Key Engineering Materials
ISSN: 1662-9795, Vol. 1048, pp 35-50
doi:10.4028/p-gzNY4X

Submitted: 2025-12-18
Revised: 2026-02-13
Accepted: 2026-02-27
Online: 2026-04-14

Compensating Property Fluctuations in Cold Extrusion Using Adaptive Dies

Christian Siedbürger[1,a*], Peter Groche[1,b]

[1]Technical University of Darmstadt, Institute for Production Engineering and Forming Machines, Otto-Berndt-Strasse 2, 64287 Darmstadt, Germany

[a]christian.siedbuerger@ptu.tu-darmstadt.de, [b]peter.groche@ptu.tu-darmstadt.de

Keywords: Cold extrusion; adjustable dies; dimensional accuracy; adaptive tooling; AI surrogate

Abstract This study investigates an adaptive die concept for cold extrusion that actively modulates radial preload during the main forming and ejection phases. A Gaussian process regression (GPR) surrogate, trained on fewer than 400 finite-element simulations, provides a highly data-efficient model capable of accurately predicting geometric tolerances, residual stresses, and process forces. Experimental spot measurements validate the physical trends captured by the surrogate, demonstrating reliable reproduction of the underlying mechanical interactions. The results show that increased preload during forming enables micrometer-level calibration of final diameters, while higher preload during ejection promotes beneficial compressive residual stresses at the cost of elevated ejector forces. A part-to-part control strategy effectively improves accuracy by independently steering two target properties through separate preload adjustments. Furthermore, a reinforcement learning-based controller, enhanced by flow stress estimates derived from hardness measurements, reduces variance and compensates for stochastic fluctuations in material and friction conditions. Overall, the adaptive die system, combined with surrogate- and RL-based control provides a robust foundation for achieving high dimensional precision and stable product properties under future variability scenarios, such as green steel and sustainable lubrication systems.

Introduction

Cold extrusion is widely used for high-volume production of load-bearing components [1]. A representative industrial process is full forward extrusion, in which an axisymmetric billet is pushed through a die to achieve a defined diameter reduction [2,3]. Typical advantages include short cycle times, high reproducibility, strain hardening, and tight dimensional tolerances down to IT7 [2]. Applications range from automotive parts to standardized fasteners and high-strength machine components [4]. These benefits rely on highly stiff tooling systems. However, the same stiffness limits flexibility and the ability to react to changing boundary conditions [5]. In conventional die architectures, the forming insert is shrink-fitted into a prestressed reinforcement and the forming zone is largely inaccessible, which hampers adjustments and the integration of adaptive elements [2].

In current industrial practice, process variability is mainly addressed by restricting the allowable scatter of input variables along the process chain [5]. Yet foreseeable developments suggest that future cold extrusion will face increased input scatter [6,7]. Decarbonization-driven transitions toward scrap-based “green steel” routes can amplify batch-to-batch variability due to fluctuating residual alloying elements, notably copper from recycling streams, which may affect formability and flow stress [6,8–13]. Additional uncertainty arises from supply-chain changes and evolving auxiliaries. Switching suppliers can shift chemistry windows for example between European 16MnCr5 (EN 10084; 1.7131) and the U.S. equivalent SAE 5115 (ASTM A29/A29M) which alters hardenability and can influence room-temperature flow stress [14–16]. Moreover, more sustainable, phosphate-free lubrication systems may increase friction scatter compared with conventional multi-layer systems [17,18]. The combined effect can be elevated scrap and reduced cost efficiency, motivating a shift from robust but passive tooling toward adaptive tooling that actively compensates disturbances while maintaining high reproducibility [5,19,20].

In this study, adaptive tooling refers to forming dies or die subsystems that adjust boundary conditions or effective tool geometry during a stroke or from one stroke to the next (Fig. 1) [20]. Adjustments are executed by actuators under supervisory rule-based or feedback control, aiming to keep part properties within targets despite disturbances such as material scatter. Typical manipulated variables in forming include blank-holder force, drawbead penetration, die/cavity size and preload, and the effective compliance of punch or die [19,20].

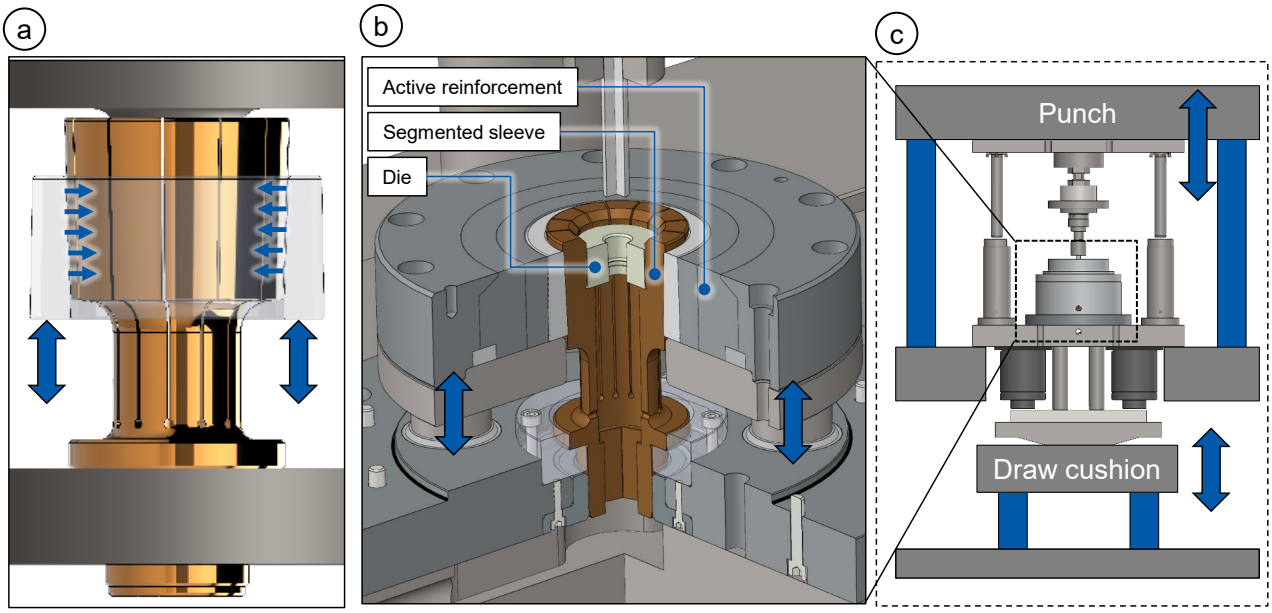

Fig. 1. Overview of the adaptive tooling system ("active die") used in this work: (a) Mechanism converting axial actuation into radial die contraction; (b) Active-die architecture; (c) Operation on a press with a spindle-driven punch and a draw cushion.

For cold forward extrusion, three complementary strategies are relevant. First, an adaptive punch can modify bottom dead center per stroke to improve length accuracy, potentially supported by in-line measurement and data-driven set-point selection [21,22]. Second, counterpressure via the ejector pin can increase hydrostatic stress and homogenize material flow, reducing tensile surface stresses and mitigating damage (e.g., chevron cracking), albeit at the cost of higher press forces and increased die contact pressure [23–25]. Third, an adaptive die can adjust its inner diameter by varying radial preload, enabling micrometer-scale die shrinkage [19,25]. This can be realized, for example in Fig 1, by axial actuation of a conical reinforcement that converts axial input into a radial adjustment of the die [26,27]. An implementation at PtU (TU Darmstadt) in cooperation with Strecon A/S uses a slotted conical sleeve that generates variable external pressure on the die via a spindle-driven draw cushion [28]. Prior work has shown that such adaptive dies can influence dimensional accuracy, forming-induced residual stresses, and ejector forces [25,28]. Although adaptive punches and counterpressure concepts can improve specific aspects of the process, the active die provides the most direct and scalable lever for controlling the critical forming-zone boundary conditions in cold extrusion. By adjusting die preload separately during forming and ejection, the concept offers two practical control inputs that can be integrated into conventional die sets.

Against this background, the research question is how to use an active die to compensate variability arising from green-steel routes and more sustainable lubrication systems, particularly when fluctuations are uncertain and their propagation to part properties is complex. To address this, a machine-learning surrogate is trained on a large ensemble of finite-element simulations and validated against representative experiments. To emulate real production scatter, an uncertainty generator models the dominant disturbance sources as probabilistic variations of the input features (material, friction and elastic properties), which are then propagated through the surrogate. The surrogate enables efficient Monte Carlo studies of adaptive steering policies under realistic distributions of material and tribological scatter, providing (i) a quantitative assessment of candidate strategies across

fluctuation bands and (ii) a transparent mapping of interactions between steering parameters and key product properties. Ultimately, this supports production-scale deployment of active die concepts for stable dimensional accuracy and residual-stress targets under realistic variability.

