

# Sensitivity Analysis of Factors Controlling Earth Fissures Due To Excessive Groundwater Pumping

Yueting Li (✉ [yueting.li@studenti.unipd.it](mailto:yueting.li@studenti.unipd.it))

University of Padova

Noemi Friedman

Institute for Computer Science and Control (SZTAKI)

Pietro Teatini

University of Padova

Andras Benczur

Institute for Computer Science and Control (SZTAKI)

Shujun Ye

Nanjing University

Lin Zhu

Capital Normal University

Claudia Zoccarato

University of Padova

---

## Research Article

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

**Posted Date:** January 25th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1232687/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

1 **Sensitivity analysis of factors controlling earth fissures**  
2 **due to excessive groundwater pumping**

3 **Yueting Li · Noemi Friedman · Pietro**  
4 **Teatini · Andras Benczur · Shujun Ye ·**  
5 **Lin Zhu · Claudia Zoccarato**

6 Received: date / Accepted: date

7 **Abstract** Aseismic earth fissures are complex consequences of groundwa-  
8 ter withdrawal and natural hydrogeologic conditions. This paper aims to im-  
9 prove the understanding of the mechanism of earth fissuring and investi-  
10 gate the relative importance of various factors to fissure activity, including  
11 bedrock geometry, piezometric depletion, compressibility and thickness of the  
12 exploited aquifer. For these purposes, a test case characterized by an imper-  
13 meable and incompressible rock ridge in a subsiding basin is developed, where  
14 stress/displacement analyses and fissure state are predicted using an interface-  
15 finite element model. Three different methods for global sensitivity analysis are

---

Yueting Li  
Department of Civil, Environmental and Architectural Engineering, University of Padova,  
Padova, Italy  
School of Earth Sciences and Engineering, Nanjing University, Nanjing, China  
E-mail: yueting.li@phd.unipd.it

Pietro Teatini · Claudia Zoccarato  
Department of Civil, Environmental and Architectural Engineering, University of Padova,  
Padova, Italy

Noemi Friedman · Andras Benczur  
Informatics Laboratory, Institute for Computer Science and Control (SZTAKI)

Shujun Ye  
School of Earth Sciences and Engineering, Nanjing University, Nanjing, China

Lin Zhu  
College of Resource Environment and Tourism, Capital Normal University, Beijing, China

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;

42 Ochoa-González et al, 2018). Different hydrogeological settings favoring the  
43 occurrence of earth fissures have been conceptualized based on field studies,  
44 including buried undulating bedrock, pre-existing fault and abrupt heteroge-  
45 neous thickness of aquifer (Sheng and Helm, 1998; Sheng et al, 2003). The  
46 features of the earth fissures such as density, shape, length, aperture, depth,  
47 and dislocation vary greatly in different settings, which also implies different  
48 driving mechanisms.

49 This work aims to improve the understanding of the earth fissure mecha-  
50 nism with the presence of buried bedrock ridges in subsiding basins. Knowl-  
51 edge of the mechanisms driving these hazards may help to predict and there-  
52 fore limit significant damages to buildings, streets, highways, railroads, earth  
53 dams, water wells, and other engineering structures. Note that earth fissures  
54 that coincide with fault scarps and abrupt thickness change may also be re-  
55 lated to seismicity (Carreón-Freyre et al, 2016; Peng et al, 2013), but this is  
56 beyond the scope of this work.

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

66 The complexity of these systems typically give rise to many uncertainties  
67 due to the geologic configuration, the pore-pressure distribution, the hydro-  
68 geomechanical parameters along with the mathematical and numerical ap-  
69 proximation of the physical problem (Frigo et al, 2019; Sheng et al, 2003). In

70 this context, a global sensitivity analysis (GSA) is fundamental to evaluate  
71 the susceptibility of input variables to fissure formation and propagation, con-  
72 sidering their possible mutual interactions (Iooss and Le Maître, 2015; Saltelli  
73 and Annoni, 2010). A variance-based GSA is employed based on the func-  
74 tional decomposition of the output variance, providing the Sobol’ indices that  
75 quantify the input contribution to the output variance (Sobol’, 1993, 2001).

76 First, we compute the indices using an efficient Monte Carlo sampling de-  
77 sign, employing the Sobol’ sequence to generate a uniformly distributed sample  
78 over the uncertain input domain (Sobol’ et al, 2011). However, a large number  
79 of samples are needed, in particular when interaction factors are investigated.  
80 This means a computationally prohibitive cost for large scale models, as it is  
81 the case of earth fissure modelling. For this reason, the use of surrogate (or  
82 proxy) models, which are approximations of the forward model built from a  
83 limited number of runs of the full model, is seen as a prominent approach to  
84 reduce the overall computational cost of the sensitivity analysis. Among sur-  
85rogate methods, polynomial chaos expansion (GPC) is a probabilistic method  
86 which uses orthogonal polynomial projections of the input random variables to  
87 build the stochastic model output (Ghanem and Spanos, 1991). This technique  
88 provides a straightforward way to derive Sobol’ indices from model represen-  
89tation coefficients (Crestaux et al, 2009). Thanks to these advantages, GPC  
90 surrogates have been recently applied for GSA in environmental modelling  
91 (Sochala and Le Maître, 2013; Couaillier and Savin, 2019; Kaintura et al,  
92 2018; Zoccarato et al, 2020; Friedman et al, 2021).

93 However, difficulties may rise when the quantity of interest presents some  
94 discontinuities with respect to the model parameters (Sochala and Le Maître,  
95 2013; Le Maître et al, 2004). In case of earth fissuring simulation, this oc-  
96 currence occurs when the discontinuity develops within the continuous porous  
97 medium. To overcome this problem, we elected to employ a decision tree-based

98 method such as the gradient boosting tree (GBT) that uses an ensemble of  
99 decision trees to approximate the solution, in particular for non-linear models  
100 with arbitrary inputs (Friedman, 2001; Louppe, 2014). Although tree-based  
101 models are considered as “black box”, many interpretation methods, such as  
102 Shapley Additive Explanations and Mean Decrease Accuracy (MDA) were  
103 designed to assess feature importance according to their relevance for the cor-  
104 responding estimator, similarly to the key insights of GSA (Breiman, 2001;  
105 Carvalho et al, 2019; Lundberg and Lee, 2017).

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

## 113 **2 Numerical model**

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

### 119 **2.1 Continuum model**

120 Stress and strain fields caused in a 3D continuous porous medium  $\Omega$  are quanti-  
121 fied by means of the classical poroelasticity theory (Biot, 1941). The governing

122 equations read:

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

123 where  $\boldsymbol{\sigma} = (\mathbf{C} : \nabla^s \mathbf{u} - b p \mathbf{1})$  is the total Cauchy stress tensor, with  $\nabla^s =$   
 124  $\frac{1}{2} (\nabla + \nabla^T)$  the symmetric gradient operator,  $\mathbf{C}$  the rank-4 elasticity tensor,  
 125  $b$  the Biot coefficient, and  $\mathbf{1}$  the rank-2 identity tensor. The displacement field  
 126  $\mathbf{u}$  is the primary unknown. The pore pressure  $p$  is a known forcing term,  
 127 either imposed according to previous physical knowledge and measurements  
 128 or provided by a groundwater flow model (Ye et al, 2018).

129 Without loosing the general validity of the approach, the constitutive re-  
 130 lationship between stresses and strains used in this study is assumed linear  
 131 elastic. Consequently, the soil compressibility  $C_m$  and Poisson ratio  $\nu$  are con-  
 132 stant and do not vary with the pressure (i.e., stress) change. Moreover, as  
 133 usually implemented in the geomechanical application related to aquifer over-  
 134 exploitation (Hernandez-Marin and Burbey, 2012; Ye et al, 2016; Zhu et al,  
 135 2020), the small strain hypothesis is adopted. A tetrahedral finite element  
 136 (FE) discretization is used.

