



Prediction of Mechanical Response of Special Thread Connections under Multiple Load States Based on XGBoost

Qi LI¹²³, Qingying Yuan¹²³, Mengke Sun⁴⁵, Jiarui Tang⁴⁵, Bosheng Qiao⁶, Shuaijun Yin⁶

¹College of Petroleum Engineering, Xi'an Petroleum University, Xi'an China, 710065

²Key Laboratory of Exploration and Development of Complex and Difficult-to-Produce Oil & Gas Reservoirs (Xi'an Petroleum University), Ministry of Education, Xi'an China, 710065

³Chinese National Engineering Research Center for Petroleum and Natural Gas Tubular Goods Co., Ltd., Xi'an 710000, Shaanxi, China

⁴CNPC Baoji Petroleum Steel Pipe Co., Ltd., Baoji 710000, Shaanxi, China

⁵Shaanxi Key Laboratory of Advanced Stimulation Technology for Oil & Gas Reservoirs, Xi'an China, 710065

⁶Key Laboratory of Road Construction Technology and Equipment of Ministry of Education, Chang'an University, Xi'an Shaanxi 710064, China

Correspondence to: Qingying Yuan(21111010007@stumail.xsyu.edu.cn)

Abstract: To address the issues of large finite element computation volume and strong nonlinear response relationships in special thread connections under multi-parameter, multi-load state conditions, a two-dimensional axisymmetric parametric finite element model was established. Latin hypercube sampling was employed to generate combinations of radial interference, load flank angle, taper percentage, and friction coefficient, and Abaqus script interface was invoked via Python to automate composite load application and response data extraction. With four model-level parameters, temperature, axial load, internal/external pressure, and contact area as inputs, XGBoost regression models were independently established for effective contact length, maximum contact pressure, maximum Mises stress, and maximum equivalent plastic strain. A data partitioning method based on model ID grouping was



30 adopted to evaluate the models' prediction capability for unseen parameter
31 combinations. The results show that the coefficient of determination (R^2) of the four
32 models on the independent test set were 0.999944, 0.812840, 0.780395, and 0.983697,
33 respectively, and the cross-validation prediction results on the training set were
34 generally consistent with the independent test results. Complete load spectrum
35 analysis reveals that the extrema of different sealing and structural responses
36 correspond to different load states, and a single loading condition is insufficient to
37 fully characterize the connection service response. SHAP analysis indicates that taper
38 percentage is the machining parameter with the highest contribution across all four
39 responses, followed by radial interference in its influence on maximum contact
40 pressure and maximum Mises stress. The established models can be used for state-
41 augmented prediction of existing finite element databases and provide a basis for
42 control of special thread machining parameters and dimensional inspection.

43 **Keywords:** Special thread; Parametric finite element; Seal integrity; XGBoost
44 surrogate model

45 Introduction

46 As oil and gas exploration and development extend into deeper formations,
47 wellbore environments are characterized by high temperature, high pressure, and
48 complex concurrent loads, making the seal integrity of pipe connections a critical
49 factor affecting wellbore service safety (Liu et al., 2024). Special thread connections
50 primarily rely on metal-to-metal interference contact to generate contact pressure,
51 thereby preventing fluid leakage along the connection interface (Wu et al., 2017).
52 Under complex operating conditions such as fracturing production and high-rate
53 injection/production in underground gas storage, the string is subjected to
54 continuously varying axial loads and internal/external pressures, and the seal face
55 contact state changes accordingly. Existing studies have shown that axial tension can
56 reduce contact stress in some sealing structures, and external pressure may also
57 degrade sealing performance (Li et al., 2017, Song et al., 2024). Therefore, it is
58 difficult to comprehensively evaluate the sealing and structural response of special
59 thread connections under complex service histories based solely on nominal
60 dimensions or single load states.



61 Currently, the evaluation of sealing and mechanical performance of special
62 thread connections mainly relies on finite element analysis and experimental
63 validation. Previous studies have conducted related analyses on connection sealing
64 performance, connection strength, and design parameter optimization (Zhang et al.,
65 2021) Other studies have evaluated the local contact and structural response of special
66 thread connections from perspectives such as thread parameter sensitivity under
67 bending moment loads, three-dimensional make-up processes, and sealing
68 performance after repeated make-up/break-out cycles (Zhu et al., 2012, Dou et al.,
69 2015, Shan et al., 2023). These studies indicate that geometric parameters, assembly
70 conditions, and load conditions all affect the contact pressure, stress distribution, and
71 sealing performance of thread connections. However, existing research has mostly
72 focused on a small number of deterministic cases, single parameter variations, or
73 preset extreme conditions. When radial interference, load flank angle, taper
74 percentage, and friction coefficient vary simultaneously, and each parameter set must
75 undergo multiple load states, the workload of sequential finite element calculations
76 increases significantly, making it difficult to efficiently analyze response patterns
77 under multi-parameter, multi-load state conditions.

78 To reduce the repetitive solution cost of multi-parameter finite element
79 calculations, design of experiments methods can be combined with machine learning
80 surrogate models to establish mapping relationships between input parameters and
81 mechanical responses using limited samples. Finite element surrogate models have
82 been used for engineering tolerance analysis and structural response prediction under
83 complex loads, and can to some extent replace repetitive high-cost numerical
84 solutions (Orosz et al., 2021, Dai et al., 2026). Latin hypercube sampling enables
85 stratified sampling within the range of each input variable, improving the uniform
86 coverage of samples in the parameter space (McKay et al., 2000.). The XGBoost
87 algorithm has been used for engineering parameter inversion and nonlinear response
88 prediction, capable of handling complex mapping relationships between multiple
89 input variables and target responses (Li et al., 2023) For special thread connections,
90 relying solely on geometric parameters and external loads is insufficient to fully
91 characterize the actual contact state of seal faces under different load states; therefore,
92 this paper further introduces the contact area obtained from finite element post-
93 processing as a state-augmented input to improve the effectiveness of multi-load state
94 sealing and structural response prediction.



95 **1 Parametric Establishment of Special Thread FEM Simulation**

96 **Model**

97 **1.1 Geometric Topology Features and Parametric Variables**

98 The seal integrity of special thread connections is closely related to metal-to-
99 metal sealing structures, contact pressure, and contact area (Gao et al., 2005) ; thread
100 parameters and thread interference affect the structural integrity, seal integrity, and
101 connection performance of the connection (Hu et al., 2018, Lu et al., 2015) .
102 Therefore, this paper examines thread radial interference, load flank angle, and taper
103 percentage as primary geometric parameters in parametric modeling. This paper takes
104 a special thread connection with asymmetric load flank profile and metal-to-metal
105 sealing structure as the research object. Considering that the connection structure,
106 material properties, and loads are all rotationally symmetric about the pipe axis, a
107 two-dimensional axisymmetric finite element model is used to describe the pin, box,
108 thread engagement zone, and seal contact zone. This simplification retains the main
109 mechanical characteristics such as radial interference, axial load transfer, and seal
110 face contact, while reducing the solution scale for multi-set parameter calculations.
111 Non-study parameters such as pipe diameter, wall thickness, pitch, thread height, seal
112 face length, and shoulder structure remain constant in the model to isolate the effects
113 of target parameter variations on connection response.

114 To achieve automatic generation of finite element models under different
115 parameter combinations, the Abaqus script interface is invoked based on Python
116 language to parametrize sketch dimensions, assembly positions, contact properties,
117 mesh division, and analysis task submission processes. After reading each set of input
118 parameters, the script sequentially completes pin and box geometry updates, contact
119 surface identification, material and contact property assignment, mesh generation,
120 load application, and solution file generation. For boundary position shifts that may
121 be caused by geometric parameter variations, predefined geometric features and
122 surface set rules are used to re-identify thread engagement surfaces and primary seal
123 faces, avoiding direct dependence on fixed node numbers.

124 The model-level input parameters include thread radial interference, load flank
125 angle, thread taper percentage, and interface friction coefficient.



Thread radial interference is denoted as δ , used to describe the radial coordinate difference between pin and box thread anchor points at the same axial reference position, defined as:

$$\delta = r_{pa} - r_{ba} \quad (1)$$

where r_{pa} is the radial coordinate of the pin thread reference anchor point, r_{ba} is the radial coordinate of the box thread reference anchor point, and δ is the thread radial interference in mm.

During parametric modeling, initial geometric interference between pin and box threads is formed by moving the box thread anchor point radially toward the axis. The thread interference described in this paper is a radial quantity, not equivalent to diametral interference, and does not include pre-set fixed assembly interference at the primary seal face and shoulder. If diametral difference is used to express thread interference, the relationship between radial interference and diametral interference is:

$$\Delta D_t = 2\delta \quad (2)$$

where ΔD_t is the diametral interference between pin and box threads.