Methodological Approach

A simulation-based workflow (Fig. 2) is employed to quantify how input uncertainties propagate to product properties and to construct a prediction model of the active-die system. The trained model serves as a computational surrogate for extensive physical trials, enabling cost- and data-efficient exploration of the tool behavior under production-relevant conditions. The overall methodology comprises seven steps:

(i) Setup of an FE model of the forming operation with the active die,
(ii) parameterization over wide but physically plausible ranges using Monte Carlo sampling to ensure domain coverage and to reduce extrapolation risk for the surrogate,
(iii) training and generalization of the surrogate model,
(iv) plausibility checks of the surrogate by representative spot experiments, verifying that the target variables can be systematically manipulated by active-die preload settings,
(v) generation of uncertainty scenarios based on expected scatter of input variables and fast propagation through the surrogate,
(vi) evaluation of different control strategies using the adjustable die preload during forming and ejection as control variables, and
(vii) assessment of the resulting distributions of selected product properties.

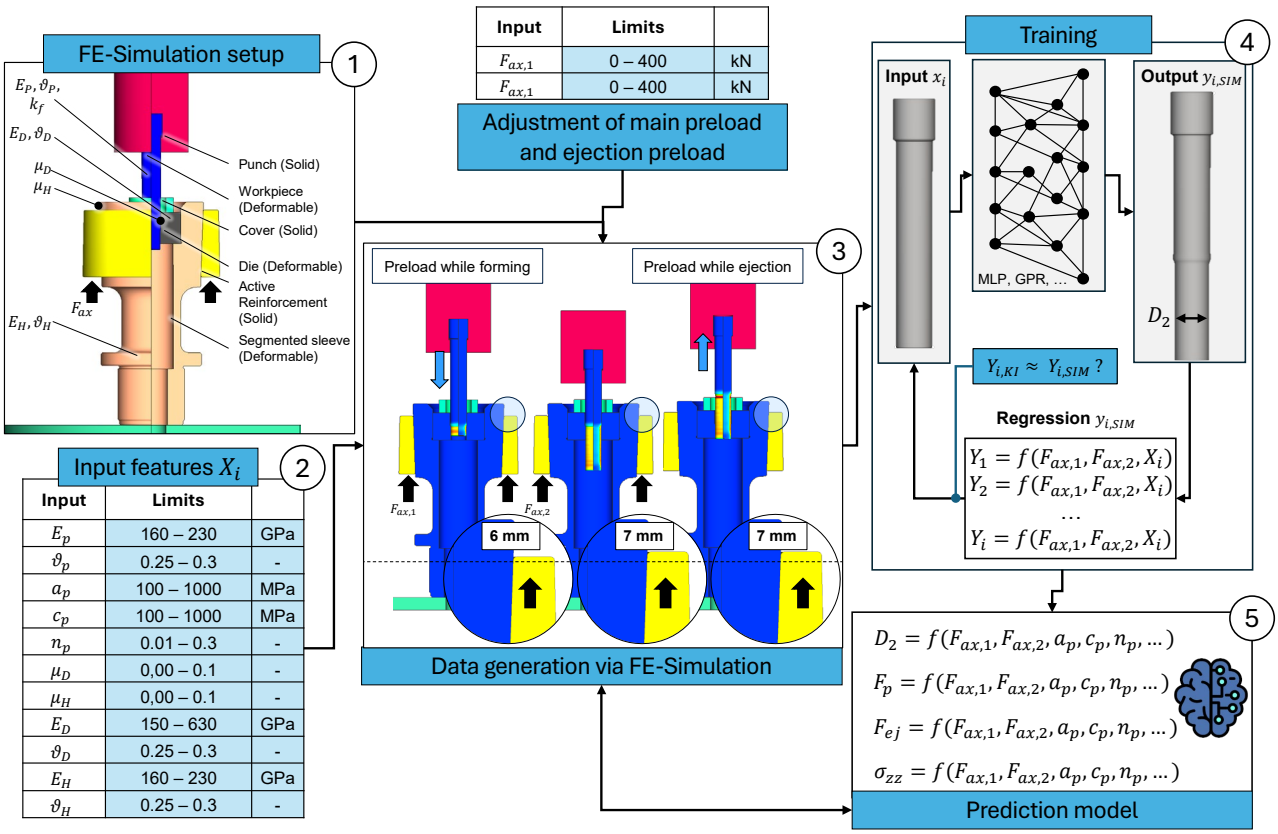

Fig.2. Overview of the workflow used to create the surrogate model for the active die system: 1.) Parametric FE simulation model,2.) Input features defining material, friction, and tool properties, 3.) Forming process with adaptive preload adjustment during forming and ejection, 4.) Training of surrogate models, 5.) Prediction model

FE model setup: The case study is a forward extrusion reduction in which an axisymmetric billet is reduced from an initial diameter of 13.000 mm to 12.000 mm. The two-dimensional (axisymmetric) FE model was implemented in SimufactForming using an implicit solver and follows a previously validated setup with consistent discretization, contact formulation, and solver settings [25]. The punch motion is prescribed by a tabulated time–stroke profile. Since the specimens are clamped at the punch, the punch is also used to represent the ejection phase within the same kinematic sequence (Fig. 2.3). The billet is modeled as plastically deformable and discretized with quadrilateral elements. To represent the preload-induced change of the die inner diameter, the die insert and the segmented sleeve are modeled as elastic bodies, whereas the remaining components are treated as rigid (Fig. 2.1). Radial preload is controlled by axially advancing the conical reinforcement. In the model this is implemented as a tabulated spring force over process time in the range of 0–400 kN, applied during the forming $F_{ax,1}$ and ejection phases $F_{ax,2}$ (Fig. 2.1). A mesh-sensitivity study was performed to ensure numerical resolution of micrometer-scale diameter changes. For element sizes of 0.3 mm and below, no relevant changes were observed in either the simulated ejection force or the mean diameter evaluated in the reduced zone. Further mesh refinement led to fluctuations below 0.1 kN in ejection force and below 0,5 µm in the averaged diameter metric. To stabilize the contact conditions, the same mesh resolution was applied to the elastically deformable die insert and the segmented sleeve, ensuring consistent contact discretization across the interfaces. To keep the evaluation of the reduced zone reproducible across the wide parameter variation, remeshing was disabled, since remeshing events can locally perturb the diameter averaging and thereby introduce additional numerical scatter that would deteriorate surrogate generalization.

Experimental setup. To verify the accuracy and physical plausibility of the simulation-based prediction, spot experiments were conducted on the physical tool system by systematically varying the axial set forces $F_{ax,1}$ (main forming preload) and $F_{ax,2}$ (ejection preload), cf. Fig. 1. Cylindrical billets of 16MnCr5 in the soft-annealed condition (8 h holding time at 700 °C) were used. To improve surface conditioning and reproducibility, all contact surfaces were sandblasted prior to applying a polymer-based lubricant. The billets were subsequently processed in the active-die forward extrusion setup (Fig. 2.3) and reduced from the initial to the target diameter at a stroke rate of 1 stroke/s. Force variations in $F_{ax,1}$ and $F_{ax,2}$ were imposed via the draw cushion during forming and ejection (Fig. 1). In total, seven experiments were conducted with different combinations of $F_{ax,1}$ and $F_{ax,2}$ to compare the measured responses against the simulation-based predictions across the control-input space.After cleaning and cooling, the final diameters were measured using a Keyence TMX-40 camera system. For each part, 50 individual measurements were taken along the reduced zone and averaged to obtain the reported diameter. The system resolution is 1,5 µm, and the specified reproducibility of the TMX-40 is below 0.1 µm. Press and ejection forces were recorded using an in-die HBM force sensor with an accuracy of 0.2 kN. Since ejection is realized as pull-out in this tool, the same sensor signal was used to evaluate both the press-force and ejection-force responses.

Material modeling. For the flow-stress description of 16MnCr5 used in Fig. 3, a simplified isotropic hardening law $k_f = a_p + c_p \cdot \varphi^{n_p}$ was adopted, since the investigated process was conducted at 1 stroke/min. Consequently, temperature effects and strain-rate sensitivity were neglected. The flow-curve parameters a_p, c_p, n_p were identified from upsetting tests via an iterative calibration approach: Experimental compression data were reproduced numerically, and the coefficients of the flow-stress law a_p, c_p, n_p were adjusted until the simulated force–stroke curve matched the experimental reference. The resulting mean coefficients a_p, c_p, n_p of the calibrated flow curve for 16MnCr5 are summarized in Fig. 3 (top left).

