

Full Length Article

Surrogate-Based probabilistic assessment of hydrogen leakage through abandoned wells in depleted gas reservoirs for underground hydrogen storage[☆]

Qiu hao Chang^{a, *}, David Dempsey^{a, *}, Guilong Xu^{a, b, *}, Liehui Zhang^c, Yulong Zhao^c

^a Civil and Environmental Engineering Department, University of Canterbury, Christchurch 8041, New Zealand

^b The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China

^c National Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, China

ABSTRACT

Hydrogen leakage through abandoned wells represents a critical risk for underground hydrogen storage (UHS) in depleted gas reservoirs, potentially leading to safety hazards and economic losses. Although reservoir-scale numerical simulations can capture leakage processes in detail, their computational cost limits their use in rapid scenario screening and early-stage design assessment. In this work, a probabilistic surrogate framework is developed to evaluate leakage likelihood under varying geological and operational conditions. The framework is intended as an early-stage screening tool to identify configurations that may pose a leakage risk and therefore warrant further detailed investigation. The surrogate is constructed using multiphase reservoir simulation results and incorporates six continuous variables together with ten families of H₂-water relative permeability curves spanning a range of low to high mobility that collectively represent structural uncertainty in gas–water flow. Validation against an independent simulation dataset shows that the surrogate achieves an overall prediction accuracy of 0.967 with an area under the ROC curve (AUC) of 0.995. Results indicate that, within the investigated ranges and under homogeneous reservoir conditions, well spacing and injection rate are the principal controls on leakage probability. Formation permeability does not show a stable directional influence under relatively high-permeability conditions (100–1000 mD), while reservoir depth and porosity are negatively associated with leakage probability. In addition, variability in relative permeability significantly affects predicted risk, with high-mobility curve families increasing leakage likelihood relative to intermediate conditions. This framework provides an efficient basis for leakage-aware screening and supports risk-informed evaluation during the early design stage of UHS deployment.

1. Introduction

Hydrogen is increasingly recognised as an important energy carrier for decarbonisation, including the replacement of fossil-derived hydrogen in chemical industries, energy system balancing, and strategic reserve applications [1]. With growing targets for hydrogen production and use, the demand for hydrogen is expected to rise in the coming decades [2]. In this context, the development of large-scale hydrogen storage technologies supports the scaling of a hydrogen economy [3,4]. Hydrogen can be stored in several forms, including compressed gas, liquid hydrogen storage, as well as solid-state storage technologies such as metal hydrides, where recent studies have explored advanced thermal management and reaction-driven heating strategies to improve performance [5,6]. Among various options, underground hydrogen storage (UHS) has attracted increasing attention due to its ability to accommodate high volumes at relatively low levelized cost while leveraging existing subsurface infrastructure [7,8]. By storing

hydrogen in geological formations such as salt caverns, aquifers, or depleted hydrocarbon reservoirs, UHS offers a technically feasible and secure pathway for balancing production and consumption at the system level [8,9]. Compared to salt caverns and aquifers, depleted hydrocarbon reservoirs offer significant advantages such as well-characterized geology, existing infrastructure, and reliable sealing performance, making them a potential option for large-scale hydrogen storage [10,11].

Despite these advantages, implementing UHS in depleted hydrocarbon reservoirs raises several concerns regarding potential hydrogen leakage. Possible migration pathways include diffusion or advection through caprock formations, lateral or vertical flow through faults and fractures, and leakage through existing wells, particularly abandoned or improperly sealed ones. To date, several studies have primarily focused on H₂ migration through caprock at the molecular scale. For example, using molecular simulation techniques such as Molecular Dynamics and Monte Carlo methods, Xie et al. [12], Zhang et al. [13], Liu et al. [14],

[☆] This article is part of a special issue entitled: 'Subsurface Gas Storage' published in Fuel.

* Corresponding authors.

E-mail addresses: david.dempsey@canterbury.ac.nz (D. Dempsey), xuguilong@tongji.edu.cn (G. Xu).

Chang et al. [15], and Kahzadvand et al. [16] have investigated how factors such as pressure, temperature, gas composition, and pore aperture influence H_2 diffusion and adsorption in nanopores, exploring the fundamental mechanisms of confined H_2 transport within caprock. In addition, Li et al. [17] observed that residual pore water can seal the nano cracks in the caprock reducing H_2 leakage. Compared to methane (CH_4), carbon dioxide (CO_2) as a cushion gas more effectively promotes water-assisted sealing of H_2 leakage through cracks. Talukdar et al. [18] demonstrated that dynamic pressure from hydrogen injection can trigger caprock fracturing and fault reactivation, with coupled leakage pathways influencing gas migration.

Due to hydrogen's small molecular size and high diffusivity, even minor defects in wellbore cement or casing could lead to sustained leakage. This risk is particularly pronounced for abandoned wells, where incomplete plugging, legacy materials, or long-term degradation may significantly weaken sealing performance. Consequently, leakage associated with abandoned wells is a credible failure mode in UHS systems deserving close scrutiny. Similar leakage pathways have been observed in other subsurface storage contexts. For example, at the Decatur CO_2 storage project in the United States, gas migration occurred through a monitoring well and was subsequently detected [19]. In another case at Lake Boehmer, corrosion in an old oil and gas well led to uncontrolled upward migration of pressurised groundwater [20]. Several studies have been conducted that systematically investigate leakage through abandoned wells in the context of CO_2 subsurface storage. Kopp et al. [21] developed a probabilistic framework for assessing CO_2 leakage risk through abandoned wells and showed that leakage is primarily controlled by buoyancy driven migration. Qiao et al. [22] presented a semi-analytical pressure-transient model to evaluate CO_2 leakage through abandoned wells, indicating that well spacing controls leakage primarily by delaying pressure propagation and the onset of leakage. Postma et al. [23] demonstrated that CO_2 leakage through abandoned wells is primarily dominated by the spatial extent of the CO_2 plume and the density of nearby wells. These studies provide a reference for understanding the operational and geological controls on gas leakage through abandoned wells in subsurface gas storage.

Drawing on insights from studies of CO_2 leakage through abandoned wells in CO_2 subsurface storage, recent studies have investigated H_2 leakage in UHS with attention to migration through wells. For example, Shoushtari et al. [24] developed a semi-analytical model to detect H_2 leakage in multi-layered reservoirs, where gas migrates from a target to a non-target layer through wellbore crossflow, and demonstrated that leakage rates can reach up to 3.5 million standard cubic feet (MMSCF)/day, primarily governed by interlayer pressure differences and flow capacities. Guo et al. [25] conducted numerical reservoir simulations of H_2 leakage through an abandoned well in a depleted gas reservoir, incorporating geological and operational variables such as wellbore integrity, reservoir permeability, and shut-in time, and demonstrated that wellbore integrity is the dominant control on leakage, with worst-case scenarios yielding hydrogen losses up to 1.06×10^5 kg after 10 years of operation. Current studies primarily rely on reservoir-scale numerical simulations to assess H_2 leakage through abandoned wells. While these simulations capture complex subsurface flow and geological features, they are computationally expensive and unsuitable for rapid scenario screening or early-stage planning. This limitation motivates the development of computationally efficient surrogate models for probabilistic leakage assessment.

Surrogate modelling has been explored in subsurface gas storage, particularly in CCUS applications, to enable rapid evaluation of leakage risk. For example, He et al. [26] combined deep learning-based surrogate models with Bayesian inversion to identify potential CO_2 leakage locations from pressure response data, achieving good accuracy while substantially reducing computational cost. Baek et al. [27] developed a deep learning-based surrogate model to predict time dependent CO_2 leakage rates through abandoned wells for well scale risk assessment, covering a range of reservoir and operational parameters as inputs

including reservoir permeability, porosity, well length, and injection rate. Allan et al. [28] proposed a statistical surrogate model based on survival analysis to predict the time to CO_2 leakage under uncertainty, showing that probabilistic frameworks can effectively capture long-term leakage risks in CO_2 sequestration systems. Chen et al. [29] built reduced-order surrogate models to predict CO_2 and oil component leakage through abandoned wells in CO_2 -EOR systems, with sensitivity analysis showing that leakage increases with reservoir permeability and oil composition, while decreases with net-to-gross ratio and residual oil saturation. Despite these advances, surrogate modelling studies for leakage risk assessment have largely focused on subsurface CO_2 systems, and comparable work addressing UHS remains limited. In addition, uncertainty in multiphase flow behaviour is typically simplified, with relative permeability represented by a single set of fixed parameters, despite the need for different curve shapes to describe varying reservoir conditions. This simplification may overlook an important source of uncertainty in gas migration and leakage prediction.

In this study, we develop a logistic surrogate model to predict the likelihood of H_2 leakage from abandoned wells in depleted gas reservoirs, paying specific attention to the issue of uncertainty in relative permeability behavior. This is addressed through a categorical representation of relative permeability using multiple physically consistent curve families, which allows variability in multiphase flow behaviour to be efficiently represented within the surrogate modelling framework. The model is trained on outputs from a physics-based reservoir simulator that cover not only typical geological and operational conditions, but also variations in relative permeability, and predicts whether injected H_2 will migrate through an abandoned well into overlying formations. The model is intended to support early-stage screening, where the objective is to identify UHS configurations that may be associated with a plausible leakage risk rather than to quantify leakage rate or volume. At the screening stage, reservoir-scale parameters are generally available, whereas detailed well integrity information required for quantitative leakage prediction might only be obtained once a leakage risk is identified. Hence, leakage rate and volume are comparatively difficult to assess outside of direct investigation. This is the first surrogate model developed to predict leakage risk from abandoned wells at the reservoir scale in the context of UHS, providing direct support for leakage-aware site selection and operational planning. The rest of the paper is organized as follows. The reservoir simulation framework is introduced in Section 2, followed by the development of the surrogate model in Section 3. The results and key insights are discussed in Section 4, and the conclusions are presented in Section 5.

2. Reservoir simulation framework

All simulations in this study were performed using the reservoir simulator Cirrus (formerly known as PFLOTTRAN-OGS), developed by OpenGoSim. Cirrus is designed to model multiphase, multicomponent flow and transport processes in porous media and is built upon the open-source PFLOTTRAN framework. Previous benchmarking studies have demonstrated close agreement between Cirrus and commercial simulators such as CMG in compositional simulations of gas–water flow in depleted reservoirs, confirming its robustness and reliability [30,31]. To date, Cirrus has been widely used in studies investigating UHS under diverse reservoir conditions [32–39]. Therefore, in this study, Cirrus serves as the reference model for generating the simulation dataset used for surrogate training and validation.

