Title
Model Development to Predict Power Output for a Novel Laser Test System

Author
To, Lynette

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Model Development to Predict Power Output for a Novel Laser Test System

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Applied Statistics

by

Lynette Ho Ching To

2023

ABSTRACT OF THE THESIS

Model Development to Predict Power Output for a Novel Laser Test System

by

Lynette Ho Ching To
Master of Applied Statistics
University of California, Los Angeles, 2023
Professor Hongquan Xu, Chair

Optical spectroscopy technology has evolved significantly, from its use in research to detect new materials to its use in the medical field to diagnose diseases. Applications of this technology have been used in wearables that provide insights to promote health and wellness. Rockley Photonics has prototyped a novel wearable via a miniaturized laser system, and the characterization of the laser output power is important for algorithm development. The functional data of interest includes repeated measurements across multiple laser systems operating at various current and temperature ranges. The ability to develop a repeatable and reliable test system is crucial, and thus this paper aims to apply statistical methods to propose a framework in developing models for power prediction. Prototype laser systems (N=5) were tested using a standard procedure and laser output power was characterized to understand significant test factors. Preliminary findings suggest that operator and repeated measurement factors are not significant to the output power and that models can be developed on an individual laser channel bases using an ordinary least squares (OLS) model with third order current effects. The application of using the same base model for individual lasers on each tested device is important for continued development of algorithms to be used

to detect various biomarkers in the health sensing space.

The thesis of Lynette Ho Ching To is approved.

Yingnian Wu

Jingyi Li

Hongquan Xu, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

Spectroscopy is the study of light and its effect on matter. Its application extends to many different disciplines including quantum research, medical imaging, and detection of distant asteroids in space. An everyday application of this technology includes wearable technology, which most consumers would recognize as heart rate monitoring via a photoplethysmogram. The fundamentals behind the photoplethysmogram involves the emission of LED light (green, red, infrared) to detect changes in volumetric changes in blood circulation.[1] Rockley Photonics, a company striving to develop a next generation wearable sensor, has taken this principle a step further using a proprietery laser sensing technology. Applications of the lasing technology can be used to detect various biomarkers or health metrics such as core body temperature. [2] This paper defines the "laser system" as the prototype that contains 36 laser channels and an on-board photodetector that measures laser signal. Laser signal is measured as a voltage, which can be correlated to laser output power. For the purposes of discussion in this report, laser power refers to the voltage signal measured from the lasers. While laser power is critical in developing various biomarker detection algorithms, there are other aspects of the laser system that need to be understood to ensure that prototypes produce repeatable and reliable measurements.

A standard protocol was developed to measure laser power in a controlled setting. Figure 1.1 provides an overview of the test system, which is defined as the test environment and

[1]Castaneda, Esparaza, Gharami, Soltanpur, 2018

[2]Rockley Photonics, 2021

procedure in which the lasers system is operating. A current driver controls the current, ranging from 20 mA to 250 mA in 10 mA steps, at which the lasers operate. When the test begins, a temperature controller of the test environment begins heating the laser system. The test sequence then proceeds where the lasers are turned on in order of channel number (i.e. from Laser 1 to Laser 36). The light that is emitted from the modules is reflected off of a control medium and detected by the on-board photodetector. For the entire duration of the test, an on-board thermistor is measuring the temperature of the laser system.

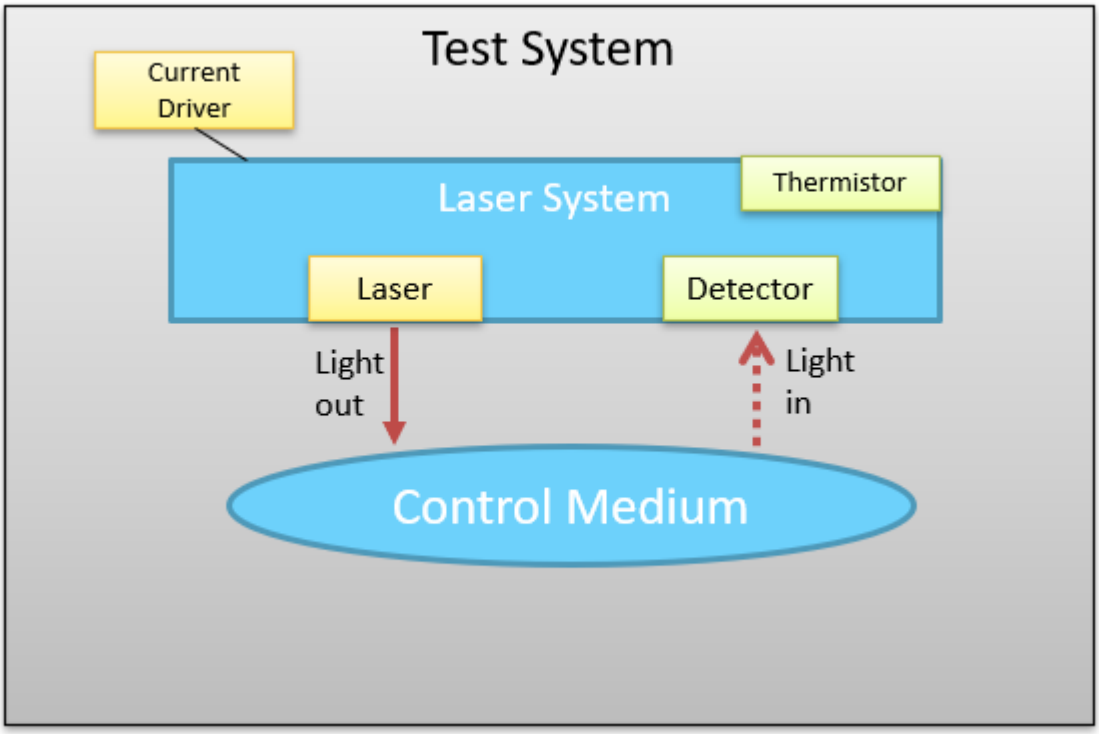

Figure 1.1: Test System Overview

Independent variables such as drive current and laser channel will contribute to the amount of laser power that is detected. Similarly, the test system's temperature as measured by the on-board thermistor may also affect laser power. The goal of this paper is to determine how much variability is contributed by these different sources and to propose a framework

for developing power prediction models that can be used in a manufacturing setting.

CHAPTER 2

Dataset of Interest

The dataset of interest was collected in Pasadena, California in January 2023 on a prototype test system. This chapter will describe an overview of how the data was collected and the motivating question for this thesis.

2.1 Measurement Data Explained

Three operators took repeated measurements on each of the 5 devices under test. The sequence of events for each measurement on the test system involves a temperature ramp (controlled by a temperature controller) and a "sweep" through the currents from 20 mA to 250 mA in 10 mA steps (i.e. the current starts at 20 mA, then 30 mA, then 40 mA... then 250 mA, and back to 20 mA). As a result, an identifier called "Measurement Number" is used to identify a full "current sweep" for each of the lasers as they are pulsed in order. Figure2.1 illustrates a simplified representation of how the current is swept across the temperature ramp for each measurement. At each current set point, all 36 lasers are pulsed one at a time. For instance, when the device is driving at 20 mA, Laser 1 turns on followed by Laser 2 followed by Laser 3... up to Laser 36. The drive current then increases to 30 mA and Laser 1 turns on followed by Laser 2 etc.

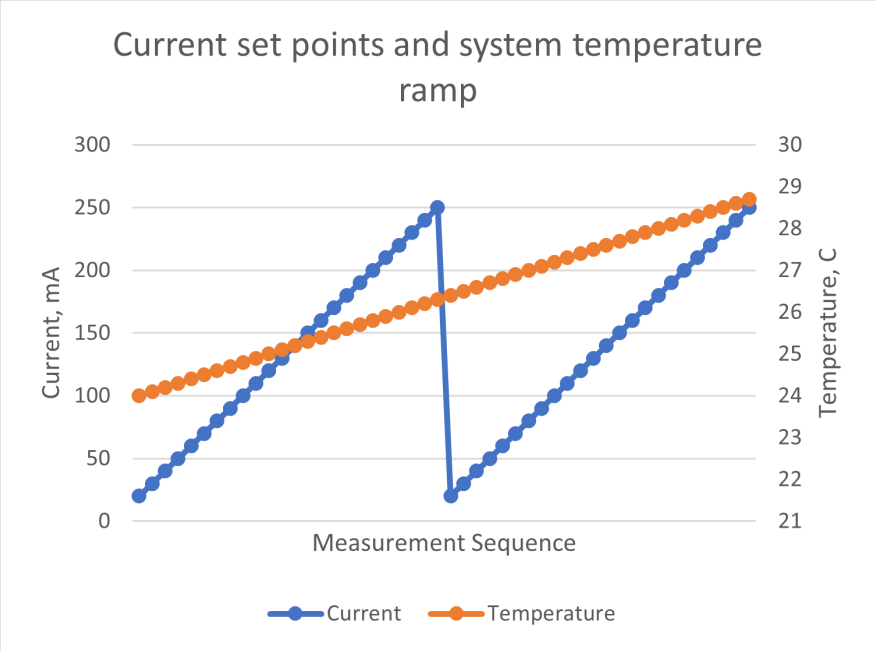

Figure 2.1: Simplified schematic of current sweep and temperature ramp

The dependent variable of interest is power measured by the on-board photodiode. Figure 2.2 shows a representative example of measured power over a full current sweep for Laser 1 on a single measurement. Aggregating all the measurements for the same laser across the entire measurement (i.e. full temperature range), Figure 2.3 shows the overlaid curve for the same laser. As the temperature increases, the signatures of each sweep is observed to shift to the right. The movement of the power curves within a measurement is a challenge when comparing across repeated measurements when assessing repeatability. Recall that there are 36 lasers for each measurement, increasing the dimensions of the data collected for each measurement.

