geosciences

MDPI

Article

Full Waveform Inversion via Uncertainty-Aware Deep Learning for CO_2 Plume Detection

Alejandro Duitama Leal [1,2,3,*], Luis A. Montes Vides [1], Héctor J. Hortúa Orjuela [3,4] and Luis A. Castillo López [5]

[1] Departamento de Geociencias, Universidad Nacional de Colombia, Bogotá 111321, Colombia; lamontesv@unal.edu.co
[2] Escuela de Ingeniería y Ciencias Básicas, Institución Universitaria Politécnico Grancolombiano, Bogotá 110231, Colombia
[3] Grupo Signos, Universidad El Bosque, Bogotá 110121, Colombia; hhortuao@unbosque.edu.co
[4] Instituto de Neurociencias, Universidad El Bosque, Bogotá 110121, Colombia
[5] Laboratory of Parallel Architectures for Signal Processing (LAPPS), Instituto Metrópole Digital (IMD), Universidade Federal do Rio Grande do Norte, Natal 59078-970, Brazil; luis.lopez.highres@imd.ufrn.br
* Correspondence: aduitama@poligran.edu.co

Abstract

This study presents a probabilistic deep learning framework developed for full waveform inversion (FWI) to detect and characterise subsurface velocity anomalies induced by stored CO_2. Three convolutional neural network architectures—LinkNet, UNet, and a feature pyramid network (FPN)—were trained and evaluated on the Kimberlina CO_2 synthetic dataset from the OpenFWI repository, which simulates CO_2 injection scenarios within stratified reservoirs. The FPN architecture with an SE-ResNet50 backbone achieved the best performance ($R^2 = 0.9386$, $MSE = 3.1676 \times 10^{-4}$, $MAE = 4.5249 \times 10^{-3}$). Uncertainty quantification was performed using variational inference with the Flipout estimator, which enables the separation of aleatoric and epistemic uncertainty. The predicted velocity models accurately reconstruct the background stratigraphy and the low-velocity anomaly associated with the CO_2 plume. The aleatoric uncertainty remains at around 10^1 m/s and is spatially localised to delineate the geometry of the low-velocity zone. In contrast, the epistemic uncertainty is several orders of magnitude smaller, at around 10^{-3} m/s, and increases specifically over the lowest-velocity region of the anomaly. This structured distribution indicates where the velocity reconstruction is least reliable. Calibration substantially reduced the overall miscalibration area, decreasing the magnitude of the aleatoric component while slightly increasing that of the epistemic uncertainty over the CO_2 plume anomaly, although the spatial structure of both components remained unchanged. This uncertainty decomposition and calibration help to distinguish between data noise and model limitations. It provides direct information for designing acquisitions and assessing risks, which is an area that is not widely covered in the current literature.

Keywords: seismic inversion; feature pyramid network; CO_2; epistemic uncertainty; aleatoric uncertainty

Check for updates

Academic Editor: Enrico Priolo

Received: 25 July 2026
Revised: 13 September 2026
Accepted: 30 September 2026
Published: 21 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Data-driven and deep learning frameworks have a wide range of applications in subsurface characterisation from seismic data, including the prediction of petrophysical properties such as porosity, lithology, and elastic impedance [1–6], facies characterisation [7–10],

prediction of fluid and lithofacies distributions [11–13], characterisation of carbonate reservoirs [14,15], three-dimensional modelling from multiple geophysical sources [16], and related tasks such as fault segmentation and AVO inversion [17–19].

One of the fields where deep learning has been widely applied is in inversion problems. Seismic inversion transforms reflection data into quantitative models of rock properties, providing a fundamental basis for reservoir characterisation. Of the available methodologies, full waveform inversion (FWI) is notable for its ability to estimate parameters such as velocity and density directly from the full waveform simultaneously [20–24]. This capability has, in turn, enabled the discrimination of different fluids, including natural gas, CO_2 and hydrogen [25]. In recent years, artificial intelligence has been integrated into these workflows through the use of convolutional neural networks for FWI. This improves the detection of velocity anomalies associated with gases and enables their characterisation [26–32].

Unlike traditional FWI, which is limited by factors such as high computational cost, convergence to local minima, and dependence on an accurate initial model, deep neural networks have enabled faster inversion, modelling of nonlinear relationships, and quantification of uncertainties through Bayesian frameworks [33–44]. Due to the limited availability of field data, particularly for CO_2 characterisation, these studies have relied on synthetic databases generated through seismic modelling [45–53].

CO_2 injection modifies the properties of subsurface rock, reducing its density and increasing its compressibility. This results in a decrease in seismic velocities, which facilitates the detection of gas-saturated zones and enables the estimation of reserves through FWI [22,24,54–58]. Gases of interest, such as methane, ethane and CO_2, together with emerging resources such as geological hydrogen, tend to accumulate in structural traps. These gases exhibit distinctive seismic signatures that can be detected using advanced inversion techniques [59–63]. Geophysical monitoring is essential for ensuring the safety of gas storage, CO_2 sequestration and geothermal projects [64–68].

Monitoring the injection and migration of CO_2 in the subsurface is essential to ensure the safety, efficiency, and long-term integrity of storage operations. In this context, static, single-baseline FWI is a key tool for characterising the velocity model of the subsurface at a given point in time. This is complemented by approaches based on refracted-wave analysis and well-data integration [69–72]. Furthermore, data-driven and physics-constrained deep learning models demonstrate strong capabilities in subsurface fluid dynamics across stratified media [73] and simultaneously estimating petrophysical properties and reservoir dynamic changes [74,75]. Such static inversions are a necessary methodological step towards more advanced time-lapse (4D) monitoring frameworks, which track changes in the velocity field over time. Due to the limited availability of real monitoring data, synthetic databases comprising single-baseline and time-lapse datasets have also been suggested for training and validating the relevant models [76–79].

Due to the ill-posed nature of seismic inversion, various approaches have been adopted to quantify uncertainty, including Bayesian methods, ensemble-based techniques, and Monte Carlo dropout integrated with full wavefield inversion (FWI) [75,80–83]. Beyond FWI, uncertainty quantification based on deep learning has also been applied to a range of related geophysical inversion and interpretation tasks, including fault detection, AVO inversion, refraction tomography, and lithological modelling [16–18,84]. However, most of these studies, including those focused on CO_2 monitoring, only estimate total uncertainty without distinguishing between its epistemic (reducible and associated with model knowledge) and aleatoric (irreducible and associated with data noise) components [80–83,85]. Even in other fields, few studies have quantified these components separately, and even fewer have assessed whether the resulting uncertainty estimates are properly calibrated. In seismics, the number of studies addressing the separation and

calibration of aleatoric and epistemic uncertainty is even smaller, which limits our ability to focus our efforts on acquiring more data, better characterising the system's intrinsic noise, or trusting the reported uncertainty as a statistically meaningful measure of confidence.

In several of these fields, efforts have specifically focused on decoupling epistemic from aleatoric uncertainty using approximate Bayesian approaches, such as Monte Carlo dropout, deep ensembles, and invertible neural networks [86–88]. This body of work confirms that decomposing predictive uncertainty into its two components is feasible and informative in diverse geophysical settings. Doing so improves model interpretability and supports calibration assessment. Nevertheless, few studies in any field have assessed whether the resulting uncertainty estimates are properly calibrated. In seismology, particularly in relation to CO_2 monitoring, the number of studies addressing the separation and calibration of aleatoric and epistemic uncertainty is especially limited. This gap constrains the ability to focus efforts on acquiring more data, characterising the system's intrinsic noise, or trusting the reported uncertainty as a statistically meaningful measure of confidence. It is this gap that this study aims to address.

The main contribution of this study is twofold. Firstly, it comparatively evaluates three convolutional neural network architectures—UNet, LinkNet and FPN—for full waveform inversion to reconstruct CO_2-induced velocity anomalies. Second, the architecture that performs best is adapted into a probabilistic framework that can estimate predictive uncertainty. This uncertainty is then decoupled into epistemic and aleatoric components and recalibrated pixel by pixel to improve the reliability of these estimates. This combination of a systematically selected inversion architecture and decoupled, recalibrated uncertainty quantification provides accurate velocity reconstructions and a principled, statistically reliable means of distinguishing between reducible model limitations and irreducible data noise in CO_2 seismic monitoring.

2. Dataset

The Kimberlina-CO_2 seismic dataset is based on numerical simulations of the 2D acoustic wave equation [27,33,34,89–91]. The wave equation is expressed as:

$$\nabla^2\Phi(x,z,t) - \frac{1}{v^2}\frac{\partial^2\Phi(x,z,t)}{\partial t^2} = S(x,z,t), \quad (1)$$

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}$, is the Laplacian, v is the velocity model, $\Phi(x,z,t)$ is the wavefield, and $S(x,z,t)$ represents the source term. Since a controlled-source method is employed, the source term S is known a priori, making the direct modeling problem involve computing the wavefield Φ given the velocity model v.

The Kimberlina-CO_2, dataset, available in OpenFWI, comprises 19,430 pairs of seismic records and corresponding reference velocity models [89,92]. These models were generated to simulate geological CO_2 storage scenarios in the subsurface. The 2D velocity models feature low-velocity anomalies associated with CO_2, embedded within a stratified background. Velocities range from 947 m/s to 2546 m/s. Each model is defined on a grid of 141×401 nodes, with a spatial resolution $\Delta x = \Delta y = 10$ m, covering a horizontal extent of 4 km and a depth of 1.4 km.

Seismic acquisition modelling involves nine shot records distributed along the surface at 400 m intervals. Each shot gathers data from 101 receivers spaced every 40 m. The seismic line extends over 3.6 km. The data sampling rate is $\Delta t = 2$ ms, with a record length of 2.5 s, resulting in 1251 samples per trace. Figure 1a shows a random model from the dataset, displaying the subsurface structure used to generate the synthetic data. Figure 1b shows the corresponding shot records, illustrating the seismic traces recorded at the nine source positions.

