Optimal monitoring location for tracking evolving risks to infrastructure systems: Theory and application to tunneling excavation risk

Zeyu Wang[1], Abdollah Shafieezadeh[2], Xiong Xiao[1], Xiaowei Wang[3] and Quanwang Li[1*]

[1] School of Civil Engineering, Tsinghua University, Beijing 100084, China

[2] The Ohio State University, Columbus 43202, United States

[3] Department of Bridge Engineering, Tongji University, Shanghai 200092, China

*Corresponding author, Email address: li_quanwang@tsinghua.edu.cn

ABSTRACT

Structural health monitoring (SHM) technologies offer ever-increasing opportunities to continually observe various responses and states of structures, such as settlement-induced building damage. Recent advances in reliability updating have enabled estimating the probability of failing to meet a prescribed objective for systems using various types of information including those acquired from SHM. However, reliability updates are sensitive to monitoring location, especially when the risks are evolving. Therefore, there may exist optimal locations in a system for monitoring that yield maximum value for reliability updating. This paper proposes a computational framework for optimal monitoring location based on an innovative metric called sensitivity of information (*SOI*). This metric quantifies the change in unconditional and conditional reliability indexes, which subsequently facilitates fast exploration of optimal monitoring location by parameterizing an optimization function. A state-of-the-practice case related to assessing evolving risks posed by tunneling-induced settlement to buildings is explored in-depth with respect to the progression of tunneling. Simulation results showcase that the proposed framework can successfully find the monitoring location that is the most impactful to the accuracy of the updated reliability.

Key words: *Infrastructure monitoring; reliability updating; reliability analysis; Machine Learning; surrogate models; Tunneling excavation*

1. Introduction

Infrastructure systems are often subject to various forms of stressors that can threaten their functionality and safety [1]. To capture those potentially unsafe conditions that may cause future catastrophic events, structural reliability analysis for structures is indispensable. As sensors and, more broadly, monitoring technologies advance, valuable information can be acquired without much effort. This brings new opportunities for risk analysis. It also introduces new challenges for the integration of data and re-evaluation of the established risk assessment processes to update risk assessments. Grounded in the Bayesian updating theory, the emergence of reliability updating technique fills this gap by updating the probability of failure. In this context, let F denote the failure event and Z denote the observed information. Reliability updating aims to estimate the conditional probability of failure, $\Pr(F|Z)$, which can be formulized as [2],

$$\Pr(F|Z) = \frac{\Pr(F \cap Z)}{\Pr(Z)} \tag{1}$$

where $\Pr(F|Z)$ is the conditional probability of failure given information Z (or the so-called posterior probability of failure) and $\Pr(F \cap Z)$ is the probability of the joint event $F \cap Z$. The information Z can be generally classified into two groups that are inclusive of equality and inequality types. Computation of reliability updating with equality information is typically non-trivial through the conventional approaches such as surface integral [2] [3] and Bayesian networks [4]–[6]. This computational challenge has been addressed by subtly introducing an auxiliary random variable to transform the equality information into an inequality one [2]. However, the computation of $\Pr(F|Z)$ requires the probability of a joint event, $F \cap Z$, which is typically a very rare event. This probability can be estimated through subset simulation (*SS*) to improve the computational efficiency [7]–[10]. Moreover, by decomposing $\Pr(F \cap Z)$ into two more frequent probabilities $\Pr(Z)$ and $\Pr(F|Z)$ and training a surrogate model for the limit state function, metamodel-based approaches can facilitate fast estimation of $\Pr(F|Z)$ [11], [12].

Reliability updating has been recently applied in engineering for solving various types of practical problems. For example, field data and soil characteristics have been used to accurately estimate the reliability of a shallow foundation in a silty soil with spatially variable properties simulated via random fields [13]. Moreover, metamodeling techniques have been integrated to analyze the prior and posterior failure probabilities of a sheet pile wall in a dyke [14]. This work demonstrated the computational capability of metamodel-based reliability updating in estimating $\Pr(F|Z)$. To ensure the safety of buildings in vicinity of a tunnel line, the settlement monitoring data at different locations were used to update the reliability of tunneling-induced settlement during excavation [15].

This technique can better assist in risk management decisions if the ability of the planned tunneling line to satisfy the safety requirement can be checked in real-time through settlement monitoring. Analogous to this case, the deformation measurements of an excavation in sandy trench with a sheet pile retaining wall were also used to update the reliability of a construction site at its full excavation status [8]. This work can be viewed as a practical case for engineers in construction sites in avoiding catastrophic trench collapse. Additionally, to improve alarming system of a flood defense infrastructure, reliability updating together with head monitoring information were implemented in [16] to mitigate the risk of piping-induced levee failure in the presence of highly uncertain geohydrological properties. This work represents the potential capability of reliability updating in strengthening risk-informed warning systems against natural hazards. To achieve the largest benefits, Klerk et al. [17] also used reliability updating and *VOI* (value of information) to seek for optimal structural health monitoring of flood defense systems from a set of representative case studies. Huang et al. [18] are among the very first authors who proposed an adaptive reliability updating of bridges using structural properties derived from nondestructive testing . Using Bayesian inference, Jin et al. [19] proposed an adaptive approach to seek for soil parameters that correspond to the measured deformation on site, which facilitates the prediction of deflections. Subsequently, Jin et al. [20] applied a Bayesian method to explore most probable parameters and demonstrated a process for obtaining those parameters. Moreover, reliability updating has also been implemented in performance assessment of deteriorating reinforced concrete structures [21], slope stability [22], [23], structural inspection and repair of infrastructures [24], system identification [25], life-cycle analysis [6] and other applications [26]–[32].

The reviewed literature showcases the high capability of reliability updating in successfully tracking the risk to infrastructures by incorporating the monitoring information within the existing computational scheme of reliability evaluation. For improving risk assessment, it is also necessary to properly select the monitoring location. Jiang et al. [33] proposed exploring the location of boreholes for site investigation for a slope by maximizing *VOI*. Hu et al. [34] proposed an efficient method to reduce the computational cost for site investigation of slope stability assessment through *VOI* analysis. These efforts are grounded in theory of *VOI*, which is tailored to minimize the economic cost considering possible structural failure, maintenance and rehabilitation. However, the estimate of *VOI* index can be trapped in a dilemma when the cost of consequences (e.g., structural failure) is unknown or probability distribution of information is unavailable. Moreover, VOI is not appropriate anymore for exploring the optimal monitoring location when safety consideration is more important than economy. Therefore, it can be inferred that there must exist optimal monitoring locations, where the updated reliability can be utmost sensitive to the obtained information. To this end, this paper develops a method that efficiently determines the optimal monitoring location by introducing a novel metric called sensitivity of information (*SOI*) that measures the amplitude of the sensitivity at any location. Without knowing the cost of consequences, *SOI* is defined as the change in updated and prior reliability index, which facilitates the quantitative measurement of sensitivity of updated reliability index to the new information at a specific location. To improve the computational efficiency of estimating $\Pr(F|Z)$, *SS* along with foregoing presented decomposition of $\Pr(F \cap Z)$ are integrated within the proposed computational framework. Moreover, the proposed *SOI* index subsequently parameterizes an objective function that is designed to find the optimal monitoring location by searching for its maxima based on a surrogate-assisted optimization. To examine the computational efficiency, a state-of-the-practical case of tunneling-induced settlement to building damage is investigated.

The rest of this article is mainly organized in 5 sections. Section 2 briefly introduces the concept of reliability updating. Section 3 presents the proposed *SOI* index together with the framework for determining the optimal monitoring location. Subsequently, section 4 presents the procedures of analyzing *SOI* and exploring the optimal settlement monitoring location for a practical case that investigates the risk posed by tunneling-induced settlements. Conclusive remarks are drawn in section 5.

2 Reliability updating with equality information

Generally, the main difference between reliability analysis and updating lies in whether the observational information is available or not. Reliability analysis focuses on the computation of unconditional probability of failure $\Pr(F)$ while reliability updating estimates the conditional probability of failure $\Pr(F|Z)$. Let $g(\boldsymbol{X})$ denote the performance function, the response of which determines the condition of the system: $g(\boldsymbol{X}) \leq 0$ indicates failure and $g(\boldsymbol{X}) > 0$ means safe state; the boundary region where $g(\boldsymbol{X}) = 0$ is called the limit state. Thus, the unconditional probability of failure can be defined as:

$$\Pr(F) = \Pr(g(\boldsymbol{X}) \leq 0) \tag{2}$$

Methods for computing $\Pr(F)$ include but are not limited to: crude Monte-Carlo simulation (*MCS*) [35], [36], first- or second-order reliability analysis method (*FORM* & *SORM*) [37], [38], importance sampling (*IS*) [39], [40], *SS* [41]–[43] and surrogate-based methods [44]–[47]. As Eq. (1) shows, the estimate of $\Pr(F|Z)$ needs to compute $\Pr(Z)$ and $\Pr(F \cap Z)$. According to [2], the probability of information $\Pr(Z)$ can be computed as follows,

$$\Pr(Z) = \int_{\uptheta \in \Omega_\uptheta} \Pr(Z|\Theta(\boldsymbol{X}) = \uptheta)\, f(\uptheta) d\uptheta \tag{3}$$

where $\boldsymbol{X}$ denotes the vector of random variables, $\Theta(\boldsymbol{X})$ denotes a function parameterized by $\boldsymbol{X}$ with the realization notation $\uptheta$, that can be the uncertainty of the system characteristic, $\Theta_s(\boldsymbol{X})$, or the external loadings, $\Theta_e(\boldsymbol{X})$. Moreover, $f(\cdot)$ represents the probability density function (PDF) and $\Omega_\uptheta$ is the probabilistic space of $\Theta(\boldsymbol{X})$. In this context, the probability of the joint event $\Pr(F \cap Z)$ can be derived as,

$$\Pr(F \cap Z) = \int_{\uptheta \in \Omega_\uptheta} \Pr(F|\Theta(\boldsymbol{X}) = \uptheta) \Pr(Z|\Theta(\boldsymbol{X}) = \uptheta)\, f(\uptheta) d\uptheta \tag{4}$$

For any likelihood functions, $L(\boldsymbol{x})$, the following identity holds true [2]:

$$L(\boldsymbol{x}) = \frac{1}{c} \Pr\{U - \Phi^{-1}[cL(\boldsymbol{x})] \leq 0\} \tag{5}$$

where c is a constant satisfying $0 \leq cL(\boldsymbol{x}) \leq 1$, Φ^{-1} denotes the inverse standard normal cumulative distribution function, and U represents a standard normal variable. By reformulation the equality information into inequality one, $\Pr(Z)$ can be estimated by introducing the auxiliary random variable, U, and define an augmented Limit State Function (LSF),

$$\Pr(Z) = \alpha \Pr(h(U, \boldsymbol{X}) \leq 0) \tag{6}$$

where $\alpha = \frac{\Pr(Z|\boldsymbol{X} = \mathbf{x})}{L(\mathbf{x})}$ is an introduced proportionality constant [2], $h(U, \boldsymbol{X})$ is the augmented limit state function with an auxiliary standard normal random variable, U [2],

$$h(U, \boldsymbol{X}) = U - \Phi^{-1}[cL(\boldsymbol{X})] \tag{7}$$

Similarly, $\Pr(F \cap Z)$ can be computed by defining a limit state function that takes the maximum value of $g(\boldsymbol{X})$ and $h(U, \boldsymbol{X})$,

$$\Pr(F \cap Z) = \alpha \Pr(\max[g(\boldsymbol{X}), h(P, \boldsymbol{X})] \leq 0) \quad (8)$$

Derivation of Eq. (8) is not elaborated in this paper for the sake of brevity. Detailed derivation can be found in [2]. Note that U is not necessarily a standard normal random variable, it can be as simple as a standard uniform distributed random variable. Therefore, one can rewrite Eq. (7) as,

$$h(U, \boldsymbol{X}) = P - cL(\boldsymbol{X}) \quad (9)$$

However, the adoption of standard normal random variable can improve the smoothness of the responses of the function. To increase the readability, the computational scheme based on Eq. (9) is used throughout the paper. Combining Eq. (6) and (8), the conditional probability of failure can be obtained by canceling out the constant α,

$$\Pr(F|Z) = \frac{\Pr(\mathrm{J}(U, \boldsymbol{X}) \leq 0)}{\Pr(h(U, \boldsymbol{X}) \leq 0)} \quad (10)$$

where $\mathrm{J}(U, \boldsymbol{X}) = \max[g(\boldsymbol{X}), h(U, \boldsymbol{X})]$. Eq. (10) enables fast reliability updating by solving two structural reliability problems. Typically, the numerator in Eq. (10) is very small, which requires powerful structural reliability methods such as subset simulation [7], [8]. In the following context, an efficient and robust approach for the estimation of $\Pr(F|Z)$ is presented, which facilities the localization of optimal monitoring location.

3 Optimal monitoring location analysis with *SOI*

Data measured at different locations of structures and infrastructure systems may have distinct impacts on the updated reliability. To precisely quantify this difference, a concept of sensitivity of information for the updated reliability is proposed in this paper. Moreover, the proposed concept can be further leveraged to identify the optimal monitoring location that makes the most significant contribution to the change of updated reliability. In this section, the concepts of sensitivity of information (*SOI*) are elaborated. By maximizing the objective function involving *SOI*, the optimal monitoring location can be derived with the goal of risk tracking for structures and infrastructure systems.

3.1 Sensitivity of information analysis for reliability updating

In practical engineering, acquiring information is typically costly; therefore, engineers should prudently select a worthwhile location for structural monitoring and diagnosis. However, information collected in some locations has very neglectable impact on the change of updated reliability. On the other hand, the updated reliability is very sensitive to the information stemming from very valuable locations. Therefore, the level of sensitivity of updated reliability to the change of information should be mathematically quantified. In this paper, the foregoing concept is denoted as sensitivity of information.

One should note that *SOI* here is substantially different in concept from Value of Information. *VOI* is tailored to evaluate the monetary value of acquired information with the consideration of possible structural failure, maintenance and rehabilitation. In other words, the objective of *VOI* is to establish a value system for acquired information primarily from an economic cost perspective. In contrary, *SOI* is aimed at evaluating the sensitivity of risk updates to monitoring location. The objective here is to compare the power of different monitoring locations (topology) for risk tracking and the focus is on system safety. In practical engineering, the probability distributions of many variables are technically imprecise or unavailable, and engineers often only know the approximate range of possible outcomes of random variable. In this context, *SOI* is a practical risk-informed metric that supports decisions for strategic placement of monitoring systems.

