



# Sensitivity analysis of factors controlling earth fissures due to excessive groundwater pumping

Yueting Li<sup>1,2</sup> · Noemi Friedman<sup>3</sup> · Pietro Teatini<sup>1</sup> · Andras Benczur<sup>3</sup> · Shujun Ye<sup>2</sup> · Lin Zhu<sup>4</sup> · Claudia Zoccarato<sup>1</sup>

Accepted: 8 April 2022 / Published online: 6 May 2022  
© The Author(s) 2022, corrected publication 2022

## Abstract

Aseismic earth fissures are complex consequences of groundwater withdrawal and natural hydrogeologic conditions. This paper aims to improve the understanding of the mechanism of earth fissuring and investigate the relative importance of various factors to fissure activity, including bedrock geometry, piezometric depletion, compressibility and thickness of the exploited aquifer. For these purposes, a test case characterized by an impermeable and incompressible rock ridge in a subsiding basin is developed, where stress/displacement analyses and fissure state are predicted using an interface-finite element model. Three different methods for global sensitivity analysis are used to quantify the extent of the fissure opening to the aforementioned factors. The conventional sampling based Sobol' sensitivity analysis is compared to two surrogate based methods, the general polynomial chaos expansion based Sobol' analysis and a feature importance evaluation of a gradient boosting decision tree model. Numerical results indicate that earth fissure is forming in response to tensile stress accumulation above the ridge associated to pore-pressure depletion, inducing the fissure opening at land surface with further downward propagation. Sensitivity analysis highlights that the geometry of bedrock ridge is the most influential feature. Specifically, the fissure grows more when the ridge is steeper and closer to the land surface. Pore pressure depletion is a secondary feature and required to reach a certain threshold to activate the fissure. As for this specific application, the gradient boosting tree is the most suitable method for its better performance in capturing fissure characteristics.

**Keywords** Earth fissure · Global sensitivity analysis · Surrogate models · Sobol indices · MDA

## 1 Introduction

Aquifer over-exploitation has led to land deformation in several semiarid basins worldwide and land subsidence is one of the major impacts on the earth surface. However, in certain cases, the accumulated deformation results in earth fissuring. This geological hazard has caused negative

impacts on economic activities, social security, and environment protection, thus raising greater attention in the last decades. So far, land subsidence can be accurately simulated and predicted by numerical models (Janna et al 2012; Teatini et al 2005; Ye et al 2016), whereas the mechanism of earth fissure is more complex and difficult to simulate (Budhu 2011; Hernandez-Marin and Burbey 2010; Ochoa-González et al 2018). Different hydrogeological settings favoring the occurrence of earth fissures have been conceptualized based on field studies, including buried undulating bedrock, pre-existing fault and abrupt heterogeneous thickness of aquifer (Sheng and Helm 1998; Sheng et al 2003). The features of the earth fissures such as density, shape, length, aperture, depth, and dislocation vary greatly in different settings, which also implies different driving mechanisms.

This work aims to improve the understanding of the earth fissure mechanism with the presence of buried

✉ Yueting Li  
yueting.li@phd.unipd.it

<sup>1</sup> Department of Civil, Environmental and Architectural Engineering, University of Padova, Padova, Italy

<sup>2</sup> School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

<sup>3</sup> Informatics Laboratory, Institute for Computer Science and Control (SZTAKI), Budapest, Hungary

<sup>4</sup> College of Resource Environment and Tourism, Capital Normal University, Beijing, China

bedrock ridges in subsiding basins. Knowledge of the mechanisms driving these hazards may help to predict and therefore limit significant damages to buildings, streets, highways, railroads, earth dams, water wells, and other engineering structures. Note that earth fissures that coincide with fault scarps and abrupt thickness change may also be related to seismicity (Carreón-Freyre et al 2016; Peng et al 2013), but this is beyond the scope of this work.

Modelling the behavior of earth fissure requires a deep understanding of contact mechanism and various numerical methods were developed to delineate the physics of this problem (Hernandez-Marin and Burbey 2010; Liu et al 2019; Wang et al 2015). The FE-IE (finite element-interface element) numerical method developed by Franceschini et al (2016) is a prominent approach which exhibits stable and accurate performances on quantifying fissure characteristics (Franceschini et al 2019; Frigo et al 2019; Ye et al 2018; Li et al 2021). In this study, it is adopted to simulate the fissuring process in a subsiding basin with buried bedrock ridges.

The complexity of these systems typically give rise to many uncertainties due to the geologic configuration, the pore-pressure distribution, the hydro-geomechanical parameters along with the mathematical and numerical approximation of the physical problem (Frigo et al 2019; Sheng et al 2003). In this context, a global sensitivity analysis (GSA) is fundamental to evaluate the susceptibility of input variables to fissure formation and propagation, considering their possible mutual interactions (Iooss and Le Maître 2015; Saltelli and Annoni 2010). A variance-based GSA is employed based on the functional decomposition of the output variance, providing the Sobol' indices that quantify the input contribution to the output variance (Sobol' 1993, 2001).

First, we compute the indices using an efficient Monte Carlo sampling design, employing the Sobol' sequence to generate a uniformly distributed sample over the uncertain input domain (Sobol' et al 2011). However, a large number of samples are needed, in particular when interaction factors are investigated. This means a computationally prohibitive cost for large scale models, as it is the case of earth fissure modelling. For this reason, the use of surrogate (or proxy) models, which are approximations of the forward model built from a limited number of runs of the full model, is seen as a prominent approach to reduce the overall computational cost of the sensitivity analysis. Among surrogate methods, polynomial chaos expansion (GPC) is a probabilistic method which uses orthogonal polynomial projections of the input random variables to build the stochastic model output (Ghanem and Spanos 1991). This technique provides a straightforward way to derive Sobol' indices from model representation coefficients (Crestaux et al 2009). Thanks to these advantages,

GPC surrogates have been recently applied for GSA in environmental modelling (Ciriello et al 2013; Couaillier and Savin 2019; Friedman et al 2021; Sochala and Le Maître 2013; Kaintura et al 2018; Zoccarato et al 2020).

However, difficulties may rise when the quantity of interest presents some discontinuities with respect to the model parameters (Sochala and Le Maître 2013; Le Maître et al 2004). In case of earth fissuring simulation, this occurrence occurs when the discontinuity develops within the continuous porous medium. To overcome this problem, we elected to employ a decision tree-based method such as the gradient boosting tree (GBT) that uses an ensemble of decision trees to approximate the solution, in particular for non-linear models with arbitrary inputs (Friedman 2001; Louppe 2014). Although tree-based models are considered as "black box", many interpretation methods, such as Shapley Additive Explanations and Mean Decrease Accuracy (MDA) were designed to assess feature importance according to their relevance for the corresponding estimator, similarly to the key insights of GSA (Arabameri et al 2022; Breiman 2001; Carvalho et al 2019; Lundberg and Lee 2017).

The paper is structured as follows. At the beginning a brief background of the geomechanical modelling approach is provided. Then, the GPC and GBT methods are described with their corresponding importance indices (Sobol' and MDA). The setup of forward model and parameterization of interest are presented in the next section. The results of the numerical simulations and the statistical analyses are then discussed in detail with a list of main conclusions that close the paper.

## 2 Numerical model

The numerical model consists of a continuous model and a contact mechanism model, where the former provides the stress field analysis while the latter describes the generation and propagation of fissures. Note that when the fissure location is unknown, the stress field is fundamental to identify the potential location.

### 2.1 Continuum model

Stress and strain fields caused in a 3D continuous porous medium  $\Omega$  are quantified by means of the classical poroelasticity theory (Biot 1941). The governing equations read:

$$-\text{div } \boldsymbol{\sigma} = \mathbf{0} \quad (1)$$

where  $\boldsymbol{\sigma} = (\mathbf{C} : \nabla^s \mathbf{u} - b p \mathbf{1})$  is the total Cauchy stress tensor, with  $\nabla^s = \frac{1}{2}(\nabla + \nabla^T)$  the symmetric gradient

operator,  $\mathbf{C}$  the rank-4 elasticity tensor,  $b$  the Biot coefficient, and  $\mathbf{I}$  the rank-2 identity tensor. The displacement field  $\mathbf{u}$  is the primary unknown. The pore pressure  $p$  is a known forcing term, either imposed according to previous physical knowledge and measurements or provided by a groundwater flow model (Ye et al 2018).

Without loosing the general validity of the approach, the constitutive relationship between stresses and strains used in this study is assumed linear elastic. Consequently, the soil compressibility  $C_m$  and Poisson ratio  $\nu$  are constant and do not vary with the pressure (i.e., stress) change. Moreover, as usually implemented in the geomechanical application related to aquifer over-exploitation (Hernandez-Marin and Burbey 2012; Ye et al 2016; Zhu et al 2020), the small strain hypothesis is adopted. A tetrahedral finite element (FE) discretization is used.

The stress analysis in the continuous model is used to locate the zones where shear and tension accumulate, i.e. they are more prone to fissuring. These are the sites where the earth fissure model is “inserted” to check the actual occurrence of discontinuity development and growth.

### 2.2 Earth fissure (EF) model

From a mathematical standpoint, a geological discontinuity such as an earth fissure can be represented as a pair of friction surfaces, possibly in contact with each other, embedded within  $\Omega$ . The model must ensure the normal contact constraint, namely the impenetrability of the two portions of the porous body detected by the discontinuity.

The discrete fracture model proposed by Karimi-Fard et al (2003) and Garipov et al (2016) is used to describe the contact mechanics. More specifically, we take advantage of the model implementation proposed by Franceschini et al (2016) and Franceschini et al (2019) where the fracture network is discretized by interface elements (IEs), which are zero-thickness FEs with shape functions compatible to those of the surrounding FEs.

The fissure is considered as a boundary  $\Gamma_f$  within  $\Omega$ , with a contact condition acting on the opposed surfaces  $\Gamma_f^1$  and  $\Gamma_f^2$  that allows for a relative displacement (opening and sliding) between corresponding points whenever the stress state violates a certain failure criterion. In this modeling approach we elect to rely to a failure criterion based on the classical Mohr-Coulomb framework, which imposes the following condition on  $\Gamma_f^1$  and  $\Gamma_f^2$ :

$$f(\mathbf{t}) = \|\mathbf{t}_T\|_2 - (c - t_N \tan(\varphi)) \leq 0 \tag{2}$$

where  $\mathbf{t} = \boldsymbol{\sigma} \cdot \mathbf{n}$  is the contact stress, with  $\mathbf{t}_T = \mathbf{t} - t_N \mathbf{n}$  and  $t_N = \mathbf{t} \cdot \mathbf{n}$  the tangential and normal components, respectively. The unit vector  $\mathbf{n}$  denotes the normal vector for the surface pair  $\Gamma_f^1$  and  $\Gamma_f^2$ . In the Coulomb criterion,  $c$  and  $\varphi$

are the cohesion and friction angle, respectively. The impenetrability of solid bodies is prescribed by the normal contact condition:

$$t_N g_N = 0 \tag{3}$$

where  $g_N$  is the normal component of  $\mathbf{g}$ , i.e.  $g_N = \mathbf{g} \cdot \mathbf{n}$ , representing the relative displacement between  $\Gamma_f^1$  and  $\Gamma_f^2$ .  $\mathbf{g}$  is defined as:

$$\mathbf{g} = g_N \mathbf{n} + \mathbf{g}_T = \llbracket \mathbf{u} \rrbracket = \mathbf{u}|_{\Gamma_f^1} - \mathbf{u}|_{\Gamma_f^2} \tag{4}$$

where  $\mathbf{u}$  is the global displacement in  $\Omega$ , as consistently computed through Eq. (1), and  $\mathbf{u}|_{\Gamma_f^i}$  the restriction to  $\Gamma_f^i$ .

