Data Science Competition

## Semiconductor Process Performance Prediction

Competition Overview

- Objective:

  - Predict 49 measurement points on semiconductor wafers.

  - Use process data to improve efficiency and product quality.

- Key Focus Areas:

  - Real-world manufacturing data analysis.

  - Machine learning for process optimisation.

Extrapolated Wafer Profile - Run ID 00fa8670-f0d6-58bd-8b4a-fb70e84bca61
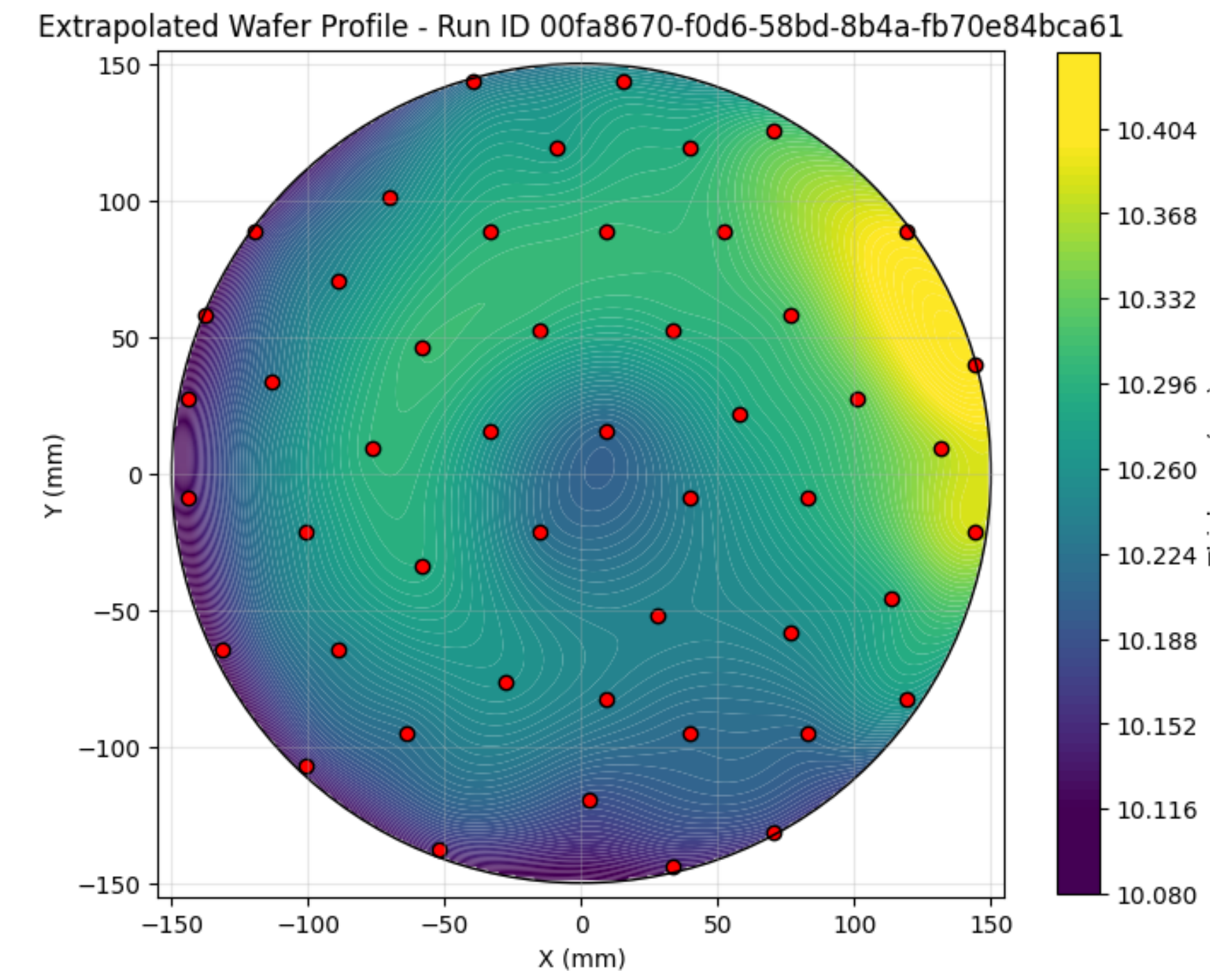
Dataset Structure

- Three Key Folders:

  - Train Data: Run Data + Incoming Run Data + Metrology Data (ground truth).

