

# SIMSHIFT: Technical Supplementary Material

## 1 Introduction

This document provides the technical supplementary material for the SIMSHIFT framework. Given that our work lies at the intersection of the machine learning research community and industrial engineering practice, we include this additional information to enhance accessibility, clarity and practical adoption by both communities.

Section 2 offers detailed explanations of the numerical simulations used to generate each dataset. This section is intended to help machine learning researchers better understand the physical models, assumptions and limitations underlying our datasets.

In Section 3, we provide additional qualitative evaluations through visual comparisons between model predictions and numerical ground truth solutions. These visualizations are particularly valuable for industrial practitioners aiming to assess model performance and reliability under distribution shifts.

## 2 Dataset Generation Details

This section offers comprehensive details on the numerical simulation processes for our datasets. We outline the assumptions, approximations, software tools and parameters used for data generation. Additionally, we report computational resources and simulation runtimes to increase transparency.

### 2.1 Hot Rolling

The *hot rolling* dataset represents a hot rolling process in which a metal slab undergoes plastic deformation to form a sheet metal product. The model considers a plane-strain representation of a heated steel slab segment with a core temperature  $T_{\text{core}}$  and a surface temperature  $T_{\text{surf}}$ , initially at thickness  $t$ , passing through a simplified roll stand with a nominal roll gap  $g$  (see Figure 1). This roll gap effectively matches the exit thickness of the workpiece. Given the material properties, the initial temperature distribution over the slab thickness and the specified pass reduction, the model aims to capture the evolution of the thermo-mechanical state of the workpiece as it traverses the roll gap.

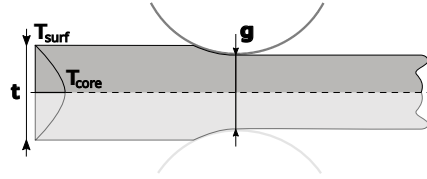


Figure 1: Illustration of the *hot rolling* simulation setup. The parameters correspond to those in Table 1, with the reduction being defined as:  $\text{reduction} = \frac{t-g}{t}$ .

To reduce computational complexity, the analysis is confined to the vertical midplane along the rolling direction based on a plane-strain assumption. This is well justified by the high width-to-thickness ratio characteristic of the workpiece. Additionally, vertical symmetry is also exploited. Consequently, only the upper half of the workpiece and the upper work roll are modeled.

The workpiece is discretized using plane-strain, reduced-integration, quadrilateral elements. Mesh generation is fully automated, with the element size calibrated according to findings from a mesh convergence study. In terms of mechanical behavior, the workpiece is modeled as elasto-plastic with isotropic hardening, employing tabulated flow curves representative for a titanium alloy [1, 2]. The elastic modulus and flow stress are temperature dependent, with the latter also influenced by the plastic strain rate. In contrast, material density and Poisson's ratio are assumed to remain constant. The work roll with a diameter of 1000 mm is idealized as an analytically defined rigid body.

In addition to the mechanical behavior, the elements also feature a temperature degree of freedom that captures thermal phenomena, which are in turn fully coupled with the mechanical field. Heat conduction within the workpiece is governed by temperature dependent thermal conductivity and specific heat capacity. Heat transfer at the interface between the workpiece and the roll is modeled

as proportional to the temperature difference between the contacting surfaces, using a heat transfer coefficient of  $5 \text{ mW/mm}^2\text{K}$ . The model also accounts for internal heat generation due to plastic deformation, based on the standard assumption that 90% of plastic work is converted into heat. Additionally, all frictional energy is assumed to be fully transformed into heat and evenly divided between the workpiece and the roll. However, since the analysis focuses on the workpiece, only the portion of this heat entering the workpiece is considered.

The Finite Element (FE) simulation is performed with the *Abaqus*<sup>1</sup> explicit solver using a relatively high mass scaling factor of 100. This mass scaling proved to be a suitable choice for maintaining both computational efficiency and solution accuracy. The pre-processing, evaluation and post-processing of the simulations was automated in Python. A full factorial design of experiments was conducted by varying the parameters outlined in Table 1. Simulation outputs from Abaqus (.odb files) were converted to a more suitable .h5 format in post-processing, enabling seamless integration into the SIMSHIFT framework. All simulations were run on a Gigabyte Aorus 15P KD consumer laptop equipped with an Intel Core i7-11800H CPU (8 cores, 16 threads, 2.30–4.60 GHz), 16 GB DDR4 RAM at 3200 MHz and a 1 TB NVMe SSD. The single-core CPU time for one simulation was 25 seconds on average, depending on the mesh size and convergence speed.

Table 1: Input parameter ranges for the *hot rolling* simulations. Samples are generated by equally spacing each parameter within the specified range using the indicated number of steps, resulting in  $5 \times 19 \times 10 \times 5 = 4750$  total samples.

Parameter	Description	Min	Max	Steps
$t$ (mm)	Initial slab thickness.	50.0	183.3	5
reduction (–)	Reduction of initial slab thickness.	0.01	0.15	19
$T_{\text{core}}$ ( $^{\circ}\text{C}$ )	Core slab temperature.	900.0	1000.0	10
$T_{\text{surf}}$ ( $^{\circ}\text{C}$ )	Surface slab temperature.	900.0	1077.77	5

## 2.2 Sheet Metal Forming

For the *sheet metal forming* dataset, a w-shaped bending process was selected due to its complex contact interactions and the highly nonlinear progression of bending forces. For this purpose, a parameterized 2D FE model of the process was developed using the commercial Finite Element Method (FEM) software *Abaqus* and its implicit solver, with the simulation pipeline implemented in Python. The initial configuration of the finite element model is shown in Figure 2 and described below.

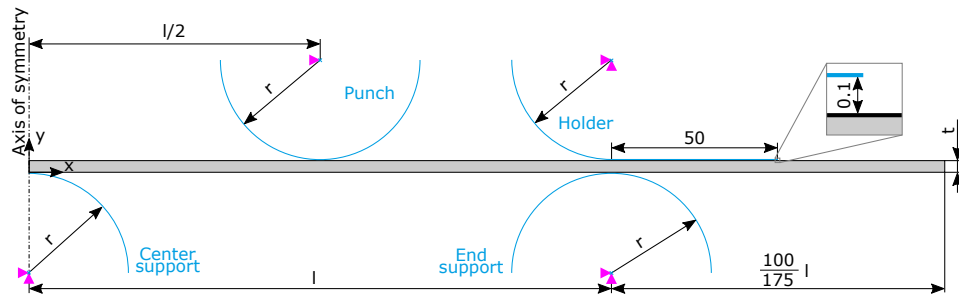


Figure 2: Abstraction of bending process layout used for FE calculations in its initial configuration.

Due to geometric and loading symmetry, only the right half of the sheet with a thickness  $t$  was modeled. The die and punch were idealized as rigid circular segments with a shared radius  $r$ . Additionally, a rigid blank holder comprising an arc and a straight segment was positioned 0.1 mm above the sheet to maintain contact and restrain vertical motion. The required sheet length was determined by the support span  $l$ , enabling material flow toward the center in response to the downward motion of the punch.