The application of the friction law (Eq. (2)) and the principle of impenetrability of solid bodies (Eq. (3)) subdivide the inner boundary  $\Gamma_f$  into three portions:

1.  $f < 0$  and  $t_N < 0$ : the fissure is in a stick state, i.e. the discontinuity is fully closed and behaves as a part of the continuum;
2.  $f = 0$  and  $t_N < 0$ : the fissure is in a slip state, i.e. a slip displacement is freely allowed at a fixed tangential traction  $\tau_{max} = c - t_N \tan(\varphi)$ ;
3.  $t_N = 0$ : the fissure is in an open state, i.e. both opening and slip displacements are freely allowed with zero traction.

The main challenges to find the solution in terms of  $\mathbf{u}$  and  $\mathbf{t}$  is the identification of stick, slip and open portions of the fissure surfaces. While the maximum extent of the fissure is fixed during the discretization phase, corresponding to the whole surface discretized through IEs, the subdivision into the three different states and the corresponding constraints evolve during the simulation. Details about the discretization and solution strategy can be found in Franceschini et al (2016) and Franceschini et al (2019).

### 3 Sensitivity analysis

A sensitivity analysis framework is implemented to investigate the most important factors (and their interactions) controlling earth fissuring in subsiding basis. In this section, we first provide the mathematical framework of two different types of surrogate models (GPC and GBT) used to reduce the computational burden by approximating the full forward model. Then, Sobol’ indices for sensitivity analysis are introduced with specific reference to their numerical computation based on Sobol’ and GPC approaches. For GBT, the mean decrease accuracy metric (MDA) is presented as a measure of the relative factor importance. The notion of partial dependence is also introduced to characterize the dependence of the model response on individual factors.

### 3.1 Surrogate models

#### 3.1.1 Generalized polynomial chaos expansion (GPC)

Running the forward geomechanical model multiple times for large and complex systems can be a very demanding task, both in terms of CPU and memory requirements. A GPC approach (Wiener 1938; Xiu 2007) is therefore proposed to approximate the outcome of the deterministic simulator as a function of the uncertain input parameters with the help of polynomials. With such approximation, propagation of the input uncertainties to the model output can be efficiently computed and statistics such as mean, variance, and quantiles can be easily determined.

The main idea of GPC surrogate models is based on using orthogonal polynomial approximations of the random input to project the stochastic model output. In the following, we provide the basic mathematical framework as derived in Xiu (2007). Let us consider the random model output  $U \in \mathbb{R}$  written as a function of the random vector  $\mathbf{Z}$  of  $n$  mutually independent random variables  $\mathbf{Z} = (Z_1, \dots, Z_n)$  and distribution function  $F_Z(z_1, \dots, z_n) = \mathcal{P}(Z_1 \leq z_1, \dots, Z_n \leq z_n)$ . We are considering a stochastic process in the probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  with space of events  $\Omega$ ,  $\sigma$ -algebra  $\mathcal{F}$  and probability measure  $\mathcal{P}$  on  $\mathcal{F}$ , see e.g. Xiu (2010).  $\mathbf{Z}$  can directly be the vector of the input random variables, or more usually a set of independent random variables, the so called ‘germs’, by which the input variables can be described.

As usual, the independence assumption implies  $F_Z(\mathbf{z}) = \prod_{i=1}^n F_{Z_i}(z_i)$ , where  $F_{Z_i}(z_i) = \mathcal{P}(Z_i \leq z_i)$  is the marginal distribution function with  $i = 1, \dots, n$ . Since any random variable may be represented as a series of polynomials in uncorrelated and independent Gaussian variables (Wiener 1938) and, in its generalized extension, in non-Gaussian measures, GPC basis functions of a univariate random variable  $Z_i$  are defined as the polynomials  $\{\phi_k(Z_i)\}_{k=0}^N$  of  $N$ th-degree satisfying the orthogonality conditions:

$$\mathbb{E}[\phi_s(Z_i)\phi_r(Z_i)] = \int_{\Sigma_i} \phi_s(z_i)\phi_r(z_i)dF_{Z_i}(z_i) = \gamma_s\delta_{s,r} \quad (5)$$

$0 \leq s, r \leq N$

with  $\gamma_s = \mathbb{E}[\phi_s^2(Z_i)]$  the normalization factors,  $\delta_{s,r}$  the Kronecker delta function and  $\Sigma_i$  is the support of  $Z_i$ . In the multivariate case, the GPC basis functions  $\Phi_\alpha(\mathbf{Z})$  of degree up to  $N$  are products of the univariate orthogonal polynomials:

$$\Phi_\alpha(\mathbf{Z}) = \phi_{\alpha_1}(Z_1) \dots \phi_{\alpha_n}(Z_n) \quad \text{with } 0 \leq |\alpha| \leq N \quad (6)$$

where  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$  is a multi-index with  $|\alpha| = \alpha_1 + \dots + \alpha_n$ . The multivariate basis functions are

orthogonal polynomials in  $L^2_{dF_Z}$ , that is, the space of all mean-square integrable functions of  $\mathbf{Z}$  with respect to the inner product based on the measure  $dF_Z$ :

$$\mathbb{E}[\Phi_\alpha(\mathbf{Z})\Phi_\lambda(\mathbf{Z})] = \int_{\Sigma} \Phi_\alpha(\mathbf{z})\Phi_\lambda(\mathbf{z})dF_Z(\mathbf{z}) = \gamma_\alpha\delta_{\alpha\lambda} \quad (7)$$

where  $\Sigma$  is defined by  $\Sigma = \Sigma_1 \times \Sigma_2 \dots \times \Sigma_n$ . As a consequence, the class of orthogonal polynomials is selected according to the measure  $F_Z$ .

In the GPC context, we aim at finding an approximation  $\tilde{U}_{\text{GPC},N}(\mathbf{Z})$  of the random function  $U(\mathbf{Z}) \in \mathbb{R}$  in the  $N$ -th degree polynomial space generated by the basis functions  $\Phi_\alpha(\mathbf{Z})$ :

$$U(\mathbf{Z}) \approx \tilde{U}_{\text{GPC},N}(\mathbf{Z}) = \sum_{|\alpha| \leq N} c_\alpha \Phi_\alpha(\mathbf{Z}) \quad (8)$$

where  $c_\alpha$  are the coefficients of the expansion. For  $U(\mathbf{Z}) \in L^2_{dF_Z}$ , the coefficients  $c_\alpha$  can be computed by defining  $\tilde{U}_{\text{GPC},N}$  as the orthogonal projection of  $U$  onto the polynomial space  $\mathcal{Z} = \text{span}\{\Phi_\alpha\}$ . By prescribing the orthogonality condition  $U - \tilde{U}_{\text{GPC},N} \perp \text{span}\{\Phi_\alpha\}$ :

$$\int_{\Sigma} [U(\mathbf{Z}) - \tilde{U}_{\text{GPC},N}(\mathbf{Z})]\Phi_\alpha dF_Z = 0 \quad (9)$$

The coefficients  $c_\alpha$  read:

$$c_\alpha = \frac{1}{\gamma_\alpha} \mathbb{E}[U(\mathbf{Z})\Phi_\alpha(\mathbf{Z})] = \frac{1}{\gamma_\alpha} \int_{\Sigma} U(\mathbf{z})\Phi_\alpha(\mathbf{z})dF_Z(\mathbf{z}) \quad |\alpha| \leq N \quad (10)$$

i.e., they can be computed by numerically evaluating the integral of the product of  $\Phi_\alpha$  and  $U$ . The expansion terms of Eq. (10) guarantees the optimal approximation of  $U$  in the sense of the norm defined in  $L^2_{dF_Z}$ .

The coefficients  $c_\alpha$  of the approximating GPC are numerically computed by a non-intrusive approach, that is without having to touch the finite element computation, and computing the coefficients only from samples  $\mathbf{z}^j$  of the parameters  $\mathbf{Z}$  and the corresponding  $U(\mathbf{z}^j)$  values. We use a pseudospectral projection, with the integral term approximated by a high-dimensional quadrature rule:

$$c_\alpha \approx \tilde{c}_\alpha = \sum_{j=1}^q U(\mathbf{z}^j)\Phi_\alpha(\mathbf{z}^j)w(\mathbf{z}^j) \quad (11)$$

with  $\mathbf{z}^j$  and  $w(\mathbf{z}^j)$  the  $q$  integration nodes and weights, respectively. Since  $\Phi_\alpha$  is at most of degree  $N$ , the integrand function has at most degree  $2N$ . In the univariate case, this requires the use of a  $(q = N + 1)$ -point Gaussian quadrature rule, while in the multivariate case with  $n$  random variables the number of points grows up to  $q = (N + 1)^n$ . Using this approximation, the surrogate model needs the

evaluation of  $U$  through the numerical solver of the forward model at the  $q$  integration points  $\mathbf{z}^j$ .

Another approach to compute the coefficients  $c_\alpha$  of the expansion is by regression, that is, by minimizing the (unweighted) mean squared  $L^2$  error

$$\sum_{j=1}^q (U(\mathbf{z}^j) - \tilde{U}_{\text{GPC},N}(\mathbf{z}^j))^2 = \sum_{j=1}^q \left( U(\mathbf{z}^j) - \sum_{|\alpha| \leq N} c_\alpha \Phi_\alpha(\mathbf{z}^j) \right)^2 \tag{12}$$

of the expansion. The sum attains its minimum, where the gradient is zero, that is, where

$$-2 \sum_{j=1}^q \left( U(\mathbf{z}^j) - \sum_{|\alpha| \leq N} c_\alpha \Phi_\alpha(\mathbf{z}^j) \right) \Phi_\beta(\mathbf{z}^j) = 0 \tag{13}$$

for all  $|\beta| < N$ . The system of equations can be solved for the coefficients  $c_\alpha$ .

### 3.1.2 Gradient boosting tree (GBT)

Convergence of the GPC approximation is especially favorable when the dependence of the model output on the given uncertain input parameters have sufficient smoothness. As this is not necessarily the case for the given geomechanical model, we have also tested a gradient boosting tree (GBT) approximation of the model output to explore the relationship between the input parameters and the model output.

