Understanding earthquake location uncertainties using global sensitivity analysis framework

Marine Laporte, Jean Letort, Michaël Bertin, Laurent Bollinger

Geophysical Journal International

Geophys. J. Int. (2024) **237,** 1048–1060
Advance Access publication 2024 March 6
GJI Seismology

https://doi.org/10.1093/gji/ggae093

Understanding earthquake location uncertainties using global sensitivity analysis framework

Marine Laporte [1,2] Jean Letort [3] Michaël Bertin[1] and Laurent Bollinger[1]

[1]*Commissariat à l'énergie atomique et aux énergies alternatives, Laboratoire de Géophysique et Aléas, F-75015 Paris, Île-de-France, France.*
Email: marinelaporte.ensg@gmail.com
[2]*Laboratoire de Géologie, Département de géosciences, Ecole normale supérieure, Université PSL, CNRS, F-75005 Paris, France*
[3]*Observatoire Midi Pyrénées, IRAP, CNRS UMR 5277, Université Paul Sabatier,* 31062*, Toulouse, France*

Accepted 2024 March 5. Received 2024 January 27; in original form 2023 August 21

SUMMARY

Earthquake hypocentres are routinely obtained by a common inversion problem of *P*- and *S*-phase arrivals observed on a seismological network. Improving our understanding of the uncertainties associated with the hypocentral parameters is crucial for reliable seismological analysis, understanding of tectonic processes and seismic hazard assessment. However, current methods often overlook uncertainties in velocity models and variable trade-offs during inversion. Here, we propose to unravel the effects of the main sources of uncertainty in the location process using techniques derived from the Global Sensitivity Analysis (GSA) framework. These techniques provide a quantification of the effects of selected variables on the variance of the earthquake location using an iterative model that challenges the inversion scheme. Specifically, we consider the main and combined effects of (1) variable network geometry, (2) the presence of errors in the analyst's observations and (3) errors in velocity parameters from a 1-D velocity model. These multiple sources of uncertainty are described by a dozen of random variables in our model. Using a Monte Carlo sampling approach, we explore the model configurations and analyse the differences between the initial reference location and 100 000 resulting hypocentral locations. The GSA approach using Sobol's variance decomposition allows us to quantify the relative importance of our choice of variables. It highlights the critical importance of the velocity model approximation and provides a new objective and quantitative insight into understanding the sources of uncertainty in the inversion process.

Key words: Monte Carlo methods; Spatial analysis; Statistical methods; Earthquake parametrization.

1 INTRODUCTION

Quantification of uncertainties in earthquake location, especially with respect to hypocentral depth, is necessary for a reliable analysis of the spatiotemporal properties of the seismicity and associated seismotectonic interpretation (e.g. Kagan *et al.* 2003; Husen & Hardebeck 2010; Zaliapin & Ben-Zion. 2015; Turquet *et al.* 2019). This issue is often raised in deciphering seismicity source mechanisms (e.g. Mousavi *et al.* 2023) as well as in the monitoring of induced seismicity (e.g. Phillips *et al.* 2000; Diehl *et al.* 2017; Garcia-Aristizabal *et al.* 2020). For natural seismicity, the quality of the earthquake catalogue, including the accuracy and precision of the hypocentral location of earthquakes used in the construction of regional seismotectonic models, has been shown to have an impact on the computation of the probabilistic seismic hazard assessment (e.g. Beauval & Scotti. 2004; Allen *et al.* 2020; Gounon *et al.* 2022). The seismogenic thickness, which can be constrained by the distribution of seismicity at depth, is also a critical factor in determining the maximum magnitude of earthquakes in a given region, which is essential for understanding the seismic cycle (e.g. Nazareth & Hauksson 2004; Chiarabba & De Gori 2016; Wu *et al.* 2017; Grevemeyer *et al.* 2019; Petrucelli *et al.* 2019; Zuza & Cao 2020).

In practice, regional earthquake locations are based on a nonlinear inversion problem that aims to estimate the hypocentral parameters (epicentral location, depth, origin time), assuming others are known (velocity model parameters), from distant observations (e.g. Tarantola. 2005). These parameters, as well as the observations (usually direct *P*- and *S*-wave arrival times picked up on a given regional seismic network), inherently possess uncertainties, either aleatory or epistemic.

Among these uncertainties, the uncertainty associated with the measurement of *P*- and *S*-phase arrival times (e.g. Gomberg *et al.* 1990; Billings & Sambridge. 1994; Phillips *et al.* 2000), in addition

to the uncertainty in their traveltime estimate from an approximated crustal velocity model, has been shown in several studies to have a significant impact on the accuracy of earthquake location (e.g. Noble *et al.* 2014; Pavlis 1986; Billings & Sambridge. 1994; Dreger *et al.* 1998; Bondár *et al.* 2004; Michelini & Lomax 2004). This has led them to propose empirical criteria for identifying well-located earthquakes (e.g. Gomberg *et al.* 1990; Bondar & McLaughlin. 2009) when the uncertainties given by the location algorithms are not fully realistic (e.g. Boyd & Snoke. 1984).

The quantification of uncertainties is a first step to understand the variability of the hypocentral parameters. It can be realized by propagating known uncertainties in the earthquake location problem (e.g. Billings & Sambridge 1994; Gesret *et al.* 2015; Mousavi *et al.* 2023). Tarantola & Valette (1982) have introduced the way of representing the parameter uncertainties as a posterior density function assuming linearly independent error terms following a Gaussian distribution. Bayesian methods are now commonly used in seismology to represent the full posterior density distribution of the source parameters (x, y, z, t) including their uncertainties. However, these classical methods do not provide any insight into how this uncertainty associated with the hypocentral parameters can be apportioned to the effects of the different sources of uncertainty in the inputs or the effects of their interactions in the inversion process.

In this study, we use the framework of Global Sensitivity Analysis (GSA) to categorise and quantify the independent and correlated effects of uncertain parameters and different network configurations that may affect the inverse problem of earthquake location. Indeed, GSA is an important tool for understanding how the output of a model varies in response to source uncertainties in its inputs (Saltelli. 2002; Saltelli *et al.* 2010). The input parameters (data uncertainties, velocity model uncertainties) are defined by a range of values that can represent either knowledge or lack of knowledge. Sensitivity analysis helps to decipher which parameters have the most influence on the outputs and which parameters have so little influence that they can be neglected. It is thus an effective tool for reducing uncertainty by better understanding the unknown parameters and their influence in the model (e.g. Sobol 1993; Homma & Saltelli 1996). It is a well-established technique in environmental science, engineering, finance and epidemiology (e.g. Borras-Mora *et al.* 2021; Edeling *et al.* 2021) because of its ability to deal with nonlinearity in complex models with many input parameters (e.g. Homma & Saltelli. 1996). It has been used to evaluate important measures in seismic risk, which is affected by both random parameters and uncertain models (Rohmer & Foerster. 2011; Gehl *et al.* 2013; Rohmer. 2015; Di Maio *et al.* 2023). However, these approaches are not yet widely used in seismology (e.g. Franczyk. 2019) due to the significant computational resources they require and their limited applicability to deterministic problems (a unique solution for the same set of parameters). The classical trade-off between the depth and the origin-time makes the earthquake location problem non-deterministic.

This paper presents a first application of a GSA method applied to the assessment of earthquake location uncertainties at the local scale. We propose a simplified model that aims to re-locate a seismic event sited in a 1-D crustal velocity model and recorded by a local network of 10 seismological stations. We use common earthquake location algorithms that provide a unique solution for the same input parameters m. The model $m = (\textit{random input variables})$ passes through the forward function $g(m)$ (starting from the reference location and returns P- and S-phase arrival times at every station) and the inverse function $g^{-1}(m)$ (starting from the P and S arrival times at every station and returns the new location) adding relevant uncertainties at each step. In the first section, we describe the framework of the sensitivity analysis, which includes the description of the method, the design of the deterministic model and the selection of the input random variable (RV) that are expected to add uncertainty in the output. The independent and correlated effects of each input RV, representing measurement errors, velocity model errors and variation in the seismological network geometry, are analysed and discussed. We then discuss the comparison of the importance measures of these three families of RV for two models using two widely used earthquake location algorithms, respectively.

2 METHODOLOGY

2.1. Variance decomposition method

Among the many tools proposed in GSA, we focus on the variance decomposition method introduced and developed by Sobol (1993, 2001), which is widely used in GSA because of its robustness (e.g. Saltelli *et al.* 2010). The method decomposes the variance of model outputs into several fractions attributable to each independent input RV and correlations between groups of inputs. Analytically, it corresponds to a decomposition of the model function f into a sum of integrals of increasing dimension. The method interrogates the model in a probabilistic manner using a Monte Carlo approach. At each iteration, it takes as input a deterministic set of values randomly drawn from their prior distribution (which, in our study, can be uniform or Gaussian). The non-dimensional Sobol indices summarise the results of the variance decomposition method in a simple way.

Considering a random input variable X_j, the first-order Sobol index S_{1j} characterizes the proportion of the variance of the model output Z that arises solely from the effect of X_j when it varies over its domain of uncertainty or variability. This quantitative output is also known as the 'main effect' and is expressed as:

$$S_{1j} = \frac{\mathrm{Var}\left[E\left(Z|X_j\right)\right]}{\mathrm{Var}\left(Z\right)},$$

$\mathrm{VAR}\left(Z\right)$: total variance of Z,

$E\left(Z|X_j\right)$: expectancy of Z knowing X_j.

In addition, second-order Sobol indices S_{ji} quantify the level of interactions between couples of variables (X_j, X_i). Higher-order indices can also be computed to define all possible interactions between X_j and sets of variables of increasing dimensions. Our focus is on determining the main effect of X_j and the total Sobol index S_{Tj}, which describes all effects of variable X_j on Z, including the effects due to complex interactions. The total effect provides the proportion of the variance of Z that can be fully attributed to X_j.

