Regression Analyses for Data Synchronization between In-process Monitoring Sensors in Laser Powder Bed Fusion Process

Nazmul Hasan (nh202@arizona.edu)

Systems and Industrial Engineering, University of Arizona

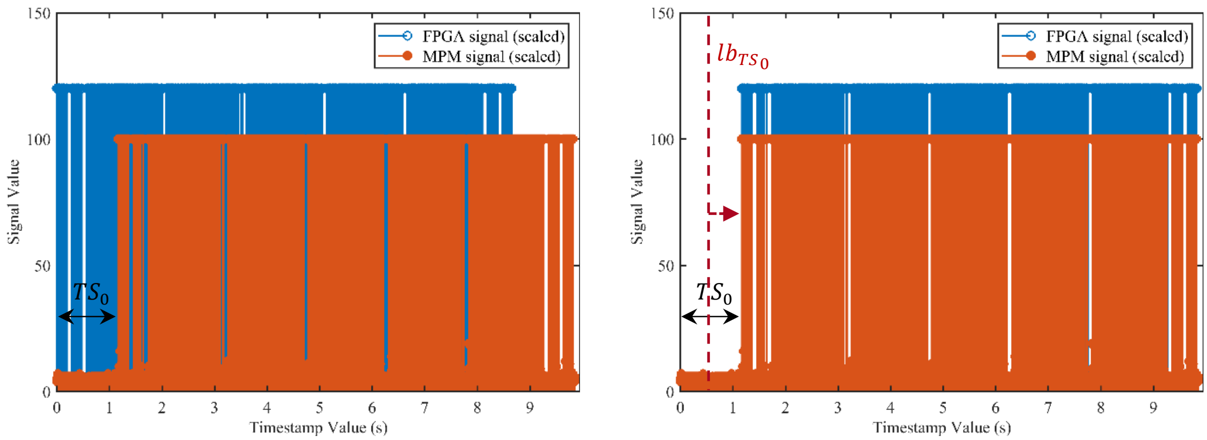
Executive Summary

This study aims to solve the data synchronization problem between two in-process monitoring sensors used in the Laser Powder Bed Fusion (LBPF) process: the Field-Programmable Gate Array (FPGA) sensor and the Melt Pool Monitoring (MPM) camera. The study investigates the statistical relationships among five variables: FPGA timestamp, initial time-shift, layer index, part index, and time-shift adjustment. The study uses regression analysis methods, such as multiple linear regression with interaction terms, LASSO-penalized model selection, and step-wise model selection. The study also employs bootstrapping for uncertainty quantification, as the normality assumption of the residuals is not valid for the fitted models. The main findings of the study are: (1) layer index and part index have statistically significant main and interaction effects on initial time-shift, meaning that initial time-shift varies from layer to layer and part to part; (2) FPGA timestamp, initial time-shift, layer index, and part index have statistically significant effects on time-shift adjustment, meaning that these variables can be used to predict the time-shift adjustment; (3) the best model for predicting the time-shift adjustment is the one that uses FPGA timestamp, and layer index as predictors, with the highest prediction accuracy. The main implications and recommendations of the study are: (1) A lower bound on initial time-shift can substantially reduce the computational burden of data synchronization between the sensors, as it can help skip frames from the beginning of the MPM camera video; (2) a confidence interval on the time-shift adjustment prediction-error can inform the accommodations for accurate synchronization of FPGA and MPM data.

Introduction

LPBF is a metal 3D printing technology that fabricates metal parts based on a digital 3D model by selectively melting fine metal powders using a laser in a layer-by-layer fashion. LPBF is used to manufacture complex and intricate parts in critical industries like aerospace and health care. Due to its critical application, to ensure close quality control, various types of sensors/cameras are used to collect data during the manufacturing process for monitoring purposes. These data are known as in-process monitoring data (Zhang et al., 2023).

