TOWARDS INTELLIGENT MANUFACTURING: SPATIO-TEMPORAL LEARNING OF PROCESS–MATERIAL DYNAMICS WITH ATTENTION-DRIVEN NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose an uncertainty-aware Attention-based Spatio-Temporal Neural Operator (ASNO) framework for additive manufacturing, extending the previously developed ASNO model with uncertainty quantification and generalization across material property space. While the original ASNO learns complex spatiotemporal dynamics in scientific and engineering systems, it does not provide predictive uncertainty, which is crucial for reliability, safety, and interpretability in data-sparse and high-stakes manufacturing. To address this gap, we integrate a Laplace-based approximation to quantify epistemic uncertainty within operator learning. The architecture preserves an implicit–explicit formulation motivated by the Backward Differentiation Formula, decoupling temporal and spatial modeling through a Transformer encoder for long-horizon evolution and a Nonlocal Attention Operator for spatial interactions. This design enables stable, accurate, and computationally efficient forecasting of spatiotemporal fields. Across chaotic and PDE benchmarks, including the Lorenz system, Darcy flow, and Navier–Stokes equations, ASNO achieves strong accuracy and efficiency while delivering calibrated uncertainty, with 94% prediction interval coverage probability and narrow intervals. We further demonstrate performance on Directed Energy Deposition by predicting full-field temperature distributions, generalizing across alloy material, and enabling inference of melt pool characteristics. Beyond forward prediction, the learned operator and its epistemic uncertainty provide an interface for inverse identification and accelerated material discovery, for example estimating temperature-dependent thermal properties from sparse observations. Overall, ASNO advances uncertainty-aware surrogate modeling for intelligent manufacturing and digital twin deployment.

1 Introduction

Metal additive manufacturing (AM) enables the fabrication of complex, high-performance components with tailored geometry and material properties Beaman et al. (2020); Frazier (2014); Zhang et al. (2019). Beyond shape complexity, AM offers a unique opportunity for accelerated materials discovery, where process conditions and thermal histories can be leveraged to design new alloys, graded compositions, and microstructures in situ. Despite this promise, AM processes are governed by tightly coupled thermal, fluid, and mechanical phenomena, making accurate modeling and real-time process control highly challenging Gunasegaram et al. (2024); Chen et al. (2025). High-fidelity numerical simulations, such as finite element models, can capture these multiphysics interactions, but they are computationally expensive and unsuitable for online design, closed-loop control, or digital twin deployment Karkaria et al. (2024); Liao et al. (2023b); Cao & Ayalew (2019).

Machine learning (ML) has emerged as a practical surrogate for high-fidelity simulations by providing faster predictions that support process optimization Jiang et al. (2022); Qin et al. (2022). However, most existing ML models are task-specific and require extensive retraining when materials, geometries, or process parameters change. In practice, the material itself may also be uncertain or evolving, for example across new alloys, modified powder batches, or temperature-dependent thermophysical properties, motivating not only forward prediction but also inverse material discovery, where digital twins infer latent material descriptors directly from observed thermal fields. Many existing approaches rely on low-dimensional quantities of interest, such as melt pool temperature or depth, rather than predicting full-field physics. This reduction discards critical spatial and temporal information needed

to understand defect formation, microstructural evolution, and residual stress Ko et al. (2022). These limitations motivate more general spatio-temporal models that can learn transferable representations across diverse AM process–material regimes Bommasani et al. (2021).

Scientific machine learning (SciML) and neural operators offer a promising path forward by learning mappings between function spaces rather than fixed grids Karniadakis et al. (2021); Kovachki et al. (2023). Recent operator architectures such as Fourier Neural Operators (FNO) You et al. (2022), Graph Neural Operator Transformers (GNOT) Hao et al. (2023b), and Transolver Wu et al. (2024) have improved the modeling of PDE-governed systems. Nevertheless, many of these methods entangle temporal and spatial dynamics, limiting interpretability and stability in long-horizon prediction. They also provide limited mechanisms for systematic uncertainty quantification. Meanwhile, Transformers capture temporal dependencies effectively but are not inherently designed to model complex spatial interactions in nonlinear physical systems such as turbulent transport or phase transitions.

To address these challenges, we propose the Attention-based Spatio-Temporal Neural Operator (ASNO). ASNO combines a Transformer Encoder for temporal extrapolation with a Nonlocal Attention Operator (NAO) for spatial refinement, inspired by an implicit–explicit (IMEX) interpretation of the Backward Differentiation Formula. This separation improves numerical stability, interpretability, and scalability for long-term predictions. In addition, ASNO incorporates a Linear Laplace Approximation to estimate epistemic uncertainty, providing calibrated confidence bounds that are essential for safe decision-making and uncertainty-guided exploration of new process–material conditions.

We validate ASNO on standard scientific benchmarks, including the Lorenz system, Darcy flow, and Navier–Stokes equations, and demonstrate its practical utility in Directed Energy Deposition (DED) by predicting full-field temperature maps and extracting melt pool characteristics.

- We develop a spatio-temporal neural operator that enables full-field surrogate modeling of evolving physical processes, moving beyond low-dimensional proxies and supporting transferable prediction across diverse manufacturing and scientific regimes.
- We integrate Laplace-based uncertainty quantification to produce calibrated confidence intervals by quantifying epistemic uncertainty.
- We show strong generalization across process–material regimes in DED, supporting genelizability accross material-process characterization for AM.

The remainder of the paper is organized as follows. Section 2 reviews recent advances in transformers and neural operators for scientific modeling. Section 3 presents the ASNO architecture, including temporal extrapolation, spatial refinement, and uncertainty quantification. Section 4 reports benchmark performance on Lorenz, Darcy, and Navier–Stokes systems. Section 4 evaluates uncertainty calibration. Section 5 applies ASNO to a DED case study, and Section 6 concludes with future research directions.

2 Technical Backgrounds

2.1 Transformers for Time-Series in Manufacturing

Transformers have quickly become a central tool for time-series modeling in manufacturing because of their ability to capture long-range dependencies, fuse multimodal sensor streams, and adapt under shifting operating conditions. On shop floors, processes often involve irregular sampling, noisy signals, and multiple sensing modalities, making attention-based models particularly effective. For instance, transformer-augmented physics models have improved milling wear prediction by embedding mechanistic priors into the attention mechanism, leading to robust generalization across cutting conditions Hao et al. (2023a). Similarly, hierarchical temporal transformers enhance cutter wear recognition by modeling multi-scale temporal patterns Xue et al. (2023), while denoising transformer auto-encoders have shown strong performance in fusing acoustic, vibration, and current signals for condition monitoring in turning and milling cells Wang et al. (2024). At the plant level, Informer and Temporal Fusion Transformer-style (TFT) architectures extend the utility of transformers from single-machine monitoring to factory-wide predictive maintenance and production forecasting. For example, predictive maintenance systems have leveraged transformers with entity disambiguation to handle Industrial IoT streams at scale Qi et al. (2024), and Temporal Fusion Transformers have delivered accurate, interpretable forecasts of electricity demand in automotive production lines by capturing calendar effects and production schedules Lenk et al. (2025). Further innovations, such as CNN–Informer hybrids, reduce false alarms in tool-wear monitoring by leveraging sparse attention

over long horizons Xie et al. (2023), and physics-informed evolutionary transformers track tool conditions with greater stability under distribution shifts Liu et al. (2023).

Despite their promise, transformers face critical limitations that hinder their deployment in real-world manufacturing. Training becomes computationally expensive as both sequence length and sensor density increase, and although sparse or compressed attention alleviates some of the burden, scalability remains an issue for very long sequences. In addition, domain shifts caused by changing tools, materials, and machine states often degrade performance, requiring hybrid approaches that incorporate physics-based constraints such as cutting-force models or energy balance laws Hao et al. (2023a); Liu et al. (2023). Interpretability is also limited, since attention weights typically highlight features like thermal cycles or transient cut segments without grounding them in physically meaningful coordinates such as scan paths, machine states, or batch identifiers. Finally, most transformer pipelines lack structured uncertainty quantification, which is essential for robust and trustworthy decision-making on the shop floor Lenk et al. (2025); Qi et al. (2024). These challenges point to the need for a new architecture that is computationally efficient, physically informed, interpretable, and uncertainty-aware, motivating the development of ASNO.

2.2 Neural Operators for Manufacturing

While transformers address time-series challenges, neural operators (NOs) target a different bottleneck: the fast and resolution-independent prediction of high-dimensional fields governed by process physics. In additive manufacturing (AM), Fourier Neural Operators (FNOs) and related architectures have been used to predict transient, full-field temperature maps during laser deposition, achieving significant speedups compared to high-fidelity solvers while maintaining generalization across geometries Chen et al. (2024). Here, the comparison refers specifically to finite element and finite volume based high-fidelity thermal solvers, which typically require hours of computation, whereas NOs can generate field-level predictions in seconds. Building on these foundations, neural operator enabled digital twins for laser powder bed fusion now combine FNO surrogates with optimization, calibration, and uncertainty modules, supporting closed-loop process monitoring and control Liu et al. (2024a;b). Beyond single-physics systems, extended DeepONet formulations can learn coupled multiphysics fields under varied loading conditions Kushwaha et al. (2024), while sequential DeepONet variants model history-dependent processes through operator compositions He et al. (2024). Applications have also expanded beyond metals. For example, physics-guided neural operators have been applied in composites manufacturing, where they predict cure cycle dependent temperature fields under physical constraints using limited training data Chen et al. (2023).

Compared to physics-informed neural networks (PINNs), which enforce pointwise residuals and have shown promise for single-track thermal analysis in LPBF Hosseini et al. (2023), they often suffer from slow convergence and scalability issues when extended to large-scale, high-dimensional fields. In contrast, neural operators map entire functions to functions, making them naturally mesh and resolution independent and therefore more transferable across geometries, scan strategies, and process settings. While best practices are emerging, such as incorporating physically meaningful conditioning variables like absorptivity or hatch spacing, embedding uncertainty quantification for robustness under out-of-recipe operations, and adopting multi-fidelity training that blends simulations with in-situ data Liu et al. (2024a;b); Chen et al. (2023); Kushwaha et al. (2024), several gaps remain. Existing neural operators often struggle to disentangle temporal and spatial dependencies, they lack principled uncertainty quantification, and they can be difficult to adapt reliably under evolving manufacturing conditions. These challenges highlight the need for a spatio-temporal operator framework that unifies temporal modeling, spatial field prediction, and uncertainty awareness, motivating the development of ASNO.

3 Model Architecture

This section presents the Attention based Spatio Temporal Neural Operator (ASNO) for scientific modeling in manufacturing. ASNO is designed around a numerically motivated principle, the Backward Differentiation Formula (BDF), which naturally decomposes stiff dynamics into an explicit extrapolation term and an implicit correction term. We map this IMEX viewpoint into a neural architecture with two complementary components. A Transformer Encoder performs the explicit temporal extrapolation from a short history window, and a Nonlocal Attention Operator performs the implicit spatial refinement conditioned on process and material features. Finally, we quantify epistemic uncertainty using a Laplace based posterior approximation around the trained parameters.

3.1 BDF MOTIVATED IMEX DECOMPOSITION AND ASNO OVERVIEW

Additive manufacturing exhibits stiff and multi timescale behavior due to steep thermal gradients, abrupt phase transitions, and rapid changes in material response. Purely explicit numerical updates can become unstable unless extremely small time steps are used, which is computationally prohibitive. Implicit methods such as BDF stabilize the fast dynamics by evaluating the forcing at the next time step, while still leveraging past states. This motivates our ASNO design.

