



SCHOOL of  
GRADUATE STUDIES  
EAST TENNESSEE STATE UNIVERSITY

East Tennessee State University  
Digital Commons @ East  
Tennessee State University

---

Electronic Theses and Dissertations

Student Works

---

5-2008

# Methods for the Analysis of Developmental Respiration Patterns.

Justin Tyler Peyton

*East Tennessee State University*

Follow this and additional works at: <https://dc.etsu.edu/etd>



Part of the [Longitudinal Data Analysis and Time Series Commons](#)

---

## Recommended Citation

Peyton, Justin Tyler, "Methods for the Analysis of Developmental Respiration Patterns." (2008). *Electronic Theses and Dissertations*. Paper 1931. <https://dc.etsu.edu/etd/1931>

This Thesis - Open Access is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact [digilib@etsu.edu](mailto:digilib@etsu.edu).

Methods for the Analysis of Developmental Respiration Patterns

---

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

---

by

Justin Peyton

May 2008

---

Edith Seier, Ph.D., Chair

Karl Joplin, Ph.D.

Robert Gardner, Ph.D.

Keywords: time series, programs, change points, *Sarcophaga crassipalpis*.

## ABSTRACT

Methods for the Analysis of Developmental Respiration Patterns

by

Justin Peyton

This thesis looks at the problem of developmental respiration in *Sarcophaga crassipalpis* Macquart from the biological and instrumental points of view and adapts mathematical and statistical tools in order to analyze the data gathered. The biological motivation and current state of research is given as well as instrumental considerations and problems in the measurement of carbon dioxide production. A wide set of mathematical and statistical tools are used to analyze the time series produced in the laboratory. The objective is to assemble a methodology for the production and analysis of data that can be used in further developmental respiration research.

Copyright by  
Justin Peyton 2008

## DEDICATION

To Jess: A Refreshing Fall Breeze

## ACKNOWLEDGMENTS

The author would like to acknowledge that this thesis would not have been possible without the significant contribution of the following people: Dr. Darrell Moore, Dr. Lyndell Kerley, Dr. Jeff Knisley and the committee members. The author would like to extend special thanks to Dr. Richard Lee and Miami University for an equipment loan. Partial support was provided through Howard Hughes Medical Institute grant #52005872.

## CONTENTS

ABSTRACT . . . . .	2
DEDICATION . . . . .	4
ACKNOWLEDGMENTS . . . . .	5
LIST OF TABLES . . . . .	8
LIST OF FIGURES . . . . .	11
1 INTRODUCTION . . . . .	12
2 DEVELOPMENTAL RESPIRATION . . . . .	14
2.1 Lifecycle of <i>Sarcophaga crassipalpis</i> . . . . .	14
2.2 Respiration as a Measure of Physiological Change . . . . .	15
2.3 Biological Goals . . . . .	16
3 INSTRUMENTATION . . . . .	17
3.1 Instrumental Considerations . . . . .	17
3.2 Materials and Methods . . . . .	18
3.3 Technical Concerns . . . . .	19
3.3.1 Negative Control . . . . .	19
3.3.2 Flow Rate . . . . .	19
3.3.3 Caustic Chemicals . . . . .	20
4 DEVELOPMENT OF MATHEMATICAL AND STATISTICAL TOOLS	22
4.1 Objectives . . . . .	22
4.2 Negative Control Data . . . . .	23
4.3 Respiration Data . . . . .	26
4.3.1 Identifying Trend . . . . .	27

4.3.2	Identifying Features in Trend . . . . .	31
4.3.3	Variability . . . . .	36
4.3.4	Periodicity . . . . .	38
5	RESULTS . . . . .	50
	BIBLIOGRAPHY . . . . .	55
	APPENDICES . . . . .	58
	Appendix A . . . . .	58
	Individuals . . . . .	58
	Splines . . . . .	65
	Spline Residuals . . . . .	72
	MWLR . . . . .	79
	Spectrograms . . . . .	86
	Appendix B . . . . .	93
	SplinerD . . . . .	93
	MWLR . . . . .	94
	MaxT and MinT . . . . .	96
	ZeroI . . . . .	98
	spec . . . . .	99
	Smax . . . . .	101
	per . . . . .	103
	VITA . . . . .	105



## LIST OF TABLES

1	Change Points . . . . .	50
2	Segment Lengths . . . . .	51
3	Segment Length Statistics . . . . .	52
4	Segment Conclusion . . . . .	53
5	Segment Conclusion Statistics . . . . .	54

## LIST OF FIGURES

1	Puparial Development . . . . .	14
2	Flow Control . . . . .	20
3	Blank and Individual 6 . . . . .	23
4	Blank Residuals and Autocorrelation . . . . .	24
5	Finite Differences and Autocorrelation . . . . .	25
6	Individual 6 . . . . .	27
7	Moving Average and Residuals . . . . .	28
8	$A(t)$ . . . . .	29
9	Spline and Residuals . . . . .	30
10	Smoothing Capabilities Comparison . . . . .	31
11	Excerpt of Individual 6 . . . . .	32
12	Spline Derivate and MWLR Comparison . . . . .	33
13	Exponential Smoothing Error . . . . .	34
14	Derivative Estimation . . . . .	36
15	Variability Threshold . . . . .	37
16	Population Exploration . . . . .	37
17	Example Periodogram . . . . .	39
18	Rectangular Window . . . . .	41
19	Hamming Window . . . . .	43
20	Blackman Window . . . . .	44
21	Spectrogram and Periodogram Comparison . . . . .	45
22	Blank Spectrogram . . . . .	47

23	Individual 6 Spectrogram . . . . .	48
24	Program Output . . . . .	49
25	Segment Lengths . . . . .	53
26	Individual 1 . . . . .	58
27	Individual 2 . . . . .	59
28	Individual 3 . . . . .	60
29	Individual 4 . . . . .	61
30	Individual 5 . . . . .	62
31	Individual 6 . . . . .	63
32	Individual 7 . . . . .	64
33	Spline of Individual 1 . . . . .	65
34	Spline of Individual 2 . . . . .	66
35	Spline of Individual 3 . . . . .	67
36	Spline of Individual 4 . . . . .	68
37	Spline of Individual 5 . . . . .	69
38	Spline of Individual 6 . . . . .	70
39	Spline of Individual 7 . . . . .	71
40	Residuals of the Spline of Individual 1 . . . . .	72
41	Residuals of the Spline of Individual 2 . . . . .	73
42	Residuals of the Spline of Individual 3 . . . . .	74
43	Residuals of the Spline of Individual 4 . . . . .	75
44	Residuals of the Spline of Individual 5 . . . . .	76
45	Residuals of the Spline of Individual 6 . . . . .	77

46	Residuals of the Spline of Individual 7 . . . . .	78
47	MWLR of Individual 1 . . . . .	79
48	MWLR of Individual 2 . . . . .	80
49	MWLR of Individual 3 . . . . .	81
50	MWLR of Individual 4 . . . . .	82
51	MWLR of Individual 5 . . . . .	83
52	MWLR of Individual 6 . . . . .	84
53	MWLR of Individual 7 . . . . .	85
54	Spectrogram of Individual 1 . . . . .	86
55	Spectrogram of Individual 2 . . . . .	87
56	Spectrogram of Individual 3 . . . . .	88
57	Spectrogram of Individual 4 . . . . .	89
58	Spectrogram of Individual 5 . . . . .	90
59	Spectrogram of Individual 6 . . . . .	91
60	Spectrogram of Individual 7 . . . . .	92

## 1 INTRODUCTION

This thesis looks at the problem of developmental respiration in *Sarcophaga crassipalpis* Macquart from the biological and instrumental points of view and adapts mathematical and statistical tools in order to analyze the data gathered. The biological motivation and current state of research is given as well as instrumental considerations and problems in the measurement of carbon dioxide production. A wide set of mathematical and statistical tools were used to analyze the time series produced in the laboratory. The objective was to assemble a methodology for the production and analysis of data that can be used in further developmental respiration research.

Chapter 2, which is devoted to developmental respiration, gives a brief introduction to the life cycle of the model species *S. crassipalpis*, with emphasis on the larval to pupal development. An overview of insect respiratory measurement and research is covered as well as the basic biological goals motivating the project.

Chapter 3 covers the instrumental concerns and considerations of the project. The basic idea behind how certain kinds of respirometry systems differ is discussed. The materials and methods of the data collection are covered as well as a few technical problems encountered including flow rate control and caustic chemical handling.

The development of mathematical tools for the segmentation of the time series produced in the laboratory is covered in Chapter 4. Tools were developed in order to detect changes in the trend, variability, and periodicity. Many of the tools developed come in the form of programs written in Maple and are included in Appendix B. Special consideration was given to negative control data to ensure that the measure-

ments and change points described were not instrumental artifacts. The conversion from data produced in the laboratory to the dataset that was actually analyzed is also discussed.

The final chapter presents the results when the tools developed in Chapter 4 are applied to the experimental data. Biological implications of this thesis are also examined.

## 2 DEVELOPMENTAL RESPIRATION

### 2.1 Lifecycle of *Sarcophaga crassipalpis*

Holometabolous insects undergo a complex transition during their life cycle, collectively described as a complete metamorphosis. After embryonic development, larvae go through a series of molts, the last larval stage enters a pupal stage and the adult emerges, becomes sexually active and reproduces [14]. In Dipteran species and some other holometabolous insects, the last larval stage undergoes a process where the larval epidermis smooths and then sclerotizes to form the puparial case (Figure 1). Inside this case, the pupal cuticle is produced and pupal apolysis occurs [6].



Figure 1: Pupal Development: wandering larvae (left), and successive puparial development (from left to right) showing the barreling and tanning process.

The life cycle of the flesh fly, *Sarcophaga crassipalpis* Macquart, consists of the female fly depositing pharate larvae directly onto a food source where the larvae feed and undergo 3 larval molts. Late in the third larval instar, the larvae leave the food source and migrate to find a pupariation site (wandering larvae, WL). During pupariation, the larvae become immobile and barrel, with the epidermis contracting into a smooth structure. Over a period of hours the larval cuticle hardens and sclerotizes

to become the puparial case (Figure 1). The hard brittle puparial case is opaque and obscures developmental processes from visual examination. The developing fly does not leave the puparial case until it emerges as a fully developed adult.

Several major developmental milestones occur between pupariation and eclosion including larval-pupal apolysis, head eversion, and pupal-adult apolysis. Larval-pupal apolysis corresponds to the transition from the encapsulated larva to a cryptocephalic or head hidden pupa. The cryptocephalic pupa is similar to the larva in that its head is still inverted and the appendages are still undifferentiated. These features change around the time of head eversion as the imaginal discs evaginate and expand to form their final pupal shapes. Pupal-adult apolysis corresponds to the transition from pupa to pharate adult [6]. Although the general outline of these processes are known, there may be many complex changes that are still unknown. One of the processes that has not received adequate attention is the development of respiratory patterns associated with these changes.

## 2.2 Respiration as a Measure of Physiological Change

The insect respiratory system consists of a series of branching tubes (trachea) that deliver oxygen to individual cells and tissues and carry waste gases, carbon dioxide and water, from the cells and out of the body [4]. This kind of respiratory system precludes the need for a complex circulatory system with an oxygen or carbon dioxide carrying pigment, such as hemoglobin in vertebrates, but has issues associated with a high surface area to volume ratio. Water loss is a severe problem in these small terrestrial organisms. Insects have addressed this problem by using spiracles to limit



the flow of gas and water loss.

Respiration is reportedly controlled by several methods in insects ranging from passive diffusive flow [8, 11], discontinuous gas exchange (DGC) [9], the coelopulse [17], and changing tracheal volume [21]. The time scale associated with these physiological functions range from hours (DGC) to less than 1 sec (tracheal volume change). Although there has been no consensus about physiological control mechanisms, the indications are that the processes are complex and are rich areas for further studies.

Respirometry could be a valuable tool to gain insights into physiological and developmental changes given the development of this organism. It is also important to note that, since a developing fly is non-motile, no adjustments or exceptions need to be made for activities such as eating or crawling; all changes in respirometry data not attributed to the instrument itself are assumed to be developmental in nature.

### 2.3 Biological Goals

The biological goals of this thesis involve using a noninvasive technique as a window for examining the physiological and developmental processes that occur during the time after pupariation. Morphological staging can only determine that the larval epidermis sclerotizes during this time and examines snapshots of development by histological sectioning. Due to the largely descriptive nature of this approach, it is difficult to delineate where developmental phases begin and end while also measuring variation in individuals. This study is the first description of the complex changes that can be seen using respiration, and presumably reflects undescribed developmental and physiological processes.

## 3 INSTRUMENTATION

### 3.1 Instrumental Considerations

Considerations when choosing a respirometry system include resolution, sensitivity, and stability. Early attempts to measure respiration used closed-system manometer measurements, such as adaptations of a Gilson respirometer [10, 5]. This type of respirometry system has limitations on sensitivity and resolution when used on insects. This method depends on measurements of changes in gas volume due to oxygen consumption. Given the low mass of insects, resolutions ranging from a single measurement of oxygen consumption per hour to one measurement per day are not uncommon [5].

The advent of flow-through or open-system respirometry for measuring carbon dioxide release allows for much higher resolution measurements. This type of system depends on the absorption of infrared light by carbon dioxide and can be sampled several times a second. This technological advance theoretically allows researchers to examine small scale changes in respiration that would reflect the physiological changes associated with development, but comes with some cost. To increase the fidelity of the system, carbon dioxide and water are chemically removed from the air stream which requires the use of caustic chemicals such as sodium hydroxide and magnesium perchlorate. The measurement of oxygen consumption cannot be measured with the same resolution and sensitivity as carbon dioxide production in a flow-through system. This is because oxygen consumption is measuring a small depletion relative to the atmospheric concentration of oxygen, while carbon dioxide production measurements

are measuring deviations from near zero concentrations [10].

### 3.2 Materials and Methods

Newly pupariated larvae within a half hour of barreling, consistent with the first and second pupariation stages seen in Figure 1, were placed into a sample chamber of a flow-through respirometry system. The air for the respirometry system was drawn out of an unsealed twenty liter carboy, which buffered the system from large changes in ambient carbon dioxide concentration. The air was drawn through a one liter Dryrite-sodalime-Dryrite column to chemically scrub out carbon dioxide and water. After passing through the pump the air stream was pushed through a smaller 50 milliliter secondary Dryrite-Ascerite-Dryrite column to further scrub water and carbon dioxide that can enter the system through the pump's semi-permeable parts. The airstream then passed through a Sierra mass-flow driver (Sierra Instruments, Monterey, CA) which controls the system's flow rate by automatically adjusting for standard temperature and pressure. The air stream then passed through the reference cell of the LI-COR 6251 differential carbon dioxide analyzer (LI-COR Biosciences, Lincoln, NE) and into the sample chamber inside a Percival environmental chamber (Percival Scientific, Perry, IA). The airstream was then passed through a column of magnesium perchlorate, to eliminate any water released by the organism. The air then flowed through the sample cell of the LI-COR and back into the carboy. The incubation chamber is set to twenty-five degrees centigrade with a photoperiod of 15:9 L:D. Quarter inch external diameter Bev-A-Line V tubing (Thermoplastic Processes, Stirling, NJ) was used to connect the components of this system. Every

time the air stream enters a piece of equipment a filter has been installed to protect the instrumentation from dust and chemicals. Signals from the mass flow controller and the LI-COR were acquired using a Sable Systems UI-2 interfacing unit and Expedata software (Sable Systems, Las Vegas, NV). Data was exported from Expedata for analysis in Maple and MINITAB.

Larvae were reared in a laboratory colony at twenty-five degrees centigrade with a photoperiod of 15:9 L:D. The colony has been maintained at East Tennessee State University by Dr. Karl Joplin since 1994 when it was established from a long term colony maintained at The Ohio State University.

### 3.3 Technical Concerns

#### 3.3.1 Negative Control

Frequent measurements of very low changes in gas concentration from an approximately 150 mg organism over a period of days make extreme demands on the stability of the instrumentation. At this low level, the instrumentation is operating near the limits of the technology and these limitations must be verified to yield a reliable source of information. Consequently a negative control or blank sample was recorded and analyzed to determine properties of the data that can be attributed to the machine.

#### 3.3.2 Flow Rate

One of the more difficult problems associated with open or flow-through respirometry is control of the flow rate. Fluctuations in flow rate can have effects on the measurement of carbon dioxide. First, fluctuations in flow rate cause the air to pass

more quickly or slowly over the sample, resulting in a more diluted or concentrated air stream respectively. This effect is accounted for when the actual time series is constructed from the flow rate and concentration streams of data (see section 4.3). Large cyclic fluctuations in flow can, however, cause noticeable artifacts in the final time series. Many samples had to be regarded as preliminary data until this problem was consistently controlled. An example of poor flow and the kind of artifact generated by poor flow control can be seen in Figure 2.

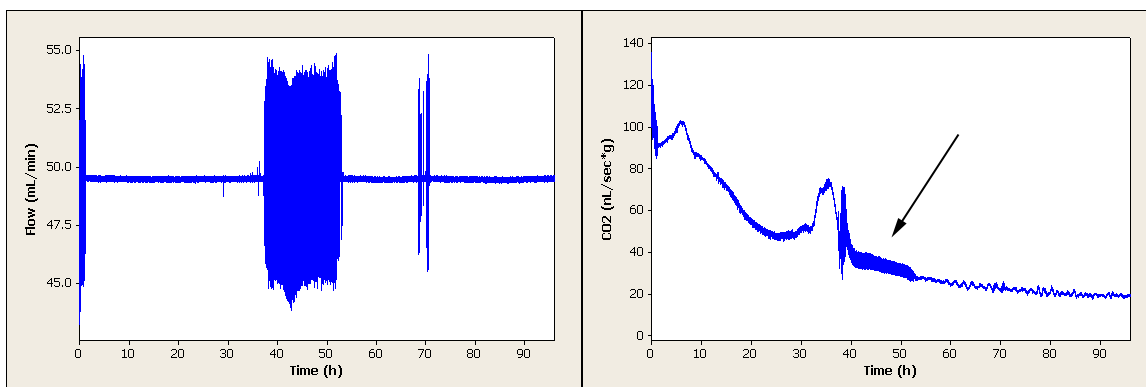


Figure 2: Flow Control: preliminary data with poor flow control (A), and artifact (Arrow) created (B).

### 3.3.3 Caustic Chemicals

It is very important to maintain proper control of the caustic chemicals used to chemically scrub the air. Improper care when handling these chemicals cannot only lead to wasted supplies, but also damage to the instrumentation and cause physical harm to lab personnel. While Dryrite is innocuous enough, Ascerite and magnesium perchlorate are highly corrosive and should be handled with care. Ascerite, which is mostly sodium hydroxide, has the added problem of being hygroscopic so that if

exposed to too much water vapor it will turn into a corrosive sludge. Insuring that chemical dust is properly contained is another important concern. Filters should be placed into the air stream so that any chemical dust does not find its way into the instrumentation.

## 4 DEVELOPMENT OF MATHEMATICAL AND STATISTICAL TOOLS

### 4.1 Objectives

One of the major objectives of this thesis is to adapt and develop analytical procedures to identify and characterize features of this and similar time series. The main features of interest to be identified and defined in the time series are long term trend, short term variability, periodicity; and identification of the moments in which these three characteristics undergo change.

Some of these characteristics and turning points might be apparent to the naked eye but a major goal of this work attempts to define these points objectively and reduce the number of subjective decisions in the identification of these features. In addition, some of these points might not be apparent, but are hidden by the prevalence of varying degrees of noise. One way to avoid subjective decisions is to develop analytical tools that clearly identify features. These tools will also enable others to easily reproduce the analysis on similar time series. The development of these analytical procedures will be described in this section and programs for analyzing this data in Maple will be presented in Appendix B.

In order to explore these time series, a single individual was chosen by dice throw as a representative. Any exploration needed, as in section 4.3, was done on this time series and decisions based on it will then applied to the other time series. In order to minimize in-text graphs, most of the graphs of experimental data are of this individual. The interested reader is pointed to Appendix A where graphs are presented for all the individuals in a larger format. An eight sided die was used to select the

representative sample randomly. The first throw showed 8, so the die was rerolled. On the second throw it showed 6 indicating individual six as the representative.

## 4.2 Negative Control Data

Before the respiration data can be properly understood, it is important to establish what kind of data the respirometry system can produce with no organism present. A negative control or blank is a sample taken under the exact same conditions, but with no organism in the sample chamber. Figure 3 shows the blank and a sample in the same units and scale.

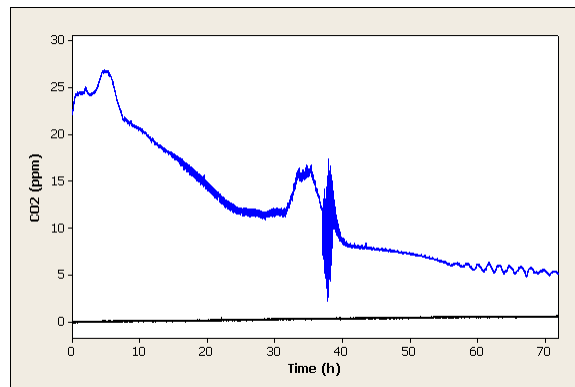


Figure 3: Blank and Individual 6: Blank (black) and Individual 6 (blue).

From this view the blank seems to be a straight line that slowly moves from zero parts per million to around 0.6 parts per million. One strategy is to fit a line over the blank and subtract that line from the blank and all other readings to correct for instrument drift. This however may not be the best method. If a line was subtracted from the blank some structure would still be left. Figure 4A shows the residuals from a linear regression or what remains of the blank after a best fit line has been



subtracted. Figure 4B shows the first 150 coefficients of the serial autocorrelation function of the residuals.

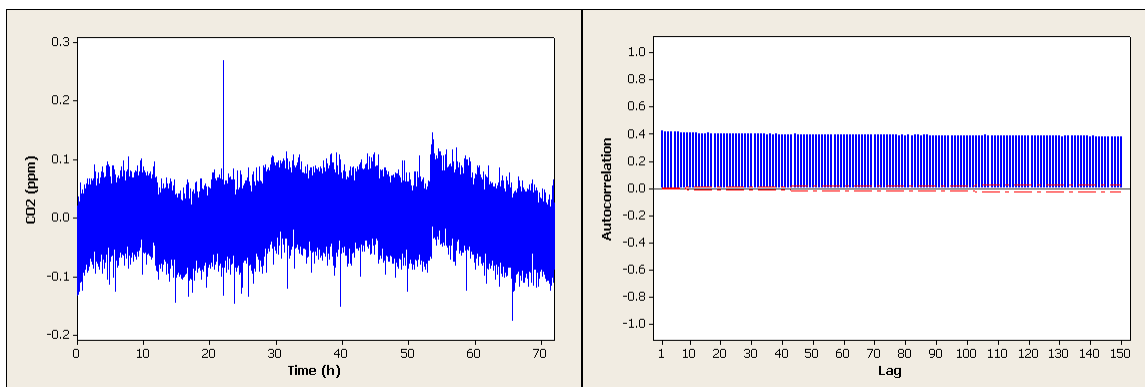


Figure 4: Blank Residuals and Autocorrelation: residuals of a linear regression of the blank (A) and the first 150 autocorrelation coefficients (B).

The autocorrelation coefficients show that the new time series is not stationary. Subtracting a line representing machine drift from all the time series would not remove all the structure induced by the machine, and it is unknown what structure might be induced by subtracting it. It is better to look at the finite differences of the original time series.

Figure 5A shows  $\nabla y_t$ , the finite difference of lag one computed from the blank data and Figure 5B displays the serial autocorrelation function. Figure 5B shows a single significant autocorrelation  $r_1 = 0.497585$ . This suggests that an ARIMA(0,1,1) model seen in Equation 1 is appropriate to describe the random process of the empty respirometry system [2].

$$Y_t = Y_{t-1} + a_t - \theta * a_{t-1} \quad (1)$$

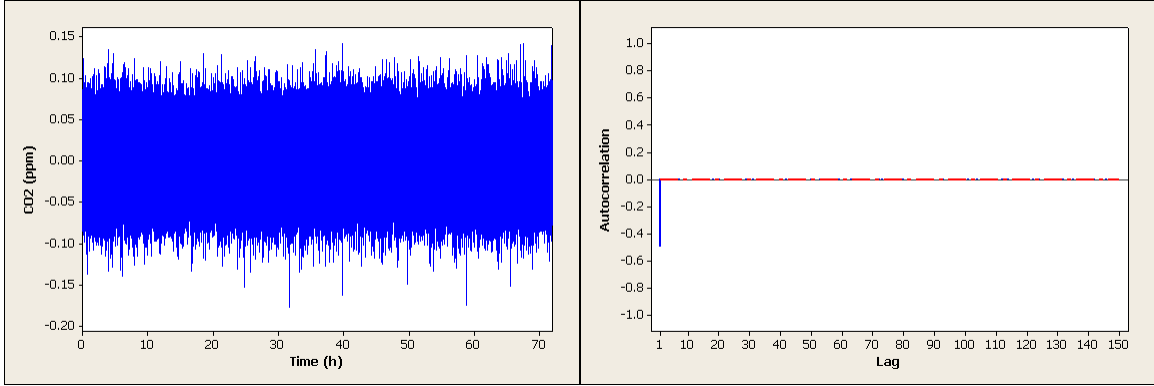


Figure 5: Finite Differences and Autocorrelation: finite differences of lag one of the blank (A) and the first 150 autocorrelation coefficients (B).