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

## 141 2.2 Earth fissure (EF) model

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

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

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

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

159 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  
 160 tangential and normal components, respectively. The unit vector  $\mathbf{n}$  denotes  
 161 the normal vector for the surface pair  $\Gamma_f^1$  and  $\Gamma_f^2$ . In the Coulomb criterion,  
 162  $c$  and  $\varphi$  are the cohesion and friction angle, respectively. The impenetrability  
 163 of solid bodies is prescribed by the normal contact condition:

$$t_N g_N = 0 \quad (3)$$

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

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

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

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

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

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

### 184 3 Sensitivity Analysis

185 A sensitivity analysis framework is implemented to investigate the most im-  
 186 portant factors (and their interactions) controlling earth fissuring in subsid-  
 187 ing basis. In this section, we first provide the mathematical framework of  
 188 two different types of surrogate models (GPC and GBT) used to reduce the  
 189 computational burden by approximating the full forward model. Then, Sobol'  
 190 indices for sensitivity analysis are introduced with specific reference to their  
 191 numerical computation based on Sobol' and GPC approaches. For GBT, the

192 mean decrease accuracy metric (MDA) is presented as a measure of the rela-  
193 tive factor importance. The notion of partial dependence is also introduced to  
194 characterize the dependence of the model response on individual factors.

### 195 3.1 Surrogate Models

#### 196 3.1.1 Generalized polynomial chaos expansion (GPC)

197 Running the forward geomechanical model multiple times for large and com-  
198 plex systems can be a very demanding task, both in terms of CPU and mem-  
199 ory requirements. A GPC approach (Wiener, 1938; Xiu, 2007) is therefore pro-  
200 posed to approximate the outcome of the deterministic simulator as a function  
201 of the uncertain input parameters with the help of polynomials. With such ap-  
202 proximation, propagation of the input uncertainties to the model output can  
203 be efficiently computed and statistics such as mean, variance, and quantiles  
204 can be easily determined.

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

216 As usual, the independence assumption implies  $F_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^n F_{Z_i}(z_i)$ ,  
217 where  $F_{Z_i}(z_i) = \mathcal{P}(Z_i \leq z_i)$  is the marginal distribution function with  $i =$

218  $1, \dots, n$ . Since any random variable may be represented as a series of polyno-  
 219 mials in uncorrelated and independent Gaussian variables (Wiener, 1938) and,  
 220 in its generalized extension, in non-Gaussian measures, GPC basis functions  
 221 of a univariate random variable  $Z_i$  are defined as the polynomials  $\{\phi_k(Z_i)\}_{k=0}^N$   
 222 of  $N$ -th-degree satisfying the orthogonality conditions:

$$\mathbb{E}[\phi_s(Z_i)\phi_r(Z_i)] = \int_{\Sigma_i} \phi_s(z_i)\phi_r(z_i)dF_{Z_i}(z_i) = \gamma_s\delta_{s,r} \quad 0 \leq s, r \leq N \quad (5)$$

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

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

227 where  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$  is a multi-index with  $|\alpha| = \alpha_1 + \dots + \alpha_n$ . The  
 228 multivariate basis functions are orthogonal polynomials in  $L^2_{dF_{\mathbf{z}}}$ , that is, the  
 229 space of all mean-square integrable functions of  $\mathbf{Z}$  with respect to the inner  
 230 product based on the measure  $dF_{\mathbf{Z}}$ :

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

231 where  $\Sigma$  is defined by  $\Sigma = \Sigma_1 \times \Sigma_2 \dots \times \Sigma_n$ . As a consequence, the class of  
 232 orthogonal polynomials is selected according to the measure  $F_{Z_i}$ .

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

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

236 where  $c_{\alpha}$  are the coefficients of the expansion. For  $U(\mathbf{Z}) \in L^2_{dF_Z}$ , the coeffi-  
 237 cients  $c_{\alpha}$  can be computed by defining  $\tilde{U}_{\text{GPC},N}$  as the orthogonal projection of  
 238  $U$  onto the polynomial space  $\mathcal{Z} = \text{span}\{\Phi_{\alpha}\}$ . By prescribing the orthogonality  
 239 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:

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

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

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

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

249 with  $\mathbf{z}^j$  and  $w(\mathbf{z}^j)$  the  $q$  integration nodes and weights, respectively. Since  
 250  $\Phi_{\alpha}$  is at most of degree  $N$ , the integrand function has at most degree  $2N$ .  
 251 In the univariate case, this requires the use of a  $(q = N + 1)$ -point Gaussian  
 252 quadrature rule, while in the multivariate case with  $n$  random variables the  
 253 number of points grows up to  $q = (N + 1)^n$ . Using this approximation, the  
 254 surrogate model needs the evaluation of  $U$  through the numerical solver of the  
 255 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

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

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

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

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

### 259 3.1.2 Gradient Boosting Tree (GBT)

260 Convergence of the GPC approximation is especially favorable when the de-  
261 pendence of the model output on the given uncertain input parameters have  
262 sufficient smoothness. As this is not necessarily the case for the given geome-  
263 chanical model, we have also tested a gradient boosting tree (GBT) approx-  
264 imation of the model output to explore the relationship between the input  
265 parameters and the model output.

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

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

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

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

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:

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

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  $\mathbf{1}_{R_{im}}$  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  $(\mathbf{z}^j, r_{jm})$ :

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

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

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

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

275 Therefore, the model can be updated by:

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

276 Moreover, regularization methods impose the constrains on fitting pro-  
 277 cedure to prevent the overfitting, that is when the surrogate model exactly  
 278 describes the training data and fails to fit unseen data. For example, the max-  
 279 imum stage of gradient boosting  $M$  in Equation (15) is a natural regular-  
 280 ization parameter which discourages learning more complex model to avoid  
 281 overfitting. However, it has been found that regularization through shrinkage  
 282 provides superior results to that obtained by restricting the maximum stage  
 283 (Copas, 1983), hence a simple shrinkage strategy is added to Equation (20):

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

284 Under this form, two regularization parameters are used in the gradient boost-  
 285 ing algorithm: the learning rate  $\nu$  and the number of boosting stage  $M$ .

## 286 3.2 Variable importance metrics

### 287 3.2.1 Variance-based Sobol' indices

288 Consider the model under investigation is described as a function  $U = f(\mathbf{Z})$ ,  
 289 where  $U$  is a scalar and the input  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$  is defined over the n-  
 290 dimensional hypercube  $I^n$  with mutually independent components. Assuming  
 291  $U$  to be a square integrable function, the Sobol' functional decomposition  
 292 scheme reads:

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

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

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

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

297 The decomposition allows to attribute the variances to the different param-  
 298 eters or their various degree interactions. These can be given by the partial  
 299 variances:

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

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

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

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

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

305 and similarly for the higher order sensitivity indices. The first order indices  
 306  $\{S_i\}_{i=1}^n$  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:

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

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  $S_{T_i}$  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  $\mathbf{A}_{\mathbf{B}^i}$  by taking matrix  $\mathbf{A}$  and replacing its  $i$ -th column with the corresponding column of  $\mathbf{B}$ . The estimators:

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

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

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  $f(\mathbf{A}_{\mathbf{B}^i})$ . Instead, by using the GPC or GBT

327 surrogates as model proxy, the model output can be directly computed from  
 328 the proxy model in a computationally cheap manner.