The load flank angle is denoted as α , representing the inclination angle of the thread load flank surface relative to the model reference direction. This parameter directly participates in thread profile coordinate calculation, and its variation changes the geometric direction of the load flank surface as well as the connection positions of adjacent crest and root. The load flank angle may affect the load transfer method between thread flanks and the local stress distribution near the thread root, but its degree of influence needs to be determined based on subsequent finite element response and surrogate model analysis.

Thread taper percentage is denoted as T , defined as the percentage of the ratio of thread diameter change to axial length per unit axial length:

$$T = (\Delta D / L) \times 100\% \quad (3)$$

where ΔD is the thread diameter change within the axial measurement length L , and T is the thread diametral taper percentage.]

The parametric script first converts the thread taper percentage to the thread taper half-angle, with the conversion relationship:

$$\tan \theta_t = T / 200 \quad (4)$$



157
$$\theta_t = \arctan(T / 200)$$
 (5)

158 where θ_t is the angle between the thread taper generatrix and the connection axis.

159 Subsequently, the coordinate increments of adjacent thread reference points in
160 radial and axial directions are calculated based on the fixed pitch:

161
$$\Delta r = P \sin \theta_t$$
 (6)

162
$$\Delta z = P \cos \theta_t$$
 (7)

163 where Δr is the radial coordinate increment, Δz is the axial coordinate
164 increment of adjacent thread reference points, and p is the thread pitch, taken as 6.35
165 in this paper. This generates continuous pin and box thread profiles distributed along
166 the taper direction. Therefore, the taper described in this paper is the diametral taper
167 percentage, not the thread taper angle.

168 The interface friction coefficient is denoted as μ , used to define Coulomb friction
169 behavior at thread surfaces, primary seal faces, and shoulder contact interfaces.
170 Thread radial interference, load flank angle, and thread taper percentage belong to
171 geometric or assembly parameters, while the friction coefficient belongs to interface
172 contact parameters.

173 To avoid confusion of different geometric parameters in coordinate definition
174 and subsequent script calls, Figure 1 gives the definition method of main geometric
175 parameters in the pin-box local thread profile. Among them, thread radial interference
176 and load flank angle directly participate in local profile construction, taper percentage
177 is used to determine coordinate increments of adjacent profile reference points along
178 radial and axial directions, and pitch participates in coordinate generation as a fixed
179 structural parameter.

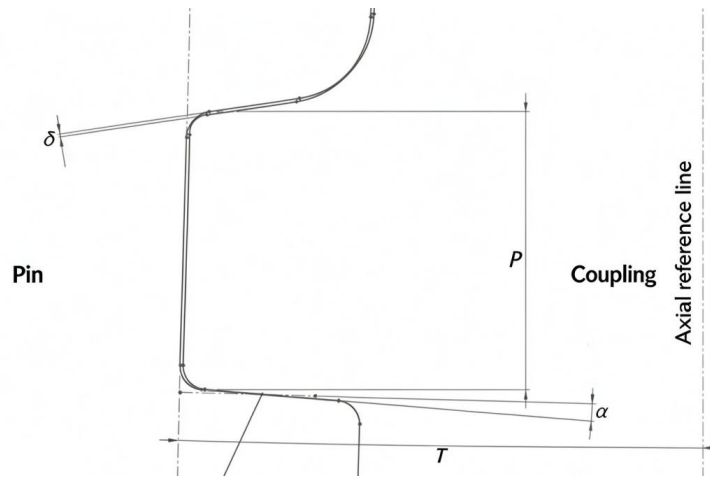


Figure 1 Schematic diagram of local thread profile geometric parameter definition for special thread connections

1.2 Material Constitutive, Assembly Steps and Composite Load Boundary

1.2.1 Elastoplastic Material Constitutive in Variable Temperature Environment

In deep oil and gas well service environments, the seal faces and thread flank regions of casing pin and box threads experience significant thermal softening effects under extreme bottomhole temperatures. To accurately characterize the mechanical response of connections under variable temperature conditions, the finite element model incorporates a temperature-dependent isotropic hardening Mises yield criterion.

The pipe body materials adopted in this study include three high-strength OCTG steels: N80, P110 and V125. High-grade alloy materials exhibit thermal softening under high-temperature environments, which leads to substantial differences in their true stress – strain evolution laws compared with those at room temperature. In the finite element solution, the solver invokes corresponding material parameters according to the temperature field of each analysis step and performs interpolation between the given temperature data points to characterize the influence of temperature variation on the elastoplastic behavior of materials. Taking the V125 grade as an example, the key plastic constitutive input parameters of the model at room temperature and high-temperature limits are listed in Table 1.



202 **Table 1 V125 steel true plastic characteristic input parameters at boundary temperatures**

Room Temperature Condition (20°C)		Extreme High Temperature Condition (180°C)	
True Plastic Strain	True Stress/MPa	True Plastic Strain	True Stress/MPa
0	959.7278	0	863.7551
0.0142	980.2254	0.0149	882.2028
0.0194	991.2867	0.0199	892.1580
0.0368	1029.6873	0.0328	918.7289
0.0526	1052.5051	0.0531	947.2546

203 **1.2.2 Composite Loading History, Boundary Conditions and Load State**
 204 **Parametric**

205 Special thread connections simultaneously bear axial tension or compression,
 206 internal/external pressures, and temperature loads during downhole service, and their
 207 seal face contact state exhibits obvious load path dependency. To characterize the
 208 mechanical response of connections under different composite load combinations, this
 209 paper, after completing initial assembly simulation, continuously applies temperature,
 210 axial load, and internal/external pressure according to preset load spectra, and extracts
 211 sealing and structural responses at each target load state.

212 Finite element analysis first establishes the initial contact state after connection
 213 assembly. The assembly process sequentially includes thread engagement, primary
 214 seal face contact, and shoulder contact. The thread engagement stage uses shrink fit
 215 method to eliminate initial radial interference between pin and box threads;
 216 subsequently, interference contact between primary seal faces and shoulders is
 217 established. Each assembly stage uses static implicit solution with geometric
 218 nonlinearity enabled, allowing contact pressure, local plastic deformation, and
 219 residual stress generated during assembly to be transferred to subsequent composite
 220 load calculations.

221 After completing assembly, composite service loads are applied while
 222 maintaining continuous contact relationships and deformation history. The box end
 223 restricts axial displacement, equivalent axial stress is applied at the pin end; internal
 224 and external pressures act on the inner and outer surfaces of the connection
 225 respectively. The primary seal face is set with pressure penetration boundary
 226 conditions to simulate the process of pressure medium continuing to propagate along
 227 the seal face after local separation of the contact interface. Axial load, internal
 228 pressure, external pressure, and temperature are all continuously applied through



229 tabular amplitude curves to avoid artificial resetting of stress and contact state
230 between different load states.

231 The load spectrum is calculated based on pipe body dimensions, material
232 reference strength, and temperature reduction factors. Load paths include high-
233 temperature composite load envelopes, temperature cyclic loads, and room-
234 temperature composite load states at different load levels, covering typical
235 combinations such as axial tension-internal pressure, axial compression-internal
236 pressure, pure axial load, and axial load-external pressure. Under high-temperature
237 load states, the temperature field and mechanically loaded loads (after temperature
238 reduction) are applied synchronously, thereby simultaneously considering material
239 stiffness changes, thermal expansion deformation, and load-bearing capacity changes
240 caused by temperature variations.

241 During finite element solving, analysis step numbers and load point numbers are
242 only used to control load application sequence and identify result output locations,
243 and are not used as input variables for the surrogate model. To eliminate the
244 restriction of different load spectrum numbers on model training, this paper re-
245 expresses each finite element result state as load features composed of temperature,
246 axial load, and internal/external pressures. For the same thread radial interference,
247 load flank angle, taper percentage, and friction coefficient, different load states
248 separately constitute independent data records.

249 After finite element result organization, each parametric model retains the 35
250 most critical load steps from the loading history for load states used in surrogate
251 model training. Each data record contains 4 model-level parameters, temperature,
252 axial load, internal/external pressures, and corresponding contact area characteristic
253 integral values, with effective contact length, maximum contact pressure, maximum
254 stress, and maximum equivalent plastic strain as prediction targets. This data
255 organization method enables the surrogate model to learn the mapping relationships
256 between geometric parameters, interface parameters, load states, and mechanical
257 responses, rather than learning results corresponding to specific analysis step numbers.