The first serial autocorrelation  $r_1$  can be used as an initial estimator of the autocorrelation of the process,  $\rho_1$ . An ARIMA(0,1,1) process has one parameter  $\theta$  and it is related to  $\rho$  by Equation 2.

$$\rho = \frac{-\theta}{1 + \theta^2} \quad (2)$$

If the first autocorrelation of the process was  $\rho = 0.497585$ , then the parameter of the model would be  $\theta = 0.9062104487$ . However a better estimation of the model can be achieved by applying maximum likelihood estimation. There is no closed form expression for this estimation; the estimators have to be obtained by numerical approximation. This was done using SAS and the estimated model can be seen in Equation 3.

$$Y_t = Y_{t-1} + a_t - 0.97313 * a_{t-1} \quad (3)$$

An interpretation of this model is that the machine has white noise error, but about three percent of this error accumulates producing something similar to a random walk. A  $\theta$  value of one would have no accumulation and would be white noise. A  $\theta$  value of zero would correspond to a true random walk.

The Fourier transform can be used to explore another aspect of the blank data, but will be covered in Subsection 4.3.3.

### 4.3 Respiration Data

Larvae were selected from the colony based upon the most recently puparited larva at the time of recording. All larvae were immobilized, had red spiracles and were only slightly tanned consistent with the second larva from the left in Figure 1. Expedata was set up to sample the carbon dioxide and flow signals sixty times a second and then average those sixty values and record that value for that second. The system was set up to record for three days producing two streams of 259,200 time points.

The first step in the analytical process was to convert the two streams of data into biologically relevant units. Each data point must be converted using the Equation 4.

$$S_t = C_t * \frac{1 \text{ part } CO_2}{1000000 \text{ total parts}} * F_t * \frac{1000000 \text{ nanoliter}}{1 \text{ milliliter}} * \frac{1 \text{ min}}{60 \text{ sec}} * \frac{1}{m} \quad (4)$$

Where  $C_t$  is carbon dioxide measured in parts per million from the LI-COR and  $F_t$  is the flow measured in milliliters per minute by the mass flow controller at time  $t$ . The mass in grams of the individual pupa at time of insertion is represented by  $m$ .

The other factors are necessary for unit conversion. The process produces a single stream of 259200 time points with units of nanoliter/second\*gram (Figure 6).

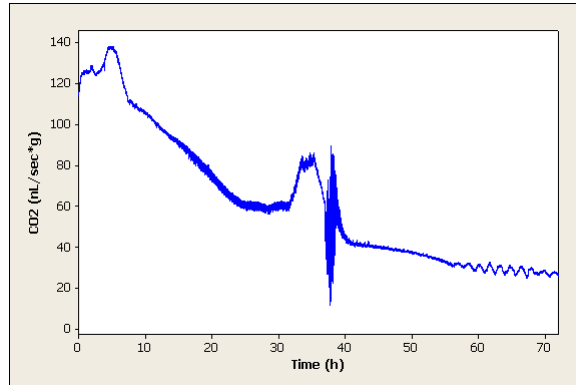


Figure 6: Individual 6

#### 4.3.1 Identifying Trend

The trend is the long term behavior of a variable through time. If the behavior of the variable through time is such that it can be described as random variations around a mean, the process is said to be stationary and the trend would be constant. If the trend is very consistent it may be described with a polynomial or non-linear function of time using regression methods. However, the trend can be quite irregular and a regression model would make little sense. In this situation it is preferable to apply a smoothing method to the time series in order to be able to describe the trend. After the trend is identified and represented either by a function of time or numerically by a smoothed version of the series, the trend can be removed from the original time series as to observe any other patterns that the variable might exhibit.

It is clear from Figure 6 that this developmental time series is non-stationary with

an irregular trend. A commonly used technique for smoothing a time series with an irregular trend is moving averages. In a moving average the time series is smoothed by averaging each point by some of its neighbors. The longer the moving average length, the more neighbors are included and the smoother the result. Equation 5 gives the formula for a centered moving average of length L.

$$S_{j+\frac{L+1}{2}} = \frac{1}{L} * \sum_{k=j}^{i+L-1} a_k \quad (5)$$

There can be negative impacts if the moving average length is too long. First the time series is shortened by the moving average length. Shortening the time series is not a problem in this case because of its length. Another negative effect is seen in the residuals. If the moving average length is too long, the residuals can have a trend. This trend takes the form of long sequences of residuals with a mean above or below zero. For example, Figure 7A shows a very desirable smoothed time series generated using a moving average length of 5000, but Figure 7B shows how the residuals can

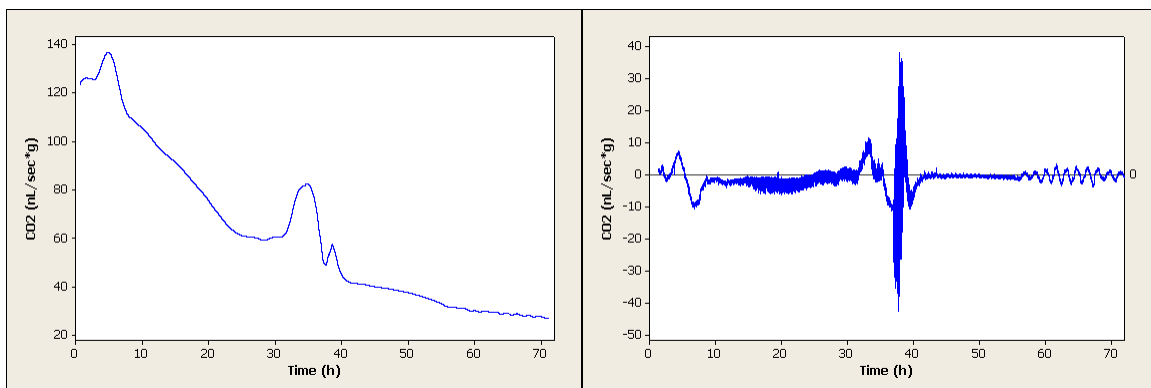


Figure 7: Moving Average and Residuals: smoothed time series using moving average of length 5000 of Individual 6 (A) and the residuals (B).

deviate from zero for long periods of time. This is because the moving average is calculated with respect to values too far away in time and no longer represents accurately the local behavior. This effect is not desirable because the residuals will be used later in the analysis to explore variability and periodicity.

Another option is to use block averaging in conjunction with cubic spline interpolation. In this method, large sections of the data are averaged together and assigned to the average of these times creating  $A(t)$ , a mapping defined only over the averaged times. In this example, the first 900 data points are averaged together and assigned to  $t_1 = 450.5$  sec and the next point is not defined until  $t_2 = 1350.5$  sec.  $A(t_1)$  is about equal to 120 (nL/sec\*g) as seen in Figure 8.

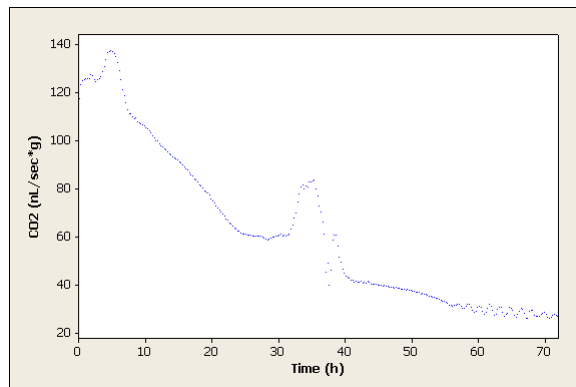


Figure 8:  $A(t)$ : of Individual 6 generated by block averaging of length 900.

Once  $A(t)$  is constructed a piecewise defined continuous cubic polynomial  $S(t)$  is constructed that goes through all the points of  $A(t)$  (Figure 9). Each piece  $S_i(t)$  of this polynomial is defined between two of our data points  $t_i$  and  $t_{i+1}$  with  $S_i(t_i) = A(t_i)$  and  $S_i(t_{i+1}) = A(t_{i+1})$ . In addition,  $S'_i(t_i) = S'_{i-1}(t_i)$  and  $S''_i(t_i) = S''_{i-1}(t_i)$ . Natural boundary conditions were chosen for this application which means that the first and

second derivative of  $S(t)$  is equal to zero at its endpoints.  $S(t)$  and its first and second derivative are continuous. It is also the piecewise continuous polynomial with the smallest curvature that goes through all the points [7]. The cubic spline is a special case of the B-spline used in Functional Data Analysis [13]. SplinerD is a program written in Maple and is available in Appendix B for computing, evaluating, and exporting a spline or its derivatives.

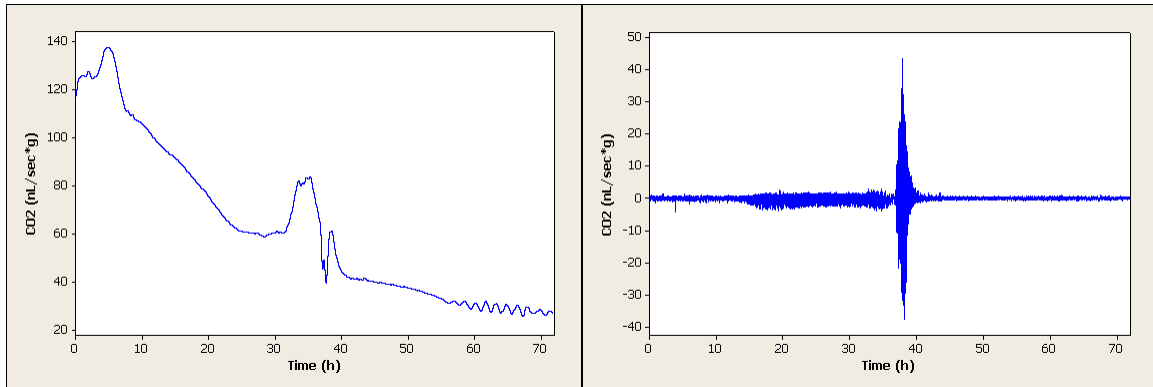


Figure 9: Spline and Residuals: smoothed time series using the spline of  $A(t)$  of Individual 6 (A) and the residuals (B).

It is important to note that in one way the comparison between a moving average of length 5000 and a block averaging of 900 is lopsided. If the spline had been computed using block averaging of 5000, its residuals would have the same problems the moving average had. The reason for the above comparison is based on the relative smoothing capabilities of the two methods. If a moving average of length 900 is compared to block averaging 900 with a cubic spline the difference in relative smoothing can be seen. Figure 10 shows an excerpt of Individual 6's trend estimated by moving average 900 and block averaging 900 with cubic spline.

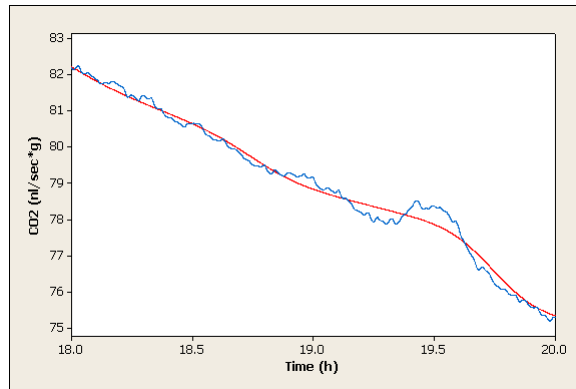


Figure 10: Smoothing Capabilities Comparison: spline based on block averaging of length 900(red) moving average of length 900 (blue).

Because block averaging coupled with cubic spline interpolation produces a smoother estimation of trend than a moving average of the same length, a smaller length block average and spline can be used to gain the same smoothness. By using a spline of a shorter length, the problem of the residuals deviating from zero for long periods of time is avoided. Thus, the time series can be decomposed into two separate time series. The spline represents the long term trend of the series, and the residuals represent the short term variation.

#### 4.3.2 Identifying Features in Trend

Although the time series has an irregular trend, the trend still has a general shape. The features of the long term trend of the time series can be isolated as changes in the first and second derivative of the time series. There are many choices available for estimating the derivative of the time series. Taking the derivative of the spline is one possibility. The problem with this method as well as many other estimators is



related to the high sampling frequency. Because of the high sampling frequency it is possible to describe or detect events that are very small in the time domain. While these features are of interest, the goal here is to describe the general shape of the time series. An example of a problematic section is given in Figure 11.

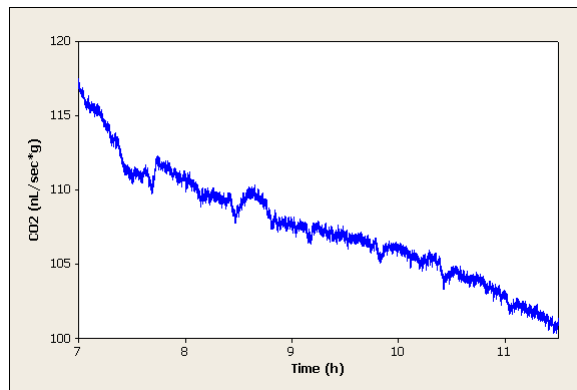


Figure 11: Excerpt of Individual 6.

Somewhere between 8.5 hours and 10.5 hours there is a change in concavity from concave up to concave down. This would correspond to the second derivative being equal to zero and the first derivative being at a maximum. The search for this structurally significant point is complicated by the presence of as yet to be described but biologically significant dips in the carbon dioxide reading about once every hour and other noise. These dips make estimations based on the original data or the spline ineffective which can be seen in Figure 12 . What can be used however, is a modified form of exponential smoothing named Moving Weighted Linear Regression (WMLR) coupled with block averaging. In this method the first  $n$  block averages are used to compute a weighted linear regression [12] and the slope of this regression is stored. The algorithm then moves over one block average and starts over. The result

is a smoothed estimation of the derivative. MWLR, a program written in Maple, is provided to compute the WMLR of a time series and is available in Appendix B.

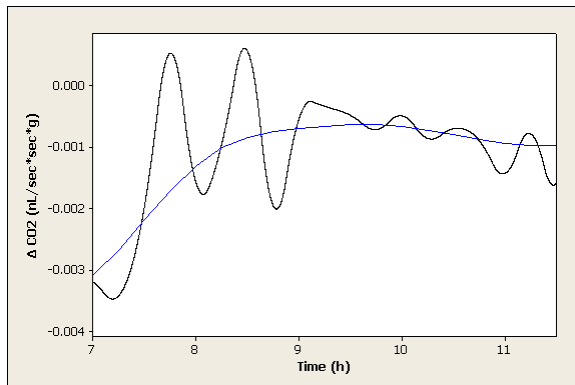


Figure 12: Spline Derivate and MWLR Comparison: spline derivate (black) and moving weighted linear regression (blue).

Choice of weighting function is dependant on analytical goals. Since a smooth estimator of the derivative is desired, a weight function related to the normal distribution was chosen as a place to start. After generating preliminary weighting values using Equation 6, the weights were divided by their mean to produce weighting values with mean equal to one.

$$W(n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{18(n-\frac{L}{2})^2}{L^2}} \quad (6)$$

The use of these particular weighting values makes the process similar to exponential smoothing except that this algorithm is two-sided and takes the future into account. Consequently the algorithm would not be useful for prediction. The trade-off for the loss of prediction capabilities is that MWLR does not suffer from some of the problems that double exponential smoothing does. Because double exponential

smoothing is one-sided the predictions for when change points are going to be are always later than actuality.

This error would not be critical if the values always deviated by a constant since a simple correction could be used. However, how far the estimation deviates from reality is based on the behavior of the time series as seen in Figure 13. If a minimum or maximum is approached slowly, similar to the point at 25.8 hours, then the estimation for when it is going to be further off than if the minimum or maximum is approached quickly, as in the point at 39.8 hours. The error is 1.4 and 0.9 for the points at 25.8 hours and 39.8 hours, respectively. This is a difference of half an hour. This level of ambiguity is not acceptable for this analysis.

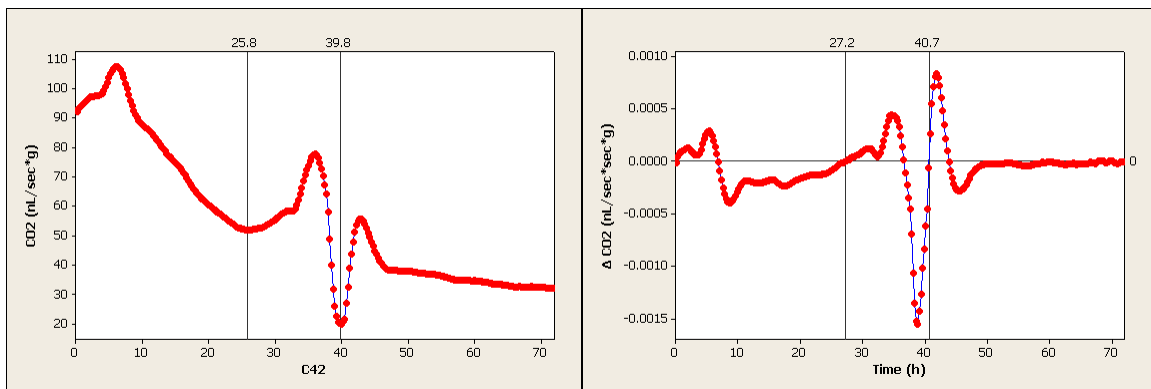


Figure 13: Exponential Smoothing Error: displays discrepancy between actual trend (A) and exponential smoothing trend estimation (B) from preliminary data

Once the smoothed derivative is estimated, change points need to be extracted. Programs were written in Maple, and are provided in Appendix B, that will identify minimums in the first derivative which correspond to changes from concave up to concave down, maximums in the first derivative which correspond to changes from

concave down to concave up, and zeros which usually correspond to local minimums or maximums in the time series. If too many features are present to move forward with analysis, a special set of these features may be chosen by some criteria. The key to this process being objective is the criterions.

A list of criterions can be assembled in many ways. The simplest would be to include all features, but this can become cumbersome. To avoid this problem a minimum time between features limit can be chosen. Features can also be chosen according to size or magnitude [16]. These are great ways to filter a large list of features especially if reconstruction is a goal. Another possibility for time series that are extremely similar is a more descriptive approach. This method has the added benefit of matching features from different individual time series. The first step is to choose a major or landmark feature that is present in all the individuals, and describe what it is and where it can be found. Now other features can be described relative to this feature. This method has the issue of not being viable if the landmark feature is not present in a subsequent individual. Consequently careful selection is need when defining any land mark features.

For this application, the first landmark feature will be called Max1 and is a maximum between .001 and .002 nL/sec\*sec\*g and happens in the first ten hours of the time series. Max1 can be seen in Figure 14; it is marked with an arrow. From Max1, several other features can be defined. If there are any zero crossings before Max1, Zero1 is the first zero crossing before Max1. The first zero crossing after Max1 is Zero2 and the first minimum after Max1 is Min1. In this way all the desired features can be enumerated. The above examples are simple and extra rules can be added if

there is any ambiguity. If no point meets the criteria for a particular feature, then the time series is said to “not have” that feature. For example, Individual 1 does not have a Zero1.

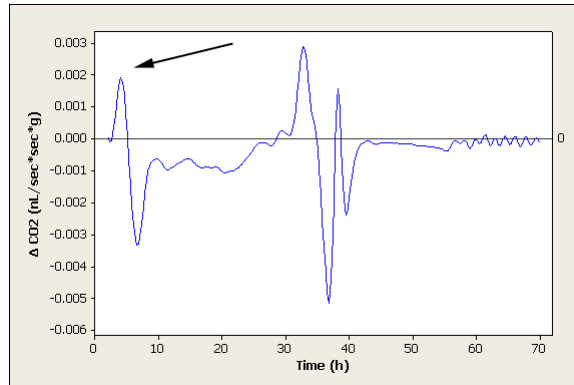


Figure 14: Derivative Estimation: of Individual 6 computed using moving weighted linear regression, indicates Max1 (Arrow).

### 4.3.3 Variability

In section 4.3.1 the separation or decomposition of the long term trend from the short term variability was covered. The goal now is to identify features based the short term variability. When the absolute value of the residuals from the spline are grafted (Figure 15) one major feature dominates the variability. The beginning and end of this Period of High Variability (PHV) could be identified by the use of a threshold value. The question remains, however: What threshold value should be used?

To explore the possibility of two distinct populations of variability, a histogram is used. Figure 16A shows an asymmetric distribution with high kurtosis. To explore the data further, a histogram of the natural log of the residuals was created. The

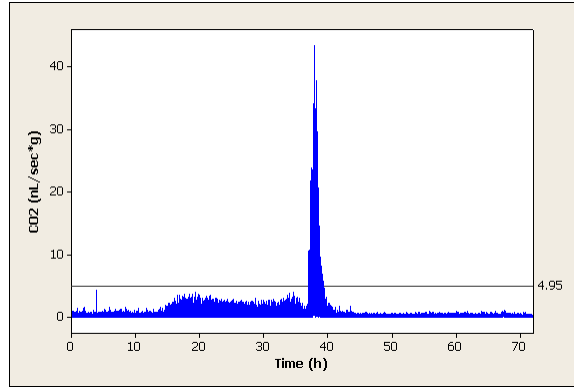


Figure 15: Variability Threshold: absolute residuals from the spline with threshold value displayed.

shape of the histogram in Figure 16B suggests that the distribution is a mixture of distributions. This information will be used to set a threshold value. A value of 1.6

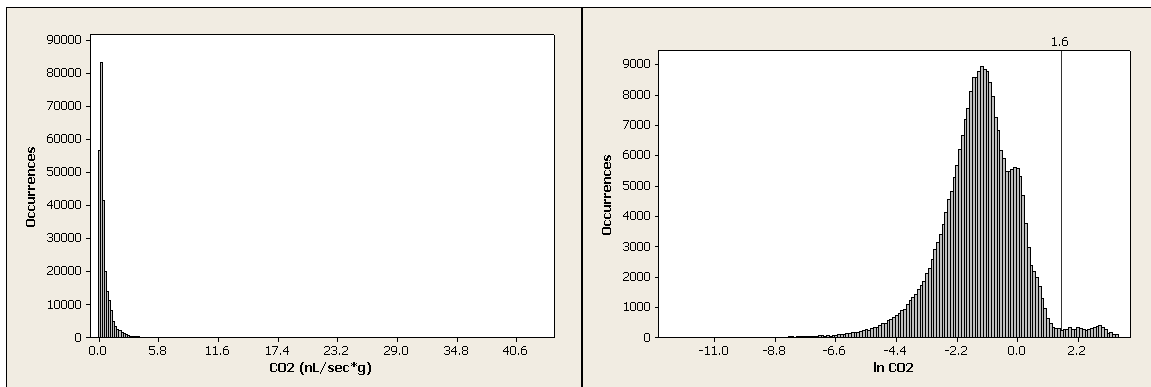


Figure 16: Population Exploration: histogram of the absolute value of the residuals from the spline (A) and histogram of the natural log of the absolute value of the residuals from the spline (B).

was chosen to separate the population to the far left from the rest of the data. This value corresponds to a value of 4.95 for absolute residuals and will be used to separate this feature on all the individuals.

#### 4.3.4 Periodicity

The frequency domain approach for the analysis of time series is an approach used for the detection of cyclic patterns in a variable. It is clear from the other subsections in 4.3 that the developmental time series data are not stationary in trend or variability. It is also likely that any periodic behavior, if present, would change over time. The detection and characterization of cyclic patterns and the changes that they might exhibit are the purpose of this subsection and the next.

When a time series is stationary, a periodogram is the tool commonly used to identify the important frequencies in the behavior of a variable. The classical spectrum is defined for stationary time series, thus when the trend or variability of a time series changes over time, the series is usually converted into a stationary time series and then a periodogram is computed. However there are processes where the periodic behavior changes over time and more complex tools, such as the spectrogram, are needed [3, 15].