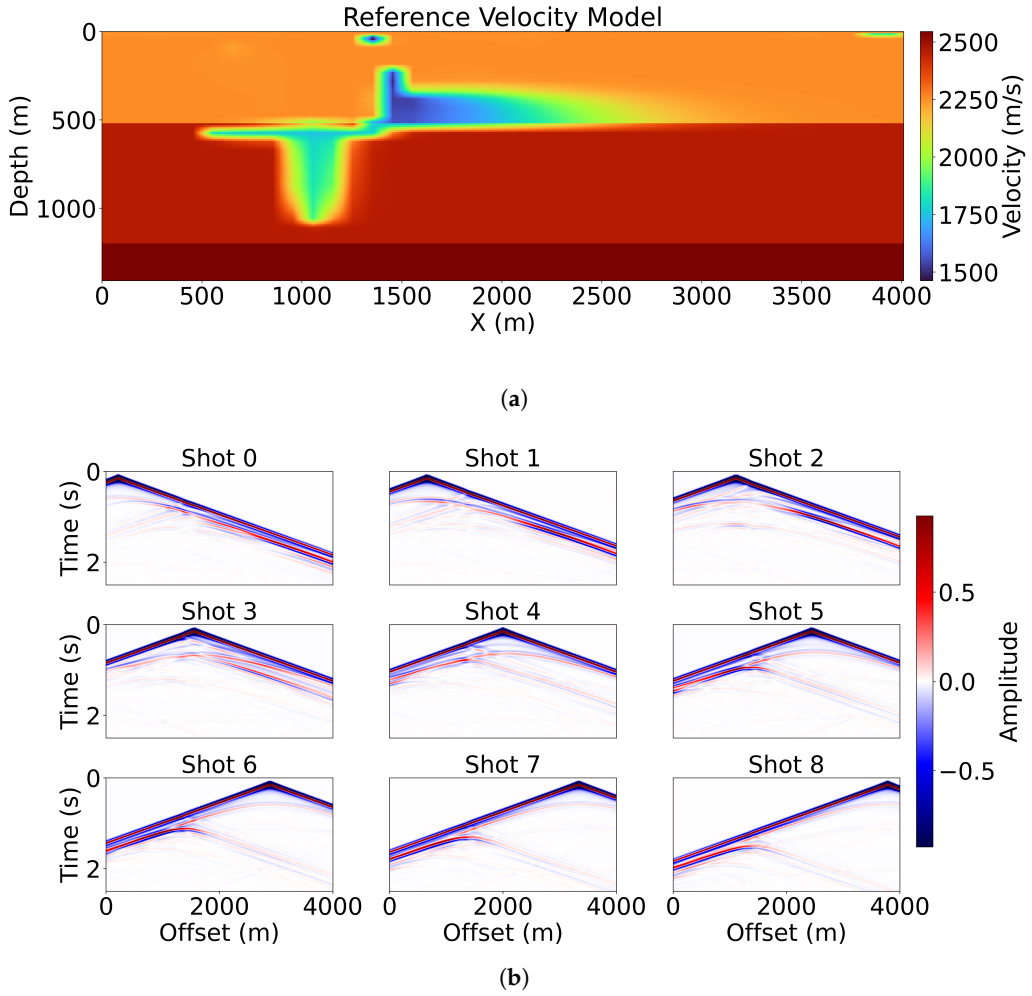


Figure 1. Representative example of the Kimberlina-CO_2 dataset, randomly selected. (**a**) Reference velocity model, displaying low-velocity anomalies associated with CO_2 presence, visible as a transition zone between approximately $x \approx 1500$ m/s and $x \approx 2000$ m/s. (**b**) Shot records corresponding to 9 shots distributed across the surface, each consisting of 101 traces spaced every 40 m, with a recording duration of 2.5 s.

Prior to training, velocity values were normalized using the minimum and maximum of the training dataset via min–max scaling:

$$v_{\text{norm}} = \frac{v - v_{\text{min}}}{v_{\text{max}} - v_{\text{min}}}, \tag{2}$$

where v_{min} and v_{max} correspond to the minimum and maximum velocity values observed in the training set. As a result, v_{norm} is a dimensionless quantity in the range $[0, 1]$.

3. Methodology

Quantifying uncertainty is central to geophysical modeling, particularly in seismic inversion, where the indirect nature of observations allows multiple subsurface configurations to produce similar seismic responses [61,84]. This ambiguity, characteristic of ill-posed inverse problems, calls for statistical frameworks that can characterize and propagate uncertainty throughout the inversion process [16,17,86,93,94]. Within this framework, distinguishing epistemic uncertainty arising from incomplete knowledge of the system and reducible through methodological refinement or additional data [18,84,95–97]. From aleatoric uncertainty, arising from inherent variability in the physical system and measurement process, and largely irreducible [17,84,87,98], is essential, since it determines

which sources of uncertainty can be actively minimized and which must be accepted as fundamental limits on subsurface predictions. Capturing both uncertainties in a neural network requires treating its weights as random variables rather than fixed values, which we formalize next through a Bayesian neural network (BNNs) formulation. Along with the introduction of BNNs, we also discuss the Flipout approach that will be used in our Bayesian inversion network, on the strength of the advantages outlined in detail in the following section. As we will show, it removes the gradient-variance plateau that otherwise limits BNN training under large mini-batches, it does so at a modest and parallelizable compute overhead, and it applies uniformly across the convolutional and dense layers that compose our architecture, leaving the aleatoric output head and NLL function unaffected [99]. report that this variance reduction translates into more reliable convergence of the variational posterior and, in turn, into better-behaved uncertainty estimates than those obtained with naively shared weight perturbations, a property we likewise observe in our experiments, where the resulting epistemic uncertainty maps show improved calibration against held-out seismic responses compared to a shared-perturbation baseline. This benefit is particularly pronounced for our dataset because seismic gathers are naturally organized into large batches (whole shots or survey patches), so training without Flipout would leave the posterior variance poorly estimated precisely in the large-batch regime our data require, while Flipout near-ideal $1/N$ variance scaling lets us train at these realistic batch sizes without trading away the fidelity of the recovered epistemic uncertainty.

3.1. Bayesian Neural Networks

To estimate the epistemic component of uncertainty in the inversion network, we adopt a Bayesian neural network (BNN) formulation in which the network weights are treated as random variables rather than fixed point estimates [100–102]. Let $f(x, W)$ denote the output of a network with weights W given input x (e.g., the seismic trace), and let $q_\theta(W)$ denote a variational distribution over the weights, parameterized by θ. Following [103,104], training proceeds by maximizing the evidence lower bound (ELBO)

$$\mathcal{F}(\theta) = \mathbb{E}_{W\sim q_\theta}\big[\log p(\mathcal{D} \mid W)\big] - D_{\mathrm{KL}}\big(q_\theta \,\|\, p(W)\big), \tag{3}$$

where $p(W)$ is a prior over the weights and $\mathcal{D}$ denotes the observed (seismic) data. The first term rewards data fit, while the KL term regularizes q_θ toward the prior. It is common to use as prior an independent, zero-mean Gaussian over every weight in the network

$$p(W_{ij}) = \mathcal{N}(0, I). \tag{4}$$

This is the standard weakly informative choice for variational BNNs [104], it does not favor any particular sign or direction for the weights, and it penalizes unnecessary model complexity in the absence of supporting data precisely the role played by the KL term in the ELBO. On the other hand, we approximate the true posterior $p(W \mid \mathcal{D})$ with a fully factorized Gaussian,

$$q_\theta(W_{ij}) = \mathcal{N}(\mu_{ij}, \sigma_{ij}^2), \tag{5}$$

so that $\theta = \{\mu_{ij}, \sigma_{ij}\}_{i,j}$ are the variational parameters learned during training (in the notation used below, $\mu_{ij} \equiv \overline{W}_{ij}$). To guarantee $\sigma_{ij} > 0$ while remaining differentiable, we parameterize it through a softplus transform of an unconstrained variable ρ_{ij}

$$\sigma_{ij} = \log\big(1 + \exp(\rho_{ij})\big), \tag{6}$$

and optimize ρ_{ij} directly via backpropagation through the a fully factorized Gaussian posterior, under which each weight is sampled via the reparameterization trick as

$$W_{ij} = \overline{W}_{ij} + \sigma_{ij}\,\epsilon_{ij}, \qquad \epsilon_{ij} \sim \mathcal{N}(0,1), \tag{7}$$

which scales the perturbation with the magnitude of the mean weight $\overline{W}_{ij}$ useful here because the dynamic range of weights across layers of a seismic inversion network (e.g., velocity fields) can vary substantially. In both cases, $\overline{W}$ represents the posterior mean weights and σ_{ij}^2 the posterior variance, which directly encodes epistemic uncertainty: it shrinks as the network is exposed to more informative data and is, in principle, reducible.

3.1.1. The Mini-Batch Variance Problem and Flipout Strategy

Sampling network weights stochastically at every forward pass is what allows the BNN to propagate epistemic uncertainty through to the output (e.g., posterior distributions over subsurface properties). However, naively implementing this is computationally prohibitive: a separate weight sample for every example in a mini-batch requires storing and applying a distinct perturbed weight tensor per example, far more expensive than the single shared weight matrix used in standard deterministic training. In practice, this forces most implementations to draw a single perturbation ΔW and share it across all N examples in a mini-batch. This sharing induces correlations among the per-example gradients, so the variance of the mini-batch gradient estimator does not decay as $1/N$. For uncertainty-quantification workflows on geophysical data, where mini-batches are often necessarily large (2D survey patches), this is a serious bottleneck: training converges slowly, and the resulting posterior over weights can be poorly estimated, degrading the quality of the epistemic uncertainty maps the inversion is meant to deliver. Flipout [99,105] resolves this without sacrificing the efficiency of shared-weight computation. It relies on the observation that, for any perturbation distribution that factorizes across weights and is symmetric about zero, which the Gaussian posteriors used in our BNN satisfy, the distribution is invariant under elementwise multiplication by a random sign matrix. Exploiting this, Flipout draws a single base perturbation $\widehat{\Delta W} \sim q_\theta$ shared across the batch, but assigns each example n its own pseudo-independent perturbation via a rank-one sign matrix