Consider the initial value problem

$$\dot{X}(t) = F(t, X(t)), \quad X(t_0) = X_0, \tag{1}$$

where $X(t)$ denotes the evolving state, for example a temperature or multi field state over a spatial mesh, and F denotes the governing dynamics and conditioning inputs. A general n step BDF discretization at time index $m+1$ is

$$\sum_{k=1}^{n} \alpha_k \, X_{m-k+1} + X_{m+1} = \Delta t \, \beta \, F\big((m+1)\Delta t, X_{m+1}\big), \tag{2}$$

where Δt is the time step, and the coefficients $\{\alpha_k\}$ and β depend on the order of the scheme. Rearranging yields

$$X_{m+1} = \underbrace{-\sum_{k=1}^{n} \alpha_k \, X_{m-k+1}}_{\tilde{X}_{m+1}} + \Delta t \, \beta \, F\big((m+1)\Delta t, X_{m+1}\big). \tag{3}$$

The first term $\tilde{X}_{m+1}$ is an extrapolated state computed from history, and the second term is an implicit correction that depends on the unknown X_{m+1}.

ASNO follows this decomposition directly. Given a history window $\{X_m, X_{m-1}, \ldots, X_{m-n+1}\}$ and conditioning features at the next time step F_{m+1}, including process inputs and material descriptors (for example low dimensional parameterizations of $k(T)$ and $C_p(T)$), ASNO computes

$$X^{\text{out}}_{m+1} = \text{NAO}\Big(\text{TE}(X_m, X_{m-1}, \ldots, X_{m-n+1}), F_{m+1}\Big), \tag{4}$$

where TE realizes the explicit temporal extrapolation and NAO realizes the implicit spatial refinement. Figure 1 illustrates this separation.

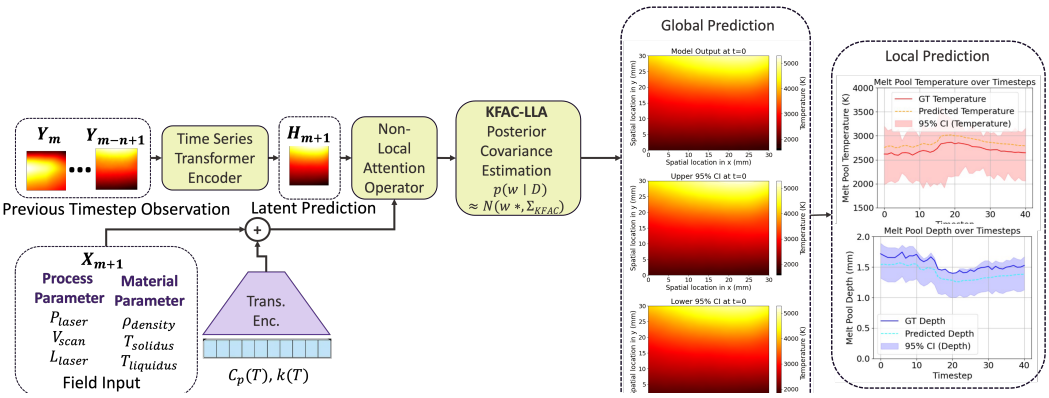

Figure 1: ASNO architecture motivated by an IMEX interpretation of BDF. The Transformer Encoder (TE) extrapolates temporal dynamics from a short history, producing $\tilde{X}_{m+1}$, and the Nonlocal Attention Operator (NAO) refines the prediction using spatial interactions and conditioning features F_{m+1}.

3.2 Explicit step, temporal extrapolation with a Transformer Encoder

The explicit term in equation 3 predicts forward using history only. In ASNO, we implement this term using a time series Transformer Encoder that maps a sequence of past states

$$\{X_m, X_{m-1}, \ldots, X_{m-n+1}\}, \quad X_k \in \mathbb{R}^{N\times d}, \tag{5}$$

to a latent extrapolated representation $\tilde{X}_{m+1}$, where N is the number of spatial points and d is the number of channels.

Each state is embedded into a model dimension d_{model} and augmented with positional encodings to preserve temporal order,

$$E_k = X_k W_E + P_k, \tag{6}$$

where $W_E \in \mathbb{R}^{d\times d_{\text{model}}}$ is learnable and P_k encodes the time index. The Transformer applies multi head self attention to learn long range temporal dependencies in the history window. For a token representation E_t, the queries, keys, and values are

$$Q = E_t W_{Tq}, \quad K = E_t W_{Tk}, \quad V = E_t W_{Tv}, \tag{7}$$

and the attention weights over the history are

$$\alpha_{m,i} = \frac{\exp\!\Big(\frac{Q_m K_{m-i}^{\top}}{\sqrt{d_k}}\Big)}{\sum_{j=0}^{n-1}\exp\!\Big(\frac{Q_m K_{m-j}^{\top}}{\sqrt{d_k}}\Big)}, \quad i = 0, \ldots, n-1, \tag{8}$$

where d_k is the key dimension. The Transformer output is a latent extrapolated state

$$\tilde{X}_{m+1} = \mathrm{TE}(X_m, X_{m-1}, \ldots, X_{m-n+1}), \tag{9}$$

which advances the system using past information only, before any spatial or material dependent correction.

3.3 Implicit step, spatial refinement with a Nonlocal Attention Operator

The implicit correction term in equation 3 captures spatial interactions and conditioning effects that are crucial in additive manufacturing, such as heat conduction, localized energy input, and variations induced by material response. ASNO models this correction using a Nonlocal Attention Operator (NAO). NAO takes the extrapolated state $\tilde{X}_{m+1}$ and the conditioning features F_{m+1}, for example laser power, scan speed, beam location, or material property descriptors, and produces the refined next state X_{m+1}^{out} via a learned nonlocal operator.

In the main paper, we describe NAO at the operator level. NAO parameterizes a nonlocal kernel K that couples the conditioning field and the extrapolated state across spatial locations, so that the refined field is computed as

$$X_{m+1}^{\text{out}}(y) = \int K\big[\tilde{X}_{m+1}, F_{m+1}\big](y,z)\, F_{m+1}(z)\,\mathrm{d}z, \tag{10}$$

where y and z denote spatial coordinates. The kernel is implemented with attention layers that learn long range interactions, and residual connections preserve stability across depth. This design enables mesh independent modeling and captures global spatial dependencies that arise in manufacturing physics.

3.4 Epistemic uncertainty via Linear Laplace Approximation

Reliable manufacturing digital twins must provide quantitative confidence measures, especially in regimes with limited labeled data, distribution shift, or ambiguity arising from coupled process and material variability. Without epistemic uncertainty, neural operator surrogates may become overconfident when extrapolating to unseen operating conditions. To address this need, we estimate epistemic uncertainty in ASNO through a Laplace approximation around the trained parameters. Our objective is to approximate the posterior over operator weights after training, enabling uncertainty-aware full-field predictions. Let $\mathcal{G}_\theta$ denote the ASNO operator mapping with parameters θ, obtained

by minimizing the regularized empirical risk

$$\theta_{\mathrm{MAP}} = \arg\min_{\theta} \mathcal{R}(\mathcal{D}, \theta) = \arg\min_{\theta} \Big(l(\mathcal{D}, \theta) + r(\theta) \Big), \tag{11}$$

where l denotes the data fidelity loss and r serves as a negative log prior over parameters. The resulting posterior distribution is

$$p(\theta \mid \mathcal{D}) \propto \exp\big(-\mathcal{R}(\mathcal{D}, \theta)\big). \tag{12}$$

We proceed by locally approximating $\mathcal{R}$ through a second-order Taylor expansion around θ_{MAP}, yielding a Gaussian posterior

$$\theta \sim \mathcal{N}(\theta_{\mathrm{MAP}}, \Sigma), \quad \Sigma = H^{-1}, \tag{13}$$

where H is the Hessian of $\mathcal{R}$ evaluated at θ_{MAP} and captures the local sensitivity of the trained operator to parameter perturbations. To ensure scalability in high-dimensional neural operator settings, we employ a diagonal approximation to H^{-1}. Propagating parameter uncertainty into the output space yields the predictive covariance

$$\mathrm{Cov}\big[X_{m+1}^{\mathrm{out}}\big] \approx J_{\theta}^{(m+1)} \Sigma (J_{\theta}^{(m+1)})^{\top}, \tag{14}$$

where $J_{\theta}^{(m+1)} = \partial X_{m+1}^{\mathrm{out}} / \partial \theta$ denotes the Jacobian of the ASNO prediction with respect to parameters. This enables epistemic uncertainty maps over the full spatio-temporal field, as well as uncertainty intervals for derived manufacturing-relevant quantities of interest.

Importantly, uncertainty estimation is integrated directly within ASNO's operator structure rather than treated as a post-processing step. ASNO implements a BDF-motivated IMEX decomposition within a neural operator framework: the Transformer Encoder advances system evolution through temporal extrapolation from historical trajectories, while the Nonlocal Attention Operator refines predictions via nonlocal spatial coupling conditioned on process and material descriptors. Thus, the Laplace approximation equips ASNO with practical epistemic uncertainty estimates, supporting reliable inference and decision-making in manufacturing settings.

4 Benchmark Performance and Uncertainty Evaluation

We evaluate ASNO on three representative benchmark systems that jointly stress the core requirements of scientific surrogate modeling, namely long horizon temporal stability, multiscale spatial coupling, and robustness under distribution shift. Although these benchmarks are canonical in SciML, they also reflect physical behaviors central to materials discovery workflows, where transport dynamics, heterogeneous coefficients, and nonlinear feedback govern process–structure–property evolution. The Lorenz system is a low dimensional chaotic ODE that tests whether a model can remain stable under sensitive dependence on initial conditions. Darcy flow is an elliptic PDE that evaluates spatial operator learning under heterogeneous coefficients and varying forcing, analogous to discovering material behavior across uncertain permeability or conductivity fields. Navier–Stokes is a high dimensional nonlinear PDE for incompressible fluid flow that tests whether a model can preserve coherent structures while avoiding numerical dissipation over time, representative of complex melt pool convection and instability in AM. Together, these benchmarks capture behaviors commonly seen in scientific and engineering digital twins, including stiffness, long range interactions, nonlinear feedback, and variability induced by changing material properties.

Across all benchmarks, each model receives a fixed history window of five states and a forcing term, and predicts the next state through a learned operator. Specifically, for time index m we provide

$$\{X_{m-4}, X_{m-3}, X_{m-2}, X_{m-1}, X_m\}, \quad F_m, \tag{15}$$

and predict

$$X_{m+1} = \mathcal{M}\big(X_{m-4}, X_{m-3}, X_{m-2}, X_{m-1}, X_m, F_m\big). \tag{16}$$

To promote robust evaluation, we augment training trajectories using random spatial permutations that preserve global structure and boundary conditions. Full benchmark definitions, solvers, discretization details, and augmentation rules are provided in Appendix B.

We compare ASNO against widely used operator learning baselines, including DeepONet, Fourier Neural Operator (FNO), a Transformer Encoder baseline, Transolver, Graph Neural Operator Transformer (GNOT), U-Net, and a Linear Encoder plus NAO ablation. Table 1 summarizes unified results

in terms of accuracy and efficiency, highlighting the ability of ASNO to serve as a transferable surrogate backbone across diverse process–material regimes.

