AUTOMATIC ANALYSIS OF *IN SITU* CYCLIC VOLTAMMETRY DATA

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

INFORMATION AND COMPUTER SCIENCES

DECEMBER 2009

By
Brian P. Jaress
jaress@hawaii.edu

Thesis Committee:

Kim Binsted, Chairperson
Henri Casanova
Phillip Johnson
Brian Glazer

We certify that we have read this thesis and that, in our opinion, it is satisfactory in scope and quality as a thesis for the degree of Master of Science in Information and Computer Sciences.

THESIS COMMITTEE

_____
Chairperson

_____

_____

_____

iii

Dedicated in memory of R. Patrick Jaress, who taught by creating experience.

Conquer the angry man by love.
Conquer the ill-natured man by goodness.
Conquer the miser with generosity.
Conquer the liar with truth.

— *The Dhammapada*

Acknowledgments

Abstract

Cyclic voltammetry is a useful way of measuring the concentration of certain chemicals in the ocean, but the required analysis of raw data is currently labor intensive. Raw electrical data needs to be converted into chemical concentrations. That analysis currently involves guidance from a human expert. Researchers at the University of Hawai'i are trying to automate the process by using software that only requires human auditing after the fact.

As part of that effort, this thesis presents the results of developing and testing several fully automatic methods for analysis of *in situ* cyclic voltammetry data.

Several of the methods tested were improvements over the existing prototype aevum. Aside from aevum, they are the first fully automatic methods for multi-chemical analysis of *in situ* cyclic voltammetry data, and their accuracies have been measured numerically.

Further testing with better data should be done to help choose amongst the methods and improve on them. The difference, for testing purposes, between calibration and field scans may be larger than expected.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Importance of Research

Voltammetry is a valuable technique for measuring the concentration of certain chemicals in water. [20, 14] Cyclic voltammetry is useful for *in situ* measurement (measurement taken in the environment; See Section 1.2.2.) because it can measure a variety of chemicals at once. [11] In particular, scientists at the University of Hawai'i are using cyclic voltammetry to detect the activity of living things which use chemical reactions as a source of energy.

Collection of data is automatic and can be performed by a device left unattended in the ocean, but analysis is currently at most semi-automatic. In existing solutions, the user must guide the software on where to look for a peak, how many peaks to look for, or even where to locate the top and bottom of the peak. This creates a bottleneck.

1.1.1 The Glazer-Binsted Collaboration

Brian Glazer is an assistant professor of oceanography at the University of Hawai'i. He researches the relationship between microbes and oxidation-reduction chemical reactions, especially the mutual influence between the concentration of redox chemicals and the presence of microbes which use the chemical reactions as an energy source. He leads a team of scientists at UH and collaborates with others around the world. Current projects using voltammetry include coastal biogeochemistry, observatory instrumentation, and seafloor hydrothermal geomicrobiology.

Kim Binsted is an associate professor of computer science, also at the University of Hawai'i, and a co-investigator of the NASA Astrobiology Institute at UH. She researches computational astrobiology and intelligent user interfaces.

Glazer and Binsted are collaborating to improve the analysis of raw voltammetric data with software. To analyze the large volume of data produced in the course of Glazer's environmental research, the Glazer-Binsted collaboration is attempting to move away from the human-guided approach of current software toward methods in which a human simply audits the software's analysis. They are also interested in improving overall usability and efficiency.

The research proposed here is a part of that collaboration.

1.2 Problem

1.2.1 Voltammetry

Voltammetry is a family of techniques for detecting certain types of chemicals and their concentrations in water by measuring an electrical current while varying the voltage. [25]

Cyclic Voltammetry

In cyclic voltammetry, one adjusts the voltage from a starting voltage to another voltage then back to the starting voltage at a steady rate, recording the current at regular intervals. Each cycle is a separate graph, called a scan, that can be analyzed to produce a measurement. Certain chemicals create peaks; the horizontal peak location identifies the chemical and the peak height is related to the concentration. [14]

Analysis by a Human

Traditionally, a human will visually detect tops and bottoms of peaks on a graph, manually measure the height, and assume that the relationship between peak height and concentration is linear.

Calibration

The relationship of peak height to concentration depends on, among other things, physical characteristics of the electrode such as manufacturing defects, corrosion, and buildup. To establish the relationship, peak heights are measured for known concentrations of a chemical. [2]

This calibration is done separately for each chemical, so there are two types of scans. Calibration scans are of water mixed in a laboratory to contain a known amount of a single chemical. Field scans are of water in the environment containing unknown concentrations to be measured. Figures 1.1 and 1.2 show two typical calibration scans. Figure 1.3 shows a typical field scan.

Figure 1.1. Typical Calibration Scan (Sulfide 61.6 Micromolars)
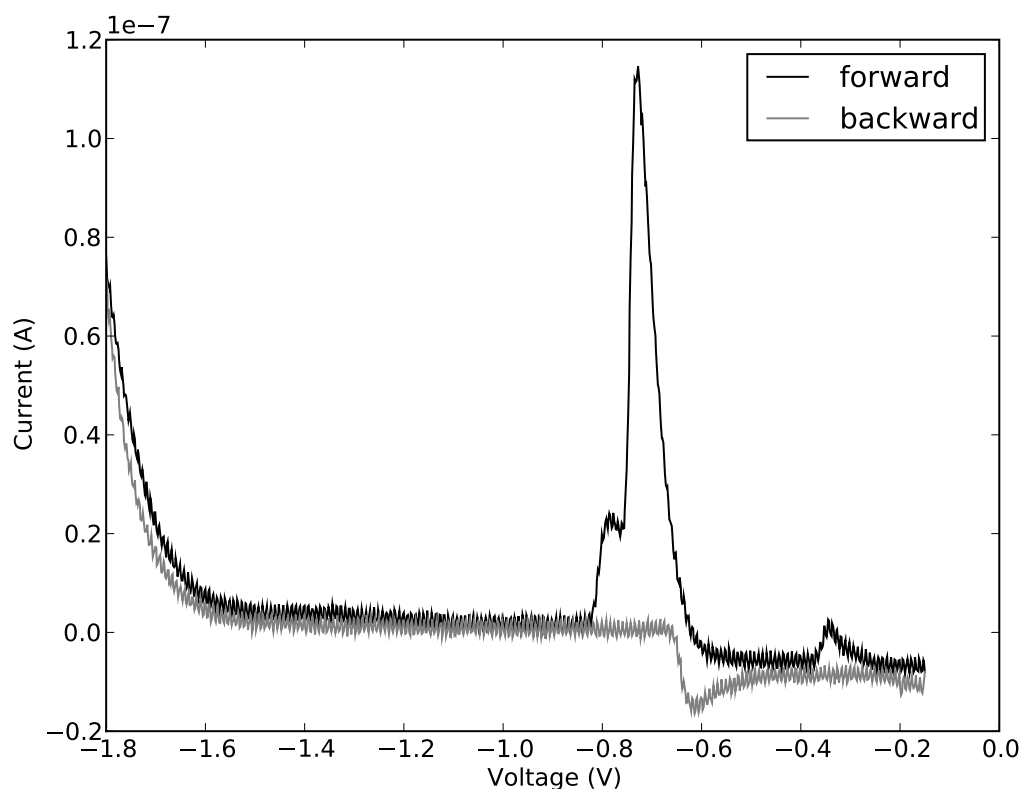
Some chemicals create a large, almost symmetrical peak on the forward part of the scan when scanned alone. These chemicals typically have a small, upside-down peak on the backward part of the scan (returning the voltage to zero). In this case, the reaction which produces the peak is called "reversible."

A scan can only be analyzed to produce measurements of chemical concentration in the context of a set of calibration scans.

1.2.2 *In Situ*

One important use for cyclic voltammetry is *in situ* measurement. [18] That is, measurements taken by placing electrodes into the environment rather than bringing a sample of water back to the lab. This use is the focus of the Glazer-Binsted collaboration. (See Section 1.1.1.)

