## Project Summary

## Motivation

Long-duration exploration missions require a paradigm shift in the design and manufacturing of space architectures. The ability to perform In-Space Manufacturing (ISM) provides a solution toward sustainable, flexible missions (both in-transit and on-surface) through on-demand fabrication, repair, and recycling capabilities for critical systems, habitats, and mission logistics and maintenance. This effort calls for innovative manufacturing processes that can produce electrical components in space.

The Electrohydrodynamic Inkjet (EHD) Printing process employs electrical force to deposit liquid ink droplets and achieve 3D printing in zero-gravity environments, making it a promising Additive Manufacturing technique for fabricating electrical components in micro and nano scales. In contrast to the highly specialized manufacturing equipment required for producing semiconductor components, EHD printing offers a flexible solution for creating conductive structures.

However, the precise process parameters required to create a Taylor Cone, a crucial physics phenomenon enabling nanoscale printing, are currently set manually. Our long-term objective is to develop AI-enabled, closed-loop, autonomous techniques for remotely controlling EHD printing. To achieve this, we first aim to investigate the relationship between droplet deposition and voltage level using images captured by high-speed cameras observing the Taylor Cone. The height of the Taylor Cone is extracted from each image and the ‘peaks’ in Taylor Cone height can be defined as where the Taylor Cone deposits a droplet of ink. The frequency pattern of droplet deposition is dependent on many process parameters, but this project aims to address how voltage affects printing behavior, and all other factors held constant. We plan to investigate if it is possible to predict droplet deposition from past printing behavior and how voltage affects the accuracy of these predictions.

## Problem Statement

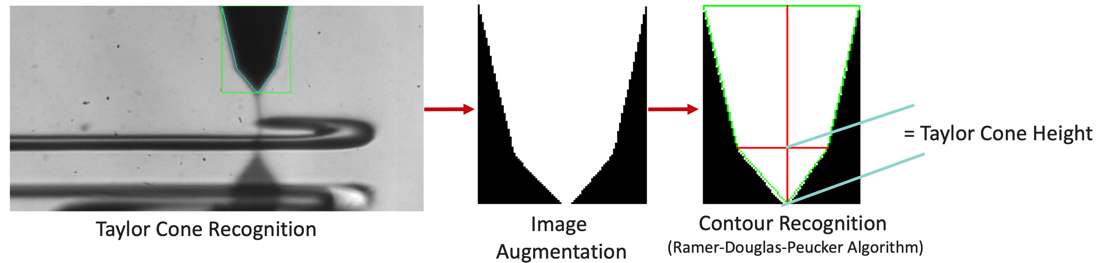
First, we investigate if frequency can be detected from the temporal behavior of the Taylor Cone. Then we will determine the effect that varying voltage levels has on droplet deposition. Lastly, based on these observations we will determine if it is possible to predict droplet deposition from past printing behavior.