In this project, in-process monitoring data collected from two sensors are investigated – a Field-Programmable Gate Array (FPGA) sensor, and a melt pool monitoring (MPM) camera. The FPGA sensor tracks the position of the laser across the build plate during the manufacturing process and their associated timestamps. The FPGA sensor also records analog digital converter (ADC) values at each timestamp indicating whether the laser is on or off. The MPM camera captures video of the melt pool created by the laser and includes timestamps associated with each frame of the video. These two sensors run independently and therefore their timestamps need to be synchronized for downstream analysis. Figure 1(a) shows a plot of the scaled FPGA signal data (recorded ADC values) and MPM signal data (maximum intensity in the frame) across their associated timestamps, for a single layer within a part. An initial time-shift (TS_0) is observed between the two data streams. Figure 1(b) shows the result of the initial time-shift adjustment of the FPGA timestamps. We are interested in investigating whether the initial time-shift varies across different layers and parts. Additionally, establishing a lower bound for the initial time-shift (lb_{TS_0}) can help optimize computational efficiency. This lower bound aids in determining a starting point (see Figure 1(b)) for the search for jumps in the MPM signal, streamlining the process by avoiding a search from the very beginning. Note that a higher MPM signal value corresponds to a higher maximum intensity in the video frame, indicating the activation of the laser.



(a) Before initial time-shift adjustment

(a) After initial time-shift adjustment

Figure 1. FPGA and MPM data for a single layer within a part.

The initial time-shift adjustment aligns the FPGA and MPM data at the beginning but not across the entire timeline. Figure 2 illustrates the variation in the time-shift across different blocks of the data. Therefore, additional time-shift adjustments (τ_i) for the FPGA data are required after the initial time-shift adjustment to synchronize the entire data, as expressed in Eq. (1).

$$TS_i = TS_0 + \tau_i \tag{1}$$

Here, TS_i is the adjusted time-shift, TS_0 is the initial time-shift, and τ_i is the time-shift adjustment. Figure 2 hints at a monotonic trend for the time-shift adjustment (τ_i). We are interested in modeling the statistical relationship of the time-shift adjustment with the variables – FPGA timestamp, initial time-shift, layer index, and part index. Next, we aim to predict the time-shift adjustments after identifying the set of appropriate predictor variables. Finally, we want to obtain a confidence interval on the time-shift adjustment prediction-error which can inform the accommodations for accurate synchronization of FPGA and MPM data.

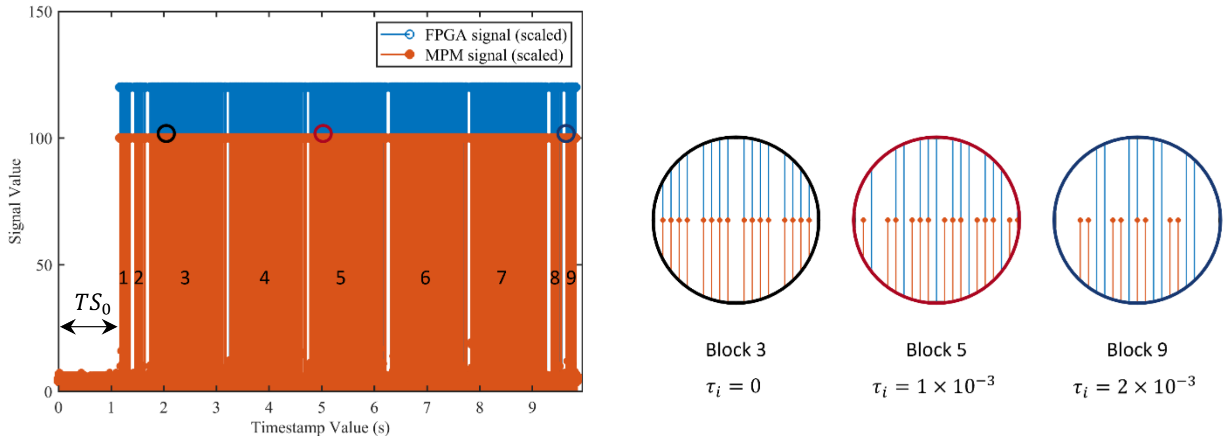


Figure 2. Illustration of the variation of time-shift across different timestamp values.

