2024 Scienceline Publication

Journal of Civil Engineering and Urbanism

Volume 14, Issue 3S: 212-219; September 15, 2024

ISSN-2252-0430

DOI: https://dx.doi.org/10.54203/jceu.2024.22

Integrating Prophet Forecasting with Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) for Early Warning System in Dam Deformation Monitoring

Thalosang Tshireletso[1,2], and **Pilate Moyo**[2]

[1]*Department of Civil Engineering, University of Botswana, Gaborone, Botswana*
[2]*Department of Civil Engineering, University of Cape Town, Cape Town, South Africa*

Corresponding author's Email: tshireletsot@ub.ac.bw

RESEARCH ARTICLE
PII: S225204302400022-14
Received: June 25, 2024
Revised: September 02, 2024
Accepted: September 05, 2024

ABSTRACT

Ensuring dam safety requires a monitoring system that can predict deformations and detect anomalies in real-time. This study combines the forecasting capabilities of the Prophet model with the real-time anomaly detection of a Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) framework. The Prophet model analyses historical deformation data to forecast future deformations, enabling early issue identification. The GMM-HMM framework continuously monitors incoming data to detect deviations from predictions. Results shows that the GMM-HMM, with 10 components and a Mahalanobis distance threshold of 0.1, achieved a precision of 0.602, recall of 1.0, and F-1 score of 0.751, ensuring high sensitivity and accurate anomaly detection. The GMM-HMM was then used to detect anomalies on Prophet forecasted radial deformations. Anomalies were detected on upper limit and lower limit deformations. This combined approach enhances dam safety by integrating predictive and real-time monitoring capabilities, offering a comprehensive early warning system for dam infrastructure.

Keywords: Gaussian Mixture Model, Hidden Markov Model, Prophet Model, Dam Deformation Forecasting

INTRODUCTION

Dam deformation monitoring is a critical aspect of ensuring the structural integrity and safety of dams. These large-scale structures are key components in managing water resources, controlling floods, and generating hydroelectric power (Rong et al., 2024). Over the past decades, a plethora of techniques and methodologies have been developed and refined to monitor dam deformations and detect any potential risks of failure (Fraley and Raftery, 2002). However, given the rapidly evolving environmental conditions and the inevitable aging of infrastructure, the urgent need for more advanced, consistent, and reliable early warning systems has become increasingly evident (Šakić et al., 2022).

Without a doubt, early warning systems play a pivotal, arguably indispensable role in mitigating the risks associated with dam failures. They achieve this by providing timely alerts to any potential deformations or structural instabilities, essentially acting as the first line of defence (Pang et al., 2023). With the data from these systems, authorities are enabled to take proactive measures, such as initiating evacuation procedures or implementing reinforcement measures, which are crucial in preventing catastrophic consequences and preserving life and property (Bahrami et al., 2021).

In the broader field of infrastructure monitoring, Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) has found applications in detecting anomalies in time series data from a diverse range of sources. These include, but are not limited to, structural health monitoring sensors, traffic flow sensors, and environmental sensors. For example, in the context of bridge health monitoring, GMM-HMM has been effectively used to identify abnormal vibrations or structural changes that are indicative of potential defects or damage (Coraça et al., 2022). Similarly, in the realm of railway infrastructure monitoring, GMM-HMM has been employed to detect anomalies in track conditions, such as rail irregularities or track geometry deviations, which could pose serious safety hazards.

Despite its demonstrated efficacy in these and other domains of infrastructure monitoring, GMM-HMM has not yet been directly applied to dam deformation anomaly detection. While dams represent critical and high-value infrastructure assets requiring robust and reliable monitoring systems, the application of GMM-HMM specifically to dam deformation data remains an unexplored frontier. Nonetheless, the principles and methodologies of GMM-HMM are conceptually well-suited to the challenges of dam deformation monitoring, where the detection of subtle anomalies in time-varying deformation patterns is crucial for ensuring dam safety.