  - Test Data: Run Data + Incoming Run Data (no Metrology).

  - Submission: Metrology Data (Measurement column empty).

All files

ICQSR Data Challenge ⚙

🕘 Recents   ⭐ Starred

Name ↑

📁 submission
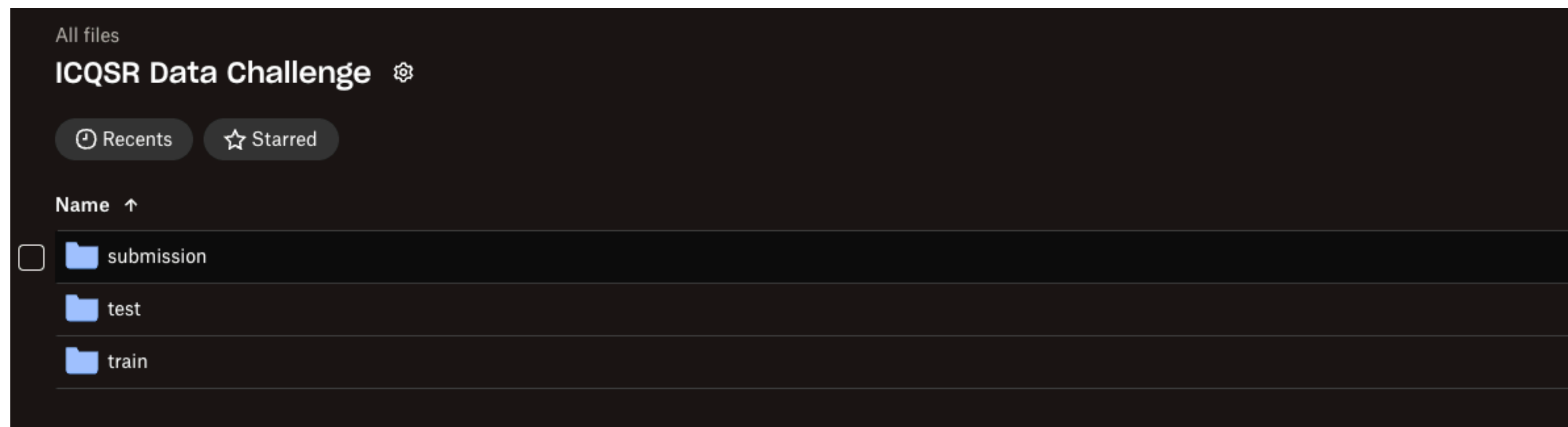
📁 test

📁 train

Data Description

Run Data - Overview

- Run Data (run_data_{file index}.parquet)

  - Represents: Current process sensor readings collected during the process step.

  - Location: Train & Test folders

  - Total Files: 20

  - Average Shape: (2,239,380 rows, 10 columns)