Data generation. The parameterization of the FE setup exposes eleven inputs X_i that are freely configurable in the simulation (Fig. 2.2). Five inputs describe the workpiece material: three parameters a_p, c_p, n_p of an isotropic flow law together with the elastic modulus E_p and Poisson's ratio ϑ_p. Two inputs define friction according to Coulomb's law, namely μ_D at the die–workpiece interface and μ_H at the interface between the conical reinforcement and the segmented sleeve. The remaining four inputs are the elastic modulus and Poisson's ratio of the reinforcement and of the

sleeve. An in-house Python interface to SimufactForming automates model generation, execution, and post-processing for the relevant targets. To transform the simulation into a data-driven prediction model, the input features listed in Fig. 2.2 are sampled to create a dataset for regression of process and product responses. While the baseline values are derived from the experimentel setup, the bounds were defined to represent realistic industrial variability in material, friction, and elastic properties (e.g., batch-to-batch scatter of soft-annealed 16MnCr5, lubrication/surface-condition changes, and additional uncertainty expected for "green-steel" supply chains). For the flow-law parameters a_p, c_p, n_p, the ranges span plausible variations in flow-stress level and hardening behavior and include a conservative safety margin to avoid surrogate extrapolation when multiple inputs simultaneously approach extreme values. Therefore, the parameter ranges in Fig. 2.2 were intentionally chosen comparatively wide to ensure sufficient feature-space coverage and robust surrogate predictions within the intended domain.

Sampling strategy. Sampling follows a Monte Carlo strategy within the specified bounds to generate diverse input combinations and thereby reduce the need for extrapolation by the surrogate. This choice aligns with the subsequent analyses, which evaluate the surrogate under randomly varying material and tribological conditions, and it allows straightforward incremental extension as additional FE simulations become available. Broad bounds extend the usable domain and further reduce extrapolation risk, whereas tighter bounds can improve local accuracy in well-known operating regions. While space-filling designs such as Latin hypercube sampling or Sobol sequences can yield more uniform coverage for a fixed simulation budget, Monte Carlo was selected here as a pragmatic and scalable approach in the high-dimensional setting. Moreover, since the later robustness analyses assume approximately Gaussian process variability, Monte Carlo sampling provides a direct and consistent representation of these distributions.

Surrogate model. Multi-output regression models are trained to map the FE inputs X_i and the two control settings $F_{ax,1}$ (main forming) and $F_{ax,2}$ (ejection) to the responses Y_i (Fig. 2.4). The candidate learners include a multilayer perceptron (MLP), Gaussian process regression (GPR), and a random forest (RF) regressor. The outputs Y_i comprise the final diameter in the forming zone, press force, ejector force, and forming-induced residual stresses (Fig. 2.5). Accordingly, the surrogate mapping is formulated as a multi-output function

$$Y_i = f(F_{ax,1}, F_{ax,1}, X_i) \tag{1}$$

These quantities are chosen because they are directly influenced by the active die system and have been demonstrated to be adjustable in prior studies [19,25]. The dataset of FE simulations, generated under the assumptions in Fig. 2, was split into training and test subsets using a conventional 80/20 partition. The choice of N = 400 was guided by an initial learning-curve study using Monte Carlo sampling, in which the surrogate was trained with increasing dataset sizes until the predictive performance on a held-out set plateaued. Beyond N ≈ 400, additional simulations yielded only marginal improvements. Model quality was assessed with standard regression metrics (MSE, RMSE, MAE, R^2). The response surfaces exhibit predominantly smooth, low-noise behavior across all targets, favoring kernel methods and well-regularized neural networks for surrogate modeling.

Uncertainty propagation. Replacing the high-fidelity FE model with a validated multi-output regression surrogate enables deterministic prediction of the responses in a fraction of the time required by the FE setup (on the order of milliseconds to seconds per query versus approximately one hour per configuration for the verified FE model). This acceleration permits efficient exploration of large combinations of input features and preload set points, provided that the AI-based predicition model (Fig. 2.5) reproduces the FE ground truth with sufficient accuracy. In real production, the forming system is subject to numerous disturbance sources that translate into uncertain inputs and, consequently, uncertain outputs. Typical contributors include temperature variations, surface roughness and lubrication variability in the contact zone, part-to-part inhomogeneities, geometric and positional tolerances in the tool system, and actuation variability during the press stroke, which can be quantified by ΔX. [19,29]

Let a generic product property be denoted by the random variable Y_i. For a large number of trials, Y_i is modeled as approximately Gaussian, consistent with common engineering processes, so that

$$Y_i = \bar{Y}_i + \Delta Y_i \tag{1}$$

In this notation, the nominal level is $\bar{Y}_i$ and the uncertainty is summarized by ΔY_i. A process is termed robust when ΔY_i is small enough that Y_i remains within the specified tolerance window with the desired probability. Equivalently, given an allowable spread ΔY_{Limit} (derived from the tolerance and the chosen capability index), the process is robust if $\Delta Y_{Limit} \geq Y_i$. Within this work, the adaptive tooling system is treated as a flexible process because the two control inputs (the radial preload during the main forming phase and during ejection) can be used to influence both $\bar{Y}_i$ and ΔY_i of the relevant properties [20,29]. In particular:

1. by appropriate choice of $(F_{ax1},\ F_{ax2})$ the mean $\bar{Y}_i$ of each target property (final diameter in the forming zone, residual-stress descriptors, ejector force) can be shifted toward its set point while simultaneously reducing the spread ΔY_i; and
2. as a consequence, the interval $[\bar{Y}_i - \Delta Y_i, \bar{Y}_i + \Delta Y_i]$ can be tightened and centered within the tolerance limits, enabling higher conformance rates and, where applicable, narrower achievable quality bands.

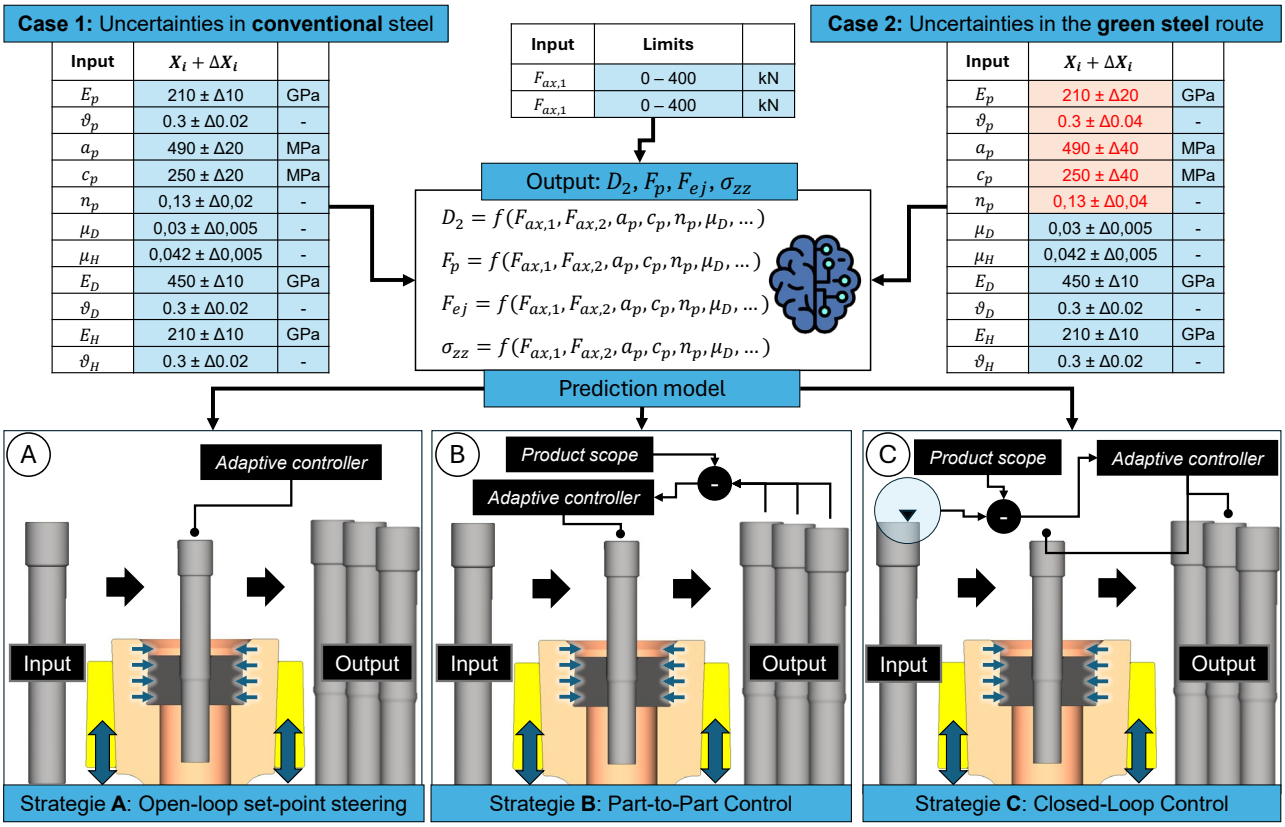

Fig.3. Input parameter uncertainties for the conventional-steel and green-steel scenarios used in the prediction model, together with the three adaptive control strategies (A–C).