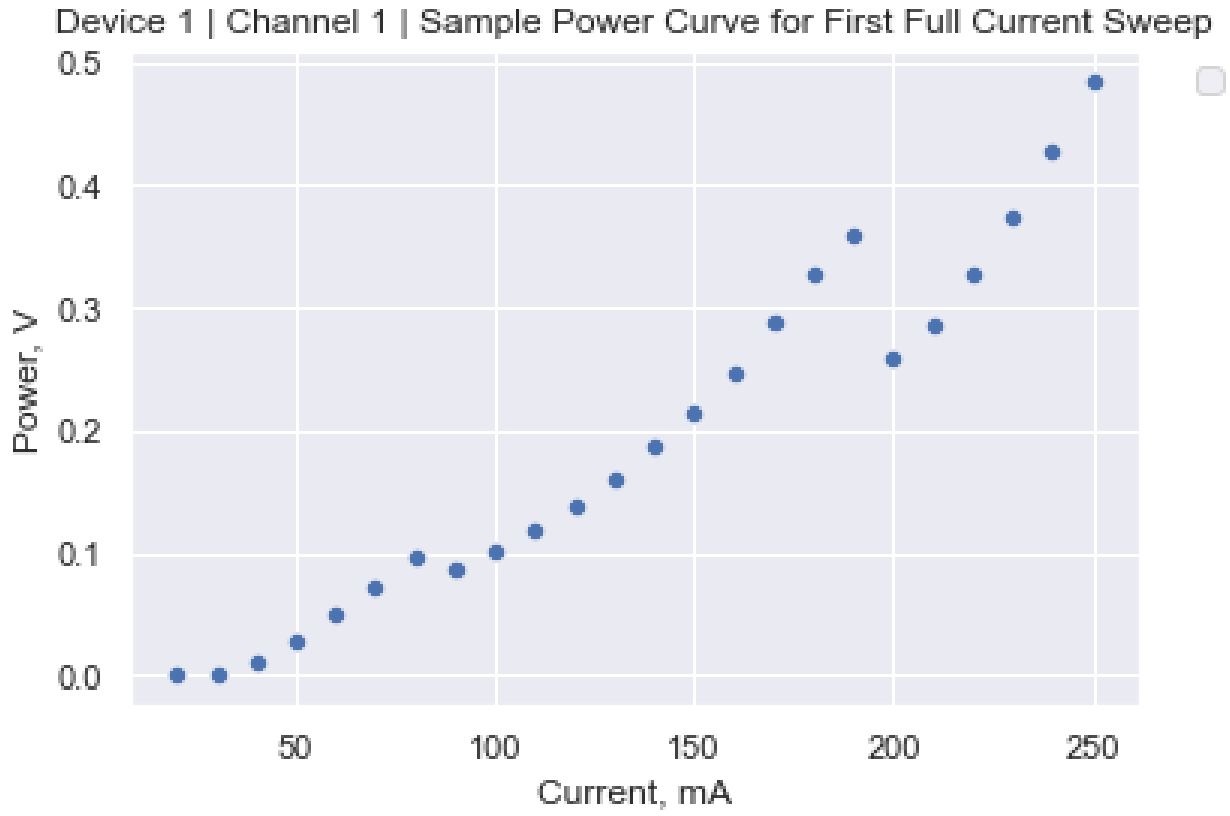

Figure 2.2: Single current sweep for one channel on one device

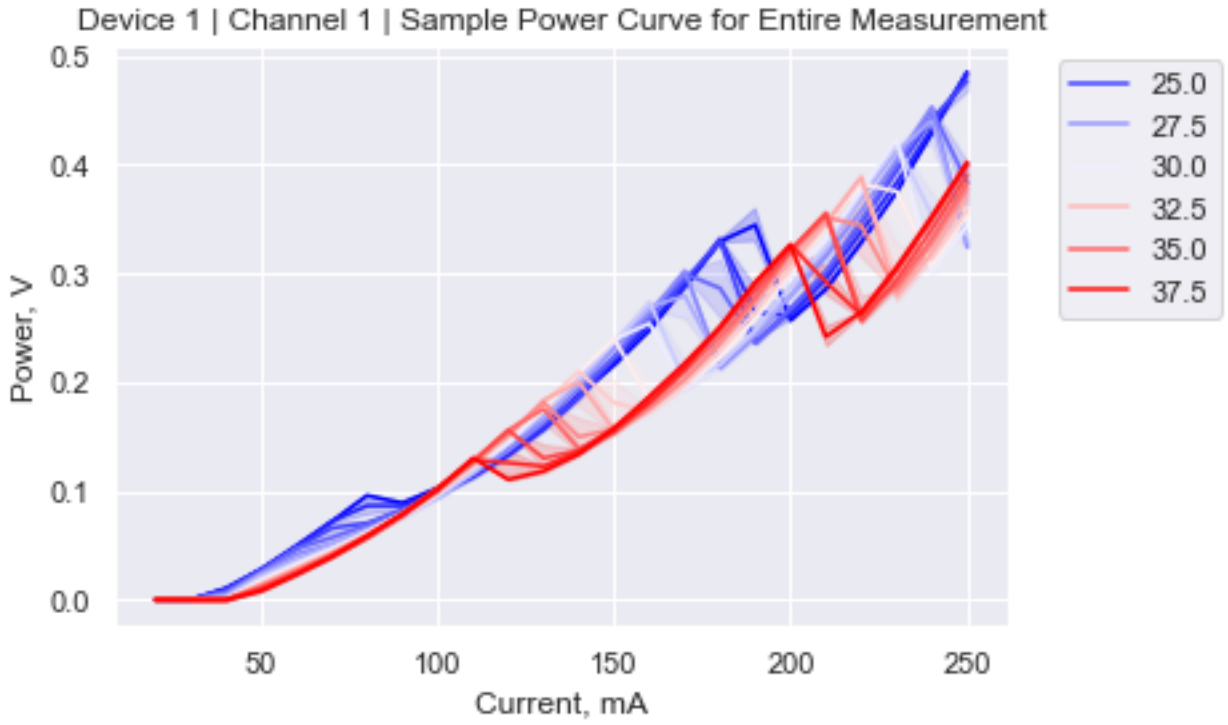

Figure 2.3: Overlaid current sweep for one channel on one device across temperature ramp

The variability in the measurements can be shown by the different shapes of the power measurements for a single laser and one operator. The different colored curves in Figure 2.4 demonstrates challenges in the repeatability of the curve with repeated measures for a single operator. Further, the measured temperature of the devices are also variable as seen in Figure 2.5 with the varying start and end temperatures for each repeated measurement. It is also important to note the duration of each measurement; each measurement is aproximately 300 seconds in duration which corresponds to a total test time of approximately 7 minutes given the time associated to cool and warm the test system. It is favorable to have a shorter test time with a more narrow temperature ramp such that the associated test time is minimized and increasing throughput of tested devices.