  - Use Case: Understanding real-time process conditions that impact measurements.
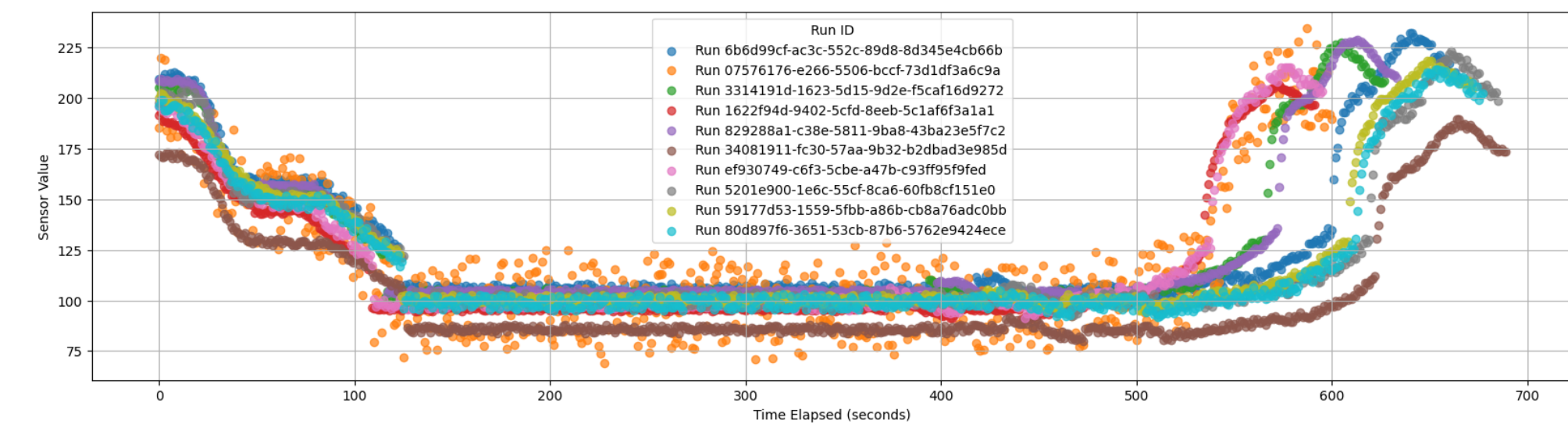