Because the measurements are *in situ*, there can be no preparation of the water that makes measurement easier, nor is it possible to scan the same volume of water again. The chemical composition of the water as a whole potentially affects the entire scan [25], including chemicals that we are not measuring and factors such as pH and temperature. [11]

3

Figure 1.2. Typical Calibration Scan (Oxygen Saturated)
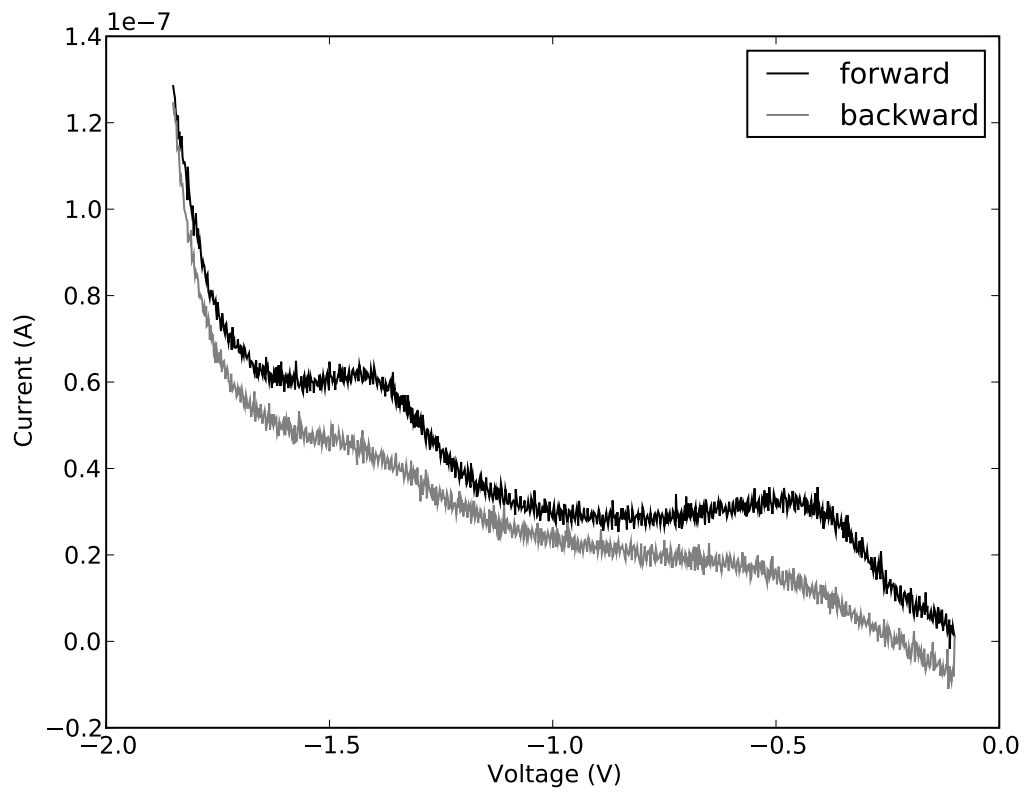
Some chemicals create peaks that are closer to plateaus on the forward part. These chemicals typically do not have peaks on the backward part. In this case, the reaction which produces the peak is called "irreversible."

Oxygen creates two peaks because the reaction which creates the first peak also produces peroxide.

Figure 1.3. Field Scan (Unknown)
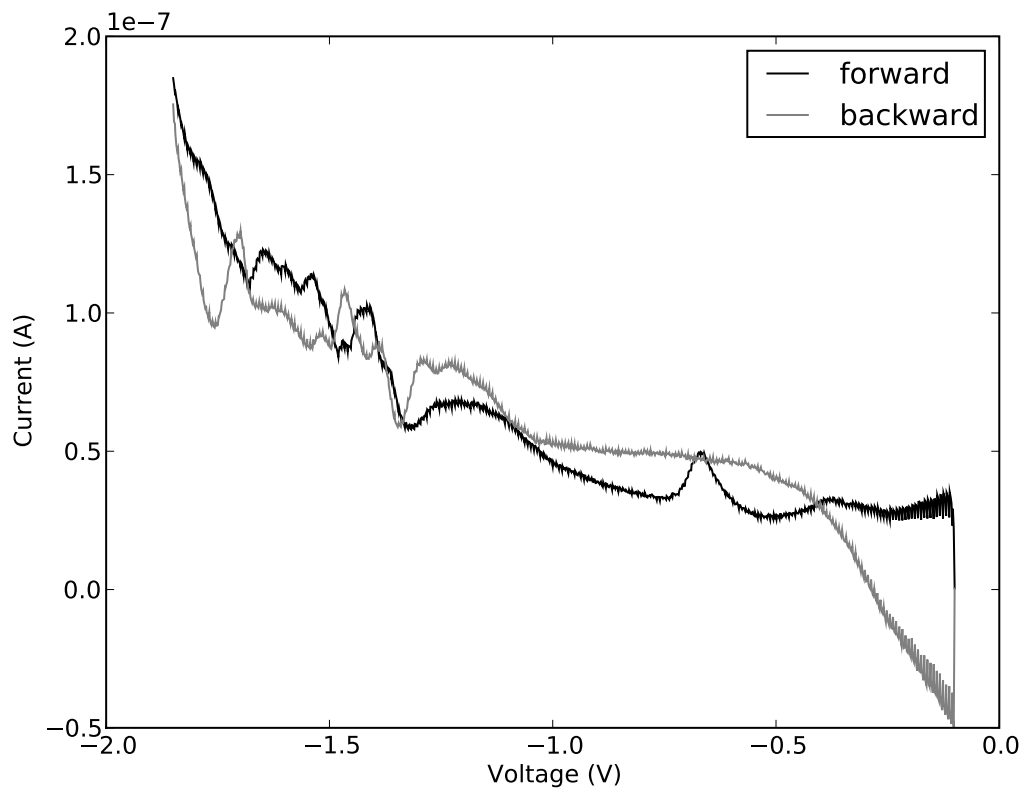
Field scans are the product of many interacting factors, some of which are to be measured.

1.2.3 Automatic Analysis

We want a method of analysis that is fully automatic. That means the user gives voltage and current data from the field to a computer program, together with calibration data, and gets back chemical concentrations as output.

This type of automation would make it much easier to obtain large data sets by reducing the labor involved (See Section 2.3.2.) and could eventually lead to machines that can respond[1] directly to important changes in dissolved chemical concentrations.

1.2.4 Difficulties

Noisy, Time-Sampled Data

The data in each scan consists of discrete points and includes noise. That makes it more difficult to find the derivative or special points like maxima and minima.

No Parametric Mathematical Formula

A single chemical by itself will follow the formula [25]:

$$I_{tot} = I_{far} + I_{chg} + I_{ads} + I_{cross}.$$

Where I_{far} is the portion of the current being analyzed by measuring peaks, while I_{chg} and I_{ads} are affected by the electrode. I_{cross} represents an interaction between the other three terms and prevents recovery of I_{far} through subtraction of a blank scan.

In some situations, there is a formula for I_{far}, but there is in general no formula for I_{cross}. (See Section 2.1.2.)

Chemicals Interfere with Each Other

I_{cross} depends on the chemical composition of the water as a whole [25], including chemicals that we are not measuring and factors such as pH and temperature. [11] Adding a new chemical to the water can raise the baseline of other chemicals' peaks by an amount that is hard to predict from the value at that position on a graph of the new chemical in isolation.

---

[1]For example, by taking a water sample.

6

1.3   Research Approach

I have developed and directly compared multiple methods using sample data from a domain expert. There were two rounds of testing, one analyzing calibration scans and the other analyzing field scans.

1.3.1   Considerations

The nature of the problem creates several considerations for direct comparison of methods.