This time series shows interesting periodic properties and the purpose of this section is to explore them. A discussion of the Fourier transform is needed to understand the interpretation of the following section. The Fourier transform maps data from the form of a function known as the time domain into an infinite sum of functions of the form of Equation 7 known as the frequency domain.

$$f = \sum_{j=1}^{\infty} a_j e^{i\omega(j)t} \quad (7)$$

Since it is not practical to work with an infinite sum, the Fast Fourier Transform (FFT) can be used to approximate the Fourier Transform with a finite sum of functions. With the use of this simplification as well as Euler's identity a simplified version (Equation 8) can be used that is more intuitive [19].

$$f = \sum_{j=1}^n a_j (\cos(\omega(j)t) + i \sin(\omega(j)t)) \quad (8)$$

Equation 8 makes it clearer that the Fourier transform is writing the data as a sum of multiple sine and cosine waves. By graphing  $|a_j|$ , which represents the amplitude of each of these waves, it is possible to get an idea of the strength of each frequency. This kind of graph is called a periodogram and can be seen in Figure 17B. Because of

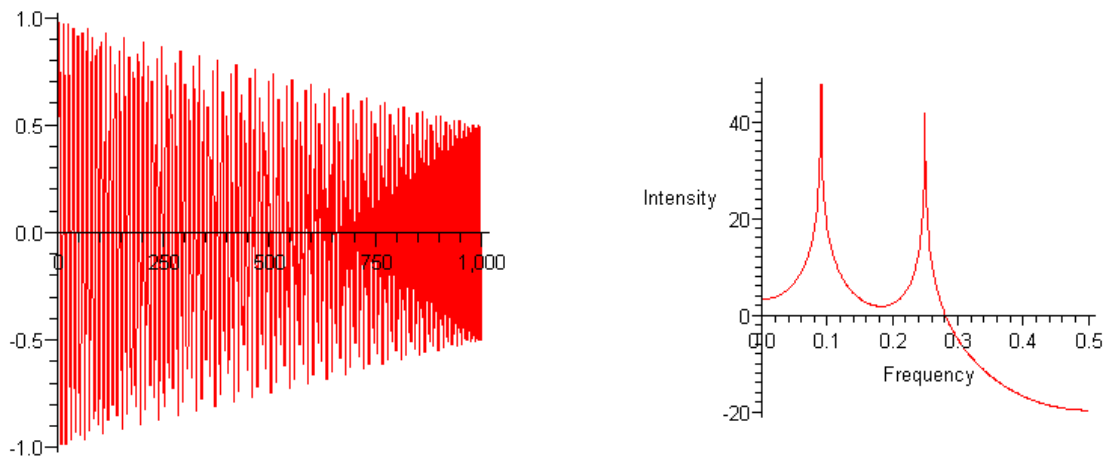


Figure 17: Example Periodogram: example time series (A) and its periodogram (B)



Figure 17 shows a time series (A) and its periodogram (B). This time series is generated by Equation 9. Notice the peaks at 0.25 and 0.091. These two peaks should, in theory, be the only data present, this is not the case. This error is known as side lobe error, and is caused by window function being used [20].

$$f(n) = \frac{n}{1000} \sin\left(\frac{2\pi n}{11}\right) + \frac{1000 - n}{2000} \sin\left(\frac{\pi n}{2}\right) \quad (9)$$

A window function is a function that is zero outside a given interval (the window) that is multiplied by the original time series before the FFT is computed. The simplest window function, the rectangular window, was used in the above example. The rectangular window is one inside the interval and zero outside it. Whole families of window functions exist and the choice of which to use is non-trivial. There are two main considerations when choosing a window function: the main lobe width and the height of the largest side-lobe. These values will correspond to resolving power and significance of side lobe error respectively [18]. In order to compare windows easily, a normalized spectrum of the window can be used. The windows that will be covered in this thesis are Rectangular, Hamming, Blackman, and Truncated Gaussian windows.

Figure 18 shows the rectangular window (A) its performance when used to compute the spectrogram of the time series created by Equation 9 (B) and its normalized spectrum (C). The interpretation of Figure 18C starts with its main lob width which is just two bins wide. This means that when the spectrogram is computed that the main peak of a signal will be only two bins thick or has high resolution. The second feature of importance is the highest side lobe is only 13 decibels below main lobe or has high side lobe error [20]. These features can be seen in Figure 18 B in that the

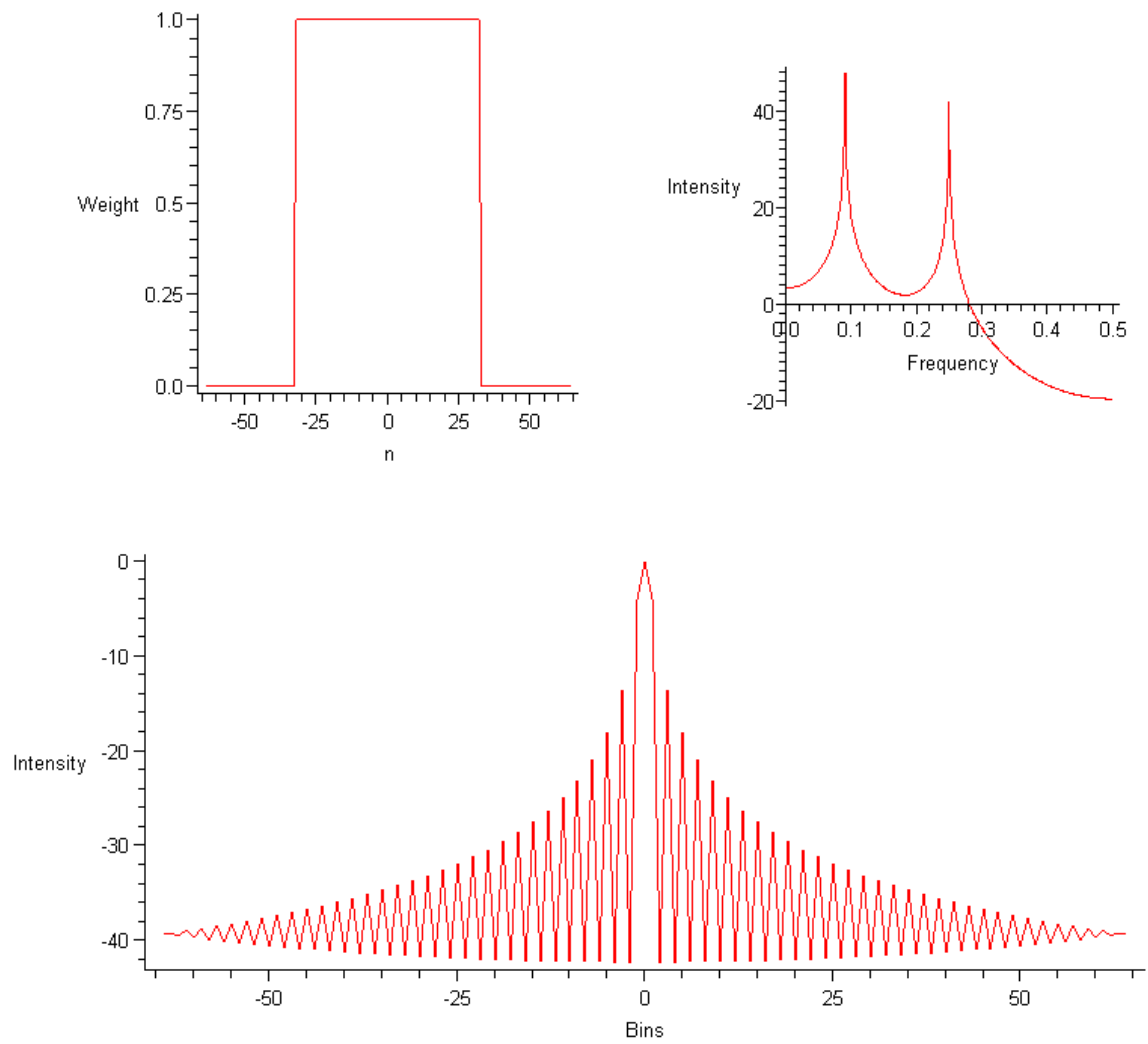


Figure 18: Rectangular Window: window (A), periodogram performance (B), and normalized spectrum(C)

peaks at 0.09 and 0.25 are thin, but high side lobe error is evidenced by how much data is above the intensity equal to zero line.

The Hamming window has a four bin wide main lobe, which means that it has less resolving power than the rectangular window, but the Hamming window has a much lower highest side lobe at 42 decibels below the main lobe [20]. This can be seen in Figure 19B where the wider peaks cross the intensity equal to zero line before side lobe error is seen.

The Blackman window (Figure 20) takes one more step in sacrificing resolution for minimized side lobe error. Its highest side lobe is 58 decibels below, but its main lobe is eight bins wide. This kind of window is good for audio encoding because its side lobe error falls below the noise floor of the system. The final window to be considered here, the Gaussian window, has a very interesting property. The Fourier Transform of the Gaussian window is the Gaussian distribution itself. Any side lobe error is due strictly to truncation. Truncation is necessary because the window is theoretically of infinite length. By manipulation of the parameters of the window and the truncation of the window, arbitrarily small side lobe error can be achieved, but at the cost of a large main lobe width [20]. This will not work well for this application and is not considered further.

The astute reader will have noticed that Equation 9 produces a signal with properties that change over time. It is important to note the FFT does not take change over time into account. Frequency intensities are given, but no information is generated about when or if dominate frequencies change. To overcome this, a spectrogram can be used. A spectrogram is generated by computing several Short-Time Fourier

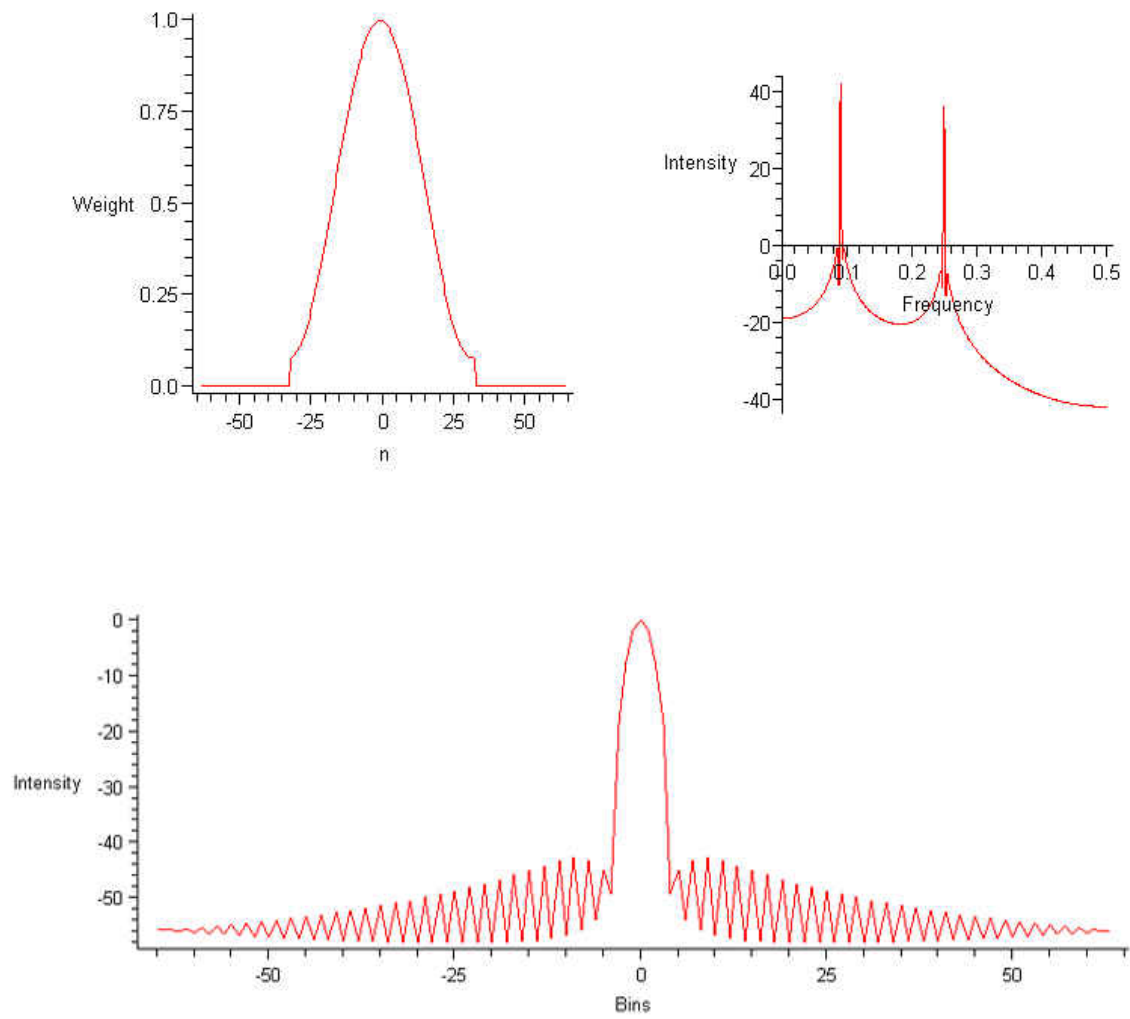


Figure 19: Hamming Window: window (A), periodogram performance (B), and normalized spectrum(C)

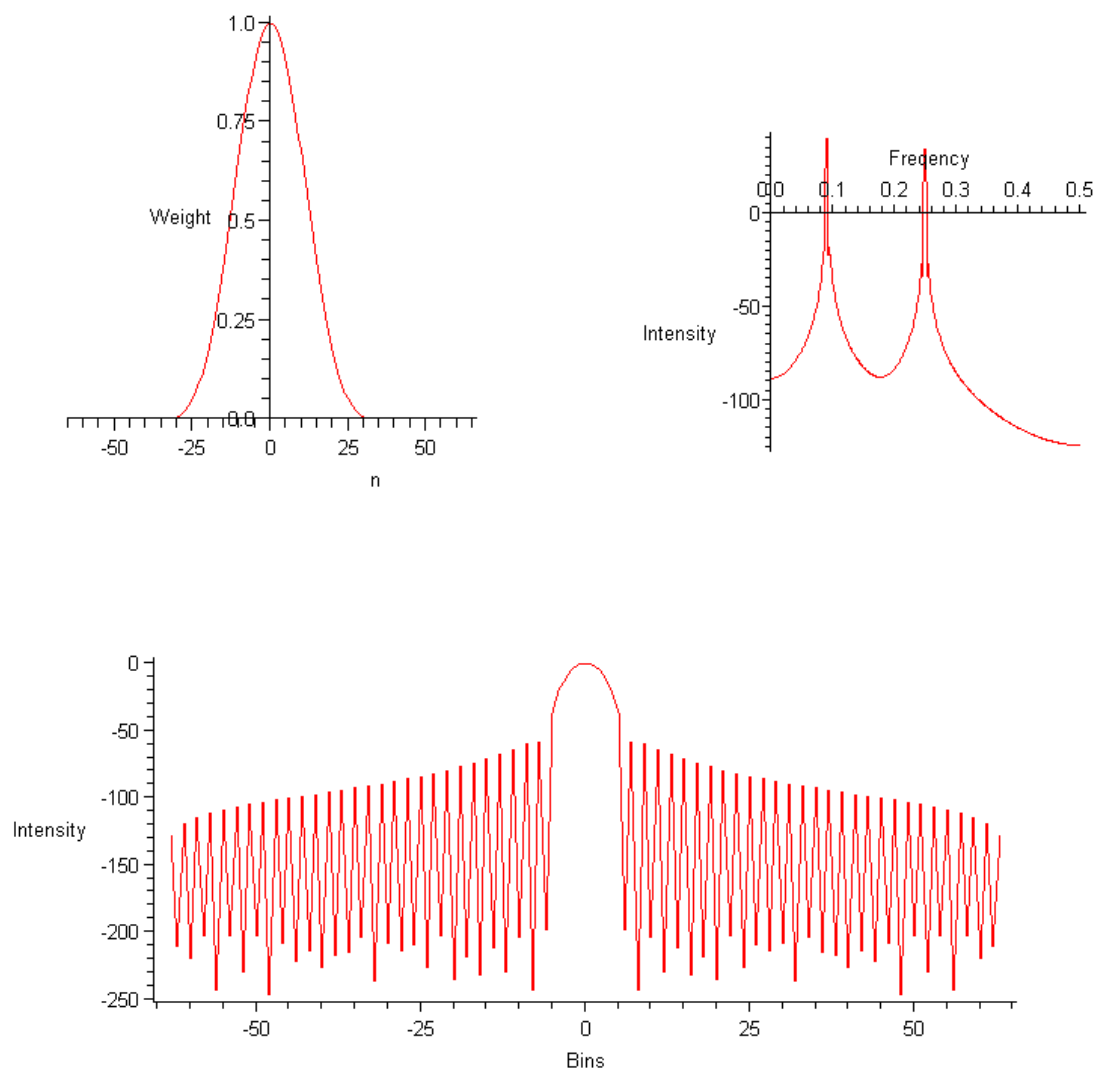


Figure 20: Blackman Window: window (A), periodogram performance (B), and normalized spectrum(C)

Transforms (STFT). A STFT is a FFT that is computed by using a window that focuses on a small piece of time series. This provides a time localized estimation of the periodicity of the time series. By sliding the window over a little or “hopping” and computing more STFTs, it is possible to gain insight into how periodicity changes over time. It is important to note that this process depends on the assumption that the any non-stationary characteristics of the time series are changing slowly over time. [18, 1]. Figure 21 shows the periodogram and spectrogram of the time series generated by Equation 9. The spectrogram was created with a program for matlab from Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications, Second Edition [19].

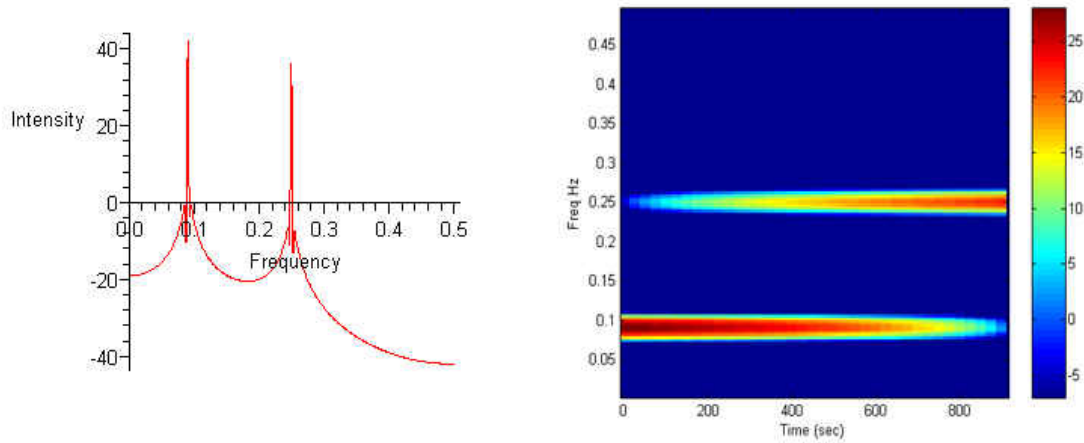


Figure 21: Spectrogram and Periodogram Comparison: shows periodogram (A) and spectrogram (B)

The spectrogram is a three-dimensional construction. The  $x$  axis is time and moves linearly with early observations on the left and later observations to the right. The  $y$  axis represents frequency and moves linearly from high frequencies at the top

to low frequencies at the bottom. The third axis represents intensity or amplitude and is in a  $20 * \log_{10}$  scale or dB scale. The third dimension is represented here as colors. The color bar shows the scale in relation to the different colors. It is possible to see in Figure 21B how the two different frequencies created by Equation 9 change over time. The intensity of the 0.9 frequency is waning, while the intensity of the 0.25 frequency is waxing.

The spectrogram has the same considerations as the periodogram with respect to how much zero padding to use and what windowing function to use, but also has the added considerations of window length and hop size. It is desirable to have as large a window as possible that still represents local conditions as it can help with frequency resolution and expands the largest identifiable period. If too long of a window is used, however, a non-local estimation would occur possibly violating assumptions about the nearly stationary nature of the time series inside the window. This could produce a spectrogram that does not accurately reflect the behavior of the time series at that time. A short hop distance relative the window length is desirable, but the shorter the hop length the more computationally expensive the spectrogram becomes. Several suggestions are given in the literature [18] as to reasonable minimums for different applications. The hop distance used here is shorter than these suggestions as to provide better resolution in the time domain for identification of features.

It is important, when considering the spectrogram as a tool for experimental analysis of this kind, that the periodic properties of the machine are checked. It is common for electronic equipment and pumps to contribute periodic noise to the data. The spectrogram of the finite differences of the blank can be seen in Figure 22

A. It shows a featureless gradient with no frequencies dominating. Had there been any frequencies present, those frequencies would have to be either filtered out of the individual data or ignored in the analysis attributing it to the machine. The fact that higher frequencies are more prevalent is an effect of the finite differences.

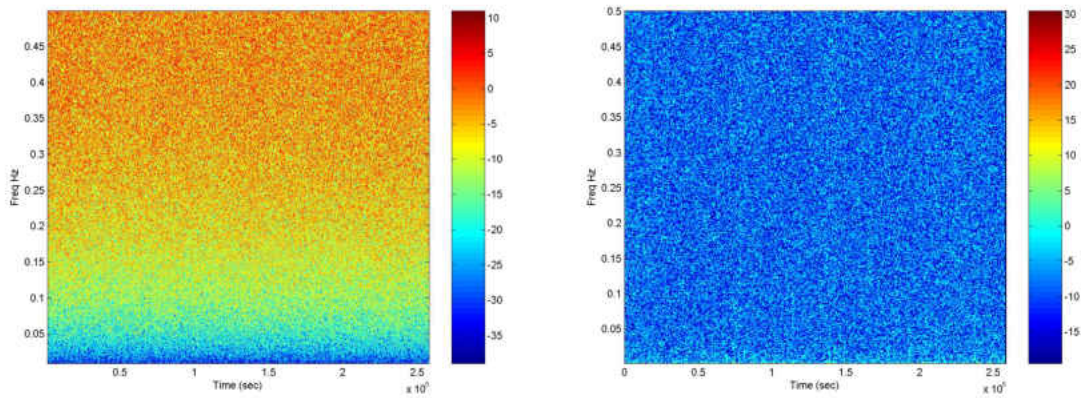


Figure 22: Blank Spectrogram: from finite differences (A) and from residuals from a linear regression (B).

The spectrogram of the residuals from a linear regression of the blank are presented in Figure 22B. The gradient effect caused by the finite differences is not masking any underlying frequencies. The program used to generate these spectrograms [19] was modified in order to display several images in the same intensity scale. The original algorithm automatically chose the scaling of the intensity scale based on the maximum intensities. The modified one lets the user select the scale manually.

The parameters for computing the following spectrogram, as well as those in Appendix A, were as follows: a Hamming window of length 2001 sec, zero padding to 4096 sec, and a hop length of 75 sec were used. Figure 23A shows the whole spectrogram while Figure 23B shows a close-up of the lower frequencies. Two major



features can be seen, first a transition from lower frequencies to slightly higher frequencies around 55000 sec and a highly intense section that will not be considered for segmentation as it coincides with the period of high variability. To explore the first feature in more depth a few programs were written in Maple designed to compliment the spectrogram.

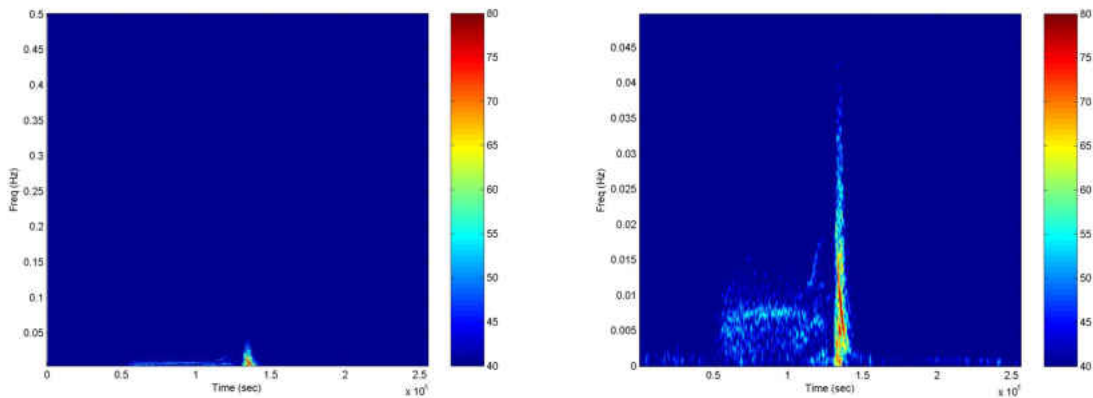


Figure 23: Individual 6 Spectrogram: standard view (A) and close up of lower frequencies (B)

The first program was designed to compute the values for the spectrogram in the same way as the program that produced the spectrogram [19], but instead of making an image out of the data, the program saves the data as a text file for importing into other programs or reloaded into Maple for further analysis. The second program written in Maple finds the most intense frequency at every time point and saves it as a text file. The output was then imported into Minitab and seen below in Figure 24 A.