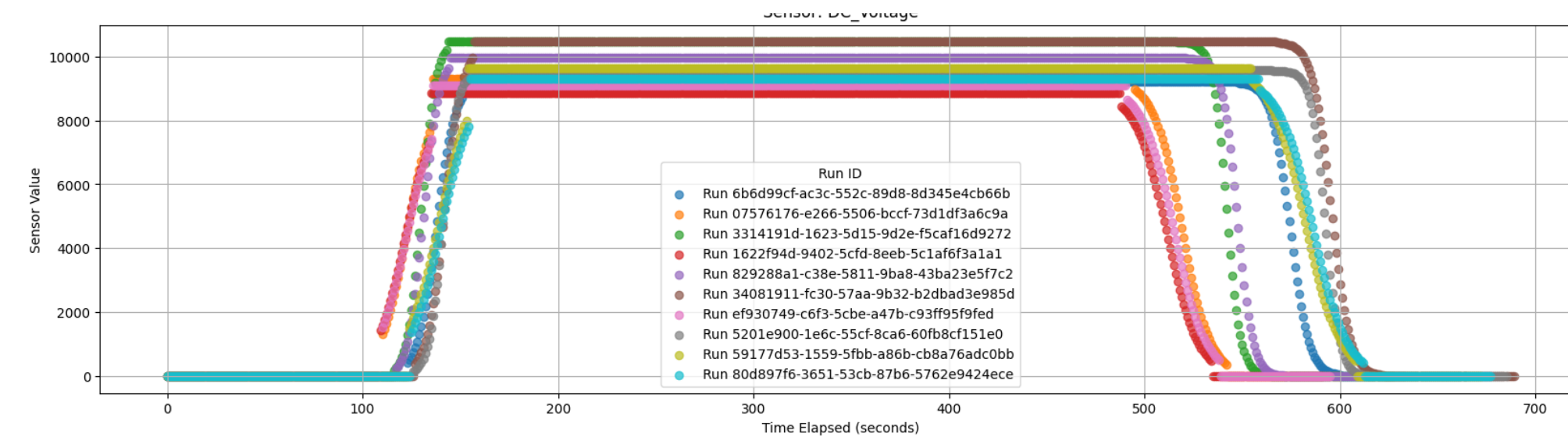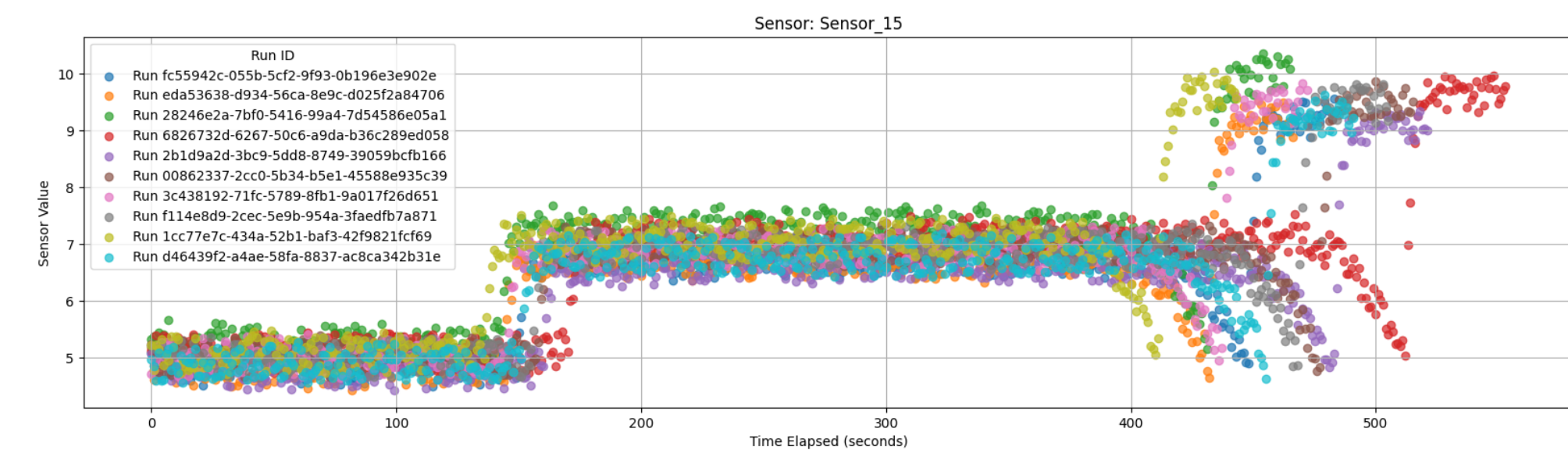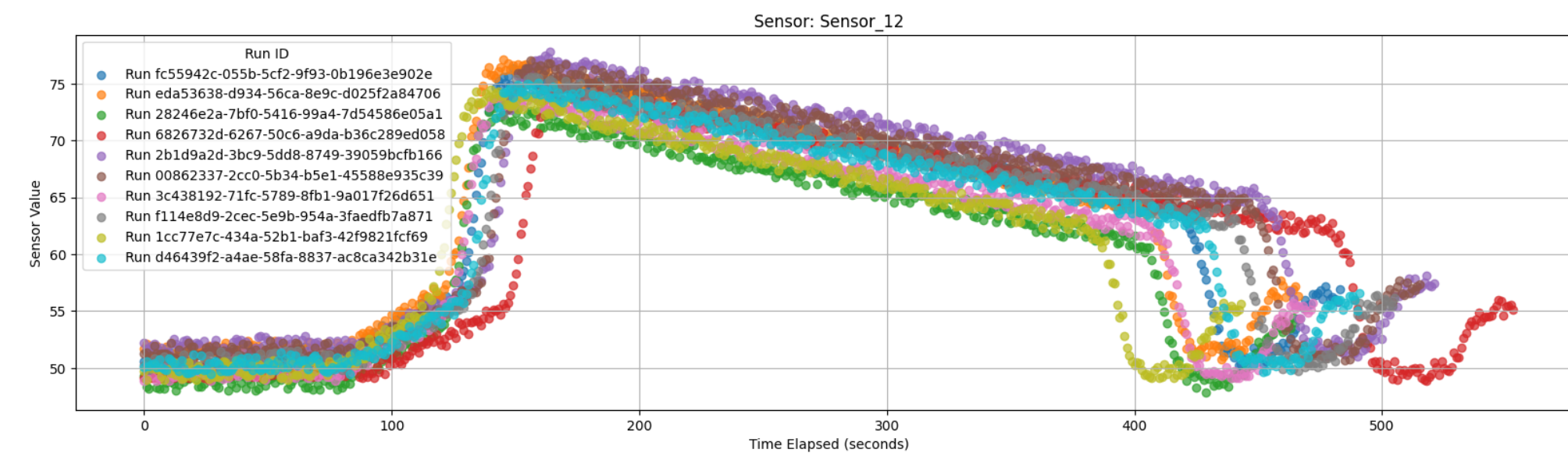
Data Description

