

1 Convolutional Surrogate for 3D Discrete Fracture-Matrix Tensor 2 Upscaling

3 Martin Špetlík^a, Jan Březina^a

4 ^aInstitute of New Technologies and Applied Informatics, Faculty of Mechatronics, Informatics and Interdisciplinary Studies, Technical University
5 of Liberec, Studentská 1402/2, 461 17 Liberec, Czech Republic

6 ARTICLE INFO

7 Keywords:

8 Deep learning surrogate
9 3D DFM models
10 Numerical homogenization
11 Equivalent hydraulic conductivity
12 3D Convolutional neural network

ABSTRACT

Modeling groundwater flow in three-dimensional fractured crystalline media requires capturing the spatial heterogeneity introduced by fractures. Direct numerical simulations using fine-scale discrete fracture-matrix (DFM) models are computationally demanding, particularly when repeated evaluations are needed. We aim to use a multilevel Monte Carlo (MLMC) method in the future to reduce computational cost while retaining accuracy. When transitioning between accuracy levels, numerical homogenization is used to upscale the impact of the hydraulic conductivity of sub-resolution fractures. To reduce the computational cost of conventional 3D numerical homogenization, we develop a surrogate model that predicts the equivalent hydraulic conductivity tensor \mathbf{K}^{eq} from a voxelized 3D domain representing a tensor-valued random field of matrix and fracture hydraulic conductivities. Fracture properties, including size, orientation, and aperture, are sampled from distributions informed by natural observations. The surrogate architecture combines a 3D convolutional neural network with feed-forward layers to capture both local spatial patterns and global interactions. Three surrogates are trained on data generated by discrete fracture-matrix (DFM) simulations, each corresponding to a different fracture-to-matrix conductivity ratio. Their performance is evaluated across varying fracture network parameters and correlation lengths of the matrix field. The trained surrogates achieve high prediction accuracy (NRMSE < 0.22) in a wide range of test scenarios. To demonstrate practical applicability, we compare conductivities upscaled by numerical homogenization and by our surrogates in two macro-scale problems: computation of equivalent tensors of hydraulic conductivity and prediction of outflow from a constrained 3D area. In both cases, the surrogate-based approach preserves accuracy while substantially reducing computational cost. Surrogate-based upscaling achieves speedups exceeding 100× when inference is performed on a GPU.

33 CRediT authorship contribution statement

34 **Martin Špetlík:** Writing – original draft, Software, Experimentation. **Jan Březina:** Supervision, Writing – review
35 & editing, Software.

36 1. Introduction

37 Understanding groundwater flow in fractured rock is essential for assessing the long-term safety of deep geological
38 repositories for radioactive waste disposal. However, direct numerical simulation (DNS) of such highly heterogeneous
39 media is usually computationally infeasible, especially when the flow is coupled with thermal, chemical, or mechanical
40 processes. A common remedy is to upscale the numerous small-scale fractures into an equivalent hydraulic conduc-
41 tivity tensor. Yet, the fractal nature of fracture networks leaves no distinct scale-separation threshold to guide that
42 upscaling.

43 The discrete fracture–matrix (DFM) framework addresses this challenge by coupling an explicit discrete fracture
44 network (DFN) with a surrounding continuum representation. This hybrid description enables the imposition of an ar-

ORCID(s): 0000-0002-5310-2663 (M. Špetlík)

45 arbitrary fracture-size cutoff, allowing for a smooth transition from full DNS to coarser, continuum-only models. When
 46 paired with fast and robust numerical homogenization, the resulting coarse model offers a low-cost, low-fidelity approx-
 47 imation - an essential tool for overcoming the subsurface data gap through inverse analysis and rigorous uncertainty
 48 quantification.

49 Our primary motivation is the multilevel Monte Carlo (MLMC) method, a computationally efficient strategy for
 50 uncertainty propagation and sensitivity analysis. MLMC relies on constructing a hierarchy of correlated fine- and
 51 coarse-grid solution pairs. However, applying MLMC directly to DFM models presents a significant challenge: the
 52 coarse model lacks the resolution needed to represent small-scale fractures. To capture their impact, an efficient yet
 53 accurate homogenization of their hydraulic conductivity contribution into the matrix properties of the coarse DFM
 54 model is required.

55 In our previous work Špetlík et al. (2024), we introduced deep learning-based surrogates to accelerate the homog-
 56 enization process for 2D DFM models with non-constant hydraulic conductivities in both the matrix and fractures. In
 57 this paper, we extend the 2D framework to the more realistic 3D setting. The transition introduces additional chal-
 58 lenges, including the increased geometric complexity of fractures and the higher dimensionality of both inputs and
 59 outputs for the surrogate models. The generation of tensor-valued spatial random fields (SRFs) representing matrix
 60 hydraulic conductivity at the micro-scale DFM models must be extended to three dimensions. Moreover, fracture
 61 properties are informed by natural data, such as the distribution of fracture orientations, which were previously as-
 62 sumed to be uniform in the 2D case study. To accommodate the increased complexity of 3D data, the surrogate model
 63 architecture, including its hyperparameters, requires adjustments.

64 The following review focuses specifically on studies relevant to 3D numerical homogenization and its neural
 65 network-based surrogates. For a broader overview, see Špetlík et al. (2024). To homogenize the hydraulic conductivity
 66 tensor from 3D DFM models, we build not only on our previous research but also on research conducted by Bogdanov
 67 et al. (2003). The authors homogenize a 3D cubic domain of a rock matrix containing planar fractures of a single
 68 characteristic size. Their model assumes homogeneous, isotropic matrix properties and distinct permeabilities for the
 69 matrix and fractures. Using tetrahedral meshes that explicitly resolve fractures, they compute the effective hydraulic
 70 conductivity by solving Darcy flow separately in each domain. Their work was later extended by Bogdanov et al.
 71 (2007) to account for power-law distributions in fracture sizes. Subsequently, Lang et al. (2014) investigate 3D DFM
 72 models and employ pressure gradient combined with flux averaging to compute equivalent hydraulic conductivity ten-
 73 sors on unstructured grids. Azizmohammadi and Matthäi (2017) continue in this research direction and investigate the
 74 anisotropy and scale dependence of equivalent permeability tensors.

75 In parallel, deep learning techniques, particularly convolutional neural networks (CNNs), have been increasingly
 76 used to infer properties of rock media from geometric or statistical descriptions. CNNs process volumetric data ef-

77 ficiently and capture 3D spatial dependencies. Hong and Liu (2020) train 3D CNNs to predict directional and mean
 78 permeabilities from binary images representing pore-scale geometries. More recently, Meng et al. (2023) combine
 79 CNNs with transformer architectures to enhance feature extraction from 3D porous structures. Applications of deep
 80 learning to 3D DFM models remain relatively scarce. Some studies, such as Rao and Liu (2020); Vasilyeva and
 81 Tyrylgin (2021); Wang et al. (2023), address permeability prediction in realistic 3D scenarios, but primarily focus on
 82 pore-scale imaging rather than DFM representations. Others, like Stepanov et al. (2023); Pal et al. (2023), employ
 83 feed-forward neural networks to predict equivalent permeability from micro-scale simulations, while Cai et al. (2023);
 84 Ferreira et al. (2022) investigate graph-based and generative models. Although upscaling of DFM models using deep
 85 learning has been explored in He et al. (2021); Andrianov (2022), these works assume constant hydraulic conduc-
 86 tivities. In contrast, our approach targets surrogate modeling of equivalent conductivity tensors derived from fully
 87 resolved 3D DFM simulations with stochastic parameters, making it compatible with MLMC frameworks.

88 The remainder of this article is organized as follows. Section 2 introduces the discrete fracture-matrix (DFM)
 89 model and outlines the parameters used to define the discrete fracture network. It then describes the generation of
 90 spatial random fields (SRFs) for a matrix at a micro-scale, followed by an overview of test macro-scale problems.
 91 This section also presents the numerical homogenization procedure within the multiscale DFM framework. Section 3
 92 describes the datasets used for training the surrogate and the applied preprocessing methods. Section 4 details the
 93 architecture of the surrogate. The results are presented and discussed in Section 5, while Section 6 concludes the
 94 paper.

95 2. Multiscale models of fractured media

96 This section first provides a stochastic description of fractured rock using (i) a discrete fracture network (DFN)
 97 model and (ii) an equivalent continuous medium (ECM) representation with a spatially correlated hydraulic conductiv-
 98 ity tensor field. Next, we present their coupling through the DFM formulation. Finally, we formulate two steady-state,
 99 macro-scale test problems and detail the numerical homogenization procedure.

100 2.1. Discrete fracture network

101 We consider a spatially uncorrelated discrete fracture network (DFN) described by the distributions of fracture size,
 102 aperture, orientation, and density. Fractures are assumed to be square-shaped, with size f_s drawn from a power-law
 103 distribution (Bonnet et al. (2001)):

$$f_s \sim Cr^{-\alpha}, \quad C = \frac{1 - \alpha}{\bar{r}^{1-\alpha} - \underline{r}^{1-\alpha}}, \quad (1)$$

104 where α is the power-law exponent, and \underline{r} and \bar{r} denote the fracture size bounds. The effective fracture aperture
 105 is considered a linear function of the size: $\delta = af_s$. Each fracture is assigned an isotropic hydraulic conductivity
 106 according to the cubic law:

$$k_f = \frac{g\rho_w}{12\mu} \delta^2, \quad (2)$$

107 with gravitational acceleration g , water density ρ_w , and dynamic viscosity μ . Fracture orientations follow the Fisher
 108 distribution (Adler and Thovert (1999)), parameterized by trend (horizontal projection angle), plunge (angle from the
 109 horizontal plane), and concentration. Fractures are spatially uncorrelated, placed using a Poisson process. The fracture
 110 density is controlled via P_{30} - the number of fractures per unit volume.

111 2.2. Spatially correlated random field on the matrix

112 To describe the impact of fractures that are beyond the resolution of a given model, we adopt the equivalent contin-
 113 uous medium (ECM) approach (Hadgu et al. (2017); Kottwitz et al. (2021)). Hydraulic properties are represented by a
 114 random hydraulic conductivity tensor field $\mathbf{K}(x)$, which can be defined implicitly by homogenizing a DFN realization
 115 (see Section 2.3) or explicitly through its marginal (pointwise) distribution and correlation structure. In the latter case,
 116 we represent the conductivity tensor field as:

$$\mathbf{K} = \mathbf{Q}^T \boldsymbol{\Lambda} \mathbf{Q}, \quad \boldsymbol{\Lambda} = \text{diag}(k_x, k_y, k_z). \quad (3)$$

117 We assume isotropically unresolved fractures modeled by the uniformly distributed rotation matrix $\mathbf{Q} = \mathbf{Q}_N \mathbf{Q}_R$. To
 118 impose rotations with a correlation structure, we use normalized vector fields $\mathbf{R} = \tilde{\mathbf{R}} / \|\tilde{\mathbf{R}}\|_2$ and $\mathbf{N} = \tilde{\mathbf{N}} / \|\tilde{\mathbf{N}}\|_2$, where
 119 both $\tilde{\mathbf{R}} = (\tilde{R}_x, \tilde{R}_y)$ and $\tilde{\mathbf{N}} = (\tilde{N}_x, \tilde{N}_y, \tilde{N}_z)$ are composed of independent Gaussian fields with zero mean, unit variance,
 120 and a specified correlation length λ . The vector field \mathbf{N} determines the rotation of the z -axis, while \mathbf{R} specifies the
 121 azimuthal orientation of the x -axis in the xy -plane.

