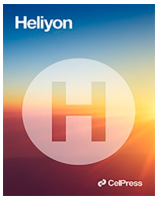
Heliyon 10 (2024) e40044

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Sensitivity analysis of parameters for carbon sequestration: Symbolic regression models based on open porous media reservoir simulators predictions

Pavel Praks [a], Atgeirr Rasmussen [b], Kjetil Olsen Lye [b], Jan Martinovič [a], Renata Praksová [a], Francesca Watson [b], Dejan Brkić [a,c,*]

[a] IT4Innovations, VSB - Technical University of Ostrava, 708 00, Ostrava, Czech Republic
[b] SINTEF Digital, 0373, Oslo, Norway
[c] Faculty of Electronic Engineering, University of Niš, 18000, Niš, Serbia

ARTICLE INFO

Keywords:
Open porous media (OPM) flow
Carbon sequestration
Sensitivity analysis
Petroleum engineering
Machine learning modeling
Symbolic regression

ABSTRACT

Open Porous Media (OPM) Flow is an open-source reservoir simulator used for solving subsurface porous media flow problems. Focus is placed here on carbon sequestration and the modeling of fluid flow within underground porous reservoirs. In this study, a sensitivity analysis of some input parameters for carbon sequestration is performed using six different uncertain parameters. An ensemble of model realizations is simulated using OPM Flow, and the model output is then calculated based on the values of the six input parameters mentioned above. CO_2 injection is simulated for a period of 15 years, while the post-injection migration of CO_2 in the saline storage aquifer is simulated for a subsequent period of 200 years, leading to a final analysis after 215 years. The input parameter values are generated using the quasi-Monte Carlo (QMC) method in the region of interest, following specified patterns suitable for analysis. The optimal convergence rate for quasi-Monte Carlo is observed. The aim of this study is to identify important input parameters contributing significantly to the model output, which is accomplished using sensitivity analysis and verified through symbolic regression modeling based on machine learning. Global sensitivity analysis using the Sobol sequence identifies input parameter 3, 'Permeability of shale between sand layers,' as having the most influence on the model output 'Secondary Trapped CO2.' All regression models, including the simplest and least accurate ones, incorporate parameter 3, confirming its significance. These approximations are valid within the designated area of interest for the input parameters and are easily interpretable for human experts. Sensitivity analysis of the developed time-dependent carbon sequestration model shows that the significance of each physical parameter changes over time: Sand porosity is more significant than shale permeability for roughly the first 120 years. Consequently, the presented results show that simulation timescales of at least 200 years are necessary for carbon sequestration evaluation.

* Corresponding author. IT4Innovations, VSB - Technical University of Ostrava, 708 00, Ostrava, Czech Republic.
E-mail addresses: pavel.praks@vsb.cz (P. Praks), Atgeirr.Rasmussen@sintef.no (A. Rasmussen), Kjetil.Olsen.Lye@sintef.no (K.O. Lye), jan.martinovic@vsb.cz (J. Martinovič), renata.praksova@vsb.cz (R. Praksová), Francesca.Watson@sintef.no (F. Watson), dejan.brkic@vsb.cz (D. Brkić).

https://doi.org/10.1016/j.heliyon.2024.e40044
Received 5 July 2024; Received in revised form 29 October 2024; Accepted 31 October 2024
Available online 1 November 2024

1. Introduction

Simulations are crucial in various fields of petroleum engineering, hydrogeology, and environmental science, helping researchers and engineers understand phenomena like groundwater flow, oil reservoir behavior, and contaminant transport. In this study we consider the simulation of CO_2 sequestration and perform a sensitivity analysis of a range of input parameters defining an ensemble of models. Using the input parameters of influence, symbolic regression models are established based on simulations done with the Open Porous Media (OPM) Flow reservoir simulator [1]. OPM Flow used in this study is an open-source software tool for simulation of fluid flow through porous materials like rocks or soil, which is extensively used in petroleum engineering but also for other purposes such as hydrogen storage [2]. It has been developed by a consortium including both research institutes and industry actors. This tool is freely available for anyone to modify and use. OPM is implemented using a flexible automatic differentiation (AD) approach to allow for easy extension with new fluid models. Existing fluid models include black-oil, polymer, solvent, and CO_2 capabilities.

CO_2 is a greenhouse gas which has influence on climate and global temperature in complex cause and effect relationships [3–7]. For this motivation, CO_2 is injected into subsurface porous formations - Geological Carbon Sequestration (including marine geological sequestration [8,9]), effectively reducing the amount that would otherwise be emitted into the atmosphere [10–12]. CO_2 can be injected in salt caverns and saline domes and aquifers [13–15], underground coal deposits [16,17] (including fractured coal deposits [18]), depleted oil and gas reservoirs [19–21], sandstones [8], etc. The efficiency of sequestration in former underground porous reservoirs of oil and gas with finished exploitation or in saline aquifers depends on various factors, such as relative permeability [22], capillary pressure, well orientation, etc. [21,23]. CO_2 can also be biologically sequestered in the soil as a result of agricultural practices [24], in forests and biomass [25,26], etc. Further, it can be sequestered in gas hydrates [27,28], into ocean water forming bicarbonates [29], concrete [30,31], in soils enriched by biochar remains from pyrolysis of waste [32], fly coal ash [33], etc. New CO_2 capture methods based on oxy-fuel combustion and chemical looping combustion are also available [34]. CO_2 has favorable thermal properties and can be efficiently used in subsurface geothermal energy storages [35]. It can be also used simultaneously for sequestration and enhanced oil recovery [36,37].

This paper is organized as follows.

1. Section 1 explains why we have chosen to use OPM flow for simulation of CO_2 sequestration in underground saline aquifers [38, 39]. Explanation of types of CO_2 sequestration and rationalities for choosing between them are also given in Section 1.
2. Various studies explored numerous input parameters for CO_2 sequestration; for example Pavan and Govindarajan [40] examined the impact of porosity, permeability, injection velocity, permeability anisotropy, residual water saturation, and residual gas saturation; while Chen et al [41] explored the uncertainty of where the optimal location to place the monitoring well(s) is and what type of data (pressure, temperature, CO_2 saturation, etc.) are of influence for sequestration in geological formations, etc. [42]. In a similar way, carefully selected input parameters for this study are shown in Section 2 of this paper and will be further analyzed. The evaluated dataset has 1024 combinations of each parameter in the range of interest, all selected using Sobol's quasi-random sequence [43–46]. Using these inputs, an ensemble of reservoir models based on the Sleipner Benchmark model [47] was created; these were simulated by the OPM Flow software tool, then in post-processing a single output parameter was calculated that represents the most stable part of the trapped CO_2. The simulation involved injecting CO_2 over a 15-year period, with subsequent analysis extending over 200 years to model its behavior within the porous media of underground storage. The final analysis assessed the conditions after a total period of 215 years.
3. The data set of input parameters chosen for this study and the model output provided by OPM Flow were analyzed by a robust global sensitivity analysis method to evaluate which of the input parameters has the largest influence on the output [48–52]. The sensitivity analysis is shown in Section 3.
4. The same data set of inputs and output was used to feed a machine learning (ML) model [53] to provide a fast symbolic regression approximation of the OPM Flow results of this use case. The obtained symbolic regression formulas are shown in Section 4 of this paper. Symbolic regression is a computational technique based on machine learning that aims to find mathematical expressions or formulas that best fit a given dataset, typically without any prior knowledge of the underlying equations [54,55] with various applications [56–61].
5. Finally, conclusions are given in Section 5.

The data analysis within OPM Flow and post-processing of output was conducted by SINTEF in Oslo, Norway, utilizing patterns of data supplied by IT4Innovations, the National Center for Supercomputing of the Czech Republic in Ostrava, where sensitivity analysis and regression modeling were also further carried out. The ensemble of models was simulated on the Karolina cluster of the National Center for Supercomputing. The post-processing scripts, including the calculation of the secondary trapped CO_2 quantity, is available at https://github.com/kjetilly/ensemble_run_across.

2. Model, inputs and outputs

The model used as a base for this analysis is a coarsened version of the Sleipner CO_2 benchmark model (CODataShare, 2020; [62–64]). The Sleipner CO_2 storage site has been operated since 1996 by Equinor to avoid emissions from the production of natural gas in the Sleipner area, on the Norwegian continental shelf off the western coast of Norway. The CO_2 has been injected at a rate of approximately 1 Mt/yr into a saline aquifer that is part of the larger Utsira Formation. The 2019 benchmark model we have used as a base for our study incorporates baseline seismic, well logs, and time-lapse seismic data.

The model has been coarsened in the horizontal direction by a factor of 3 in both x- and y-directions, yielding an approximately 9 times smaller model than the original, with a total of 211 508 grid cells. This was done to save resources and speed up the runtime of the ensemble. We have not studied if this has an impact on the analysis results but leave that to future work.

The injected CO_2 can be entrapped through four distinct mechanisms: 1. Structural trapping, 2. Capillary trapping, 3. Solubility trapping, or 4. Mineral trapping. However, our simulation model omits consideration of mineral trapping, as it typically occurs over significantly longer durations compared to the first three mechanisms. In our final analysis, we concentrate on a single variable, termed 'Secondary Trapped CO_2', which represents the proportion of injected CO_2 that undergoes capillary entrapment or dissolution. While our model accounts for the impact of structural trapping, we do not specifically analyze CO_2 trapped through this mechanism in this study. The case is set up with 15 years of initial injection, in which approximately 13.3 million tons of CO_2 is injected, followed by a 200-year migration period. Our analysis focuses on the state at the end of the 215-year schedule.

Fig. 1 shows the time evolution of the secondary trapping, with capillary trapped CO_2 in orange, dissolved CO_2 in blue, and their sum ('Secondary Trapped CO_2') in green. The colored fields indicate the spread of values, with the thicker line indicating the mean and standard deviations shown as vertical bars. The rest of the CO_2 injected is structurally trapped.

From the output data several parameters may be explored in an uncertainty analysis of CO_2 sequestration, such as: amount of injected CO_2, amount of free-flowing CO_2, amount of leaked CO_2 from the structure, capillary trapped CO_2, dissolved CO_2, etc. The single output variable we have chosen to study is 'Secondary Trapped CO_2', defined as the mass fraction of injected CO_2 that is either capillary trapped or dissolved. The rest of the injected CO_2 will be structurally trapped in the model scenarios considered. In the models studied, we observe no or very little leakage of CO_2 from the storage area, and this in part is what motivates our choice of output variable, as it exhibits more interesting behavior for our ensemble. In case a leak in the caprock exists or develops, the structurally trapped but mobile CO_2 may be at risk of leaking out, whereas the capillary trapped or dissolved CO_2 is not. Therefore, our chosen variable may be used to differentiate between high-quality storage scenarios. For an overview of trapping mechanisms and more information about capillary trapping, see e.g. Ref. [65].

In our analysis this output parameter depends on the input parameters: 1. 'Temperature gradient', 2. 'Permeability of sand layers', 3. 'Permeability of shale between sand layers', 4. 'Permeability of feeder chimneys connecting sand layers', 5. 'Porosities of sand, shale, and feeders' and 6. 'Topography of top surface'. These 6 input parameters are used to generate a model ensemble, and then each model is simulated with OPM Flow. Then a post-processing calculation is done to find the 'Secondary Trapped CO_2' output parameter (dimensionless, and falling between 0 and 1, indicating the fraction of capillary trapped and dissolved CO_2; A value of 0 would mean that the injected CO_2 is only structurally trapped, while a value of 1 would mean that everything has been capillary trapped or dissolved in the brine). The input parameters for the CO_2 sequestration for the sensitivity analysis in the range of interest for this study are summarized in Table 1[1]

The theoretical aspects of variance-based sensitivity analysis were proposed by I. M. Sobol' [43]. An integral formulation can be found in Ref. [66]:

Given a mathematical model $y = f(x)$ for an n-dimensional input vector x denoting the parameters, where f is an integrable function from R^n to R^n and x lies within the unit hyper-cube H^n, defined as $\boldsymbol{H^n} = [0,1]^n \subset \mathbf{R}^n$ see Ref. [43,66]:

According to Sobol, $f(x)$ can be decomposed as a sum of elementary functions by a high-dimensional model representation (HMDR), Eq. (1):

$$f(x) = f_0 + \sum_{i}^{n} f_i(x_i) + \sum_{1 \le i < j \le n} f_{ij}(x_i, x_j) + \sum_{1 \le i_1 < i_2 < \ldots < i_s \le n} f_{i_1 i_2 \ldots i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) + \ldots + f_{12\ldots n}(x_1, x_2, \ldots, x_n) \quad \text{Eq. (1)}$$

Each term in this sum can be evaluated using multiple integrals as given in Eqs (2)–(4):

$$f_0 = \int_{\mathbf{H}^n} f(x) dx \quad \text{Eq. (2)}$$

$$f_i(x_i) = \int f(x) \prod_{k \ne i} dx_k - f_0 \quad \text{Eq. (3)}$$

$$f_{ij}(x_i, x_j) = \int f(x) \prod_{k \ne i,j} x_k - f_i(x_i) - f_j(x_j) - f_0 \quad \text{Eq. (4)}$$

[1] Permeability is given here in millidarcy (mD or md), which is not an SI unit but which is widely used in petroleum engineering. According to "The SI Metric System of Units and SPE Metric Standard" of the Society of Petroleum Engineers: http://www.spe.org/authors/docs/metric_standard.pdf (accessed on June 11, 2024), permeability has unit of the square of the length, where 1D = 0.9869233 $\mu m^2 \approx 1\ \mu m^2$. A medium with a permeability of 1D permits a flow of 1 cm^3/s of a fluid with viscosity 1 cP i.e. 1 cP (=1 mPa s) under a pressure gradient of 1 atm/cm acting across an area of 1 cm^2, where the shown value of conversion factor 0.9869233 is the same as reciprocal of 1.013250—the conversion factor from atmospheres to bar. However, SI definition of permeability in m^2 is: permeability of 1 m squared will permit a flow of 1 m^3/s of fluid of 1 Pa s viscosity through an area of 1 m^2 under a pressure gradient of 1 Pa/m. Darcy and the square micrometer (μm^2) can be considered equivalent when high accuracy is not needed or implied, with implication that 1.1D ~ 1.1 μm^2 and 5D ~ 5 μm^2. Porosity is dimensionless, and fall between 0 and 1, and is expressed as a fraction of pore volume and total volume.