By leveraging the capabilities of GMM-HMM in modelling temporal dynamics and detecting anomalies, there is a significant potential opportunity to enhance existing dam deformation monitoring systems. Moreover, the development of more effective early warning systems for dam safety could be realized. The integration of GMM-HMM with other predictive modelling techniques, like Prophet Forecasting, could further enhance the accuracy and reliability of anomaly detection in dam deformation data.

The motivation behind integrating Prophet Forecasting with GMM-HMM stems from the complementary strengths of these two approaches in time series analysis and pattern recognition, respectively (Hassani et al., 2019). Prophet Forecasting, a system developed by Facebook, has gained prominence for its ability to handle time series data with irregularities, such as seasonality, holidays, and outliers. On the other hand, GMM-HMM is well-suited for modelling complex temporal patterns and detecting anomalies (Benoit et al., 2013).

By integrating Prophet Forecasting with GMM-HMM, we aim to harness the predictive capabilities of Prophet to forecast dam deformation trends accurately. Subsequently, the GMM-HMM component will leverage these forecasts to detect deviations from expected patterns and issue early warnings in real-time. This hybrid approach offers a synergistic solution that combines the robustness of statistical forecasting with the flexibility of probabilistic modelling, thereby enhancing the reliability and effectiveness of dam deformation monitoring systems.

Our research aims to pioneer the development of a groundbreaking framework that seamlessly integrates Prophet Forecasting with Gaussian Mixture Model Hidden Markov Model (GMM-HMM) to bolster early warning systems tailored specifically for dam deformation monitoring. This innovative integration harnesses the predictive capabilities of Prophet Forecasting, renowned for its adeptness in handling irregular time series data, to forecast dam deformation trends with heightened accuracy. Leveraging the robustness of GMM-HMM, which excels in discerning complex temporal patterns and anomalies, our framework promises a synergistic solution for detecting deviations from expected deformation patterns in real-time.

MATERIALS AND METHODS

The study was based on Roode Elsberg dam, located in Western Cape, South Africa. The monitoring system included measurement of environmental and operational conditions, water temperatures, measured by thermometers embedded on the dam wall; water level; deformations; and accelerations. The monitoring was carried out by Concrete Materials and Structural Integrity Research Unit (CoMSIRU)/UCT and the Department of Water and Sanitation.

Roode elsberg dam

Roode Elsberg dam is in the Western Cape province, South Africa, about 130 km northeast of Cape Town near the town of de Doorns, at coordinates (33.4361°S, 19.5680°E), Figure 1. Construction of the dam was completed in 1969, and its main purpose is irrigation of vineyards in the surrounding farms and limited domestic use via a 7-km tunnel. The dam is a double curvature concrete arch dam with a centrally located spillway and gross capacity of 8.21 million m^3. The height to the lowest foundation point is 72 m, and the length of the crest is 274 m, consisting of two galleries, one following the foundation level and one top instrumentation gallery located about 20 m above the foundation level.

Figure 1. Roode Elsberg dam

Roode Elsberg dam-monitoring system

To understand the behaviour of Roode Elsberg, two monitoring systems were installed on the dam. These included a continuously monitored GPS system at four survey beacons in 2010 and a dynamic monitoring system at the dam crest in 2013. In addition, environmental and operational conditions were measured, i.e., water level measured using staff gauges and water temperatures; a weather station; and a suite of thermometers located at different water levels: 26.23 m, 46.62 m, and 47.30 m. Figure 2 shows the layout of the Roode Elsberg GPS monitoring systems, where blue indicates control stations P01 and P02 while red indicates rover stations P203 and P206 on the left and right flanks, respectively.

Figure 2. Roode Elsberg GPS monitoring system

Measurement of environmental and operational factors

Results from ambient vibration surveys, carried out on the dam, indicated that there was a need to install a continuous dynamic monitoring system to capture any changes in the behaviour of the dam. This system consists of three forced balanced accelerometers installed on the dam crest and a data acquisition system in the upper gallery of the dam (Figure 3).