122 The principal hydraulic conductivities are given by the Gaussian vector $(k_x, k_y, k_z) = \exp(\boldsymbol{\mu} + \sqrt{\boldsymbol{\Sigma}} \tilde{\mathbf{k}})$, where $\boldsymbol{\mu}$
 123 is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. Consequently, the tensor field is defined by eight independent scalar
 124 fields: $\tilde{R}_x, \tilde{R}_y, \tilde{N}_x, \tilde{N}_y, \tilde{N}_z, \tilde{k}_x, \tilde{k}_y$, and \tilde{k}_z , each following an $N(0, 1)$ distribution with Gaussian correlation of length
 125 λ . These fields are sampled using the GSTools library (Müller and Schüller (2019)).

126 2.3. Discrete fracture-matrix model

127 The discrete fracture-matrix (DFM) approach couples a DFN with a surrounding rock matrix represented by a
 128 heterogeneous ECM conductivity field. We shall provide a minimal formulation; for details and related concepts, see
 129 Sandve et al. (2012); Berrone et al. (2013); Březina and Stebel (2016). In 3D, the matrix occupies a volumetric domain

130 Ω_m , while fractures are represented as embedded 2D surfaces Ω_f , forming a mixed-dimensional mesh combining 3D
 131 and 2D finite elements. Within Ω_m , we consider the Darcy flow equation:

$$\operatorname{div} \delta_m \mathbf{u}_m = 0, \quad \mathbf{u}_m = -\mathbf{K}_m \nabla h_m, \quad (4)$$

132 where $\mathbf{u}_m(\mathbf{x})$ denotes the Darcy velocity [m/s] and $\mathbf{K}_m(\mathbf{x})$ is the hydraulic conductivity tensor [m/s]. The principal
 133 unknown is the piezometric head h_m [m]. For consistency with the fracture domain equations, we set $\delta_m = 1$. The
 134 corresponding equation on Ω_f is given by:

$$\operatorname{div} \delta_f \mathbf{u}_f = q^+ + q^-, \quad \mathbf{u}_f = -k_f \nabla h_f, \quad (5)$$

135 where $\delta_f(\mathbf{x}) = ar_i$ [m] is the aperture of fracture i , and $k_f(\mathbf{x}) = k_f^i$ denotes the isotropic hydraulic conductivity of the
 136 fracture fill. The sources q^\pm denote outflows from the two aligned boundaries of Ω_m with normal vectors \mathbf{n}^\pm , coupling
 137 the fracture and matrix domain through Robin-like boundary conditions:

$$-\mathbf{K}_m \nabla h_m \cdot \mathbf{n}^\pm = q^\pm := \frac{k_f}{2\delta} (h_m^\pm - h_f) \quad \text{on } \Omega_f. \quad (6)$$

138 The system is solved using the Flow123d simulator (Březina et al. (2011–2025)), based on a mixed finite element
 139 method.

140 We now introduce two benchmark problems designed to evaluate the performance of the proposed homogenization
 141 surrogates. The second benchmark additionally serves as the micro-scale homogenization problem that will later be
 142 approximated by a surrogate. The multiscale nature of this approach is further elaborated in the subsequent section.

143 2.3.1. Constraint problem

144 We examine fluid flow through a cubic domain $\Omega = (0, L)^3$. Dirichlet boundary conditions are imposed by setting
 145 the pressure head $h = H$ at $x = 0$ and $h = 0$ at $x = L$. No-flow boundary conditions are applied on the remaining
 146 boundaries (Y and Z faces).

147 2.3.2. Anisotropy problem

148 To obtain the full hydraulic conductivity tensor, we solve three problems on $\Omega = (0, L)^3$, applying the boundary
 149 pressure head $P^1 = x$, $P^2 = y$, and $P^3 = z$, respectively. Mixed formulation with RT_0 finite elements provides
 150 the velocity vector $\mathbf{u}(e, P^j)$ on element e for boundary condition P^j , $j = 1, 2, 3$. We get the corresponding pressure

151 gradient $\nabla h(e, P) = -\mathbf{K}_e^{-1}\mathbf{u}(e, P)$. For a vector quantity $\mathbf{v}(e, P)$ on elements, we introduce a weighted average:

$$\langle \mathbf{v} \rangle^j = \frac{\sum_{e \in \mathcal{T}} |e| \delta_e \mathbf{v}(e, P^j)}{\sum_{e \in \mathcal{T}} |e| \delta_e}, \quad j = 1, 2, 3 \quad (7)$$

152 With this notation, we are able to write down the least squares problem for the components of the symmetric equivalent
153 tensor written in Voigt notation:

154 $\mathbf{K}^{eq} = (k_{xx}, k_{yy}, k_{zz}, k_{yz}, k_{xz}, k_{xy})$:

$$-\begin{bmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \mathbf{A}^3 \end{bmatrix} (\mathbf{K}^{eq})^T = \begin{bmatrix} \langle \mathbf{u} \rangle^1 \\ \langle \mathbf{u} \rangle^2 \\ \langle \mathbf{u} \rangle^3 \end{bmatrix},$$

155 with blocks:

$$A^i = \begin{bmatrix} \langle \nabla h \rangle_x^j & 0 & 0 & 0 & \langle \nabla h \rangle_z^j & \langle \nabla h \rangle_y^j \\ 0 & \langle \nabla h \rangle_y^j & 0 & \langle \nabla h \rangle_z^j & 0 & \langle \nabla h \rangle_x^j \\ 0 & 0 & \langle \nabla h \rangle_z^j & \langle \nabla h \rangle_y^j & \langle \nabla h \rangle_x^j & 0 \end{bmatrix}.$$

156 2.4. Numerical homogenization in the context of multiscale models

157 A single level of MLMC consists of a pair of correlated fine and coarse resolution models. To this end, we adopt
158 a two-scale strategy where the coarse model is a macro-scale model using equivalent hydraulic conductivity tensors
159 resulting from block numerical homogenization of micro-scale models.

160 We start from the fine model. Its mesh discretization uses elements of maximum size h and can explicitly capture
161 fractures $\mathcal{F}_{h,L}$ in the size range (h, L) . The effect of fractures smaller than h is represented by a realization of a spatially
162 correlated random field of hydraulic conductivity \mathbf{K}_h (see Section 2.2).

163 The coarse model is discretized with a larger element size H ($H > h$), and explicitly resolves fractures $\mathcal{F}_{H,L}$ in the
164 size range (H, L) . The effect of smaller fractures $\mathcal{F}_{h,H}$ is incorporated in the hydraulic conductivity tensor field \mathbf{K}_H ,
165 which is determined by numerical homogenization in terms of the micro-scale instance of the Anisotropy problem.

166 The following procedure is adopted:

- 167 1. Center points \mathbf{x} of homogenization blocks are determined within domain of size (L, L, L) . Homogenization
168 blocks are of size $l \times l \times l$ with overlap $l/2$, where $l = 1.5H$.
- 169 2. The fracture network and the hydraulic conductivity field \mathbf{K}_h are generated on an enlarged domain of side $L+2l$,
170 enabling homogenization blocks to capture boundary regions.
- 171 3. For each block, the intersection of the fractures $\mathcal{F}_{h,H}$ with the block is determined, a compatible mesh with a

172 maximum element size of h is constructed, and the field \mathbf{K}_h is interpolated to the mesh.

173 4. A field of equivalent hydraulic conductivity tensors $\mathbf{K}_H(\mathbf{x}) = P_h(\mathbf{x}, \mathcal{F}_{h,H}, \mathbf{K}_h)$ is calculated and linearly inter-
174 polated to the unstructured coarse grid cells.

175 Figure 1 shows an example of $\mathcal{F}_{h,L}$ (left) and the corresponding $\mathcal{F}_{H,L}$ (right), for $L = 60$, $H = 10$, and $h = 5$.
176 Large fractures from the surrounding volume that intersect the domain of interest can produce small fragments, which
177 remain visible even on the coarse DFM mesh. To illustrate the homogenization effect, Figure 2 presents a slice in the
178 ZY-plane through the center of the cube. The homogenized representation of $\mathcal{F}_{h,H}$ (left, shown in orange) exhibits
179 reduced variability in hydraulic conductivity in the corresponding coarse model (right).

180 To ease compatibility with convolutional neural networks, \mathbf{K}_h is interpolated from \mathbf{K}_G , which is a hydraulic con-
181 ductivity SRF generated on a regular grid slightly larger than L . This approach allows us to use \mathbf{K}_G for voxelization
182 to avoid costly mesh generation. Extension of this homogenization scheme to the full MLMC is a topic of ongoing
183 research.

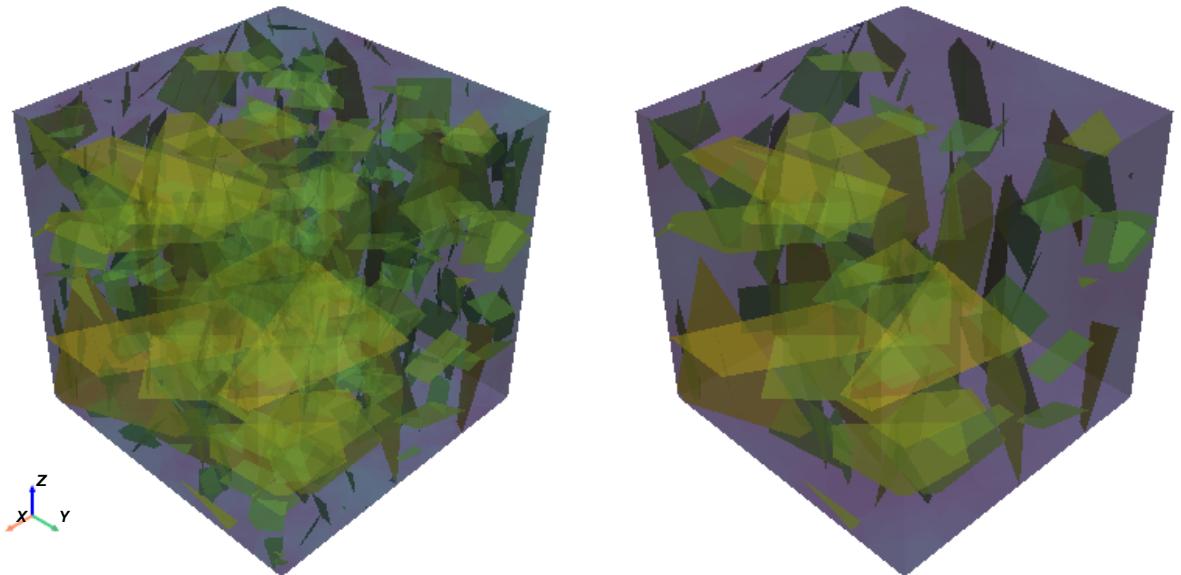


Figure 1: Comparison of the fine DFM model fracture network $\mathcal{F}_{h,L}$, with $h < 5$ (left), and the corresponding coarse DFM model fracture network $\mathcal{F}_{H,L}$, with $H < 10$ (right).