A conceptual model was developed to simulate hydrogen flow within a reservoir, as well as its leakage from an abandoned well into an overlying aquifer, as illustrated in Fig. 1. The model comprises a reservoir, an impermeable caprock, an overlying aquifer, one injection well (which also functions as a production well), and an abandoned well. The model domain is discretized into $100 \times 100 \times 100$ grid blocks, resulting in overall dimensions of $5000 \text{ m} \times 5000 \text{ m} \times 2000 \text{ m}$ along the x, y, and z directions, respectively. This discretization scheme is adopted based on

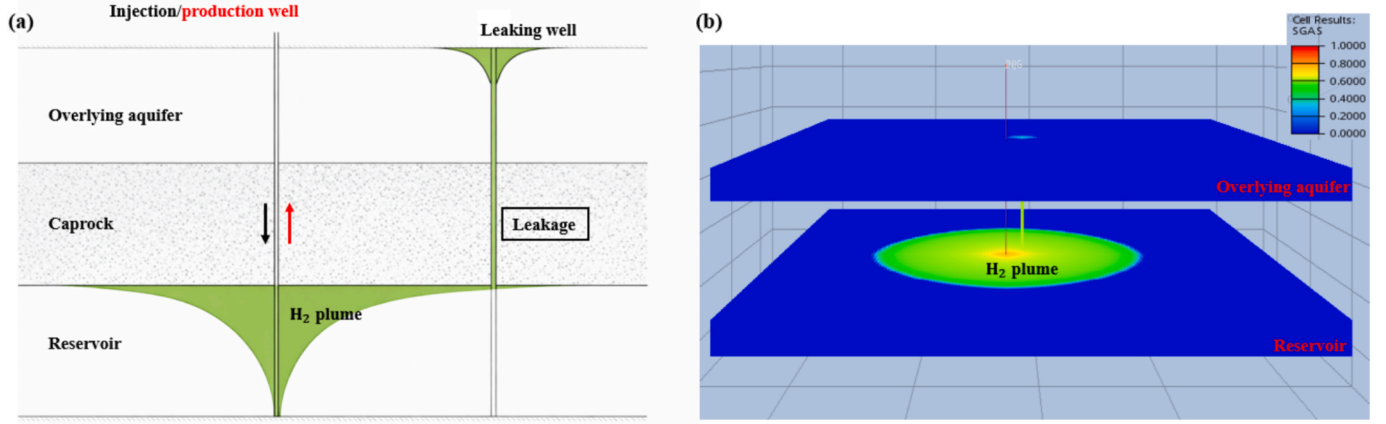


Fig. 1. (a) Schematic illustration of H_2 leakage from a depleted reservoir into an overlying aquifer through a leaking well. (b) Corresponding reservoir model constructed in this study, with the impermeable caprock hidden for clarity. Leakage occurrence is defined as the presence of H_2 in the overlying aquifer.

previous studies, ensuring simulation accuracy [25,29,40]. The overlying aquifer and the caprock each have a thickness of 500 m, while the reservoir has a thickness of 1000 m.

The numerical simulations were performed by solving the governing equations for mass conservation, as shown below [40]:

$$\frac{\partial}{\partial t} \left(\phi \sum_{j=1}^{n_p} x_{ij} S_j \eta_j \right) - \nabla \cdot \sum_{j=1}^{n_p} (x_{ij} u_j \eta_j + \phi S_j D_j \nabla x_{ij}) + Q_i = 0 \quad (2.1)$$

Where ϕ represents the porosity, n_p is the number of phases, j denotes the phase index, and n_c is the number of components. The quantity x_{ij} represents the mole fraction of component i in phase j , and S is the saturation of phase j . The term η_j is the molar density of phase j , while u_j denotes the flow velocity of phase j . The parameter D_j is the diffusion coefficient of phase j , and Q_i represents the source/sink term of component i . The velocity of phase j in the reservoir is described by Darcy's law:

$$u_j = -\frac{k k_{rj}}{\mu_j} (\nabla P_j - \rho_j g \hat{z}) \quad (2.2).$$

Where k is the absolute permeability, k_{rj} is the relative permeability of phase j , μ_j represents the viscosity of phase j , ρ_j is the mass density of phase j , g is the gravitational acceleration, and $\hat{z}$ is a unit vector in the direction of gravity.

Phase densities and viscosities were computed using the GERG 2008 equation of state. In line with the methodologies reported by Smith et al. [41] and Mahdi et al. [42], H_2 was adopted as the cushion gas to maintain reservoir pressure during the initial stage of storage. After this pressure establishment phase, the system was operated under cyclic H_2 injection and production. The analysis therefore centers on the behavior of H_2 -water two-phase flow in porous media. For model tractability, several simplifying assumptions were made. The reservoir was considered homogeneous, with properties that do not vary spatially or temporally. This is a deliberate design choice aligned with the role of the framework as an early-stage screening tool, rather than an attempt to reproduce site-specific geological complexity. Reservoir heterogeneity is highly site-specific and requires detailed characterisation, which is typically not available at the screening stage considered in this study. Incorporating such information would require site-specific calibration and would reduce the general applicability of the model. Such a modelling strategy is consistent with common practice in surrogate-based reservoir simulation studies, where simplified or homogeneous representations are often adopted to enable efficient exploration of parameter space, uncertainty analysis, and rapid decision support [43–47].

In addition, bio-chemical reactions and geo-mechanical processes were not included, as capturing their effects would expand the parameter space in a way that makes the surrogate problem intractable.

Instead, these issues should be addressed in follow-up simulations if a project moves to developed design. The system was treated strictly as a gas-water two-phase system, with no oil phase involved. As H_2 dissolution in water is very small [48], previous studies indicate that the associated hydrogen loss due to dissolution and diffusion is typically on the order of 0.1 ~ 1% over the lifetime of UHS [49]. Even at the upper bound of this range, such a small reduction in hydrogen volume would not meaningfully affect the spatial extent of the gas plume. Consequently, plume migration behaviour and its potential to reach and interact with abandoned wells would remain effectively unchanged. On this basis, H_2 dissolution is not expected to influence the prediction of leakage occurrence in the present framework, and was therefore not included in the simulations.

The influencing variables considered in this study and subsequently used for surrogate model development include the reservoir depth (H_r), formation permeability (K), porosity (ϕ), H_2 injection rate (Q), the distance between injection and abandoned wells (L), and the number of injection-production cycles (N_c). Each cycle consists of six months of H_2 injection followed by six months of production, a configuration that has been widely adopted in previous studies [50–54]. In addition, the selected variable ranges were informed by published studies [33,34,50–57] and were intentionally defined to span a wide range of conditions that may be encountered in future UHS applications. The corresponding ranges are summarized in Table 1. Both the reservoir and aquifer formations were assumed to exhibit permeability anisotropy, a common feature of natural geological media [25]. A typical anisotropy ratio of 1:10 was adopted in this study, with vertical permeability defined as one-tenth of the horizontal permeability [25,58]. An abandoned wellbore was not modelled explicitly with cement and casing components due to the complexity of fluid transport within such systems.

Instead, consistent with previous studies that treat leaky wells as high-permeability vertical pathways [23,25,29,59–61], the abandoned well is modelled as a high-permeability channel. A fixed permeability value of 4000 mD is adopted and this value is not intended to represent any specific abandoned well. Instead, it is used as a screening

Table 1
Variation ranges of the six continuous variables.

variable	Lower bound	Upper bound
H_r , (m)	1000	3000
ϕ	0.1	0.5
K , (md)	100	1000
Q , (Sm ³ /d)	10^4	10^6
L , (m)	70	700
N_c	1	10

assumption to represent a degraded well condition [23,25,62]. In practice, well integrity will depend on cement and casing conditions, and is generally not known in advance for candidate wells. The purpose of this assumption is to assess whether leakage could occur under a conservative degraded-well scenario, and to identify configurations that may require further investigation. The model is therefore designed to address a screening question, rather than to predict leakage behaviour for a specific well. For this reason, the abandoned-well permeability is kept constant and is not treated as an uncertain variable in the surrogate model. This assumption has been further assessed through additional simulations to evaluate its impact on leakage prediction, as presented in the Results and discussion section. Prior to model fitting, several continuous variables were rescaled to reduce differences in numerical magnitude. Reservoir depth, well spacing, and formation permeability were divided by 10^3 , while the hydrogen injection rate was divided by 10^6 . Porosity and the number of cycles were retained in their original forms. These transformations were introduced solely to facilitate model fitting and did not affect the physical interpretation of the variables.

In addition to the six continuous variables, the H_2 -water relative permeability was treated as a categorical variable in the surrogate model to account for uncertainty in the H_2 -water relative permeability curves. Ten distinct sets of H_2 -water relative permeability (K_r) curves, together with the corresponding capillary pressure (P_c) curves, were considered. In the numerical simulations, each set of relative permeability curves was described using a modified Brooks Corey (BC) model, as adopted in previous studies for UHS [40,63,64]. Table 2 lists the parameters used to determine the ten K_r curves based on the modified BC model, which is expressed as follows.

$$k_{rw} = k_{rw,Max} \left(\frac{S_w - S_{wr}}{1 - S_{wr} - S_{gr}} \right)^n \quad (2.3)$$

$$k_{rg} = k_{rg,Max} \left(1 - \frac{S_w - S_{wr}}{1 - S_{wr} - S_{gr}} \right)^m \quad (2.4)$$

$$P_c = \frac{1}{\alpha} \left(\frac{S_w - S_{wr}}{1 - S_{wr}} \right)^{-1/\lambda} \quad (2.5)$$

k_{rw} denotes the relative permeability of water, and $k_{rw,Max}$ represents the maximum relative permeability of water. S_w is the water saturation, and S_{wr} is the residual water saturation. Similarly, k_{rg} denotes the relative permeability of gas, and $k_{rg,Max}$ is the maximum relative permeability of gas. S_g represents the gas saturation, and S_{gr} is the residual gas saturation. The parameters n and m are the exponents associated with the water and gas phases, respectively. The capillary pressure is capped to a maximum pressure value, $P_{c,Max}$, as water saturation approaches the residual water saturation. The term of $\frac{1}{\alpha}$ represents the critical capillary pressure, while λ characterizes the shape of the P_c curve.

The parameters of these K_r curves were derived from published

H_2 -water relative permeability functions [65–71], which report considerable variability in both relative permeability and capillary pressure characteristics. The parameters defining relative permeability functions are typically correlated and constrained by experimental observations, reflecting underlying rock properties. As a result, treating these parameters as independent continuous inputs may generate curve shapes that are physically unrealistic. To address this, relative permeability is represented using discrete curve sets derived from the literature, where each set corresponds to a complete and internally consistent relative permeability and capillary pressure relationship. This ensures that all sampled cases represent physically plausible multiphase flow behaviour, while remaining consistent with the objective of the surrogate model, which focuses on predicting leakage occurrence rather than reconstructing detailed flow functions. This representation also differs from conventional reservoir simulation studies, where a single relative permeability curve is typically adopted for a given reservoir scenario. While this is appropriate for site-specific analysis, it does not capture the variability reported in experimental measurements when exploring a broader parameter space. The present approach therefore enables this variability to be incorporated while maintaining physical consistency across all simulated cases.

Ten K_r -curve families were selected to provide representative coverage of the variability reported in published studies while maintaining a manageable model complexity for surrogate training and analysis. Using fewer curves would compress the range of flow behaviours, while a substantially larger set would reduce the number of simulations available per curve family and limit the ability of the surrogate model to resolve curve-specific effects. For ease of discussion, the curves are grouped into three categories: low, intermediate, and high mobility. However, noticeable differences still exist within each group, and using only three categories would not adequately capture this variability. Therefore, all ten curves are retained as separate inputs in the surrogate model. As shown in Table 2, high mobility curves (1, 5, and 9) correspond to higher gas mobility, characterized by low residual gas saturation and relatively high gas endpoint permeability. Intermediate mobility curves (2, 3, 4, and 8) exhibit intermediate mobility with moderate residual gas effects. Low mobility curves (6, 7, and 10) reflect lower gas mobility, associated with higher residual gas saturation and stronger capillary resistance. The ten K_r curves are shown in Fig. 2.