Given that the model output  $U$  is a function of the input variables  $\mathbf{Z} = (Z_1, \dots, Z_n)$ , gradient boosting method assumes that the approximation  $\tilde{U}_{\text{GBT}}(\mathbf{Z})$  is represented by an ensemble of base learners (e.g., weak basic models) which minimizes the average value of a specified loss function  $L(U(\mathbf{Z}), \hat{U}(\mathbf{Z}))$  such that:

$$\tilde{U}_{\text{GBT}}(\mathbf{Z}) = \arg \min_{\hat{U}(\mathbf{Z})} \mathbb{E}[L(U(\mathbf{Z}), \hat{U}(\mathbf{Z}))] \tag{14}$$

evaluated at the sample points, where  $\hat{U}(\mathbf{Z})$  is the predicted values of the observed values  $U$  and can be written in the form:

$$\hat{U}(\mathbf{Z}) = \sum_{m=0}^M \beta_m h_m(\mathbf{Z}) \tag{15}$$

with  $h_m(\mathbf{Z})$  the base learner at  $m$ -th stage of the boosting algorithm characterized by a fixed size of stages  $M$ . In particular, GBT uses the decision tree as base learner, thus  $h_m(\mathbf{Z})$  can be written as:

$$h_m(\mathbf{Z}) = \sum_{i=1}^{I_m} b_{im} \mathbf{I}_{R_{im}} \tag{16}$$

where  $I_m$  refers to the number of leaves at stage  $m$ , subscript  $i$  is the index for each leaf in the tree,  $b_{im}$  is the predicted value of the terminal region  $R_{im}$  with  $\mathbf{I}_{R_{im}}$  the indicator function, which takes value 1 if  $\mathbf{Z}$  lies in the subset  $R_{im}$  otherwise takes value 0. Then, a steepest descent step is commonly applied to fit  $h_m(\mathbf{Z})$  to the pseudo-residuals  $r_{jm}$  with the training set  $\{\mathbf{z}^j, U(\mathbf{z}^j)\}_{j=1}^q$ , i.e., intermediate error terms at  $m$ -th stage, for  $j$ -th sample point  $(\mathbf{z}^j, r_{jm})$ :

$$L_j = \frac{1}{2} (U(\mathbf{z}^j) - \hat{U}_{m-1}(\mathbf{z}^j))^2 \tag{17}$$

$$r_{jm} = - \frac{\partial L_j}{\partial \hat{U}_{m-1}} = U(\mathbf{z}^j) - \hat{U}_{m-1}(\mathbf{z}^j) \tag{18}$$

where the mean squared error (MSE) is used as loss function. Afterwards, the expansion coefficient  $\beta_m$  can be optimized:

$$\beta_m = \arg \min_{\beta} \sum_{j=1}^q [U(\mathbf{z}^j) - (\hat{U}_{m-1}(\mathbf{z}^j) + \beta h_m(\mathbf{z}^j))] \tag{19}$$

Therefore, the model can be updated by:

$$\hat{U}_m(\mathbf{Z}) = \hat{U}_{m-1}(\mathbf{Z}) + \beta_m h_m(\mathbf{Z}) \tag{20}$$

Moreover, regularization methods impose the constrains on fitting procedure to prevent the overfitting, that is when the surrogate model exactly describes the training data and fails to fit unseen data. For example, the maximum stage of gradient boosting  $M$  in Eq. (15) is a natural regularization parameter which discourages learning more complex model to avoid overfitting. However, it has been found that regularization through shrinkage provides superior results to that obtained by restricting the maximum stage (Copas 1983), hence a simple shrinkage strategy is added to Eq. (20):

$$\hat{U}_m(\mathbf{Z}) = \hat{U}_{m-1}(\mathbf{Z}) + v \beta_m h_m(\mathbf{Z}) \tag{21}$$

Under this form, two regularization parameters are used in the gradient boosting algorithm: the learning rate  $v$  and the number of boosting stage  $M$ .

## 3.2 Variable importance metrics

### 3.2.1 Variance-based Sobol' indices

Consider the model under investigation is described as a function  $U = f(\mathbf{Z})$ , where  $U$  is a scalar and the input  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$  is defined over the  $n$ -dimensional hypercube  $I^n$  with mutually independent components. Assuming  $U$  to

be a square integrable function, the Sobol’ functional decomposition scheme reads:

$$f(\mathbf{Z}) = f_0 + \sum_{i=1}^n f_i(Z_i) + \sum_{i=1}^n \sum_{j>i}^n f_{ij}(Z_i, Z_j) + \dots + f_{12\dots n}(Z_1, Z_2, \dots, Z_n) \tag{22}$$

where each term is square integrable over  $I^n$  and mutually orthogonal. The terms in the decomposition may be derived by:

$$\begin{aligned} f_0 &= \mathbb{E}(U) \\ f_i &= \mathbb{E}_{\mathbf{Z}_{\sim i}}(U|Z_i) - \mathbb{E}(U) \\ f_{ij} &= \mathbb{E}_{\mathbf{Z}_{\sim ij}}(U|Z_i, Z_j) - f_i - f_j - f_0 \end{aligned} \tag{23}$$

and similarly for higher degree terms. In Eq. (23)  $\mathbb{E}_{\mathbf{Z}_{\sim i}}$  denotes the expected value over all elements of the parameters  $\mathbf{Z}$  except the  $i$ -th one.

The decomposition allows to attribute the variances to the different parameters or their various degree interactions. These can be given by the partial variances:

$$\begin{aligned} V_i &= \text{Var}(f_i(Z_i)) \\ V_{ij} &= \text{Var}(f_{ij}(Z_i, Z_j)) \end{aligned} \tag{24}$$

and similarly for higher order terms. With the help of the partial variances, the total variance can be decomposed:

$$\text{Var}(U) = \sum_{i=1}^n V_i + \sum_{i=1}^n \sum_{j>i}^n V_{ij} + \dots + V_{12\dots n} \tag{25}$$

The variance based sensitivity is described by the ratio of the partial variances and the total variance. The first and second-order Sobol’ indices are defined as:

$$\begin{aligned} S_i &= \frac{V_i}{\text{Var}(U)} \\ S_{ij} &= \frac{V_{ij}}{\text{Var}(U)} \end{aligned} \tag{26}$$

and similarly for the higher order sensitivity indices. The first order indices  $\{S_i\}_{i=1}^n$  measure the effect on the output variance of factor  $Z_i$  alone. Higher-order indices represents the combined effect of the group of factors  $Z_1, Z_2, \dots, Z_n$  on the variance of the model output.

Another sensitivity measure is the total index of the  $i$ -th factor:

$$S_{Ti} = 1 - \frac{\text{Var}_{\mathbf{Z}_{\sim i}}(\mathbb{E}_{Z_i}(U|\mathbf{Z}_{\sim i}))}{\text{Var}(U)} \tag{27}$$

where  $\text{Var}_{\mathbf{Z}_{\sim i}}(\mathbb{E}_{Z_i}(U|\mathbf{Z}_{\sim i}))$  can be regarded as the first order index of  $\mathbf{Z}_{\sim i}$ , so that  $S_{Ti}$  measures the contribution to the output variance of all terms which contain factor  $Z_i$ .

These indices can be calculated by Monte Carlo method (Saltelli 2002; Saltelli and Annoni 2010). The

procedure is as follows. Generate  $q \times 2n$  sample matrix of the input random variables  $\mathbf{Z}$ . The first  $n$  columns are gathered as matrix  $\mathbf{A}$  and the second  $n$  columns are used similarly as matrix  $\mathbf{B}$ . From these two matrices we generate  $n$  further  $q \times n$  matrices  $\mathbf{A}_B^i$  by taking matrix  $\mathbf{A}$  and replacing its  $i$ -th column with the corresponding column of  $\mathbf{B}$ . The estimators:

$$\begin{aligned} V_i(\mathbb{E}_{\mathbf{Z}_{\sim i}}(U|Z_i)) &\approx \frac{1}{q} \sum_{j=1}^q f(\mathbf{B})_j (f(\mathbf{A}_B^i)_j - f(\mathbf{A})_j) \\ \mathbb{E}_{\mathbf{Z}_{\sim i}}(V_{Z_i}(U|\mathbf{Z}_{\sim i})) &\approx \frac{1}{2q} \sum_{j=1}^q (f(\mathbf{A})_j - f(\mathbf{A}_B^i)_j)^2 \end{aligned} \tag{28}$$

used in Eqs. (26) and (27) allow to compute the indices  $S_i$  and  $S_{Ti}$ .

In this work, we use the Sobol’ sequence, i.e., a quasi-random low discrepancy sequence, to generate the samples  $\mathbf{z}^j$  (Saltelli 2002; Saltelli and Annoni 2010). The difference with the ordinary Monte Carlo is that quasi-Monte Carlo substitutes random points with low discrepancy sequences, thus improving the convergence of the estimator. The main problem with this sampling-based method is the cost of computing  $f(\mathbf{A}_B^i)$ . Instead, by using the GPC or GBT surrogates as model proxy, the model output can be directly computed from the proxy model in a computationally cheap manner.

One of the great advantages of the GPC surrogate model is that the Sobol’ indices can be computed analytically without using the sampling based approximation given in Eq. (28). Due to the orthogonality condition, the mean and the total variance of the GPC can be directly computed from the coefficients of the expansion as follows:

$$\mathbb{E}(f(\mathbf{Z})) \approx \mathbb{E}(\tilde{f}) = c_{\alpha}|_{|\alpha|=0}, \tag{29}$$

$$\text{Var}(f(\mathbf{Z})) \approx \text{Var}(\tilde{f}) = \sum_{0 < |\alpha| \leq N} c_{\alpha}^2 \gamma_{\alpha}. \tag{30}$$

According to Sudret (2008), if we introduce the set of  $\alpha$  tuples  $\mathcal{J}_{i_1, \dots, i_s}$  such way that only the indices  $(i_1, \dots, i_s)$  are nonzero:

$$\mathcal{J}_{i_1, \dots, i_s} = \left\{ \alpha : \begin{aligned} \alpha_k &> 0 \quad \forall k = 1, \dots, s, \quad k \in (i_1, \dots, i_s) \\ \alpha_j &= 0 \quad \forall k = 1, \dots, n, \quad k \notin (i_1, \dots, i_s) \end{aligned} \right\}, \tag{31}$$

than  $\mathcal{J}_i$  is defined as a set of all multi-indices that corresponds to the polynomials depending only on parameter  $Z_i$ . Consequently, the Sobol’ decomposition (see Eq. (22)) of the GPC approximation is straightforward:

$$\begin{aligned} \tilde{f}(Z) = & f_0 + \sum_{i=1}^n \sum_{\alpha \in \mathcal{J}_i} c_\alpha \Phi_\alpha(Z_i) \\ & + \sum_{1 \leq i_1 < i_2 \leq n} \sum_{\alpha \in \mathcal{J}_{i_1, i_2}} c_\alpha \Phi_\alpha(Z_{i_1}, Z_{i_2}) \\ & + \dots + \sum_{\alpha \in \mathcal{J}_{1, 2, \dots, n}} c_\alpha \Phi_\alpha(Z_1, Z_2, \dots, Z_n) \end{aligned} \tag{32}$$

and thus any element of the decomposition can be written as:

$$f_{i_1, \dots, i_s} = \sum_{\alpha \in \mathcal{J}_{i_1, \dots, i_s}} c_\alpha \Phi_\alpha(Z_{i_1}, \dots, Z_{i_s}) \tag{33}$$

The partial variances can be also easily computed from:

$$V_{i_1, \dots, i_s} = \text{Var} \left( \sum_{\alpha \in \mathcal{J}_{i_1, \dots, i_s}} c_\alpha \Phi_\alpha(Z_{i_1}, \dots, Z_{i_s}) \right) = \sum_{\alpha \in \mathcal{J}_{i_1, \dots, i_s}} c_\alpha^2 \eta_\alpha \tag{34}$$

that is, the coefficients corresponding to the polynomials that have dependence only on the selected variables have to be collected, squared, multiplied with its norm and summed up. For the sensitivity index this expression has to be divided by the total variance given in Eq. (30).