The research questions (RQs) pursued in this study and the intended methodologies are summarized below:

- RQ1. Does a statistical relationship exist among the variables "Initial time-shift, TS_0", "Layer index, L", and "Part index, P"?
  - A multiple linear regression model with interaction terms will be fitted and the p-values associated with the model coefficients will be examined to confirm the statistical relationship among the variables.
- RQ2. Does a statistical relationship exist among the variables "Time-shift adjustment, τ“, "FPGA timestamp, T_{FPGA}", "Initial time-shift, TS_0", "Layer index, L", and "Part index, P"?
  - A multiple linear regression model with interaction terms will be fitted and the p-values associated with the model coefficients will be examined to confirm the statistical relationship among the variables.
- RQ3. Can "Time-shift adjustment, τ" be predicted reliably by using its statistical relationship with the predictor variables: "FPGA timestamp, T_{FPGA}", "Initial time-shift, TS_0", and "Layer index, L"?
  - 3 candidate models will be explored, and their prediction accuracy will be evaluated using 10-fold cross-validation to select the best model. The candidate models are –Full model, LASSO-penalized model, and Step-wise regression model. Details on these three models are discussed in the “Statistical Analyses” section.

Exploratory Data Analysis

To answer research question 1, initial time-shift (TS_0) values are collected from the two parts (P) for all layers (L). Figure 3 shows generalized pairs plots for the dataset collected for the analysis of initial time-shift values. The plots are obtained using the "GGally" package which is an extension to "ggplot2". The discrete numeric variable "Layer_ID" indicates the index of each layer. "Part_ID" is a categorical variable with two levels that identifies the part index. The continuous numeric variable "InitialTimeShift" measures the initial time-shift (TS_0) values. Note that Part 1 includes 321 layers and Part 2 includes 401 layers. The plot of "InitialTimeShift" vs "Layer_ID" reveals a weak positive correlation between the initial time-shift values and the layer indices. The data points are more dispersed for higher layer indices, indicating more variability in the initial time-shift values. The box plots of initial time-shift values across the two parts show that Part 1 has a lower median and a smaller interquartile range than Part 2, suggesting that Part 1 has lower and more consistent initial time-shift values. Part 2 also has several outliers that are much higher than the rest of the data.