Figure 3. Dynamic monitoring system

Measurement of environmental and operational factors

The environmental and operational factors include ambient temperature (AT), water temperature, and water level. Ambient temperature is measured by the weather station installed on the dam crest, and water levels are measured by staff gauges. There are also thermometers that are embedded into the dam wall at different levels to measure water temperatures; Figure 3 below shows the position of water temperatures on the dam wall. There are 6 thermometers on each side of the wall; avg1-R indicates that thermometer 1 is on the right flank and avg1-L indicates that thermometer 1 is on the left flank. Avg1-R and avg2-R are on the same level, avg3-R and avg4-R are on the same level, and avg5-R and avg6-R are on the same level, as indicated by Figure 4. This applies to thermometers on the left flank.

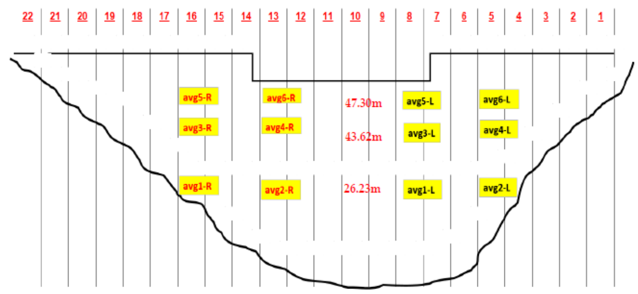

Figure 4. Wall thermometers measuring water temperature

Prophet forecasting model

Prophet is a powerful and comprehensive forecasting model that has been designed specifically to manoeuvre time series data with non-regular patterns such as seasonal variations and holidays (Taylor and Letham, 2018). Unlike other forecasting models, Prophet employs an additive model that deconstructs the time series data into its basic components, which include trend, seasonality, and holiday factors. This methodical separation of components allows for precise prediction capabilities, even when the data set contains outliers or incomplete data.

To capture non-linear trends that develop over time, the model employs a piecewise linear or logistic growth curve. This is particularly useful in situations where the data does not follow a consistent linear pattern. Furthermore, Prophet incorporates Fourier series to model recurring patterns, adding another layer of accuracy to the forecast. In an innovative move, Prophet integrates uncertainty estimation directly into the forecasting process. It achieves this by incorporating Bayesian inference methods, an approach that is grounded in the principles of Bayesian statistics. This allows for a more comprehensive evaluation of prediction uncertainty, which can be crucial in strategic decision making.

The additive model that is the cornerstone of Prophet's forecasting capabilities can be represented as follows:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

where g(t) represents the trend component, s(t) represents the seasonal component, h(t) represents the holiday component and ε_t is the error term. y(t) represents the value of the time series at time t. Here, t stands for the specific time at which the observation is made.

Prophet is a robust forecasting model developed by Facebook that decomposes time series data into trend, seasonality, and holiday effects. This decomposition enables the model to capture and interpret the underlying patterns in the data.

Trend component

Prophet models the trend component using a piecewise linear or logistic growth model. For piecewise linear growth, the trend is given by:

$$g(t) = k + \sum_{i=1}^{n} a_i (t - t_i) 1_{\{t \geq t_i\}} \quad (2)$$

where $T(t)$ is the trend at time *t*, *k* is the initial growth rate, a_i are the rate changes at changepoints t_i, and $1_{\{t \geq t_i\}}$ is an indicator function that is 1 if $t \geq t_i$ and 0 otherwise (Taylor & Letham, 2018). For logistic growth, the trend component is modelled as:

$$g(t) = \frac{C}{1 + exp(-k(t-m))} \quad (3)$$

where *C* is the carrying capacity, *k* is the growth rate, and *m* is the midpoint of the growth (Taylor and Letham, 2018).

Seasonality component

Seasonality in Prophet is modelled using Fourier series:

$$s(t) = \sum_{k=1}^{N} \left(a_k \cos\left(\frac{2\pi kt}{P}\right) + b_k \sin\left(\frac{2\pi kt}{P}\right) \right) \quad (4)$$

where *S(t)* is the seasonal component at time *t*, *P* is the period (e.g., 365.25 for yearly seasonality), and a_k , b_k are the coefficients of the Fourier series (Taylor and Letham, 2018).