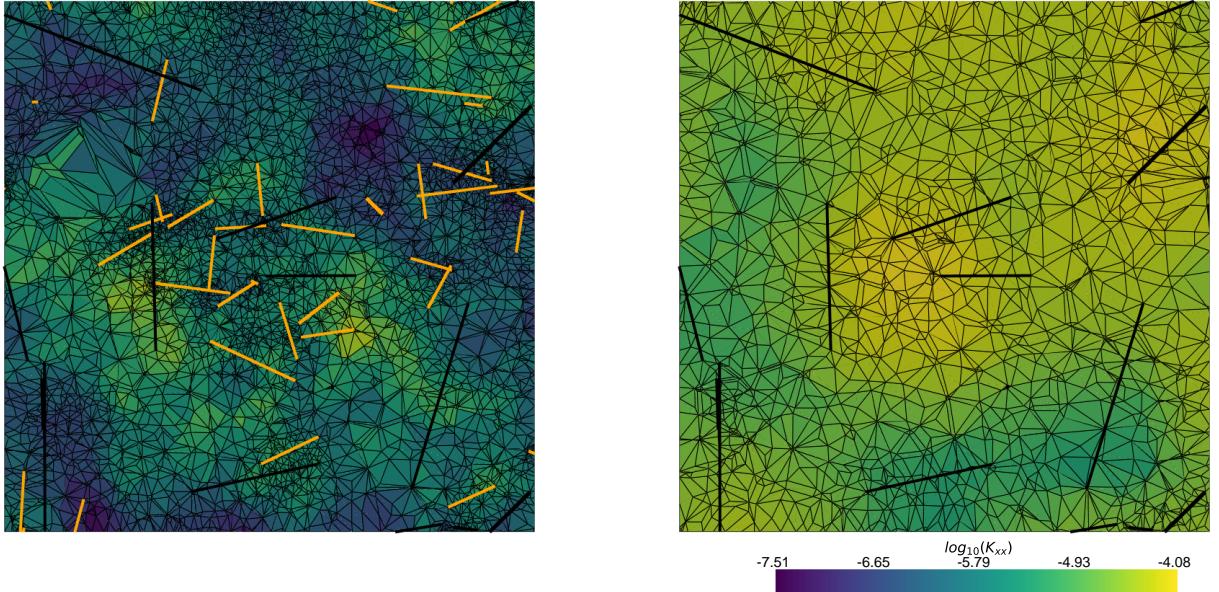


Figure 2: Left: The fine DFM model ($h < 5$). Orange fractures and the hydraulic conductivity tensor field are homogenized using overlapping square blocks of size 15. **Right:** Corresponding coarse DFM model ($H < 10$) with homogenized hydraulic conductivity tensor component K_{xx} . Notice the reduced range in the upscaled field. Only the K_{xx} component of the hydraulic conductivity tensor is shown.

184 3. Datasets

185 Let $\mathcal{D} = \{(\mathbf{X}_j, \mathbf{K}_j^{eq})\}_{j=1}^D$ be a dataset of independent and identically distributed samples. Here $\mathbf{X}_j \in \mathbb{R}^{6 \times F}$ is a
 186 vector of input features (voxelized upper triangle of hydraulic conductivity tensors of bulk and fractures, F - number
 187 of voxels) and $\mathbf{K}_j^{eq} \in \mathbb{R}^6$ is the vector of components: k_{xx} , k_{yy} , k_{zz} , k_{yz} , k_{xz} , and k_{xy} of the corresponding \mathbf{K}^{eq}
 188 obtained by numerical homogenization described in Section 2.4. Individual homogenization samples are generated
 189 independently on a fixed domain $\Omega_o = (0, 15)^3$. The DFN is generated with parameters listed in Table 1 - adopted
 190 from (Joyce et al., 2009, p. 23). The fracture density P_{30} (number of fractures per unit volume) is derived from k_r
 191 and P_{32} . To achieve a target average P_{30} of the whole population, the parameter r_0 is adjusted. The size distribution
 192 parameters are set as $\alpha = 2$, $r = 5$, and $\bar{r} = 100$, while the aperture parameter is $a = 1 \times 10^{-4}$.

193 Similarly to our 2D study (Špetlík et al. (2024)), we examine various fracture-to-matrix hydraulic conductivity
 194 ratios $K_f/K_m \in \{1 \times 10^3, 1 \times 10^5, 1 \times 10^7\}$, forming three datasets:

195 • **Dataset A:** $K_f/K_m = 1 \times 10^3$,

196 • **Dataset B:** $K_f/K_m = 1 \times 10^5$,

197 • **Dataset C:** $K_f/K_m = 1 \times 10^7$.

Table 1
DFN parameters

Fracture set name	Orientation (trend, plunge), concentration	Size model power-law k_r	P_{32} (mean fracture surface area) (1, 564 m)	P_{30} (fracture intensity) (1, 564 m)
NS	(292, 1) 17.8	2.50	0.094	0.0196
NE	(326, 2) 14.3	2.70	0.163	0.0427
NW	(60, 6) 12.9	3.10	0.098	0.0348
EW	(15, 2) 14.0	3.10	0.039	0.0138
HZ	(5, 86) 15.2	2.38	0.141	0.0247

¹⁹⁸ The K_m field is generated with a covariance matrix

$$\Sigma = \begin{bmatrix} 0.25 & 0.2 & 0.2 \\ 0.2 & 0.25 & 0.2 \\ 0.2 & 0.2 & 0.25 \end{bmatrix},$$

¹⁹⁹ and dataset-specific mean vectors: $\mu = [-4.0, -3.8, -3.9]$ for Dataset \mathcal{A} , $\mu = [-6.0, -5.8, -5.9]$ for Dataset \mathcal{B} , and

²⁰⁰ $\mu = [-8.0, -7.8, -7.9]$ for Dataset \mathcal{C} .

²⁰¹ Each dataset contains 75,000 samples, evenly split between $P_{30} = 0.001$ and $P_{30} = 0.0025$. To enable generalization ²⁰² across different matrix correlation lengths, each dataset includes equal representation of $\lambda \in \{0, 10, 25\}$. Samples ²⁰³ are split into 80% training set \mathcal{L} and 20% test set \mathcal{T} . The training set is further split into 80% training and 20% validation ²⁰⁴ samples (\mathcal{V}).

²⁰⁵ In order to form these datasets of homogenization blocks, the original input hydraulic conductivities \mathbf{K}_G prescribed ²⁰⁶ on a regular grid and fractures $\mathcal{F}_{h,H}$ within the homogenization block are voxelized by the Bgem library (Březina et al. ²⁰⁷ (2025)). Figure 3 shows an input mesh with fractures and its voxelized counterpart, illustrating the K_{xx} component ²⁰⁸ of the hydraulic conductivity tensors. The plots are adjusted to enhance fracture visibility. As the voxelization quality ²⁰⁹ required for accurate neural network predictions is not established, no systematic assessment was performed. To ²¹⁰ simplify and speed up, the voxelization algorithm considers disc-shaped fractures of equivalent area. According to ²¹¹ our experiments, this approximation is able to capture key patterns and provides sufficient information to the surrogate ²¹² model.

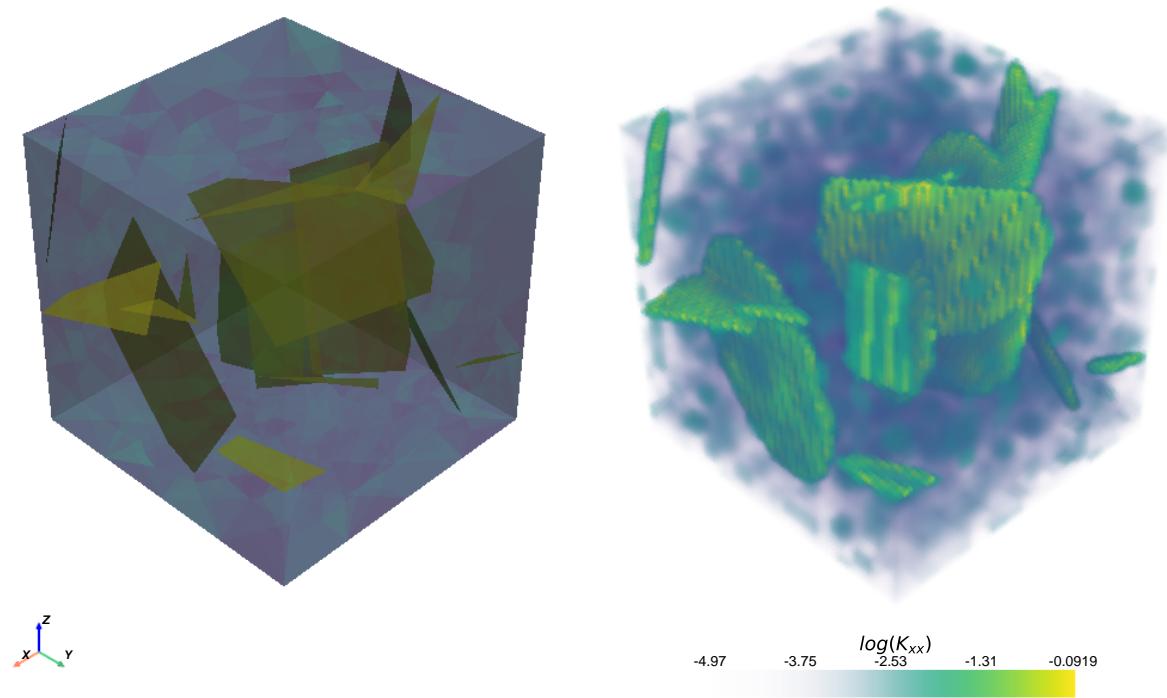
K_{xx} (bulk surface + fractures) K_{xx} voxelized

Figure 3: An illustration of an input hydraulic conductivity tensors (K_{xx} component) on a mesh and its voxelized representation.

213 3.1. Datasets analysis

214 The distributions of components of equivalent hydraulic conductivity tensors for datasets \mathcal{A} , \mathcal{B} , and \mathcal{C} are depicted
 215 in Figure 4. Focusing on the diagonal components: k_{xx} , k_{yy} , k_{zz} we observe that, for Dataset \mathcal{A} , the equivalent tensor
 216 is primarily influenced by matrix elements characterized by $\mu = [-4, -3.8, -3.9]$. The peak values are attributed
 217 to samples with $\lambda = 0$. For Dataset \mathcal{B} , we note an increased impact of fractures. There is a heavier left tail of the
 218 distributions that accounts for samples with a low number of fractures that are not fully connected. As the ratio K_f/K_m
 219 increases, the left tail of the distribution becomes heavier (K_m gets lower in our setting). The maximum values in \mathbf{K}^{eq}
 220 do not increase, which indicates that even for $K_f/K_m = 1 \times 10^5$, there are cases where the impact of bulk on \mathbf{K}^{eq} is
 221 negligible. This is more pronounced for Dataset \mathcal{C} . Regarding off-diagonal components: k_{yz} , k_{xz} , k_{xy} , the data shape
 222 is very similar for Dataset \mathcal{B} and Dataset \mathcal{C} . While for Dataset \mathcal{A} , the off-diagonal values are more concentrated, but
 223 of the order of magnitude higher.