Let $\Pr(F|Z = z, L = l)$ represent the conditional probability of failure given the specific equality information z and the monitoring location l, which can be calculated based on Eq. (10). The difference of the conditional reliability index, β_{post}, compared to the unconditional reliability index, β_{prior}, can be calculated as,

$$\begin{aligned} d_{up}(Z = z, L = l) &= \left|\beta_{post} - \beta_{prior}\right| = \left|-\Phi^{-1}[\Pr(F|Z = z, L = l)] - (-\Phi^{-1}[\Pr(F)])\right| \\ &= \left|\Phi^{-1}[\Pr(F)] - \Phi^{-1}[\Pr(F|Z = z, L = l)]\right| \end{aligned} \tag{11}$$

where d_{up} denotes the change in reliability. However, the information z is typically unknown before it is measured at the location l. In fact, z can be any number from $-\infty$ to $+\infty$ without any prior knowledges. However, some ranges can be unrealistic. Therefore, it is assumed that z is uniformly distributed over the interval $[Z_{lob}, Z_{upb}]$, where Z_{lob} and Z_{upb} represent the lower and upper bounds of possible information which can be determined by engineering judgement. Therefore, the expected value of $d_{up}(Z, L = l)$ can be adopted to reflect the magnitude of $d_{up}(Z, L = l)$. In this paper, the sensitivity of information at location l is computed as,

$$SOI(\,L = l) = \int_{-\infty}^{+\infty} d_{up}(Z = z, L = l) f_u(z) dz \approx \frac{1}{Z_{upb} - Z_{lob}} \int_{Z_{lob}}^{Z_{upb}} r_{up}(Z = z, L = l) dz \tag{12}$$

It can be inferred from Eq. (12) that SOI varies with location L. If Z is a vector, Eq. (12) becomes a multiple integral with Z integrated over all dimensions. Moreover, a monitoring location with a large SOI tends to have significant impact on the change of reliability index while monitoring location with a small SOI indicates that the monitoring action is not valuable. The computation of Eq. (12) involves an operation of integral, which requires numerical discretization. Hence, the computational complexity depends on the scheme of such numerical discretization. Assume that the integral space is discretized into n_{dis} pieces. Subsequently, Eq. (12) can be calculated as,

$$SOI(\, L = l) \approx \frac{1}{Z_{upb} - Z_{lob}} \sum_{i=1}^{n_{dis}} d_{up}(Z = z_i, L = l)\, \Delta_z \tag{13}$$

where $z_i = (2i - 1)\Delta_z$ is the point centered at the integral pieces and $\Delta_z = \left(Z_{upb} - Z_{lob}\right)/n_{dis}$. Eq. (12) needs to investigate the estimate of reliability updating n_{dis} times, which is computationally very intensive and not practical. Concerning this issue, the computation of $d_{up}(Z = z, L = l)$ in Eq. (11) needs to be optimized.

3.2 Computational details of *SOI*

The computation of $d_{up}(Z = z, L = l)$ needs to investigate the estimates of $\Pr(F)$ and $\Pr(F|Z)$ for n_d times. These probabilities can be possibly rare for some cases. To enhance the computational efficiency and robustness of the estimates of $\Pr(F)$ and $\Pr(F|Z)$, *SS* along with a strategy of decomposing $\Pr(F \cap Z)$ into $\Pr(Z|F) \cdot \Pr(F)$ is utilized in this paper. Therefore, the following equation is represented to estimate $\Pr(F|Z)$,

$$\Pr(F|Z) = \frac{\Pr(Z|F) \cdot \Pr(F)}{\Pr(Z)} \tag{14}$$

Eq. (14) optimizes the computation of Eq. (10) by decomposing $\Pr(F \cap Z)$ into $\Pr(Z|F)$ and $\Pr(F)$ via Bayes' theorem. This strategy completely avoids the computation of the probability of the rare event of $\Pr(F \cap Z)$. Integrating with *SS*, Eq. (14) can be rewritten as,

$$\Pr(F|Z) = \frac{\Pr(Z|F)}{\Pr(Z)} P\left(\bigcap_{i=1}^{m} F_i\right) = \frac{\Pr(Z|F)}{\Pr(Z)} P(F_1) \prod_{i=1}^{m-1} P(F_{i+1}|F_i) \tag{15}$$

where F_i denotes the intermediate failure event of $g(\boldsymbol{X})$, m denotes the number of subsets and F_m is the target failure event. Given that $F = F_m$, Eq. (15) can be further simplified as,

$$\Pr(F|Z) = \frac{\Pr(Z|F_m)}{\Pr(Z)} P(F_1) \prod_{i=1}^{m-1} P(F_{i+1}|F_i) \quad (16)$$

This indicates that the computation of $\Pr(F|Z)$ only relies on the estimates of $\Pr(Z)$ and $\Pr(Z|F_m)$ once the estimate of $\Pr(F)$ is completed through *SS*. For different information, the estimate of $\Pr(F|Z)$ can be as simple as reevaluating $\Pr(Z)$ and $\Pr(Z|F_m)$ based on the samples remained in the last target subset. However, we often encounter the situation where $\Pr(F|Z)$ cannot be estimated with sufficient accuracy due to the insufficient samples. This inaccuracy can lead to the associated inaccurate computation of $SOI(\ L = l)$. To overcome this limitation, samples generated through Markov Chain Monte Carlo simulation (*MCMC*) in each subset $S^i, i = 1,2, \ldots, m$ should be sufficient so that $\mathrm{COV}_{P_{F|Z}}$ is smaller than COV_{thr}, where $\mathrm{COV}_{P_{F|Z}}$ and COV_{thr} denote the coefficient of variation (COV) of $\Pr(F|Z)$ and the prescribed threshold, respectively. Toward this goal, the number of intermediate failure samples for *SS* is redefined in an adaptive way to facilitate the robust estimation of $\Pr(F|Z)$. Therefore, procedures for estimating $\Pr(F|Z)$ through the adaptive adjustment of N_{ss} is summarized in the following procedures:

- **Step 1**: Define initial parameters COV_{thr}, N_{ss} and p_0 for *SS*. In this paper, the UQLab toolbox with *Reliability* module in MATLAB® software is used. Other sets for performing *SS* follows the default settings in UQLab [48], [49].
- **Step 2**: Perform *SS* and record the computational results such as P_F, t_is, COV_{P_F} and S^i etc. In this step, the proposal distribution for *MCMC* is selected to be uniform. Moreover, a conceptual illustration of this computation for a 2D problem is presented in Fig 1. For this step, all the failure samples are kept for the sake of computing $\Pr(Z|F)$ in step 4.
- **Step 3:** Estimate $\Pr(Z)$ with the following limit state function,

$$h_1(p, \boldsymbol{x}) = P - c_1 L(\boldsymbol{X}) \quad (17)$$

In most cases, $\Pr(Z)$ can be estimated through *MCS* if the estimate of $L(\boldsymbol{x})$ is model free. Otherwise, $\Pr(Z)$ can be estimated through *SS*.

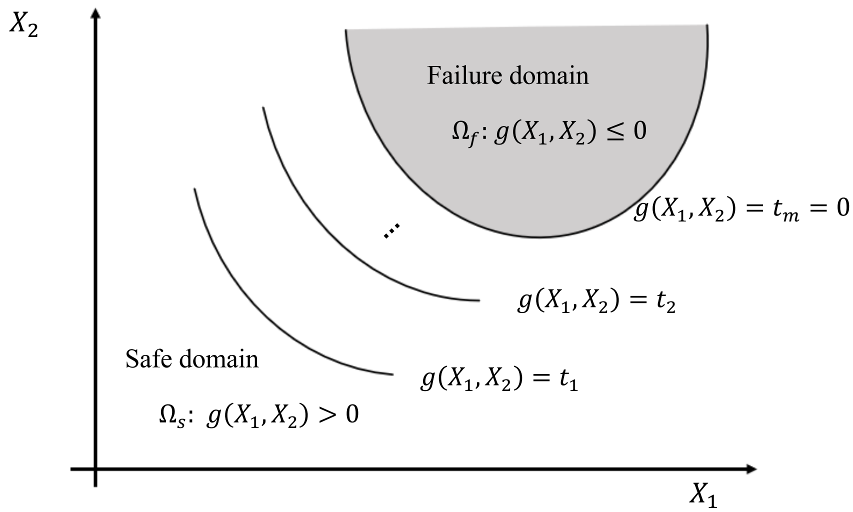


Fig 1. Illustration of SS for a 2D example with safe and failure domains and the limit state $g(X_1, X_2) = 0$

- **Step 4:** Estimate $\Pr(Z|F)$ based on the kept failure samples in step 2 with the following limit state function,

$$h_2(P, \boldsymbol{X}) = P - c_2 L(\boldsymbol{X}) \tag{18}$$

- **Step 5:** Check if $\mathrm{COV}_{P_{F|Z}} \leq COV_{thr}$. Go to Step 6 if satisfied; otherwise, reset $N_{ss} = N_{ss}^{last} + \Delta N_{ss}$ and go back to Step 2, where N_{ss}^{last} denotes the number of intermediate failure samples in each subset in the last iteration.
- **Step 6:** Output $\Pr(F)$ and $\Pr(F|Z)$.

Essentially, step 5 investigates the computation of $\mathrm{COV}_{P_{F|Z}}$ which impacts the computational robustness of the updated reliability. Let P_F, P_Z, $P_{Z|F}$ and $P_{F|Z}$ denote $\Pr(\mathrm{F})$, $\Pr(\mathrm{Z})$, $\Pr(Z|F)$ and $\Pr(F|Z)$for the sake of readability of this manuscript. To this end, $\mathrm{COV}_{P_{F|Z}}$ is computed in the following context. In virtue of the equality $\mathrm{Var}(AB) = [\mathrm{E}(A)]^2\mathrm{Var}(B) + [\mathrm{E}(B)]^2\mathrm{Var}(A) + \mathrm{Var}(A)\mathrm{Var}(B)$, where A and B are two mutually independent random variables, the following equation holds true,

$$\begin{aligned}\mathrm{COV}_{P_{F|Z}} &= \frac{Var\left(P_{F\cap Z}\frac{1}{P_Z}\right)^{\frac{1}{2}}}{E\left(P_{F\cap Z}\frac{1}{P_Z}\right)} \\ &= \frac{\left[[\mathrm{E}(P_{F\cap Z})]^2\mathrm{Var}\left(\frac{1}{P_Z}\right) + \left[\mathrm{E}\left(\frac{1}{P_Z}\right)\right]^2\mathrm{Var}(P_{F\cap Z}) + \mathrm{Var}(P_{F\cap Z})\mathrm{Var}\left(\frac{1}{P_Z}\right)\right]^{\frac{1}{2}}}{E(P_{Z|F})E(P_F)E\left(\frac{1}{P_Z}\right)}\end{aligned} \tag{19}$$

where P_F, $P_{F\cap Z}$ and $P_{Z|F}$ denote $\Pr(F)$, $\Pr(F \cap Z)$ and $\Pr(Z|F)$, $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$ represent the operations of mean and variance. Moreover, $\mathrm{E}(P_{F\cap Z}) = E(P_{Z|F})E(P_F)$. The computation of $\mathrm{Var}(P_{F\cap Z})$, $E\left(\frac{1}{P_Z}\right)$ and $\mathrm{Var}\left(\frac{1}{P_Z}\right)$ is elaborated next. First, $\mathrm{Var}(P_{F\cap Z})$ can be estimated according to the following equation,

$$\begin{gathered}\mathrm{Var}(P_{F\cap Z}) = \mathrm{Var}(P_F P_{Z|F}) \\ = [\mathrm{E}(P_F)]^2\mathrm{Var}(P_{Z|F}) + [\mathrm{E}(P_{Z|F})]^2\mathrm{Var}(P_F) + \mathrm{Var}(P_F)\mathrm{Var}(P_{Z|F})\end{gathered} \tag{20}$$

If N_{ss} is sufficiently large, the following equation holds true,

$$\lim_{N_{ss}\to 0} \mathrm{E}(P_F) \cong \tilde{P}_F \tag{21}$$

where $\tilde{P}_F$ denotes the ground truth of the unconditional probability of failure. The variance of P_F can be correspondingly calculated as,

$$\mathrm{Var}(P_F) = \mathrm{COV}_{P_F}^2[\mathrm{E}(P_F)]^2 \tag{22}$$

and the COV of P_f, COV_{P_F}, is calculated as,

$$\mathrm{COV}_{P_F} = \sqrt{\sum_{i=1}^{m} \mathrm{COV}_{\hat{p}_i}^2}, \text{for } i = 2,3,\dots,m \tag{23}$$

Generally, the COV of each $\hat{p}_i$ can be estimated as,

$$\mathrm{COV}_{\hat{p}_1} = \sqrt{\frac{1-P_i}{P_i N}}, \qquad \text{for } i = 1 \tag{24}$$

and

$$\mathrm{COV}_{\hat{p}_i} = \sqrt{\frac{1-P_i}{P_i N}(1+\gamma_i)}, \qquad \text{for } i = 2,3,\dots,m \tag{25}$$

where γ_i is a computational index that can be determined as,

$$\gamma_i = 2\sum_{k=1}^{N/N_c - 1}\left(1 - \frac{kN_c}{N}\right)\rho_i(k) \tag{26}$$

where $\rho_i(k)$ denotes the correlation coefficient at lag k of the stationary sequence $\left\{I_{j,k}^{\{i\}}: k = 1, \dots, N/N_c\right\}$, which can be calculated as,

$$\rho_i(k) = R_i(k)/R_i(0) \tag{27}$$

The covariance sequence $\{R_i(k): i = 0, \dots, N/N_c - 1\}$ can be estimated based on Markov chain samples as follows,

$$R_i(k) \cong \left(\frac{1}{N_{ss} - kN_c} \sum_{j=1}^{N_c} \sum_{i=1}^{N_{ss}/N_c - k} I_{j,l}^{\{i\}} I_{j,l+k}^{\{i\}}\right) - p_i^2 \tag{28}$$

where N_c denotes the number of Markov chains and $I_{j,l}^{\{i\}}$ denotes the failure indicator for the kth sample in the jth Markov chain simulation level i. Typically, COV_{P_F} based on real simulations is slightly larger than the theoretical one. To ensure the accuracy, the threshold of COV_{P_F} can be set slightly stricter than the desired level. Moreover, the mean and the variance of $P_{Z|F}$ are estimated as,

$$\mathrm{E}\left(P_{Z|F}\right) = P_{Z|F} \tag{29}$$

and

$$\mathrm{Var}\left(P_{Z|F}\right) = NP_{Z|F}\left(1 - P_{Z|F}\right) \tag{30}$$

Therefore, $\mathrm{Var}(P_{F\cap Z})$ can be obtained by combining Eq. (19) ~ (30). Moreover, $\mathrm{E}\left(\frac{1}{P_Z}\right)$ and $\mathrm{Var}\left(\frac{1}{P_Z}\right)$ can be estimated through numerical simulation by taking the reciprocal of a normal random variable (The Central Limit Theorem) after its COV and mean are acquired. Importantly, the computation of COV of P_Z depends on the type of reliability method (i.e., *MCS* or *SS*). One should note that the proposed approach overperforms those approaches that rely on the limit state function of joint event (i.e., $\mathrm{J}(U, \boldsymbol{X})$ defined in Eq. (10)) [7], [8]. Different from the computational scheme that needs to estimate $\Pr(F \cap Z)$, $\Pr(Z|F)$ only focuses on the failure domain of the performance function, which completely avoids unnecessary computational efforts of estimating $\Pr(Z|F)$. To clarify this point, let Pr(F)= 10^{-3} and Pr(Z)= 10^{-2} and F and Z be mutually independent. Then the joint event Pr(F ∩ Z) can be as small as 10^{-5}. This means that the total number of simulations based on F ∩ Z can be larger than 25000 if the batch size is set as 5000. However, the proposed approach only needs to estimate Pr(F) and subsequently estimate Pr(F|Z) based on the failure samples from Pr(F). Because these failure samples are already calculated through performance function in the procedure of estimating Pr(F), estimation of Pr(F|Z) does not require any simulation. This indicates that the total number of simulations is around 15000. Therefore, the cost of simulation through proposed approach relies on Pr(F) but not Z.

