POLITECNICO DI TORINO
Repository ISTITUZIONALE

Efficient analysis of composites manufacturing using multi-fidelity simulation and probabilistic machine learning

Original
Efficient analysis of composites manufacturing using multi-fidelity simulation and probabilistic machine learning / Schoenholz, C., Zappino, E., Petrolo, M., Zobeiry, N.. - In: COMPOSITES. PART B, ENGINEERING. - ISSN 1359-8368. - ELETTRONICO. - 280:(2024). [10.1016/j.compositesb.2024.111499]

Availability:
This version is available at: 11583/2988922 since: 2024-05-22T13:42:45Z

Publisher:
Elsevier

Published
DOI:10.1016/j.compositesb.2024.111499

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Elsevier postprint/Author's Accepted Manuscript

© 2024. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/.The final authenticated version is available online at:
http://dx.doi.org/10.1016/j.compositesb.2024.111499

(Article begins on next page)

21 November 2026

REVISED Manuscript (revised text UNMARKED) Click here to view linked References

Efficient Analysis of Composites Manufacturing Using Multi-fidelity Simulation and Probabilistic Machine Learning

Caleb Schoenholz [a], Enrico Zappino [b], Marco Petrolo [b], Navid Zobeiry [a,*]

[a] *Materials Science & Engineering Department, University of Washington*
302 Roberts Hall, Box 352120
Seattle, WA, 98195-2120, U.S.A.

[b] *MUL2 Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino*
Corso Duca degli Abruzzi 24
10129 Torino, Italy

[*] *Corresponding author, navidz@uw.edu*

Abstract

This paper introduces an innovative approach for the efficient analysis of composites manufacturing processes and phenomena. The method combines low- and high-fidelity simulation schemes with limited amounts of experimental data to train surrogate machine learning (ML) models. Guided by a novel approach, Spatially Weighted Gaussian Process Regression (SWGPR), a predictive model is efficiently constructed and calibrated by assigning datapoint-dependent noise levels to simulation points, establishing a multi-scale data-driven uncertainty structure. This study demonstrates the effectiveness of the method in accurately predicting process-induced deformations (PIDs) for L-shaped cross-ply laminates using minimal experimental efforts. The presented method aims to provide a cost-effective and broadly applicable framework for understanding and improving the design, development, and manufacturing of composites.

Keywords: A. Polymer-matrix composites (PMCs), B. Residual/internal stress, C. Statistical properties/methods, E. Autoclave, OTHER: Probabilistic machine learning

1. Introduction

While carbon fiber-reinforced polymer (CFRP) composites are widely utilized in the aerospace industry, manufacturers continue to face several challenges, one of which is mitigating process-induced deformations (PIDs), which significantly affect aerostructure assembly timelines [1–3]. During high-temperature/pressure processing (e.g., autoclave) of composites, residual stresses develop due to complex, interdependent, and multi-scale phenomena [1]. Upon demolding, PIDs such as variations in a part's enclosed angles (e.g., spring-in) or warpage of initially flat sections, may emerge to alleviate some stresses [4]. Figure 1 schematically illustrates these PIDs for an L-shaped part, a common geometry utilized for composite aerostructure components. Both spring-in and warpage can lead to assembly joining gaps, extended production timelines and costs, and compromised mechanical performance in the final structure [5].

Although PIDs are recognized among manufacturers, minimizing undesired deformations via process (e.g., cure cycle) optimization remains challenging. These difficulties primarily stem from limitations of traditional methods used for manufacturing analyses and PID predictions, typically classified into three categories, as depicted in Figure 1: low-fidelity simulation, high-fidelity simulation, and experimentation. One notable constraint of these methods is the trade-off between fidelity/accuracy and time/cost [6].

At the bottom end of the cost/accuracy spectrum, low-fidelity simulation methods, including analytical solutions and/or reduced-order (i.e., 1D/2D) finite element (FE) strategies, often provide rapid and easy-to-interpret analyses while requiring minimal material characterization and computational efforts [7–9]. However, many of these approaches often fail to describe complex processing phenomena (e.g., tool-part interaction

[10]), leading to deviations from reality. Positioned in the middle, high-fidelity simulation methods (i.e., 3D FE), often provide more precise predictions but require substantial time commitments for material characterization, calibration, and validation [11–14]. Moreover, despite their improved accuracy, high-fidelity simulations typically neglect processing uncertainty and are impractical for optimization problems with large design spaces due to higher computational requirements. At the highest cost/accuracy level lies experimentation, involving the hands-on manufacturing and analysis of composites. This approach offers several advantages, including the ability to bypass material characterization and capture the probabilistic nature of composites, but can be extremely time-consuming and expensive for large-scale composite structures.

The cost/accuracy trade-offs pose a dilemma for manufacturers when selecting methods to predict process phenomena and outcomes. One common strategy involves generating an extensive set of simulation data alongside a smaller experimental dataset. Attempts are then made to calibrate the simulations by connecting virtual and real-world domains using deterministic methods (e.g., least-squares). However, due to the difficulties and expenses in obtaining large data, the virtual-to-real connection and calibration attempts often rely on suboptimal amounts of data, leading to inaccurate manufacturing assessments and inefficient process optimization attempts.

Given the outlined challenges, there exists an opportunity to explore alternative and more efficient process analysis approaches for composites. This paper introduces one such approach, which follows the subsequent workflow for a case study focused on PIDs of L-shaped composite parts. First, a low-fidelity FE simulation scheme, based on 1D thermo-chemical and 2D thermo-mechanical analyses, is used to rapidly compute PIDs for composite

laminates in a defined design space. The low-fidelity virtual data is then mapped to a theory-guided domain using Classical Laminated Plate Theory (CLPT), Bayesian Information Criterion (BIC), and the probabilistic machine learning (ML) technique, Gaussian Process Regression (GPR). The theory-guided GPR model is then iteratively calibrated by incorporating limited amounts of high-fidelity 3D simulation data, based on 1D models and the Carrera Unified Formulation (CUF), and select experimental results. Each calibration step involves assigning noise levels to simulation datapoints through a novel Gaussian distance-decay weighing mechanism, creating an adaptive probabilistic model with a data-driven uncertainty structure. The effectiveness of the ML method is assessed by predicting PIDs of L-shaped laminates made from Toray's T800S/3900-2B aerospace-grade material system. The strategies presented offer an alternative, cost-efficient, and broadly applicable framework for solving manufacturing problems of composites using multi-fidelity simulation and limited experimental data.