$$\Delta W_n = \Delta\widehat{W} \odot r_n s_n^\top, \tag{8}$$

for a convolutional layer, r_n and s_n act on the input and output channel dimensions, respectively, while their resulting sign pattern is broadcast over the spatial dimensions of the convolutional kernel. For a 2D convolutional layer, let $X \in \mathbb{R}^{N\times H\times W\times C_{\text{in}}}$ denote the input activations, where N is the mini-batch size, H and W are the spatial dimensions, and C_{in} is the number of input channels. The posterior mean kernel $\bar{W}$ and the shared base perturbation $\Delta\widehat{W}$ have shape $\bar{W}, \Delta\widehat{W} \in \mathbb{R}^{k_h\times k_w\times C_{\text{in}}\times C_{\text{out}}}$, where $k_h \times k_w$ is the convolutional kernel size and C_{out} is the number of output channels. Here, $\bar{W}$ denotes the posterior mean of the convolutional weights, whereas $Delta\widehat{W}$ denotes the zero-mean base perturbation sampled from the variational posterior. For each example n, Flipout draws independent Rademacher sign vectors $r_n \in \{-1,+1\}^{C_{\text{in}}}$ and $s_n \in \{-1,+1\}^{C_{\text{out}}}$. Let R and S stack the per-example sign vectors r_n and s_n row-wise, respectively, $R \in \{-1,+1\}^{N\times C_{\text{in}}}$, and $S \in \{-1,+1\}^{N\times C_{\text{out}}}$. For the convolutional operation, these tensors are reshaped as

$$R \to \mathbb{R}^{N\times 1\times 1\times C_{\text{in}}}, \qquad S \to \mathbb{R}^{N\times 1\times 1\times C_{\text{out}}},$$

so that they are broadcast over the spatial dimensions. The Flipout forward pass is therefore

$$Y = \phi\Big[\mathrm{Conv2D}(X, \bar{W}) + \mathrm{Conv2D}(X \odot R, \Delta\widehat{W}) \odot S\Big], \tag{9}$$

where ϕ denotes the activation function and $Y \in \mathbb{R}^{N\times H'\times W'\times C_{\text{out}}}$. Thus, R is broadcast over the input spatial dimensions (H, W), whereas S is broadcast over the output spatial dimensions (H', W'), ensuring tensor compatibility across convolutional layers with different spatial resolutions and channel dimensions. This requires only one additional matrix multiplication relative to a shared-perturbation forward pass, roughly 2× overhead, parallelizable, and substantially cheaper than full per-example weight sampling. Plugging the resulting Flipout perturbations into the variance decomposition, the authors arrive at

$$\mathrm{Var}(\mathcal{G}_{\mathcal{B}}) = \frac{1}{N}\alpha + \frac{N-1}{N}\gamma, \tag{10}$$

where α is the per-example gradient variance, and γ is the covariance term arising from the shared base perturbation. Finally, we model the aleatoric part directly at the network output by replacing the conventional deterministic output layer with a probabilistic output layer in TensorFlow Probability (https://www.tensorflow.org/probability, accessed on 1 July 2026), which maps the network's final hidden activations to the parameters of a predictive distribution rather than to a single point estimate. In fact, for a target y (e.g., a subsurface property at a given location), the network outputs both a mean and a standard deviation

$$\mu(x, W) = h_\mu(x, W), \qquad \sigma_y(x, W) = \mathrm{softplus}\big(h_\sigma(x, W)\big), \tag{11}$$

and the predictive distribution, conditioned on a given weight sample W, is taken to be Gaussian

$$p(y \mid x, W) = \mathcal{N}\big(y;\, \mu(x, W),\, \sigma_y^2(x, W)\big). \tag{12}$$

The softplus transform on h_σ guarantees a strictly positive variance while allowing it to be learned as a function of the input, so the network can express heteroscedastic noise, i.e., predict larger aleatoric uncertainty in regions of the input space (e.g., noisier or more ambiguous seismic responses) and smaller uncertainty elsewhere. This $\sigma_y(x, W)$ is learned jointly with the rest of the network but remains conceptually distinct from the posterior variance σ_{ij}^2 of the latter reflects uncertainty about the weights, while $\sigma_y^2(x, W)$ reflects irreducible noise in y given a fixed, fully specified model.

3.1.2. Negative Log-Likelihood as the ELBO Loss

With the predictive distribution $p(y \mid x, W)$ defined above, the data-fit term of the ELBO is evaluated by drawing a weight sample W_n via Flipout for each example n in the mini-batch and computing the log-likelihood of the observed target under the resulting Gaussian predictive distribution. For a single training example (x, y), this is [106]:

$$\log p(y \mid x, W) = -\frac{1}{2}\log\left(2\pi\sigma_y^2(x, W)\right) - \frac{\big(y - \mu(x, W)\big)^2}{2\sigma_y^2(x, W)}. \tag{13}$$

Minimizing the negative of this quantity defines the negative log-likelihood (NLL) loss used during training,

$$\mathcal{L}_{\text{NLL}}(x, y, W) = \frac{1}{2}\log\left(2\pi\sigma_y^2(x, W)\right) + \frac{\big(y - \mu(x, W)\big)^2}{2\sigma_y^2(x, W)}, \tag{14}$$

averaged over a mini-batch $\mathcal{B}$ of size N to approximate the expectation over $\mathcal{D}$. This loss automatically penalizes overconfident predictions: if $\sigma_y^2(x, W)$ is small but the squared residual is large, the second term grows sharply, while shrinking $\sigma_y^2(x, W)$ too aggressively is discouraged by the logarithmic term, so the network is incentivized to learn an aleatoric variance that genuinely reflects the spread of plausible targets, rather than collapsing to an arbitrarily narrow or wide interval. Combining this NLL with the closed-form KL term, the full training objective minimized over θ is the negative ELBO [107].

$$\widehat{\mathcal{L}}(\theta) = \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}_{\mathrm{NLL}}(x_n, y_n, W_n) + \frac{1}{|\mathcal{D}|} D_{\mathrm{KL}}\big(q_\theta \,\|\, p\big), \tag{15}$$

where the KL term is rescaled by $1/|\mathcal{D}$ (or, equivalently, by the number of mini-batches per epoch) so that its relative weight does not dominate the data-fit term as the dataset size grows. In this formulation, the two predictive sources of uncertainty are cleanly separated by construction: epistemic uncertainty is obtained by drawing multiple weight samples $W^{(k)} \sim q_\theta$ at inference time and examining the spread of the resulting means $\mu(x, W^{(k)})$, while aleatoric uncertainty is read directly from the learned $\sigma_y(x, W^{(k)})$ for each draw together yielding a full predictive distribution that decomposes total uncertainty into its reducible and irreducible components.

3.1.3. Calibration of Uncertainty Estimates

Beyond decomposing the predictive uncertainty into its epistemic and aleatoric components, it is essential that the resulting distributions be calibrated: a model is calibrated when its predicted confidence levels match the empirically observed frequencies, so that, for example, 90% credible intervals contain the true value in 90% of cases. Formally, denoting by F_x the predictive cumulative distribution function for input x, calibration requires

$$\mathbb{P}\Big(y \le F_x^{-1}(p)\Big) = p, \qquad \forall\, p \in [0,1], \tag{16}$$

a condition that can be assessed empirically through reliability diagrams and summarised by the miscalibration area, i.e., the area enclosed between the empirical calibration curve and the ideal diagonal.

Although the Flipout estimator yields low-variance, in our application the statistical properties of the data vary strongly across the spatial domain such as differences in reflectivity contrasts, uneven source–receiver artifacts, and reduced seismic sensitivity at depth cause the network to be systematically overconfident in some regions and underconfident in others. Calibration quality is therefore spatially heterogeneous, and a single global correction would be insufficient.

To address this, we apply a post-hoc recalibration procedure based on isotonic regression, following Kuleshov et al. [108]. Using a held-out validation set, the empirically observed proportion of true values falling below each predicted quantile is computed, and an auxiliary isotonic (monotone, non-parametric) regressor $R : [0,1] \to [0,1]$ is fitted to map the predicted cumulative probabilities onto their observed frequencies. The recalibrated predictive distribution is then obtained as $R \circ F_x$. Because isotonic regression is monotone, it preserves the ordering of the predicted quantiles and the spatial structure of the uncertainty maps, adjusting only the widths of the credible intervals. Given the spatial variability of the miscalibration described above, this procedure was applied independently to each pixel of the output grid, using the uncertainty-toolbox implementation.

3.2. Network Architectures

Three convolutional neural network architectures were selected for comparison: the Feature Pyramid Network (FPN) [109], LinkNet [110], and UNet [78], illustrated in Figure 2a–c respectively.

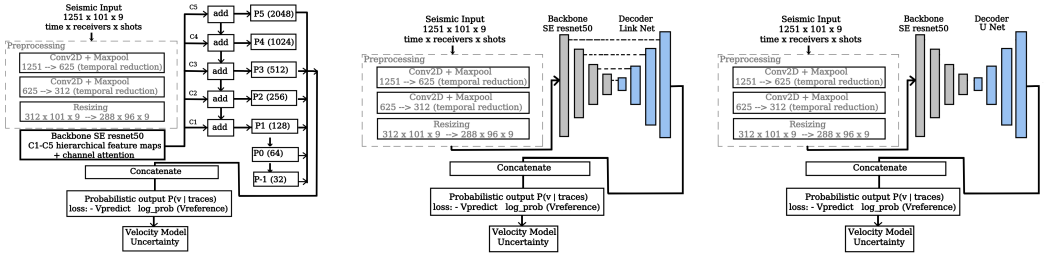


Figure 2. Proposed network architectures for probabilistic full-waveform seismic inversion. The input consists of seismic records from nine shots (101 receivers and 1251 time samples), which are reduced via two stages of convolution and max pooling before being resized to 288 × 96. An SE-ResNet50 backbone extracts hierarchical feature maps with per-channel attention, which are fused by a decoder and concatenated to produce a probabilistic output $P(v \mid \text{traces})$, from which the predicted velocity model and its associated uncertainty are obtained. (**a**) FPN decoder, combining backbone feature maps ($C1$–$C5$) via lateral connections and a top–down path with additive fusion at each pyramid level (P_{-1}–$P5$). (**b**) LinkNet decoder, which fuses encoder and decoder features through direct skip connections at each resolution level. (**c**) U-Net decoder, which fuses encoder and decoder features through concatenation-based skip connections at each resolution level.

The three architectures used SE-ResNet50 [111] for feature extraction and were trained and evaluated under the same conditions. Based on this analysis, the FPN-based architecture was selected.