### 3.2.2 Mean decrease accuracy (MDA)

For the GBT surrogate model instead of computing the Sobol’ indices with the help of the MC estimation, we use a different sensitivity measure that can be efficiently computed by GBT models. Each feature importance is here evaluated through a permutation-based measure following the idea of Breiman (2001) and the application in Jaxa-Rozen and Kwakkel (2018). Given the  $q \times n$  matrix  $\mathbf{A}$  of the random input variables  $\mathbf{Z}$ , the MDA index of the  $i$ -th feature measures the decrease of the estimator accuracy by randomly permuting the values of  $Z_i$  ( $i$ -th column of input variables matrix) for  $K$  times in total and for each repetition re-computing the ensemble tree predictions with the  $k$ -th (for  $k = 1, \dots, K$ ) permuted column  $\mathbf{A}_i$ . The higher the inaccuracy, the most important is the feature for the particular model. MDA of the  $i$ -th feature is defined as

$$\text{MDA}_i = s - \frac{1}{K} \sum_{k=1}^K s_{k,i} \tag{35}$$

where  $s$  is the reference score and  $s_{k,i}$  the score for the  $k$ -th permutation of feature  $Z_i$ , where the score is obtained by computing the mean square error between predictions and observations. The feature is important if permuting its values causes a large drop on the model performance.

### 3.3 Convergence criterion of importance indices

Here we apply the convergence criterion proposed by Roustant et al (2014) to evaluate the stability of the important indices. The vector  $\mathbf{V}_q = (v_1, \dots, v_n)$  of the variable importance indices is estimated from a sample size of  $q$  observation points, where  $n$  is the number of input features. Specifically, the Euclidean norm of the vector is taken into account rather than the individual indices so that the more influential indices have more effect on the convergence measurement. The importance indices are computed sequentially over an increasing sample size at intervals of  $\Delta q$ . Then the convergence criterion  $k_q$  is computed by:

$$k_q = \frac{1/t \sum_{i=1}^t \|\mathbf{V}_q - \mathbf{V}_{q-i\Delta q}\|}{\|\mathbf{V}_q\|} \tag{36}$$

where  $\|\cdot\|$  is the Euclidean norm and  $t$  is the number of total intervals. The values of  $\Delta q$  and  $t$  are case-dependent. This criterion will be imposed on the total Sobol’ indices  $S_T$  and MDA.

### 3.4 Partial dependence

Compared to Sobol’ indices and MDA, partial dependence is more similar to one-at-a-time (OAT) sensitivity analysis, which assumes the model response is a function of one or two input variables and characterizes the average marginal effect on model prediction (Goldstein et al 2015). Owing to this feature, partial dependence plot can visually depicts the relationship between model prediction and the variables of interest. The partial dependence function  $\tilde{U}_i$  reads:

$$\tilde{U}_i(z_i) = \mathbb{E}_{\mathbf{Z}_j} [\tilde{U}(z_i, \mathbf{Z}_j)] = \int \tilde{U}(z_i, \mathbf{Z}_j) d\mathcal{P}(\mathbf{Z}_j) \tag{37}$$

where  $z_i$  and  $\mathbf{Z}_j$  are respectively the feature set of interest and complement used in the surrogate models  $\tilde{U}$ . Generally,  $z_i$  accounts for two components at most. The partial dependence can be computed by a Monte or quasi Monte Carlo method with the help of a surrogate model.

### 3.5 Software availability

In this work, the forward geomechanical models are carried out by the GEPS3D simulator (Isotton et al 2019; Franceschini et al 2016). The reference global sensitivity analysis, that is Sobol’ technique, is implemented by the *SALib* library in Python environment (Herman and Usher 2017). *SGLib* library (Zander 2020; Friedman and Zander 2020) is used to compute the polynomial chaos expansion and Sobol’ indices (Vittek et al 2006). Gradient boosting tree algorithm and mean decrease accuracy computation

are carried out by the *scikit – learn* module in Python with gradient boosting regression estimator and permutation feature importance function (Pedregosa et al 2011).

#### 4 Model setup and parameterization

The investigated configuration conceptualizes the geological setting in Wuxi, China, where the compressible deposits of the Yangtze River cover an undulating bedrock.

The numerical simulation is developed on a quasi-3D domain 2000-m long ( $x$ -direction), 50-m thick ( $y$ -direction), and 500-m deep ( $z$ -direction). A traction-free top surface and a fixed bottom surface are considered (Fig. 1). On the lateral surfaces the horizontal displacements are precluded in the orthogonal direction. The conceptual model is composed of three hydrostratigraphic units: an upper aquitard, a bottom aquifer, and a buried triangular bedrock. For sake of simplicity, each material is assumed to behave elastically with the same Poisson ratio  $\nu = 0.25$ . Cohesion  $c$  and friction angle  $\varphi$  are set equal to 0.01 MPa and  $30^\circ$  respectively. A piezometric drop lineally varies from 0 to  $\Delta p$  in 10 years and is uniformly assigned to the bottom aquifer, meanwhile the upper aquitard is regarded as an hydraulically “inactive” unit where the pore pressure propagation from the underlying sandy layer is negligible. The initial stress field is computed based on the gradient density ( $\sigma_v = 1200 \text{ kg/m}^2/\text{m}$ ) and the minimum-to-maximum stress ratio reads  $\sigma_h/\sigma_v = \nu/(1 - \nu)$ .

The vertical size of tetrahedral FE elements is 10 m and the horizontal dimension is in range between 5.5 and 20 m, slightly varying according to the ridge geometry. Previous studies have proved that earth fissures are most prone to generate from the ground surface above the ridge tip downward, with a stress state characterized by high tension above the ridge (Ye et al 2018; Frigo et al 2019). Moreover, fissures cannot propagate within the bedrock where pressure does not change and the stress field variation is negligible. Therefore, a IE alignment is vertically introduced from the land surface to the ridge tip as highlighted by the white line in Fig. 1. The triangular IE discretization

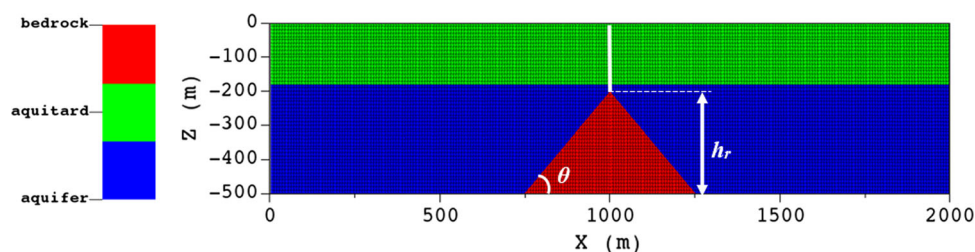
is consistent with FE discretization. Stress distribution and magnitude depend on the ridge geometry, the aquifer thickness, and differential subsidence. This latter is primarily dependent on the pore pressure change and sediment compressibility. Therefore, these four variables, i.e. ridge geometry, aquifer thickness, pressure change and sediment compressibility, are selected as input features for GSA.

Here, the ridge geometry is characterized by the slope of the bedrock ridge ( $\tan \theta$ ). Note that the length of ridge basement is fixed at 500 m. The fraction of the aquifer thickness over the domain thickness (500 m) is denoted  $\zeta$ . The selected ranges for  $\tan \theta$  and  $\zeta$  are determined by the domain dimensions and discretization. The bounds for the aquifer compressibility ( $C_m$ ) and the maximum piezometric decline ( $\Delta p$ ) are based on available literature data on exploited aquifer systems (Burbey 2002; Conway 2016; Ochoa-González et al 2018; Ye et al 2018; Zhu et al 2020). The piezometric decline reaches the maximum value with a linear behavior in 10 years. The variability ranges of the parameters used in this study are summarized in Table 1. A uniform probability distribution is assumed for each variable.

## 5 Results

### 5.1 Deterministic model run

We randomly choose one experiment designed for GSA to present the numerical outcomes. The simulated temporal evolution of tensile stress on a vertical section of the domain is shown in Fig. 2. Tensile stress  $\sigma_h$  initially accumulates around the ridge tip. As piezometric level declines, a tensile zone also occurs at the land surface above the apex of ridge. Once tensile stress exceeds the tensile strength, i.e.  $t_N = 0$  (see Sect. 2.2), IEs change from a stick to an open state with the discontinuity that develops at the land surface and propagates downwards. Simultaneously, tensile stress dissipates due to the crack opening. Notice that the porous medium directly above the ridge tip



**Fig. 1** FE grid of one numerical experiment in sensitivity analysis, where  $\tan \theta = 1.2$ ,  $\zeta = 0.65$ . The three colors (red, green and blue) represent the hydrogeologic units.  $h_r$  represents bedrock ridge height

which is the function of  $\theta$ . The IE alignment is highlighted by a white line whose length equal to  $500 - h_r$



**Table 1** Range of the input features for GSA: the four random variables  $\tan \theta$ ,  $\zeta$ ,  $C_m$ , and  $\Delta p$  are uniformly distributed

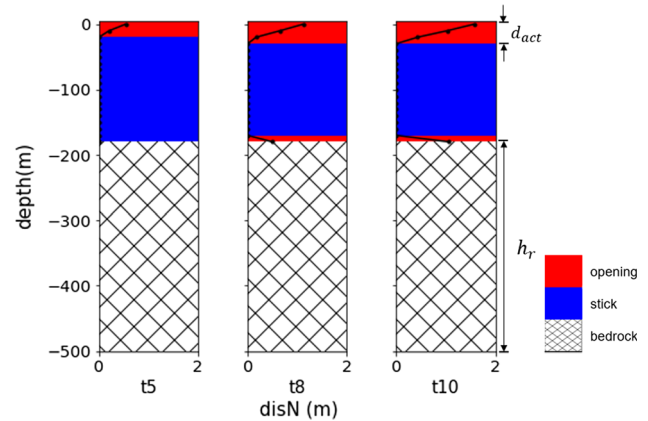
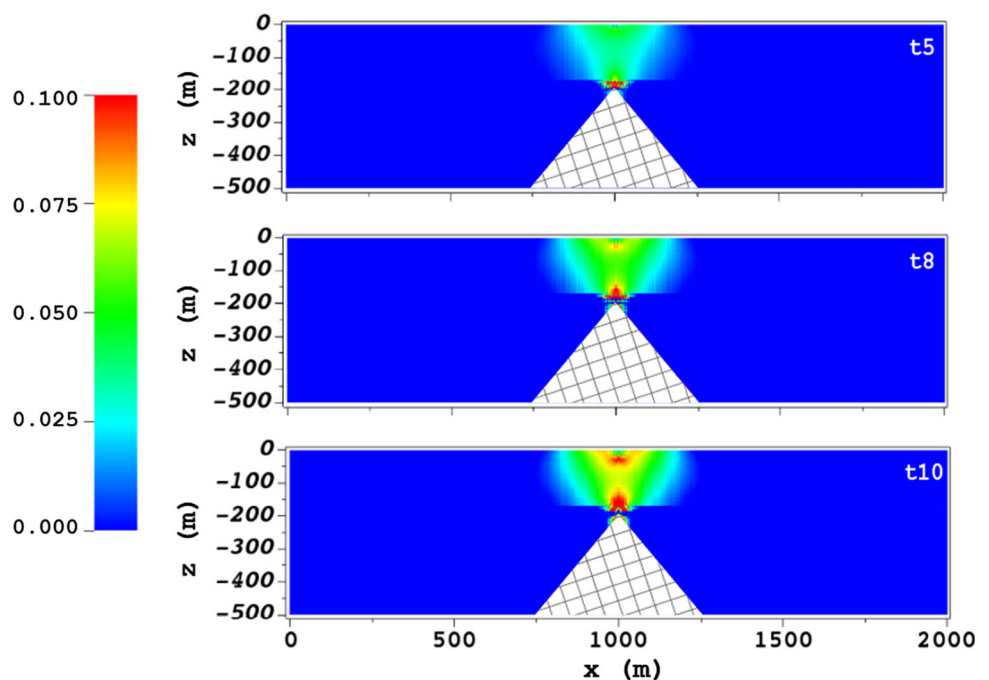
Feature	Min.	Max.
$\tan \theta$	$5.0 \times 10^{-1}$	$1.9 \times 10^{-0}$
$\zeta$	$4.0 \times 10^{-1}$	$9.0 \times 10^{-1}$
$C_m$ (MPa <sup>-1</sup> )	$5.0 \times 10^{-3}$	$5.0 \times 10^{-2}$
$\Delta p$ (MPa)	$-1.0 \times 10^{-0}$	$0.0 \times 10^{-0}$

does not experience any shear stress due to the symmetric configuration. Therefore, only fissure opening develops with this setting.