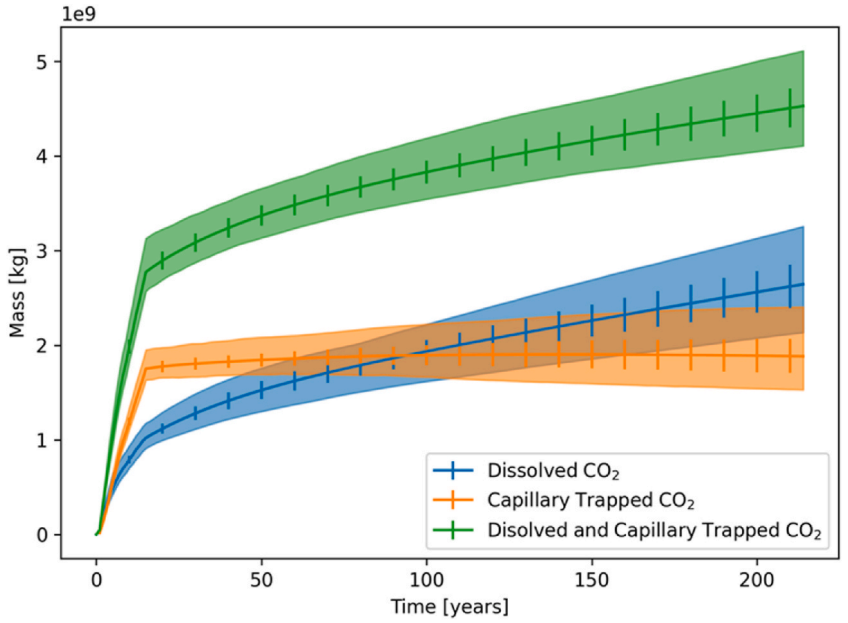


Fig. 1. Mass of dissolved CO_2, capillary trapped CO_2 and total secondary trapped CO_2 over time.

Table 1
Input parameters of the CO_2 sequestration in this study for sensitivity analysis, uncertainty analysis and machine learning predictions.

No.	Parameter	Notation	Minimum	Maximum
1	Temperature gradient (tegrad)	X_1	30 °C	40 °C
2	Permeability of sand layers (pesand)	X_2	1100 mD	5000 mD
3	Permeability of shale between sand layers (peshale)	X_3	0.00075 mD	0.0015 mD
4	Permeability of feeder chimneys connecting sand layers (pefeeder)	X_4	1100 mD	5000 mD
5	Porosities of sand, shale, and feeders (posand)	X_5	0.27	0.4
6	Topography of top surface (totopsur)	X_6	−10 m	+10 m

Sobol's global sensitivity method uses the ratio of the partial variances to the total variance to estimate the influences of the model variables on the model responses. The total variance of the model outputs and the partial variances of each elementary function are calculated as follows, respectively as given in Eqs. (5) and (6):

$$D = \int_{H^n} f^2(x)dx - f_0^2 \quad \text{Eq. (5)}$$

$$D_{i_1 i_2 \ldots i_s} = \int_0^1 f^2_{i_1 i_2 \ldots i_s}(x_{i_1}, x_{i_2}, \ldots x_{i_s}) \prod dx_{i_k}, 1 \le k \le s \quad \text{Eq. (6)}$$

By squaring and integrating over H^n, we can obtain Eq. (7)

$$D = \sum_{i=1}^{n} D_i + \sum_{1 \le i < j \le n} D_{ij} + \ldots + D_{12\ldots n} \quad \text{Eq. (7)}$$

The Sobol' sensitivity index is defined in Eq. (8):

$$S_{i_1 i_2 \ldots i_s} = D_{i_1 i_2 \ldots i_s}/D \quad \text{Eq. (8)}$$

The first-order sensitivity index $S_i = D_i/D$ represents the variance contribution of the variable x_i to the variance of the model responses, whereas the second-order index S_{ij} reflects the variance contribution of the variables x_i and x_j to the total variance of the model output. A detailed timeline of sensitivity analysis, including its development from local to global methods, can be found in Tarantola et al. [67].

The quasi-Monte Carlo method is used via the SALib software [68] with parameters of Table 1 used for selection of the presented ranges of values of parameters. An example of generated Sobol's quasi-random points is presented in Table 2 (a complete version is given as an Electronic Appendix of this paper). A quasi-Monte Carlo Sobol's sample in 1024 points ([43]; Sobol' et al., 2011; [45,46]) is used in the given region of interest for all 6 input parameters to generate the output 'Secondary Trapped CO_2'; i.e. ranges are set to

Table 2
Example of input parameters generated by the quasi-Monte Carlo method[a]. The column 'dummy' presents a non-significant parameter, which is used in sensitivity analysis to distinguish between significant and non-significant input parameters. The last column 'tco2' represents the output parameter extracted from OPM Flow results ('Secondary Trapped CO_2').

No.	X_1 tegrad	X_2 pesand	X_3 peshale	X_4 pefeeder	X_5 posand	X_6 totopsur	X_7 dummy	Y tco2
1	30.01465	2568.213	0.001086	2998.584	0.342427	6.884766	0.241699	0.333244732
2	35.87402	2568.213	0.001086	2998.584	0.342427	6.884766	0.241699	0.336624054
3	30.01465	3817.432	0.001086	2998.584	0.342427	6.884766	0.241699	0.336522887
4	30.01465	2568.213	0.001254	2998.584	0.342427	6.884766	0.241699	0.343364373
5	30.01465	2568.213	0.001086	4304.932	0.342427	6.884766	0.241699	0.330077595
6	30.01465	2568.213	0.001086	2998.584	0.38978	6.884766	0.241699	0.342296732
7	30.01465	2568.213	0.001086	2998.584	0.342427	4.130859	0.241699	0.333458577
8	30.01465	2568.213	0.001086	2998.584	0.342427	6.884766	0.338379	0.333399068
9	30.01465	3817.432	0.001254	4304.932	0.38978	4.130859	0.338379	0.350168533
10	35.87402	2568.213	0.001254	4304.932	0.38978	4.130859	0.338379	0.348800606
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1024	36.96777	2659.619	0.000937	4000.244	0.306499	−5.556640	0.572753	0.319556864

[a] Electronic Appendix of this paper provides a complete dataset (available there with more decimal places), i.e. 1024 rows of the inputs X_1 to X_6, of the "dummy" parameter and of the output Y.

cover the domain of variation of all input parameters. SINTEF provided IT4Innovations the data set 'Secondary Trapped CO_2' extracted from 1024 OPM Flow simulations, marked as Y in Table 2. The parameter labeled as X_7 in Table 2, referred to as the "dummy" parameter, holds no statistical significance in affecting the output parameter [69]. It serves solely for additional scrutiny of the performed sensitivity analysis. The sensitivity index of the "dummy" parameter is proportional to the numerical errors in the sensitivity analysis. This helps to distinguish among important and non-important input parameters [70–72].

The OPM Flow model is evaluated for the given sample, which represents a different combination of input parameters given in Table 2. To automate the evaluation of OPM Flow for various input parameters from Table 1, SINTEF developed the OPM-runner[2] python program, which is a handy tool to run OPM Flow with various parameters. This tool has been successfully used for automating the sensitivity analysis procedure, as the repeated evaluation of OPM Flow is a time-consuming procedure. Consequently, the output parameter 'Secondary Trapped CO_2' has been automatically extracted from the OPM Flow results, in Table 2 denoted by the symbol "tco2" as the output parameter Y.

The used quasi-Monte Carlo Sobol's sample is not random; instead, it is rather quasi-random ensuring that the same sequence consistently emerges in repeated experiments. An added benefit is its comprehensive coverage of the entire data space, without gaps (and especially without any huge random gaps). It is expected that the input parameters have uniform distribution (Fig. 2a–l) to maximize the information regarding the response of the OPM Flow model. The same applies for the "dummy" parameter (Fig. 3). The input parameter 'dummy' is modeled by a uniform distribution with a minimum of 0 and a maximum of 1, see also Table 2. We note that even though the histograms of the dummy parameter and the "totopsur" parameter look similar, they are in fact distinct variables with different scatters.

3. Sensitivity analysis

The uncertainty of the model response, shortly "uncertainty analysis" aims to quantify the variability of the model output due to the variability of the model inputs [52]. The quantification is usually performed by the summarization of the model output data by exploratory data analysis [73].

The uncertainty analysis involves several steps: Defining the uncertainty of input variables, such us setting the minimum and maximum values of the input parameters for sampling (Table 1 represents input parameters of the carbon sequestration case study); generating a sample of the model inputs, as described in Section 2; generating an ensemble of simulation models corresponding to the input parameters and simulating each ensemble member with OPM Flow; evaluating the simulation results for each ensemble model to calculate the output Y with the resulting dataset of X_1-X_7 and Y shown partially in Table 2 and completely in the Electronic Appendix; and representing the model output.

The output of the uncertainty analysis is the unconditional distribution p(y), which represents the overall uncertainty of the model output. Fig. 4 shows the overall uncertainty of the model output Y: 'Secondary Trapped CO_2', which is represented by a histogram.

Statistical results of the output parameter Y: 'Secondary Trapped CO_2' are summarized in Table 3. The model output varies from circa 0.309 to 0.385.

Fig. 5a–f presents results of the sensitivity analysis by sensitivity plots where on the x-axis are input parameters X_1 to X_6 from Table 2, whereas the model output Y: 'Secondary Trapped CO_2' is on the y-axis. The trend has been fitted by a polynomial regression model. Fig. 6 provides the same for the "dummy" parameter X_7.

The concentration of dots in Fig. 5 clearly shows that the variance of the output Y:' Trapped CO_2' is mainly explained by X_3: 'peshale' and X_5: 'posand'. X_3: 'peshale' is directly proportional to the output Y. Low values of X_3: 'peshale' have a trend to cause small

[2] A small and simple tool to run OPM Flow with various parameters: https://github.com/kjetilly/opm-runner (accessed on June 12, 2024).

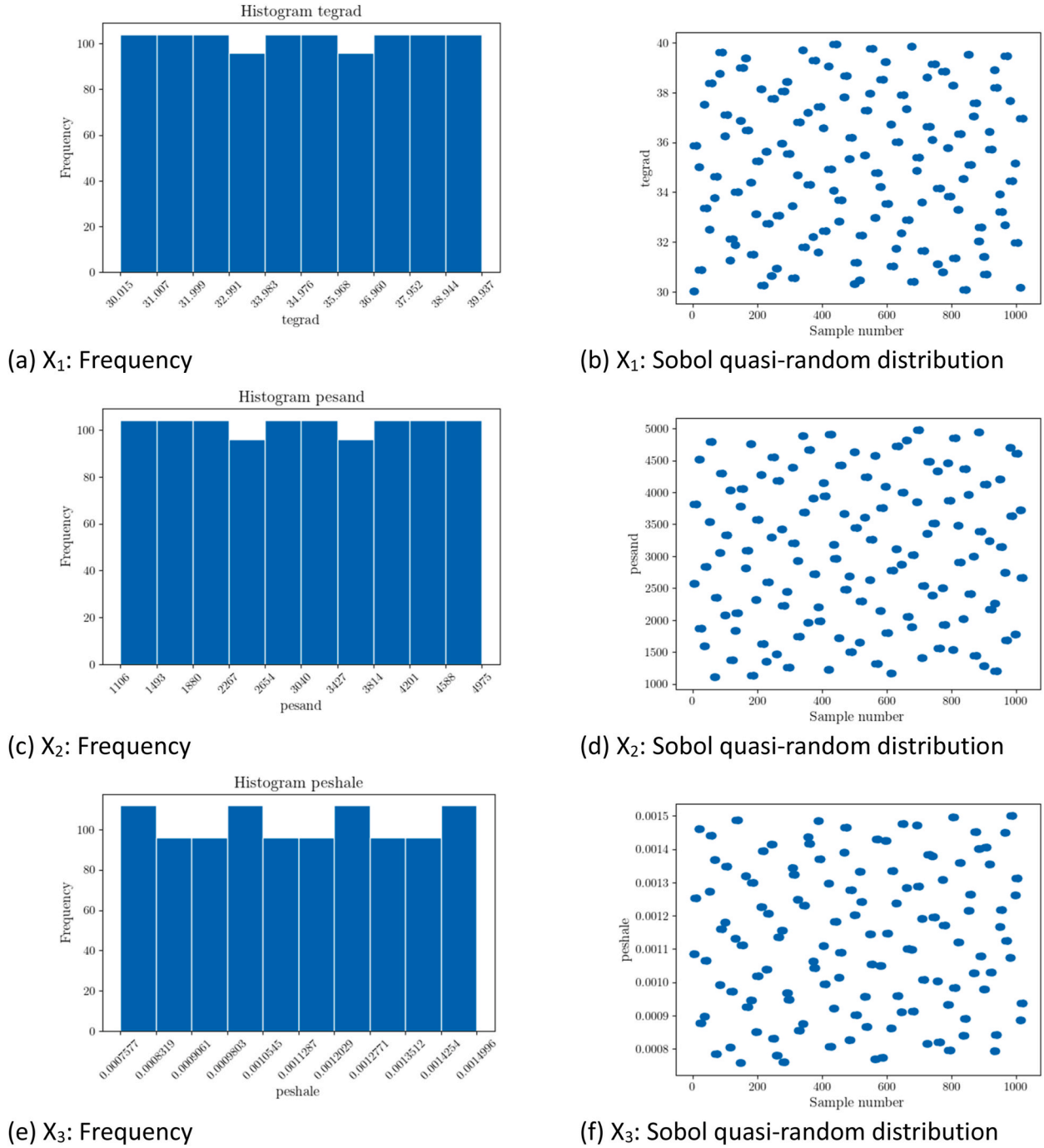
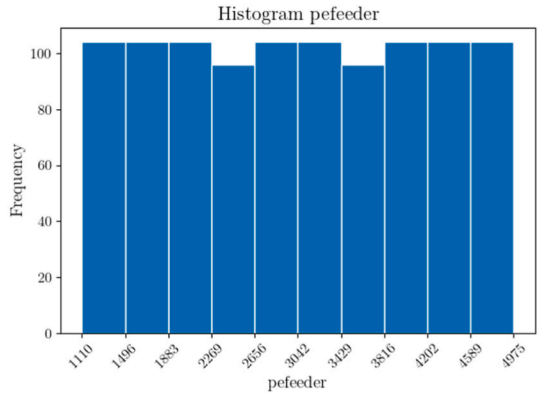


Fig. 2. Distribution of the input parameters X_1-X_6.

