Front. Eng. Manag. 2025, 12(3): 447–466
https://doi.org/10.1007/s42524-024-0082-1

RESEARCH ARTICLE

Limao ZHANG, Ying WANG, Xianlei FU, Xieqing SONG, Penghui LIN

Geological risk prediction under uncertainty in tunnel excavation using online learning and hidden Markov model

Abstract The accurate estimation of geological risks is essential for preventing geohazards, and ensuring efficient and safe construction processes. This study proposes a method, the online hidden Markov model (OHMM), which combines online learning with the hidden Markov model to estimate geological risks. The OHMM is tailored for the continuous nature of observational data, allowing it to adaptively update with each new piece of data. To address the challenge of limited data in the early stages of construction, we use pre-construction borehole samples as additional data. This approach extends the short sequence of observed data to match the length of a complete sequence through an observation extension mechanism. The effectiveness of the OHMM, equipped with this observation extension mechanism, is demonstrated in a case study that models geological risks for a tunnel excavation project in Singapore. The OHMM outperforms traditional methods, including the hidden Markov model, long short-term memory network, neural network, and support vector machine, in predicting geological risks ahead of the tunnel boring machine. Notably, the OHMM can accurately forecast geological risks in areas yet to be constructed, using limited observational and site investigation data. This research advances geological risk prediction models by offering an online updating capability for tunnel excavation and construction projects. It enables early-stage risk prediction and provides long-term forecasts with minimal historical data requirements, maximizing the use of site investigation data.

Keywords geological risk prediction, machine learning, online learning, hidden Markov model, borehole logging

1 Introduction

Geohazards, including collapse, water inrush, and landslide, have a significant impact on tunnel construction, resulting in project delays, cost overruns, and safety concerns. From 2002 to 2018, there were a total of 97 geohazard events during tunnel construction in mountainous regions of China, resulting in 393 fatalities, 467 injuries, and 51 missing persons (Wang et al., 2020a). Prior knowledge of geological conditions is crucial in preventing such adverse events and minimizing associated losses. This information is valuable at various stages of tunnel construction (Lu et al., 2019), including the selection of excavation methods and optimization of tunnel design in the pre-construction phase (Wang and Zhang, 2023; Zhang and Lin, 2021), prediction of operational parameters during construction (Fu and Zhang, 2021), and allocation of emergency resources in response to geohazards (Wan et al., 2021).

Traditionally, two types of techniques are commonly used to gather geological data: invasive methods, such as drilling, and non-invasive methods, like ground penetrating radar (GPR). Invasive methods are typically employed during earth expeditions to extract core samples for geological analysis or to conduct *in situ* measurements (Wang et al., 2020b). Drilling has the advantage of being able to reach the greatest depths, with scientific drilling going as deep as 12 km (Savenok et al., 2020). On the other hand, non-invasive methods, which are based on the geophysical properties of the ground, offer high

Received Aug. 20, 2023; revised Jan. 8, 2024; accepted Feb. 17, 2024

Limao ZHANG, Xieqing SONG
School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Ying WANG, Xianlei FU, Penghui LIN (✉)
School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore
E-mail: penghui003@e.ntu.edu.sg

This work is supported in part by the National Natural Science Foundation of China (Grant No. 72271101) and the Outstanding Youth Fund of Hubei Province (Grant No. 2022CFA062).

spatial resolution. For instance, seismic methods like vertical seismic profiling (VSP), horizontal seismic profiling (HSP), and tunnel seismic prediction (TSP) utilize seismic reflection properties (Li et al., 2017). A detailed comparison of common geological prospecting techniques is presented in Table 1. However, none of the methods reviewed can provide a viable solution for accurately detecting volumes of the areas that would be covered by planned tunnel routes. In particular, borehole drilling yields relatively accurate information but at a sparse sampling rate and with limited detection resolution. Non-invasive methods offer high spatial resolution but lack the same level of accuracy due to measurement errors and errors resulting from signal processing. Therefore, the tunneling industry currently needs a method to describe geological information and accurately assess corresponding risks to support tunnel construction effectively.

In the past decade, there has been an explosion of artificial intelligence (AI) applications in construction engineering and risk management (Zhou et al., 2021). In particular, a multitude of machine learning methods have been employed to address the challenge of predicting geological conditions ahead of tunnel boring machines (TBMs). However, the efficiency of machine learning techniques is often limited by the assumption that there are sufficient measurements available for the system being modeled. Tunneling projects are massive undertakings that can extend over several or even tens of years, which is significantly longer than the time needed for model updating. This creates a dilemma where high-resolution geological conditions are desired before excavation, but data about these conditions arrive incrementally as the excavation progresses. It has been shown that additional data from the excavation of a 200-m tunnel section at the inlet and outlet locations can greatly enhance the prediction performance of rock mass rating (RMR) for each chainage of the tunnel (Mahmoodzadeh et al., 2021). However, the updating process was only performed once and relied on data that could be updated with chainage-specific information. Furthermore, accurately estimating the underlying geological risk remains challenging despite the ease of recording geological observations.

The hidden Markov model (HMM) is commonly used to analyze hidden states that are difficult to measure based on easily measurable observations (Li et al., 2022). This study focuses on utilizing streaming chainage data to infer hidden geological risk using an online hidden Markov model (OHMM). To overcome limited data during the early construction stage, an observation extension mechanism is proposed to expand short observation sequences to the length of the full observation sequence, which characterizes the geological risk of the tunnel coverage region. This study advances the field of geological risk prediction by providing a high-resolution method (per chainage/ring) with the following contributions: (1) the systematic OHMM approach allows for frequent model updating based on new data for each chainage/ring, and (2) the observation extension mechanism incorporates feedback between the updated OHMM prediction and true states from sparse borehole samples, making it suitable for early construction applications.

The remainder of the paper is organized as follows: Section 2 provides a review of related studies on geological risk prediction. Section 3 presents the proposed approach, including data accumulation, online learning of HMM, feedback from sparse borehole samples, and performance evaluation. In Section 4, a case study is conducted to verify the proposed approach, focusing on modeling the geological risk of a TBM-excavated tunnel in Singapore. Section 5 compares the proposed model with other prediction models, examines the effect of a key OHMM hyper-parameter (number of rings ahead) on forward accuracy, and evaluates the model's performance on limited data. Finally, Section 6 concludes the study and outlines potential future directions.

2 Literature review

Tunnel excavation plays a crucial role in infrastructure development as it enables passage through various geological formations. However, the geological risks associated with tunnel construction present significant challenges, including delays, increased costs, and safety concerns. Traditionally, borehole logging has been used to record geological information. However, this method has limited spatial resolution in the horizontal plane and

Table 1 Comparison of ahead geological prospecting techniques

Methods	Invasive	Applications	Detection range	Spatial resolution
Borehole logging (Savenok et al., 2020)	Yes	Vertical mapping	Up to 12 km	Depending on the sampling distance, usually sparse
Seismic methods (Gandhi, 2016; Li et al., 2017)	No	Oil exploration, reservoir characterization, subsurface stratigraphy	500 – 2000 m	High
Electromagnetic methods (Wolfsberg, 1997; Yazdani et al., 2018)	No	Obtaining structural heterogeneities and lithological interfaces, evaluating geo-mechanical parameters	3 – 10 m	Very high
Electrical and induced polarization methods (Dakir et al., 2019; Li et al., 2017)	No	Determining the geological discontinuities	Around 54 m	High, decreases rapidly with depth

is restricted by the sampling distance.

To address these limitations, conventional geological models have been developed for geology modeling and risk assessment through mapping techniques. For instance, Xiong et al. (2018) proposed a geological model that visualizes geological boundaries, such as strata and discontinuities, using data from regional field surveys. In addition to earlier 2D geological models, which focus on cross-sections, more recent 3D geological models connect multiple cross-sections to provide an additional dimension along the path connecting separate cross-sections (Lin et al., 2017; Peng and Li, 2021). These visualization models offer a clear understanding of the geological distribution, including information about strata and discontinuities.

In recent years, the advancement of machine learning techniques has greatly motivated research in geological hazard analysis from presentation to prediction. For instance, a long short-term memory (LSTM) network with a global attention mechanism was utilized to classify rock mass and optimize TBM operational parameters based on TBM data (Liu et al., 2021). A variant of the K-means algorithm was also developed to identify geology clusters by analyzing various TBM data channels, such as cutter head speed, cutter head torque, thrust, and TBM advance rate (Erharter and Marcher, 2020). Gaussian process regression, support vector machine (SVM), and decision tree (DT) techniques were employed to estimate rock quality for tunnel construction (Mahmoodzadeh et al., 2021). These methods have demonstrated high applicability and performance in accurately predicting geological conditions during tunneling. However, predicting geological conditions ahead of the TBM involves not only identifying the geological conditions but also determining the size and location of changes to support on-site decision-making (Sheil et al., 2020). Most geological prediction models rely on pre-excavation data and do not account for continuously arriving data, which is recorded with the spatial resolution of a chainage during excavation. Failure to consider this new data can lead to a failure to identify valuable information that helps constrain uncertainties.

When it comes to modeling potential uncertainties, the HMM stands out as a cutting-edge approach for analyzing sequential data. It has been successfully applied to a variety of fields including audio segmentation (Bietti et al., 2015), network attack prediction (Chadza et al., 2020), rainfall monitoring (Stoner and Economou, 2020), and petroleum reservoir detection (Huang et al., 2017). The HMM is a generative model that focuses on understanding the underlying mechanisms and dynamics of observation generation driven by hidden states (Sicking et al., 2020).

For example, in their study, Chis and Harrison (2015) proposed an OHMM method that combines a sliding HMM and a multi-input HMM to tackle the challenge of predicting infrequent additional loads in computer systems. This approach significantly reduces complexity and learning time while maintaining cost-effectiveness and prediction accuracy. Although the HMM method has shown promise in computer systems, its potential in tunnel construction remains unexplored and worth investigating. Instead of focusing solely on improving computing efficiency, prioritizing the enhancement of multistep forecasting ability can provide sufficient time for the operations to be responded to by the TBM.

Online learning, a subset of machine learning methods, offers adaptability to streaming data and real-time prediction by leveraging historical data (Blum, 1998). Unlike batch learning, which processes data as a whole set, online learning utilizes each sample as it is observed. This updating process not only improves efficiency but also enables scalability in large-scale applications (Hoi et al., 2021). The scenario of online learning aligns perfectly with the need to predicting geological risks ahead of TBM tunnel excavation by continuously utilizing incoming data. To take advantage of these benefits, Cao et al. (2021) developed an online method for detecting geological anomalies based on TBM parameters. Similarly, Erharter et al. (2020) proposed an online artificial neural network (ANN) method that classifies rock mass behavior using geological information extracted from TBM cutterhead behavior. However, both approaches heavily rely on the availability of geological information, which limits their usability.

In summary, a significant limitation in predicting geological risk in tunnel excavation is the availability and utilization of data. Traditionally, machine learning methods have failed to address uncertainties stemming from new data chainages. Additionally, while borehole data remains the most direct and accurate way to describe geological conditions, effectively utilizing this data remains an ongoing concern in this field.

3 Methodology

The safe construction of TBM-excavated tunnels relies on knowledge of the geological conditions ahead of the TBM and the perceived geological risk. However, accurate prediction of geological risk ahead of the TBM requires historical data, which is typically obtained during the excavation process. This delay in data collection creates a dilemma between the desired informative time (before excavation) and the actual informative time as determined by data availability (post-excavation). It is important to note that tunnel construction, being a megaproject, often takes months to years to complete. Therefore, it is practical to have the capability to inform geological risk ahead of the TBM with a short forecasting distance of several meters. Furthermore, predicting geological risk is challenging due to its dependence on multiple factors, such as

soil components (Xu et al., 2021a), environmental loading and unloading (Xu et al., 2021b), and ambient moisture (Vereecken et al., 2022). To address these challenges, we deploy and update an HMM to inform the hidden geological risk of the TBM face based on available geological observations, including on-site observations and sampling analysis results. This developed method is referred to as OHMM. Moreover, an observation extension mechanism is proposed to overcome the issue of limited data at an early construction stage by utilizing sparse borehole samples. The framework of OHMM, which includes in situ data accumulation, the OHMM learning algorithm, the OHMM feedback from sparse samples (i.e., observation extension mechanism), and performance evaluation, is illustrated in Fig. 1.