329 One of the great advantages of the GPC surrogate model is that the Sobol'  
 330 indices can be computed analytically without using the sampling based ap-  
 331 proximation given in Equation (28). Due to the orthogonality condition, the  
 332 mean and the total variance of the GPC can be directly computed from the  
 333 coefficients of the expansion as follows:

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

334

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

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

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

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

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

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

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

338 The partial variances can be also easily computed from:

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

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

### 343 3.2.2 Mean Decrease Accuracy (MDA)

344 For the GBT surrogate model instead of computing the Sobol' indices with  
 345 the help of the MC estimation, we use a different sensitivity measure that  
 346 can be efficiently computed by GBT models. Each feature importance is here  
 347 evaluated through a permutation-based measure following the idea of Breiman  
 348 (2001) and the application in Jaxa-Rozen and Kwakkel (2018). Given the  $q \times n$   
 349 matrix  $\mathbf{A}$  of the random input variables  $\mathbf{Z}$ , the MDA index of the  $i$ -th feature  
 350 measures the decrease of the estimator accuracy by randomly permuting the  
 351 values of  $Z_i$  ( $i$ -th column of input variables matrix) for  $K$  times in total and  
 352 for each repetition re-computing the ensemble tree predictions with the  $k$ -th  
 353 (for  $k = 1, \dots, K$ ) permuted column  $\mathbf{A}_i$ . The higher the inaccuracy, the most  
 354 important is the feature for the particular model. MDA of the  $i$ -th feature is  
 355 defined as

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

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

### 360 3.3 Convergence criterion of importance indices

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

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

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

### 373 3.4 Partial Dependence

374 Compared to Sobol' indices and MDA, partial dependence is more similar to  
 375 one-at-a-time (OAT) sensitivity analysis, which assumes the model response is  
 376 a function of one or two input variables and characterizes the average marginal  
 377 effect on model prediction (Goldstein et al, 2015). Owing to this feature, partial  
 378 dependence plot can visually depicts the relationship between model prediction

379 and the variables of interest. The partial dependence function  $\tilde{U}_i$  reads:

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

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

### 384 3.5 Software availability

385 In this work, the forward geomechanical models are carried out by the GEPS3D  
 386 simulator (Isotton et al, 2019; Franceschini et al, 2016). The reference global  
 387 sensitivity analysis, that is Sobol' technique, is implemented by the *SALib* li-  
 388 brary in Python environment (Herman and Usher, 2017). *SGLib* library (Zan-  
 389 der, 2020; Friedman and Zander, 2020) is used to compute the polynomial  
 390 chaos expansion and Sobol' indices (Vittekk et al, 2006). Gradient boosting  
 391 tree algorithm and mean decrease accuracy computation are carried out by  
 392 the *scikit - learn* module in Python with gradient boosting regression esti-  
 393 mator and permutation feature importance function (Pedregosa et al, 2011).

## 394 4 Model Setup and Parameterization

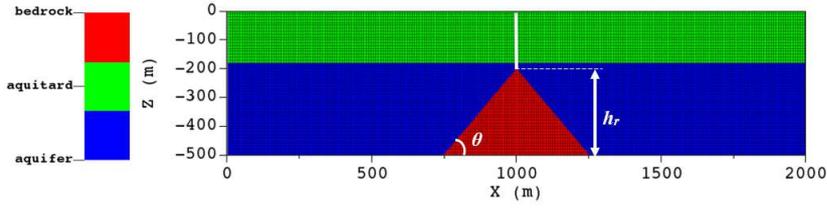
395 The investigated configuration conceptualizes the geological setting in Wuxi,  
 396 China, where the compressible deposits of the Yangtze River cover an undu-  
 397 lating bedrock.

398 The numerical simulation is developed on a quasi-3D domain 2000-m long  
 399 ( $x$ -direction), 50-m thick ( $y$ -direction), and 500-m deep ( $z$ -direction). A traction-  
 400 free top surface and a fixed bottom surface are considered (Fig.1). On the

401 lateral surfaces the horizontal displacements are precluded in the orthogo-  
402 nal direction. The conceptual model is composed of three hydrostratigraphic  
403 units: an upper aquitard, a bottom aquifer, and a buried triangular bedrock.  
404 For sake of simplicity, each material is assumed to behave elastically with the  
405 same Poisson ratio  $\nu = 0.25$ . Cohesion  $c$  and friction angle  $\varphi$  are set equal  
406 to 0.01 MPa and  $30^\circ$  respectively. A piezometric drop  $\Delta p$  is uniformly as-  
407 signed to the bottom aquifer, meanwhile the upper aquitard is regarded as an  
408 hydraulically “inactive” unit where the pore pressure propagation from the  
409 underlying sandy layer is negligible. The initial stress field is computed based  
410 on the gradient density ( $\sigma_v = 1200 \text{ kg/m}^2/\text{m}$ ) and the minimum-to-maximum  
411 stress ratio reads  $\sigma_h/\sigma_v = \nu/(1 - \nu)$ .

412 The vertical size of tetrahedral FE elements is 10 m and the horizontal  
413 dimension is in range between 5.5 and 20 m, slightly varying according to  
414 the ridge geometry. Previous studies have proved that earth fissures are most  
415 prone to generate from the ground surface above the ridge tip downward, with  
416 a stress state characterized by high tension above the ridge (Ye et al, 2018;  
417 Frigo et al, 2019). Moreover, fissures cannot propagate within the bedrock  
418 where pressure does not change and the stress field variation is negligible.  
419 Therefore, a IE alignment is vertically introduced from the land surface to  
420 the ridge tip as highlighted by the white line in Fig. 1. The triangular IE dis-  
421 cretization is consistent with FE discretization. Stress distribution and mag-  
422 nitude depend on the ridge geometry, the aquifer thickness, and differential  
423 subsidence. This latter is primarily dependent on the pore pressure change and  
424 sediment compressibility. Therefore, these four variables, i.e. ridge geometry,  
425 aquifer thickness, pressure change and sediment compressibility, are selected  
426 as input features for GSA.

427 Here, the ridge geometry is characterized by the slope of the bedrock ridge  
428 ( $\tan \theta$ ). Note that the length of ridge basement is fixed at 500 m. The frac-



**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.

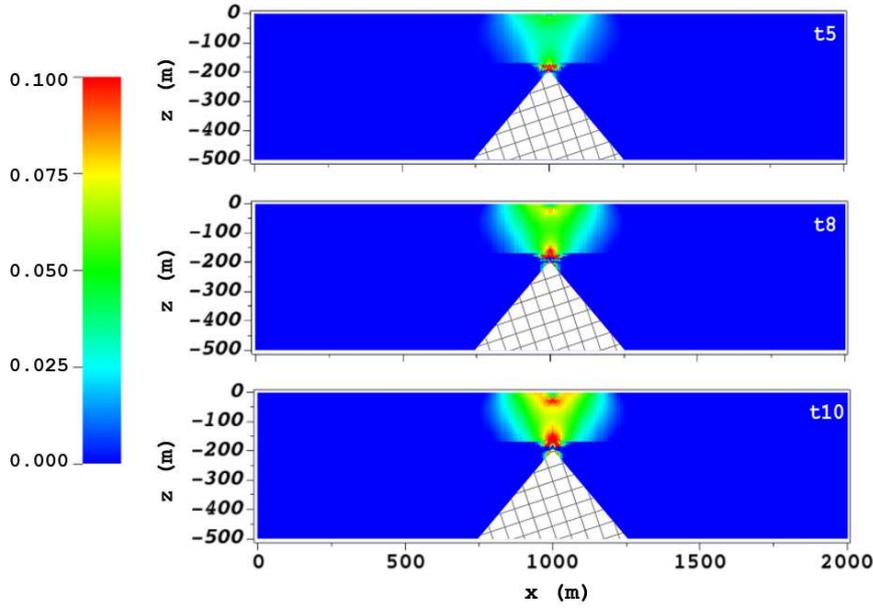
Feature	Min	Max
$\tan\theta$	$5.0 \times 10^{-1}$	1.9
$\zeta$	$4.0 \times 10^{-1}$	$9.0 \times 10^{-1}$
$C_m$ (MPa $^{-1}$ )	$5.0 \times 10^{-3}$	$5.0 \times 10^{-2}$
$\Delta p$ (MPa)	-1.0	0.0

tion 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 dimen-  
sions 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 time steps. The variability ranges  
of the parameters used in this study are summarized in Tab. 1. A uniform  
probability distribution is assumed for each variable.