Based on the above combinations of continuous variables and curve classifications, a total of 800 cases were generated using Latin hypercube sampling (LHS) within the predefined parameter ranges to ensure representative coverage of the multi-dimensional parameter space [53]. For each simulation case, the output variable is defined as a binary indicator of leakage occurrence. Leakage is considered to have occurred when injected hydrogen migrates through the abandoned well and enters the overlying layer within the simulation period. Each case is therefore assigned a value of 1 (leakage occurred) or 0 (no leakage). This binary formulation is adopted to support the screening objective of the

Table 2

Parameters of the ten K_r curves based on the modified BC model. The mobility classification corresponding to each curve is listed in the first column. The curves are constructed to reflect H_2 -water relative permeability behaviour reported in the literature, rather than reproducing individual datasets. Each curve represents a consistent set of relative permeability and capillary pressure functions, and together they cover a range of gas mobility conditions from high mobility to strong trapping.

Curve ID	m	S_{gr}	$K_{rg,Max}$	n	S_{wr}	$K_{rw,Max}$	α	λ	$P_{c,Max}$	Source
1 (High)	2	0.02	0.95	2	0.2	0.98	0.12	0.6	0.3	[66]
2 (Intermediate)	3.5	0.03	0.9	3	0.25	0.95	0.14	0.65	0.5	[66]
3 (Intermediate)	5	0.04	0.85	4	0.3	0.93	0.16	0.7	0.7	[65]
4 (Intermediate)	7	0.06	0.78	5.5	0.35	0.9	0.18	0.75	0.9	[65]
5 (High)	3	0.02	0.98	7	0.18	0.99	0.13	0.62	0.4	[66]
6 (Low)	6	0.08	0.7	2.5	0.5	0.88	0.2	0.8	1.2	[65]
7 (Low)	9	0.1	0.6	6	0.6	0.85	0.24	0.85	1.5	[67]
8 (Intermediate)	5.5	0.07	0.75	8.5	0.4	0.92	0.22	0.78	1	[69]
9 (High)	1.6	0.01	0.9	1.6	0.15	1	0.15	0.72	0.6	[70]
10 (Low)	9.5	0.09	0.55	9	0.7	0.8	0.3	0.9	2	[67]

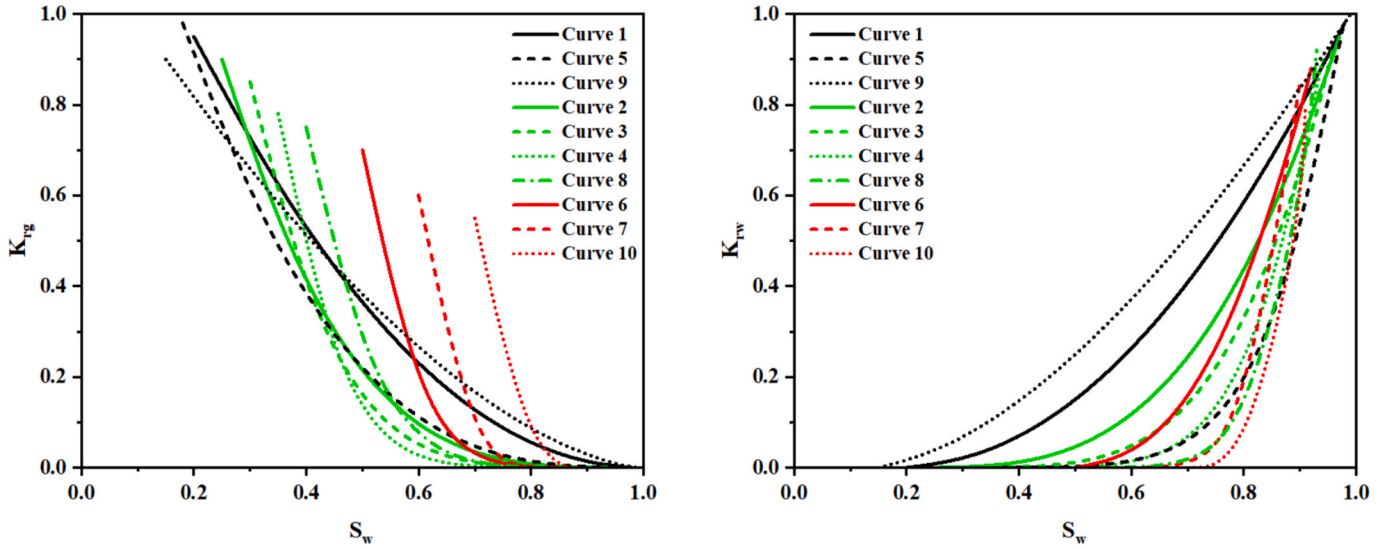


Fig. 2. Gas (H_2) and water relative permeability curves (K_{rg} and K_{rw}) for the ten Kr curve families. Colours indicate mobility class: black (high), green (intermediate), and red (low). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model, where the primary question is whether a given configuration represents a plausible leakage risk. The resulting dataset was used for surrogate model training.

3. Surrogate model development

To investigate the factors contributing to the occurrence of H_2 leakage, logistic regression was adopted in this study. It was selected to provide an interpretable relationship between input variables and leakage probability, enabling direct analysis of how variations in input variables influence predicted outcomes for screening applications. In addition, logistic regression offers a high level of portability, as the model can be expressed as a closed-form equation consisting of a linear combination of inputs followed by a logistic transformation. This allows the model to be implemented directly in a spreadsheet environment without requiring specialised software or machine learning frameworks. This is particularly important for the intended users of the model, including reservoir engineers and regulators, who may not have access to or familiarity with detailed modelling pipelines. Given the high predictive performance achieved in this study, logistic regression represents a suitable and practical choice for the intended purpose.

As a discrete outcome analysis method, logistic regression has been widely used by previous studies for the treatment of category dependent variables [72–75]. However, in the traditional logistic model, all regression coefficients are assumed to be constant across all observations. This assumption in the standard logistic model cannot represent situations where the influence of a variable varies between cases [74,76–78]. To account for unobserved case-level heterogeneity, that is, situations where the effect of a variable may differ between cases, this study proposed two Bayesian logistic regression models: Random effect logistic model (RELM) and Random parameter logistic model (RPLM). RELM introduces a random effect term to capture unobserved heterogeneity, i.e., variability in how variables influence leakage across different cases, while RPLM allows all parameters to vary across observed cases for improving modelling accuracy. In all model specifications, the continuous parameters were included in the transformed forms defined earlier to ensure comparable scales across variables. The categorical parameter, Kr-curve (Curve_ID), was introduced through one-hot encoding, with Curve 3 taken as the reference category, as it represents an intermediate mobility condition, so that the effects of other curve families are interpreted relative to it. These three logistic models are introduced as follows:

3.1. Standard logistic model

In the standard logistic model, the H_2 leakage outcome is marked as $y_i = 1$ and $y_i = 0$ indicates no leakage for the i th observed case. Then y_i Bernoulli(p_i)

$$\text{logit}(p_i) = \ln \left[\frac{p_i}{1 - p_i} \right] = \beta_0 + \mathbf{X} \cdot \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k} \quad (3.1)$$

where $\mathbf{Y} = [y_1, y_2, \dots, y_n]$ denotes outcomes in all observed cases; p_i denotes the probability of H_2 leakage occurring for i th observed case ($i = 1, 2, \dots, n$); β_0 is the model constant; $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_k]$ is the parameter vector and $\mathbf{X} = [x_{i,1}, x_{i,2}, \dots, x_{i,k}]$ is a vector of explanatory variables; k is the number of explanatory variables. Eq.(3.1) can be used to link the relationship between probability of H_2 leakage and contributing factors (explanatory variables).

Through the Eq.(3.1), p_i can be written by

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k})}{1 + \exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k})} \quad (3.2)$$

Then the likelihood function can be given as

$$\begin{aligned} f(\mathbf{Y}|\boldsymbol{\beta}) &= \prod_{i=1}^n [p_i^{y_i} (1 - p_i)^{1-y_i}] \\ &= \prod_{i=1}^n \left[\left(\frac{\exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k})}{1 + \exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k})} \right)^{y_i} \right. \\ &\quad \left. \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k})} \right)^{1-y_i} \right] \end{aligned} \quad (3.3)$$

To estimate parameters in Eq.(3.3), this study employs the Bayesian inference approach. The Markov Chain Monte Carlo (MCMC) method is adopted to determine the posterior distribution of the model parameters. Based on Bayes' theorem, these posterior distributions can be estimated by providing prior distribution, which is shown as

$$f(\boldsymbol{\beta}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta}) \quad (3.4)$$

where $f(\boldsymbol{\beta}|\mathbf{Y})$ is the joint posterior distribution of estimated parameters $\boldsymbol{\beta}$ based on dataset $\mathbf{Y}$. $f(\mathbf{Y}|\boldsymbol{\beta})$ denotes the likelihood function of $\mathbf{Y}$ condi-

tional on parameters β . $\pi(\beta)$ is the prior distribution of the estimated parameters β . The priors are specified as normal distributions with large variances and are therefore effectively non-informative:

$$\beta \sim N(0_k, 10^6 I_k) \quad (3.5)$$

where 0_k is the zero vector with dimension of k . I_k is the identity matrix with dimension of $k \times k$. Then joint posterior distribution can be derived by

$$f(\beta|Y) \propto f(Y|\beta)\pi(\beta) = \prod_{i=1}^n \left[\left(\frac{\exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k})}{1 + \exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k})} \right)^{y_i} \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k})} \right)^{1-y_i} \right] \times \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}10^3} \exp\left(-\frac{1}{2} \frac{(\beta_k)^2}{10^6}\right) \right] \quad (3.6)$$

The joint posterior distribution of $f(\beta|Y)$ in Eq.(3.6) can be generated through Metropolis–Hastings sampling method. It should be noted that the inference is derived from the remaining draws after burning some previous samples.

3.2. Random effect logistic model (RELM)

To account for the unobserved heterogeneity in the standard logistic model, RELM allows one error term to vary across cases. Compared to standard logistic in Eq.(3.1), RELM can be denoted by

$$\text{logit}(p_i) = \ln\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \mathbf{X}\beta + \varepsilon_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k} + \varepsilon_i \quad (3.7)$$

where ε_i is an unstructured error term, assumed to follow an independent normal distribution with zero mean and variance σ_ε^2 , which can be written as

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2), i = 1, 2, \dots, n$$

ε_i is a varying distribution instead of a certain value so RELM in Eq.(3.7) can capture potential heterogeneity more flexibly.