## Results

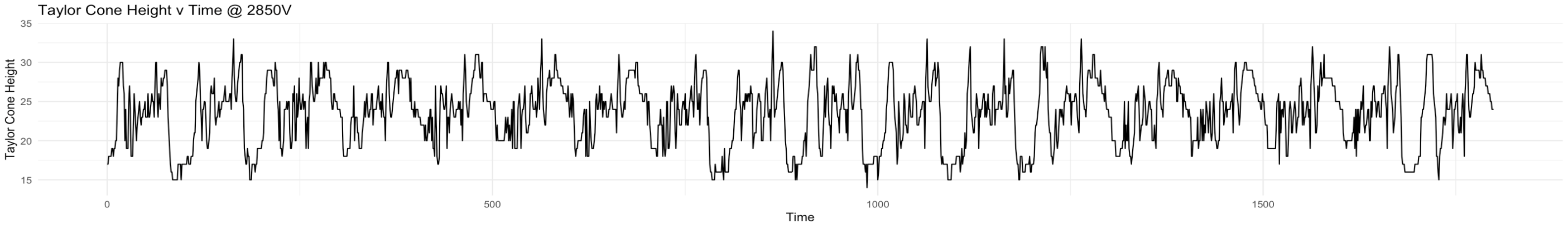
## Data

## Pre-processing

The datasets used in this project were obtained from video footage of EHD printing of silver nano ink during a zero-gravity flight test. The video footage was recorded using a high-speed camera that records at 10,000 frames per second. Thus, images are obtained 0.01 seconds apart. Python was used to extract the Taylor Cone height from each image (Figure 1), which is a critical parameter that depends on the ink type, voltage, and frequency. The frequency was set to 100Hz, and was held constant during the experiment. The datasets consist of Taylor Cone height measurements and time values, collected over a 0.18-second duration (1800 frames) for each of five voltage levels: 1950V, 2100V, 2150V, 2200V, and 2850V. The Taylor Cone height measurements are expressed in pixel units, with each pixel approximately corresponding to 3.33 nanometers. Each dataset comprises 1800 measurements, and the time series graphs for voltage level 2850V can be seen in Figure 2, and the time series graphs for all voltage levels can be found in Table 1 of Appendix A.



*Figure 1: Process of extracting the Taylor Cone Height from each image.*

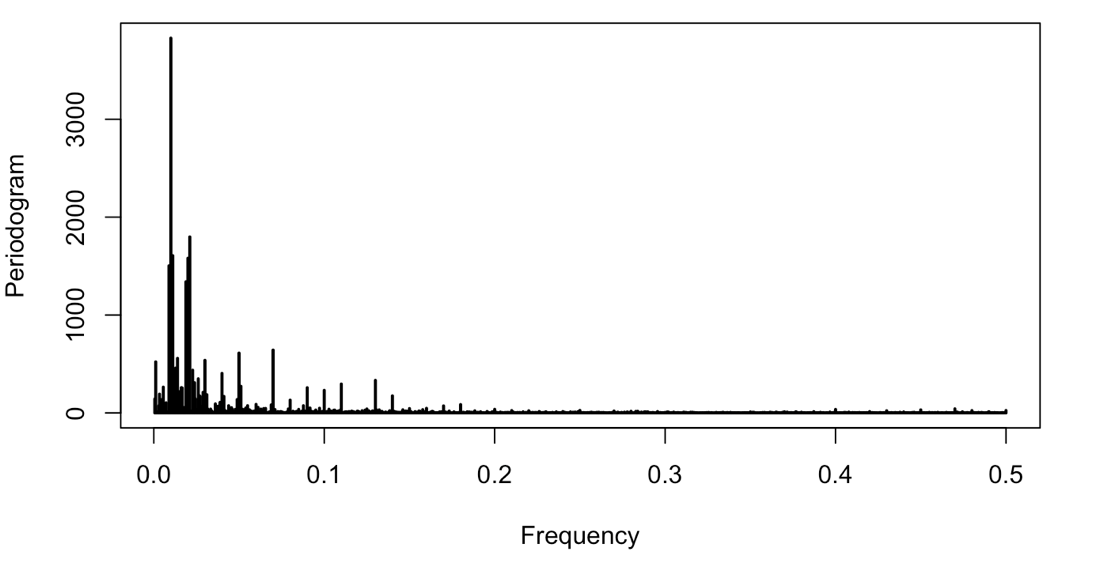
*Figure 2: Taylor Cone Height at 2850 V.*