First-order and total Sobol indices can be estimated using Monte Carlo-based sampling. In this paper, we use a quasi-Monte Carlo sampling method, which consists in selecting a number of representative sample sets (input scenarios), described as low-discrepancy samples that achieve a near uniformity over the input space Ω (e.g. Sobol. 2001; Saltelli *et al.* 2010; Campolongo *et al.* 2011). Considering k random variables that are distributed in the space Ω of dimension N, a robust estimation of the main and total effects requires a minimum number of $N(k + 2)$ iterations. Since N can be very large, the computational cost remains very high. Some recent studies have proposed to use polynomial chaos expansion or machine-learning techniques in order to reduce the computational cost of the screening process (e.g. Lucay *et al.* 2022). Here, we choose to first test our experimental design by a quick and simple

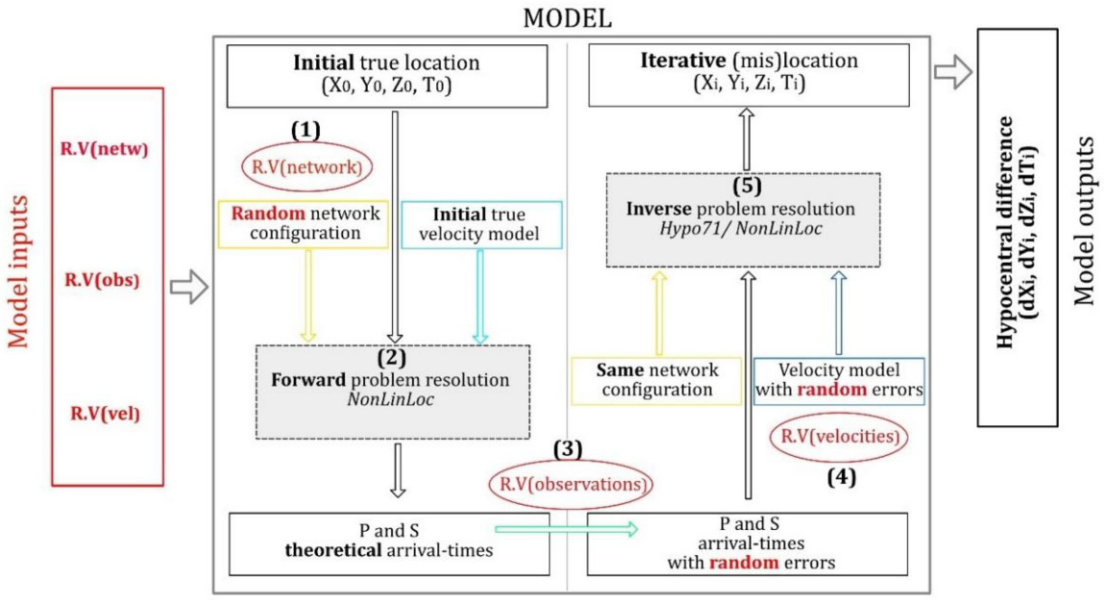

Figure 1. Scheme of the forward–inverse relocation function using random input variables for the network configuration, for velocity model errors and for errors in arrival times.

sensitivity measure of a 'one at a time' experiment using the elementary effects screening method (Morris. 1991; Nabi *et al.* 2021). The method and its results are described in the Supporting Information (Sup.1). We then apply the Sobol method in a second step. We then perform the Sobol analysis on an $N = 2^{13} = 8192$ randomly drawn samples for each input. We generate our sample sets over the input space Ω using the Saltelli extension of the Sobol sequence (Campolongo *et al.* 2011). To compute second-order Sobol indices in addition to the main and total Sobol indices, we need $N(2k + 2)$ iterations to sample the space Ω, corresponding to approximately 245 000 experiences of the model.

2.2. Experimental design

Basically, sensitivity analysis is based on a model, its inputs and its outputs. To apply this method, we have developed a deterministic model that represents an earthquake relocation function, taking into account multiple input random variables X_j regrouped into three families, including varying seismological network geometry, varying errors in the velocity model approximation and varying errors for P and S arrival-time observations.

Our deterministic model is constructed to relocate a synthetic earthquake sited in a reference 1-D crustal velocity model and detected by a local network of 10 stations. We fix the reference location of this earthquake at a depth of 25 km. The largest station–event distance is 150 km, the synthetic event can thus be referred to any $M > 2$ earthquake. The model outputs the hypocentral deviation and the difference in origin time (dx, dy, dz, dt) resulting from changes in the set of values for the input variables. To explore changes in the network geometry, the model solves both the forward problem to compute the theoretical P and S arrival times at each station in the network, and the inverse problem using a standard procedure and earthquake location algorithm. Each step of the model relocation function is schematically described in Fig. 1. There are five sequential steps:

(i) Generation of a network geometry using network-defined random variables (RV netw, Fig. 1, see Section 2.3.1);

(ii) Resolution of the forward problem for the fixed seismological network by computing theoretical P and S seismic wave arrival times at each station in the local network. We use the program Time2Eq from the NonLinLoc earthquake location algorithm, which computes theoretical traveltimes from a synthetic source and seismic stations located on a 3-D grid. The grid is centred at (0,0), with a width of 200 km in (x, y) and 100 km in depth, and a distance of 500 m between nodes;

(iii) Adding errors, from observation variables (RV obs, Fig. 1, see section 2.3.2), to the theoretical arrival times estimated in step 2 to simulate observed P- and S-wave arrivals that may have been incorrectly picked;

(iv) Calculation of the approximate 1-D velocity model with random errors in the velocities as well as in the layer depth from the random velocity variables. (RV vel, Fig. 1, see Section 2.3.3);

(v) Resolution of the inverse problem for the defined network, the simulated observed arrival times and the new velocity model.

The sensitivity analysis is performed twice using two different inversion methods: Geiger iterative linearization using the Hypo71 algorithm (Lee & Lahr. 1972) and grid-search nonlinear inversion using NonLinLoc algorithm (Lomax *et al.* 2000). We refer to M_1 and M_2 as the two models using the NonLinLoc and Hypo71 inversion algorithms, respectively. The goal is to compare their sensitivity regarding the random variables. One important difference between the two inversion methods is that Hypo71 needs an initial location from which the algorithm will attempt to converge towards the optimal location. The convergence is therefore highly dependent on this initial location; for this study, we choose an initial depth of 5 km. The observations are weighted according to the distance and phase type. To completely eliminate the dependence on the initial location, the same sensitivity analysis is performed using NonLinLoc grid search algorithm, using the grid already defined in step 2 and a grid-search optimization using the Octree algorithm. The grid inversion is performed using the Tarantola and Valette method (Tarantola & Valette 1982).

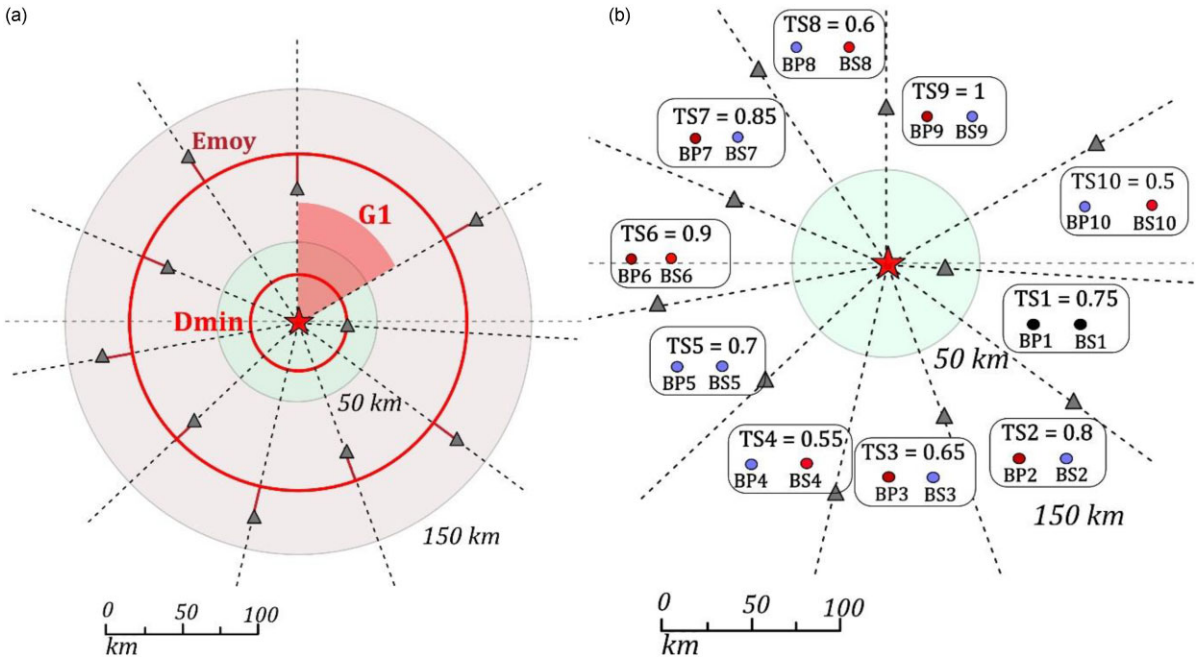

Figure 2. (a) Configuration scheme of a seismic network of 10 stations (stations are shown as triangles) around the reference event (central star) defined by the four fixed random variables. G1: primary azimuthal gap; Dmoy: average source–station distance; Dmin: minimum source–station distance; Emoy: deviation from the average distance. (b) Representation of the station error terms for picking arrival times. Red badges correspond to a positive coefficient (larger picking error) BP*i* or BS*i* for the *P* or *S* phase, respectively. Blue badges correspond to a negative coefficient (smaller picking error). Each station is associated with a probability of observing an *S* phase: TS*i*.

Table 1 Definition of the azimuth and the source–station distance according to the values of the random variables (in bold) and the station number *i*.