Two types of scans Calibration scans are of water with a single chemical, mixed in the lab. They are used to calibrate, though they may also be measured. Field scans are of water from the environment, with many chemicals. Their actual chemical contents are only known through measurement, and they are not used for calibration.

No ideal test cases Because of the division into two types of scans, there are no scans with both multiple chemicals and a composition known independently of any measurement.[2]

Scans grouped by electrode Physical characteristics of the electrode are being calibrated for, so calibrations are only valid for a particular set of field scans made using the same electrode. Because electrodes wear out, only limited field scans are available for any calibration set.

Location dependent data The actual contents of field scans will depend on where the scans were taken. Calibration scans contain only the chemicals anticipated in the corresponding (same electrode) field scans.

1.3.2   Two Rounds of Testing

In light of the considerations above, there were two rounds of testing.

First Round

The first round used calibration scans only. In other words the "field scans" used in the test were actually calibration scans. Each calibration scan was measured based on a calibration derived from the other scans in the same calibration set (same electrode).

---

[2]It would be possible to create mixtures especially to meet these criteria, but that was not in the data provided by our domain expert. Calibrations with multiple chemicals have been done, for example by Luther et al. [14], but do not reflect the typical usage of the domain expert.

Second Round

The second round used actual field scans as well as calibration scans. Each field scan was measured based on a calibration derived from the corresponding calibration set (all calibration scans from the same electrode). Calibration scans were not measured.

1.3.3 Comparison of Methods

For each method of measurement, once each scan was measured it was assigned an error. The errors of the individual scans were combined into three summary numbers indicating how well the particular method performed. Those three numbers were evaluated by a domain expert after each round.

Error of Individual Scans

Scans had actual concentrations assigned by a domain expert. In the case of calibration scans, these came from the expert putting the chemicals into the water. In the case of field scans, they came from analysis using a combination of software and expert judgment.

The error for an individual scan the magnitude of the difference between the system-determined and actual or expert-determined concentrations, treating them as vectors with each chemical as a dimension.

Observed Error

The observed error of a measurement method is the root mean squared error ("RMSE") of all measured scans. In other words, it is the square root of the mean of the squares of the errors. RMSE is a common indicator of goodness-of-fit.

Estimated Error

The estimated error is a bootstrap estimate of RMSE: the mean of the RMSEs of 10,000 bootstrap data sets generated from the underlying set of all measured scans. Each bootstrap set was the same size as the underlying set and generated by randomly selecting individual scans, with replacement.

Bootstrap estimation is a simple method for inferential statistics that often works well and does not require a formula relating the sample statistic to the estimate. [7]

Threshold Error

  The threshold error is the top of a bootstrap estimate of a 95% confidence interval of RMSE. This was derived from the same 10,000 bootstrap RMSEs as the estimated error, but by taking the 250th largest rather than mean.

Chapter 2

Literature Review

2.1   History and Theory of Cyclic Voltammetry

2.1.1   History

Voltammetry is a family of techniques in which electricity is run through a chemical solution and voltage is varied while current is measured.

The oldest form of voltammetry is called polarography and was invented in the early 1920s by Jaroslav Heyrovský. Some authors distinguish voltammetry from polarography, referring to "polarography and voltammetry," but the main difference is merely that polarography uses an electrode of liquid mercury[1] while voltammetry uses a variety of solid metals or carbon. As recently as the late 1980s, liquid mercury was still the electrode of choice. [17, 25]

Kalvoda [17] mentions cyclic voltammetry in passing, as an example of how techniques for solid electrode voltammetry can be simpler and more informative but also more difficult than liquid mercury polarography to analyze.[2] For that reason, he focuses on polarography.

Vassos and Ewing [25] devote a chapter to cyclic and linear sweep voltammetry. In linear sweep voltammetry, the voltage is increased linearly, so that a graph of voltage vs time looks like a line. In cyclic voltammetry, the voltage is linearly increased then linearly decreased, so that the graph of voltage vs time looks like a triangle. They are essentially the same technique.

By the late 1980s and early 1990s, Wipf and Wightman [28, 27] were detailing the benefits of microdisk electrodes, in which the electrode surface is in the shape of a very small circle. This type of electrode is made by cutting and polishing a very thin, insulated wire. Using a microdisk

---

[1]The mercury flows from a capillary, hanging at the opening in a droplet that is periodically replaced.

[2]Polarography techniques generally vary the voltage stepwise, in sync with the replacement of mercury drops. Some solid-electrode voltammetry techniques are also stepwise, but cyclic voltammetry is not.

electrode for cyclic voltammetry reduces the distortion that occurs at high scan rates by reducing impedance and voltage drop. Higher scan rates help the reaction of interest "outrun" other factors affecting the current and can, if distortion is prevented, provide a more accurate measurement. Wipf and Wightman report scans as fast as 10^6 V/s while tracking 9-fluorenone.

2.1.2 Theory

Vassos and Ewing give the overall formula for a single chemical:

$$I_{tot} = I_{far} + I_{chg} + I_{ads} + I_{cross}.$$

Where I_{far} is the portion of the current being analyzed by measuring peaks, while I_{chg} and I_{ads} are affected by the electrode. I_{cross} represents an interaction between the other three terms and prevents recovery of I_{far} through subtraction of a blank scan.

I_{cross} depends on the chemical composition of the water as a whole [25], including chemicals not being measured and factors such as pH and temperature. [11] Adding a new chemical to the water can raise the baseline of other chemicals' peaks by an amount that is hard to predict from the value at that position on a graph of the new chemical in isolation.

First Principles

Most theoretical work uses Fick's Laws [21] and the Nernst Equation [21] as "first principles."

Fick's Law covers diffusion, the physical movement of the dissolved chemical through the water. The law shows how movement is from areas of high concentration to areas of low concentration, with the difference in concentration affecting speed:

$$J = -D\nabla C,$$

$$\frac{\partial C}{\partial t} = D\nabla^2 C.$$

The second equation above can be derived from the first. They are sometimes called Fick's First Law and Fick's Second Law and sometimes they are together called Fick's Law. The first equation directly relates the chemical flow J to the concentration C, while the second expresses flow as the partial derivative of concentration with respect to time. In both cases, D is a constant determined by, among other things, the particular chemical.

11

The Nernst Equation deals with voltages. The voltage at which we expect a peak E_{red} is related to a standard voltage E_{red}^{Θ} by the number of electrons used, the concentrations of the two forms of the chemical (before and after the reaction that produces the peak) the temperature, and several physical constants:

$$E_{red} = E_{red}^{\Theta} - \frac{RT}{nF} \ln \frac{C_R}{C_O}.$$

Theoretical Analysis

Attempts to give a detailed analytical formula have focused on simplified special cases with one or two reactions and ideal electrodes. This has produced different formulas even for cases that, in physical reality, are similar and lie along a continuum.

Nicholson [21] used variations on the Nernst Equation and Fick's Law (See Section 2.1.2.) to derive equations for several specific cases involving ideal planar and spherical electrodes. Each case dealt with a single chemical to be detected, and the cases were separated by specific properties of the chemical. Many of Nicholson's assumptions appeal to the experimenter's ability to adjust parameters, such as the scan rate, until a usable scan is achieved. The equations are then solved using a combination of techniques, including numerical integration.

More recently, Amatore [1] examined the generally assumed equivalence between hemisphere and disk microelectrodes, as well as between band and hemicylinder microelectrodes. Specific special cases were considered, leading to different equations. The final result was that there is no equivalence, and there is approximate equivalence in fewer cases for the hemisphere and disk than for the hemicylinder and band.

Simulation

Gavaghan [8, 9, 10] developed a simulation technique specifically to overcome problems with equations for the physical movement of dissolved chemicals near electrodes. Looking specifically at the case of a disk-shaped electrode, there is a singularity (a discontinuity in the concentration gradient) at the edge of the disk. Under existing simulation techniques, this singularity became an overpowering source of error.