## 5 Results

### 5.1 Deterministic model run

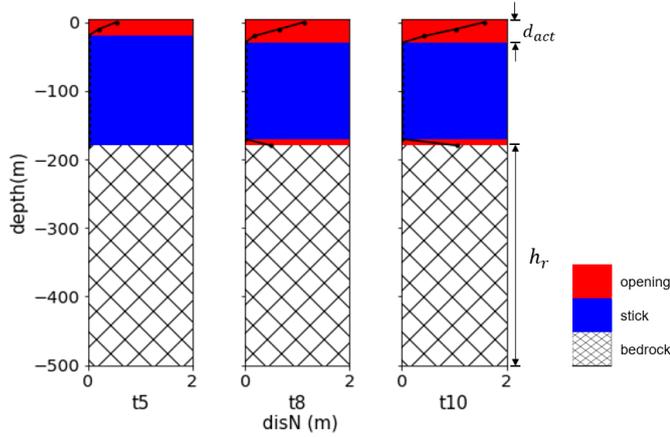
We randomly choose one experiment designed for GSA to present the numer-  
ical outcomes. The simulated temporal evolution of tensile stress on a vertical



**Fig. 2** Sequential evolution of the dimensionless horizontal stress  $\sigma_h^* = \sigma_h / \Delta p$  at the 5th, 8th and 10th loading steps 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  $\Delta p = -0.89 \text{ MPa}$ .

442 section of the domain is shown in Fig. 2. Tensile stress  $\sigma_h$  initially accumulates  
 443 around the ridge tip. As piezometric level declines, a tensile zone also occurs  
 444 at the land surface above the apex of ridge. Once tensile stress exceeds the  
 445 tensile strength, i.e.  $t_N = 0$  (see Section 2.2), IEs change from a stick to an  
 446 open state with the discontinuity that develops at the land surface and prop-  
 447 agates downwards. Simultaneously, tensile stress dissipates due to the crack  
 448 opening. Notice that the porous medium directly above the ridge tip does not  
 449 experience any shear stress due to the symmetric configuration. Therefore,  
 450 only fissure opening develops with this setting.

451 Fig. 3 shows the evolution of the earth fissure as provided by the IE model.  
 452 Fissure initially originates at the land surface where the tensile strength is the  
 453 lowest and later develops at depth too, just above the ridge tip, in response to  
 454 the concentration of tensile stress. The upper fissure keeps enlarging horizon-



**Fig. 3** Sequential evolution of fissure opening at the 5th, 8th and 10th loading steps: opening  $g_N$  (dotted black line), opening  $I_{open}$  and close  $I_{stick}$  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  $\Delta p = -0.89 \text{ MPa}$ .

455 tally and extending downwards. At the 10th step, the fissure reaches a depth  
 456 about 30 m with a maximum opening equal to 1.7 m. Conversely, the bottom  
 457 discontinuity remains confined at the depth without a significant development.

458 Generally, the energy is dissipated after fissuring mainly at land surface.  
 459 Moreover, the overburden stress due to sediment load usually limits tensile  
 460 fissuring at depth (Budhu and Shelke, 2008). Therefore, the bottom activated  
 461 zone is not included in the quantity of interest  $d_{act}$  that is defined as the fissure  
 462 depth from the land surface. It is worth mentioning that the size of activated  
 463 depth is controlled by the vertical length of the IE alignment, thus the relative  
 464 activated depth  $d_{r,act} = d_{act}/(500 - h_r)$  is introduced to have comparable  
 465 results when varying the model geometry (Fig. 1). In this case,  $d_{r,act}$  equals  
 466 0.132.

## 467 5.2 Approximation by surrogate models

### 468 5.2.1 GPC surrogate

469 The computation of the GPC coefficients has been initially carried out by the  
 470 pseudospectral approach (Equation (11)) and increasing step by step the GPC  
 471 degree of the polynomial response. The surrogate model  $\tilde{U}_{\text{GPC}}$  is developed to  
 472 approximate  $d_{r,act}$  with input parameters  $\mathbf{Z} = \{\tan \theta, \zeta, C_m, \Delta p\}$ . The valida-  
 473 tion of the fitted surrogate model is carried out by employing 7000 samples,  
 474 that is the available set of points used to compute Sobol' indices with the quasi  
 475 Monte Carlo approach. Moreover, train and validation of the GPC surrogate is  
 476 also obtained from the 7000 points using 80% for regression and the remaining  
 477 20% for validation.

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

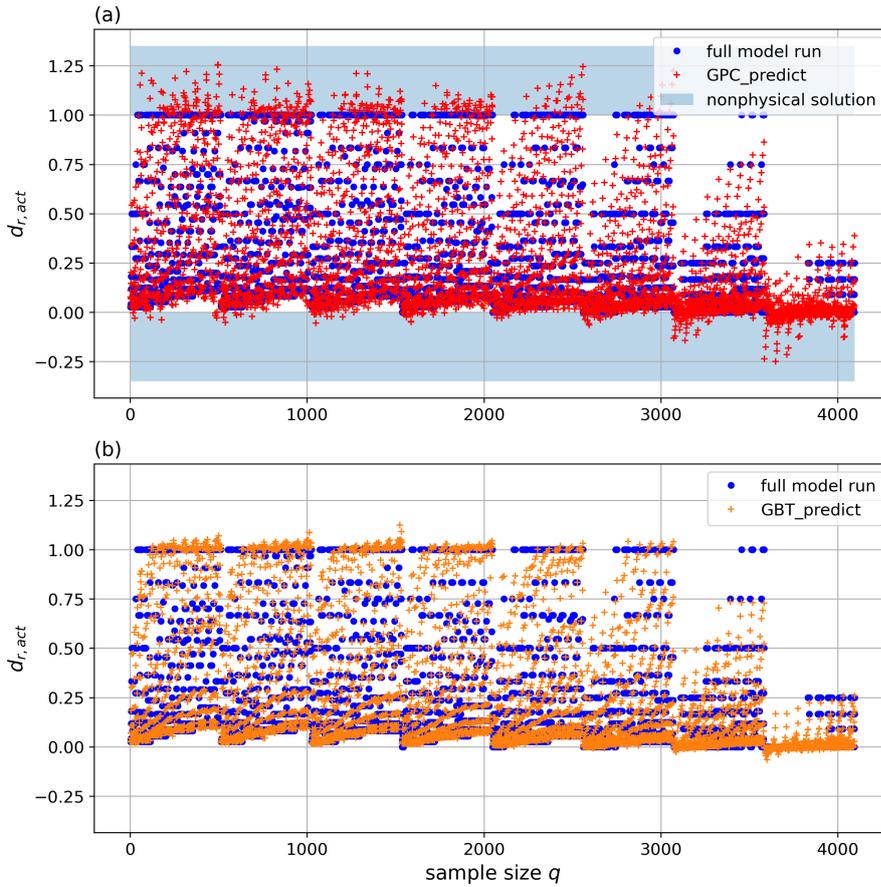
**Table 2** Validation of the GPC at increasing value of the maximum total degree of the GPC expansion. The coefficient of determination  $R^2$  is used to assess the goodness of fit.  $R_{PS}^2$  refers to GPC coefficients computed using the pseudospectral approach through Gauss quadrature, whereas  $R_{RG}^2$  is obtained by regression. The number of points indicated are those used to train the surrogate models.