Station	1	2	3	i = 4 to n
Azimuth (°N) *source/station*	0	**G1**	$\mathbf{G1} + \frac{(360-\mathbf{G1})}{n-1}$	$\mathbf{G1} + (i-2).\frac{(360-\mathbf{G1})}{n-1}$
Distance (km) *source/station*	**Dmoy** + A1**Emoy**	**Dmoy** + A2**Emoy**	**Dmin**	**Dmoy** + A$_i$**Emoy**

2.3 Definition of random variables

GSA requires the selection of independent RV inputs, which we group into the three families (network, observations and velocities). In addition, the problem must remain deterministic, with a unique solution per set of inputs. The input variables take values from either a normal distribution, denoted as $N(\mu, \sigma)$, or a uniform distribution, denoted as $U([inf, sup])$.

2.3.1 Random variables for the network geometry

Two issues arise when defining random input variables for the network (RV netw). First, many networks can satisfy the standard criteria used in the literature to characterize network geometry, including the number of stations, primary and secondary azimuthal gaps, average and minimum source–station distances and interstation distances (e.g. Bondàr *et al.* 2004). Secondly, these network variables are highly interrelated, as the value of a secondary azimuthal gap depends on the primary azimuthal gap, whose limits depend on the number of stations, making it difficult to satisfy the condition of independent variables needed for a proper analysis. To overcome these challenges, an integrated strategy is adopted that allows only one network configuration from the values of the selected network inputs RV. In this study, the number of stations per simulation is fixed, and the secondary azimuthal gap is not investigated. Therefore, only four independent network variables are considered (Fig. 2a):

- G1, the primary azimuthal gap (in degrees relative to geographic north);
- Dmin, the distance to the nearest station (in km);
- Dmoy, the average distance between stations (in km);
- Emoy, the deviation from the average distance that introduces controlled heterogeneity in the distances, following a scheme [A1,A2,. . . ,An] with randomly selected A*i* (A*i* = 1, -1) by an equiprobable draw prior to the analysis.

The network defined by these few variables corresponds to a unique configuration if the stations are placed in accordance with the conditions presented in Table 1.

The generated networks do not explore the effect of the secondary azimuthal gap or the distance between stations nor do they explore the full set of possible local network geometries to characterize the source. However, they are constructed to be representative of a large number of networks or parts of networks in densely covered regions (Europe, America, Japan. . .). The maximum source–station distance of 150 km which we allow corresponds to the approximate maximum detection distance of *Pg* and *Sg* waves (Crotwell *et al.* 1999; Diehl *et al.* 2021). We exclude seismic phases with travel paths that are not direct (Pn, Sn).

Table 2 Distribution laws for the random variables characterizing the network geometry.

RV	G1	Dmin	Dmoy	Emoy
Law	U([360/*nsta* ; 270])	U([2 ; 40])	U([61 ; 125])	U([0 ; 20])
Unit	Deg	Km	Km	Km

Since the number of stations n is fixed before the analysis, the minimum azimuthal gap of the simulation is $G1_{min} = 360/n$. We limit the maximum azimuthal gap to $G1_{max} = 270^{\circ}$ in order to avoid considering configurations where the network geometry is too degraded for which the inversion by iterative linearized methods may not provide a solution. For the simulations, the primary azimuthal gap is always placed in the northeastern quadrant and each station is placed with an increasing azimuth $\Delta az = \frac{360-G_1}{n-1}$ so as to be evenly distributed.

We set the distance bounds of our model variables Dmin, Dmoy and Emoy consistent with the Hypo71 parameters Xnear and Xfar, which define the minimum and maximum distances at which observations are accepted and rejected, respectively, in the Hypo71 inversion. The weight assigned to P and S observations decreases linearly with the distance to the source between Xnear and Xfar. To ensure that our nearest station is always within the Xnear distance and that all other stations are within the Xfar distance, we choose our minimum and maximum variables to satisfy the following assumptions:

(i) Max(Dmin) $\leq$ Xnear
(ii) Min(Dmin) $>$ 0
(iii) Max(Dmoy) $+$ Max(Emoy) $\leq$ Xfar
(iv) Min(Dmoy) $-$ Max(Emoy) $>$ Max(Dmin)

The range of values and the choice of distributions are described in Table 2.

2.3.2 *Random variables for P and S observations*

Uncertainties in P- and S-wave arrival times can be represented by a random error term around the theoretical arrivals (e.g. Diehl *et al.* 2009; Husen & Hardebeck. 2010; Retailleau *et al.* 2022). Ideally, this error should be estimated independently for each observation, which greatly increases the number of random variables in the analysis (20 error terms for 10 stations). To reduce the computational cost, we limit the number of random variables to common errors errP and errS for all stations.

We added fixed additional error terms at each station for P-phase picks (BP1, ... BP10) and S-phase picks (BS1, ... BS10). The station picking error is the combination of the common error errP and the additional error term. The coefficients BPi and BSi were chosen randomly before the simulation, from a uniform distribution between 0.8 and 1.2, in order to change the picking errors of the stations to a maximum of $\pm$20%. These error terms can be positive and represent an increase of the common picking error (low SNR at the station) or negative (high SNR at the station). The signs of BS*i* for each station are represented in blue or red badges respectively in Fig. 2(b).

In addition, in seismology, it has been proposed that the hypocentral depth of an earthquake can be particularly well constrained by the S–P difference between arrival times (e.g. Uhrammer. 1982). In practice, it is not always easy to pick S-wave arrivals on a seismogram which may be hidden in the P-wave energy or not well recorded at stations with only a vertical component. Therefore, not all stations may provide an S-phase pick.

To represent this difference in the number of S-phase picks from one simulation to another, we define a random variable associated with the number of S-phases: n_S. The number of S phase picks varies from 1 to 10. Each station is associated with a higher or lower probability of picking an S-wave defined by T_{Si}, which is defined for each station before the analysis. This station characteristic value is between 0.5 and 1 and all values are evenly distributed according to the number n of stations (Fig. 2b). The condition of having an S-phase pick is the combination of the RV n_S and the station probability T_{Si}.

S condition : $n_S.T_{Si} \geq 0.5$,

where n_S is uniformly drawn between 0.5 and 1 so that:

(i) If $n_S = 1$, the condition of S is validated for all stations.
(ii) If $n_S = 0.5$, the nearest station which has a characteristic value of 1 will validate the condition of S.
(iii) If $0.5 \leq n_S \leq 1$, the number of observations of S increases linearly with n_S .

Thus, five independent variables characterizing errors of observations in the seismogram are defined for the model (Table 3):

- ErrP: average error on the observation of P-wave
- ErrS: average error on the observation of the S-wave
- ErrPc: error on the P-wave observation at the nearest station
- ErrSc: error on the S-wave observation at the nearest station
- nS: condition of S-wave observation

We sample the errors of the normal distributions Gp and Gs centred on 0 with a standard deviation of 0.05 s for the errors of the P-wave arrivals and 0.15 s for the errors of the S-wave arrivals. These values have been estimated by comparing several manual and automatic pickings of P and S waves in Mayotte, without considering misidentifications for seismic phases, which have a lower probability of occurrence in seismic bulletins (Retailleau *et al.* 2022).

2.3.3 *Random variables for the velocity model*

The uncertainty arising from the poor approximation of the crustal velocity model is one of the main issues in quantifying uncertainties in the earthquake location problem since some heterogeneities or local variations can be missed in the 1-D approximation. These errors are very difficult to identify and therefore to quantify. The P- and S-wave velocities can be influenced by several factors, including the density and elastic parameters of the crustal layer, which in turn are determined by the lithology and pressure–temperature conditions. On the other hand, the Vp/Vs ratio is influenced by Poisson's ratio, which is related to physical processes (Eberhart-Phillips *et al.* 2017).

Some earthquake location algorithms now employ 3-D velocity models that better capture the spatial heterogeneities of seismic phase velocities in the crust (e.g. Lomax *et al.* 2000; Diehl *et al.* 2021). However, to apply our model to a 3-D velocity grid, velocity errors should also be estimated independently at each node of the grid for the P-wave velocity and for the Vp/Vs ratio. In this study, we choose to represent a classical location problem using a 1-D reference velocity model with three horizontal layers. The model is based on a continental crust with the Moho at 55 km and the mid-crustal layer at 20 km. The average Vp/Vs ratio for sedimentary layers (1.73) and typical P-wave velocities for continental crust are used. To derive an approximate model for each inversion,

Table 3 Distribution laws for the random variables characterizing *P*- and *S*-wave observations.

VA	ErrP	ErrS	ErrPc	ErrSc	nS
Law	N(0 ; 0.05)	N(0 ; 0.15)	N(0 ; 0.05)	N(0 ; 0.15)	U([0.5 ; 1])
Units	s	s	s	s	s

we consider five independent random variables representing errors associated with the reference model.

- Vp1: error on the *P*-wave velocity in the layer 1
- Vp2: error on the *P*-wave velocity in the layer 2
- Vp3: error on the *P*-wave velocity in the layer 3
- Vp/Vs: error on the Vp/Vs ratio (identical for all layers)
- H1: thickness of the layer 1

In order to obtain a realistic velocity model even after perturbing the reference model, we impose that the velocity gradient between layers 2 and 3 is sufficiently steep and that the *P*-wave velocity always increases with depth:

(i) Realistic boundary condition for the Moho:

$V_{P2} \leq 6.9\ \mathrm{km\,s^{-1}}$ and $V_{P3} \geq 7.6\ \mathrm{km\,s^{-1}}$

(ii) Boundary condition Layer 1/Layer 2: $V_{P1} \leq 5.9\ \mathrm{km\,s^{-1}} \leq V_{P2}$

(iii) Boundary condition Layer 0/Layer 1: $V_{P0} \leq 4.9\ \mathrm{km\,s^{-1}}$

The scheme for building the (approximate) velocity model is shown in Fig. 3 and the probability distributions of RV velocities in Table 4. The boundaries and standard deviation values are chosen by considering large but likely variations of the Vp velocities and Vp/Vs ratio from earthquake tomographic models (e.g. Paul *et al.* 2001; Chevrot *et al.* 2022).