Run Data - Key columns

- **Tool ID:** Identifies the specific tool used for the process.

- **Run Start Time / Run End Time:** Timestamps indicating when a particular process run began and ended.

- **Run ID:** A unique identifier for each process run.

- **Process Step:** A unique identifier for the specific step within the process.

- **Consumable Life:** A numerical value representing the remaining or used life of a consumable component.

- **Step ID:** The identifier for a specific processing step within the run.

- **Time Stamp:** The exact time a specific sensor measurement was recorded.

- **Sensor Name:** The name of the sensor collecting the measurement (e.g., "Sensor_a").

- **Sensor Value:** The numerical reading from the sensor.

Data Description

Incoming Run Data - Overview

- Incoming Run Data (incoming_run_data_{file index}.parquet)

  - Represents: Sensor readings from incoming process steps before the current process.

  - Location: Train & Test folders

  - Total Files: 20

  - Average Shape: (4,472,116 rows, 9 columns)

  - Similar to Run Data but does not include Consumable Life.

  - Use Case: Provides additional insights into the influence of prior processing steps on the current process.
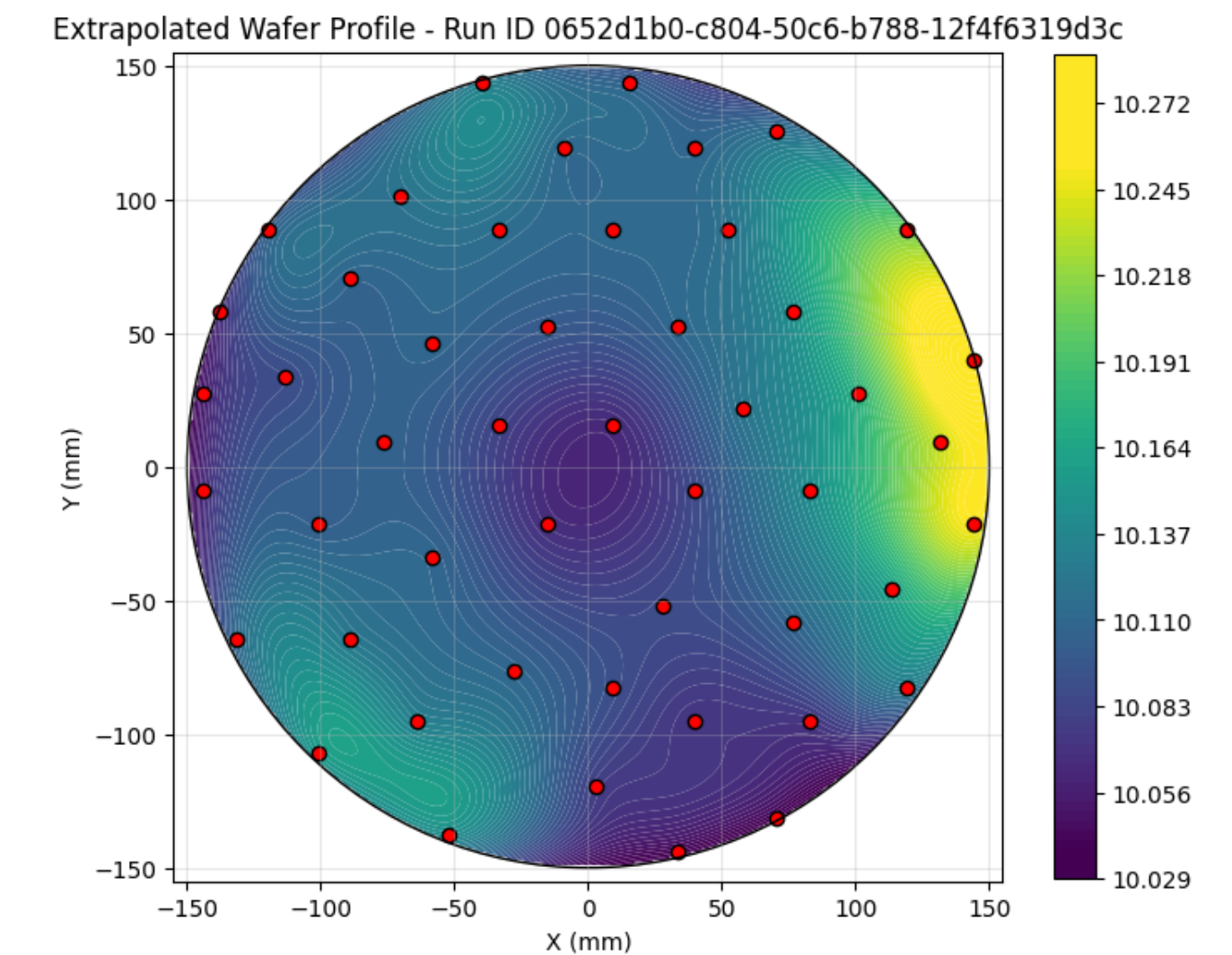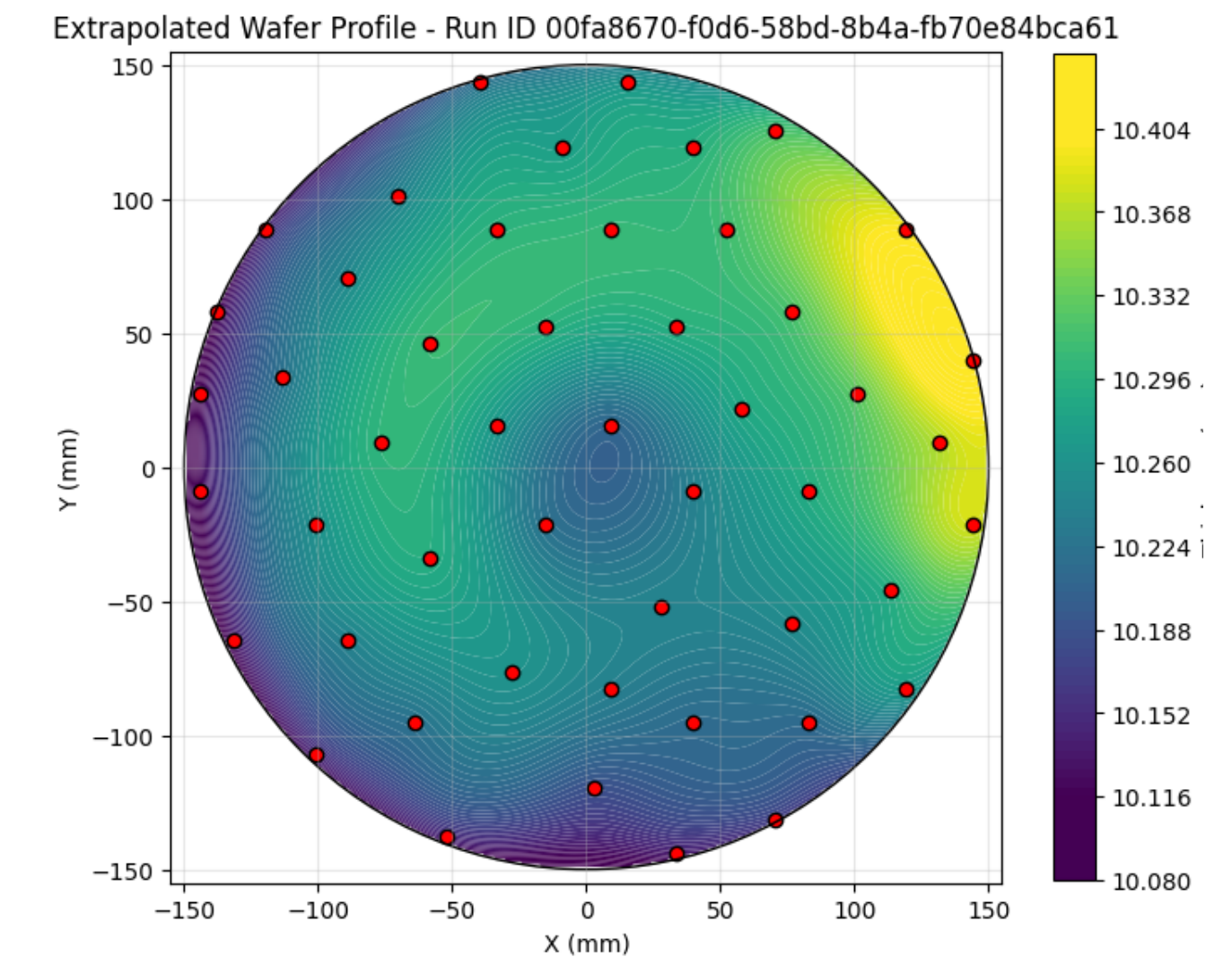
Data Description