Figure 3 shows the evolution of the earth fissure as provided by the IE model. Fissure initially originates at the land surface where the tensile strength is the lowest and later develops at depth too, just above the ridge tip, in response to the concentration of tensile stress. The upper fissure which initially is very narrow, keeps enlarging horizontally and extending downwards as the aquifer pressure continues to decrease. At the 10th year, i.e. in correspondence of a piezometric decline in the confined aquifer equal to 1 MPa, the fissure reaches a depth about 30 m with a maximum opening equal to 1.7 m at the land surface and tapers with depth. Conversely, the bottom discontinuity remains confined at the depth without a significant development.

Generally, the energy is dissipated after fissuring mainly at land surface. Moreover, the overburden stress due to sediment load usually limits tensile fissuring at depth

**Fig. 2** Sequential evolution of the dimensionless horizontal stress  $\sigma_h^* = \sigma_h/\Delta p$  at the 5th, 8th and 10th years as simulated with the EF model. The results are obtained using  $\tan \theta = 1.2$ ,  $\zeta = 0.65$ ,  $C_m = 0.05 \text{ MPa}^{-1}$  and  $\Delta p = -0.89 \text{ MPa}$



**Fig. 3** Sequential evolution of fissure opening at the 5th, 8th and 10th years: opening  $g_N$  (dotted black line), opening  $\Gamma_{open}$  and close  $\Gamma_{stick}$  areas (red and blue zones) are provided. The results are obtained using  $\tan \theta = 1.2$ ,  $\zeta = 0.65$ ,  $C_m = 0.05 \text{ MPa}^{-1}$  and  $\Delta p = -0.89 \text{ MPa}$

(Budhu and Shelke 2008). Therefore, the bottom activated zone is not included in the quantity of interest  $d_{act}$  that is defined as the fissure depth from the land surface. It is worth mentioning that the size of activated depth is controlled by the vertical length of the IE alignment, thus the relative activated depth  $d_{r,act} = d_{act}/(500 - h_r)$  is introduced to have comparable results when varying the model geometry (Fig. 1). In this case,  $d_{r,act}$  equals 0.132.

## 5.2 Approximation by surrogate models

### 5.2.1 GPC surrogate

The computation of the GPC coefficients has been initially carried out by the pseudospectral approach (Eq. (11)) and increasing step by step the GPC degree of the polynomial response. The surrogate model  $\tilde{U}_{GPC}$  is developed to approximate  $d_{r,act}$  with input parameters  $\mathbf{Z} = \{\tan \theta, \zeta, C_m, \Delta p\}$ . The validation of the fitted surrogate model is carried out by employing 7000 samples, that is the available set of points used to compute Sobol' indices with the quasi Monte Carlo approach. Moreover, train and validation of the GPC surrogate is also obtained from the 7000 points using 80% for regression and the remaining 20% for validation.

The results are shown in Table 2 with the coefficient of determination  $R^2$  used to assess the fit goodness of the surrogate model to the full problem and computed by means of the leave-one-out cross-validation. Increasing the GPC degree both approaches, i.e. pseudospectral and regression, provide similar results with increasing values of the coefficient of determination  $R^2$ . Obviously, the computational cost of regression is much higher, in particular at low GPC degrees. The maximum  $R^2$  is close to 0.80. A visual comparison of the full model results and the surrogate solution is shown in Fig. 4a. The higher discrepancy is obtained at the boundary of the solution where the surrogate solution provides results larger than 1.0 or lower than 0.0. These solutions correspond to the nonphysical response meaning (i) a fissure reaches and propagates within bedrock ( $d_{r,act} > 1.0$ ) and (ii) a negative opening ( $d_{r,act} < 0.0$ ), representing non-penetration that is not admitted by the model hypothesis (see Eq. (3)).

**Table 2** Validation of the GPC at increasing value of the maximum total degree of the GPC expansion

GPC degree $N$	# Gauss points	$R_{pS}^2$	# Regression points	$R_{RG}^2$
2	81	0.64	5600	0.64
3	256	0.69	5600	0.68
4	625	0.72	5600	0.73
5	1296	0.77	5600	0.76
6	2401	0.78	5600	0.78
7	4096	0.79	5600	0.79

The coefficient of determination  $R^2$  is used to assess the goodness of fit.  $R_{pS}^2$  refers to GPC coefficients computed using the pseudospectral approach through Gauss quadrature, whereas  $R_{RG}^2$  is obtained by regression. The number of points indicated are those used to train the surrogate models

### 5.2.2 GBT surrogate

Gradient boosting algorithm is also implemented with increasing size of input data to check the convergence, thereof 80% is used to fit model with remaining 20% for validation. Hyperparameter tuning is carried out by a Grid Search method which enumerates all the possible combinations of hyperparameters and gets optimal values based on the corresponding coefficient of determination  $R_{GB}^2$ . In this application, only the learning rate  $\nu$  has been tuned given it's the most important hyperparameter for GBT estimator (Probst et al 2019).

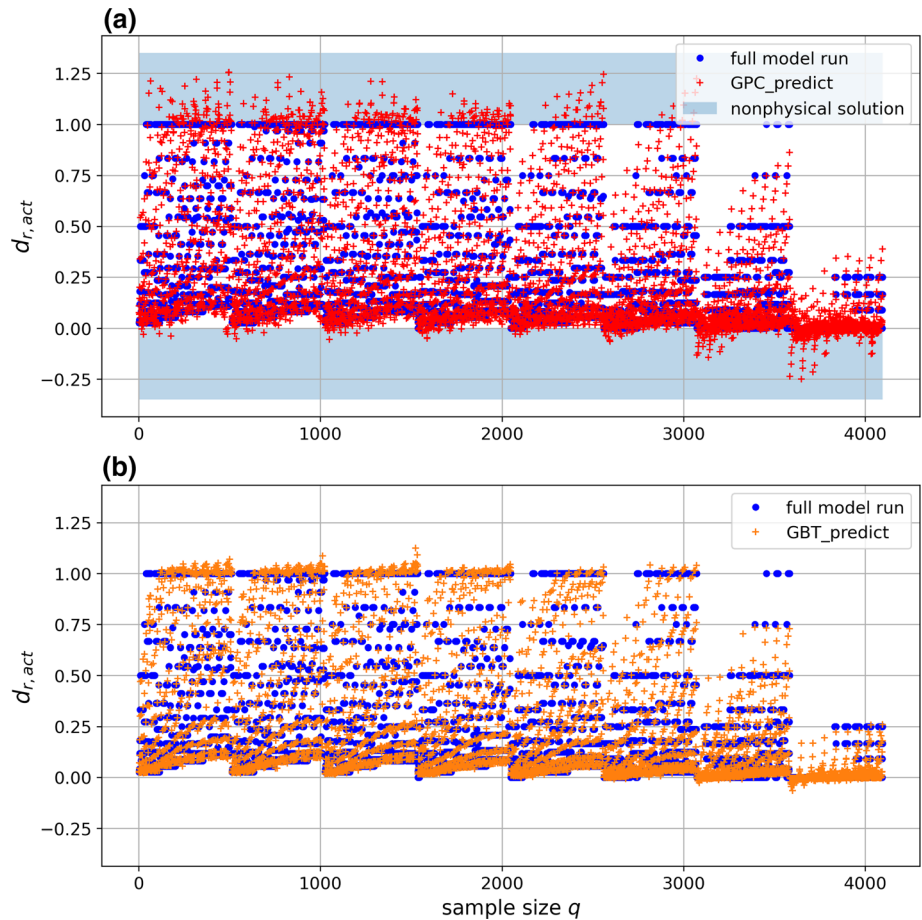
Table 3 shows the model goodness that stabilizes when the data size reaches 5000. In addition to Sobol' samples, the Gauss points (4096) for the GPC method are also used to validate the regression tree obtained from maximum sample size (7000) (Fig. 4b). The regression tree also has some nonphysical predictions ( $d_{r,act} > 1.0$ ), however the absolute values of discrepancy are much less than that of GPC solutions. Moreover,  $R_{GB}^2$  suggests GBT algorithm outperforms GPC algorithm with respect to the prediction accuracy ( $R_{GB}^2=0.96$  vs  $R_{RG}^2=0.79$ ). Nevertheless, both two surrogate models fail to capture the characteristic that the fissure opening depth  $d_{r,act}$  keeps constant within some  $q$  sub-domains irrespective of parameters variation.

## 5.3 Importance metrics

Sobol' technique is chosen as the "reference" sensitivity analysis with the input samples generated by the Sobol' sequence. As mentioned in Sect. 5.2, these samples are reused to train and validate GPC and GBT surrogates. Knowing that the reliability of sensitivity measures obtained by metamodels are dependent on their predictive power, we can assure GPC and GBT are good surrogates and the "overfitting" problem can be excluded according to the  $R^2$  values obtained from the cross validation.

Figure 5 shows that Sobol' and GPC methods estimate similar results of first order indices  $S_i$  and total Sobol' indices ( $S_{Ti}$ ), with the GPC algorithm showing a quicker convergence with respect to the quasi Monte Carlo method. For each variable,  $S_i$  accounts for the larger proportion of the corresponding  $S_{Ti}$ , indicating a minor contribution from interactions between the  $i$ -th variable and the other input factors. The second order indices from Sobol' technique and GPC, both computed with 7000 samples, are listed in Table 4. It is evident the small interaction between input factors. Note that Sobol' technique gives some negative indices for the non-influential terms indicating some computation errors which could not be eliminated with the current sample size (Herman and Usher 2017).