3.1 *In situ* data accumulation

Information regarding the geology ahead of the TBM is crucial in minimizing construction accidents and machine defects during the construction process (Cao et al., 2021). During the pre-construction phase, samples from boreholes are carefully chosen and analyzed to provide preliminary geological information about the conditions in the area. However, these cylindrical drilling holes, typically with a diameter as small as 0.15m (Thalund-Hansen et al., 2023), only represent a limited section in the horizontal plane where the tunnel lies. This is illustrated by the small intersection part compared to the large horizontal plane in Fig. 2(a). Once the construction process commences, the TBM excavates along the planned tunnel path, with the tunnel being built ring by ring. The arrangement of rings for a particular tunnel section is shown in Fig. 2(b). The sparse drilling holes, obtained before excavation, are insufficient in characterizing the geological risks that guide the excavation process. Therefore, this study incorporates continuous in situ geological observations, gathered during the excavation process, into the problem of inferring and predicting geological risks with fine resolution (i.e., one ring).

In practice, the construction process is diligently recorded in daily reports. Following good practice, these reports adhere to a consistent data input structure, which includes detailed information about the schedule, technical parameters, soil types, on-site photos, and more. Geological observations can be extracted accordingly. However, these reports typically do not include information regarding geological risks. Geological risks are influenced by the

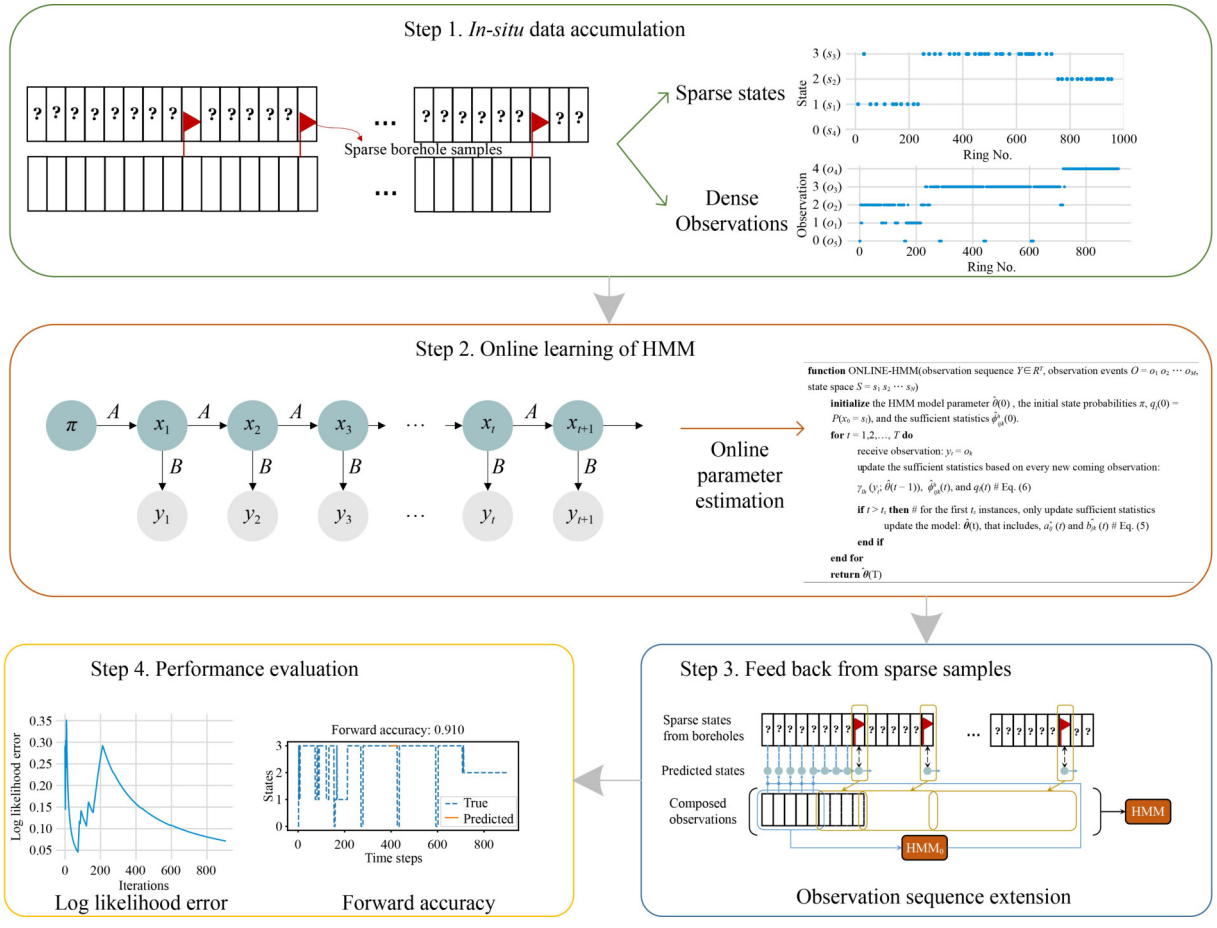

Fig. 1 A framework of the proposed OHMM in predicting geological risk.

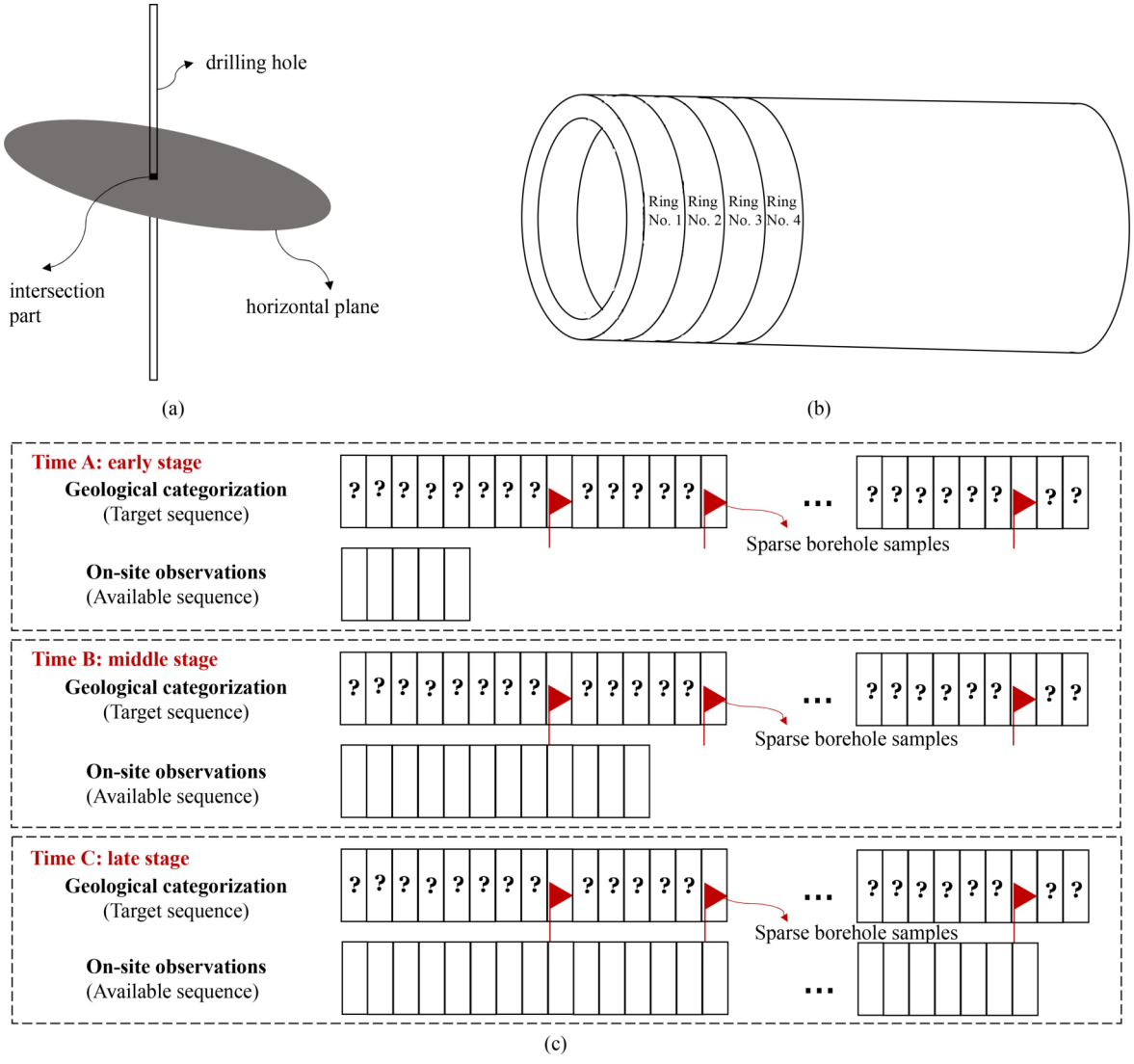

Fig. 2 Illustration of the problem: geological risk inference and prediction to inform excavation. (a) Schematic of limited geological information from drilling hole due to sparse samples; (b) Schematic of tunnel rings; (c) Gradually increased data volume of observations to infer desired yet unknown states from the early to the middle, and further to the late construction stages. Both sequences are recorded with one-ring (indicated by one rectangle) resolution, where empty rectangles indicate available data while question marks indicate unknown states; red flags indicate known states from borehole analysis, which has a multi-ring resolution and is sparser than the required.

characteristics and properties of in situ rocks and soils (Golpasand et al., 2018). Assessing geological risks, even with some guiding standards, is indeed a challenging task that necessitates expert knowledge and engineering experience. Throughout the construction process, the number of observations steadily increases while the geological risks remain unknown. The scenarios corresponding to three stages (early, middle, and late stages) are depicted in Fig. 2(c). Among all the rings, only a small number that intersect with the boreholes have their geological risks informed through borehole analysis, while the remainder remains unknown. In this study, the geological risks corresponding to the tunnel rings are explicitly labeled for validation purposes by employees closely involved in the project.

The proposed method aims to infer the unknown states of geological risks based on available observations and further extend the information from excavated rings to rings that are yet to be excavated. The model is one-dimensional, following the direction of the tunnel alignment, and operates at a fine spatial resolution. The necessary data, retrievable from construction daily reports adhering to a standard format, consist of (i) points pertaining to geological risks obtained from borehole analysis; and (ii) a sequence of geological observations whose length varies with the volume of excavation. The three assumptions of the model are listed below.

(i) The model does not include the depth of the tunnel.

(ii) Geological conditions for each ring in the cylindrical space are consistent, meaning that there is no variation.

Typically, the ring has a diameter of several meters and a length of 1 m (Cho et al., 2017).

(iii) Changes in geological conditions only occur between rings along the alignment direction, which is the direction of TBM advance if it is perfectly aligned.

3.2 Online learning of HMM

A Markov chain, which serves as the foundation for HMM methods, is a memoryless stochastic process where the probability of a state depends only on the previous state in a first-order Markov chain. A higher-order Markov chain retains more memory about the states and can be either time-homogeneous or time-inhomogeneous. The Markov assumption for a k^{th} order Markov model is modified and stated such that the probability of the current states, given all the previous states, is the same as the probability of the current state given the previous k states. In a time-homogeneous Markov chain, the probability of any state transition is independent of the time steps. In a time-inhomogeneous Markov chain, the transition probability is non-stationary and may vary with the steps. Transition matrices possess the property that a multi-step transition is equal to the product of step-by-step transition matrices. Mathematically, the k-step transition matrix follows Eq. (1), and its expanded form is shown in Eq. (2):

$$P_t^{(m)} = P_t \cdot P_{(t+1)} \cdots P_{(t+m-1)}, \tag{1}$$

$$P_t^{(m)} = \begin{pmatrix} P(x_{t+m}=s_1 \mid x_t=s_1) & \cdots & P(x_{t+m}=s_N \mid x_t=s_1) \\ \vdots & & \vdots \\ P(x_{t+m}=s_1 \mid x_t=s_N) & \cdots & P(x_{t+m}=s_N \mid x_t=s_N) \end{pmatrix}, \tag{2}$$

where $P_t^{(m)}$ is the m-step transition matrix from state x_t to state x_{t+m}, $P_t, P_{t+1}, \ldots, P_{t+m-1}$ are the transition matrices at steps $t, t+1, \ldots, t+m-1$, and the set $\{s_1, s_2, \ldots, s_N\}$ is the state space.