3.3. Random parameter logistic model (RPLM)

Another way to accommodate the unobserved heterogeneity is to allow some or all parameters to vary across observed cases. This type of logistic model is defined as the RPLM, which is given as

$$\text{logit}(p_i) = \ln\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_{i,1} x_{i,1} + \beta_{i,2} x_{i,2} + \dots + \beta_{i,k} x_{i,k} + \varepsilon_i \quad (3.8)$$

where $\beta = [\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,k}]$ indicates a variable random parameter vector across cases. These random parameters are assumed to follow the independent and normal distribution as

$$\beta_{i,j} \sim N(\beta_j, \sigma_j^2), j = 1, 2, \dots, k$$

For prediction of new samples, the leakage probability was calculated using the posterior mean parameter estimates obtained from the MCMC chains.

3.4. Model comparison

In the Bayesian inference framework, the deviance information criterion (DIC) is a commonly used index to compare candidate models [79–82]. This index is an Akaike's information criteria, considering complexity and fitting of models, which is defined as

$$\text{DIC} = \hat{D} + 2(\bar{D} - \hat{D}) \quad (3.9)$$

where the deviance is defined as -2 times the log-likelihood of the model, $\bar{D}$ denotes the mean deviance over posterior distribution, and $\hat{D}$ is the deviance derived from a model that uses the posterior mean parameters. According to Spiegelhalter et al. [83], a difference of more than 5 in DIC is considered to indicate a meaningful improvement in model fit.

A schematic overview of the probabilistic surrogate framework developed in this study is provided in Fig. S1 in the [Supplementary Material](#) to illustrate the overall workflow, from input characterization to model utilization.

4. Results and discussion

4.1. Reservoir simulation analysis and validation

In this section, a case study is presented as an example to illustrate the character of the reservoir simulation and the cyclic operation process. This case allows the full leakage pathway to be clearly observed, from gas flow within the reservoir, upward through the abandoned well, and into the overlying formation. The continuous variables for this case are set as follows: a depth of 2000 m, a porosity of 0.29, a formation permeability of 570 mD, an injection rate of $10^6 \text{ m}^3/\text{day}$, a spacing of 283 m between injection and leaking well, and 10 operation cycles. Relative permeability curve 5 is selected for this case.

Fig. 3 shows the evolution of H_2 saturation obtained from the reservoir simulations under different storage cycles (the 1st and 5th cycles shown as representative examples). During the injection period, H_2 spreads outward from the well due to the pressure gradient. At the same time, it moves upward under buoyancy and accumulates near the top of the reservoir. As a result, a high saturation region forms in the upper part of the formation and extends laterally along the top. This distribution reflects the density difference between H_2 and water and shows that the numerical model is capturing gravity-driven migration. During the production period, the gas saturation region recedes toward the well. Gas flows into the well, and the pressure near the well drops. This reduces the driving force for outward expansion. The alternating expansion and recession of the gas plume during injection and production agrees with the H_2 -water flow behavior previously simulated [53].

As the storage cycles increase, the leakage plume grows and the gas saturation in the upper formation increases and expands. This trend appears in both injection and production periods. The main reason is that more H_2 remains in the reservoir after each cycle. The average gas saturation increases over time, which maintains a vertical pressure difference along the leakage pathway. Even during production, the pressure does not fully recover and some gas remains trapped. As a result, upward migration continues. The predicted leakage behavior is consistent with the results reported by Guo et al [25].

4.2. Surrogate model analysis

The MultiBUGS software was used to conduct Full Bayesian (FB) inference for model estimation. Based on simulated data and prior information, FB inference will generate the posterior distribution of the estimated parameters. All variances were given a prior distribution of $\text{gamma}(0.001, 0.001)$. It should be noted that the variances will be discussed when the estimated means are statistically significant in 95% confidential level. Markov Chain Monte Carlo (MCMC) simulation was performed in MultiBUGS for Bayesian inference. Convergence was assessed using the Brooks-Gelman-Rubin (BGR) diagnostic. Two chains with different starting points were run for 20,000 iterations, of which the first 5,000 were discarded as burn-in. Convergence was considered to be achieved when the BGR statistic was less than 1.2. The trace plots of posterior samples for all parameters in the RPLM model (as an example) exhibit stable behaviour without noticeable trends, indicating satisfactory mixing and convergence. These plots are provided in the

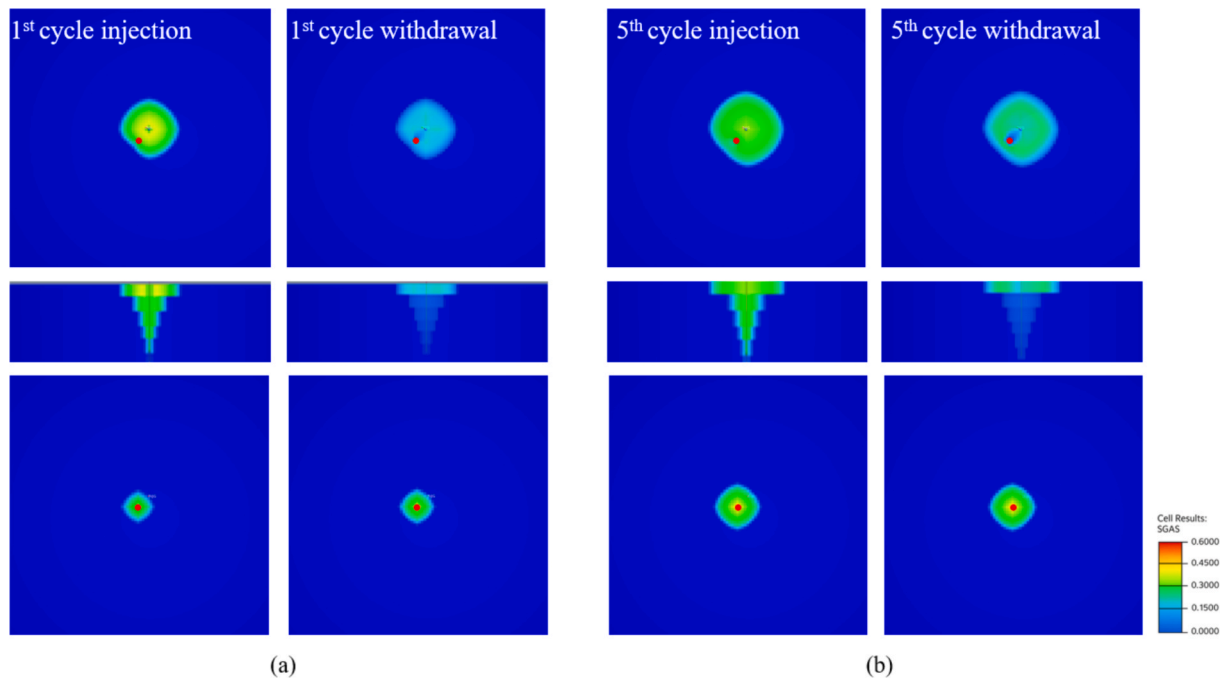


Fig. 3. Evolution of H₂ saturation during cyclic underground hydrogen storage. The first row presents the plan-view distribution in the reservoir, and the second row shows the cross-section through the injection well plane. In addition, the third row shows the plan-view of H₂ leakage into the overlying layer. The red dots in the first and third rows highlight the location of the abandoned well. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Results of the standard logistic model.

Variables	Parameter	mean	sd	Val 2.5pc	Val 97.5pc
	β_0	0.2776	0.5108	-0.731	1.27
Depth	β_1	-0.4836	0.163	-0.8087	-0.1712
Distance	β_2	-6.737	0.5785	-7.892	-5.635
Permeability	β_3	0.3167	0.4365	-0.5397	1.158
Porosity	β_4	-1.745	0.711	-3.14	-0.3427
Injection rate	β_5	3.598	0.4473	2.727	4.472
Number of cycles	β_6	0.01884	0.04674	-0.07294	0.11
Curve-1	β_7	1.814	0.3735	1.087	2.548
Curve-2	β_8	0.905	0.4629	-0.01706	1.793
Curve-4	β_9	0.2677	0.403	-0.5298	1.047
Curve-5	β_{10}	1.663	0.3843	0.9072	2.403
Curve-6	β_{11}	-0.1428	0.5209	-1.206	0.8579
Curve-7	β_{12}	-2.764	0.6028	-3.987	-1.618
Curve-8	β_{13}	-0.00323	0.4226	-0.8448	0.8034
Curve-9	β_{14}	1.978	0.4308	1.133	2.819
Curve-10	β_{15}	-2.699	0.6012	-3.921	-1.555
	DIC	382.4			

Table 4

Results of the RELM, with variables defined as in Table 3.

Parameter	mean	sd	Val 2.5pc	Val 97.5pc
β_0	0.2493	0.5182	-0.776	1.273
β_1	-0.4841	0.1658	-0.808	-0.1628
β_2	-6.75	0.5867	-7.912	-5.624
β_3	0.3256	0.447	-0.5566	1.208
β_4	-1.748	0.7179	-3.151	-0.3458
β_5	3.621	0.453	2.754	4.514
β_6	0.01894	0.04722	-0.07353	0.1112
β_7	1.826	0.3751	1.086	2.546
β_8	0.8963	0.4658	-0.03772	1.786
β_9	0.2604	0.4092	-0.561	1.034
β_{10}	1.67	0.3818	0.9263	2.413
β_{11}	-0.1513	0.521	-1.203	0.8464
β_{12}	-2.773	0.6045	-4.004	-1.64
β_{13}	-0.01856	0.4267	-0.8786	0.8114
β_{14}	1.995	0.4314	1.159	2.833
β_{15}	-2.699	0.6017	-3.918	-1.564
σ_e	0.09755	0.06207	0.02418	0.2531
DIC	335.7			

Fig. S2 and S3 in the Supplementary Material.

The model estimation results are shown in Table 3–Table 5. Among the three models, the RPLM yields the lowest DIC value (330.5), compared to 382.4 for the standard logistic model and 335.7 for the RELM. The difference exceeds the threshold of 5, indicating a substantially better fit. This improvement reflects the ability of the RPLM to account for unobserved heterogeneity by allowing model parameters to vary across cases. It is also noted that the signs and relative magnitudes of the key coefficients are consistent across the three models, which provides additional confidence in the robustness of the results. Accordingly, the subsequent analysis is based on the RPLM.