A shift from extremely low frequencies to slightly higher frequencies is seen in Figure 24A, which shows how the most intense frequencies change over time. In

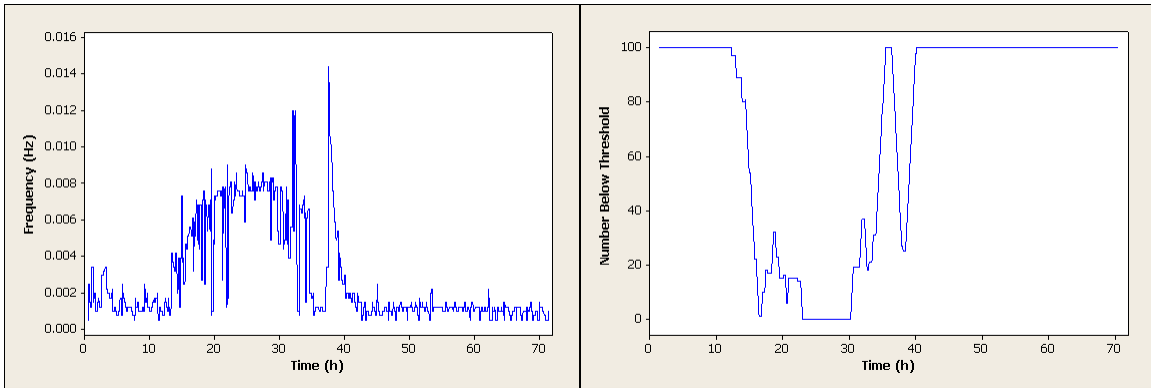


Figure 24: Program Output:  $S_{\max}(A)$  and  $\text{per}(B)$

order to identify this change point, a program was written in Maple to compute how many time points in a particular neighborhood had a frequency below a threshold. A graphical representation is given in Figure 24B. In the example given, one hundred time points were considered at a time and the threshold was set to 0.004 Hz. By examining when 60 of the 100 values are below the threshold, we can get an idea of when the transition occurs. This tool identifies a change in the periodic nature of the time series at time 53500 sec or about 14.9 hours.

## 5 RESULTS

The procedures outlined in Chapter 3 produce 21 change points with respect to trend, variability, and periodicity. These change points are displayed in Table 1 and define 20 separate developmental time periods. Given the large number of segments and the complexity contained in each one, an in depth analysis of each segment is outside the scope of this thesis. With that being said, some basic analysis of these time periods is useful for illustrating the value of the system as a whole.

Table 1: Change Points: Identified from experimental data.

	IND 1	IND 2	IND 3	IND 4	IND 5	IND 6	IND 7
Zero1	*	7517	7667	*	9030	9277	8972
Max1	10800	17100	15300	13500	14400	14400	15300
Zero2	14733	20691	19134	17161	18476	18100	18864
Min1	20700	25200	25200	23400	24300	23400	24300
Max2	29700	33300	33300	31500	32400	34200	33300
Min2	37800	40500	42300	42300	41400	41400	42300
Max3	46800	49500	51300	49500	49500	52200	49500
Perio	52675	51400	85750	59800	74575	53500	63625
Zero3	88747	99408	97095	102452	102787	102514	104933
Max4	101700	105300	109800	107100	108000	105300	108900
Min3	105300	108900	114300	110700	109800	109800	113400
Max5	114300	115200	121500	119700	118800	117900	121500
Zero4	120885	123535	130084	127480	125509	125425	129425
Min4	128700	129600	135900	134100	133200	132300	135900
PHVB	129512	130412	137048	136539	133842	132970	136783
Zero5	132295	133238	140891	137862	136370	135975	139673
Max6	134100	135000	141300	139500	137700	137700	141300
Zero6	135381	136764	142133	141756	139590	139314	142592
PHVE	135973	136982	143698	142236	142032	141261	144067
Min5	138600	139500	145800	144900	142200	142200	145800
Zero7	206420	209946	219282	209703	207774	219674	220908

A is defined as the period between Zero1 and Max1, B is defined as the period between Max1 and Zero2, and the rest are labeled in a similar manner. Some basic statistics describing the length of each segment is presented in Table 3. While there is a fair amount of variation in how long each segment, it is amazing that a biological

measurement could be so complex yet so consistently measured. Figure 25 shows dot plots to illustrate the distribution of lengths of select segments.

Table 2: Segment Lengths: Identified from experimental data.

	IND 1	IND 2	IND 3	IND 4	IND 5	IND 6	IND 7
A	*	9583	7633	*	5370	5123	6328
B	3933	3591	3834	3661	4076	3700	3564
C	5967	4509	6066	6239	5824	5300	5436
D	9000	8100	8100	8100	8100	10800	9000
E	8100	7200	9000	10800	9000	7200	9000
F	9000	9000	9000	7200	8100	10800	7200
G	5875	1900	34450	10300	25075	1300	14125
H	36072	48008	11345	42652	28212	49014	41308
I	12953	5892	12705	4648	5213	2786	3967
J	3600	3600	4500	3600	1800	4500	4500
K	9000	6300	7200	9000	9000	8100	8100
L	6585	8335	8584	7780	6709	7525	7925
M	7815	6065	5816	6620	7691	6875	6475
N	812	812	1148	2439	642	670	883
O	2783	2826	3843	1323	2528	3005	2890
P	1805	1762	409	1638	1330	1725	1627
Q	1281	1764	833	2256	1890	1614	1292
R	592	218	1565	480	2442	1947	1475
S	2627	2518	2102	2664	168	939	1733
T	67820	70446	73482	64803	65574	77474	75108

What is even more astounding than the fact that this many features can be observed in so many samples is how regular is the end and onset of each stage. Table 4 shows the time from Max1 to the conclusion of the segment named. It is interesting that the progression from Max1 through the end of segment S is about 35.6 hours long with a 95% confidence interval of about plus or minus 45 minutes. Basic statistics of the data presented in Table 4 are presented in Table 5.

The tools developed in this thesis are well suited to the analysis of this time series, but could also be used in other applications with highly variable time series. Continued work can and needs to be done on these time series. An in-depth analysis of each segment may reveal more biologically significant features. The further char-

Table 3: Segment Length Statistics

	N	Mean	StDev	Min	Medn	Max	Range	CoefVar
A	5	6807.0	1838.0	5123	6328	9583	4460	0.270016
B	7	3765.6	189.5	3564	3700	4076	512	0.050324
C	7	5620.0	594.0	4509	5824	6239	1730	0.105694
D	7	8743.0	1001.0	8100	8100	10800	2700	0.114492
E	7	8614.0	1258.0	7200	9000	10800	3600	0.146041
F	7	8614.0	1258.0	7200	9000	10800	3600	0.146041
G	7	13289.0	12415.0	1300	10300	34450	33150	0.934231
H	7	36659.0	13238.0	11345	41308	49014	37669	0.361112
I	7	6881.0	4179.0	2786	5213	12953	10167	0.607325
J	7	3729.0	962.0	1800	3600	4500	2700	0.257978
K	7	8100.0	1039.0	6300	8100	9000	2700	0.128272
L	7	7635.0	760.0	6585	7780	8584	1999	0.099542
M	7	6765.0	760.0	5816	6620	7815	1999	0.112343
N	7	1058.0	631.0	642	812	2439	1797	0.596408
O	7	2743.0	750.0	1323	2826	3843	2520	0.273423
P	7	1471.0	493.0	409	1638	1805	1396	0.335146
Q	7	1561.0	468.0	833	1614	2256	1423	0.299808
R	7	1246.0	831.0	218	1475	2442	2224	0.666934
S	7	1822.0	952.0	168	2102	2664	2496	0.522503
T	7	70672.0	4873.0	64803	70446	77474	12671	0.068952

acterization would also be useful in ongoing biological research. Another interesting avenue for future research would be to create an average or representative time series that incorporated all the individual time series using function data analysis. The end result of this thesis is not just a set of tools that bring the analysis of this depth within the reach of more people, but an excellent model system to better understand development in flies. Experimental manipulation of this system could include changes in temperature, insecticide treatments, and physiological manipulations just to name a few.

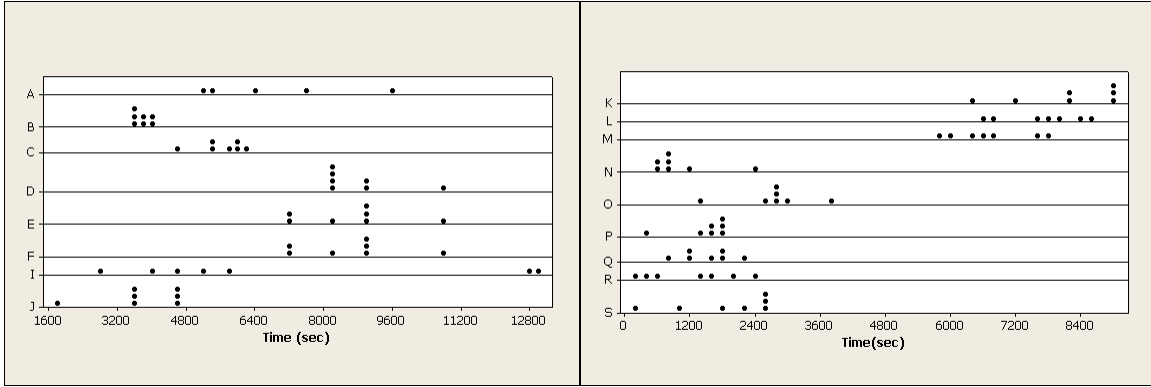


Figure 25: Segment Lengths

Table 4: Segment Conclusion

	IND 1	IND 2	IND 3	IND 4	IND 5	IND 6	IND 7
B	3933	3591	3834	3661	4076	3700	3564
C	9900	8100	9900	9900	9900	9000	9000
D	18900	16200	18000	18000	18000	19800	18000
E	27000	23400	27000	28800	27000	27000	27000
F	36000	32400	36000	36000	35100	37800	34200
G	41875	34300	70450	46300	60175	39100	48325
H	77947	82308	81795	88952	88387	88114	89633
I	90900	88200	94500	93600	93600	90900	93600
J	94500	91800	99000	97200	95400	95400	98100
K	103500	98100	106200	106200	104400	103500	106200
L	110085	106435	114784	113980	111109	111025	114125
M	117900	112500	120600	120600	118800	117900	120600
N	118712	113312	121748	123039	119442	118570	121483
O	121495	116138	125591	124362	121970	121575	124373
P	123300	117900	126000	126000	123300	123300	126000
Q	124581	119664	126833	128256	125190	124914	127292
R	125173	119882	128398	128736	127632	126861	128767
S	127800	122400	130500	131400	127800	127800	130500
T	195620	192846	203982	196203	193374	205274	205608

Table 5: Segment Conclusion Statistics

	N	Mean	StDev	Min	Medn	Max	Range	CoefVar
B	7	3766	189.5	3564	3700	4076	512	0.050324
C	7	9386	708.0	8100	9900	9900	1800	0.075431
D	7	18129	1093.0	16200	18000	19800	3600	0.060290
E	7	26743	1620.0	23400	27000	28800	5400	0.060577
F	7	35357	1701.0	32400	36000	37800	5400	0.048109
G	7	48646	12627.0	34300	46300	70450	36150	0.259569
H	7	85305	4562.0	77947	88114	89633	11686	0.053479
I	7	92186	2256.0	88200	93600	94500	6300	0.024472
J	7	95914	2429.0	91800	95400	99000	7200	0.025325
K	7	104014	2886.0	98100	104400	106200	8100	0.027746
L	7	111649	2936.0	106435	111109	114784	8349	0.026297
M	7	118414	2886.0	112500	118800	120600	8100	0.024372
N	7	119472	3196.0	113312	119442	123039	9727	0.026751
O	7	122215	3125.0	116138	121970	125591	9453	0.025570
P	7	123686	2886.0	117900	123300	126000	8100	0.023333
Q	7	125247	2813.0	119664	125190	128256	8592	0.022460
R	7	126493	3181.0	119882	127632	128767	8885	0.025148
S	7	128314	3023.0	122400	127800	131400	9000	0.023559
T	7	198987	5724.0	192846	196203	205608	12762	0.028766

## BIBLIOGRAPHY

- [1] Allen, Jont B., Lawrence R. Rabiner. "A Unified Approach to Short-Time Fourier Analysis and Synthesis." *Proceedings of the IEEE* 65 (1977):1558-1564.
- [2] Box, George E. P., Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Revised Edition. San Francisco: Holden-Day, 1976.
- [3] Bruscato, Adriana, Clélia M. C. Tolo. "Spectral Analysis of Non-stationary Processes Using the Fourier Transform." *Brazilian Journal of Probability and Statistics* 18 (2004):69-102.
- [4] Chapman, R. F. *The Insects: Structure and Function*. Fourth Edition. Cambridge: Cambridge University Press, 1998.
- [5] Denlinger, D. L., J. H. Willis, G. Fraenkel. "Rates and Cycles of Oxygen Consumption During Pupal Diapause in *Sarcophaga* Flesh Flies," *Journal of Insect Physiology* 18 (1972):871-882.
- [6] Denlinger, D. L., Jan Žďárek. "Metamorphosis Behavior of Flies." *Annual Review of Entomology* 39 (1984):234-66.
- [7] Greville, T. N. E. "Numerical Procedures for Interpolation by Spline Functions." *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis* 1 (1964):53-68.
- [8] Krough, A. *The Comparative Physiology of Respiratory Mechanisms*. Philadelphia: University of Pennsylvania Press, 1941.



- [9] Lighton, John R. B. “Discontinuous Ventilation in Terrestrial Insects.” *Physiological Zoology* 67 (1994):142-162.
- [10] Lighton, J. R. B. “Measurements on Insects,” *Concise Encyclopedia on Biological and Biomedical Systems*. Ed. P. A. Payne. Oxford: Pergammon, 1991.
- [11] Miller P. L. “The Regulation of Breathing in Insects.” *Advances in Insect Physiology* 3 (1996):273-345.
- [12] Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, William Wasserman. *Applied Linear Statistical Models*. Fourth Edition. Chicago: Irwin, 1996.
- [13] Ramsay, J. O., B. W. Silverman. *Functional Data Analysis*. Second Edition. New York: Springer, 2005.
- [14] Romoster, W. S., J. G. Stoffolano Jr. *The Science of Entomology*. Fourth Edition. Boston: McGraw-Hill, 1998.
- [15] Shah, S. I., L. F. Chaparro, A. El-Jaroudi, J. M. Furman. “Evolutionary Maximum Entropy Spectral Estimation and Heart Rate Variability Analysis.” *Multi-dimensional Systems and Signal Processing* 9 (1998):453-458.
- [16] Sharifzadeh, Mehdi, Farnaz Azmoodeh, Cyrus Shahabi. “Change Detection in Time Series Data Using Wavelet Footprints.” Technical Report: 05-855, 2005. March 2008 <<http://www.cs.usc.edu/Research/ReportsList.htm>>.
- [17] Sláma, K. “Regulation of Respiratory Acidemia by the Autonomic Nervous System (Coelopulse) in Insects and Ticks.” *Physiological Zoology* 67 (1994):163-174.

- [18] Smith, Julius O. III., Xavier Serra. “PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation.” Technical Report: CCRMA STAN-M-43, 1987. March 2008 <<http://ccrma.stanford.edu/stanms/stanm43/>>.
- [19] Smith, Julius O. III. *Mathematics of the Discrete Fourier Transform (DFT): With Audio Applications*. Second Edition. Online Book, 2007. March 2008 <<http://ccrma.stanford.edu/~jos/mdft/mdft.html>>.
- [20] Smith, Julius O. III. *Spectral Audio Signal Processing*. Draft. Online Book, 2007. March 2008 <<http://ccrma.stanford.edu/~jos/sasp/sasp.html>>.
- [21] Westneat, Mark W., Oliver Betz, Richard W. Blob, Kamel Fezzaa, W. James Cooper, Wah-Keat Lee. “Tracheal Respiration in Insects Visualized with Synchrotron X-ray Imaging.” *Science* 299 (2003):558-60.