<sup>1</sup><https://www.3ds.com/products/simulia/abaqus>

The sheet was discretized using bilinear, plane-strain quadrilateral elements with reduced integration and hourglass control (Abaqus element type CPE4R). A prior mesh convergence study indicated that accurate simulation results require a minimum of 10 element rows across the sheet thickness. The element size was fixed at  $0.125 \times 0.1$  mm to ensure a uniform aspect ratio, constraining the sheet thickness to  $t > 1$  mm.

The sheet material was modeled as elastoplastic with von Mises plasticity and linear isotropic hardening. The following properties were assigned: Young’s modulus of 210 GPa, Poisson’s ratio of 0.3, yield stress of 410 MPa, and hardening modulus of 2268 MPa.

For all contact interfaces, a normal contact formulation with surface-to-surface discretization, penalty enforcement, and finite-sliding tracking was employed. Tangential contact was modeled via a Coulomb friction law with a coefficient  $\mu$ .

The supports and blank holder were fixed by constraining horizontal and vertical translations as well as in-plane rotations. These constraints were applied at the centroid of each arc segment, representing the reference point for the respective rigid body. The punch was similarly constrained against horizontal movement and rotation but retained vertical mobility. The deformed configuration following a vertical displacement  $U$  of the punch is illustrated in Figure 3.

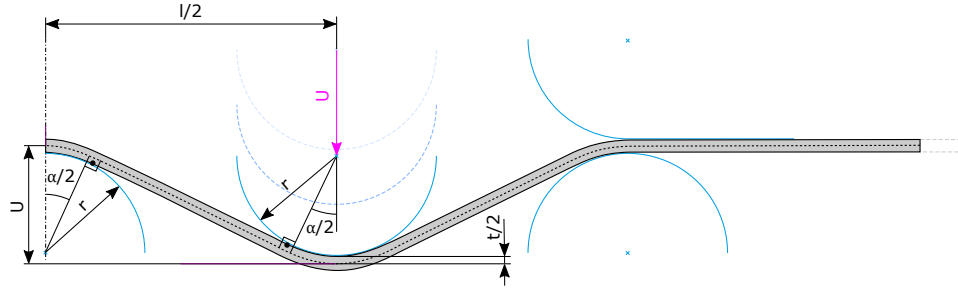


Figure 3: Bending process abstraction, deformed configuration.

A full factorial design of experiments was conducted by varying the parameters outlined in Table 2. As for the hot rolling simulations, outputs from Abaqus (.odb files) were converted to .h5 format in post-processing, to integrate them into the SIMSHIFT framework. All simulations were run on a Gigabyte Aorus 15P KD consumer laptop equipped with an Intel Core i7-11800H CPU (8 cores, 16 threads, 2.30–4.60 GHz), 16 GB DDR4 RAM at 3200 MHz and a 1 TB NVMe SSD. The single-core CPU time for one simulation run was 300 seconds on average, depending on mesh size and convergence speed.

Table 2: Input parameter ranges for the *sheet metal forming* simulations. Samples are generated by equally spacing each parameter within the specified range using the indicated number of steps, resulting in  $17 \times 13 \times 3 \times 5 = 3315$  total samples.

Parameter	Description	Min	Max	Steps
$r$ (mm)	Roll radius.	10.0	50.0	17
$t$ (mm)	Sheet thickness.	2.0	5.0	13
$l$ (mm)	Sheet length.	175.0	350.0	3
$\mu$ (—)	Friction coefficient between the sheet and the rolls.	0.1	0.5	5

## 2.3 Electric Motor Design

The *electric motor design* dataset includes a structural FE simulation of a rotor within electric machinery, subjected to mechanical loading at burst speed. The rotor topology is modeled after the motor architecture of the 2010 Toyota Prius [3], an industry-recognized benchmark frequently used for validation and comparison in academic and industrial research. The Prius rotor topology is based on a V-shaped magnet configuration as shown in Figure 4.

98 Structural rotor simulations are essential in multi-physics design optimization, where motor per-  
 99 formance is evaluated across multiple domains including electromagnetic, thermal, acoustic, and  
 100 structural. Using a design optimization framework, stator and rotor design are iteratively refined to  
 101 identify Pareto-optimal solutions based on objectives such as efficiency, torque, weight, and speed.  
 102 In this process, the structural FE model predicts stress and deformation due to loading ensuring the  
 103 rotor’s structural integrity.

104 The set up and execution of the structural simulations for this dataset are automated and implemented  
 105 in the open source design optimization framework *SyMSpace*<sup>2</sup>. The FE simulation of the rotor is  
 106 performed using a mixed 2D plane stress and plane strain formulation with triangular elements. To  
 107 enhance computational efficiency, geometric symmetry is exploited and only a 1/16 sector of the full  
 108 rotor is modeled. The mechanical simulation is static and evaluates the rotor under centrifugal loading,  
 109 incorporating press-fit conditions between the rotor core and shaft, as well as contact interactions  
 110 between the rotor core and embedded magnets.

111 An elastic material behavior is employed for all components, including the rotor core, shaft, and  
 112 magnets. Material properties are summarized in Table 3. Based on the parametrized CAD model of  
 113 the rotor topology, the geometry is automatically meshed using *Netgen*<sup>3</sup>. The design optimization  
 114 tool also automatically identifies nodes for boundary conditions and contact surfaces and applies  
 115 the corresponding constraints and interactions required for the simulation. The implicit FE solver  
 116 *HOTINT*<sup>4</sup> is used to compute the quasi-static response of the system, providing local stress and strain  
 117 fields across the rotor topology.

Table 3: Material parameters for the structural *electric motor design* simulations.

	Rotor Core	Rotor Shaft	Permanent Magnet
Material	NO27-14 Y420HP	42CrMo4	BMN-40SH
Density (kg/dm <sup>3</sup> )	7.6	7.72	7.55
Possions ratio (-)	0.29	0.3	0.24
Young’s Modulus (kN/mm <sup>2</sup> )	185.0	210.0	175.0
Tensile Strength (kN/mm <sup>2</sup> )	550.0	850.0	250.0

118 To generate the electric motor dataset, a comprehensive motor optimization study was conducted  
 119 using SyMSpace, based on design specifications of the 2010 Toyota Prius. The optimization aimed  
 120 to minimize multiple performance metrics, including motor mass, material costs, rotor torque  
 121 ripple, motor losses, coil temperature, stator terminal current, and elastic rotor deformation. A  
 122 genetic algorithm was employed to explore the design space and identify Pareto-optimal solutions.  
 123 In the process, 3,196 motor configurations were evaluated by varying, among other factors, the  
 124 rotor’s topological parameters within the bounds specified in Table 4. The outputs of the structural  
 125 simulations were generated in .vtk format and then stored in .h5 files, allowing direct integration  
 126 into the SIMSHIFT framework. Each structural simulation required approximately 4 to 5 minutes of  
 127 single-core CPU time on a Intel Core i9-14900KS processor (24 Cores, 3200 MHz), depending on  
 128 convergence speed of the contact algorithm.