Gavaghan's solution was to exponentially subdivide the space near the singularity. The results were consistent with the only case for which an analytical result was known: a microdisk electrode with the chemical concentrations in a steady state.

Even for simple, narrow questions, researchers have needed to resort to simulation. Davies [4] used two-dimensional simulation to examine the effect of the distance between electrodes in arrays of microdisk electrodes. Electrode arrays are not common in *in situ* voltammetry, but the need for Davies's work, and the other theoretical work above, shows how constrained our theoretical knowledge of cyclic voltammetry is. What is known does not generalize or extend very well, even the equations for simple situations can be resistant to analytic solution, and much of what works in practice has little or no theoretical basis.

2.2 *In Situ* Cyclic Voltammetry

2.2.1 Technical Issues

Brendel and Luther [2] developed a gold and mercury amalgam microelectrode which made *in situ* cyclic voltammetry practical. The electrode is made of a gold wire (usually 100 μm in width) held in a glass tube with epoxy and electroplated with mercury. The small size and composition of the microelectrode make it accurate and self-cleaning[3].

Outside factors can affect the measured current, and should ideally be either controlled in the calibration or corrected for in the calculations. Brendel and Luther [2] identified several: chloride metal, salinity, and pH. Later, Luther [14], et al. added temperature and flow rate. There appears to be no universal rule on which factors are worth compensating for.

2.2.2 Cyclic Voltammetry in Action

An atypical but revealing use of Brendel and Luther's electrode is presented in a 2007 paper by Ma, et al. [18] which describes the use of cyclic voltammetry to study seasonal fish kills (many fish dying at once) and algae blooms (sudden excessive growth of algae) in Delaware.

Dredging for raw material to build homes had created artificial, dead-end canals on the coast of Delaware. These canals were shallow, but had large "holes" (isolated deeper regions). Since their creation, these canals were the site of seasonal fish kills and algae blooms.

In 2001, Ma et al. used cyclic voltammetry and gold-mercury microelectrodes to confirm the cause of the problem: too little oxygen was getting into the holes. Without oxygen, the decay of organic material led to a buildup of hydrogen sulfide, phosphorous, and nitrogen. Seasonal storms

---

[3]Normal use creates buildup, but the buildup can be removed with a steady -0.8V current.

brought these accumulated chemicals to the surface waters. Surface fish died from the flood of hydrogen sulfide, and algae bloomed on the nitrogen and phosphorous.

In 2007, Ma returned to study the effectiveness of recently installed solar-powered water circulators. The circulators were intended to end the fish kills and algae blooms by circulating water between the surface and the bottoms of the holes. Cyclic voltammetry showed that the circulators had very little effect on oxygen levels, correctly indicating that the fish kills and algae blooms would continue.

The study by Ma is atypical because of the focus on fish. Most *in situ* cyclic voltammetry studies focus on microorganisms. For example, the buildup of hydrogen sulfide from decay in the absence of oxygen is driven by the interaction between microorganisms and chemicals.

2.2.3 The Role of Cyclic Voltammetry in Environmental Studies

A 2008 overview of cyclic voltammetry in environmental studies by Luther et al. [14] details the link between cyclic voltammetry and microorganisms. Ocean microorganisms, especially those living in or just above sediment, are closely tied to changes in chemical concentrations over small distances. Cyclic voltammetry can give fine-grained measurements in both water and sediment.[4] Another benefit is the large number of chemicals that can be detected–seventeen, including nearly all those which microorganisms can use as simple energy sources.

A good example of these benefits is a study by Luther et al. from 1998 [13] comparing cyclic voltammetry measurements against an alternative method for measuring chemicals in sediment. The alternative method was to take a core sample, extract the water from it, and perform various chemical-specific tests on portions of of the water. The cyclic voltammetry was performed by inserting electrodes directly into the core samples before extracting the water. Results between the two methods were consistent, but the extraction method only gave an average for each core sample. The voltammetry showed gradients within the sample down to a millimeter. That gave a clearer picture of which chemicals overlapped and could be used by microorganisms in the same reaction.

There are several recent overviews of ocean sensors, each with its own perspective of cyclic voltammetry. Johnson [15] has very little to say about voltammetry, but Moore [19] has an even better description of how the electrodes are used than most voltammetry-specific papers.[5]

---

[4]The Luther overview emphasizes the use of cyclic voltammetry in sediment and even discusses directly inserting electrodes into microbial cultures.

[5]There are actually three electrodes: working, counter, and reference. Only the working electrode needs to be a gold-mercury microelectrode, and the reference electrode is often far from the other two. The voltage is a relative voltage between the working and reference electrodes, and the current flows between the working and counter electrodes.

Viollier's overview [26] mentions discoveries of general interest made using using cyclic voltammetry, including a connection between the nitrogen cycle and metals and a connection between sulfur and iron cycles. Cyclic voltammetry is well-suited to discovering these types of connections because it scans a spectrum of voltage for a whole list of chemicals.

2.3 Computer-Aided Analysis for Cyclic Voltammetry

2.3.1 Prior Work Used in Publications

The machines that supply the voltage and record the current, called potentiostats, come with software that requires the user to specify the top and bottom of each peak. The software merely displays the graph and performs subtraction, making it the most accurate but also the most time-consuming option.

Bristow and Taillefert [3] provide a semi-automated analysis, in which a human expert and a computer cooperate. The computer suggests a peak's top and bottom, which the human can modify. The human also has an opportunity to specify that the computer deconvolute a region of the graph, attempting to separate overlapping peaks. Bristow and Taillefert do not provide a discussion of accuracy, except to list experimental techniques that seem to affect accuracy.

Brian Glazer created an oxygen-specific program to assist with comparing voltammetry to optical fluorescence quenching for detection of oxygen. [11] The program has acceptable accuracy but only works for a single chemical and requires the user to roughly specify the location and width of the peak.

Moore [20], et al. used another custom oxygen-specific program based on preset voltage windows. The program initially worked by subtracting the current near zero and the current after the peak (oxygen having a plateau shaped peak). Because of electrical noise, the program was modified to find the current halfway up the peak and use that as a baseline instead of the current near zero voltage.

2.3.2 Automation

The programs discussed above all require some degree of human guidance on peak detection, but there is another barrier to automation. The software discussed above does not output chemical concentrations or track the linear relationship that is used as the final step in producing them. Therefore, in addition to guiding the program to find peaks, the user must:

1. Run the program to get a peak height for each calibration scan;

2. Calibrate by fitting a line to those points with a separate program;

3. Run the program to get a peak height on each field scan; and then

4. Apply the linear relationship to each of those points with a separate program.

At first, this may seem like a conceptually minor problem, though very inconvenient. In fact, the problem is large. The linear relationship is not calculated and applied by rote as the summary above may suggest. The process enumerated above gives the user many opportunities to alter the result, for example:

- Factors that affect the relationship, such as temperature, can be corrected for or not, depending on whether they were controlled for during the calibration scans and on the user's judgment.

- The relationship for one chemical can be inferred, either from relationships of other chemicals or from anything else of the user's choosing, rather than derived from calibration scans.

- Because the existing programs require guidance on peak detection, the user can treat a subset of scans as a special case, giving different guidance and maintaining a separate linear relationship between current and concentration.

- Concentrations can be declared by fiat. The user can look at the output of the program, which gives a current, and declare a particular value for the concentration for any reason at all.[6]

Expert users take advantage of these opportunities to make sure that the final answers are correct. Therefore, it is not known how dependent the process is on subjective human expertise. Discussions with domain experts indicate that none of the programs described above are accurate enough without careful human correction.

The time spent by the human varies greatly, but the domain expert who provided the test data in Section 3.1 recalls spending roughly eighty hours on a batch of 45,000 scans taken from the same location.

---

[6]Users try to avoid doing this, but it was done, for example, with some of the data provided by our domain expert.