3.3 Analysis of optimal monitoring location

The metric *SOI* can be leveraged to derive the optimal monitoring location that has the most significant impact on the change of updated reliability index. Generally, the optimal monitoring location can be identified according to the following equation,

$$\boldsymbol{l}^* = \arg\max_{l \in \Gamma} SOI\ (\boldsymbol{L} = \boldsymbol{l}) \tag{31}$$

where $\boldsymbol{l}^*$ denotes the vector of optimal monitoring location ($\boldsymbol{l}^*$ is not bold as l^* if it denotes one location), Γ represents the domain of the global feasible monitoring locations and $SOI(\boldsymbol{L} = \boldsymbol{l})$ represents the sensitivity of information at location $\boldsymbol{l}$. However, the optimization problem represented in Eq. (31) can be computationally prohibitive due to the complex topology of monitoring location with large dimension or discretization. To further interpret this point, let N_T denote the total number of discretized mapping points, it can be calculated as,

$$N_T = \prod_{i=1}^{N_{dim}} N_i^d \tag{32}$$

where N_{dim} is the number of the dimension of Γ and N_i^d represents the number of discretized points in the i^{th} direction. For example, N_T can be as large as 10^6 for topology with three dimensions if it is discretized into 100 pieces in each dimension. The optimization defined in Eq. (31) becomes computationally intractable if *SOIs* of all these discretized points are calculated. To efficiently solve the optimization problem in Eq. (31), a surrogate model-based optimization solution is adopted to find l^*. In this paper, the Kriging surrogate model with noisy responses is adopted to tackle the inconsistent estimate of *SOI* presented in section 3.2. Based on Kriging surrogate model with noisy response, *SOI* for each discretized sample l can be represented as:

$$\widehat{SOI}(\boldsymbol{l}) = \boldsymbol{F}(\boldsymbol{l}, \boldsymbol{\beta}) + \Psi(\boldsymbol{l}) + \epsilon_k = \boldsymbol{f}^T(\boldsymbol{l})\boldsymbol{\beta} + \Psi(\boldsymbol{l}) + \epsilon_k, \tag{33}$$

where $\widehat{SOI}(\boldsymbol{l})$ denotes the estimated value of *SOI* at $\boldsymbol{L} = \boldsymbol{l}$ estimated through the Kriging surrogate model, $\Psi(\boldsymbol{l})$ denotes the Gaussian process, ϵ_k is the additive noise of response which follows a zero-mean Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}_{\mathrm{n}}$, and $F(\boldsymbol{l}, \boldsymbol{\beta})$ is the so-called regression basis denoting the Kriging trend, which can be a constant, a polynomial term or any mathematical form. Moreover, $\boldsymbol{f}(\boldsymbol{l})$ is the vector of Kriging basis and $\boldsymbol{\beta}$ is the vector of regression coefficients. Specifically, $\boldsymbol{f}^T(\boldsymbol{l})\boldsymbol{\beta}$ often takes the form of ordinary (β_0), linear (β_0+$\sum_{n=1}^{N_{dim}}[\beta]_n[x]_n$) or quadratic ($\beta_0$+$\sum_{n=1}^{N_{dim}}[\beta]_n[x]_n$+$\sum_{n=1}^{N_{dim}}\sum_{k=n}^{N_{dim}}[\beta]_{nk}[x]_n[x]_k$), where $[\beta]_n$ and $[x]_n$ denote the n^th^ component of $\boldsymbol{\beta}$ and $\boldsymbol{l}$, respectively. Moreover, $\Psi(\boldsymbol{l})$ has a zero mean and a covariance matrix between two points, $\boldsymbol{l}_i$ and $\boldsymbol{l}_j$:

$$\mathrm{COV}\left(\Psi(\boldsymbol{l}_i), \Psi(\boldsymbol{l}_j)\right) = \sigma^2 R\left(l_i, l_j; \boldsymbol{\theta}\right), \tag{34}$$

where σ^2 is the process variance or the generalized mean square error from the regression part and $R\left(\boldsymbol{l}_i, \boldsymbol{l}_j; \boldsymbol{\theta}\right)$ is the correlation function or the kernel function representing the correlation function of the process with hyper-parameter $\boldsymbol{\theta}$. Multiple types of correlation functions are available for Kriging models including linear, exponential, Gaussian, Matérn models, among others [50]. In this paper, the Gaussian kernel function is implemented:

$$R\left(\boldsymbol{l}_i, \boldsymbol{l}_j; \boldsymbol{\theta}\right) = \prod_{n=1}^{N_{dim}} \exp\left(-[\theta]_n \left([\boldsymbol{l}_i]_n - [\boldsymbol{l}_j]_n\right)^2\right), \tag{35}$$

where $[\boldsymbol{l}_i]_n$ is the n^th^ component of the realization $\boldsymbol{l}_i$, $\boldsymbol{\theta}$ denotes the hyper-parameter that can be estimated via maximum likelihood estimation (MLE) or cross-validation [50]. It is shown that the Kriging prediction is very sensitive to the value of $\boldsymbol{\theta}$ [51]–[53]. In this article, the optimal hyper-parameter $\boldsymbol{\theta}^*$ is searched through MLE:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta} \epsilon \mathbb{R}}{\operatorname{argmin}} \frac{1}{2}\left[\log\left(\det\left(R\left(\boldsymbol{l}_i, \boldsymbol{l}_j; \boldsymbol{\theta}\right)\right)\right) + n_{des}\log(2\pi\sigma^2) + n_{des}\right], \tag{36}$$

where n_{des} is the number of design-of-experiment (DoE) points. Thus, for a number of DoE (training) points, $S_{DoE} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \dots, \boldsymbol{l}_m]$, and the corresponding responses from the performance function $\mathbf{Y} = [SOI(\boldsymbol{l}_1), SOI(\boldsymbol{l}_2), \dots, SOI(\boldsymbol{l}_m)]$, the traditional BLUP (Best Linear Unbiased Predictor) estimation of Kriging prediction for a group of testing points, $S_t = [\boldsymbol{l}_1, \boldsymbol{l}_2, \dots, \boldsymbol{l}_{N_t}]$ gives:

$$\mu_{\hat{k}}(\boldsymbol{l}_t) = \boldsymbol{f}^T(\boldsymbol{l}_t)\widetilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{r}}(\boldsymbol{l}_t)^T\boldsymbol{\gamma}, \qquad \boldsymbol{l}_t \in S_t. \tag{37}$$

where $\boldsymbol{l}_t$ denotes testing samples. Moreover, let $\boldsymbol{C} = \sigma^2\boldsymbol{R} + \boldsymbol{\Sigma}_{\mathrm{n}}$, $\boldsymbol{\Sigma}_{\mathrm{n}} = \sigma_n^2\boldsymbol{I}$ (where $\boldsymbol{I}$ is an identity matrix and σ_n^2 is the variance of noise of SOI) and $\tau = \sigma^2/(\sigma_n^2 + \sigma^2)$, where σ^2 is the Gaussian process variance, and $\boldsymbol{u}(\boldsymbol{l}_t)$ are:

$$\sigma^2 = \frac{1}{m}(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta})^T \boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta}) \tag{38}$$

The parameters presented in Eq. (37) can be calculated as,

$$\begin{gathered}
\widetilde{\boldsymbol{\beta}} = (\boldsymbol{F}^T\boldsymbol{C}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{C}^{-1}\boldsymbol{Y}, \\
\boldsymbol{\gamma} = \boldsymbol{C}^{-1}\left(\boldsymbol{Y} - \boldsymbol{F}\widetilde{\boldsymbol{\beta}}\right), \\
\boldsymbol{r}(\boldsymbol{l}_t) = [R(\boldsymbol{l}_1, \boldsymbol{l}_t; \boldsymbol{\theta}), \dots . R(\boldsymbol{l_m}, \boldsymbol{l}_t; \boldsymbol{\theta})]_{1\times m}^T, \boldsymbol{l}_t \in S_t \\
\tilde{\boldsymbol{r}}(\boldsymbol{l}_t) = (1-\tau)\boldsymbol{r}(\boldsymbol{l}_t) \\
\widetilde{\boldsymbol{R}} = (1-\tau)\boldsymbol{R} + \tau\boldsymbol{I} \\
\boldsymbol{F} = [\boldsymbol{f}(\boldsymbol{l}_1), \boldsymbol{f}(\boldsymbol{l}_2), \dots \boldsymbol{f}(\boldsymbol{l}_m)]^T.
\end{gathered} \tag{39}$$

Then, the mean-square error (MSE) of $\widehat{SOI}(\boldsymbol{l}_t)$ can be calculated by:

$$\sigma_{\hat{k}}^2(\boldsymbol{l}_t) = (\sigma_n^2 + \sigma^2)\left(1 + \boldsymbol{u}^T(\boldsymbol{l}_t)\left(\boldsymbol{F}^T\widetilde{\boldsymbol{R}}^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{u}(\boldsymbol{l}_t) - \tilde{\boldsymbol{r}}^T(\boldsymbol{l}_t)\widetilde{\boldsymbol{R}}^{-1}\tilde{\boldsymbol{r}}(\boldsymbol{l}_t)\right), \tag{40}$$

where $\boldsymbol{u}(\boldsymbol{l}_t) = \boldsymbol{F}^T\boldsymbol{C}^{-1}\boldsymbol{r}(\boldsymbol{l}_t) - \boldsymbol{f}(\boldsymbol{l}_t)$. According to Kriging theory, for all testing points, S_t, the outputs $\boldsymbol{Y} = [g(\boldsymbol{l}_1), g(\boldsymbol{l}_2), \dots, g(\boldsymbol{l}_t)]$ from the Kriging model are parameterized with the mean, $\mu_{\hat{g}}(\boldsymbol{l}_t)$, and the variance, $\sigma_{\hat{g}}^2(\boldsymbol{l}_t)$:

$$\widehat{SOI}(\boldsymbol{l}_t) \sim N\left(\mu_{\hat{k}}(\boldsymbol{l}_t), \sigma_{\hat{k}}^2(\boldsymbol{l}_t)\right), \quad \boldsymbol{l}_t \in S_t. \tag{41}$$

The general principle of surrogate-based optimization is to start with a small number of training points that compute SOI to build a surrogate for $\widehat{SOI}(\boldsymbol{L} = \boldsymbol{l})$ and subsequently refine the Kriging surrogate model by adaptively adding new training samples until the target $\boldsymbol{l}^*$ is steadily identified. The procedure discussed above is elaborated in the following steps:

- **Step 1**: Discretizing the regions of observation, Ω_{ob}, into discretized points and denote these samples as S_{ob}.
- **Step 2**: Select a limited number of points from S_Γ as initial training points $\boldsymbol{l}_{in}$ for Kriging construction. As suggested by [44], the number of $\boldsymbol{l}_{in}$ should be greater than $\frac{(N_{dim}+1)(N_{dim}+2)}{2}$. Note that $\boldsymbol{l}_{tr}$ can change upon every iteration of active learning but it is equal to $\boldsymbol{l}_{in}$ in the first iteration.
- **Step 3**: Construct the Kriging model with current training points $\boldsymbol{l}_{tr}$. Denote the Kriging model as $\widehat{SOI}(\boldsymbol{l})$. Construction is based on UQLab toolbox in MATLAB®, with ordinary Kriging basis and Gaussian correlation function. The model type is selected as prediction with noisy responses and other parameters follow default settings.

Subsequently, the Kriging responses $\mu_{\hat{k}}(\boldsymbol{l})$ and variances $\sigma_{\hat{k}}^2(\boldsymbol{l})$ can be acquired from UQLab toolbox [50].

- **Step 4**: To search for the maximum value of SOI, the expected improvement learning function (EI) for global optimization is adopted. The next training point is selected according to the following criterion and is denoted as $\boldsymbol{l}_{tr}^*$.

$$\boldsymbol{l}_{tr}^* = \arg\max_{\boldsymbol{l} \in S_\Gamma} \mathrm{EI}(\boldsymbol{l}) \tag{42}$$

where,

$$\mathrm{EI}(\boldsymbol{l}) = \left(\mu_{\hat{k}}(\boldsymbol{l}) - \widehat{SOI}_{max}^*\right)\Phi\left(\frac{\mu_{\hat{k}}(\boldsymbol{l}) - \widehat{SOI}_{max}^*}{\sigma_{\hat{k}}(\boldsymbol{l})}\right) + \sigma_{\hat{k}}(\boldsymbol{l})\varphi\left(\frac{\mu_{\hat{k}}(\boldsymbol{l}) - \widehat{SOI}_{max}^*}{\sigma_{\hat{k}}(\boldsymbol{l})}\right) \tag{43}$$

where $\widehat{SOI}_{max}^*$ denotes the maximum *SOI* among $\boldsymbol{l}_{tr}$ in the current iteration.

- **Step 5**: Determine if the stopping criterion ($max(EI) \leq EI_{thr}$) has been satisfied in the current iteration, where EI_{thr} denotes the threshold value. In this paper, EI_{thr} is set as 10^{-5} and the maximum number of iterations is set as 100. Go to Step 6 if satisfied; otherwise, go back to Step 3.
- **Step 6:** Output $\boldsymbol{l}^*$ and $\widehat{SOI}$ for S_{ob}.

The above procedure presents an efficient approach for evaluating *SOI* for all potential monitoring points. However, the computational complexity of constructing Kriging surrogate model increases substantially as N_{dim} grows. This is known as the 'curse of dimensionality', which can be further optimized in the future. To explore the performance of the proposed framework, a geotechnical case that investigates tunneling-induced settlement to building damage is investigated in the next section.

4 Case study

4.1 Description of the physical model

Settlements caused by the construction of tunnel can threaten the functionalities and integrity of structures and infrastructure systems aboveground. This process is illustrated in Fig 2, where the *y* axis follows the reverse direction of tunnel advance and *x* axis is perpendicular to the tunnel longitudinal axis [15], [54]. Origin is the intersection of the extensional line of building wall and the *y*-axis. Moreover, the positive degree refers to alignments counterclockwise with respect to the *x*-axis. Starting from tunnel portal $y = y_f$ and adaptively advancing towards $y = -\infty$ with the tunnel boring machine (TBM), a tunnel is under construction with the tunnel face located at $y = y_s$. In this paper, y_f is assumed to originate from infinity with $y_f = +\infty$. A building wall of length l_{build}, denoted by a reference point $\hat{A}$, is located at a distance d_{orig} from the origin and aligned θ_r degrees with respect to the tunnel transverse plane. To better interpret the concept, a 3D plot of this physical model is showcased in Fig 3, where d and z_0 denote the diameter of tunnel and the depth from surface of ground $(z = 0)$ to the center of tunnel. The objective is to identify the optimal monitoring location on the feasible grounds to better tracking risks of building subsidence as the tunneling excavation proceeds. As introduced later, these feasible grounds are typically off the location of TBM. The formulas that predict the tunneling-induced subsidence of building, analysis of reliability updating and exploration of the optimal monitoring location are subsequently introduced in the following context.

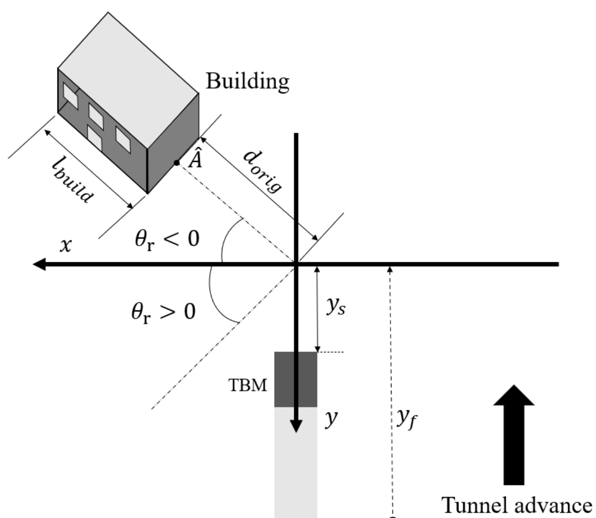


Fig 2. Illustration of tunneling-induced settlements.