224 The equivalent tensor distributions differ notably in both scale and shape, making it difficult to train a single surro-
 225 gate model across varying K_f/K_m ratios. To address this, we train separate surrogate models: Surrogate A for Dataset
 226 \mathcal{A} , Surrogate B for Dataset \mathcal{B} , and Surrogate C for Dataset \mathcal{C} . For each input \mathbf{X}_j , the surrogate corresponding to its
 227 K_f/K_m ratio is used to predict \mathbf{K}_j^{eq} .

228 **3.2. Dataset preprocessing**

229 To facilitate surrogate training, inputs \mathbf{X}_j and outputs \mathbf{K}_j^{eq} are preprocessed. For each sample j , the average matrix
 230 hydraulic conductivity $\bar{\mathbf{X}}_j^m$ is computed as a scalar by averaging over all tensor components and spatial locations. Inputs
 231 and outputs are then normalized to allow generalization across different conductivity scales for the same K_f/K_m ratio:

$$\mathbf{X}_j = \frac{\mathbf{X}_j}{\bar{\mathbf{X}}_j^m}, \quad \mathbf{K}_j = \frac{\mathbf{K}_j}{\bar{\mathbf{X}}_j^m}.$$

232 Second, data is standardized independently for input and output hydraulic conductivities. Each component $k_{\alpha\beta}^j$ of the
 233 hydraulic conductivity tensor (where $\alpha, \beta \in \{x, y, z\}$) is standardized:

$$\tilde{k}_{\alpha\beta}^j = \begin{cases} \frac{\log(k_{\alpha\alpha}^j) - \hat{\mu}_{\mathcal{L}}(\log k_{\alpha\alpha})}{\hat{\sigma}_{\mathcal{L}}(\log k_{\alpha\alpha})}, & \text{if } \alpha = \beta \text{ (diagonal components)} \\ \frac{k_{\alpha\beta}^j - \hat{\mu}_{\mathcal{L}}(k_{\alpha\beta})}{\hat{\sigma}_{\mathcal{L}}(k_{\alpha\beta})}, & \text{if } \alpha \neq \beta \text{ (off-diagonal components)} \end{cases}$$

234 where the empirical mean $\hat{\mu}_{\mathcal{L}}(k_{\alpha\beta}) = \frac{1}{|\mathcal{L}|} \sum_{j \in \mathcal{L}} k_{\alpha\beta}^j$ and empirical standard deviation

235 $\hat{\sigma}_{\mathcal{L}}(k_{\alpha\beta}) = \sqrt{\frac{1}{|\mathcal{L}|-1} \sum_{j \in \mathcal{L}} (k_{\alpha\beta}^j - \hat{\mu}_{\mathcal{L}}(k_{\alpha\beta}))^2}$ are estimated from the training set \mathcal{L} .

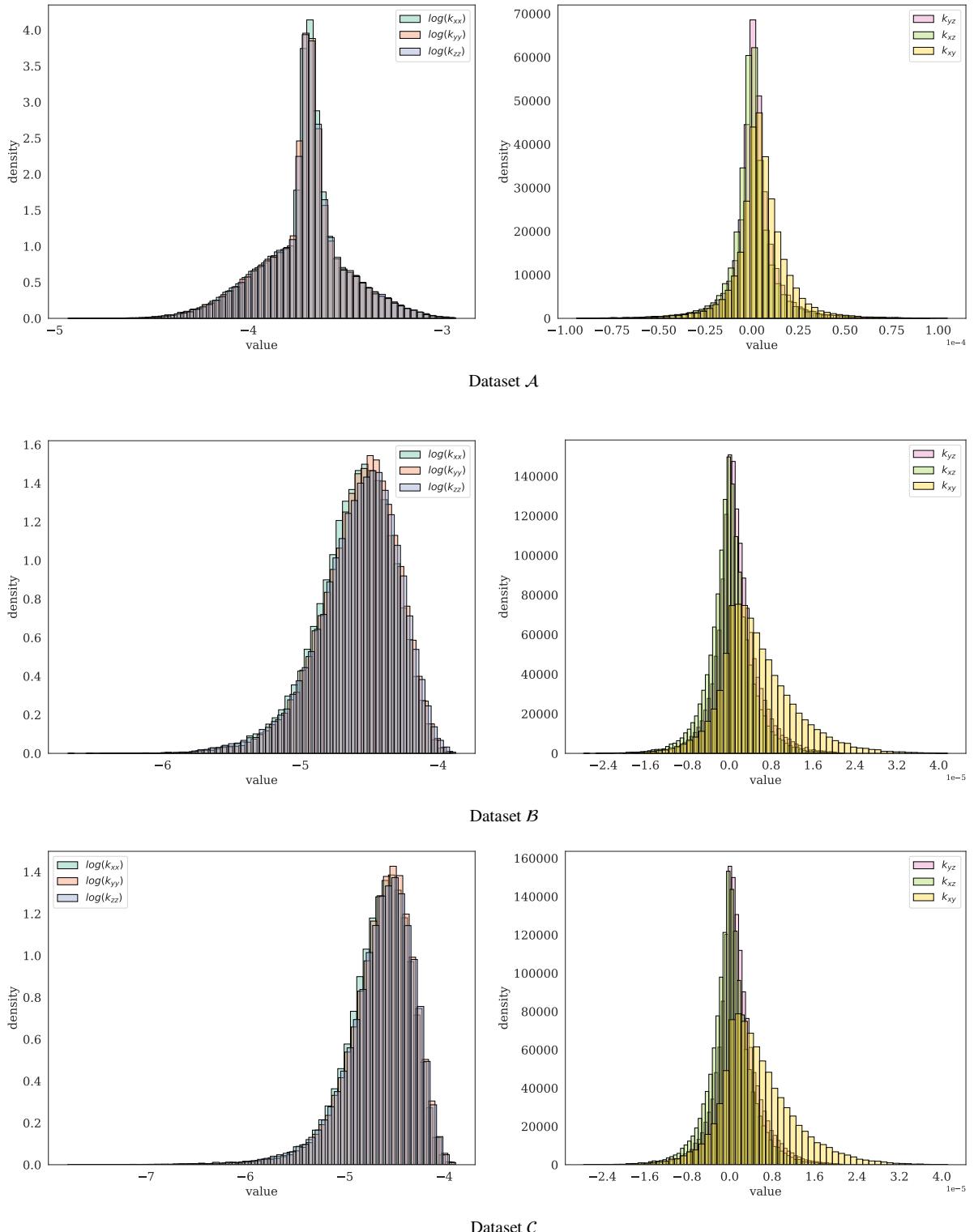


Figure 4: Distributions of \mathbf{K}^{eq} components for datasets of different K_f/K_m , Dataset \mathcal{A} for $K_f/K_m = 1 \times 10^3$, Dataset \mathcal{B} for $K_f/K_m = 1 \times 10^5$, and Dataset \mathcal{C} for $K_f/K_m = 1 \times 10^7$. Diagonal components are shown on a \log_{10} scale.

236 4. Surrogate architecture

237 The architecture of the surrogates is closely guided by the structure of the dataset samples. Given the nature of
238 the input data, the surrogate combines a convolutional neural network (CNN), which acts as a feature extractor (see
239 Section (Goodfellow et al., 2016, ch. 9)), with a feed-forward neural network (FNN) for the regression task (see (Good-
240 fellow et al., 2016, p. 328)). The CNN part comprises four Conv3D layers, each integrating 3D convolution and max
241 pooling. The convolution uses a kernel of size (3, 3, 3), stride 1, and padding 1, followed by max pooling with size
242 (2, 2, 2) and stride 2. Batch normalization is applied after each convolution to stabilize training. Starting from input
243 data of shape $64 \times 64 \times 64 \times 6$, the CNN reduces it to a feature map of size $4 \times 4 \times 4 \times 1296$, which is then flattened
244 using global average pooling. The resulting vector passes through fully connected layers with 2048, 2048, and 1024
245 neurons. So far, each layer uses ReLU activation to introduce nonlinearity. The final output layer has 6 neurons with
246 an identity activation function. The full architecture is detailed in Table 2.

Table 2

The Architecture of the surrogate

Layer	Type	Output size	Learnable parameters
	Input	$64 \times 64 \times 64 \times 6$	
1	Conv3D	$32 \times 32 \times 32 \times 48$	7,824
2	Conv3D	$16 \times 16 \times 16 \times 144$	186,768
3	Conv3D	$8 \times 8 \times 8 \times 432$	1,680,048
4	Conv3D	$4 \times 4 \times 4 \times 1296$	15,117,840
5	AvgPool	$1 \times 1 \times 1 \times 1296$	
6	Linear	2048	2,656,256
7	Linear	2048	4,196,352
8	Linear	1024	2,098,176
9	LinearOutput	6	6,150

247 Surrogates are trained for 125 epochs, with the version yielding the lowest validation loss retained for further
 248 use. The initial learning rate $\alpha = 0.0025$ is adaptively reduced by 50% if the validation loss does not improve for 10
 249 consecutive epochs. Training is performed using the Adam optimizer with a batch size of 64.

250 Mean squared error (MSE) is used as the loss function, monitored across training, validation, and test datasets.
 251 To compare models, we report the normalized root mean squared error (NRMSE), which expresses prediction error
 252 relative to target variability. For component \mathbf{K}_i^{eq} , $i = 1, \dots, 6$ NRMSE is defined as:

$$\text{NRMSE}_i = \frac{\sqrt{\frac{1}{|\mathcal{M}|} \sum_{\mathcal{M}} (\mathbf{K}_i^{eq} - f_i(\mathbf{X}))^2}}{\text{std}_{\mathcal{M}}(\mathbf{K}_i^{eq})}, \quad (8)$$

253 where the numerator is the root mean squared error, and the denominator is the empirical standard deviation of the
 254 target values over the test set \mathcal{M} . This normalization enables meaningful comparisons of NRMSE across components
 255 and test datasets with varying output variances. For further analysis, we also compute the average NRMSE across all
 256 components:

$$\overline{\text{NRMSE}} = \frac{1}{6} \sum_{i=1}^6 \text{NRMSE}_i. \quad (9)$$

257 5. Results

258 5.1. Prediction accuracy on test datasets

259 This section evaluates the prediction accuracy of the trained surrogates on the test datasets \mathcal{T}_A , \mathcal{T}_B , and \mathcal{T}_C . Figure 5
 260 displays target-versus-prediction plots and NRMSE values for each surrogate and every component of the predicted
 261 \mathbf{K}^{eq} .