2.3.3 Aevum

The only existing fully automatic program is a prototype produced several years ago by the Glazer-Binsted collaboration called aevum. The prototype is fully automatic in that it requires only:

- calibration scans,

- the known concentrations of the calibration scans,

- scan metadata, and

- field scans to which it assigns measured concentrations.

The problem with aevum is that it gives wrong answers.[7] There is, therefore, gap between all programs used in published work, which are accurate but subjective and labor-intensive, and aevum, which has the opposite problem.

---

[7]When used to analyze oxygen calibration scans, aevum gives concentration values of roughly 1.4 μM for both scans of with no oxygen and oxygen-saturated scans.

Chapter 3

Methods Tested and Experimental Setup

3.1   Test Data

A domain expert provided scans for five electrodes, shown in Table 3.1.

3.1.1   Data Set Size and Organization

The group of scans for each electrode is subdivided into calibration scans and field scans. Calibration scans are of water mixed in a laboratory to contain a known amount of a single chemical each. Field scans are of water in the environment containing unknown concentrations to be measured.

Table 3.1. Scan count by electrode

Electrode	Calibration	Field
W121-Feb	70	7650
W121-March	60	4690
W140-Feb	60	7650
W141-April	15	8700
W142-April	15	7543

Any given scan can only be interpreted–processed to compute chemical concentrations–in terms of a group of calibration scans from the same electrode as the given scan. In cyclic voltammetry, a group of calibration scans is used to establish a relationship which is then applied to a field scan to determine the concentrations of chemicals in the field scan.

One challenge to cyclic voltammetry is that calibration and field scans are different in kind. Having only a single chemical, calibration scans do not reflect interactions between chemi-

cals or overlapping peaks. This makes many popular machine learning techniques, which assume training data comparable to test data, inappropriate.

3.1.2 Origins

These are actual scans and actual concentration numbers, generated in the course of research by the domain expert. The field scans come from the Kilo Nalu Observatory off the south shore of Oahu and were taken *in situ*. The calibration scans come from a laboratory at the University of Hawai'i.

In both cases, the scans were taken using an In Situ Electrochemical Analyzer made by Analytical Instrument Systems, Inc.

As is typical, the expert took calibration scans for the specific chemicals to be measured, oxygen and sulfide in this case. The expert reported making no corrections for factors such as temperature or pH and used 215 μM as an estimate for the concentration of calibration samples containing oxygen.[1]

3.1.3 Preprocessing

One challenge here is that the "true" concentration values provided by the domain expert for the field scans are not certain. This is not a problem with the calibration scans. Each calibration scan is associated with a certain concentration of a certain chemical because the expert created the sample by adding a specific amount of chemical to pure water. Field scans, however, have "true" concentrations deduced from the scans by a combination of software and human judgment.

The concentrations provided by the domain expert for field scans are partially aggregated. The values given are means-without-outliers for sets of five scans, meaning:

1. The sequence of scans for each electrode was broken into groups of five.

2. For each group of five, the mean and the standard deviation were computed.

3. Each group mean was recomputed without scans more than one standard deviation from the original mean.

Some of these means were further modified by coercing negative and error values to zero.

---

[1]It is difficult to prepare water samples with specific intermediate amounts of oxygen in the lab. The oxygen calibration scans in the data set all have either no oxygen or as much oxygen as the water would absorb from the air.
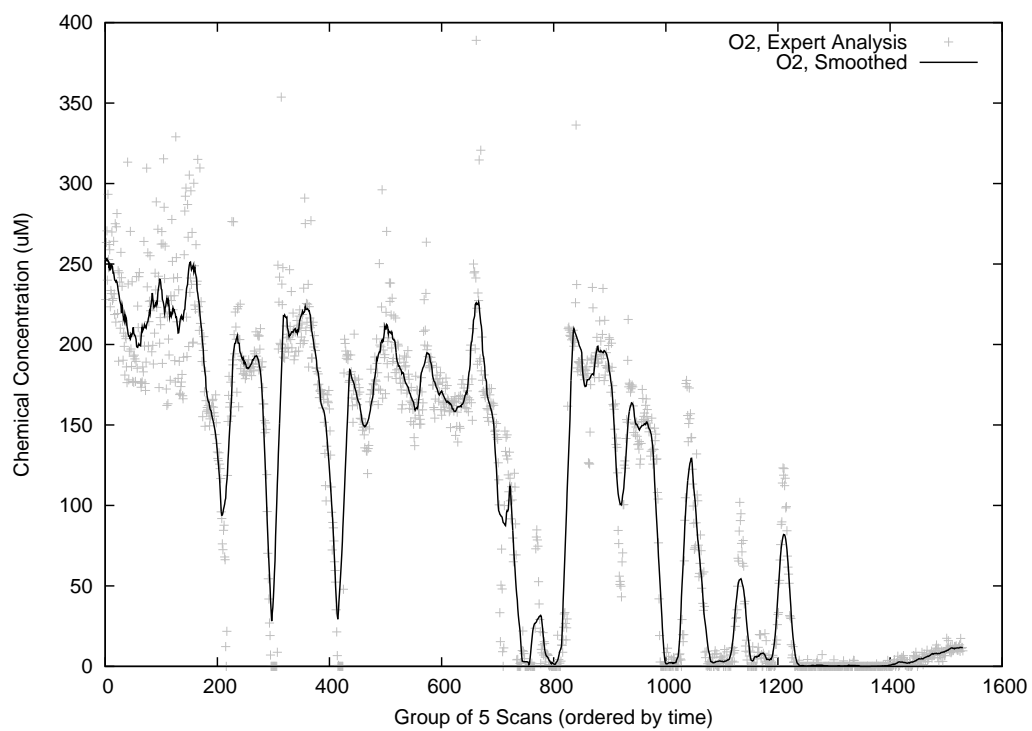
Figure 3.1. 5-Scan Means, Electrode W121, Feb.