3 RESULTS OF THE GLOBAL SENSITIVITY ANALYSIS

3.1 Variations of the model outputs

As presented in Fig. 1, the model outputs the difference between the reference point-source parameter and its relocation:

$$dx_i = x_0 - x_i; dy_i = y_0 - y_i; \quad dz_i = z_0 - z_i; \quad dt_i = t_0 - t_i$$

representing differences in the *x*- and *y*-axes, depth as well as the origin time, respectively. In addition, we also look at the variations of the root-mean-square (RMS) residuals to better understand the relation between the RMS, which is sometimes used as a criterion for location accuracy, and the actual uncertainty of the solution. There is as many solutions as sets of inputs tested by the Monte Carlo sampling, thus we consider the variations within the 245 000 solutions.

We first look at the variations of the model M_1 outputs individually before looking at their variance decomposition. Pair-wise relationships between every M_1 outputs are shown in Fig. 4. In the figure, we visually separate the solutions between those with high origin time errors ($dt > 0.4$ s in orange) and those with low errors in order to highlight the trade-off between hypocentral depth errors (dz) and origin-time errors (dt) (Fig. 4. D3). For the three location parameters (dx, dy, dz), we plot the tolerance ellipse at 1σ (68 per cent), 2σ (95 per cent) and 3σ (99.7 per cent) from the covariance matrixes. The 68 per cent confidence interval is 80 and 200 m for the dx and dy errors and 4.2 km for the depth error for the model M_1. With very low occurrences, some experiments of the model M_1 produced depth errors up to $\pm$18 km.

From Fig. 4, it can be seen that large depth errors ($dz > 7\ km$) mostly coincide with large origin-time errors ($dt > 0.4\ s$). This observation cannot be extended to epicentral (dx, dy) errors which are more randomly distributed with respect to the other outputs. Moreover, the epicentral errors are very small (< 2 km) compared to the depth errors. Fig. 4 also highlights that, for our model design, the RMS values of NonLinLoc do not exceed 0.3 s while the dt errors reach $\pm$2 s with a standard deviation of $\pm$0.66 s. Thus, the RMS does not correct for more than 10 per cent of dt errors. The RMS value takes into account the errors in the observations, but it cannot represent errors due to a bad approximation of the velocity model. This has been pointed out for other linearized inversions, such as centroid moment tensor inversions (e.g. Valentine & Trampert 2012; Hejrani *et al.* 2017; Vasyura-Bathke *et al.* 2021). Therefore, it should not be the only criterion used to characterize the location accuracy of a NonLinLoc inversion.

Much larger epicentral errors of 4.6 and 3.8 km are obtained with the inversion using the Hypo71 algorithm. The 68 per cent dz standard deviation is within the same order of $\pm$4.9 km, but the depth errors distribution does not follow a strictly Gaussian distribution with a tendency to make the depth more superficial (closer to the initial depth of the inversion). However, the dt error shows a standard deviation of $\pm$0.6 s slightly lower than for the NonLinLoc inversion and probably better corrected by the RMS that has a much larger distribution with a standard deviation of 1.3 s which is a large value for an earthquake catalogue. For the Hypo71 inversion, results of the variations of the model M_2 outputs are presented in Supplementary Materials (Sup.2). The figure shows that the large errors in epicentral locations are linearly correlated with each other's as well as with the depth error dz and the dt error. Thus, the trade-off problem is shared between all source parameters and large depth errors are not only explained by large dt errors. It is shown that the algorithm can often converge in a 'local' optimum in presence of a bad set of inputs (meaning large velocity model errors, bad network configuration etc.). As an example, we also see the strong dependence of the depth distribution to the initial depth used for the Hypo71. Hypo71 fixes the depth at 5 km (dz=20 km) when the constraints given by the input sets are too loose and the depth errors do not follow a strictly Gaussian distribution. The Gaussian tail describes the tendency of Hypo71 to locate at shallower depths (due to the shallow initial depth). In addition, we note that although the RMS are much larger, the RMS distribution does not follow the mislocation linearly, so high RMS are not indicative of a large mislocation.

From this first-order comparison, we show that the two inversion methods have very different output variations in response to our model design. Non-LinLoc epicentres are very robust (less than 1 km error) to changes in network geometry, velocity model or phase observations. The criterium for epicentre accuracy is probably too restrictive for this inversion method (Bondar *et al.* 2004). Nevertheless, the trade-off between depth and origin time is very

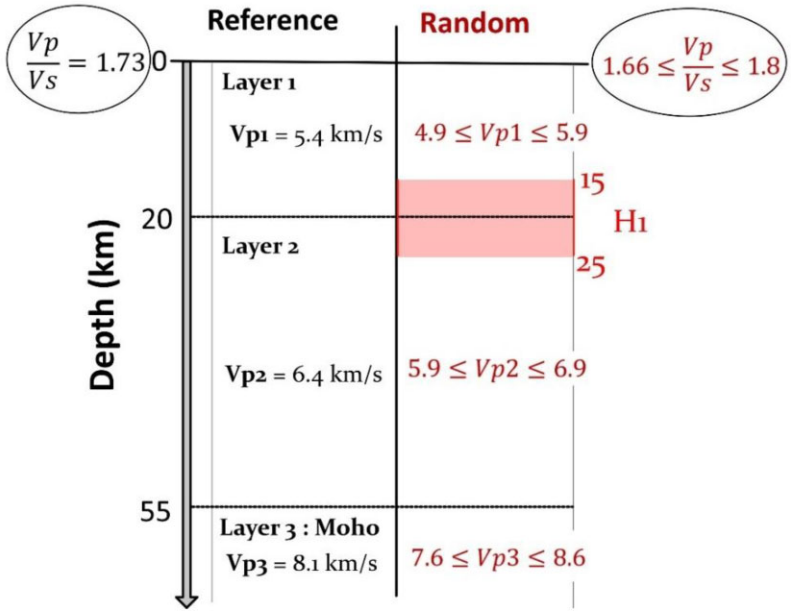

Figure 3. Left-side, the reference three layers 1-D velocity model. Right-side, the random velocity model, wrongly estimated, which will be used in the inverse problem.

Table 4 Distribution laws for the random variables characterizing errors in the velocity model.

VA	Vp1	Vp2	Vp3	$\frac{Vp}{Vs}$	H1
Loi	U([-0.5 ;0.5])	U([-0.5 ;0.5])	U([-0.5 ;0.5])	N(0 ;0.023)	U([-5 ;5])
Unités	km s^{-1}	km s^{-1}	km s^{-1}	SU	km

pronounced, and the depth error is of the same order as that obtained using the Hypo71 inversion and cannot be deciphered using the RMS.

3.2 Main and total Sobol indices

As described in Section 2.1, the Sobol method deciphers the amount of variation in the output parameters that can be attributed to one or more inputs of our model. A preliminary ranking of the importance of each of the 14 input parameters $X_j \in \{X_{netw}; X_{obs}; X_{vel}\}$ was provided by the preliminary analysis using the elementary effect method (Supporting Information Sup.1). This preliminary analysis showed the primary importance of the velocity of the layer in which the earthquake is actually located (Vp2) for depth variations, on the one hand. On the other hand, for errors in epicentral location, it showed the relative importance of inputs characterizing the network geometry such as the distance of the stations and primary azimuthal gap. These preliminary results were consistent with the most common criterion used to characterize the accuracy of epicentral locations (Bondar *et al.* 2004). However, the elementary effects method did not provide any quantitative insights into the importance ranking and does only provide a qualitative information on the presence or absence of correlated errors.

In this study, we focus mainly on the epicentral locations and depth errors ($de = \sqrt{dx^2 + dy^2}$, dz) with a look at the origin-time errors and RMS variations in the discussion. Overall, the Sobol analysis finds a similar hierarchy to that proposed by the Morris analysis corresponding to the relative values of total (S_T) Sobol indices between each input variable X_j (red bars in Fig. 5). In addition, the Sobol method provides the proportion of the total effect that is due solely to the variation of the variable without considering the effect of its interaction with other inputs as showed as the main Sobol index (S_1) (blue bars in Fig. 5). The difference between the total and the main effect indicates the proportion of the outputs variance that is due to the interaction of the input variable X_j with the other ones. It should be noted that the sum of all ST indices is always greater than one in presence of correlated interactions, as these interactions are shared over several inputs.

For the depth errors, the effects of interactions between input variables are not dominant (Fig. 5a). The good approximation of the crustal velocity model accounts for 99 per cent of the depth error in our model. In particular, the velocity of the layer where the earthquake is located (Vp2) accounts for 71 ± 2 and $83 \pm 3\%$ of the variance in the hypocentral depth, considering its main and total effects, respectively. In second and third places are the thickness of the layer (H1: 10%±1 by main effect S_1) and the velocity of the superficial layer (Vp1: 5%±1 by main effect S_1). The inputs describing the geometry of the local network are not important for the depth estimation alone, but they have a small influence in the presence of a poor approximation of the velocity model (up to 5±1 % for S_T indices). Six of the fourteen inputs appear to have negligible effects on the depth error with Sobol S_T indices below 0.5%. These include the observation inputs (ErrP, ErrS, ErrS_c, nS) and the Vp/Vs ratio (Vp/Vs). With the same trend, the variations in the earthquake origin time depend quasi exclusively on the velocity model inputs (Vp2, Vp1, H1), with very few interactions and almost no effect from the observation errors and the network geometry (Fig. 5c).