Figure 2a shows the first architecture, in which the FPN decoder fuses the backbone feature maps through lateral connections and a top–down pathway. This produces seven feature levels ($P5$–P_{-1}) with filters ranging from 2048 to 32 channels. These levels are combined additively and concatenated into a unified tensor. The second architecture (Figure 2b) uses a LinkNet decoder that fuses encoder and decoder features at each resolution level via skip connections and element-wise addition. This approach keeps the number of channels constant, reduces computational cost, and preserves spatial detail. The third architecture (Figure 2c) uses a U-Net decoder to combine features from the encoder and decoder at each resolution level via skip connections based on concatenation rather than element-wise addition. While this increases the number of channels, parameters and computational cost, it retains more complete information from the encoder. In three architectures, this process yields a probability distribution, denoted $p(v|d)$, which is optimised using the negative log-likelihood (NLL) loss function. The estimated velocity model $\mu(x, z)$ and its associated uncertainty are derived from this distribution. All models were trained on an NVIDIA GeForce RTX 3070 Ti GPU with an adaptive learning rate scheduler initialised at 0.01, and early stopping was applied to prevent overfitting and determine the optimal number of training epochs. The total training time ranged from approximately 8 to 10 h per model.

The probabilistic output layer was built using TensorFlow Probability [112] for all three architectures (https://www.tensorflow.org/probability, accessed on 1 July 2026). To ensure that the confidence intervals reflect the true observed frequencies as accurately as possible, the predicted uncertainty was calibrated and subsequently recalibrated using the uncertainty-toolbox library (https://github.com/uncertainty-toolbox/uncertainty-toolbox/blob/main/examples/viz_recalibrate.py, accessed on 1 July 2026).

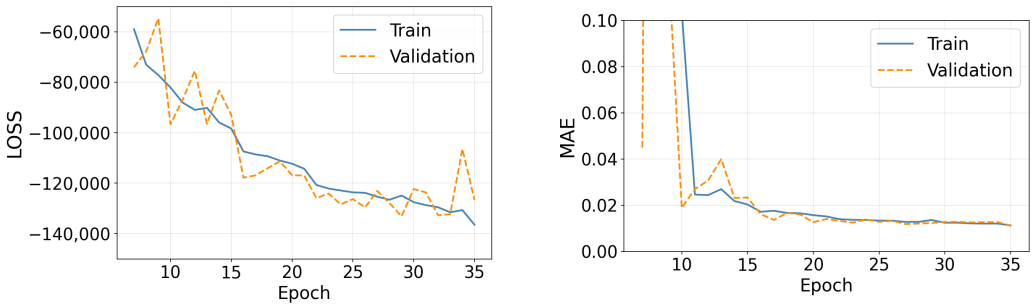
4. Results

A comparison of the three architectures on the test set is summarised in Table 1. The FPN-based network achieved the highest coefficient of determination and the lowest mean squared error, outperforming the LinkNet and U-Net variants. Although the mean absolute error (MAE) of LinkNet was marginally lower than that of FPN, FPN exhibited more consistent performance across metrics, particularly in terms of explained variance and squared error. Based on these results, the FPN architecture was selected as the best-performing model and adopted for the rest of the study.

Table 1. Evaluation metrics on the test set for the three explored architectures (FPN, LinkNet, and UNet), including the coefficient of determination (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE). MAE and RMSE values are normalized.

Métric	FPN	LinkNet	UNet
R^2	0.9386	0.9242	0.911
MAE	4.5249×10^{-3}	4.2469×10^{-3}	6.7375×10^{-3}
MSE	3.1676×10^{-4}	3.9087×10^{-4}	3.7363×10^{-4}

As shown in Figure 3, the training NLL and MAE curves decline sharply during the initial epochs and then stabilise at low levels. The validation curve also declines initially but fluctuates throughout training, with a noticeable spike at epoch 12. From approximately epoch 13 onwards, the training and validation curves converge and remain close at low values, with no strong evidence of sustained overfitting.



(**a**) NLL (**b**) Mean Absolute Error - MAE

Figure 3. Evolution of the NLL and Mean Squared Error (MAE) for training and validation sets across epochs during model training.

The proposed FPN-based model was benchmarked against two established deterministic approaches, VelocityGAN and InversionNet [89], on the Kimberlina-CO_2 dataset (see Table 2). The model achieved a mean absolute error (MAE = 0.0045) and root mean squared error (RMSE = 0.0176) lower than those obtained by both benchmark models.

Table 2. Benchmark comparison on the Kimberlina-CO_2 dataset between VelocityGAN, Inversion Net and the proposed FPN-based model. MAE and RMSE values are normalized and expressed in km/s.

Model	MAE	RMSE
VelocityGAN (ℓ_1) [89]	0.0122	0.0574
VelocityGAN (ℓ_2) [89]	0.0119	0.0387
Proposed model (FPN)	0.0045	0.0176
Inversion Net (ℓ_1) [89]	0.0061	0.0374
Inversion Net (ℓ_2) [89]	0.0098	0.04

The framework attains this level of accuracy while also providing calibrated, pixel-wise epistemic and aleatoric uncertainty estimates. This offers a measure of confidence that is unavailable in VelocityGAN or InversionNet. These results suggest that the FPN-based Bayesian formulation is a favourable candidate for CO_2-monitoring applications requiring both estimation accuracy and uncertainty quantification.

Probabilistic models often produce poorly calibrated uncertainty estimates, meaning that the predicted confidence intervals do not accurately reflect the true frequency of observed outcomes. Recalibration addresses this issue by adjusting the predicted distribution to better match empirical coverage, thereby making the resulting uncertainty estimates statistically meaningful. As the FPN model provides a probabilistic output for each pixel in the velocity model, the recalibration procedure was applied to each pixel independently.

Figure 4 illustrates the result of this procedure for a randomly selected pixel. Before recalibration, the proportion of true values falling within the predicted intervals is substantially higher than the diagonal for most of the range, indicating overconfidence in the model at this pixel: the predicted uncertainty intervals are narrower than required for nominal coverage.

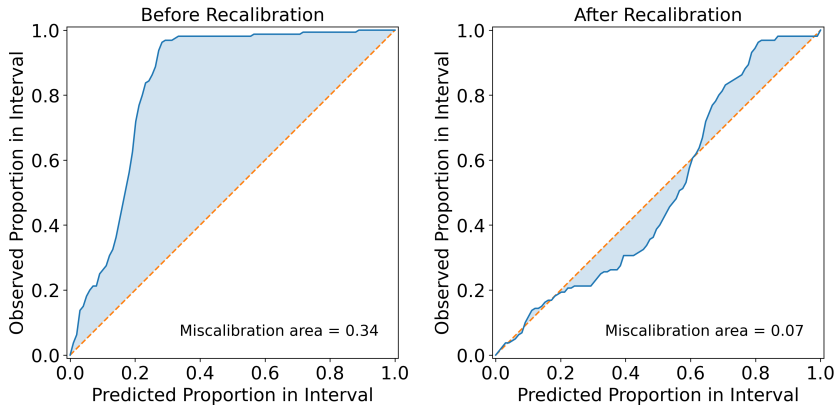


Figure 4. Calibration curves for the FPN model before (**left**) and after (**right**) recalibration, showing the observed proportion of true values falling within each predicted confidence interval as a function of the predicted proportion. The dashed diagonal line represents perfect calibration. The miscalibration area, computed as the area between the calibration curve and the diagonal, decreases from 0.34 before recalibration to 0.07 after recalibration.

This is reflected in a miscalibration area of 0.34. After recalibration, however, the calibration curve lies much closer to the diagonal across the entire range of predicted proportions and the miscalibration area is reduced to 0.07, representing an approximately fivefold improvement. This behaviour was consistent across most pixels in the output grid, indicating that the recalibration procedure substantially improves the reliability of uncertainty estimates throughout the estimated velocity model. This allows the predicted confidence intervals to be interpreted as meaningful probabilistic bounds.

Figure 5 shows the spatial distribution of the miscalibration area across the predicted velocity model, comparing calibration and recalibration results. In the calibration, the miscalibration area is heterogeneous, with high values exceeding 0.4 concentrated in the shallow part of the model ($y < 1000$ m) and in the left-hand portion of the domain ($x < 1000$ m). A distinct horizontal band of high miscalibration is also visible at $y = 500$ m. This pattern indicates that the uncertainty estimates produced by the FPN model are considerably less reliable in these regions, likely corresponding to areas of greater structural complexity or reduced seismic sensitivity at depth.

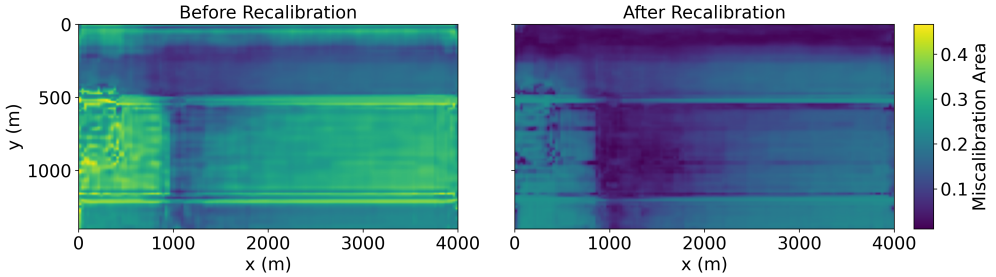


Figure 5. The miscalibration area is computed across the predicted velocity model before (**left**) and after (**right**) applying pixel-wise recalibration. The value at each location represents the area enclosed between the local calibration curve and the diagonal corresponding to perfect calibration. Darker regions, which are associated with lower values, indicate a more accurate estimate of uncertainty.

Recalibration substantially reduces the miscalibration area, which becomes markedly more homogeneous throughout the domain, with most values falling below 0.2. Although a region of slightly higher miscalibration remains near the shallow, left-hand side of the model, its magnitude is considerably reduced relative to the uncalibrated case. These results demonstrate that the pixel-wise recalibration procedure improves the overall reliability of the uncertainty estimates and reduces the spatial heterogeneity in calibration quality across the estimated velocity model, as previously illustrated for a single pixel.

Figure 6 presents the results of the probabilistic inversion for for two representative velocity models, Model 1 and Model 2 (top and bottom, respectively), randomly selected from the held-out test set used to evaluate the trained network. In both cases, panel (a) shows the reference velocity model, panel (b) the predicted velocity model, panels (c,d) the calibrated and recalibrated aleatoric uncertainty, and panels (e,f) the calibrated and recalibrated epistemic uncertainty.