2. Material and Methods

2.1. Process Specifications

The composite material utilized in this work was Toray T800S/3900-2B unidirectional (UD) prepreg with a resin content of 35.5% by weight, a common primary structural material in major aircraft such as the Boeing 787 [15,16]. Figure 2 presents the geometry and terminology used for composite parts in this study. Before processing, each part was laid-up in an L-shaped configuration with eight T800S/3900-2B plies, a flange length of 154.2 mm, a width of 50.8 mm, a corner radius of 15.875 mm, and a corner angle of 90°. All parts in this study were processed according to the Manufacturer's Recommended Cure Cycle (MRCC), consisting of heating to 180 °C at 2 °C/min, holding at 180 °C for 120

minutes, then cooling to room temperature at 2.78 °C/min. Curing was performed under a combined autoclave and vacuum pressure of approximately 0.7 MPa.

After curing and demolding, L-shaped parts may deform into various configurations with diverse spring-in and warpage magnitudes and directions [17]. Figure 2 provides two examples of cured and deformed L-shaped parts to serve as reference throughout this paper. As illustrated, positive spring-in values represent angle enlargements between flanges, while negative values signify angle enclosures. Likewise, positive and negative warpages represent concave-down and concave-up flange distortions, respectively. The maximum warpage (w_{max}) for each part corresponds to the greatest absolute out-of-plane distortion along the flange. In this study, the 1 and 0° directions denote the longitudinal fiber orientation, 2 and 90° signify the transverse fiber orientation, and 3 pertains to the out-of-plane fiber dimension. Lastly, directions x, y, and z denote the longitudinal, transverse, and out-of-plane L-shaped laminate dimensions.

2.2. Low-fidelity Simulation

In this section, we outline the low-fidelity numerical model utilized for predicting PIDs, drawing on an analytical strategy from Takagaki et al. [7]. The model uses the in-situ shear and bending moduli of a laminate, free strains during cure, and geometric parameters of L-shaped parts as inputs to estimate incremental stresses and deformations during processing. Input properties were derived by simulating the degree of cure (DoC) and glass transition temperature (T_g) throughout the MRCC using a well-established cure kinetics model [18], then computing thermo-mechanical properties stemming from bi-material beam (BMB) tests [14]. The Cure Hardening Instantaneously Linear Elastic (CHILE) assumption was then applied to predict the final corner spring-in, tip spring-in, and maximum warpage

of the L-shaped parts. The entire curing and deformation analysis was executed using a custom Python [19] code. An overview of the thermo-mechanical equations and algorithms implemented in the code can be referenced in Appendix B.

2.3. High-fidelity Simulation

This section presents the high-fidelity simulation approach utilized for predicting PIDs. Since the method is well-established and validated, only a brief overview is provided, while a more detailed description can be found in [20]. This high-fidelity scheme utilizes a refined 1D kinematic model rooted in the Carrera Unified Formulation (CUF) [21,22] and CHILE assumptions [13], for efficient layer-wise (LW) modeling and accurate 3D representations of residual stresses and PIDs. The model utilizes bending and shear moduli (E and G), Poisson's ratio (ν), coefficient of thermal expansion (α), and cure shrinkage-induced strain ($\Delta\varepsilon_{cs}$) in each of the three principal directions, sourced from published literature [15,18,20,23] and a BMB validation test. After each high-fidelity simulation, geometric analyses were performed to derive corner spring-in, tip spring-in, and maximum warpage at the center and edge of each laminate.

2.4. Experimentation

This section outlines procedures used to generate experimental PID data. The process involved fabricating L-shaped parts using an autoclave, then quantifying PIDs using laser profilometry. Each experimental round included the fabrication of three identical parts to integrate the effects of material and processing uncertainties into the prediction scheme. A 6.35 mm-thick A-36 steel tool covered with one layer of fluorinated ethylene propylene (FEP) release film was used as the layup mold. Three parts with dimensions specified in

Figure 2 were laid up evenly across the tool's width, vacuum sealed using standard bagging procedures, placed in an autoclave, and subjected to the MRCC.

After each autoclave cycle, the cured L-shaped parts were demolded and 2D spatial profiles were obtained at three locations using a Keyence LJ-X8400 laser scanner. One profile was scanned just inside each edge, and another was captured at the center of each part. Schematic representations of the layup and laser scanning procedures can be referenced in [17]. Following the scans, profiles were overlaid onto plots featuring the premeasured tool profile at the respective location. A custom Python code was then used to extract spring-in and warpage values at the center and edges of each part.

2.5. TGML Prediction Methodology

In this section, we present a novel methodology for efficient analysis of composites manufacturing phenomena using multi-fidelity simulation data, experimental data, and theory-guided machine learning (TGML) [6,17,24–27]. This paper focuses on introducing the method through a relatively simple case study of predicting PIDs for L-shaped parts with eight-ply layups comprised solely of zero- and ninety-degree plies, cured according to the MRCC. The outlined procedures are designed to be general, allowing for potential expansion to include other manufacturing methods and variables.