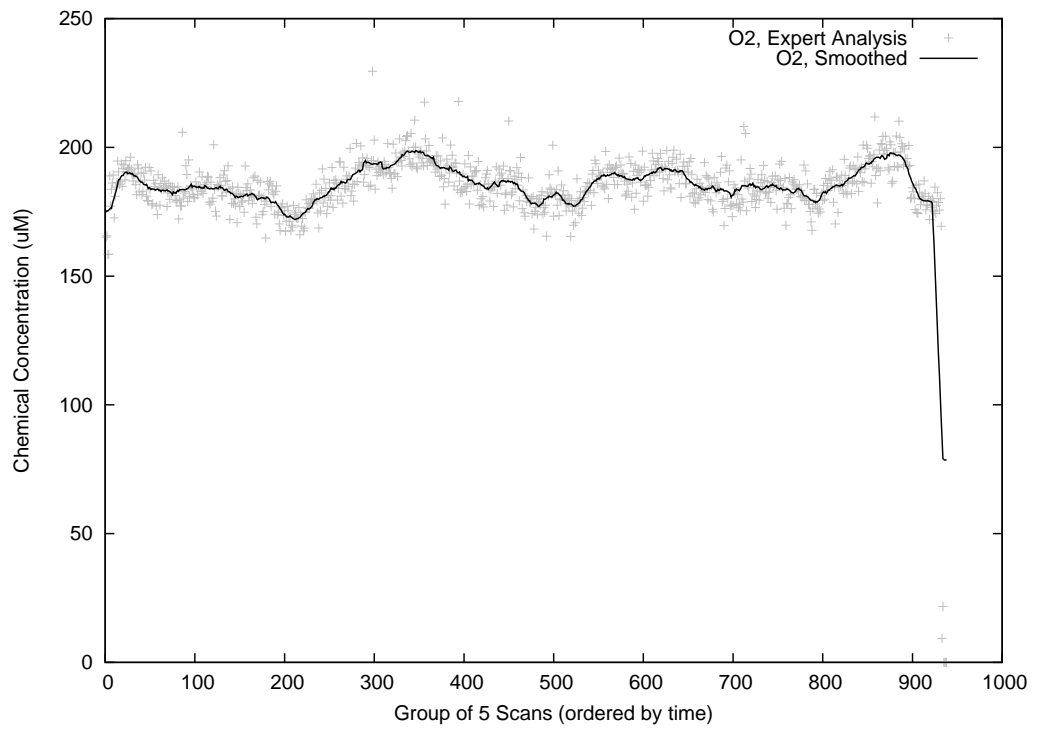Figure 3.2. 5-Scan Means, Electrode W121, March
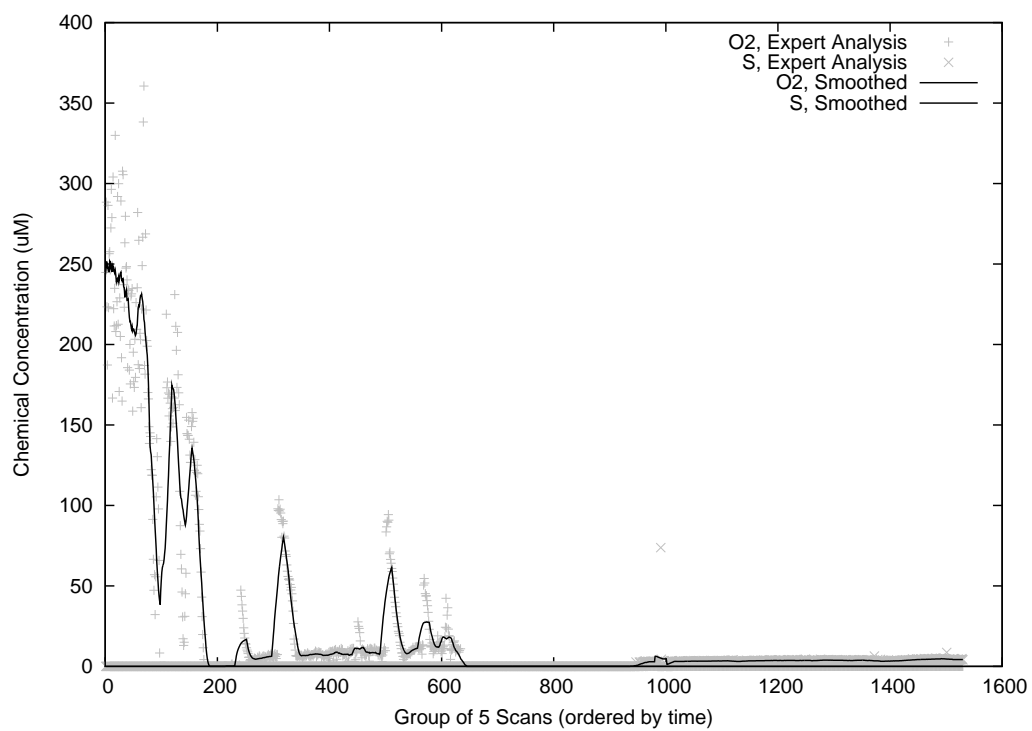
Figure 3.3. 5-Scan Means, Electrode W140, Feb.

Figure 3.4. 5-Scan Means, Electrode W141, April

23
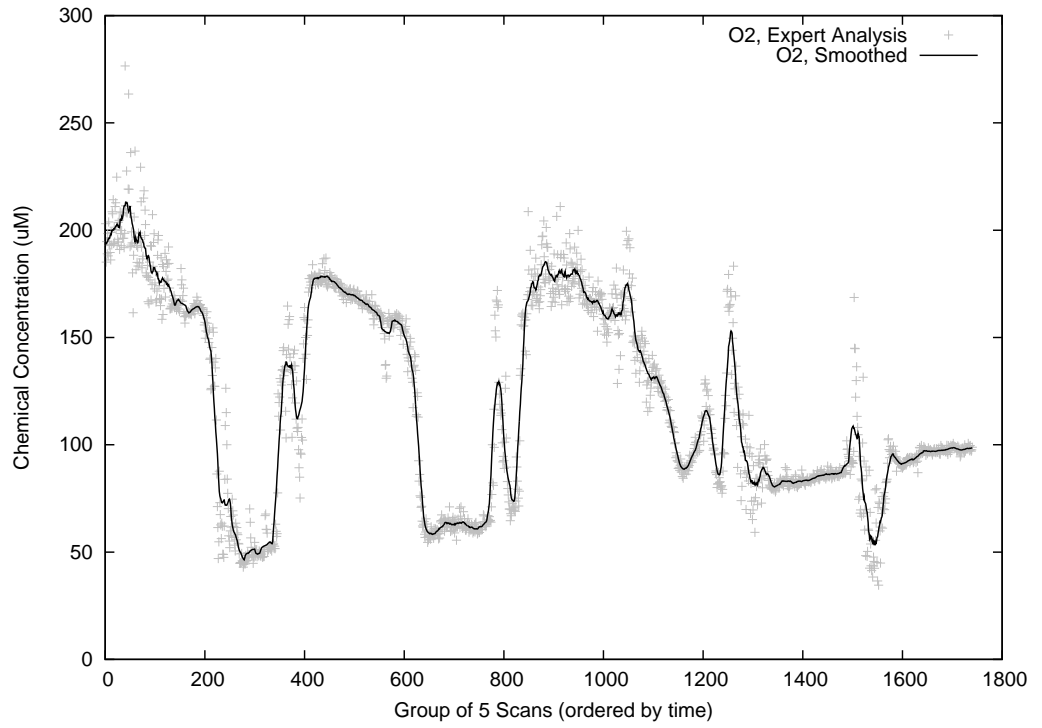
Figure 3.5. 5-Scan Means, Electrode W142, April
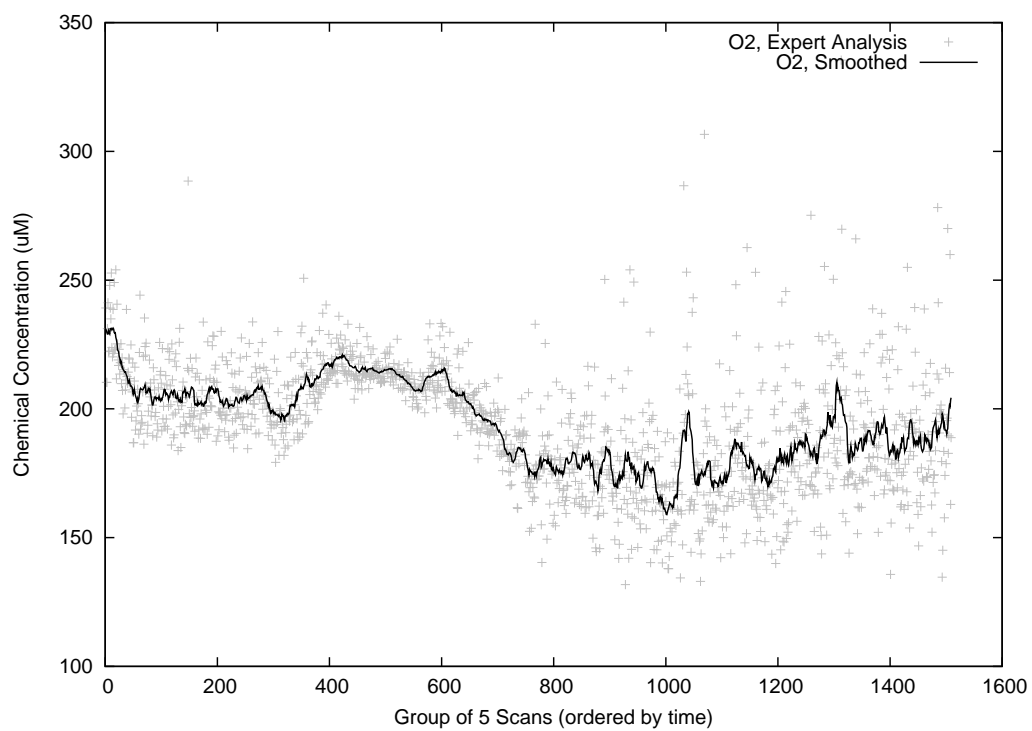
The individual concentrations, before taking any means, were the result of a linear relationship applied to the output of the ScanVolt program used in [11]. The relationship was computed and applied by the expert, using separate general-purpose software.[2]

Even after this aggregation, the resulting values appear to contain significant noise. A graph of the means is suitable for detecting trends, but many individual points are not credible "true" values for the five scans in the group associated with the point.[3]

Therefore, the five-scan means provided by the expert were smoothed with 21-point averaging, giving the values considered "true" for the purpose of testing the accuracy of the automatic methods in the second round of testing. Figures 3.1, 3.2, 3.3, 3.4, and 3.5 show the five-scan means plotted together with the smoothed values for each of the five electrodes.