When states cannot be directly observed, the concept of hidden states is introduced alongside observable events in a parallel manner. This gives rise to a double-sequence model with inference capability known as HMM. Similar to a Markov chain, an HMM simulates a memoryless process that allows for multi-step transitions between hidden states and connections between hidden states and observations. The mathematical representation of an HMM is as follows: (i) state set $S = \{s_1, s_2, \ldots, s_N\}$; (ii) transition probability matrix $A = [a_{11} \cdots a_{ij} \cdots a_{NN}]$ with a_{ij} representing the probability of moving from state s_i to state s_j, and $\sum_{j=1}^{N} a_{ij} = 1$; (iii) observation set $O = \{o_1, o_2, \ldots, o_M\}$; (iv) emission probability matrix $B = [b_{11} \ldots b_{jk} \ldots b_{NM}]$ with b_{jk} being the probability of generating an observation o_k from a state s_j; and (v) initial probability distribution $\pi = [\pi_1 \pi_2 \ldots \pi_N]$. The HMM explains the generation process of observations, often denoted by observation sequence $Y = y_1, y_2, \ldots, y_T$. Usually, S and O are specified according to the problem context; Y is the collected data; and A, B, π specify the model with inference capability.

The OHMM is defined by the same parameters as HMM including state space S, observation space O, transition probability matrix A, emission probability matrix B, and initial probability distribution π. An explanation of the parameters is shown in the preliminary Section 2.2. The difference between OHMM and HMM lies in the updating rules specified by recursion equations. A graphical illustration of the HMM structures is shown in Fig. 3. Besides the defining parameters, extra learning parameters are introduced for model updating of HMM as well as OHMM. The learning problem of OHMM is stated as below: given the sequence of observations $Y = y_1, y_2, \ldots, y_T$, to infer the corresponding hidden states $X = x_1, x_2, \ldots, x_T$, the model parameters $\theta = (A, B, \pi)$ are to be learned. It should be mentioned that the parameters $\theta = (A, B, \pi)$ are sufficient to specify the observation generation process (i.e., observing Y). Details of updating rules are described below.

The HMM updates are based on the offline algorithm called the forward-backward algorithm. While OHMM updates its parameters by recursively updating the intermediate sufficient statistics (Mongillo and Deneve, 2008). The transition probability of OHMM is estimated as the fraction of the probability of transiting from the state s_i to state s_j and the probability of transiting from state s_i to any state. Similarly, the emission probability of OHMM is estimated as the fraction of the probability of observing o_k at state s_j and the probability of observing any observation at the state s_j. Their mathematical forms are shown in Eq. (3).

$$\hat{a}_{ij}(T) = \frac{\sum_k \sum_h \hat{\phi}_{ijk}^h(T)}{\sum_{j,k} \sum_h \hat{\phi}_{ijk}^h(T)}, \quad \hat{b}_{jk}(T) = \frac{\sum_i \sum_h \hat{\phi}_{ijk}^h(T)}{\sum_{i,k} \sum_h \hat{\phi}_{ijk}^h(T)}, \tag{3}$$

where $\hat{a}_{ij}(T)$ is the estimated transition probability from state s_i to state s_j based on a sequence of length T, $\hat{b}_{jk}(T)$ is the estimated emission probability of observing o_k from state s_j based on a sequence of length T, and $\hat{\phi}_{ijk}^h(T)$ is the sufficient statistics representing the sum over the time index of the probability of being at the state s_h and observing o_k at time index t with some previous state

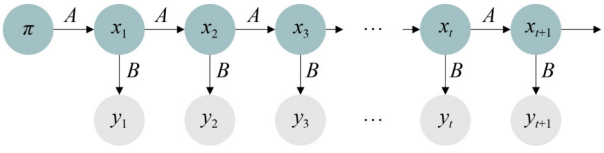

Fig. 3 Structure of a generic hidden Markov model.

transition from state s_i to state s_j conditioned by observing the sequence of length T, i.e., $Y=y_1,y_2,\ldots,y_T$. The mathematical form of $\hat{\phi}_{ijk}^{h}(T)$ is expressed as:

$$\hat{\phi}_{ijk}^{h}(T)=\frac{1}{T}\sum_{t=1}^{T}\delta(y_t-o_k)\cdot P(x_{t-1}=s_i,\,x_t=s_j,\,x_T=s_h\,|\,Y),$$

where the pre-factor $\frac{1}{T}$ makes sure that its value does not diverge for an infinitely long sequence, T is the length of the sequence, $\delta(\cdot)$ is the Kronecker delta function whose output is 1 for zero arguments and 0 otherwise, y_t is the observation at time index t, O_k is the value of y_t, and x_{t-1},x_t,x_T are the states corresponding to the time index $(t-1)$, t, and T, respectively. The 4 indices of $\hat{\phi}_{ijk}^{h}(T)$ are the consecutive historical states s_i to s_j, the historical observation O_k from s_j, and the latest state s_h. Any irrelevant index is eliminated by summing over the full index range, that is, i,j,h from 1 to N, and k from 1 to M in the parameter estimation process.

The parameters are updated as the sufficient statistics are updated. The recurrence relation for the sufficient statistics corresponding to two consecutive time indices is given by Eq. (4), which represents the update based on a new sample. For a detailed derivation, refer to (Mongillo and Deneve, 2008). Model parameters are closely related to the introduced statistical variables.

$$\begin{aligned}\hat{\phi}_{ijk}^{h}(T)=&\sum_l\gamma_{ln}\left(y_T;\hat{\theta}(T-1)\right)\times\left\{\hat{\phi}_{ijk}^{l}(T-1)+\eta(T)\left[\delta(y_T-O_k)\right.\right.\\&\left.\left.\cdot g_{ij}(l,h)\cdot q_l(T-1)-\hat{\phi}_{ijk}^{l}(T-1)\right]\right\},\end{aligned}\tag{4}$$

where $\hat{\phi}_{ijk}^{h}(T)$ and $\hat{\phi}_{ijk}^{l}(T-1)$ are the sufficient statistics corresponding to two consecutive time indices, $\gamma_{lh}\left(y_T;\hat{\theta}(T-1)\right)$ is a complicated probability term that considers the one-step transition of states, emission from the state, and the addition of a new observation. Its value can be calculated based on the latest model parameters. It is defined and computed as

$$\begin{aligned}\gamma_{lh}\left(y_T;\hat{\theta}(T-1)\right)&\equiv\frac{P(y_T=O_k|x_T=s_h)\,P(x_T=s_h|x_{T-1}=s_l)}{P(y_T|y_1,y_2,\ldots,y_{T-1})}\\&=\frac{\hat{\alpha}_{lh}(T-1)\cdot\hat{b}_{hk}(T-1)}{\sum_{m,n}\hat{\alpha}_{mn}(T-1)\cdot\hat{b}_{nk}(T-1)\cdot q_m(T-1)},\end{aligned}$$

with $q_m(T-1)$ being the probability of being state s_m given an observation sequence of length $T-1$, i.e., $y_1,y_2,\ldots,y_{T-1}$. It can be calculated recursively (will be given soon) and is defined as:

$$q_m(T-1)\equiv P(x_{T-1}=s_m\,|\,y_1,y_2,\ldots,y_{T-1}),$$

$\eta(T)$ is a time-dependent discount factor, for example,

$$\eta(T)=\frac{1}{T},$$

$\delta(\cdot)$ is the Kronecker delta function whose output is 1 for zero arguments and 0 otherwise, y_T is the latest observation with time index T, O_k is the value of y_T, $g_{ij}(l,h)$ is a function defined as:

$$g_{ij}(l,h)\equiv\delta(i-l)\cdot\delta(j-h),$$

and $q_l(T)$ is the probability of being state s_l given an observation sequence of length T, i.e., $y_1,y_2,\ldots,y_T$. It is updated recursively by:

$$\begin{aligned}q_l(T-1)&\equiv P(x_{T-1}=s_l|y_1,y_2,\ldots,y_{T-1})\\&=\sum_m\gamma_{ml}\left(y_{T-1};\hat{\theta}(T-2)\right)q_m(T-2).\end{aligned}$$

The initialization and updating process for OHMM learning is described in Algorithm 1. In contrast, Algorithm 2 outlines the offline parameter estimation process of HMM, specifically the forward-backward algorithm, also known as the Baum-Welch algorithm. It is evident that the online algorithm updates sufficient statistics based on each new observation and the most recent model, which is then used to update the model itself. Conversely, the offline algorithm calculates sufficient statistics utilizing the entire sequence. Consequently,

Algorithm 1. Online parameter estimation for the OHMM.

function ONLINE-HMM(observation sequence $Y\in R^{\mathrm{T}}$, observation events $O=o_1o_2\cdots o_M$, state space $S=s_1s_2\cdots s_N$)

initialize the HMM model parameter $\hat{\theta}(0)$, the initial state probabilities π, $q_l(0)=P(x_0=s_l)$, and the sufficient statistics $\hat{\phi}_{ijk}^{h}(0)$.

for $t=1,2,\ldots,T$ **do**

receive observation: $y_t=o_k$

update the sufficient statistics based on every new coming observation: $\gamma_{lh}\left(y_t;\hat{\theta}(t-1)\right)$, $\hat{\phi}_{ijk}^{h}(t)$, and $q_l(t)$ # Eq. (6)

if $t>t_s$ **then** # for the first t_s instances, only update sufficient statistics

update the model: $\hat{\theta}(t)$, that includes, $\hat{a}_{ij}(t)$ and $\hat{b}_{jk}(t)$ # Eq. (5)

end if

end for

return $\hat{\theta}(\mathrm{T})$

Algorithm 2. The offline forward-backward algorithm for the HMM.

function FORWARD-BACKWARD (observation sequence $Y \in R^n$, observation events $O = o_1 o_2 \cdots o_M$, state space $S = s_1 s_2 \cdots s_N$)

initialize A, B, and π

iterate until convergence

E-step

$$\xi_t(i,j) = \frac{f_t^i a_{ij} b_j(y_{t+1}) b_{t+1}^j}{\sum_{j=1}^{N} f_t^j b_t^j} \forall t,\ i, \text{ and } j$$

$$\zeta_t(j) = \frac{f_t^j b_t^j}{\sum_{j=1}^{N} f_t^j b_t^j} \forall t \text{ and } j$$

where f_t^i is the forward probability at time t and state s_i, b_t^j is the backward probability at time t and state s_j, and $b_j(y_{t+1})$ is the emission probability at state s_j with observation y_{t+1}

M-step

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{n-1} \xi_t(i,j)}{\sum_{t=1}^{n-1} \sum_{k=1}^{N} \xi_t(i,k)}$$

$$\hat{b}_{jk} = \frac{\sum_{t=1 s.t. y_t = o_k}^{n} \zeta_t(j)}{\sum_{t=1}^{n} \zeta_t(j)}$$

return A, B

online learning of HMM is more efficient and scalable than offline learning in handling long, continuous sequences.

3.3 Feedback from sparse samples

During the early stages of construction, the observation sequence may not be long enough for the OHMM to converge. This can lead to overfitting problems due to the limited amount of observation data, thus adversely affecting performance in prediction. To address this issue and ensure the model is suitable for observation sequences of varying lengths that correspond to different construction stages, we propose a mechanism that explicitly extends short observation sequences to match the full sequence. This mechanism utilizes the most updated model parameter estimates and the sparse samples in advance. The convergence checking and observation sequence extension mechanism are detailed below.