In contrast, Fig. 5(b) shows that a very large part of the epicentral location errors (de) comes from these interactions between the

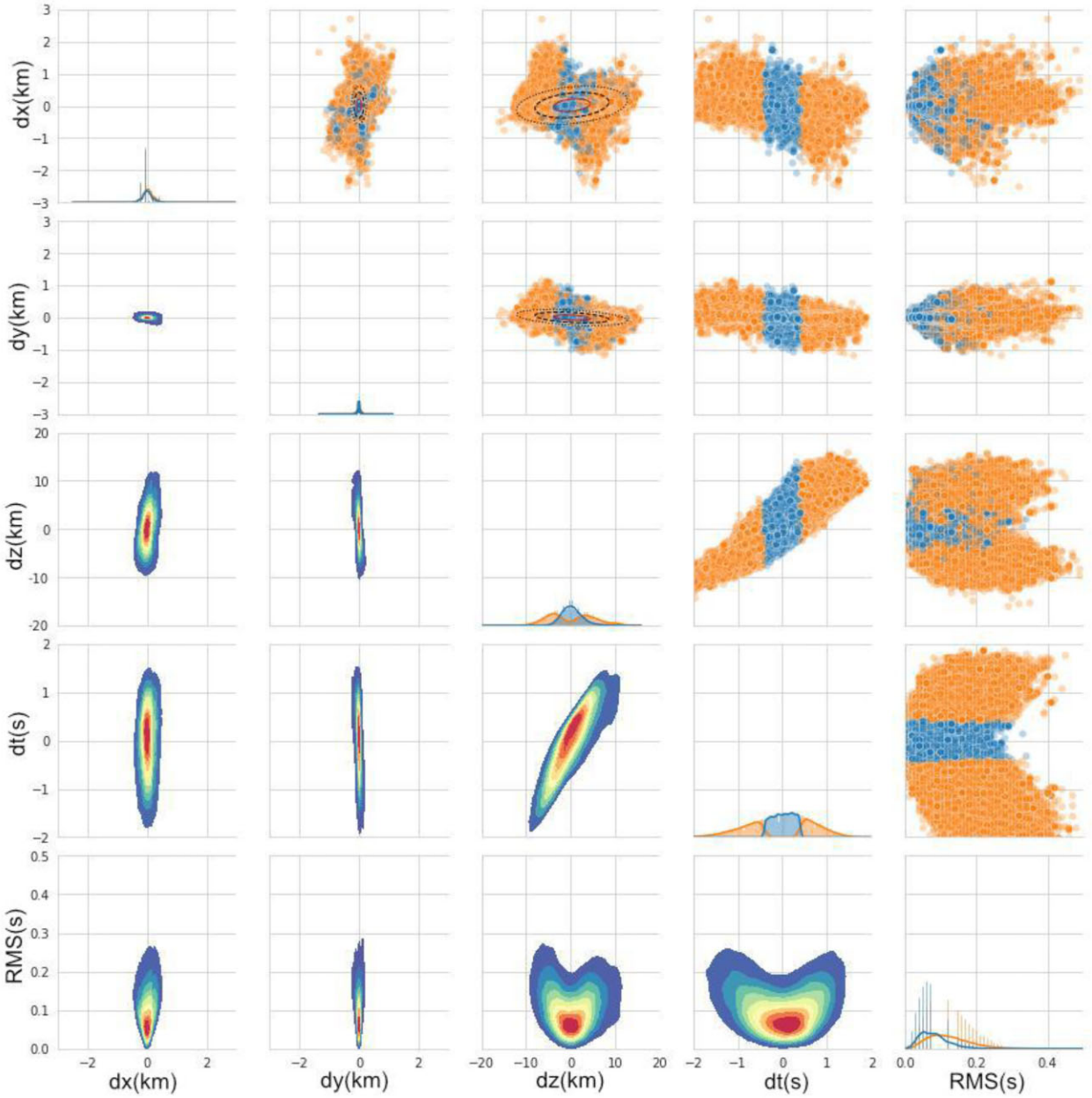

Figure 4. Representation of the 5 × 5 symmetric matrix of the model's $\boldsymbol{M}_1$ output distributions using the NonLinLoc inversion. From top to bottom and left to right: *X*-axis error *dx* (km); *Y*-axis error *dy* (km); depth error *dz* (km); time origin error *dt* (s) and RMS residuals (s), respectively. On the diagonal, the density probability distribution of each output with respect to its *xy*-axis. Upper triangle: distributions of the 245 000 source solutions projected on respective output's axis. We coloured in orange all the solutions with large *dt* errors ($dt > 0.4$ s). Lower triangle: 2-D kernel density estimates of the distributions of the 245 000 source solutions projected on the respective outputs axis. Large to low densities of points are represented from red to blue colours, respectively.

inputs of the model M_1 (red bars in Fig. 5b). The three main input variables that share about 50% of the variance in the epicentral locations are the mean distance between the source stations (Dmoy: 53%±13 per cent), the deviation of the mean distance that allows us to generate more chaotic geometries with various interstation distances (Emoy: 60±6 per cent) and the velocity of the layer where the earthquake is actually sited (Vp2: 48%±4 per cent). In the fourth rank comes the primary azimuthal gap with 27%±5 of shared influence. The predominant effect of the primary azimuthal gap has been used in several studies to derive high-quality locations in seismic catalogues (e.g. Bondar *et al.* 2004; Bondar & McLaughlin. 2009; Lomax *et al.* 2001).

Finally, the Sobol indices of observation error for the *P* and *S* phases for the nearest station show that they play a negligible role in the epicentral error. The velocity of the layer below the reference location (Vp3) has no effect on any of the output parameters, since the seismic ray path does not pass through this layer, regardless of which inversion algorithm we use. All other input parameters have a small influence of less than 10 per cent of the variance decomposition of the epicentral errors. Errors in the observations, as well as the effect of the number of detected *S* phases (nS) and the Vp/Vs ratio, only play a role in the final RMS value, meaning that these errors are well handled by the earthquake location inversion method (Fig. 5d).

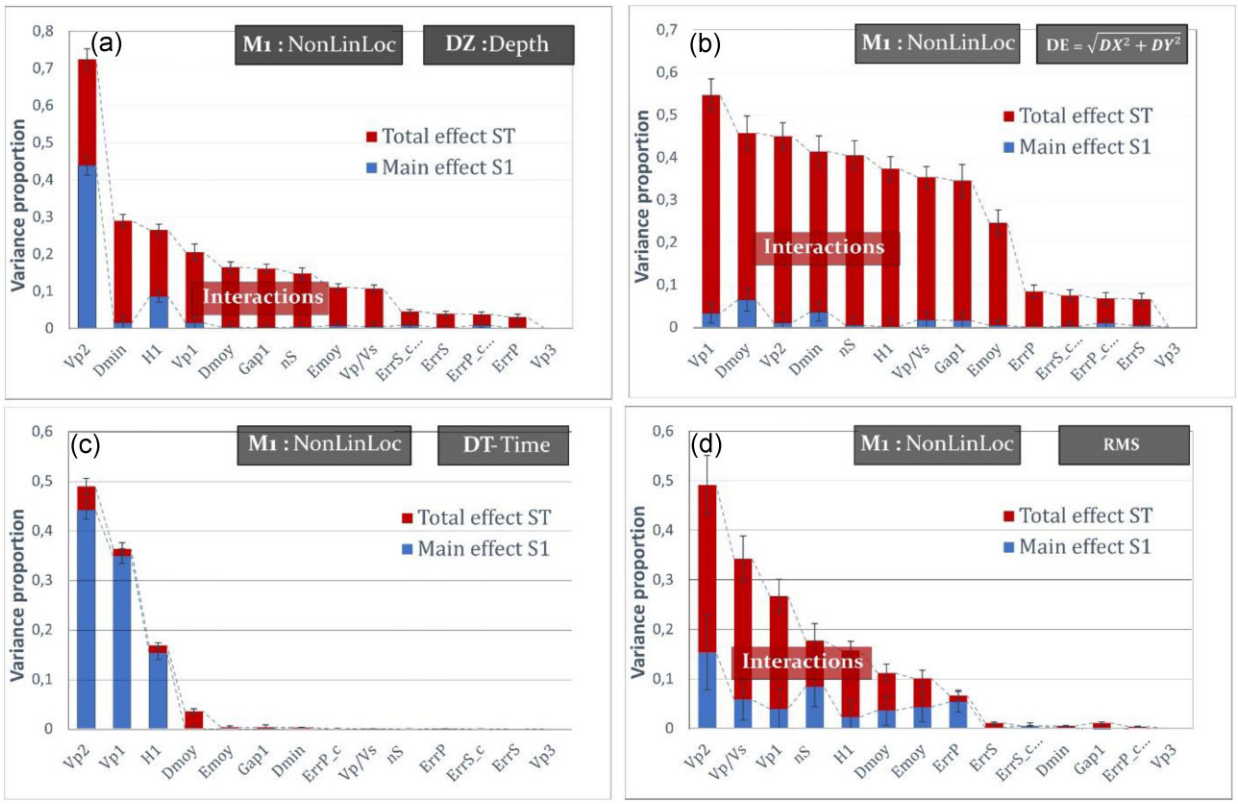

Figure 5. Histograms of main (blue bars) and total (red bars) Sobol indices of the 14 model's M1 inputs with regards to: (a) the depth error ***dz***, (b) the epicentral error ***de***, (c) the error in origin time ***dt*** and (d) the ***RMS***. The model M1 corresponds to the simulation with a NonLinLoc inversion.

Sobol indices obtained for the model using M_2 with the Hypo71 inversion are governed by the interactions between the input variables (Supporting Information Sup.3). Regarding the variations of their outputs, the importance rankings and Sobol indices for the output source parameters of the two inversion methods are very different despite the very close model design. The five outputs (dx, dy, dz, dt and RMS) are sensitive to all the input variables (except Vp3) with varying importance rankings. For NonLinLoc inversion, it seems that all errors relative to the *S*-phases are not involved in the depth estimation. Hypocentral locations using a Hypo71 inversion are especially more sensitive to errors in observations and to the number of *S* seismic phases used. The origin time errors are not only due to the velocity model inputs but also to the number of *S* phases used and to the network geometry. Here, the application of the Sobol-Monte-Carlo analysis provides keys to understanding the advantages and the disadvantages of an inversion method.