<sup>2</sup><https://symspace.lcm.at/>

<sup>3</sup><https://ngsolve.org/>

<sup>4</sup><https://hotint.lcm.at/>



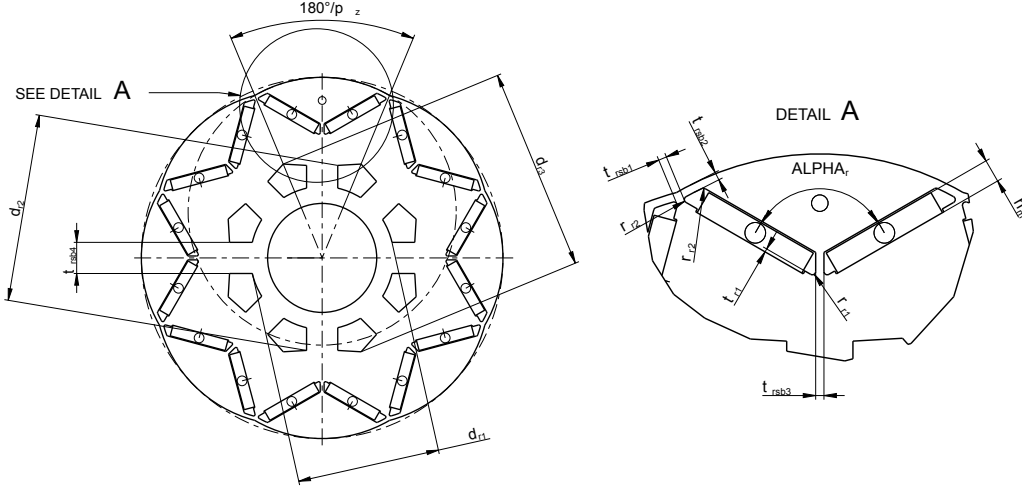


Figure 4: Technical drawing of the electrical motor.

Table 4: Input parameters for the *electric motor design* simulations. Since the design space was explored by a genetic algorithm, the parameters are not uniformly sampled as in the previous simulation scenarios. In total, 3196 simulations were performed.

Parameter	Description	Min	Max
$d_{si}$ (mm)	Stator inner diameter.	150.0	180.0
$h_m$ (mm)	Magnet height.	6.0	9.0
$\alpha_r$ ( $^\circ$ )	Angle between magnets.	120.0	160.0
$t_{r1}$ (mm)	Magnet step.	1.0	5.0
$r_{r1}$ (mm)	Rotor slot fillet radius 1.	0.5	2.5
$r_{r2}$ (mm)	Rotor slot fillet radius 2.	0.5	3.5
$r_{r3}$ (mm)	Rotor slot fillet radius 3.	0.5	5.0
$r_{r4}$ (mm)	Rotor slot fillet radius 4.	0.5	3.0
$t_{rsb1}$ (mm)	Thickness saturation bar 1.	4.0	12.0
$t_{rsb2}$ (mm)	Thickness saturation bar 2.	1.0	3.0
$t_{rsb3}$ (mm)	Thickness saturation bar 3.	1.2	4.0
$t_{rsb4}$ (mm)	Thickness saturation bar 4.	5.0	12.0
$d_{r1}$ (mm)	Rotor slot diameter 1.	60.0	80.0
$d_{r2}$ (mm)	Rotor slot diameter 2.	80.0	120.0
$d_{r3}$ (mm)	Rotor slot diameter 3.	100.0	125.0

## 2.4 Heatsink Design

The *heatsink design* dataset consists of heatsink geometries similar to the example shown in Figure 5, placed centrally at the bottom of a surrounding box-shaped domain filled with air. The dimensions of the surrounding enclosure are 0.14 m  $\times$  0.14 m  $\times$  0.5 m (length  $\times$  width  $\times$  height).

The geometric configuration of each heatsink is defined by several parameters, which were varied within specified bounds for the design study. These parameters and their corresponding value ranges are summarized in Table 5. A total of 460 simulation cases were generated, with non-uniform sampling across the parameter space.

The dataset was generated using Computational Fluid Dynamics (CFD) simulations based on the Reynolds-Averaged Navier-Stokes (RANS) equations coupled with the energy equation. All simulations were conducted in the open-source CFD suite OpenFOAM <sup>5</sup>.

<sup>5</sup><https://openfoam.org/>

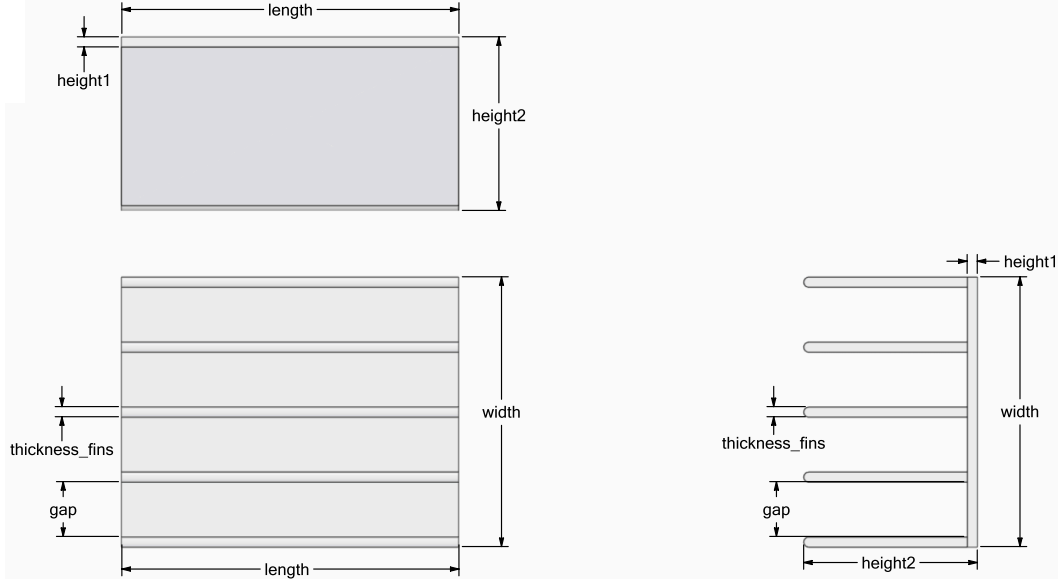


Figure 5: Technical drawing of the solid body in the *heatsink design* dataset. Some of the shown parameters are varied for data generation (see Table 5).

Table 5: Geometric and physical parameters of the *heatsink design* simulations. The variable parameters were not uniformly sampled. In total, 460 simulations were performed.

Parameter	Description	fixed Value	Min	Max
length (m)	Heatsink length	0.1	-	-
width (m)	Heatsink width	0.08	-	-
height1 (m)	Baseplate height	0.003	-	-
T(amb) (K)	Ambient Temperature	300	-	-
fins (-)	Number of fins	-	5	14
gap (m)	Gap between fins	-	0.0023	0.01625
thickness_fins (m)	Thickness of fins	-	0.003	0.004
height2 (m)	Heatsink height	-	0.053	0.083
T (solid) (K)	Temperature of the solid fins	-	340	400

140 The computational domain was discretized using a finite volume method with second-order spatial  
141 discretization schemes. A structured hexahedral background mesh was generated with the blockMesh  
142 utility in OpenFOAM, followed by mesh refinement using snappyHexMesh to accurately resolve the  
143 heatsink structure defined in STL format.