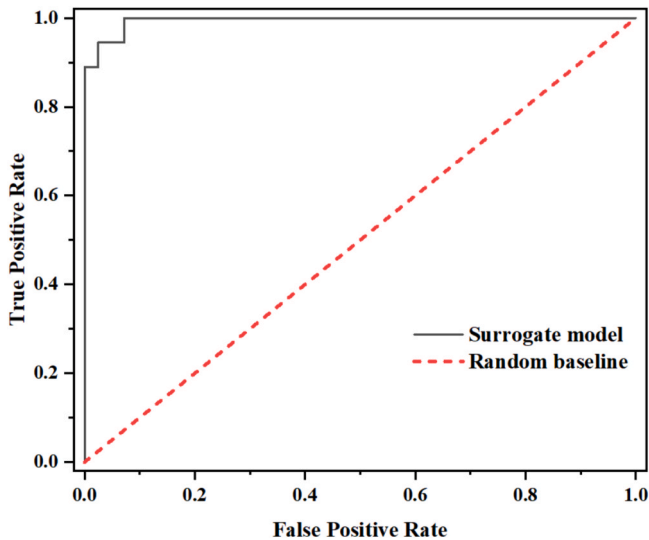
For RPLM, an independent validation set comprising 60 additional cases was generated within the predefined variable ranges using LHS to ensure representative coverage of the multi-dimensional parameter space. These cases were not involved in the surrogate model training

(out-of-sample). For each case, leakage occurrence was determined through reservoir simulation, and the predicted leakage probability was obtained from the surrogate model. The reported performance metrics therefore reflect agreement between the surrogate model and the underlying reservoir simulator, rather than direct comparison with experimental or field observations. Using a probability threshold of 0.5, the surrogate model achieved an overall accuracy of 0.967 (95% Wilson confidence interval: 0.886–0.991), while the recall for leakage events was 0.889 (95% Wilson confidence interval: 0.672–0.969). The corresponding confusion matrix contained 16 true positives, 2 false negatives, 0 false positives, and 42 true negatives. These confidence intervals were included to reflect the statistical uncertainty associated with the limited number of leakage cases in the validation dataset. With a validation set of 60 cases, this corresponds to a small number of missed leakage events, which is an important consideration in safety-oriented applications. The

Table 5

Results of the RPLM, with variables defined as in Table 3.

Parameter	mean	sd	Val_2.5pc	Val_97.5pc
β_0	0.2525	0.5516	-0.911	1.273
β_1	-0.5862	0.193	-0.9771	-0.2256
β_2	-7.234	0.5653	-8.352	-6.124
β_3	0.4168	0.5949	-0.5584	1.581
β_4	-1.653	0.6104	-2.844	-0.4886
β_5	3.863	0.439	3.036	4.736
β_6	-0.01438	0.05411	-0.1156	0.09415
β_7	2.037	0.42	1.263	2.866
β_8	0.8356	0.5623	-0.3162	1.922
β_9	0.2266	0.4641	-0.6867	1.115
β_{10}	1.961	0.406	1.135	2.771
β_{11}	-0.2915	0.6452	-1.61	0.9294
β_{12}	-2.868	0.6403	-4.204	-1.672
β_{13}	-0.09476	0.4905	-1.131	0.8041
β_{14}	2.14	0.5143	1.107	3.106
β_{15}	-2.841	0.6293	-4.114	-1.612
σ_1	0.05763	0.03073	0.02104	0.1416
σ_2	0.3459	0.3869	0.03543	1.442
σ_3	0.2664	0.2294	0.04101	0.8455
σ_4	0.5088	0.6785	0.04738	2.569
σ_5	0.2776	0.3174	0.04105	1.229
σ_6	0.01748	0.005369	0.00936	0.03054
σ_7	0.3406	0.4039	0.03943	1.44
σ_8	0.9922	1.744	0.04621	5.187
σ_9	0.5385	0.6922	0.04168	2.429
σ_{10}	0.3355	0.3697	0.03766	1.349
σ_{11}	1.039	1.619	0.04991	5.575
σ_{12}	0.6355	0.7736	0.04226	2.927
σ_{13}	0.7029	0.934	0.04948	3.383
σ_{14}	1.066	2.258	0.04718	5.874
σ_{15}	0.5594	0.6606	0.04845	2.439
σ_e	0.1183	0.0905	0.02728	0.355
DIC	330.5			

**Fig. 4.** ROC curve for the independent validation dataset of the RPLM surrogate model.

receiver operating characteristic (ROC) curve for the validation dataset is presented in Fig. 4. The curve illustrates the discrimination capability of the surrogate across varying probability thresholds. The corresponding area under the curve (AUC) is 0.995, with a 95% stratified bootstrap confidence interval of 0.980–1.000 based on 10,000 resamples, indicating clear separation between leakage and non-leakage conditions across the investigated scenarios. To further examine this aspect, the effect of adjusting the classification threshold was evaluated. Reducing the threshold from 0.5 to 0.3 allows all leakage cases in the validation set to be correctly identified (Recall = 1.0), at the expense of

an increase in false positives (Precision = 0.783). This reduction in precision is an expected consequence of adopting a more conservative threshold. The performance of the model under different classification thresholds is summarised in Table S1 in the [Supplementary Material](#).

In addition to the DIC comparison presented earlier, the predictive performance of the standard logistic model, the RELM, and the RPLM was also evaluated using the independent validation dataset. The comparison results are summarised in Table S2 in the [Supplementary Material](#). Among the three models, the RPLM achieved the highest accuracy and AUC values, supporting its selection as the final surrogate model. Overall, the close agreement between predicted probabilities and simulation outcomes suggests that the surrogate captures the dominant trends governing leakage occurrence within the explored parameter domain. While a small number of leakage cases are not identified at the default threshold, this can be mitigated in practical applications by adopting a more conservative classification threshold, or by subjecting identified high-risk configurations to further detailed analysis. The model is therefore intended as a screening tool, where threshold selection can be adapted to reflect the relative cost of false negatives and false positives.

4.3. Regression-based analysis of variables

In this section, we analyze the posterior means of the regression coefficients along with their 95% credible intervals to assess how each variable influences leakage probability. In addition, the results are interpreted alongside the distributions shown in Fig. 5, which provide additional insight into how each variable differentiates between leakage and non-leakage cases. By considering both the estimated effect and its associated uncertainty, we determine whether a given variable tends to increase or decrease leakage within the explored parameter range. The results offer a clearer interpretation of the model behavior and provide further evidence that the surrogate model is exploiting physically realistic parameter relationships in the synthetic training data.

The mean coefficient for injection rate is 3.863, with a 95% credible interval from 3.036 to 4.736, entirely above zero, indicating a strong positive effect. Leakage probability rises consistently with increasing injection rate, as higher rates generate stronger pressure buildup near the well and strengthen the driving force for gas migration. As a result, the plume expands more rapidly and can reach the abandoned well earlier in the storage cycle. Fig. 5 shows that leakage cases are associated with higher injection rates, with a clear upward shift in both median and interquartile range. Although some overlap persists, the separation is pronounced, indicating that injection rate is a dominant operational control. A similarly strong but opposite trend is observed for well distance. The mean coefficient for distance is -7.234, with a 95% credible interval from -8.352 to -6.124. Leakage probability declines markedly as the spacing between the injection and abandoned wells increases, which is expected. Fig. 5 shows a clear separation between leakage and non-leakage cases, with leakage predominantly occurring at shorter distances. The limited overlap indicates that well spacing provides strong discriminatory power, reflecting its direct control on the migration pathway.

The mean coefficient for depth is -0.586, with a 95% credible interval from -0.977 to -0.226, indicating a consistent negative effect. Leakage probability decreases as reservoir depth increases. Deeper formations are subject to higher in-situ pressure, under which H_2 is more compressed and occupies less volume. For example, with a hydrostatic pressure gradient of approximately 10 MPa/km and a geothermal gradient of about 30 °C/km, reservoir conditions are on the order of 10 MPa and 323 K at 1 km depth and 30 MPa and 383 K at 3 km depth. Under these conditions, the density of H_2 increases from about 7.1 kg/m³ to 16.5 kg/m³, corresponding to a reduction in volume of roughly 57% for the same injected mass. This limits plume expansion and reduces the likelihood of the gas reaching the abandoned well. A similar tendency has been reported in previous studies for CO_2 subsurface

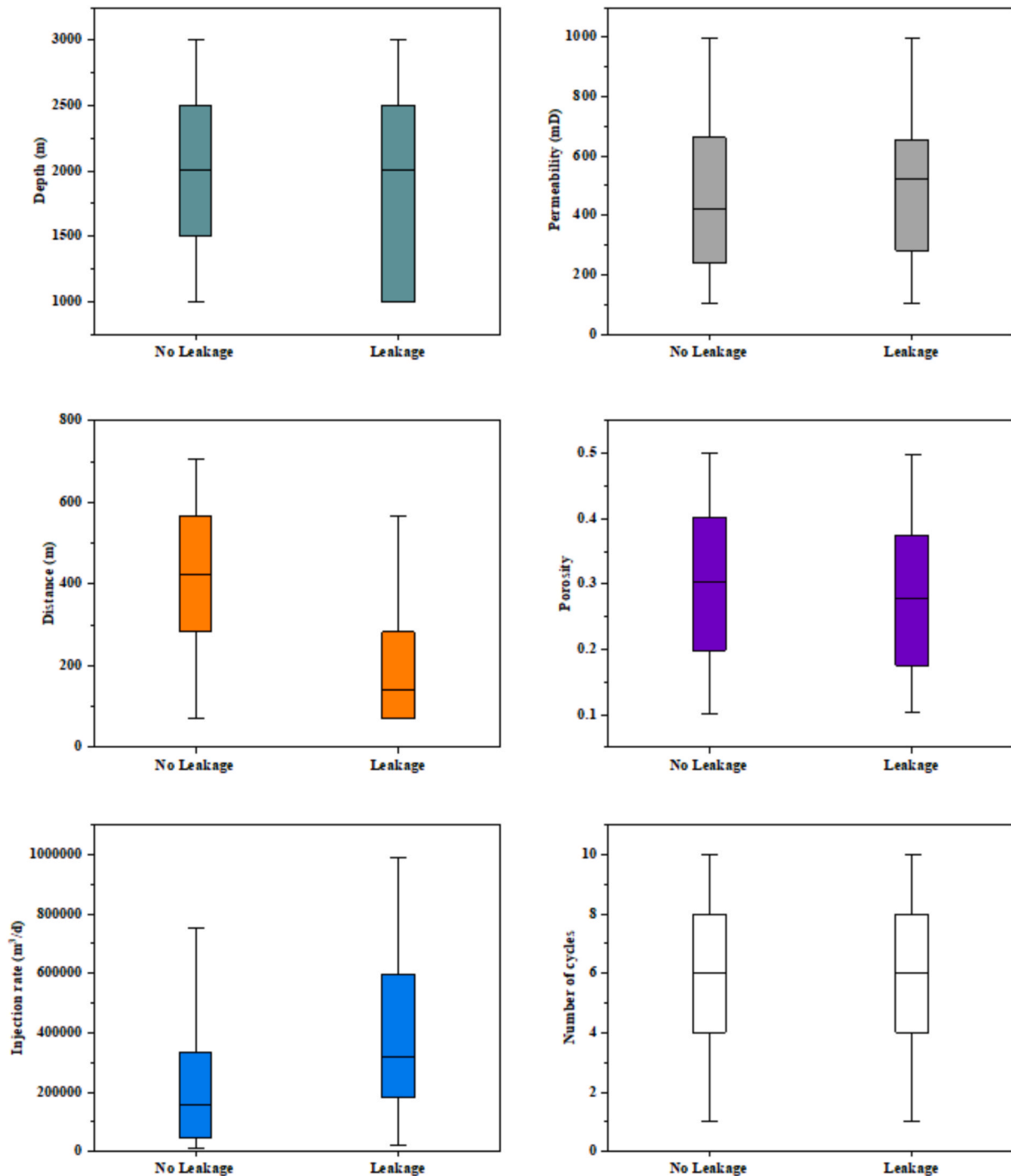


Fig. 5. Distributions of the six continuous variables for leakage and non-leakage cases. The box represents the interquartile range (25th–75th percentiles), and the central line denotes the median value. Variables showing clearer separation between the two groups correspond to stronger effects in the surrogate model, whereas substantial overlap indicates weaker or non-significant influence.

storage, where shallower reservoirs exhibit earlier leakage [21]. Consistent with Fig. 5, leakage cases tend to extend toward lower depth values, although substantial overlap exists between the two groups, indicating a consistent but moderate effect.