3.3 Details on the interactions: second-order indices

As shown previously, the hypocentral locations obtained from the sensitivity analysis for a NonLinLoc inversion (M_1) are less subject to interactions between inputs than those for a Hypo71 inversion (M_2). On the one hand, for M_1, most of these interactions concerns the epicentral error (dx and dy) and the RMS variations which are in fact very robust ($de < 1km$). However, it is still interesting to understand the few interactions that are present. The second-order Sobol indices give us a closer look at the pair-wise relationships between the inputs in the model. In particular, for the two models and for each input, the greatest interactions concern either the velocities random variables (RV vel) with some network geometry inputs (RV netw) (Fig. 6). For the depth estimation, the main interactions are from the interaction of the minimum distance, mean distance, and mean elevation of the stations with the velocity of the main layer (Vp2) (Fig. 6a). DT errors are influenced by only four input variables (Fig. 6c), and the mean distance of all stations (Dmoy) to the source is the only network geometry variable that acts through the mean of interactions with random velocity variables (RV vel). Finally, the observation variables (RV obs) have a very little effect on the hypocentral location (dx, dy, dz, dt) but have a main effect on the RMS value.

On the other hand, for the Hypo71 inversion M_2, the main interaction is clearly between the velocity of the second layer and the minimum source-station distance (Supporting Information Sup.4) for the depth estimation. Also, errors in the observations (RV obs) have a much larger impact on all outputs (dx, dy, dz, dt, rms), particularly on the dt errors and the RMS and their effects are more complex due to interactions. This implies that these errors are probably much more difficult to correct using the Hypo71 inversion algorithm even when they are known. Small changes in phase picking arrival times can affect the depth and origin time resolution when using Hypo71, but less so when using NonLinLoc. NonLinLoc would therefore give more stable results when picking is difficult (note however that we have assumed that picking errors are normally distributed).

3.4 Insights on the outputs' variations

To examine the covariance effects of the inputs on the depth distribution, we project the distributions of dz for the 245 000 experiments onto the domain of definition of each of the 14 inputs of the model

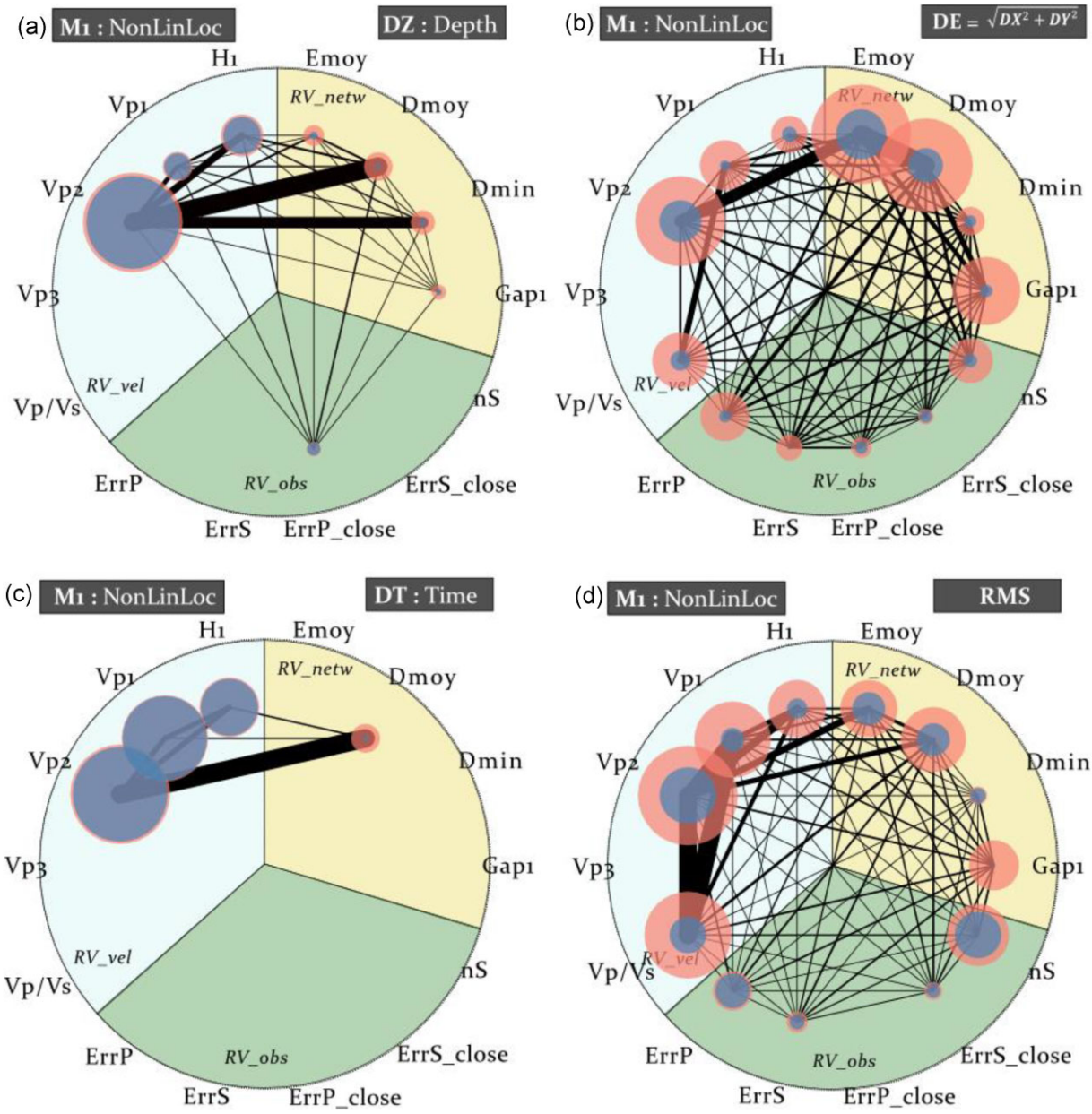

Figure 6. Radial plot showing the first, total and secondary Sobol indices above 1 per cent for (a) the depth error $\boldsymbol{dz}$, (b) the epicentral location error $\boldsymbol{de}$, (c) the error on the time origin $\boldsymbol{dt}$ and (d) the $\boldsymbol{RMS}$.Inner blue circles correspond to the main effect and outer red circles to the total effect including all the interactions. Their size is relative to their share in the variance of the output. Second order indices are represented by the black lines between pairwise inputs. The line width is relative to the value of the second order Sobol indices.

(Fig. 7). These projections provide additional insight into the overall effect of each variable on dz. It should be noticed that the projections display the uniform distributions that we designed for variables $\{Gap1,\ Emoy,\ Dmin,\ Dmoy,\ Vp1,\ Vp2,\ Vp3,\ H1,\ nS\}$, and Gaussian distributions for other variables (Fig. 7) The total distribution of dz (on all y-axes of Fig. 7 subfigures) remains unchanged from one subfigure to another and accounts for the total effects of all inputs variables.

Fig. 7 confirms that some inputs variables have no clear impact on the depth distribution ($ErrP,\ ErrPc,\ ErrS,\ ErrSc,\ Vp/Vs,\ nS,\ Vp3,\ Gap1$). For the other ones, we can better understand the real effect that the input variable produces on the depth estimation. As depicted by Sobol indices (Fig. 6a), the error on Vp2 has a monotonic and quasi-linear impact on dz. More precisely, it shows that an underestimation of Vp2 of 0.5 km s^{-1} implies a negative dz bias of 5 km ($= z_0 - z_i$ meaning that the experience has overestimated the depth) (Fig. 7.a3). The same observation is also evidenced for the M2 scenario (Supporting Information Sup.7). A lower velocity within a layer above (Vp1 or Vp2) indicates a delayed arrival time of the seismic phase at each station, resulting in a greater inferred depth for a given origin time. This trend is depicted in the same manner by looking at the covariance of dt varying of ± 1 s on average with regards to velocities variation (Supporting Information Sup.6). Finally, looking at the x-axis errors dx we see that, even if the variations are smaller than the ones observed