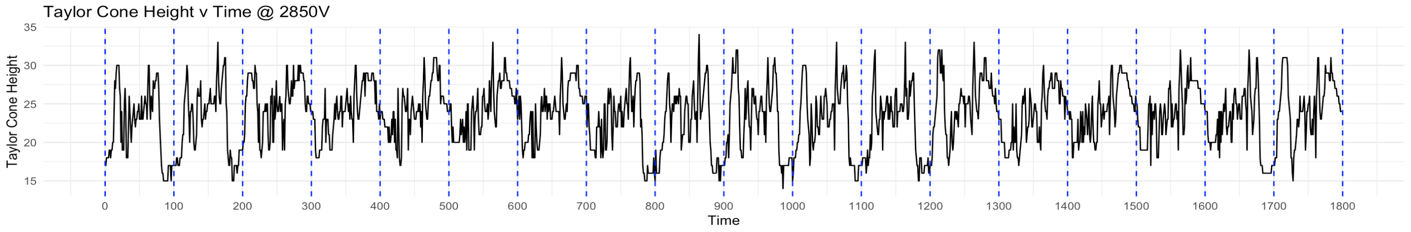
## Seasonality Extraction

Seasonality is an important attribute of this analysis because it represents when a droplet should be theoretically deposited. We would like to see if the datasets display the seasonality that corresponds to the frequency printing parameter. Theoretically, if the algorithm used to measure the Taylor Cone functions correctly, each dataset should exhibit a certain seasonality that corresponds to when the Taylor Cone elongates to deposit a droplet of ink. The set frequency is 100Hz, so the data should represent this seasonal pattern.

To determine the seasonality in the data, a spectral analysis was conducted using the Fourier Transform algorithm [1]. The objective was to represent seasonality in terms of sine and cosine waves with different frequencies. The frequency with the most dominant correlation of the sine and cosine waves to the peaks in the data was determined to be the seasonality. The periodogram() function in R was used to calculate the power spectral density using the fft() function. This function returns a graph of estimated power spectral density values at different frequencies. The frequency with the highest power corresponds to the most significant correlation, thereby allowing the seasonality of the data to be calculated accordingly.

After applying this method to the 2850V dataset, a seasonality of 100 was found. This was evident from the highest peak on the periodogram, as seen in Figure 1. The maximum spectrum corresponded to a frequency of 0.01. By taking the reciprocal of this number, a seasonality of 100 was obtained. This estimate was confirmed by visually graphing the seasonality over the time series, as shown in Figure 7. It was found that the pattern of the Taylor Cone Height, also known as jetting frequency, repeats approximately every 0.18 seconds.



*Figure 1: Spectral density periodogram graph to determine frequency of Taylor Cone Height at 2850V and visually displaying the detected seasonality over the data.*

The same method was utilized for voltage levels of 1950V, 2100V, 2150V, and 2200V. The outcomes of the analysis are summarized in Table 1. The periodogram for each dataset can be found in Table 1 of Appendix B. It was observed that the peaks on the spectral density graphs were less prominent and there were more of them, leading to the conclusion that the datasets had a weaker seasonal pattern than that found in the 2850V dataset. Voltage levels 2200V and 2850V exhibited the expected seasonality: 100. Voltage levels 1950V, 2100V, and 2150V exhibited a lower seasonality. However, the detected frequency for Voltage levels 2100V and 2150V were 33.33 and 50 respectively, both are multiples of 100, meaning there may be some sub-seasonality of droplet deposition at those voltage levels. There may be several reasons for this. At lower voltage levels, there is less electrical force to drive droplets to be released from the nozzle, because of this there may not be enough force to overcome the surface tension at the tip. The Taylor Cone may ‘pulse’ meaning it will build up 1-2 additional droplets at the tip before the surface tension can be overcome to release a droplet defining a different seasonality than the set parameter. Other possible reasons: there may be another factor that contributes to droplet deposition at lower voltage levels, the algorithm that measures the Taylor Cone height is not sensitive enough to detect droplet behavior, or the printing behavior is not stable, i.e., electrospraying may be occurring. This observation and explanation are accompanied by reasonable doubt. The behavior of EHD printing and the relationship between printing parameters in zero-gravity conditions has yet to be defined by physics equations. The project is further continued with the aim of exploring and discovering new findings. 100Hz will be the frequency used in the remainder of the report, as it was the set printing parameter at all voltage levels.

|  |  |
| --- | --- |
| **Dataset** | **Detected Seasonality** |
| 1950V | 81.82 |
| 2100V | 33.33 |
| 2150V | 50 |
| 2200V | 100 |
| 2850V | 100 |

*Table 1: Detected seasonal component for each dataset, 1950V, 2100V, 2150V, 2200V, 2850V.*