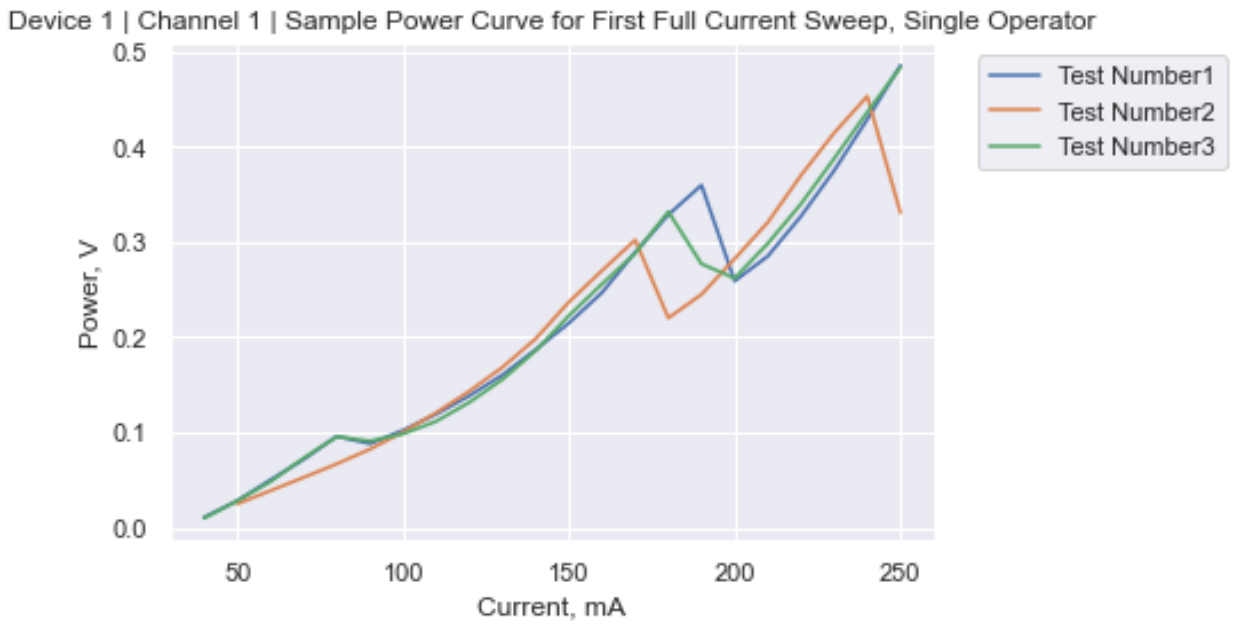

Figure 2.4: Overlaid current sweep for one channel across 3 repeated measurements

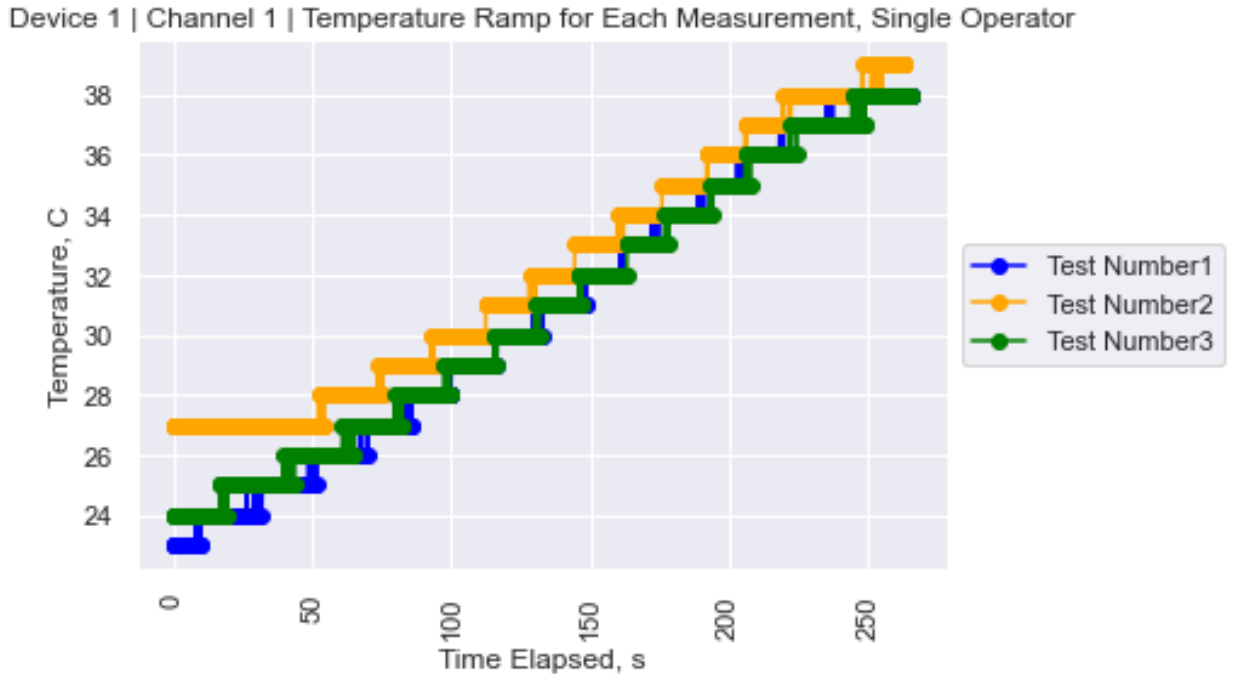

Figure 2.5: Example of temperature ramp for 3 repeated measurements

The final variable to consider is the individual operator. The measured dataset includes measurements from three different operators. In addition to the variability of measurements within operators, there is also variability contributed across operators. The different colored curves in Figure 2.6 shows the power curve for the same channel for a single measurement of each of the operators. Again, there is a challenge of how similar the curves are given the extra variable of different operators.

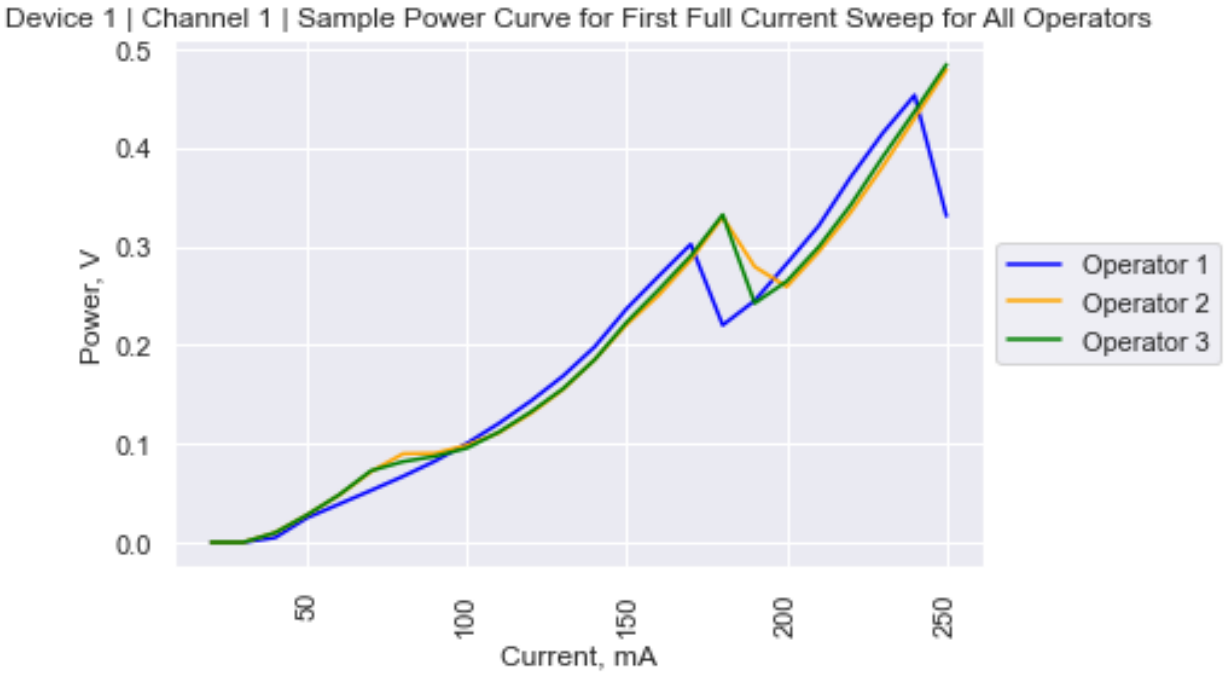

Figure 2.6: Example of single current sweep for three operators

The objective of this thesis is thus to answer the following questions:

- What is an appropriate framework to assess repeatability of the measurements given the various factors (current, temperature, operator, repeated measurements) on the measured output (power)?
- Using this framework, how can models be developed to approximate measurement curves?

2.2 Dataset Overview

The columns of data are described as follows:

- Device: the identifier of each unique device [unitless]
- Current: the drive current of the prototype laser system [mA]
- Power: the power measured on the on-board photodetector [V]
- Temperature: the temperature of the laser system measured on the on-board thermistor [C]
- Start Time: the date and time of the measurement [MM/DD/YYYY HH:MM]
- Test Number: Encoded test repeat number from start time [unitless]
- Operator: the identifier of the unique operators [unitless]
- Channel: the identifier of the laser channel number [unitless]
- Measurement Number: the identifier for the measurement of each "sweep" of the lasers [unitless]

CHAPTER 3

Exploratory Data Analysis

3.1 Data Inclusion

The dataset of interest was collected based on the availability of the operators and the test station, which resulted in an unbalanced count of measurements for each device and each operator. Methodologies for randomization and design of experiments could be implemented to further improve the findings from the analysis. [1]. A total of individual measurements were taken across the 3 operators and 5 devices. The counts of measurements per device and operator is summarized in Figure 3.1.