The prediction process (Figure 3) starts with generating data using the fastest available data source, recognizing that this data may contain errors due to the source's speed-enhancing numerical simplifications (Appendix B). Focusing only on eight-ply cross-ply layups in this case study, the low-fidelity simulation scheme (Section 2.2) was employed for all $2^8 = 256$ predictions of corner spring-in, tip spring-in, and maximum warpage. Throughout the low-fidelity procedures, all nine elements of the extensional (A), coupling (B), and

bending (D) stiffness matrices were also computed using Classical Laminated Plate Theory (CLPT) [28]. Thus, the low-fidelity prediction process yielded a high-dimensional dataset of 27 stiffness coefficients and three PID predictions (i.e., corner spring-in, tip spring-in, warpage) for each cross-ply laminate (Figure 3a). The entire low-fidelity data generation process was completed in approximately 24 minutes on a standard desktop. For more complex problems with larger datasets, the efficiency of this low-fidelity simulation phase may be improved by limiting simulations to a representative subset of the design space.

After constructing the low-fidelity dataset, initialization of predictive models for each PID type begins. To start this process, numerical values serving as parametric inputs for each lamination must be obtained. The primary objective of the parametrization phase is to establish optimal feature representations (i.e., input parameters) for modeling PIDs using the most cost-effective data source. Although the low-fidelity data may contain errors, it is assumed that the method's accuracy is adequate for identifying optimal parameters for subsequent modeling [7,13,18].

In this case study, laminates defined by similar input parameters should yield similar PID outputs to facilitate "smoothness" in the model's predictions. Using methods from prior studies [17], we aim to achieve this smoothness by directly integrating stiffness coefficients obtained during low-fidelity simulation procedures as input parameters. Since previous research has established connections between CLPT and PIDs of composites [7–9,28], utilizing stiffness coefficients as parametric inputs serves to provide physics-based "guidance" to the domain and enhance the model's accuracy. This parametrization approach forms the basis of our "theory-guided" machine learning methodology [6,17,24,27].

Training a model on a highly parametrized design space with a large input-to-output ratio is prone to overfitting, computationally demanding, and challenging in terms of interpretation. To address these constraints, input parameters must be refined to an optimal number. This task was accomplished using a strategy inspired by [29,30] and termed Focused Bayesian Information Criterion (FBIC). To balance accuracy and interpretability of the model, we first employ the Bayesian Information Criterion (BIC) [29], a well-established parameter selection method. BIC assesses the appropriateness of input subsets by maximizing the likelihood function on the training data and penalizing the number of parameters to prevent overfitting. However, in our simple cross-ply design space, numerous input subsets would yield identical BIC values due to similarities among layups and stiffness coefficients. To overcome this challenge, we introduce an additional term to the BIC, prioritizing predictive efficiency (i.e., computational time) and creating a custom parameter selection method, the FBIC:

$$FBIC = -2\ln(\hat{L}) + kt\ln(n) \tag{1}$$

where $\hat{L}$ is the model's maximized likelihood function (~ accuracy on training data), k is the number of parameters, n is the number of datapoints, t is the computational time needed for model training, and models with lower FBIC values are preferred.

Applying (1), training was conducted on all potential combinations of two-input parameters (e.g., A_{11} and B_{22}) to predict low-fidelity simulated PIDs. Enforcing the constraint $k = 3$ throughout this process limited the evaluation to two-input subsets, allowing the models to be visualizable and interpretable as 3D surfaces. Gaussian Process Regression (GPR), a probabilistic machine learning (ML) technique, was employed for model fitting [31]. During GPR training, an underlying function is predicted as the mean of a distribution over functions

that is dependent on the values of inputs, outputs, and Gaussian measurement noise (i.e., uncertainty). This initial GPR training phase was completed using a summation of the Radial Basis Function (RBF) and white noise kernels in Python's scikit-learn library [19,32].

With 27 potential input parameters, there were 351 potential subset combinations for each PID output, all of which GPR modeled in approximately four minutes. Afterward, employing (1), the FBIC was computed for each model. The two-input GPR model yielding the lowest FBIC was considered optimal (Figure 3b,c), leading to the exclusion of all other contenders. It is critical to note that input parameter subsets were not equally optimal for all three PID types and thus the GPR initialization process was repeated for each.

After constructing GPR models, a calibration process begins by iteratively replacing low-fidelity simulation data with high-fidelity simulation and experimental data (Figure 3d). The primary assumption underlying this phase is that while simulation data may enhance the performance of a GPR model, certain simulation data could be unhelpful or even detrimental due to inaccuracies. Conversely, experimental data will exclusively enhance GPR performance as they represent the truth. To incorporate these assumptions into the calibration process, noise levels were assigned to different data types in the model based on the proximity of low- and high-fidelity simulation points to the experimental data. This methodology enables the GPR model to consider data from various sources differently, placing greater "trust" in and near experimental points. Conversely, in regions lacking experiments, the model exhibits heighted uncertainty. This establishes a GPR model with a spatially weighted uncertainty structure, termed Spatially Weighted Gaussian Process Regression (SWGPR). Another perspective on SWGPR is to consider it akin to instance-transfer learning (TL),

where data weights are adjusted to mitigate the impact of "bad" virtual data while promoting the contribution of "good" virtual data in the predictive model [33].

To illustrate the SWGPR approach, consider a simple space with data from one experiment and one lower-fidelity virtual source, where the outputs (e.g., PIDs) are assumed to be a function of some arbitrary parameter (Figure 4a). Similar to assuming that PID magnitudes are alike for nearby inputs, a smoothness in noise or uncertainty can also be assumed. To incorporate this, we declare that as the distance between a lower-fidelity and experimental datapoint increases, the influence or weight of the low-fidelity point on the model's predictions decreases following a Gaussian distance-decay function:

$$w_{ij} = e^{-d_{ij}^2/2(h\times d_{ij,max})^2} \tag{2}$$

where w_{ij} is the weight of each lower-fidelity datapoint and d_{ij} is the Euclidean distance between a lower-fidelity and experimental datapoint, included in the numerator to relate spatial separation to Gaussian-like decay. The parameter h is a decay factor influencing the rate of weight decay and $d_{ij,max}$ is the maximum distance between two datapoints in the domain, included in the denominator to minimize weights at this juncture. The inclusion of 2 in the function's denominator moderates weight decays, preventing overly rapid reductions and minimizing overlaps between different data fidelities across the design space.