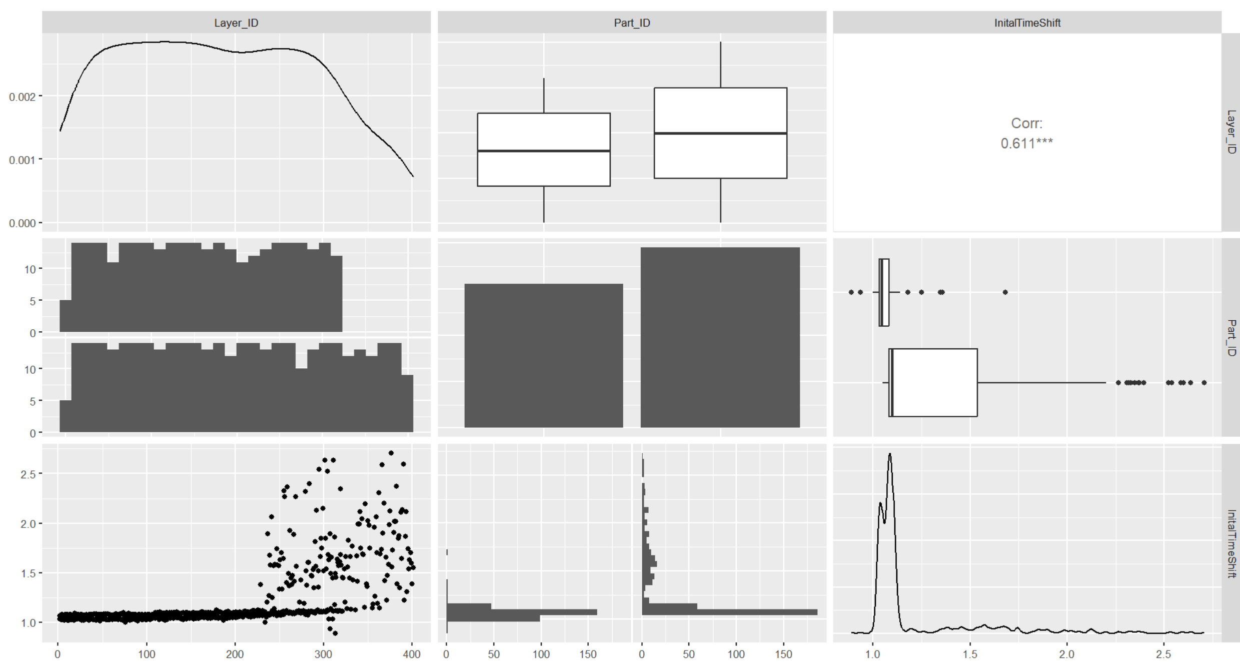


Figure 3. Generalized pairs plot of the dataset collected for the regression analysis of initial time-shift values.

To answer research questions 2 and 3, time-shift adjustments (τ) values across FPGA timestamps (T_{FPGA}) are collected from the two parts (P) for selected layers (L) along with associated initial time-shift (TS_0) values. Figure 4 shows generalized pair plots for the dataset collected for the analysis of time-shift adjustment values. In this dataset, we have two additional variables – the numeric variables "FPGATimeStamps" and "TimeShiftAdjustments" denote the recorded FPGA timestamps and the time-shift adjustment (τ) values, respectively. The pair plot and the correlation metric (-0.999) show a strong negative correlation between the two variables. Figure 5 zooms in on the plot of time-shift adjustments vs FPGA timestamps, where the discrete nature of the time-shift adjustment values is evident. Hence, the predicted values from the regression model need to be discretized.