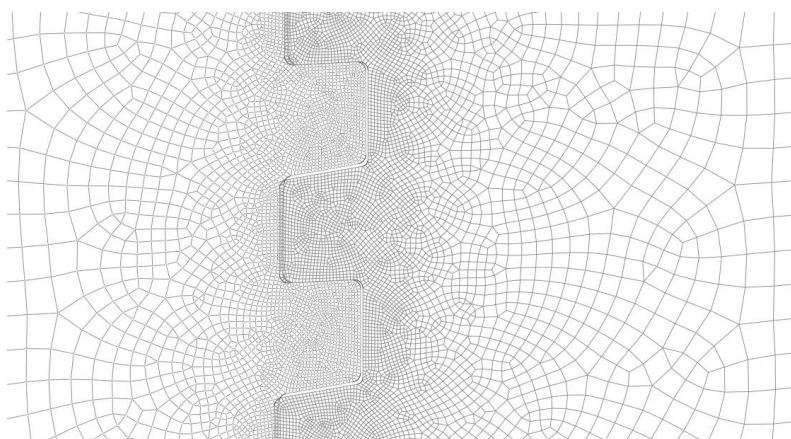
258 Through the above processing, the same set of geometric parameters can
259 correspond to multiple different composite load states, and the surrogate model
260 thereby gains the ability to predict connection sealing and structural response under
261 given parameter combinations and load conditions. At the same time, all load states
262 under the same finite element model number are kept as a whole during data set



263 partitioning, to avoid the same geometric parameter set entering both training and test
 264 sets simultaneously.

265 **1.3 FEM Mesh Division and Local Refinement**

266 To balance the computational accuracy of nonlinear contact with high-
 267 throughput solving efficiency for massive samples, forced local refinement is applied
 268 to thread flank engagement zones and seal face contact bands where stress gradients
 269 are severe, with characteristic mesh size controlled at 0.1 mm; the macroscopic pipe
 270 body far from contact zones uses 1.0 mm coarse mesh for smooth transition. Mesh
 271 topology details are shown in Figure 2.



272

273 **Figure 2 Local mesh refinement details of special thread connection finite element model**

274 **1.4 Contact Nonlinearity Characterization and High-Throughput Solving** 275 **Robustness Control**

276 The sealing performance and load transfer capability of special thread
 277 connections are closely related to the contact state of thread surfaces, metal seal faces,
 278 and shoulder surfaces. To describe separation, closure, and relative sliding of contact
 279 regions under assembly and composite load actions, the model adopts surface-to-
 280 surface contact discretization with finite slip tracking enabled.

281 The normal behavior of contact surfaces uses a hard contact model allowing
 282 separation to restrict normal penetration of contact surfaces; the tangential behavior
 283 uses a Coulomb friction model, with stick and slip of contact surfaces described
 284 through isotropic penalty method. The maximum elastic slip rate in the penalty
 285 method is set to 0.005 to control the elastic slip range before the contact interface



transitions from stick to slip state. Thread surfaces, metal seal faces, and shoulder surfaces use the same friction coefficient, which varies as a parametric variable within preset ranges.

The assembly process establishes initial states of different contact regions in stages. First, initial geometric interference between pin and box threads is eliminated through shrink fit to form contact pressure in the thread engagement zone; subsequently, interference contact of metal seal faces and shoulder surfaces is established in sequence. After completing assembly, existing contact states and deformation history are maintained, and temperature, axial load, and internal/external pressures are applied in subsequent analyses.

Since the model simultaneously contains material nonlinearity, geometric nonlinearity, and contact nonlinearity, rapid contact state transitions and significant stiffness matrix variations may occur under certain parameter combinations and load states, thereby increasing the convergence difficulty of implicit solving. To improve solving stability for batch parametric calculations, automatic time increment control is adopted in the analysis, with initial time increment set to 0.01, minimum time increment set to 1×10^{-8} , and maximum allowed increments per analysis step set to 10000. When equilibrium iteration within a certain increment cannot satisfy convergence conditions, the solver automatically reduces the time increment and recalculates until convergence requirements are met or the preset minimum increment limit is reached. The automatic increment adjustment process for static analysis steps under typical parameter combinations is shown in Figure 3.

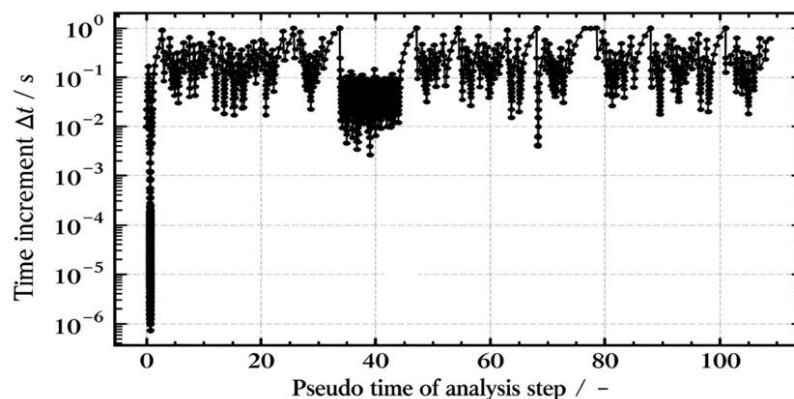


Figure 3 Automatic increment adjustment process for static analysis steps under typical parameter combinations



As shown in Figure 3, during stages where local contact states undergo significant changes, the convergence difficulty of equilibrium iteration increases, and the solver re-performs nonlinear iteration by automatically reducing time increments, with some increments decreasing to the 10^{-6} level. Automatic increment control can improve solving stability under complex contact conditions, reduce calculation interruptions caused by extreme parameter combinations, and provide support for continuous solving of batch parametric models and acquisition of valid samples.

2 Cross-Scale Feature Space and High-Throughput Sample Set Construction

2.1 Multi-Level Feature Space and Variable Definition

To establish mapping relationships between geometric parameters, load states, and mechanical responses of special thread connections, this paper divides the surrogate model inputs into three levels: model-level parameters, load state parameters, and finite element state-augmented parameters.

Model-level parameters include thread radial interference, load flank angle, taper percentage, and interface friction coefficient. Combined with connection design dimensions and preset machining fluctuation ranges, the value boundaries of each parameter are determined, as shown in Table 2. The above four parameters remain unchanged throughout the complete load history of the same finite element model, used to characterize different machining parameter combinations and interface conditions.

Table 2 LHS sampling ranges of model-level parameters

Variable Name	Variable Symbol	Unit	Range
Thread radial interference	δ_t	-	[0.032, 0.042]
Load flank angle	α_b	°	[2.8, 3.3]
Taper percentage	k_t	-	[4.7, 5.0]
Friction coefficient	μ	-	[0.02, 0.06]

Latin hypercube sampling method is used to generate model-level parameter combinations. The initial sample size is 540 sets, meaning the value range of each variable is divided into 540 equal-probability sub-intervals, with 1 sample randomly drawn from each sub-interval, followed by random combination of sample sequences



for different variables. This method can improve stratified coverage of each variable within its boundary range and reduce local clustering produced by simple random sampling.

Load state parameters include temperature, axial load, and internal/external pressures. The above variables are determined by preset composite load spectra and do not participate in LHS sampling. The same model-level parameter combination forms multiple data records under different load states, enabling the surrogate model to identify mechanical responses when geometric parameters and load conditions change simultaneously.

On this basis, the contact area state quantity obtained from finite element post-processing - seal face contact area integral value - is used as state-augmented input. This variable is used to characterize the overall state of the contact interface under current geometric parameters and load conditions, and together with 4 model-level parameters and 3 load state parameters forms an 8-input feature set. Model outputs are effective contact length, maximum contact pressure, maximum stress, and maximum equivalent plastic strain, as shown in Figure 4.