262 For the diagonal components (left column), Surrogate A demonstrates the highest accuracy (NRMSE < 0.04),

which can be attributed to the narrow distribution of these components. With training datasets of equal size, narrower distributions tend to be better represented, resulting in improved model performance. As the data distributions become broader and more complex (Figure 4), the predictive accuracy of the surrogate models declines (Figure 5). This trend is especially pronounced for Surrogate C, where predictions for samples in the left tail of the distribution are considerably less accurate, likely due to their underrepresentation in the training set. This issue becomes more severe with increasing K_f/K_m .

In contrast, for the off-diagonal components, Surrogate A performs the worst ($\text{NRMSE} \approx 0.14$), reflecting the broader and more complex nature of these components' distributions. As a result, the overall prediction accuracy of Surrogate A is largely constrained by its ability to estimate off-diagonal components. Surrogate B achieves a more uniform accuracy across all components of \mathbf{K}^{eq} . For Surrogate C, the off-diagonal components are predicted with even slightly higher accuracy than the diagonal ones.

Although underrepresentation in certain regions of the training distribution adversely affects surrogate performance, this issue could potentially be alleviated through targeted dataset augmentation. However, we do not have a straightforward method for generating DFM homogenization blocks that correspond to a prescribed \mathbf{K}^{eq} .

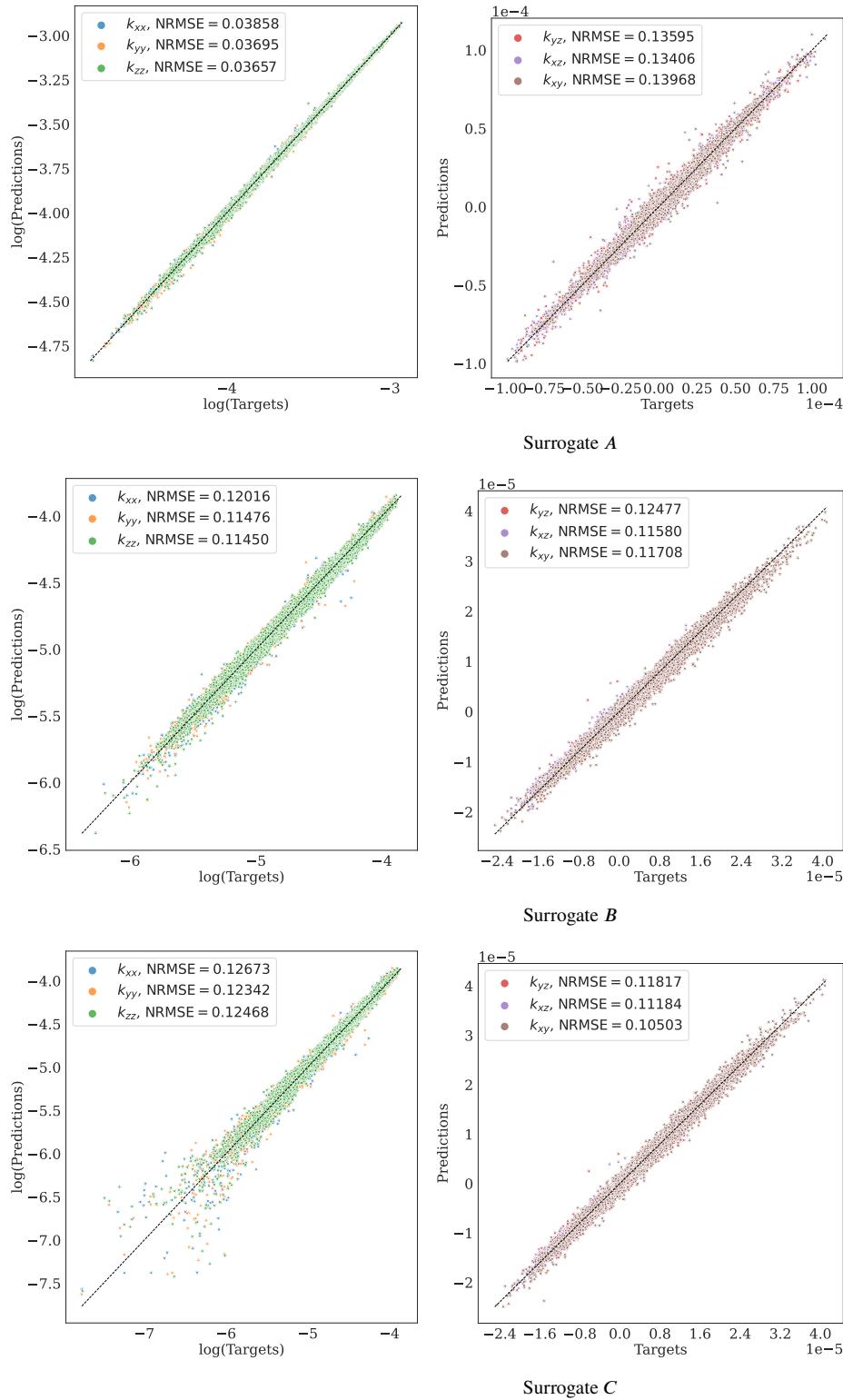


Figure 5: The prediction accuracy of the trained surrogates for the components of \mathbf{K}^{eq} evaluated on test datasets. Diagonal components are shown on a \log_{10} scale.

277 **5.2. Impact of correlation length of SRF on prediction accuracy**

278 We evaluated the effect of the matrix SRF correlation length λ on surrogate accuracy using test datasets of 3,500
 279 samples with fixed fracture density ($P_{30} = 0.0015$) and $\lambda \in \{0, 5, 10, 15, 25, 50, 75, 100, 250, 500, 1000\}$. Surrogates
 280 were trained on data with $\lambda \in \{0, 10, 25\}$ equally represented.

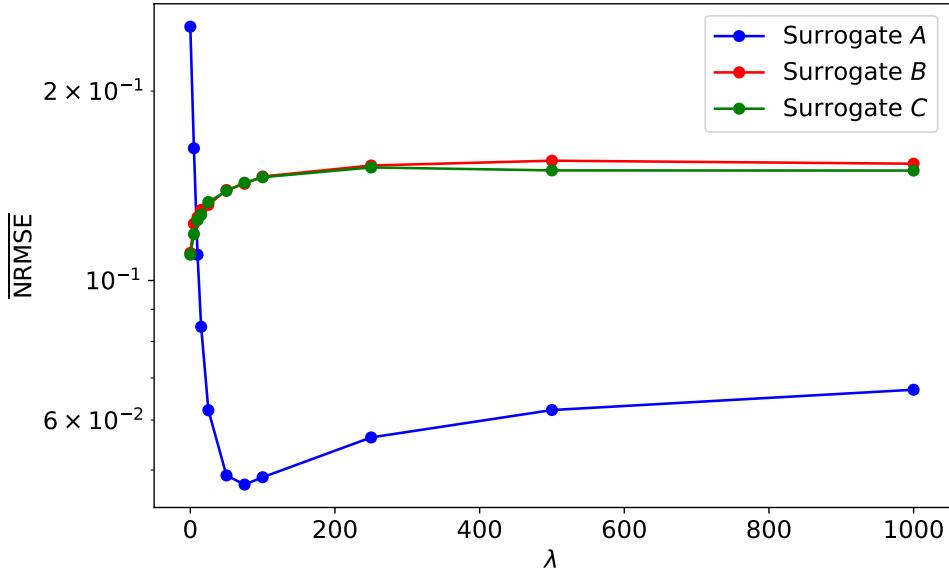


Figure 6: Effect of the SRF correlation length λ on the prediction accuracy of the surrogate model.

281 As shown in Figure 6, the prediction accuracy (NRMSE) of Surrogate A improves with increasing λ up to 75,
 282 consistent with reduced SRF variability at longer correlation lengths and limited fracture influence, which simplifies
 283 the learning task. Accuracy slightly declines beyond $\lambda = 75$, likely due to limited training data. In contrast, Surrogates
 284 B and C show decreasing accuracy with increasing λ . Trained on high fracture–matrix conductivity ratios ($K_f/K_m =$
 285 1×10^5 and 1×10^7), they struggle to distinguish samples with similar DFNs but varying SRFs, particularly at large λ
 286 values underrepresented in training datasets. Both perform best for uncorrelated fields ($\lambda = 0$).

287 All surrogates generalize well to intermediate $\lambda = \{5, 15\}$, indicating that training on $\lambda = \{0, 10, 25\}$ sufficiently
 288 captures the range of relevant SRF correlation lengths within the $15 \times 15 \times 15$ homogenization block. Similar trends
 289 were observed in our 2D study Špetlík et al. (2024). Henceforth, $\lambda = 10$ is used for testing, as it yields consistent
 290 accuracy across all surrogates.

291 **5.3. Impact of number of fractures**

292 We further analyzed the prediction accuracy of the surrogates for test datasets of a different number of fractures.
 293 Originally, surrogates were trained on the dataset with samples of $P_{30} = 0.0010$ and $P_{30} = 0.0025$ equally represented.

294 Test datasets of 3,500 samples, $\lambda = 10$ and $P_{30} \in \{0.0004, 0.0006, 0.0008, 0.0012, 0.0015, 0.0020, 0.0025, 0.0030, 0.0050,$
 295 $0.0075, 0.0100, 0.0150, 0.0200\}$ were formed.

296 Figure 7 shows the development of $\overline{\text{NRMSE}}$ depending on P_{30} . All three surrogates achieve their best prediction ac-
 297 curacy at $P_{30} = 0.0008$. For Surrogate A, $\overline{\text{NRMSE}}$ remains nearly constant for $P_{30} \in \{0.0004, 0.0006, 0.0008, 0.0012\}$,
 298 followed by a sharp increase, except at $P_{30} = 0.0025$, which was explicitly included for training. Even for $P_{30} = 0.0200$,
 299 Surrogate A maintains a reasonable accuracy with $\overline{\text{NRMSE}} = 0.2151$. Surrogates B and C show a different trend: their
 300 accuracy slightly improves from $P_{30} = 0.0004$ to $P_{30} = 0.0008$, followed by an increase that drops at $P_{30} = 0.0020$.
 301 Surrogate A is the most sensitive to P_{30} values outside its training range. Surrogate B demonstrates greater robustness
 302 to changes in P_{30} . The sharper increase in $\overline{\text{NRMSE}}$ for $P_{30} > 0.0020$ for Surrogate C compared to Surrogate B is
 303 driven by errors in predicting the diagonal components.

304 Overall, the surrogates achieve prediction accuracy with $\overline{\text{NRMSE}} < 0.22$ across all investigated P_{30} values. This
 305 level of accuracy is likely sufficient for a wide variety of applications. If higher accuracy is desired, the training dataset
 306 can be rebalanced by incorporating samples covering a broader range of P_{30} values.