The mean coefficient for formation permeability is 0.417, while its 95% credible interval extends from -0.558 to 1.581 and crosses zero. This indicates that formation permeability does not exert a consistent directional effect under the modelled conditions. Although permeability governs gas flow capacity in porous media, its effect on leakage occurrence is limited within the range considered in this study (100–1000

mD). It may be that the tested range, which is based on the characteristics of prospective storage sites, is simply too small to produce a large model effect. As illustrated in Fig. 5, the permeability distributions of leakage and non-leakage cases largely overlap, with no distinct separation between the two groups. Both outcomes span nearly the full sampled interval, and their median values are comparable. This pattern suggests that, under these relatively high-permeability conditions, variations in permeability are insufficient to alter whether gas breakthrough occurs during the modelled operational period. Instead, permeability mainly influences migration rate, whereas leakage

occurrence within the investigated ranges and under homogeneous reservoir conditions is more strongly associated with geometric spacing and injection intensity. Accordingly, its contribution in probability space appears weak and statistically non-significant, and this observation may not generalise to more heterogeneous or lower-permeability reservoirs.

The mean coefficient for porosity is -1.653 , and its 95% credible interval ranges from -2.844 to -0.489 , indicating a negative effect. Increasing porosity is associated with a reduction in leakage probability, as larger pore volume allows gas to be accommodated more gradually along the migration pathway. This delays the formation of a continuous gas phase near the abandoned well. In Fig. 5, leakage cases are slightly shifted toward lower porosity values, although substantial overlap remains. This indicates a secondary influence rather than a dominant control. The mean coefficient for cycle number is -0.014 , and its 95% credible interval ranges from -0.116 to 0.094 , spanning zero. This indicates that cycle number does not have a statistically stable effect within the examined range. While repeated cycling may gradually increase residual gas saturation and promote connectivity, production phases reduce reservoir pressure and draw gas back toward the injection well. These competing effects offset each other, resulting in no clear trend. Fig. 5 confirms this result, as the distributions for leakage and non-leakage cases are nearly identical across the entire range. The opposing effects of gas accumulation and pressure depletion during cyclic operation likely offset each other, resulting in no consistent trend. In conclusion, the variables that display clear separation between leakage and non-leakage cases (well spacing and injection rate) correspond to the strongest regression effects within the investigated ranges and modelling assumptions, whereas variables with substantial or complete overlap (permeability, porosity and cycle number) show limited influence under these conditions. Depth occupies an intermediate role, contributing consistent but less discriminative effects.

The relative permeability curves are treated as a categorical variable and are interpreted based on regression coefficients rather than distributional comparison. Curves 1, 5, and 9 show consistently positive coefficients, with 95% credible intervals entirely above zero, indicating a statistically robust increase in leakage probability relative to the reference curve (Curve 3). This pattern is consistent with their higher gas mobility characteristics. In contrast, Curves 7 and 10 exhibit negative coefficients with credible intervals fully below zero, indicating a stable reduction in leakage probability. These curves correspond to lower gas mobility and stronger capillary restraint, which limit gas connectivity toward the abandoned well. The remaining curves (2, 4, 6, and 8) show credible intervals that overlap zero, suggesting that their influence is conditional and depends on operating conditions and interactions with other variables. Overall, the directional trends align with the predefined mobility classification of the curve families, indicating that the surrogate model captures the underlying physical differences embedded in the relative permeability design. Compared with some of the continuous variables, whose effects are weak within the tested ranges, the differences between curve families are more distinct. This suggests that the representation of relative permeability can influence the predicted leakage outcome even when other conditions are held constant.

In addition to the posterior mean coefficients (β), the estimated standard deviations (σ) of the random parameters quantify the extent to which the corresponding effects vary across the investigated cases. Non-zero σ values indicate the presence of coefficient heterogeneity, implying that the influence of a given variable on leakage probability is not identical across all investigated cases, but varies depending on the geological and operational conditions considered. From a physical perspective, such variability is expected because leakage is governed by the interaction of multiple reservoir properties and operational parameters. Variables such as well spacing, permeability, porosity, injection rate, and relative permeability influence different aspects of plume development, migration, and plume–well interaction, and their effects may therefore vary under different reservoir conditions. Consequently,

the sensitivity of leakage probability to a given variable is not fixed, but may change under different combinations of geological and operational conditions. For the Kr curves, which were introduced as categorical variables using Curve 3 as the reference condition, the corresponding σ values describe the extent to which the differences between a given Kr realisation and the reference curve vary across the investigated cases.

However, the interpretation of this heterogeneity should be considered together with the corresponding β values. Within the random parameter logistic model, each random coefficient is assumed to follow a normal distribution,

$$\beta_i \sim N(\beta_i, \sigma_i^2)$$

where β determines the location of the coefficient distribution and σ controls its spread. To illustrate this concept, Fig. S4 in the Supplementary Material presents the estimated coefficient distributions for distance and permeability as representative examples. As shown in Fig. S4, the β for distance is -7.234 with a σ of 0.346 , whereas permeability has a β value of 0.417 and a σ of 0.266 . Although the σ for distance is slightly larger, the substantially greater magnitude of its β results in a coefficient distribution that remains entirely within the negative region and well separated from zero. In contrast, the permeability coefficient distribution extends much closer to the zero-effect region. This comparison also provides insight into the relative importance of the estimated heterogeneity. While non-zero σ values indicate that the effects of both variables vary across the investigated cases, the practical significance of this variability depends on its magnitude relative to the corresponding mean coefficient. For distance, the estimated coefficient distribution remains concentrated around a strongly negative mean coefficient, indicating that the overall effect is dominated by the average relationship captured by the model. Although some variability exists across cases, it is unlikely to substantially alter the predicted influence of distance on leakage probability. In contrast, permeability exhibits a much weaker average effect while retaining a comparable level of variability. As a result, the estimated coefficient distribution extends much closer to the zero-effect region, suggesting that the influence of permeability is more strongly affected by interactions among geological and operational variables that are not explicitly represented within the linear predictor. Consequently, the effect of permeability on leakage probability is more condition-dependent, and a single average coefficient provides a less complete description of its influence across the investigated scenarios.

4.4. Practical influence on leakage probability

To further compare the relative influence of the continuous variables, the calibrated surrogate model was used to evaluate how leakage probability varies across the investigated range of each variable. Although the regression coefficients provide information on the direction and statistical significance of each effect as discussed in Sec 4.3, they are defined in log-odds space and are therefore not directly comparable across variables with different units and ranges. To obtain a more interpretable measure, each variable was varied from its lower to upper bound, while all other variables were fixed at representative

Table 6
Change in predicted leakage probability across the investigated range of each continuous variable.

Variable	ΔP
Distance	-0.53
Injection rate	0.44
Depth	-0.12
Porosity	-0.06
Permeability	0.04
Cycle number	-0.01

midpoint values to provide a consistent reference condition. The resulting change in predicted leakage probability was used as an indicator of practical influence, defined as

$$\Delta P_i = P(x_i^{max}) - P(x_i^{min}) \quad (4.1)$$

Where P denotes the leakage probability predicted by the surrogate model, and x_i^{min} and x_i^{max} are the lower and upper bounds of variable i , respectively. The corresponding results are summarized in Table 6.

Consistent with the analysis presented in Sec 4.3, the results show that well spacing and injection rate produce the largest changes in leakage probability across their respective ranges. Increasing well spacing leads to a substantial reduction in leakage likelihood, whereas increasing injection rate results in a pronounced increase. In comparison, depth has a moderate effect, while porosity, formation permeability, and cycle number induce relatively small changes within the investigated ranges. These results suggest that, within the investigated ranges and modelling assumptions, operational variables have a more pronounced influence on leakage occurrence than intrinsic reservoir properties.

In addition to the continuous variables, the influence of relative permeability was evaluated by varying the Kr-curve family while keeping all continuous variables fixed at their midpoint values. The predicted leakage probabilities are shown in Fig. 6, with Curve 3 taken as the reference case. A clear difference is observed among the curve families. Curves associated with higher mobility (1, 5, and 9) produce leakage probabilities of approximately 0.47–0.51, compared with 0.11 for the reference case. In contrast, Curves 7 and 10 yield very small leakage probabilities. The remaining curves fall within an intermediate range. Although Curve 6 is also classified as low mobility, it shows a weaker reduction in leakage probability and lies closer to the intermediate range. This reflects that the mobility classification represents a continuum rather than discrete levels. At the same time, the predicted probability for Curve 6 remains lower than all intermediate curves, indicating that the overall ordering across mobility classes is preserved. These results show that changing only the relative permeability representation can lead to substantial differences in predicted leakage probability under otherwise identical conditions. This effect is therefore relevant for practical screening and should be considered alongside the continuous geological and operational variables.

The analysis presented above is based on a simplified representation of the abandoned well as a high-permeability pathway with a fixed

permeability value of 4000 mD. To further examine the influence of this assumption on leakage prediction, 18 additional simulations were performed by varying the abandoned-well permeability over three orders of magnitude (400, 4000, and 40,000 mD), while keeping all other variables unchanged. As shown in Table S3 in supplementary material, Cases 1–6 represent clear leakage cases, Cases 7–9 represent a near-threshold leakage case, Cases 10–12 represent a near-threshold non-leakage case, and Cases 13–18 represent clear non-leakage cases. The results show that the leakage classification remains unchanged for the clear leakage and clear non-leakage cases. The near-threshold non-leakage case also remains non-leaking even at 40,000 mD, whereas the near-threshold leakage case changes to non-leakage only at the lowest permeability value of 400 mD. This suggests that abandoned-well permeability affects the binary leakage outcome only under marginal near-threshold conditions and when the well conductivity approaches the permeability level of the surrounding formation. Since 400 mD is already within the same order of magnitude as the formation permeability range considered in this study, it is better regarded as a low-permeability sensitivity case rather than a representative degraded abandoned-well condition. The timing of leakage varies, with higher permeability leading to earlier arrival of gas in the overlying aquifer. Overall, the results support the interpretation that leakage occurrence is primarily controlled by plume migration at the reservoir scale, while abandoned-well permeability mainly influences the rate and completion of vertical transport after plume arrival.

5. Conclusions

This study presents a reservoir-scale probabilistic framework for assessing hydrogen leakage through abandoned wells in depleted gas reservoirs. A logistic surrogate model was developed based on multi-phase reservoir simulations, enabling rapid evaluation of leakage likelihood across a broad range of geological and operational conditions.

Within the investigated variable ranges and under homogeneous reservoir conditions, well spacing and injection rate emerge as the primary controls on leakage probability. In contrast, formation permeability does not exhibit a consistent directional effect under the relatively high-permeability conditions considered (100–1000 mD). Reservoir depth and porosity are negatively associated with leakage probability, reflecting their influence on in-situ pressure conditions and storage capacity.