An OHMM with discrete observation values can be characterized by a transition matrix A and emission matrix B, which is usually denoted as $\theta = (A, B)$. To track the model updates due to the filled observation sequence, matrix distance (MD) is used to quantify the change of the two key matrices (Langfield-Smith and Wirth, 1992). The sum of the difference is divided by the number of distinct states in the state space (i.e., N) so that the resultant value is in the range $[0, 1]$, where 0 means no change and 1 means that two matrices are completely different. The change in the transition matrix and emission matrix is defined in Eqs. (5)–(6), respectively. The normalized matrix distance is utilized to monitor the change in parameter estimates. A default criterion of 0.0005 is employed to determine if the OHMM has converged.

$$\mathrm{d}(A(t+1), A(t)) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| a_{ij}(t+1) - a_{ij}(t) \right|, \quad (5)$$

where $d(A(t+1), A(t))$ is the change of the transition matrix from time step t to $t+1$.

$$\mathrm{d}(B(t+1), B(t)) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} \left| b_{ij}(t+1) - b_{ij}(t) \right|, \quad (6)$$

where $d(B(t+1), B(t))$ is the change of the emission matrix from time step t to $t+1$.

During the early stages of construction, when the full observation sequence is not yet available, a preliminary HMM is trained using the limited short observation sequence that is available. This preliminary model is then used to infer the most probable states. At the locations where boreholes are situated and the states are known, the inferred states are compared to the known states. The comparison results are used as feedback to extend the short observation sequence to the length of the full observation sequence. The composed observation sequence consists of two parts: the available observations and the filled observations based on the feedback from the sparse states obtained from borehole samples and the inferred states from the preliminary HMM. This composed observation sequence is then used to train the HMM model for prediction purposes. The feedback process from the sparse borehole samples is illustrated in Fig. 4.

The principle of minimality, which is a primary focus in data repair and data cleaning (Arieli et al., 2007; Rekatsinas et al., 2017; Zhang et al., 2017), is employed in the observation sequence filling process. The filling of the observation sequence follows the principle of minimal change. In the context of HMM, the rule for data filling is as follows: if the next state (which may be several steps ahead) remains the same as the current state, the observation remains unchanged. However, if the next state is different from the current state, there is a single jump in the observation value at a random step from the current time index

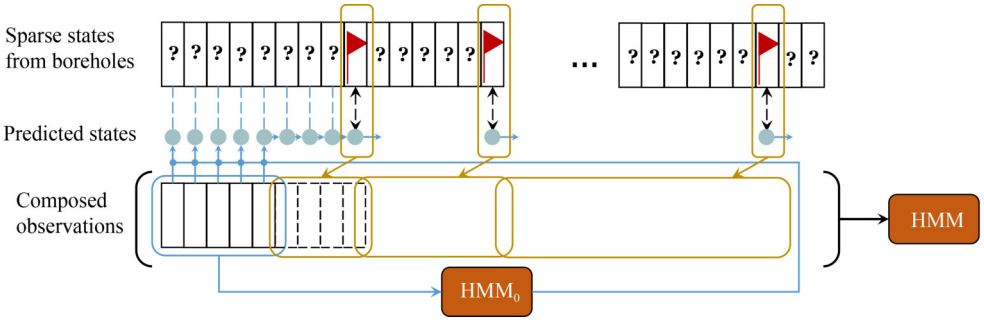

Fig. 4 A schematic illustrating the feedback from sparse samples. Available observation is used to train a preliminary HMM (denoted as HMM_0). For the rings with observations (indicated by rectangles with solid lines), the model HMM_0 is used to infer the states (indicated by non-connected circles) from the observations; for the rings without observations (indicated by rectangles with dashed lines), the most possible states (indicated by arrow-connected circles) are directly predicted based on the most recent state and the transition matrix of HMM_0. At the borehole position (indicated by red flags), the predicted state is compared to the known state, whose results are used to fill in unavailable observations. Available observations (indicated by a blue rectangle) and filled observations (indicated by golden rectangles) together form the composed observations, which are then used to train the HMM model for prediction purposes.

to the time corresponding to the next known state. The desired observation symbol is the one that is observed, and the next known state is most likely to occur at the corresponding time step. The mechanism for extending the observation sequence by leveraging HMM inference and known sparse states is shown in Algorithm 3.

3.4 Performance evaluation

To monitor the online learning process of OHMM, a log-likelihood error is defined to quantify the difference between the most probable states and the true states, as shown in Eq. (7). At each time step, using the updated model $\hat{\theta}(t)$ and uniform initial probabilities, the most possible hidden state sequence (up to time step t) is discovered using the Viterbi algorithm. The obtained posterior probabilities are used to calculate the log-likelihood, which is compared to the log-likelihood corresponding to the true state sequence.

$$\lambda(t) = L(t) - L_{\text{true}}, \tag{7}$$

where $\lambda(t)$ is the log-likelihood error, $L(t)$ is the average log-likelihood corresponding to the most probable states inferred from the model $\hat{\theta}(t)$, and L_{true} is the average log-likelihood corresponding to the true state sequence.

The developed OHMM can be deployed to infer the hidden geological risk based on available geological observations, as well as to predict the hidden geological risk for future time steps. These two tasks are referred to as inference and prediction, respectively. Inference ability is evaluated using backward accuracy, which involves inferring the hidden states based on the observation sequence. Prediction ability is assessed using forward accuracy, which extends the interpretation of hidden states to the next ΔT steps where observations are unavailable. Further details of these deployment scenarios are provided in Table 2. However, in practical terms, predicting future geological risks holds much greater significance. Early alerting of geological risks provides operators with sufficient time. Forward accuracy evaluates the states of future ΔT steps, as shown in Eq. (8).

$$\begin{aligned} Acc_{\text{forward}}(t,\Delta T) &= f(\hat{x}_{t+1},\hat{x}_{t+2},\ldots,\hat{x}_{t+\Delta T};x_{t+1},x_{t+2},\ldots,x_{t+\Delta T}) \\ &= 1-\frac{\sum_{i=t+1}^{t+\Delta T}|\hat{x}_i - x_i|}{\Delta T}, \end{aligned} \tag{8}$$

where $f(\cdot)$ is the accuracy score for the predicted sequence and true sequence, ΔT is the steps ahead in forward prediction, $\hat{x}_{t+1},\hat{x}_{t+2},\ldots,\hat{x}_{t+\Delta T}$ denotes the predicted states of the future ΔT steps, and $x_{t+1},x_{t+2},\ldots,x_{t+\Delta T}$ denotes the true states of the future ΔT steps.

4 Case study

To demonstrate the effectiveness and capability of the proposed approach, real data has been collected from one of the tunnel projects in Singapore for the case study. Further details about the background, model development, and result analysis are provided below.

4.1 Background

Geological risk refers to the potential occurrence of adverse events associated with natural and/or human-induced geohazards, such as sinkholes, landslides, land subsidence, and waterlogging (Eremina et al., 2018). Geomaterials exhibit unique geochemical and mechanical properties that vary during the complex process of loading and unloading, which critically influence failure mechanisms, as seen in tunnel excavation (Xiang et al., 2021). Insufficient information about the heterogeneous soil conditions, for instance, has led to increased uncertainty in estimating geological risks for tunneling projects (Sundell et al., 2019). Current approaches to assessing

Algorithm 3. Generation of composed observation sequence.

```
function extend_obs(model, Y_observed, pre_ring_no_idx, pre_ring_no_state)
    # model is an object of the initialized HMM model with defined functions
    seq_len=len(Y_observed) # seq_len in the range [1,T-1]
    # pre_ring_no_idx is the ring no. indices corresponding to sparse borehole samples
    # pre_ring_no_state is the states obtained from sparse borehole samples
    # find the index of pre_ring_no_idx where its element is large than seq_len for the first
    time
    for j, idx in enumerate (pre_ring_no_idx) do
        if idx > seq_len
            start=j
            break
        end if
    model.fit(Y_observed) # model learning
    # get the ob so that the next state will be the same as the given state
    next_ob=model.fill_next_ob_from_state(pre_ring_no_state[start])
    Y_filled= repeat(next_ob, idx-seq_len+1) # get the filled obs
    Y=concatenate((Y_observed, Y_filled)) # get the composed seq
    for j=start, …, len(pre_ring_no_idx)-1 do
        if pre_ring_no_state[j+1]== pre_ring_no_state[j]
            Y_filled_new=repeat(Y_filled[-1],
            pre_ring_no_idx[j+1]-pre_ring_no_idx[j])
        else
            model.fit(Y_observed)
            next_ob=model.fill_next_ob_from_state(pre_ring_no_state[j+1])
            # randomly pick the change position
            jump_pos=randrange(pre_ring_no_idx[j], pre_ring_no_idx[j+1])
            Y_filled_p1= repeat(Y_filled[-1], jump_pos- pre_ring_no_idx[j])
            Y_filled_p2= repeat(next_ob, pre_ring_no_idx[j+1]-jump_pos)
            Y_filled_new=concatenate(Y_filled_p1, Y_filled_p2)
        Y_filled=concatenate((Y_filled, Y_filled_new))
        Y=concatenate((Y_observed, Y_filled))
        end if
    end for
    return Y
```

Table 2 The two deployment scenarios of the developed OHMM

	Scenario 1	Scenario 2
Name	Inference	Prediction
Description	Interpreting the hidden geological risk based on observation	Extending the interpreted geological risk from now to the near future
Input	$y_1, y_2, \ldots, y_t$	$y_1, y_2, \ldots, y_t$
Output	$\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_t$	$\hat{x}_{t+1}, \hat{x}_{t+2}, \ldots, \hat{x}_{t+\Delta T}$
Performance metrics	Backward accuracy $Acc_{\text{backward}}(t)$	Forward accuracy $Acc_{\text{forward}}(t, \Delta T)$
Practical significance	Not that significant; To evaluate the practice of excavated sections, may contribute to engineering experience accumulation	Significant; To guide the excavation

geological risk, such as geological maps, vulnerability maps of urban environments (Eremina et al., 2018), and soil stratigraphy (Sundell et al., 2019), lack the level of detail necessary to guide TBM-excavated tunneling practices. For example, while geological profiles obtained from the soil stratification model offer a precise vertical resolution (0.1 m) based on borehole analysis, the horizontal resolution derived from interpolation between boreholes is much more coarse (i.e., 10 m × 10 m) (Sundell et al., 2019). The proposed method aims to develop a geological risk model with a fine resolution (per ring level, approximately 1 m), which would inform the excavation process and serve as an early warning system for identifying high-risk regions during construction.

The tunnel studied in this research is located in Singapore and comprises 915 rings. In geological investigations, earth materials at the site are typically described and classified based on their engineering uses (Reddi et al., 2012). The widely recognized unified soil classification system (USCS) classifies soils according to their physical

properties (e.g., grain size) and behavioral characteristics. The system consists of three major groups: (i) highly organic soils (e.g., peat), (ii) fine-grained soils (e.g., silts and clays), and (iii) coarse-grained soils (e.g., gravels and sands). Additionally, Einstein (2004) classified geological conditions into three categories: "little," "medium," and "intense." In this study, soil types corresponding to sparse boreholes and per-ring records were obtained from the daily reports. The soil samples were classified as fluvial sand (FS), old alluvium (OA), fluvial clay (FC), marine clay (MC), and organic clay (OC), based on grain size and susceptibility to deformation under pressure. The observations are defined as shown in Table 3. Geo-materials with varying properties, including strength, density, and static and dynamic responses, demonstrate diverse stress paths and fracture behaviors that can pose challenges to tunnel construction (Mishra et al., 2017; Zhang et al., 2021). However, the relationship between soils and their geological risk is not well understood and is difficult to determine due to inter-class and intra-class variations. To facilitate interpretation, the geological risk is categorized into three levels: (i) Category 1 (low risk) consists of relatively hard soils, such as sands; (ii) Category 2 (medium risk) consists of fine-grained soils; and (iii) Category 3 (high risk) consists of soft soils. To account for uncertainties in interpreting geological risk from soil observations, such as missing or multiple records, an additional category called "uncertain" is included. The hidden states representing geological risk are defined as shown in Table 4.

The horizontal resolution of borehole samples (i.e., tens of rings) is coarser than that of tunnel rings (i.e., per ring), as illustrated in the spatial distribution of sparse drilling holes and tunnel rings. The training data in Fig. 5 display both sparse hidden states representing geological risk obtained from borehole analysis and dense observations obtained from on-site excavation. In the absence of excavation observations, geological risk state information is only available at the boreholes. As depicted in Fig. 5(a), the state information is discrete in the unconstructed area, which is one of the initial challenges addressed in this paper using the sequence extension mechanism. Dense hidden states representing geological risk per ring are also collected for evaluation purposes; however, during the construction process, per-ring state information is not available. Consequently, the problem of predicting geological risk, which corresponds to the hidden states in the HMM framework, cannot be treated simply as a single time series prediction problem since the states of geological risk are typically not accessible.