**Fig. 4** Visual comparison between the full model run and the surrogate solutions. **a** GPC model with regression, the blue shaded areas imply nonphysical solutions, i.e.,  $d_{r,act}$  larger than 1.0 representing a fissure that extends within the buried bedrock, and lower than 0 indicating interpenetration of solid bodies. **b** GBT model, where the predictions are basically within the rational range



**Table 3** Learning rate and validation of GBT at increasing sample size

Sample size $q$	Learning rate $\nu$	$R^2_{GB}$
1000	0.23	0.90
2000	0.14	0.94
3000	0.12	0.95
4000	0.16	0.95
5000	0.10	0.96
6000	0.10	0.96
7000	0.10	0.96

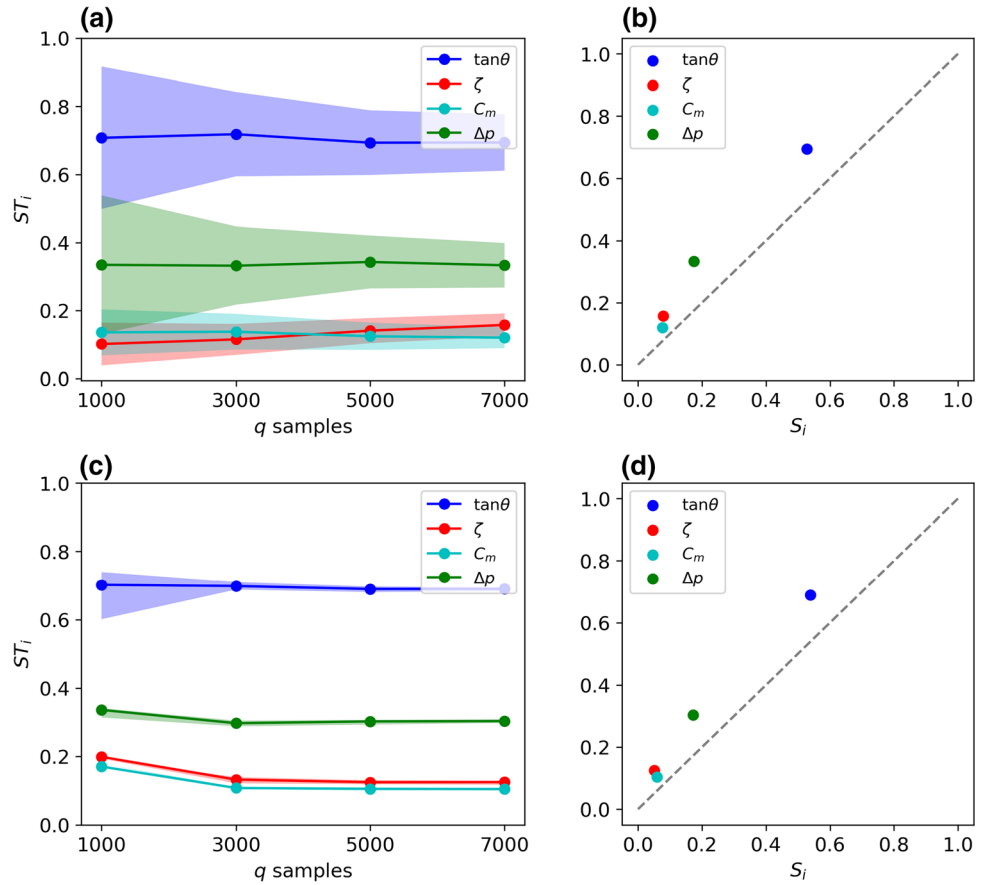
Figure 6a depicts MDA importance indices obtained from the GBT model with default repetition  $K = 100$ . A comparison with Sobol’ method measures is presented in Fig. 6b. Note that Sobol’ indices and MDA importance indices measure different quantities, thus a min-max scaling for each value is employed for direct comparison of the indices. According to Eq. (36), both GPC and GBT surrogate models reach the convergence criteria  $k_q < 0.05$  at  $q = 7000$  (with  $\Delta q = 2000$  samples and  $t = 3$  intervals). Conversely, the reference Sobol’ method fails to converge

at 7000 sample size which proves surrogate models can reduce the overall computational cost of analysis with respect to Sobol’ method. Moreover, GBT not only ranks the variables in the same way as Sobol’ method but also provides basically identical proportional indices with respect to the total effect. The importance metrics obtained from the three methods highlight the ridge geometry is the most influential variable for the fissure opening, with the pressure variation also having a non-negligible impact on the fissure development. The contributions from the other two variables are smaller.

### 5.4 Partial dependence

We also employ partial dependence to investigate the surrogate model response to the variable changes (Fig. 7). A number of 50 samples from the validation set are used to illustrate how the model prediction to one variable changes, keeping fixed the other features. Note that each sample is represented by one thin line. The thicker lines represent the partial dependence calculated from the whole validation set (20% of 7000 samples). Although there are some

**Fig. 5** Sobol’ indices with reference Sobol’ method (top panels) and GPC (bottom panels) method. Left panels present the convergence of Sobol’ total indices  $ST_{T,i}$ , with the shaded areas in **a** representing the 95% confidence intervals of the indices. The right panels shows the relationships between first order and total indices



**Table 4** Second order indices with reference Sobol’ method and GPC surrogate model

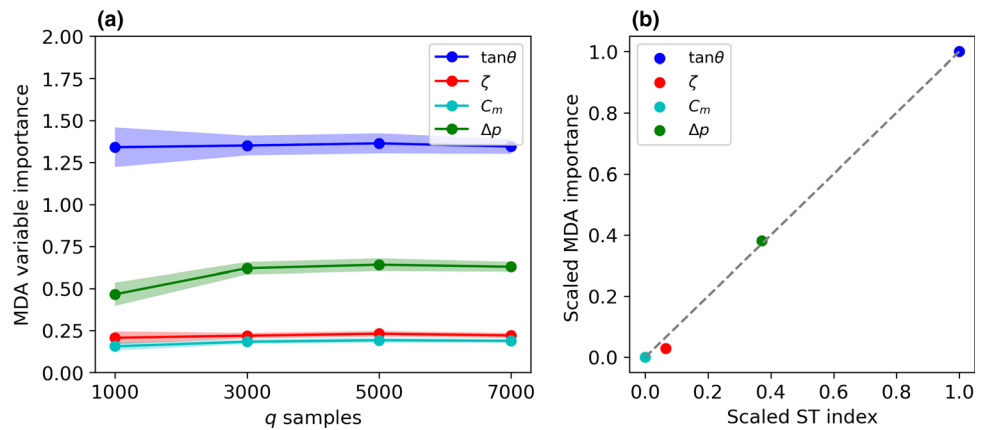
	$S_{12}$	$S_{13}$	$S_{14}$	$S_{23}$	$S_{24}$	$S_{34}$
Sobol’	0.013	0.033	0.059	-0.005	-0.005	0.010
GPC	0.026	0.020	0.070	0.002	0.017	0.007

discrepancies between GPC and GBT models results, the trend of partial dependence for each variable is similar.

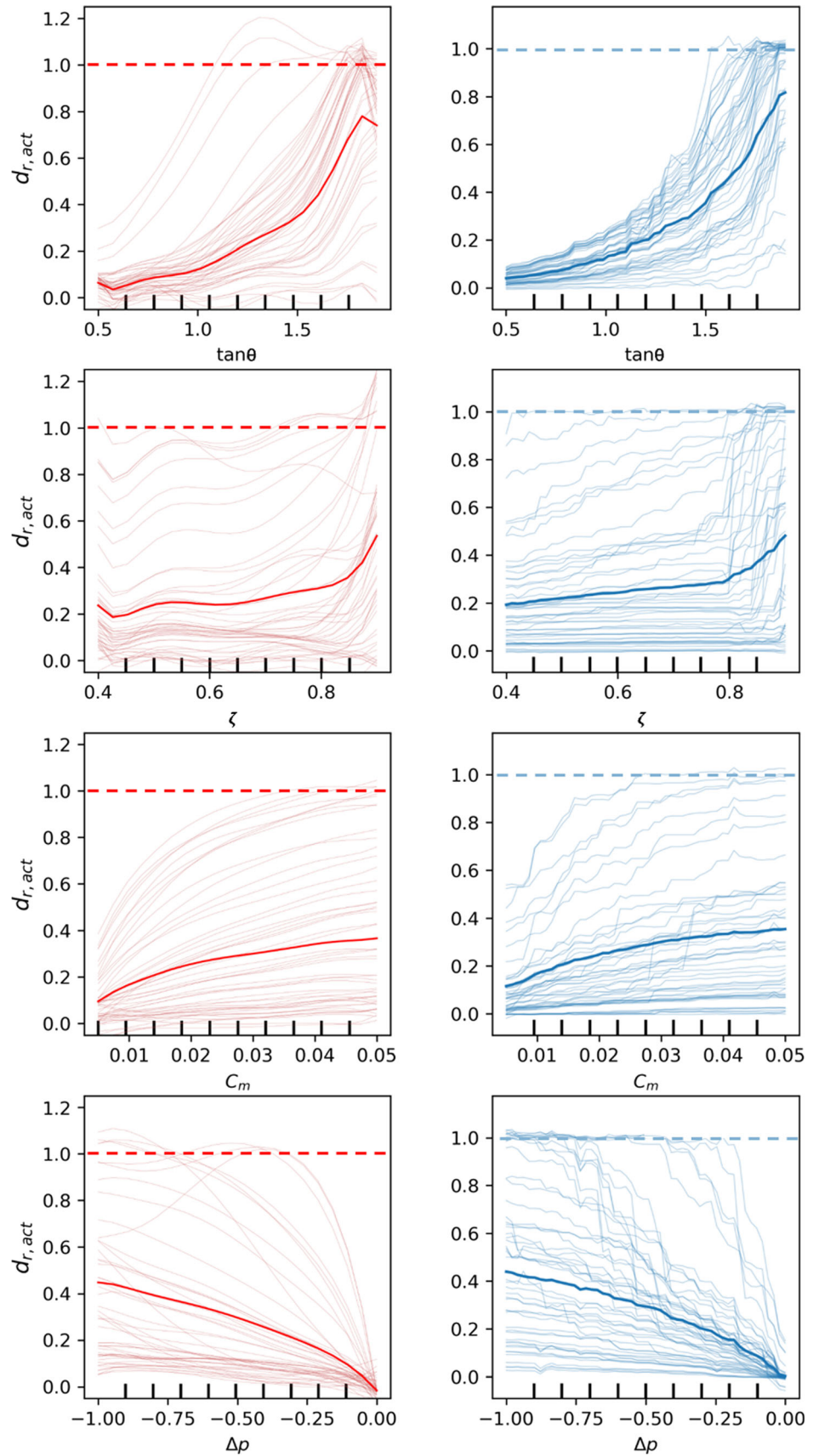
The impact is limited when  $\tan\theta < 1$ , however, once exceeding 1.47, the relative activated depth boosts

significantly. The model response remains almost constant until  $\zeta > 0.8$  when the average line slope that abruptly increases, although the contribution of thicker aquifers is still limited. According to the gradient variation,  $d_{r,act}$  is more sensitive with  $C_m$  up to  $0.02 \text{ MPa}^{-1}$  and a larger compressibility does not favor a much larger fissure development.  $\Delta p$  causes a relative larger output variation with respect to  $\zeta$  and  $C_m$ , which is consistent with the variable importance ranking. The slope of partial dependence gently decreases when the absolute value of  $\Delta p$  reduces.