Figure 4a illustrates the Gaussian distance-decay function for various h values, where the magnitude is a direct reflection of the uncertainty in a data source. If there is high uncertainty in modeling, material properties, or other variables, this knowledge can be implemented by specifying a smaller h value. One approach for selecting h is employing an objective technique such as Silverman's rule of thumb [34] to estimate a bandwidth for the data, then adjusting it based on the perceived uncertainty relative to other source fidelities.

Once decay rates and weights are determined, a distance-weighted sum is calculated when multiple experimental points are present. These weights are then directly translated into noise levels using the following equation:

$$\alpha_{ij} = \ln(1/w_{ij}) \tag{3}$$

where α_{ij} is considered the variance of additional Gaussian noise surrounding a datapoint. Figure 4b depicts normal distributions corresponding to Gaussian noises for various decay factors and interpoint distances. After establishing α_{ij} for each lower-fidelity point, these values are incorporated into the GPR framework, altering its prediction scheme to:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x') + \alpha(d)I) \tag{4}$$

where $f(x)$ is a function to be predicted, $\mathcal{GP}$ is the Gaussian Process, $m(x)$ is the mean function, $k(x, x')$ is the kernel function describing covariance between points x and x', $\alpha(d)$ is an array of distance-dependent noise levels for each datapoint, and I is the identity matrix.

During fitting, α values are added to the kernel matrix's diagonal, introducing point-specific noise to each lower-fidelity datapoint based on its distance to experimental data. This approach systematically establishes a multi-scale uncertainty structure in the model. Figure 4c illustrates SWGPR predictions for various decay factors, where dashed lines indicate its mean response, shaded areas represent 95% confidence bounds, and global covariance is modeled as the summed RBF and white noise kernels. For one extreme (h = 0.1), there is high uncertainty in lower-fidelity data, resulting in strong reliance and tight convergence around the experiment. Conversely, when h = 1.0, uncertainties are similar among fidelities and SWGPR predicts closely with the traditional GPR without $\alpha(d)$ terms.

Building on prior discussions, a three-fidelity TGML prediction scheme requires specifying h values for each virtual data source. To achieve this, we assume that uncertainties

in simulation datapoints primarily stem from material properties referenced during modeling. In simpler terms, inaccuracies in virtual data are hypothesized to be mostly attributed to numerical inputs rather than modeling capabilities, given the challenging nature of composites characterization and using publicly-sourced properties. Consequently, we set h equal to the inverse of the number of material properties used as inputs: $h = 1/3 = 0.33$ for low-fidelity and $h = 1/15 = 0.07$ for high-fidelity simulation. It is crucial to note that a smaller h value for high-fidelity simulation does not imply lower accuracy but acknowledges a larger margin for error due to a higher number of inputs.

After establishing h for each virtual data source, the TGML prediction method enters its final stage (Figure 3d). At this point in the case study, a GPR model trained on 256 low-fidelity simulations is available for each PID type. The last step involves calibrating these models by conducting targeted high-fidelity simulations or experiments until a specified accuracy is achieved. The overarching goal is to attain desired accuracy with the minimum number of experiments and high-fidelity simulations to keep the model cost-effective.

To initiate calibration, one experiment is performed with parameters exhibiting the highest uncertainty in the GPR model. The experimental results replace the low-fidelity datapoint with the same input parameters, and other virtual points are weighed per (2) and (3), assigning each low-fidelity point a unique weight based on its distance to the experimental data. The SWGPR model is then retrained to predict PIDs and identify a new location with the highest distance-weighted uncertainty. Next, the parameters for which SWGPR is most uncertain are fed into the high-fidelity simulation scheme, PIDs are predicted, and the model is retrained. If the addition of high-fidelity simulation data improves accuracy, the process repeats with the high-fidelity simulation scheme. To fully optimize this

process for maximum efficiency, an automatic loop should be implemented, cycling through target identification, data generation, and model updating. However, if accuracy remains unchanged or decreases throughout a cycle at any point in the loop, the process must halt while another experiment is conducted and added to the model. The automatic high-fidelity simulation loop may then be reinstated until its next accuracy plateau or until the model meets accuracy requirements. Throughout the iterative calibration, as the low-dimensional input domain remains consistent despite the addition of multi-fidelity data, another form of TL emerges from the SWGPR method, specifically a feature-representation transfer [33].

In this study, SWGPR accuracy is evaluated at each calibration step by comparing its predictions to experimental PID values obtained from six testing laminates listed in Table 1. These laminates included a wide range of various properties, including UD, symmetric, and asymmetric layups, with diverse in-plane/out-of-plane coupling coefficients, bending stiffnesses, and zero- to ninety-degree ply ratios. The diversity of the testing dataset ensured comprehensive evaluation of the SWGPR models. Therefore, if the models meet accuracy requirements for the testing dataset, it is expected to generalize well to other laminates within the design space. Accuracy requirements in this study stipulate that the SWGPR model must predict PIDs within the standard deviation of the six testing laminates, which were fabricated following procedures outlined in Section 2.4.

3. Results and Discussion

In this section, we initially illustrate the training of an SWGPR model as we aim to predict the tip spring-in at the center of six L-shaped composite parts with layups and input parameters listed in Table 1. Then, we extend the SWGPR approach to predict post-curing deformed L-shapes. The L-shape prediction involved predicting all PID types (corner spring-

in, tip spring-in, and warpage at the center and edge of each part), then generating interpolated profiles from these values.