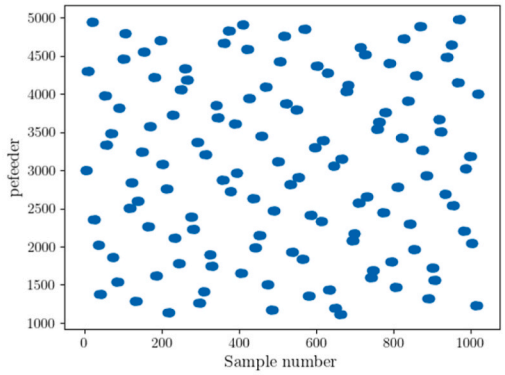
values of Y and vice versa. The same observation (but less visible) is valid also for X_5: 'posand'. In contrast, 'Secondary Trapped CO_2' is inversely proportional to X_4: 'pefeeder'. The influence of X_2: 'pesand' to the output Y exists but is very limited.

The scatter of dots in all other cases is large, while the associated trendlines are almost always horizontal and with a very small curvature indicating that these input parameters have little influence on output.

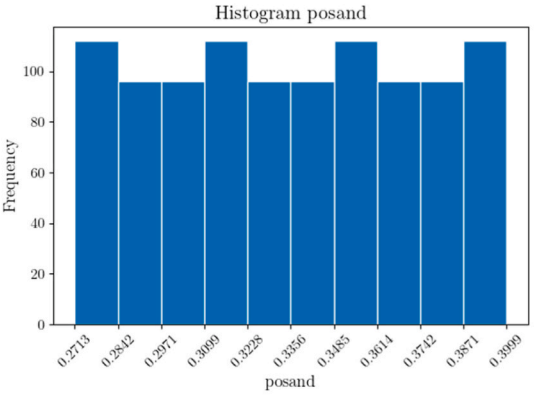
The open-source SALib software is used for Sobol's sensitivity analysis, which computes the first-order and the total sensitivity indices. The sensitivity indices in Sobol's sensitivity analysis measure the contribution of a single variable or a group of variables to the total variance [43]. These indices are fractions of partial and total variances, and they range from 0 to 1 [45]. First-order indices,



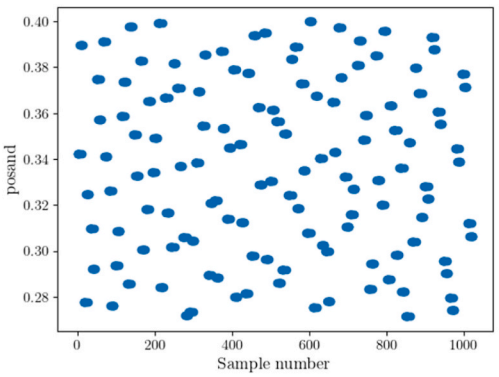
(g) X_4: Frequency



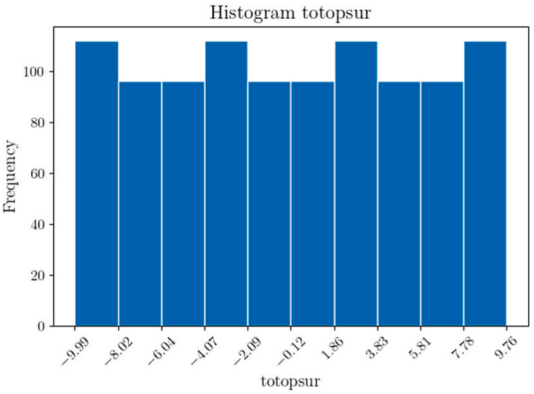
(h) X_4: Sobol quasi-random distribution



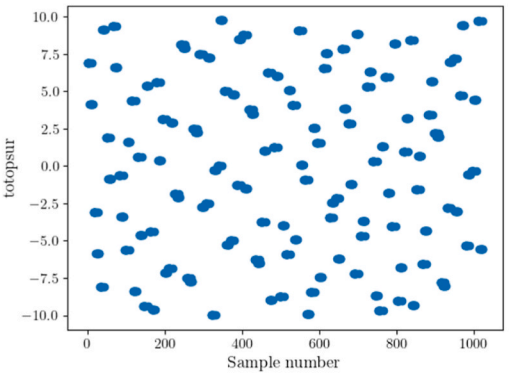
(i) X_5: Frequency



(j) X_5: Sobol quasi-random distribution



(k) X_6: Frequency



(l) X_6: Sobol quasi-random distribution

Fig. 2. (*continued*).

sometimes also called “main effect indices”, measures the contribution to the output variance by a single model input alone. As discussed in SALib,[3] the total-order index measures the contribution to the output variance caused by a model input, including both its first-order effects (the input varying alone) and all higher-order interactions. The first-order sensitivity index is very close to the total sensitivity index. This means that the model is additive with no significant interactions among input parameters. Sensitivity indices

[3] SALib - Sensitivity Analysis Library in Python – Basics: https://salib.readthedocs.io/en/latest/user_guide/basics.html#what-is-sensitivity-analysis (accessed on October 29, 2024).

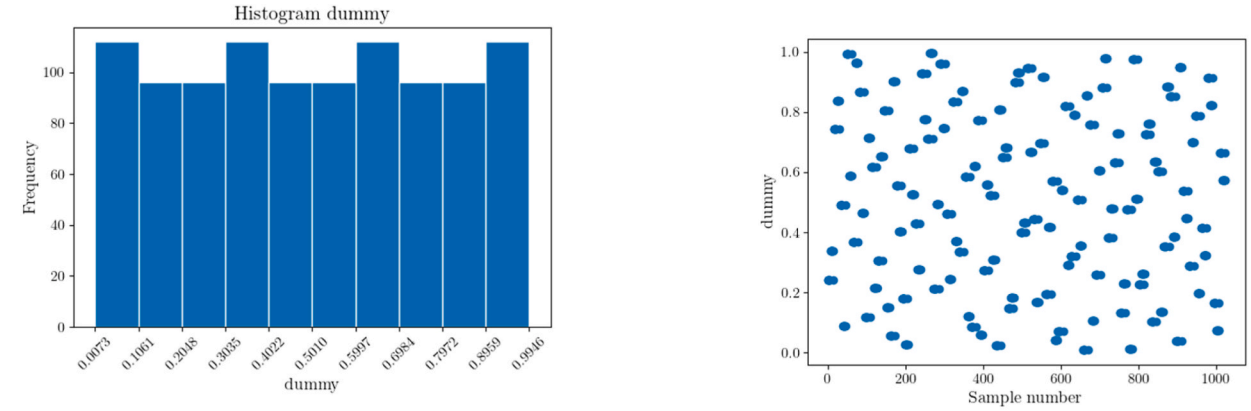


Fig. 3. Distribution of the "dummy" parameter X_7.

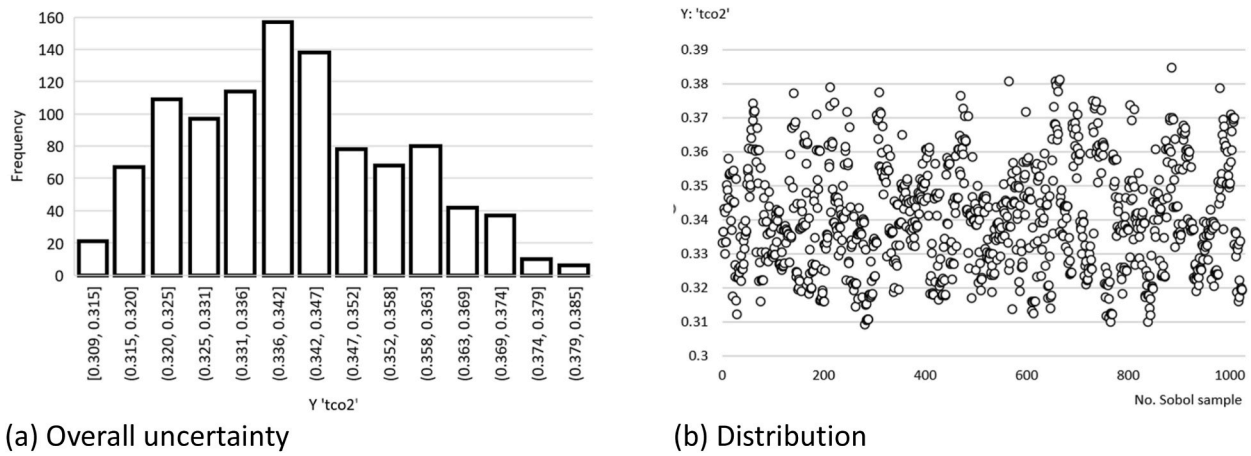


Fig. 4. Model output 'Secondary Trapped CO_2' represented by the symbol Y.

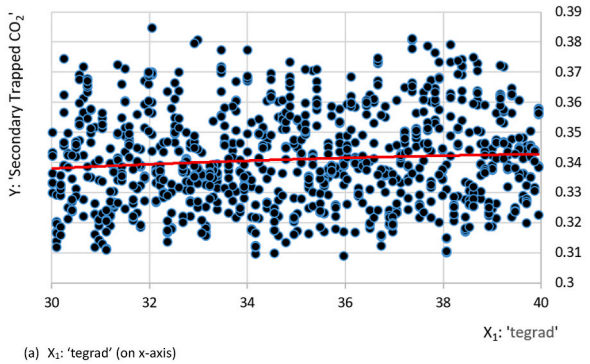
Table 3
Statistical analysis of the model output Y: 'Secondary Trapped CO_2' for 1024 Sobol's quasi-Monte Carlo points.

count	mean	standard deviation	min	0.25 quantile	0.5 quantile	0.75 quantile	max
1024	0.340913	0.015648	0.309241	0.328218	0.339245	0.351542	0.384821

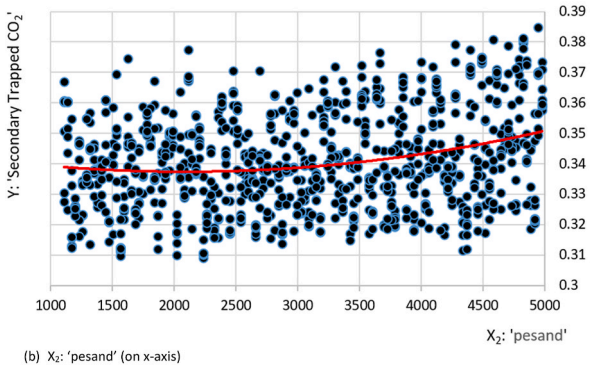
very close to zero indicate a negligible effect to the model output.

Fig. 7 summarizes the results of the sensitivity analysis.

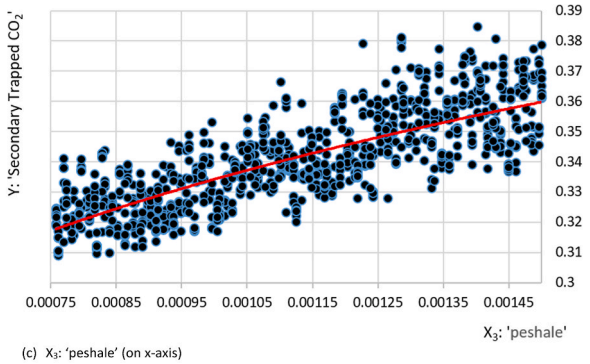
1. The most important parameter is X_3: 'peshale'. The total sensitivity index (0.61) is the same as the first-order sensitivity index. This means that the output variance is fully explained by the variance in X_3: 'peshale', without any additional variance caused by the interaction between them (negligible interactions).
2. The parameters X_5:'posand', X_4: 'pefeeder' and X_2 'pesand' have a much lower importance, as their total sensitivity indices are 0.2, 0.14 and 0.08 respectively. First-order sensitivity indices of these parameters, 0.17, 0.12 and 0.06 respectively, are close to the total order indices which indicates that interactions exist, but they are limited.
3. The input parameter X_1: 'tegrad' has negligible importance.
4. The parameter X_6:'totopsur' is not important at all for the ensemble under study. The same can be concluded also for the parameter X_7:'dummy', by its definition.