In consideration of the online manner, the update of soil conditions in the environment is taken into account along with the excavation. During the excavation process, the operation of the TBM may disturb the soil. However, this disturbance primarily affects the area passed by the TBM. When forecasting the soil condition ahead of the

Table 3 Observation events of soil conditions and their denotation symbols

Site record description (Observation events)	Symbols
A mix of FS and MC	o_1
FC	o_2
MC, a mix of OA and OC	o_3
OA	o_4
Missing or multiple records	o_5

Table 4 States of geological risk and their denotation symbols

Geological risk (States)	Symbols
Low	s_1
Medium	s_2
High	s_3
Uncertain	s_4

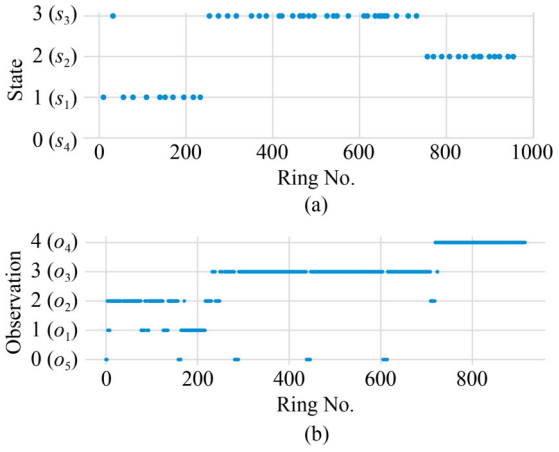

Fig. 5 Symbols and the encoded numeric values for (a) sparse states obtained from borehole analysis; (b) dense observations per tunnel ring obtained from excavation.

TBM, the disturbance is not significant and becomes increasingly negligible with further forecasting. Furthermore, as the goal of this work is to forecast geological risks based on soil condition classification, changes in geological conditions will not have a significant impact on the results compared to a regression problem that involves precise parameters. Finally, based on the site investigation information, the data collected from the project does not include changes in soil conditions. Therefore, this work does not consider the change in the environment over time.

4.2 Model development

Based on the defined observations (Table 3) and hidden states (Table 4), an OHMM is constructed to relate the on-site observations to the hidden states of geological risk. The state diagram and its emission patterns are

shown in Fig. 6. Mathematically, the developed OHMM is specified by:

(i) State space of 4 states $S = \{s_1, s_2, s_3, s_4\}$;

(ii) A transition probability matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix};$$

(iii) Observation space of 5 observable events $O = \{o_1, o_2, o_3, o_4, o_5\}$;

(iv) Emission probability matrix:

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \end{bmatrix};$$

(v) An initial probability distribution $\pi = \pi_1\pi_2\pi_3\pi_4$.

The learning of OHMM concerns updating the parameters $\theta = (A, B)$ till convergence. Three model parameters (i.e., A, B, π) and two sufficient statistics (i.e., $q_l(0)$ and $\hat{\phi}_{ijk}^{h}(0)$) are to be initialized for the training of the OHMM. Since the true values of the parameters are unknown, the initial probability distribution is set using a uniform distribution, the transition probability matrix is set with equal probabilities for any state, and the emission probability matrix is set using a random matrix with a specific seed for consistency. The initial emission probabilities are represented as:

$$B_0 = \begin{bmatrix} 0.136 & 0.093 & 0.053 & 0.299 & 0.420 \\ 0.066 & 0.272 & 0.233 & 0.019 & 0.410 \\ 0.175 & 0.335 & 0.347 & 0.070 & 0.072 \\ 0.264 & 0.106 & 0.048 & 0.259 & 0.322 \end{bmatrix}.$$

Besides, the sufficient statistics $q(0)$ is initialized to be the same as π. The sufficient statistics $\phi(0)$ is initialized as zeros with the shape (N, N, N, len_obs), where N is the number of states in the state space and len_obs is the length of the observation sequence. The hyperparameter t_s, mentioned in Algorithm 1, is adjusted to ensure that sufficient statistics gather sufficient information about the system before being used to update the model parameters. By default, a value of 250 is used, which keeps the model static for the first 250 samples and then updates it for every subsequent sample after the 250$^{\text{th}}$ step.

4.3 Analysis of results

To model the geological risk of the TBM face, the constructed OHMM is updated based on each observation representing the soil components per ring of the tunnel. To illustrate the OHMM learning process, the case of full observations is used as an example. The log-likelihood error over time is depicted in Fig. 7, while Fig. 8 presents the online estimates of the OHMM parameters. Additionally, Fig. 9 shows the probabilistic version of the inferred states obtained from the trained OHMM. To demonstrate the superior performance of the OHMM, a comparison is made with two traditional machine learning algorithms (namely neural network (NN) and SVM), an LSTM, and another generative model (the normal HMM).

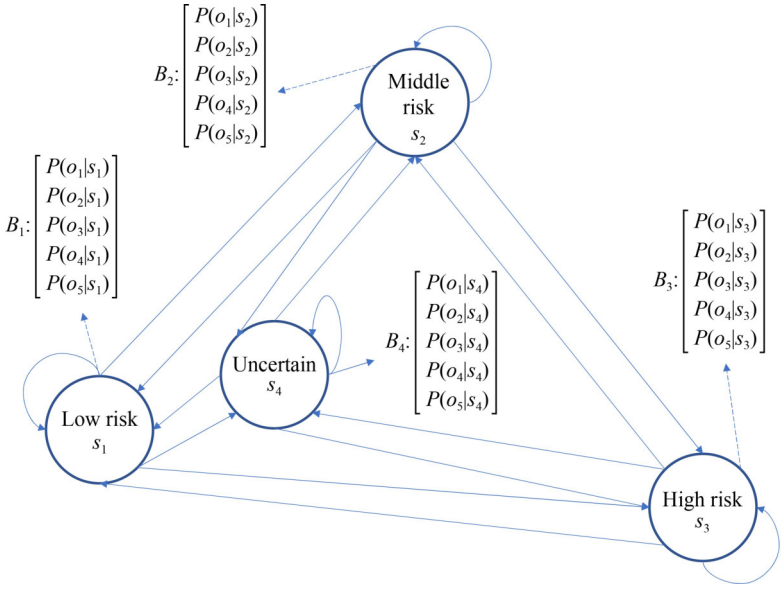

Fig. 6 State diagram of the OHMM with 5 observation events and 4 hidden states.

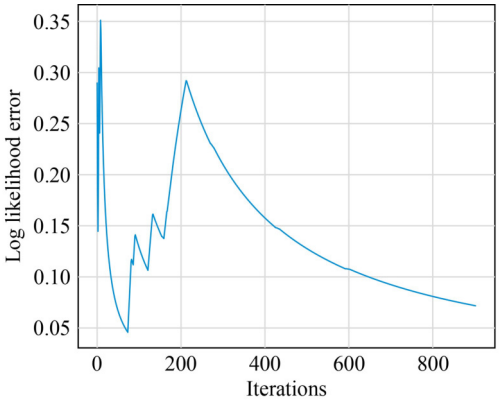

Fig. 7 Time course of the log-likelihood error along with the training iteration.

Furthermore, two instances with 300 observed rings (#300) and 600 observed rings (#600) are tested to evaluate the models' performance with limited observation data. All models are designed to predict 30 steps ahead as a baseline. The OHMM's performance is thoroughly assessed in terms of its online nature and the mechanism for extending the observation sequence, in comparison to the selected competitors.

Regarding the training of competitive models, the NN is structured with an input layer featuring 3 neurons representing three historical states. Additionally, it consists of 3 hidden layers with 16 neurons each, utilizing ReLU activation, and an output layer with 4 neurons corresponding to 4 states, employing softmax activation. The NNs are trained by employing a categorical cross-entropy loss function, with a batch size of 16 and an epoch count of 100. The two parameters of SVM, the regularization parameter and the kernel coefficient, are determined and optimized by grid search in the range $[10^{-10}, 10^{5}]$ with 16 values spaced evenly on a log scale. The optimized SVM for instance #300 has a regularization parameter $C = 100$ and a kernel coefficient $\gamma = 0.01$ for all 30 random trials. The optimized SVM for instance #600 has a set of parameters of $C = 1$ and $\gamma = 1$. For the training of the LSTM model, the network consists of an embedding layer incorporating 10 historical steps, a spatial dropout layer with a dropout rate of 0.2, an LSTM layer including 16 neurons with respective dropout rates of 0.6 and 0.5 for regular dropout and recurrent dropout, and finally a dense layer equipped with the softmax activation function, serving the purpose of classification. The LSTM model is trained by utilizing a categorical cross-entropy loss function, with a batch size of 32 and an epoch count of 100, employing the Adam optimizer. The average accuracy scores achieved throughout the 30 random trials, accompanied by significance tests, are presented in Table 5. Additionally, Fig. 9 provides a visual representation of the confusion matrices showcasing the predictions made by these algorithms, corresponding to the time instances 300 and 600, respectively. The categorical sensitivity and specificity of these prediction results can be observed in Table 6.

(1) The average likelihood of the learned OHMM approaches the average likelihood of the true model as the log-likelihood error tends to zero over time. In the context of OHMM, the mapping patterns between geological risk and soil components are represented by a space-varying emission probability matrix. Similarly, the transition of underlying states from one location to another is represented by a space-varying transition probability matrix. Converged OHMM can effectively represent the transition and emission behaviors of the system, which may vary from region to region in the case of geological risk prediction. The initialization process plays a crucial role in the convergence speed, as it depends on the distance between the initial model and the true model. Under the chosen initialization schema, the variables are uniformly initialized, with the exception of the emission probability matrix, which is randomly drawn. Figure 7 illustrates the time course of the log-likelihood error for a single trial, whose initial values are provided in Section 4.2. The irregular trend observed during the first 250 iterations can be attributed to the control parameter t_s (Algorithm 1), which is set to 250. During this phase, the model remains static, and only sufficient statistics are updated. This can be observed from the iteration of parameters shown in Fig. 8. Early freezing of model parameters aids in stabilizing the training process, allowing sufficient statistics to gather enough information about the system before being used to update the model. From iteration 250 onwards, the log-likelihood error consistently decreases at a slower rate, indicating convergence. (2) The trained OHMM is capable of accurately inferring the hidden states in certain cases, providing uncertain states when confronted with short sub-series that include new observation events. The raw probabilities associated with the defined four states vary with the ring number, as depicted in Fig. 9(a), thus indicating a changing geological risk along the tunnel. By analyzing the OHMM, it is possible to retrieve the maximum probabilities for each tunnel ring, enabling a judgment of the state. Figure 9(c) illustrates a comparison between the OHMM's state determination and the ground truth of the states. Compared to the ground truth of the states (Fig. 9 (c)), the state s_3 (i.e., high risk) can always be identified with high probability, which is of high significance as high risk is always major concern in excavation processes. The states s_1 (i.e., low risk) and s_2 (i.e., medium risk) do not refer to the exact state but are semantically correct, especially in the last 100 rings when observing new events that do not last long and the probabilities of all states are relatively equal. Tracing back to the observations (Fig. 5 (b)), the states s_1 and s_2 corresponds to observations o_1 and o_4, respectively. By the definition (Table 3), categories o_1 and o_3 have some common soil components as o_3, where engineers distinguish them with certainty in the