Table 1: Unified benchmark performance of ASNO and baselines across Lorenz, Darcy, and Navier–Stokes.

Model	Lorenz			Darcy			Navier–Stokes		
	Params	Time	Loss	Params	GPU	Loss	Params	GPU	Loss
ASNO	258K	1.55s	**0.00079**	760K	181MB	**0.0368**	4.66M	880MB	**0.0213**
Transolver	396K	1.46s	0.00083	811K	422MB	0.0428	4.14M	911MB	0.0234
DeepONet	266K	1.74s	0.00175	6.23M	2146MB	0.0537	5.10M	3100MB	0.0921
Transformer	258K	1.18s	0.00182	1.62M	173MB	0.0559	5.19M	961MB	0.0967
FNO	–	–	–	900K	214MB	0.0768	4.10M	846MB	0.1186
U-Net	–	–	–	821K	123MB	0.1150	5.02M	991MB	0.1940
GNOT	401K	1.99s	0.00219	760K	208MB	0.0516	5.25M	1024MB	0.0322
Linear+NAO	306K	1.29s	0.00529	720K	165MB	0.0547	4.05M	791MB	0.0328

4.1 Key findings across benchmarks

Across benchmarks, ASNO demonstrates strong accuracy, stability, and efficiency. On the Lorenz system, ASNO achieves the lowest loss and remains stable in long-horizon autoregressive rollouts, consistent with its IMEX-inspired design in which temporal extrapolation preserves trajectory memory while spatial refinement captures nonlinear feedback, with long-horizon comparisons provided in Appendix Figure 3. For Darcy flow, ASNO attains the lowest loss and shows greater robustness than baselines that tend to oversmooth or fail to capture long-range spatial correlations induced by heterogeneous permeability fields. This robustness is essential when generalizing across uncertain material coefficients in discovery settings, where transport properties vary across candidate alloys or processing conditions. On Navier–Stokes, ASNO again achieves the best loss while preserving fine-scale vorticity and coherent flow structures with reduced numerical dissipation, relevant for capturing instability and convection effects in melt pool dynamics. Finally, as shown in Table 1, ASNO attains favorable memory and parameter efficiency relative to strong baselines, making it well-suited for deployment in uncertainty-aware digital twins that support materials exploration under practical runtime constraints. To quantify long horizon stability, we report cumulative error over rollout time,

$$E_T = \sum_{t=1}^{T} \left\| X_t^{\text{true}} - X_t^{\text{pred}} \right\|_{L^2}. \tag{17}$$

Cumulative error curves for Darcy and Navier–Stokes are provided in Appendix Figure 2, where ASNO exhibits minimal error growth over time, confirming stable long term behavior.

4.2 Integrated uncertainty evaluation

In manufacturing and materials modeling, accuracy alone is insufficient. A reliable digital twin must also report calibrated confidence to guide adaptive experimentation, safe decision making, and uncertainty-aware materials screening. We therefore evaluate ASNO uncertainty estimates on the Darcy benchmark using two standard metrics, Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW). We compute these metrics on a held out Darcy test set with 100 samples and 95 time steps over the full 21×21 grid, averaging across space and time. PICP measures calibration by checking whether ground truth values fall within the predicted 95% intervals. MPIW measures sharpness, lower is better when calibration is preserved. Formal definitions of both metrics are provided in Appendix C.

Table 2: Uncertainty quantification metrics on Darcy.

Metric	Value
PICP (Coverage%)	94.00 %
MPIW (Mean interval width)	0.3046

Table 2 shows that ASNO achieves coverage close to the nominal 95% while maintaining modest interval widths. This indicates that the uncertainty estimates are both calibrated and informative

for downstream tasks such as adaptive sampling, Bayesian optimization, and risk-aware exploration of new material and process regimes. The uncertainty intervals are derived from a diagonal Linear Laplace approximation applied to the parameter posterior. Full mathematical derivation, Hessian approximation strategy, and implementation details are provided in Appendix C.

5 Melt Pool Temperature Field Prediction in Additive Manufacturing

Table 3 compares ASNO against strong neural operator and sequence baselines on the DED full-field prediction task. ASNO achieves the lowest test loss (0.0140), substantially outperforming widely used operator architectures such as FNO (0.0357) and DeepONet (0.0456), as well as convolutional surrogates such as U-Net (0.0652). This gain is especially significant in DED, where thermal evolution is governed by stiff moving-heat-source dynamics, sharp melt pool gradients, and strong coupling between localized energy input, conduction, and temperature-dependent material properties. ASNO's IMEX-inspired separation directly aligns with this physics: the Transformer Encoder captures long-range temporal accumulation of heat, while the Nonlocal Attention Operator performs spatial correction that resolves nonlocal transport and melt pool localization, yielding more stable and accurate full-field predictions.

Beyond accuracy, ASNO remains computationally efficient, requiring only 6.17s per epoch with a moderate parameter count (2.83M), comparable to other competitive baselines. While the Transformer Encoder alone trains faster (3.95s per epoch), it performs substantially worse (0.0563), indicating that temporal extrapolation without explicit spatial refinement cannot capture the coupled thermal transport mechanisms that govern melt pool shape and microstructure evolution. Overall, ASNO provides a high-fidelity uncertainty-aware surrogate that balances predictive performance and efficiency, supporting reliable process monitoring and data-efficient process–material optimization in additive manufacturing.

5.1 DED DATASET AND LEARNING OBJECTIVE

Training data are generated using GAMMA, an in-house GPU-accelerated finite element thermal simulator Liao et al. (2023a). GAMMA solves transient heat conduction with a moving laser flux, producing high-fidelity temperature fields over time. At each timestep, we record the full-field temperature map together with process inputs (laser power, scan speed, location) and material descriptors, including temperature-dependent thermal conductivity $k(T)$, specific heat capacity $C_p(T)$, density ρ, and laser absorption and heat-loss parameters (η, ϵ, h). Conditioning on these thermophysical properties enables ASNO to generalize across manufacturing recipes and alloy settings relevant to accelerated materials discovery. Such calibrated surrogates are essential for self-driving laboratories, where rapid screening of new materials requires both accurate prediction and reliable uncertainty estimates. Detailed simulation setup and dataset construction are provided in Appendix D.

Formally, let $T_{\text{pool}}(m) \in \mathbb{R}^N$ denote the temperature field and let F_{m+1} denote conditioning features at the next step, including process and material parameters. ASNO predicts

$$\hat{T}_{\text{pool}}(m+1) = \text{ASNO}\Big(F_{m+1},\, T_{\text{pool}}(m-4:m)\Big), \tag{18}$$

where a Transformer Encoder extrapolates temporal dynamics and a Nonlocal Attention Operator refines the spatial field (Figure 1). ASNO further outputs epistemic uncertainty via a diagonal Linear Laplace approximation, yielding calibrated 95% confidence intervals for full-field predictions and derived melt pool features. Training and uncertainty details are provided in Appendix E and Appendix F.

5.2 FULL-FIELD TEMPERATURE PREDICTION

Table 3 compares ASNO against strong neural operator and sequence baselines on the DED full-field prediction task. ASNO achieves the lowest test loss (0.0140), substantially outperforming widely used operator architectures such as FNO (0.0357) and DeepONet (0.0456), as well as convolutional surrogates such as U-Net (0.0652). This gain is especially significant in DED, where thermal evolution is governed by stiff moving-heat-source dynamics, sharp melt pool gradients, and strong coupling between localized energy input, conduction, and temperature-dependent material properties. ASNO's IMEX-inspired separation directly aligns with this physics: the Transformer Encoder captures long-range temporal accumulation of heat, while the Nonlocal Attention Operator performs spatial

Table 3: Comparison of models on the Additive Manufacturing dataset.

Model Type	Trainable Params	Time per Epoch (s)	Best Test Loss
ASNO	2,825,444	6.17	**0.0140**
FNO	3,654,312	7.10	0.0357
DeepONet	3,385,605	5.25	0.0456
UNet	4,564,157	6.64	0.0652
GNOT	**2,442,678**	6.60	0.0332
Transolver	2,636,9158	8.71	0.0502
Linear + NAO	2,943,435	4.75	0.0397
Transformer Encoder (TE)	2,784,920	**3.95**	0.0563

correction that resolves nonlocal transport and melt pool localization, yielding more stable and accurate full-field predictions.

Beyond accuracy, ASNO remains computationally efficient, requiring only 6.17s per epoch with a moderate parameter count (2.83M), comparable to other competitive baselines. While the Transformer Encoder alone trains faster (3.95s per epoch), it performs substantially worse (0.0563), indicating that temporal extrapolation without explicit spatial refinement cannot capture the coupled thermal transport mechanisms that govern melt pool shape and microstructure evolution. Overall, ASNO provides a high-fidelity uncertainty-aware surrogate that balances predictive performance and efficiency, supporting reliable process monitoring and data-efficient process–material optimization in additive manufacturing.

5.3 Local feature prediction: melt pool temperature and depth

Beyond full-field prediction, we extract melt pool temperature and depth, which are key thermal descriptors linked to microstructure evolution and part quality. As shown in Figure 1, these features are tracked over timesteps with calibrated confidence intervals. We define the molten region using a liquidus threshold T_{thr} and compute melt pool statistics from the predicted fields, with complete definitions, discretization details, and extraction procedures provided in Appendix G. We evaluate constant-power, laser shutoff, and high-power out-of-distribution scenarios, and the corresponding quantitative feature trajectories with 95% uncertainty bounds are reported in Appendix Figure 6. Across all cases, ASNO consistently tracks melt pool evolution while uncertainty expands appropriately under unseen regimes. This is critical for guiding adaptive experimentation, since melt pool uncertainties indicate where additional data collection is needed when exploring new materials. Taken together, these results demonstrate that ASNO provides an uncertainty-aware surrogate for DED that supports real-time monitoring and data-efficient exploration of new process–material settings, aligning with the goals of automated materials discovery and intelligent manufacturing.

6 Conclusion

In this paper we introduce the Attention-based Spatio-Temporal Neural Operator (ASNO), a novel hybrid framework that leverages the implicit–explicit structure of the Backward Differentiation Formula to separately model temporal and spatial dependencies. ASNO combines a Transformer encoder for time evolution with a Nonlocal Attention Operator for spatial refinement, enabling accurate and scalable modeling across chaotic systems and physical processes governed by PDEs. Benchmark evaluations on the Lorenz system, Darcy flow, and Navier–Stokes equations demonstrate that ASNO surpasses baseline neural operators in both predictive accuracy and computational efficiency. Furthermore, ASNO incorporates uncertainty quantification through a diagonal Laplace approximation, achieving a PICP of 94% with a modest prediction interval width (MPIW = 0.30), along with a negative log-likelihood of -1.58 and a CRPS of 0.022, which together highlight its ability to produce sharp and well-calibrated confidence estimates.

In a practical manufacturing setting, ASNO successfully predicts full-field temperature maps in DED, supporting the extraction of local thermal features such as melt pool depth and cooling rates, while also generalizing across material property regimes in the additive manufacturing material space. These capabilities make ASNO suitable for both large-scale process optimization and fine-grained quality monitoring, aligning with the workshop focus on uncertainty-aware process–material modeling and accelerated materials discovery. Potential future extensions include the modeling of multi-physics interactions, such as phase transitions and fluid–solid coupling, which could further expand ASNO's scope of application in concurrent process–structure–property design.