## Methods

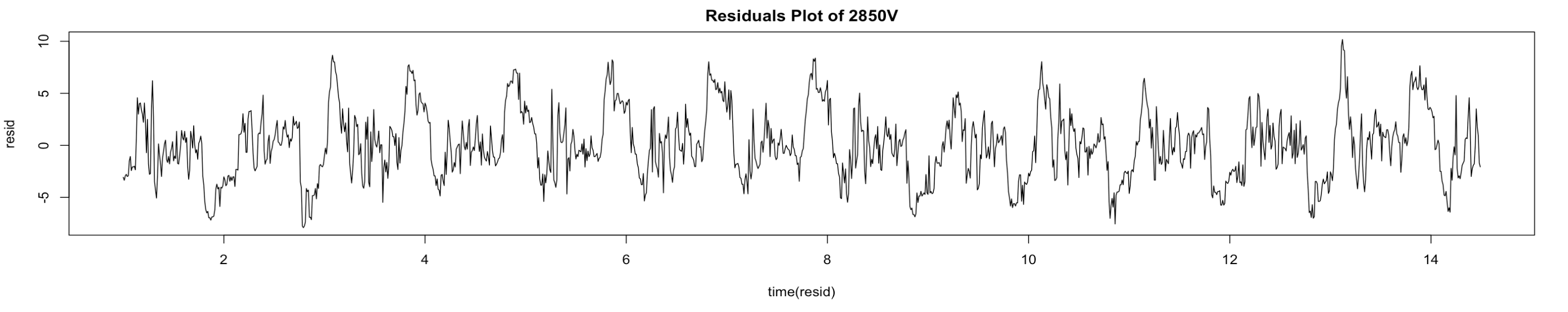
The following models were used to predict future droplet deposition: [LIST MODELS HERE]. The dataset consisted of 1800 observations, corresponding to a 0.18 second period, with the first 1350 observations (0.135 seconds) used for training and the remaining 450 observations (0.0450 seconds) utilized as a validation set. The accuracy of the models were evaluated using the root mean squared error (RMSE) and mean absolute percent error (MAPE), and the results are presented in Table 2 of Appendix B. The actual and predicted values of the models for 2850 Voltage are shown in Table 2 as time series plots. The red line indicates the prediction for the training period and the blue line represents the prediction for the validation period.

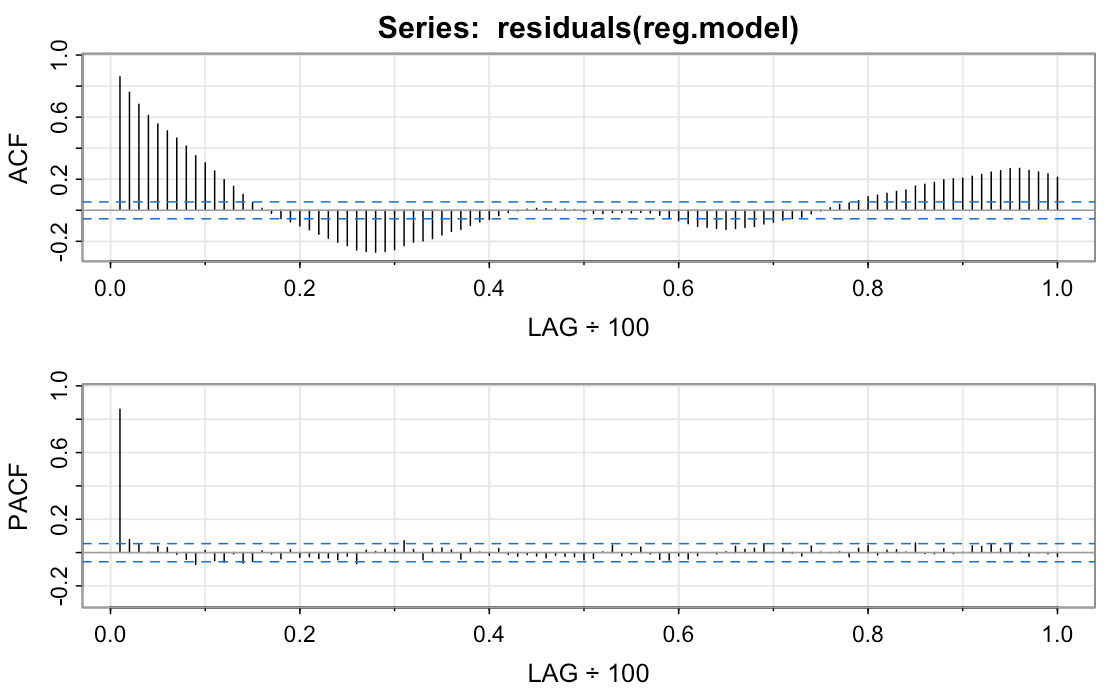
## Linear Regression

Describe Linear Regression Method/R-Code and summarize RMSE and MAPE results for varying voltage levels. + Bar graphs of RMSE and MAPE