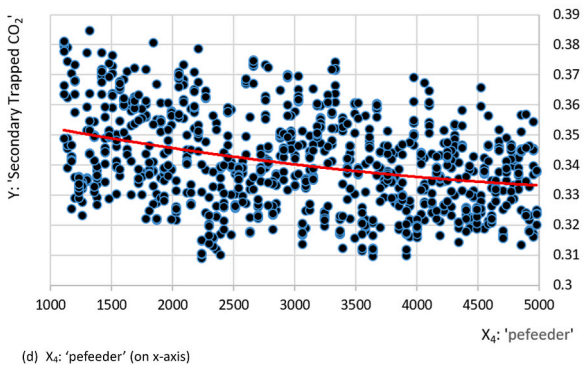
(a) X_1: 'tegrad' (on x-axis)



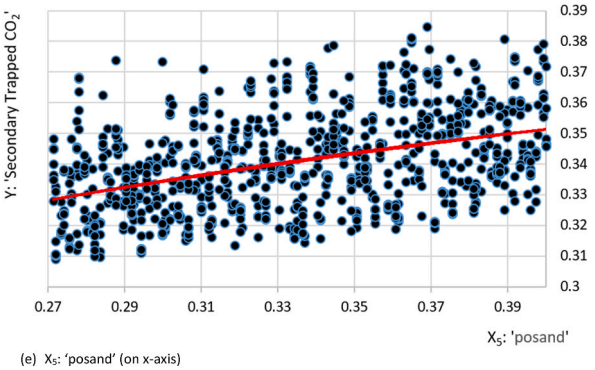
(b) X_2: 'pesand' (on x-axis)



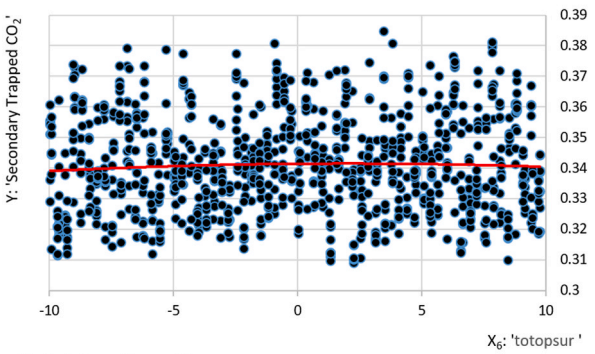
(c) X_3: 'peshale' (on x-axis)



(d) X_4: 'pefeeder' (on x-axis)



(e) X_5: 'posand' (on x-axis)



(f) X_6: 'totopsur' (on x-axis)

Fig. 5. A plot of the chosen input variables X (on x-axis) and the model output Y: 'Secondary Trapped CO_2' (on y-axis): (a) X_1: 'tegrad', (b) X_2: 'pesand' (c) X_3: 'peshale', (d) X_4: 'pefeeder', (e) X_5: 'posand', (f) X_6: 'totopsur'; Scatter plot represents real data while line represents trend.

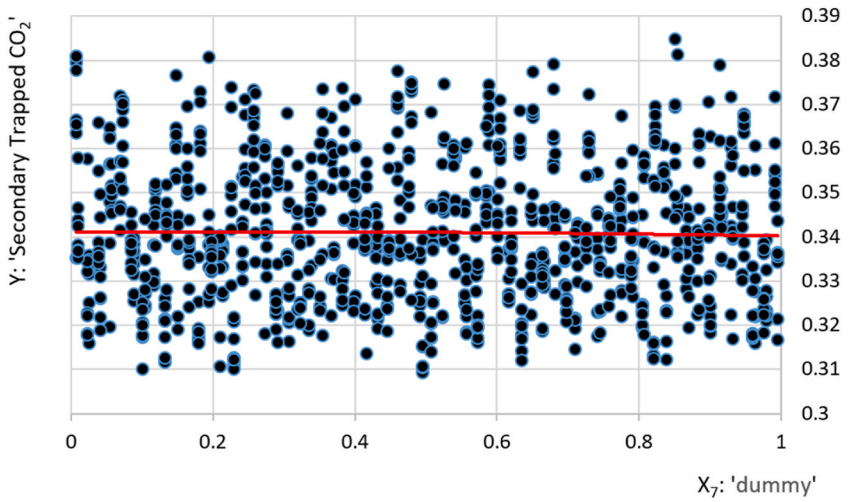


Fig. 6. A plot of the "dummy" input variable X_7 (on x-axis) and the model output Y: 'Secondary Trapped CO_2' (on y-axis); Scatter plot represents real data while line represents trend.

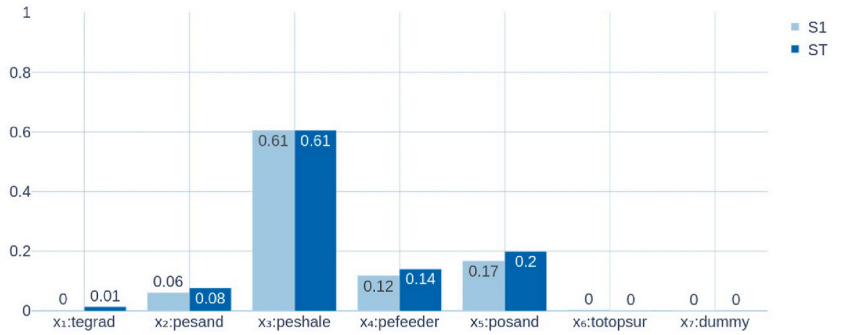


Fig. 7. Model sensitivity indices: (S1) first-order and (ST) total.

3.1. Convergence studies

While there are convergence guarantees for quasi-Monte Carlo, and specifically the Sobol sequence, said convergence guarantees rely on the Hardy-Krause variation of the function, which in most practical cases is not available [74]. To remedy this for the current problem, we have run convergence studies for both the expectation and the first order sensitivity indices to confirm the convergence rate and applicability of QMC for the given problem.

We first generate a reference solution using 2753 samples. We then plot the error (deviation) between the solution using M samples and the reference solution as a function of M for time 107 years and 214 years. A linear regression between log(M) and log(error) is done to estimate the convergence rate for both timesteps. We carry out this convergence study for the mean (expectation) in Fig. 8, where we see an overall convergence of order approximately 1, which is to be expected for smooth observables. To gauge convergence of the first order Sobol sensitivity index (S1), we consider the deviation from the reference solution Sobol S1 index for each parameter (pesand, peshale, etc) and take the $\widehat{\ell}\infty$-norm of the errors to measure joint convergence, the result of which is displayed in Fig. 9. We note that all obtained convergence rates measured are close to 1, meaning we are close to the optimal convergence rate for quasi-Monte Carlo. Furthermore, we note that the errors obtained at the sample size we are using for the rest of the paper, 1024 samples, are well within the tolerances required to order the parameters from least to most significant.

3.2. Time evolution of the sensitivity

For each year between 0 and 214, we plot the computed first Sobol sensitivity index (S1), and plot this as a function of time in Fig. 10. As we can clearly see from this plot, the relative sensitivity of each parameter changes over time. Indeed, the permeability of the shale layers is among the least important parameters even after 50 years, and only becomes dominant after more than 100 years of simulation.

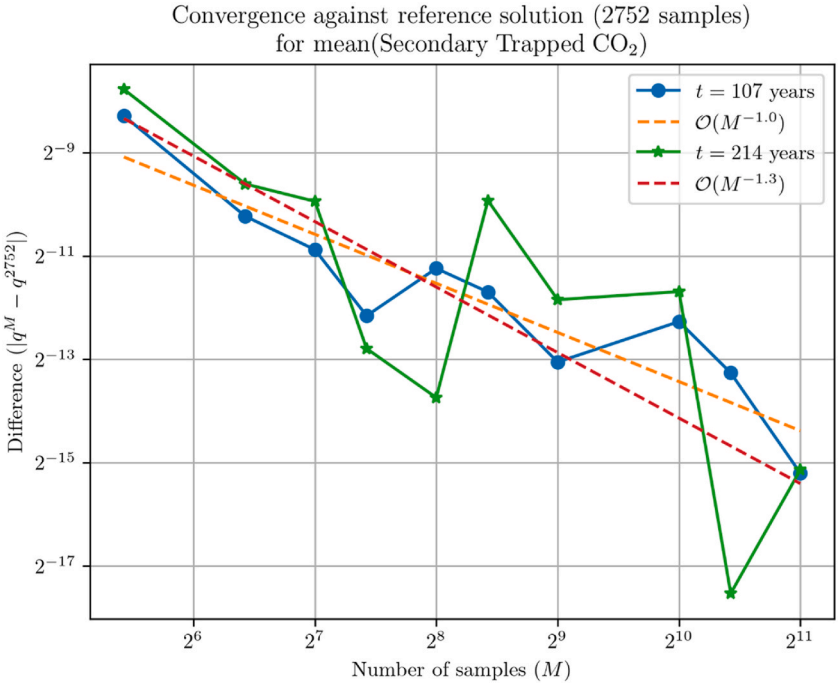


Fig. 8. Convergence of the mean of secondary trapped CO_2 against a reference solution with 2752 samples.

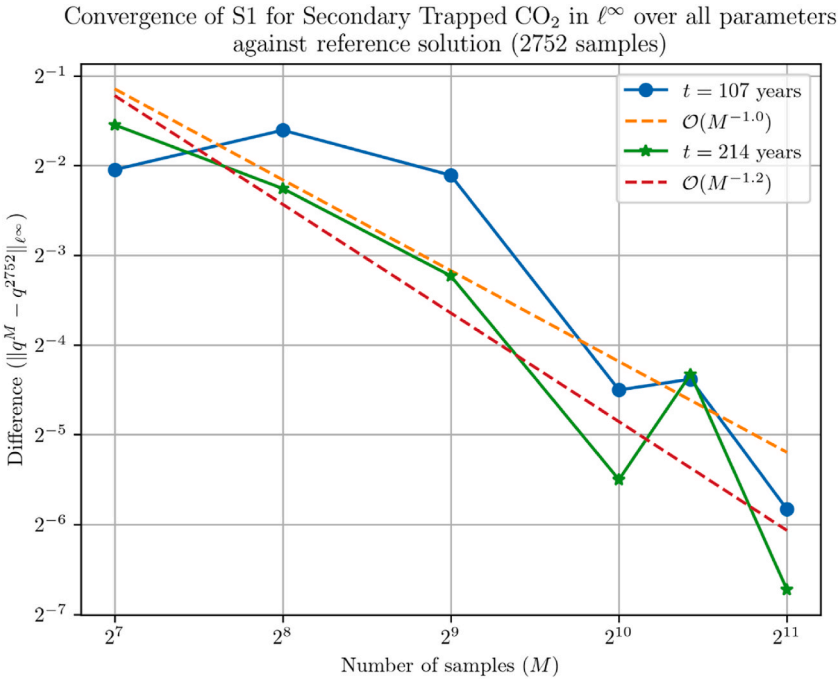


Fig. 9. Convergence of the first order Sobol index for the observable secondary trapped CO_2 in the infinity-norm over all parameters.

4. Regression models

4.1. Symbolic regression

Symbolic regression is a computational technique that aims to discover mathematical expressions or equations that best fit a given

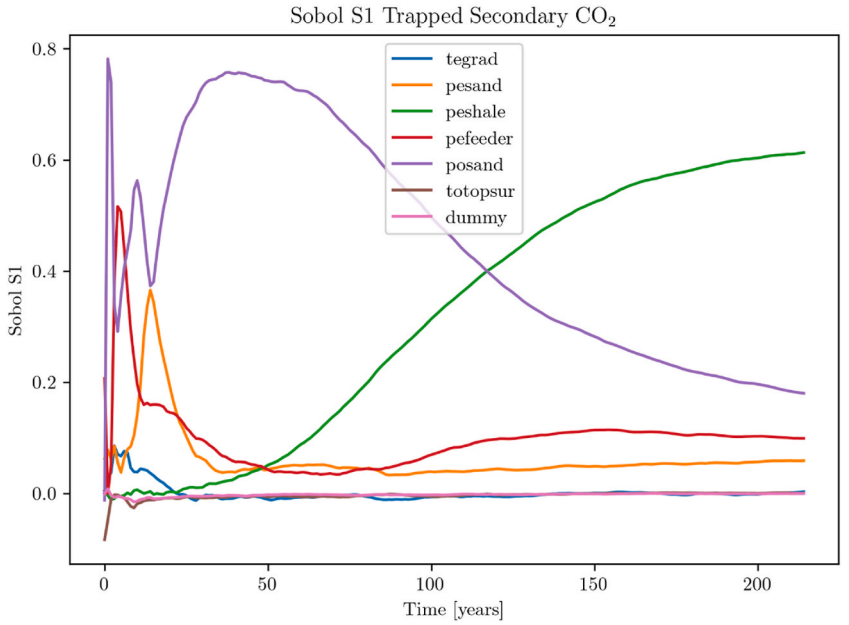


Fig. 10. Time evolution of the Sobol indices for trapped secondary CO_2.

dataset, without requiring any predefined mathematical model structure [54]. It essentially seeks to find symbolic expressions that represent the relationships within the data, providing explainable (symbolic) models.