References

JJ Beaman, David L Bourell, CC Seepersad, and DJJOMS Kovar. Additive manufacturing review: early past to current practice. *Journal of Manufacturing Science and Engineering*, 142(11):110812, 2020.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Xiaoqing Cao and Beshah Ayalew. Robust multivariable predictive control for laser-aided powder deposition processes. *Journal of the Franklin Institute*, 356(5):2505–2529, 2019.

Gengxiang Chen, Yingguang Li, Xu Liu, Charyar Mehdi-Souzani, Qinglu Meng, Jing Zhou, and Xiaozhong Hao. Physics-guided neural operator for data-driven composites manufacturing process modelling. *Journal of Manufacturing Systems*, 2023.

Jiangce Chen, Wenzhuo Xu, Martha Baldwin, Björn Nijhuis, Ton van den Boogaard, Noelia Grande Gutiérrez, Sneha Prabha Narra, and Christopher McComb. Capturing local temperature evolution during additive manufacturing through fourier neural operators. *Journal of Manufacturing Science and Engineering*, 2024.

Yi-Ping Chen, Vispi Karkaria, Ying-Kuan Tsai, Faith Rolark, Daniel Quispe, Robert X Gao, Jian Cao, and Wei Chen. Real-time decision-making for digital twin in additive manufacturing with model predictive control using time-series deep neural networks. *arXiv preprint arXiv:2501.07601*, 2025.

William E Frazier. Metal additive manufacturing: a review. *Journal of Materials Engineering and performance*, 23:1917–1928, 2014.

Christoph Fredebeul. A-bdf: a generalization of the backward differentiation formulae. *SIAM journal on numerical analysis*, 35(5):1917–1938, 1998.

DR Gunasegaram, AS Barnard, MJ Matthews, BH Jared, AM Andreaco, K Bartsch, and AB Murphy. Machine learning-assisted in-situ adaptive strategies for the control of defects and anomalies in metal additive manufacturing. *Additive Manufacturing*, pp. 104013, 2024.

Caihua Hao, Xinyong Mao, Tao Ma, Songping He, Bin Li, Hongqi Liu, Fangyu Peng, and Lei Zhang. Intelligent milling tool wear prediction model based on transformer informed physics. *Advanced Engineering Informatics*, 2023a.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *International Conference on Machine Learning*, pp. 12556–12569. PMLR, 2023b.

Chao He, Yonghao Zhang, Qiang Li, and Xiang Chen. Sequential deep operator networks (s-deeponet) for history-dependent field predictions. *Engineering Applications of Artificial Intelligence*, 2024.

Ehsan Hosseini, Philipp Scheel, Oliver Müller, Roberto Molinaro, and Siddhartha Mishra. Single-track thermal analysis of laser powder bed fusion process: Parametric solution through physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 2023.

Jingchao Jiang, Yi Xiong, Zhiyuan Zhang, and David W Rosen. Machine learning integrated design for additive manufacturing. *Journal of Intelligent Manufacturing*, 33(4):1073–1086, 2022.

Vispi Karkaria, Anthony Goeckner, Rujing Zha, Jie Chen, Jianjing Zhang, Qi Zhu, Jian Cao, Robert X Gao, and Wei Chen. Towards a digital twin framework in additive manufacturing: Machine learning and bayesian optimization for time series process optimization. *Journal of Manufacturing Systems*, 2024.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021. doi: 10.1038/s42254-021-00314-5. URL https://doi.org/10.1038/s42254-021-00314-5.

Hyunwoong Ko, Jaehyuk Kim, Yan Lu, Dongmin Shin, Zhuo Yang, and Yosep Oh. Spatial-temporal modeling using deep learning for real-time monitoring of additive manufacturing. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 86212, pp. V002T02A019. American Society of Mechanical Engineers, 2022.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Karthik Kushwaha, Sushil Mangalapally, Venu Varanasi, and Ramana P. Grandhi. Advanced deep operator networks to predict multiphysics solution fields under variable loading histories and geometries. *Engineering Applications of Artificial Intelligence*, 2024.

Katharina Lenk, Jens Vogt, and Karl Herrmann. An approach to predicting energy demand within automobile production using the temporal fusion transformer. *Energies*, 2025.

Shuheng Liao, Ashkan Golgoon, Mojtaba Mozaffar, and Jian Cao. Efficient gpu-accelerated thermomechanical solver for residual stress prediction in additive manufacturing. *Computational Mechanics*, 71(5):879–893, 2023a.

Shuheng Liao, Jihoon Jeong, Rujing Zha, Tianju Xue, and Jian Cao. Simulation-guided feedforward-feedback control of melt pool temperature in directed energy deposition. *CIRP Annals*, 72(1): 157–160, 2023b.

Ning Liu, Xuxiao Li, Manoj R. Rajanna, Edward W. Reutzel, Brady Sawyer, Prahalada Rao, Jim Lua, Nam Phan, and Yue Yu. Deep neural operator enabled digital twin modeling for additive manufacturing. *AIMS Applied Control and Signal Processing*, 2024a.

Ning Liu, Xuxiao Li, Manoj R. Rajanna, Edward W. Reutzel, Brady Sawyer, Prahalada Rao, Jim Lua, Nam Phan, and Yue Yu. Deep neural operator enabled digital twin modeling for additive manufacturing. *arXiv preprint arXiv:2405.09572*, 2024b.

Yuekai Liu, Tianyang Wang, Shilin Sun, Qi Li, and Fulei Chu. Physics-informed scaling evolutionary transformer for in-situ tool condition monitoring. *IEEE/ASME Transactions on Mechatronics*, 2023.

Liqiang Qi, Mohammad A. Tayebi, Shyqyri Haxha, Xiaoxia Wei, Alessandro Bogliolo, Seyed Jalaleddin Mousavirad, Habib Ullah, and Sepehr Gordan. Predictive maintenance based on identity resolution and transformer models. *Future Internet*, 2024.

Jian Qin, Fu Hu, Ying Liu, Paul Witherell, Charlie CL Wang, David W Rosen, Timothy W Simpson, Yan Lu, and Qian Tang. Research and application of machine learning for additive manufacturing. *Additive Manufacturing*, 52:102691, 2022.

Hui Wang, Shuhui Wang, Weifang Sun, and Jiawei Xiang. Multi-sensor signal fusion for tool wear condition monitoring using a denoising transformer auto-encoder. *Journal of Manufacturing Processes*, 2024.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. *arXiv preprint arXiv:2402.02366*, 2024.

Y. Xie, Y. Huang, M. Sun, G. Li, and H. Liu. Intelligent tool wear monitoring using cnn and informer. *Lubricants*, 2023.

Zhongling Xue, Ni Chen, Youling Wu, and Liang Li. Hierarchical temporal transformer network for tool wear state recognition. *Advanced Engineering Informatics*, 2023.

Huaiqian You, Quinn Zhang, Colton J Ross, Chung-Hao Lee, and Yue Yu. Learning deep implicit fourier neural operators (IFNOs) with applications to heterogeneous material modeling. *Computer Methods in Applied Mechanics and Engineering*, 398:115296, 2022. doi: https://doi.org/10.1016/j.cma.2022.115296.

Chi Zhang, Fei Chen, Zhifeng Huang, Mingyong Jia, Guiyi Chen, Yongqiang Ye, Yaojun Lin, Wei Liu, Bingqing Chen, Qiang Shen, et al. Additive manufacturing of functionally graded materials: A review. *Materials Science and Engineering: A*, 764:138209, 2019.

A Appendix, Detailed Model Architecture

This appendix provides a complete technical specification of ASNO, with emphasis on its role as a surrogate backbone for process–material discovery in additive manufacturing. We formalize the Backward Differentiation Formula (BDF) viewpoint that motivates separating temporal extrapolation from spatial correction in stiff multiphysics systems. This decomposition is particularly relevant for accelerated materials design, where thermal fields evolve differently across candidate alloys, temperature-dependent properties, and uncertain process recipes. We then present the explicit and implicit interpretations that map directly to the Transformer Encoder (TE) and Nonlocal Attention Operator (NAO). Next, we detail the NAO kernel parameterization, its discretization on a spatial mesh, and the iterative attention refinement that produces corrected full-field predictions conditioned on process and material descriptors. Finally, we derive the Laplace approximation used for epistemic uncertainty quantification, introduce scalable curvature approximations suitable for high-dimensional networks, and show how uncertainty is propagated to obtain predictive covariances that guide safe screening of new process–material regimes.

A.1 BDF DETAILS, COEFFICIENTS AND ORDER

For the n step BDF scheme in equation 2, the coefficients $\{\alpha_k\}_{k=1}^n$ and β are determined by enforcing exactness on polynomials up to degree p through Taylor expansion around $(m+1)\Delta t$. This yields truncation error $\mathcal{O}(\Delta t^{p+1})$ for a method of order p and provides the standard stability properties of BDF methods used for stiff systems (see, for example, Fredebeul (1998)). The BDF update can be written in the rearranged form

$$X_{m+1} = -\sum_{k=1}^{n} \alpha_k X_{m-k+1} + \Delta t\, \beta\, F\big((m+1)\Delta t, X_{m+1}\big), \tag{19}$$

which makes an implicit–explicit (IMEX) decomposition explicit. The first term aggregates information from history, while the second term introduces forcing and state-dependent correction through F. Defining the history-based extrapolated state

$$\tilde{X}_{m+1} = -\sum_{k=1}^{n} \alpha_k X_{m-k+1}, \tag{20}$$

we recover

$$X_{m+1} = \tilde{X}_{m+1} + \Delta t\, \beta\, F\big((m+1)\Delta t, X_{m+1}\big). \tag{21}$$

This form clarifies the modeling objective in ASNO: first produce a stable temporal extrapolation $\tilde{X}_{m+1}$, then refine it using a spatially coupled operator that accounts for nonlinear transport, localized energy input, and variations induced by changing material properties. When $F \equiv 0$, $\tilde{X}_{m+1}$ matches X_{m+1}, highlighting that correction is only required when external forcing or material-dependent effects are present.

A.2 EXPLICIT STEP INTERPRETATION

The explicit step corresponds to computing $\tilde{X}_{m+1}$ purely from history,

$$-\sum_{k=1}^{n} \alpha_k\, X_{m-k+1} = \tilde{X}_{m+1}. \tag{22}$$

In ASNO, the Transformer Encoder represents this extrapolation map. Rather than implementing a fixed linear multistep combination, TE learns a data-driven extrapolator that captures nonlinear temporal dependencies, regime shifts, and nonstationary behavior. This is essential in additive manufacturing, where temporal evolution depends not only on scan strategy and boundary effects, but also on alloy-dependent conductivity, heat capacity, and absorption, which vary across candidate material systems in discovery workflows.

A.3 IMPLICIT STEP INTERPRETATION

The implicit correction solves for X_{m+1} through

$$\tilde{X}_{m+1} - X_{m+1} + \Delta t\, \beta\, F\big((m+1)\Delta t, X_{m+1}\big) = 0. \tag{23}$$

This equation emphasizes that the correction depends on the unknown state X_{m+1} and must represent state-dependent coupling between the field, forcing, and material response. In ASNO, this motivates a spatial refinement operator that incorporates long-range spatial interactions beyond temporal extrapolation alone. The NAO plays the role of a learned solver step, taking $\tilde{X}_{m+1}$ and F_{m+1} as inputs and producing a corrected field consistent with the implicit structure of equation 23, enabling generalization across process–material variations.