[1] C. F. J. Wu and M. S. Hamada (2009)

Device	operator	
Device 1	Operator 1	3
	Operator 2	1
	Operator 3	2
Device 2	Operator 1	2
	Operator 2	1
	Operator 3	3
Device 3	Operator 1	2
	Operator 2	1
	Operator 3	2
Device 4	Operator 1	3
	Operator 2	1
	Operator 3	2
Device 5	Operator 1	2
	Operator 2	2
	Operator 3	4

Figure 3.1: Test counts for each device and operator

The distribution of power output demonstrated the need for data exclusion. Due to the system operating at low currents such that power output was undetectable, there was a disproportionate amount of data that was measured as zero power or close to zero power. As a result, data where output power measured less than 0.05 V was removed from the datset. Figure 3.2 illustrates the probability distribution before removal of data (top) and after removal of low power points (bottom).

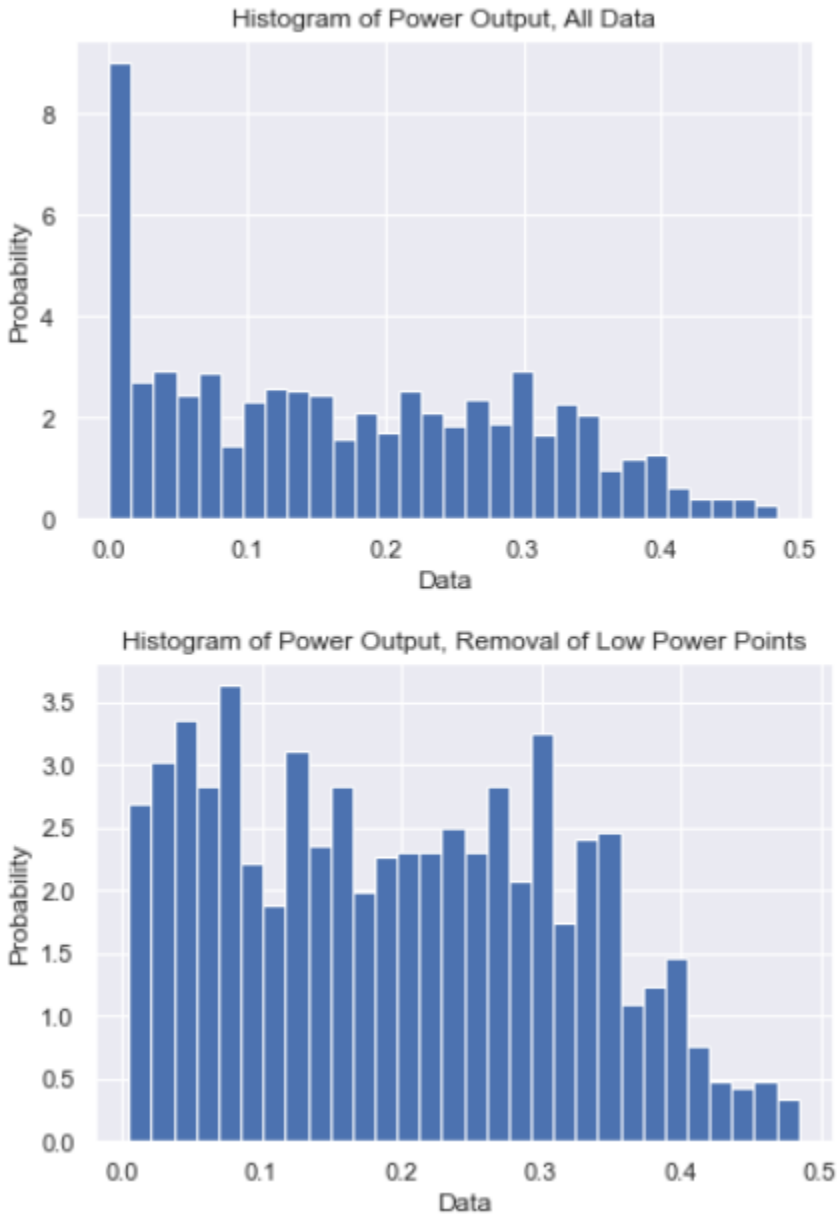

Figure 3.2: Probability density histogram for power data

The response variable of interest is power output and the predictors of interest are current and temperature. The relationship of these variables are summarized in Figure 3.3 for a single laser channel (Channel 1). The relationship between current and power suggests a higher-order polynomial relationship where power current increases with temperature. The

relationship between temperature and power is not as clear and will be explored in the future sections.

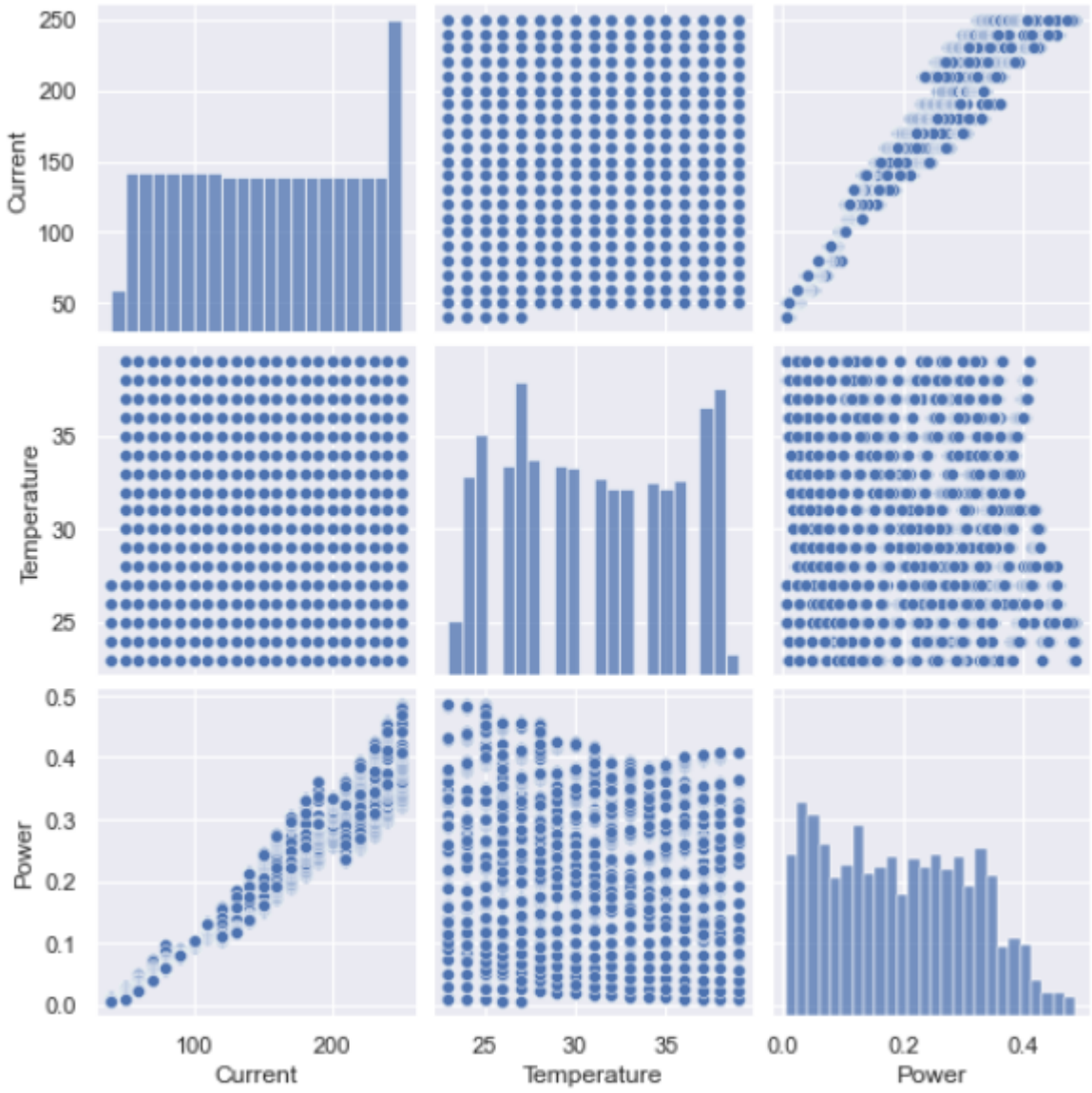

Figure 3.3: Summary of response and predictors for Channel 1

Each device has 36 lasers, all of which have a potentially unique response during the test sequence. To show a simple representation of the exploratory data analysis, the remainder of this chapter will consider a truncated dataset with one operator, one test measurement,

one device, and one channel (Channel 1).

3.2 Multiple Linear Regression

To study the relationship of each predictor on the response, a simple multiple linear regression[2] was applied to the truncated dataset with one operator, one test measurement, one device, and one channel (Channel 1) . Equation 3.1 displays the multiple linear regression expression where y refers to output power, x_1 refers to temperature, x_2 refers to current, and the β terms refer to the respective coefficients or vectors of coefficients (β_1, β_2) and intercept (β_0).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{3.1}$$

The need for a transformation was also investigated.[3] The top plot in Figure 3.4 shows characteristics of a bimodal distribution of the raw output power, and the bottom plot shows a more normally distributed distribution of the transformed (square root) power output. The remainder of this section will explore the multiple linear regression on both of these outputs.