Symbolic regression finds the best model for the given dataset by exploring different mathematical expressions. Symbolic regression does not have a fixed model to begin with. It creates symbolic expressions by randomly mixing mathematical elements, such as operators, functions, and constants. Thus, symbolic regression can be viewed as an automated machine learning method. We used the same input data set for the machine learning task as for the previous uncertainty and sensitivity analysis tasks.

The aim of the machine learning task in this study is to find a regression model, which estimates the relations between input parameters of the output parameter provided by the OPM Flow software. An additional task is to verify the outcomes of the sensitivity analysis, as symbolic regression possesses the capability to incorporate solely the input parameters that significantly impact the output, while disregarding those with minimal or negligible influence [75,76].

The symbolic regression tool Eureqa has been used to analyze the data from Table 2 (a full data set is in Electronic Appendix of this paper), where input parameters are X_1 to X_6 while the output is Y: 'Secondary Trapped CO_2'. (The 'dummy' parameter was used only for sensitivity analysis, not for regression.) Eureqa is a software tool designed for symbolic regression, aiming to discover mathematical formulas that accurately represent relationships within data [77]. All obtained machine learning models were trained using 512 simulations and successfully validated by 512 simulations of the OPM flow. The simulations were split randomly for training and testing. The symbolic regression procedure generated a variety of symbolic models with different complexity and precision from the analyzed data set. The most suitable approximations of the output parameter Y: 'Secondary Trapped CO_2' with increasing number of the input parameters X (1D to 6D models) obtained by symbolic regression are summarized in Table 4.

Symbolic regression models from Table 4 are nonlinear. All models (except 1D model) have interactions, for example ($X_3 \times X_5$) and $X_3 \times X_2^3$. However, to be consistent with the results of the sensitivity analysis, we show in the next subsection "Multiple linear regression", that these interactions are not very significant.

Fig. 11 presents the scatter plot of the Y: 'Secondary Trapped CO_2' values from the OPM flow software on x-axis – Data and machine

Table 4
Obtained symbolic regression models.

Number of input parameters[a]	Approximations[b]	Maximal absolute error	R-squared
1 – 1D model	$0.415–7.638e\text{-}5/X_3$	0.027	0.59
2 – 2D model	$0.414–2.454e\text{-}5/(X_3 \times X_5)$	0.020	0.76
3 – 3D model	$0.301 + 162.727 \times X_3 \times X_5\text{-}6.591{\cdot}10^{-6} \times X_4$	0.013	0.89
4 – 4D model	$0.274 + 158.407 \times X_3 \times X_5 + 0.007 \times X_2/X_4$	0.010	0.95
5 – 5D model	$0.269 + 155.313 \times X_3 \times X_5 + (X_1 + 2.13{\cdot}10^{-7} \times X_3 \times X_2^3)/(456.48 + X_4)$	0.007	0.98
6 – 6D model	$0.269 + 154.788 \times X_3 \times X_5 + (X_1 + 2.139{\cdot}10^{-7} \times X_3 \times X_2^3 + \cos(X_6))/(480.002 + X_4)$	0.006	0.98

[a] Dimensions D are correlated with Fig. 11.
[b] Automatized tool is written in Python to transform the coefficients of the symbolic models to the desired precision. We use 3 digits for coefficients, as the precision of symbolic models remains approximately the same. Models with 3-digit coefficients are better for the human eyes. Coefficients with 3 digits provide a good compromise between complexity and accuracy.

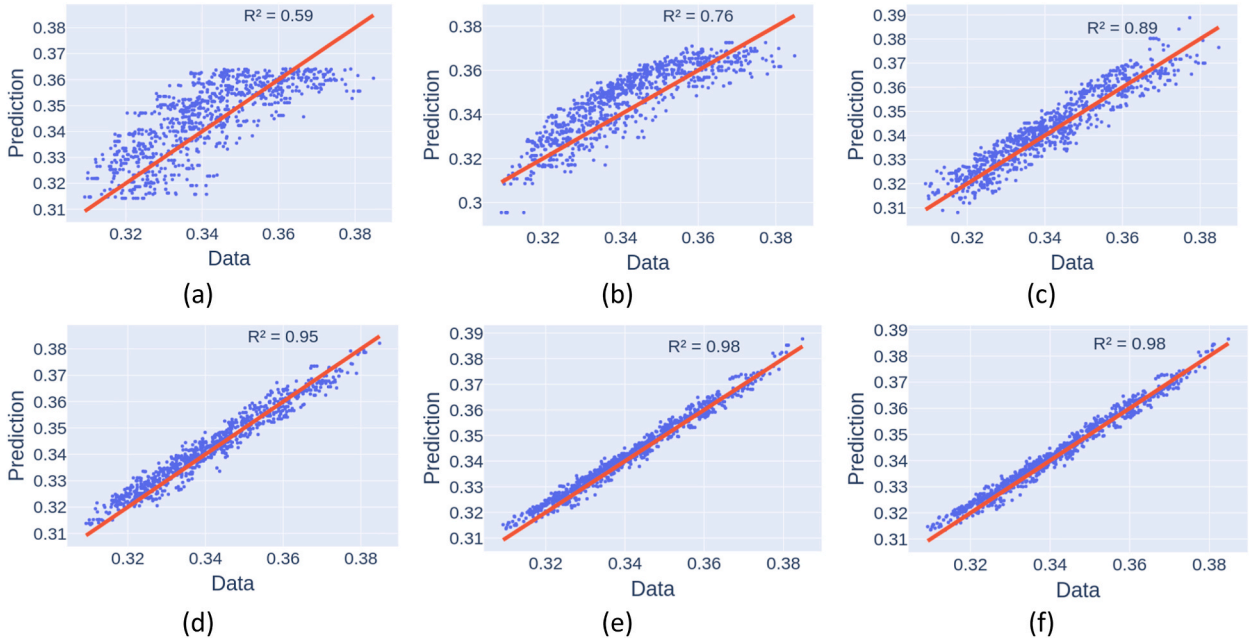


Fig. 11. Scatter plot and trendline of observed OPM Flow Y: 'Secondary Trapped CO_2' values from the OPM Flow software on x-axis – Data and machine learning predictions on y-axis – Prediction of (a) one-parameter model – 1D model, (b) two-parameter model – 2D model, (c) three-parameter model – 3D model, (d) four-parameter model – 4D model, (e) five-parameter model – 5D model, and (f) six-parameter model – 6D model (Dimensions D are correlated with the number of input parameters from Table 4).

learning predictions on y-axis – Prediction of (a) one-parameter model – 1D model, (b) two-parameter model – 2D model, (c) three-parameter model – 3D model, (d) four-parameter model – 4D model, (e) five-parameter model – 5D model, and (f) six-parameter model – 6D model. The R^2 value, also known as the coefficient of determination, measures the proportion of variation in the dependent variable explained by the independent variable or how well the regression model fits the data. The R^2 value ranges from 0 to 1, and a higher value indicates a better fit.

Developed symbolic regression models with 1 and 2 parameters have a poor fit, see Fig. 11. However, machine learning models developed with 3 or more parameters fit the analyzed carbon sequestration use case very well. For example, Fig. 11 shows that the coefficients of determination (R-squared) are very close to one: for the 3D model, $R^2 = 0.89$, whereas for the 4D model, $R^2 = 0.95$. Fig. 11 also shows that prediction errors are very small, see Fig. 11 (d), (e), (f).

4.2. *Multiple linear regression*

The results of the sensitivity analysis indicate that the model of 'Secondary Trapped CO_2' is additive with no significant interactions among input parameters. Consequently, we can expect that the model of 'Secondary Trapped CO_2' can be accurately expressed by approximations, which are linear in input parameters. For this reason, this section tests the ability of multiple linear regression to predict the output parameter Y: 'Secondary Trapped CO_2'

The Python open-source machine-learning library scikit-learn is used for multiple linear regressions. The most suitable linear regression approximations of the output parameter Y: 'Secondary Trapped CO_2' with increasing number of the input parameters X (1D to 6D models), are summarized in Table 5.

4.3. *Validation of sensitivity analysis results with machine learning models*

The discovered machine learning approximations with at least 3 parameters have acceptable maximum absolute errors for doing quick what-if analyses. Moreover, our results show that regression agrees with sensitivity analysis. Thus, regression validates sensitivity analysis, which is interesting: It indicates that regression can be a useful model for surrogate modeling in this context.

Both symbolic regression and linear regression models were able to accurately identify the most important input parameters according to the sensitivity analysis results. The r-squared values are approximately equal to the sum of the first-order indices of the input parameters. Findings based on the results from Tables 4 and 5 can be summarized.

1. For the models with one input parameter – 1D models, we consider the most important input parameter X_3:'peshale' from sensitivity analysis. The obtained 1D model depends only on X_3:'peshale' and fits the given input data set with R, which is close to its

Table 5
Multiple linear regression models.

Number of input parameters	Approximations[a]	Maximal absolute error	R-squared
1 – 1D model	$0.277 + 56.73 \times X_3$	0.033	0.62
2 – 2D model	$0.218 + 56.73 \times X_3 + 0.1768 \times X_5$	0.026	0.80
3 – 3D model	$0.232 + 56.71 \times X_3 + 0.176 \times X_5 - 0.000004687 \times X_4$	0.018	0.91
4 – 4D model	$0.224 + 56.5 \times X_3 + 0.1761 \times X_5 - 0.000004683 \times X_4 + 0.000002865 \times X_2$	0.012	0.96
5 – 5D model	$0.206 + 56.51 \times X_3 + 0.1761 \times X_5 - 0.000004708 \times X_4 + 0.000002859 \times X_2 + 0.0005133 \times X_1$	0.011	0.96
6 – 6D model	$0.206 + 56.5 \times X_3 + 0.1763 \times X_5 - 0.000004709 \times X_4 + 0.000002855 \times X_2 + 0.000514 \times X_1 + 0.00006851 \times X_6$	0.011	0.96

[a] as in Tables 4 and 3 digits for coefficients are used; precision of linear models remains approximately the same.

first-order and total order indices (0.61). This indicates that the discovered 1D linear approximation has R-squared close to the theoretical maximum identified by sensitivity analysis.

2. 2D regression models depend on X_3 and X_5; Their r-squared values are close to the sum of first-order indices of these two input parameters (0.78). It is not a paradox, as sensitivity indices (such as Sobol indices) are estimated from a finite sample of model evaluations [48].
3. 3D regression models depend on X_3, X_5 and X_4; Their r-squared values are approximately the sum of first-order indices of these three input parameters (0.9).
4. 4D regression models depend on X_3, X_5, X_4 and X_2; Their r-squared values are the sum of first-order indices of these four input parameters (0.96).
5. 5D regression models depend on X_3, X_5, X_4, X_2 and X_1; Their r-squared values are close to the sum of first-order indices of these five input parameters (0.97).
6. 6D regression models depend on all input parameters; Their r-squared values are close to the sum of first-order indices of these six parameters (0.97). Inclusion of X_6 does not increase the accuracy but only complexity.

We also evaluated the performance of multiple linear regression in predicting the secondary trapped CO_2 using the training sets ranging from 8 to 1024 quasi-Monte Carlo samples. To streamline the evaluation process, we incremented the training set size by 2, resulting in the evaluation of 508 linear models. We used the Python package scikit-learn for regression modelling. The statistical metrics R-squared and the maximum residual error (MAE) were calculated for each linear model by comparing the secondary trapped CO_2 predictions against the entire testing dataset, which comprised 1024 Quasi Monte-Carlo samples, see Figs. 12 and 13.

Fig. 12 shows the relationship between the size of the training set (x-axis) and the R-squared value (y-axis) of the testing data set. As the size of the training set increases, the R-squared value, which measures the goodness of fit, tends to increase. For example, the R-squared coefficient reaches the value of 0.96, when the training set has 354 Quasi Monte Carlo simulations. Beyond this point, the R-squared value tends to stabilize above 0.96: In only 49 out of 670 cases (7.3 %), the R-squared value slightly dips below 0.96.

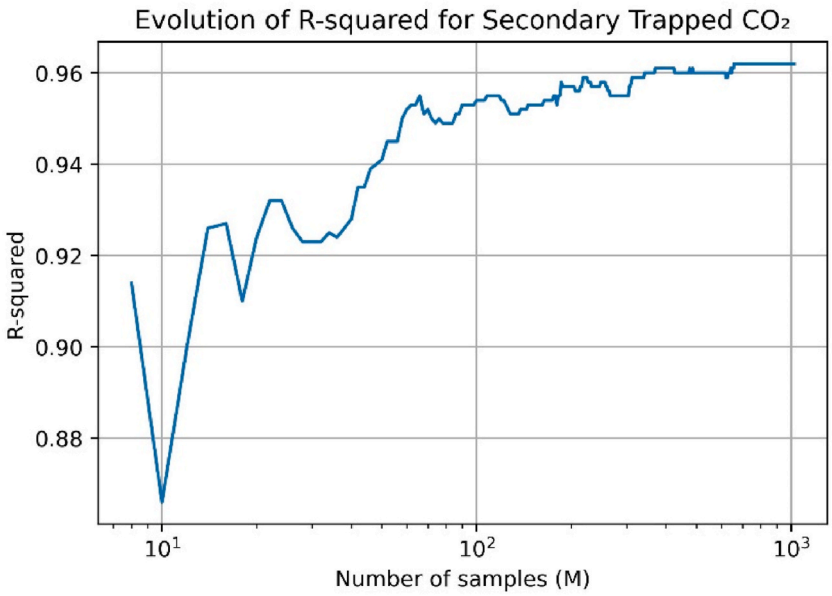


Fig. 12. Effect of training set size on R-Squared.