Uncertainty scenarios ΔX_i and control-policy evaluation. Figure 3 summarizes the assumed input uncertainties ΔX_i and how they enter the prediction model. Two scenarios are considered. The first scenario represents a reference steel 16MnCr5. The assumed variations for the material coefficients, the elastic modulus E_p and Poisson's ratio ϑ_p reflect typical datasheet scatter for the annealed condition of this grade. The two friction coefficients are treated as Coulomb coefficients. The coefficient μ_D at the die–workpiece contact is derived from sliding compression tests. The

coefficient μ_H at the reinforcement–sleeve contact is identified by matching the characteristic force–stroke curve from experiments in Figure 1 to FE simulations with different μ_H, which yields $\mu_H \approx 0.03$ for the best fit. The uncertainty $\Delta\mu_i$, is set conservatively to 0.02 based on repeat tests with the polymer lubricant used in the experimental setup. For the carbide die (grade G40) and the segmented sleeve (tool steel 1.2343), E, ϑ and their ΔX_i, are estimated using industrial datasheets.

The second scenario ("green steel") inherits all uncertainties from the reference case but enlarges the workpiece-material uncertainties to reflect copper variability in scrap-based metallurgy. The assumption is that the copper content may vary between 0 and 1 percent. Reported trends indicate that copper contamination near 1 percent can increase yield strength and ultimate tensile strength by roughly 200 MPa compared with copper-free conditions [13]. To represent this elevated scatter without committing to a specific metallurgy model, the ΔX_i for the workpiece material $a_p, c_p, n_p E_p, \nu_p$ are doubled in this scenario.

Target setting. For a given scenario, Monte-Carlo sampling draws X_i from the specified bounds $X_i + \Delta X_i$. The surrogate model then evaluates the mapping $(X_i, F_{ax,1}, F_{ax,2}) \rightarrow Y_i$. With N = 1000 samples, the empirical mean $\bar{Y}_i$ and standard deviation ΔY_i, are obtained for each response. The multi-objective target is defined by a diameter set-point of $D_{target} = 12.000\,mm$ and by ejection conditions that promote compressive near-surface residual stresses. Previous investigations have shown a clear correlation between the ejector force and the forming-induced residual-stress state, which can be explained physically by the additional deformation imposed during pull-out: increasing the ejection preload $F_{ax,2}$ raises the radial confinement of the die during ejection, thereby intensifying near-surface plastic straining and reducing tensile residual stresses, potentially shifting them into compression. This mechanism and the associated ejector-force–residual-stress relationship have been quantified in prior work [X]. Furthermore, the FE simulations indicate that the dependence of the relevant residual-stress descriptor on F_{ej} can be approximated well by a quadratic function within the studied operating range. To deliberately drive the residual-stress state toward compression beneficial for fatigue performance and reduced distortion an ejector-force target of $F_{ej,target} = -5{,}5\,kN$ is selected.

Control strategies. Three control strategies are applied to the surrogate to assess compensation capability under both scenarios:

Open-loop set-point steering. A constant pair $(F_{ax,1}, F_{ax,2})$ is selected from the surrogate without feedback from previous strokes.

Part-to-part adaptive control. After each stroke, the diameter and ejector force are compared with their targets and the set points are updated by discrete increments ΔF within the bounds, which implements a run-to-run correction using only the last part's measurements.

Inline closed-loop control with hardness-based feedforward and Q-learning. Before forming, an in-process hardness measurement provides an estimate of the flow stress $k_f(0.3) = a_p + c_p \cdot 0.3^{n_p}$. This estimate supplements the state used for decision-making. In MATLAB, a neural action-value function is trained using episodic rollouts of the prediction model. The state includes the hardness-based flow-stress estimate, the last set points, and the previous part's diameter and ejector-force errors. The actions are the discrete changes, $F_{ax,1}$ and $F_{ax,2}$, within the allowable range. An ε-greedy policy with replay and target-network stabilization is used. The reward penalizes the squared deviations from D_{target} and $F_{ej,target}$ adds soft penalties for exceeding the press and ejector-force limits. Training proceeds until the policy stabilizes on validation seeds. Then, the learned policy is deployed stroke-wise.

The resulting output distributions are compared across strategies to demonstrate how each policy maintains its targets under the increased variability of the green-steel scenario. This allows us to quantify improvements in the mean level and the spread of the key properties.

Results

Surrogate accuracy and model selection.

Based on the N = 400 Monte-Carlo-generated FE dataset (Fig 2), three candidate surrogate learners (GPR, MLP, and RF) were benchmarked to identify the most suitable prediction model for subsequent uncertainty propagation. Gaussian process regression (GPR) delivers the best or near-best generalization across three of the four targets and, critically, provides predictive uncertainty via its posterior variance directly usable for Monte-Carlo propagation, robust set-point selection, and capability assertions. GPR also trains stably on moderate data volumes with minimal tuning and yields interpretable length-scales that reflect process sensitivities. For diameter, it attains RMSE = 0.0014 mm, R^2 = 0.978 on the test set. The multilayer perceptron (MLP) is a strong alternative that proves competitive overall. In particular, it achieves the best ejector-force prediction with RMSE = 1.25 kN, R^2 = 0.782, whereas GPR yields 1.49 kN, R^2 = 0.694 for that target.

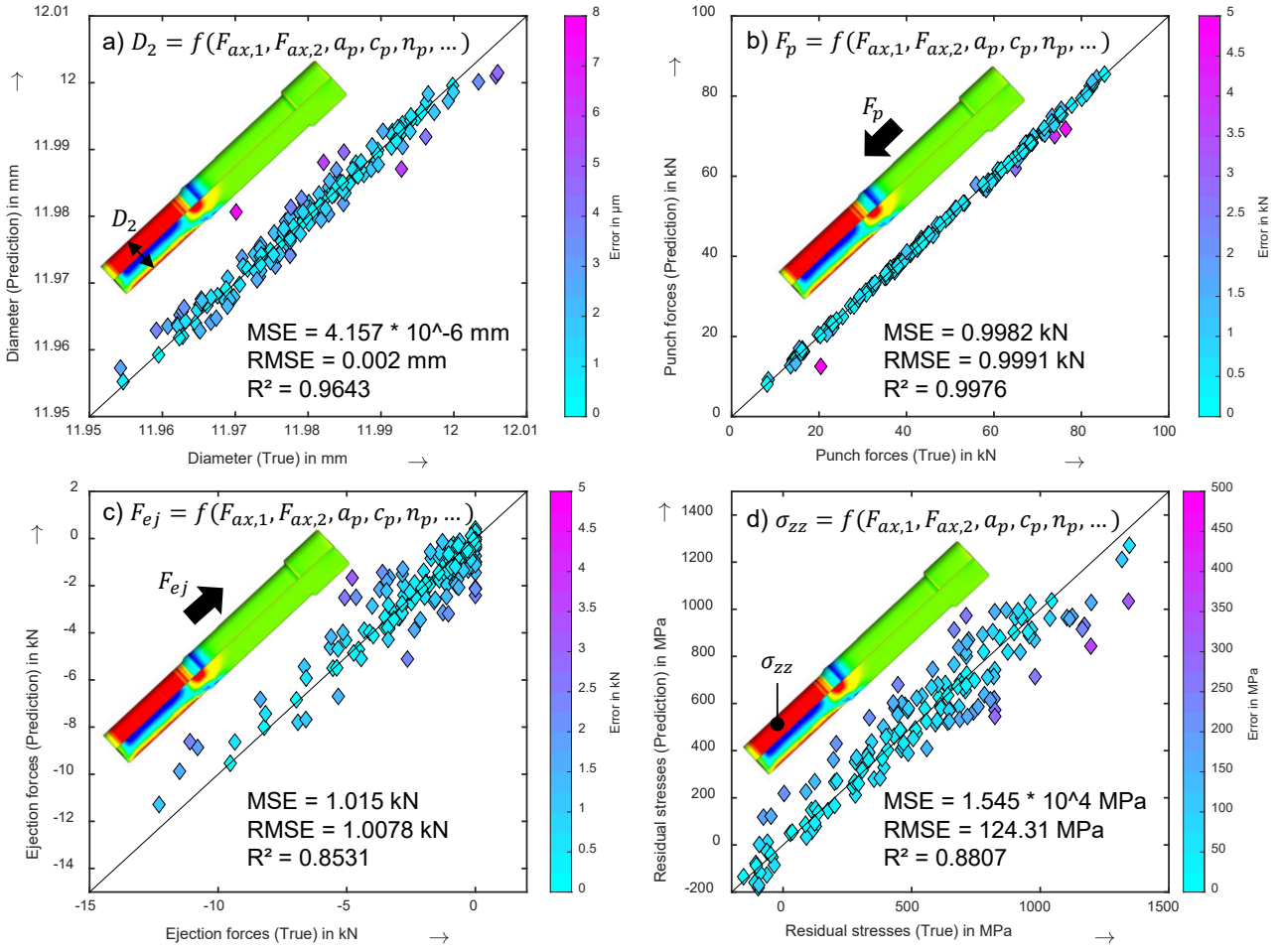

Fig.4. Parity plots of the refined GPR surrogate for (a) diameter, (b) press force, (c) ejector force, and (d) residual stresses; color encodes absolute error.