The two reference models (see panels 1a y 2a) differ in terms of the geometry of the low-velocity anomaly within the layered background. Model 1 features an intrusion that extends from the surface to approximately $y = 1000$ m and is centred around $x = 1000$–1500 m. This intrusion is accompanied by a horizontal discontinuity near $y = 500$ m. Model 2 features a more localised anomaly that is centred around $x = 1000$ m and $y = 1000$–1500 m and extends laterally with decreasing intensity. In both cases, the predicted velocity model (panels 1b and 2b) accurately reproduces the overall shape and position of the anomaly; however, its boundaries and lateral extent are depicted with a smoother transition than in the reference model.

The uncertainty estimates follow a consistent pattern across both models. Aleatoric uncertainty (panels c,d) is concentrated around the main anomaly in each case, which is consistent with these regions being the dominant source of variability in the seismic response. By contrast, epistemic uncertainty (panels e,f) is more sharply localised: in Model 1, it is concentrated at the top of the intrusion near $x = 1000$ m and $y = 500$ m, where the anomaly's boundary is least constrained by the surrounding structure; in Model 2, it is confined to a compact region centred on the anomaly. In both cases, its magnitude remains comparatively low elsewhere in the domain, indicating that the network has effectively learnt the surrounding structure. In both models, recalibration (panels d,f) preserves the spatial pattern of both uncertainty components, redistributing their magnitude without altering their overall structure.

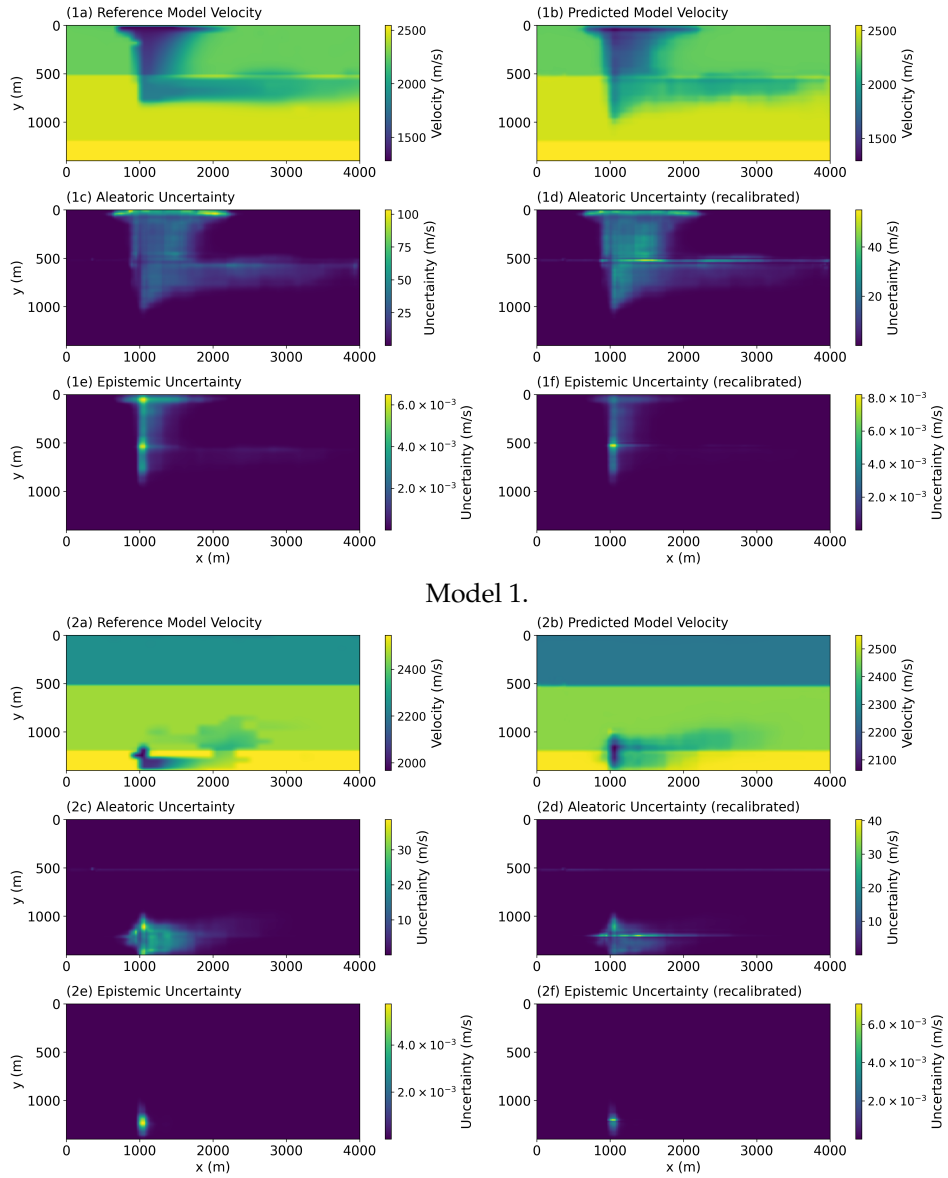


Figure 6. Bayesian seismic inversion results for Model 1 and Model 2: (**1a**, **2a**) reference velocity; (**1b**, **2b**) predicted velocity; (**1c**, **2c**) aleatoric uncertainty; (**1d**, **2d**) recalibrated aleatoric uncertainty; (**1e**, **2e**) epistemic uncertainty; (**1f**, **2f**) recalibrated epistemic uncertainty.

In regions where the reference and predicted models differ, discrepancies may be due to limitations in the quality of the available data. For example, low-velocity anomalies may be less pronounced in relation to their surroundings, and insufficient seismic data resolution or quality may prevent the model from fully capturing their lateral extent. This issue is exacerbated by the effect of recalibration. As previously mentioned, recalibration only corrects the model's overall overconfidence or underconfidence; it does not provide any new information. Consequently, when epistemic and aleatoric uncertainty are rescaled, low-velocity zones with a stronger contrast relative to the layered background become more apparent, whereas those with a less pronounced contrast remain comparatively smoothed out. This combination of factors explains why this extension of the anomaly is not clearly reflected in the epistemic and aleatoric uncertainty maps.

5. Discussion

The main objective of this study was to develop a bayesian deep learning framework that achieves accurate inversion while simultaneously estimating, from the output of the FPN architecture, the irreducible noise arising from the seismic response of the CO_2 plume (aleatoric uncertainty) and the reducible uncertainty associated with incomplete model knowledge (epistemic uncertainty). This is relevant given that most seismic inversion studies typically report total uncertainty without distinguishing between its two components (see, for example, the following references: [80,81,83,85]).

Among the three architectures, FPN was selected as the basis for the proposed Bayesian framework. Although LinkNet achieved a marginally lower MAE (see Table 1), FPN showed more consistent performance across metrics, particularly in terms of explained variance and squared error. This is probably due to the decoder's multi-scale, top–down fusion of feature maps, which combines fine spatial detail with broader contextual information—a property that is particularly important for identifying the low-velocity, low-contrast CO_2 plume structures within a stratified background-. The proposed framework outperforms the deterministic benchmarks such as VelocityGAN and InversionNet [89] in terms of MAE and RMSE (see Table 2), while also providing calibrated aleatoric and epistemic uncertainty estimates, which the aforementioned baselines do not offer.

In our work, we adopt Flipout, which is a variational inference (VI) method, and we deliberately chose this path over MCMC and deep ensembles for the following practical reasons. MCMC is prohibitively expensive for high-dimensional inversion problems. As discussed in the literature [105], MCMC samplers face severe challenges in scaling to high dimensions, requiring an enormous number of likelihood evaluations to achieve adequate exploration and convergence. More concretely, a previous study [113] reported that for a problem with approximately 11,000 parameters, their MCMC algorithm required 38 h on a 32-core cluster for 10,000 iterations. In contrast, our FPN architecture contains 80 million parameters over 7000 times larger, making standard MCMC sampling completely infeasible in practice. As explicitly stated in [113]: "MCMC sampling of high-dimensional problems still requires a massive amount of computing time and resources." Moreover, the authors of [105] highlight that adaptive MCMC methods, while improving convergence, still struggle with the computational cost as dimensionality grows, and they advocate for hybrid approaches such as using Bayesian neural networks to accelerate proposals precisely because of these limitations. On the other hand, Deep ensembles, while effective for uncertainty estimation, require training multiple independent neural networks from different random initializations. In our case, the FPN architecture already contains approximately 80 million parameters. Maintaining an ensemble of, say, 20 such models would demand roughly 20X the GPU memory and training time, which is impractical in most realistic hardware setup and would severely limit the applicability of our method in field scenarios. Therefore, Variational inference with Flipout offers a favorable middle ground since it provides a principled Bayesian approximation to the posterior over network weights, capturing epistemic uncertainty through a learned distribution, while requiring only a modest overhead over a single deterministic training run.

Pixel-wise calibration, following the isotonic regression approach of Kuleshov et al. [108], substantially improved the statistical reliability of the reported confidence intervals. This confirms that raw variational posteriors are often overconfident and require post-hoc correction to yield statistically meaningful uncertainty estimates [108]. Recalibration preserved the spatial structure of both uncertainty components while adjusting only their magnitude, indicating that the network's learned spatial pattern is informative and only the absolute calibration of the intervals needs correction.

The results suggest that the Flipout estimator [99] was effective, producing spatially coherent uncertainty maps for both test scenarios (see Figure 6). In these maps, epistemic uncertainty is concentrated within the anomaly, specifically in the region of lowest velocity. In contrast, aleatoric uncertainty is distributed more broadly across the shallow, low-velocity layer associated with the anomaly.

6. Conclusions

This framework validates the efficacy of probabilistic seismic inversion using deep neural networks to detect and localise subsurface CO_2 plume anomalies. By enabling the simultaneous estimation of velocity and the decoupling of aleatoric and epistemic uncertainties, this approach provides critical insights for monitoring tasks. The framework achieves an accurate reconstruction of both the subsurface stratigraphy and the CO_2 plume geometry, with results that closely match the reference model; its ability to resolve these minor, low-velocity structures within stratified media demonstrates the robustness of the approach for the operational monitoring of geological carbon storage. Aleatoric (random) uncertainty is concentrated in the region occupied by the CO_2 plume and closely follows its shape, suggesting that this component is linked to the anomaly itself, and calibration and recalibration do not alter this spatial pattern but rather adjust the magnitude of the estimates so that they better reflect the true prediction errors. Epistemic uncertainty, in contrast, does not consistently reproduce the shape of the anomaly; instead, it is concentrated within the region of lowest velocity inside the anomaly without extending throughout the domain, while the flat, layered background exhibits consistently low values, indicating that the model has learned this structure effectively. This decomposition allows us to determine whether limitations in plume characterisation stem from data noise or from deficiencies in the inversion model, and calibration ensures that the reported uncertainty magnitudes are statistically reliable; distinguishing between these two factors and providing calibrated estimates of each is essential for operational decision-making and for developing risk mitigation strategies in CO_2 storage projects.