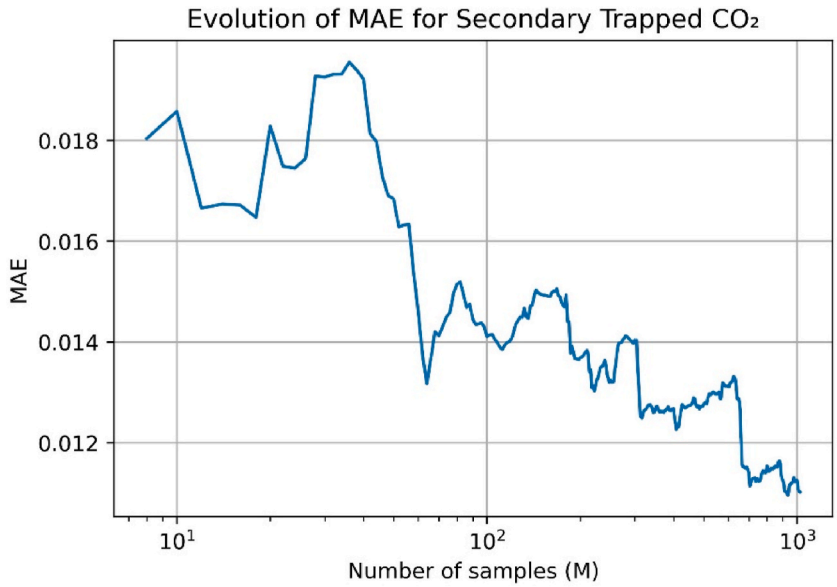


Fig. 13. Effect of training set size on MAE.

Additionally, the maximum residual error (MAE) metric was analyzed to assess the accuracy of the predictions. As shown in Fig. 13, the MAE decreases as the size of the training set grows, indicating improved prediction accuracy. These analyses help in understanding how the amount of training data impacts the ML model's performance for predicting the secondary trapped CO_2.

4.4. Regression-based sensitivity measures

In this section, we will analyze the ability of regression-based sensitivity measures to perform sensitivity analysis. Results of the multiple regression analysis show that the model response is linear and the coefficient of determination is large. For this reason, we test regression-based sensitivity measures [78,79] for the carbon sequestration case study.

4.4.1. Ordinary regression coefficient (ORC)

Let us assume the multiple linear model, Eq. (9):

$$y = b_0 + \sum_{i=1}^{N} b_i x_i, \qquad \text{Eq. (9)}$$

where b_0 is called intercept and ORC regression coefficients b_i can be considered as absolute sensitivity measures reflecting the direct influence of each input parameter on the output [78], (Iooss, B., Lemaître, P., 2015).

We used Pingouin[4] open-source library for estimating regression-based sensitivity measures. Multiple linear regression results of the carbon sequestration case study for the first 8 Sobol's quasi-random points are summarized in Table 6.

The results confirm the ability of the multiple linear model to predict the secondary trapped CO_2, as both the R-squared (0.999921) and adjusted R-squared (0.999445) coefficients are very close to 1. As we analyze here regression-based sensitivity measures, regression coefficients are sorted by the absolute values: x_3:peshale is on the top, followed by x_5:posand. Thus, only the two most important parameters are successfully detected by ORC, as ranking of the remaining parameters is totally different than the ranking of global sensitivity analysis.

The values of standard errors (se) estimate the standard deviation of the coefficients: a small standard error indicates a more precise estimates of the coefficient, and vice versa. T-values are the coefficients divided by their standard errors (T = coef/SE). Higher absolute T-values indicate that the corresponding coefficient is more significantly different from zero.

P-value for regression analysis represents the probability of obtaining a test statistic at least as extreme as the one observed, assuming that the null hypothesis (which states the coefficient is zero), is true. Very small p-values (typically less than 0.05) indicate that the corresponding regression coefficient is statistically significant from zero. We can see that even 8 Sobol samples are enough to confirm that the following input parameters have a significant effect on the output parameter: peshale:$x_3 = 59.87477$ with p-value of 5.98×10^{-8}, followed by posand:$x_5 = 0.189528$ (p-value = 0.00894) and the intercept of the model (intercept = 0.187702, p-value is 0.00484). These statistically significant parameters are highlighted in Table 6 in bold.

[4] https://pingouin-stats.org/.

Table 6
Summarized results of multiple linear regression for 8 Sobol's quasi-random points.

i	names	coef	se	T	Pval	CI[2.5 %]	CI[97.5 %]
3	**peshale**	59.87477	5.63E-06	1.06E+07	**5.98E-08**	59.8747	59.87485
5	**posand**	0.189528	2.66E-03	7.12E+01	**8.94E-03**	0.155715	0.223341
0	**Intercept**	0.187702	1.43E-03	1.31E+02	**4.84E-03**	0.169551	0.205853
1	tegrad	0.000564	2.15E-05	2.62E+01	2.43E-02	0.00029	0.000837
2	pesand	0.000003	1.01E-07	2.54E+01	2.50E-02	0.000001	0.000004
4	pefeeder	-2E-06	9.65E-08	−2.57E+01	2.47E-02	-4E-06	-1E-06
6	totopsur	−0.00005	4.58E-05	−1.08E+00	4.74E-01	−0.00063	0.000532

Although ORC is easy to compute and interpret, the method is not very suitable for sensitivity analysis, as ORC coefficients are influenced by the units of x_i. Consequently, ORC coefficients cannot be used for measuring the relative importance [78], (Iooss, B., Lemaître, P., 2015).

However, regression results from 8 quasi-random samples show that the estimated upper confidence limit (CL) of the intercept (0.205853) is very close to the intercept of the regression model trained for 1024 simulations (0.206).

4.4.2. *Standardized regression coefficient (SRC)*

To reduce the issue of differing units in the ORC, standardization can be an effective solution. The standardization involves adjusting the data so that the mean is zero and the variance is one, Eq. (10):

$$\mathrm{SRC}(y, x_i) = b_i \frac{\sigma_i}{\sigma_y}, \qquad \text{Eq. (10)}$$

where b_i is the ORC coefficient, σ_i is the standard deviation of the input parameter x_i and σ_y is the standard deviation of the model output y, respectively.

For independent input variables, the standardized regression coefficients (SRCs) can be used as relative sensitivity measures, as they measure the fraction of the uncertainty (variance) in y contributed by x_i [78], (Iooss, B., Lemaître, P., 2015). SRCs tend to perform well when the correlation between the input variables is weak and the model coefficient of determination R_y^2 approaches 1.

Results for the first 10 Sobol's Quasi Monte-Carlo samples show that the absolute values of standardized regression coefficients correlate to global sensitivity indices. Thus, sorted absolute values SRCs can be used in this case for ranking the importance of input parameters, see Table 7:

These results confirm the ability of the standardized multiple linear model to predict the secondary trapped CO_2, as both the R-squared (0.997155) and adjusted R-squared (0.991464) coefficients are very close to 1.

Moreover, when the absolute values of SRC coefficients are sorted in descending order, the ranking is compatible with the global sensitivity method: The sorted order of SRCs for all input parameters estimated using 10 Sobol's samples is exactly the same as the sorted order of global sensitivity indices for 1024 Sobol's samples.

We also evaluated the performance of SRC coefficients in predicting the ranking of the most important input parameters for the secondary trapped CO_2 regression model using the training sets ranking from 8 to 1024 quasi-Monte Carlo samples. First, we analyzed how the absolute values of standardized regression coefficients evolve for various Quasi-Monte Carlo sample sizes. The SRC results show a consistent pattern for the analyzed Sobol sequences within the analyzed range: When the absolute values of SRCs coefficients are sorted in descending order, their values correlate with global sensitivity indices, see Fig. 14.

This straightforward convergence of SRCs rankings occurs because the secondary trapped CO_2 model can be accurately approximated by a multiple linear model for the given input parameters (Iooss, B., Lemaître, P., 2015), as the R-squared of the linear model is very close to 1. Although the absolute values of SRCs corresponding to the most important parameters tend to increase, the ranking of the coefficients remains approximately the same, see Fig. 15: The parameter x_3:peshale is the most important in all cases, followed by x_5:posand, x_4:pefeeder, x_2:pesand, x_1:tegrad and x_6:totopsur.

The advantage of multiple linear regression is that the method is simple and fast. These results confirm that SRC is a useful and fast alternative for ranking input variables in terms of their sensitivity for the secondary trapped CO_2, even with very small quasi-Monte Carlo sizes.

Table 7
Summarized results of the standardized linear regression for 10 Sobol's quasi-random points sorted by absolute values of SRC coefficients (the column "coef").

i	names	coef	se	T	pval	CI[2.5 %]	CI[97.5 %]
3	peshale	6.48E-01	0.04011	1.62E+01	0.000515	0.520635	0.775929
5	posand	5.74E-01	0.04011	1.43E+01	0.000738	0.446614	0.701908
4	pefeeder	−2.73E-01	0.04011	−6.80E+00	0.006492	−0.40055	−0.14526
2	pesand	1.81E-01	0.034167	5.30E+00	0.013118	0.07231	0.289778
1	tegrad	1.43E-01	0.034167	4.18E+00	0.024993	0.03398	0.251448
6	totopsur	3.85E-02	0.04011	9.60E-01	0.407933	−0.08915	0.166147
0	Intercept	−1.76E-16	0.030798	−5.70E-15	1	−0.09801	0.098011

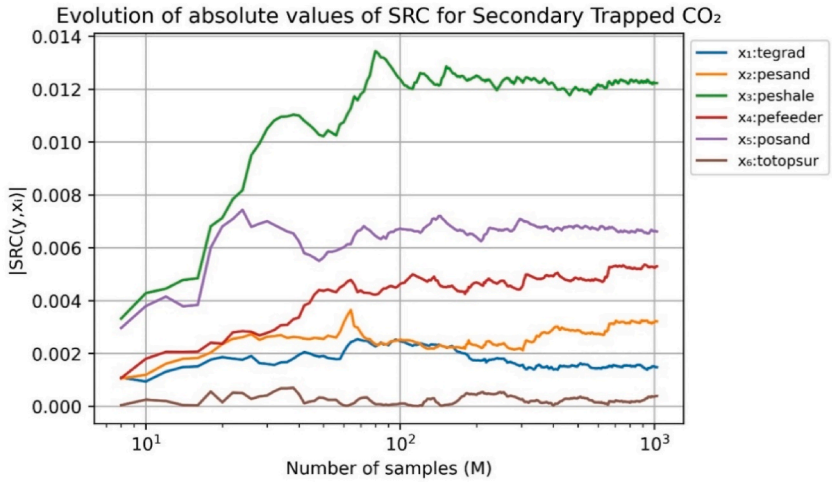


Fig. 14. Evolution of absolute values of standardized regression coefficients for secondary trapped CO_2 over all input parameters.

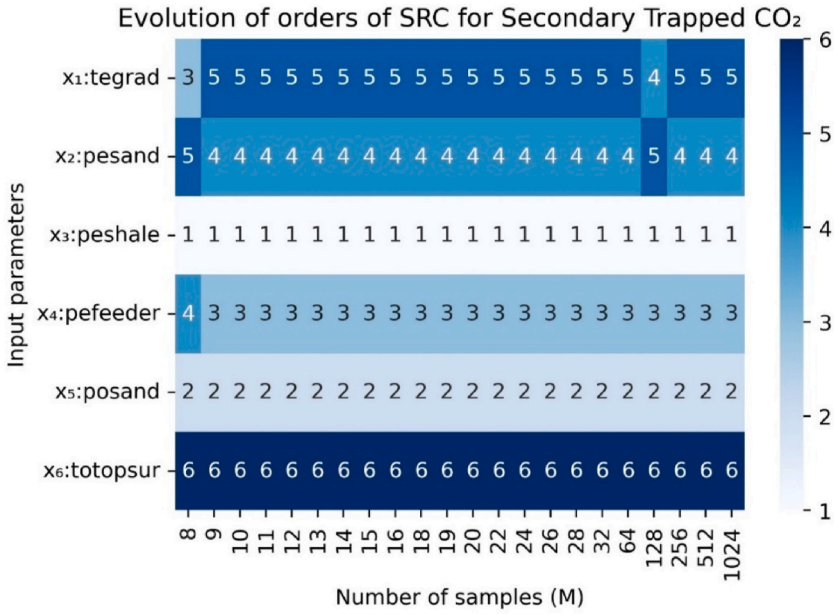


Fig. 15. Evolution of orders of absolute values of standardized regression coefficients for secondary trapped CO_2 over all input parameters (Order 1 corresponds to the top SRC).

5. Conclusions

In this study we considered the simulation of CO_2 sequestration and performed a sensitivity analysis of a range of input parameters defining an ensemble of models. The open-source reservoir simulator Open Porous Media Flow (OPM Flow) is used in this study for simulating the model ensembles.

In our study we defined six input parameters (Table 1), then used a quasi-Monte Carlo method to generate input samples and evaluated the resulting cases with OPM Flow. We condensed the full output to a single variable of interest and then performed a sensitivity analysis and created sets of increasingly complex symbolic regression models and multilinear models.