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

### 492 5.2.2 GBT surrogate

493 Gradient boosting algorithm is also implemented with increasing size of input  
 494 data to check the convergence, thereof 80% is used to fit model with remaining  
 495 20% for validation. Hyperparameter tuning is carried out by a Grid Search  
 496 method which enumerates all the possible combinations of hyperparameters  
 497 and gets optimal values based on the corresponding coefficient of determination  
 498  $R_{GB}^2$ . In this application, only the learning rate  $\nu$  has been tuned given it's  
 499 the most important hyperparameter for GBT estimator (Probst et al, 2019).

500 Tab. 3 shows the model goodness that stabilizes when the data size reaches  
 501 5000. In addition to Sobol' samples, the Gauss points (4096) for the GPC  
 502 method are also used to validate the regression tree obtained from maximum  
 503 sample size (7000) (Fig. 4(b)). The regression tree also has some nonphysical  
 504 predictions ( $d_{r,act} > 1.0$ ), however the absolute values of discrepancy are much  
 505 less than that of GPC solutions. Moreover,  $R_{GB}^2$  suggests GBT algorithm out-  
 506 performs GPC algorithm with respect to the prediction accuracy ( $R_{GB}^2=0.96$   
 507 vs  $R_{RG}^2=0.79$ ). Nevertheless, both two surrogate models fail to capture the  
 508 characteristic that the fissure opening depth  $d_{r,act}$  keeps constant within some  
 509  $q$  sub-domains irrespective of parameters variation.



**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.

### 510 5.3 Importance metrics

511 Sobol' technique is chosen as the "reference" sensitivity analysis with the input  
 512 samples generated by the Sobol' sequence. As mentioned in Section 5.2, these  
 513 samples are reused to train and validate GPC and GBT surrogates. Know-  
 514 ing that the reliability of sensitivity measures obtained by metamodels are  
 515 dependent on their predictive power, we can assure GPC and GBT are good

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

sample size $q$	learning rate $\nu$	$R_{GB}^2$
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

516 surrogates and the “overfitting” problem can be excluded according to the  $R^2$   
 517 values obtained from the cross validation.

518 Fig. 5 shows that Sobol’ and GPC methods estimate similar results of first  
 519 order indices  $S_i$  and total Sobol’ indices ( $S_{Ti}$ ), with the GPC algorithm show-  
 520 ing a quicker convergence with respect to the quasi Monte Carlo method. For  
 521 each variable,  $S_i$  accounts for the larger proportion of the corresponding  $S_{Ti}$ ,  
 522 indicating a minor contribution from interactions between the  $i$ -th variable  
 523 and the other input factors. The second order indices from Sobol’ technique  
 524 and GPC, both computed with 7000 samples, are listed in Tab. 4. It is evi-  
 525 dent the small interaction between input factors. Note that Sobol’ technique  
 526 gives some negative indices for the non-influential terms indicating some com-  
 527 putation errors which could not be eliminated with the current sample size  
 528 (Herman and Usher, 2017).

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

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

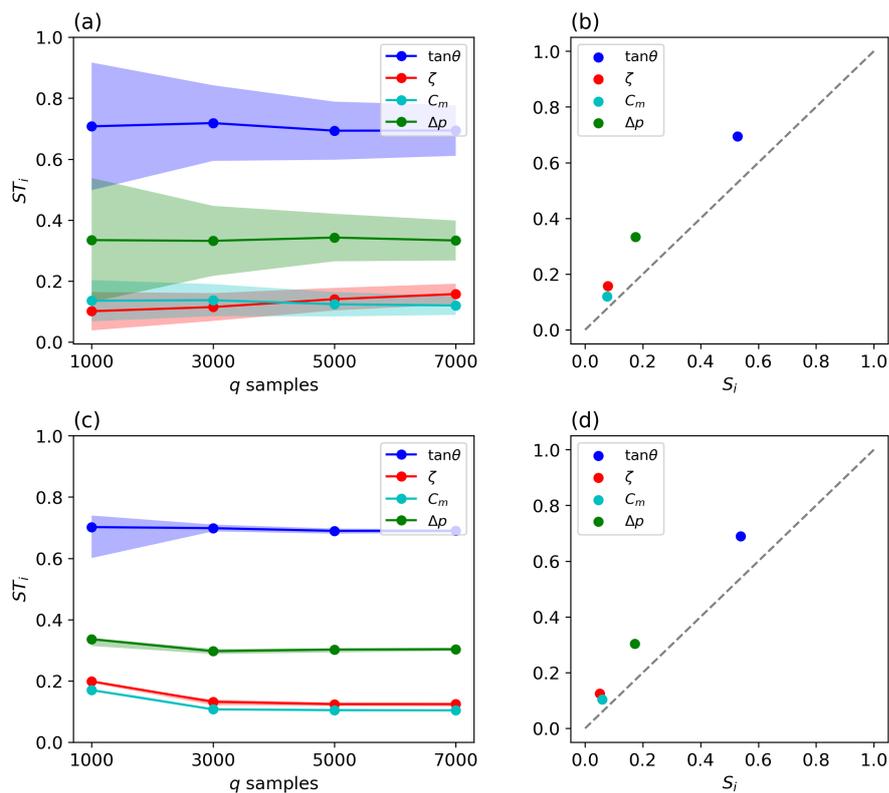
-	$S_{12}$	$S_{13}$	$S_{14}$	$S_{23}$	$S_{24}$	$S_{34}$
Sobol'	0.013	0.033	0.059	-0.005	-0.005	0.010
GPC	0.026	0.020	0.070	0.002	0.017	0.007

537 can reduce the overall computational cost of analysis with respect to Sobol'  
538 method. Moreover, GBT not only ranks the variables in the same way as Sobol'  
539 method but also provides basically identical proportional indices with respect  
540 to the total effect. The importance metrics obtained from the three methods  
541 highlight the ridge geometry is the most influential variable for the fissure  
542 opening, with the pressure variation also having a non-negligible impact on  
543 the fissure development. The contributions from the other two variables are  
544 smaller.

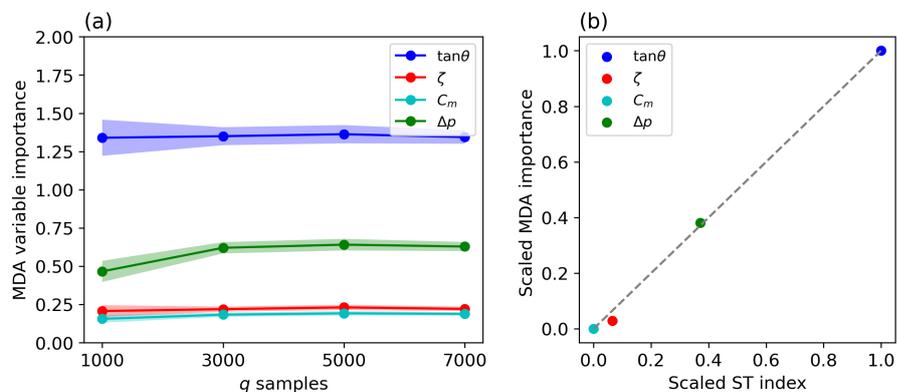
#### 545 5.4 Partial Dependence

546 We also employ partial dependence to investigate the surrogate model response  
547 to the variable changes(Fig.7). A number of 50 samples from the validation set  
548 are used to illustrate how the model prediction to one variable changes, keeping  
549 fixed the other features. Note that each sample is represented by one thin line.  
550 The thicker lines represent the partial dependence calculated from the whole  
551 validation set (20% of 7000 samples). Although there are some discrepancies  
552 between GPC and GBT models results, the trend of partial dependence for  
553 each variable is similar.