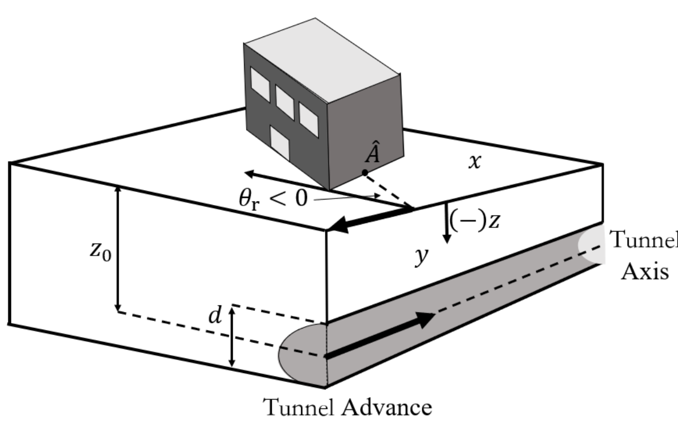


Fig 3. 3D illustration of tunnel and building wall positions.

The soil over the excavated underground space can be viewed as a distributed loading with the other ending node fixed at $y = +\infty$. According to [15], [55], [56], the settlement of ground can be calculated as,

$$S(x, y, z, d, y_s, y_0, y_f, z_0, V_L, K_x, K_y) = -1000 \cdot S_{max} \cdot \exp\left[-\frac{x^2}{2 \cdot K_x^2 \cdot (z_0 - z)^2}\right] \cdot \left[\Phi\left(\frac{y - (y_s + y_0)}{K_y \cdot (z_0 - z)}\right) - \Phi\left(\frac{y - y_f}{K_y \cdot (z_0 - z)}\right)\right] \quad (44)$$

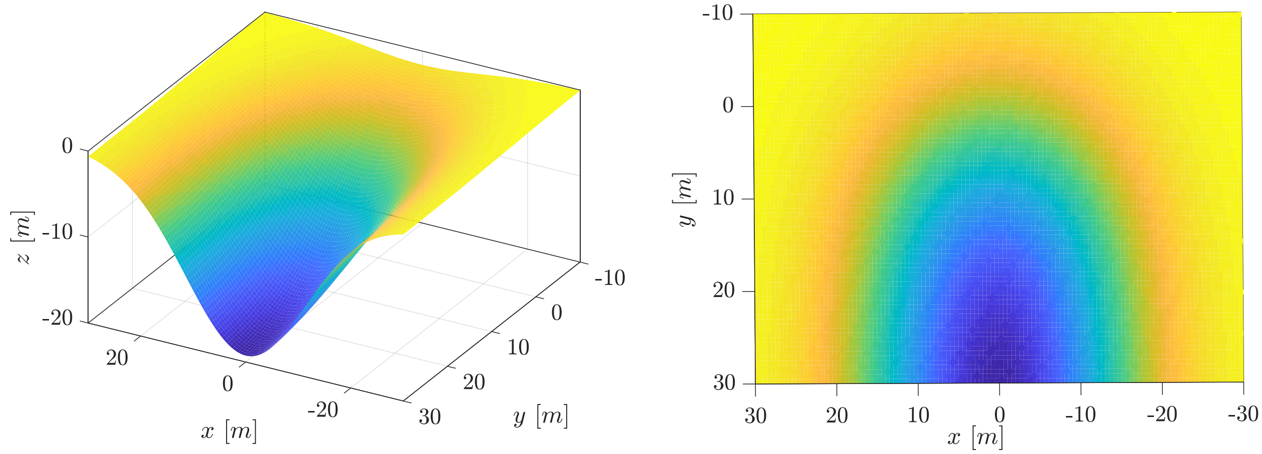
where V_L is the volume ground loss per unit and K_x and K_y are non-dimensional through width parameters reflecting characteristics of the soil and describing the Gaussian settlement profiles in the transverse and longitudinal direction. It is typically assumed that $K_x = K_y = K$ [57]. Moreover, S_{max} denotes the absolute value of maximum settlement at y ($y \geq y_s$) and can be calculated as,

$$S_{max} = \frac{V_L \cdot \pi \cdot d^2}{\sqrt{2\pi} \cdot K_x \cdot (z_0 - z) \cdot 4} \quad (45)$$

where y_0 in Eq. (44) is the horizontal shift of the longitudinal settlement profile with respect to the tunnel face and can be calculated as,

$$y_0 = -\Phi^{-1}(\delta) \cdot K \cdot z_0 \quad (46)$$

where δ represents the ratio between the surface settlement above the tunnel face and S_{max} at $y = +\infty$. In this paper, δ is defined as 0.3 for the sake of practical consideration [40], [41]. The shape of settlement is presented in Fig 4. It can be observed that the shape of settlement along the x-axis follows the PDF of Gaussian distribution, while along the y-axis, the shape is close to the CDF of Gaussian distribution. The settlement reaches the highest value at $(0,30,0)$.



(a) (b)

Fig 4. Settlement produced by tunnel excavation in (a) the 3D surface plot and (b) the x-y view with $d = 12\text{m}$, $y_s = 0$, $z_0 = 23\text{m}$, $V_L = 0.5\%$ and $K = 0.5$.

By treating the building wall as a weightless linear elastic rectangular beam, the response of the building to the settlement is modeled through the equivalent beam method [60]. The distribution of tensile strains along the beam is governed by the shape of the deflection and the mode of deformation. The extreme fiber strains caused by bending and shear ε_{br} and ε_{dr} can be calculated according to the following equations,

$$\varepsilon_{br}\left(V_L, K, \frac{E}{G}\right) = (\varepsilon_{b\,max} + \varepsilon_h) \cdot E_{\varepsilon_{br}} \tag{47}$$

$$\varepsilon_{dr}\left(V_L, K, \frac{E}{G}\right) = \left[\varepsilon_h\left(1 - \frac{E}{4G}\right) + \sqrt{\frac{\varepsilon_h^2}{16}\left(\frac{E}{G}\right)^2 + \varepsilon_{d\,max}^2}\right] \cdot E_{\varepsilon_{dr}} \tag{48}$$

where $\frac{E}{G}$ represents the ratio between the Young's modulus and the shear modulus of the building material and $E_{\varepsilon_{br}}$ and $E_{\varepsilon_{dr}}$ are multiplicative model errors. In this paper, $\frac{E}{G}$, $E_{\varepsilon_{br}}$ and $E_{\varepsilon_{dr}}$ are modeled as random variables. Moreover, $\varepsilon_{b\,max}$ and $\varepsilon_{d\,max}$ are the maximum bending and shear strains due to the deflection. Specifically, $\varepsilon_{b\,max}$ and $\varepsilon_{d\,max}$ are calculated separately for the different zones of the building. The building zones that have settlements induced by the tunnel can be typically classified into two types: the sagging and hogging deflections. As shown in Fig 5, the main difference of them lies in the position of the profile curvature change: sagging deflection represents upwards concavity while hogging deflection indicates downwards concavity. To better illustrate the foregoing difference, Fig 6 showcases the sagging and hogging deflections in the different zones of a building.

The number of inflection points along the building depends on the three parameters l_{build}, d_{orig} and θ_r. Moreover, the type of deflection of a building depends on the number of inflection points, which are summarized in Table 1. A conceptual plot for the last case in Table 1 is shown in Fig 6, where the building is divided in to three zones: one sagging zone and two hogging zones. Let l_{ref} denote the horizontal distance between two reference points and Δ_{ref} be the relative deflection; the deflection ratio Δ_{ref}/l_{ref} for different deflection types can be represented as Δ_{sag}/l_{sag} and Δ_{hog}/l_{hog}. The maximum bending and shear strains, $\varepsilon_{b\,max}$ and $\varepsilon_{d\,max}$, for a given zone (sagging or hogging) can be calculated as follows [61],

$$\varepsilon_{b\,max} = \frac{\Delta_{ref}/l_{ref}}{\left(\frac{l_{ref}}{12t} + \frac{3I}{2al_{ref}H}\frac{E}{G}\right)} \tag{49}$$

and

$$\varepsilon_{d\,max} = \frac{\Delta_{ref}/l_{ref}}{\left(1 + \frac{Hl_{ref}^2}{18I}\frac{G}{E}\right)} \tag{50}$$

where H denotes the height of building, $I = H^3/12$ is the inertia per unit length, t is depth of neutral axis and $a = t$ is the location of the fiber where strains are calculated. For sagging and hogging deflections, $t = H/2$ and H, respectively. Moreover, the resultant horizontal strain in the ground surface along the base of the team, ε_h, in Eq. (47) and (48) can be computed according to the following equation:

$$\varepsilon_h(x, y, z, V_L, K) \equiv \cos^2\theta_r \cdot \varepsilon_{h,xx} + \sin^2\theta_r \cdot \varepsilon_{h,yy} + 2 \cdot \cos\theta_r \sin\theta_r \cdot \varepsilon_{h,xy} \tag{51}$$

where $\varepsilon_{h,xx}$, $\varepsilon_{h,yy}$ and $\varepsilon_{h,xy}$ are the fields of strain in the ground. The maximum strain of the building ε_{max} can be determine according to the six parameters,

$$\varepsilon_{max} = \max\left[\varepsilon_{br}^{sag}, \varepsilon_{dr}^{sag}, \varepsilon_{br}^{hog,1}, \varepsilon_{dr}^{hog,1}, \varepsilon_{br}^{hog,2}, \varepsilon_{dr}^{hog,2}\right] \tag{52}$$

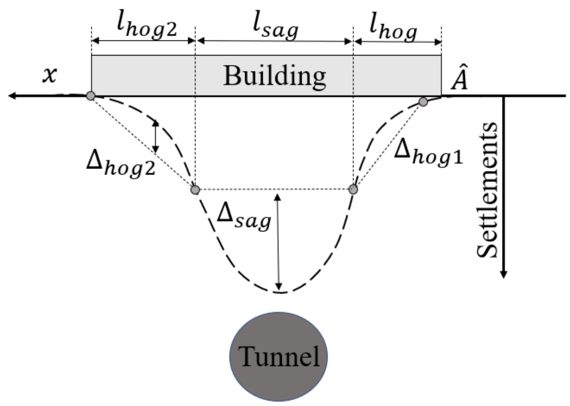


Fig 5. Conceptual illustration of sagging and hogging deflections in different zones of a building.

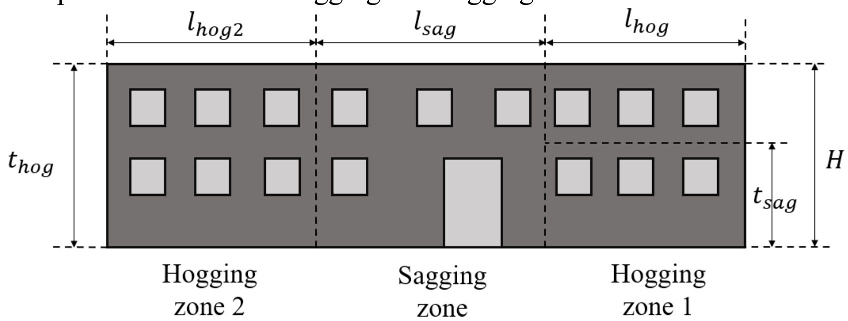


Fig 6. A building that is subjected to tunneling-induced settlement with 1 sagging and 2 hogging zones.

The location of the building determines the number of extreme fiber strains. It indicates that the building can be divided into sagging and hogging zones. Therefore, six notations including ε_{br}^{sag}, ε_{dr}^{sag}, $\varepsilon_{br}^{hog,1}$, $\varepsilon_{dr}^{hog,1}$, $\varepsilon_{br}^{hog,2}$ and $\varepsilon_{dr}^{hog,2}$ are sufficient to represent the 4 cases described in Table 1, where ε_{br}^{sag}, $\varepsilon_{br}^{hog,1}$ and $\varepsilon_{br}^{hog,2}$ are the maximum bending strains in sagging zone and ε_{dr}^{sag}, $\varepsilon_{dr}^{hog,1}$ and $\varepsilon_{dr}^{hog,2}$ are the maximum shear strains in hogging zone. For the first case, the last four terms are equal to zero due to the existence of one sagging zone; For the second case, one hogging zone indicates that the third and fourth terms are non-zero; For the third case, the last two terms are equal to zero while all the six terms are non-zero for the last case. Finally, the geometry of the cracks can be estimated. Moreover, the classification of damage due the cracks is summarized in Table 2, where ε_{lim} denotes the limit tensile strain.

Table 1. Types of deflection of building located in a specific location

Location of building	Number of inflections	Types of Deflection
Above the tunnel axis	0	1 sagging
Far from the tunnel axis	0	1 hogging

Starts in the sagging zone and reaches the hogging zone	1	1 sagging and 1 hogging
Central part in the sagging zone and lateral parts in the hogging zone	2	1 sagging and 2 hogging

Table 2. Classification of damage [61].