144 To simulate buoyancy-driven natural convection, the buoyantSimpleFoam solver was employed.  
145 This solver is designed for steady-state, compressible, buoyant flows, using the SIMPLE algorithm  
146 for pressure-momentum coupling, extended with under-relaxation techniques to enhance numerical  
147 stability and robust convergence.

148 Boundary conditions were applied as follows:

- 149 • Walls of the surrounding: no-slip velocity condition with fixed ambient temperature as  
150 defined in Table 5.
- 151 • Walls of the heatsink: no-slip velocity condition with solid temperature within the range  
152 specified for parameter T (solid) in Table 5.

153 Given the turbulent nature of the flow, the RANS equations were closed using the SST  $k-\omega$  turbulence  
154 model [4]. Near-wall regions were modeled using a  $y^+$ -insensitive near-wall treatment, allowing  
155 accurate resolution of boundary layers without the need for excessively fine meshes.

156 A mesh convergence study was conducted to ensure numerical accuracy. Depending on mesh  
 157 resolution, each simulation required approximately 11 to 18 hours of single-core CPU time on an  
 158 Intel Core i9-14900KS processor (24 cores, 3.2 GHz).

### 159 3 Visual Evaluation

160 While the presented quantitative metrics in the main paper offer a high level summary of model  
 161 performance, industry practitioners often need a more fine grained picture to assess the neural  
 162 surrogate’s capabilities under distribution shifts. To address this, we include additional visualizations  
 163 that qualitatively assess model performance.

164 First, we provide error distribution histograms to better illustrate the difficulty of the domain shift  
 165 occurring in each dataset. Additionally, we present fringe and scatter plots comparing model  
 166 predictions with the respective ground truth numerical solutions. For each dataset, we display  
 167 example predictions from the best performing model, selected based on average Normalized Root  
 168 Mean Squared Error (NRMSE) on the target domain, alongside their corresponding ground truth  
 169 fields.

#### 170 3.1 Hot Rolling

171 Figure 6 highlights the substantial distribution shift between the source and target domain of the *hot*  
 172 *rolling* dataset. Errors in the target domain are noticeably larger, almost up to an order of magnitude  
 173 higher than those observed in the source domain.

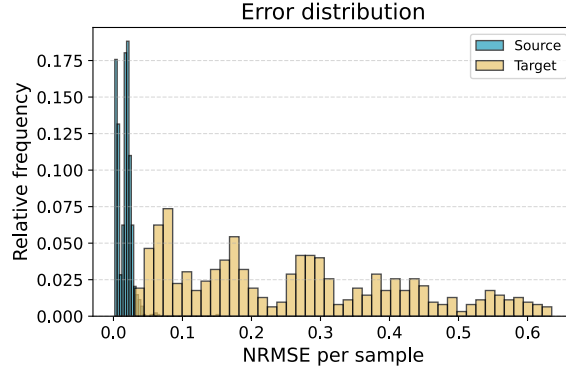


Figure 6: Distribution of NRMSE (averaged across all fields) for the test sets of the source (blue) and target (yellow) domains in the *hot rolling* dataset. Bar height indicates the relative frequency of samples within each bin.

174 To further illustrate the model’s performance, we analyze two representative samples, one from the  
 175 source and one from the target domain. Since the most critical field for downstream applications is  
 176 the Equivalent Plastic Strain (PEEQ), we restrict our analysis on this scalar field only.

177 Table 6 presents a summary of the absolute PEEQ prediction errors for the selected source and target  
 178 samples. Additionally, Figure 7 and Figure 8 visualize the ground truth, predictions, and absolute  
 179 errors for these samples using fringe plots.

Table 6: Absolute error of PEEQ predictions for representative samples from the source and target domain of the *hot rolling* dataset.

Metric	Source	Target
Mean	<b>2.07e-04</b>	1.46e-02
Std	<b>1.87e-04</b>	2.73e-03
Median	<b>1.66e-04</b>	1.49e-02
Q <sub>01</sub>	<b>7.45e-09</b>	6.82e-03
Q <sub>25</sub>	<b>6.38e-05</b>	1.37e-02
Q <sub>75</sub>	<b>2.99e-04</b>	1.58e-02
Q <sub>99</sub>	<b>7.61e-04</b>	2.14e-02

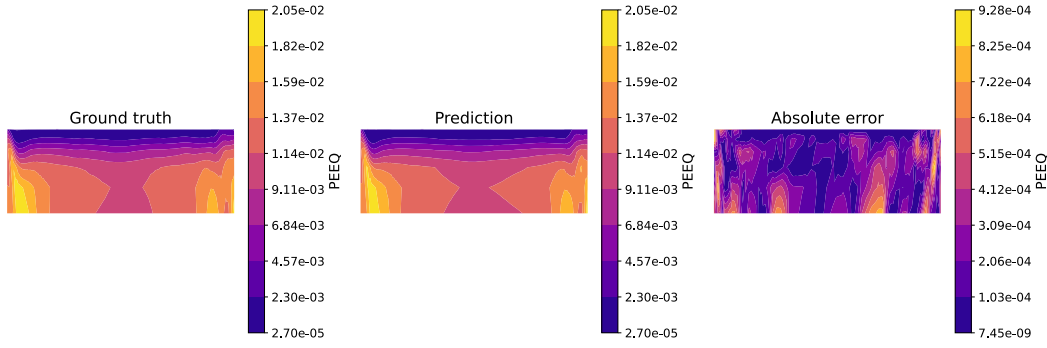


Figure 7: Fringe plot of the *hot rolling* dataset (representative source sample). Shown is the ground truth (left) and predicted (middle) PEEQ, as well as the absolute error (right).

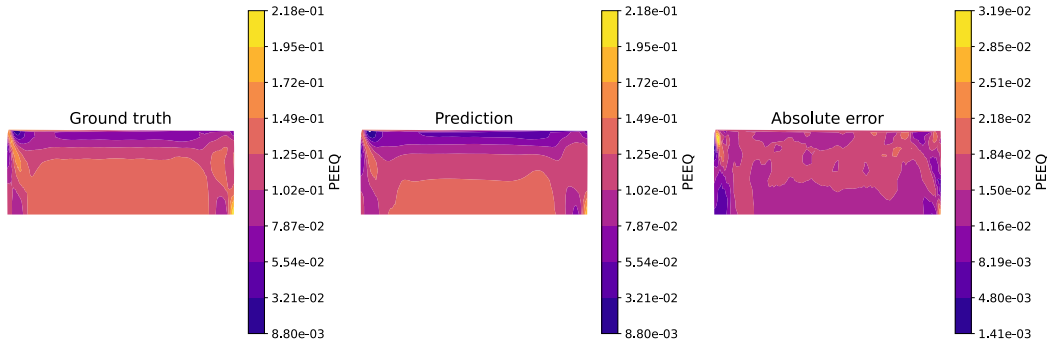


Figure 8: Fringe plot of the *hot rolling* dataset (representative target sample). Shown is the ground truth (left) and predicted (middle) PEEQ, as well as the absolute error (right).