Incoming Run Data - Key columns

- **Tool ID:** Identifies the specific equipment used for the process.

- **Run Start Time / Run End Time:** Timestamps marking the beginning and end of a process run.

- **Run ID:** A unique identifier for each process run.

- **Process Step:** Identifies the process step within the run.

- **Step ID:** A specific identifier for a step in the process.

- **Time Stamp:** The exact time when a sensor measurement was recorded.

- **Sensor Name:** The name of the sensor collecting the measurement.

- **Sensor Value:** The numerical reading from the sensor.

Data Description
Metrology Data - Overview

- Metrology Data (metrology_data{file index}.parquet)

  - Represents: Performance measurement of the process (actual performed profile after processing).

  - Location: Train & Submission folders

  - Total Files: 20

  - Average Shape: (11,025 rows, 9 columns)

  - Use Case: Provides ground truth for training and serves as the target for submission.

Extrapolated Wafer Profile - Run ID 00fa8670-f0d6-58bd-8b4a-fb70e84bca61

Extrapolated Wafer Profile - Run ID 0652d1b0-c804-50c6-b788-12f4f6319d3c

Data Description

Metrology Data - Key columns

- **Run ID:** A unique identifier for each process run.

- **Run Start Time / Run End Time:** Timestamps marking the beginning and end of a process **run.**

- **X_index / Y_index:** Grid indices representing measurement locations on the wafer.

- **X / Y:** The actual spatial coordinates of the measurement points.

- **Point Index:** An identifier for each measurement point.

- **Measurement:** The measured process performance at the specified location.

Competition Goal

Your Challenge: Predict Process Resultant Profile for All 49 Points

- Use Train data to build a model that predicts the full perfomance profile at all 49 measurement locations.

- Apply the trained model to Test data to generate predictions.

- Fill the predicted values into the Submission folder's Metrology Data, where the Measurement column is empty.

- Improve semiconductor manufacturing efficiency and yield.

Evaluation & Submission

- Model Performance Metrics:

  1. Point-wise Average RMSE (Root Mean Square Error)

  2. Additional Metrics: MAE, R-squared, etc.

- Submission Format:

  1. The Submission folder's Metrology Data must be filled with predicted Measurement values.

  2. Submit the updated Metrology Data file.

- Final Ranking: Based on prediction accuracy and robustness.

Get Started!

Ready to Compete?

- Download the dataset.

- Train your model using the Train folder.

- Generate predictions using the Test folder.

- Fill in the missing measurement values in the Submission folder.

- Stay tuned for updates and discussions.