Figure 5 displays an SWGPR model and its predictions for tip spring-in of L-shaped laminates (a) before and (b) after undergoing calibration. The model's progressive evolution at various phases can also be referenced in Video 1. Red points in the left-side plots represent low-fidelity simulations, blue squares are high-fidelity simulations, and black stars are experiments used for model training. Multi-colored surfaces depict the SWGPR models' mean predictions, while grey surfaces are 95% confidence bounds. Tip spring-in is plotted against B_{11} and D_{22}, identified as optimal parameters using the FBIC. In Video 1, high-fidelity simulations are incrementally added in multiples of twenty for clarity. The right side of the graphics display tip spring-in predictions for the six testing laminates, with black outlined bars and error bars representing average experimental values and standard deviations, respectively.

Initially, the SWGPR model trained on 256 low-fidelity simulations without spatial weighing exhibited a large root mean squared error (RMSE) of 27.7° in predicting tip spring-in for the six testing laminates. Upon adding the first experiment, the model's mean response and confidence bounds converged around the experiment and loosely followed low-fidelity data with additional Gaussian noises. Integrating high-fidelity simulation data then led to an evolving structure with increasing accuracy until approximately 40 simulations were added and the RMSE plateaued at 4.4°. Another experiment was then required near the model's bounds, suggesting inaccuracies in the high-fidelity simulation scheme and the need for multiple experiments in the domain. After calibration, the SWGPR model incorporated 114 low-fidelity simulations, 138 high-fidelity simulations, 4 experiments, achieved a minimal

RMSE of 0.1°, and predicted within the standard deviation for all six testing laminates. Most notably, the calibrated SWGPR structure was dominated by virtual simulation data and only required a minimal number of experiments for accuracy convergence. Overall, this illustrates that the SWGPR approach can effectively utilize virtual and real-world datatypes for accurate and cost-efficient analysis of composites manufacturing phenomena.

Figure 6 shows average experimental profiles for the cured L-shaped parts compared with SWGPR predictions after the calibration process. As previously mentioned, experimental profiles were generated through laser scanning the L-shaped coupons, while SWGPR profiles were derived by repeating the calibration process (Video 1) for all PID types and locations. Throughout the SWGPR calibrations, optimal input parameters were determined as B_{22} and D_{22} for corner spring-in and B_{11} and D_{11} for warpage, indicating the strong influence of coupling and bending on all PID types and an interconnectedness. All PID types necessitated four experiments to meet accuracy specifications, with minimal variation in the number of high-fidelity simulations. Overall, Figure 6 demonstrates that similar to the convergence observed in tip spring-in predictions, the SWGPR method achieves highly accurate predictions of the L-shaped parts with minimal material characterization and experimental efforts, showing promise in better understanding PIDs and other facets of composites manufacturing.

4. Summary and Conclusions

This study introduced an innovative approach for accurate and efficient analysis of composites manufacturing by combining multi-fidelity simulations, experimentation, and theory-guided machine learning (TGML). The method initiates with a low-fidelity simulation scheme to rapidly predict processing outcomes in a defined design space. These solutions are

then mapped to a reduced-order domain using closed-form physical theory and Gaussian Process Regression (GPR). The theory-guided GPR model then undergoes iterative calibrations using high-fidelity simulation and targeted experimental results. During calibration, numerical datapoints are assigned point-specific noise levels using a novel Gaussian distance-decay weighing mechanism, creating a Spatially Weighted Gaussian Process Regression (SWGPR) model. The SWGPR approach accurately predicts spring-in and final deformed shapes of cured L-shaped laminates after integrating just four experiments and bypassing complex material characterization. The strategies presented hold promise for further investigation, understanding, and improvement of composites manufacturing.

CRediT Authorship Contribution Statement

Caleb Schoenholz: Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Enrico Zappino:** Formal analysis, Investigation, Data curation. **Marco Petrolo:** Conceptualization, Methodology, Supervision. **Navid Zobeiry:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Toray Composite Materials America, Inc. and by the project "An AI-assisted virtual manufacturing approach to mitigate defects in advanced composites" funded by the Italian Minister of Foreign Affairs and International Cooperation (MAECI) [grant number US23GR12].

Appendix A. Supplementary Data

S1. Numerical and experimental data produced in this study.

Appendix B. Equations and Algorithms

S2. Overview of the "low-fidelity simulation" code employed for predicting process-induced deformations (PIDs) in L-shaped composite parts.

References

[1] Zobeiry N, Poursartip A. The origins of residual stress and its evaluation in composite materials. Structural Integrity and Durability of Advanced Composites: Innovative Modelling Methods and Intelligent Design, Woodhead Publishing; 2015, p. 43–72. https://doi.org/10.1016/B978-0-08-100137-0.00003-1.

[2] Zobeiry N, Forghani A, Li C, Gordnian K, Thorpe R, Vaziri R, et al. Multiscale characterization and representation of composite materials during processing. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2016;374. https://doi.org/10.1098/RSTA.2015.0278.

[3] Fernlund G, Mobuchon C, Zobeiry N. 2.3 Autoclave Processing. Comprehensive Composite Materials II, Elsevier; 2018, p. 42–62. https://doi.org/10.1016/B978-0-12-803581-8.09899-4.

[4] Albert C, Fernlund G. Spring-in and warpage of angled composite laminates. Compos Sci Technol 2002;62:1895–912. https://doi.org/10.1016/S0266-3538(02)00105-7.

[5] Manohar K, Hogan T, Buttrick J, Banerjee AG, Kutz JN, Brunton SL. Predicting shim gaps in aircraft assembly with machine learning and sparse sensing. J Manuf Syst 2018;48:87–95. https://doi.org/10.1016/J.JMSY.2018.01.011.

[6] Zobeiry N, Poursartip A. Theory-Guided Machine Learning for Process Simulation of Advanced Composites 2021.

[7] Takagaki K, Minakuchi S, Takeda N. Process-induced strain and distortion in curved composites. Part I: Development of fiber-optic strain monitoring technique and analytical methods. Compos Part A Appl Sci Manuf 2017;103:236–51. https://doi.org/10.1016/J.COMPOSITESA.2017.09.020.