### 3.2 Sheet Metal Forming

Figure 9 shows a moderate distribution shift between the source and target domain of the *sheet metal forming* dataset. The overall increase in error from the source to the target domain is less grave than in the *hot rolling* scenario.

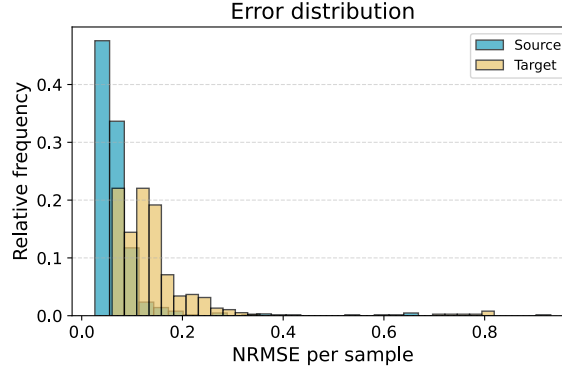


Figure 9: Distribution of NRMSE (averaged across all fields) for the test sets of the source (blue) and target (yellow) domains in the *sheet metal forming* dataset. Bar height indicates the relative frequency of samples within each bin.

To better understand the model’s predictive behavior in this setting, we examine the best and worst performing samples from both the source and target domains. As with the previous case, we focus our analysis on the PEEQ field due to its relevance in downstream applications and workflows.

Table 7 provides a statistical summary of the absolute PEEQ prediction errors across the selected cases. The corresponding fringe plots are shown in Figures 10 to 13, comparing ground truth and predicted fields along with the resulting absolute errors. These visualizations emphasize that while the best prediction in the target domain remains reasonably accurate, others (e.g., the worst case sample) exhibit notable discrepancies in the localized regions around the bends that we are most interested in.

Table 7: Absolute error of PEEQ predictions for the best and worst samples from the source and target domain of the *sheet metal forming* dataset.

Metric	Source		Target	
	Best	Worst	Best	Worst
Mean	<b>5.47e-05</b>	1.68e-04	1.72e-04	1.86e-03
Std	<b>1.35e-04</b>	3.80e-04	4.52e-04	5.97e-03
Median	<b>1.96e-05</b>	4.35e-05	5.58e-05	2.85e-04
Q <sub>01</sub>	<b>2.43e-07</b>	6.52e-07	1.03e-06	6.66e-06
Q <sub>25</sub>	<b>7.49e-06</b>	1.84e-05	2.60e-05	1.53e-04
Q <sub>75</sub>	<b>4.05e-05</b>	1.05e-04	9.10e-05	4.67e-04
Q <sub>99</sub>	<b>7.37e-04</b>	1.96e-03	2.56e-03	3.40e-02

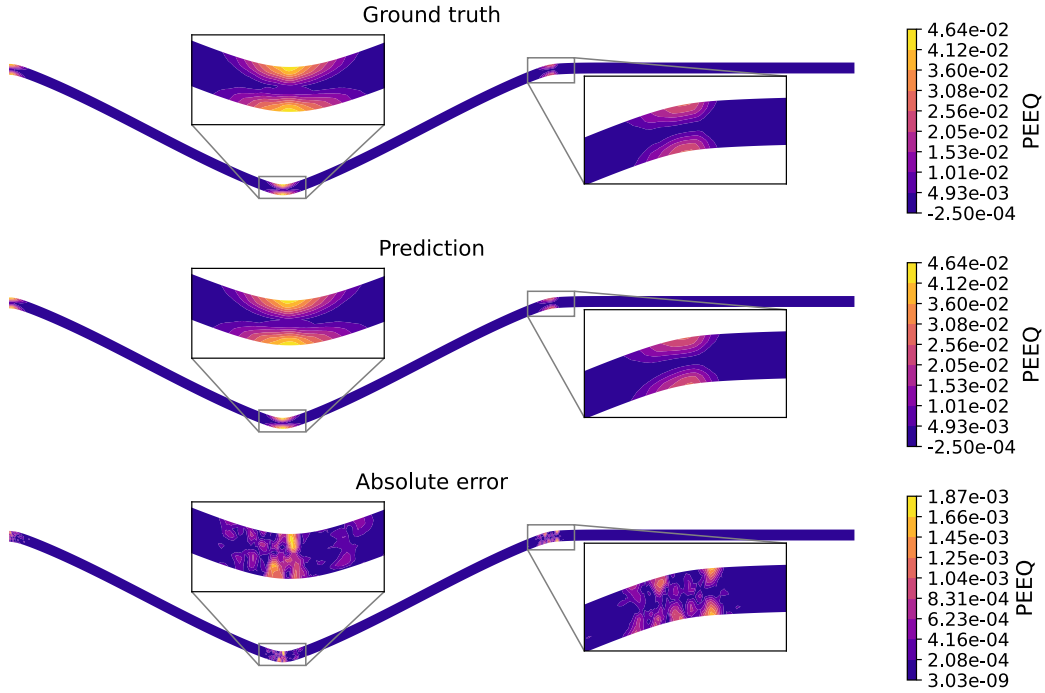


Figure 10: Fringe plot of the *sheet metal forming* dataset (best source sample). Shown is the ground truth (top) and predicted (middle) PEEQ, as well as the absolute error (bottom).

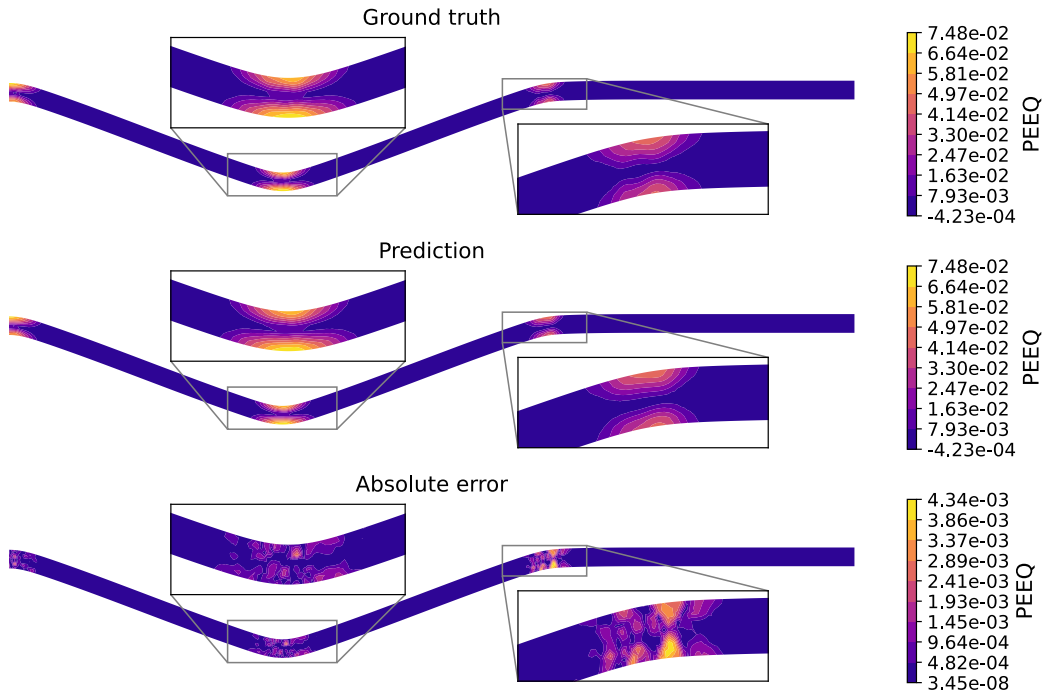


Figure 11: Fringe plot of the *sheet metal forming* dataset (worst source sample). Shown is the ground truth (top) and predicted (middle) PEEQ, as well as the absolute error (bottom).