Category of damage	Normal degree of severity	Typical damage	Tensile strain ε_{max}(%)	Limiting strain ε_{lim}(%)
0	Negligible	< 0.1 mm	0-0.050	0.050
1	Very slight	< 1.0 mm	0.050-0.075	0.075
2	Slight	< 5.0 mm	0.075-0.150	0.150
3	Moderate	< 15.0 mm	0.150-0.300	0.300
4	Severe	< 25.0 mm	>0.300	-
5	Very Severe	> 25.0 mm	-	-

To calculate $\varepsilon_{h,xx}$, $\varepsilon_{h,yy}$ and $\varepsilon_{h,xy}$, let U_x and U_y denote the horizontal displacements in [mm] in the transvers and longitudinal direction, respectively, at a certain position with coordinate x, y, z in [m]. U_x and U_y can be calculated as,

$$U_x = \frac{x}{z_0 - z} \cdot S \tag{53}$$

and

$$U_y = 1000 \cdot \frac{V_L \cdot d^2}{8 \cdot (z_0 - z)} \cdot \left[\exp\left(\frac{-\left(y - (y_s + y_0)\right)^2 - x^2}{2 \cdot K_y^2 \cdot (z_0 - z)^2}\right) - \exp\left(\frac{-\left(y - y_f\right)^2 - x^2}{2 \cdot K_y^2 \cdot (z_0 - z)^2}\right)\right] \tag{54}$$

Therefore $\varepsilon_{h,xx}$, $\varepsilon_{h,yy}$ and $\varepsilon_{h,xy}$ can be calculated based on U_x and U_y,

$$\varepsilon_{h,xx} = \frac{\partial U_x}{\partial x} = \frac{S/1000}{z_0 - z} \cdot \left(1 - \left(\frac{x^2}{K_x^2 \cdot (z_0 - z)^2}\right)\right) \tag{55}$$

and

$$\varepsilon_{h,yy} = \frac{\partial U_y}{\partial y} = \frac{V_L \cdot d^2}{8 \cdot (z_0 - z)} \cdot \left[\begin{array}{c} \left(\frac{-2y + 2(y_s + y_0)}{2 \cdot K_y^2 \cdot (z_0 - z)^2}\right) \exp\left(\frac{-\left(y - (y_s + y_0)\right)^2 - x^2}{2 \cdot K_y^2 \cdot (z_0 - z)^2}\right) \\ -\left(\frac{-2y + 2y_f}{2 \cdot K_y^2 \cdot (z_0 - z)^2}\right) \exp\left(\frac{-\left(y - y_f\right)^2 - x^2}{2 \cdot K_y^2 \cdot (z_0 - z)^2}\right) \end{array} \right] \tag{56}$$

and

$$\varepsilon_{h,xy} = \frac{1}{2}\left(\frac{\partial U_x}{\partial y} + \frac{\partial U_y}{\partial x}\right) \tag{57}$$

where $\frac{\partial U_x}{\partial y}$ and $\frac{\partial U_y}{\partial x}$ read as,

$$\frac{\partial U_x}{\partial y} = \frac{x}{z_0 - z} \cdot \left(-\frac{V_L \cdot \pi \cdot d^2}{\sqrt{2\pi} \cdot K_x \cdot (z_0 - z) \cdot 4}\right) \cdot \left(\begin{array}{c} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(\frac{y-(y_s+y_0)}{K_y(z_0-z)}\right)^2}{2}} \cdot \left(\frac{1}{K_y(z_0 - z)}\right) \\ -\frac{1}{\sqrt{2\pi}} e^{-\frac{\left(\frac{y-y_f}{K_y(z_0-z)}\right)^2}{2}} \cdot \left(\frac{1}{K_y(z_0 - z)}\right) \cdot \exp\left(-\frac{x^2}{2 \cdot K_x^2 \cdot (z_0 - z)^2}\right) \end{array} \right) \tag{58}$$

and

$$\frac{\partial U_y}{\partial x} = \frac{V_L \cdot d^2}{8 \cdot (z_0 - z)} \cdot \frac{(-2x)}{2 \cdot K_x^2 \cdot (z_0 - z)^2} \left[\begin{array}{c} \exp\left(\frac{-\left(y - (y_s + y_0)\right)^2 - x^2}{2 \cdot K_y^2 \cdot (z_0 - z)^2}\right) \\ -\exp\left(\frac{-\left(y - y_f\right)^2 - x^2}{2 \cdot K_y^2 \cdot (z_0 - z)^2}\right) \end{array} \right] \tag{59}$$

4.2 Sensitivity of information (*SOI*) analysis

In this subsection, the computational procedure for the estimation of *SOI* for different locations is elaborated. According to [15], the event that tunnel excavation-caused building crack exceeds 0.1mm ($\varepsilon_{\mathrm{lim}} = 0.05\%$) is defined as the limit state, indicating damage level 1 in Table 2. Accordingly, the limiting strain for this case can be set as $\varepsilon_{lim} = 0.05\%$, leading to the limit state function (LSF), $g(\boldsymbol{X})$ for this case,

$$g(\boldsymbol{X}) = \varepsilon_{lim} - \varepsilon_{max}(\boldsymbol{X}) \tag{60}$$

where $\boldsymbol{X}$ denotes the vector of random variables. In this context, $\boldsymbol{X} = \left[V_L; K; \frac{E}{G}; E_{\varepsilon_{br}}^{sag}; E_{\varepsilon_{br}}^{hog,1}; E_{\varepsilon_{br}}^{hog,2}; E_{\varepsilon_{dr}}^{sag}; E_{\varepsilon_{dr}}^{hog,1}; E_{\varepsilon_{dr}}^{hog,2}\right]$, where $E_{\varepsilon_{br}}^{sag}, E_{\varepsilon_{br}}^{hog,1}, E_{\varepsilon_{br}}^{hog,2}, E_{\varepsilon_{dr}}^{sag}, E_{\varepsilon_{dr}}^{hog,1}$ and $E_{\varepsilon_{dr}}^{hog,2}$ are the errors of the equivalent beam model of Eq. (47) and (48) in the sagging and hogging zones, respectively. The probabilistic distribution of these 9 random variables is summarized in Table 3. Moreover, the failure domain Ω_f can be defined as,

$$\Omega_f = \{g(\boldsymbol{x}) \leq 0\} \tag{61}$$

where $\boldsymbol{x}$ is a stochastic realization from $\boldsymbol{X}$. To precisely assess and track the risk of the tunnelling-induced settlement to the building, the measurement of settlement at the location l_m, $s_m(l_m)$, is conducted over the region of observation, Ω_{ob}. Z_{lob} and Z_{upb} are equal to 5 [m] and 15 [m], respectively, based on engineering experience. Fig 7 illustratively interprets this strategy, where the light green region is represented as Ω_{ob}. Moreover, the relation between the measured and ground truth settlement can be read as:

$$S_m = S(x_m^i, y_m^i, z_m^i, V_l, K) + E_f + E_m = S(x_m^i, y_m^i, z_m^i, V_l, K) + E_E \tag{62}$$

where E_f is the model error interpreting the potential inaccuracy of the Gaussian settlement shape and E_m is the measurement error stemming from the manmade imprecision, imperfection of instruments etc. Let $E_E = E_f + E_m$, the likelihood function for this case is defined as:

$$L(v_l, k) = f_E\left(s_m^i - S(x_m^i, y_m^i, z_m^i, v_l, k)\right) \tag{63}$$

where v_l and k are the realizations of random variables V_l and K, and f_E is the PDF of the integrated error E_E. Therefore, the two augmented limit state function $h_1(P, \boldsymbol{X})$ and $h_2(P, \boldsymbol{X})$ can be sequentially defined based on Eq. (63). Before exploring l^* over Ω_{ob}, the procedures of estimating P(F) and P(F|Z) are elaborated. Therein, four measurements, $s_m(x_m^i, y_m^i, z_m^i), i = 1,2,3$ and 4 are ready to update P(F|Z) from P(F). The corresponding data and simulation results are reported in Table 4 together with the corresponding interpretative figures illustrated in Fig 8.

Table 3. Probabilistic distribution of random variables [15].

Random variable	Description	Type of distribution	Mean	Standard deviation
K(-)	Trough width parameter	Lognormal	0.3	0.06
V_L(%)	Volume loss	Lognormal	0.4	0.16
$\frac{E}{G}$(-)	Material ratio	Beta	2.5	0.045

$\begin{bmatrix} E_{\varepsilon_{br}}^{sag}; E_{\varepsilon_{br}}^{hog,1}; E_{\varepsilon_{br}}^{hog,2}; \\ E_{\varepsilon_{dr}}^{sag}; E_{\varepsilon_{dr}}^{hog,1}; E_{\varepsilon_{dr}}^{hog,2} \end{bmatrix}$(-)	Equivalent beam model errors	Lognormal	1.0	0.05
E_m(mm)	Measurement error	Normal	0	1
E_f(mm)	Settlement model error	Normal	0	2

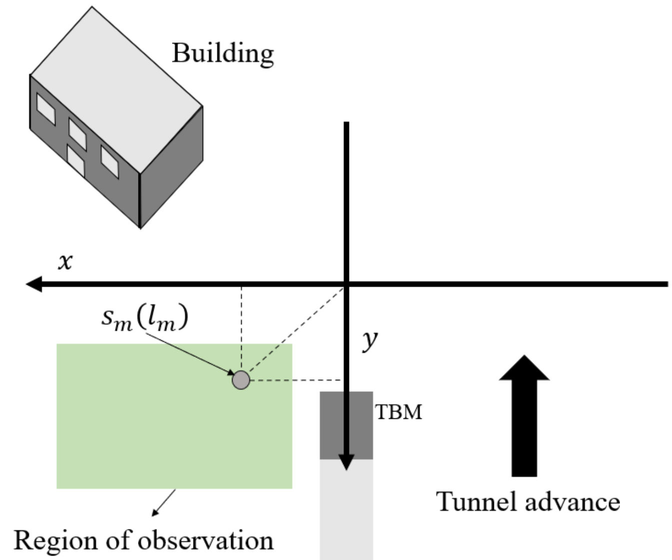


Fig 7. Conceptual illustration of Ω_{ob} and the monitoring measurement $s_m(l_m)$, where $l_m = (x_m, y_m, z_m)$ and $l_m \in \Omega_{ob}$.

Table 4. Simulation results of case study through the proposed reliability updating method, where $\text{COV}_{\text{thr}} = 0.05$ and $N_{in} = 10^4$.

Information	Position	Pr(F)	Pr(F|Z)	d_{up}	COV_{P_F}	$\text{COV}_{P_{F\|Z}}$	N_{eva}
$s_m(l_1) = 10$	(10,10,0)	8.26×10^{-3}	1.29×10^{-2}	0.069	0.0449	0.0471	56000
$s_m(l_2) = 10$	(15,15,0)	8.40×10^{-3}	8.47×10^{-2}	0.425	0.0375	0.0497	84000
$s_m(l_3) = 10$	(20,20,0)	8.31×10^{-3}	2.23×10^{-2}	0.161	0.0448	0.0494	56000
$s_m(l_4) = 10$	(25,25,0)	8.36×10^{-3}	9.84×10^{-3}	0.025	0.0453	0.0462	56000

Table 5. Simulation results via *MCS*, *FORM/SORM*, *IS* and the proposed method.

Method	$\Pr(F\|Z)$	$\Pr(F)$	N_{sim}
MCS	9.76×10^{-2}	8.35×10^{-3}	10^6
FORM [2]	N/A	2.13×10^{-2}	100
SORM [2]	N/A	8.27×10^{-2}	132
IS [2]`	5.34×10^{-3}	6.15×10^{-3}	5100
The proposed method	9.84×10^{-2}	8.26×10^{-3}	22498

By setting K and V_L as the x and y axis and starting with $N_{ss} = 10^4$, Fig 8(a) illustrates the estimate of Pr(F) through *SS* with information $s_m(l_4)$, where S_1, S_2 and S_3 denote samples located in the three intermediate subsets. However, the initial set of N_{ss} is insufficient so that COV_{P_F} is estimated as large as 0.0657. Therefore, N_{ss} is adaptively increased to 2×10^4 and Pr(F) is finally estimated as 8.26×10^{-3} with COV_{P_F} equal to 0.0449. Fig 8(b) showcases the estimate of Pr(Z) based on the augmented limit state function $h_1(P, \boldsymbol{X}) = 0$, where the darker dots denote the accepted samples, S_{acc}, and the brighter ones represent the rejected samples, S_{rej}. In this step, Pr(Z) is estimated as 8.87×10^{-6} and c_1 is equal to 1.78×10^{-5}. Fig 8(c) showcases the estimate of Pr(Z|F) through the augmented limit state function $h_2(P, \boldsymbol{X}) = 0$, where the darker dots denote the accepted samples, S_{acc}^{last}, and the brighter ones represent the rejected samples, S_{rej}^{last}. One should note that $\left[S_{acc}^{last}, S_{rej}^{last}\right] \in S^{last}$, where S^{last} is the last sample set in Fig 8(a). In this step, the two terms are estimated as $\mathrm{Pr(Z|F)} = 1.05 \times 10^{-5}$ and $c_2 = 2.02 \times 10^{-5}$. Fig 8(d) exhibits the conventional procedure represented in Eq. (10) that relies on the computational scheme $P(F \cap Z)/P(Z)$ with the joint limit state function $J(P, \boldsymbol{X}) = \max[g(\boldsymbol{X}), h_1(P, \boldsymbol{X})]$. The conventional approach results in the simulation data with $P(F \cap Z) = 2.92 \times 10^{-3}$, which is significantly smaller than Pr(F). This implies that more evaluations of $g(\boldsymbol{X})$ should be conducted compared to the proposed approach, which further demonstrate the computational efficiency of the proposed reliability updating approach. Moreover, the size of samples in each subset is adaptively increased to guarantee the sufficiently consistency of Pr(Z|F), which facilitates the computational robustness of *SOI* and the exploration of l^*. To further demonstrate the computational efficiency of the proposed method, the computational performance via *MCS*, *FORM*, *SORM* and *IS* is summarized in Table 5. All implementations are conducted through UQLab package in MATLAB with default settings. By treating the result of *MCS* as the benchmark, all the four methods are able to estimate the prior failure probability. However, the posterior failure probability cannot be estimated through *FORM* or *SORM.* It is shown that the estimated posterior failure probability is more accurate than the importance sampling method used in literature [2]. According to Table 5, $\Pr(F|Z)$ is estimated as 9.76×10^{-2} through *MCS.* Comparing this with estimates of 9.84×10^{-2} and

5.34×10^{-3} using the proposed method and *IS*, respectively, it is evident that the proposed method overperforms *IS*.

SOI can be understood as a metric that considers the change of $\Pr(F)$ due to all possible information that could be acquired at a specific location. Moreover, it can be concluded from the simulation results that *SOI* is not solely determined by $\Pr(F)$; it is determined by both $\Pr(F)$ and $\Pr(F|Z)$. A monitoring system placed at a location with a large *SOI* improves the accuracy of $\Pr(F)$ estimation significantly. Therefore, acquiring information at that location is necessary if one needs to have a more accurate $\Pr(F|Z)$. Estimation of *SOI* in turn requires investigating an integral and evaluation of the limit state function multiple times.

Fig 9 showcases the relation of d_{up} versus Z at locations l_1 and l_3. One can infer that $d_{up}(l_1)$ reaches 0 when $Z = 8.7$, which indicates that the updated reliability deviates substantially from the prior one when the information is involved because it is beyond the expectation of prior knowledge. However, $d_{up}(l_3)$ almost increases linearly over the interval $[Z_{lob}, Z_{upb}]$. The next subsection elaborates the procedures of exploring l^* over different regions of observation and excavation stage.

It can be observed from Table 4 that d_{up} with $s_m(l_2) = 10$ is apparently the largest one, which also indicates the largest change for the update of reliability when information $s_m(l_2)$ is available. As the location transits from l_2 to l_4 and the settlement information keeps unchanged, r_{up} decreases significantly from 0.2099 to 0.0129. This is in attributed to the location of l_4 is further from both the tunnel axis and building facade compared to location l_2. Therefore, r_{up} can be an efficient metric for quantifying the contribution of the change of updated reliability for different source of information. Moreover, the significance of information at some locations cannot be interpreted by intuition, therefore, the metric d_{up} can reflect this effect. For example, d_{up} is estimated as 0.1219 at location l_1, which is less significant than location l_3 because the settlement close to TBM becomes smaller, thereby it has less influence on d_{up}. This point, however, does not indicate that l_1 is less valuable than l_3 in terms of *SOI* because Table 6 presents that *SOI* for location l_1 is greater than location l_3 . Nevertheless, l_2 is deemed to be the most significant location among three selected points according to *SOI*.