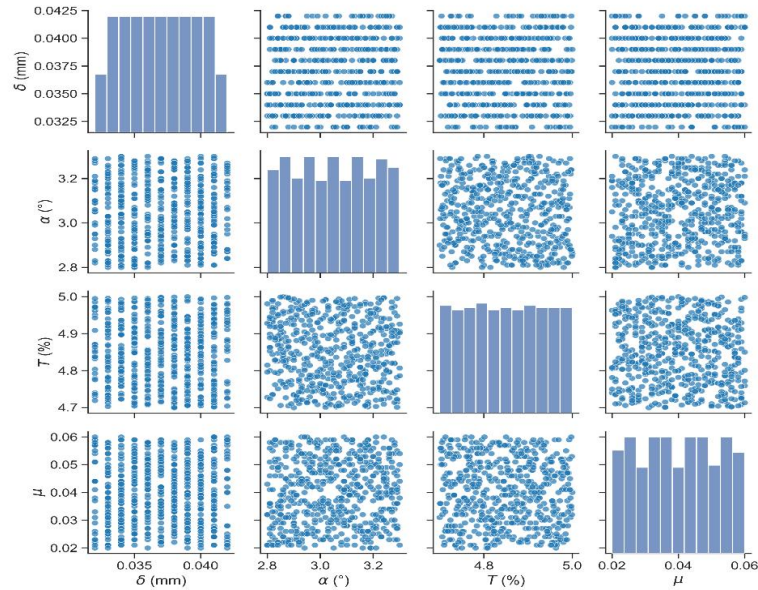


Figure 4 LHS sampling model parameter space distribution

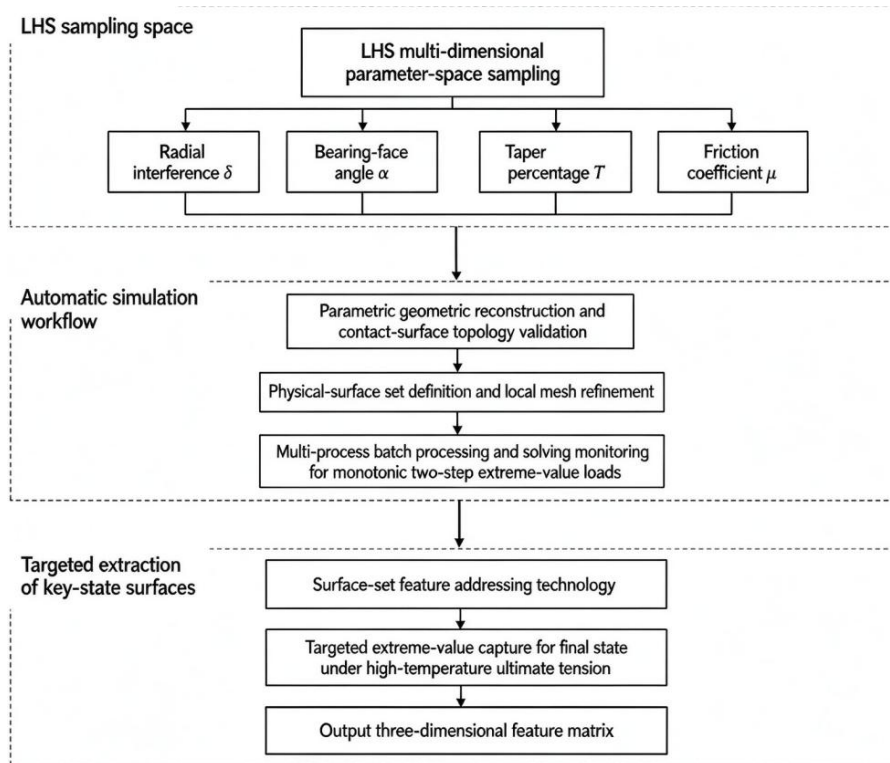


2.2 Seal Surface Targeted Automated Simulation and Feature Extraction

Based on Python-Abaqus secondary development program, parameter reading, geometry updates, contact surface identification, mesh generation, load application, task submission, and result extraction are sequentially completed. The master script reads thread radial interference, load flank angle, taper percentage, and friction coefficient combinations generated by Latin hypercube sampling, automatically reconstructs pin and box models, and re-identifies thread engagement surfaces, metal seal faces, and shoulder contact zones based on geometric positions, to ensure consistency of result extraction regions under different parameter combinations.

After completing assembly simulation, temperature, axial load, and internal/external pressures are continuously applied according to preset load spectra. Unlike extracting only a single limit analysis step, this paper retains calculation results of each parametric model at multiple representative load states. After solving state and data integrity checks, a total of 538 valid model IDs were obtained, with 35 load states retained for each model.

For each load state, the script automatically extracts temperature, axial load, internal/external pressures, and contact area state quantities, and calculates effective contact length and maximum contact pressure of the metal seal region, as well as maximum stress and maximum equivalent plastic strain in the model. The above data are associated row by row with corresponding 4 model-level parameters by model ID and load state, forming a structured data set for XGBoost model training. The parametric modeling, composite load calculation, and response extraction flow are shown in Figure 5.



378

379 **Figure 5 Cross-scale parametric modeling and targeted extraction flow of key state features**

380 **3 XGBoost-Based Special Thread Multi-State Mechanical Response** 381 **Prediction**

382 The sealing and structural response of special thread connections are jointly
 383 influenced by machining parameters, interface conditions, and composite loads, with
 384 significant nonlinear coupling relationships among variables. To achieve rapid
 385 prediction of finite element calculation results, this paper uses thread geometric
 386 parameters, friction coefficient, load states, and finite element seal face contact area
 387 state quantities as inputs, and independently establishes XGBoost regression models
 388 for effective contact length, maximum contact pressure, maximum stress, and
 389 maximum equivalent plastic strain.

390 **3.1 XGBoost Algorithm Principle and Multi-Model Modeling Framework**

391 XGBoost is an ensemble learning algorithm based on residual approximation,
 392 which fits the prediction residuals of the previous round by continuously adding new



393 regression trees. For a data set containing m samples, the prediction value for the m -th
 394 sample at the t -th iteration can be expressed as:

$$395 \quad \hat{y}_i^{(t)} = \sum_{j=1}^t f_j(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (8)$$

396 where $\hat{y}_i^{(t-1)}$ is the accumulated value of previous round predictions, $f_t(x_i)$ is the
 397 newly added regression tree model at the t -th round, and x_i is the input feature vector.]

398 To control model complexity and suppress overfitting in small sample training,
 399 XGBoost performs second-order Taylor expansion of the loss function and forcibly
 400 introduces a structural risk regularization term in the objective function. Its core
 401 objective function is defined as:

$$402 \quad obj^{(t)} = \sum_{i=1}^m l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (9)$$

403 where l is a loss function used to quantify the error between the true calculated
 404 extreme value y_i from finite element analysis and the predicted value $\hat{y}_i^{(t)}$, and
 405 $\Omega(f_k)$ is the regularization function penalizing model complexity.

$$406 \quad \Omega(f_k) = \gamma T_k + \frac{1}{2} \lambda \|\omega\|^2 \quad (10)$$

407 where T_k is the total number of leaf nodes of the k -th tree, ω denotes the weight
 408 corresponding to the leaf node, and γ and λ are the regularization penalty
 409 coefficients controlling the node splitting threshold and leaf weight, respectively. By
 410 minimizing the above objective function, the surrogate model can adaptively
 411 compress the topological depth of decision trees while ensuring prediction accuracy
 412 of mechanical indicators.

413 Since seal face effective contact length, maximum contact pressure, maximum
 414 stress, and maximum equivalent plastic strain have different dimensions, numerical
 415 distributions, and nonlinear response patterns, this paper does not adopt a unified
 416 multi-output model, but independently establishes separate XGBoost regression
 417 models for 4 prediction targets. Each model uses the same 8 input features but
 418 performs hyperparameter optimization and error evaluation separately, to avoid
 419 mutual interference of different target scales on the model training process.

420 Among them, the seal face contact area integral value is extracted from finite
 421 element results, used to supplement characterization of the overall contact state under



specific geometric parameters and load conditions. Therefore, the models established in this paper belong to mechanical response prediction models enhanced with finite element state information, and their main purpose is to rapidly obtain multiple sealing and structural responses under known contact state quantities.

3.2 Data Grouping and Model Training Method

In the finite element data set, the same model corresponds to the same combination of thread radial interference, load flank angle, taper percentage, and friction coefficient, but contains multiple state records under different temperature, axial load, and internal/external pressure conditions. If conventional random partitioning is used, different load states of the same geometric model may simultaneously enter the training and test sets, leading to overestimation of the model's generalization capability for unseen parameter combinations. To avoid this type of data leakage, this paper partitions the data set by model ID as the grouping unit, and all load states under the same ID can only enter the training set or independent test set as a whole.

Data partitioning uses the GroupShuffleSplit method, randomly selecting 20% by model ID as the independent test set, with the remaining data as the training set. The independent test set does not participate in computation during hyperparameter search and model training stages, and is only used for final generalization performance evaluation. The training set internally uses 5-fold GroupKFold cross-validation, with no duplicates between each fold's training and test sets, thereby ensuring model validation is performed on geometric parameter combinations not involved in training. And 4 independent XGBoost regression models are trained.

Before model training, integrity and seal state checks are performed on finite element output data. Existing studies have used the ratio of seal face contact pressure integral value to the absolute values of internal and external pressures to evaluate the sealing capability of special thread connections, and took the ratio greater than 4.5 as the criterion for good seal integrity. Combined with enterprise engineering application experience and existing experimental data calibration results, this paper sets the seal state effectiveness threshold at 5.02, meaning only calculation records with the ratio of seal face contact area integral value to the absolute values of internal/external pressures no less than 5.02 are retained. For states where the absolute values of internal and external pressures are 0 and effective contact has been formed, this ratio



455 criterion is not used to determine failure. After screening, valid samples meeting seal
 456 state requirements with complete response data were obtained.

