Demonstration of the Data Fusion Labeler (dFL) for Additive Manufacturing

Sophelio LLC / dFL Project Team

Abstract

This document presents a detailed demonstration of the *Data Fusion Labeler (dFL)* platform applied to additive manufacturing (AM) process data. The example illustrates how multimodal, high-frequency layer data from a laser powder bed fusion (LPBF) system can be harmonized, visualized, labeled, and analyzed within dFL. The workflow serves as a generalizable template for similar data fusion efforts in manufacturing research, process optimization, and predictive analytics.

1 Introduction

This demonstration showcases the application of the *Data Fusion Labeler (dFL)* within an Additive Manufacturing (AM) context, leveraging the NIST Laser Powder Bed Fusion (LPBF) dataset [1]. The curated dFL workspace is pre-loaded with layer-level process data and custom visualization modules, allowing users to conduct exploratory analysis and automated labeling in a fully reproducible environment.

The demonstration is designed as a *template workflow* that researchers can extend for deeper studies in additive manufacturing—including model development, uncertainty quantification, and machine learning-based prediction.

2 Challenge and Role of dFL

Manufacturing engineers face a fundamental challenge: translating high-frequency, heterogeneous sensor data into actionable process intelligence. Conventional post-processing methods are manual and fragmented, often requiring extensive scripting to extract insights.

The dFL platform directly addresses this gap. It integrates data ingestion, visualization, harmonization, and autolabeling into a single loop, transforming raw data into structured knowledge in minutes. Its operator-aware transformation pipeline ensures consistent ordering of operations—Trim → Fill → Resample → Smooth → Normalize—thereby preserving physical interpretability and reproducibility across campaigns [2].

3 Dataset Description

The demonstration uses the NIST Additive Manufacturing Process Monitoring dataset [1], which provides fully registered in-situ and ex-situ data from an LPBF build of four cubic parts (5 mm × 5 mm × 9 mm, 250 layers each). The dataset includes:

- Commanded and measured machine control signals (laser position, power, scan speed);
- In-situ melt pool imaging metrics at thresholds 80, 100, and 120;

- Post-process X-ray Computed Tomography (XCT) voxel data for density evaluation.

Each record is labeled as `partXX_LYYYY`, representing one layer per part. This structure enables both intra- and inter-part analyses of stability, drift, and build quality.

4 Visual Analysis in dFL

The demonstration begins with several built-in and custom visualization modes implemented through the `additive_manufacturing_data_provider.py` module.

4.1 Machine Fidelity Diagnostics

The default grapher compares commanded versus real laser trajectories and power levels. The overlay of dual traces immediately reveals tracking performance and quantifies control loop fidelity.

4.2 Laser Path Map

The *Laser Path Map* visualizer projects the laser's 2D scan trajectory while coloring each point by a selected feature (e.g., melt pool area). It provides spatial awareness of scan strategy, motion irregularities, and potential thermal accumulation. This mode bridges motion control signals and thermal signatures into a single coherent visualization.

4.3 Layer Trend by Part

This visualization aggregates a selected feature (e.g., mean XCT voxel value) across layers for each of the four parts. The resulting trends reveal startup phases, steady-state behavior, and subtle layer-dependent drifts. In this demonstration, a minor dip near layer 50 indicates a possible transient in process thermal stability.

4.4 Layer Histogram Heatmap

The *Layer Histogram Heatmap* tracks the statistical distribution of any feature across build layers. Horizontal shifts indicate process drift; vertical bands imply stability; and multi-modal regions signal potential regime changes. This is particularly useful for monitoring gradual degradation, powder spreading issues, or equipment variability.

5 Introducing Autolabeling

Manual inspection provides intuition, but automated surveillance is essential for scale. dFL's *autolabeling* engine enables users to convert manual insights into reusable labeling logic. Each autolabeler is implemented as a small Python function, evaluated layer-by-layer or across an entire dataset.

Two example autolabelers are constructed:

1. **XCT Porosity Detector**: Flags low-density regions detected in voxel data as potential porosity.

2. **Power Deviation Detector**: Flags segments where commanded and real laser power diverge beyond a threshold—an early indicator of process instability or hardware malfunction.

6 Building with AI

Using an AI-assisted development environment (Cursor IDE), users can rapidly generate these autolabeling functions through natural-language prompting. Once created, the functions appear dynamically in the dFL interface under the Autolabeling panel.

This highlights dFL's capacity for human-AI collaboration: domain knowledge remains encoded in user-defined rules, while AI accelerates their construction and integration.

7 Demonstrating the Autolabeler

The **Power Deviation Detector** is executed on a single layer to highlight anomalies, which are then visually overlaid on the power time series. Batch execution across all layers (e.g., 250 layers $\times$ 4 parts) generates a comprehensive anomaly map.

These labeled datasets can be:

- Exported for quality assurance review,
- Shared with collaborators for reproducibility,
- Fed into machine learning models for defect prediction.

8 Conclusion and Outlook

The demonstration concludes with a complete end-to-end cycle: data ingestion, harmonized visualization, AI-assisted autolabeling, and automated anomaly detection. This workflow demonstrates how dFL acts not merely as a preprocessing utility, but as a core component in intelligent, self-improving process monitoring pipelines.

Future extensions of this workflow include:

- Predictive modeling of melt pool geometry and defect probability,
- Cross-correlation of in-situ sensor data with ex-situ quality metrics,
- Adaptive feedback control based on real-time labeled signals,
- Multi-material or multi-machine harmonization for cross-platform learning.

By treating this additive manufacturing example as a canonical case, researchers can replicate its structure to design harmonized pipelines for any high-throughput, multimodal dataset—whether in manufacturing, fusion energy, or other complex engineering systems.

Acknowledgements

This work integrates public datasets from the National Institute of Standards and Technology (NIST) and extends methods developed for fusion energy data harmonization under the Data Fusion Labeler (dFL) initiative.

References

[1] Zhuo Yang, Yan Lu, Brandon Lane, and Ho Yeung. A fully registered in-situ and ex-situ dataset for metal powder bed fusion additive manufacturing: Data processing, feature extraction, registration, and uncertainties, 2025.

[2] Craig Michoski, Matthew Waller, Brian Sammuli, Zeyu Li, Tapan Ganatma Nakkina, Raffi Nazikian, Sterling Smith, David Orozco, Dongyang Kuang, Martin Foltin, Erik Olofsson, Mike Fredrickson, Jerry Louis-Jeune, David R. Hatch, Todd A. Oliver, Mitchell Clark, and Steph-Yves Louis. The data fusion labeler (dfl): Challenges and solutions to data harmonization, labeling, and provenance in fusion energy, 2025.