Table 6. *SOI*s at locations l_1, l_2, l_3 and l_4

Number	Location	*SOI*
1	l_1	0.1219
2	l_2	0.2099
3	l_3	0.0801
4	l_4	0.0129

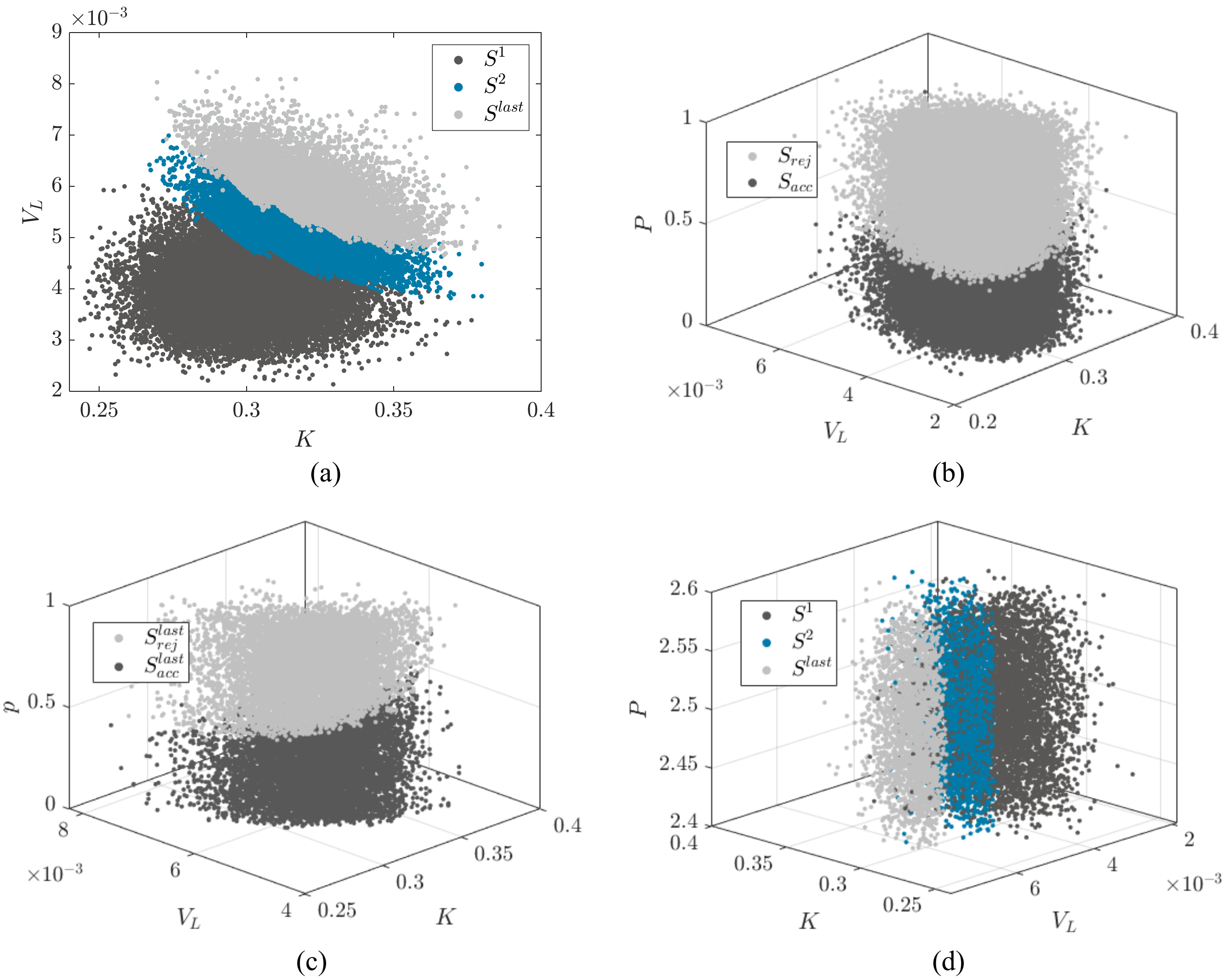
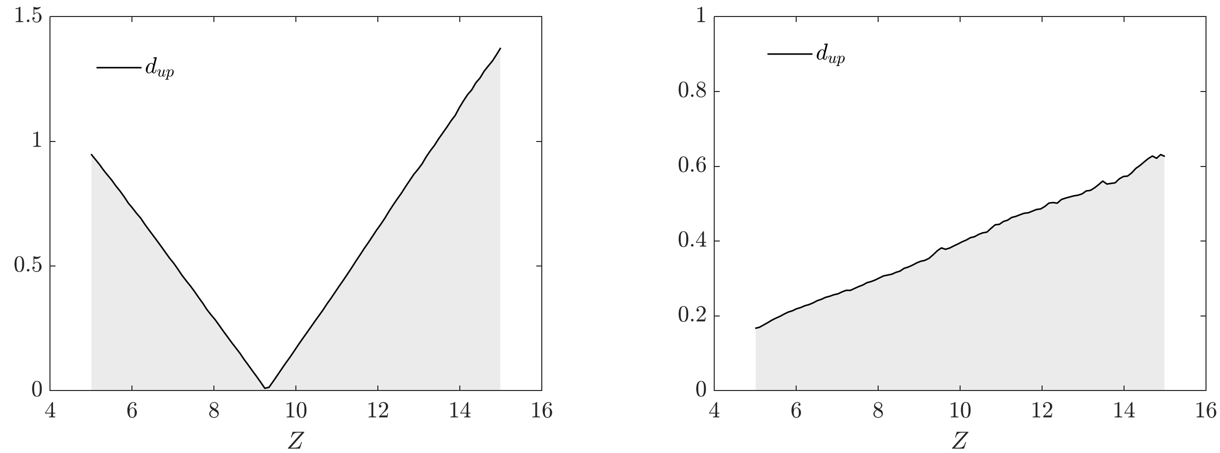


Fig 8. Use information $s_m(l_4)$ via *SS* to estimate (a) Pr(F) based on $g(\boldsymbol{X})$; (b) Pr(Z) based on $h_1(P, \boldsymbol{X})$ at the location; (c) Pr(Z|F) based on $h_2(P, \boldsymbol{X}')$, where $\boldsymbol{X}' \in \Omega_f$ and (d) $\Pr(Z \cap F)$ based on $J(P, \boldsymbol{X}) = \max[g(\boldsymbol{X}), h_1(P, \boldsymbol{X})]$.



(a) (b)

Fig 9. d_{up} versus Z at the location of (a) l_1 and (b) l_3

4.3 Optimal settlement monitoring location

In this subsection, l^* is explored over the whole region of observations, Ω_{ob}, globally, where the corresponding steps are depicted in Fig10. In Fig 10(a), the tunnel is plotted with gray region, the light red square showcases where the building façade locates, C_s represents the contour of tunneling excavation-induced settlement and Ω_{ob} is represented by light blue square. Initially, 81 equally distributed training samples (locations), l_{in}, represented with black star dots are ready to training a surrogate model for $\widehat{SOI}(l)$ over $\Omega_{\text{ob}} = [x_{lim}^1, x_{lim}^2; y_{lim}^1, y_{lim}^2;]$, where $[x_{lim}^1, x_{lim}^2; y_{lim}^1, y_{lim}^2;]$ denotes the x and y limits of axis of Ω_{ob}. For example, Ω_{ob} is parameterized by $[10{,}30; 10{,}30;]$ in the Fig 10(a) and the true responses of the 81 discretized training samples are estimated, which facilitates the initial construction of $\widehat{SOI}(l)$. Subsequently, extra training samples are adaptively enriched through the EI active learning function and terminates until the stopping criterion is satisfied. For this case, 7 extra training samples are finally added and the surface plot of $\widehat{SOI}(l)$ based on the Kriging surrogate model is represented in Fig 10(b). Fig 10(c) showcases the x-y view of $\widehat{SOI}(l)$, where $\widehat{SOI}$ increases as the regions transits from blue to yellow. Moreover, the initial optimal location, l_{in}^*, among l_{in} is identified as $(15{,}27.5{,}0)$ and finally transits to the final optimal location, l^*, where $l^* = (14.9{,}24.8{,}0)$ with *SOI* estimated as 0.2356, as highlighted in Fig 10(d).

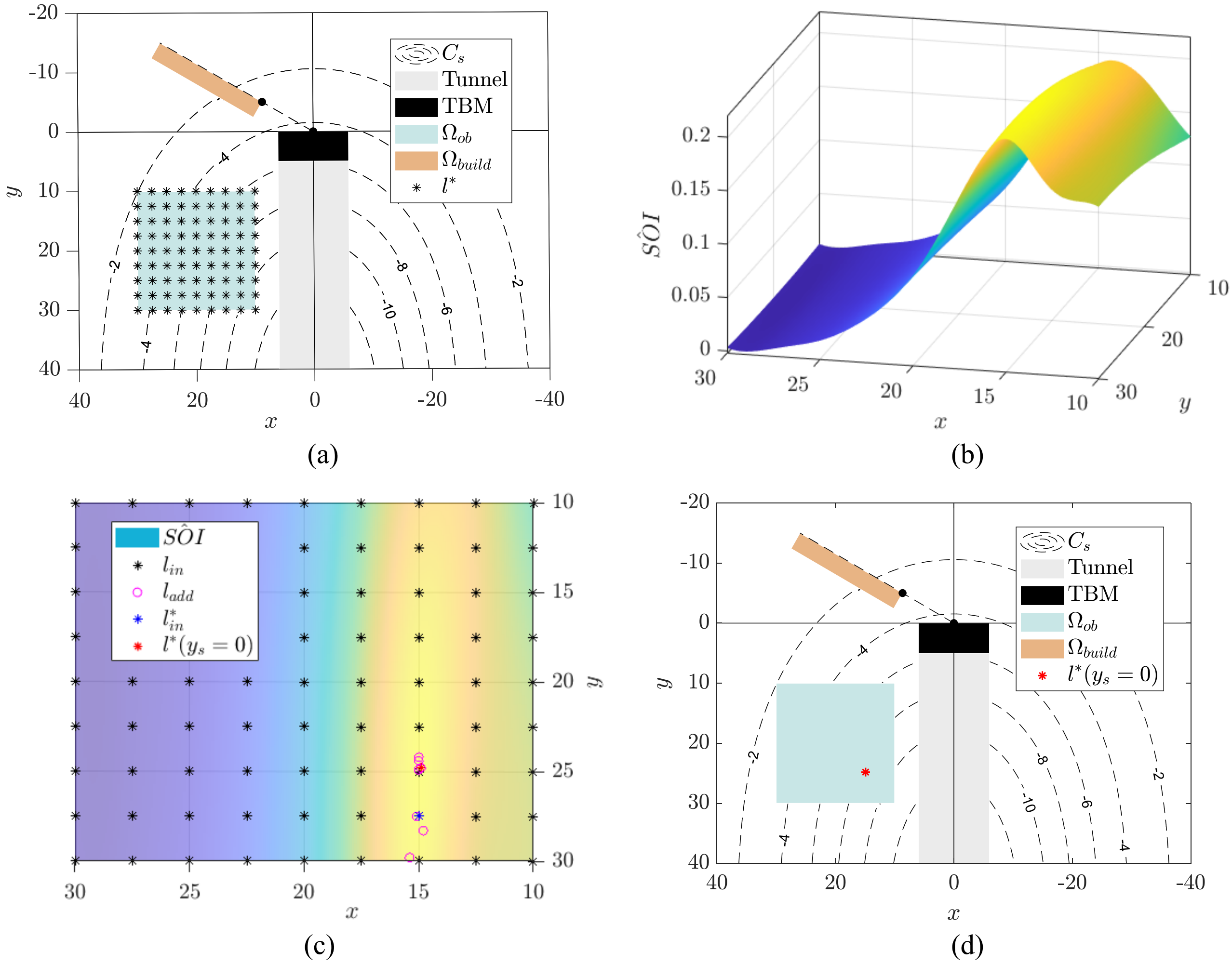
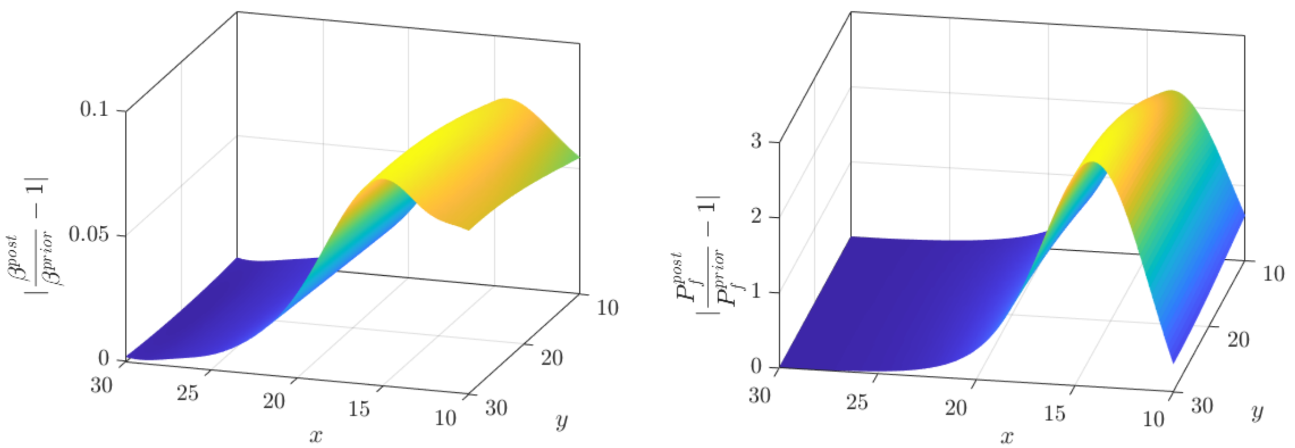


Fig 10. Procedure of exploring l^* through surrogate-based optimization with (a) the initial training samples located in $\Omega_{\text{ob}} = [10{,}30; 10{,}30;]$, where $y_s = 0$; (b) surface plot of $\widehat{SOI}$; (c) the addition of training samples through active learning and the identified l^* and (d) an overview of l^* in the process of tunneling excavation.



(a) (b)

Fig 11. Surface plots based on metrics defined with (a) $\left|\frac{\beta_{post}}{\beta_{post}} - 1\right|$ and (b) $\left|\frac{P_f^{post}}{P_f^{prior}} - 1\right|$

One should note that *SOI* can take different forms; it can be also defined as $\left|\frac{\beta_{post}}{\beta_{prior}} - 1\right|$, $\left|\frac{P_f^{post}}{P_f^{prior}} - 1\right|$, among other possibilities. Fig 11 showcases the surface plots of $\left|\frac{\beta_{post}}{\beta_{prior}} - 1\right|$ and $\left|\frac{P_f^{post}}{P_f^{prior}} - 1\right|$ versus x and y. It can be observed from Fig 11 and Fig 10(b) that all these three metrics can efficiently quantify the sensitivity of the information. By comparing Fig 11 with Fig 10(b), the metric based on $\left|\frac{\beta_{post}}{\beta_{prior}} - 1\right|$ indicates the least variation, while $\left|\frac{P_f^{post}}{P_f^{prior}} - 1\right|$ indicates the highest variation. Moreover, due to the smooth property of the metric based on $|\beta_{post} - \beta_{prior}|$, it is more efficient for integration with surrogate models. Moreover, the metric based on $|\beta_{post} - \beta_{prior}|$ is more intuitive for understanding and communication of the concept of *SOI*. According to this experimental study, while the shape of *SOI* is affected the functional form, the optimal monitoring location remains the same.