457 Model hyperparameters are determined using random search method, with
 458 minimum root mean square error of 5-fold GroupKFold validation set as the
 459 parameter optimization criterion. Search parameters include number of regression
 460 trees, learning rate, maximum tree depth, minimum child node weight, row sampling
 461 ratio, column sampling ratio, node splitting threshold, and regularization coefficients.
 462 Each prediction target randomly searches 24 parameter combinations, with each
 463 combination undergoing 5-fold group cross-validation, completing 120 fits in total.
 464 After obtaining optimal parameters, the model is retrained using the entire training set,
 465 and cross-validation prediction results on the training set and independent test set
 466 prediction results are calculated separately.

467 Model prediction performance is evaluated using coefficient of determination,
 468 root mean square error, mean absolute error, and mean absolute percentage error.
 469 Considering that the 4 prediction targets have different dimensions and numerical
 470 ranges, evaluation metrics of each output model are reported separately, without using
 471 a single comprehensive metric to summarize model performance.

472 **Table 3 Optimal hyperparameters of four XGBoost models**

Prediction Target	Num. Trees	Learning Rate	Max Depth	Min. Weight	Row Sample	Col Sample	Min. Loss	L1 Reg.	L2 Reg.
Effective contact length	1100	0.02	8	12	0.90	1.00	0.50	0.00	1.00
Maximum contact pressure	500	0.02	6	8	0.65	0.90	0.01	0.10	0.50
Maximum Mises stress	500	0.02	6	8	0.65	0.90	0.01	0.10	0.50
Maximum equivalent plastic strain	2200	0.05	8	3	1.00	0.90	0.05	0.01	1.50

473 **Table 4 Prediction accuracy comparison of four XGBoost models**

Prediction Target	Dataset	R ²	RMSE	MAE	MAPE
Effective contact length	Training (out-of-fold)	0.999896	1.448972	0.461501	0.293529
Effective contact length	Independent test	0.999944	1.069613	0.412197	0.279772
Maximum contact pressure	Training (out-of-fold)	0.828528	213.121017	140.832569	8.696553
Maximum contact pressure	Independent test	0.812840	203.352281	134.875812	8.384899
Maximum stress	Training (out-of-fold)	0.781087	42.400140	19.923222	2.208761
Maximum stress	Independent test	0.780395	41.195226	19.540259	2.147105
Maximum equivalent plastic strain	Training (out-of-fold)	0.979535	0.025892	0.014687	6.868372
Maximum equivalent plastic strain	Independent test	0.983697	0.022832	0.014812	7.340534



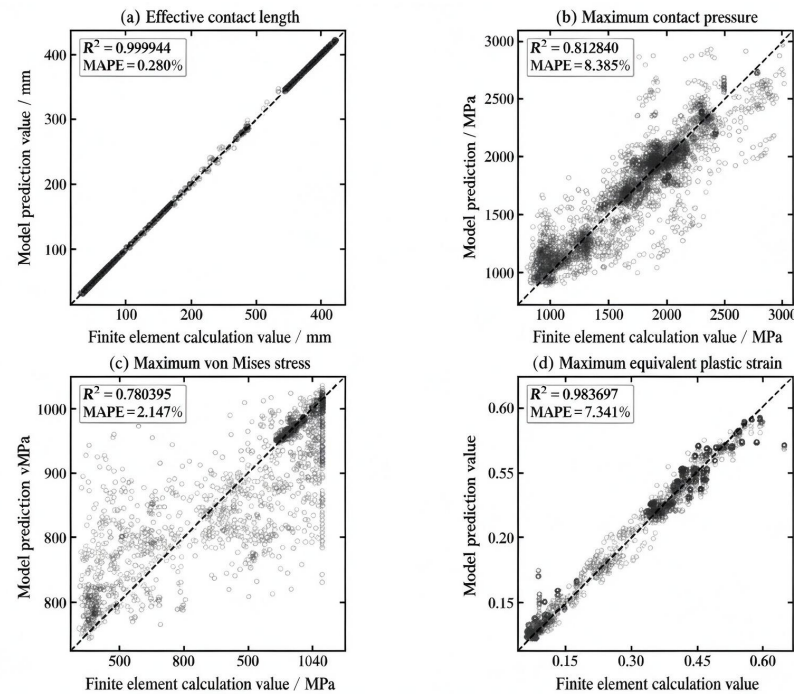
474 As shown in Table 3, different optimal hyperparameters exist for different
475 mechanical responses. Among them, the effective contact length model uses more
476 decision trees and greater tree depth to fit the nonlinear relationship between contact
477 state and effective contact length; the maximum contact pressure and maximum stress
478 models obtained the same optimal parameter combination; the maximum equivalent
479 plastic strain model uses 2200 decision trees and a learning rate of 0.05. The above
480 results indicate that different prediction targets have different response patterns and
481 data distributions, making separate hyperparameter optimization necessary.

482 As shown in Table 4, the effective contact length model has the highest
483 prediction accuracy, with an independent test set R^2 of 0.999944 and MAPE of
484 0.280%. The independent test set R^2 values for the maximum contact pressure model
485 and maximum Mises stress model are 0.812840 and 0.780395, with corresponding
486 MAPE values of 8.385% and 2.147%. The independent test set of the maximum
487 equivalent plastic strain model reaches 0.983697, but due to some sample values
488 being close to zero, its MAPE is relatively high at 7.341%. The evaluation metrics of
489 the four models on training set cross-validation and independent test set are generally
490 close, indicating that the models do not show significant accuracy degradation on
491 parameter combinations not involved in training.

492 It should be noted that the dimensions and numerical ranges of different
493 prediction targets vary, so their RMSE and MAE cannot be directly compared
494 horizontally. Therefore, this paper mainly combines R^2 and MAPE to evaluate the
495 prediction effects of different models.

496 3.3 Independent Test Set Prediction Result Analysis

497 Table 4 evaluates the prediction accuracy of each model from a statistical
498 indicator perspective, but R^2 and average error are difficult to reflect the dispersion of
499 predicted values within different response ranges. To further observe the consistency
500 between model predictions and finite element calculations, scatter plots of predicted
501 values versus finite element calculated values are drawn using 3080 state samples
502 from the independent test set, as shown in Figure 6. The dashed line in the figure
503 represents the ideal relationship when predicted values exactly match finite element
504 calculated values.



505

506 **Figure 6 Comparison of predicted values and finite element calculated values on independent test**
 507 **set**

508 As shown in Figure 6(a), the predicted points for effective contact length are
 509 basically distributed near the ideal reference line, with only a few samples showing
 510 slight deviation. The model achieves an R^2 of 0.999944 on the independent test set
 511 with MAPE of 0.280%, indicating that under known contact area state quantities, the
 512 model can accurately characterize the variation of effective contact length under
 513 different geometric parameters and load states.

514 In Figure 6(b), the predicted points for maximum contact pressure are generally
 515 distributed along the ideal reference line, but with significantly higher dispersion than
 516 the effective contact length model, and some high contact pressure samples show
 517 increased prediction errors. Its independent test set R^2 is 0.812840 with MAPE of
 518 8.385%. Maximum contact pressure is a local contact response, relatively sensitive to
 519 contact boundary position, local mesh, and load state variations, so its prediction
 520 difficulty is relatively high.

521 In Figure 6(c), the predicted points for maximum Mises stress also show
 522 noticeable dispersion, with an independent test set R^2 of 0.780395. Although this



metric is lower than the other three models, its MAPE is only 2.147%, indicating that the prediction errors remain in a relatively low range compared to the overall stress level. The relatively low R^2 mainly reflects the model's ability to explain stress fluctuation amplitude between samples, and cannot be used alone to judge engineering error magnitude.

In Figure 6(d), the prediction results for maximum equivalent plastic strain generally agree with finite element calculation results, with an independent test set R^2 of 0.983697. The model's MAPE is 7.341%, higher than the effective contact length and maximum Mises stress models, mainly because the equivalent plastic strain of some samples is close to zero, and small absolute deviations also produce relatively large relative errors. Therefore, the performance evaluation of the maximum equivalent plastic strain model should combine R^2 , RMSE, and MAE for comprehensive judgment.

Combining Table 4 and Figure 6, all four XGBoost models maintain prediction accuracy on the independent test set comparable to their training set cross-validation results. Among them, effective contact length and maximum equivalent plastic strain have higher overall fitting degrees; the prediction points for maximum contact pressure and maximum Mises stress show relatively larger dispersion, but their error levels can still reflect main response variation patterns. The above results indicate that XGBoost models built on finite element contact area state quantities can achieve rapid prediction of sealing and structural responses under multi-load states.

3.4 Prediction Error Distribution and Model Stability Analysis

To further analyze the prediction error distribution within different response ranges, the prediction residual is defined as the difference between model predicted values and finite element calculated values. Residuals greater than 0 indicate over-prediction, while residuals less than 0 indicate under-prediction. Based on all 3080 state samples from the independent test set, the distribution of prediction residuals versus finite element calculated values is plotted, as shown in Figure 7. The dashed line in the figure represents the zero-residual position, and the solid line represents the average residual of the corresponding model.