APPENDICES

Appendix A

Individuals

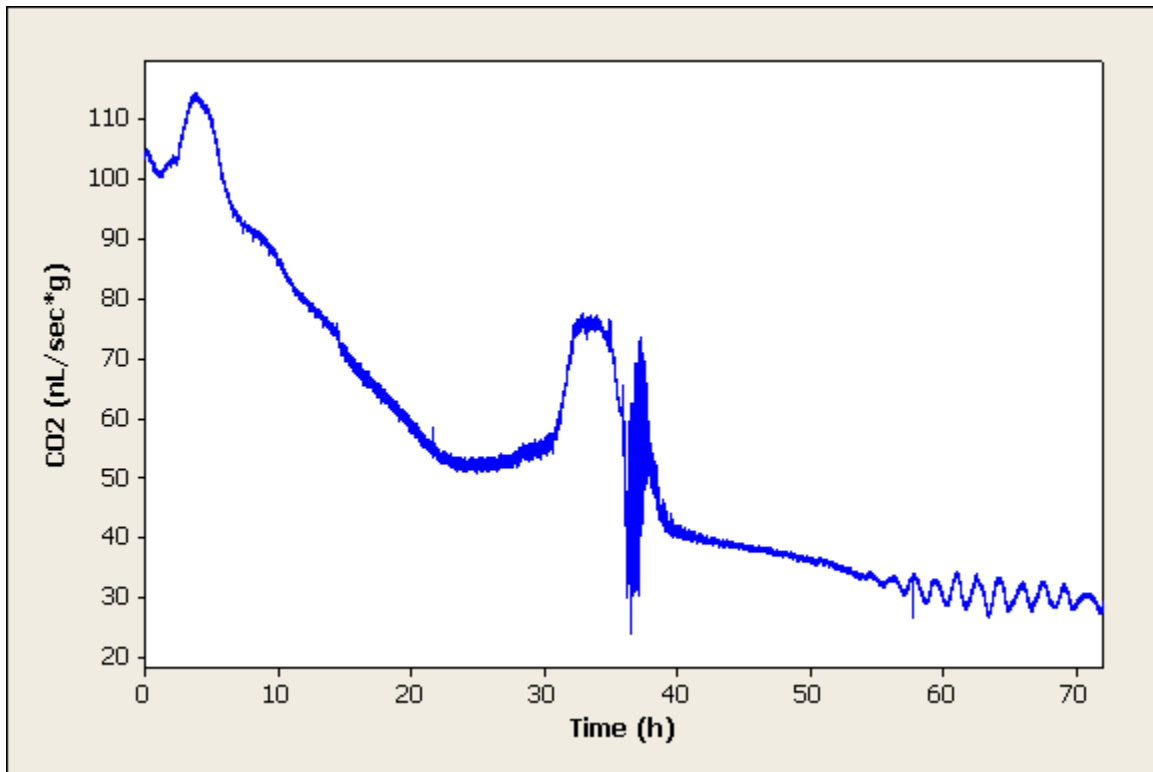


Figure 26: Individual 1

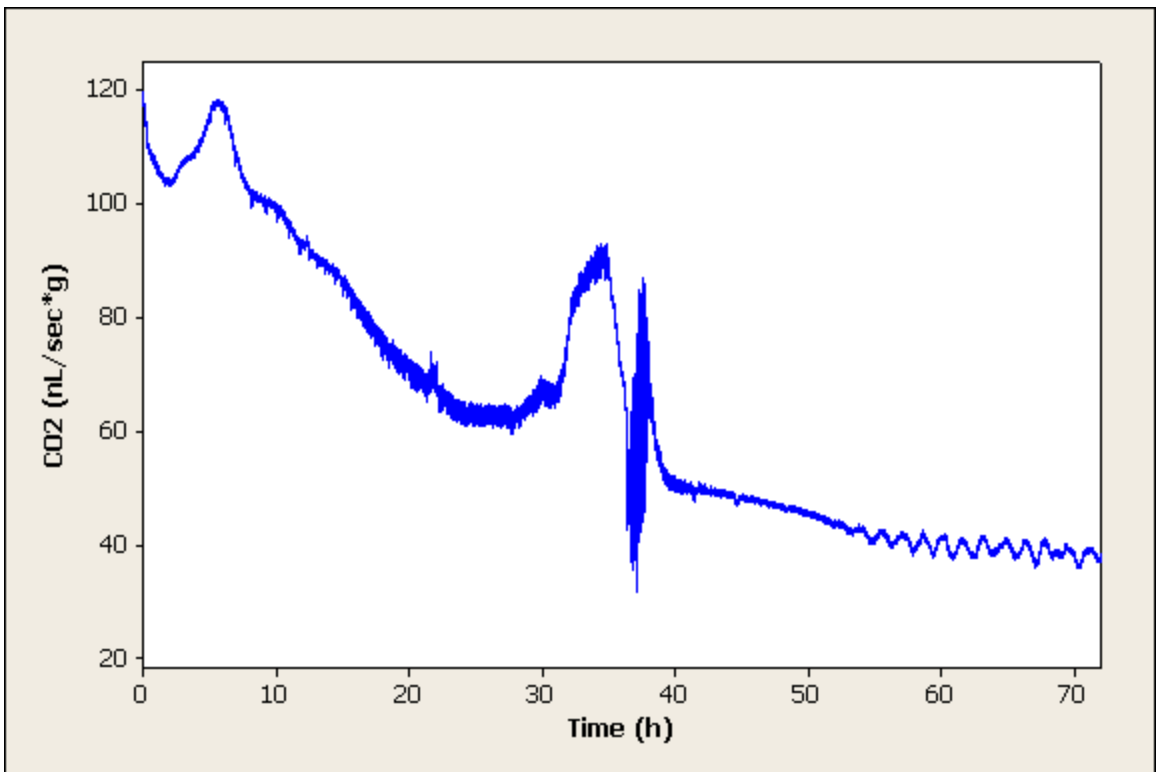


Figure 27: Individual 2

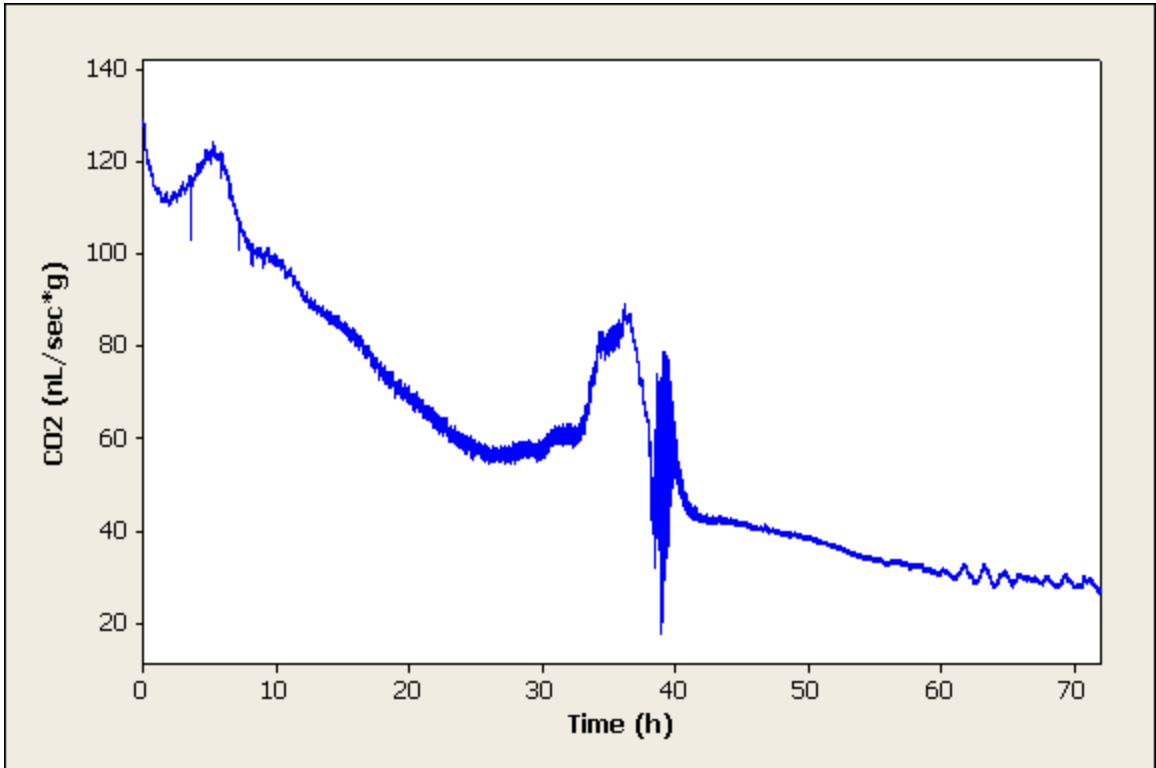


Figure 28: Individual 3

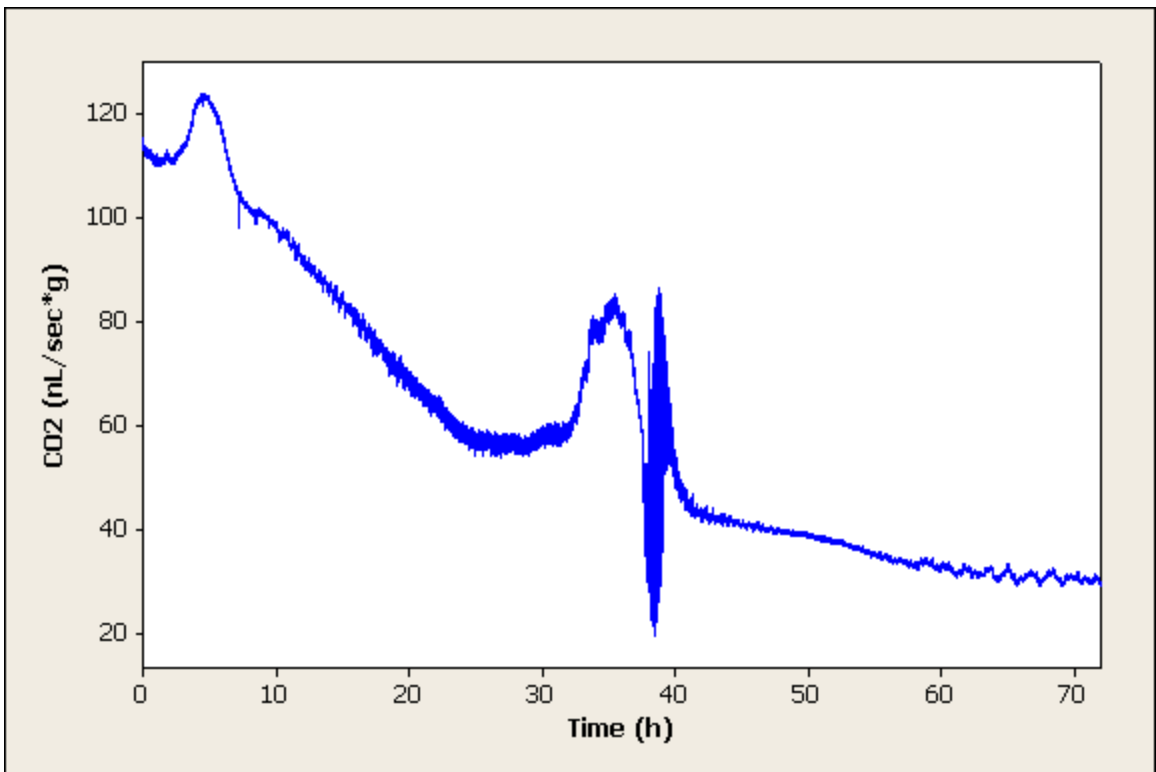


Figure 29: Individual 4

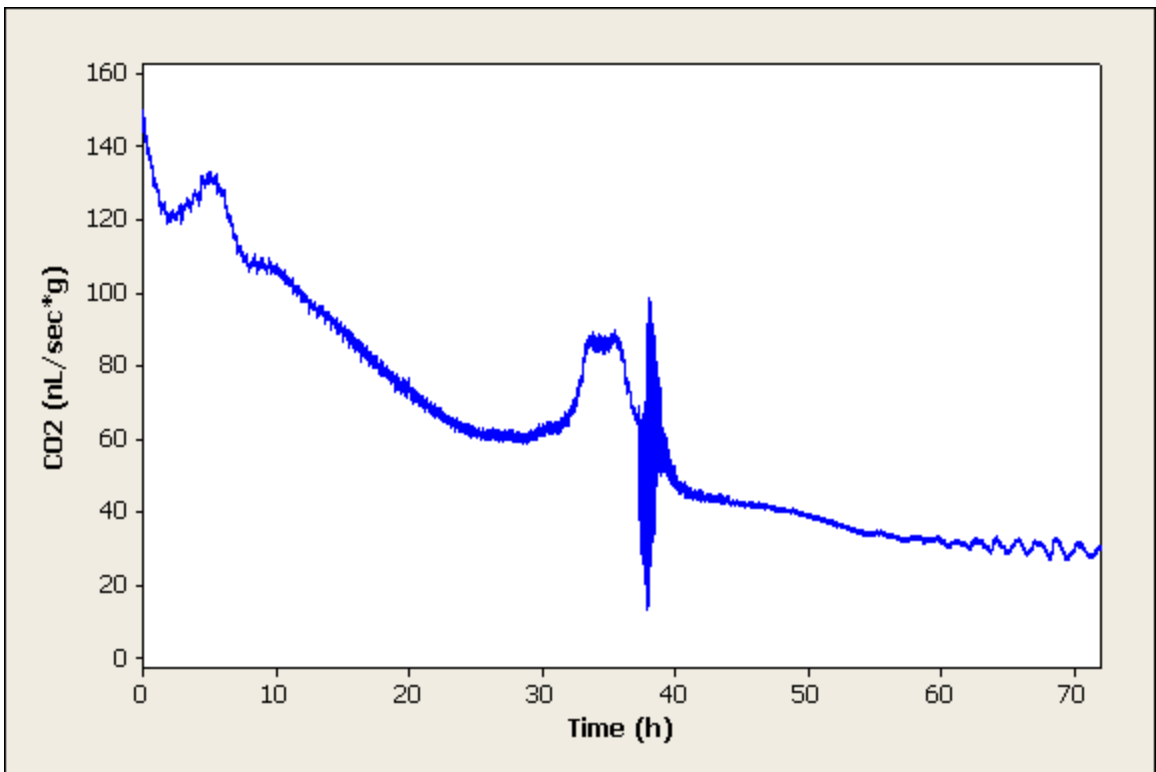


Figure 30: Individual 5

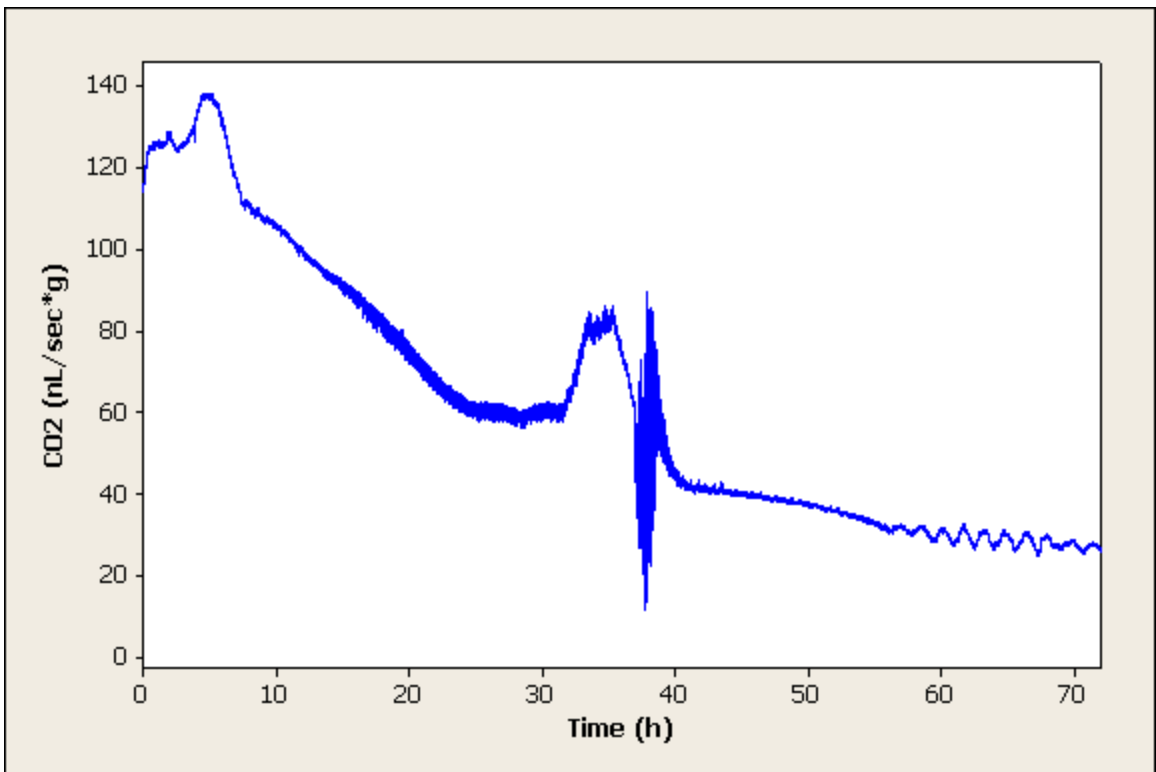


Figure 31: Individual 6



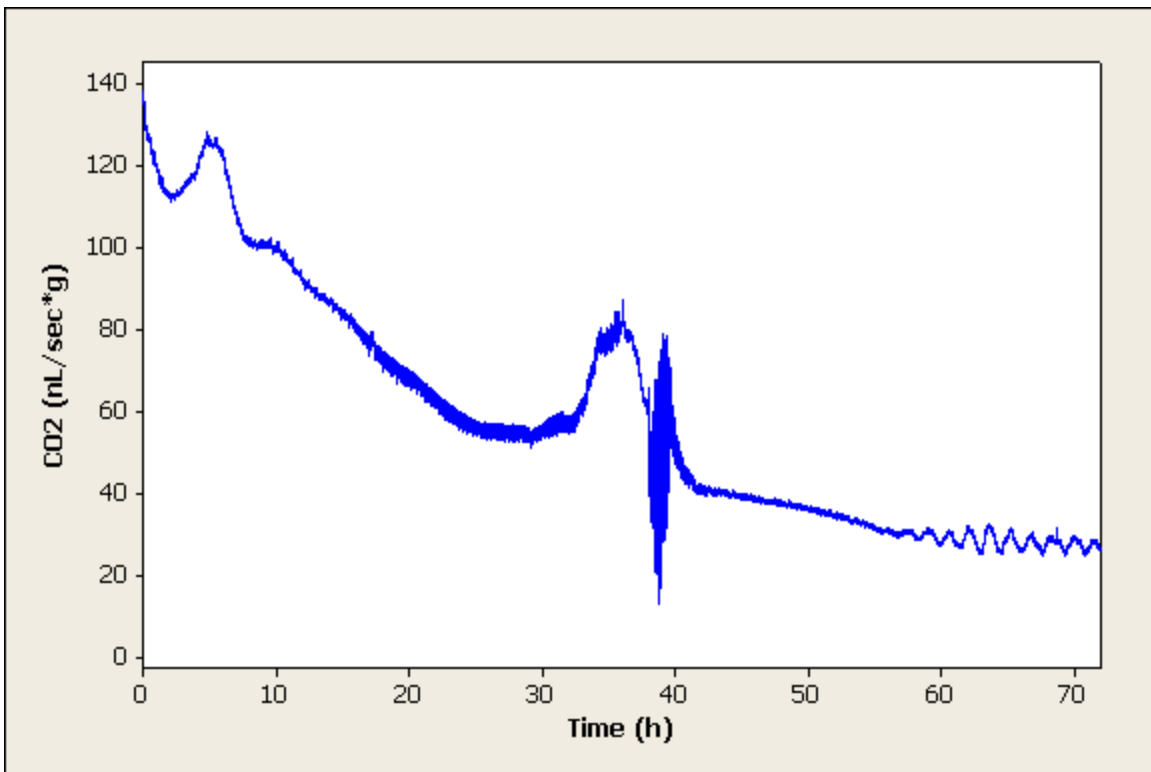


Figure 32: Individual 7

# Splines

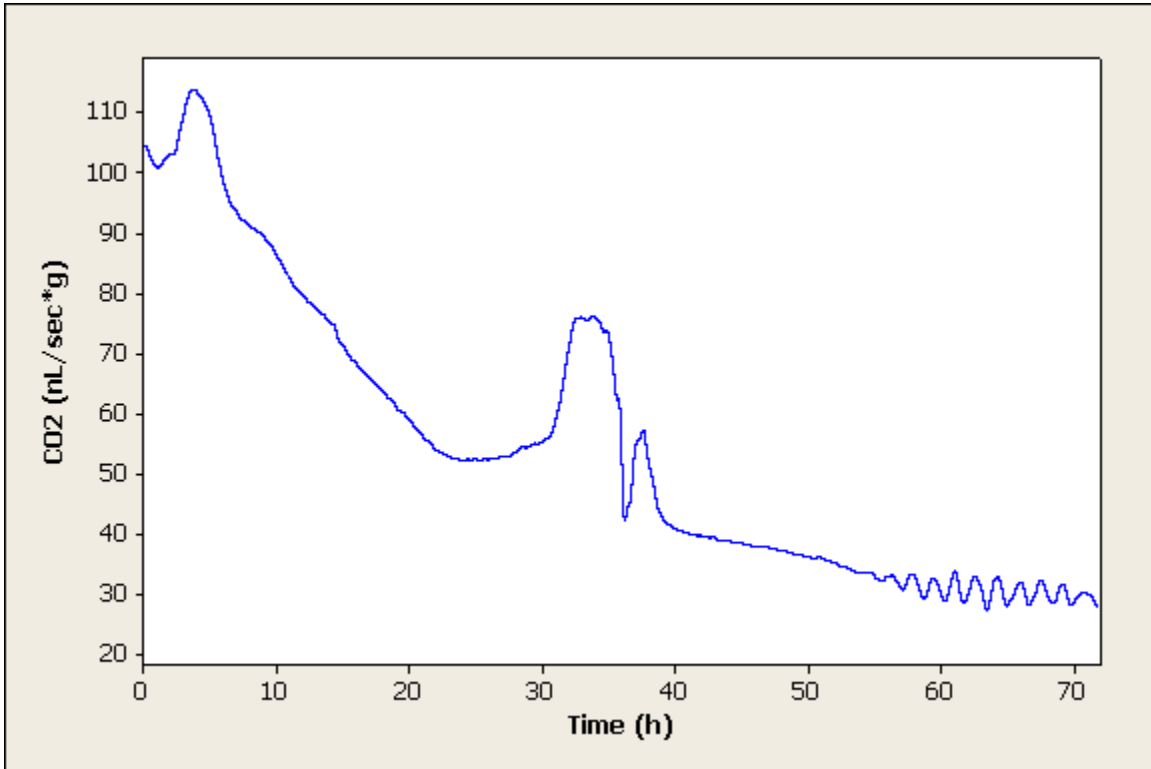


Figure 33: Spline of Individual 1

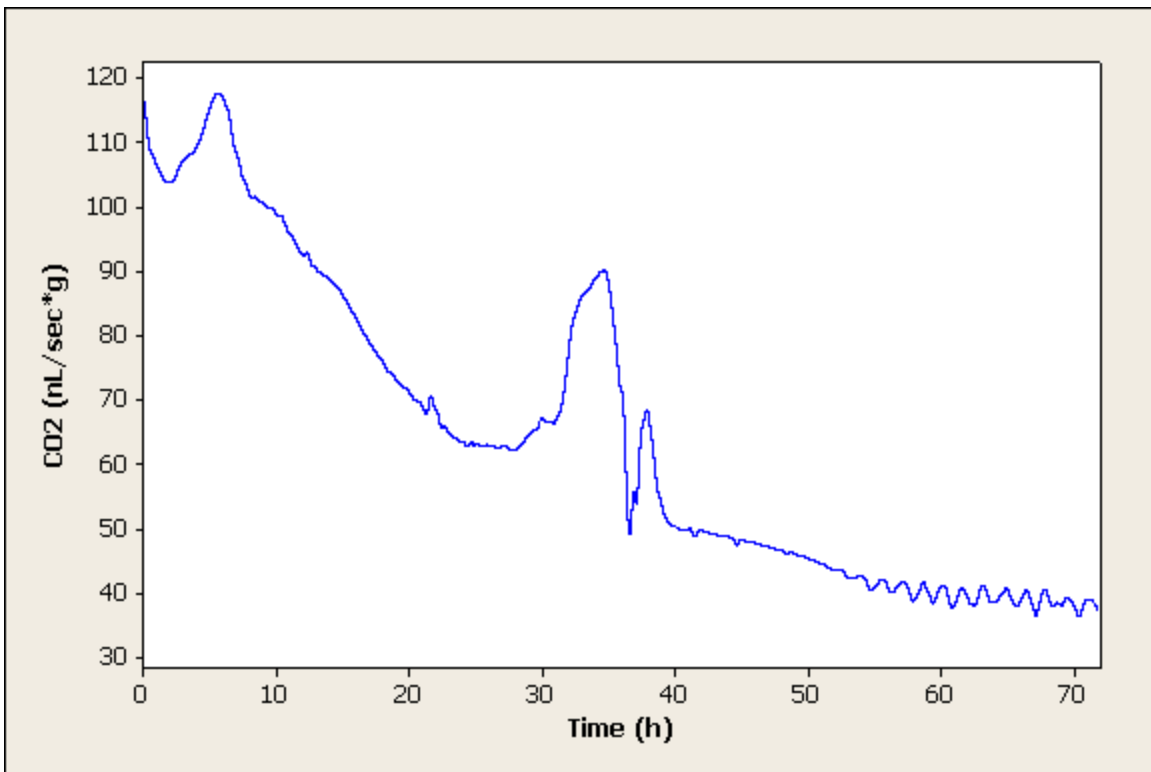


Figure 34: Spline of Individual 2

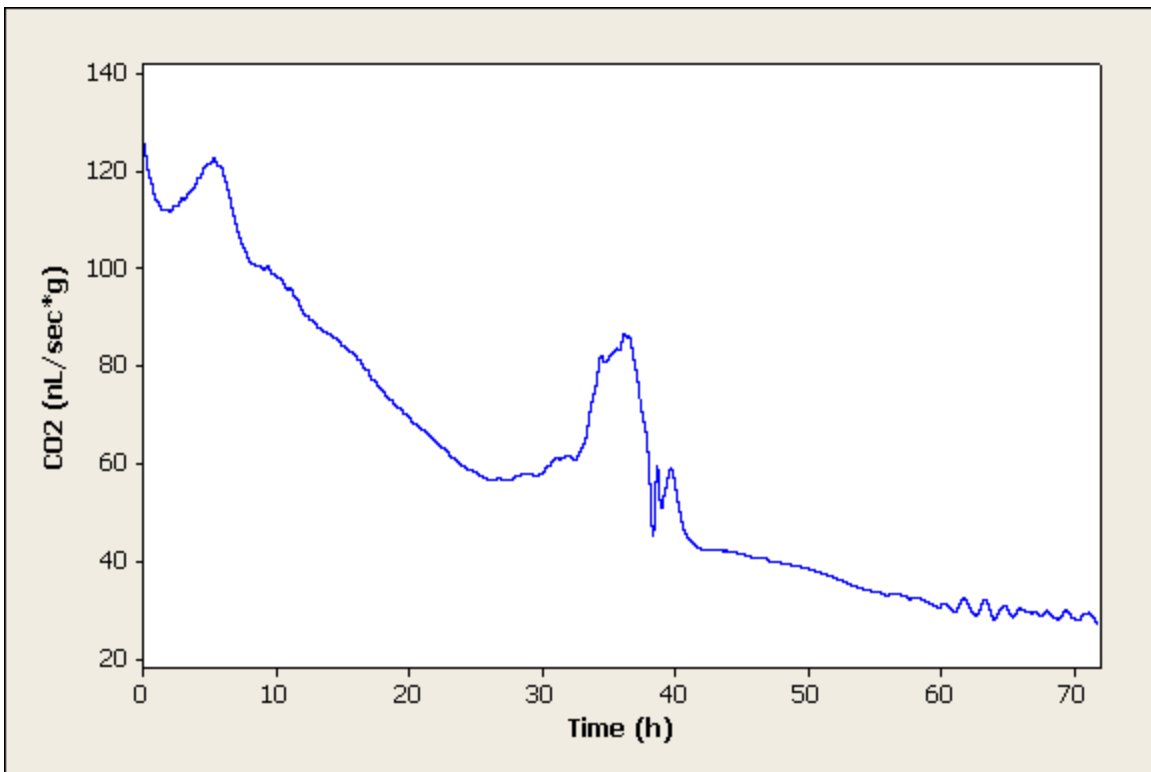


Figure 35: Spline of Individual 3

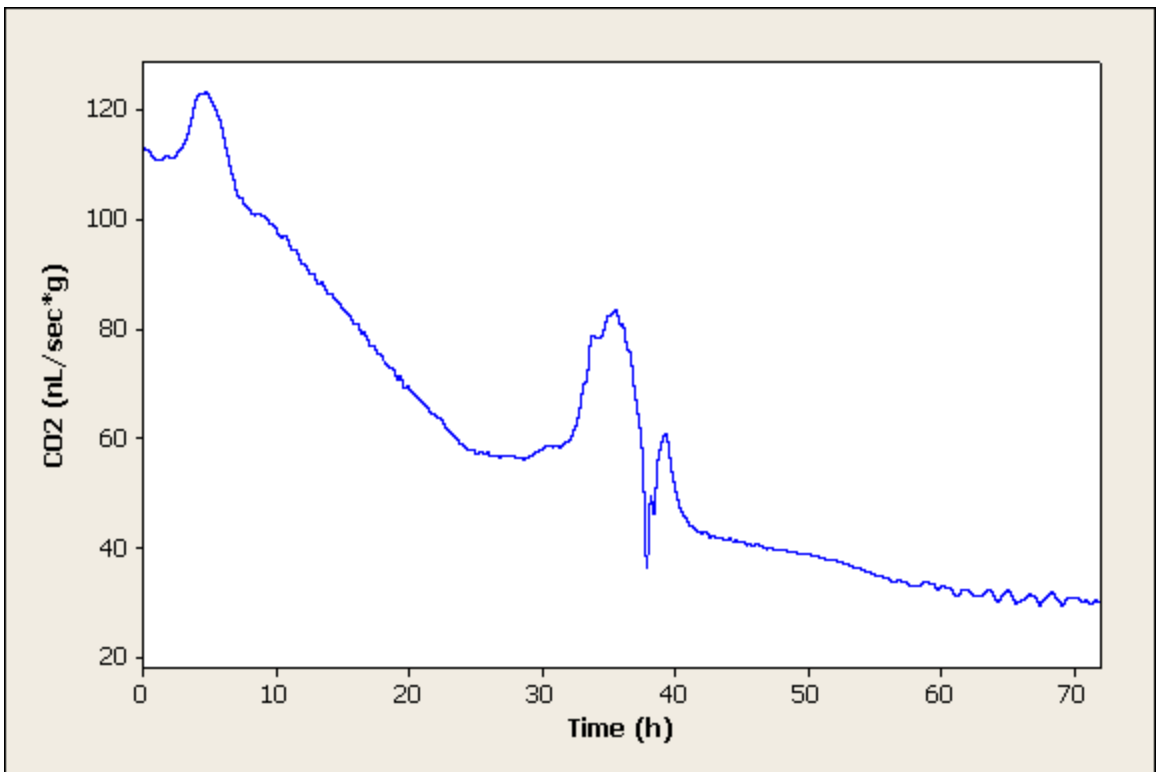


Figure 36: Spline of Individual 4

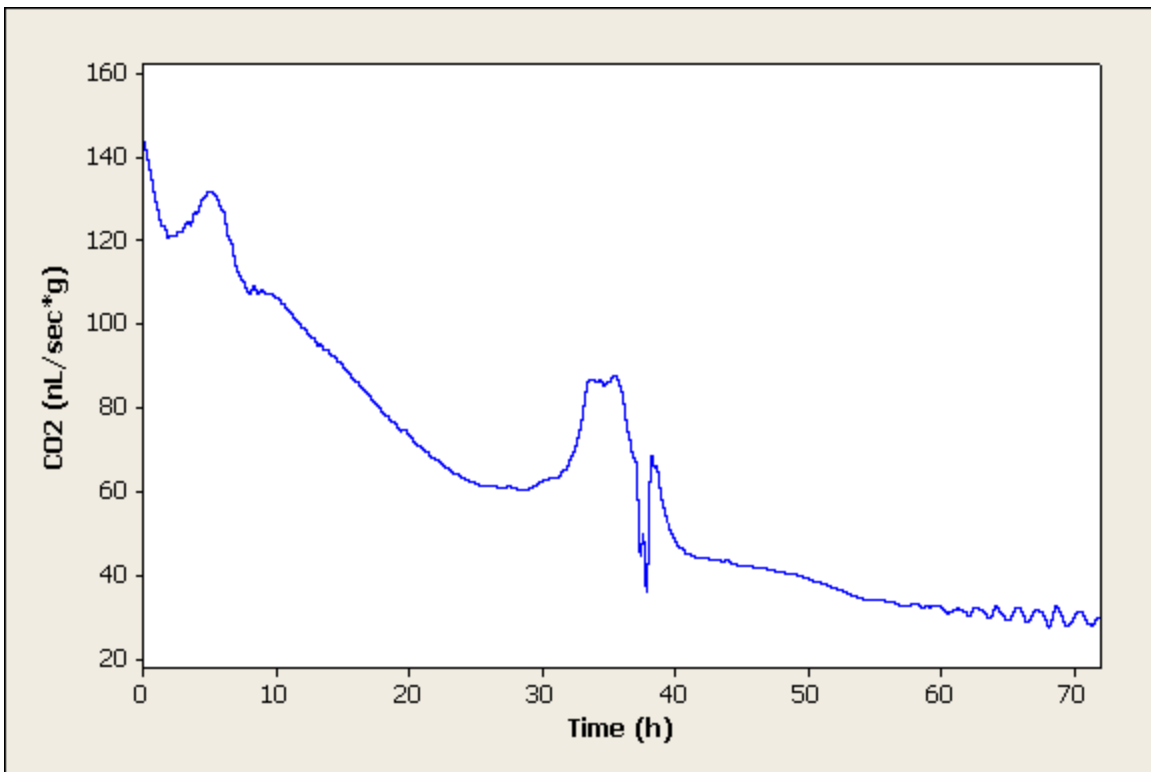


Figure 37: Spline of Individual 5

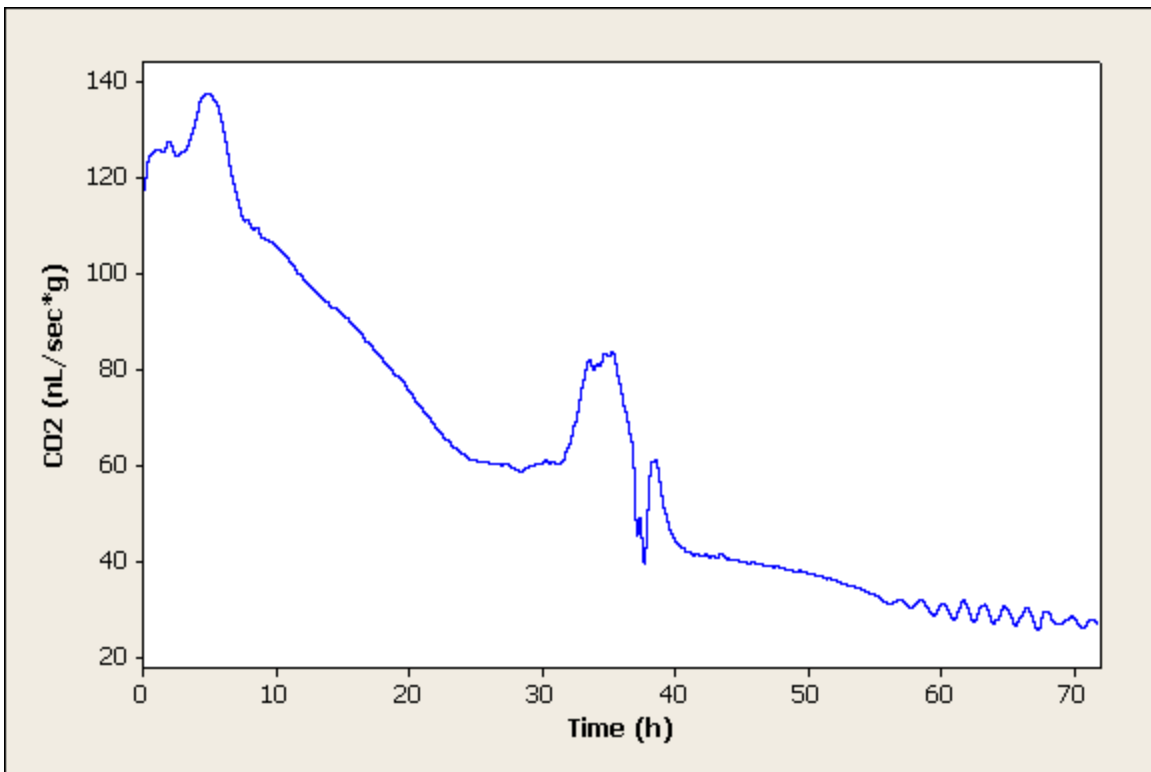


Figure 38: Spline of Individual 6

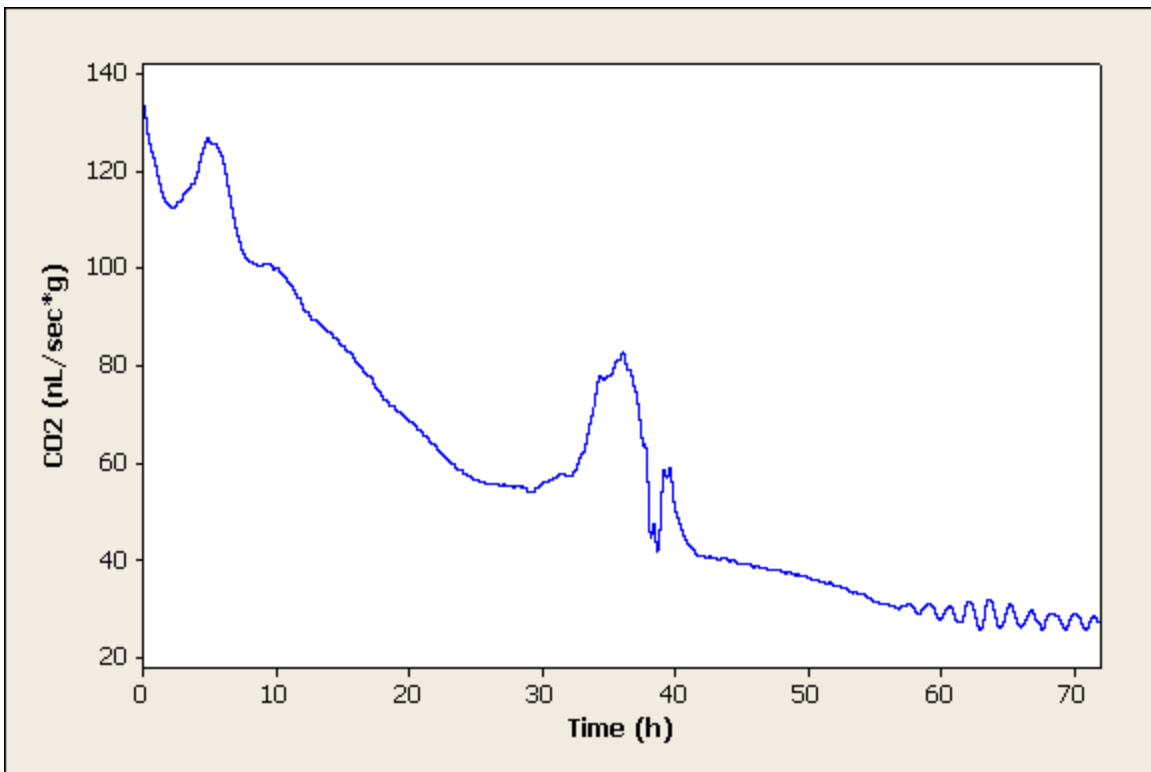


Figure 39: Spline of Individual 7



## Spline Residuals

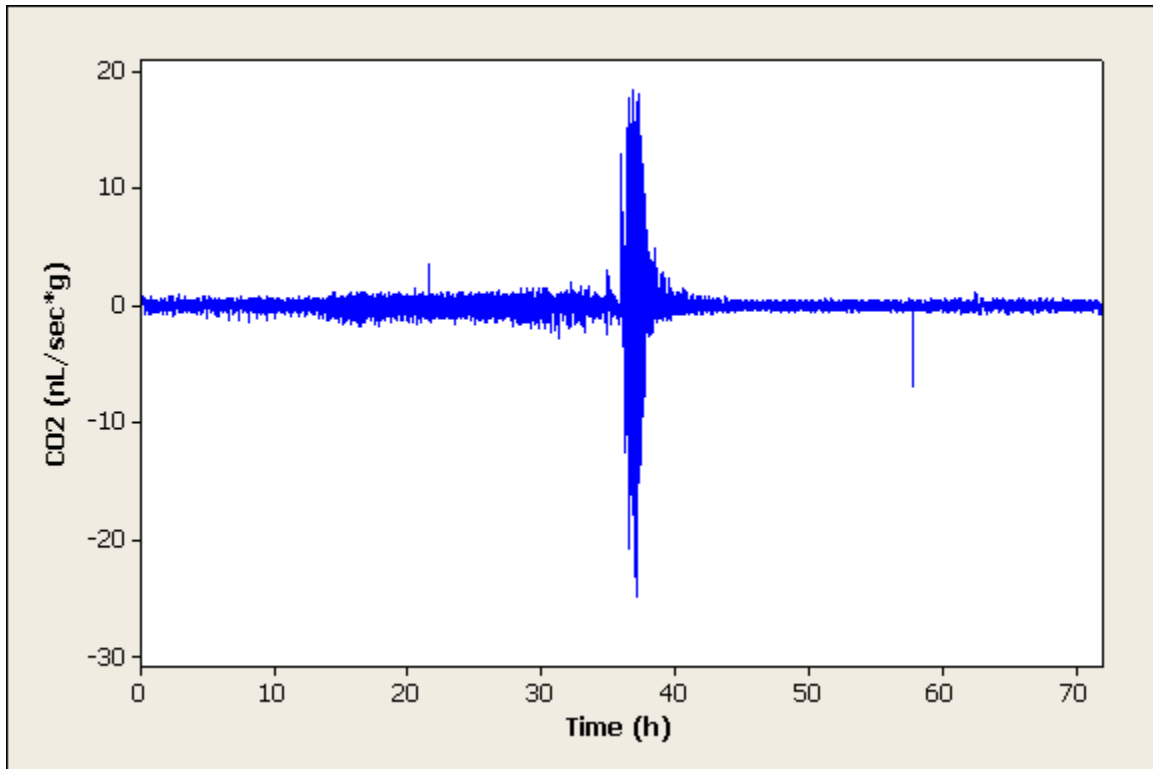


Figure 40: Residuals of the Spline of Individual 1

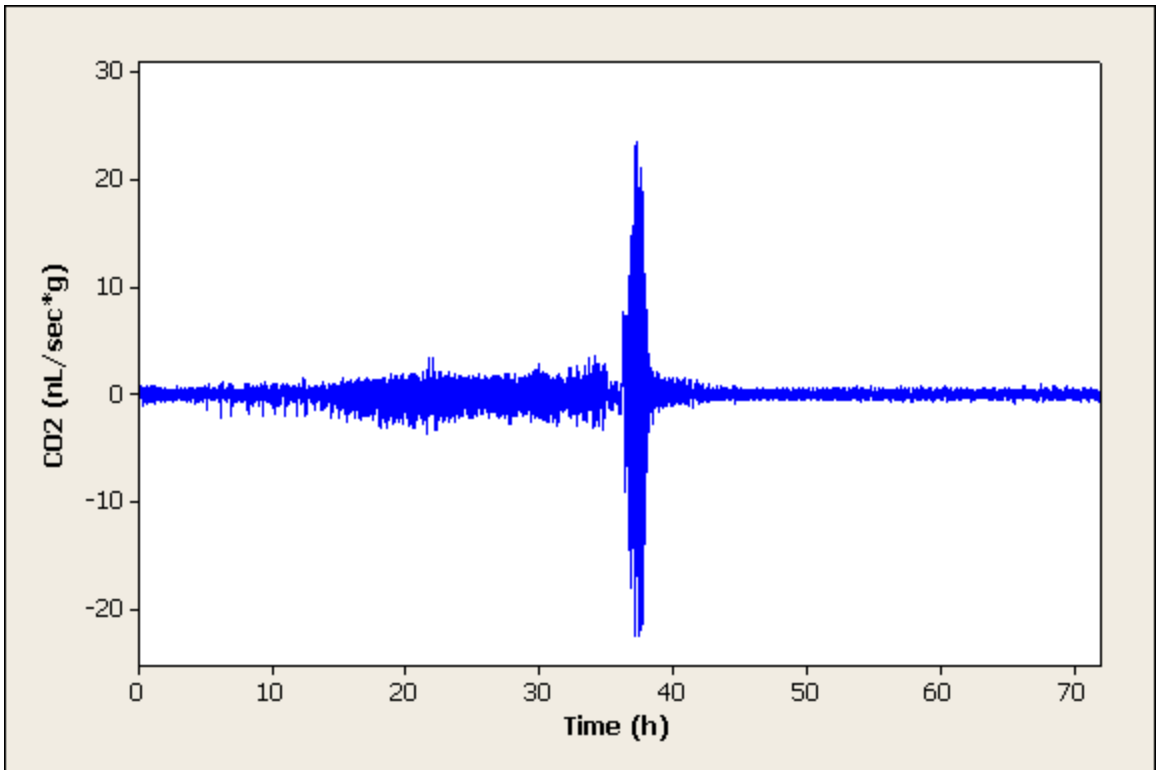


Figure 41: Residuals of the Spline of Individual 2

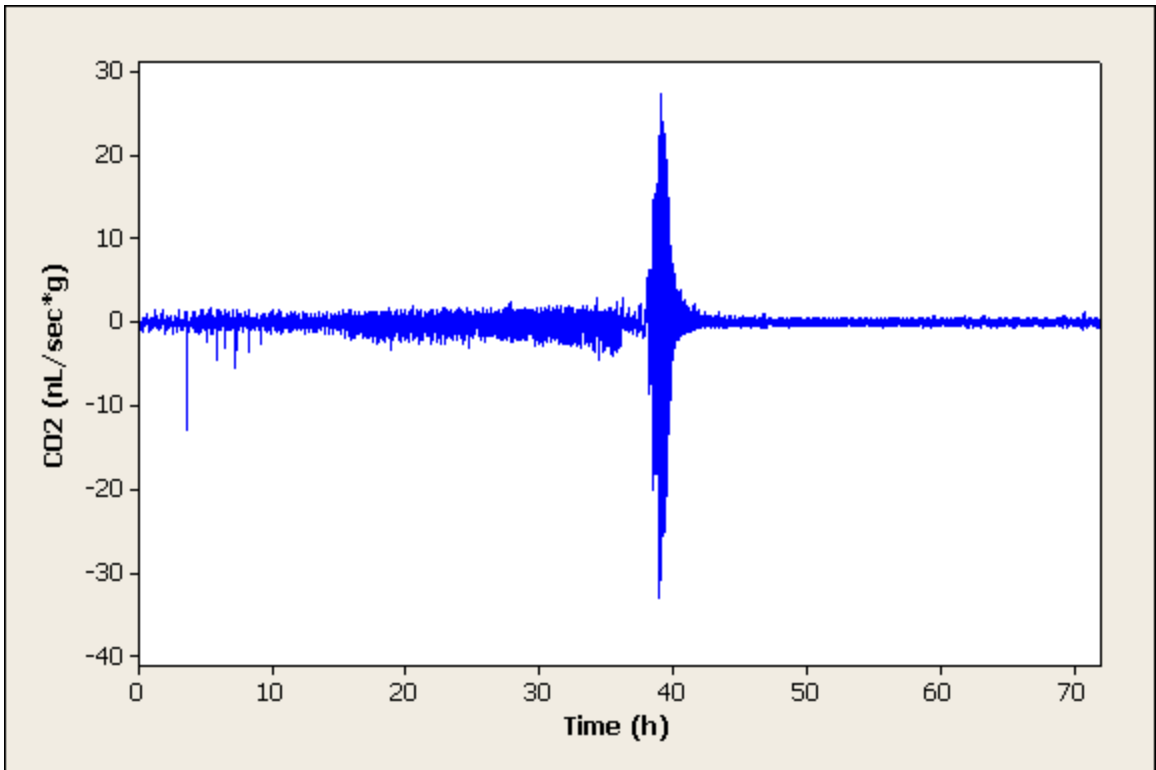


Figure 42: Residuals of the Spline of Individual 3

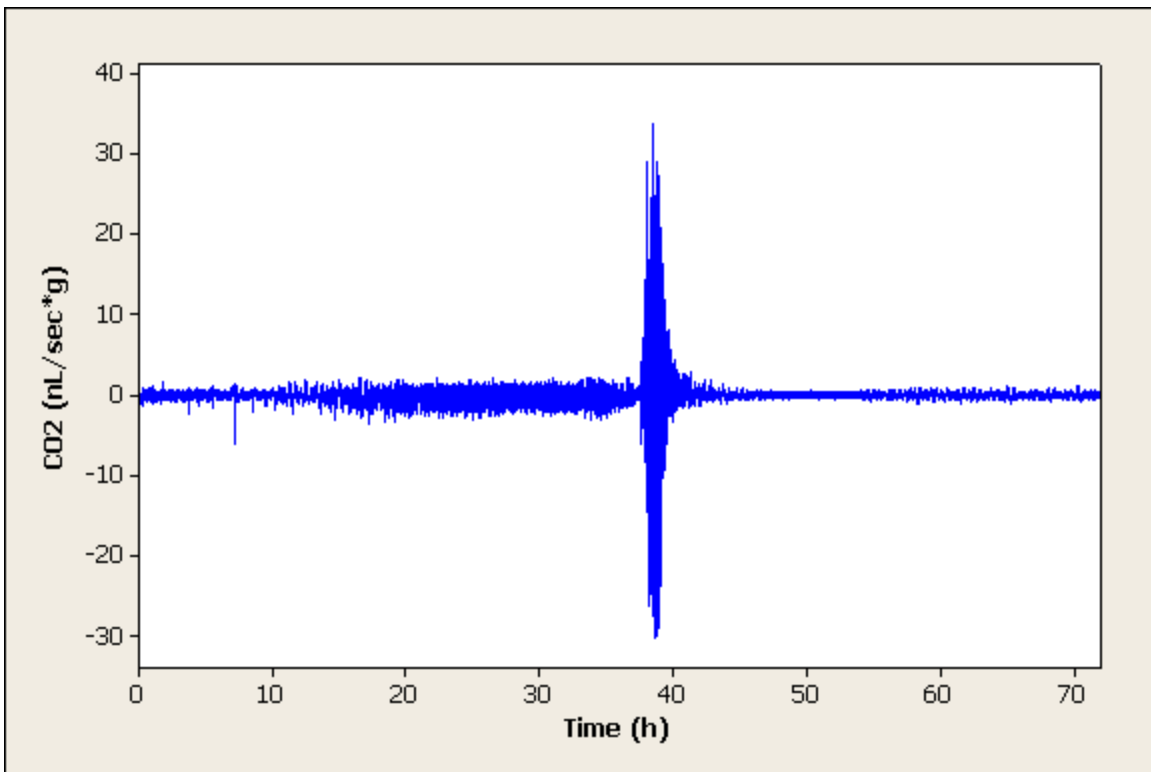


Figure 43: Residuals of the Spline of Individual 4

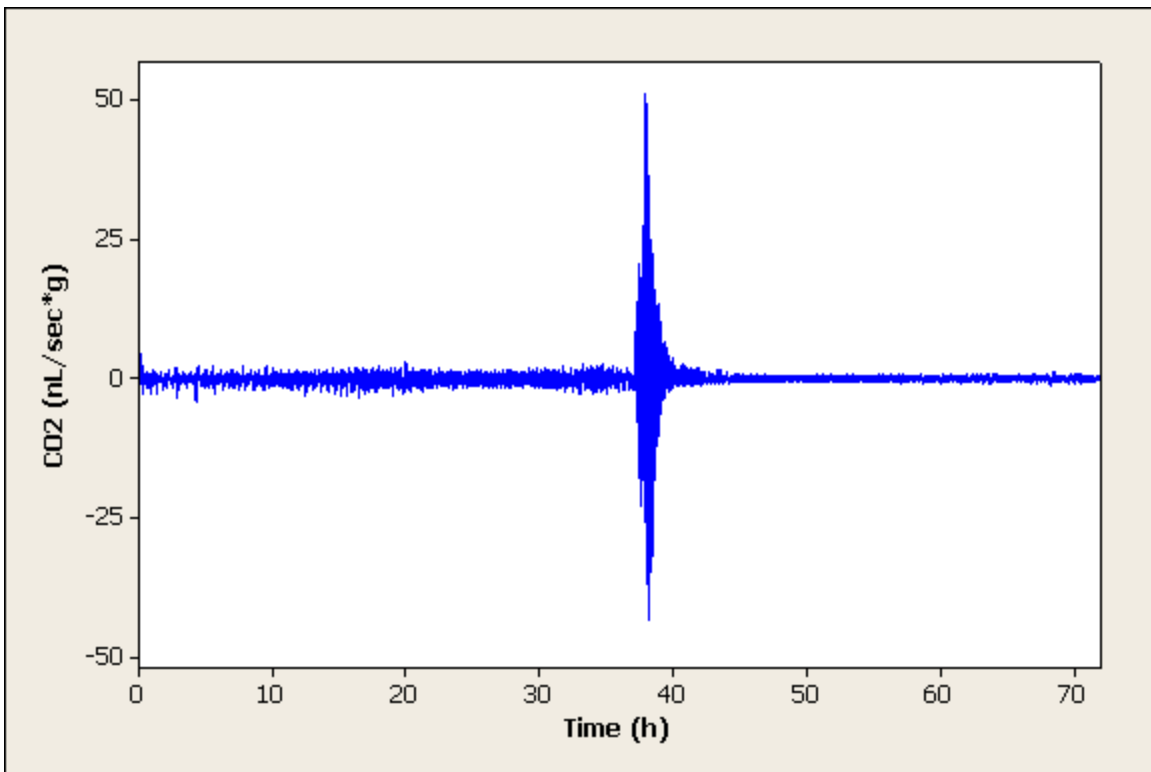


Figure 44: Residuals of the Spline of Individual 5

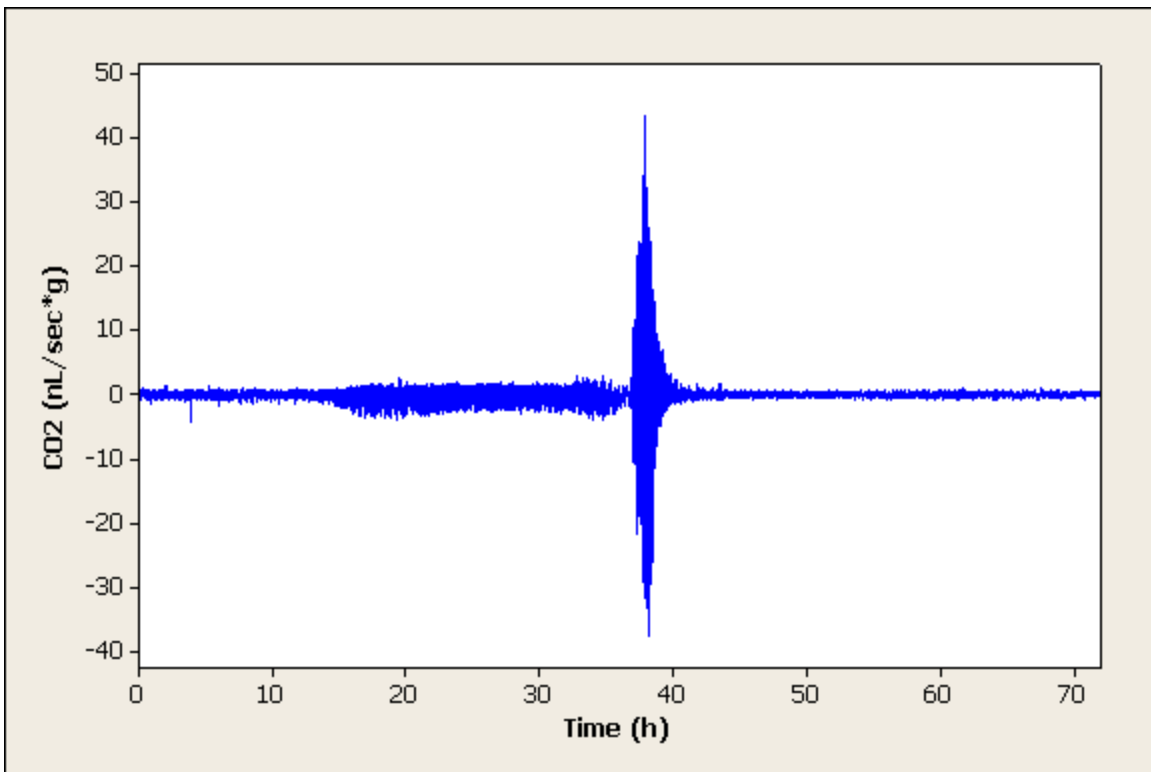


Figure 45: Residuals of the Spline of Individual 6

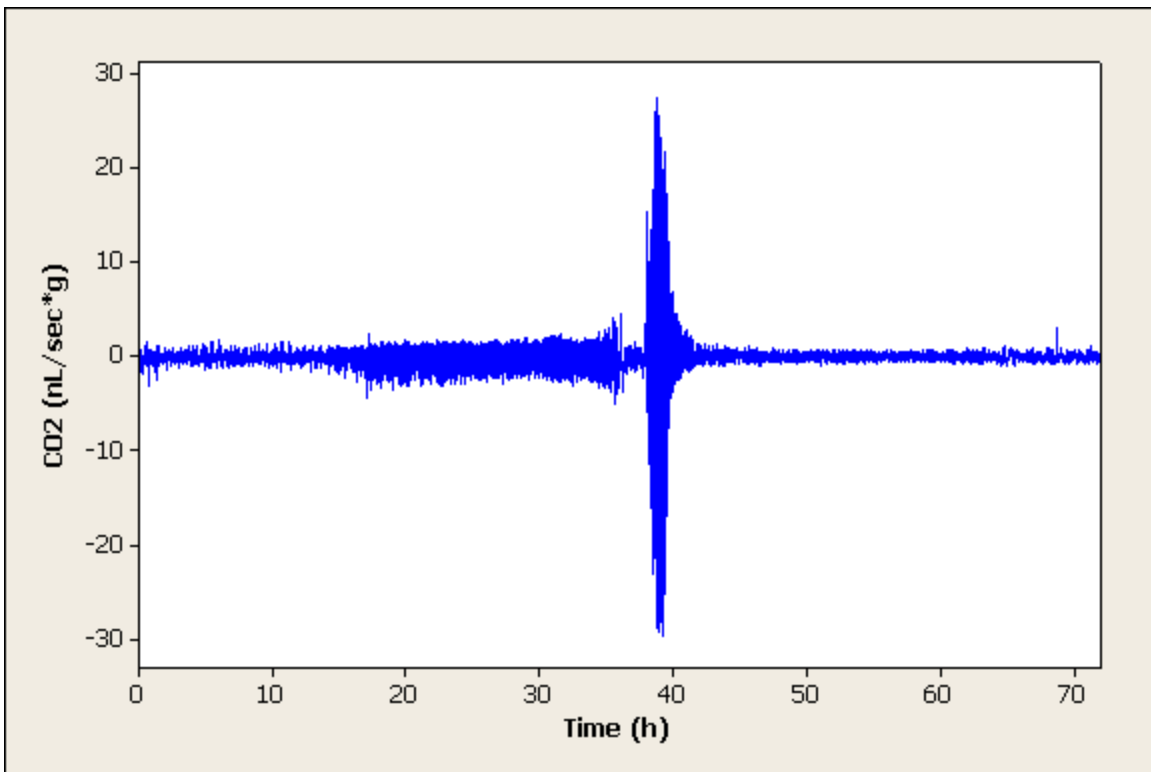


Figure 46: Residuals of the Spline of Individual 7

# MWLR

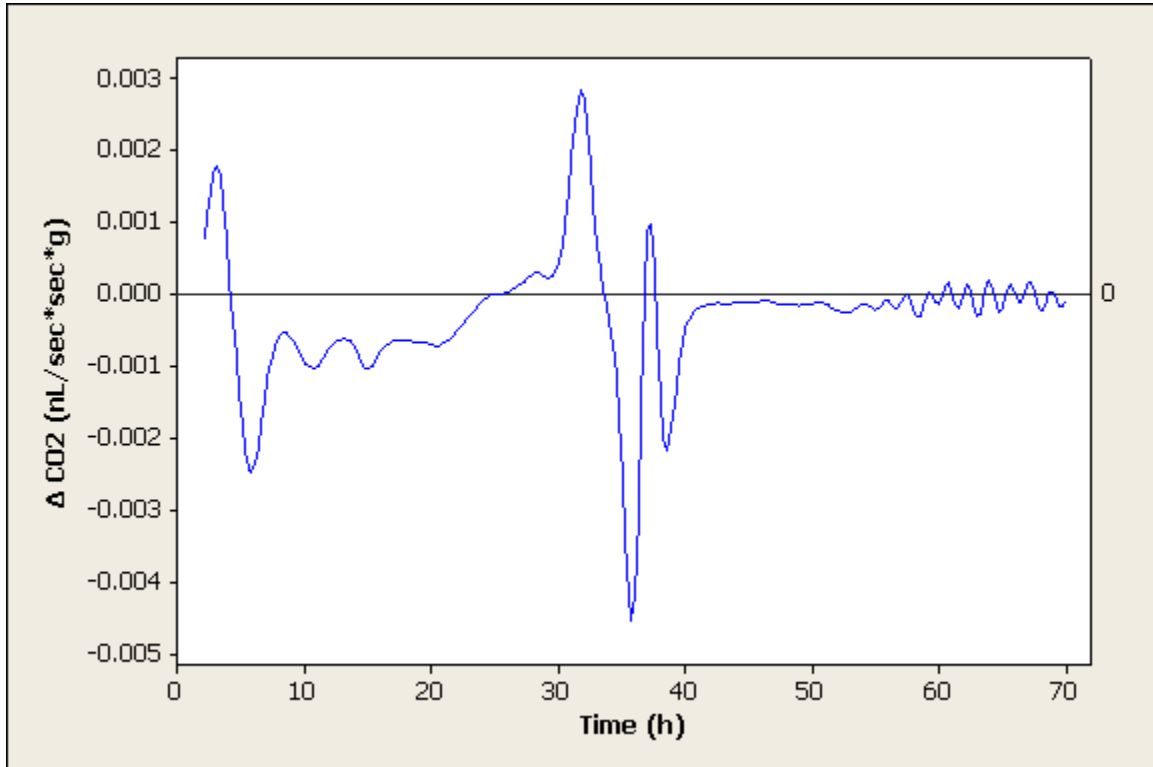


Figure 47: MWLR of Individual 1



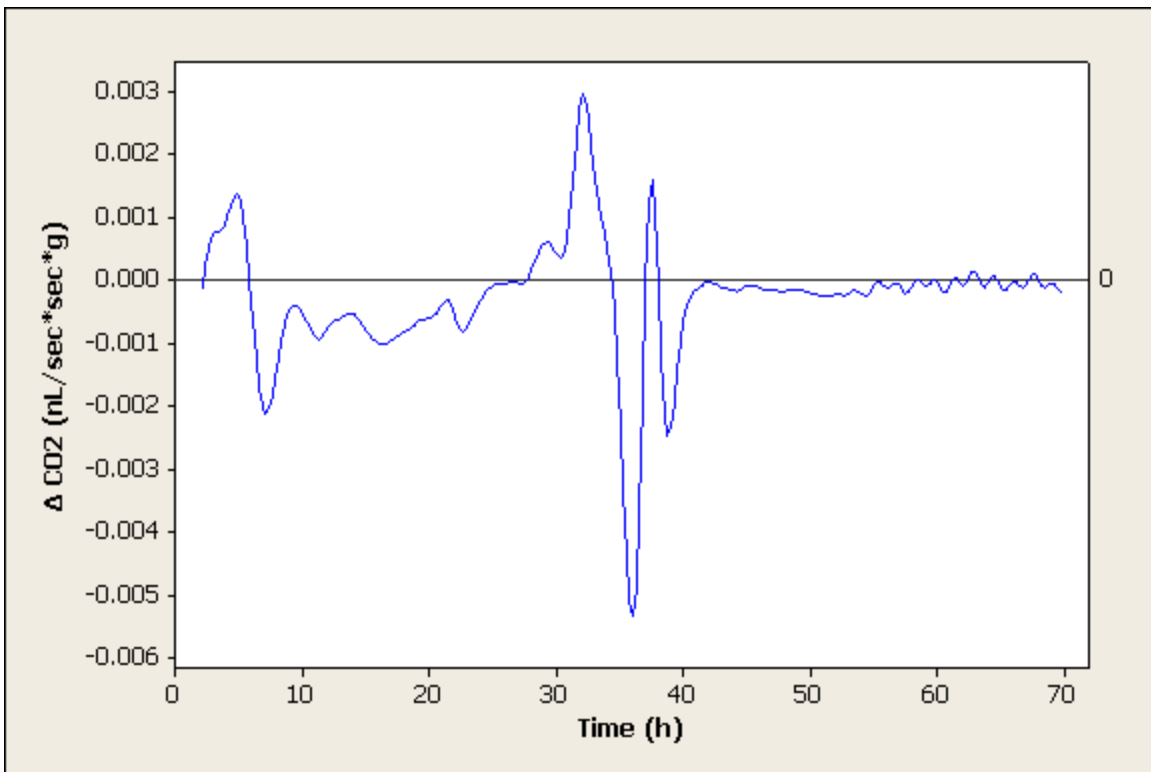


Figure 48: MWLR of Individual 2

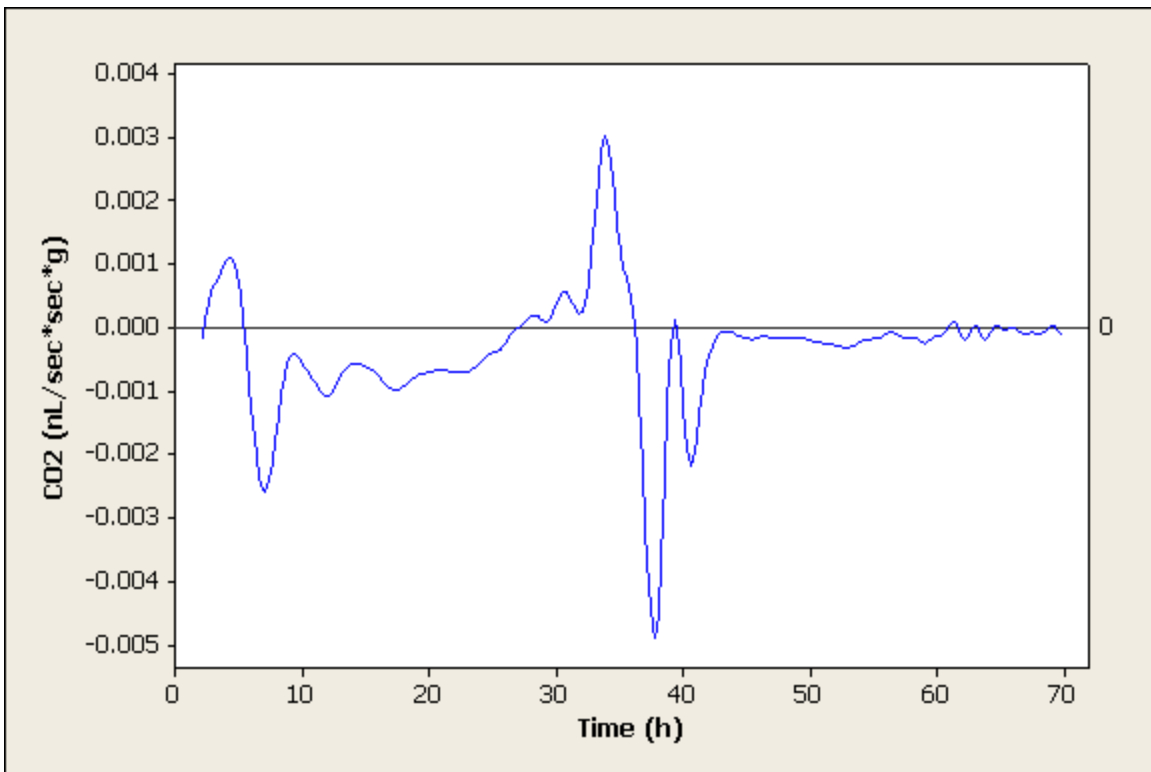


Figure 49: MWLR of Individual 3

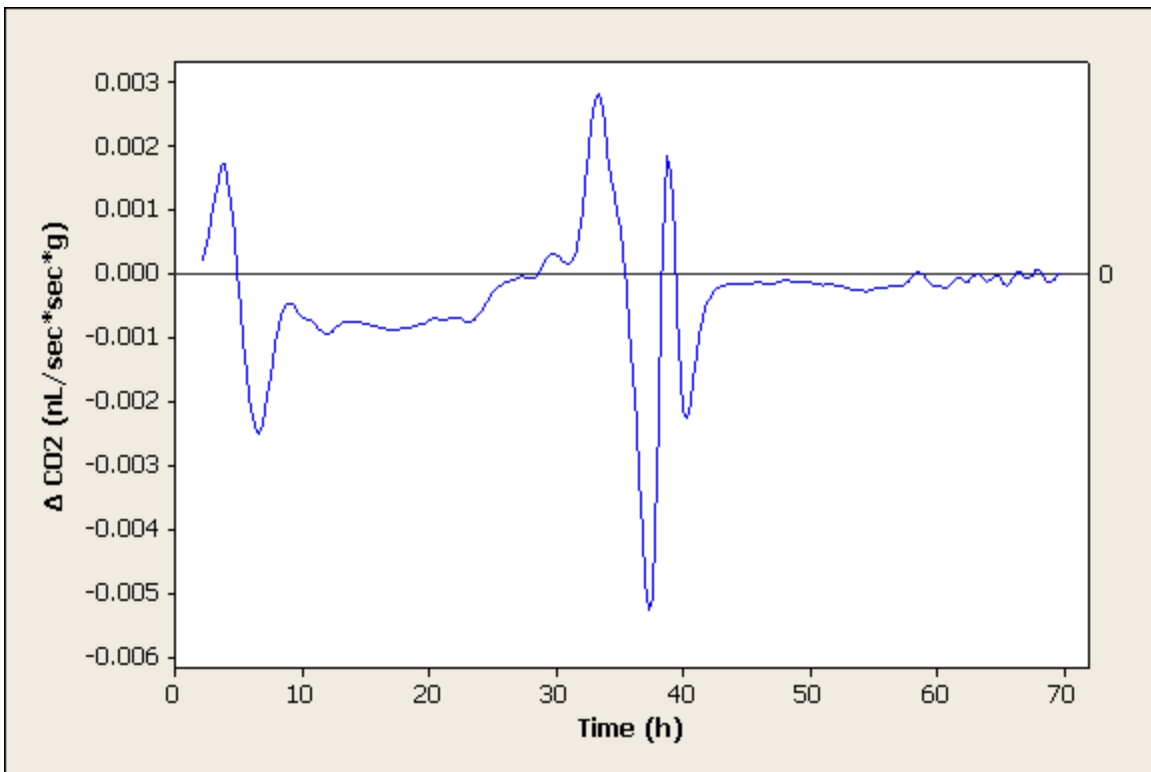


Figure 50: MWLR of Individual 4

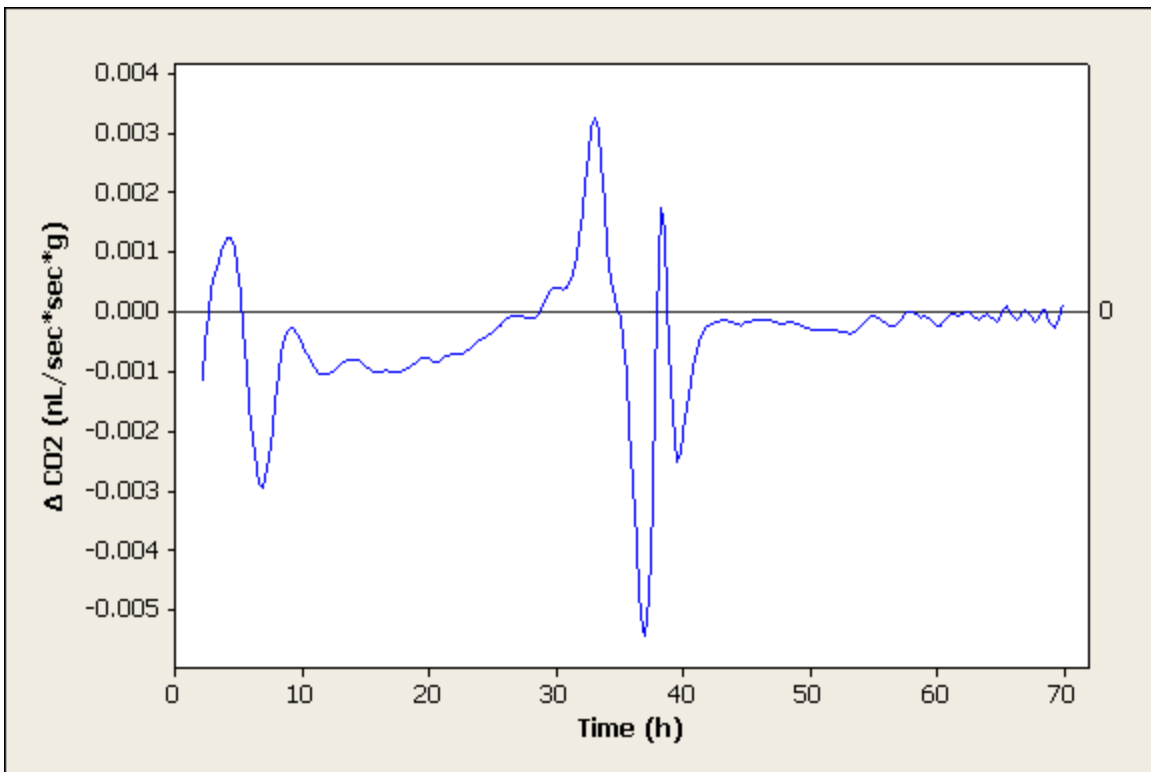


Figure 51: MWLR of Individual 5

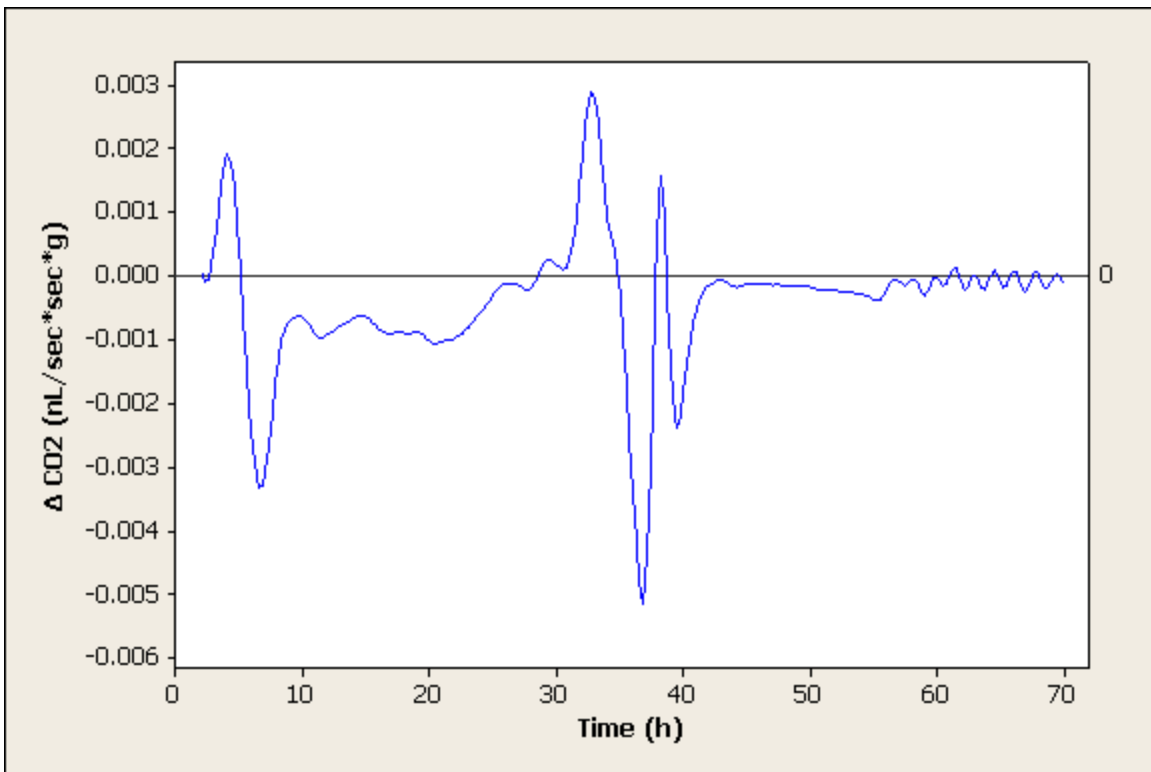


Figure 52: MWLR of Individual 6

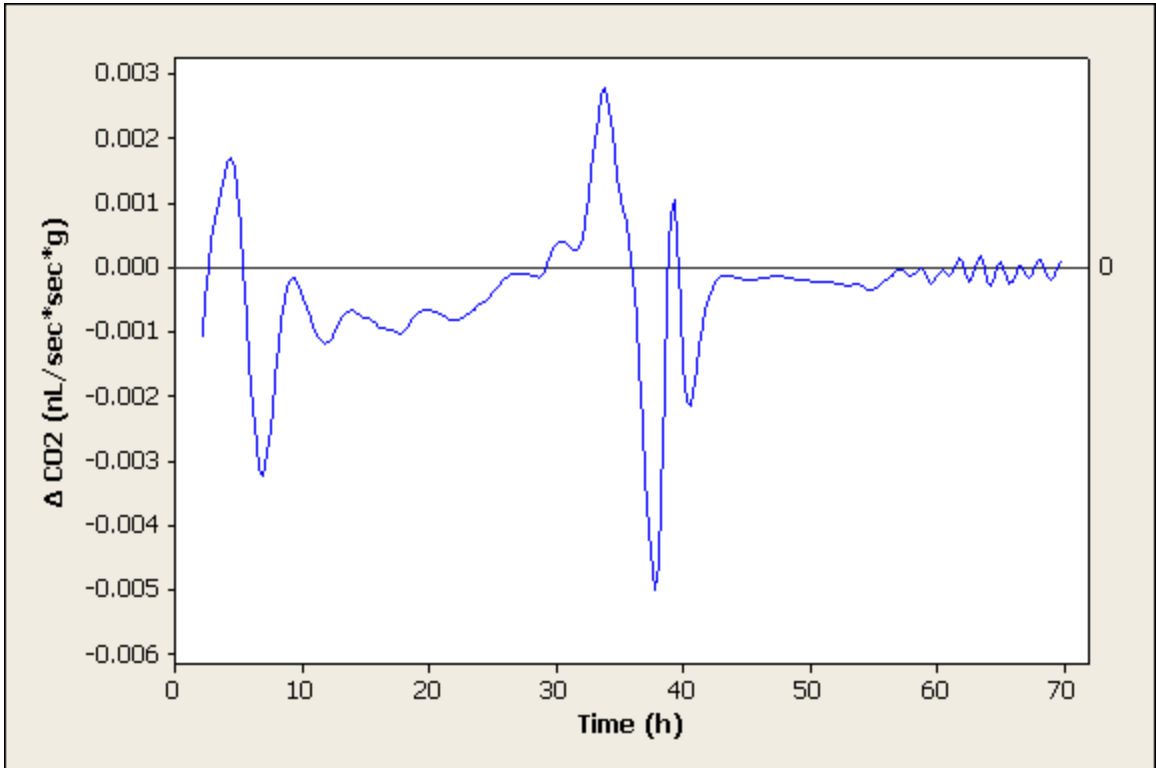


Figure 53: MWLR of Individual 7

## Spectrograms

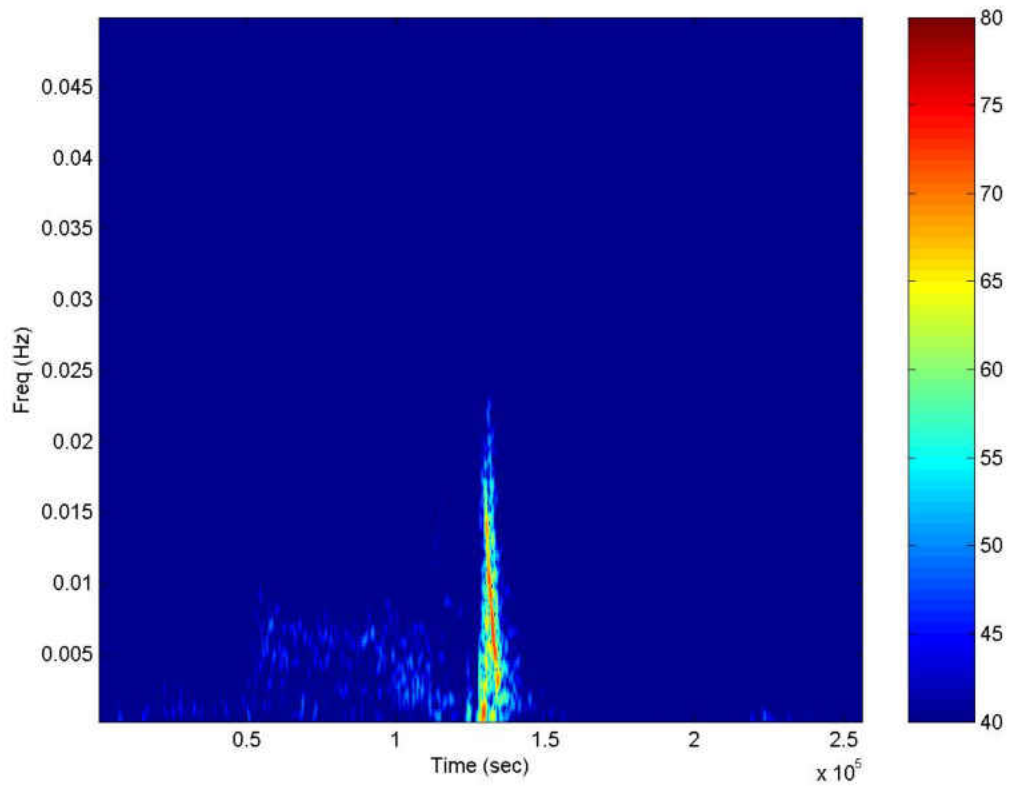


Figure 54: Spectrogram of Individual 1

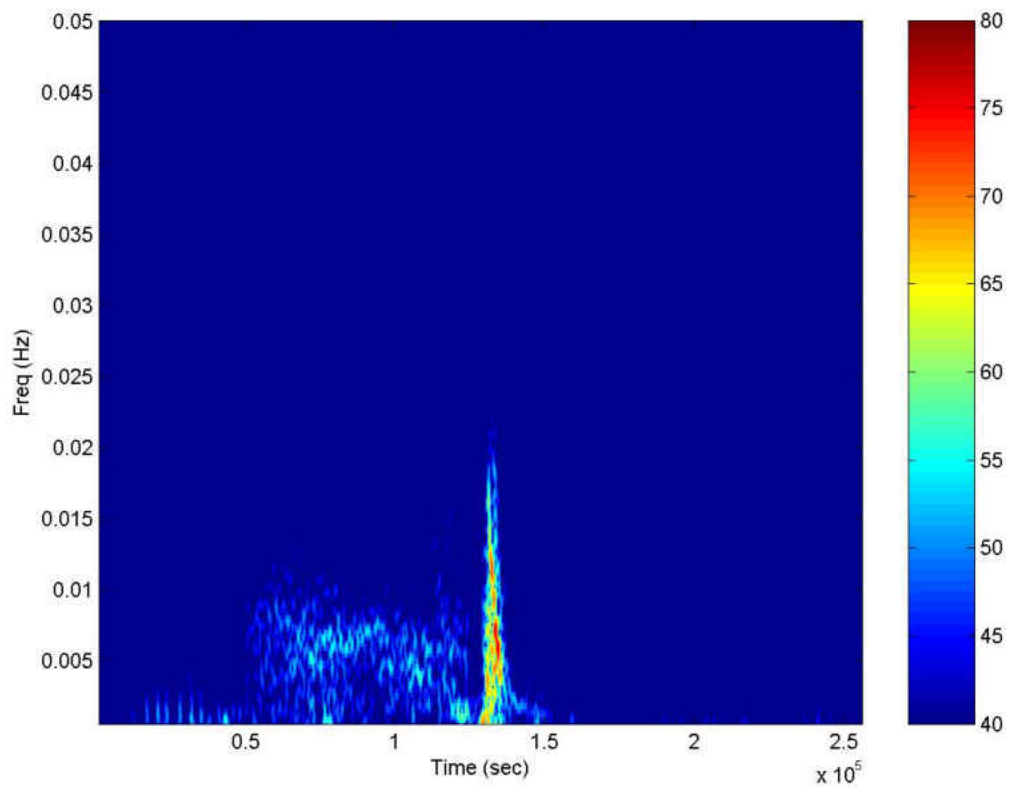


Figure 55: Spectrogram of Individual 2



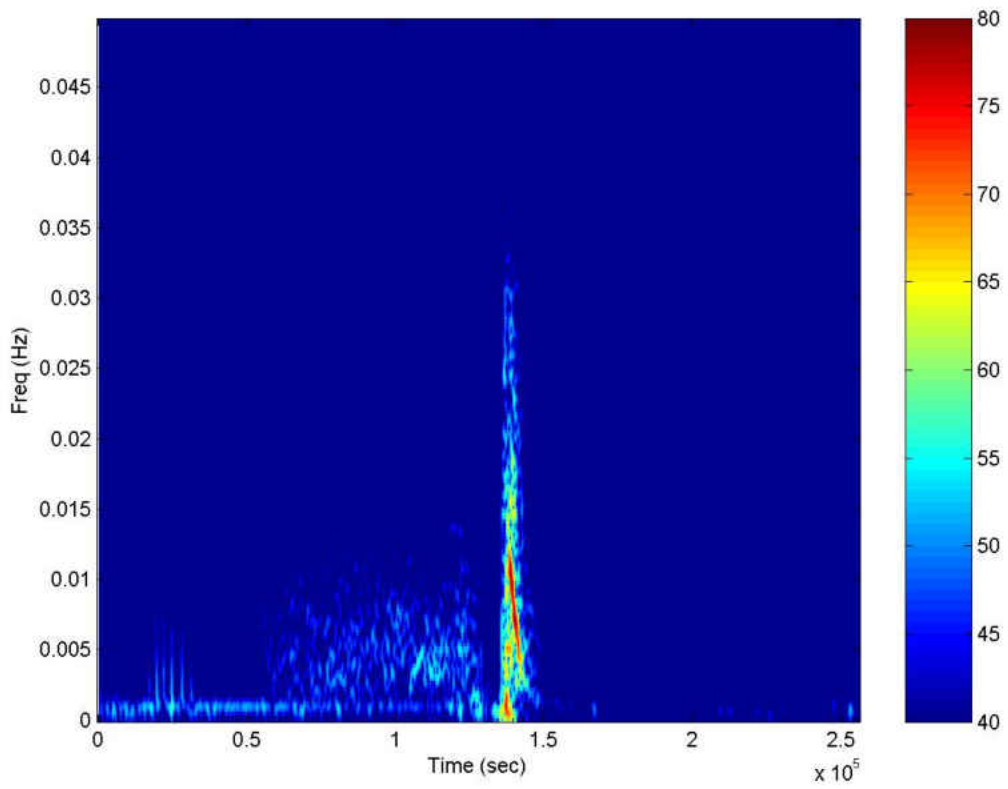


Figure 56: Spectrogram of Individual 3

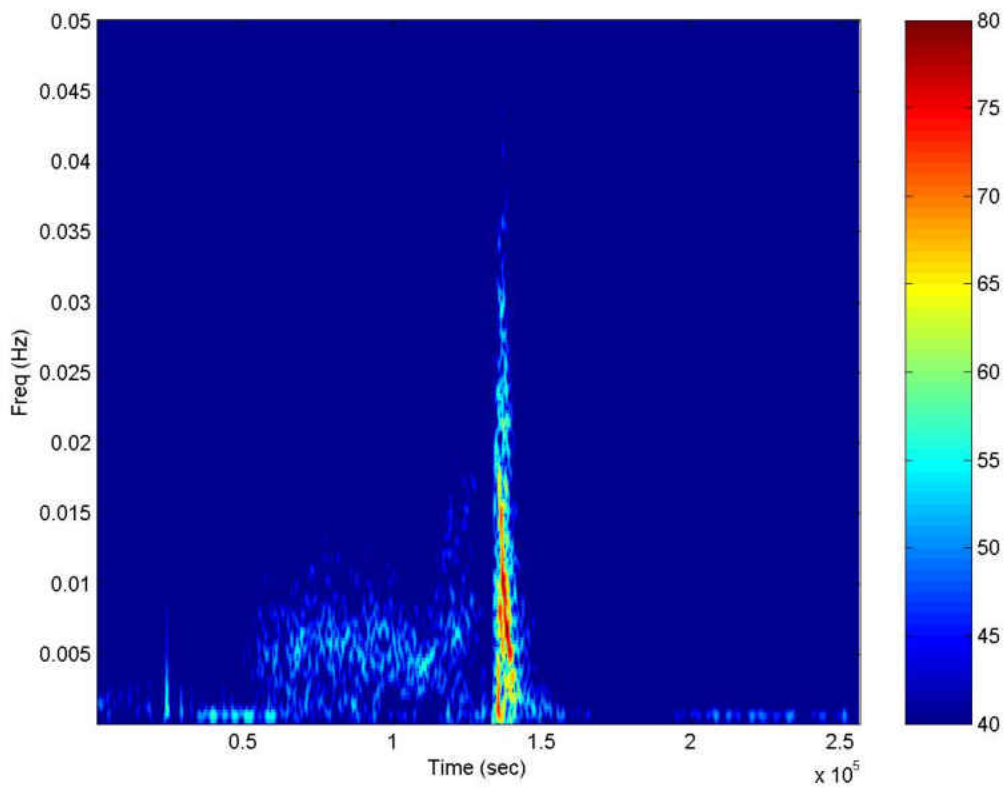


Figure 57: Spectrogram of Individual 4

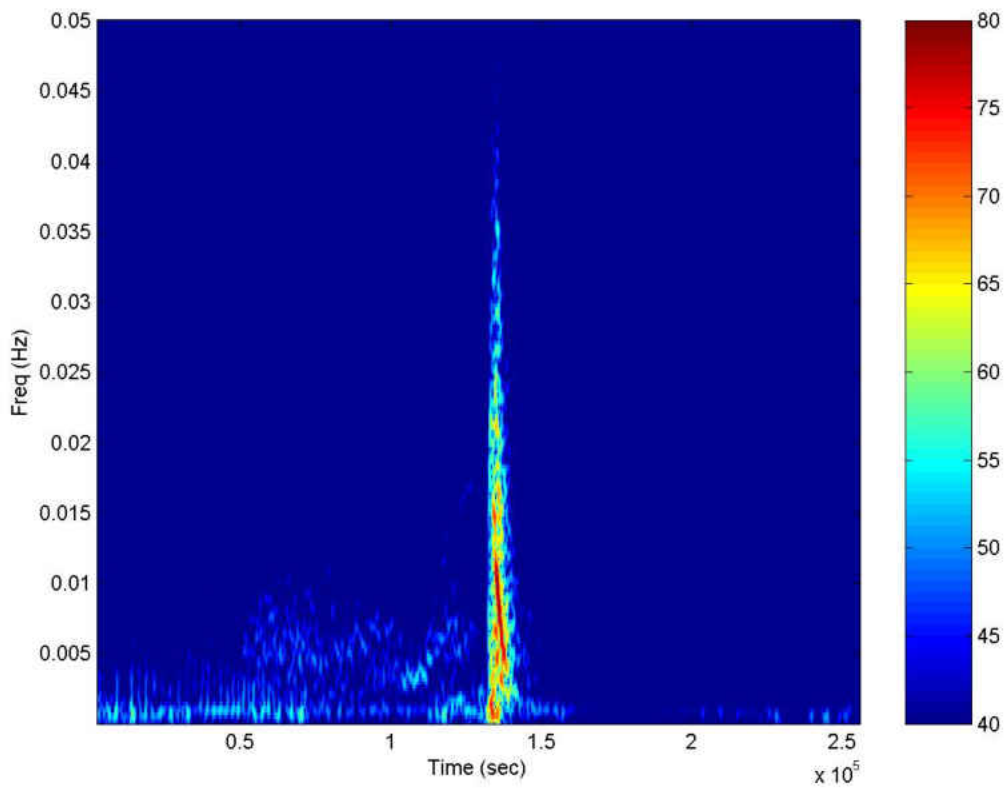


Figure 58: Spectrogram of Individual 5

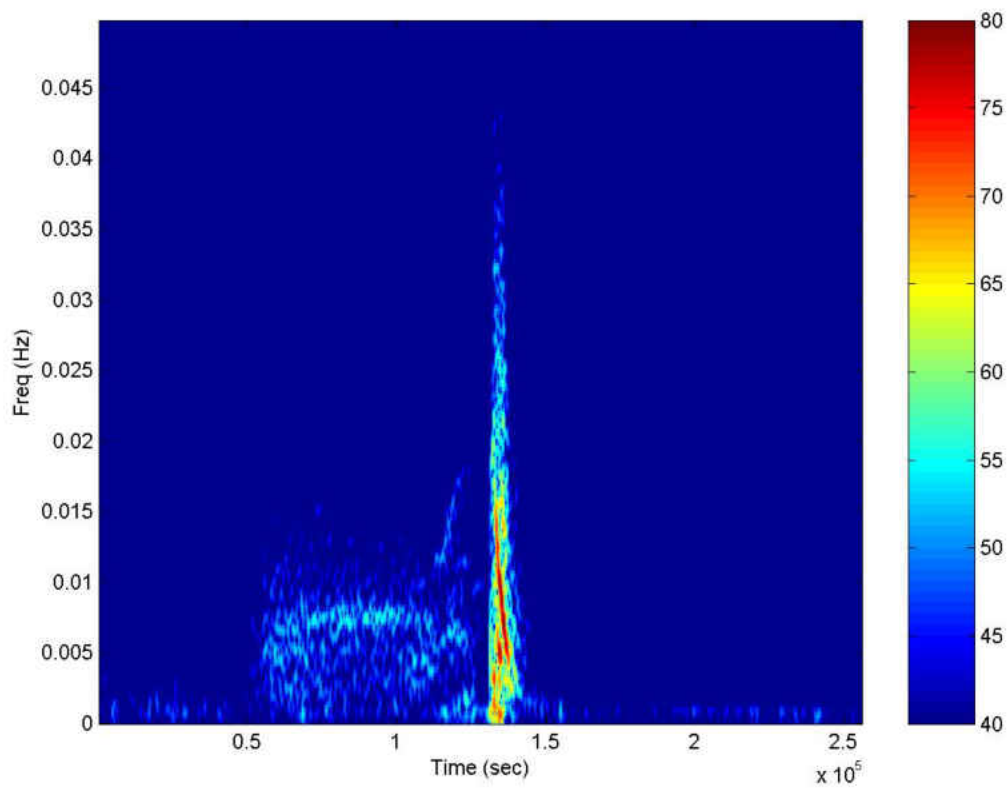


Figure 59: Spectrogram of Individual 6

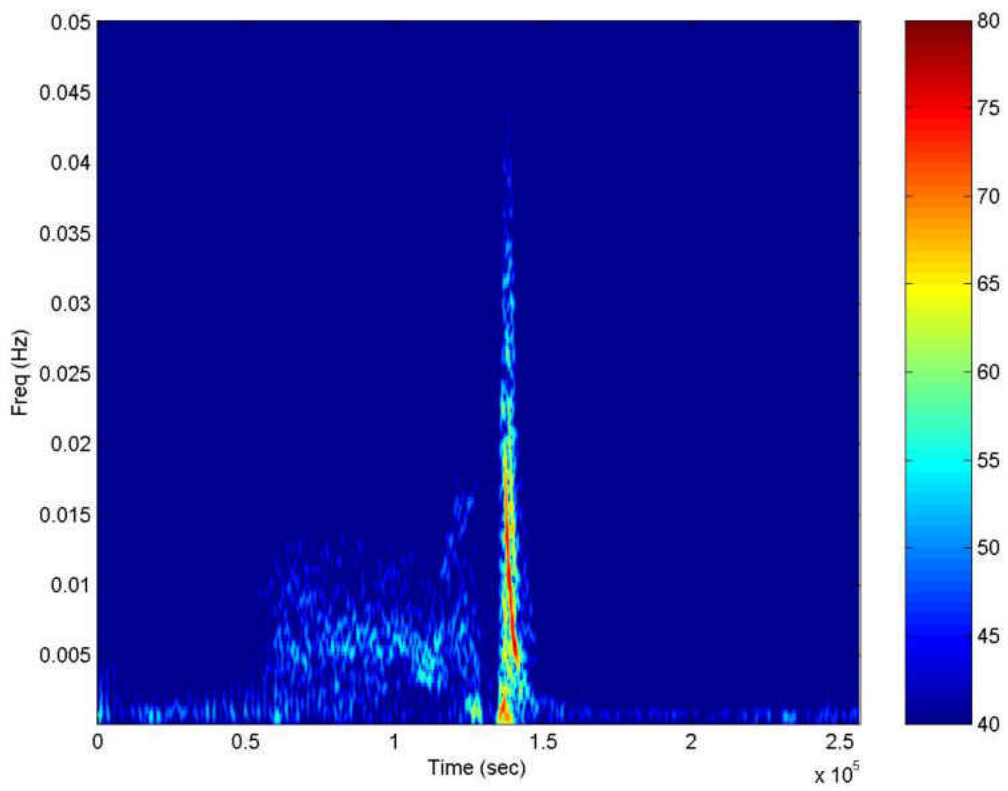


Figure 60: Spectrogram of Individual 7