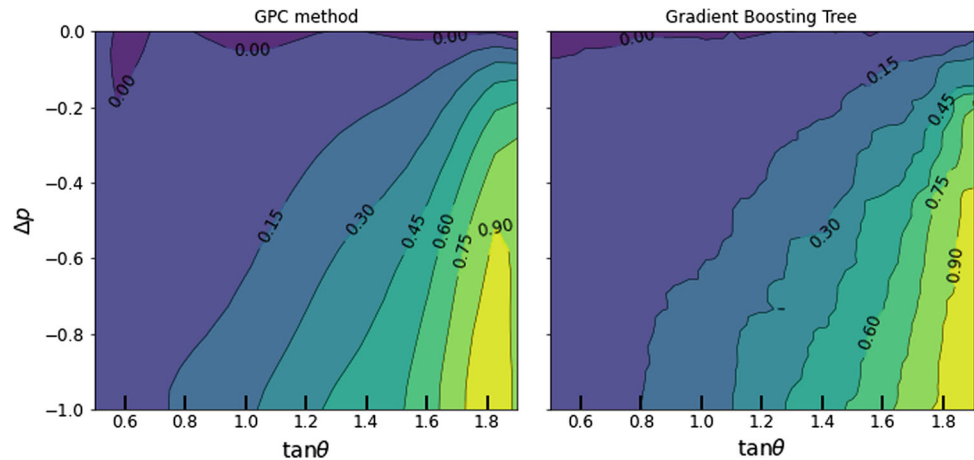
**Fig. 6** MDA metric with the GBT method. **a** Convergence of MDA and **b** comparison of total effect between Sobol’ and GBT methods. The importance indices are obtained from the sample size  $q = 7000$



**Fig. 7** Marginal effect of GPC (left panel) and GBT (right panel) methods on the model parameters  $\tan\theta$ ,  $\zeta$ ,  $C_m$ , and  $\Delta\rho$ . Each subplot contains 50 samples which are represented by the thin lines. Partial dependence is highlighted by the thick line



**Fig. 8** Partial dependence of  $\tan \theta$  and  $\Delta p$  with GPC (left panel) and GBT (right panel)



Based on these analyses, we conclude that the ridge geometry and the pressure variation are the first and secondary variables influencing fissure generation and propagation. Therefore, we plot the partial dependence of these two variables as shown in Fig. 8. The outcomes of two surrogate models are mainly consistent with respect to the model output distribution. In general, the size of ridge slope controls the upper limit of fissure opening. While a certain amount of  $\Delta p$  is necessary for the fissure occurrence.

## 6 Discussion

Earth fissuring accompanying differential subsidence above buried bedrock ridges is becoming a worldwide hazard. Since 1950s, this typical fissure occurrence has been reported, for example, in Casa Grande in Arizona, USA (Jachens and Holzer 1979), Yangzi Delta in China (Wang et al 2010), Najran Basin in Saudi Arabia (Youssef et al 2014). These studies have pointed out that fissure formation is induced by groundwater depletion and buried geological structures, but the undergoing physical process is not well known owing to little information and limited modelling technique.

The general consensus is that the pore pressure depletion causes a variation of the in-situ stress field, which is responsible for aseismic formation of earth fissures. Opening and sliding are induced by tensile and shearing, respectively (Hernandez-Marin and Burbey 2010; Budhu 2011). Here, we limit the investigation on the depth of the fissure opening that occurs when the stress normal component becomes greater than zero. This causes the vanishing of contact between the pair of surfaces constituting the IEs inserted above the ridge tip. However, we are aware of the possible formation of fissures due to sliding condition only, such as in the case addressed by Li et al (2021),

where multi-fissure formation is simulated for the hydro-geologic setting at Guangming village, Wuxi, China. These are more complex cases, where discontinuities develop where the stress field reaches the yield surface (Eq. (2)), and require an appropriate analyses of the stress field in the continuous body prior to insert the IEs in the most critical zones of the 3D mesh. Notice that in regions with bare land surface, e.g., in Arizona (Cook 2011, 2013), initial (thin) fissures caused by stress accumulation can be enlarged at the land surface by erosion and collapses due to surface flow during significant rainfall events. These processes are not accounted for in our modelling approach.

With more fissure appearances over the last decades, researches started focusing on the quantitative analyses of the fissure formation mechanism. Sheng et al (2003) defined the ratio of tensile stress over tensile strength as an indication for fissure inception and carried out an one-at-a-time sensitivity analysis which suggested the confining stress and, secondarily, the depth of aquifer as the key parameters for fissure formation. Unfortunately, the impact of the ridge geometry was not taken account into this analysis. Frigo et al (2019) applied a multivariate regression to fit the depth of fissure opening as function of the pressure variation and the ratio between exploited aquifer thickness and ridge tip depth. The regression surface consists of a pair of planes with discontinuous joint, highlighting that fissure is more prone to occur when the depth of the ridge tip is shallow. It is also found that pressure depletion plays an important role by controlling the differential subsidence. However, multivariate analysis standardizes regression coefficients as direct measures of sensitivity, which is more suitable for linear problems. Moreover, the number of evaluated variables was restricted to two in order to derive a regression surface in a 3D setting. Thus, the ridge depth and aquifer geometry were combined in a single parameter.

In this work, to systematically investigate the model sensitivity to input parameters and model geometry, we perform a global sensitivity analysis using a variance-based approach. Sobol' and total effects indices result in ranking the input factors, with a priority of importance assigned to the geometry of the ridge and the pore pressure drop of the system. The aquifer thickness and the compressibility are less influential with respect to the increase of the probability of fissure opening. The interactions between factors is one order of magnitude lower than the main indices, indicating a second order effect on the output variation. Results are mostly according to the mentioned previous studies, except for the less importance assigned to the aquifer thickness. However, this parameter was considered in a combined form with the ridge depth in Frigo et al (2019), probably causing an overestimation effect. Another reason may be attributable to the selection of the bounds values of the uniform distributions from which each parameter is sampled (Wagener and Pianosi 2019). For example, the slope of the bedrock ridge ( $\tan \theta$ ) is assumed variable between 0.5 and 1.9 due to some model grid constrains, discarding all  $\theta$  values lower than  $\sim 27^\circ$ . This may result in a not sufficiently wide choice of the parameter space, cutting out the possible influence of the other parameters at lower  $\theta$  values. We also point out that further analyses are needed to consider the possibility that variance is not a good measure of the output uncertainty, for example using indices that consider moment independence (Borogonovo 2007; Pianosi and Wagener 2015; Dell'Oca et al 2017).

We advocate the use of surrogate models to reduce the computational cost of the generation of the output samples for the computation of the importance metrics (Saltelli et al 2010). Surrogates based on GPC techniques are prominent because the easy derivation of the Sobol' indices at no additional computational burden. However, it is observed that increasing the level of problem non linearity, e.g., in proximity of the fissure opening, the method fails to provide a good model proxy (e.g., the predicted  $d_{r,act} > 1$ , see Fig. 4a). For this reason, we also employed the GBT algorithm for the estimation of total sensitivity measures, i.e., the mean decrease accuracy estimates MDA. Compared to Sobol' indices, MDA importance lacks of a straightforward interpretation, as they are computed based on the model prediction accuracy rather than the effects on the output variance. Thus, MDA is limited to assess the interaction effects between the input variables. However, this limit can be compensated by using other interpretability methods like SHAP (Lundberg and Lee 2017) or by using them as interpretation of the Sobol' total effects. The counterpart of this methodology, which seems more suitable for applications on discontinuous problems

compared to GPC, is that GBT regression tree spends more time on tuning hyperparameters to optimize the model performance, increasing the overall computational burden.

## 7 Conclusion

The present work investigates the relative importance of various hydrogeologic features to the formation and propagation of aseismic fissures above the crest of buried bedrock. The conceptual model used for numerical simulation is idealised and simplified from the field case reported in Wuxi, China. Earth fissures develop only with the simultaneous occurrence of an undulating bedrock that intercept a thick compressible aquifer where pressure decline takes place.

An advanced geomechanical simulator is used to analyse the stress field and quantify the fissure opening. The numerical simulations show how the pore pressure depletion results in the accumulation of tensile stress above the ridge tip, favoring the development of an earth fissure at the land surface. The fissure deepens as the pressure decline increases.

The numerical results are processed by GSA to assess the variable (bedrock ridge geometry, aquifer thickness and compressibility, and pore pressure variation) importance to fissure activation and propagation. Sobol' provided with Monte Carlo approach are taken as the reference and compared to Sobol' indices derived from surrogates of the forward model. GPC and GBT algorithms are applied to fit the numerical solution and then estimate the factor importance based on surrogate model prediction. The following main conclusions can be drawn:

- The three methods, i.e. Sobol', GPC, and GBT, rank the four variables consistently and provide similar importance measurements, thus supporting the validity of the achievement;
- The aquifer thickness and compressibility are less-influential variables to fissure opening;
- Marginal effect and surface response plots highlight that the probability of significant fissuring (deep and with large opening) is higher when the buried ridge is steeper and its tip closer to the land surface, with sufficiently large pore pressure depletion.

Finally, we have assessed the computational performances of three techniques on this application. Sobol' technique requires a larger sample size to converge, which makes it computationally expensive. Compared to GPC-based model, GBT performs better on approximating the discontinuous solution but requires a larger computational burden.

**Acknowledgements** The Authors wish to thank the University of Padova Strategic Research Infrastructure Grant 2017 “CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l’Innovazione” for the availability of the computational resources used in this work and acknowledge the partial support of the Hungarian Ministry of Innovation and Technology NRDI Office within the framework of the Artificial Intelligence National Laboratory Program and the Hungarian National Research, Development and Innovation Office (SNN 134368). The cooperation project “Inversion of SAR-based measurements to constrain land subsidence model in typical Beijing area” between the Capital Normal University (China) and the University of Padova (Italy) and the Project 41877180 of the National Natural Science Foundation of China are also kindly acknowledged. The first author was supported by the European Union PRIMA Programm under grant agreement No. 1924, project RESERVOIR.