Holiday effects

Prophet incorporates holiday effects by adding an additional component:

$$\mathrm{h(t)} = \sum_{\mathrm{i=1}}^{\mathrm{m}} \lambda_{\mathrm{i}} 1_{\{\mathrm{t} \in \mathrm{H_i}\}} \quad (5)$$

where *h(t)* represents the holiday effect, λi are the holiday effects coefficients, and H_i are the holiday intervals (Taylor and Letham, 2018).

Gaussian mixture model hidden Markov model (GMM-HMM)

GMM-HMM combines the advantages of Gaussian Mixture Models (GMM) and Hidden Markov Models (HMM) to model intricate temporal patterns and identifies anomalies in time series data (Bicego et al., 2019). The GMM component uses a mixture of Gaussian distributions to model the probability distribution of observations at each time step, allowing for a flexible representation of data with multiple underlying patterns. On the other hand, the HMM component captures the temporal dependencies between observations by defining a set of hidden states and transition probabilities between these states. By jointly learning the parameters of the GMM and HMM using Expectation-Maximization (EM) algorithms, GMM-HMM can effectively spot abnormal patterns and anomalies in the time series.

The joint probability distribution of the observations and hidden states in GMM-HMM can be represented as:

$$p(X, Z|\theta) = p(z_1|\pi) \prod_{t=2}^{T} p(z_t|z_{t-1}, A) \prod_{t=1}^{T} p(x_t|z_t, \mu, \Sigma) \quad (6)$$

where:

- $X = \{x_1, x_2, \dots, x_T\}$ are the observed data points,
- $Z = \{z_1, z_2, \dots, z_T\}$ are the hidden states,
- θ represents the model parameters,
- π is the initial state probability vector,
- A is the state transition probability matrix,
- μ and Σ are the mean and covariance matrices of the Gaussian components.

Integration of models for early warning system

The dam's radial deformations (RD206) situated on the right flank, are the target variable while ambient temperature, water level and six variables that represents the water temperatures on the right flank (avg1-R, avg2-R, avg3-R, avg4-R, avg5-R and avg6-R) are the features.

Therefore, let $W(t)$ represent the vector of all features (including RD206) at time t. The Prophet model generates forecasts for all features at time t+1 based on historical data to time t, represented as:

$$\widehat{W}(t+1) = Prophet\_Forecast(W(1:t)) \quad (7)$$

where, Ŵ(t+1) represents the forecasted values of all features at time t+1. Given the forecasted values Ŵ(t+1) for all features at time t+1, including RD206, the data is input into the GMM-HMM model for anomaly detection.

Let Z(t+1) represents the hidden states corresponding to the forecasted values Ŵ(t+1) at time t+1. The likelihood of observing forecasted data Ŵ(t+1) given a learnt GMM-HMM parameters θ can be computed using the forward algorithm:

$$p(\widehat{W}(t+1)\theta = \Sigma zp(\widehat{W}(t+1), Z|\theta) \quad (8)$$

θ (Theta) refers to the set of parameters in the machine learning model, specifically in the GMM-HMM framework. These parameters are essential because they determine how the model fits the data. In this context, θ includes:

GMM Parameters: In the Gaussian Mixture Model, θ consists of the weights, means, and covariances of the Gaussian components that describe the distribution of data. These parameters help in modelling the probability distribution of the observed data.

- **Weights (π)**: Represent the proportion of each Gaussian component in the mixture.
- **Means (μ)**: The centre of each Gaussian component, indicating where the data points are likely to cluster.
- **Covariances (Σ)**: Capture the spread and orientation of each Gaussian component, reflecting how much variability there is in different directions.

HMM Parameters: In the Hidden Markov Model, θ also includes parameters that describe the hidden state dynamics:

- **Transition probabilities (A)**: The probability of moving from one hidden state to another.
- **Emission probabilities (B)**: The probability of observing a particular set of data given the current hidden state.