## Linear Regression and ARIMA

After modeling the data using the linear regression model, the residuals were graphed to investigate any remaining seasonality. The graph of the residuals at 2850V can be seen in Figure X. The residuals show significant seasonality, so ARIMA was used to model the error between the predicted and observed Taylor Cone Heights. The ACF and PACF were graphed to show significant autocorrelation between the lag values. Shown in Figure X, there is a significant spike at 2 on the PACF graph corresponding to an ARIMA(2,0,0)(0,0,0) model. The auto.arima() function in R was also tested and returned the optimal model of ARIMA(2,0,2)(0,0,0). Neither method returned evidence for a seasonal component in the ARIMA model; however, the graph of the residuals displayed significant seasonality, thus we tested several other ARIMA models. It was found that removing the nonseasonal components and focusing on only the seasonal components proved to be more successful than the previous models. The results from all ARIMA models were added to the linear regression predicted values and their accuracies are summarized in Table X. The best model was ARIMA(0,0,0)(0,0,2) with a RMSE of 3.393 and MAPE of 11.623 for the 2850V dataset. In practice, this tuning should be performed for each dataset individually; however, for simplicity, we will use the linear regression + ARIMA(0,0,0)(0,0,2) model for all voltage levels. The RMSE and MAPE can be summarized in Figure X using this method for each of the 5 voltage levels. After fitting these models, the residuals did not display any seasonality and resembled white noise.