In contrast, the configured random forest (RF) surrogate underperforms. Bagged trees with large minimum leaf sizes produce piecewise-constant approximations that do not align well with the smooth, continuous sensitivities of the active-die system, leading to high bias. Simply increasing the number of trees does not remedy this structural limitation and reducing leaf size at the available data scale risks excessive variance. Given its consistently strong fidelity, native uncertainty quantification, and data efficiency, GPR is selected as the primary surrogate for the prediction model used in the uncertainty analysis, robust set-point design, and control-policy studies.

After systematic hyperparameter sweeps (length-scale, signal variance, noise level, mean function), a squared-exponential GPR with constant mean and automatic noise scaling was selected. The parity plots in Fig. 4 show that the surrogate reproduces the global trends for all targets.

A few outliers remain at the sparsely sampled edges of the domain. This is consistent with the deliberately broad training ranges and with locally higher posterior variance. These deviations are physically plausible, since the responses become sensitive to contact transitions and friction at extreme settings. Despite additional tuning, the prediction error for ejector force could not be reduced below about 1 kN. This level is comparable to the error observed for punch force and is mainly driven by sensitivity to the reinforcement–sleeve friction and by stick–slip or lift-off phenomena, which introduce heteroscedastic label noise in the FE data. Residual stresses show a similar saturation because of steep near-surface gradients and path dependence during ejection. Further improvement would therefore require targeted enrichment of the dataset in high-variance regions or additional features that encode contact state.

For the purposes of this work, the surrogate is sufficient. Since the goal is not to reproduce an industry-grade forming process in all details, but to capture the sensitivity of the relevant responses Y_i to the two control variables $F_{ax,1}$ and $F_{ax,2}$ on the basis of an FE model validated for plausibility. Diameter and press-force predictions provide the fidelity required for robust set-point design, while ejector-force and residual-stress predictions capture the relevant trends and magnitudes and are accompanied by GPR uncertainty that is propagated in the Monte-Carlo studies. Further improvements (e.g., larger datasets, refined friction modeling, or explicit geometric uncertainties) could increase external validity but are beyond the scope here. Therefore, the refined GPR is adopted as the primary surrogate for the subsequent uncertainty and control analyses.

Plausibility check of the prediction model. For comparison between experimentel spot experiments and the results in surgerate model, the GPR surrogate was evaluated on a $(F_{ax,1}, F_{ax,2})$ grid to form 3D response maps, and experimental points were overlaid with color indicating absolute error (Fig. 5). For these maps, all remaining input features X_i were fixed to the mean values reported in Fig. 3 for 16MnCr5 and the corresponding experimental setup. Only $F_{ax,1}$ and $F_{ax,2}$ were varied to span the response surface shown in Fig. 5. The agreement is good for diameter and ejection force (diameter: RMSE = 1.8 µm, R^2 = 0.85, ejection force: RMSE = 0.369 kN, R^2 = 0.762). Larger deviations occur for punch force. These outliers are attributed to tribological variability of the single-layer lubricant and minor stock inhomogeneity. In addition, the main-stroke preload $F_{ax,1}$ alters the die by only about 15 µm, so the overall forming degree and thus the press force is dominated by friction and material scatter rather than by the control variables $F_{ax,1}$ and $F_{ax,2}$.

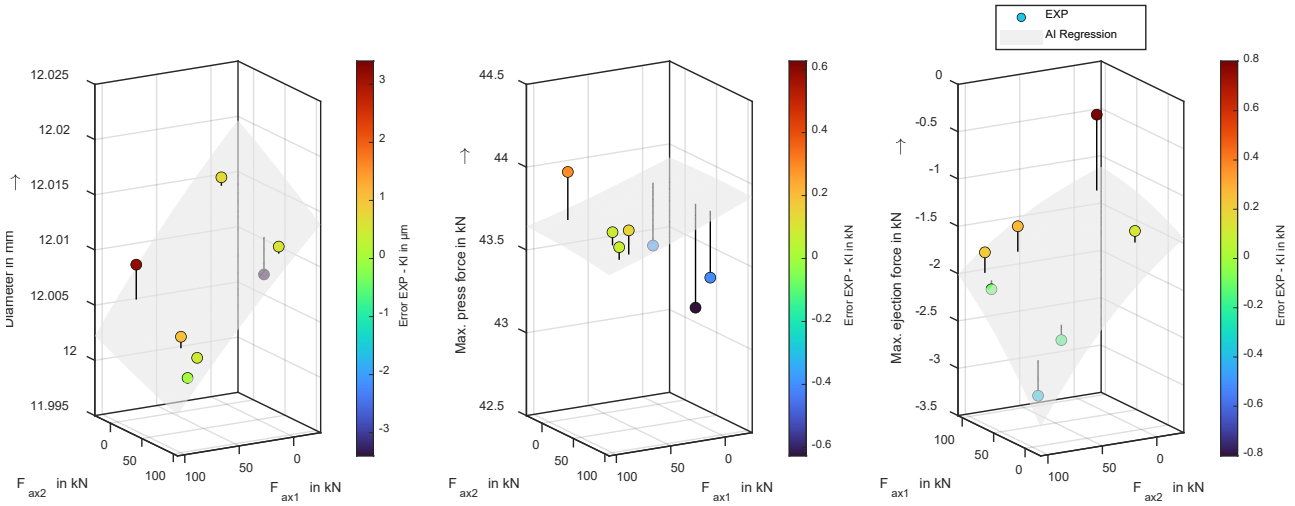

Fig.5. Experimental verification of the surrogate model: parity surfaces from the GPR surrogate with overlaid experiments for (left) diameter, (center) press force, (right) ejection force; color encodes absolute error for each metric.

The maps also clarify the process trends. Increasing either $F_{ax,1}$ or $F_{ax,2}$ reduces the final diameter in the forming zone. $F_{ax,1}$ tightens the die before forming, $F_{ax,2}$ raises clamping pressure during ejection and enhances calibration. Ejection force is governed primarily by $F_{ax,2}$ and by the difference ΔF_{ax}. Low $F_{ax,1}$ with high $F_{ax,2}$ yields the highest ejection forces, while high $F_{ax,1}$ with low $F_{ax,2}$ produces the lowest. Higher ejection forces increase local near-surface plasticity in the calibration zone, shifting residual stresses toward compression (beneficial for fatigue and distortion), at the expense of higher contact pressures and tool load. This is an explicit trade-off to be managed in the control strategies. [25,28]

The comparison in Fig. 5 reflects not only the surrogate error (cf. Fig. 4), but also accuracy limits of the underlying FE model and the experiments. In this workflow, the FE simulations serve as ground truth within the investigated domain, mainly to learn sensitivities with respect to $F_{ax,1}$, $F_{ax,2}$ and X_i. Absolute agreement at the micrometer level is difficult because the predicted diameter changes are comparable to (i) mesh and contact discretization effects, (ii) solver tolerances, and (iii) experimental scatter caused by surface roughness, local adhesion, and lubrication variability. Therefore, the FE–surrogate chain is assessed primarily by trends and response gradients rather than absolute diameters. This is also why an offset was applied when comparing the idealized FE geometry (12.000 mm) to the measured tool dimension (~12.018 mm ± 2 µm). Reducing the remaining discrepancy would require further model refinements (e.g., improved friction characterization and modeling, material calibration, and contact or geometric-uncertainty modeling), which is beyond the scope of this contribution. For the intended uncertainty propagation and adaptive control design, the fidelity is sufficient because the model reproduces the direction and magnitude of the controllable effects of $F_{ax,1}$ and $F_{ax,2}$ on the relevant responses Y_i.

Application of Control Strategy A. Building on the trends in Fig. 5, strategy A fixes constant set points for the two actuators and evaluates N = 1000 parts per scenario via Monte-Carlo sampling of the input uncertainties. Fig. 6 reports probability distributions for absolute diameter, IT class, ejection force, and axial residual stress.