[8] Arafath ARA, Vaziri R, Poursartip A. Closed-form solution for process-induced stresses and deformation of a composite part cured on a solid tool: Part II – Curved geometries. Compos Part A Appl Sci Manuf 2009;40:1545–57. https://doi.org/10.1016/J.COMPOSITESA.2009.01.009.

[9] Wisnom MR, Potter KD, Ersoy N. Shear-lag Analysis of the Effect of Thickness on Spring-in of Curved Composites. J Compos Mater 2006;41:1311–24. https://doi.org/10.1177/0021998306068072.

[10] Schoenholz C, Zobeiry N. Investigating the impacts of processing variability on tool-part interaction for interply-toughened aerospace composites using a novel shear technique. Compos Part A Appl Sci Manuf 2024;178:107973. https://doi.org/10.1016/J.COMPOSITESA.2023.107973.

[11] Li D, Li X, Dai J. Process modelling of curing process-induced internal stress and deformation of composite laminate structure with elastic and viscoelastic models.

Applied Composite Materials 2017;25:527–44. https://doi.org/10.1007/S10443-017-9633-5/FIGURES/14.
[12] Ding A, Li S, Wang J, Zu L. A three-dimensional thermo-viscoelastic analysis of process-induced residual stress in composite laminates. Compos Struct 2015;129:60–9. https://doi.org/10.1016/J.COMPSTRUCT.2015.03.034.
[13] Johnston AA, Vaziri R, Profile S, Poursartip A. A Plane Strain Model for Process-Induced Deformation of Laminated Composite Structures. J Compos Mater 2001;35:1435–69. https://doi.org/10.1106/YXEA-5MH9-76J5-BACK.
[14] Thorpe RJ. Experimental characterization of the viscoelastic behavior of a curing epoxy matrix composite from pre-gelation to full cure. Master of Applied Science. University of British Columbia, 2013. https://doi.org/10.14288/1.0073820.
[15] 3900 Prepreg System | Toray Composite Materials America, Inc. 2020.
[16] Odagiri N, Kishi H, Yamashita M. Development of TORAYCA prepreg P2302 carbon fiber reinforced plastic for aircraft primary structural materials. Advanced Composite Materials 1996;5:249–54. https://doi.org/10.1163/156855196X00301.
[17] Schoenholz C, Zobeiry N. An Accelerated Process Optimization Method to Minimize Deformations in Composites Using Theory-guided Probabilistic Machine Learning. Compos Part A Appl Sci Manuf 2024;176:107842. https://doi.org/10.1016/J.COMPOSITESA.2023.107842.
[18] Dykeman D. Minimizing uncertainty in cure modeling for composites manufacturing. Doctor of Philosophy. University of British Columbia, 2008. https://doi.org/10.14288/1.0066334.
[19] Van Rossum G, Drake Jr F. Python. Version 3.10. 2021.
[20] Zappino E, Zobeiry N, Petrolo M, Vaziri R, Carrera E, Poursartip A. Analysis of process-induced deformations and residual stresses in curved composite parts considering transverse shear stress and thickness stretching. Compos Struct 2020;241:112057. https://doi.org/10.1016/J.COMPSTRUCT.2020.112057.
[21] Carrera E, Cinefra M, Zappino E, Petrolo M. Finite Element Analysis of Structures through Unified Formulation. vol. 9781119941217. Wiley Blackwell; 2014. https://doi.org/10.1002/9781118536643.
[22] Scano D, Carrera E, Petrolo M. Use of the 3D Equilibrium Equations in the Free-Edge Analyses for Laminated Structures with the Variable Kinematics Approach. Aerotecnica Missili & Spazio 2023 2023;1:1–17. https://doi.org/10.1007/S42496-023-00177-2.
[23] Chen C, Poursartip A, Fernlund G. A novel method to measure laminate shear modulus development of interlayer toughened composite laminates during the curing process. Proceedings of the American Society for Composites - 34th Technical Conference, ASC 2019 2019. https://doi.org/10.12783/ASC34/31316.
[24] Zobeiry N, Reiner J, Vaziri R. Theory-guided machine learning for damage characterization of composites. Compos Struct 2020;246:112407. https://doi.org/10.1016/J.COMPSTRUCT.2020.112407.
[25] Wagner N, Rondinelli JM. Theory-guided machine learning in materials science. Front Mater 2016;3:203425. https://doi.org/10.3389/FMATS.2016.00028/BIBTEX.
[26] Karpatne A, Atluri G, Faghmous JH, Steinbach M, Banerjee A, Ganguly A, et al. Theory-guided data science: A new paradigm for scientific discovery from data. IEEE

Trans Knowl Data Eng 2017;29:2318–31. https://doi.org/10.1109/TKDE.2017.2720168.
[27] Zobeiry N, Profile S, Poursartip A. Theory-Guided Machine Learning Composites Processing Modelling for Manufacturability Assessment in Preliminary Design, Quebec City, Canada: NAFEMS 17th World Congress; 2019.
[28] Kassapoglou C. 3. Review of Classical Laminated Plate Theory. Design and Analysis of Composite Structures: With Applications to Aerospace Structures. Second Edition, West Sussex, United Kingdom: John Wiley & Sons, Ltd.; 2013, p. 33–53. https://doi.org/https://doi.org/10.1002/9781118536933.ch3.
[29] Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics 1978;6:461–4.
[30] Claeskens G, Hjort NL. Model selection and model averaging. Cambridge University Press; 2008. https://doi.org/10.1017/CBO9780511790485.
[31] Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. The MIT Press; 2006.
[32] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825–30.
[33] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22:1345–59. https://doi.org/10.1109/TKDE.2009.191.
[34] Silverman BW. Density Estimation for Statistics and Data Analysis. New York: CRC Press; 1998. https://doi.org/10.1201/9781315140919/DENSITY-ESTIMATION-STATISTICS-DATA-ANALYSIS-BERNARD-SILVERMAN.