554 The impact is limited when  $\tan \theta < 1$ , however, once exceeding 1.47, the rel-  
555 ative activated depth boosts significantly. The model response remains almost  
556 constant until  $\zeta > 0.8$  when the average line slope that abruptly increases,  
557 although the contribution of thicker aquifers is still limited. According to the  
558 gradient variation,  $d_{r,act}$  is more sensitive with  $C_m$  up to  $0.02 \text{ MPa}^{-1}$  and a



**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 show the relationships between first order and total indices.



**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  $q = 7000$ .

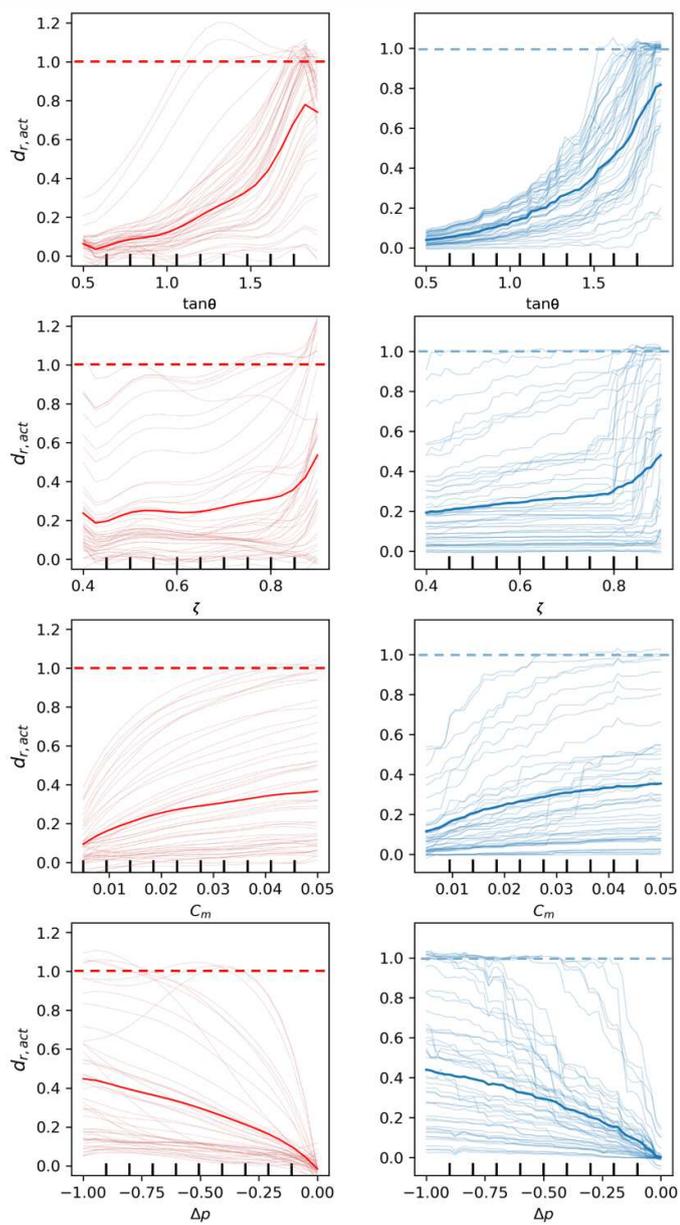
larger compressibility does not favor a much larger fissure development.  $\Delta p$  causes a relative larger output variation with respect to  $\zeta$  and  $C_m$ , which is consistent with the variable importance ranking. The slope of partial dependence gently decreases when the absolute value of  $\Delta p$  reduces.

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.

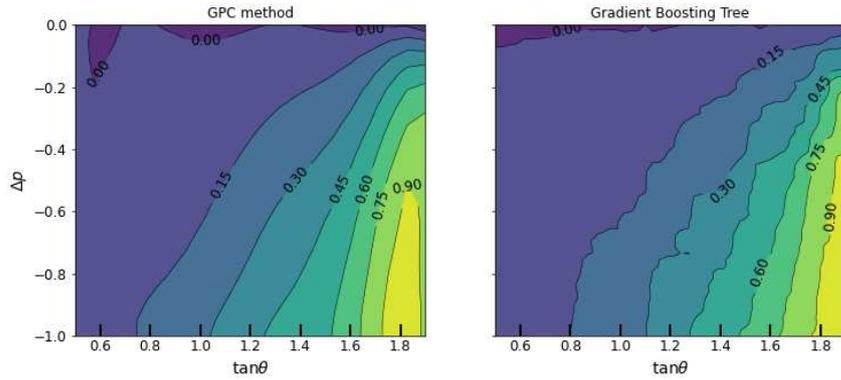
## 6 Discussion

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



**Fig. 7** Marginal effect of GPC (left panel) and GBT (right panel) methods on the model parameters  $\tan\theta$ ,  $\zeta$ ,  $C_m$ , and  $\Delta p$ . 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).

585 between the pair of surfaces constituting the IEs inserted above the ridge tip.  
 586 However, we are aware of the possible formation of fissures due to sliding condi-  
 587 tion only, such as in the case addressed by Li et al (2021), where multi-fissure  
 588 formation is simulated for the hydrogeologic setting at Guangming village,  
 589 Wuxi, China. These are more complex cases, where discontinuities develop  
 590 where the stress field reaches the yield surface (Equation (2)), and require an  
 591 appropriate analyses of the stress field in the continuous body prior to insert  
 592 the IEs in the most critical zones of the 3D mesh.

593 With more fissure appearances over the last decades, researches started fo-  
 594 cusing on the quantitative analyses of the fissure formation mechanism. Sheng  
 595 et al (2003) defined the ratio of tensile stress over tensile strength as an indica-  
 596 tion for fissure inception and carried out an one-at-a-time sensitivity analysis  
 597 which suggested the confining stress and, secondarily, the depth of aquifer  
 598 as the key parameters for fissure formation. Unfortunately, the impact of the  
 599 ridge geometry was not taken account into this analysis. Frigo et al (2019)  
 600 applied a multivariate regression to fit the depth of fissure opening as function  
 601 of the pressure variation and the ratio between exploited aquifer thickness and  
 602 ridge tip depth. The regression surface consists of a pair of planes with discon-

tinuous 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,

631 for example using indices that consider moment independence (Borgonovo,  
632 2007; Pianosi and Wagener, 2015; Dell’Oca et al, 2017).

633 We advocate the use of surrogate models to reduce the computational cost  
634 of the generation of the output samples for the computation of the importance  
635 metrics (Saltelli et al, 2010). Surrogates based on GPC techniques are promi-  
636 nent because the easy derivation of the Sobol’ indices at no additional compu-  
637 tational burden. However, it is observed that increasing the level of problem  
638 non linearity, e.g., in proximity of the fissure opening, the method fails to pro-  
639 vide a good model proxy (e.g., the predicted  $d_{r,act} > 1$ , see Fig. 4(a)). For this  
640 reason, we also employed the GBT algorithm for the estimation of total sensi-  
641 tivity measures, i.e., the mean decrease accuracy estimates MDA. Compared  
642 to Sobol’ indices, MDA importance lacks of a straightforward interpretation,  
643 as they are computed based on the model prediction accuracy rather than the  
644 effects on the output variance. Thus, MDA is limited to assess the interaction  
645 effects between the input variables. However, this limit can be compensated by  
646 using other interpretability methods like SHAP (Lundberg and Lee, 2017) or  
647 by using them as interpretation of the Sobol’ total effects. The counterpart of  
648 this methodology, which seems more suitable for applications on discontinuous  
649 problems compared to GPC, is that GBT regression tree spends more time  
650 on tuning hyperparameters to optimize the model performance, increasing the  
651 overall computational burden.