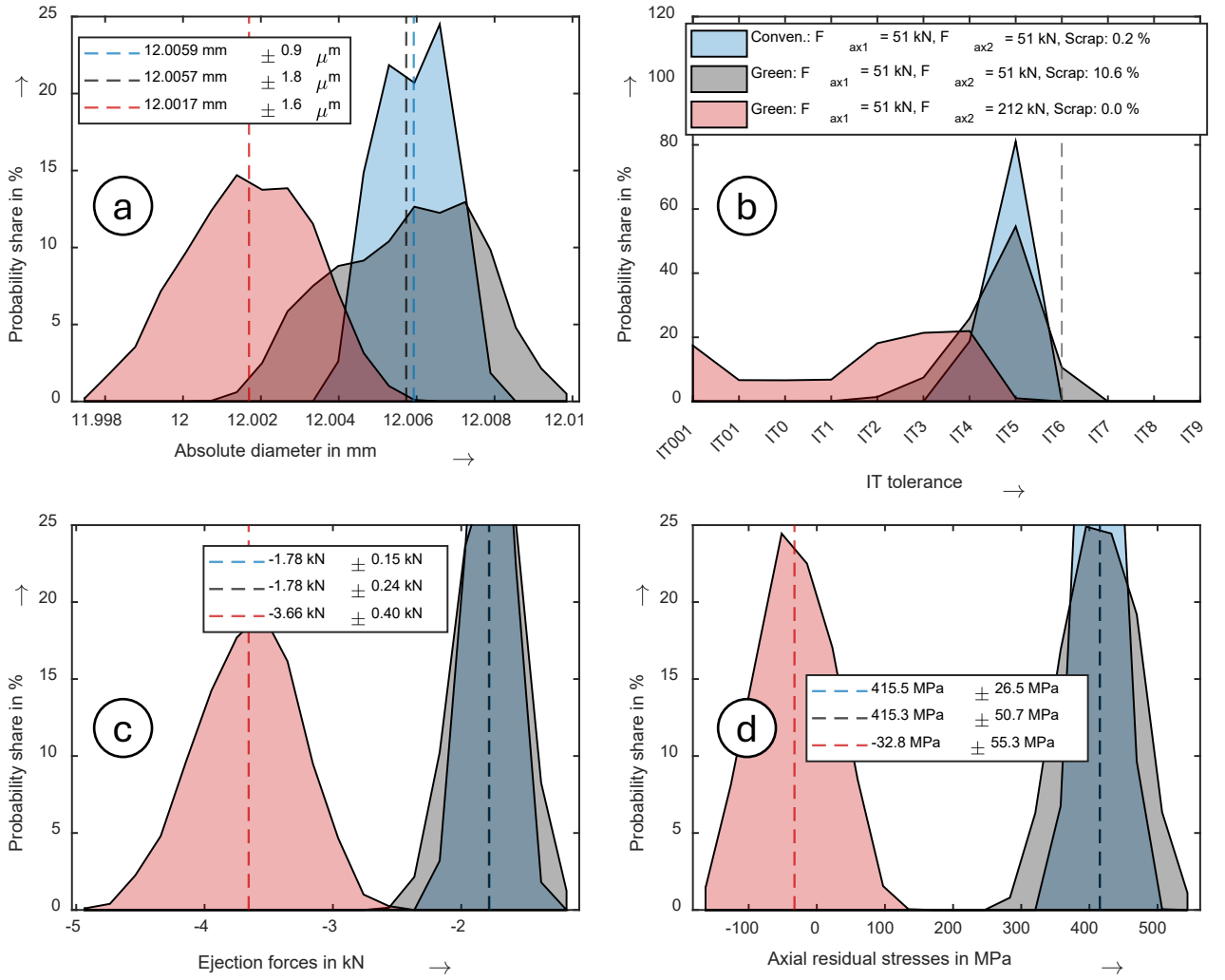

Fig.6. Effect of open-loop set-point steering (Strategy A) on product properties under conventional steel. Distributions are shown for (a) absolute diameter, (b) IT class with IT6 indicated, (c) ejection force, and (d) axial residual stress; dashed lines denote means. Each case uses N = 1000 Monte-Carlo samples.

With conventional steel (16MnCr5) at blue, the process centers at 12.0059 mm ± 0.9 µm for diameter, −1.78 kN ± 0.15 kN for ejection force, and 415.5 MPa ± 26.5 tensile axial residual stress. The share above IT6 corresponds to 0.2 % scrap. Substituting green steel while keeping the same set points (gray) leaves the means essentially unchanged (e.g., diameter 12.0057 mm), but inflates variance as expected from the larger input scatter. The diameter spread increases to ± 1.8 µm, the ejection-force spread to ± 0.24 kN, and the residual-stress spread nearly doubles to ± 50.7 MPa. The broadened IT distribution pushes scrap to 10.6 %, confirming that the dominant effect of green-steel variability is a growth in standard deviation rather than a systematic bias of the mean.

A simple set-point change can re-center the outcomes. Raising only the ejection-phase preload to $F_{ax,2} = 212\ kN$ while keeping $F_{ax,1} = 51\ kN$ shifts the distributions in the desired directions. The diameter moves closer to target, the magnitude of the ejection force increases, and the axial residual stress is driven into compression. Mechanistically, the higher $F_{ax,2}$ elevates clamping pressure in the calibration zone during ejection, inducing additional near-surface plasticity, which both reduces the final diameter and shifts residual stresses toward compression. Crucially, the IT distribution tightens around the target and the scrap fraction drops to 0.1 % despite the green-steel variability.

Overall, Strategy A is effective for mean correction under known variability, but it has limited leverage on variance. Comparing the red and gray cases shows similar standard deviations for diameter and residual stress, with only modest changes for ejection force. This motivates the use of

part-to-part (B) and inline strategies (C) in the next section, where variance reduction and mean re-centering are required.

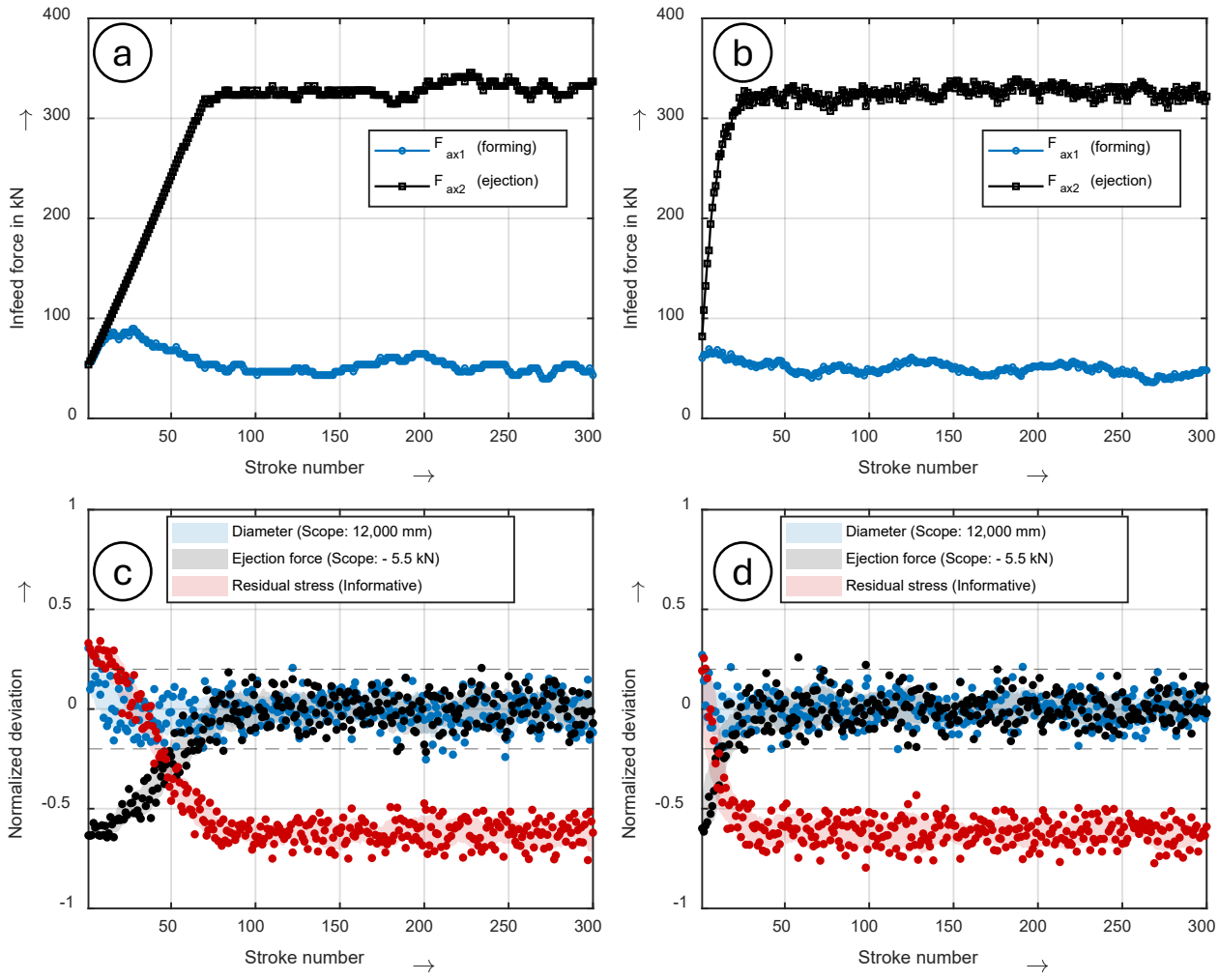

Fig.7. Comparison of the control trajectories starting from Fax1 = Fax2 = 50 kN for control strategies B and C. (a) Evolution of the infeed forces Fax1 and Fax2 used as control variables in Strategy B. (b) Evolution of the infeed forces Fax1 and Fax2 used as control variables in Strategy C. (c) Normalized target deviations for Strategy B. (d) Normalized target deviations for Strategy C.