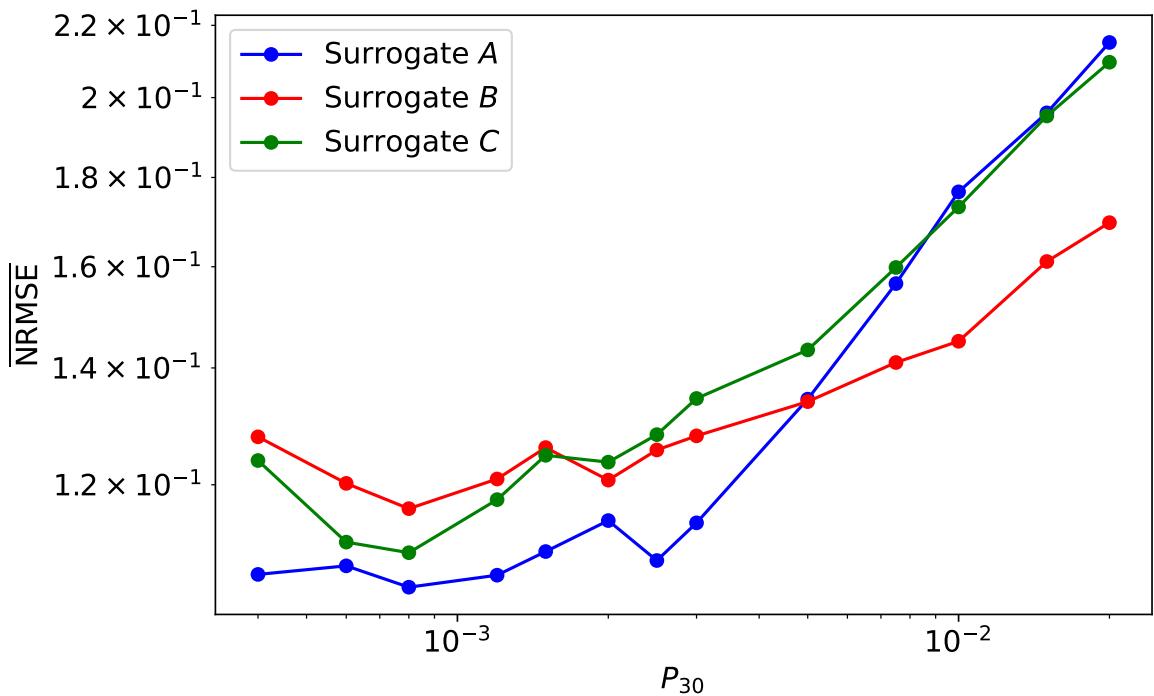


Figure 7: Impact of number of fractures (P_{30}) on surrogate prediction accuracy.

Table 3
Test DFN configurations

DFN	Fracture set name	Orientation (trend, plunge), concentration	Size model power-law k_r	P_{32} (mean fracture surface area) (1, 564 m)	P_{30} (fracture intensity) (1, 564 m)
DFN 1	NS	(292, 1) 17.8	2.50	0.073	0.0152
	NE	(326, 2) 14.3	2.70	0.319	0.0837
	NW	(60, 6) 12.9	3.10	0.107	0.0380
	EW	(15, 2) 14.0	3.10	0.088	0.0313
	HZ	(5, 86) 15.2	2.38	0.543	0.0953
DFN 2	NS	(292, 1) 17.8	2.50	0.142	0.0296
	NE	(326, 2) 14.3	2.70	0.345	0.0905
	NW	(60, 6) 12.9	3.10	0.133	0.0472
	EW	(15, 2) 14.0	3.10	0.081	0.0288
	HZ	(5, 86) 15.2	2.38	0.316	0.0554
DFN 3	NNE	(283, 6.5) 13	2.50	0.230	0.0480
	ENE	(332, 0.5) 21	2.45	0.180	0.0351
	WNW	(201, 5) 12	2.30	0.190	0.0291
	NNW	(246, 10) 12	2.30	0.050	0.0077
	Sub-H	(334, 87) 10	2.20	0.350	0.0443
DFN 4	ENE	(340.3, 1.2) 15	2.60	0.390	0.0921
	WNW	(208.9, 2.2) 10.9	2.30	0.550	0.0843
	NS	(272.8, 12) 11.5	2.70	0.320	0.0840
	SubH	(277.1, 84.3) 11.1	2.65	0.290	0.0723

5.4. Prediction accuracy on datasets of different DFN settings

307 To evaluate the prediction accuracy across varying DFN parameter settings, we generated four additional test
 308 datasets, each comprising 3,500 samples with $\lambda = 10$ and $P_{30} = 0.0015$. Table 3 summarizes the DFN configurations.
 309 DFN 1 and DFN 2 follow the hydrological DFN model by Joyce et al. (2009)[Table 2-3], developed for groundwater
 310 flow simulations in granite bedrock at the Forsmark nuclear power plant site in Sweden. DFN 3 and DFN 4, adopted
 311 from site descriptive modeling of the Laxemar area (Rhén et al. (2008)), also represent granite bedrock. DFN 3 (Rhén
 312 et al., 2008, Table 10-31) was used for upscaling studies, while DFN 4 (Rhén et al., 2008, Table 10-16) was used in a
 313 model for estimating open fracture intensity.
 314

315 Figure 8 compares prediction accuracy (NRMSE) across different DFN configurations (DFN 1–4), alongside the
 316 training setup (DFN 0). Results are shown separately for the diagonal components (k_{xx}, k_{yy}, k_{zz}) and the off-diagonal
 317 components (k_{yz}, k_{xz}, k_{xy}) of \mathbf{K}^{eq} , as predicted by Surrogates A, B, and C. Off-diagonal predictions exhibit consistent
 318 accuracy across all DFN configurations for each surrogate, with minor variations likely due to dataset size and potential
 319 outliers. In contrast, diagonal predictions vary more significantly. Surrogates show a slight increase in NRMSE for
 320 DFN 1–4. This is attributed to rare but significant fracture effects that form distribution tails, which amplify error
 321 metrics.

322 Despite the differences among DFN configurations, all surrogates maintain robust performance. This indicates
 323 that fractures within the homogenization block do not exhibit patterns that can impair neural networks' generalization

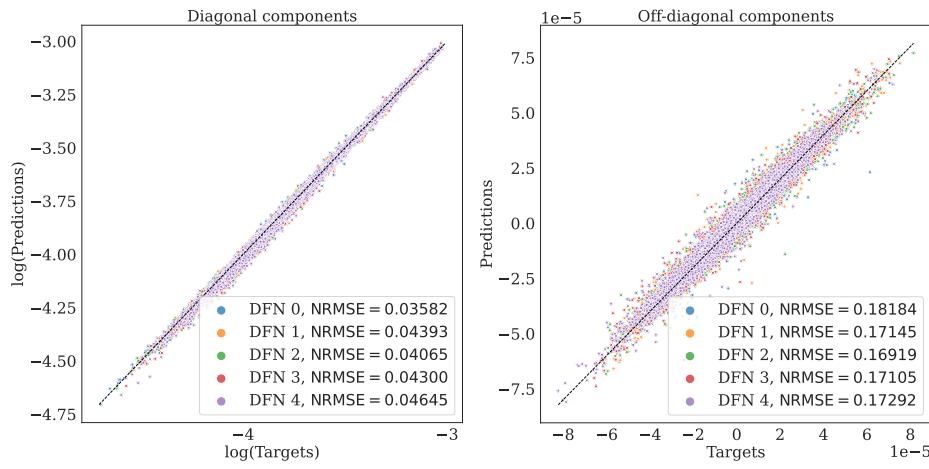
324 capability.

325 5.5. Computational cost reduction

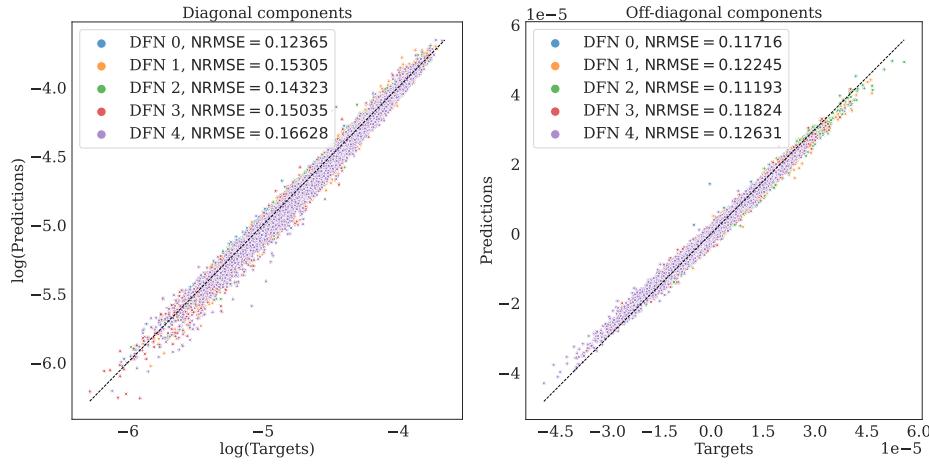
326 Surrogates are primarily employed to accelerate numerical homogenization. To quantify this benefit, we compare
 327 the computational cost of standard numerical homogenization (C_H - CPU time) with that of surrogate inference (C_S
 328 - GPU time), excluding training time. We examine cases with varying domain sizes Ω , corresponding to different
 329 numbers of homogenization blocks placed according to the procedure described in Section 2.4. Numerical homoge-
 330 nization was performed on an Intel Xeon Silver 4114 CPU (2.2 GHz, 45 GB RAM), while surrogate GPU inference
 331 used an NVIDIA Tesla T4 (16 GB). Preprocessing steps such as SRF and fracture generation are excluded from the
 332 comparison, as they are identical across cases.

333 Figure 9 presents the computational times for both approaches. As expected, the cost increases with domain size,
 334 with standard numerical homogenization being the most computationally expensive. The cost C_H includes mesh and
 335 SRF generation ($\approx 60\%$) and simulation runs for each block ($\approx 30\%$). In contrast, the cost C_S is dominated by
 336 voxelization (about 58%), with the remainder attributed to data handling; neural network inference time is negligible.
 337 The entire domain is voxelized once and split into blocks, which explains the use of 45 GB RAM. If memory becomes
 338 a limiting factor, voxelization can be performed in smaller chunks. We gained a speedup of $C_H/C_S > 100$ for all
 339 investigated cases. Inference on CPU reduces $C_H/C_S^{CPU} > 16$.

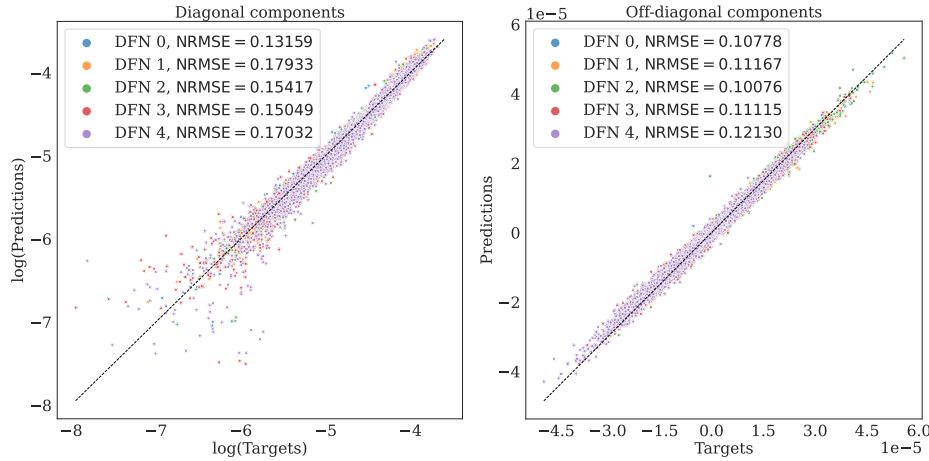
Convolutional Surrogate for 3D Discrete Fracture-Matrix Tensor Upscaling



Surrogate A



Surrogate B



Surrogate C

Figure 8: The prediction accuracy of the trained surrogates for diagonal and off-diagonal components of \mathbf{K}^{eq} evaluated on test datasets with different DFNs. Diagonal components are shown on a \log_{10} scale.