Author Contributions: Conceptualization, A.D.L. and L.A.M.V.; methodology, A.D.L., H.J.H.O. and L.A.C.L.; software, A.D.L. and L.A.M.V., H.J.H.O. and L.A.C.L.; validation, A.D.L. and L.A.M.V.; formal analysis, A.D.L., L.A.M.V., H.J.H.O. and L.A.C.L.; investigation, A.D.L., L.A.M.V., H.J.H.O. and L.A.C.L.; resources, A.D.L.; data curation, A.D.L.; writing—original draft preparation, A.D.L., L.A.M.V., H.J.H.O. and L.A.C.L.; writing—review and editing, A.D.L., H.J.H.O. and L.A.C.L.; visualization, A.D.L. and L.A.M.V., H.J.H.O. and L.A.C.L.; supervision, A.D.L. and L.A.M.V.; project administration, A.D.L. and L.A.M.V.; funding acquisition, A.D.L. and L.A.M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Institución Universitaria Politécnico Grancolombiano and the Universidad Nacional de Colombia, grant number CCI2024_CITIC_05-87398.

Data Availability Statement: The dataset used in this study is publicly available through the OpenFWI [89] repository and can be accessed at: https://edx.netl.doe.gov/dataset/kimberlina-1-2-velocity-models (accessed on 1 July 2026). The code developed for this study (version 1.0) is publicly available in a GitHub repository and can be accessed at: https://github.com/AlejandroDuitamaLeal/Full-waveform-inversion-via-uncertainty-aware-deep-learning-for-CO2-plume-detection.git (accessed on 1 July 2026).

Acknowledgments: The The authors would like to express their gratitude for the support provided by the Department of Geosciences at the Universidad Nacional de Colombia and the School of Basic Sciences at the Institución Universitaria Politécnico Grancolombiano. Their academic guidance and institutional support were instrumental in developing this work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Das, V.; Mukerji, T. Petrophysical properties prediction from prestack seismic data using convolutional neural networks. *Geophysics* **2020**, *85*, N41–N55. [CrossRef]
2. Zhang, G.; Wang, Z.; Chen, Y. Deep learning for seismic lithology prediction. *Geophys. J. Int.* **2018**, *215*, 1368–1387. [CrossRef]
3. Allo, F.; Coulon, J.P.; Formento, J.L.; Reboul, R.; Capar, L.; Darnet, M.; Issautier, B.; Marc, S.; Stopin, A. Characterization of a carbonate geothermal reservoir using rock-physics-guided deep neural networks. *Lead. Edge* **2021**, *40*, 751–758. [CrossRef]
4. Di, H.; Abubakar, A. Estimating subsurface properties using a semisupervised neural network approach. *Geophysics* **2021**, *87*, IM1–IM10. [CrossRef]
5. Di, H.; Li, Z.; Abubakar, A. Using relative geologic time to constrain convolutional neural network-based seismic interpretation and property estimation. *Geophysics* **2021**, *87*, IM25–IM35. [CrossRef]
6. Gui, J.; Gao, J.; Li, S.; Liu, B.; Guo, X.; Pan, H. Petrophysical Properties Prediction based on a Deep Learning Scheme. In Proceedings of the 85th EAGE Annual Conference & Exhibition, Oslo, Norway, 10 June 2024; pp. 1–5. [CrossRef]
7. Yenwongfai, H.; Mondol, N.H.; Lecomte, I.; Faleide, J.I.; Leutscher, J. Integrating facies-based Bayesian inversion and supervised machine learning for petro-facies characterization in the Snadd Formation of the Goliat Field, south-western Barents Sea. *Geophys. Prospect.* **2019**, *67*, 1020–1039. [CrossRef]
8. Singh, S.; Tsvankin, I.; Zabihi Naeini, E. Facies prediction with Bayesian inference: Application of supervised and semisupervised deep learning. *Interpretation* **2022**, *10*, T279–T290. [CrossRef]
9. Bhattacharya, S.; Mishra, S. Applications of machine learning for facies and fracture prediction using Bayesian Network Theory and Random Forest: Case studies from the Appalachian basin, USA. *J. Pet. Sci. Eng.* **2018**, *170*, 1005–1017. [CrossRef]
10. Grana, D.; Azevedo, L.; Liu, M. A comparison of deep machine learning and Monte Carlo methods for facies classification from seismic data. *Geophysics* **2020**, *85*, WA41–WA52. [CrossRef]
11. Zhao, L.; Zou, C.; Chen, Y.; Shen, W.; Wang, Y.; Chen, H.; Geng, J. Fluid and lithofacies prediction based on integration of well-log data and seismic inversion: A machine-learning approach. *Geophysics* **2021**, *86*, M151–M165. [CrossRef]
12. Meng, J.; Zhao, L.; Xu, M.; Chen, H. Deep learning-based seismic lithofacies prediction in sparse well areas via geology-informed pseudo-well construction and transfer learning. *J. Appl. Geophys.* **2025**, *242*, 105891. [CrossRef]
13. Li, Y.W.; Ali, S.S.S.; Maskeen, A.A. Lithofacies Based Fluid Prediction Using Machine Learning. In Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, United Arab Emirate, 31 October–3 November 2022. [CrossRef]
14. Chen, Y.; Zhao, L.; Pan, J.; Li, C.; Xu, M.; Li, K.; Zhang, F.; Geng, J. Deep carbonate reservoir characterisation using multi-seismic attributes via machine learning with physical constraints. *J. Geophys. Eng.* **2021**, *18*, 761–775. [CrossRef]
15. Zou, C.; Zhao, L.; Hong, F.; Wang, Y.; Chen, Y.; Geng, J. A comparison of machine learning methods to predict porosity in carbonate reservoirs from seismic-derived elastic properties. *Geophysics* **2023**, *88*, B101–B120. [CrossRef]
16. Zhang, Z.; Wang, G.; Carranza, E.J.M.; Liu, C.; Li, J.; Fu, C.; Liu, X.; Chen, C.; Fan, J.; Dong, Y. An integrated machine learning framework with uncertainty quantification for three-dimensional lithological modeling from multi-source geophysical data and drilling data. *Eng. Geol.* **2023**, *324*, 107255. [CrossRef]
17. Feng, R.; Grana, D.; Balling, N. Uncertainty quantification in fault detection using convolutional neural networks. *Geophysics* **2021**, *86*, M41–M48. [CrossRef]
18. Junhwan, C.; Seokmin, O.; Joongmoo, B. Uncertainty estimation in AVO inversion using Bayesian dropout based deep learning. *J. Pet. Sci. Eng.* **2022**, *208*, 109288. [CrossRef]
19. Chopra, S.; Sharma, R.K.; Marfurt, K.J.; Zhang, R.; Wen, R. Improving porosity and gamma-ray prediction for the Middle Jurassic Hugin sandstones in the southern Norwegian North Sea with the application of deep neural networks. *Interpretation* **2021**, *10*, T25–T34. [CrossRef]
20. Iturrarán-Viveros, U.; Muñoz-García, A.M.; Castillo-Reyes, O.; Shukla, K. Machine Learning as a Seismic Prior Velocity Model Building Method for Full-Waveform Inversion: A Case Study from Colombia. *Pure Appl. Geophys.* **2021**, *178*, 423–448. [CrossRef]
21. Brossier, R.; Operto, S.; Virieux, J. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics* **2009**, *74*, WCC105–WCC118. [CrossRef]
22. Zayier, Y.; Yalikun, Y.; Cheng, Y.; Yang, D. Integrating physics and machine learning for unified seismic forward modeling and reservoir property inversion. *Sci. Rep.* **2026**, *16*, 5932. [CrossRef] [PubMed]
23. Zhang, Z.; Alkhalifah, T. High-resolution reservoir characterization using deep learning aided elastic full-waveform inversion: The North Sea field data example. *Geophysics* **2019**, *85*, 1365–1369. [CrossRef]
24. Feng, S.; Lin, Y.; Wohlberg, B. Multiscale Data-Driven Seismic Full-Waveform Inversion With Field Data Study. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4506114. [CrossRef]
25. Fang, J.; Chen, H.; Zhou, H.; Rao, Y.; Sun, P.; Zhang, J. Elastic full-waveform inversion based on GPU accelerated temporal fourth-order finite-difference approximation. *Comput. Geosci.* **2020**, *135*, 104381. [CrossRef]
26. Tillero, E. Machine learning-based modelling for geologic CO_2 storage in deep saline aquifers. Case study of bunter sandstone in Southern North Sea. *Int. J. Greenh. Gas Control* **2024**, *133*, 104077. [CrossRef]