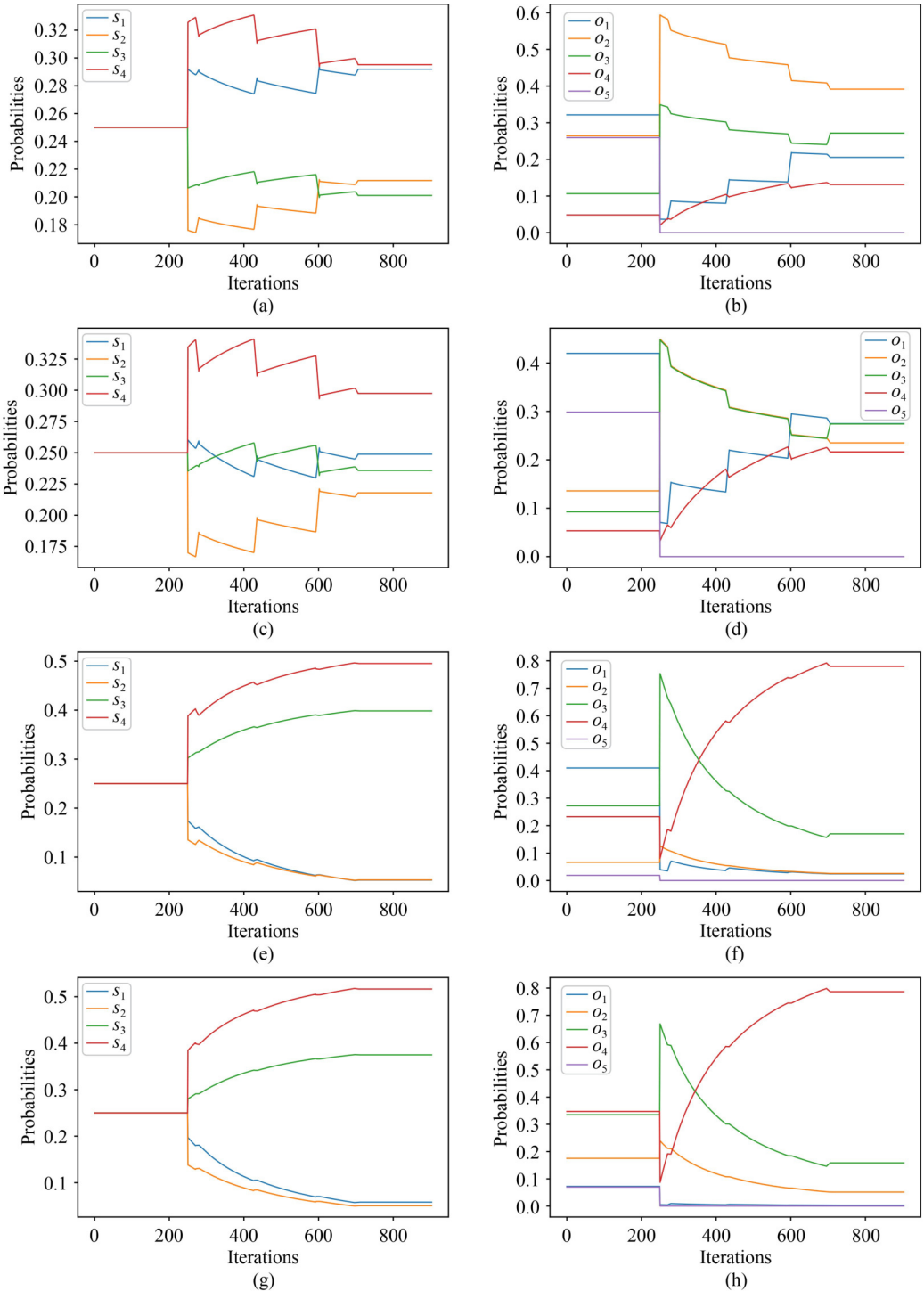

Fig. 8 Online estimates of the OHMM parameters: (a) transition probabilities from state s_1 to some state; (b) emission probabilities of observing some observation event at state s_1; (c) transition probabilities from state s_2 to some state; (d) emission probabilities of observing some observation event at state s_2; (e) transition probabilities from state s_3 to some state; (f) emission probabilities of observing some observation event at state s_3; (g) transition probabilities from state s_4 to some state; and (h) emission probabilities of observing some observation event at state s_4.

frame of categories while model gives a judgment of "uncertain" with higher probabilities. For the rings with a missing record or multiple records, the model gives a judgment with a specific risk state rather than an uncertain state.

(3) The OHMM algorithm demonstrates superior performance compared to traditional machine learning techniques due to its online consideration of manner and the mining of hidden states. Traditional methods like NN and SVM lack stable performance in classification tasks

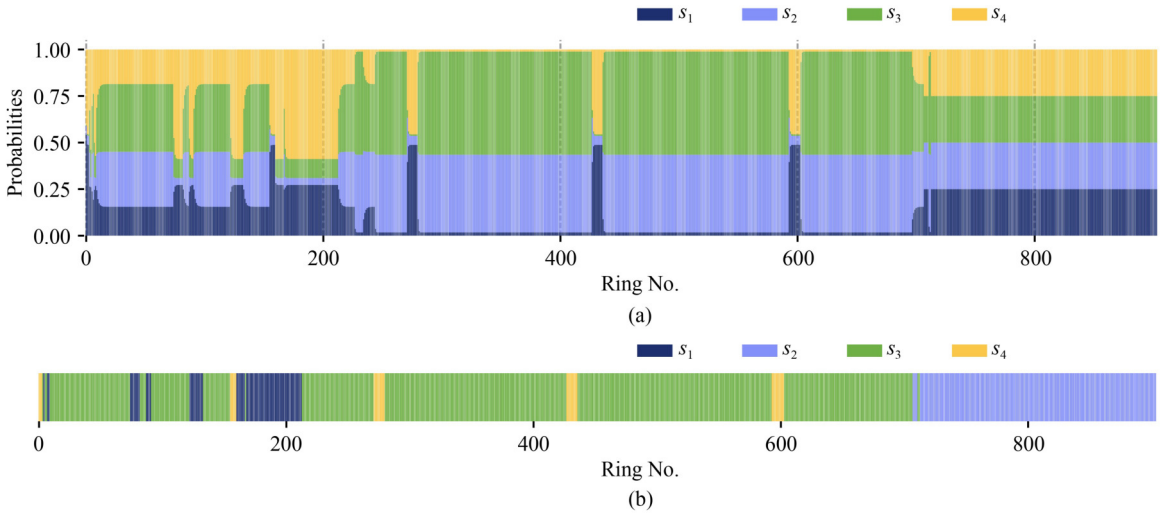

Fig. 9 Geological risk profile along the tunnel path: (a) Probability distributions over the states, and (b) Ground truth of the states.

and are highly affected by the data set. Notably, there is a significant difference in performance between forward prediction accuracy at #300 and #600. In contrast, LSTM, HMM, and OHMM consistently exhibit stable forward prediction accuracy. Crucially, OHMM surpasses LSTM and HMM in forward prediction accuracy (0.968 for #300 and 0.902 for #600) because OHMM incorporates online consideration. Moreover, OHMM and HMM are generative models capable of inferring hidden states, while NN, SVM, and LSTM require historical hidden states to predict future hidden states. This characteristic makes OHMM more practical for real projects as it can be applied after initialization steps. Thanks to the observation sequence extension mechanism, the model reliably predicts geological risks based on site investigation data.

(4) The exceptional performance of OHMM stems from its consistently high sensitivity over risk classes for the upcoming state sequence. The confusion matrices in Fig. 10 illustrate the state classification results from the 1st random trial of the algorithms corresponding to the two instances. When predicting forward from this single trial, all five algorithms (NN, SVM, LSTM, HMM, and OHMM) accurately identify the geological risk as "high risk" for time instance #300 (Figs. 10(a)–(d)). However, at time instance #600 (Figs. 10(e)–(h)), traditional methods like NN and SVM fail to identify "high risk," while LSTM, HMM, and OHMM can. Nevertheless, OHMM outperforms the others and exhibits the least misclassification. Table 6 displays the categorical sensitivity and specificity of these prediction results. In a multi-class classification problem, high sensitivity toward a specific risk level is preferred. OHMM maintains high sensitivity for the "high risk" in both instances and outperforms the other algorithms. This implies higher reliability when implemented in real projects.

(5) The OHMM demonstrates an accurate estimation of model parameters even when limited observation data are

Table 5 Comparison with other methods

Time instance	Method	Input data		Performance
		Observation sequence (Easy to be recorded) required	State sequence (Hard to judge) required	Forward accuracy*
#300	NN	No	Yes	1.000
	SVM	No	Yes	1.000
	LSTM	No	Yes	0.833
	HMM	Yes	No	0.700
	OHMM	Yes	No	0.968
#600	NN	No	Yes	0.100
	SVM	No	Yes	0.100
	LSTM	No	Yes	0.892
	HMM	Yes	No	0.763
	OHMM	Yes	No	0.902

Note: * Obtained from the average of 30 trials of training.

Table 6 Categorical sensitivity and specificity for the prediction scenarios

Time instance	Method	Uncertain		Low risk		Medium risk		High risk	
		Sen	Spe	Sen	Spe	Sen	Spe	Sen	Spe
#300	NN	Nil	1	Nil	1	Nil	1	1	Nil
	SVM	Nil	1	Nil	1	Nil	1	1	Nil
	LSTM	Nil	1	Nil	1	Nil	1	1	Nil
	HMM	Nil	1	Nil	1	Nil	1	1	Nil
	OHMM	Nil	1	Nil	1	Nil	1	1	Nil
#600	NN	1	0	Nil	1	Nil	1	0	1
	SVM	1	0	Nil	1	Nil	1	0	1
	LSTM	0	1	Nil	0.87	Nil	1	0.87	Nil
	HMM	0	1	Nil	0.67	Nil	1	0.74	1
	OHMM	0	1	Nil	1	Nil	1	1	0

Note: Sen represents Sensitivity, Spe represents Specificity.

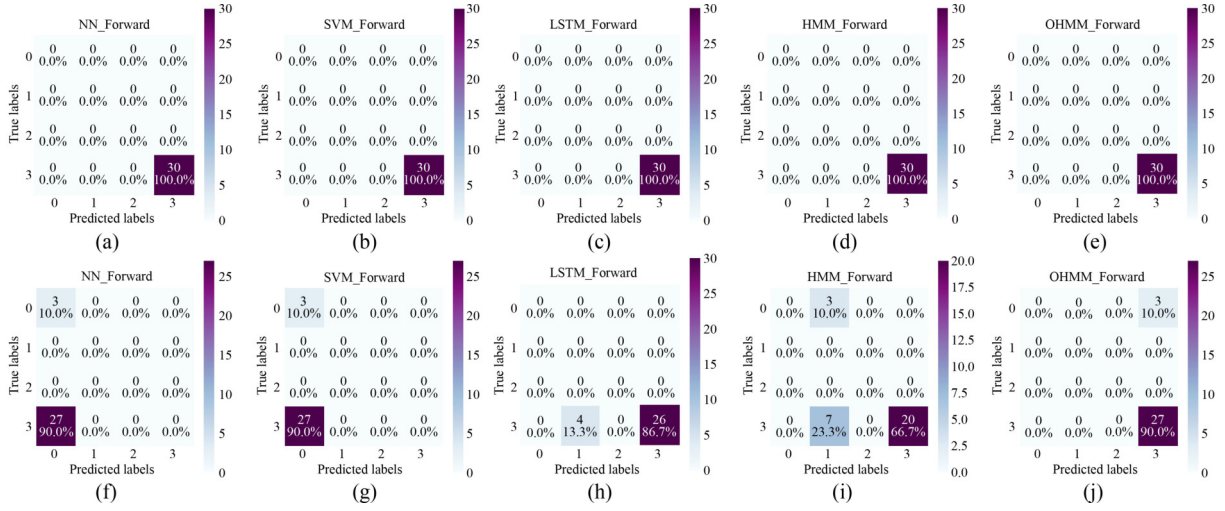

Fig. 10 Confusion matrixes for the forward predictions at time instance #300 from (a) NN; (b) SVM; (c) LSTM; (d) HMM; (e) OHMM, and time instance #600 from (f) NN; (g) SVM; (h) LSTM; (i) HMM; (j) OHMM. Note: The numeric values 0, 1, 2, and 3 represent the states "uncertain," "low risk," "medium risk," and "high risk," respectively.

available, using the observation extension mechanism. Two instances mimic engineering scenarios at different stages of excavation and construction. In Fig. 11, changes in the model parameters and the final parameter are shown. The OHMM's learned parameters from the composed observation sequence converge at a similar speed to those learned from the full observation sequence. Additionally, the final estimated parameters based on the composed observation sequence display similar patterns to those based on the full observation sequence. The similarity of the matrix falls within a 99% confidence interval, as shown in Fig. 12. The OHMM model continuously updates itself with new incoming data, eliminating the need for a large amount of initialization data. This property makes the OHMM applicable in scenarios where only limited data are available, such as the early and middle stages of the construction process. As a result, the developed OHMM greatly extends the compatibility for on-site geological condition prediction throughout the construction process.