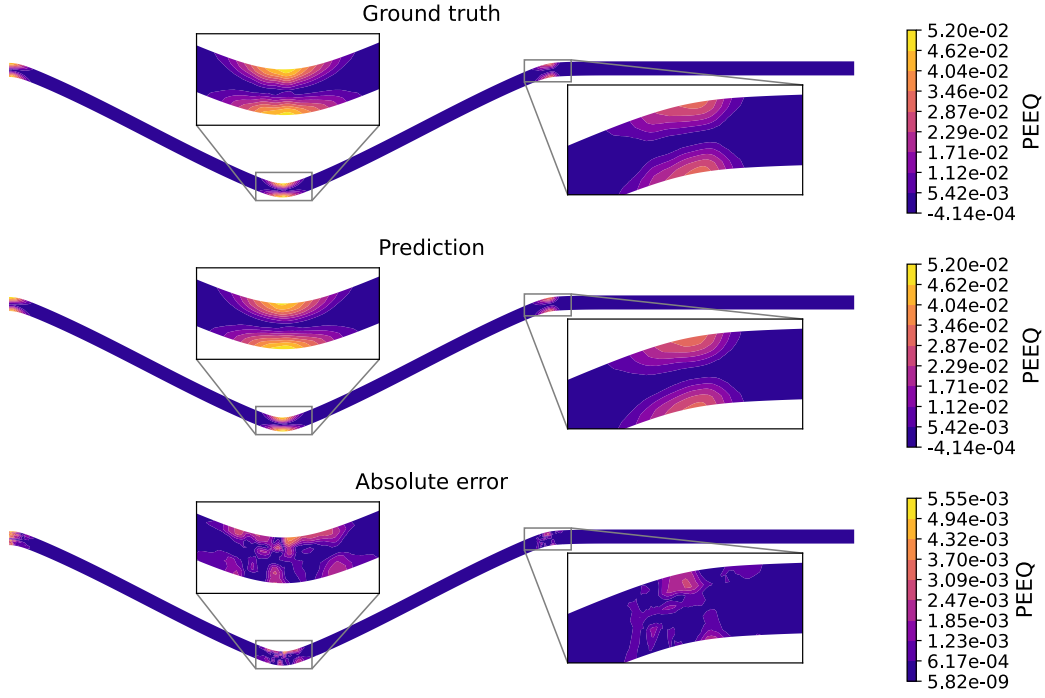


Figure 12: Fringe plot of the *sheet metal forming* dataset (best target sample). Shown is the ground truth (top) and predicted (middle) PEEQ, as well as the absolute error (bottom).

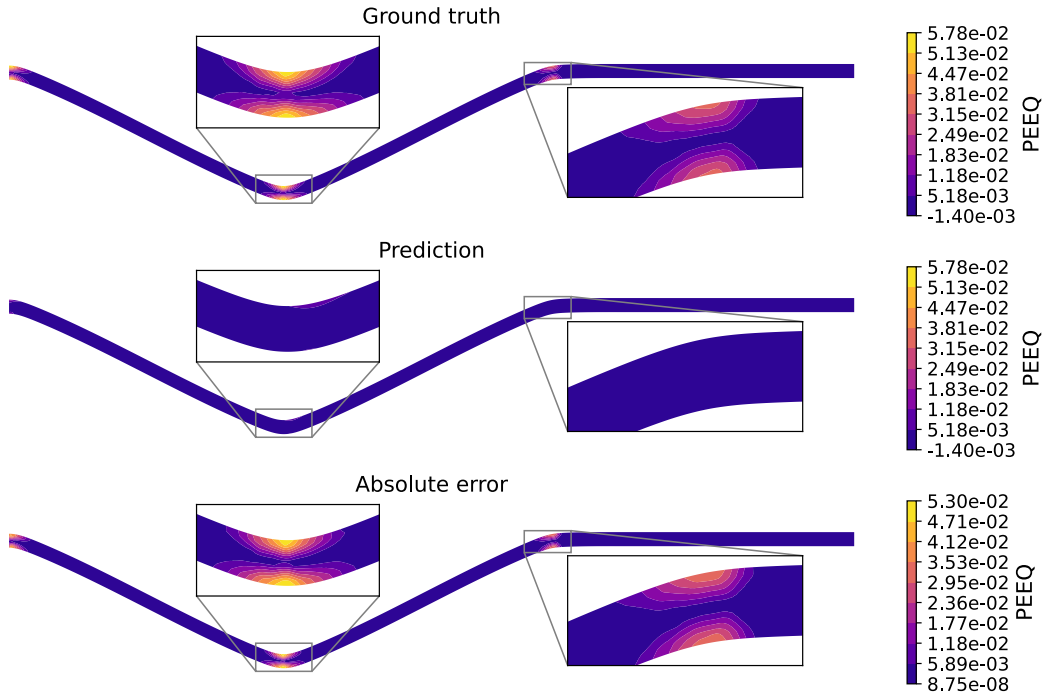


Figure 13: Fringe plot of the *sheet metal forming* dataset (worst target sample). Shown is the ground truth (top) and predicted (middle) PEEQ, as well as the absolute error (bottom).



### 3.3 Electric Motor Design

To assess the effect of the domain shift on prediction accuracy in the *electric motor design* dataset, Figure 14 shows the distribution of NRMSEs in the source and target domain. Compared to the *hot rolling* task, the increase in error when moving from the source to the target domain is less substantial.

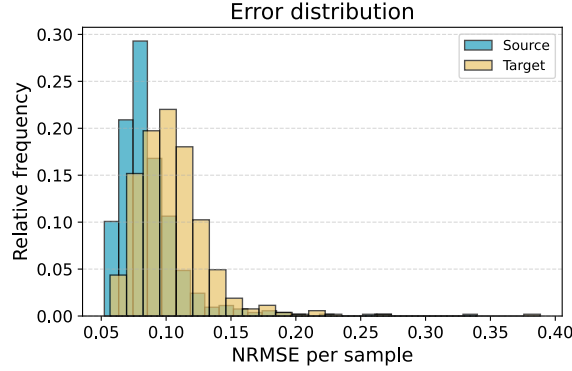


Figure 14: Distribution of NRMSE (averaged across all fields) for the test sets of the source (blue) and target (yellow) domains in the *electric motor design* dataset. Bar height indicates the relative frequency of samples within each bin.