27. Wang, Z.; Harbert, W.P.; Dilmore, R.M.; Huang, L. Modeling of time-lapse seismic monitoring using CO_2 leakage simulations for a model CO_2 storage site with realistic geology: Application in assessment of early leak-detection capabilities. *Int. J. Greenh. Gas Control* **2018**, *76*, 39–52. [CrossRef]
28. Marques, C.R.; dos Santos, V.G.; Lunelli, R.; Roisenberg, M.; Rodrigues, B.B. Analysis of Deep Learning Neural Networks for Seismic Impedance Inversion: A Benchmark Study. *Energies* **2022**, *15*, 7452. [CrossRef]
29. Romero, J.; Heidrich, W.; Luiken, N.; Ravasi, M. Seismic Reservoir Characterization With Implicit Neural Representations. *J. Geophys. Res. Mach. Learn. Comput.* **2024**, *1*, e2024JH000275. [CrossRef]
30. Schuster, G.; Chen, Y.; Feng, S. Review of Physics-Informed Machine Learning Inversion of Geophysical Data. *Geophysics* **2024**, *89*, 1–91. [CrossRef]
31. he, Q.; Wang, Y. Reparameterized full waveform inversion using deep neural networks. *Geophysics* **2020**, *86*, 1–71. [CrossRef]
32. Yang, F.; Ma, J. Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics* **2019**, *84*, R583–R599. [CrossRef]
33. Wang, Z.; Dilmore, R.M.; Harbert, W. Inferring CO_2 saturation from synthetic surface seismic and downhole monitoring data using machine learning for leakage detection at CO_2 sequestration sites. *Int. J. Greenh. Gas Control* **2020**, *100*, 103115. [CrossRef]
34. Buscheck, T.A.; Mansoor, K.; Yang, X.; Wainwright, H.M.; Carroll, S.A. Downhole pressure and chemical monitoring for CO_2 and brine leak detection in aquifers above a CO_2 storage reservoir. *Int. J. Greenh. Gas Control* **2019**, *91*, 102812. [CrossRef]
35. Ali, M.; He, C.; Wei, N.; Jiang, R.; Zhu, P.; Hao, Z.; Hussain, W.; Ashraf, U. Optimizing seismic-based reservoir property prediction: A synthetic data-driven approach using convolutional neural networks and transfer learning with real data integration. *Artif. Intell. Rev.* **2024**, *58*, 31. [CrossRef]
36. Mustafa, A.; Alfarraj, M.; AlRegib, G. Estimation of Acoustic Impedance from Seismic Data using Temporal Convolutional Network. *arXiv* **2019**, arXiv:1906.02684.
37. Zerafa, C.; Galea, P.; Sebu, C. Theory-guided Pseudo-spectral Full Waveform Inversion via Deep Neural Networks. *arXiv* **2025**, arXiv:2502.17624.
38. Yang, Y.; Gao, A.F.; Azizzadenesheli, K.; Clayton, R.W.; Ross, Z.E. Rapid Seismic Waveform Modeling and Inversion With Neural Operators. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–12. [CrossRef]
39. Vantassel, J.P.; Kumar, K.; Cox, B.R. Using convolutional neural networks to develop starting models for near-surface 2-D full waveform inversion. *Geophys. J. Int.* **2022**, *231*, 72–90. [CrossRef]
40. Pelemo-Daniels, D.; Stewart, R.R. Petrophysical Property Prediction from Seismic Inversion Attributes Using Rock Physics and Machine Learning: Volve Field, North Sea. *Appl. Sci.* **2024**, *14*, 1345. [CrossRef]
41. Sircar, A.; Yadav, K.; Rayavarapu, K.; Bist, N.; Oza, H. Application of machine learning and artificial intelligence in oil and gas industry. *Pet. Res.* **2021**, *6*, 379–391. [CrossRef]
42. Romero, J.; Heidrich, W.; Ravasi, M. Bayesian seismic inversion with implicit neural representations. *Geophys. J. Int.* **2025**, *242*, ggaf249. [CrossRef]
43. Sun, J.; Innanen, K.; Zhang, T.; Trad, D. Implicit Seismic Full Waveform Inversion With Deep Neural Representation. *J. Geophys. Res. Solid Earth* **2023**, *128*, e2022JB025964. [CrossRef]
44. Jia, A.; Sun, J.; Bo, D.; Lin, Y. Seismic Full Waveform Inversion With Uncertainty Analysis Using Unsupervised Variational Deep Learning. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1. [CrossRef]
45. Ghyselincks, S.; Okhmak, V.; Zampini, S.; Turkiyyah, G.; Keyes, D.; Haber, E. Synthetic Geology: Structural Geology Meets Deep Learning. *J. Geophys. Res. Mach. Learn. Comput.* **2026**, *3*, e2025JH000986. [CrossRef]
46. Maleki, M.; Akbari, A.; Kazemzadeh, Y.; Ranjbar, A. Machine learning models for the prediction of hydrogen solubility in aqueous systems. *Sci. Rep.* **2025**, *15*, 31487. [CrossRef] [PubMed]
47. Liu, F.; Li, Y.X.; Su, R.; Huang, J.P.; Bai, L. Deep reparameterization for full waveform inversion: Architecture benchmarking, robust inversion, and multiphysics extension. *Pet. Sci.* **2026**, *23*, 1890–1907. [CrossRef]
48. Parasyris, A.; Stankovic, L.; Stankovic, V. Synthetic Data Generation for Deep Learning-Based Inversion for Velocity Model Building. *Remote Sens.* **2023**, *15*, 2901. [CrossRef]
49. Yan, X.B.; Wu, K.; Xu, Z.Q.J.; Ma, Z. An Unsupervised Deep Learning Approach for the Wave Equation Inverse Problem. *arXiv* **2023**, arXiv:2311.04531.
50. Klinge, J.; Schippkus, S.; Walda, J.; Hadziioannou, C.; Gajewski, D. Predictive modelling of seismic wave fields: Learning the transfer function using encoder–decoder networks. *Geophys. J. Int.* **2025**, *240*, 1611–1621. [CrossRef]
51. Mosser, L.; Kimman, W.; Dramsch, J.; Purves, S.; De la Fuente Briceño, A.; Ganssle, G. Rapid Seismic Domain Transfer: Seismic Velocity Inversion and Modeling Using Deep Generative Neural Networks. In Proceedings of the 80th EAGE Conference and Exhibition 2018, EAGE Publications BV, 2018, Copenhagen, Denmark, 11–14 June 2018. [CrossRef]
52. Palgunadi, K.H.; Bergmeister, A.; Bosisio, A.; Ermert, L.A.; Koroni, M.; Perraudin, N.; Dirmeier, S.; Meier, M.A. High Resolution Seismic Waveform Generation Using Denoising Diffusion. *J. Geophys. Res. Mach. Learn. Comput.* **2025**, *2*, e2025JH000862. [CrossRef]

53. Farris, S.; Clapp, R.; Araya-Polo, M. Learning-Based Seismic Velocity Inversion with Synthetic and Field Data. *Sensors* **2023**, *23*, 8277. [CrossRef] [PubMed]
54. Anwar, S.; Kadi, H.H.E.; Hosny, A.; Reda, M.; Mostafa, T. Identifying gas bearing sand using simultaneous pre-stack seismic inversion method: Case study of the Simian field, Offshore Nile Delta, Egypt. *J. Pet. Explor. Prod. Technol.* **2025**, *15*, 72. [CrossRef]
55. Carrasquero, G.; Zambrano, O. Modelo de física de rocas para la interpretación de propiedades petrofísicas de un campo en el Golfo de México. *Rev. Tec. Fac. Ing. Univ. Zulia* **2020**, *43*, 33–40. [CrossRef]
56. Hosseinzadeh, S.; Saberi, M.R.; Haghighi, M.; Salmachi, A.; Salimzadeh, S. Seismic inversion approaches for reservoir characterization: A comprehensive review. *J. Appl. Geophys.* **2025**, *243*, 105953. [CrossRef]
57. Abdel-Fattah, M.I.; Hamdan, H.A.; Mahdi, A.Q.; Abd-Allah, Z.M.; Hammouri, N.A.; Abuzied, S.M. Integrated approach and insights into shale gas potential: Geological, petrophysical, geochemical, and geomechanical evaluation of the Khatatba Formation (Western Desert, Egypt). *Pet. Res.* **2025**, *11*, 415–430. [CrossRef]
58. Wang, Y.R.; Zong, Z.Y.; Yin, X.Y. Fluid discrimination incorporating amplitude variation with angle inversion and squirt flow of the fluid. *Pet. Sci.* **2022**, *19*, 1592–1604. [CrossRef]
59. Geng, Z.; Wang, Y. Automated design of a convolutional neural network with multi-scale filters for cost-efficient seismic data classification. *Nat. Commun.* **2020**, *11*, 3311. [CrossRef] [PubMed]
60. Shehata, A.A.; Osman, O.A.; Nabawy, B.S. Neural network application to petrophysical and lithofacies analysis based on multi-scale data: An integrated study using conventional well log, core and borehole image data. *J. Nat. Gas Sci. Eng.* **2021**, *93*, 104015. [CrossRef]
61. Qu, L.; Araya-Polo, M.; Demanet, L. Uncertainty Quantification in Seismic Inversion Through Integrated Importance Sampling and Ensemble Methods. *arXiv* **2024**, arXiv:2409.06840.
62. Blay-Roger, R.; Bach, W.; Bobadilla, L.F.; Reina, T.R.; Odriozola, J.A.; Amils, R.; Blay, V. Natural hydrogen in the energy transition: Fundamentals, promise, and enigmas. *Renew. Sustain. Energy Rev.* **2024**, *189*, 113888. [CrossRef]
63. Yang, J.; Yu, P.; Wang, S.; Sun, Z. CO_2 Storage Monitoring via Time-Lapse Full Waveform Inversion with Automatic Differentiation. *Nanomaterials* **2024**, *14*, 138. [CrossRef] [PubMed]
64. Luo, X.; Tveit, S.; Gholami, R.; Andersen, P. Underground hydrogen storage (UHS) in natural storage sites: A perspective of subsurface characterization and monitoring. *Fuel* **2024**, *364*, 131038. [CrossRef]
65. Scott, S.W.; O'Sullivan, J.P.; Maclaren, O.J.; Nicholson, R.; Covell, C.; Newson, J.; Gudjónsdóttir, M.S. Bayesian Calibration of a Natural State Geothermal Reservoir Model, Krafla, North Iceland. *Water Resour. Res.* **2022**, *58*, e2021WR031254. [CrossRef]
66. Lin, K.; Wei, N.; Zhang, Y.; Ali, M.; Chen, Q.; Wang, W.; Song, Z.; Yin, Y.; Vo Thanh, H. Advances in Machine-Learning-Driven CO_2 Geological Storage: A Comprehensive Review and Outlook. *Energy Fuels* **2025**, *39*, 13315–13343. [CrossRef]
67. Gahlot, A.P.; Orozco, R.; Yin, Z.; Bruer, G.; Herrmann, F.J. An uncertainty-aware digital shadow for underground multimodal CO_2 storage monitoring. *Geophys. J. Int.* **2025**, *242*, ggaf176. [CrossRef]
68. DeVault, B.; Clochard, V.; Spangler, L.H.; Bowen, D.W. Multicomponent Seismic Data and Joint Inversion. In *Geophysical Monitoring for Geologic Carbon Storage*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2022; Chapter 12, pp. 179–196. [CrossRef]
69. Fachtony, F.A.; Brossier, R.; Boddupalli, B.; Dupuy, B.; Métivier, L.; Romdhane, A. CO_2 monitoring at Sleipner field using reflection oriented full waveform inversion—Part 1: Baseline reconstruction. *Geophys. J. Int.* **2025**, *241*, 1226–1242. [CrossRef]
70. Queißer, M.; Singh, S.C. Full waveform inversion in the time lapse mode applied to CO_2 storage at Sleipner. *Geophys. Prospect.* **2013**, *61*, 537–555. [CrossRef]
71. Martinez, R.; Vinje, V.; Stovas, A.; Mispel, J.; Ringrose, P.; Duffaut, K.; Landrø, M. Diving-wave time-lapse delay for CO_2 thin layer detection. *Geophys. J. Int.* **2024**, *237*, 235–249. [CrossRef]
72. Barnes, C.; Charara, M.; Tsuchiya, T.; Yamada, N. CO_2 Monitoring by Using VSP-FWI—Timelapse Elastic Parameters Estimation. In Proceedings of the EAGE/SEG Research Workshop 2017, Trondheim, Norway, 28–31 August 2017. [CrossRef]
73. Wang, N.; Zhang, D.; Chang, H.; Li, H. Deep learning of subsurface flow via theory-guided neural network. *J. Hydrol.* **2020**, *584*, 124700. [CrossRef]
74. Zhan, W.; Chen, Y.; Liu, Q.; Li, J.; Sacchi, M.; Zhuang, M.; Liu, Q.H. Simultaneous prediction of petrophysical properties and formation layered thickness from acoustic logging data using a modular cascading residual neural network (MCARNN) with physical constraints. *J. Appl. Geophys.* **2024**, *224*, 105362. [CrossRef]
75. Ahmed, N.; Waldemar Weibull, W.; Grana, D.; Bhakta, T. Constrained nonlinear amplitude-variation-with-offset inversion for reservoir dynamic changes estimation from time-lapse seismic data. *Geophysics* **2023**, *89*, R1–R15. [CrossRef]
76. Alumbaugh, D.; Commer, M.; Crandall, D.; Gasperikova, E.; Feng, S.; Harbert, W.; Li, Y.; Lin, Y.; Samarasinghe, S.; Yang, X. Development of a multi-scale synthetic data set for the testing of subsurface CO_2 storage monitoring strategies. In Proceedings of the AGU Fall Meeting Abstracts, New Orleans, Louisiana, 13–17 December 2021; Volume 2021, p. S25A-0212.
77. Mosser, L.; Dubrule, O.; Blunt, M.J. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. *Math. Geosci.* **2020**, *52*, 53–79. [CrossRef]