As results of the sensitivity analysis, the model of 'Secondary Trapped CO_2' is additive with no significant interactions among input parameters. Consequently, we also provided multiple linear regression models, that accurately approximate 'Secondary Trapped CO_2'. All discovered regression models (both symbolic and linear) can be easily implemented. Moreover, discovered regression models can be easily interpreted by human experts.

The results of the convergence studies for training sets ranging from 8 to 1024 quasi-Monte Carlo samples show that regression-

based sensitivity measures are a useful alternative for both ranking input parameters in terms of their importance (for the SRC method) and also for predicting the secondary trapped CO_2, even with very small quasi-Monte Carlo sizes (for example for $M = 10$), see Figs. 12–15. This is due to the demonstrated high degree of linearity of the secondary trapped CO_2 model with respect to the analyzed input parameters.

We see from the analysis and models that the permeability of the shale layers is the most significant for predicting the secondary trapped CO_2. Our hypothesis for the cause is as follows: when the shale is more permeable, the rising CO_2 plume will to a lesser extent be forced through the feeder chimneys and therefore will spread out more. At the upper parts of the sandstone layers this wider plume results in more flow into the shale, as well as exposing larger amounts of CO_2 to unsaturated brine, into which significant amounts dissolve over time. To fully confirm this hypothesis, further studies of layered storage reservoirs should be performed, as well as higher-resolution studies. These investigations will be a subject for future research.

The significance of each physical parameter changes over time, as shown in Fig. 10. We note that terminating the simulations earlier, closer to the end of injection time, might have yielded different conclusions about the relative significance of the various inputs, as sand porosity is more significant than shale permeability for roughly the first 120 years. This confirms in a novel way the prevailing view that simulation timescales of 200 years or more are necessary for carbon sequestration evaluation.

We also investigate the usage of a surrogate model for a case having a CO_2 injection in the start followed by a long period without a CO_2 injection. Consequently, this use case is well applicable for future CO_2 reservoir simulations. Furthermore, using Quasi Monte Carlo and machine learning in this context is novel. It is not obvious that Quasi Monte Carlo should work for this setting. Finally, our new added convergence study validates that Quasi Monte Carlo works and has a better convergence ratio than the plain Monte Carlo. The petascale cluster Karolina of IT4Innovations was used for analyses.

To our best knowledge, this paper also differs from previous references: It is the first application of time dependent sensitivity analysis and machine learning in the context of CO_2 simulations using Sobol (Quasi Monte Carlo) sampling. Moreover, the discovered regression models provide accurate linear and non-linear explanations of important quantities of interest of subsurface CO_2 storage, which can be easily interpretated by human experts. The advantage for users is the time-dependent analysis of the importance of the given input parameters of the model. Moreover, the effectiveness of the regression models shows that symbolic regression can be used for surrogate modelling in this case.

The Bayesian techniques for performing model calibration introduced in Hervas-Raluy et al. [80] will be a subject of our future research. In future, we would like also to test Gaussian process regression machine learning (GPR), as the recent studies reported in Chinta et al. [81] show the superiority of GPR for earth system models compared to other surrogate modeling approaches.

CRediT authorship contribution statement

Pavel Praks: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Atgeirr Rasmussen:** Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kjetil Olsen Lye:** Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jan Martinovič:** Supervision, Resources, Project administration, Funding acquisition. **Renata Praksová:** Visualization, Software. **Francesca Watson:** Software, Methodology. **Dejan Brkić:** Writing – review & editing, Writing – original draft, Visualization, Investigation.

Data availability

The data are given in Electronic Appendix of this paper and through the following links.

- The Open Porous Media (OPM) initiative, see https://github.com/OPM
- opm-runner: A small, a handy tool to run OPM Flow with various parameters, see https://github.com/kjetilly/opm-runner
- Convergence study, time evolution of the sensitivities, calculation of secondary trapped CO2 quantities, see https://github.com/kjetilly/ensemble_run_across

Declaration of competing interest

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

Acknowledgments

This work was supported by the ACROSS project. This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No. 955648. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Italy, France, the Czech Republic, the United Kingdom, Greece, the Netherlands, Germany and Norway. This project has received funding from the Ministry of Education, Youth and Sports of the Czech Republic (ID: MC2104) and e-INFRA CZ (ID:90254). This article has been produced with the financial support of the European Union under the REFRESH - Research Excellence For REgion Sustainability and High-tech Industries project number CZ.10.03.01/00/22_003/0000048 via the Operational Programme Just Transition. In addition, Dejan Brkić received support from the Ministry of Science, Technological

Development and Innovation of the Republic of Serbia through institutional financing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e40044.

References

[1] A.F. Rasmussen, T.H. Sandve, K. Bao, A. Lauser, J. Hove, B. Skaflestad, R. Klöfkorn, M. Blatt, A.B. Rustad, O. Sævareid, K.-A. Lie, A. Thune, The open porous media flow reservoir simulator, Comput. Math. Appl. 81 (2021) 159–185, https://doi.org/10.1016/j.camwa.2020.05.014.

[2] E.R. Okoroafor, L. Sampaio, F. Gasanzade, Y.P. Claro, J.D. Zhou, S.D. Saltzer, S. Bauer, A.R. Kovscek, Intercomparison of numerical simulation models for hydrogen storage in porous media using different codes, Energy Convers. Manag. 292 (2023) 117409, https://doi.org/10.1016/j.enconman.2023.117409.

[3] G. Roe, In defense of Milankovitch, Geophys. Res. Lett. 33 (24) (2006), https://doi.org/10.1029/2006GL027817.

[4] G.E. Marsh, Interglacials, Milankovitch cycles, solar activity, and carbon dioxide, J. Climatol. (2014) 345482, https://doi.org/10.1155/2014/345482, 2014.

[5] W.J. Davis, The relationship between atmospheric carbon dioxide concentration and global temperature for the last 425 million years, Climate 5 (4) (2017) 76, https://doi.org/10.3390/cli5040076.

[6] T.M. Gür, Carbon dioxide emissions, capture, storage and utilization: review of materials, processes and technologies, Prog. Energy Combust. Sci. 89 (2022) 100965, https://doi.org/10.1016/j.pecs.2021.100965.

[7] B.M. Smirnov, Physics of the earth's glacial cycle, Foundations 2 (4) (2022) 1114–1128, https://doi.org/10.3390/foundations2040073.

[8] S.P. Rigby, A. Alsayah, Storage sites for carbon dioxide in the north sea and their particular characteristics, Energies 17 (1) (2023) 211, https://doi.org/10.3390/en17010211.

[9] X. Sun, A. Shang, P. Wu, T. Liu, Y. Li, A review of CO_2 marine geological sequestration, Processes 11 (7) (2023) 2206, https://doi.org/10.3390/pr11072206.

[10] K.S. Lackner, A guide to CO_2 sequestration, Science 300 (5626) (2003) 1677–1678, https://doi.org/10.1126/science.1079033.

[11] R. Lal, Carbon sequestration, Phil. Trans. Biol. Sci. 363 (1492) (2008) 815–830, https://doi.org/10.1098/rstb.2007.2185.

[12] N. Kumar, A. Verma, T. Ahmad, R.K. Sahu, A. Mandal, M. Mubashir, M. Ali, N. Pal, Carbon capture and sequestration technology for environmental remediation: a CO_2 utilization approach through EOR, Geoenergy Science and Engineering 212619 (2023), https://doi.org/10.1016/j.geoen.2023.212619.

[13] C. Park, J. Oh, S. Jo, I. Jang, K.S. Lee, Multi-objective optimization of CO_2 sequestration in heterogeneous saline aquifers under geological uncertainty, Appl. Sci. 11 (20) (2021) 9759, https://doi.org/10.3390/app11209759.

[14] L. Tang, J. Fan, Z. Li, J. Chen, W. Liu, A new constitutive model for salt rock under cyclic loadings based on state variables, Geoenergy Science and Engineering 233 (2024) 212433, https://doi.org/10.1016/j.geoen.2023.212433.

[15] Z. Xie, C. Cao, Y. Zhao, L. Zhang, R. Zhang, J. Li, D. Zhang, A new pressure management framework for CO_2 sequestration in deep saline aquifers based on genetic algorithm, Geoenergy Science and Engineering 212668 (2024), https://doi.org/10.1016/j.geoen.2024.212668.

[16] K. Deng, Z. Ge, H. Zhang, S. Gong, H. Zhang, Isothermal adsorption characteristics of various phases of CO_2 and CH_4 in different rank coals, Energy Sources, Part A Recovery, Util. Environ. Eff. 45 (2) (2023) 6356–6370, https://doi.org/10.1080/15567036.2023.2211035.

[17] S. Han, S. Wang, C. Guo, S. Sang, A. Xu, W. Gao, P. Zhou, Distribution of the adsorbed density of supercritical CO_2 onto the anthracite and its implication for CO_2 geologic storage in deep coal, Geoenergy Science and Engineering 234 (2024) 212624, https://doi.org/10.1016/j.geoen.2023.212624.

[18] J. Wei, Y. Chen, W. Liang, Experimental and theoretical investigation on permeability evolution of fractures in anthracite with supercritical CO_2 immersion, Int. J. Rock Mech. Min. Sci. 174 (2024) 105651, https://doi.org/10.1016/j.ijrmms.2024.105651.

[19] Y. Li, H. Tian, Development and application of a simulator for simulating the behaviors of a geological system when replacing CH_4 from hydrate-bearing reservoirs by CO_2, Energies 16 (8) (2023) 3342, https://doi.org/10.3390/en16083342.

[20] B. Wei, B. Wang, X. Li, M. Aishan, Y. Ju, CO_2 storage in depleted oil and gas reservoirs: a review, Advances in Geo-Energy Research 9 (2) (2023) 76–93, https://doi.org/10.46690/ager.2023.08.02.

[21] R.G. Heidarabad, K. Shin, Carbon capture and storage in depleted oil and gas reservoirs: the viewpoint of wellbore injectivity, Energies 17 (5) (2024) 1201, https://doi.org/10.3390/en17051201.

[22] Y. Yang, G. Liao, Y. Wu, X. Ma, Z. Dong, W. Li, J. Wei, Displacement efficiency and storage characteristics of CO_2 in low permeability reservoirs: an experimental work, Energy Explor. Exploit. 41 (2) (2023) 601–618, https://doi.org/10.1177/01445987221142664.

[23] A. Gupta, A.R. Paul, Influence of relative permeability, capillary pressure, and well orientation in the geological carbon sequestration in a saline aquifer, Int. J. Oil Gas Coal Technol. 31 (3) (2022) 263–279, https://doi.org/10.1504/IJOGCT.2022.126360.

[24] J.J. Hutchinson, C.A. Campbell, R.L. Desjardins, Some perspectives on carbon sequestration in agriculture, Agric. For. Meteorol. 142 (2–4) (2007) 288–302, https://doi.org/10.1016/j.agrformet.2006.03.030.

[25] D. Guan, J. Nie, L. Zhou, Q. Chang, J. Cao, How to simulate carbon sequestration potential of forest vegetation? A forest carbon sequestration model across a typical mountain city in China, Rem. Sens. 15 (21) (2023) 5096, https://doi.org/10.3390/rs15215096.

[26] P. Verma, P.K. Ghosh, The economics of forest carbon sequestration: a bibliometric analysis, Environ. Dev. Sustain. (2023) 1–31, https://doi.org/10.1007/s10668-023-02922-w.

[27] J. Zheng, Z.R. Chong, M.F. Qureshi, P. Linga, Carbon dioxide sequestration via gas hydrates: a potential pathway toward decarbonization, Energy & Fuels 34 (9) (2020) 10529–10546, https://doi.org/10.1021/acs.energyfuels.0c02309.

[28] X. Cao, H. Wang, K. Yang, S. Wu, Q. Chen, J. Bian, Hydrate-based CO_2 sequestration technology: feasibilities, mechanisms, influencing factors, and applications, J. Petrol. Sci. Eng. 219 (2022) 111121, https://doi.org/10.1016/j.petrol.2022.111121.

[29] G.H. Rau, K. Caldeira, Enhanced carbonate dissolution: a means of sequestering waste CO_2 as ocean bicarbonate, Energy Convers. Manag. 40 (17) (1999) 1803–1813, https://doi.org/10.1016/S0196-8904(99)00071-0.

[30] E. Ellingboe, J.H. Arehart, W.V. Srubar III, On the theoretical CO_2 sequestration potential of pervious concrete, Infrastructures 4 (1) (2019) 12, https://doi.org/10.3390/infrastructures4010012.

[31] M. Kazemian, B. Shafei, Carbon sequestration and storage in concrete: a state-of-the-art review of compositions, methods, and developments, J. CO2 Util. 70 (2023) 102443, https://doi.org/10.1016/j.jcou.2023.102443.

[32] M. Sajdak, M. Zajemska, M. Ouadi, W. Mucha, E. Misztal, C. Pieszko, G. Gałko, Perspectives of using sewage sludge char in CO_2 sequestration on degraded and brownfield sites, Energies 16 (9) (2023) 3945, https://doi.org/10.3390/en16093945.

[33] L. Jiang, L. Cheng, Y. Zhang, G. Liu, J. Sun, A review on CO_2 sequestration via mineralization of coal fly ash, Energies 16 (17) (2023) 6241, https://doi.org/10.3390/en16176241.

[34] W. Huang, R. Xu, F. Zhang, Y. Zou, P. Jiang, CO_2 capture analysis in different combustion methods for CO_2 utilisation and storage, Int. J. Oil Gas Coal Technol. 29 (3) (2022) 285–305, https://doi.org/10.1504/IJOGCT.2022.121048.