**Funding** Open access funding provided by Università degli Studi di Padova within the CRUI-CARE Agreement.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arabameri A, Santosh M, Rezaie F, Saha S, Coastache R, Roy J, Mukherjee K, Tiefenbacher J et al (2022) Application of novel ensemble models and k-fold cv approaches for land subsidence susceptibility modelling. *Stoch Env Res Risk Assess* 36(1):201–223
- Biot MA (1941) General theory of three-dimensional consolidation. *J Appl Phys* 12(2):155–164
- Borgonovo E (2007) A new uncertainty importance measure. *Reliab Eng Syst Saf* 92(6):771–784
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Budhu M (2011) Earth fissure formation from the mechanics of groundwater pumping. *Int J Geomech* 11(1):1–11
- Budhu M, Shelke A (2008) The formation of earth fissures due to groundwater decline. In: Proceedings of the 12th International conference of international association for computer methods and advances in geomechanics (IACMAG), pp 3051–3059
- Burbey TJ (2002) The influence of faults in basin-fill deposits on land subsidence, Las Vegas Valley, Nevada, USA. *Hydrogeol J* 10(5):525–538
- Carreón-Freyre D, Cerca M, Ochoa-González G, Teatini P, Zuñiga FR (2016) Shearing along faults and stratigraphic joints controlled by land subsidence in the Valley of Queretaro, Mexico. *Hydrogeol J* 24:657–674. <https://doi.org/10.1007/s10040-016-1384-0>
- Carvalho DV, Pereira EM, Cardoso JS (2019) Machine learning interpretability: a survey on methods and metrics. *Electronics* 8(8):832
- Ciriello V, Di Federico V, Riva M, Cadini F, De Sanctis J, Zio E, Guadagnini A (2013) Polynomial chaos expansion for global sensitivity analysis applied to a model of radionuclide migration in a randomly heterogeneous aquifer. *Stoch Env Res Risk Assess* 27(4):945–954
- Conway BD (2016) Land subsidence and earth fissures in south-central and southern Arizona, USA. *Hydrogeol J* 24(3):649–655
- Cook JP (2011) Why did the fissure cross the roads? new and old earth fissure activity in Cochise County, Arizona. *Arizona Geol* 41(1):Future Article
- Cook JP (2013) Revisiting earth fissures near Luke Air Force Base, central Maricopa County, Arizona. Arizona Geological Survey Open File Report, OFR-13-15
- Copas JB (1983) Regression, prediction and shrinkage. *J Roy Stat Soc: Ser B (Methodol)* 45(3):311–335
- Couaillier V, Savin É (2019) Generalized polynomial chaos for non-intrusive uncertainty quantification in computational fluid dynamics. In: Uncertainty management for robust industrial design in aeronautics, pp 123–141. Springer
- Crestaux T, Le Maître O, Martinez JM (2009) Polynomial chaos expansion for sensitivity analysis. *Reliab Eng Syst Saf* 94(7):1161–1172. <https://doi.org/10.1016/j.ress.2008.10.008>
- Dell’Oca A, Riva M, Guadagnini A (2017) Moment-based metrics for global sensitivity analysis of hydrological systems. *Hydrol Earth Syst Sci* 21(12):6219–6234. <https://doi.org/10.5194/hess-21-6219-2017>
- Franceschini A, Ferronato M, Janna C, Teatini P (2016) A novel lagrangian approach for the stable numerical simulation of fault and fracture mechanics. *J Comput Phys* 314:503–521
- Franceschini A, Castelletto N, Ferronato M (2019) Block preconditioning for fault/fracture mechanics saddle-point problems. *Comput Methods Appl Mech Eng* 344:376–401
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* pp 1189–1232
- Friedman N, Zander E (2020) Sglib collection of tutorials. <https://ezander.github.io/ParameterAndFieldIdentification/>, accessed: 2021-09-23
- Friedman N, Zoccarato C, Zander E, Matthies HG (2021) A worked-out example of surrogate-based bayesian parameter and field identification methods. In: Chiachio-Ruano J, Chiachio Ruano M, Snkararaman S (eds) Bayesian inverse problems: fundamentals and engineering applications (1st ed.), CRC Press, pp 155–203. <https://doi.org/10.1201/b22018>
- Frigo M, Ferronato M, Yu J, Ye S, Galloway D, Carreón-Freyre D, Teatini P (2019) A parametric numerical analysis of factors controlling ground ruptures caused by groundwater pumping. *Water Resour Res* 55(11):9500–9518
- Garipov TA, Karimi-Fard M, Tchelepi HA (2016) Discrete fracture model for coupled flow and geomechanics. *Comput Geosci* 20(1):149–160. <https://doi.org/10.1007/s10596-015-9554-z>
- Ghanem RG, Spanos PD (1991) Stochastic finite elements: a spectral approach. Springer Verlag, New York. (Reedited by Dover Publications, Mineola, 2003)
- Goldstein A, Kapelner A, Bleich J, Pitkin E (2015) Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *J Comput Graph Stat* 24(1):44–65
- Herman J, Usher W (2017) Salib: An open-source python library for sensitivity analysis. *J Open Source Soft* 2(9):97. <https://doi.org/10.21105/joss.00097>



- Hernandez-Marin M, Burbey TJ (2010) Controls on initiation and propagation of pumping-induced earth fissures: insights from numerical simulations. *Hydrogeol J* 18(8):1773–1785
- Hernandez-Marin M, Burbey TJ (2012) Fault-controlled deformation and stress from pumping-induced groundwater flow. *J Hydrol* 428:80–93
- Iooss B, Le Maître P (2015) A review on global sensitivity analysis methods. In: *Uncertainty management in simulation-optimization of complex systems*, pp 101–122. Springer
- Isotton G, Teatini P, Ferronato M, Janna C, Spiezia N, Mantica S, Volonte G (2019) Robust numerical implementation of a 3d rate-dependent model for reservoir geomechanical simulations. *Int J Numer Anal Meth Geomech* 43(18):2752–2771
- Jachens RC, Holzer TL (1979) Geophysical investigations of ground failure related to groundwater withdrawal - Picacho Basin, Arizona. *Ground Water* 17(6):574–585
- Janna C, Castelletto N, Ferronato M, Gambolati G, Teatini P (2012) A geomechanical transversely isotropic model of the Po River basin using PSInSAR derived horizontal displacement. *Int J Rock Mech Min Sci* 51:105–118
- Jaxa-Rozen M, Kwakkel J (2018) Tree-based ensemble methods for sensitivity analysis of environmental models: a performance comparison with sobol and morris techniques. *Environ Modell Softw* 107:245–266. <https://doi.org/10.1016/j.envsoft.2018.06.011>
- Kaintura A, Dhaene T, Spina D (2018) Review of polynomial chaos-based methods for uncertainty quantification in modern integrated circuits. *Electronics* 7(3):30
- Karimi-Fard M, Durlofsky LJ, Aziz K (2003) An efficient discrete fracture model applicable for general purpose reservoir simulators. In: *SPE reservoir simulation symposium, society of petroleum engineers*, <https://doi.org/10.2118/79699-MS>
- Le Maître O, Knio O, Najm H, Ghanem R (2004) Uncertainty propagation using wiener-haar expansions. *J Comput Phys* 197(1):28–57. <https://doi.org/10.1016/j.jcp.2003.11.033>
- Li Y, Teatini P, Yu J, Franceschini A, Frigo M, Zoccarato C, Ye S (2021) Aseismic multifissure modeling in unfaulted heavily pumped basins: mechanisms and applications. *Water Resour Res* 57(10):e2021WR030, <https://doi.org/10.1029/2021WR030127>
- Liu Y, Zhang D, Gy Wang, Liu C, Zhang Y (2019) Discrete element method-based prediction of areas prone to buried hill-controlled earth fissures. *J Zhejiang Univ Sci A* 20(10):794–803
- Loupe G (2014) Understanding random forests: From theory to practice. arXiv preprint [arXiv:14077502](https://arxiv.org/abs/1407.7502)
- Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: *Proceedings of the 31st international conference on neural information processing systems*, pp 4768–4777
- Ochoa-González G, Carreón-Freyre D, Franceschini A, Cerca M, Teatini P (2018) Overexploitation of groundwater resources in the faulted basin of querétaro, mexico: A 3d deformation and stress analysis. *Eng Geol* 245:192–206
- Ochoa-González G, Carreón-Freyre D, Franceschini A, Cerca M, Teatini P (2018) Overexploitation of groundwater resources in the faulted basin of Querétaro, Mexico: A 3D deformation and stress analysis. *Eng Geol* 245:192–206. <https://doi.org/10.1016/j.enggeo.2018.08.014>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Peng J, Chen L, Huang Q, Men Y, Fan W, Yan J (2013) Physical simulation of ground fissures triggered by underground fault activity. *Eng Geol* 155:19–30. <https://doi.org/10.1016/j.enggeo.2013.01.001>
- Pianosi F, Wagener T (2015) A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environ Modell Softw* 67:1–11. <https://doi.org/10.1016/j.envsoft.2015.01.004>
- Probst P, Boulesteix AL, Bischl B (2019) Tunability: importance of hyperparameters of machine learning algorithms. *J Mach Learn Res* 20(1):1934–1965
- Roustant O, Fruth J, Iooss B, Kuhnt S (2014) Crossed-derivative based sensitivity measures for interaction screening. *Math Comput Simul* 105:105–118
- Saltelli A (2002) Sensitivity analysis for importance assessment. *Risk Anal* 22(3):579–590
- Saltelli A, Annoni P (2010) How to avoid a perfunctory sensitivity analysis. *Environ Modell Softw* 25(12):1508–1517
- Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010) Variance based sensitivity analysis of model output design and estimator for the total sensitivity index. *Comput Phys Commun* 181(2):259–270
- Sheng Z, Helm DC (1998) Multiple steps of earth fissuring caused by ground-water withdrawal. In: *Land subsidence case studies and current research: proceedings of the Dr. Joseph F. Poland Symposium on Land Subsidence, Assoc. Eng. Geol. Special Publication, vol 8*, pp 149–154
- Sheng Z, Helm DC, Li J (2003) Mechanisms of earth fissuring caused by groundwater withdrawal. *Environ Eng Geosci* 9(4):351–362
- Sobol' IM (1993) Sensitivity analysis for non-linear mathematical models. *Math Modell Comput Exp* 1:407–414
- Sobol' IM (2001) Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Math Comput Simul* 55(1–3):271–280
- Sobol' IM, Asotsky D, Kreinin A, Kucherenko S (2011) Construction and comparison of high-dimensional sobol' generators. *Wilmott* 56:64–79
- Sochala P, Le Maître O (2013) Polynomial chaos expansion for subsurface flows with uncertain soil parameters. *Adv Water Resour* 62:139–154. <https://doi.org/10.1016/j.advwatres.2013.10.003>
- Sudret B (2008) Global sensitivity analysis using polynomial chaos expansions. *Reliab Eng Syst Saf* 93(7):964–979
- Teatini P, Ferronato M, Gambolati G, Bertoni W, Gonella M (2005) A century of land subsidence in Ravenna, Italy. *Environ Geol* 47(6):831–846
- Vittekk M, Borovansky P, Moreau PE (2006) A simple generic library for c. In: *Reuse of off-the-shelf components. Proceedings of 9th International Conference on Software Reuse, Turin, Italy*, pp 423–426. Springer
- Wagener T, Pianosi F (2019) What has global sensitivity analysis ever done for us? a systematic review to support scientific advancement and to inform policy-making in earth system modelling. *Earth Sci Rev* 194:1–18. <https://doi.org/10.1016/j.earscirev.2019.04.006>
- Wang G, You G, Shi B, Qiu Z, Li H, Tuck M (2010) Earth fissures in jiangsu province, china and geological investigation of hetang earth fissure. *Environ Earth Sci* 60(1):35–43
- Wang Z, Zhang Y, Wu J, Yu J, Gong X (2015) Numerical simulation of earth fissures due to groundwater withdrawal. *Proc Int Assoc Hydrol Sci* 372:395–398
- Wiener N (1938) The homogeneous chaos. *Am J Math* 60(4):897–936. <https://doi.org/10.2307/2371268>
- Xiu D (2007) Efficient collocational approach for parametric uncertainty analysis. *Commun Comput Phys* 2(2):293–309
- Xiu D (2010) *Numerical methods for stochastic computations. A spectral method approach*. Princeton University Press, Princeton, New Jersey

- Ye S, Luo Y, Wu J, Yan X, Wang H, Jiao X, Teatini P (2016) Three-dimensional numerical modeling of land subsidence in Shanghai, China. *Hydrogeol J* 24(3):695–709
- Ye S, Franceschini A, Zhang Y, Janna C, Gong X, Yu J, Teatini P (2018) A novel approach to model earth fissure caused by extensive aquifer exploitation and its application to the Wuxi case, China. *Water Resour Res* 54(3):2249–2269
- Youssef AM, Sabtan AA, Maerz NH, Zabramawi YA (2014) Earth fissures in Wadi Najran, Kingdom of Saudi Arabia. *Nat Hazards* 71(3):2013–2027
- Zander E (2020) Sglib github repository. <https://github.com/ezander/sglib-testing>, accessed: 2021-09-23
- Zhu L, Franceschini A, Gong H, Ferronato M, Dai Z, Ke Y, Pan Y, Li X, Wang R, Teatini P (2020) The 3-D facies and geomechanical modeling of land subsidence in the Chaobai Plain. *Water Resour Res*, Beijing. <https://doi.org/10.1029/2019WR027026>
- Zoccarato C, Gazzola L, Ferronato M, Teatini P (2020) Generalized polynomial chaos expansion for fast and accurate uncertainty quantification in geomechanical modelling. *Algorithms* 13(7):156

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.