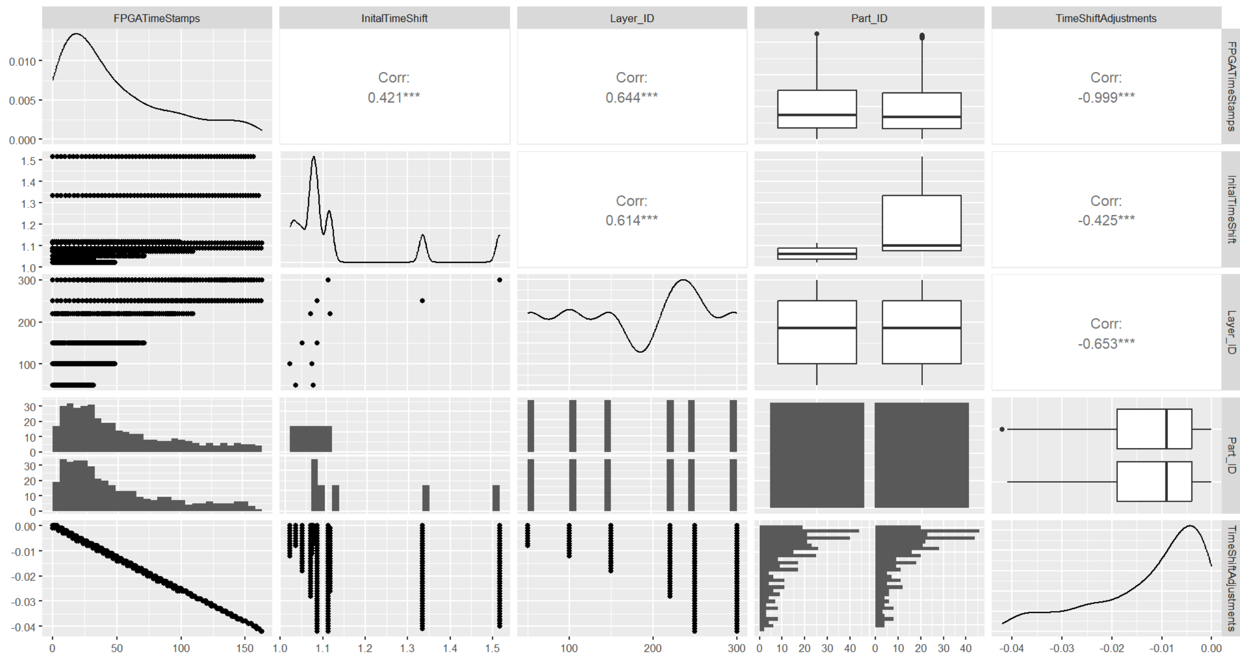


Figure 4. Generalized pairs plot of the dataset collected for the regression analysis of time-shift adjustment values.

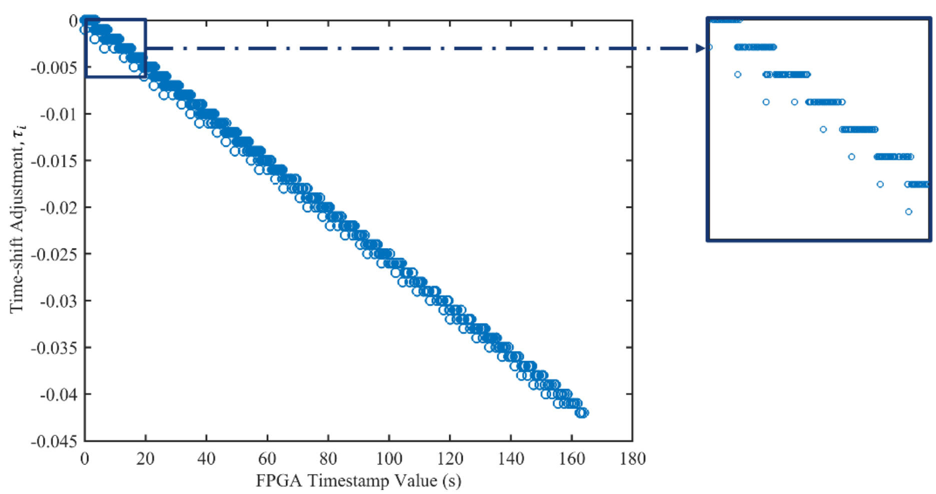


Figure 5. Time-shift adjustments vs FPGA timestamps plot.

Statistical Analyses

To address the first research question, a regression model is fitted with the "InitialTimeShift" as the response variable and the "Layer_ID" and "Part_ID" as explanatory variables. Both main and interaction effects are included in the model. The summary of the fitted model is shown in Figure

6. The model coefficients have low p-values, indicating that they are statistically significant. Hence, it can be concluded that the initial time-shift (TS_0) values differ across different layers and parts.

```
TS0_formula <- (InitalTimeShift ~ Layer_ID + Part_ID + Layer_ID:Part_ID)
model_TS0_all <- lm(TS0_formula, data = df_TS0)
summary(model_TS0_all)
```

```
##
## Call:
## lm(formula = TS0_formula, data = df_TS0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55106 -0.06097 -0.00399  0.02038  1.08003
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.0063053  0.0232055  43.365  < 2e-16 ***
## Layer_ID            0.0003313  0.0001247   2.657  0.00806 **
## Part_IDPart 2      -0.1294588  0.0310462  -4.170 3.43e-05 ***
## Layer_ID:Part_IDPart 2  0.0019283  0.0001535  12.566  < 2e-16 ***
```

Figure 6. Summary of the fitted model for the regression analysis of initial time-shift values.