[2] J. J. Faraway (2005)

[3] S. Sheather (2009)

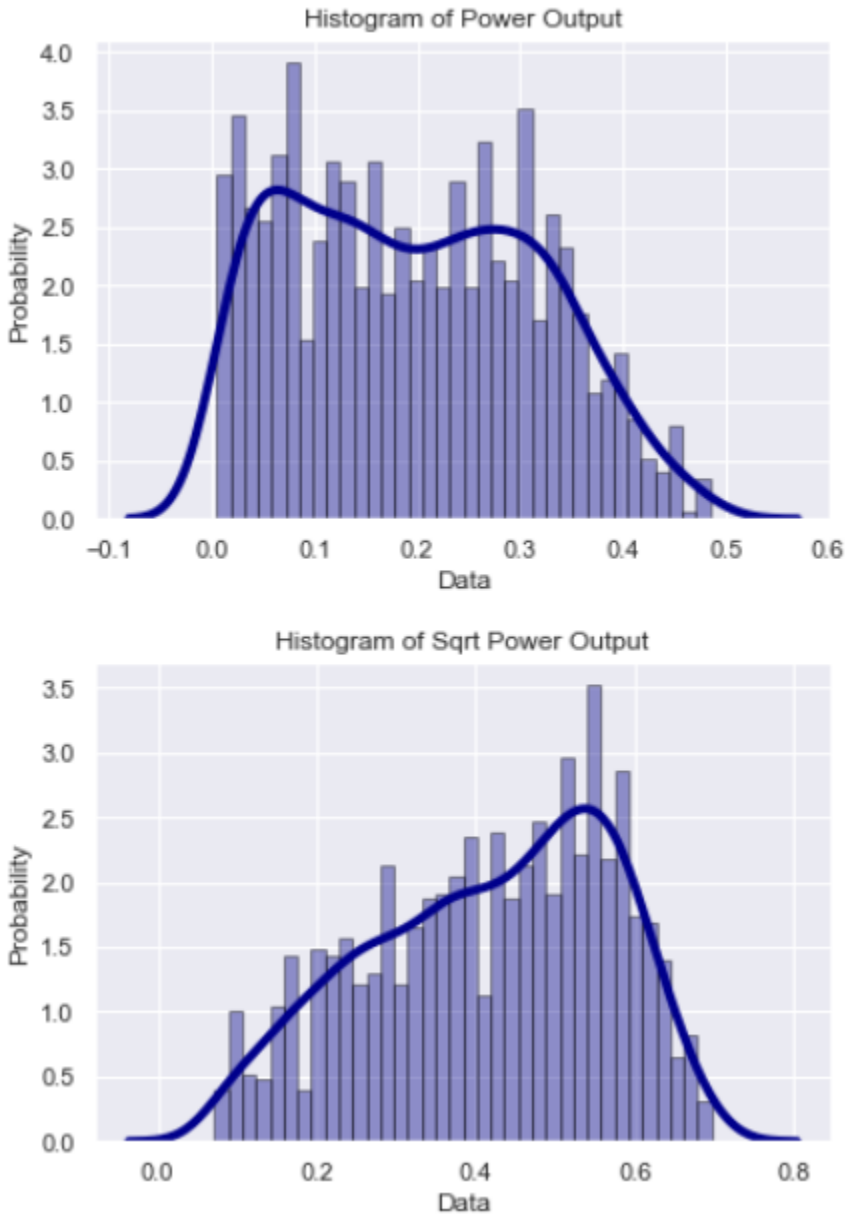

Figure 3.4: Probability density histogram of raw and transformed power

The summary for each multiple linear regression is shown in Figure 3.5. The R^2 value of regression using the raw power output (a) was 93.9% and resulted in residuals that increased with the fitted power output. The R^2 value of regression using the transformed power output (b) was 94.0% and resulted in residuals that have a parabolic shape centered around 0.4 V.

The assumption of homoskedasticity in the residuals is not met, which suggests that there is a trend of increasing variance with the fitted power (a) and trend of variance driven by a higher order factor with the transformed power (b). However, it is interesting to note that the R^2 values for both regressions are similar, so the shape of the residuals suggest the exploration of a polynomial regression. Recall from Figure 3.3 that the predictor, current, may have a higher order relationship with power output. This will be further expanded in the next section.

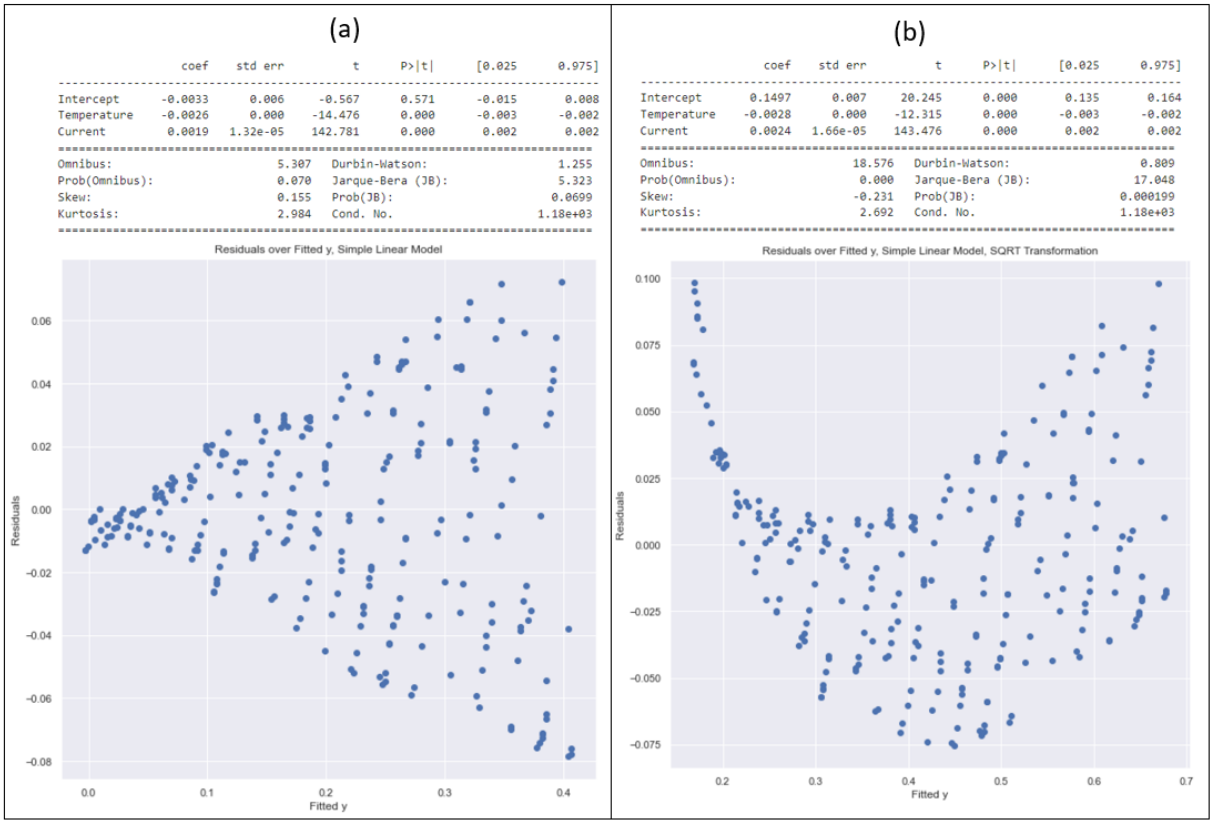

Figure 3.5: Summary of multiple linear regressions for raw (a) and transformed (b) power

The logaraithmic transformation was also explored for the multiple linear regression, which yielded unfavorable results compared to the squre root transformation. However, Section 4.5 will explore the Ordinary Least Squares (OLS) models which demonstrate the benefit of the log transformation.

3.3 Higher Order Fit for Current Term

The exploration of a higher order for the predictor, current, was done by comparing the root mean square error (RMSE) of various polynomial fits using current as the predictor and power as response. Using RMSE as an accuracy metric to compare across, polynomial fits with degrees up to 7 were used to determine identify what the recommended power should be used for the polynomial fit. Figure 3.6 (right) demonstrates that the 2nd order polynomial results in a reduction in RMSE of 0.007 and improvement with the 3rd order polynomial is an improvement in RMSE of less than 0.0005. The left plot shows that the 2nd and 3rd order fits are nearly identical, as the predicted curves are overlaid with the original data. The next chapter uses the 2nd and 3rd order terms for current as a basis for building the models to (a) predict output power and (b) determine the significance of the additional test paramters such as operator and repeat measurements.