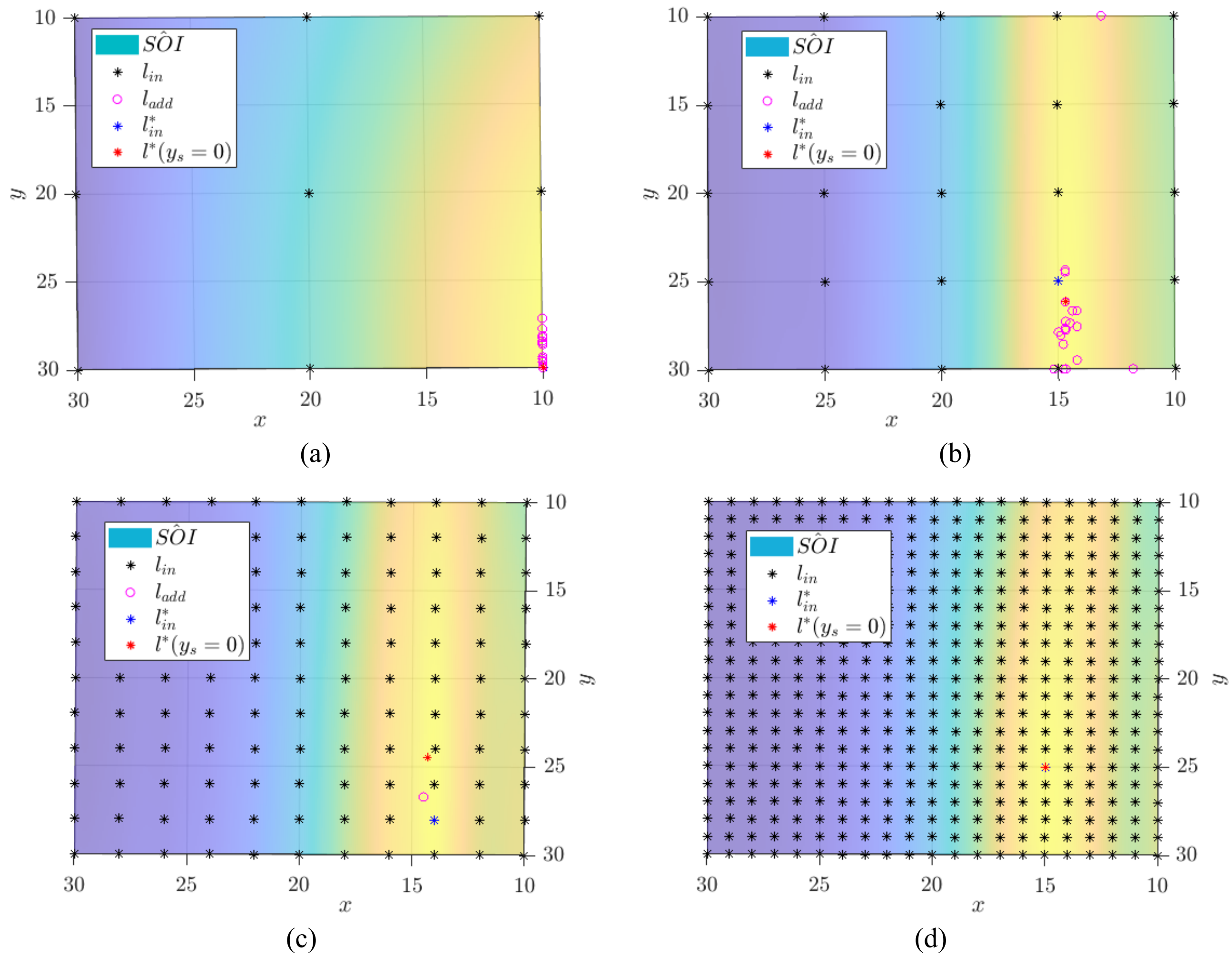


Fig 12. Investigation of simulation results with (a) 9, (b) 25, (c) 121 and (d) 441 initial training samples

Moreover, the definition of initial samples can also affect the performance of the proposed framework. If the initial training sample size is insufficient, the problem can become ill-conditioned. To explore such an effect, the number of initial training samples is increased from 9, to 25, 121, and 441. The corresponding simulation results are illustrated in Fig 12. According to Fig 12 (a), all the added training samples and the identified optimal sample are clustered at the right bottom corner. This can be attributed to the scant initial training samples that lead to large uncertainties in the active learning process of the surrogate model. The computational performance gets rid of ill condition as the number of initial training samples reaches 25 as seen in Fig 12(b). In this figure, 21 additional samples are included. As the number of initial training samples reaches 121, the number of additional training samples decreases, as seen in Fig 12(c). On the other hand, it is computationally demanding to prepare a large number of initial training samples. This is seen in Fig 12(d), where no extra training sample is needed to refine the Kriging surrogate model because the 441 initial samples are sufficient for estimating the value of *SOI*.

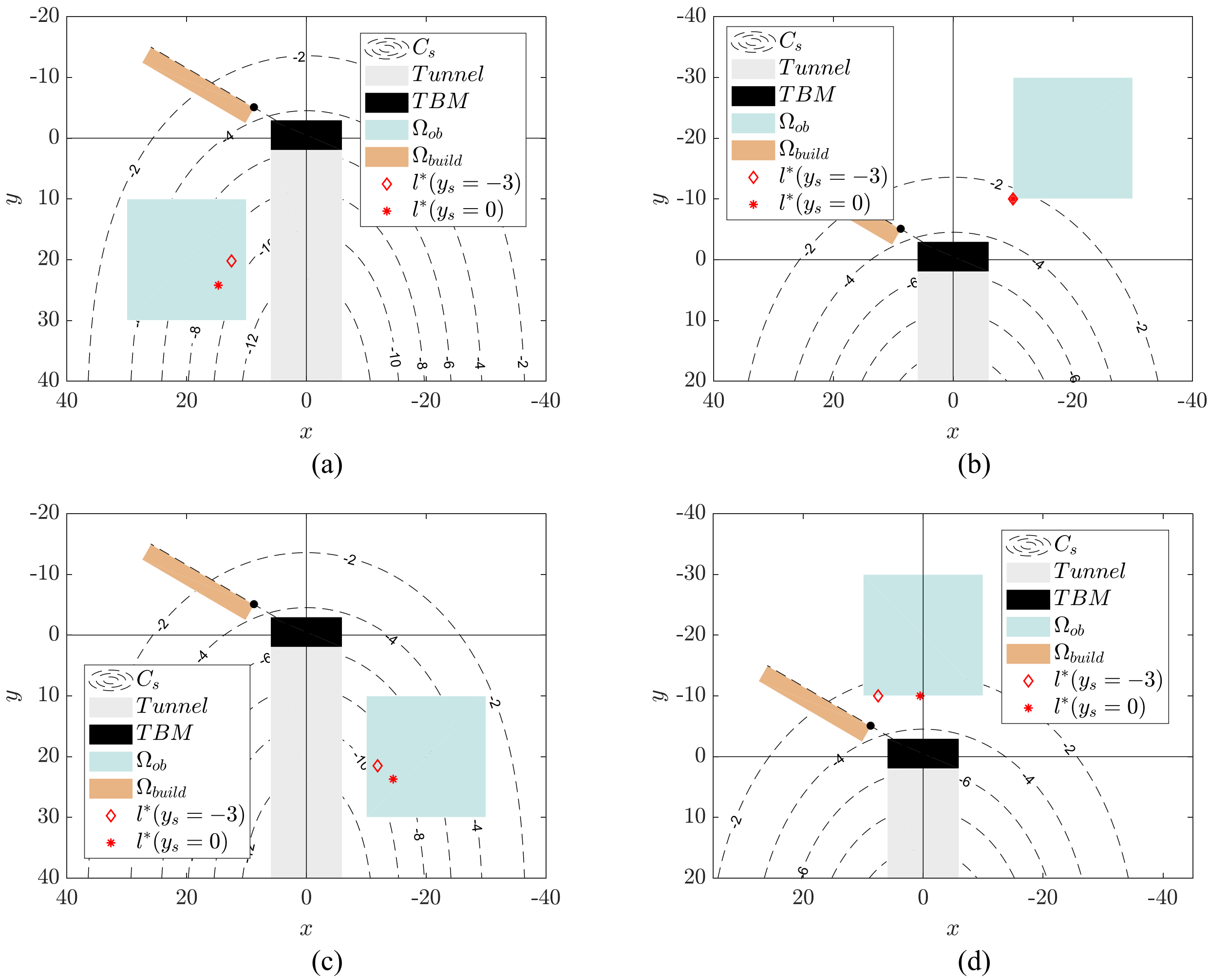


Fig 13. Illustration of identified l^*s with (a) $\Omega_{ob} = \Omega_A$, $y_s = 0$ and -3; (b) $\Omega_{ob} = \Omega_B$, $y_s = 0$ and -3 (c) $\Omega_{ob} = \Omega_C$, $y_s = 0$ and -3 and (d) $\Omega_{ob} = \Omega_D$, $y_s = 0$ and -3.

As the excavation of tunnel proceeds and Ω_{ob} changes, l^* changes accordingly. Fig 13 showcases four scenarios, $\Omega_A, \Omega_B, \Omega_C$ and Ω_d of Ω_{ob} for tunnel excavation along with the tunnel façade y_s changing from 0 to -3, of which the simulation results are summarized in Table 7. According to Fig 13, l^* changes from (14.9,24.8,0) to (12.5,20.2,0) while the excavation proceeds to $y_s = -3$ and Ω_{ob} keeps unchanged which leads to *SOI* increases from 0.2356 to 0.6042. This phenomenon can be interpreted by the settlement caused by excavation of tunnel dominates the change of the updated reliability. Moreover, Fig 13(b) represents that l^* maintains unchanged even though y_s changes from 0 to -3 along with the increase of *SOI* from 0.0327 to 0.0798. This is because the point at the very bottom left over $\Omega_{\mathrm{ob}} = \Omega_B$ is the most valuable point. As Ω_{ob} changes from Ω_A to Ω_B, $l^*(y_s = 0)$ and $l^*(y_s = -3)$ are estimated as $(-14.5,23.7,0)$ and $(-11.9,21.5,0)$ with *SOI* equals to 0.2391 and 0.7329, respectively. It can be inferred from the comparison between Fig 13(a) and (c) that above two optimal locations are closely symmetric to the two identified optimal locations when $\Omega_{\mathrm{ob}} = [10,30; 10,30;]$ along the y-axis. This can be explained by the symmetric characteristics of Gaussian settlement defined in Eq. (44). In Fig 13(d), l^* changes from $(0.5, -10,0)$ to $(7.5, -10,0)$ with the corresponding *SOI* estimated as 0.0824 and 0.0261, when $\Omega_{\mathrm{ob}} = \Omega_D$. The tunnel façade at $y_s = 0$ causes a slight deviation of l^* close to the building side when $\Omega_{\mathrm{ob}} = \Omega_D$ and this effect of deviation strengthens when $y_s = -3$.

Table 7. Identification of l^* with corresponding *SOI* based on different combinations of Ω_{ob} and y_s. 20 simulations are conducted to eliminate the uncertainty of the method, where $\bar{l}^*$(m) and $\overline{\mathrm{SOI}}$ (-) denote the mean of l^*(m) and SOI (-)

Region	Parameters (m)	y_s(m)	$\bar{l}^*$(m)	$\overline{SOI}$ (-)	COV_{SOI}
Ω_A	$[10,30; 10,30;]$	0	$(14.52,24.17,0)$	0.2389	0.112
		-3	$(12.88,19.25,0)$	0.5326	0.127
Ω_B	$[-30,-10; -30,-10;]$	0	$(-10,-10,0)$	0.0392	0.027
		-3	$(-10,-10,0)$	0.0827	0.022
Ω_C	$[-30,-10; 10,30;]$	0	$(-14.47,23.15,0)$	0.2622	0.134
		-3	$(-11.77,21.29,0)$	0.6480	0.108
Ω_D	$[-10,10; -30,-10;]$	0	$(0.52,-10.0,0)$	0.0927	0.035
		-3	$(7.58,-10.0,0)$	0.0272	0.024

Therefore, the procedures represented above showcase a systematic approach for localizing the optimal monitoring topology for the risk assessment and tracking of a tunneling-induced structural failure. Instead of focusing on the location where the largest deformation happens, this paper sheds light on utilizing probabilistic tools to account for the uncertainties involved. It can be further investigated to explore the uncertainty of the soil properties via random fields modelling and consider the paradigm that can handle multiple building over the tunneling contour. It is expected that this work can be leveraged to improve the efficiency for decision-making of structural health/risk monitoring of geo-structures.

5. Conclusions and discussions

This paper proposes a computational framework based on a novel metric called *SOI* (sensitivity of information) to determine the optimal monitoring location for risk tracking of infrastructure systems. Generally, the major contributions of this paper can be summarized as:

- A novel metric called *SOI* (Sensitivity of Information) is proposed to quantify the change in updated and prior reliability of a structure or a system at a specific location with possible new information that can be acquired through a monitoring system placed at another location. In terms of failure risk, *SOI* seeks for monitoring locations that offer the highest sensitivity of reliability update to new information. Monitoring at locations with high *SOI* can significantly improve the accuracy of updated reliability for the structure or infrastructure system of interest. Compared to *VOI*, the calculation of *SOI* is more straightforward and is purely grounded in reliability updating theory without the need to establish possible actions and costs.
- Determining the proposed *SOI* is computationally very challenging. Therefore, a novel computational framework is proposed to facilitate efficient computation of *SOI* and to explore the optimal monitoring location for infrastructure systems. This is achieved through integration of adaptively trained surrogate models based on active learning concepts in the computation of *SOI* as well as in solving the optimization model that is formulated in search of the location with maximum *SOI*.

To explore the performance of the proposed computational framework, a practical case that investigates the risk posed by tunneling-induced settlement to building damage is studied. Simulation results showcase that the optimal settlement monitoring grounded in reliability updating theory can be accurately determined. In of the context of risk analysis, this proposed framework can also be applied to other infrastructure systems whenever the identification of optimal monitoring location is needed. For example, it can be modified for optimal sensor placement for fire warning systems or for structural health monitoring application. A challenge in the application of the proposed framework is the associated computational cost of evaluating the limit state function. While this problem is addressed in this paper through integration of adaptive Kriging, for complex performance functions, e.g., high-dimensional or non-smooth limit state functions [62] additional research may be needed. Moreover, *SOI* in this paper is defined as $|\beta_{post} - \beta_{prior}|$. Other forms of *SOI* can be explored in depth in future studies.

Acknowledgements

This research was supported in part by 'Shuimu Tsinghua Scholar' Plan by Tsinghua University through award 2020SM006, China Postdoctoral Science Foundation through award YJ20210126, the National Science Foundation of China (NSFC) through awards 51778337, 51890901 and 52008155 and the U.S. National Science Foundation (NSF) through award CMMI- 2000156. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of these supports.

Reference

[1] X. Wang, R. K. Mazumder, B. Salarieh, A. M. Salman, A. Shafieezadeh, and Y. Li, "Machine Learning for Risk and Resilience Assessment in Structural Engineering: Progress and Future Trends," *J. Struct. Eng.*, vol. 148, no. 8, p. 03122003, Aug. 2022, doi: 10.1061/(ASCE)ST.1943-541X.0003392.

[2] D. Straub, "Reliability updating with equality information," *Probabilistic Eng. Mech.*, vol. 26, no. 2, pp. 254–258, Apr. 2011, doi: 10.1016/j.probengmech.2010.08.003.

[3] S. Gollwitzer, B. Kirchgäßner, R. Fischer, and R. Rackwitz, "PERMAS-RA/STRUREL system of programs for probabilistic reliability analysis," *Struct. Saf.*, vol. 28, no. 1, pp. 108–129, Jan. 2006, doi: 10.1016/j.strusafe.2005.03.008.

[4] D. Straub, "Stochastic Modeling of Deterioration Processes through Dynamic Bayesian Networks," *J. Eng. Mech.*, vol. 135, no. 10, pp. 1089–1099, Oct. 2009, doi: 10.1061/(ASCE)EM.1943-7889.0000024.

[5] Straub Daniel and Der Kiureghian Armen, "Bayesian Network Enhanced with Structural Reliability Methods: Methodology," *J. Eng. Mech.*, vol. 136, no. 10, Art. no. 10, Oct. 2010, doi: 10.1061/(ASCE)EM.1943-7889.0000173.