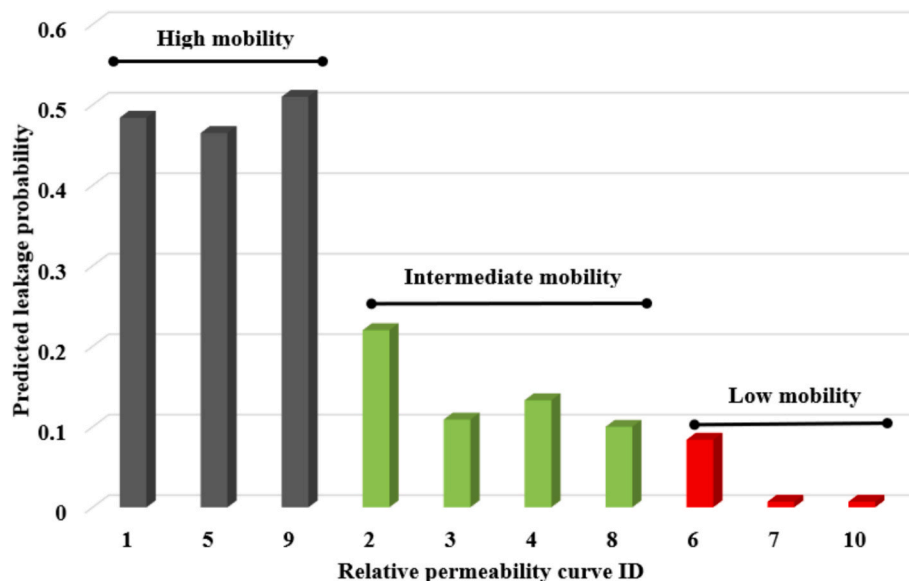


Fig. 6. Predicted leakage probability for the ten Kr curves, with all continuous variables fixed at their midpoint values. Colors indicate mobility class: black (high), green (intermediate), and red (low). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Uncertainty in H₂-water relative permeability also affects leakage predictions. Kr-curve families characterized by higher gas mobility increase leakage likelihood relative to the reference condition, whereas lower-mobility curves reduce it. These trends are consistent with the underlying physical-based flow behavior represented in the reservoir simulations and the surrogate model successfully captures these structural flow differences. The proposed framework provides an efficient approach for leakage-aware screening and operational planning in UHS projects. By integrating reservoir-scale physics with probabilistic modelling, this work supports risk-informed evaluation during the early design stage of UHS deployment.

It should be noted that the present framework adopts a simplified representation of multiphase flow behaviour by treating relative permeability as discrete curve sets, which enables efficient handling of uncertainty but does not capture continuous variability. Furthermore, field-scale datasets for hydrogen leakage through abandoned wells in UHS are not yet available, reflecting a limitation of the current state of the field. In addition, the reservoir is represented as homogeneous in this study. While real subsurface systems are inherently heterogeneous, such heterogeneity is typically site-specific and requires detailed characterisation. Incorporating these features into a general screening framework would require site-specific information that is not available at the stage considered in this study. These simplifications are consistent with the role of the framework as a screening tool, where the objective is to provide a first-order assessment of leakage risk across a wide range of conditions. Within this framework, the present model represents the first stage of a two-stage workflow, in which the objective is to screen a wide range of configurations and identify those that may pose a plausible leakage risk.

In the second stage, configurations identified as high-risk should be analysed in greater detail using more advanced simulations that incorporate reservoir heterogeneity and well-specific characteristics. This would enable more precise quantification of leakage behaviour, including leakage rate, timing and cumulative volume, as part of future work. In addition, the simulation dataset generated in this study can also be used to develop conditional surrogate models that predict these quantities specifically for configurations classified as leakage prone.

CRediT authorship contribution statement

Qiuhao Chang: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **David Dempsey:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Guilong Xu:** Writing – review & editing, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Liehui Zhang:** Writing – review & editing, Resources, Methodology, Conceptualization. **Yulong Zhao:** Writing – review & editing, Resources, Methodology, Conceptualization.

Funding

This work was supported by the New Zealand government under the MBIE contract UOCX2207 for the Endeavour research program “Pūhiko Nukutū: a green hydrogen geostorage battery in Taranaki”.

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

We acknowledge the Civil and Environmental Engineering

department at the University of Canterbury, New Zealand, for computational resources and dedicated support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fuel.2026.140374>.

Data availability

No data was used for the research described in the article.

References

- [1] Oliveira AM, Beswick RR, Yan Y. A green hydrogen economy for a renewable energy society. *Curr Opin Chem Eng* 2021;33:100701.
- [2] Wappler M, Unguder D, Lu X, Ohlmeyer H, Teschke H, Lueke W. Building the green hydrogen market—Current state and outlook on green hydrogen demand and electrolyzer manufacturing. *Int J Hydrogen Energy* 2022;47:33551–70.
- [3] Staffell I, Scamman D, Abad AV, Balcombe P, Dodds PE, Ekins P, et al. The role of hydrogen and fuel cells in the global energy system. *Energ Environ Sci* 2019;12:463–91.
- [4] Chang Q, Wu Y, Tian Z. Molecular dynamics study of H₂-oil mixture behavior in shale nanopores. *Computational Energy Science* 2025;2:85–92.
- [5] Yao J, Wu Z, Luo Q, Hu B, Li Q. Reaction-induced localized self-heating achieving fast and efficient hydrogen storage of modular Mg-Ni-Y alloy bed. *Journal of Materials Science & Technology* 2025.
- [6] Jung S, Zhuo Y, Shen Y. Numerical study of a metal hydride hydrogen storage system equipped with phase change material. *Fuel* 2026;410:137842.
- [7] Chang Q, Dempsey D, Zhang L, Zhao Y, Huang L. Molecular dynamics insights into gas-water interfacial tension: Optimizing hydrogen storage in subsurface conditions. *Int J Hydrogen Energy* 2024;64:896–905.
- [8] Zivar D, Kumar S, Foroozesh J. Underground hydrogen storage: a comprehensive review. *Int J Hydrogen Energy* 2021;46:23436–62.
- [9] Tarkowski R. Underground hydrogen storage: Characteristics and prospects. *Renew Sustain Energy Rev* 2019;105:86–94.
- [10] Perera M. A review of underground hydrogen storage in depleted gas reservoirs: Insights into various rock-fluid interaction mechanisms and their impact on the process integrity. *Fuel* 2023;334:126677.
- [11] Chang Q, Huang L, McKenzie K, Carere C, Stott M, Nicol A, et al. Influence of hydrogen sulfide on gas-water interface in underground hydrogen storage: a molecular dynamics study. *J Storage Mater* 2024;97:112766.
- [12] Xie C, Huang J, Jiang S, Li Y, Yin X, Zhao H. Hydrogen leakage through organic-rich caprock: a molecular simulation study under water and cushion gas conditions. *Fuel* 2025;402:135980.
- [13] Zhang H, Luo X, Yang D, Liu K, Xie Q, Diao R. Molecular simulation of H₂ loss by dissolution in caprock water-saturated nanopores under the nanoconfinement effect for underground hydrogen storage. *Energy Fuel* 2023;37:19357–68.
- [14] Liu J, Zhang T, Sun S. Molecular mechanisms of hydrogen leakage through caprock in moisture and residual gas conditions: a molecular dynamics–Monte Carlo study. *Phys Fluids* 2024;36.
- [15] Chang Q, Dempsey D, Huang L. Investigation of Gas Behaviors in Caprock for Understanding Gas Leakage during Underground Hydrogen Storage. *SPE Asia Pacific Oil and Gas Conference and Exhibition: SPE*. 2024.
- [16] Kahzadvand K, Ghasemi M, Yazaydin AO, Babaei M. Risk of H₂ Leakage into caprock and the role of cushion gas as a barrier in H₂ geo-storage: a molecular simulation study. *J Phys Chem C* 2024;128:16161–71.
- [17] Li Z, Li T, Meng L, Wang X, Sun H, Zhang M, et al. Molecular mechanisms of hydrogen leakage and blockage in kaolinite nano-cracks for underground hydrogen storage. *Phys Fluids* 2025;37.
- [18] Talukdar M, Zheng F, Jha B. Leakage Mechanisms for Caprock Fracturing and Fault Re-Activation in Porous Underground Hydrogen Storage Media. *SPE Reservoir Simulation Conference: SPE* 2025. p. D011S03R02.
- [19] Mazzoldi A, Oldenburg CM. Leakage Risk Assessment of CO₂ Transportation by Pipeline at the Illinois Basin Decatur Project. *Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States)*; 2013.
- [20] Kim J-W, Lu Z. Association between localized geohazards in West Texas and human activities, recognized by Sentinel-1A/B satellite radar imagery. *Sci Rep* 2018;8:4727.
- [21] Kopp A, Binning PJ, Johannsen K, Helmig R, Class H. A contribution to risk analysis for leakage through abandoned wells in geological CO₂ storage. *Adv Water Resour* 2010;33:867–79.
- [22] Qiao T, Hoteit H, Fahs M. Semi-analytical solution to assess CO₂ leakage in the subsurface through abandoned wells. *Energies* 2021;14:2452.
- [23] Postma TJ, Bandilla KW, Celia MA. Estimates of CO₂ leakage along abandoned wells constrained by new data. *Int J Greenhouse Gas Control* 2019;84:164–79.
- [24] Shoushtari S, Jafari A, Namdar H, Khoozan D. Modeling, qualification, and quantification of hydrogen leakage in multilayered reservoirs. *Int J Hydrogen Energy* 2024;91:636–48.
- [25] Guo R, Wang H, Leng J, Hosseini SA. Evaluation of Hydrogen Leakage through Abandoned Wells to Overlying Saline Aquifers during Underground Hydrogen Storage in Depleted Natural Gas Reservoirs. *Gas Sci Eng* 2025;205659.