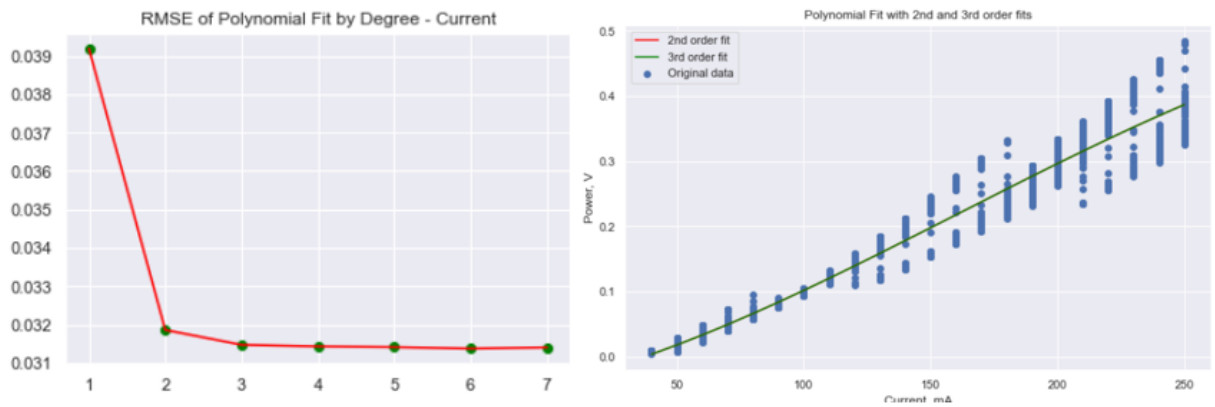

Figure 3.6: Exploration of degree for current term

CHAPTER 4

Model Development

4.1 Framework Overview

The onset of this thesis aimed to define a framework to determine the effects of various factors of a multidimensional experimental dataset. Due to the nested nature of each test measurement, meaning one device is measured numerous times by numerous operators, and each device has power data collected over a range of currents and temperature, this chapter proposes a framework to break down the model development such that the final models can serve as a predictor of output power.

The first phase of the framework is to determine the independence of channels. Given that there are 36 channels on each device, determining how similar or different channel output power is would direct the subsequent phases. The second phase is then to pool all of the data (across all devices, all operators, all repeated measurements) on a single channel. The purpose of this analysis is to assess whether data collected on a given channel is significantly affected by the operator or repeated measurement. The third phase is to further down-select the data, limiting to one device. Isolating the analysis within a single device is motivated by the intrinsic variability due to device-to-device differences and removes an additional source of variation. The final phase is the the most narrowed dataset, which is representative of the potential use case in which laser systems are tested in a manufacturing setting and the models would be used for further development. During this phase, the comparison of models on the same laser channel for different devices will also enlighten the need for models to be

developed on an individual device basis. All models assume that there is no interaction effect of current and temperature on operator, device, and repeat measurement (i.e. *start_time*).

4.2 Phase 1: Comparison of All Channels

The standard model development approach is to feed all factors into a giant model and down-select factors based on the statistical significance of each factor. However, given that each device has 36 laser channels, this approach may not be valid given that there are system designs that define the need for differences in laser channels based on the needs of the laser system application. Ordinary Least Squares (OLS) was used to determine the significance of the predictors on the response.[1] Equation 4.1 defines the base model where y refers to output power, x_1 refers to temperature and x_2 refers to current. *Channel*, *Device*, and *Operator* refer to the the corresponding categorical variables, and the β terms refer to the respective coefficients (β_1 to β_8) and intercept (β_0).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2 + \beta_4 x_2^2 + \beta_5 x_2^3 + \beta_6 Channel + \beta_7 Device + \beta_8 Operator \quad (4.1)$$

The R^2 for this model with all laser channels was 92.8% and the ANOVA summary (Figure 4.1) provides evidence to show that the laser channel is significant. The Channel Number factor contributes the most amount of variance (approximately 24.6% of the total variance) while the operator factor contributes the last amount of variance (approximately 0.00019% of the total variance). It is observed that the sum of squares of the residual (RSS) is 2521.2, which is disproportionately high (about 52.13%) compared to the other model parameters. This suggests that much of the variance is explained by the residuals, indicating a model that can be further improved. The next phase will thus investigate a single channel to compare across all devices and all test measurements.

[1] D. C. Montgomery (2005-2019)

	sum_sq	df	F	PR(>F)
ChannelNum	1188.420782	35.0	24412.677406	0.000000e+00
Device	871.888708	4.0	156716.340162	0.000000e+00
operator	0.009425	2.0	3.388048	3.377476e-02
Temperature	143.000551	1.0	102813.685827	0.000000e+00
Current	80.478272	1.0	57861.789614	0.000000e+00
Temperature:Current	30.086963	1.0	21631.745719	0.000000e+00
I(Current ** 2)	0.837976	1.0	602.483029	5.076107e-133
I(Current ** 3)	0.128848	1.0	92.638137	6.286177e-22
Residual	2521.249530	1812713.0	NaN	NaN

Figure 4.1: ANOVA summary for model from Equation 4.1

4.3 Phase 2: All Devices, All Measurements, One Channel

The effect of the device number was determined by defining another OLS model. The categorical variables *Device* and *Operator* were included in the base model defined in Equation 4.2, where the remainder of the variables are defined the same as Equation 4.1. The R^2 for this model was 93%, and the ANOVA table in Figure 4.2 provides evidence that the device number is significant, which suggests that each device should have its own model to predict output power. The *Operator* term is determined to not be significant at an alpha-level of 0.05, which provides early evidence that operator variability is negligible. This is expected given that the early stages of the prototype development may result in device to device differences, which are shown in this result. It is also observed that the sum of squares of the residual (RSS) is 59.4, which is disproportionately high (about 46.5%) compared to the other model parameters.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2 + \beta_4 x_2^2 + \beta_5 x_2^3 + \beta_6 Device + \beta_7 Operator \tag{4.2}$$

	sum_sq	df	F	PR(>F)
operator	0.000656	2.0	0.271044	7.625842e-01
Device	54.670489	4.0	11299.874059	0.000000e+00
Temperature	8.779900	1.0	7258.890244	0.000000e+00
Current	1.325214	1.0	1095.636598	1.223795e-237
Temperature:Current	3.476442	1.0	2874.191399	0.000000e+00
I(Current ** 2)	0.099028	1.0	81.872787	1.502917e-19
I(Current ** 3)	0.056776	1.0	46.939950	7.405567e-12
Residual	59.387085	49099.0	NaN	NaN

Figure 4.2: ANOVA summary for model from Equation 4.2

4.4 Phase 3: One Device, All Operators, One Channel

The third phase of the model development is determining the effect of operator and repeated measurement on the response. Recall that Figure 3.1 summarized the count of measurements for each device and each operator. For any given device, each operator had taken either one, two, or three measurements. The repeated measurement refers to any measurement that was taken two or more times. As a result, the variable *start_time* in the dataset refers to the categorical variable of each repeated measurement. To illustrate the effect of operator and repeated measurements on the output power, the different colored curves in Figure 4.3 correspond to each individual measurement taken by the corresponding operator. For example, there are two unique shapes of the measurements taken by Operator 1 (blue) where one curve has the first peak at 125 mA and another curve has its first peak at around 150 mA.

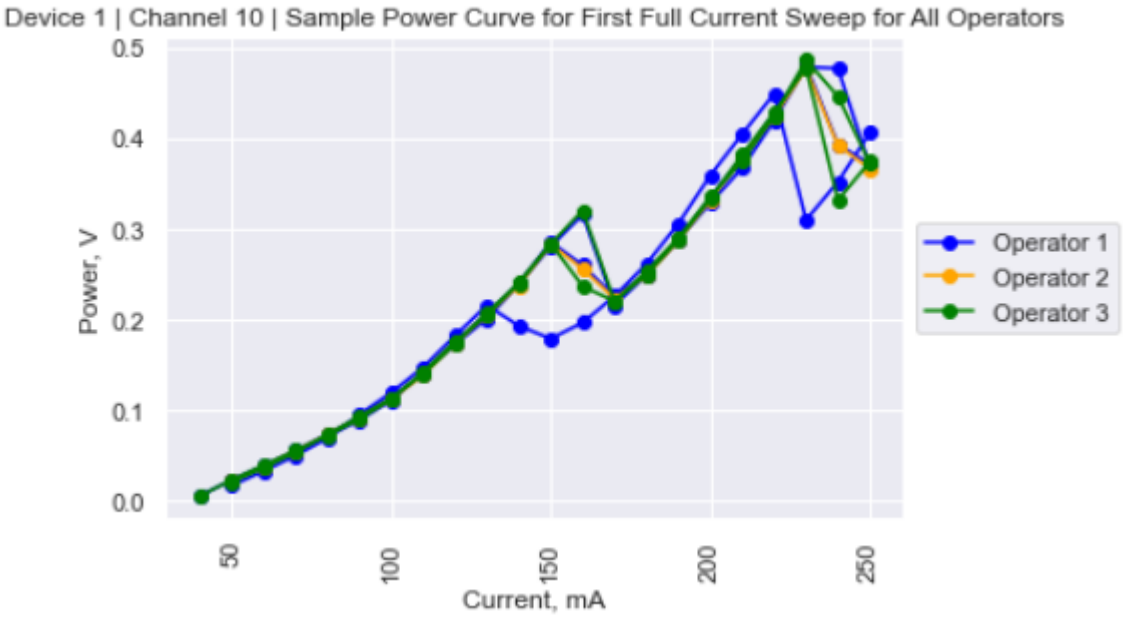

Figure 4.3: Overlaid curve for the first sweep of each measurement for Device 1, Channel 10