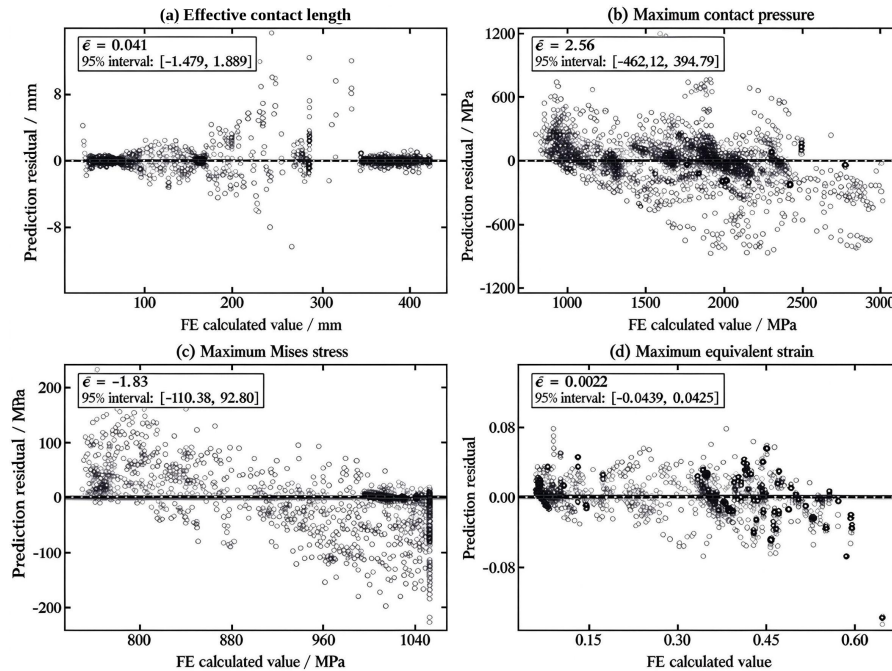


Figure 7 Prediction residual distribution on independent test set

As shown in Figure 7(a), the residuals of the effective contact length model are mainly concentrated near zero, with an average residual of 0.041 mm, median residual close to zero, and 95% empirical percentile interval of -1.479 to 1.889 mm. This result indicates that the model does not show obvious overall over-prediction or under-prediction for effective contact length, consistent with the higher R^2 and lower MAPE in Table 3-2.

In Figure 7(b), the average residual of the maximum contact pressure model is 2.555 MPa, with 95% empirical percentile interval of -462.119 to 394.787 MPa. Its average residual is small relative to the overall contact pressure level, but the residual distribution range is significantly wider than the effective contact length model, indicating relatively stronger prediction dispersion of local maximum contact pressure. This is related to the fact that maximum contact pressure is greatly affected by local contact position, contact state changes, and pressure peak fluctuations.

In Figure 7(c), the average residual of the maximum Mises stress model is -1.825 MPa, with 95% empirical percentile interval of -110.384 to 92.798 MPa. Residuals are generally distributed on both sides of zero, and the model does not show obvious



571 persistent unidirectional bias. Although the R^2 of this model is relatively low, its
572 MAPE is only 2.147%, indicating that prediction errors remain small relative to the
573 overall stress level.

574 In Figure 7(d), the average residual of the maximum equivalent plastic strain
575 model is 0.0022, with 95% empirical percentile interval of -0.0439 to 0.0425. Positive
576 and negative residuals are generally balanced. Since the equivalent plastic strain of
577 some samples is close to zero, small absolute deviations may also form relatively
578 large percentage errors, so the accuracy of this model should be evaluated
579 comprehensively by combining R^2 , RMSE, and residual absolute values.

580 Combining Figure 7, the average residuals of all four models are close to zero,
581 with no obvious overall over-prediction or under-prediction. Among them, the
582 residuals of the effective contact length model are most concentrated; the residuals of
583 maximum contact pressure and maximum Mises stress models show relatively larger
584 dispersion; although the maximum equivalent plastic strain model has relatively high
585 MAPE, its absolute residuals remain at a low level. The residual analysis results are
586 basically consistent with the model prediction performance reflected in Table 3-2 and
587 Figure 6.

588 Combining Table 4, Figure 6, and Figure 7, the 4 established XGBoost models
589 all maintain relatively stable prediction accuracy on the independent test set, with no
590 obvious overall over-prediction or under-prediction. The models can simultaneously
591 obtain sealing contact and structural strength related responses under given machining
592 parameters, load conditions, and contact state quantities, providing a computational
593 basis for subsequent extreme value identification, parameter response analysis, and
594 machining parameter range screening under multi-load states.

595 **4 Extreme Value Identification and Parameter Response Analysis** 596 **under Multiple Load States**

597 **4.1 Multi-Load State Response Extreme Value Identification**

598 To identify unfavorable response states of different machining parameter
599 combinations throughout the complete load history, the solving integrity of the finite
600 element database is first checked. The original data contains 538 Model_IDs, each
601 with 35 load states set. Among them, all 35 states of 445 Model_IDs solved



602 successfully, 5 Model_IDs obtained only partial successful states, and the remaining
 603 88 Model_IDs did not obtain valid results. To avoid extreme value misjudgment
 604 caused by missing load states, this section only uses 445 parameter combinations with
 605 all-state successful solving for statistics, meaning each response indicator is extracted
 606 based on finite element state records.

607 Extreme value extraction uses Model_ID as the basic statistical unit. For each
 608 parameter combination, effective contact length takes the minimum value among 35
 609 load states to characterize the most unfavorable sealing contact range during the load
 610 history; maximum contact pressure, maximum Mises stress, and maximum equivalent
 611 plastic strain each take the maximum value among 35 states to characterize local
 612 contact pressure peaks, structural stress peaks, and plastic deformation accumulation
 613 levels. Thus, each type of response obtains 445 mutually independent model-level
 614 extrema, with statistical results shown in Table 5.

615 **Table 5 Statistical results of response extrema**

Response Extremum	Mean	Median	Min	Max	Tied Count	Tied Ratio
Min. effective contact length	99.5726	44	28.8	388.6	73	16.4045
Max. contact pressure	2375.7435	2360.9939	1661.2512	3551.8469	8	1.7978
Max. Mises stress	1052.5669	1052.5839	1052.51	1052.6151	125	28.0899
Max. equivalent plastic strain	0.4134	0.4187	0.0715	0.6477	359	80.6742

616 As shown in Table 5, the distribution range of the minimum effective contact
 617 length is relatively wide, and its median is significantly lower than the mean,
 618 indicating that there are considerable differences in the sealing contact retention
 619 capability under different combinations of machining parameters. Maximum contact
 620 pressure and maximum equivalent plastic strain also show relatively noticeable inter-
 621 model fluctuations. In contrast, maximum Mises stress extrema are concentrated in a
 622 relatively narrow interval, indicating that this indicator has limited discrimination
 623 ability for machining parameter differences under the current load path and result
 624 extraction method.

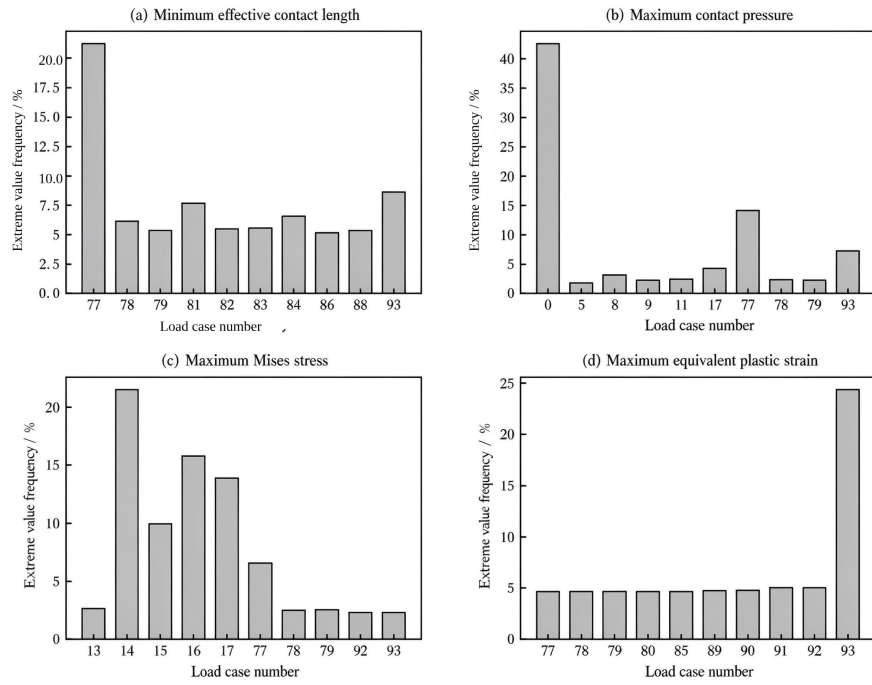


Figure 8 Load state distribution corresponding to different response extrema

Figure 8 gives the frequency distribution of load states corresponding to each response extremum. Minimum effective contact length is dispersed across multiple load states, maximum contact pressure extrema mostly appear in initial assembly states, and maximum Mises stress extrema are mainly concentrated in high-temperature load stages. Maximum equivalent plastic strain has multiple states reaching the maximum value in parallel in some models, reflecting the path accumulation characteristics of plastic response. Therefore, the controlling load states for different responses are not consistent, and subsequent machining parameter analysis should use the full load history extrema of each Model_ID separately; a single load state cannot simultaneously represent sealing and structural responses.