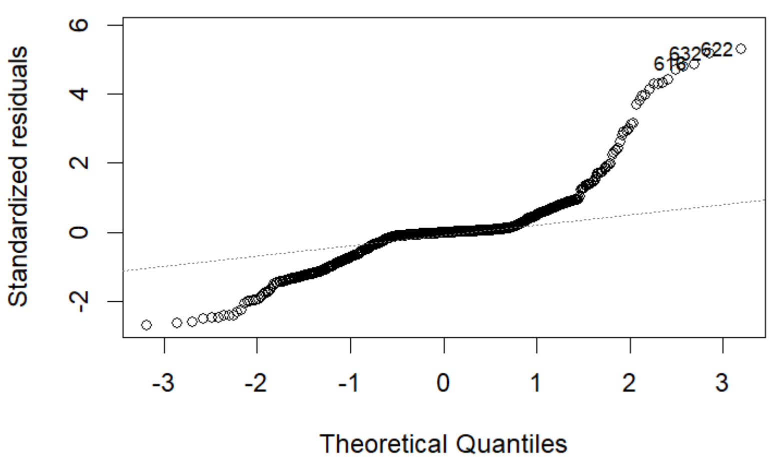


Figure 7. Q-Q plot of the residuals of the fitted model for the regression analysis of initial time-shift values.

The Q-Q plot in Figure 7 reveals that the normality assumption for the residuals of the fitted model is not valid. Hence, bootstrapping with 1000 iterations is used to obtain a 99% confidence band for the fitted values of initial time-shift. Note that a 99% confidence interval is used in this study to obtain conservative results. The lowest band value is 0.881 seconds which can be used as a starting point (lb_{TS_0}) to determine the initial time-shift. Thus, 881 frames from the MPM camera video (1000 fps) can be skipped, and the computational burden can be reduced to a great extent.

To address the second research question, a regression model is fitted with the response variable "TimeShiftAdjustments", and the explanatory variables "FPGATimeStamps", "InitialTimeShift", "Layer_ID", and "Part_ID". Both main and interaction effects are included in the model. The summary of the fitted model is shown in Figure 8. The p-values associated with model coefficients indicate either the main or interaction effect of all the explanatory variables. The variables "FPGATimeStamps", "InitialTimeShift", and "Layer_ID" will be used for the prediction of time-shift adjustments. Although part index (Part_ID) is found to be an important factor affecting time-shift adjustments, it is not a suitable variable for prediction purposes.

```
tau_formula_initial <- (TimeShiftAdjustments ~ FPGATimeStamps + InitalTimeShift +
                  Layer_ID + Part_ID + FPGATimeStamps:InitalTimeShift +
                  FPGATimeStamps:Layer_ID + FPGATimeStamps:Part_ID +
                  InitalTimeShift:Layer_ID + InitalTimeShift:Part_ID +
                  Layer_ID:Part_ID )

model_initial <- lm(tau_formula_initial, data = df_tau)
summary(model_initial)
```

```
##
## Call:
## lm(formula = tau_formula_initial, data = df_tau)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -8.031e-04 -2.559e-04 -4.170e-06  2.521e-04  8.370e-04
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -2.754e-03  2.197e-03  -1.254 0.210260
## FPGATimeStamps                 -2.659e-04  3.829e-06 -69.464  < 2e-16 ***
## InitalTimeShift                 2.981e-03  2.158e-03   1.381 0.167571
## Layer_ID                        9.558e-06  4.283e-06   2.232 0.025931 *
## Part_IDPart 2                  -1.622e-03  2.015e-03  -0.805 0.421068
## FPGATimeStamps:InitalTimeShift  1.555e-05  4.410e-06   3.527 0.000446 ***
## FPGATimeStamps:Layer_ID        -2.634e-08  8.534e-09  -3.086 0.002100 **
## FPGATimeStamps:Part_IDPart 2   -3.422e-06  1.326e-06  -2.581 0.010034 *
## InitalTimeShift:Layer_ID       -9.845e-06  4.077e-06  -2.415 0.015990 *
## InitalTimeShift:Part_IDPart 2   1.898e-03  1.990e-03   0.953 0.340680
## Layer_ID:Part_IDPart 2         -4.970e-06  7.995e-07  -6.217 8.39e-10 ***
```

Figure 8. Summary of the fitted model for the regression analysis of time-shift adjustment values.