5 Discussion

The previous section of this report discussed the overall performance of the OHMM. In this section, we will focus on studying the OHMM's ability to forecast multiple steps ahead and its performance at different stages of excavation. Initially, we select 6 equidistant time instances (ranging from 100 to 600 rings) to predict 30 rings ahead. The results of this prediction are plotted in Fig. 13. It is worth noting that the key difference between these time instances lies in the volume of available data, with data volume increasing as the time index increases. At each time instance, the OHMM is trained using the available observations up to that point, with the length of the available observations corresponding to the number of excavated rings.

Next, we use the model trained at each time instance to forecast different numbers of steps ahead (20, 30, 50, 100, and 150). The forward accuracy, aggregated by the number of rings ahead, is presented in Fig. 14 and summarized in Table 7. Further details are discussed below.

(1) With the assistance of the observation sequence extension mechanism, the OHMM trained on the composed sequence demonstrates high accuracy in predicting scenarios after 200 rings. As the time index increases, the length of the observation sequence also increases. Specifically, OHMMs trained using time instances $t = 300, 400, 500, 600$ can successfully predict the next 30 samples, achieving forward accuracy scores of 1, 0.867, 1, and 0.933, respectively. For the time instances $t = 100, 200$, the low forward accuracy scores are due to the frozen treatment of the OHMM. Additionally, it can be told that the models trained with $t = 100, 200$ have very high instability in forward predicting based on the results in Fig. 14. Therefore, even with the assistance of the observation sequence extension mechanism, a minimum amount of data for initialization is still necessary. This study has shown that this minimum amount is achieved at 200 rings. Therefore, if sufficient information is obtained during the early stages of construction, the OHMM can be effectively utilized to guide excavation in construction engineering.

(2) The OHMM demonstrates high accuracy in predicting geological risk up to a distance of 100 steps ahead, while maintaining good performance in terms of stability and overfitting. The number of steps ahead can be considered as an indicator of the model's ability to predict future results. A larger number of steps indicates

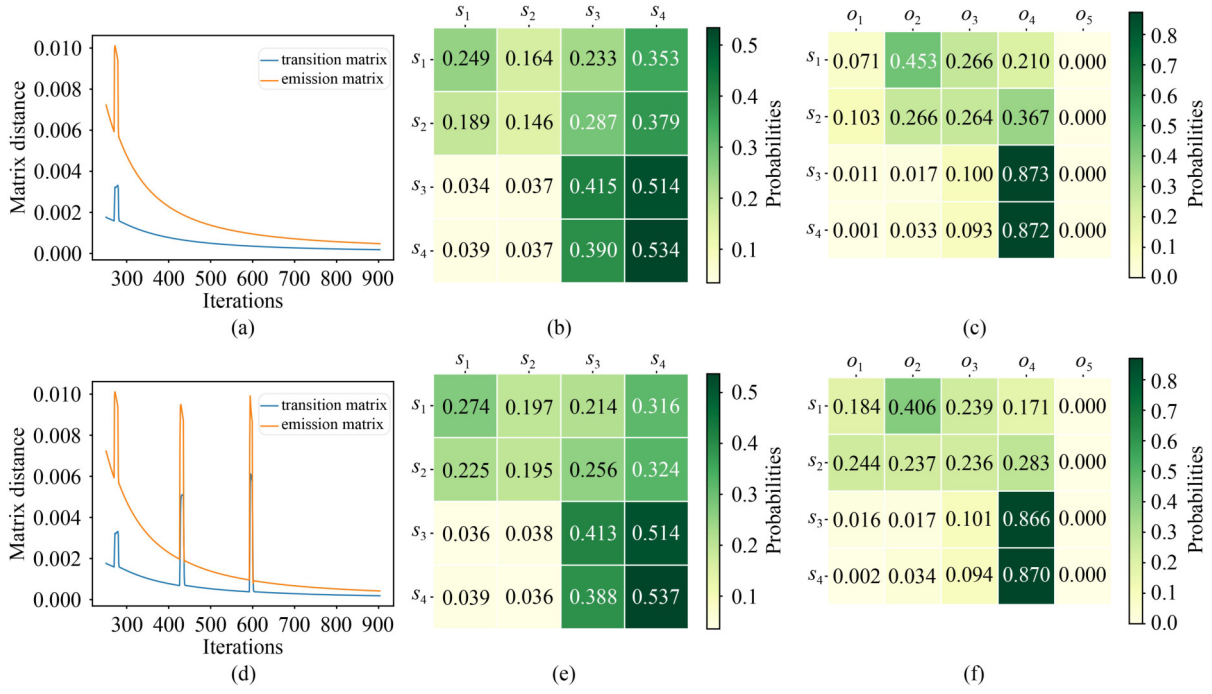

Fig. 11 OHMM parameter learning for several time instances $t = 300, 600$ corresponding to observation sequences with decreasing lengths. (a) matrix distance between two consecutive parameter estimates for $t = 300$; (b) converged transition matrix for $t = 300$; (c) converged emission matrix for $t = 300$; (d) matrix distance between two consecutive parameter estimates for $t = 600$; (e) converged transition matrix for $t = 600$; (f) converged emission matrix for $t = 600$.

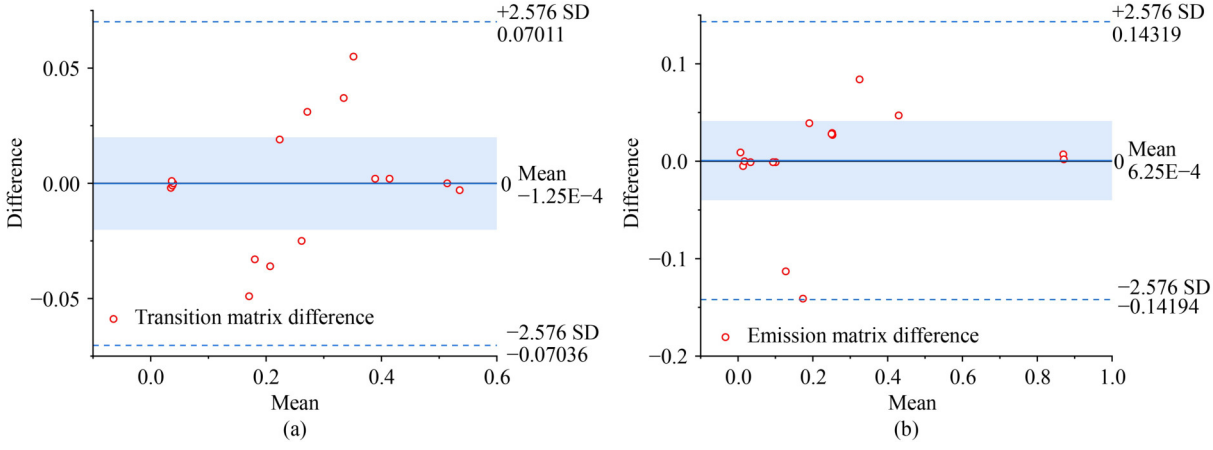

Fig. 12 Bland-Altman plot of (a) transition matrix difference and (b) emission matrix difference with $t = 300$ and $t = 600$ historical data.

that the model can forecast over a greater distance, whereas a smaller number of steps allows for more precise patterns in each sample. The stability and overfitting of our OHMM model when forecasting different numbers of steps ahead using varying historical time steps are presented in Table 7. From the results, it can be observed that once over 200 historical data points are collected, the predictions of future patterns can achieve high accuracy even up to 150 steps ahead. The performance remains stable regardless of the changes in historical steps and the number of steps ahead. For a larger number of steps ahead, the forward accuracy preserves predictions for a larger range of samples and resists changes based on a few incoming samples. Selecting a number of steps ahead of 30 strikes a balance between the extent of foresight and the preservation of individual features in the forecasted samples.

6 Conclusions and future work

This study proposes a framework for applying an online version of HMM, known as OHMM, in the prediction of geological risk in TBM-excavated tunnels. The proposed OHMM is supplemented with an observation extension mechanism, which facilitates effective learning from

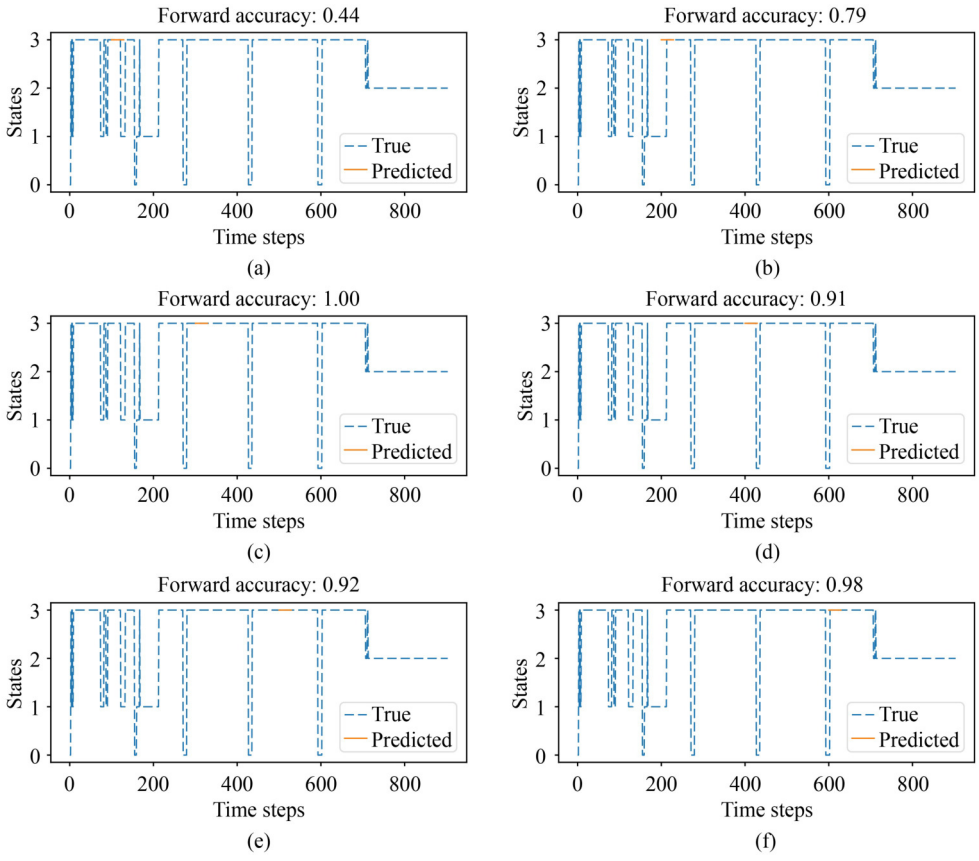

Fig. 13 Predicted hidden states for the next 30 samples from the OHMM at several time instances: (a) $t = 100$; (b) $t = 200$; (c) $t = 300$; (d) $t = 400$; (e) $t = 500$; (f) $t = 600$.

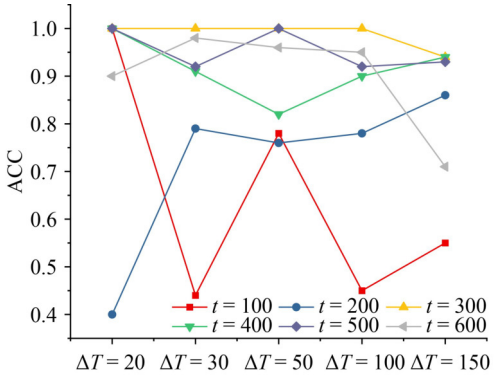

Fig. 14 Effects of the number of rings ahead (ΔT) on forward accuracy from the OHMM trained on full observations.