Optimization with Strategies B and C. Building on the trends derived in Fig. 5, Strategies B and C are evaluated under green-steel variability to enhance the accuracy and robustness previously achieved with Strategy A. For interpretability, Fig. 7 presents stacked trajectories in which the actuator set points $F_{ax,1}$ (forming) and $F_{ax,2}$ (ejection) are displayed together with the normalized outputs for diameter, ejection force, and axial residual stress in subfigures (c) and (d). Normalization maps each response to its target. Residual stress is referenced to 400 MPa, so values near zero indicate high accuracy. Strategy B implements a bounded part-to-part adjustment law with discrete actions $\Delta F \in \{-5, 0, +5\}$ kN. After each stroke, $F_{ax,1}$ is updated based solely on the previous part's diameter error until the normalized deviation enters a ± 5 % band, after which the actuator is held constant. $F_{ax,2}$ is corrected analogously from the deviation in ejection force by increasing the difference $F_{ax,1} - F_{ax,2}$ whenever the magnitude of the measured ejector force is too small, again with a ± 5 % deadband. Starting from $F_{ax,1} = F_{ax,2} = 50\ kN$, this run-to-run policy converges within approximately 75 strokes to operating points fluctuating around $F_{ax,1} = 50\ kN$ and $F_{ax,2} = 350\ kN$. Strategy C augments this scheme with inline information and a learned decision layer: before each stroke a hardness-based flow-stress estimate $k_f(0{,}3)$ is fed forward into a Q-learning controller.

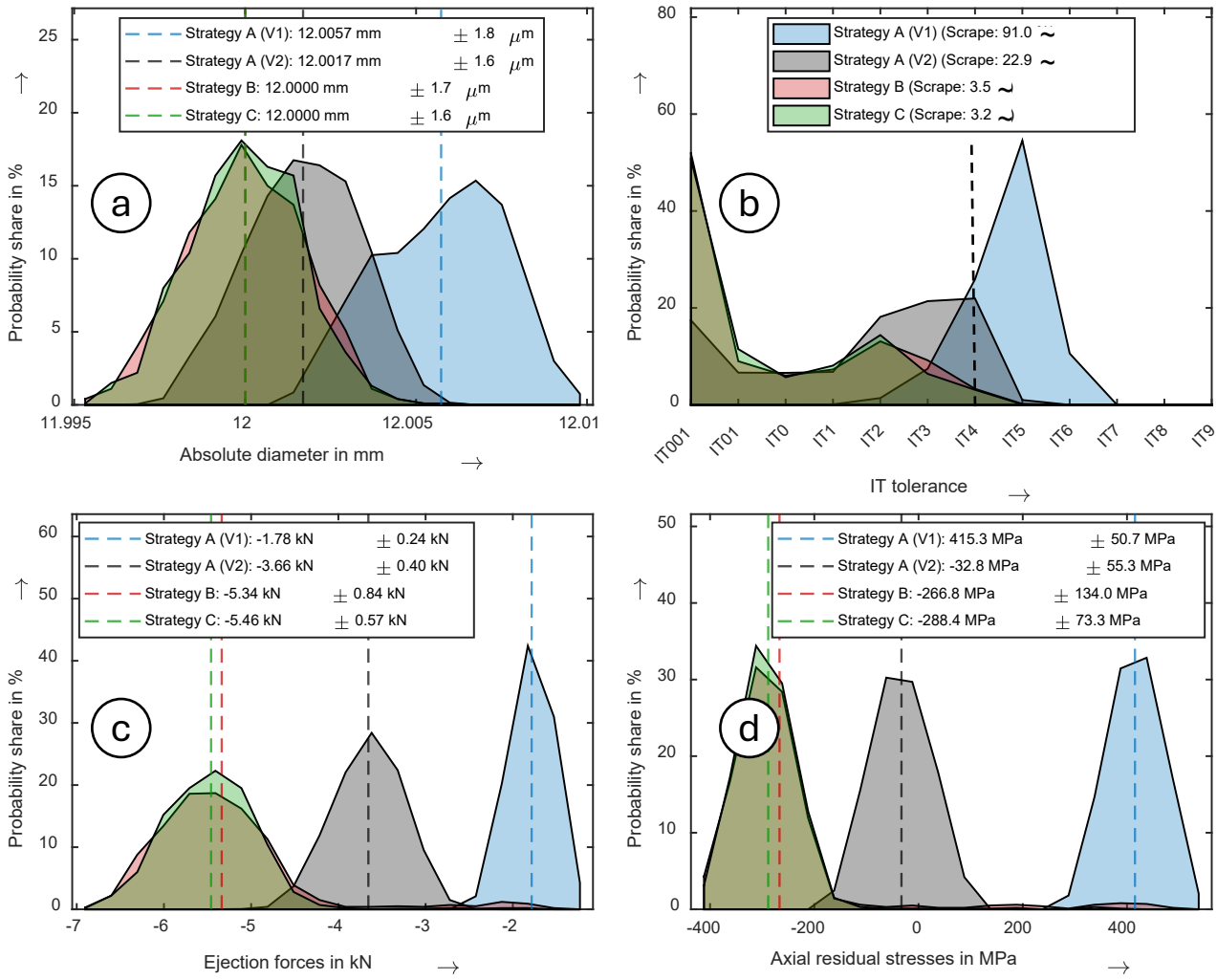

Fig.8. Comparison of the results for control strategies A to C for the green-steel scenario. Probability distributions are shown for (a) the absolute diameter, (b) the resulting IT classes, where IT4 and above are classified as scrap, (c) the ejector forces, and (d) the axial residual stresses. Dashed lines denote the respective mean values.

The state vector stacks the last part's diameter and ejection-force errors, the current set points, and the flow-stress proxy. To better compensate material scatter, the action set is expanded to $\Delta F \in \{-15, -5, 0, +5, +15\}$ kN providing stronger corrective authority. A neural action-value function with two hidden layers (32 units each) is trained for 500 epochs in the prediction environment using an ε-greedy policy, a replay buffer, and a target network. Accuracy stabilizes at roughly epoch 300. When deployed online from the same initial set points, the controller converges toward $F_{ax,1} = 50\ kN$ and $F_{ax,2} = 350\ kN$ within only 25 strokes. As visible in Fig. 7 (a–b), the control signals exhibit higher dynamic activity than in Strategy B because the policy actively counteracts incoming material fluctuations rather than reacting one part later.

The resulting performance is summarized in Fig. 8(a–d). Relative to Strategy A (constant set points), Strategy B efficiently re-centers the means of diameter, IT class, and ejection force, though with only modest variance reduction. Strategy C achieves significantly smaller standard deviations, approximately a factor of two for ejection forces and residual stresses by exploiting the hardness-derived flow-stress estimate to anticipate variability. For geometric tolerances, however, the improvement is marginal compared with Strategy B. This indicates that, despite the additional information from $k_f(0,3)$, Strategy C provides little further benefit for the diameter distribution. The underlying reason is that other, unmeasured inputs such as tool elastic properties or friction contribute substantially to the remaining variance in diameter, limiting the achievable reduction (Fig. 3).

Even though the decrease in standard deviation is small for geometric tolerances, Strategy C still reduces scrap from 3.5 % to 3.2 % under the criterion that only parts below IT4 are accepted. This demonstrates that even slight reductions in variability can meaningfully improve the robustness of a cold extrusion process. It also highlights Strategy C's potential of for high-value or tight-tolerance applications. At the same time, Strategy B remains a sensor-lean and computationally efficient option that already achieves high accuracy and competitive precision when aggressive variance reduction is not mandatory.