## Appendix B

### SplinerD

SplinerD is a program designed to create and evaluate a spline or its derivatives and then save the results as a text file. SplinerD has four arguments.  $A$ , the first argument, is expected to be a two column Matrix with the first column representing the increasing time increments and the second representing values of  $A(t)$ . The second argument  $deg$  is expected to be a positive integer corresponding to the degree of the polynomials to be used in the spline. The third argument,  $dir$ , corresponds to the derivate that should be computed. For example if the second derivative is desired a 2 should be entered. Enter 0 if the spline itself is desired. The final argument,  $name$ , is expected to be a combination of text and numbers (stay away from symbols) and will be the root of the name of the file this program saves. The full file name will be “name”.txt and will be saved in the same folder as the worksheet being used. Maple output is a one column Matrix of the evaluated values.

```
SplinerD:=proc(A,deg,dir,name)
local g,m,n,o,f,F,i,S,M;
    g:=CurveFitting[Spline](A,x,degree=deg):
    if dir < 1 then
        f:=g
    else
        f:=diff(g,x$dir);
```

```

end if;

F:=unapply(f,x):

m:=A[1,1];

n:=linalg[rowdim](A);

o:=A[n,1];

S:=[seq(evalhf(F(a)), a = m..o)];

M:=Matrix(S);

ExportMatrix(cat(name, ".txt"),M,target=delimited,delimiter="\ t"):

M;

end proc:

```

## MWLR

MWLR is a program designed to compute a moving weighted linear regression and export the result to a text file. WMLR has four arguments. The first argument  $y$  is expected to be a two column Matrix with the first column representing the increasing time increments and the second representing values from either the time series or a block averaging.  $W$ , the second argument, is expected to be an  $n$  by  $n$  matrix with the weights along the main diagonal. How far the algorithm moves down the time series before recomputing the weighted linear regression is the third argument: `hop`. The final argument, `name`, is expected to be a combination of text and numbers (stay away from symbols) and will be the root of the name of the file this program saves.

The full file name will be “name”MWLR.txt and will be saved in the same folder as the worksheet being used. Maple output is a one dimensional Array of the MWLR.

```
MWLR:=proc(y,W,hop,Name)
localw,q,n,A,i,M,Y,X,Xp,StepOne,StepTwo,StepThree,StepFour,FinalVector,a;

w:=LinearAlgebra[RowDimension](W);
q:=LinearAlgebra[RowDimension](y);
n:=q-w;
A:=<seq(0,i=1..n,hop)>;
for i from 1 to n by hop do
    Y:=Matrix([seq([y[d,2]],d=i..i+w-1)]);
    X:=Matrix([seq([1,y[d,1]],d=i..i+w-1)]);
    Xp:=Transpose(X);
    StepOne:=LinearAlgebra[MatrixMatrixMultiply](Xp,W);
    StepTwo:=LinearAlgebra[MatrixMatrixMultiply](StepOne,X);
    StepThree:=LinearAlgebra[MatrixInverse](StepTwo);
    StepFour:=LinearAlgebra[MatrixMatrixMultiply](StepOne,Y);
    FinalVector:=LinearAlgebra[MatrixMatrixMultiply](StepThree,StepFour);
    a:=FinalVector[2,1];
    A[i]:=a;
end do;

M:=Matrix(A);
```



```

ExportMatrix(cat(Name, "MWLR", ".txt"),M,target=delimited,delimiter="\ t"):
A;
end proc:

```

### MaxT and MinT

MaxT and MinT are very similar programs designed to compliment MWLR by finding local maxima and minima of its output respectively. Both programs have the same three arguments. V the second argument is intended to be the output of MWLR. P its first argument is expected to be a one dimensional array of the same length as the V that represents the times each MWLR value corresponds to. The correct values for P are going to depend on the length of the weighting function, choice of hop size and sampling frequency of the data entered into MWLR. The final argument, name, is expected to be a combination of text and numbers (stay away from symbols) and will be the root of the name of the file this program saves. The full file name will be "name"MAX.txt or "name"Min.txt and will be saved in the same folder as the worksheet being used. Maple output is a two column Matrix of the MWLR local maxima or minima and their position in time.