## 652 **7 Conclusion**

653 The present work investigates the relative importance of various hydrogeologic  
654 features to the formation and propagation of aseismic fissures above the crest  
655 of buried bedrock. The conceptual model used for numerical simulation is ide-  
656 alised and simplified from the field case reported in Wuxi, China. Earth fissures

657 develop only with the simultaneous occurrence of an undulating bedrock that  
658 intercept a thick compressible aquifer where pressure decline takes place.

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

664 The numerical results are processed by GSA to assess the variable (bedrock  
665 ridge geometry, aquifer thickness and compressibility, and pore pressure vari-  
666 ation) importance to fissure activation and propagation. Sobol' provided with  
667 Monte Carlo approach are taken as the reference and compared to Sobol' in-  
668 dices derived from surrogates of the forward model. GPC and GBT algorithms  
669 are applied to fit the numerical solution and then estimate the factor impor-  
670 tance based on surrogate model prediction. The following main conclusions  
671 can be drawn:

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

681 Finally, we have assessed the computational performances of three techniques  
682 on this application. Sobol' technique requires a larger sample size to converge,  
683 which makes it computationally expensive. Compared to GPC-based model,

684 GBT performs better on approximating the discontinuous solution but requires  
685 a larger computational burden.

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

---

**References**

- 697 **References**
- 698 Biot MA (1941) General theory of three-dimensional consolidation. *Journal of*  
699 *Applied Physics* 12(2):155–164
- 700 Borgonovo E (2007) A new uncertainty importance measure. *Reliability En-*  
701 *gineering & System Safety* 92(6):771–784
- 702 Breiman L (2001) Random forests. *Machine Learning* 45(1):5–32
- 703 Budhu M (2011) Earth fissure formation from the mechanics of groundwater  
704 pumping. *International Journal of Geomechanics* 11(1):1–11
- 705 Budhu M, Shelke A (2008) The formation of earth fissures due to groundwa-  
706 ter decline. In: *Proceedings of the 12th International Conference of Interna-*  
707 *tional Association for Computer Methods and Advances in Geomechanics*  
708 (IACMAG), pp 3051–3059
- 709 Burbey TJ (2002) The influence of faults in basin-fill deposits on land subsi-  
710 dence, Las Vegas Valley, Nevada, USA. *Hydrogeology Journal* 10(5):525–538
- 711 Carreón-Freyre D, Cerca M, Ochoa-González G, Teatini P, Zuñiga FR (2016)  
712 Shearing along faults and stratigraphic joints controlled by land subsidence  
713 in the Valley of Queretaro, Mexico. *Hydrogeology Journal* 24:657–674, DOI  
714 10.1007/s10040-016-1384-0
- 715 Carvalho DV, Pereira EM, Cardoso JS (2019) Machine learning interpretabil-  
716 ity: A survey on methods and metrics. *Electronics* 8(8):832
- 717 Conway BD (2016) Land subsidence and earth fissures in south-central and  
718 southern Arizona, USA. *Hydrogeology Journal* 24(3):649–655
- 719 Copas JB (1983) Regression, prediction and shrinkage. *Journal of the Royal*  
720 *Statistical Society: Series B (Methodological)* 45(3):311–335
- 721 Couaillier V, Savin É (2019) Generalized polynomial chaos for non-intrusive  
722 uncertainty quantification in computational fluid dynamics. In: *Uncertainty*  
723 *Management for Robust Industrial Design in Aeronautics*, Springer, pp 123–

724 141

725 Crestaux T, Le Maître O, Martinez JM (2009) Polynomial chaos expansion for  
726 sensitivity analysis. *Reliability Engineering and System Safety* 94(7):1161–  
727 1172, DOI 10.1016/j.ress.2008.10.008

728 Dell’Oca A, Riva M, Guadagnini A (2017) Moment-based metrics for global  
729 sensitivity analysis of hydrological systems. *Hydrology and Earth System  
730 Sciences* 21(12):6219–6234, DOI 10.5194/hess-21-6219-2017

731 Franceschini A, Ferronato M, Janna C, Teatini P (2016) A novel lagrangian  
732 approach for the stable numerical simulation of fault and fracture mechanics.  
733 *Journal of Computational Physics* 314:503–521

734 Franceschini A, Castelletto N, Ferronato M (2019) Block preconditioning for  
735 fault/fracture mechanics saddle-point problems. *Computer Methods in Ap-  
736 plied Mechanics and Engineering* 344:376–401

737 Friedman JH (2001) Greedy function approximation: a gradient boosting ma-  
738 chine. *Annals of Statistics* pp 1189–1232

739 Friedman N, Zander E (2020) Sglib collection of tutorials. [https://ezander.  
740 github.io/ParameterAndFieldIdentification/](https://ezander.github.io/ParameterAndFieldIdentification/), accessed: 2021-09-23

741 Friedman N, Zoccarato C, Zander E, Matthies HG (2021) A worked-out exam-  
742 ple of surrogate-based bayesian parameter and field identification methods.  
743 In: Chiachio-Ruano J, Chiachio Ruano M, Snkararaman S (eds) *Bayesian  
744 Inverse Problems: Fundamentals and Engineering Applications* (1st ed.),  
745 CRC Press, pp 155–203, DOI 10.1201/b22018

746 Frigo M, Ferronato M, Yu J, Ye S, Galloway D, Carreón-Freyre D, Teatini  
747 P (2019) A parametric numerical analysis of factors controlling ground  
748 ruptures caused by groundwater pumping. *Water Resources Research*  
749 55(11):9500–9518

750 Garipov TA, Karimi-Fard M, Tchelepi HA (2016) Discrete fracture model for  
751 coupled flow and geomechanics. *Computational Geosciences* 20(1):149–160,

DOI 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. *Journal of Computational and Graphical Statistics* 24(1):44–65

Herman J, Usher W (2017) Salib: An open-source python library for sensitivity analysis. *Journal of Open Source Software* 2(9):97, DOI 10.21105/joss.00097, URL <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. *Hydrogeology Journal* 18(8):1773–1785

Hernandez-Marin M, Burbey TJ (2012) Fault-controlled deformation and stress from pumping-induced groundwater flow. *Journal of Hydrology* 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*, Springer, pp 101–122

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. *International Journal for Numerical and Analytical Methods in Geomechanics* 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

- 780 derived horizontal displacement. *International Journal of Rock Mechanics*  
781 *and Mining Sciences* 51:105–118
- 782 Jaxa-Rozen M, Kwakkel J (2018) Tree-based ensemble methods for sensitivity  
783 analysis of environmental models: A performance comparison with sobol  
784 and morris techniques. *Environmental Modelling & Software* 107:245–266,  
785 DOI 10.1016/j.envsoft.2018.06.011
- 786 Kaintura A, Dhaene T, Spina D (2018) Review of polynomial chaos-based  
787 methods for uncertainty quantification in modern integrated circuits. *Elec-*  
788 *tronics* 7(3):30
- 789 Karimi-Fard M, Durlofsky LJ, Aziz K (2003) An efficient discrete frac-  
790 ture model applicable for general purpose reservoir simulators. In: *SPE*  
791 *Reservoir Simulation Symposium*, Society of Petroleum Engineers, DOI  
792 10.2118/79699-MS
- 793 Le Maître O, Knio O, Najm H, Ghanem R (2004) Uncertainty propagation  
794 using wiener–haar expansions. *Journal of Computational Physics* 197(1):28–  
795 57, DOI 10.1016/j.jcp.2003.11.033
- 796 Li Y, Teatini P, Yu J, Franceschini A, Frigo M, Zoccarato C, Ye S (2021)  
797 Aseismic multifissure modeling in unfaulted heavily pumped basins: Mecha-  
798 nisms and applications. *Water Resources Research* 57(10):e2021WR030127,  
799 DOI 10.1029/2021WR030127
- 800 Liu Y, Zhang D, Wang Gy, Liu C, Zhang Y (2019) Discrete element method-  
801 based prediction of areas prone to buried hill-controlled earth fissures. *Jour-*  
802 *nal of Zhejiang University - SCIENCE A* 20(10):794–803
- 803 Louppe G (2014) Understanding random forests: From theory to practice.  
804 arXiv preprint arXiv:14077502
- 805 Lundberg SM, Lee SI (2017) A unified approach to interpreting model pre-  
806 dictions. In: *Proceedings of the 31st International Conference on Neural*  
807 *Information Processing Systems*, pp 4768–4777