4.2 Influence of Input Features on Mechanical Response

To compare the contribution of different input variables to model prediction results, SHAP values are obtained using the model's native contribution value calculation method, and the global importance of each feature is characterized by the mean absolute SHAP value of each feature. The calculation uses the 15666 valid state



records screened after model training, without re-partitioning the data set or retraining
the model. To eliminate the influence of different prediction target dimensions, the
mean absolute SHAP values of each feature are normalized separately by target
model, with results shown in Figure 9. SHAP values reflect the contribution of
variables to model prediction results and do not represent independent causal
relationships between variables and finite element responses.

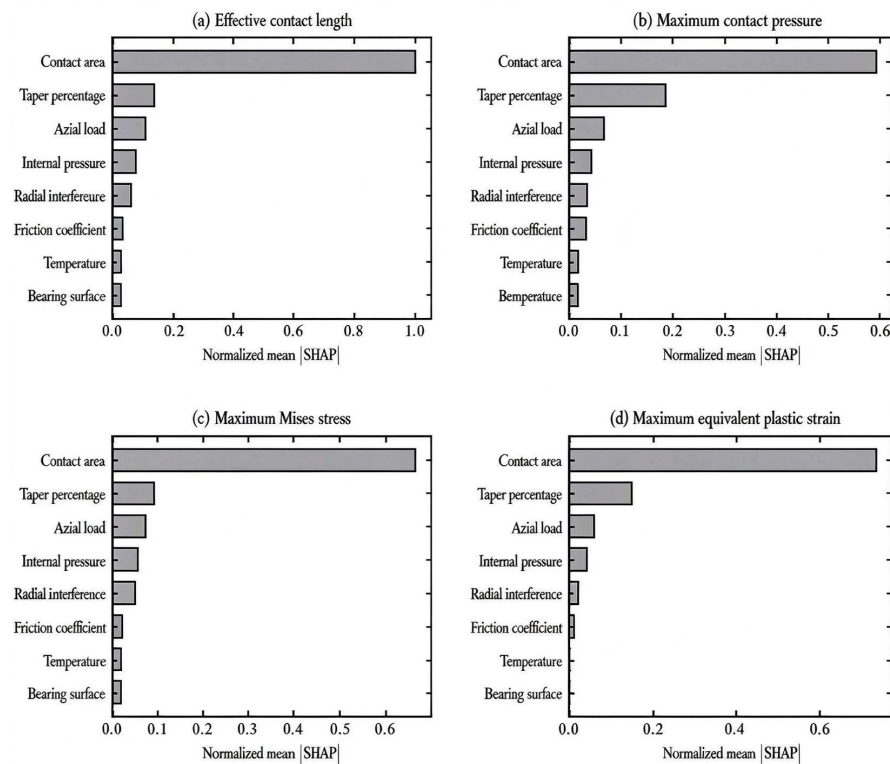


Figure 9 Global importance of main input variables (SHAP)

As shown in Figure 9, contact area has high global importance in all 4 prediction
models, indicating that the finite element contact area state quantity contains the main
contact state information under current geometric parameters and load conditions.
Among them, this variable's contribution to the effective contact length model is most
concentrated, while in the maximum contact pressure, maximum Mises stress, and
maximum equivalent plastic strain models, other geometric parameters and load
variables still retain certain contributions. This result further indicates that the models
established in this paper belong to finite element state information enhanced



prediction models, with the prerequisite that the corresponding contact area under the state has been obtained.

Contact area, temperature, axial load, and internal/external pressures do not belong to machining dimension parameters and cannot be directly used for machining tolerance sensitivity ranking. To compare the relative contributions of controllable machining parameters, SHAP importance of radial interference, load flank angle, taper percentage, and friction coefficient is further extracted, with results shown in Table 6.

Table 6 SHAP importance of machining parameters under different prediction targets

Prediction Target	Radial Interference	Flank Angle	Taper %	Friction Coeff.	Top Machining Parameter
Effective contact length	0.0002	0.0003	0.0007	0.0003	Taper percentage
Maximum contact pressure	0.0440	0.0170	0.1865	0.0187	Taper percentage
Maximum Mises stress	0.0494	0.0198	0.0909	0.0227	Taper percentage
Maximum equivalent plastic strain	0.0053	0.0022	0.1493	0.0014	Taper percentage

As shown in Table 6, taper percentage is the machining parameter with the highest contribution across all 4 prediction targets, especially with notable contributions to maximum contact pressure, maximum Mises stress, and maximum equivalent plastic strain. Radial interference ranks second among machining parameters in importance for maximum contact pressure (0.0440) and maximum Mises stress (0.0494). The global contributions of load flank angle and friction coefficient are relatively low within the current value ranges.

For effective contact length, the importance of all 4 machining parameters is below 0.001. This does not mean that machining parameters do not physically affect effective contact length, but rather that under the condition where contact area is already used as model input, machining parameters provide less additional information for prediction results. The contact area itself comprehensively reflects the combined effects of geometric parameters, assembly conditions, and external loads on the seal contact region.

SHAP importance can only reflect the overall contribution of variables to model output, cannot give the direction of response change when machining parameters increase or decrease, and cannot directly identify local mutation intervals. Therefore, further analysis of the distribution relationships between 4 machining parameters and



key responses is performed based on the response extrema of 445 complete load history models, with results shown in Figure 10.

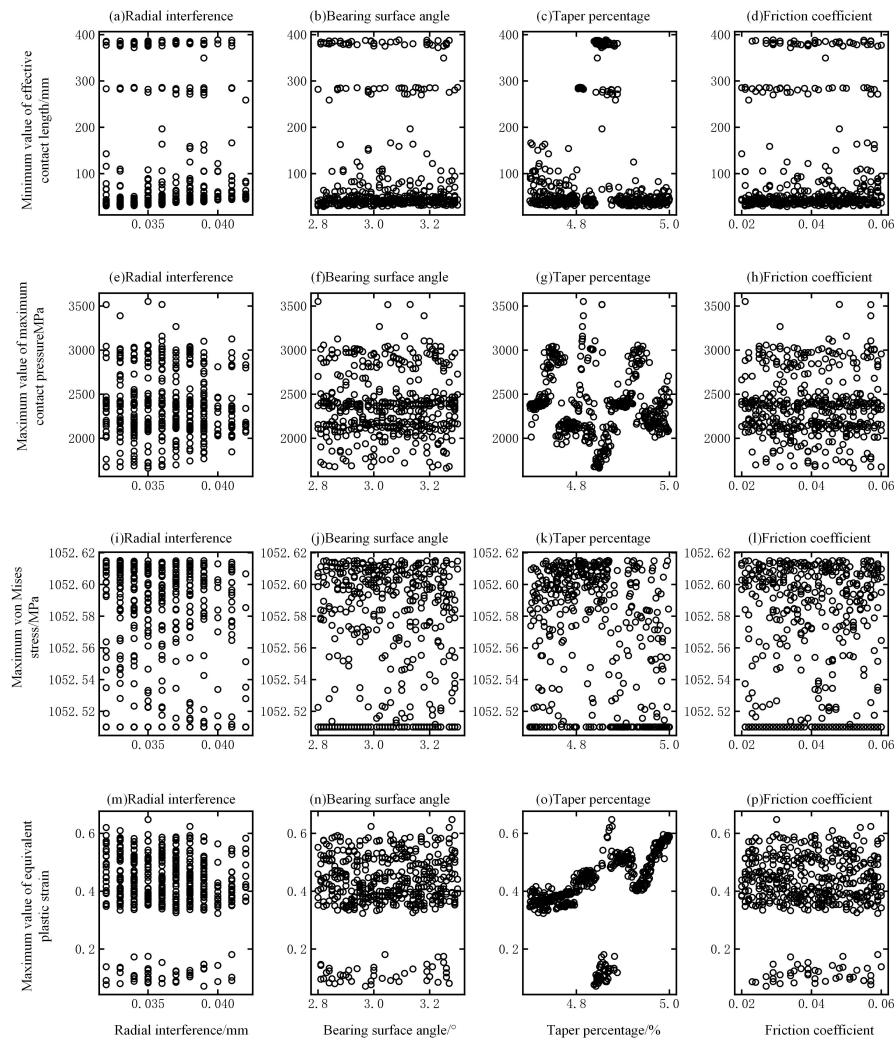


Figure 10 Relationships between machining parameters and key response extrema

As shown in Figure 10, radial interference is positively correlated with minimum effective contact length, with Spearman correlation coefficient of 0.411, indicating that within the parameter range of this paper, increasing radial interference generally helps maintain the effective contact length of seal faces. Its correlations with maximum contact pressure, maximum Mises stress, and maximum equivalent plastic



694 strain are all weak, indicating that radial interference does not show stable monotonic
695 influence on the above responses.

