while decreasing the TDA compute load.

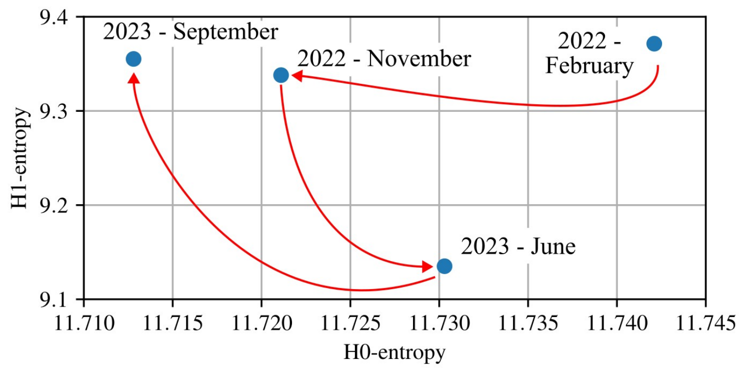

Fig. 4. Results for the TDA analysis in the H1 and H2 plane where the red arrow denotes the passage of time. June 2023 is an outlier in terms of H1 entropy, likely caused by the increased moisture present in the slope during the summer.

4 Conclusion

In this work we endeavor to use powerful Topological Data Analysis (TDA) approaches to analyze the 3D point clouds. To avoid the prohibitively high computing cost, $\mathcal{O}(n^3)$, demanded by TDA, we resorted to a random sampling consensus algorithm to reduce the time complexity by several orders of magnitudes: from millions to thousands, which renders the corresponding computing feasible and efficient. To our knowledge, this is the first time TDA was conducted on a real infrastructure dataset considering computing efficiency and performance. The team is currently in the process of collecting more data with corresponding in situ sensor measurements to more rigorously evaluate the proposed RANSAC-TDA algorithm algorithm.

Acknowledgments. This work is supported by the National Science Foundation Grant numbers 2152896 and 2324052. This work was also partially supported by the Air Force Research Laboratory Faculty Fellowship program which is gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the United States Air Force, or the US government.

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

References

1. Blasch, E.P., Darema, F., Bernstein, D.: Introduction to the Dynamic Data Driven Applications Systems (DDDAS) Paradigm, pp. 1–32. Springer International Publishing (2022). https://doi.org/10.1007/978-3-030-74568-4_1
2. Chazal, F., Michel, B.: An introduction to topological data analysis: Fundamental and practical aspects for data scientists. Frontiers in Artificial Intelligence **4** (2021). https://doi.org/10.3389/frai.2021.667963
3. Chen, C., Kerber, M.: Persistent homology computation with a twist. In: Proceedings 27th European workshop on computational geometry. vol. 11, pp. 197–200 (2011)
4. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. Discrete Comput. Geom. **28**(4), 511–533 (nov 2002)
5. Edelsbrunner, H., Harer, J.: Computational Topology - an Introduction. American Mathematical Society (2010)
6. Gupta, B., Sharma, K., Koirala, T., Sharma, S.: IoT-Based Landslide Monitoring System Using LiDAR Sensors: A Review, pp. 363–379. Springer Nature Singapore (2023). https://doi.org/10.1007/978-981-99-4284-8_30
7. Khan, M.S., Amini, F., Nobahar, M., et al.: Performance evaluation of highway slopes on yazoo clay. Tech. rep., Mississippi. Dept. of Transportation (2020)
8. Khan, S., Zohuruzzaman, A., Wambi, D., Downey, A.R.: Dataset-slope-lidar-embankment-(slide) (Apr 2024), https://github.com/ARTS-Laboratory/Dataset-Slope-LiDAR-Embankment-SLidE
9. Nobahar, M., Khan, M.S., Ivoke, J.: Combined effect of rainfall and shear strength on the stability of highway embankments made of yazoo clay in mississippi. Geotechnical and Geological Engineering **38**(3), 2787–2802 (Jan 2020). https://doi.org/10.1007/s10706-020-01187-8
10. Parida, S., Sett, K., Singla, P.: A model data fusion for statistical characterization of constitutive parameters: Applications to site characterization and seismic performance evaluation. Handbook on Dynamic Data Driven Applications Systems (Volume 3)
11. Puppala, A.J.: Performance evaluation of infrastructure on problematic expansive soils: Characterization challenges, innovative stabilization designs, and monitoring methods. Journal of Geotechnical and Geoenvironmental Engineering **147**(8) (Aug 2021). https://doi.org/10.1061/(asce)gt.1943-5606.0002518
12. Schrader, P.T.: Topological multimodal sensor data analytics for target recognition and information exploitation in contested environments. In: Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII. vol. 12547, p. 125470C. SPIE (2023). https://doi.org/10.1117/12.2662383
13. Stolz, B.J., Harrington, H.A., Porter, M.A.: Persistent homology of time-dependent functional networks constructed from coupled time series. Chaos **27**(4) (2017)
14. Tauzin, G., Lupo, U., Tunstall, L., Pérez, J.B., Caorsi, M., Medina-Mardones, A.M., Dassatti, A., Hess, K.: giotto-tda: A topological data analysis toolkit for machine learning and data exploration. Journal of Machine Learning Research **22**(39), 1–6 (2021), http://jmlr.org/papers/v22/20-325.html
15. Zohuruzzaman, A., Wamai, D.P., Feng, W., Khan, S., Downey, A.R.J., Wei, J., Blasch, E., Schrader, P.T.: Highway slope monitoring using 3d laser scanning at different seasons. In: Palaniappan, K., Seetharaman, G. (eds.) Geospatial Informatics XIV. SPIE (Jun 2024). https://doi.org/10.1117/12.3016172