3.2   Testing the Methods

The methods presented here are fully automatic (See Section 1.2.3.) and could be tested automatically.[4]

The methods and the code to test them were written in the Python language using the SciPy library. [16] Running times indicated with test results are from a 2.1 GHz desktop computer.

Any validation procedure is inherently arbitrary only intended to demonstrate satisfactory accuracy within an applicable domain. [23]

3.2.1   Two Rounds of Testing

Because there are two types of scans, there were two rounds of testing. In the first, calibration scans were analyzed. In the second, field scans were analyzed. Analyzing calibration scans has the advantage that true concentrations are known with certainty and the disadvantage that samples are artificial, lacking the chemical interactions that can happen in the field. Field scans are based on a realistic water sample, but their true concentrations are only inferred.

---

[2]MATLAB was used to fit a line to the output of ScanVolt when run on calibration scans, and then to apply that linear relationship to ScanVolt output for field scans.

[3]Each group of five scans represents a time span of about three or four minutes, so successive group means are expected to be similar.

[4]Code is publicly available at `https://sourceforge.net/projects/instans/develop`

First Round

The first round used calibration scans only. In other words the "field scans" used in the test were actually calibration scans. Each calibration scan was measured based on a calibration derived from the other scans in the same calibration set (same electrode).

For example, the test data contains 70 calibration scans for the electrode `W121-Feb`. In the first round, the first of the 70 was analyzed based on a calibration derived from the other 69. Then the second of the 70 was analyzed based on a calibration derived from the first and the other 68. This pattern was continued to produce chemical concentration values for each of the 70.

Then, the first of the 60 calibration scans for electrode `W121-March` was analyzed based on a calibration derived from the other 59, and so on.

In this way, the first round produced computed concentrations for each calibration scan in the sample data. The computed values were then compared to the known values provided by the domain expert.

Second Round

The second round analyzed actual field scans, using calibration scans only to calibrate. For example, the first of the 7,650 field scans for electrode `W121-Feb` was analyzed based on a calibration derived from all 70 calibration scans for `W121-Feb`. Then the second of the 7,650 was analyzed based on that same calibration.

The first of the 4,690 field scans for electrode `W121-March` was then analyzed based on a calibration derived from all 60 calibration scans for electode `W121-March`.

In this way, the second round produced computed concentrations for each field scan in the sample data.

3.2.2   Comparison of Methods

For each method of measurement, once each scan was analyzed it was assigned an error. The errors of the individual scans were combined into three summary numbers indicating how well the particular method performed. Those three numbers were evaluated by a domain expert after each round.

Error of Individual Scans

Scans had actual concentrations assigned to them for the purpose of calculating the error. (See Section 3.1.3.)

The error was the magnitude of the difference between the system-determined concentrations and actual or expert-determined concentrations, treating them as vectors with each chemical as a dimension.

Observed Error

The observed error of a measurement method is the root mean squared error ("RMSE") of all measured scans. In other words, it is the square root of the mean of the squares of the errors. RMSE is a common indicator of goodness-of-fit. [22]

The purpose of the observed error is to indicate how well each method performed on the sample data.

Estimated Error

The estimated error is a bootstrap estimate of RMSE: the mean of the RMSEs of 10,000 bootstrap data sets generated from the underlying set of all measured scans. Each bootstrap set was the same size as the underlying set and generated by randomly selecting individual scans, with replacement.[5]

Bootstrap estimation is a simple method for inferential statistics that often works well and does not require a formula relating the sample statistic to the estimate. [7, 12]

The purpose of the estimated error is to estimate how well each method would perform on data which is similar to, but not exactly the same as, the sample data.

Threshold Error

The threshold error is the top of a bootstrap estimate of a 95% confidence interval of RMSE, derived from the same 10,000 bootstrap RMSEs as the estimated error by taking the 250th largest rather than mean.

The purpose of the threshold error is to indicate whether a method is consistent enough.

---

[5]To save time, resampling was actually done on the collection of squared errors rather than the collection of scans.

3.2.3 Points of Reference

A domain expert can judge the acceptability of methods based on the three measures of error from each round of testing. There are, however, some points of reference which a non-expert can compare to the results for a particular method:

- All the oxygen calibration scans have either 0 μM or 215 μM of oxygen. (See Section 3.1.2.) If the first round were run on only oxygen scans (it was not) a method that always reported the same concentration could achieve 107.5 μM of error on every scan.

- Treating the smoothed expert-provided concentrations as "true" values (as is done in round two) gives the *unsmoothed* expert-provided concentrations an RMSE of about 17 μM. This is not a theoretically significant value[6] but it can be seen as a practical estimate of a desirable level of error in round two.

- The existing solution aevum has an RMSE of roughly 150 μM when used to measure oxygen calibration scans.

3.3 Spline-Slopes Method

A spline is a piecewise polynomial with at least $n - 1$ continuous derivatives (even at the borders between pieces) where n is the degree of each polynomial piece. [5]

This method does a least-squares fit of a spline to the raw data, then assumes a linear relationship between the chemical concentration and the slope at its steepest point on the side of the corresponding peak.

3.3.1 Method

To measure the concentration of a chemical in a scan:

1. Fit the forward scan to a spline.

2. Take the derivative of the spline at a voltage previously determined during calibration.

3. Use a linear relationship, determined during calibration, to convert the derivative value into a concentration.

---

[6]This is because the smoothed values were actually derived from the unsmoothed values and the unsmoothed values had already been cleaned of outliers, averaged, and had some points coerced to zero

To calibrate for a chemical:

1. Fit the forward portion of each calibration scan to a spline.

2. Find the local minimum of the derivative on the peak in each spline with nonzero concentration.

3. Take the mean of the voltages of these minima. This is the voltage passed on for use in measurement.

4. Use that same voltage for splines from scans with zero concentration. (Evaluate the derivative of the spline at that voltage.)

5. This produces a (voltage, derivative) point for each calibration scan: at a local minimum of the derivative for nonzero scans and at the expected voltage for zero scans.

6. Fit the points to a line, using a least-squares fit.

3.3.2 Explanation

Although the concentration of the individual chemical is not the only factor affecting the derivative, its relative influence is likely to be strongest approximately at the minimum of the derivative because it is the concentration that creates the peak. Most methods rely on peak height, but the formula for the faradaic current of a single chemical can be written as a sum of exponentials in base e, one dominant during the upslope and the other during the downslope, so there is some basis for hoping that the derivative will, at that point, be roughly proportional to the concentration. [25] Also, an initial increase in current is a common feature of all peaks, so it may be a good feature to look for in peak detection.[7]

Fitting to a spline provides a function that can be used with standard numerical methods for finding minima and reduces noise. The minimum was found using SciPy's `fmin`, and the spline was fitted with SciPy's `splrep`.

---

[7]Some peaks, such as that for oxygen, are shaped more like plateaus and have little or no decrease after the peak.