[6] Straub Daniel and Der Kiureghian Armen, "Bayesian Network Enhanced with Structural Reliability Methods: Application," *J. Eng. Mech.*, vol. 136, no. 10, Art. no. 10, Oct. 2010, doi: 10.1061/(ASCE)EM.1943-7889.0000170.

[7] J. Luque and D. Straub, "Reliability analysis and updating of deteriorating systems with dynamic Bayesian networks," *Struct. Saf.*, vol. 62, pp. 34–46, Sep. 2016, doi: 10.1016/j.strusafe.2016.03.004.

[8] I. Papaioannou and D. Straub, “Reliability updating in geotechnical engineering including spatial variability of soil,” *Comput. Geotech.*, vol. 42, pp. 44–51, May 2012, doi: 10.1016/j.compgeo.2011.12.004.

[9] Z. Wang and A. Shafieezadeh, “Metamodel-based subset simulation adaptable to target computational capacities: the case for high-dimensional and rare event reliability analysis,” *Struct. Multidiscip. Optim.*, Apr. 2021, doi: 10.1007/s00158-021-02864-9.

[10] Y. Zhao and Z. Wang, “Subset simulation with adaptable intermediate failure probability for robust reliability analysis: an unsupervised learning-based approach,” *Struct. Multidiscip. Optim.*, vol. 65, no. 6, p. 172, Jun. 2022, doi: 10.1007/s00158-022-03260-7.

[11] Z. Wang and A. Shafieezadeh, “Real time High-fidelity Reliability updating with equality information using adaptive Kriging,” *Reliab. Eng. Syst. Saf.*, 2019.

[12] C. Zhang, Z. Wang, and A. Shafieezadeh, “Error Quantification and Control for Adaptive Kriging-Based Reliability Updating with Equality Information,” *Reliab. Eng. Syst. Saf.*, vol. 207, p. 107323, Mar. 2021, doi: 10.1016/j.ress.2020.107323.

[13] I. Papaioannou and D. Straub, “Learning soil parameters and updating geotechnical reliability estimates under spatial variability – theory and application to shallow foundations,” *Georisk Assess. Manag. Risk Eng. Syst. Geohazards*, vol. 11, no. 1, pp. 116–128, Jan. 2017, doi: 10.1080/17499518.2016.1250280.

[14] A. P. van den Eijnden, T. Schweckendiek, and M. A. Hicks, “Metamodelling for geotechnical reliability analysis with noisy and incomplete models,” *Georisk Assess. Manag. Risk Eng. Syst. Geohazards*, vol. 0, no. 0, pp. 1–18, Jul. 2021, doi: 10.1080/17499518.2021.1952611.

[15] C. Camós, O. Špačková, D. Straub, and C. Molins, “Probabilistic approach to assessing and monitoring settlements caused by tunneling,” *Tunn. Undergr. Space Technol.*, vol. 51, pp. 313–325, Jan. 2016, doi: 10.1016/j.tust.2015.10.041.

[16] T. Schweckendiek and A. C. W. M. Vrouwenvelder, “Reliability updating and decision analysis for head monitoring of levees,” *Georisk*, vol. 7, no. 2, Art. no. 2, 2013, doi: 10.1080/17499518.2013.791034.

[17] W. J. Klerk, T. Schweckendiek, F. den Heijer, and M. Kok, “Value of Information of Structural Health Monitoring in Asset Management of Flood Defences,” *Infrastructures*, vol. 4, no. 3, 2019, doi: 10.3390/infrastructures4030056.

[18] Q. Huang, P. Gardoni, and S. Hurlebaus, “Adaptive Reliability Analysis of Reinforced Concrete Bridges Subject to Seismic Loading Using Nondestructive Testing,” *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part Civ. Eng.*, vol. 1, no. 4, p. 04015014, Dec. 2015, doi: 10.1061/AJRUA6.0000835.

[19] Y. Jin, G. Biscontin, and P. Gardoni, “Adaptive prediction of wall movement during excavation using Bayesian inference,” *Comput. Geotech.*, vol. 137, p. 104249, Sep. 2021, doi: 10.1016/j.compgeo.2021.104249.

[20] Y. Jin, G. Biscontin, and P. Gardoni, “A Bayesian definition of ‘most probable’ parameters,” *Geotech. Res.*, vol. 5, no. 3, pp. 130–142, Sep. 2018, doi: 10.1680/jgere.18.00027.

[21] J. Hackl and J. Kohler, “Reliability assessment of deteriorating reinforced concrete structures by representing the coupled effect of corrosion initiation and progression by Bayesian networks,” *Struct. Saf.*, vol. 62, pp. 12–23, Sep. 2016, doi: 10.1016/j.strusafe.2016.05.005.

[22] S.-H. Jiang, J. Huang, X.-H. Qi, and C.-B. Zhou, “Efficient probabilistic back analysis of spatially varying soil parameters for slope reliability assessment,” *Eng. Geol.*, vol. 271, p. 105597, Jun. 2020, doi: 10.1016/j.enggeo.2020.105597.

[23] S.-H. Jiang, I. Papaioannou, and D. Straub, “Bayesian updating of slope reliability in spatially variable soils with in-situ measurements,” *Eng. Geol.*, vol. 239, pp. 310–320, May 2018, doi: 10.1016/j.enggeo.2018.03.021.

[24] D. Y. Yang and D. M. Frangopol, “Probabilistic optimization framework for inspection/repair planning of fatigue-critical details using dynamic Bayesian networks,” *Comput. Struct.*, vol. 198, pp. 40–50, Mar. 2018, doi: 10.1016/j.compstruc.2018.01.006.

[25] S.-H. Lee and J. Song, “Bayesian-network-based system identification of spatial distribution of structural parameters,” *Eng. Struct.*, vol. 127, pp. 260–277, Nov. 2016, doi: 10.1016/j.engstruct.2016.08.029.

[26] D. J. Jerez, H. A. Jensen, and M. Beer, “An effective implementation of reliability methods for Bayesian model updating of structural dynamic models with multiple uncertain parameters,” *Reliab. Eng. Syst. Saf.*, p. 108634, Jun. 2022, doi: 10.1016/j.ress.2022.108634.

[27] H. A. Jensen and D. J. Jerez, “A Bayesian model updating approach for detection-related problems in water distribution networks,” *Reliab. Eng. Syst. Saf.*, vol. 185, pp. 100–112, May 2019, doi: 10.1016/j.ress.2018.12.014.

[28] S. Shuto and T. Amemiya, “Sequential Bayesian inference for Weibull distribution parameters with initial hyperparameter optimization for system reliability estimation,” *Reliab. Eng. Syst. Saf.*, vol. 224, p. 108516, Aug. 2022, doi: 10.1016/j.ress.2022.108516.

[29] Y. Zhao, W. Gao, and C. Smidts, “Sequential Bayesian inference of transition rates in the hidden Markov model for multi-state system degradation,” *Reliab. Eng. Syst. Saf.*, vol. 214, p. 107662, Oct. 2021, doi: 10.1016/j.ress.2021.107662.

[30] K. Kim, G. Lee, K. Park, S. Park, and W. B. Lee, “Adaptive approach for estimation of pipeline corrosion defects via Bayesian inference,” *Reliab. Eng. Syst. Saf.*, vol. 216, p. 107998, Dec. 2021, doi: 10.1016/j.ress.2021.107998.

[31] Z. Pang, X. Si, C. Hu, D. Du, and H. Pei, “A Bayesian Inference for Remaining Useful Life Estimation by Fusing Accelerated Degradation Data and Condition Monitoring Data,” *Reliab. Eng. Syst. Saf.*, vol. 208, p. 107341, Apr. 2021, doi: 10.1016/j.ress.2020.107341.

[32] X. Xiao, Q. Li, and Z. Wang, “A novel adaptive importance sampling algorithm for Bayesian model updating,” *Struct. Saf.*, vol. 97, p. 102230, Jul. 2022, doi: 10.1016/j.strusafe.2022.102230.

[33] S.-H. Jiang, I. Papaioannou, and D. Straub, “Optimization of Site-Exploration Programs for Slope-Reliability Assessment,” *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part Civ. Eng.*, vol. 6, no. 1, p. 04020004, Mar. 2020, doi: 10.1061/AJRUA6.0001042.

[34] J. Z. Hu, J. Zhang, H. W. Huang, and J. G. Zheng, “Value of information analysis of site investigation program for slope design,” *Comput. Geotech.*, vol. 131, p. 103938, Mar. 2021, doi: 10.1016/j.compgeo.2020.103938.

[35] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2016. Accessed: May 12, 2017.

[36] G. Fishman, *Monte Carlo: concepts, algorithms, and applications*. Springer Science & Business Media, 2013. Accessed: May 12, 2017.

[37] R. Rackwitz and B. Flessler, “Structural reliability under combined random load sequences,” *Comput. Struct.*, vol. 9, no. 5, pp. 489–494, Nov. 1978, doi: 10.1016/0045-7949(78)90046-9.

[38] Kiureghian Armen Der and Stefano Mario De, “Efficient Algorithm for Second-Order Reliability Analysis,” *J. Eng. Mech.*, vol. 117, no. 12, pp. 2904–2923, Dec. 1991, doi: 10.1061/(ASCE)0733-9399(1991)117:12(2904).

[39] S. K. Au and J. L. Beck, “A new adaptive importance sampling scheme for reliability calculations,” *Struct. Saf.*, vol. 21, no. 2, pp. 135–158, 1999.

[40] M. Hohenbichler and R. Rackwitz, “Improvement of second-order reliability estimates by importance sampling,” *J. Eng. Mech.*, vol. 114, no. 12, Art. no. 12, 1988.

[41] S.-K. Au and J. L. Beck, “Estimation of small failure probabilities in high dimensions by subset simulation,” *Probabilistic Eng. Mech.*, vol. 16, no. 4, Art. no. 4, Oct. 2001, doi: 10.1016/S0266-8920(01)00019-4.

[42] S. K. Au and J. L. Beck, “Subset simulation and its application to seismic risk based on dynamic analysis,” *J. Eng. Mech.*, vol. 129, no. 8, Art. no. 8, 2003.

[43] I. Papaioannou, W. Betz, K. Zwirglmaier, and D. Straub, “MCMC algorithms for Subset Simulation,” *Probabilistic Eng. Mech.*, vol. 41, pp. 89–103, Jul. 2015, doi: 10.1016/j.probengmech.2015.06.006.

[44] B. Echard, N. Gayton, and M. Lemaire, “AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation,” *Struct. Saf.*, vol. 33, no. 2, Art. no. 2, 2011.

[45] Z. Wang and A. Shafieezadeh, “REAK: Reliability analysis through Error rate-based Adaptive Kriging,” *Reliab. Eng. Syst. Saf.*, vol. 182, pp. 33–45, Feb. 2019, doi: 10.1016/j.ress.2018.10.004.

[46] Z. Wang and A. Shafieezadeh, “ESC: an efficient error-based stopping criterion for kriging-based reliability analysis methods,” *Struct. Multidiscip. Optim.*, vol. 59, no. 5, Art. no. 5, May 2019, doi: 10.1007/s00158-018-2150-9.

[47] Z. Wang and A. Shafieezadeh, “On confidence intervals for failure probability estimates in Kriging-based reliability analysis,” *Reliab. Eng. Syst. Saf.*, vol. 196, p. 106758, Apr. 2020, doi: 10.1016/j.ress.2019.106758.

[48] S. Marelli, R. Schöbi, and B. Sudret, “UQLab User Manual – Structural Reliability (Rare Events Estimation),” p. 60.

[49] "UQLab input manual," *UQLab, the Framework for Uncertainty Quantification*, May 13, 2017. http://www.uqlab.com/userguideinput (accessed May 14, 2017).

[50] "UQLab Kriging (Gaussian process modelling) manual," *UQLab, the Framework for Uncertainty Quantification*. http://www.uqlab.com/userguidekriging (accessed May 13, 2017).

[51] I. Kaymaz, "Application of kriging method to structural reliability problems," *Struct. Saf.*, vol. 27, no. 2, pp. 133–151, 2005.

[52] J. Wang, Z. Sun, Q. Yang, and R. Li, "Two accuracy measures of the Kriging model for structural reliability analysis," *Reliab. Eng. Syst. Saf.*, vol. 167, pp. 494–505, Nov. 2017, doi: 10.1016/j.ress.2017.06.028.

[53] Z. Wen, H. Pei, H. Liu, and Z. Yue, "A Sequential Kriging reliability analysis method with characteristics of adaptive sampling regions and parallelizability," *Reliab. Eng. Syst. Saf.*, vol. 153, pp. 170–179, 2016.

[54] C. Camós and C. Molins, "3D analytical prediction of building damage due to ground subsidence produced by tunneling," *Tunn. Undergr. Space Technol.*, vol. 50, pp. 424–437, Aug. 2015, doi: 10.1016/j.tust.2015.08.012.

[55] P. B. Attewell and J. P. Woodman, "PREDICTING THE DYNAMICS OF GROUND SETTLEMENT AND ITS DERIVITIVES CAUSED BY TUNNELLING IN SOIL," *GROUND Eng.*, vol. 15, no. 8, Nov. 1982, Accessed: Dec. 28, 2021. [Online]. Available: https://trid.trb.org/view/188316

[56] M. P. O'reilly and B. M. New, "SETTLEMENTS ABOVE TUNNELS IN THE UNITED KINGDOM - THEIR MAGNITUDE AND PREDICTION," *Publ. Inst. Min. Metall.*, Art. no. Monograph, 1982, Accessed: Dec. 28, 2021. [Online]. Available: https://trid.trb.org/view/186714

[57] P. B. Attewell, J. Yeates, and A. R. Selby, "SOIL MOVEMENTS INDUCED BY TUNNELLING AND THEIR EFFECTS ON PIPELINES AND STRUCTURES," 1986, Accessed: Dec. 28, 2021. [Online]. Available: https://trid.trb.org/view/247405

[58] V. Fargnoli, D. Boldini, and A. Amorosi, "TBM tunnelling-induced settlements in coarse-grained soils: The case of the new Milan underground line 5," *Tunn. Undergr. Space Technol.*, vol. 38, pp. 336–347, Sep. 2013, doi: 10.1016/j.tust.2013.07.015.

[59] T. Sugiyama *et al.*, “Observations of Ground Movements during Tunnel Construction by Slurry Shield Method at the Docklands Light Railway Lewisham Extension—East London,” *Soils Found.*, vol. 39, no. 3, pp. 99–112, Jun. 1999, doi: 10.3208/sandf.39.3_99.

[60] C. Camós, C. Molins, and O. Arnau, “Case Study of Damage on Masonry Buildings Produced by Tunneling Induced Settlements,” *Int. J. Archit. Herit.*, vol. 8, no. 4, pp. 602–625, Jul. 2014, doi: 10.1080/15583058.2012.704479.

[61] J. B. Burland and C. P. Wroth, “SETTLEMENT OF BUILDINGS AND ASSOCIATED DAMAGE,” Art. no. CP 33/75, Apr. 1975, Accessed: Dec. 29, 2021. [Online]. Available: https://trid.trb.org/view/29202

[62] Y. Pang and X. Wang, “Cloud-IDA-MSA Conversion of Fragility Curves for Efficient and High-Fidelity Resilience Assessment,” *J. Struct. Eng.*, vol. 147, no. 5, Art. no. 5, May 2021, doi: 10.1061/(ASCE)ST.1943-541X.0002998.