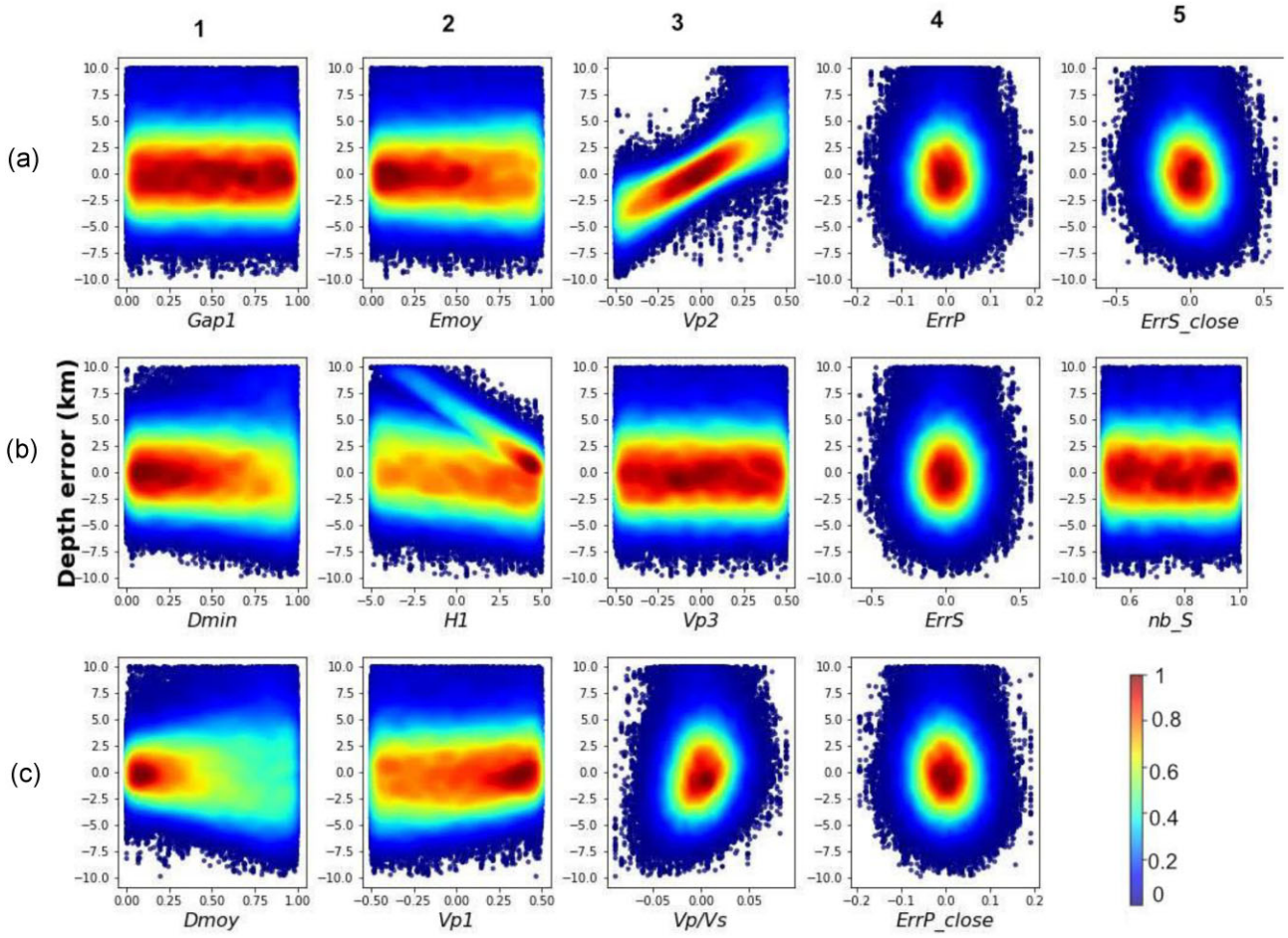

Figure 7. The distribution of depth errors dz (on all y-axes) for the model M1 is projected onto the domain of definition of each input variable (varying on the x-axis of each sub-figure). The 245 000 output depth values from M1 are displayed using a colour scale that represents the density of points.

for the depth estimations, there is also a bias due to the under or overestimation of velocities (Supporting Information Sup.5). NonLinLoc algorithm might compensate the depth increase induced by the velocity error of the second layer by putting the epicentre a few hundred meters to the west. This compensation is even more pronounced for the model M2 (Supporting Information Sup.8) with an epicentre bias of ± 2.5 km in average for a change in 0.5 km s^{-1} in $Vp2$.

The two models share as well another bias due to the layers' discontinuities of the 1-D velocity model. Both models show a tendency to assign the earthquake depth onto the layer discontinuity, resulting in a clustering of relocated events that correspond to changes in the discontinuity (Fig. 7.b2). For the M2 model, this depth bias due to the position of the layer is even more common than the bias of the M2 model to fix the depth at its initial value of 5 km (dz=20 km) (Supporting Information Sup.8). Once again, the NonLinLoc inversion seems more stable.

Finally, the three variables associated to source-station distance ($Dmin$, $Emoy$, $Dmoy$) present a similar impact on dz. A more robust depth estimation is obtained for closer and less chaotically distributed stations (Figs 7.a2, b1 and c1). This impact is as well depicted for dx errors. These input variables do not change the mean location error but increase their dispersion. Far away is the closest station ($Dmin$) or the mean source-station distance ($Dmoy$) and larger is the standard deviation of location errors.

4 DISCUSSION AND CONCLUSION

This study is a first application of GSA to quantify and a better understand earthquake location uncertainties. We have proposed here a simple model design for an earthquake located at 25 km depth in a 1-D velocity model and detected by 10 stations at local distance (<150 km). It provides some quantitative insights into the ability of two earthquake location algorithms routinely used in seismology, Hypo71 and NonLinLoc, to estimate an accurate and precise location. We show that the uncertainty in the epicentral location is highly dependent on the inversion method. The NonLinLoc algorithm is able to find an accurate and ± 1 km precise epicentral location for all experiences of our model even in the presence of large errors in the crustal velocities. The Hypo71 algorithm is much more sensitive to interactions between errors in arrival-time observations, the network geometry and the poor approximation of the crustal velocity model. Even though they are not sensitive to the same input variables or not even in the same way (with or without interactions), both inversion methods show approximately the same standard deviation for the depth estimation errors of approximately ± 9 km. This analysis confirms that NonLinLoc provides a stable and very robust epicentral location. Nevertheless, it proves that the error on the depth estimation is mainly due to the error of ± 0.5 km s^{-1} in the velocity of the layer in which the earthquake is sited. This affirmation is supported by some tests of M2 model for different depth of the reference event (Supporting Information Sup.9), which

shows that if the earthquake is sited in the third layer, the velocity of the third layer has the largest total effect. We propose a way to quantify this strong dependence of depth estimation on the accuracy of the velocity model. Therefore, in order to reduce the uncertainties in earthquake location, it seems crucial to focus on improving the accuracy of velocity models and to promote the use of local 3-D velocity models. This is already promoted by numerous seismological studies (e.g. Lomax *et al.* 2009; Husen *et al.* 2003; Wagner *et al.* 2013; Turquet *et al.* 2019). Improving the network geometry or the earthquake location algorithm seems to have less impact on the depth resolution, but still remains crucial for constraining the epicentral locations.

Our analysis does not attempt to reproduce errors arising from the heterogeneity of the Earth model. The use of a 3-D velocity as a reference will add a lot of complexity and is likely to increase the influence of the velocity model on hypocentral errors. However, numerous runs of GSA can be performed by slightly changing the model design to better match a specific configuration (a context with a known seismological network but larger uncertainties in the velocity model, or a context with very few three-component stations). This global sensitivity approach is easily adaptable to deciphering earthquake location uncertainties for fixed seismological networks at local and regional scale. It can be adapted to worst-case scenarios where a lot of seismic phases are misinterpreted with larger errors in the seismic phase arrival-times observations. For a more realistic view of the uncertainties due to errors in the velocity model approximation, it would be interesting to perform a GSA for a synthetic event located in a 3-D velocity model that has been interpreted as 1-D in the inversion process.

ACKNOWLEDGMENTS

This work detailed in this manuscript is part of ML PhD thesis, a thesis funded by CEA. Ivstan Bondar, Babak Hejrani and the editor Andrew Valentine provided thoughtful constructive comments that helped improve the original manuscript.

DATA AVAILABILITY

Python scripts of the GSA will be made available on request.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

suppl_data

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

REFERENCES

Allen, T.I., Griffin, J.D., Leonard, M., Clark, D.J. & Ghasemi, H., 2020. The 2018 national seismic hazard assessment of Australia: quantifying hazard changes and model uncertainties. *Earthq. Spectra,* **36**(1_suppl), 5–43.

Beauval, C. & Scotti, O., 2004. Quantifying sensitivities of PSHA for France to earthquake catalog uncertainties, truncation of ground-motion variability, and magnitude limits. *Bull. seism. Soc. Am.,* **94**(5), 1579–1594.

Billings, S.D., Sambridge, M.S. & Kennett, B.L.N., 1994. Errors in hypocenter location: picking, model, and magnitude dependence. *Bull. seism. Soc. Am.,* **84**(6), 1978–1990.

Bonda´r, I. & McLaughlin, K.L., 2009. A new ground truth data set for seismic studies. *Seismol. Res. Lett.,* **80**(3), 465–472.

Bondár, I., Myers, S.C., Engdahl, E.R. & Bergman, E.A., 2004. Epicentre accuracy based on seismic network criteria. *Geophys. J. Int.,* **156**(3), 483–496.

Borras Mora, E., Spelling, J. & van der Weijde, A.H., 2021. Global sensitivity analysis for offshore wind cost modelling. *Wind Energy,* **24**(9), 974–990.

Boyd, T.M. & Snoke, J.A., 1984. Error estimates in some commonly used earthquake location programs. *Earthq. Notes,* **55**(2), 3–6.

Campolongo, F., Saltelli, A. & Cariboni, J., 2011.From screening to quantitative sensitivity analysis.A unified approach. *Comput. Phys. Commun.,* **182**, 978–988.

Chevrot, S., Sylvander, M., Villaseñor, A., Díaz, J., Stehly, L., Boué, P. & Vidal, O., 2022. Passive imaging of collisional orogens: a review of a decade of geophysical studies in the PyrénéesImagerie passive des orogènes collisionnels: une revue d'une décennie d'études géophysiques dans les Pyrénées. *Bull. Soc. Géol. Fr.,* **193**(1), 1. doi: 10.1051/bsgf/2021049.

Chiarabba, C., De Gori, P., Faccenna, C., Speranza, F., Seccia, D., Dionicio, V. & Prieto, G.A., 2016. Subduction system and flat slab beneath the E astern C ordillera of C olombia. *Geochem. Geophys. Geosyst.,* **17**(1), 16–27.

Crotwell, H.P., Owens, T.J. & Ritsema, J., 1999. The TauP toolkit: flexible seismic travel-time and ray-path utilities. *Seismol. Res. Lett.,* **70,** 154–160

Di Maio, F., Gallo, N., Arcangeli, D., Taroni, M., Selva, J. & Zio, E., 2023. A bootstrapped modularised method of Global sensitivity analysis applied to probabilistic seismic hazard assessment. *Struct. Saf.,* **101,** 102312.

Diehl, T., Kissling, E., Herwegh, M. & Schmid, S.M., 2021. Improving absolute hypocenter accuracy with 3D pg and Sg body-wave inversion procedures and application to earthquakes in the Central Alps region. *J. geophys. Res.,* **126**(12), e2021JB022155.

Diehl, T., Kissling, E., Husen, S. & Aldersons, F., 2009. Consistent phase picking for regional tomography models: application to the greater Alpine region. *Geophys. J. Int.,* **176**(2), 542–554.