A.4 NAO AS A SURROGATE STATIC PDE SOLVER

We model the spatial refinement as solving a nonlinear static PDE surrogate at each time step. Let $h \in \mathcal{H}$ denote the latent state and $g \in \mathcal{F}$ denote the forcing field, including process inputs and thermophysical material descriptors. We write the operator form

$$L_K[h] + \epsilon = g, \tag{24}$$

where K is a learned nonlocal kernel and ϵ captures model mismatch. This interpretation treats NAO as an operator that converts extrapolated dynamics into a physically consistent corrected field by learning the long-range interactions induced by heat diffusion, convection, and alloy-dependent transport.

Kernel parameterization. NAO constructs the kernel via a learned attention map

$$\begin{aligned} K\big[h_{1:d}, g_{1:d}; \theta\big] = W_{P,h}\, \sigma\Big(\tfrac{1}{\sqrt{d_k}} (H_t)^\top W_{Q_t} (W_{K_t})^\top H_t\Big) \\ + W_{P,f}\, \sigma\Big(\tfrac{1}{\sqrt{d_k}} (F_t)^\top W_{Q_t} (W_{K_t})^\top H_t\Big), \end{aligned} \tag{25}$$

where $\theta = \{W_{P,h}, W_{P,f}, W_{Q_t}, W_{K_t}\}$ are trainable parameters and H_t and F_t are latent features from state and forcing. The first term models state–state coupling, while the second injects forcing–state interactions so refinement adapts to both process inputs and material response.

Discretization. Let the spatial mesh be $\{y_k\}_{k=1}^N$. The discretized inputs are

$$\tilde{X}_{1:d} = \big(\tilde{X}_j(y_k)\big)_{1\le j\le d,\ 1\le k\le N}, \quad F_{1:d} = \big(F_j(y_k)\big)_{1\le j\le d,\ 1\le k\le N}. \tag{26}$$

Iterative attention refinement. NAO processes features through T attention layers with residual connections

$$H_t = \mathrm{Attn}\big(H_{t-1}; \theta_t\big)\, H_{t-1} + H_{t-1}, \quad 1 \le t \le T, \tag{27}$$

which stabilizes refinement while incrementally adding physically meaningful long-range transport effects.

Integral operator output. After T layers, NAO outputs

$$X^{\mathrm{out}}_{1:d}(y) = \int K\big[\tilde{X}_{1:d}, F_{1:d}\big](y, z)\, F_{1:d}(z)\, \mathrm{d}z, \tag{28}$$

computed numerically on the mesh, producing alloy- and process-conditioned full-field corrections.

Training objective across S systems. To learn an operator transferable across regimes, we minimize

$$\mathcal{L} = \sum_{\eta=1}^{S} \Big\| \int K[\tilde{X}_{1:d}, F_{1:d}](y, z)\, F_{1:d}(z)\, \mathrm{d}z - X_{1:d}(y) \Big\|^2_{L^2}, \tag{29}$$

encouraging the kernel to capture shared spatial physics while adapting to heterogeneous forcing and material coefficients across candidate alloys.

A.5 Laplace approximation details and scalable Hessian approximations

We approximate the posterior around θ_{MAP} obtained from

$$\theta_{\mathrm{MAP}} = \arg\min_{\theta} \mathcal{R}(\mathcal{D}, \theta) = \arg\min_{\theta} \Big(l(\mathcal{D}, \theta) + r(\theta) \Big). \tag{30}$$

The posterior is

$$p(\theta \mid \mathcal{D}) \propto \exp\big(-\mathcal{R}(\mathcal{D}, \theta)\big). \tag{31}$$

A second order Taylor expansion gives

$$\theta \sim \mathcal{N}(\theta_{\mathrm{MAP}}, \Sigma), \quad \Sigma = H^{-1}, \tag{32}$$

which enables tractable propagation of epistemic uncertainty.

Practical approximations. Computing H exactly is infeasible for large networks. We adopt a diagonal approximation,

$$\Sigma \approx \mathrm{diag}\Big(\tfrac{1}{\lambda_1}, \ldots, \tfrac{1}{\lambda_n} \Big), \tag{33}$$

retaining scalability while capturing dominant sensitivities that govern uncertainty under sparse data and unseen material settings.

Predictive covariance. For the ASNO mapping

$$X_{m+1}^{\mathrm{out}} = \mathrm{NAO}\Big(\mathrm{TE}(X_m, \ldots, X_{m-n+1}), F_{m+1} \Big), \tag{34}$$

the predictive covariance is approximated by

$$\mathrm{Cov}\big[X_{m+1}^{\mathrm{out}}\big] \approx J_\theta^{(m+1)} \Sigma (J_\theta^{(m+1)})^\top, \tag{35}$$

providing spatially resolved confidence bounds that guide adaptive sampling and safe screening in materials discovery.

A.6 End to end algorithmic summary

Given history $\{X_m, \ldots, X_{m-n+1}\}$ and forcing F_{m+1}, ASNO computes $\tilde{X}_{m+1} = \mathrm{TE}(\cdot)$, refines it via $X_{m+1}^{\mathrm{out}} = \mathrm{NAO}(\tilde{X}_{m+1}, F_{m+1})$, and quantifies epistemic uncertainty by constructing a Gaussian posterior through Laplace approximation and propagating it through equation 35. This yields a transferable operator surrogate that predicts full spatio-temporal fields while communicating confidence, a key requirement for uncertainty-aware process–material discovery pipelines.

B Appendix, Benchmark Definitions and Experimental Details

This appendix specifies the benchmark problems and experimental procedures used to evaluate ASNO and all baselines on Lorenz, Darcy, and Navier–Stokes. While these benchmarks are canonical in scientific machine learning, they also capture core behaviors that arise in process–material discovery workflows, where transport dynamics, heterogeneous coefficients, and nonlinear feedback govern process–structure–property evolution. We describe the governing equations, numerical solvers, and data generation settings that define each dataset, and we detail the shared training and evaluation protocol used to ensure fair comparison across models. In addition, we include extended qualitative diagnostics that support interpretation of the quantitative results, including long-horizon rollout behavior and representative error patterns across regimes relevant to digital twins for manufacturing and accelerated materials exploration.

B.1 Common evaluation protocol

All models receive five historical states and forcing (when applicable) and predict the next state according to equation 16. We evaluate both one-step accuracy and multi-step stability using a consistent autoregressive rollout protocol for long-horizon evaluation, where model outputs are recursively fed back as inputs. This rollout setting directly measures robustness to compounding error, which is a dominant failure mode in chaotic, advection-dominated, and stiff transport systems that also appear in melt pool evolution and thermophysical process modeling.

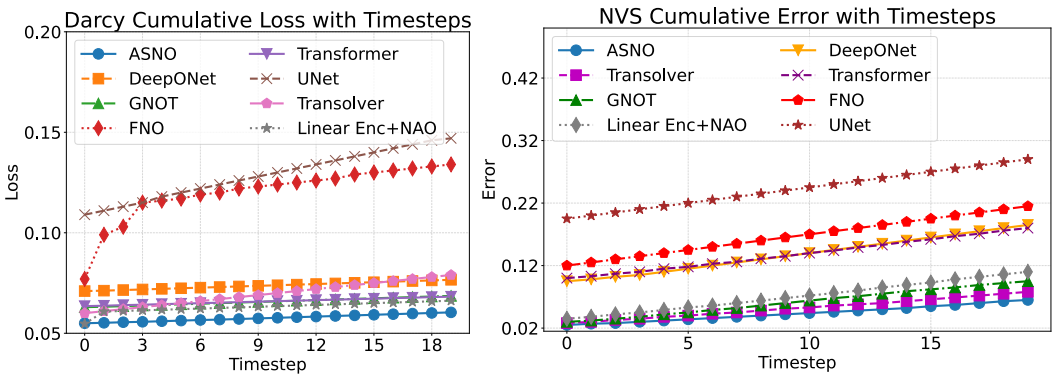

Figure 2: Cumulative L^2 error over time for Darcy and Navier–Stokes benchmarks. ASNO exhibits slower error growth, indicating more stable long-horizon performance.

To promote generalization across regimes, we augment the training data by applying 20 random permutations per trajectory. These permutations correspond to spatial shuffles that preserve global structure and boundary conditions, ensuring physical validity of the augmented states. This augmentation increases diversity while avoiding leakage of unphysical artifacts, and it encourages the learned operator to rely on transferable physical dependencies rather than trajectory-specific coincidences. Such robustness is essential when surrogates are deployed to screen new material coefficients or unseen processing conditions.

To assess long-horizon stability, we compute cumulative L^2 error over time using equation 17. This metric emphasizes error growth behavior rather than isolated step-wise deviations, and it distinguishes models that remain stable from those that drift due to small systematic biases. Figure 2 shows cumulative error curves for Darcy and Navier–Stokes, where ASNO exhibits consistently slower error growth compared to baselines, indicating improved stability for iterative surrogate rollouts in process–material exploration settings.

B.2 Lorenz System

The Lorenz benchmark is generated using the canonical chaotic ODE

$$\frac{dx}{dt} = \sigma(y - x), \quad \frac{dy}{dt} = x(\rho - z) - y, \quad \frac{dz}{dt} = xy - \beta z, \tag{36}$$

with $\sigma = 10$, $\rho = 28$, and $\beta = \frac{8}{3}$. This system exhibits sensitive dependence on initial conditions and therefore provides a stringent test of long-horizon predictive stability. In a materials context, such chaotic sensitivity reflects regimes where small process perturbations or uncertain material response can lead to divergent thermal trajectories, motivating stable surrogate rollouts.

We generate ground truth trajectories by discretizing equation 36 with fourth-order Runge–Kutta and time step $\Delta t = 0.01$, producing time series suitable for both one-step and autoregressive evaluation.

We evaluate models on long-horizon recursive rollouts to probe stability under compounding error. In this regime, even small one-step biases can push trajectories off the attractor, causing drift or collapse. Figure 3 compares ASNO to ground truth over extended horizons. ASNO maintains alignment with the true attractor structure, whereas baseline models tend to drift or collapse to spurious states as prediction length increases, indicating reduced robustness to accumulated error.

Additional ablations on history length, noise sensitivity, and rollout horizon are included in supplementary materials, and they further characterize the stability region in which the learned dynamics remain consistent with the underlying attractor geometry.