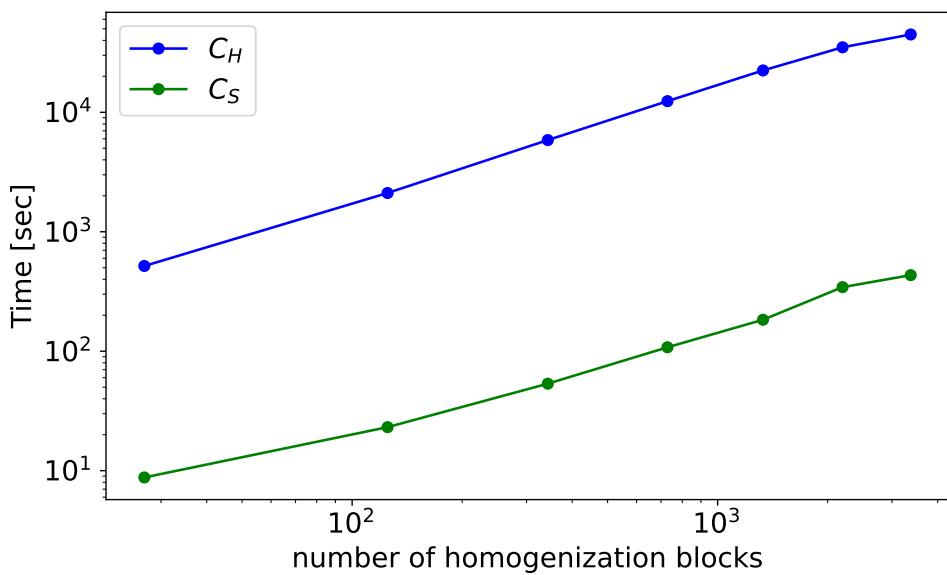


Figure 9: Comparison of computational time for numerical homogenization C_H (CPU run) and its surrogate counterpart C_S (GPU used for inference). Data shown on a \log_{10} scale.

340 **5.6. Upscaled equivalent hydraulic conductivity tensors in use**

341 Although the surrogates demonstrate high predictive accuracy (NRMSE < 0.22, corresponding to a coefficient of
 342 determination $R^2 > 0.95$), the precision required for practical applications remains uncertain. To investigate this, we
 343 examine how surrogates' predicted equivalent tensors for \mathbf{K}_H affect the Constraint and Anisotropy problems described
 344 in Section 2.3. Following the procedure outlined in Section 2.4, we interpolated the grid of homogenized/predicted
 345 \mathbf{K}_H values onto the barycenters of the bulk elements in the unstructured DFM mesh. In each case, the outputs of the
 346 macro-scale DFM models computed using purely numerical homogenization were used as target values.

347 We consider DFM samples characterized by $P_{30} = 0.0015$, DFN 0 configuration, $\lambda = 10$, a fine mesh step size
 348 of $h = 5$, and a coarse mesh step size $H = 10$. Trained Surrogate A, Surrogate B, and Surrogate C are employed
 349 to predict \mathbf{K}^{eq} , which are subsequently assembled into \mathbf{K}_H . Two domain sizes are analyzed: $\Omega_{UP} = (60, 60, 60)$ and
 350 $\Omega_{UP} = (15, 15, 15)$. The smaller domain represents a scenario in which the coarse sample contains only a small number
 351 of fractures with limited influence on the overall hydraulic conductivity. Given a homogenization block size of $l = 15$
 352 ($1.5H$), with an overlap of $l/2$, the number of homogenization blocks per sample is 729 for $\Omega_{UP} = (60, 60, 60)$ and 27
 353 for $\Omega_{UP} = (15, 15, 15)$, respectively.

354 **5.6.1. Constraint problem**

355 We first compare the upscaled DFM models for the Constraint problem described in Section 2.3.1. Table 4 reports
 356 the NRMSE between the target outflow Y (using numerical homogenization for upscaling) and the prediction outflow
 357 (upscaling by surrogates), for the domain $\Omega_{UP} = (60, 60, 60)$. In all three cases, using surrogates to determine \mathbf{K}_H
 358 produces coarse model outflow values Y that closely match the reference, with NRMSE < 0.01. Moreover, the absolute
 359 values of Y are consistent across all three K_f/K_m ratios, suggesting that remaining fractures predominantly control
 360 the outflow. As the influence of fractures increases with K_f/K_m , the NRMSE further decreases, ignoring different
 361 surrogates' predictive performance.

362 Table 5 presents results for the smaller domain, $\Omega_{UP} = (15, 15, 15)$. The NRMSE values for the coarse model
 363 outflow Y - comparing numerical homogenization and surrogate-based upscaling - reflect trends in the prediction
 364 accuracy of the surrogates on the test data (see Section 5.1). The closest agreement occurs for $K_f/K_m = 1 \times 10^3$, where
 365 Surrogate A also achieved the highest accuracy on the test set. For this smaller domain, the upscaled conductivities
 366 have a more pronounced impact on the outflow Y , resulting in generally worse NRMSE values compared to the larger
 367 domain case $\Omega_{UP} = (60, 60, 60)$.

368 **5.6.2. Anisotropy problem**

369 Second, we compare the equivalent hydraulic conductivity tensors of the coarse DFM models, upscaled by numeri-
 370 cal homogenization and by the surrogates. The equivalent hydraulic conductivity tensor \mathbf{K}_{UP}^{eq} is calculated as described

Table 4

NRMSE comparison of Constraint problem outflow Y obtained from coarse DFM models, evaluated for different fracture-to-matrix conductivity ratios. For each ratio, numerical homogenization vs. surrogate-based upscaling effect is compared. Parameters: $H = 10$, $h = 5$, $\Omega_{\text{UP}} = (60, 60, 60)$, and 250 samples.

K_f/K_m	NRMSE	
	\bar{Y}	Y
Surrogate A	1×10^3	0.00968
Surrogate B	1×10^5	0.00418
Surrogate C	1×10^7	0.00381

Table 5

NRMSE comparison of the coarse Constraint problem outflow Y , evaluated for different fracture-to-matrix conductivity ratios. For each ratio, numerical homogenization vs. surrogate-based upscaling effect is compared. Parameters: $H = 10$, $h = 5$, $\Omega_{\text{UP}} = (15, 15, 15)$, and 250 samples.

K_f/K_m	NRMSE	
	\bar{Y}	Y
Surrogate A	1×10^3	0.01323
Surrogate B	1×10^5	0.06235
Surrogate C	1×10^7	0.06419

Table 6

NRMSE comparison of the components of the equivalent hydraulic conductivity tensor $\mathbf{K}_{\text{UP}}^{\text{eq}}$ obtained from the coarse DFM models for different fracture-to-matrix conductivity ratios. Surrogate-based upscaling is compared against numerical homogenization for each ratio. Parameters: $H = 10$, $h = 5$, domain size $\Omega_{\text{UP}} = (60, 60, 60)$, and 250 samples.

K_f/K_m	NRMSE					
	\mathbf{k}_{xx}	\mathbf{k}_{yy}	\mathbf{k}_{zz}	\mathbf{k}_{yz}	\mathbf{k}_{xz}	\mathbf{k}_{xy}
Surrogate A	1×10^3	0.00133	0.00054	0.00046	0.00057	0.00084
Surrogate B	1×10^5	0.00038	0.00022	0.00030	0.00064	0.00037
Surrogate C	1×10^7	0.00022	0.00022	0.00033	0.00041	0.00022

371 in Section 2.3.2, but for the entire domain Ω_{UP} . As shown in Table 6, the NRMSE values for $\Omega_{\text{UP}} = (60, 60, 60)$
 372 case (see Table) demonstrate that all surrogates provide highly accurate predictions of \mathbf{K}^{eq} , regardless of the value
 373 of $K_f/K_m \in \{1 \times 10^3, 1 \times 10^5, 1 \times 10^7\}$ used. Notably, the overall NRMSE is an order of magnitude lower than
 374 in the Constraint problem. This observation aligns with findings from the 2D case (Špetlík et al. (2024)), where the
 375 Constraint problem exhibits greater sensitivity to matrix conductivity values than the Anisotropy problem.

376 As shown in Table 7, the Anisotropy problem also reveals a decline in the accuracy of the $\mathbf{K}_{\text{UP}}^{\text{eq}}$ components when
 377 the influence of fractures resolved by the coarse model is limited, as in the smaller domain $\Omega_{\text{UP}} = (15, 15, 15)$. In-
 378 terestingly, the accuracy of the off-diagonal components improves with increasing K_f/K_m , which can be attributed
 379 to the decreasing relative importance of surrogate-predicted matrix values as K_f/K_m increases, leading to reduced
 380 sensitivity to prediction errors.

Table 7

NRMSE comparison of the components of the equivalent hydraulic conductivity tensor $\mathbf{K}_{\text{UP}}^{\text{eq}}$ obtained from the coarse DFM models for different fracture-to-matrix conductivity ratios. Surrogate-based upscaling is compared against numerical homogenization for each ratio. Parameters: $H = 10$, $h = 5$, domain size $\Omega_{\text{UP}} = (15, 15, 15)$, and 250 samples.

K_f/K_m	NRMSE					
	\mathbf{k}_{xx}	\mathbf{k}_{yy}	\mathbf{k}_{zz}	\mathbf{k}_{yz}	\mathbf{k}_{xz}	\mathbf{k}_{xy}
Surrogate A	1×10^3	0.01222	0.01372	0.01793	0.03631	0.04027
Surrogate B	1×10^5	0.01483	0.01297	0.01289	0.02076	0.01229
Surrogate C	1×10^7	0.01543	0.01228	0.01247	0.01562	0.01272

381 For both problems presented, we observe that even in cases where the influence of the upscaled conductivity is
 382 significant, the trained surrogates produce results that remain in close agreement with those obtained via standard
 383 numerical homogenization.

384 6. Conclusions

385 This study introduced a deep learning surrogate designed to predict the equivalent hydraulic conductivity tensor
 386 \mathbf{K}^{eq} from 3D discrete fracture-matrix (DFM) models. We evaluated the surrogate's predictive performance as well as
 387 its utility in macro-scale (coarse model) hydrogeological simulations. Three datasets were formed corresponding to
 388 fracture-to-matrix hydraulic conductivity ratios of $K_f/K_m \in \{1 \times 10^3, 1 \times 10^5, 1 \times 10^7\}$. Surrogates trained on these
 389 datasets achieved high predictive accuracy on the test dataset, with $\text{NRMSE} < 0.22$ for all tensor components.