Diehl, T., Kraft, T., Kissling, E. & Wiemer, S., 2017. The induced earthquake sequence related to the St. Gallen deep geothermal project (Switzerland): fault reactivation and fluid interactions imaged by microseismicity. *J. geophys. Res.,* **122**(9), 7272–7290.

Dreger, D., Uhrhammer, R., Pasyanos, M., Franck, J. & Romanowicz, B., 1998. Regional and far-regional earthquake locations and source parameters using sparse broadband networks: a test on the Ridgecrest sequence. *Bull. seism. Soc. Am.,* **88** *(6),* 1353–1362.

Eberhart-Phillips, D., Bannister, S. & Reyners, M., 2017. Deciphering the 3-D distribution of fluid along the shallow Hikurangi subduction zone using P-and S-wave attenuation. *Geophys. J. Int.,* **211** *(2),* 1032–1045.

Edeling, W., Arabnejad, H., Sinclair, R., Suleimenova, D., Gopalakrishnan, K., Bosak, B. & Coveney, P.V., 2021. The impact of uncertainty on predictions of the CovidSim epidemiological code. *Nat. Comput. Sci.,,* **1**(2), 128–135 R.C. Smith, *Uncertainty Quantification*, 1st edn. (SIAM, Philadelphia, 2014)

Franczyk, A., 2019. Using the Morris sensitivity analysis method to assess the importance of input variables on time-reversal imaging of seismic sources. *Acta Geophys.,* **67**(6), 1525–1533.

Garcia-Aristizabal, A., Danesi, S., Braun, T., Anselmi, M., Zaccarelli, L., Famiani, D. & Morelli, A., 2020. Epistemic uncertainties in local earthquake locations and implications for managing induced seismicity. *Bull. seism. Soc. Am.,* **110**(5), 2423–2440.

Gehl, P., Ulrich, T., Rohmer, J., Negulescu, C., Ducellier, A. & Douglas, J., 2013. Ranking of epistemic uncertainties in scenario-based seismic risk evaluations, in *11th Int. Conf. Structural Safety & Reliability: ICOSSAR 2013,* pp. 1–7.

Gesret, A., Desassis, N., Noble, M., Romary, T. & Maisons, C., 2015. Propagation of the velocity model uncertainties to the seismic event location. *Geophys. J. Int.,* **200** *(1),* 52–66.

Gomberg, J.S., Shedlock, K.M. & Roecker, S.W., 1990. The effect of S-wave arrival times on the accuracy of hypocenter estimation. *Bull. seism. Soc. Am.,* **80**(6A), 1605–1628.

Gounon, A., Letort, J., Cotton, F., Weatherill, G., Sylvander, M. & Latour, S., 2022. Improving depth estimations of African earthquakes using

teleseismic data, and influence for the East-African rift seismic hazard characterization. *Geophys. J. Int.,* **228**(1), 447–460.

Grevemeyer, I., Hayman, N.W., Lange, D., Peirce, C., Papenberg, C., Van Avendonk, H.J. & Dannowski, A., 2019. Constraining the maximum depth of brittle deformation at slow- and ultraslow-spreading ridges using microseismicity. *Geology,* **47**(11), 1069–1073.

Hejrani, B., Tkalčić, H. & Fichtner, A., 2017. Centroid moment tensor catalogue using a 3-D continental scale earth model: application to earthquakes in Papua New Guinea and the Solomon Islands. *J. geophys. Res.,* **122,** 5517–5543.

Homma, T. & Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.,* **52**(1), 1–17.

Husen, S. & Hardebeck, J., 2010. Earthquake location accuracy, Community Online Resource for Statistical Seismicity Analysis

Husen, S., Kissling, E., Deichmann, N., Wiemer, S., Giardini, D. & Baer, M., 2003. Probabilistic earthquake location in complex three-dimensional velocity models: application to Switzerland. *J. geophys. Res.,* **108**(B2). doi: 10.1029/2002JB001778.

Kagan, Y.Y., 2003. Accuracy of modern global earthquake catalogs. *Phys. Earth planet. Inter.,* **135**(2-3), 173–209.

Lee, W.H.K. & Lahr, J.C., 1972. *HYPO71: A Computer Program for Determining Hypocenter, Magnitude, and First Motion Pattern of Local Earthquakes,* p. 100, US Department of the Interior, Geological Survey, National Center for Earthquake Research.

Lomax, A., Michelini, A., Curtis, A. & Meyers, R.A. 2009. Earthquake location, direct, global-search methods, in *Encyclopedia of Complexity and Systems Science,* pp. 2449–2473. Springer.

Lomax, A., Virieux, J., Volant, P. & Berge-Thierry, C., 2000. Probabilistic earthquake location in 3D and layered models: introduction of a Metropolis-Gibbs method and comparison with linear locations, in *Advances in Seismic Event Location,* pp. 101–134. Springer.

Lomax, A., Zollo, A., Capuano, P. & Virieux, J. 2001. Precise, absolute earthquake location under Somma–Vesuvius volcano using a new three-dimensional velocity model. *Geophys. J. Int.,* **146**(2), 313–331.

Lucay, F.A., 2022. Accelerating global sensitivity analysis via supervised machine learning tools: case studies for mineral processing models. *Minerals,* **12**(6), 750.

Michelini, A. & Lomax, A., 2004. The effect of velocity structure errors on double-difference earthquake location. *Geophys. Res. Lett.,* 31(9).

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics,* **33**(2), 161–174.

Mousavi, S., Hejrani, B., Miller, M.S. & Salmon, M., 2023. Hypocenter, fault plane, and rupture characterization of Australian earthquakes: application to the September 2021 M_w 5.9 woods point earthquake. *Seismol. Soc. Am.,* **94**(4), 1761–1774.

Nabi, S., Ahanger, M.A. & Dar, A.Q., 2021. Investigating the potential of Morris algorithm for improving the computational constraints of global sensitivity analysis. *Environ. Sci. Pollut. Res.,* **28**(43), 60900–60912.

Nazareth, J.J. & Hauksson, E., 2004. The seismogenic thickness of the southern California crust. *Bull. seism. Soc. Am.,* **94**(3), 940–960.

Noble, M., Gesret, A. & Belayouni, N., 2014. Accurate 3-D finite difference computation of traveltimes in strongly heterogeneous media. *Geophys. J. Int.,* **199**(3), 1572–1585.

Paul, A., Cattaneo, M., Thouvenot, F., Spallarossa, D., Béthoux, N. & Fréchet, J., 2001. A three-dimensional crustal velocity model of the southwestern Alps from local earthquake tomography. *J. geophys. Res.,* **106**(B9), 19367–19389.

Pavlis, G.L., 1986. Appraising earthquake hypocenter location errors: a complete, practical approach for single-event locations. *Bull. seism. Soc. Am.,* **76**(6), 1699–1717.

Petruccelli, A., Schorlemmer, D., Tormann, T., Rinaldi, A.P., Wiemer, S., Gasperini, P. & Vannucci, G., 2019. The influence of faulting style on the size-distribution of global earthquakes. *Earth Planet. Sci. Lett.,* **527,** 115791. doi: 10.1016/j.epsl.2019.115791.

Phillips, W.S., 2000. Precise microearthquake locations and fluid flow in the geothermal reservoir at Soultz-sous-Forêts, France. *Bull. seism. Soc. Am.,* **90**(1), 212–228.

Retailleau, L., Saurel, J.M., Laporte, M., Lavayssière, A., Ferrazzini, V., Zhu, W. & Team, O.V.P.F., 2022. Automatic detection for a comprehensive view of Mayotte seismicity. *C. R Géosci.,* **354**(S2), 1–18.

Rohmer, J. & Foerster, E., 2011. Global sensitivity analysis of large-scale numerical landslide models based on Gaussian-process meta-modeling. *Comput. Geosci.,* **37**(7), 917–927.

Rohmer, J., 2015. Importance Ranking of Parameter Uncertainties in Geohazard Assessments, *Doctoral dissertation*, Université de Lorraine.

Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.,* **145**(2), 280–297.

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.

Sobol, I.M., 1993. Sensitivity analysis for non-linear mathematical models. *Math. Model. Comput. Exp.,* **1,** 407–414.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.,* **55**(1-3), 271–280.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys.,* **20**(2), 219–232.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.

Turquet, A.L., Bodin, T., Arroucau, P., Sylvander, M. & Manchuel, K., 2019. Quantifying location uncertainties in seismicity catalogues: application to the Pyrenees. *J. Seismol.,* **23,** 1097–1113.

Uhrhammer, R.A., 1982. The optimal estimation of earthquake parameters. *Phys. Earth planet. Inter.,* **30**(2-3), 105–118.

Valentine, A.P. & Trampert, J., 2012. Assessing the uncertainties on seismic source parameters:towards realistic error estimates for centroid-moment-tensor determinations. *Phys. Earth planet. Inter.,* **210-211,** 36–49.

Vasyura-Bathke, J.D., Dutta, R., Mai, P.M. & Jónsson, S., 2021. Accounting for theory errors with empirical bayesian noise models in nonlinear centroidmoment tensor estimation. *Geophys. J. Int.,* **225**(2), 1412–1431.

Wagner, M., Husen, S., Lomax, A., Kissling, E. & Giardini, D., 2013. High-precision earthquake locations in Switzerland using regional secondary arrivals in a 3-D velocity model. *Geophys. J. Int.,* **193**(3), 1589–1607.

Wu, W.N., Yen, Y.T., Hsu, Y.J., Wu, Y.M., Lin, J.Y. & Hsu, S.K., 2017. Spatial variation of seismogenic depths of crustal earthquakes in the Taiwan region: implications for seismic hazard assessment. *Tectonophysics,* **708,** 81–95.

Zaliapin, I. & Ben-Zion, Y., 2015. Artefacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California. *Geophys. J. Int.,* **202**(3), 1949–1968.

Zuza, A.V. & Cao, W., 2020. Seismogenic thickness of California: implications for thermal structure and seismic hazard. *Tectonophysics,* **782-783,** 228426.