Anomaly detection is then computed on the probability of observing the forecasted data given the learnt GMM-HMM parameters.

Mahalanobis distance

The Mahalanobis distance was introduced into the GMM-HMM to detect anomalies. The Mahalanobis distance can be used to identify outliers or anomalies by comparing the distance of each data point to the distribution of normal data points. Data points with Mahalanobis distances exceeding a certain threshold are considered anomalies, as they deviate significantly from the expected distribution (Mahalanobis, 1936). The Mahalanobis distance measures how many standard deviations away a point is from the mean of the distribution along each dimension, considering the correlation between dimensions:

$$D(x,\mu,\Sigma) = \sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)} \quad (9)$$

where x is the vector of observed values (data point) for which we want to compute the Mahalanobis distance, μ is the mean vector of the distribution of the dataset, it represents the average values of each feature. Σ is The covariance matrix of the distribution of the dataset. It represents the variances and covariances between each pair of features.

RESULTS AND DISCUSSION

Model tuning

The process of tuning the Gaussian Mixture Model - Hidden Markov Model (GMM-HMM) to achieve its optimal performance was primarily done by adjusting the number of components and the Mahalanobis distance threshold. It was observed that the number of components had a considerable influence on the precision of the model but had minimal impact on its recall and F1-score, figure 4. The highest level of precision was achieved when the number of components was set to 5 and 10. On the other hand, adjusting the Mahalanobis distance threshold seemed to have a more profound impact on all the evaluation metrics, figure 5. A higher recall and F1 score were observed at a threshold of 0.1. Following this, the number of components was further adjusted using the values 5 and 10 for more fine-tuning. The best performance was observed when the number of components was set to 10, table 1.

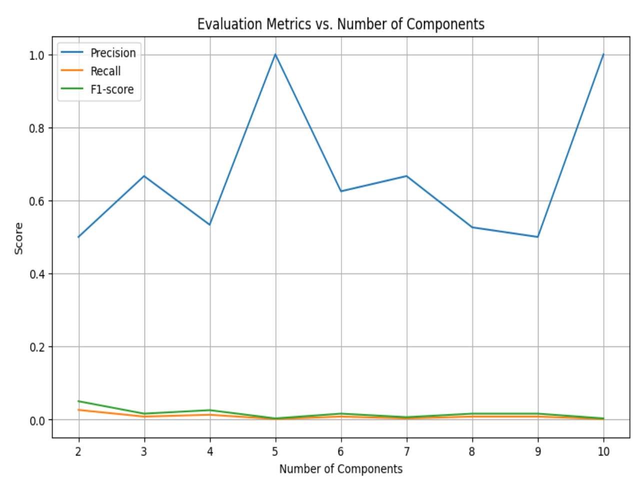

Figure 5. Evaluation Metrics against number of components

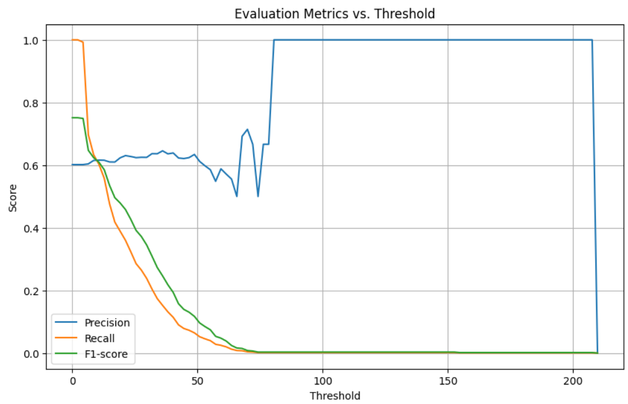

Figure 6. Evaluation metrices against threshold.

Table 1. Best performance.