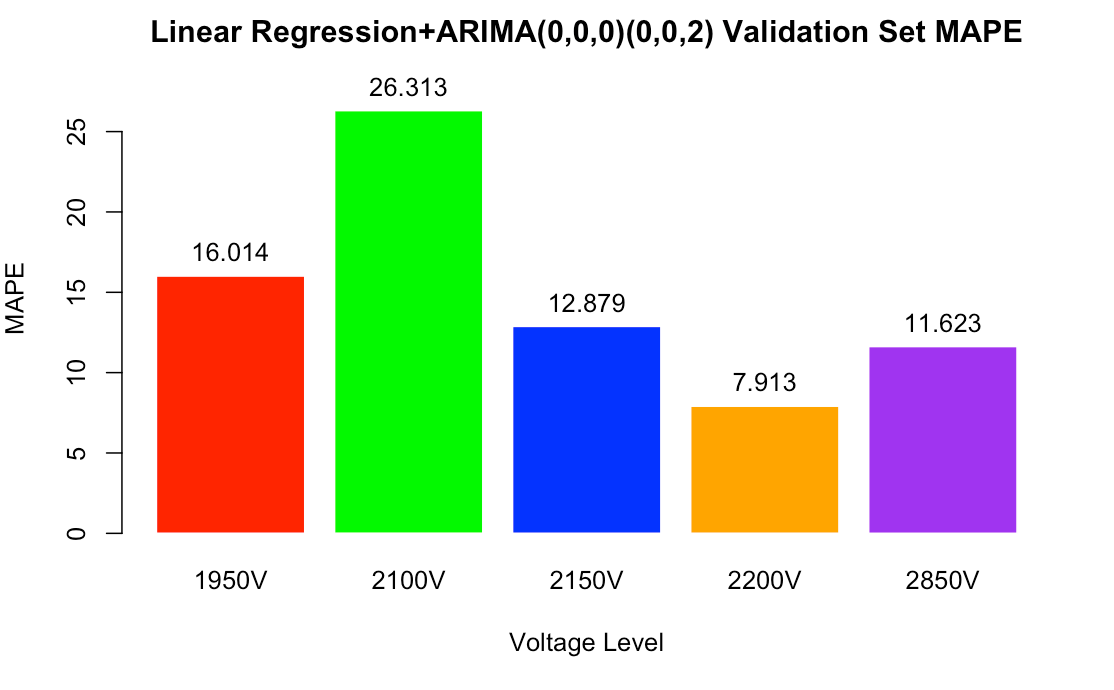
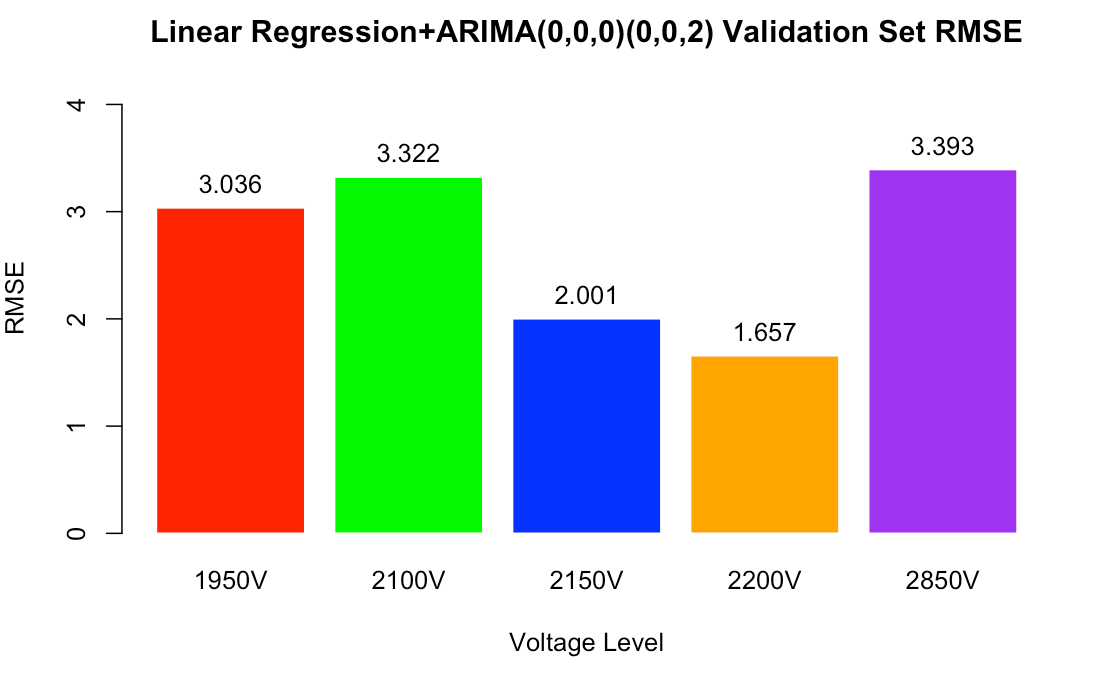
*Figure X: Residuals from the Linear Regression model at 2850V.*



*Figure X: ACF and PACF graphs depicting the autocorrelation between lagged values of the residuals at 2850V.*

|  |  |  |
| --- | --- | --- |
| **Model** | **RMSE** | **MAPE** |
| Linear Regression | 3.541512 | 12.10918 |
| Linear Regression + ARIMA(2,0,0)(0,0,0) | 3.546187 | 12.12071 |
| Linear Regression + ARIMA(2,0,2)(0,0,0) | 3.547515 | 12.11787 |
| Linear Regression + ARIMA(0,0,0)(1,0,0) | 3.410815 | 11.6643 |
| Linear Regression + ARIMA(0,0,0)(0,0,1) | 3.433305 | 11.70699 |
| Linear Regression + ARIMA(0,0,0)(1,0,1) | 3.410821 | 11.65852 |
| Linear Regression + ARIMA(0,0,0)(2,0,0) | 3.410966 | 11.65133 |
| Linear Regression + ARIMA(0,0,0)(0,0,2) | 3.392988 | 11.62316 |

*Table X: RMSE and MAPE of various Linear Regression models using ARIMA to predict the residuals at 2850V.*