B.3 Darcy Flow

For Darcy flow, the goal is to learn the solution $u(x)$ to the elliptic PDE

$$-\nabla \cdot (a(x)\nabla u(x)) = f(x), \quad x \in \Omega \subset \mathbb{R}^2, \tag{37}$$

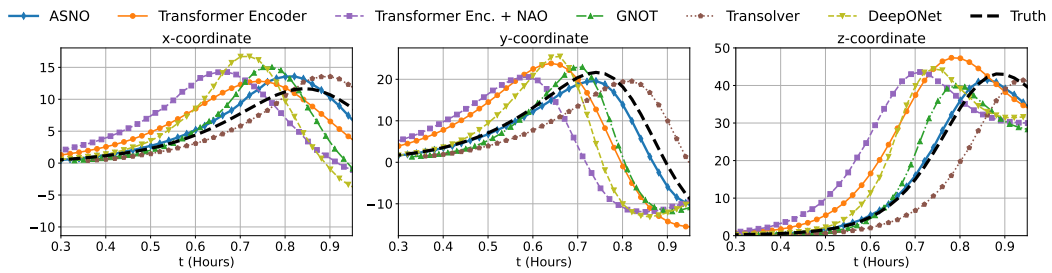

Figure 3: Long-horizon prediction on the Lorenz system. ASNO preserves the attractor structure and remains aligned with ground truth over extended rollouts.

subject to Dirichlet boundary conditions

$$u(x) = 0, \quad x \in \partial\Omega. \tag{38}$$

This benchmark emphasizes learning a spatial operator under strong heterogeneity, where long-range correlations arise from the coefficient field and boundary constraints. Here $a(x)$ is a heterogeneous permeability field and $f(x)$ is a known forcing term. In process–material discovery, this benchmark is directly analogous to learning transport response under uncertain conductivity or permeability landscapes across candidate materials.

Data are generated by solving equation 37 with varying realizations of $a(x)$ using a finite volume solver on a 21×21 grid, yielding paired samples of coefficients and corresponding solutions.

We evaluate both in-distribution and out-of-distribution generalization by varying the statistics of $a(x)$ and the forcing term $f(x)$. This probes whether a learned operator captures elliptic structure rather than overfitting to a narrow coefficient distribution, a key requirement when screening new material property settings. Representative qualitative solution snapshots, error maps, and resolution sensitivity studies are provided in supplementary figures to illustrate typical failure modes, including oversmoothing and loss of long-range spatial correlations.

B.4 Navier–Stokes equations

For Navier–Stokes, we use the 2D incompressible vorticity–streamfunction formulation

$$\frac{\partial \omega}{\partial t} + \nabla^{\perp}\psi \cdot \nabla\omega = \nu\Delta\omega, \quad -\Delta\psi = \omega, \tag{39}$$

where ω is vorticity, ψ is streamfunction, and ν is kinematic viscosity. This formulation isolates the evolution of vorticity under advection and diffusion while enforcing incompressibility through the Poisson solve for ψ. In additive manufacturing, closely related convection and instability mechanisms arise in melt pool flow, where preserving coherent structures is essential for predicting defects and microstructural variability.

Simulations are conducted on a periodic domain using spectral discretization, which provides high-fidelity reference trajectories and supports stable long rollouts.

We evaluate long-horizon rollouts, cumulative errors, and qualitative preservation of coherent structures such as vortical filaments and large-scale circulation patterns. In this benchmark, overly dissipative models can appear accurate at short horizons but progressively wash out fine-scale vorticity under autoregressive iteration. Additional figures are provided in the supplementary material, including comparisons of dissipation behavior, stability trends across rollout length, and sensitivity to different viscosities.

C Appendix, Uncertainty Quantification Verification Details

This appendix provides a detailed description of how uncertainty quantification (UQ) is evaluated on the Darcy benchmark, including dataset construction, spatial resolution, metric computation, and statistical aggregation. Although Darcy flow is a canonical SciML testbed, it also serves as a controlled proxy for process–material discovery settings, where transport response depends on uncertain or varying coefficients (e.g., permeability, conductivity) and reliable surrogates must report confidence when screening new material property regimes. Beyond reporting numerical values, we clarify

how the UQ metrics relate to calibration and decision utility, and why this protocol is appropriate for validating ASNO as an uncertainty-aware neural operator that can support uncertainty-guided exploration and risk-aware optimization.

C.1 Dataset and aggregation

We compute Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW) on a held-out Darcy test set consisting of 100 independent samples and 95 time steps per sample, using the full $21 \times 21 = 441$ spatial grid. This spatial resolution matches the training and evaluation setup in Appendix B.3, ensuring that uncertainty is assessed at the same field fidelity as the learned operator. Evaluating uncertainty at full-field resolution is important for coefficient-driven transport systems because uncertainty is typically spatially inhomogeneous, for example near sharp transitions or regions where heterogeneous coefficients induce localized sensitivity.

Metrics are computed per time step and per spatial location rather than collapsing the field into a single scalar, enabling spatially resolved assessment of where the model is confident versus uncertain. After computing PICP and MPIW at each spatial point and time step, we average across all test samples and time steps to obtain the aggregate statistics reported in Table 2. This aggregation balances local calibration behavior with global representativeness, which is aligned with downstream discovery use cases where a surrogate must remain reliable across an entire domain rather than only at a few monitored points.

C.2 Interpretation of calibration and sharpness

PICP measures calibration, namely the empirical frequency with which the ground-truth value lies within the predicted nominal 95% interval. Values below 0.95 indicate under-coverage, meaning intervals are too narrow and the model is overconfident, which is unsafe for exploration and optimization because it can lead to incorrect ranking of candidate material settings. Values above 0.95 indicate over-coverage, meaning intervals are conservative and may be safe but potentially reduce decision efficiency by failing to discriminate among candidates.

MPIW measures sharpness, the average width of the predicted intervals. Smaller MPIW corresponds to tighter bounds and higher decision resolution, but sharpness must be interpreted jointly with calibration. A model with very small MPIW but poor PICP is unreliable, whereas a model with high PICP but extremely large MPIW can be uninformative for selecting among process–material candidates.

ASNO achieves PICP close to the nominal 95% level while maintaining a modest MPIW, indicating that the uncertainty intervals are neither overconfident nor excessively conservative. This balance suggests that the Laplace-based uncertainty mechanism captures meaningful parameter sensitivity while still providing usable resolution for downstream tasks such as uncertainty-aware ranking, adaptive sampling, and Bayesian optimization over coefficient regimes.

C.3 Epistemic versus aleatoric uncertainty

The uncertainty intervals reported by ASNO primarily reflect epistemic uncertainty arising from limited training coverage, parameter ambiguity, and model form mismatch between the learned operator and the true governing PDE. This uncertainty is quantified through a Laplace approximation of the parameter posterior, which captures how predictions change under local perturbations of the learned weights around θ_{MAP}. Epistemic uncertainty is reducible, meaning it can be lowered by collecting more training data, improving operator expressivity, or incorporating stronger physics priors.

In process–material discovery, epistemic uncertainty is particularly valuable because it signals where the surrogate is extrapolating across coefficient space, for example when screening permeability or conductivity settings that are not well represented in the training distribution. In such regimes, wider intervals appropriately indicate low confidence and can be used to trigger additional simulation or experimental acquisition through active learning.

In contrast, aleatoric uncertainty arises from irreducible stochasticity or observation noise and cannot be eliminated simply by collecting more data. While this work emphasizes epistemic uncertainty, the calibration behavior indicates that ASNO's confidence intervals remain sensible across both familiar and challenging regimes. Future work can incorporate explicit aleatoric modeling, for example

via heteroscedastic likelihoods or conformal prediction, to more explicitly separate epistemic and aleatoric contributions and further strengthen uncertainty-guided materials screening and robust decision making.

D Appendix, DED Data Generation and Simulation Setup

This appendix provides the full technical specification of the GAMMA simulation setup used to generate training and test data for DED temperature field prediction. Beyond serving as a manufacturing benchmark, this dataset is constructed to support process–material discovery, where digital twins must rapidly evaluate how changes in thermophysical properties and processing conditions influence melt pool dynamics, microstructure, and part performance. The description follows the physical modeling pipeline, from governing equations and laser physics to discretization, geometry, material properties, and dataset construction, clarifying how the synthetic data represents realistic DED thermal behavior while enabling large-scale neural operator training.

D.1 Governing equation and constitutive relations

GAMMA solves the transient heat conduction equation

$$\rho\, C_p(T)\, \frac{\partial T}{\partial t} + \nabla \cdot q = 0, \tag{40}$$

where T is temperature (K), ρ is density (g/mm^3), and $C_p(T)$ is the temperature-dependent specific heat capacity (J/g/K). This formulation captures internal energy storage and transport in the evolving thermal field.

Heat flux is computed using Fourier's law

$$q = -k(T)\, \nabla T, \tag{41}$$

with temperature-dependent thermal conductivity $k(T)$ (W/m·K). The dependence of both $C_p(T)$ and $k(T)$ on temperature is essential for capturing nonlinear thermal response near melting, and it provides a mechanism for representing alloy-to-alloy variability in discovery settings. These relations define the core physics that ASNO learns to emulate as a surrogate for rapid screening of process and material regimes.

D.2 Laser boundary flux and Gaussian intensity profile

The laser is modeled through a boundary condition on the outward normal heat flux,

$$q \cdot n = -\frac{2\eta\, P_{\text{laser}}}{\pi r_{\text{laser}}^2} \exp\left(-\frac{2r^2}{r_{\text{laser}}^2}\right) + \sigma\, \epsilon\, (T^4 - T_0^4) + h\, (T - T_0), \tag{42}$$

where P_{laser} is laser power, r_{laser} is beam radius, and r is radial distance to the beam center. Here η is absorption coefficient, h is convective heat transfer coefficient, σ is the Stefan–Boltzmann constant, ϵ is emissivity, and T_0 is ambient temperature.

This boundary condition couples laser energy input with radiation and convection losses, ensuring that both heating and cooling mechanisms are represented consistently. In process optimization and material discovery, these terms determine how candidate alloys respond to identical energy input through different absorption or heat-loss behavior.

The corresponding Gaussian laser intensity is

$$I_{\text{laser}}(r) = \frac{2P_{\text{laser}}}{\pi r_{\text{laser}}^2} \exp\left(-\frac{2r^2}{r_{\text{laser}}^2}\right), \tag{43}$$

which concentrates energy near the beam center while decaying smoothly with distance, generating physically meaningful melt pool transients.

D.3 Moving laser protocol, sampling rate, and dataset size

GAMMA simulates a moving laser over the build geometry. At each timestep m, laser power $P_{\text{laser}}(m)$, scanning rate $V_{\text{scan}}(m)$, and beam location $L_{\text{laser}}(m)$ are specified, producing a spatio-

temporal dataset that couples process inputs with thermal evolution. This variability is critical for learning surrogates that generalize across manufacturing recipes and enable data-efficient exploration of new process–material settings.

Data are collected every 0.02 seconds, yielding 15,000 timesteps per simulation. This temporal resolution captures rapid melt pool heating and cooling transients needed for accurate operator learning.

Data are generated for 100 distinct laser power profiles to introduce diverse thermal histories and prevent overfitting to a single regime. A 5 second sliding window corresponds to 250 timesteps, defining the history length used by ASNO.

With stride one, each simulation yields

$$15{,}000 - 250 + 1 = 14{,}751 \tag{44}$$

training samples, leading to a total of

$$14{,}751 \times 100 = 1{,}475{,}100 \tag{45}$$

datapoints. This scale provides broad coverage of process trajectories and supports stable training of a high-capacity surrogate suitable for accelerated discovery workflows.

D.4 Geometry, grid, and scan path

Simulations are performed on a 21×21 spatial grid over a thin-wall structure with dimensions 30 mm in height and 22.5 mm in width. This resolution balances computational tractability with sufficient spatial detail to resolve melt pool features that influence microstructure and defect formation.

The laser follows a predetermined path along the x axis, alternating direction between passes to mimic realistic DED toolpaths. After each pass, the beam shifts upward by 0.75 mm along the z axis, depositing a new layer. This protocol introduces strong spatial and temporal variability in thermal history, which is essential when learning transferable representations across geometries and candidate material conditions.