Evaluation Matrices	Best value
Threshold	0.1
Precision	0.602
Recall	1.0
F-1 Score	0.751

The evaluation matrices in table 1 provided insightful information about the anomaly detection model's performance. Notably, a 0.1 threshold, demonstrating a balanced trade-off between precision and recall. The precision, approximately 0.602, shows that a significant portion of the flagged anomalies were indeed real, highlighting the model's aptitude for identifying true positives. Meanwhile, a recall score of 1.0 indicates the model's exceptional ability to detect all actual anomalies, reducing the chance of missing critical events. The F-1 score of 0.751 confirms the model's overall effectiveness, achieving a balance between precision and recall. This equilibrium is vital in practical applications, where missing anomalies or false alarms can have significant consequences. These optimal outcomes emphasize the model's efficacy in accurately identifying anomalies while maintaining high sensitivity. This boosts confidence in its use for important tasks, such as dam safety management.

Anomaly detection for forecasted radial deformations

The Prophet Model was utilized to generate forecasts under three distinct scenarios, each representing a different possible outcome. These scenarios were the forecasted lower boundary, the forecasted average, and the forecasted higher boundary. Upon careful examination of these predictions, anomalies were identified in only the forecasted upper scenario and the forecasted lower scenario. These anomalies are represented by black circles within the data visualization. The detected anomalies were discovered to fall within a specific range of radial displacements, specifically, between -0.01 and 0.01, figure 6. This range corresponds to the period when the dam is moving in the downstream direction, figure 7. It's important to note that this downstream movement is a result of the dam filling up. Interestingly, instances of anomalies have only been detected when the water levels within the dam dropped to low levels.

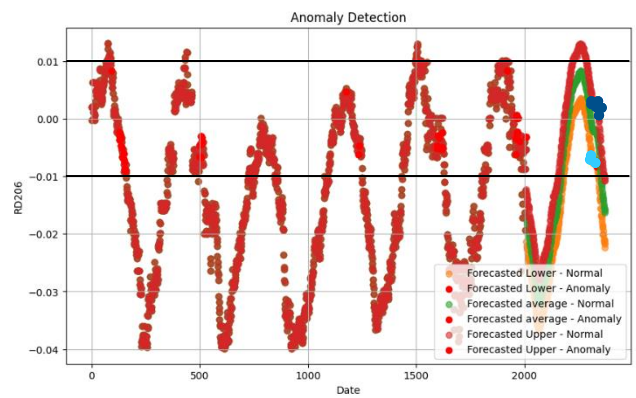

Figure 7. Forecasted anomalies.

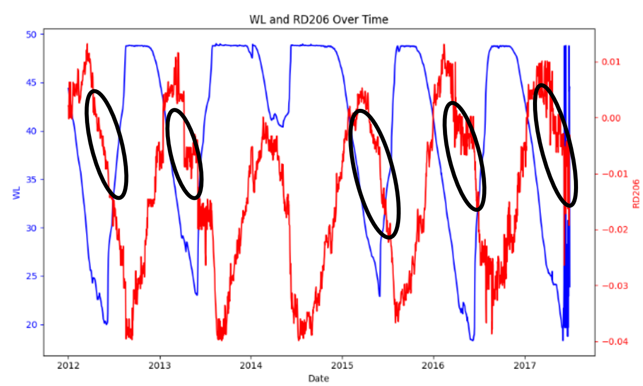

Figure 8. Regions where most anomalies occur.

CONCLUSIONS

The research has made significant strides in the field of dam deformation monitoring by leveraging the Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) framework in tandem with Prophet Forecasting. By integrating the predictive capabilities of Prophet Forecasting, which is renowned for its accuracy in predicting trends, with the anomaly detection strengths of the GMM-HMM, we provided an innovative, robust, and effective solution for real-time detection of deviations from expected deformation patterns. This synergistic approach holds considerable promise for augmenting the accuracy and reliability of dam deformation monitoring systems, thereby significantly mitigating risks associated with dam failure and ensuring the safety of these critical infrastructures. The results of the evaluation demonstrated promising performance of the integrated framework in anomaly detection for dam deformation data. The precision score of approximately 0.602 indicated a significant proportion of correctly identified anomalies, thereby highlighting the model's effectiveness in distinguishing between normal and anomalous deformation patterns. Furthermore, the recall score of 1.0 reflected the model's exceptional sensitivity and capacity to capture all actual anomalies without missing any true positive instances. In addition to these, the F1-score, which is a harmonic mean of precision and recall, stood at 0.751, affirming the balanced performance of the model in accurately identifying deviations from expected deformation patterns without favouring either recall or precision disproportionately. These evaluation metrics collectively underscore the effectiveness of the integrated GMM-HMM and Prophet Forecasting framework in enhancing early warning systems for dam safety and demonstrate its potential for broader applications in infrastructure monitoring.