Conclusion

This study demonstrates that adaptive die systems effectively compensate for uncertainties arising from fluctuating material and lubrication conditions, such as those expected in green steel production or sustainable tribological systems, while maintaining tight and reproducible product tolerances. Overall, the results reveal a broad potential for next-generation cold extrusion strategies in which process forces, dimensional accuracy, and residual stresses can be actively shaped to meet both sustainability and quality demands. Key insights include:

1. A Gaussian process regression (GPR) model trained with fewer than 400 random finite-element simulations is sufficient to build a reliable surrogate model, which is capable of predicting all relevant product and process states with high physical fidelity.
2. Comparison with experimental spot measurements verifies the model's predictive trends and provides evidence that the surrogate captures the dominant mechanical interactions of the adaptive die system.
3. Increasing the die preload during the main forming phase enables micrometer-scale calibration of final diameters, thereby reducing the sensitivity of geometric tolerances to variations in flow stress.
4. Raising the preload during the ejection phase promotes near-surface plasticity and shifts residual stresses into compression while simultaneously increasing the required ejector forces.
5. A part-to-part control strategy (Strategy B) significantly improves accuracy in achieving target properties by adjusting the preload between strokes, allowing two independent target quantities to be steered through separate modulation of forming- and ejection-phase preload.
6. To reduce the standard deviation under fluctuating material conditions, a reinforcement-learning-based controller (Strategy C) can incorporate additional in-process features, such as hardness-derived flow-stress estimates, thereby jointly minimizing offset and variability in the key product properties.

Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the framework of the priority program SPP 2013 („Residual stresses“) by a grant OE 558/16-2, GR 1818/63-2 and GR 1818/86-1 (531874276).

References

[1] Y. Yuan, Y. Ren, Q. Wu, Research on Improvement of Cold Extrusion Technology of Connecting Screw, J. Phys.: Conf. Ser. 2160 (2022) 12055. https://doi.org/10.1088/1742-6596/2160/1/012055.

[2] T. Altan, G. Ngaile, G. Shen, Cold and hot forging: Fundamentals and applications, 1st ed.

[3] Handbook of metal forming, 1st ed., Society of Manufacturing Engineers, Dearborn, Mich, 2011.

[4] P. Groche, Trends in der Kaltmassivumformung, VDI-Z 162 (2020) 58–59. https://doi.org/10.37544/0042-1766-2020-01-02-58.

[5] D.Y. Yang, M. Bambach, J. Cao, J.R. Duflou, P. Groche, T. Kuboki, A. Sterzing, A.E. Tekkaya, C.W. Lee, Flexibility in metal forming, CIRP Annals 67 (2018) 743–765. https://doi.org/10.1016/j.cirp.2018.05.004.

[6] J.K. Singh, A.K. Rout, Advances in Green Steel Making Technology - A Review, American Journal of Materials Engineering and Technology 6 (2018) 8–13. https://doi.org/10.12691/materials-6-1-2.

[7] Hasan Muslemani, Xi Liang, Katharina Kaesehage, Francisco Ascui, Jeffrey Wilson, Opportunities and challenges for decarbonizing steel production by creating markets for 'green steel' products, Journal of Cleaner Production 315 (2021) 128127. https://doi.org/10.1016/j.jclepro.2021.128127.

[8] P. Sellamuthu, P. Hodgson, N. Stanford, Effect of copper on microstructure, recrystallization and precipitation kinetics in strip cast low carbon steel, Mater. Res. Express 6 (2019) 1265j5. https://doi.org/10.1088/2053-1591/ab7310.

[9] B. Mishra, K. Kumbhar, K.S. Kumar, K.S. Prasad, M. Srinivas, Effect of copper addition on microstructure and mechanical properties of ultra high strength NiSiCrCoMo steel, Materials Science and Engineering: A 651 (2016) 177–183. https://doi.org/10.1016/j.msea.2015.10.129.

[10] A. Huang, K. Wang, Y. Zhao, W. Wang, X. Wei, J. Peng, Effect of Copper Addition on the Formability of 304L Austenitic Stainless Steel, Journal of materials engineering and performance 32 (2023) 3563–3570. https://doi.org/10.1007/s11665-022-07367-2.

[11] T. Hain, CO2-Fottprint Reduzierung in der (Kalt-)Massivumformung: Worauf kommt es an?, VDI-Jahrestreffen der Kaltmassivumformer 2023 2023 1–22.

[12] European Steel Technology Platform (ESTEP), Green Steel by EAF Route: A Sustainable Value Chain in the Steel Industry, Brussels, 2019.

[13] A. Gramlich, T. Hinrichs, H. Springer, U. Krupp, Recycling-Induced Copper Contamination of a 42CrMo4 Quench and Tempering Steel: Alterations in Transformation Behavior and Mechanical Properties, steel research int. 94 (2023). https://doi.org/10.1002/srin.202200623.

[14] R. Nkhoma, K.A. Annan, C. Siyasiya, Effect of adding Cr, Ni and Mo on quench and temper microstructure and mechanical properties of a Si-Mn spring steel, Materials Today: Proceedings 56 (2022) 1629–1634. https://doi.org/10.1016/j.matpr.2021.09.395.

[15] S. Arunkumar, M. Chandrasekaran, V. Muthuraman, T. Vinod Kumar, Study properties and mechanical behavior of the shaft material 16MnCr5, Materials Today: Proceedings 37 (2021) 2458–2461. https://doi.org/10.1016/j.matpr.2020.08.286.

[16] Handbuch der europäischen Eisen- und Stahlwerke, 16th ed., Montan- und Wirtschaftsverlag, Düsseldorf, 2012.

[17] S. Volz, J. Launhardt, N. Bay (1), C. Huc, P. Moreaud, L. Dubard, C. Nielsenb, K. Hayakawae, K. Kitamuraf, P. Groche (2) a*, International round robin test of environmentally benign lubricants for cold forging.

[18] P. Groche, S. Zang, C. Müller, D. Bodenmüller, A study on the performance of environmentally benign lubricants at elevated temperatures in bulk metal forming, Journal of Manufacturing Processes 20 (2015) 425–430. https://doi.org/10.1016/j.jmapro.2014.06.004.

[19] P. Groche, Flexible Forming – A Key Factor for Sustainable Manufacturing, Proceedings of the 9th JSTP International Seminar on Precision Forging (ISPF 2024) March 11–14, Kyoto, Japan 2024 10–18.

[20] J.M. Allwood, S.R. Duncan, J. Cao, P. Groche, G. Hirt, B. Kinsey, T. Kuboki, M. Liewald, A. Sterzing, A.E. Tekkaya, Closed-loop control of product properties in metal forming, CIRP Annals 65 (2016) 573–596. https://doi.org/10.1016/j.cirp.2016.06.002.

[21] K. Siimut, G. Bissacco, K.M. Madsen, E. Ceron, C.V. Nielsen, Tooling system for an ironing punch with adjustable diameter, Journal of Manufacturing Processes 155 (2025) 681–700. https://doi.org/10.1016/j.jmapro.2025.10.035.

[22] P. Tchasse, M. Liewald, T. Deliktas, Development of an intelligent metal forming robot and application to multi-stage cold forging, Prod. Eng. Res. Devel. 19 (2025) 731–749. https://doi.org/10.1007/s11740-025-01330-5.

[23] C. Soyarslan, Prevention of Internal Cracks in Forward Extrusion by Means of Counter Pressure: A Numerical Treatise, Steel Research International 2009 (2009). https://doi.org/10.2374/SRI08SP170.

[24] H. Hoche, A. Balser, M. Oechsner, A. Franceschi, P. Groche, Enhancement of the residual stresses of cold full-forward extruded parts by application of an active counter punch, Materialwiss. Werkstofftech. 50 (2019) 669–681. https://doi.org/10.1002/mawe.201900050.

[25] W. Volk, Targeted Use of Forming-induced Residual Stresses in Metal Components, Technical University of Munich, 2024.

[26] J. Baumgarten, Ansätze zur Kompensation der elastischen Matrizenaufweitung bei Kaltmassivumformprozessen. Dissertation, Hannover, 2003.

[27] F. Jütte, Beitrag zur Präzisionsumformen von Stirnradverzahnungen. Dissertation, Paderborn, 1986.

[28] A. Franceschi, J. Stahl, C. Kock, R. Selbmann, S. Ortmann-Ishkina, A. Jobst, M. Merklein, B. Kuhfuß, M. Bergmann, B.-A. Behrens, W. Volk, P. Groche, Strategies for residual stress adjustment in bulk metal forming, Arch Appl Mech 91 (2021) 3557–3577. https://doi.org/10.1007/s00419-021-01903-7.

[29] P. Pelz, Uncertainty in Mechanical Engineering II, Trans Tech Publications Limited, Pfaffikon, 2015.