```
MaxT:=proc(P,V,Name)
```

```
local M,n,A,i,a1,a2,a;
```

```
    n:=LinearAlgebra[Dimension](V)-1;
```

```
    A:=[];
```

```

for i from 2 to n do
    if V[i] > V[i-1] then
        if V[i] > V[i+1] then
            a1:=P[i];
            a2:=V[i];
            A:=[op(A),[a1,a2]];
        end if;
    end if;
end do;

M:=Matrix(A);

ExportMatrix(cat(Name, "MAX", ".txt"),M,target=delimited,delimiter="\ t"):

A;

end proc:

```

Similarly

```

MinT:=proc(P,V,Name)
local M,n,A,i,a1,a2,a;

n:=LinearAlgebra[Dimension](V)-1;

A:=[];

for i from 2 to n do

```

```

    if V[i] < V[i-1] then
        if V[i] < V[i+1] then
            a1:=P[i];
            a2:=V[i];
            A:=[op(A),[a1,a2]];
        end if;
    end if;
end do;

M:=Matrix(A);

ExportMatrix(cat(Name, "MIN", ".txt"),M,target=delimited,delimiter="\ t");

A;

end proc:

```

### ZeroI

ZeroI is a program designed to compliment MWLR by finding intervals that contain zeros. This program has two arguments. V the second argument is intended to be the output of MWLR. P its first argument is expected to be a one dimensional array of the same length as the V that represents the times each MWLR value corresponds to. The correct values for P are going to depend on the length of the weighting function, choice of hop size and sampling frequency of the data entered into MWLR.

No text file is saved. Maple output is a two column Matrix of the values before and after a zero.

```
ZeroI:=proc(P,V)
local A,n,i,x,a,b,c,d,M;

  A:=[];
  n:=LinearAlgebra[Dimension](V)-1;
  for i from 1 to n do
    x:=V[i]*V[i+1];
    if x < 0 then
      c:=P[i];
      d:=P[i+1];
      A:=[op(A),[c,d]];
    end if;
  end do;
  M:=Matrix(A);
  M;
end proc;
```

spec

The program spec was designed to compute the spectrogram and save it as a

text file. The program has five arguments. The first argument, Sample is expected to be a one column Matrix of Real numbers representing the data the spectrum is going to be computed for. FFTN is the number of Fourier Coefficients that will be generated. This value should be a power of two for fastest computation. Note that only FFTN/2 will be reported as this program was set up with Real numbers in mind. N is the length of the hamming window to be used to compute the STFT and must be less than FFTN. H is the hop distance or how far the window will be advanced for successive computations of the STFT. The final argument, name, is expected to be a combination of text and numbers (stay away from symbols) and will be the root of the name of the file this program saves. The full file name will be spec“name”.txt and will be saved in the same folder as the worksheet being used. Note that with(Matlab) is needed before the program is loaded.

```
with(Matlab):
```

```
spec:=proc(Sample,FFTN,N,H,Name)
```

```
local b,x,a,y,A,w,W,c,c1,c2,X,s,c3,M;
```

```
    b:=FFTN/2;
```

```
    a:=LinearAlgebra[RowDimension](Sample);
```

```
    y:=trunc((a-N)/H)+1;
```

```
    A:=Array(1..b,1..y);
```

```
    w:= 0.54 -.46*cos((2*Pi*n)/(N-1));
```

```
    W:=unapply(w,n);
```

```
    c:=a-N;
```

```

c1:=0;

c2:=1;

while c1 < c do