In this task, the Mises stress is used as a scalar summary of the multi-axial stress state and is particularly interesting for downstream analysis and optimization. We therefore focus our closer inspection on this field.

Table 8 presents a comparison of absolute Mises stress errors for the best and worst samples from both the source and target test sets. The corresponding fringe plots are shown in Figures 15 to 18, comparing the ground truth and predicted fields alongside their absolute errors. They show that the best samples are predicted very well, whereas the worst sample of the source domain visually appears slightly worse than the one of the target domain. On average, however, it is still predicted more accurately than the worst sample of the target domain, as shown in Table 8.

Table 8: Absolute error (MPa) of Mises stress predictions for the best and worst samples from the source and target domain of the *electric motor design* dataset.

Metric	Source		Target	
	Best	Worst	Best	Worst
Mean	<b>2.00</b>	20.50	2.67	23.21
Std	<b>2.73</b>	39.50	3.13	23.32
Median	<b>1.26</b>	13.63	1.68	13.09
Q <sub>01</sub>	<b>0.02</b>	0.18	0.03	0.11
Q <sub>25</sub>	<b>0.60</b>	5.08	0.75	2.87
Q <sub>75</sub>	<b>2.24</b>	24.05	3.50	41.97
Q <sub>99</sub>	<b>13.43</b>	140.17	15.05	78.28

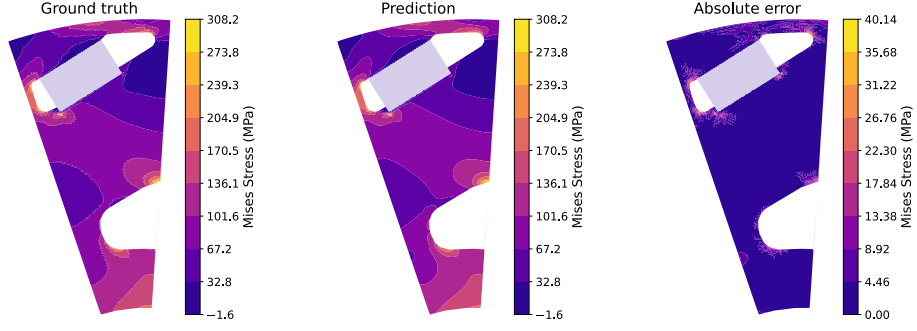


Figure 15: Fringe plot of the *electric motor design* dataset (best source sample). Shown is the ground truth (left) and predicted (middle) Mises stress, as well as the absolute error (right).

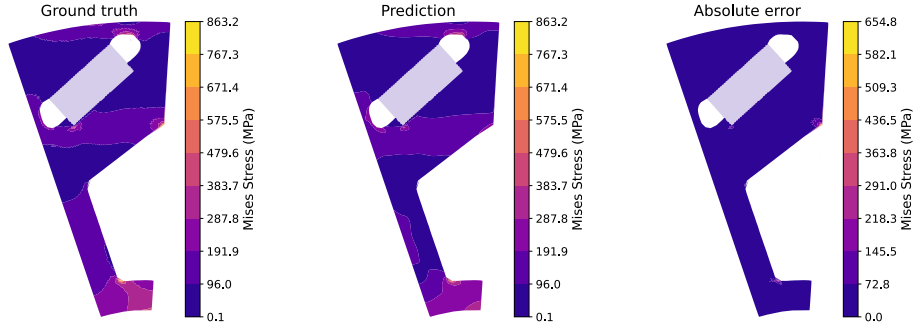


Figure 16: Fringe plot of the *electric motor design* dataset (worst source sample). Shown is the ground truth (left) and predicted (middle) Mises stress, as well as the absolute error (right).

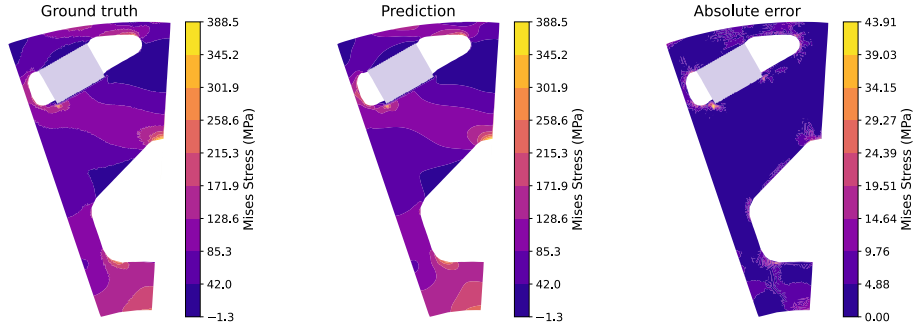


Figure 17: Fringe plot of the *electric motor design* dataset (best target sample). Shown is the ground truth (left) and predicted (middle) Mises stress, as well as the absolute error (right).

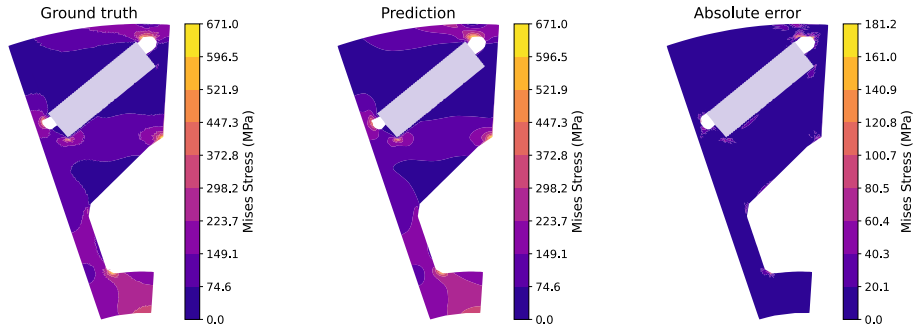


Figure 18: Fringe plot of the *electric motor design* dataset (worst target sample). Shown is the ground truth (left) and predicted (middle) Mises stress, as well as the absolute error (right).

### 3.4 Heatsink Design

Figure 19 presents the distribution of prediction errors in the source and target domain for the *heatsink design* dataset. Again, the effect of the distribution shift between the source and target domain on the prediction errors is clearly visible as errors in the target domain are noticeably higher.

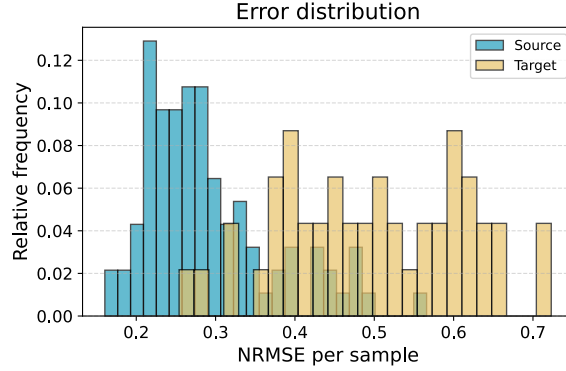


Figure 19: Distribution of NRMSE (averaged across all fields) for the test sets of the source (blue) and target (yellow) domains in the *heatsink design* dataset. Bar height indicates the relative frequency of samples within each bin.