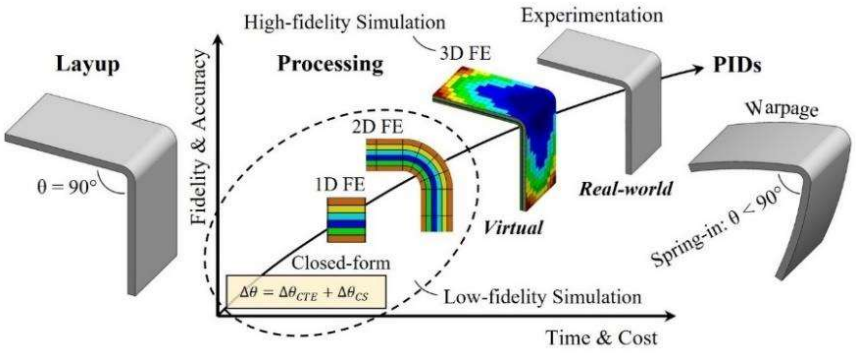


Figure 1. Schematics of process-induced deformations (PIDs) in an L-shaped composite part and the trade-off between time/cost and fidelity/accuracy in analysis methods.

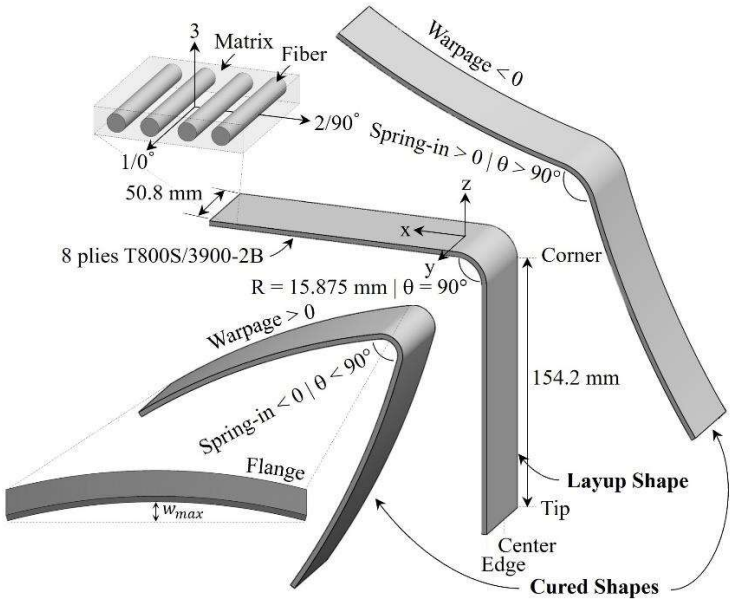


Figure 2. Geometry and terminology used to characterize L-shaped composite parts.

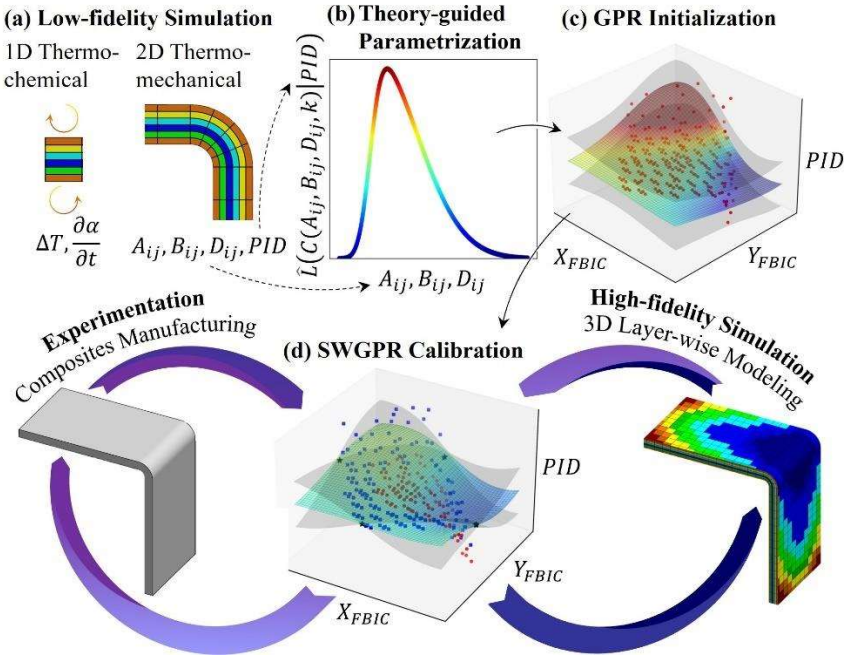


Figure 3. Flowchart of efficient composites manufacturing analysis method using multi-fidelity simulation and theory-guided machine learning.