D.5 Material properties

Thermal properties correspond to stainless steel powder with a 1018 carbon steel substrate. Thermal conductivity varies from 0.01396 W/m·K at 300 K to 0.03439 W/m·K at 1600 K, reflecting increased heat transport at elevated temperatures. Specific heat capacity varies from 0.512 J/g·K at 400 K to 0.770 J/g·K at 1800 K, capturing nonlinear response near melting.

These temperature-dependent thermophysical properties are central to melt pool prediction and represent the types of material descriptors that vary across alloys in materials discovery. Conditioning surrogates on such properties enables digital twins to rapidly screen how changes in $k(T)$ and $C_p(T)$ reshape thermal fields, melt pool stability, and downstream process–structure relationships under new candidate compositions.

E Appendix, DED Model Inputs, Architecture Settings, and Training Details

This appendix specifies the full input encoding, neural architecture settings, and training protocol used for applying ASNO to Directed Energy Deposition (DED). The goal of this setup is not only accurate thermal forecasting, but also enabling material–process discovery, where the surrogate learns how coupled laser controls and alloy thermophysical properties jointly govern melt pool evolution, cooling rates, and solidification behavior. Following the UNiCO-AM formulation, ASNO is designed as a material–process agnostic operator surrogate that generalizes across diverse alloys, while remaining sensitive to the underlying temperature-dependent constitutive physics through explicit material property conditioning.

E.1 Input and output configuration

ASNO performs full-field temperature prediction by conditioning on both historical thermal states and next-step process–material design variables. This reflects the physical structure of DED, where

the future temperature field depends not only on previous heating and diffusion history, but also on externally applied laser energy inputs and intrinsic alloy-specific heat transport properties.

At each time step m, the model receives a history window of five melt pool temperature fields together with the forcing variables at $m+1$:

$$\Big(T_{\text{pool}}(m-4:m),\ x_p(m+1),\ x_m(m+1)\Big), \tag{46}$$

where the process inputs are

$$x_p = \big[P_{\text{laser}},\ V_{\text{scan}},\ L_{\text{laser}}\big], \tag{47}$$

and the material inputs are defined following the concurrent material–process design space in UNiCO-AM:

$$x_m = \big[\rho,\ T_{\text{solidus}},\ T_{\text{liquidus}},\ C_p(T),\ k(T)\big]. \tag{48}$$

Here, $C_p(T)$ and $k(T)$ are temperature-dependent thermophysical functions discretized at seven representative temperature points across the alloy solidification interval, enabling ASNO to learn how variations in conductivity and heat capacity regulate thermal diffusion, melt pool morphology, and cooling efficiency. This parameterization supports inverse materials discovery, since the surrogate can interpolate and generalize across continuous property profiles rather than being restricted to fixed discrete alloys. :contentReference[oaicite:1]index=1

The learning target is the next full-field temperature map:

$$T_{\text{pool}}(m+1), \tag{49}$$

yielding a field-to-field operator surrogate suitable for downstream co-design and Bayesian optimization.

E.2 Transformer Encoder settings

The temporal extrapolation component of ASNO is implemented using a Time Series Transformer Encoder (TE). In DED, temporal evolution is strongly nonstationary due to laser motion, track overlap, and layerwise reheating. The TE therefore learns a nonlinear extrapolation map that stabilizes rollouts under stiff moving-heat-source dynamics.

For the DED experiments, the TE is configured with an embedding dimension of 100, 50 attention heads, and 3 transformer layers. Given the input sequence $T_{\text{pool}}(m-4:m)$, the TE produces a latent extrapolated state

$$H_{m+1} = \text{TE}\big(T_{\text{pool}}(m-4:m)\big), \tag{50}$$

which corresponds to the explicit multistep extrapolation stage in the IMEX interpretation of ASNO.

This design is critical in additive manufacturing because the melt pool thermal history exhibits long-memory effects, where previous heating episodes influence future diffusion, residual heat accumulation, and solidification kinetics.

E.3 Spatial refinement and process–material coupling via NAO

The second stage of ASNO performs spatial refinement through the Nonlocal Attention Operator (NAO), which corrects the latent extrapolated prediction produced by the Transformer Encoder by explicitly modeling spatial heat redistribution and coupled process–material interactions within the melt-pool region. This refinement step is essential in Directed Energy Deposition (DED), where the temperature field exhibits steep spatial gradients near the melt pool boundary, rapid heating–cooling cycles, and strong nonlocal transport effects driven jointly by laser forcing and alloy-dependent thermophysical response. In such regimes, temporal extrapolation alone is insufficient, since the spatial structure of the melt pool is governed by long-range conduction pathways, heterogeneous diffusion strength, and material-specific solidification behavior.

Following the UNiCO-AM formulation, ASNO predicts the evolution of a localized *melt-pool temperature window*, defined as a moving spatio-temporal region centered around the laser location that captures the dominant thermal physics of deposition. This window continuously tracks the laser path and provides a physically meaningful functional representation of melt pool dynamics across different alloys and process settings. As described in the NAMRC framework, the temperature window is discretized at high spatial resolution, enabling ASNO to learn field-to-field mappings that

reproduce transient heat transfer, cooling rates, and solidification behavior in a material-agnostic manner.

Concretely, ASNO decouples prediction into a temporal encoding step followed by NAO-based spatial correction. The Transformer Encoder produces a latent extrapolated melt-pool state

$$H_{m+1} = \mathrm{TE}\big(Y_{m-n+1:m}; \theta_{\mathrm{TE}}\big), \tag{51}$$

which captures the evolving temporal history of the melt pool. The NAO then refines this extrapolated state by conditioning on the *joint process–material design vector*

$$x = [x_p, x_m], \tag{52}$$

where the process variables encode laser forcing and the material variables encode alloy thermophysical behavior. In the DED co-design setting, these inputs are defined as

$$x_p = [P_{\mathrm{laser}}], \qquad x_m = [\rho,\ T_S,\ T_L,\ C_p(T),\ k(T)], \tag{53}$$

with density ρ, solidus and liquidus temperatures (T_S, T_L), and temperature-dependent conductivity and heat capacity profiles $k(T)$ and $C_p(T)$ discretized at representative temperature points to span the solidification interval.

The resulting NAO refinement step is given by the operator update

$$\hat{Y}_{m+1} = \mathrm{NAO}\big(H_{m+1}; x, \theta_{\mathrm{NAO}}\big) = H_{m+1} + \int_{\Omega} \alpha(x, x'; x, \theta_{\mathrm{NAO}})\, W_{\phi}\big(H_{m+1}(x')\big)\, dx', \tag{54}$$

where $\alpha(\cdot)$ is a learned attention kernel that encodes nonlocal spatial coupling, and W_ϕ is a learned feature projection. This formulation explicitly represents the physical role of NAO as a surrogate correction operator, injecting spatial consistency into the predicted melt-pool field while capturing how process energy input interacts with alloy-dependent conduction and solidification mechanisms.

Critically, conditioning NAO on functional material descriptors $k(T)$ and $C_p(T)$ enables ASNO to learn *material-process agnostic nonlocal kernels* that adapt diffusion strength, melt pool stability, and cooling dynamics across diverse thermophysical regimes. This is a central requirement for accelerated material discovery, since the optimizer is not restricted to a single fixed alloy but instead explores a continuous space of effective thermophysical behaviors. As emphasized in UNiCO-AM, such operator surrogates enable generalization to unseen material–process combinations and support the discovery of novel alloy-property profiles that maximize cooling efficiency while maintaining thermal stability.

Therefore, NAO serves as the key mechanism by which ASNO transitions from purely predictive modeling to a discovery-capable surrogate: it enforces physically consistent spatial refinement while embedding the coupled influence of laser forcing and alloy-dependent thermal transport, enabling uncertainty-aware exploration of new process–material design spaces in additive manufacturing.

E.4 Training protocol and evaluation

ASNO and all baseline surrogates are trained under a unified protocol to ensure fair comparison and to reflect the requirements of data-driven material–process discovery in Directed Energy Deposition (DED). Because thermal fields in AM span large dynamic ranges across alloys, scan strategies, and heat accumulation regimes, all temperature outputs are standardized, and all scalar process–material inputs are normalized to improve gradient conditioning and stabilize optimization across highly varying magnitudes.

Dataset organization across process–material space. Training data are constructed from high-fidelity melt-pool temperature fields generated using the GAMMA DED simulator, which provides transient solutions across multiple representative alloys and laser trajectories. Each training sample corresponds to a localized moving melt-pool window of size $10\,\mathrm{mm} \times 10\,\mathrm{mm}$ centered on the laser position and discretized into a 30×30 grid, yielding 900 spatial degrees of freedom that capture the physically dominant thermal evolution during deposition **?**. By focusing learning on this active region, ASNO effectively models heat transfer, cooling rate, and solidification behavior across diverse material–process combinations.

To ensure balanced generalization rather than memorization of a narrow regime, we adopt a stratified 80–10–10 split across joint process–material pairs, enforcing coverage across both laser power profiles

and alloy thermophysical variability during training and evaluation. This protocol is essential in AM discovery settings, where small changes in conductivity, density, or liquidus temperature can induce fundamentally different melt-pool dynamics.

Curriculum-based rollout stabilization. Since autoregressive deployment in digital twins requires stable long-horizon prediction under compounding error, ASNO is trained using a short-to-long rollout curriculum over increasing temporal horizons (10, 25, and 50 time steps). This curriculum encourages robustness of the learned operator over extended deposition sequences, where thermal accumulation and layer transitions amplify small surrogate biases. Each training input consists of five preceding melt-pool windows $Y_{m-n+1:m}$, which encode the recent spatio-temporal history of local heating and cooling in the deposition process.

Optimization and training hyperparameters. Model parameters $\theta = [\theta_{\mathrm{TE}}, \theta_{\mathrm{NAO}}]$ are learned via MAP estimation using a mean squared error objective,

$$L_{\mathrm{MSE}} = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{T_j}\sum_{m}\left\|\hat{Y}_{m+1}^{(j)} - Y_{m+1}^{(j)}\right\|_2^2, \tag{55}$$

which yields a stable reference solution $\theta^\star$ that also initializes epistemic uncertainty quantification through Laplace approximation. Training is conducted for 500 epochs using mini-batches of size 64, together with a learning-rate scheduler using decay factor $\gamma = 0.1$ to ensure stable convergence and robust cross-alloy generalization.

Quantitative surrogate fidelity and efficiency. We report training efficiency (time per epoch), trainable parameter counts, and best test loss in Table 3. ASNO achieves the lowest test loss (0.0140) and a mean absolute percentage error (MAPE) of 4.5% on in-distribution alloys, confirming high-fidelity reconstruction of transient melt-pool dynamics and accurate alignment with physics-based thermal behavior observed in GAMMA simulations. Compared against operator learning architectures such as FNO, DeepONet, UNet, and GNOT, ASNO provides the strongest balance between predictive accuracy and computational efficiency, making it particularly suitable for iterative digital twin optimization loops.