Two different models were generated to determine the effect of the operator and repeated measurement on the response variable. Equation 4.3 includes the categorical variable *Operator* and Equation 4.4 includes the categorical variable, *start_time*. Because *Opeartor* is nested within *start_time*, each term is included in each model separately, and not together.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2 + \beta_4 x_2^2 + \beta_5 x_2^3 + \beta_6 Operator \tag{4.3}$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2 + \beta_4 x_2^2 + \beta_5 x_2^3 + \beta_6 start\_time \tag{4.4}$$

The summarized ANOVA output (Figure 4.4) for each model suggests that the operator term is not significant at an alpha level of 0.05 (Model from Equation 4.3) and the repeated measurement term (i.e. *start_time*) is not statistically significant (Model from Equation 4.4) at an alpha level of 0.05. These findings suggest that the operator factor and repeated measurements are not important on the output power measurements.

	sum_sq	df	F	PR(>F)
operator	0.003329	2.0	1.640297	1.939895e-01
Temperature	1.522806	1.0	1500.824355	2.343601e-300
Current	2.708947	1.0	2669.844129	0.000000e+00
I(Current ** 2)	0.670843	1.0	661.159264	5.242633e-140
I(Current ** 3)	0.357779	1.0	352.614457	5.664529e-77
Residual	7.887860	7774.0	NaN	NaN
	sum_sq	df	F	PR(>F)
start_time	0.006421	5.0	1.265639	2.757020e-01
Temperature	1.507984	1.0	1486.225997	1.097028e-297
Current	2.709211	1.0	2670.120625	0.000000e+00
I(Current ** 2)	0.670970	1.0	661.288637	4.962951e-140
I(Current ** 3)	0.357867	1.0	352.702970	5.437027e-77
Residual	7.884768	7771.0	NaN	NaN

Figure 4.4: ANOVA summary for models from Equation 4.2 and 4.4

4.5 Phase 4: One Device, One Operator, One Channel

The final phase of the model development is to consider the lowest level dataset that is representative of what measurements taken in a manufacturing setting. For instance, were these laser systems to only be tested once, how would a model be developed for each channel? This section will explore the models for two laser channels, Channel 1, and Channel 10, on a single device (Device 1).The log transformation is used as the transformation of choice given the better performance compared to the square root transformation that was explored in previous sections. For simplificity, only comparisons with the log transformation are included in this section.

4.5.1 Device 1, Channel 1

An OLS model for the un-transformed response was first determined. The full model with all the second order temperature, third order current, and first order interactions was the starting point for model development. Equation 4.5 defines the full model where y refers

to output power, x_1 refers to temperature, x_2 refers to current, the β terms refer to the respective coefficients (β_1 to β_6) and intercept (β_0). This full model resulted in an R^2 value of 94.2% and and AIC of -5612.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_2^3 \tag{4.5}$$

A simplified model after removing non-significant terms is shown in Equation 4.6 which only includes the first order temperature factor, third order current factor, and no interaction terms. This model yielded an R^2 value of 94.0% and and AIC of -5570.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \beta_4 x_2^3 \tag{4.6}$$

As seen from the R^2 and AIC statistics, the simplified model has a 0.2% reduction in the R^2 value and a 0.7% increase in AIC. Given that the simplified model has no interaction terms and fewer factors while sacrificing marginal performance in the R^2 and AIC terms, the simplified model will be the first model of comparison using the untransformed power response variable. This simplified model thus has the coefficients as indicated in Equation 4.7.

$$y = 0.0268 + -0.00268 x_1 + 0.00112 x_2 + [5.078\text{e}{-6}] x_2^2 + [-1.049\text{e}{-8}] x_2^3 \tag{4.7}$$

A model using the form from Equation 4.5 where y is the log transformation of power was also developed and yielded an R^2 of 96.8% and AIC of -3234. A simplified model was also developed and follows the form in Equation 4.6. Simplified models for both the untransformed and transformed responses include the 3rd order current term and no interaction terms. The simplified model for the transformed, simplified model yielded an R^2 of 96.7% and AIC of -3191.

The change in the R^2 and AIC statistics between the simplified and full model for the transformed response was also 0.2% reduction in the R^2 and 0.7% increase in AIC. Similar

to the model comparisons for the untransformed response, the simplified model from 4.6 is the selected model given the comparable model performance and reduction of factors in the model. As a result, the simplified model of the transformed response has the coefficients shown in Equation 4.8, where y is the log transformation of the output power.

$$y = -2.9439 - 0.0075x_1 + 0.0371x_2 - 0.0002x_2^2 + [3.295\text{e}{-7}]x_2^3 \tag{4.8}$$

The model development for both the untransformed and transformed response were repeated using an 80-20 test-train split. The performance of the fits for the untransformed response are demonstrated in Figure 4.5 where the left figure shows the fitted points in yellow overlaid with the actual points in green. The shape of the residuals in the right figure has a shape such that the variance in the residuals is increasing with the fitted power, again suggesting that the assumption in homoskedasticity is not met.The performance of the fits for the transformed response are shown in Figure 4.6. Comparing the residuals from Figure 4.5 and Figure 4.6, it can be observed that the residuals of the transformed response have a less obvious trend than observed in the untransformed response's residuals. This suggests that the assumption on homoskedasticity of the residuals is better addressed with the transformed response except when the fitted value is small, which corresponds to the cases where the power output is small and hard to measure accurately.