To develop a reliable prediction model for time-shift adjustments, three candidate models are explored –

1. *Full model*: This model includes all the main effects and interaction effects of the predictor variables.
2. *LASSO-penalized model*: "cv.glmnet" with LASSO penalty is employed to select the model that corresponds to the minimum mean-squared error across a 10-fold cross-validation (Friedman et al., 2010).
3. *Step-wise regression model*: Step-wise regression in both directions is used to select the model based on the AIC (Akaike's information criterion) metric (Kutner et al., 2005).

The full model includes six terms. The LASSO-penalized model and Step-wise regression model retain two and five terms, respectively. The terms included in the candidate models are reported in Table 1. The model with the highest average prediction accuracy under 10-fold cross-validation will be selected for future implementation. The prediction accuracy under each iteration of cross-validation is calculated using Eq. (2).

$$\begin{gathered} e_i = \tau_i - \mathrm{round}(\hat{\tau}_i, 3) \\ \text{Accuracy (\%)} = \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(e_i = 0) \right) \times 100 \end{gathered} \tag{2}$$

Here, τ_i is the true time-shift adjustment value; $\hat{\tau}_i$ is the predicted time-shift adjustment value which is rounded to the 3rd decimal place for discretization; e_i is the prediction error; $\mathrm{I}(e_i = 0)$ denotes the indicator function that equals 1 if the condition $(e_i = 0)$ is true, and 0 otherwise; n is the number of test data, and Accuracy (%) is the prediction accuracy under an iteration of cross-validation. The average prediction accuracies of the three candidate models under 10-fold cross-validation are reported in Table 1. The LASSO-penalized model has the highest accuracy, and therefore, it is selected for time-shift adjustment prediction.

Table 1. Average cross-validation accuracies of the candidate models for time-shift adjustment prediction.

Model	Average Cross-validation Accuracy
Full model $(\tau \sim T_{FPGA} + TS_0 + L + T_{FPGA} * TS_0 + T_{FPGA} * L + TS_0 * L)$	72.80075%
LASSO-penalized model $(\tau \sim T_{FPGA} + L)$	74.09945%
Step-wise regression model $(\tau \sim T_{FPGA} + TS_0 + L + T_{FPGA} * L + TS_0 * L)$	73.06049%

Figure 9 shows the Q-Q plot of the residuals of the LASSO-penalized model. The normality assumption is not valid for the discreet residuals, as evident from the Q-Q plot. Hence, bootstrapping with 1000 iterations is employed to obtain a 99% confidence interval for the residual values. The histogram of the residuals obtained in the bootstrapping iterations is presented in Figure 10. The 99% confidence interval is [−0.001, 0.001] seconds. Hence, after obtaining the prediction for τ_i, synchronization results need to be checked for τ_i and (τ_i±0.001) to confirm the accurate synchronization. If the confidence interval were wider, we would need to search a wider neighborhood around τ_i, for accurate synchronization.