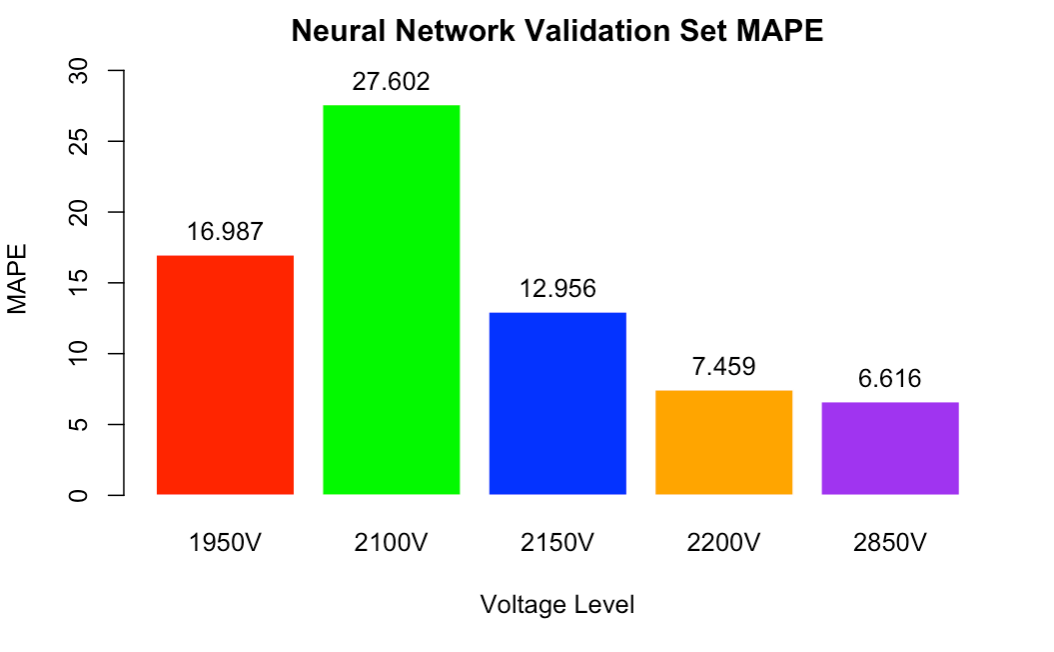
*Figure X: RMSE and MAPE of the Linear Regression + ARIMA(0,0,0)(0,0,2) model for each voltage level.*

## Auto ARIMA

Describe Auto ARIMA Method/R-Code and summarize RMSE and MAPE results for varying voltage levels. + Bar graphs of RMSE and MAPE

## Neural Network

In an attempt to model the complex printing behavior not captured by previous models, we turn to deep learning using the ‘keras’ package in R. The 10 previous observations of Taylor Cone height were used as the lags at each time step and used as input to the network. Three different structures of networks were tested, one with 5 layers, 7 layers, and 10 layers. The neural network with 5 layers performed best. It used two dense layers with 32 units each and ReLU activation. The third and fourth dense layers have 64 units each and the fifth, 1 output unit. The network is compiled using the RMSprop optimizer and mean squared error (MSE) as the loss function. The mean absolute error (MAE) is used as the performance metric while training. The early stopping callback is used to stop the training process if validation loss does not improve for 200 epochs. The network is trained for 1000 epochs with a batch size of 25. The RMSE and MAPE can be summarized in Figure X for each of the 5 voltage levels. Voltage level 2100V performed exceptionally bad compared to the other levels: RMSE 4.078 and MAPE 27.602. Voltage levels 2200V and 2850V displayed promising RMSE of 1.583 and 2.14 and MAPE of 7.459 and 6.616 respectively. A graph of the modeled and forecasted Taylor Cone heights for 2850V can be found in Table 2.