    x:=Array([seq(evalf(Sample[j+c1,1]*W(j)),j=1..N)]);

    X:=Matlab[fft](x,FFTN);

    s:=[seq(20*log10(abs(X[k])),k=FFTN/2+1..FFTN)];

    for c3 from 1 to b do

        A[c3,c2]:=s[c3];

    end do;

    c1:=c1+H;

    c2:=c2+1;

end do;

M:=Matrix(A);

ExportMatrix(cat("spec",Name,".txt"),M,target=delimited,delimiter="\ t");

M;

end proc:

```

## Smax

Smax is a program designed to compliment spec by finding the most intense frequency for every time point of its output and save as a text file. Smax has two

arguments. The first argument, A is intended to be the output of spec. The final argument, name, is expected to be a combination of text and numbers (stay away from symbols) and will be the root of the name of the file this program saves. The full file name will be Smax“name”.txt and will be saved in the same folder as the worksheet being used. Maple output is a one column Matrix of frequencies.

```
SMax:=proc(M,name)
local a,x,y,A,c1,m,c2,B;

  x:=LinearAlgebra[RowDimension](M);
  y:=LinearAlgebra[ColumnDimension](M);
  A:=Array(1..y,1..2);
  for c1 from 1 to y do
    m:=M[1,c1];
    a:=.5;
    for c2 from 1 to x do
      if m < M[c2,c1] then
        m:=M[c2,c1];
        a:=.5-((c2-2)/(2*x));
      end if;
    end do;
    A[c1,1]:=m;
    A[c1,2]:=a;
  end for;
end proc;
```

```

end do;

B:=Matrix(A);

ExportMatrix(cat("Smax",Name,".txt"),B,target=delimited,delimiter="\t"):

B;

end proc:

```

per

The program per was designed to compliment Smax and through it spec by counting how many of the frequencies reported by Smax in a particular time frame are below a threshold. The first argument of per, M, is intended to be the output of Smax. L, the second argument is the length of the time frame in which the counting is conducted. The final argument, name, is expected to be a combination of text and numbers (stay away from symbols) and will be the root of the name of the file this program saves. The full file name will be per" name".txt and will be saved in the same folder as the worksheet being used. Maple output is a one column Matrix of integers between zero and L.

```

per:=proc(M,L,T,name)

local x,B,c1,A,a,c2,Q;

x:=LinearAlgebra[RowDimension](M)-L;

B:=Array(1..x);

```



```

for c1 from 1 to x do
    A:=[seq(M[n,2],n=c1..c1+L-1)];
    a:=0;
    for c2 from 1 to L do
        if A[c2] < T then
            a:=a+1;
        end if;
    end do;
    B[c1]:=a;
end do;
Q:=Matrix(B);
ExportMatrix(cat("per",name,".txt"),Q,target=delimited,transpose=true,delimiter="\t");
Q;
end proc:

```

VITA

JUSTIN PEYTON

- Education: B.S. Chemistry, Milligan College,  
Milligan College, Tennessee 2005  
M.S. Mathematics, East Tennessee State University,  
Johnson City, Tennessee 2008
- Presentations: Justin Peyton, Karl Joplin, Edith Seier. (2007). Analysis of the pupal respiratory cycles in the flesh fly, *Sarcophaga crassipalpis*. Entomological Society of America. San Diego, CA.
- Karl Joplin, Justin Peyton, Edith Seier. (2007). Unique developmental respiratory patterns in the flesh fly, *Sarcophaga crassipalpis*: pupal respiratory cycles. Entomological Society of America. San Diego, CA.
- Karl Joplin, Edith Seier, Justin Peyton. (2007). Unique developmental respiratory patterns in the flesh fly, *Sarcophaga crassipalpis*: pupal respiratory cycles. International Symposium on the Environmental Physiology of Ectotherms and Plants. Dunedin New Zealand.