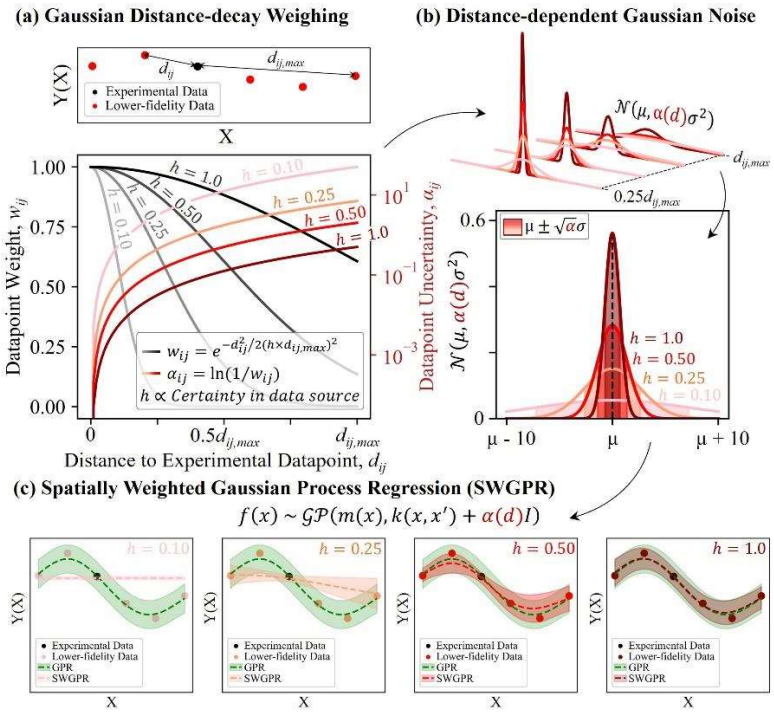


Figure 4. (a) Gaussian distance-decay weight and uncertainty, (b) Gaussian noise, and (c) SWGPR predictions for different uncertainty levels in a lower-fidelity data source.

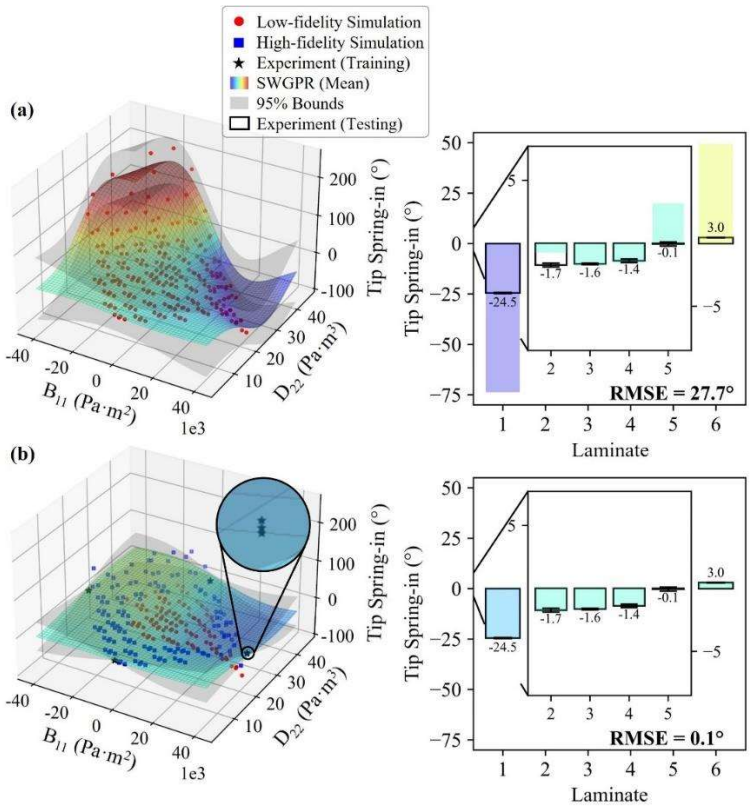


Figure 5. SWGPR model (a) before and (b) after calibration and its predictions of tip spring-in for L-shaped composite parts.

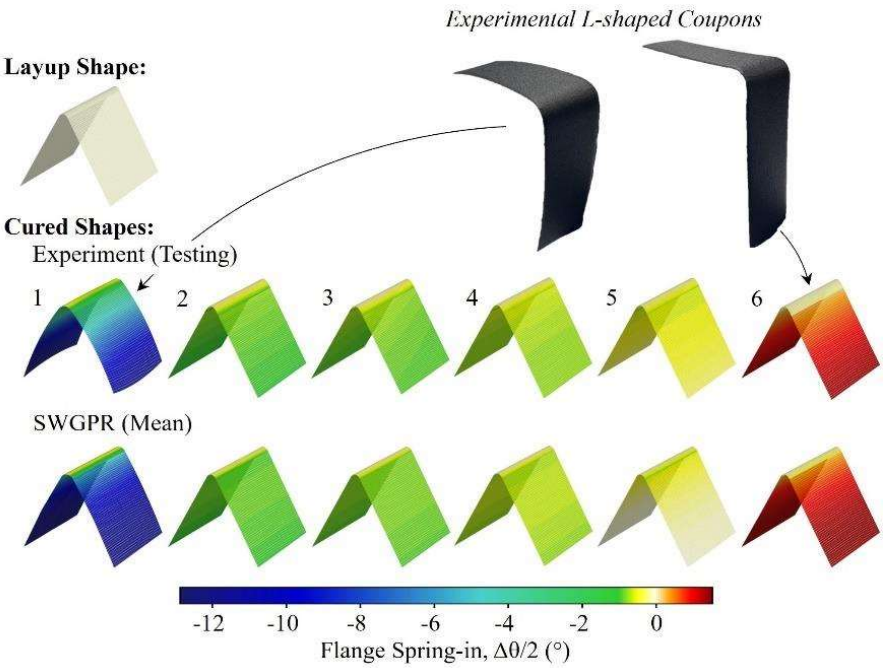


Figure 6. Comparison between experimental PIDs and predictions using calibrated SWGPR models.

Video 1. Progressive evolution of the SWGPR model and its predictions of tip spring-in throughout the calibration process.

Table 1. Layups and select properties of testing laminates employed for evaluating the accuracy of SWGPR models.

Laminate	**Layup**	**$B_{11} \times 10^{-3}$ (Pa·m^2)**	**D_{22} (Pa·m^3)**
1	[90/90/90/90/0/0/0/0]	40.05	23.23
2	[0/90/0/90/90/0/90/0]	0.00	15.63
3	[90/0/0/90/0/90/90/0]	0.00	23.24
4	[0/0/0/0/0/0/0/0]	0.00	2.95
5	[90/0/90/0/90/90/0/90]	-2.50	31.16
6	[0/0/0/0/0/0/90/90]	-30.04	20.70