78. Lin, Y.; Wu, Y. InversionNet: A real-time and accurate full waveform inversion with convolutional neural network. *J. Acoust. Soc. Am.* **2018**, *144*, 1683. [CrossRef]
79. Zhang, F.; Juhlin, C.; Cosma, C.; Tryggvason, A.; Pratt, R.G. Cross-well seismic waveform tomography for monitoring CO_2 injection: a case study from the Ketzin Site, Germany. *Geophys. J. Int.* **2012**, *189*, 629–646. [CrossRef]
80. Hunziker, J.; Laloy, E.; Linde, N. Bayesian full-waveform tomography with application to crosshole ground penetrating radar data. *Geophys. J. Int.* **2019**, *218*, 913–931. [CrossRef]
81. Grana, D.; Liu, M.; Ayani, M. Prediction of CO_2 saturation spatial distribution using geostatistical inversion of time-lapse geophysical data. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 3846–3856. [CrossRef]
82. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the International Conference on Machine Learning*; PMLR: New York, NY, USA, 2016; pp. 1050–1059.
83. Zhu, W.; Xu, K.; Darve, E.; Biondi, B.; Beroza, G.C. Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification. *Geophysics* **2022**, *87*, R93–R109. [CrossRef]
84. Sun, Y.; Williamson, P. Invertible neural networks for uncertainty quantification in refraction tomography. *Lead. Edge* **2024**, *43*, 358–366. [CrossRef]
85. Kaur, H.; Zhong, Z.; Sun, A.; Fomel, S. Time-lapse seismic data inversion for estimating reservoir parameters using deep learning. *Interpretation* **2022**, *10*, T167–T179. [CrossRef]
86. Mosser, L.; Zabihi Naeini, E. A comprehensive study of calibration and uncertainty quantification for Bayesian convolutional neural networks—An application to seismic data. *Geophysics* **2022**, *87*, IM157–IM176. [CrossRef]
87. Qu, L.; Araya-Polo, M.; Demanet, L. Importance Sampling for Deep Learning-driven Seismic Inversion Uncertainty Quantification. In Proceedings of the 86th EAGE Annual Conference & Exhibition, Toulouse, France, 2–5 June 2025; pp. 1–5. [CrossRef]
88. Ahmed, N.; Oikonomou, D.; Cardozo, N.; Weibull, W.; Grana, D.; Alaei, B.; Larsen, E. Enhanced Deep Learning Using a Convolutionally Embedded Transformer Architecture for Reservoir Elastic Property Prediction: Comparison With Physics-Based Seismic Inversion. *Geophys. Prospect.* **2026**, *74*, e70211. [CrossRef]
89. Deng, C.; Feng, S.; Wang, H.; Zhang, X.; Jin, P.; Feng, Y.; Zeng, Q.; Chen, Y.; Lin, Y. OpenFWI: Large-Scale Multi-Structural Benchmark Datasets for Seismic Full Waveform Inversion. *arXiv* **2023**, arXiv:2111.02926.
90. Zhou, Z.; Lin, Y.; Zhang, Z.; Wu, Y.; Wang, Z.; Dilmore, R.; Guthrie, G. A data-driven CO_2 leakage detection using seismic data and spatial–temporal densely connected convolutional neural networks. *Int. J. Greenh. Gas Control* **2019**, *90*, 102790. [CrossRef]
91. Yang, X.; Buscheck, T.A.; Mansoor, K.; Wang, Z.; Gao, K.; Huang, L.; Appriou, D.; Carroll, S.A. Assessment of geophysical monitoring methods for detection of brine and CO_2 leakage in drinking water aquifers. *Int. J. Greenh. Gas Control* **2019**, *90*, 102803. [CrossRef]
92. Harbert, W.; Dilmore, R.; Huang, L.; Lin, Y.; Feng, S.; Wang, Z. Kimberlina 1.2 Velocity Models and Seismic Data. 2021. Available online: https://opennetzero.org/national-energy-technology-laboratory-netl/kimberlina-1-2-velocity-models (accessed on 1 July 2026).
93. Pereira, P.; Bordignon, F.; Azevedo, L.; Nunes, R.; Soares, A. Strategies for integrating uncertainty in iterative geostatistical seismic inversion. *arXiv* **2018**, arXiv:1810.07405.
94. Thurin, J. Uncertainties estimation in Full Waveform Inversion using Ensemble methods. *arXiv* **2020**, arXiv:2409.06840v1.
95. Rahma Putra, M.H.; Hermana, M.; Yogi, I.B.S.; Hossain, T.M.; Abdurrachman, M.F.; Kadir, S.J.A. Reservoir Porosity Assessment and Anomaly Identification from Seismic Attributes Using Gaussian Process Machine Learning. *Earth Sci. Inform.* **2024**, *17*, 1315–1327. [CrossRef]
96. Filardi, M.; Bettini, C.; Schwedersky neto, G. Comparison of Deterministic and Geostatistical Inversion Results from a Petroleum Field. *Anuário Do Inst. De Geociências UFRJ* **2016**, *38*, 145. [CrossRef]
97. Liu, Y.; Zhang, X.; Tsvankin, I.; Lin, Y. Enhanced prediction accuracy with uncertainty quantification in monitoring CO_2 sequestration using convolutional neural networks. *arXiv* **2022**, arXiv:2212.04567.
98. Agudelo Cardona, V.; Barbosa, I.; Vallejo, M.; Bahamón Cortés, N.; Delgado Trejos, E. Incertidumbre epistémica y aleatoria en soft metrología: Una perspectiva desde el aseguramiento de la validez de los resultados. *Ingeniería* **2023**, *28*, e18883. [CrossRef]
99. Wen, Y.; Vicol, P.; Ba, J.; Tran, D.; Grosse, R. Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches. *arXiv* **2018**, arXiv:1803.04386.
100. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2022**, arXiv:1312.6114.
101. Hortúa, H.J.; Malagò, L.; Volpi, R. Constraining the Reionization History using Bayesian Normalizing Flows. *Mach. Learn. Sci. Technol.* **2020**, *1*, 035014. [CrossRef]
102. Hortúa, H.J.; Volpi, R.; Marinelli, D.; Malagò, L. Parameter estimation for the cosmic microwave background with Bayesian neural networks. *Phys. Rev. D* **2020**, *102*, 103509. [CrossRef]
103. Graves, A. Practical Variational Inference for Neural Networks. In *Proceedings of the Advances in Neural Information Processing Systems*; Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2011; Volume 24, pp. 2348–2356.

104. Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; Wierstra, D. Weight Uncertainty in Neural Networks. *arXiv* **2015**, arXiv:1505.05424.
105. Hortua, H.J.; Volpi, R.; Marinelli, D.; Malago, L. Accelerating MCMC algorithms through Bayesian Deep Networks. *arXiv* **2020**, arXiv:2011.14276.
106. Kendall, A.; Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Red Hook, NY, USA, 4–9 December 2017; pp. 5580–5590.
107. Chen, W.; Li, B.; Zhang, R.; Li, Y. Bayesian Computation in Deep Learning. *arXiv* **2025**, arXiv:2502.18300.
108. Kuleshov, V.; Fenner, N.; Ermon, S. Accurate Uncertainties for Deep Learning Using Calibrated Regression. *arXiv* **2018**, arXiv:1807.00263.
109. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. *arXiv* **2017**, arXiv:1612.03144.
110. Li, H.P.; Qi, R.; Hu, J.X.; Sun, Y.X. 3D gravity anomaly inversion based on LinkNet. *Appl. Geophys.* **2023**, *20*, 36–50. [CrossRef]
111. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *arXiv* **2019**, arXiv:1709.01507.
112. Dillon, J.V.; Langmore, I.; Tran, D.; Brevdo, E.; Vasudevan, S.; Moore, D.A.; Patton, B.; Alemi, A.A.; Hoffman, M.D.; Saurous, R.A. TensorFlow Distributions. *arXiv* **2017**, arXiv:1711.10604.
113. Zhang, R.; Czado, C.; Sigloch, K. Bayesian spatial modelling for high dimensional seismic inverse problems. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2016**, *65*, 187–213. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.