In this task, the temperature field is the most critical for downstream analysis and optimization, which is why we focus our detailed analysis on this field.

Table 9 compares the absolute temperature prediction errors for the best and worst samples from both the source and target test sets. The corresponding scatter plots are shown in Figures 20 to 23, comparing the ground truth and predicted temperature fields, alongside their absolute errors.

While the best source domain prediction is quite accurate, with low average and percentile errors (Table 9, Figure 20), the 99th percentile of the worst source domain prediction reaches up to 29K. Given a total temperature range of 100K, this represents a relative error of nearly 30%. The worst target domain prediction is even less accurate, showing substantial visual and quantitative deviations from the ground truth (Table 9, Figure 23).

Table 9: Absolute error (K) of temperature predictions for the best and worst samples in the source and target domain of the *heatsink design* dataset.

Metric	Source		Target	
	Best	Worst	Best	Worst
Mean	<b>1.84e+00</b>	5.79e+00	2.23e+00	1.42e+01
Std	<b>1.94e+00</b>	5.90e+00	2.85e+00	1.46e+01
Median	<b>1.25e+00</b>	4.06e+00	1.31e+00	8.84e+00
Q <sub>01</sub>	<b>2.17e-02</b>	7.51e-02	2.41e-02	1.62e-01
Q <sub>25</sub>	<b>5.49e-01</b>	1.92e+00	5.95e-01	4.49e+00
Q <sub>75</sub>	<b>2.44e+00</b>	7.48e+00	2.68e+00	1.87e+01
Q <sub>99</sub>	<b>9.26e+00</b>	2.88e+01	1.49e+01	6.61e+01

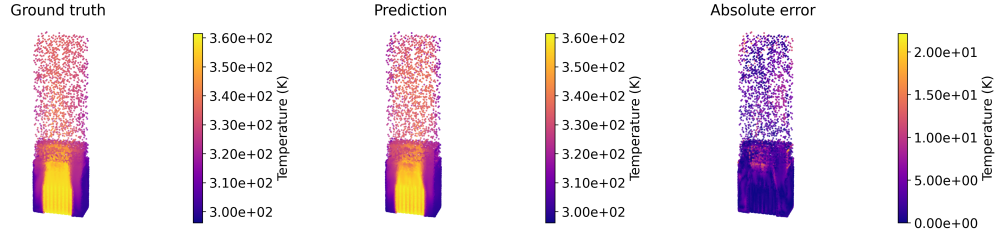


Figure 20: Sliced scatter plot of the *electric motor design* dataset (best source sample). Shown is the ground truth (left) and predicted (middle) temperature field, as well as the absolute error (right).

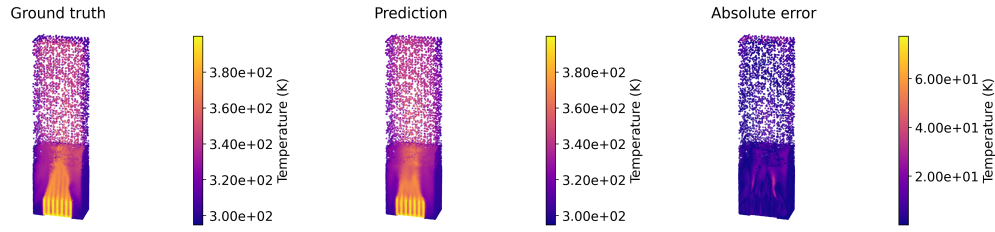


Figure 21: Sliced scatter plot of the *electric motor design* dataset (worst source sample). Shown is the ground truth (left) and predicted (middle) temperature field, as well as the absolute error (right).

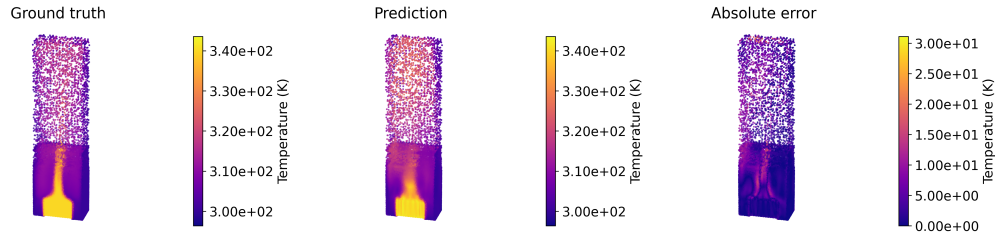


Figure 22: Sliced scatter plot of the *electric motor design* dataset (best target sample). Shown is the ground truth (left) and predicted (middle) temperature field, as well as the absolute error (right).

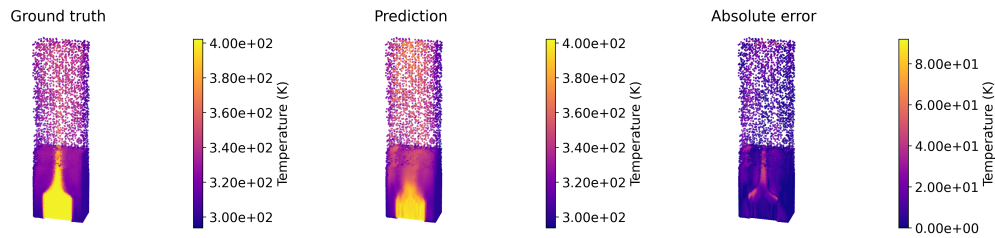


Figure 23: Sliced scatter plot of the *electric motor design* dataset (worst target sample). Shown is the ground truth (left) and predicted (middle) temperature field, as well as the absolute error (right).

220 **References**

- 221 [1] Donald R Lesuer. Experimental investigations of material models for ti-6al-4v titanium and  
222 2024-t3 aluminum. *DOT/FAA/R-00/25*, 2000.
- 223 [2] Xufei Lu, Xin Lin, Michele Chiumenti, Miguel Cervera, JunJie Li, Liang Ma, Lei Wei, Yunlong  
224 Hu, and Weidong Huang. Finite element analysis and experimental validation of the thermo-  
225 mechanical behavior in laser solid forming of ti-6al-4v. *Additive Manufacturing*, 21:30–40,  
226 2018.
- 227 [3] Timothy A Burrell, Steven L Campbell, Chester Coomer, Curtis William Ayers, Andrew A  
228 Wereszczak, Joseph Philip Cunningham, Laura D Marlino, Larry Eugene Seiber, and Hua-Tay  
229 Lin. Evaluation of the 2010 toyota prius hybrid synergy drive system. Technical report, Oak  
230 Ridge National Lab.(ORNL), Oak Ridge, TN (United States). Power . . . , 2011.
- 231 [4] Florian Menter, M. Kuntz, and RB Langtry. Ten years of industrial experience with the sst  
232 turbulence model. *Heat and Mass Transfer*, 4, 01 2003.