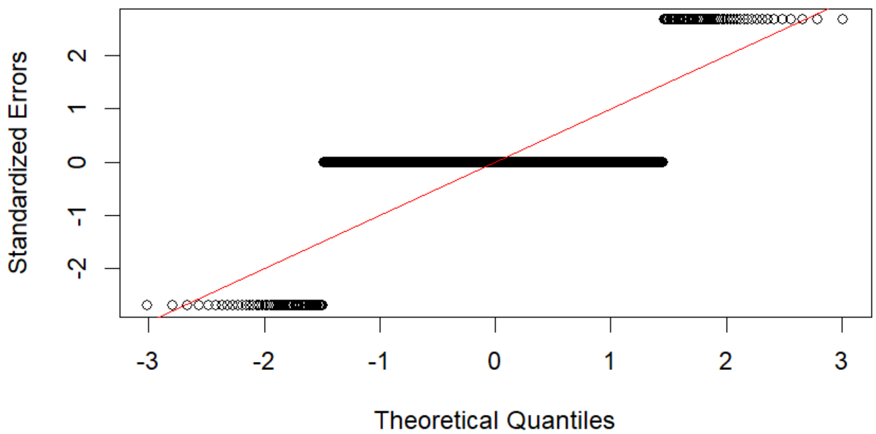


Figure 9. Q-Q plot of the discretized residuals of the LASSO-penalized model.

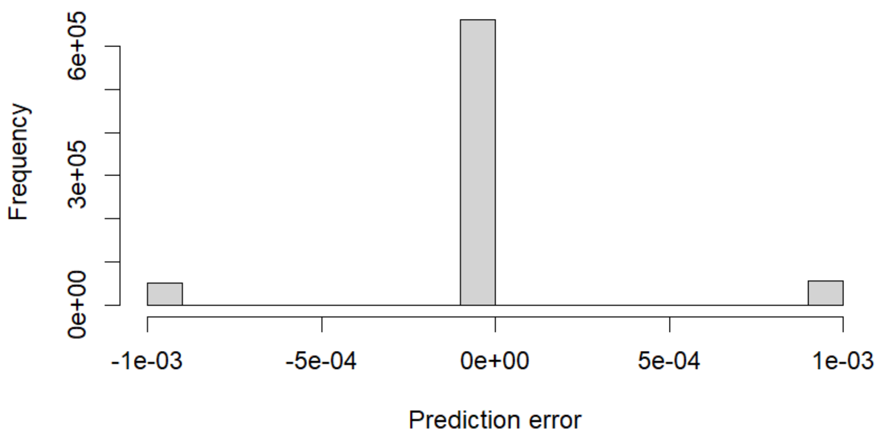


Figure 10. Histogram of the discretized residuals of the LASSO-penalized model under the bootstrapping iterations.

Conclusions

This study investigates the statistical relationships among five variables (FPGA timestamp, initial time-shift, layer index, part index, and time-shift adjustment) that are involved in the data synchronization between the FPGA sensor and the MPM camera. Conducted regression analyses reveal important insights regarding these variables. First, it is found that initial time-shift values vary across different layers and parts. Additionally, a starting point (0.881 seconds) is obtained using bootstrapping, for determining the initial time-shift for a given layer within a part. In practice, this starting point can substantially reduce computational burden as 881 frames can be skipped from the beginning of the MPM camera video. Second, time-shift adjustments are found to be affected by the factors FPGA timestamp, initial time-shift, layer index, and part index. Finally, three candidate models were compared with respect to their prediction accuracies under 10-fold cross-validation and the best model was selected. The LASSO-penalized model is found to have the highest prediction accuracy (74.09945%). Bootstrapping was used to obtain the maximum time-shift adjustment prediction error magnitude (0.001 seconds). This information can be used to examine synchronization results around the neighborhood of the predicted value, and thus ensure accurate synchronization. The regression analyses in this study were limited to the dataset collected from two parts. Data collected from more parts and across more layers can yield more statistically reliable results – which can be a potential future research direction.

References

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1 - 22. https://doi.org/10.18637/jss.v033.i01

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models*. McGraw-hill.

Zhang, Y., Hasan, N., Middendorf, J., Spears, T., Smith, T., Zhang, F., Shafae, M., & Wessman, A. (2023). Correlating Alloy Inconel 718 Solidification Microstructure to Local Thermal History Using Laser Powder Bed Fusion Process Monitoring. Proceedings of the 10th International Symposium on Superalloy 718 and Derivatives,