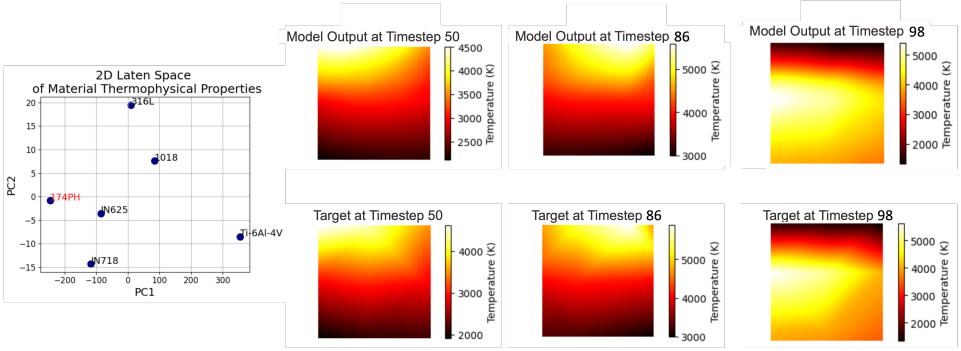

Figure 4: ASNO generalization to 17-4 PH stainless steel, showing close agreement with GAMMA simulations across time.

Cross-material generalization for alloy discovery. A key requirement for accelerated materials discovery is extrapolation to thermophysical regimes unseen during surrogate training. We therefore evaluate ASNO in a zero-shot setting on a held-out alloy (17-4 PH stainless steel), which occupies a distinct region in latent material property space due to differences in solidus/liquidus temperatures, density, and conductivity. Despite these shifts, ASNO predicts coherent melt-pool evolution, accurately capturing hot spots, melt-pool contraction, and global heat dissipation trends. Figure 4 illustrates close agreement between ASNO predictions and GAMMA simulations across time, demonstrating that the learned process–material embedding supports discovery-driven generalization beyond the training alloys. The resulting out-of-distribution test loss of 0.0356 remains acceptable for BO-guided exploration of new alloy–process combinations.

Qualitative evaluation and epistemic uncertainty. To provide qualitative evidence of full-field surrogate fidelity for process–material discovery, Figure 5 visualizes ASNO predictions against

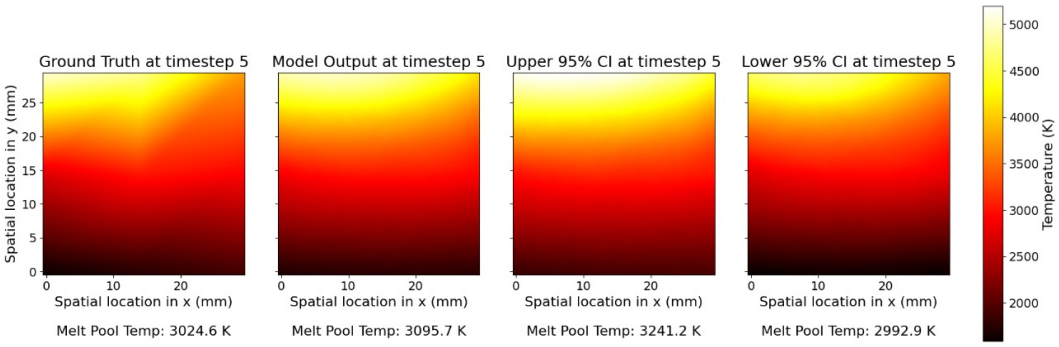

Figure 5: Appendix visualization of full-field DED temperature prediction. ASNO predictions are compared to GAMMA ground truth with 95% epistemic confidence bounds computed via Linear Laplace approximation, supporting uncertainty-guided process–material exploration.

GAMMA ground truth together with spatially resolved 95% epistemic confidence bounds obtained via diagonal Linear Laplace approximation. In DED, thermal gradients, melt pool boundaries, and cooling transients vary sharply across scan strategies and alloy thermophysical properties. ASNO accurately reconstructs these localized structures while simultaneously reporting calibrated uncertainty. Notably, uncertainty increases near layer transitions, in regions far from the laser path, and under out-of-distribution material–process regimes, reflecting reduced training density and higher epistemic ambiguity. Such uncertainty-aware field predictions are essential for safe Bayesian optimization, adaptive sampling, and closed-loop exploration of new alloys in self-driving materials laboratories, where confidence bounds directly indicate where additional experiments or simulations are most informative.

F Appendix, Uncertainty Computation for Full-Field DED Prediction

This appendix clarifies how 95% epistemic confidence bounds are computed for full-field DED predictions and how they should be interpreted in the context of digital twins for accelerated process–material discovery.

We adopt a diagonal Linear Laplace approximation over network parameters to estimate epistemic uncertainty arising from limited training coverage, parameter ambiguity, and alloy-dependent regime shifts. This provides a scalable Bayesian treatment of uncertainty without requiring expensive Monte Carlo sampling, making it suitable for real-time deployment and iterative design loops.

The predictive mean is taken as the ASNO output at the MAP estimate θ_{MAP}. Predictive variance is approximated using Jacobian-based covariance propagation as described in Appendix C, which propagates posterior parameter uncertainty through the nonlinear spatio-temporal operator mapping. This yields spatially resolved uncertainty fields over the full melt pool window.

These confidence intervals can be interpreted as the surrogate digital twin's belief over competing thermal evolutions under uncertain alloy thermophysical settings. In downstream optimization, they provide principled acquisition signals for risk-aware control and adaptive experimentation, enabling safe screening of new materials and process parameters.

G Appendix, Local Feature Extraction: Melt Pool Temperature and Depth

This appendix provides the complete definitions and discretization details for melt pool feature extraction. These local melt pool descriptors serve as physically meaningful links between predicted thermal fields and downstream structure–property outcomes, making them critical quantities for process monitoring, defect prevention, and data-efficient materials discovery.

G.1 Melt pool temperature

We define the molten region using a liquidus threshold T_{thr}. For 316L stainless steel we set $T_{\mathrm{thr}} = 1680$ K, consistent with metallurgical phase transition data.

The molten set is

$$M = \{(x, y) : T(x, y) > T_{\mathrm{thr}}\}, \tag{56}$$

and melt pool temperature is computed as

$$T_{\text{melt}} = \frac{1}{|M|} \sum_{(x,y)\in M} T(x, y), \tag{57}$$

where $|M|$ denotes the number of molten grid locations.

This feature captures the average thermal severity of deposition, directly influencing cooling rate, grain morphology, and microstructural evolution, making it an actionable target in alloy-sensitive process optimization.

G.2 Melt pool depth

For each column x in the spatial grid, let $y_{\max}(x)$ and $y_{\min}(x)$ denote the maximum and minimum y coordinates such that $T(x, y) > T_{\text{thr}}$. This identifies the vertical penetration depth of melting at each lateral position.

The melt pool depth is estimated as

$$d_{\text{melt}} = \frac{1}{N_x} \sum_{x=1}^{N_x} \big[y_{\max}(x) - y_{\min}(x)\big], \tag{58}$$

where N_x is the number of spatial columns.

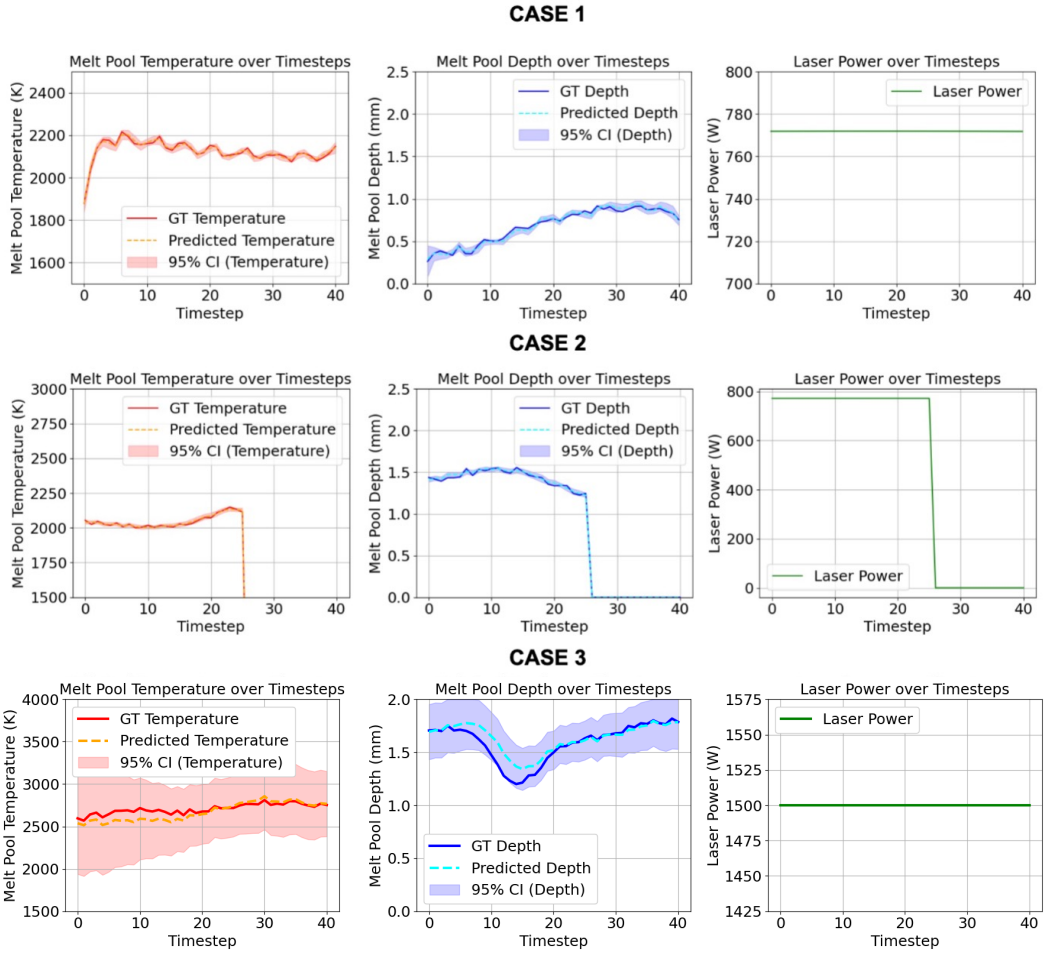

Figure 6: Local feature extraction from full-field predictions. Predicted and ground-truth melt pool temperature and depth are shown with laser power profiles and 95% confidence intervals across three DED test scenarios over 40 timesteps.

Depth is a key indicator of fusion quality, bonding integrity, and lack-of-fusion defect risk, and thus forms a central constraint in uncertainty-aware co-design of alloys and scan strategies.

G.3 Evaluation scenarios

We evaluate melt pool feature extraction under three representative DED scenarios designed to probe both steady-state thermal behavior and transient out-of-distribution dynamics:

(i) constant laser power of approximately 770 W, representing nominal deposition,

(ii) laser shutoff to simulate abrupt layer transitions and process interruptions, and

(iii) an out-of-distribution regime with elevated laser power of 1500 W, stressing generalization to unseen energy inputs and material responses.

In each case, ASNO produces full-field thermal predictions together with 95% epistemic confidence bounds via Linear Laplace approximation. Extracted melt pool temperature and depth are compared against GAMMA ground truth over 40 timesteps, as shown in Figure 6.

These scenarios collectively assess whether ASNO can robustly track melt pool evolution under normal operation, respond correctly to sharp process perturbations, and provide meaningful uncertainty signals when extrapolating to new high-energy or alloy-dependent regimes, enabling adaptive experimentation in automated materials discovery pipelines.