390 We assessed the generalizability of the surrogates across different fracture network configurations and spatial corre-
 391 lation lengths of the matrix random field. Higher fracture densities, beyond those seen in training, resulted in a decline
 392 in performance. When tested on samples from four distinct fracture network configurations, the trained surrogates
 393 exhibited only a slight reduction in predictive accuracy. The impact of changes in spatial correlation length depended
 394 on the value of K_f/K_m : prediction accuracy remained consistent for correlation lengths within the training range but
 395 deteriorated for significantly larger values.

396 The surrogates demonstrated substantial computational advantages across various numbers of homogenization
 397 blocks in use. The upscaling with the surrogate was approximately 100 \times faster than standard numerical homogeniza-
 398 tion when GPU-based inference was employed.

399 To evaluate the impact of surrogate accuracy, we embedded the upscaled \mathbf{K}^{eq} - obtained via numerical homoge-
 400 nization and surrogates - into two macro-scale hydrological problems. In both the Constraint and Anisotropy problems,
 401 the quantities of interest were only marginally affected by surrogate-based upscaling, even in scenarios where the in-
 402 fluence of the upscaled conductivities was substantial. The Constraint problem exhibited greater sensitivity to the bulk
 403 conductivities and, consequently, to the accuracy of the surrogate predictions.

404 Overall, the proposed 3D surrogate models offer a scalable and effective tool for upscaling in fractured porous
405 media, with promising generalizability across a broad range of settings. Future work will focus on integrating these
406 surrogates into multilevel Monte Carlo frameworks, including investigating the impact of the homogenization block
407 size.

408 **7. Acknowledgments**

409 The research was supported by the European Partnership on Radioactive Waste Management 2 (EURAD-2).
410 EURAD-2 is co-funded by the European Union under Grant Agreement N° 101166718. This work was (partly) sup-
411 ported by the Student Grant Scheme at the Technical University of Liberec through project nr. SGS-2025-3560. Com-
412 putational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education,
413 Youth and Sports of the Czech Republic. We acknowledge VSB – Technical University of Ostrava, IT4Innovations
414 National Supercomputing Center, Czech Republic, for awarding this project access to the LUMI supercomputer, owned
415 by the EuroHPC Joint Undertaking, hosted by CSC (Finland) and the LUMI consortium through the Ministry of Edu-
416 cation, Youth and Sports of the Czech Republic through the e-INFRA CZ (grant ID: 90254).

417 **Code availability section** The source codes are available for downloading at the link: https://github.com/martinspetlik/MLMC-DFM/tree/MS_3d

419 **References**

- 420 Adler, P.M., Thovert, J.F., 1999. Fractures and Fracture Networks. Springer Netherlands, Dordrecht. doi:10.1007/978-94-017-1599-7.
- 421 Andrianov, N., 2022. Upscaling of two-phase discrete fracture simulations using a convolutional neural network. Computational Geosciences 26, 1237 – 1259. URL: <https://api.semanticscholar.org/CorpusID:249551231>, doi:<https://doi.org/10.1007/s10596-022-10149-3>.
- 424 Azizmohammadi, S., Matthäi, S.K., 2017. Is the permeability of naturally fractured rocks scale dependent? Water Resources Research 53, 8041–8063. URL: <http://doi.wiley.com/10.1002/2016WR019764>, doi:10.1002/2016WR019764.
- 426 Berrone, S., Pieraccini, S., Scialò, S., 2013. On Simulations of Discrete Fracture Network Flows with an Optimization-Based Extended Finite Element Method. SIAM Journal on Scientific Computing 35, A908–A935. doi:10.1137/120882883.
- 428 Bogdanov, I.I., Mourzenko, V.V., Thovert, J.F., Adler, P.M., 2003. Effective permeability of fractured porous media in steady state flow. Water Resources Research 39. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001WR000756>, doi:<https://doi.org/10.1029/2001WR000756>.
- 431 Bogdanov, I.I., Mourzenko, V.V., Thovert, J.F., Adler, P.M., 2007. Effective permeability of fractured porous media with power-law distribution of fracture sizes. Phys. Rev. E 76, 036309. URL: <https://link.aps.org/doi/10.1103/PhysRevE.76.036309>, doi:10.1103/PhysRevE.76.036309.
- 434 Bonnet, E., Bour, O., Odling, N.E., Davy, P., Main, I., Cowie, P., Berkowitz, B., 2001. Scaling of fracture systems in geological media. Reviews of Geophysics 39, 347–383. doi:<https://doi.org/10.1029/1999RG000074>.
- 436 Březina, J., Stebel, J., 2016. Analysis of model error for a continuum-fracture model of porous media flow, in: Kozubek, T., Blaheta, R., Šístek, J., Rozložník, M., Čermák, M. (Eds.), High Performance Computing in Science and Engineering, Springer International Publishing, Cham. pp. 152–160. doi:10.1007/978-3-319-40361-8_11.
- 439 Březina, J., Stebel, J., Exner, P., Hybš, J., 2011–2025. Flow123d. <http://flow123d.github.com>, repository: <http://github.com/flow123d/flow123d>.
- 441 Březina, J., Kopal, J., Srb, R., Ehlerová, J., Hnídek, J., 2025. bgem: B-spline geometry modeling package. URL: <https://github.com/GeoMop/bgem>.
- 443 Cai, C., Vlassis, N., Magee, L., Ma, R., Xiong, Z., Bahmani, B., Wong, T.F., Wang, Y., Sun, W., 2023. Equivariant geometric learning for digital rock physics: estimating formation factor and effective permeability tensors from morse graph. International Journal for Multiscale Computational Engineering 21, 1–24.
- 446 Ferreira, C.A., Kadeethum, T., Bouklas, N., Nick, H.M., 2022. A framework for upscaling and modelling fluid flow for discrete fractures using conditional generative adversarial networks. Advances in Water Resources 166, 104264. URL: <https://www.sciencedirect.com/science/article/pii/S0309170822001336>, doi:<https://doi.org/10.1016/j.advwatres.2022.104264>.
- 449 Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, Massachusetts. <http://www.deeplearningbook.org>.
- 450 Hadgu, T., Karra, S., Kalinina, E., Makedonska, N., Hyman, J.D., Klise, K., Viswanathan, H.S., Wang, Y., 2017. A comparative study of discrete fracture network and equivalent continuum models for simulating flow and transport in the far field of a hypothetical nuclear waste repository in crystalline host rock. Journal of Hydrology 553, 59–70. doi:10.1016/j.jhydrol.2017.07.046.
- 453 He, X., Santoso, R., Alsinan, M., Kwak, H., Hoteit, H., 2021. Constructing Dual-Porosity Models from High-Resolution Discrete-Fracture Models

- 454 Using Deep Neural Networks, in: Proceedings of the SPE Reservoir Simulation Conference, p. D011S014R012. doi:10.2118/203901-MS.
- 455 Hong, J., Liu, J., 2020. Rapid estimation of permeability from digital rock using 3d convolutional neural network. Computational Geosciences 24, 456 1523–1539. URL: <https://link.springer.com/10.1007/s10596-020-09941-w>. doi:10.1007/s10596-020-09941-w.
- 457 Joyce, S., Simpson, T., Hartley, L., Applegate, D., Hoek, J., Jackson, P., Swan, D., Marsic, N., Follin, S., 2009. Groundwater Flow Modelling of 458 Periods with Temperate Climate Conditions – Forsmark. Technical Report SKB R-09-20. Svensk Kärnbränslehantering AB.
- 459 Kottwitz, M.O., Popov, A.A., Abe, S., Kaus, B.J.P., 2021. Investigating the effects of intersection flow localization in equivalent-continuum-based 460 upscaling of flow in discrete fracture networks. Solid Earth 12, 2235–2254. doi:10.5194/se-12-2235-2021.
- 461 Lang, P.S., Paluszny, A., Zimmerman, R.W., 2014. Permeability tensor of three-dimensional fractured porous rock and a comparison to trace map 462 predictions. Journal of Geophysical Research: Solid Earth 119, 6288–6307. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/2014JB011027>. doi:10.1002/2014JB011027.
- 463 Meng, Y., Jiang, J., Wu, J., Wang, D., 2023. Transformer-based deep learning models for predicting permeability of porous media. Advances 464 in Water Resources 179, 104520. URL: <https://www.sciencedirect.com/science/article/pii/S0309170823001549>, doi:<https://doi.org/10.1016/j.advwatres.2023.104520>.
- 465 Müller, S., Schüler, L., 2019. GSTools. <https://github.com/GeoStat-Framework/GSTools>.
- 466 Pal, M., Makauskas, P., Malik, S., 2023. Upscaling porous media using neural networks: A deep learning approach to homogenization and averaging. 467 Processes 11. URL: <https://www.mdpi.com/2227-9717/11/2/601>, doi:10.3390/pr11020601.
- 468 Rao, C., Liu, Y., 2020. Three-dimensional convolutional neural network (3d-cnn) for heterogeneous material homogenization. Computational 469 Materials Science 184, 109850. URL: <https://www.sciencedirect.com/science/article/pii/S0927025620303414>, doi:<https://doi.org/10.1016/j.commatsci.2020.109850>.
- 470 Rhén, I., Forsmark, T., Hartley, L., Jackson, P., Roberts, D., Swan, D., Gylling, B., 2008. Hydrogeological Conceptualisation and Parameterization, 471 Site Descriptive Modelling SDM-Site Laxemar. Technical Report SKB R-08-78. Svensk Kärnbränslehantering AB.
- 472 Sandve, T.H., Berre, I., Nordbotten, J.M., 2012. An efficient multi-point flux approximation method for Discrete Fracture–Matrix simulations. 473 Journal of Computational Physics 231, 3784–3800. doi:10.1016/j.jcp.2012.01.023.
- 474 Špetlík, M., Březina, J., Laloy, E., 2024. Deep learning surrogate for predicting hydraulic conductivity tensors from stochastic discrete fracture- 475 matrix models. Computational Geosciences URL: <https://link.springer.com/article/10.1007/s10596-024-10324-8>, doi:<https://doi.org/10.1007/s10596-024-10324-8>.
- 476 Stepanov, S., Spiridonov, D., Mai, T., 2023. Prediction of numerical homogenization using deep learning for the richards equation. Journal of Compu- 477 tational and Applied Mathematics 424, 114980. URL: <https://www.sciencedirect.com/science/article/pii/S0377042722005787>, doi:<https://doi.org/10.1016/j.cam.2022.114980>.
- 478 Vasilyeva, M., Tyrylgin, A., 2021. Machine learning for accelerating macroscopic parameters prediction for poroelasticity problem in stochastic 479 media. Computers & Mathematics with Applications 84, 185–202. URL: <https://www.sciencedirect.com/science/article/pii/S0898122121000067>, doi:<https://doi.org/10.1016/j.camwa.2020.09.024>.
- 480 Wang, N., Liao, Q., Chang, H., Zhang, D., 2023. Deep-learning-based upscaling method for geologic models via theory-guided convolutional neural 481 network. Computational Geosciences 27, 913–938. URL: <https://link.springer.com/10.1007/s10596-023-10233-2>, doi:10.1007/s10596-023-10233-2.