- [26] He X, Zhu W, Kwak H, Yousef A, Hoteit H. Deep learning-assisted Bayesian framework for real-time CO₂ leakage locating at geologic sequestration sites. *J Clean Prod* 2024;448:141484.
- [27] Baek S, Bacon DH, Huerta NJ. Enabling site-specific well leakage risk estimation during geologic carbon sequestration using a modular deep-learning-based wellbore leakage model. *Int J Greenhouse Gas Control* 2023;126:103903.
- [28] Gurwicz A, Chen J, Gutman D, Gildin E. Assessing risk in long-term CO₂ storage under uncertainty via survival analysis-based surrogates. *SPE J* 2025;30:2837–54.
- [29] Chen B, Mehana MZ, Pawar RJ. Quantitatively evaluating greenhouse gas leakage from CO₂ enhanced oil recovery fields. *Advances in Geo-Energy Research* 2023;7: 20–7.
- [30] Ho CK, Webb SW. Gas transport in porous media: Springer; 2006.
- [31] Nordbotten JM, Fernø MA, Flemisch B, Kovscek AR, Lie K-A, Both JW, et al. Benchmarking CO₂ storage simulations: results from the 11th Society of Petroleum Engineers Comparative Solution Project. *Int J Greenhouse Gas Control* 2025;148: 104519.
- [32] Sheikhi S, Flynn MR. Can a hydrodynamic model predict the flow evolution of a hydrogen plume in a depleted natural gas reservoir? *Int J Hydrogen Energy* 2024; 81:795–804.
- [33] Liu J, Dempsey D, Nicol A, Parker M, Passmore J. Impact of operational parameters on water invasion in underground hydrogen storage: a numerical investigation of the Ahuroa field, New Zealand. *Int J Hydrogen Energy* 2025;153:150236.
- [34] Liu J, Dempsey D, Nicol A, Parker M, Passmore J. Effects of reservoir interlayers on hydrogen migration and distribution in underground hydrogen storage. *Int J Hydrogen Energy* 2026;211:153704.
- [35] Huang Z, Maher K, Kovscek AR. Hydrogeochemical modeling of hydrogen storage in depleted gas reservoirs: Insights from local and global sensitivity analysis. *Appl Energy* 2025;391:125940.
- [36] Chang Q, Dempsey D. Numerical Reservoir Simulations of potential Gas Leakage through Abandoned Wells during Hydrogen Storage in Depleted Reservoirs. *SPE Symposium: Decommissioning and Abandonment*: SPE 2025. p. D012SR01.
- [37] Verma A, Malik S, Pal M. In: Feasibility of UHS in Baltic basin: a case study of Lithuanian reservoir in diverse geological settings. *European Association of Geoscientists & Engineers*; 2024. p. 1–5.
- [38] Huang Z, Ates H, Eichmann S, Kovscek A. Impact of Geochemical Reactions for Under-Ground Storage of Hydrogen Versus Ammonia. *SPE Middle East Oil and Gas Show and Conference*: SPE. 2025.
- [39] Dempsey D, Parker M, Liu J, Chang Q, Nicol A, Passmore J. Reservoir Simulation-Informed Cost-Benefit Analysis for Structural Assessment: a Case Study of Hydrogen Geostorage in the Ahuroa Gas Field, Aotearoa New Zealand. *SPE Asia Pacific Oil and Gas Conference and Exhibition*: SPE. 2024.
- [40] Liu J, Dempsey D, Nicol A, Parker M, Titus K. Impacts of reservoir permeability anisotropy on hydrogen migration and recovery efficiency in geological storage. *Fuel* 2026;408:137596.
- [41] Smith D, Arnold D, Hough E, Busch A. Numerical modelling of the effects of permeability contrasts on underground hydrogen storage in sandstone reservoirs. *Geoenergy* 2024;2. *geoenergy*2023 -039.
- [42] Mahdi DS, Al-Khdheawi EA, Yuan Y, Zhang Y, Iglauer S. Hydrogen underground storage efficiency in a heterogeneous sandstone reservoir. *Advances in Geo-Energy Research* 2021;5:437–43.
- [43] Chen M, Abdalla OA, Izady A, Nikoo MR, Al-Maktoumi A. Development and surrogate-based calibration of a CO₂ reservoir model. *J Hydrol* 2020;586:124798.
- [44] Li C, Wang X, Yu W, Zhou P. A General Surrogate Model for CO₂ Flooding Dynamic Prediction based on Dimensionless Features and Implicit Time. *ACS Omega* 2026; 11:18486–505.
- [45] Wang Z. A fast physics-based data-driven surrogate model for unconventional reservoirs with rapid decline and well interference. *Geoenergy Sci Eng* 2024;237: 212772.
- [46] Madani M, Alipour M. Gas-oil gravity drainage mechanism in fractured oil reservoirs: surrogate model development and sensitivity analysis. *Comput Geosci* 2022;26:1323–43.
- [47] Kandala V, Govindarajan SK. Machine learning-based surrogates for predicting low saline water flooding in sandstone reservoirs. *Energy Fuel* 2025;39:11607–24.
- [48] Wiebe R, Gaddy V. The Solubility of Hydrogen in Water at 0, 50, 75 and 100 from 25 to 1000 Atmospheres. *J Am Chem Soc* 1934;56:76–9.
- [49] Heinemann N, Alcalde J, Miocic JM, Hangx SJ, Kallmeyer J, Ostertag-Henning C, et al. Enabling large-scale hydrogen storage in porous media—the scientific challenges. *Energy Environ Sci* 2021;14:853–64.
- [50] Lysy M, Fernø MA, Ersland G. Effect of relative permeability hysteresis on reservoir simulation of underground hydrogen storage in an offshore aquifer. *J Storage Mater* 2023;64:107229.
- [51] Abdellatif M, Hashemi M, Azizmohammadi S. Large-scale underground hydrogen storage: integrated modeling of reservoir-wellbore system international journal of hydrogen energy 2023;48:19160–71.
- [52] Camargo JT, White JA, Hamon FP, Fakeye V, Buscheck TA, Huerta N. Managing reservoir dynamics when converting natural gas fields to underground hydrogen storage. *Int J Hydrogen Energy* 2024;49:1261–73.
- [53] Mao S, Chen B, Malki M, Chen F, Morales M, Ma Z, et al. Efficient prediction of hydrogen storage performance in depleted gas reservoirs using machine learning. *Appl Energy* 2024;361:122914.
- [54] Huang T, Moridis GJ, Blasingame TA. Numerical simulation of seasonal underground hydrogen storage: Role of the initial gas amount on the round-trip hydrogen recovery efficiency. *Int J Hydrogen Energy* 2024;88:289–312.
- [55] Okoroafor ER, Saltzer SD, Kovscek AR. Toward underground hydrogen storage in porous media: Reservoir engineering insights. *Int J Hydrogen Energy* 2022;47: 33781–802.
- [56] Wang J, Wu R, Zhao K, Bai B. Numerical simulation of underground hydrogen storage converted from a depleted low-permeability oil reservoir. *Int J Hydrogen Energy* 2024;69:1069–83.
- [57] Bo Z, Boon M, Hajibeygi H, Hurter S. Impact of experimentally measured relative permeability hysteresis on reservoir-scale performance of underground hydrogen storage (UHS) international journal of hydrogen energy 2023;48:13527–42.
- [58] Al-Dujaili AN, Shabani M, AL-Jawad MS. Identification of the best correlations of permeability anisotropy for Mishrif reservoir in West Qurna/1 oil Field, Southern Iraq. *Egypt J Pet* 2021;30:27–33.
- [59] Class H, Ebigo A, Helmig R, Dahle HK, Nordbotten JM, Celia MA, et al. A benchmark study on problems related to CO₂ storage in geologic formations: Summary and discussion of the results. *Comput Geosci* 2009;13:409–34.
- [60] Pawar RJ, Watson TL, Gable CW. Numerical simulation of CO₂ leakage through abandoned wells: model for an abandoned site with observed gas migration in Alberta. *Canada Energy Procedia* 2009;1:3625–32.
- [61] Das AK, Hassanzadeh H. Analytical solutions for brine invasion across caprock and leakage through an abandoned well in CO₂ storage sites. *J Hydrol* 2023;625: 130046.
- [62] Lackey G, Vasyukivskaya VS, Huerta NJ, King S, Dilmore RM. Managing well leakage risks at a geologic carbon storage site with many wells. *Int J Greenhouse Gas Control* 2019;88:182–94.
- [63] Basyouni SS, Kabbaj N, El-Amin MF. Relative Permeability Modeling for Underground Hydrogen Storage: a Comparative Analysis of Hysteresis Effects. In: 2025 22nd International Learning and Technology Conference (L&T): IEEE; 2025. p. 124–9.
- [64] Brooks RH, Corey AT. Properties of porous media affecting fluid flow. *J Irrig Drain Div* 1966;92:61–88.
- [65] Yekta A, Manceau J-C, Gaboreau S, Pichavant M, Audigane P. Determination of hydrogen–water relative permeability and capillary pressure in sandstone: application to underground hydrogen injection in sedimentary formations. *Transp Porous Media* 2018;122:333–56.
- [66] Hashemi L, Blunt M, Hajibeygi H. Pore-scale modelling and sensitivity analyses of hydrogen-brine multiphase flow in geological porous media. *Sci Rep* 2021;11 (8348):2021.
- [67] Lysy M, Føyen T, Johannessen EB, Fernø M, Ersland G. Hydrogen relative permeability hysteresis in underground storage. *Geophys Res Lett* 2022;49: e2022GL100364.
- [68] Rezaei A, Hassanpouryouzband A, Molnar I, Derikvand Z, Haszeldine RS, Edlmann K. Relative permeability of hydrogen and aqueous brines in sandstones and carbonates at reservoir conditions. *Geophys Res Lett* 2022;49: e2022GL099433.
- [69] Boon M, Hajibeygi H. Experimental characterization of H₂/water multiphase flow in heterogeneous sandstone rock at the core scale relevant for underground hydrogen storage (UHS). *Sci Rep* 2022;12:14604.
- [70] Higgs S, Da Wang Y, Sun C, Ennis-King J, Jackson SJ, Armstrong RT, et al. Direct measurement of hydrogen relative permeability hysteresis for underground hydrogen storage. *Int J Hydrogen Energy* 2024;50:524–41.
- [71] Bagheri M, Mahani H, Ayatollahi S, Zivar D. Direct numerical approach for determination of H₂-water flow functions applicable to underground hydrogen storage and production. *J Storage Mater* 2024;96:112573.
- [72] Al-Mudhafer WJ. Multinomial logistic regression for bayesian estimation of vertical facies modeling in heterogeneous sandstone reservoirs. *Offshore Technology Conference Asia: OTC 2024*. p. OTC-24732-MS.
- [73] Dong J-J, Tung Y-H, Chen C-C, Liao J-J, Pan Y-W. Logistic regression model for predicting the failure probability of a landslide dam. *Eng Geol* 2011;117:52–61.
- [74] Zhu Z, Lin C, Zhang X, Wang K, Xie J, Wei S. Evaluation of geological risk and hydrocarbon favorability using logistic regression model with case study. *Mar Pet Geol* 2018;92:65–77.
- [75] Xu G, Yang Z, Ying J, Xie S, Bai S, Qi Y. An integrated potential safety hazard assessment framework in connected car-following scenario. *Accid Anal Prev* 2025; 216:108010.
- [76] Sandi E, Cahill A, Welch L, Beckie R. Characterizing oil and gas wells with fugitive gas migration through Bayesian multilevel logistic regression. *Sci Total Environ* 2021;769:144678.
- [77] Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med* 2011;18: 1099–104.
- [78] Xu G, Yang Z, Ying J, Xie S, Zhang T. How does the driver respond to rear-end risk under time-varying driving styles: a temporal instability test using the model with heterogeneity in means and variances. *Accid Anal Prev* 2026;224:108288.
- [79] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B (Stat Methodol)* 2002;64:583–639.
- [80] Shriver D, Yi N. Deviance information criterion (DIC) in Bayesian multiple QTL mapping. *Comput Stat Data Anal* 2009;53:1850–60.
- [81] Lord D, Qin X, Geedipally SR. Highway safety analytics and modeling. Elsevier 2026.
- [82] Celeux G, Forbes F, Robert CP, Titterton DM. Deviance information criteria for missing data models. 2006.
- [83] Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS user manual. version 2003.