- 808 Ochoa-González G, Carreón-Freyre D, Franceschini A, Cerca M, Teatini P  
809 (2018) Overexploitation of groundwater resources in the faulted basin of  
810 Querétaro, Mexico: A 3d deformation and stress analysis. *Engineering Geol-*  
811 *ogy* 245:192–206
- 812 Ochoa-González G, Carreón-Freyre D, Franceschini A, Cerca M, Teatini P  
813 (2018) Overexploitation of groundwater resources in the faulted basin of  
814 Querétaro, Mexico: A 3D deformation and stress analysis. *Engineering Ge-*  
815 *ology* 245:192–206, DOI 10.1016/j.enggeo.2018.08.014
- 816 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blon-  
817 del M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cour-  
818 napeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine  
819 learning in Python. *Journal of Machine Learning Research* 12:2825–2830
- 820 Peng J, Chen L, Huang Q, Men Y, Fan W, Yan J (2013) Physical simulation of  
821 ground fissures triggered by underground fault activity. *Engineering Geology*  
822 155:19–30, DOI 10.1016/j.enggeo.2013.01.001
- 823 Pianosi F, Wagener T (2015) A simple and efficient method for global sen-  
824 sitivity analysis based on cumulative distribution functions. *Environmental*  
825 *Modelling & Software* 67:1–11, DOI 10.1016/j.envsoft.2015.01.004
- 826 Probst P, Boulesteix AL, Bischl B (2019) Tunability: importance of hyperpa-  
827 rameters of machine learning algorithms. *The Journal of Machine Learning*  
828 *Research* 20(1):1934–1965
- 829 Roustant O, Fruth J, Iooss B, Kuhnt S (2014) Crossed-derivative based sen-  
830 sitivity measures for interaction screening. *Mathematics and Computers in*  
831 *Simulation* 105:105–118
- 832 Saltelli A (2002) Sensitivity analysis for importance assessment. *Risk Analysis*  
833 22(3):579–590
- 834 Saltelli A, Annoni P (2010) How to avoid a perfunctory sensitivity analysis.  
835 *Environmental Modelling & Software* 25(12):1508–1517

- 836 Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010)  
837 Variance based sensitivity analysis of model output. design and estimator for  
838 the total sensitivity index. *Computer Physics Communications* 181(2):259–  
839 270
- 840 Sheng Z, Helm DC (1998) Multiple steps of earth fissuring caused by ground-  
841 water withdrawal. In: *Land Subsidence Case Studies and Current Research: Proceedings of the Dr. Joseph F. Poland Symposium on Land Subsidence*,  
842 Assoc. Eng. Geol. Special Publication, vol 8, pp 149–154
- 844 Sheng Z, Helm DC, Li J (2003) Mechanisms of earth fissuring caused  
845 by groundwater withdrawal. *Environmental & Engineering Geoscience*  
846 9(4):351–362
- 847 Sobol' IM (1993) Sensitivity analysis for non-linear mathematical models.  
848 *Mathematical Modelling and Computational Experiment* 1:407–414
- 849 Sobol' IM (2001) Global sensitivity indices for nonlinear mathematical models  
850 and their monte carlo estimates. *Mathematics and Computers in Simulation*  
851 55(1-3):271–280
- 852 Sobol' IM, Asotsky D, Kreinin A, Kucherenko S (2011) Construction and  
853 comparison of high-dimensional sobol'generators. *Wilmott* 2011(56):64–79
- 854 Sochala P, Le Maître O (2013) Polynomial chaos expansion for subsurface flows  
855 with uncertain soil parameters. *Advances in Water Resources* 62:139–154,  
856 DOI 10.1016/j.advwatres.2013.10.003
- 857 Sudret B (2008) Global sensitivity analysis using polynomial chaos expansions.  
858 *Reliability Engineering & System Safety* 93(7):964–979
- 859 Teatini P, Ferronato M, Gambolati G, Bertoni W, Gonella M (2005) A century  
860 of land subsidence in Ravenna, Italy. *Environmental Geology* 47(6):831–846
- 861 Vittek M, Borovansky P, Moreau PE (2006) A simple generic library for c. In:  
862 *Reuse of Off-the-Shelf Components. Proc. of 9th International Conference*  
863 *on Software Reuse*, Turin, Italy, Springer, pp 423–426

- 864 Wagener T, Pianosi F (2019) What has global sensitivity analysis ever done  
865 for us? a systematic review to support scientific advancement and to inform  
866 policy-making in earth system modelling. *Earth-Science Reviews* 194:1–18,  
867 DOI 10.1016/j.earscirev.2019.04.006
- 868 Wang G, You G, Shi B, Qiu Z, Li H, Tuck M (2010) Earth fissures in jiangsu  
869 province, china and geological investigation of hetang earth fissure. *Envi-  
870 ronmental Earth Sciences* 60(1):35–43
- 871 Wang Z, Zhang Y, Wu J, Yu J, Gong X (2015) Numerical simulation of earth  
872 fissures due to groundwater withdrawal. *Proceedings of the International  
873 Association of Hydrological Sciences* 372:395–398
- 874 Wiener N (1938) The homogeneous chaos. *American Journal of Mathematics*  
875 60(4):897–936, DOI 10.2307/2371268
- 876 Xiu D (2007) Efficient collocational approach for parametric uncertainty anal-  
877 ysis. *Communications in Computational Physics* 2(2):293–309
- 878 Xiu D (2010) *Numerical Methods for Stochastic Computations. A Spectral  
879 Method Approach*. Princeton University Press, Princeton, New Jersey
- 880 Ye S, Luo Y, Wu J, Yan X, Wang H, Jiao X, Teatini P (2016) Three-  
881 dimensional numerical modeling of land subsidence in Shanghai, China.  
882 *Hydrogeology Journal* 24(3):695–709
- 883 Ye S, Franceschini A, Zhang Y, Janna C, Gong X, Yu J, Teatini P (2018) A  
884 novel approach to model earth fissure caused by extensive aquifer exploita-  
885 tion and its application to the Wuxi case, China. *Water Resources Research*  
886 54(3):2249–2269
- 887 Youssef AM, Sabtan AA, Maerz NH, Zabramawi YA (2014) Earth fissures in  
888 Wadi Najran, Kingdom of Saudi Arabia. *Natural Hazards* 71(3):2013–2027
- 889 Zander E (2020) Sglib github repository. [https://github.com/ezander/  
890 sglib-testing](https://github.com/ezander/sglib-testing), accessed: 2021-09-23

- 
- 891 Zhu L, Franceschini A, Gong H, Ferronato M, Dai Z, Ke Y, Pan Y, Li X,  
892 Wang R, Teatini P (2020) The 3-D facies and geomechanical modeling of  
893 land subsidence in the Chaobai Plain, Beijing. *Water Resources Research*  
894 56(3), DOI 10.1029/2019WR027026
- 895 Zoccarato C, Gazzola L, Ferronato M, Teatini P (2020) Generalized poly-  
896 nomial chaos expansion for fast and accurate uncertainty quantification in  
897 geomechanical modelling. *Algorithms* 13(7):156