DECLARATIONS

Corresponding author

Correspondence and requests for materials should be addressed to Thalosang Tshireletso; E-mail: tshireletsot@ub.ac.bw; ORCID: 0000-0002-5112-8077

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Acknowledgements

We would like to acknowledge the Department of Water and Sanitation, South Africa for allowing us to use their dam, Roode Elsberg. This research was funded by Water Research Commission, South Africa under project name Climate change impacts on the safety of dams in South Africa. Project number K5-2749

Authors' contribution

First Author designed the experimental process, performed the experiments, analysed the data obtained and wrote the manuscript. Second Author revised the manuscript. Both authors read and approved the final manuscript

Competing interests

The authors declare no conflict of interest.

REFERENCES

Benoit, L., Briole, P., Martin, O., & Thom, C. (2013). Real-time deformation monitoring by a wireless network of low-cost GPS. *Journal of Applied Geodesy*, 7(3), 161-172. https://doi.org/10.1515/jag-2013-0023

Bicego, M., Lovato, P., Rota, B., & Farinelli, A. (2019). Gaussian mixture models for time series clustering: a review. Pattern Recognition, 87, 298-314. https://doi.org/10.1016/j.patcog.2018.10.020

Šakić Trogrlić, R., van den Homberg, M., Budimir, M., McQuistan, C., Sneddon, A., & Golding, B. (2022). Early warning systems and their role in disaster risk reduction. In *Towards the "Perfect" Weather Warning: Bridging Disciplinary Gaps through Partnership and Communication* (pp. 11-46). Springer. https://doi.org/10.1007/978-3-030-98989-7

Pang, Z., Jin, Q., Fan, P., Jiang, W., Lv, J., Zhang, P., Cui, X., Zhao, C., & Zhang, Z. (2023). Deformation Monitoring and Analysis of Reservoir Dams Based on SBAS-InSAR Technology—Banqiao Reservoir. *Remote Sensing*, 15(12), 3062. https://doi.org/10.3390/rs15123062

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458), 611-631. https://doi.org/10.1198/016214502760047131

Coraça, E. M., Ferreira, J. V., & Nóbrega, E. G. O. (2022). An unsupervised structural health monitoring framework based on Variational Autoencoders and Hidden Markov Models. *Reliability Engineering & System Safety*, 227, 109025. https://doi.org/10.1016/j.ress.2022.109025

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 12(1), 49-55. Google Scholar

Rong, Z., Pang, R., Xu, B., & Zhou, Y. (2024). Dam safety monitoring data anomaly recognition using multiple-point model with local outlier factor. *Automation in Construction*, 151, 105290. https://doi.org/10.1016/j.autcon.2024.105290

Hassani, H., Rua, A., Sirimal Silva, E., & Thomakos, D. (2019). Monthly forecasting of GDP with mixed-frequency multivariate singular spectrum analysis. *International Journal of Forecasting*, 35(3), 821-830. https://doi.org/10.1016/j.ijforecast.2019.03.021

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37-45. https://doi.org/10.1080/00031305.2017.1380080

Bahrami, O., Wang, W., & Lynch, J.P. (2021). Hidden Markov models for sequential damage detection of bridges. In *Bridge Maintenance, Safety, Management, Life-Cycle*

Sustainability and Innovations (1st ed., pp. 7). CRC Press. eBook ISBN: 9780429279119. https://www.taylorfrancis.com/chapters/edit/10.1201/9780429279119-209/hidden-markov-models-sequential-damage-detection-bridges-bahrami-wang-lynch