*Figure X: RMSE and MAPE of the Neural Network model for each voltage level.*

## Other Models

Describe Other Method/R-Code and summarize RMSE and MAPE results for varying voltage levels.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Forecasted Plot** | **RMSE** | **MAPE** |
| Linear Regression |  | 3.541512 | 12.10918 |
| Linear Regression + ARIMA(0,0,0)(0,0,2) |  | 3.392988 | 11.62316 |
| Auto ARIMA |  | 3.606936 | 12.69013 |
| Neural Network |  | 2.139696 | 6.616387 |
|  |  |  |  |
|  |  |  |  |

*Table #: Forecasted Time Series Plots using Linear Regression, Auto ARIMA, [OTHER MODELS] of Taylor Cone height at 2850V and their associated RMSE and MAPE for the testing sets.*

## Results

## Discussion

## Limitations

The datasets used in this project were obtained from a zero-gravity flight test, which may not be readily available in large quantities. This may limit the ability to train and test future models to validate these methods. The EHD printing process involves a complex physics phenomenon, which may be difficult to model accurately. The Taylor Cone height is one critical parameter that depends on the ink type and is affected by various factors such as voltage level, fluid properties, and environmental conditions. Understanding and accounting for these factors is challenging. The datasets used in this project were observed to have a significant amount of noise, which may affect the accuracy of the models. Preprocessing techniques are required to remove the noise and extract useful information. The models developed in this project may not provide a clear explanation of how they arrived at their predictions. This may limit their interpretability, and it may be difficult to identify and address any issues or biases in the models.

## Future Directions

## References

[1] “Fourier transform,” *Wikipedia*, 01-Apr-2023. [Online]. Available: <https://en.wikipedia.org/wiki/Fourier_transform>. [Accessed: 11-Apr-2023].

## Appendix A

|  |
| --- |
| **Time Series Graphs** |
|  |
|  |
|  |
|  |
|  |

*Table 1: Time series graphs for each voltage level: 1950V, 2100V, 2150V, 2200V, and 2850V .*

## Appendix B

|  |  |  |
| --- | --- | --- |
| **Dataset** | **Detected Seasonality** | **Spectral Density Graph** |
| 1950V | 81.82 |  |
| 2100V | 33.33 |  |
| 2150V | 50 |  |
| 2200V | 100 |  |
| 2850V | 100 |  |

*Table 1: Spectral Density Graph for each dataset to determine frequency.*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | | **Linear Regression** | | **Linear Regression + ARIMA(0,0,0)(0,0,2)** | | **Auto ARIMA** | | **Neural Network** | |
| **RMSE** | **MAPE** | **RMSE** | **MAPE** | **RMSE** | **MAPE** | **RMSE** | **MAPE** |
| **1950** | **Train** | 3.628155 | 22.57195 | 3.599503 | 137.9328 | 3.396413 | 19.87025 | 2.571651 | 13.55835 |
| **Test** | 3.040542 | 16.04217 | 3.035815 | 16.01442 | 2.931261 | 15.49805 | 3.304403 | 16.98718 |
| **2100** | **Train** | 3.170801 | 24.44505 | 3.16107 | 6.837738e+12 | 3.085741 | 22.69814 | 2.198925 | 13.14145 |
| **Test** | 3.316492 | 26.25681 | 3.322498 | 26.31287 | 3.23778 | 26.02416 | 4.078364 | 27.60189 |
| **2150** | **Train** | 2.233102 | 16.33006 | 2.211052 | 5.497713e+13 | 2.272428 | 16.20931 | 1.782321 | 10.47422 |
| **Test** | 1.995198 | 12.82626 | 2.001136 | 12.87926 | 1.958208 | 12.9104 | 2.11136 | 12.95577 |
| **2200** | **Train** | 1.327941 | 6.017763 | 1.322714 | 1.9546417e+11 | 1.25888 | 5.432147 | 1.397038 | 6.623678 |
| **Test** | 1.651289 | 7.85782 | 1.656573 | 7.912924 | 1.631874 | 7.734764 | 1.582689 | 7.458777 |
| **2850** | **Train** | 3.254894 | 11.79577 | 3.161958 | 8.433485e+13 | 1.930643 | 6.34382 | 1.706153 | 5.437034 |
| **Test** | 3.541512 | 12.10918 | 3.392988 | 11.62316 | 3.606936 | 12.69013 | 2.139696 | 6.616387 |

*Table 2. Initial Accuracy Results (RMSE&MAPE) for Linear Regression and Auto ARIMA*