Table 7 The accuracy of the OHMM predicting different numbers of rings ahead with different sizes of input data

Time instances	Number of rings ahead				
	$\Delta T = 20$	$\Delta T = 30$	$\Delta T = 50$	$\Delta T = 100$	$\Delta T = 150$
$t = 100$	1.00	0.44	0.78	0.45	0.55
$t = 200$	0.40	0.79	0.76	0.78	0.86
$t = 300$	1.00	1.00	1.00	1.00	0.94
$t = 400$	1.00	0.91	0.82	0.90	0.94
$t = 500$	1.00	0.92	1.00	0.92	0.93
$t = 600$	0.90	0.98	0.96	0.95	0.71

short observation sequences and significantly enhances the applicability of the OHMM across different construction stages. The parameters of the OHMM are estimated online based on relatively easily obtained observations. The trained OHMM can then be used to predict future hidden states based on the current state and their transition patterns. The prediction results can provide guidance for future excavation and construction in tunnel engineering. The performance of the model is evaluated based on forward prediction accuracy.

The proposed method is applied to a case study that models the geological risk for a tunnel in Singapore. Based on soil observations, an OHMM with 4 states and 5 observation events is constructed. Its parameters can be effectively updated in an online manner, with continuously decreasing log-likelihood errors. The findings from the study can be summarized as follows: (1) Assisted with the observation extension mechanism, the OHMM can be applied from ring no. 300 until the end of the project (i.e., ring no. 800 onwards), with an accuracy of 0.968. (2) The OHMM is proven to significantly outperform traditional machine learning models, thanks to its online learning

algorithm. (3) The OHMM requires a minimum amount of data (300 rings, as demonstrated in this study) to initialize and achieve high accuracy and stability. (4) The OHMM is capable of providing acceptable predictions up to 100 rings ahead. However, considering the stability of the model performance, a foresight distance of 30 rings is the recommended choice for balancing prediction accuracy in the studied tunnel.

Observation extension is a critical component for deploying the OHMM in the varying stages of the construction process, specifically for geological condition modeling and prediction. This study only considers the situation where there is only one jump between two borehole samples in the filled observations. However, in reality, the geological conditions ahead may change multiple times. To approach the real pattern of changes in geological conditions, particularly in regions with mixed and heterogeneous geological formations, an effective algorithm allowing the modeling of multiple jumps is required. Additionally, integrating TBM operational data, along with the geological observation data, can further enhance the overall performance of geological modeling and prediction. During excavation, the TBM may cause disturbances in the soil, leading to changes in soil properties over time. These changes cannot be reflected in the project data. Therefore, conducting proper simulations in the future to consider these factors is worth exploring. Furthermore, the proposed model should undergo further testing using new data sets from real projects or simulations to assess the generality of the proposed method.

Competing Interests The authors declare that they have no competing interests.

References

Arieli O, Denecker M, Bruynooghe M (2007). Distance semantics for database repair. Annals of Mathematics and Artificial Intelligence, 50(3): 389–415

Bietti A, Bach F, Cont A (2015). An online em algorithm in hidden (semi-)Markov models for audio segmentation and clustering. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1881–1885

Blum A (1998). On-line algorithms in machine learning. Online Algorithms: The State of the Art, 306–325

Cao B T, Saadallah A, Egorov A, Freitag S, Meschke G, Morik K (2021, 2021). Online geological anomaly detection using machine learning in mechanized tunneling. Paper presented at the Challenges and Innovations in Geomechanics, Cham

Chadza T, Kyriakopoulos K G, Lambotharan S (2020). Analysis of hidden Markov model learning algorithms for the detection and prediction of multi-stage network attacks. Future Generation Computer Systems, 108: 636–649

Chis T, Harrison P G (2015). Adapting hidden Markov models for online learning. Electronic Notes in Theoretical Computer Science, 318: 109–127

Cho S H, Kim J, Won J, Kim M K (2017). Effects of jack force and construction steps on the change of lining stresses in a TBM tunnel. KSCE Journal of Civil Engineering, 21(4): 1135–1146

Dakir I, Benamara A, Aassoumi H, Ouallali A, Ait Bahammou Y (2019). Application of induced polarization and resistivity to the determination of the location of metalliferous veins in the Taroucht and Tabesbaste areas. International Journal of Geophysics, 2019: 5849019

Einstein H H (2004). Decision aids for tunneling. Update., 1892(1): 199–207

Eremina O, Kozliakova I, Anisimova N, Kozhevnikova I (2018). Assessment of exogenous geological hazards in Moscow, Russia. 55(1): 133–140

Erharter G H, Marcher T (2020). MSAC: Towards data driven system behavior classification for TBM tunneling. Tunnelling and Underground Space Technology, 103: 103466

Erharter G H, Marcher T, Reinhold C (2020). Artificial neural network based online rockmass behavior classification of TBM data. Information Technology in Geo-Engineering, 178–188

Fu X, Zhang L (2021). Spatio-temporal feature fusion for real-time prediction of TBM operating parameters: A deep learning approach. Automation in Construction, 132: 103937

Gandhi S M, Sarkar B C (2016). Chapter 5 - Geophysical exploration. In: Essentials of Mineral Exploration and Evaluation. 97–123

Golpasand M B, Do N A, Dias D, Nikudel M R (2018). Effect of the lateral earth pressure coefficient on settlements during mechanized tunneling. Geomechanics and Engineering, 16(6): 643–654

Hoi S C H, Sahoo D, Lu J, Zhao P (2021). Online learning: A comprehensive survey. Neurocomputing, 459: 249–289

Huang X, Li J, Liang Y, Wang Z, Guo J, Jiao P (2017). Spatial hidden Markov chain models for estimation of petroleum reservoir categorical variables. Journal of Petroleum Exploration and Production Technology, 7(1): 11–22

Langfield-Smith K, Wirth A (1992). Measuring differences between cognitive maps. Journal of the Operational Research Society, 43(12): 1135–1150

Li K, Qiu C, Zhou X, Chen M, Lin Y, Jia X, Li B (2022). Modeling and tagging of time sequence signals in the milling process based on an improved hidden semi-Markov model. Expert Systems with Applications, 205: 117758

Li S, Liu B, Xu X, Nie L, Liu Z, Song J, Sun H, Chen L, Fan K (2017). An overview of ahead geological prospecting in tunneling. Tunnelling and Underground Space Technology, 63: 69–94

Lin B, Zhou L, Lv G, Zhu A X (2017). 3D geological modelling based on 2D geological map. Annals of GIS, 23(2): 117–129

Liu Z, Li L, Fang X, Qi W, Shen J, Zhou H, Zhang Y (2021). Hard-rock tunnel lithology prediction with TBM construction big data using a global-attention-mechanism-based LSTM network. Automation in Construction, 125: 103647

Lu C, Liu J, Liu Y, Liu Y (2019). Intelligent construction technology of railway engineering in China. Frontiers of Engineering Management, 6(4): 503–516

Mahmoodzadeh A, Mohammadi M, Daraei A, Farid Hama Ali H, Ismail Abdullah A, Kameran Al-Salihi N (2021). Forecasting tunnel geology, construction time and costs using machine learning

methods. Neural Computing & Applications, 33(1): 321–348

Mishra S, Rao K S, Gupta N K, Kumar A (2017). Damage to shallow tunnels under static and dynamic loading. Procedia Engineering, 173: 1322–1329

Mongillo G, Deneve S (2008). Online learning with hidden Markov models. Neural Computation, 20(7): 1706–1716

Peng S, Li Q (2021). Research on 3D geological modeling method of tunnel engineering computer based on geological cross-section. Journal of Physics: Conference Series, 1992(2): 022112

Reddi L N, Jain A K, Yun H B (2012). 6 - Soil materials for earth construction: properties, classification and suitability testing. In Hall M, Lindsay R, Krayenhoff M. eds. Modern Earth Buildings, Woodhead Publishing Series in Energy: 155–171

Rekatsinas T, Chu X, Ilyas I F, Ré C (2017). HoloClean: Holistic data repairs with probabilistic inference. arXiv:1702.00820

Savenok O V, Povarova L V, Kusov G V (2020). Application of superdeep drilling technology for study of the earth crust. IOP Conference Series. Earth and Environmental Science, 459(5): 052066

Sheil B B, Suryasentana S K, Mooney M A, Zhu H (2020). Machine learning to inform tunnelling operations: Recent advances and future trends. Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction, 173(4): 74–95

Sicking J, Pintz M, Akila M, Wirtz T (2020). DenseHMM: Learning hidden Markov models by learning dense representations. arXiv: 2012.09783

Stoner O, Economou T (2020). An advanced hidden Markov model for hourly rainfall time series. Computational Statistics & Data Analysis, 152: 107045

Sundell J, Haaf E, Norberg T, Alén C, Karlsson M, Rosén L (2019). Risk mapping of groundwater-drawdown-induced land subsidence in heterogeneous soils on large areas. Risk Analysis, 39(1): 105–124

Thalund-Hansen R, Troldborg M, Levy L, Christiansen A V, Bording T S, Bjerg P L (2023). Assessing contaminant mass discharge uncertainty with application of hydraulic conductivities derived from geoelectrical cross-borehole induced polarization and other methods. Water Resources Research, 59: e2022WR034360

Vereecken H, Amelung W, Bauke S L, Bogena H, Brüggemann N, Montzka C, Vanderborght J, Bechtold M, Blöschl G, Carminati A, Javaux M, Konings A G, Kusche J, Neuweiler I, Or D, Steele-Dunne S, Verhoef A, Young M, Zhang Y (2022). Soil hydrology in the Earth system. Nature Reviews. Earth & Environment, 3(9): 573–587

Wan F, Guo H, Li J, Gu M, Pan W, Ying Y (2021). A scheduling and planning method for geological disasters. Applied Soft Computing, 111: 107712

Wang B, Sun D, Chen Q, Lin W, Li A W, Cao H (2020a). Stress-state differences between sedimentary cover and basement of the Songliao Basin, NE China: *In-situ* stress measurements at 6–7 km depth of an ICDP Scientific Drilling borehole (SK-II). Tectonophysics, 777: 228337

Wang X, Lai J, Qiu J, Xu W, Wang L, Luo Y (2020b). Geohazards, reflection and challenges in mountain tunnel construction of China: A data collection from 2002 to 2018. Geomatics, Natural Hazards & Risk, 11(1): 766–785

Wang Z, Zhang B (2023). Key technical innovations in the construction of Baihetan Hydropower Station Project. Frontiers of Engineering Management, 10(2): 367–372

Wolfsberg A (1997). Rock Fractures and fluid flow. Contemporary Understanding and Applications, 78(49): 569–573

Xiang Y, Zeng Z, Xiang Y, Abi E, Zheng Y, Yuan H (2021). Tunnel failure mechanism during loading and unloading processes through physical model testing and DEM simulation. Scientific Reports, 11(1): 16753

Xiong Z, Guo J, Xia Y, Lu H, Wang M, Shi S (2018). A 3D multi-scale geology modeling method for tunnel engineering risk assessment. Tunnelling and Underground Space Technology, 73: 71–81

Xu Z, Liu F, Lin P, Shao R, Shi X (2021a). Non-destructive, *in-situ*, fast identification of adverse geology in tunnels based on anomalies analysis of element content. Tunnelling and Underground Space Technology, 118: 104146

Xu Z H, Wang W Y, Lin P, Nie L C, Wu J, Li Z M (2021b). Hard-rock TBM jamming subject to adverse geological conditions: Influencing factor, hazard mode and a case study of Gaoligongshan Tunnel. Tunnelling and Underground Space Technology, 108: 103683

Yazdani N, Garcia E C, Riad M (2018). Field assessment of concrete structures rehabilitated with FRP. Eco-Efficient Repair and Rehabilitation of Concrete Infrastructures, 171–194

Zhang A, Song S, Wang J, Yu P S (2017). Time series data cleaning: from anomaly detection to anomaly repairing. *10*(10%J Proc. VLDB Endow.), 1046–1057

Zhang L, Lin P (2021). Multi-objective optimization for limiting tunnel-induced damages considering uncertainties. Reliability Engineering & System Safety, 216: 107945

Zhang P, Lu D, Du X, Qi J (2021). A division method for shallow tunnels and deep tunnels considering soil stress path dependency. Computers and Geotechnics, 135: 104012

Zhou Y, Chen X, Wu M, Cao W (2021). Modeling and coordinated optimization method featuring coupling relationship among subsystems for improving safety and efficiency of drilling process. Applied Soft Computing, 99: 106899