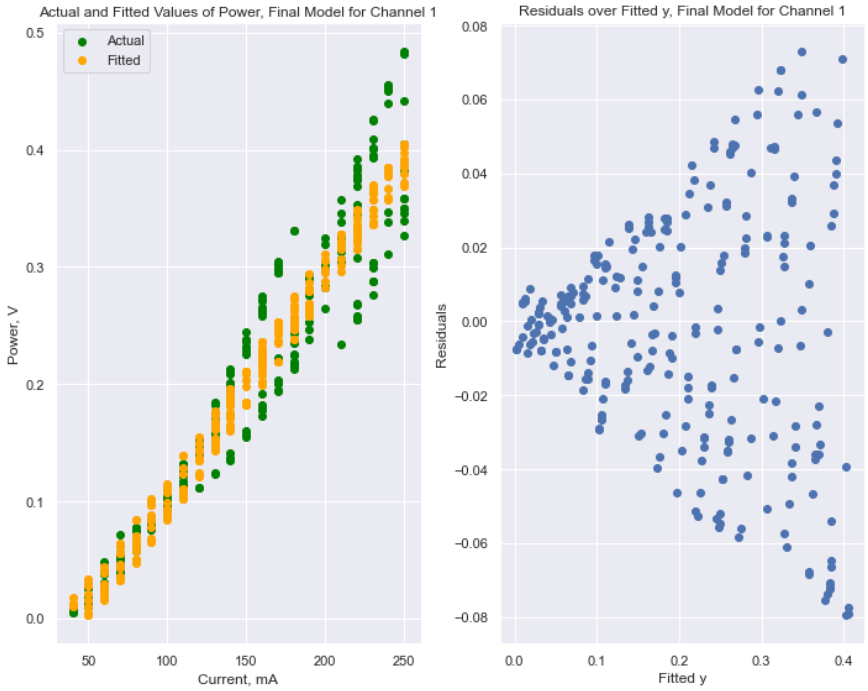

Figure 4.5: Summary of fitted power for Channel 1, untransformed response

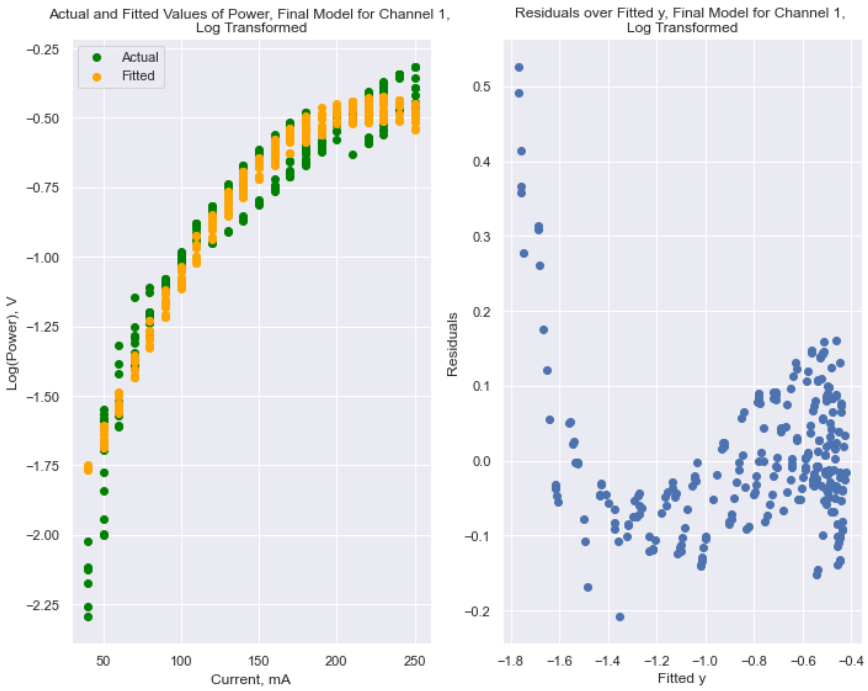

Figure 4.6: Summary of fitted power for Channel 1, transformed response

4.5.2 Device 1, Channel 10

It has also been shown that the power output is dependent on the individual channels. To explore the difference in model development per channel, Channel 10 was used to generate fits to compare to the fits from Channel 1 on the same device. The methodology from Section 4.5.1 was repeated for Channel 10, an arbitrary channel for comparison. The simplified model for the untransformed response was found to have the same model parameters as the simplified model from Channel 1 (Equation 4.6). The coefficients for this model are shown in Equation 4.9. This model has an R^2 of 93.2% and AIC of -5125. Comparing this model to the respective model for Channel 1, it is observed that the R^2 is slightly reduced from 94.0%.

$$y = 0.0165 + -0.0026x_1 + 0.0013x_2 + [4.334\text{e}{-6}]x_2^2 + [-7.978\text{e}{-9}]x_2^3 \tag{4.9}$$

Continued model development using the transformed response also resulted in the same model parameters as that found from Channel 1 (Equation 4.6). The coefficients for this model are included in Equation 4.10 where y refers to the log transformation of the power output. This model has an R^2 of 96.2% and AIC of -849.3. Similar to the findings from the models for Channel 1, the transformed response yields better performing R^2 and AIC statistics.

$$y = -7.0040 - 0.0183x_1 + 0.0906x_2 - 0.0005x_2^2 + [8.143\text{e}{-7}]x_2^3 \tag{4.10}$$

Reviewing the final models for Channel 1 and Channel 10 on the same device illustrates how models can be developed per channel on the same device. For instance, the two example channels show that the transformed response produces the best R^2 and AIC statistics for model goodness of fit. The two channels also resulted in simplified models that had the same base parameters. The differences in the models are driven by the coefficients, which can provide a stepping stone for the same base model for all of the channels.

4.5.3 Device 3, Channel 10

The final model for comparison is taking the same channel on a different device to show the device-to-device needs for model development. Device 3 was selected as an arbitrary device of comparison, and Channel 10 was identified as the channel of interest. The same model development steps in Section 4.5.2 were followed and the simplified model of the transformed response shared the same model parameters from Equation 4.6. This model resulted in an R^2 of 96.1% and AIC of -801.3. The coefficients for this model are included in Equation 4.11 where y refers to the log transformation of the power output.

$$y = -5.5580 - 0.0208x_1 + 0.0742x_2 - 0.0004x_2^2 + [6.556\text{e}{-7}]x_2^3 \tag{4.11}$$

Inspection of the model coefficients from Equation 4.10 and Equation 4.11 results in the apparent difference in the intercept value. Device 1's model has an intercept of -7.0040 while Device 3's model has an intercept of -5.5580. The difference in the measured data is shown in Figure 4.7. The measured power on Device 3 (green curve) has peaks that are higher than that of Device 1 (blue curve) which is consistent with the larger intercept. From this comparison, it is evident that the local maxima in the power curves contributes to the intercept of the coefficient and the overall model.

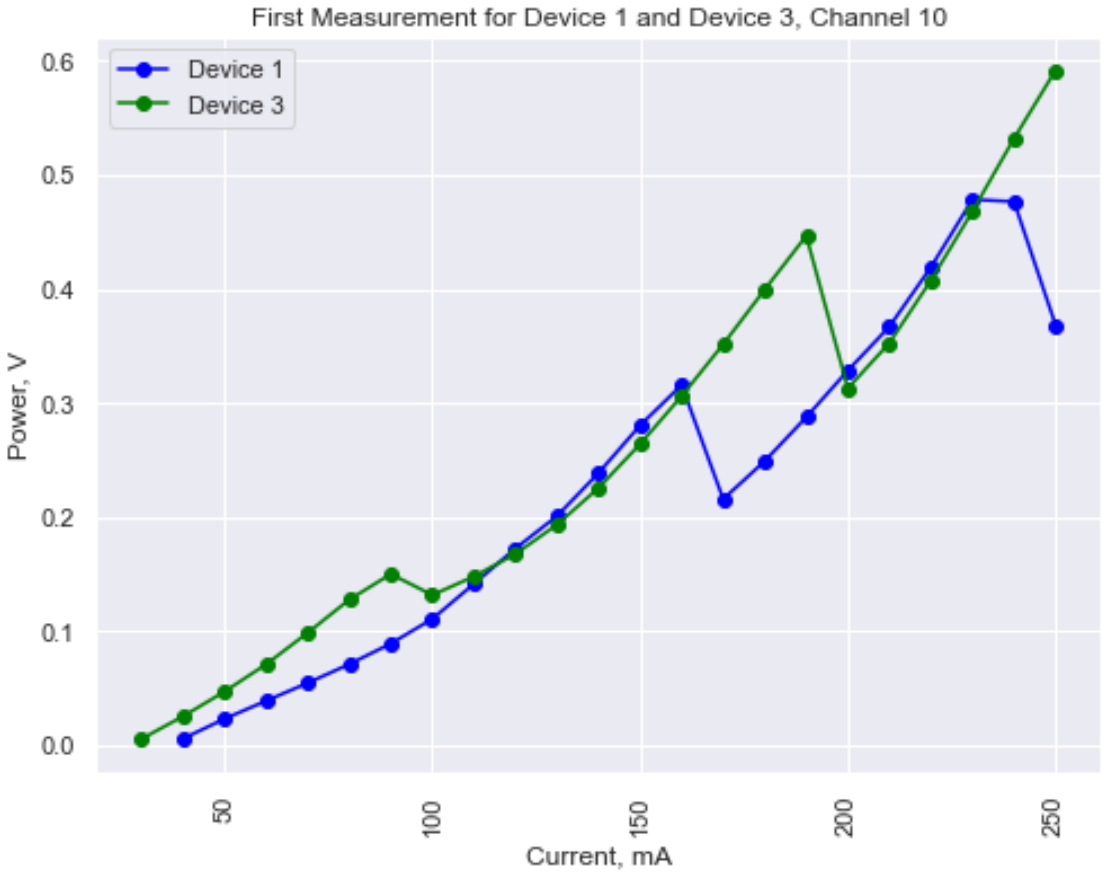

Figure 4.7: Comparison of power predictions for Device 1 and Device 3

4.5.4 Summary

This section dove into three specific examples that demonstrated three main takeaways when generating models on an individual device and channel:

- Transforming the response variable for model development yielded overall better R^2 and AIC compared to the untransformed models
- The final model parameters were the same (first order temperature, third order current, and no interaction terms) for all three examples

The presence of local maxima in the current sweeps (Figure 4.7) may have a significant effect

on the overall model.

CHAPTER 5

Conclusion

The objective of this paper was to introduce the complexity of data collection for a novel laser test system and propose a framework to generate models for power prediction that could be used in a manufacturing setting. The proposed framework started with the largest aggregate dataset including all of the model parameters such as current, temperature, operator, device, and repeated measurement. A stepwise approach was taken to determine that devices and channels are significant factors, which are expected due to the nature of the prototyping process with inherent device variability and the system design of various lasing wavelengths. In contrast, it was determined that the operator and test repeat parameters were not important, suggesting that the test system is robust enough for test repeatability. The final phase of the model development explored three examples for an individual laser on a selected device. Comparison of the final models determined that the log transformation of the response resulted in the best models, which all shared the same model parameters. This finding provides a basis for generating the same base model for each laser, which can be implemented in larger test systems that use the power models for further laser system development.

REFERENCES

Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron. 2018;4(4):195-202. doi: 10.15406/ijbsbe.2018.04.00125.

Rockley Photonics. Photonics-Based Measurement of Core Body Temperature. Rockley Photonics. 2021. https://rockleyphotonics.com/wp-content/uploads/2021/12/Rockley-Photonics-Photonics-Based-Measurement-of-Core-Body-Temperature.pdf.

F. J. Wu and M. S. Hamada (2009). Experiments: Planning, Analysis and Optimization, 2nd ed., Wiley.

D. C. Montgomery (2005-2019). Design and Analysis of Experiments, 6-10th ed., Wiley.

J. Faraway (2005). "Linear Models with R," Chapman Hall.

S. Sheather (2009) "A Modern Approach to Regression with R", Springer.