696 The correlation coefficients of load flank angle and friction coefficient with all 4
697 response extrema are relatively low, and no clear upward or downward trends are
698 formed in scatter plots. Therefore, their superiority cannot be judged solely based on
699 single-parameter correlation, and their roles still need to be comprehensively analyzed
700 in combination with other geometric parameters and load conditions.

701 Taper percentage shows the strongest correlation with maximum equivalent
702 plastic strain extrema, with Spearman correlation coefficient of 0.696, indicating that
703 taper variation has a significant effect on plastic deformation levels. Taper percentage
704 shows certain negative correlation with maximum Mises stress extrema, but the
705 overall variation range of this stress indicator is relatively small, and its engineering
706 influence degree cannot be judged solely based on correlation coefficient.

707 **4.3 Parameter Control Insights under Complete Load Spectrum**

708 The most unfavorable states of different sealing and structural responses are not
709 concentrated in the same load stage. Effective contact length, maximum contact
710 pressure, maximum Mises stress, and maximum equivalent plastic strain are each
711 controlled by different load states, indicating that the seal retention capability and
712 local structural response of special thread connections have different load sensitivities.
713 Therefore, machining parameter evaluation should be based on response extrema
714 under complete load spectra, and a single loading condition cannot fully characterize
715 connection service performance.

716 Within the parameter range set in this paper, taper percentage exhibits the
717 highest machining parameter contribution to all 4 prediction responses and is the
718 primary control variable affecting seal contact state and local structural response.
719 Radial interference has a relatively noticeable effect on maximum contact pressure
720 and maximum Mises stress, and can serve as an auxiliary parameter for assembly
721 contact degree and local stress control. The relative contributions of load flank angle
722 and friction coefficient are lower, but this conclusion only applies to the given value
723 ranges and load conditions of this paper.

724 Therefore, special thread machining and dimensional inspection should prioritize
725 control of taper percentage, combined with radial interference for coordinated control
726 of contact pressure and structural stress. Parameter adjustment cannot be based solely



on improvement of single response, but should simultaneously verify effective contact length, maximum contact pressure, maximum Mises stress, and maximum equivalent plastic strain. Existing results can determine the relative control priority of machining parameters and provide a basis for subsequent machining quality control and parameter review.

5 Conclusions

(1) Based on Abaqus parametric finite element analysis and Latin hypercube sampling, a data set containing geometric parameters, load parameters, and contact area state quantities was constructed, and 4 XGBoost prediction models were established. The independent test set coefficient of determination values were 0.999944, 0.812840, 0.780395, and 0.983697, respectively, capable of accurately predicting the sealing and structural response of special thread connections.

(2) Complete load spectrum analysis shows that the controlling load states corresponding to different response extrema are not consistent; within the parameter range of this paper, taper percentage has the highest contribution to all 4 responses, followed by radial interference in its influence on maximum contact pressure and maximum Mises stress. Therefore, machining and dimensional inspection should prioritize control of taper percentage, combined with radial interference for comprehensive control.

Data availability

The data used in this study are not publicly available due to laboratory confidentiality requirements. Relevant information may be made available by the corresponding author upon reasonable request and with permission from the laboratory.

Author contributions

Qi Li and Qingying Yuan proposed the research design and framework of the paper. Mengke Sun drafted the initial manuscript. Jiarui Tang drew the model. Bosheng Qiao and Shuaijun Yin reviewed the entire paper and provided guidance for the revisions. Qingying Yuan contributed to the literature review and model construction.

Competing interests



759 The contact author has declared that none of the authors has any competing interests.

760

761 **Disclaimer**

762 Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional
 763 claims made in the text, published maps, institutional affiliations, or any other
 764 geographical representation in this paper. The authors bear the ultimate responsibility
 765 for providing appropriate place names. Views expressed in the text are those of the
 766 authors and do not necessarily reflect the views of the publisher.

767

768 **Financial support**

769 This work was supported by the National Natural Science Foundation of China
 770 (General Program) under Grant No. 52274006.

771 **References**

- 772 Dai, C., Zhao, J., and Sun, R.: Fatigue Damage Assessment of Offshore Wind Turbine Foundation
 773 Under Coupled Wind-Wave Loading Using Surrogate Modeling, *Energies*, 19, no. 10, 2383, 2026.
- 774 Dou, Y. H., Wang, K., Yu, Y., et al.: Three-dimensional finite element analysis of make-up
 775 performance of special thread tubing connections, *Petroleum Machinery*, 43, no. 4, 99 – 104,
 776 <https://doi.org/10.16082/j.cnki.issn.1001-4578.2015.04.021>, 2015.
- 777 Gao, L. X., Jin, Y., and Zhang, J. Q.: Sealing design of special thread connections for oil country
 778 tubular goods, *Journal of Mechanical Engineering*, no. 3, 216 – 220, 2005.
- 779 Hu, Z. L., Zhang, C. G., Li, X. B., et al.: Thread parameter optimization design of special thread
 780 connections based on ABAQUS, *Steel Pipe*, 47, no. 6, 50 – 57, 2018.
- 781 Li, J., Yan, Y. J., Wang, Z. C., et al.: Study on special thread gas seal performance of integrated string
 782 tubing, *Petroleum Machinery*, 45, no. 8, 92 – 97+116, [https://doi.org/10.16082/j.cnki.issn.1001-](https://doi.org/10.16082/j.cnki.issn.1001-4578.2017.08.019)
 783 [4578.2017.08.019](https://doi.org/10.16082/j.cnki.issn.1001-4578.2017.08.019), 2017.
- 784 Li, W., Liu, J. S., Shi, Z. Q., et al.: Parameter inversion and application of rockfill south water model
 785 based on XGBoost algorithm, *Journal of Water Resources and Hydropower Engineering*, no. 3, 111 –
 786 120, 2023.
- 787 Liu, Y. S., Zhang, J. W., and Huang, H. C.: Key technologies and development directions for deep and
 788 ultra-deep drilling and completion in China, *Acta Petrolei Sinica*, 45, no. 1, 312 – 324, 2024.
- 789 Lu, B. W., Bai, H., Yuan, F. L., et al.: Effect of thread interference on performance of special thread
 790 connections, *Steel Pipe*, 44, no. 4, 49 – 53, 2015.

791



- 792 McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting
 793 values of input variables in the analysis of output from a computer code, *Technometrics*, 42, no. 1,
 794 55 – 61, 2000.
- 795 Orosz, T., Gadó, K., Katona, M., et al.: Automatic tolerance analysis of permanent magnet machines
 796 with encapsulated FEM models using digital-twin-distiller, *Processes*, 9, no. 11, 2077, 2021.
- 797 Shan, H., Gao, L. X., Wang, C., et al.: Study on sealing performance of special thread tubing
 798 connections after repeated make-up/break-out cycles, *Pressure Vessels and Piping*, 40, no. 5, 25 – 33,
 799 2023.
- 800 Song, W., Zhou, Y., Wu, Y., et al.: Study on casing thread sealing performance of high-rate
 801 injection/production underground gas storage wells, *Journal of Southwest Petroleum University*
 802 (Science & Technology Edition), 46, no. 3, 170 – 178, 2024.
- 803 Wu, X. S., Gao, L. X., Yuan, P. B., et al.: Optimization of sealing structure for special thread oil
 804 country tubular goods connections, *Petroleum Machinery*, 45, no. 6, 30 – 35,
 805 <https://doi.org/10.16082/j.cnki.issn.1001-4578.2017.06.007>, 2017.
- 806 Zhang, J. H.: Study on sealing performance and connection strength of special thread connections and
 807 optimization of design parameters, Xi'an Shiyou University,
 808 <https://doi.org/10.27400/d.cnki.gxasc.2021.000734>, 2021.
- 809 Zhu, X. H., Gao, Y., and Jia, Y. J.: Sensitivity analysis of offset trapezoidal casing connection thread
 810 parameters under bending moment load, *Engineering Mechanics*, 29, no. 10, 301 – 307, 2012.