[35] Y. Liu, T. Hu, Z. Rui, Z. Zhang, K. Du, T. Yang, B. Dindoruk, E.H. Stenby, F. Torabi, A. Afanasyev, An integrated framework for geothermal energy storage with CO_2 sequestration and utilization, Engineering 30 (2023) 121–130, https://doi.org/10.1016/j.eng.2022.12.010.

[36] Z. Dai, R. Middleton, H. Viswanathan, J. Fessenden-Rahn, J. Bauman, R. Pawar, S.Y. Lee, B. McPherson, An integrated framework for optimizing CO_2 sequestration and enhanced oil recovery, Environ. Sci. Technol. Lett. 1 (1) (2014) 49–54, https://doi.org/10.1021/ez4001033.

[37] C. Sambo, N. Liu, R. Shaibu, A.A. Ahmed, R.G. Hashish, A technical review of CO_2 for enhanced oil recovery in unconventional oil reservoirs, Geoenergy Science and Engineering 221 (2023) 111185, https://doi.org/10.1016/j.petrol.2022.111185.

[38] F.E. Watson, S.A. Mathias, S.E. Daniels, R.R. Jones, R.J. Davies, B.J. Hedley, J. van Hunen, Dynamic modelling of a UK North Sea saline formation for CO_2 sequestration, Petrol. Geosci. 20 (2) (2014) 169–185, https://doi.org/10.1144/petgeo2012-072.

[39] M.A. Celia, S. Bachu, J.M. Nordbotten, K.W. Bandilla, Status of CO_2 storage in deep saline aquifers with emphasis on modeling approaches and practical simulations, Water Resour. Res. 51 (9) (2015) 6846–6892, https://doi.org/10.1002/2015WR017609.

[40] T.N.V. Pavan, S.K. Govindarajan, Numerical investigations on performance of sc-CO_2 sequestration associated with the evolution of porosity and permeability in low permeable saline aquifers, Geoenergy Science and Engineering 225 (2023) 211681, https://doi.org/10.1016/j.geoen.2023.211681.

[41] B. Chen, D.R. Harp, Y. Lin, E.H. Keating, R.J. Pawar, Geologic CO_2 sequestration monitoring design: a machine learning and uncertainty quantification based approach, Appl. Energy 225 (2018) 332–345, https://doi.org/10.1016/j.apenergy.2018.05.044.

[42] D. Rathmaier, F. Naim, A.C. William, D. Chakraborty, C. Conwell, M. Imhof, G.M. Holmes, L.E. Zerpa, A reservoir modeling study for the evaluation of CO_2 storage upscaling at the decatur site in the Eastern Illinois Basin, Energies 17 (5) (2024) 1212, https://doi.org/10.3390/en17051212.

[43] I.M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simulat. 55 (1–3) (2001) 271–280, https://doi.org/10.1016/S0378-4754(00)00270-6.

[44] I.M. Sobol', D. Asotsky, A. Kreinin, S. Kucherenko, Construction and comparison of high-dimensional Sobol generators, Wilmott 2011 (56) (2011) 64–79, https://doi.org/10.1002/wilm.10056.

[45] I. Azzini, T. Mara, R. Rosati, Monte Carlo estimators of first-and total-orders Sobol' indices, arXiv preprint arXiv:2006.08232 (2020), https://doi.org/10.48550/arXiv.2006.08232.

[46] P. Praks, D. Brkić, Approximate flow friction factor: estimation of the accuracy using Sobol's quasi-random sampling, Axioms 11 (2) (2022) 36, https://doi.org/10.3390/axioms11020036.

[47] CO2DataShare, Sleipner 2019 Benchmark Model (2020), https://doi.org/10.11582/2020.00004.

[48] W. Becker, Uncertainty Propagation through Large Nonlinear Models, University of Sheffield, 2011. https://etheses.whiterose.ac.uk/15000/, 20-May-24.

[49] V. Kopustinskas, P. Praks, T. Mara, R. Rossati, Application of PCE sensitivity analysis method to gas transmission network, in: Stein Haugen, Anne Barros, Coen Gulijk (Eds.), Safety and Reliability–Safe Societies in a Changing World, CRC Press. Proceedings of ESREL 2018, June 17-21, 2018, Trondheim, Norway, 2018, pp. 2693–2700, https://doi.org/10.1201/9781351174664-339. Trond Kongsvik, and Jan Erik Vinnem.

[50] E. Pisoni, D. Albrecht, T.A. Mara, R. Rosati, S. Tarantola, P. Thunis, Application of uncertainty and sensitivity analysis to the air quality SHERPA modelling tool, Atmos. Environ. 183 (2018) 84–93, https://doi.org/10.1016/j.atmosenv.2018.04.006.

[51] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, Q. Wu, Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices, Environ. Model. Software 114 (2019) 29–39, https://doi.org/10.1016/j.envsoft.2019.01.012.

[52] C. Geffray, A. Gerschenfeld, P. Kudinov, I. Mickus, M. Jeltsov, K. Kööp, D. Grishchenko, D. Pointer, Verification and validation and uncertainty quantification, in: Thermal Hydraulics Aspects of Liquid Metal Cooled Nuclear Reactors, Woodhead Publishing, 2019, pp. 383–405, https://doi.org/10.1016/B978-0-08-101980-1.00008-9.

[53] E.V. Filippov, L.A. Zaharov, D.A. Martyushev, I.N. Ponomareva, Reproduction of reservoir pressure by machine learning methods and study of its influence on the cracks formation process in hydraulic fracturing, Записки Горного института/Journal of Mining Institute 258 (2022) 924–932, https://doi.org/10.31897/PMI.2022.103.

[54] M. Schmidt, H. Lipson, Distilling free-form natural laws from experimental data, Science 324 (5923) (2009) 81–85, https://doi.org/10.1126/science.1165893.

[55] B.M. de Silva, D.M. Higdon, S.L. Brunton, J.N. Kutz, Discovery of physics from data: universal laws and discrepancies, Frontiers in Artificial Intelligence 3 (2020) 25, https://doi.org/10.3389/frai.2020.00025.

[56] D. Brkić, P. Praks, Accurate and efficient explicit approximations of the Colebrook flow friction equation based on the Wright ω-function, Mathematics 7 (1) (2019) 34, https://doi.org/10.3390/math7010034.

[57] P. Praks, D. Brkić, Review of new flow friction equations: constructing Colebrook's explicit correlations accurately, Rev. Int. Métodos Numéricos Cálculo Diseño Ing. 36 (3) (2020) 41, https://doi.org/10.23967/j.rimni.2020.09.001.

[58] L. Ma, Q. Guo, X. Li, S. Xu, J. Zhou, M. Ye, Z. Liu, Drag correlations for flow past monodisperse arrays of spheres and porous spheres based on symbolic regression: effects of permeability, Chem. Eng. J. 445 (2022) 136653, https://doi.org/10.1016/j.cej.2022.136653.

[59] R. Haghpanah, D. Shade, Fitting adsorption isotherms with symbolic regression, Ind. Eng. Chem. Res. 62 (51) (2023) 22141–22148, https://doi.org/10.1021/acs.iecr.3c02900.

[60] S. Kalam, M. Arif, A. Raza, N. Lashari, M. Mahmoud, Data-driven modeling to predict adsorption of hydrogen on shale kerogen: implication for underground hydrogen storage, Int. J. Coal Geol. 280 (2023) 104386, https://doi.org/10.1016/j.coal.2023.104386.

[61] I.N. Ponomareva, D.A. Martyushev, S.K. Govindarajan, A new approach to predict the formation pressure using multiple regression analysis: case study from Sukharev oil field reservoir–Russia, Journal of King Saud University-Engineering Sciences (2022), https://doi.org/10.1016/j.jksues.2022.03.005.

[62] P.A. Witte, T. Konuk, E. Skjetne, R. Chandra, Fast CO_2 saturation simulations on large-scale geomodels with artificial intelligence-based Wavelet Neural Operators, Int. J. Greenh. Gas Control 126 (2023) 103880, https://doi.org/10.1016/j.ijggc.2023.103880.

[63] M.G. Rezk, A.F. Ibrahim, A.R. Adebayo, Influence of impurities on reactive transport of CO_2 during geo-sequestration in saline aquifers, Fuel 344 (2023) 127994, https://doi.org/10.1016/j.fuel.2023.127994.

[64] R. Feng, Physics-informed deep learning for rock physical inversion and its uncertainty analysis, Geoenergy Science and Engineering 230 (2023) 212229, https://doi.org/10.1016/j.geoen.2023.212229.

[65] Y. Teng, P. Wang, H. Xie, J. Zhu, Capillary trapping characteristics of CO_2 sequestration in fractured carbonate rock and sandstone using MRI, J. Nat. Gas Sci. Eng. 108 (2022) 104809, https://doi.org/10.1016/j.jngse.2022.104809.

[66] J. Liu, L. Tu, G. Liu, C. Jiang, Z. Zhang, An analytical structural global sensitivity analysis method based on direct integral, Inverse Problems in Science and Engineering 27 (11) (2018) 1559–1576, https://doi.org/10.1080/17415977.2018.1531856. Informa UK Limited.

[67] S. Tarantola, F. Ferretti, S. Lo Piano, M. Kozlova, A. Lachi, R. Rosati, A. Puy, P. Roy, G. Vannucci, M. Kuc-Czarnecka, A. Saltelli, An annotated timeline of sensitivity analysis, in: Environmental Modelling & Software, vol. 174, Elsevier BV, 2024 105977, https://doi.org/10.1016/j.envsoft.2024.105977.

[68] J. Herman, W. Usher, SALib: an open-source Python library for sensitivity analysis, J. Open Source Softw. 2 (9) (2017) 97, https://doi.org/10.21105/joss.00097.

[69] F. Pianosi, T. Wagener, Distribution-based sensitivity analysis from a generic input-output sample, Environ. Model. Software 108 (2018) 197–207, https://doi.org/10.1016/j.envsoft.2018.07.019.

[70] J.P. Kleijnen, An overview of the design and analysis of simulation experiments for sensitivity analysis, Eur. J. Oper. Res. 164 (2) (2005) 287–300, https://doi.org/10.1016/j.ejor.2004.02.005.

[71] T. Pujol, M. Duran-Ros, J.D. Betancur, G. Arbat, S. Cuffí, J. Pujol, F.R. de Cartagena, J. Puig-Bargués, Sensitivity analysis of a particle retention model and application to a pressurised sand bed filter for drip irrigation, Biosyst. Eng. 230 (2023) 51–70, https://doi.org/10.1016/j.biosystemseng.2023.04.006.

[72] S. Ouanes, L. Sriti, Regression-based sensitivity analysis and multi-objective optimisation of energy performance and thermal comfort: building envelope design in hot arid urban context, Build. Environ. 248 (2024) 111099, https://doi.org/10.1016/j.buildenv.2023.111099.

[73] L. Stajić, R. Praksová, D. Brkić, P. Praks, Estimation of global natural gas spot prices using big data and symbolic regression, Resour. Pol. 95 (2024) 105144, https://doi.org/10.1016/j.resourpol.2024.105144.

[74] Gunther Leobacher, Pillichshammer Friedrich, Introduction to Quasi Montecarlo Methods, vol. 2012, 2013, https://doi.org/10.1007/978-3-319-03425-6.

[75] F.M. Gomes, F.M. Pereira, A.F. Silva, M.B. Silva, Multiple response optimization: analysis of genetic programming for symbolic regression and assessment of desirability functions, Knowl. Base Syst. 179 (2019) 21–33, https://doi.org/10.1016/j.knosys.2019.05.002.

[76] G.H. Yamashita, F.S. Fogliatto, M.J. Anzanello, G.L. Tortorella, Customized prediction of attendance to soccer matches based on symbolic regression and genetic programming, Expert Syst. Appl. 187 (2022) 115912, https://doi.org/10.1016/j.eswa.2021.115912.

[77] R. Dubčáková, Eureqa: software review, Genet. Program. Evolvable Mach. 12 (2011) 173–178, https://doi.org/10.1007/s10710-010-9124-z.
[78] G. Manache, C.S. Melching, Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters, Environ. Model. Software 23 (5) (2008) 549–562, https://doi.org/10.1016/j.envsoft.2007.08.001. Elsevier BV.
[79] B. Iooss, P. Lemaître, A review on global sensitivity analysis methods, in: G. Dellino, C. Meloni (Eds.), Uncertainty Management in Simulation-Optimization of Complex Systems, Operations Research/Computer Science Interfaces Series, vol. 59, Springer, Boston, MA, 2015, https://doi.org/10.1007/978-1-4899-7547-8_5.
[80] S. Hervas-Raluy, B. Wirthl, P.E. Guerrero, G. Robalo Rei, J. Nitzler, E. Coronado, J. Font de Mora Sainz, B.A. Schrefler, M.J. Gomez-Benito, J.M. Garcia-Aznar, W.A. Wall, Tumour growth: an approach to calibrate parameters of a multiphase porous media model based on in vitro observations of Neuroblastoma spheroid growth in a hydrogel microenvironment, Comput. Biol. Med. 159 (2023) 106895, https://doi.org/10.1016/j.compbiomed.2023.106895. Elsevier BV.
[81] S. Chinta, X. Gao, Q. Zhu, Machine learning driven sensitivity analysis of E3SM land model parameters for wetland methane emissions. In journal of advances in modeling earth systems, American Geophysical Union (AGU) 16 (Issue 7) (2024), https://doi.org/10.1029/2023ms004115.