3.4   Spline-Halves Method

3.4.1   Method

This method is similar to the spline-slopes method in Section 3.3, but an additional point is found. After the minimum of the derivative is found, the following root of the derivative is found.[8]

The minima of the derivative becomes an estimate of the half-peak potential, the point halfway up the peak, and the following root becomes an estimate of the top of the peak. From these, the method estimates the half-peak height and fits a linear relationship between that height and the chemical concentration.

The strategy of analyzing the derivative to find the half-peak height was used by More, et al. [20] to handle noisy, oxygen-only scans.

3.5   Multiple Discriminant Analysis Methods

3.5.1   Method

Duda, Hart, and Stork [6] present a multicategory generalization of Fischer linear discriminants. Given data in the form of d-dimensional points divided into c categories, the method finds a projection of the points into $c - 1$ dimensional space that maximizes the scattering between categories while minimizing scattering within categories. Both types of scattering are represented by scatter matrices, defined in a way that is analogous to sums of squared distances to the mean.

The variations tested here all calibrate in the following way:

1. Treat the calibration scans as many-dimensional points.

2. Find the multi-category Fischer projection, treating concentration levels as categories.

3. Project the points.

4. Fit the projected points to a flat surface using least-squares, treating concentration as the dependent variable.

5. Save the projection and the fitted surface for measurement.

Measurement is simple:

---

[8]If the following root of the derivative cannot be found, the following maximum of the derivative is found using `fmin`.

1. Treat the field scan as a many-dimensional point.

2. Project the field scan point.

3. Apply the fitted surface to assign a concentration to the projected point.

Variations

The variations differ in how scans become points, and all variations take the number of dimensions desired (before projection) as a parameter. All variations used a 21-point averaging filter for smoothing.

Resampled Smooth the scan and select evenly spaced current values as the values along the dimensions.

FFT Smooth the scan and then apply a Fourier transform. Take a prefix of the result as the values along the dimensions.

Resampled FFT Resample as in the resampling variation, but then apply a Fourier transform. Use the complete Fourier transform of the resampled data as the values along the dimensions

Differences FFT Smooth the scan, take the discrete differences (the difference between each sample and the next) then take the Fourier transform of the differences. Use a prefix of the result as the values along the dimensions.

3.5.2 Explanation

This form of multiple discriminant analysis finds a projection that emphasizes the differences between the categories and deemphasizes the similarities between categories. That could, in theory, be useful for cyclic voltammetry.

Consider the case of a 0 μM oxygen calibration scan and a 215 μM oxygen calibration scan. Certain differences between the 0 calibration and the 215 calibration are expected to also appear as differences between field scans with 0 and 215 μM of oxygen. Similarities between the two calibration scans, however, may not be shared with either field scan or may be shared with one of them but not the other for reasons that do not involve oxygen. The fact that calibration scans and field scans are known to be different in kind makes it even more important than usual to not let similarities among all calibration scans drive the calibration.

The code tested here found the projection by using SciPy's `linalg.eig` to solve the generalized eigenvector problem given by Duda, Hart, and Stork [6]:

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i.$$

The solutions for $\mathbf{w}_i$ corresponding to the largest solutions for λ_i are the columns of the optimal projection matrix $\mathbf{W}$.[9]

$\mathbf{S}_B$ and $\mathbf{S}_W$ come from the individual (unprojected) points $\mathbf{x}$, the category means $\mathbf{m}_i$, the number n_i of points in each category $\mathcal{D}_i$, and the overall mean $\mathbf{m}$:

$$\mathbf{S}_B \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t,$$

$$\mathbf{S}_W = \sum_{i=1}^{c} \mathbf{S}_i,$$

$$\mathbf{S}_i = \sum_{x \in \mathcal{D}_i} (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^t.$$

---

[9]Multiplying the transpose of $\mathbf{W}$ by a point gives the projected point.

Chapter 4

Results

4.1 Test Results

4.1.1 Choosing Parameters

The methods that were tested take parameters. For the methods in sections 3.3 and 3.4, the parameters specify the number of pieces in and the degree of the fitted spline. The MDA variants in section 3.5 take a single parameter specifying the number of dimensions.

To choose what the parameters should be set to during testing, the methods were "pre-tested" under round one conditions (calibration scans only) with various parameter values. Tables 4.1, 4.2, and 4.3 show the results for the best values found, plus a few parameter values near the best values for comparison.

Table 4.1 shows pre-test results for the spline-slopes method from Section 3.3. The best value for the degree of the spline appeared to be the lowest.[1] The top two had 65 and 80 pieces, beating in-between values, and the outcome seems to depend on the combination of degree and number of pieces in some way that is not obvious. This method appeared less sensitive than the others to the parameter settings.

Table 4.2 shows the pre-test results for the spline-halves method from Section 3.4. Aside from a general trend that lower degrees tended to be better and a lack of an obvious pattern, the results were quite different from those for the spline-slopes method. Different numbers of pieces turned out to be best for the different methods, and even the most accurate version of spline-halves

---

[1]The degree needed to be at least three because the analysis requires a continuous second derivative. It may seem counterintuitive that degree three gives better results than higher degrees. Splines of degree three have a minimum curvature property to which some have attributed their popularity [24], and this may have played a role in the results.

Table 4.1. Spline-Slopes Method with Various Parameters

(all errors in μM)

Name	Observed	Est.	Threshold
SplineSlopes(65 pieces, degree 3)	36.93	36.90	39.89
SplineSlopes(80 pieces, degree 3)	38.52	38.48	41.33
SplineSlopes(70 pieces, degree 5)	38.73	38.70	41.70
SplineSlopes(50 pieces, degree 3)	39.02	39.01	42.44
SplineSlopes(75 pieces, degree 3)	39.23	39.21	41.98
SplineSlopes(70 pieces, degree 3)	39.73	39.71	42.79
SplineSlopes(75 pieces, degree 4)	40.02	39.97	42.69
SplineSlopes(80 pieces, degree 5)	40.02	39.99	43.14
SplineSlopes(60 pieces, degree 3)	40.33	40.29	43.75
SplineSlopes(80 pieces, degree 4)	40.50	40.47	43.61
SplineSlopes(70 pieces, degree 4)	41.84	41.81	45.01
SplineSlopes(65 pieces, degree 4)	41.95	41.94	44.91
SplineSlopes(55 pieces, degree 3)	41.98	41.97	44.99
SplineSlopes(60 pieces, degree 4)	42.74	42.69	46.28
SplineSlopes(55 pieces, degree 5)	43.41	43.40	46.74
SplineSlopes(50 pieces, degree 4)	43.57	43.56	46.45
SplineSlopes(65 pieces, degree 5)	43.68	43.64	46.86
SplineSlopes(45 pieces, degree 4)	43.96	43.87	47.97
SplineSlopes(55 pieces, degree 4)	44.52	44.51	48.19
SplineSlopes(45 pieces, degree 3)	44.96	44.94	49.58
SplineSlopes(50 pieces, degree 5)	46.54	46.48	50.27
SplineSlopes(75 pieces, degree 5)	48.17	48.10	52.38
SplineSlopes(60 pieces, degree 5)	48.21	48.11	53.49
SplineSlopes(45 pieces, degree 5)	53.88	53.79	60.14

running time: 15.50 min.

was less accurate than all versions of spline-slopes shown in table 4.1. Also spline halves was more sensitive to the parameter values.

Table 4.3 shows the pre-test results for the MDA variations with different numbers of dimensions. The different variations all had their best results close to 200 dimensions, but some variations were much more sensitive to the number of dimensions than others.

FFT with 200 dimensions had the best observed error, but the estimated error is significantly worse (most methods had a lower estimated error than observed error) and the threshold error is much worse than the threshold error of other methods that did well in terms of observed error. That suggests 200-D FFT is a very inconsistent method.

Table 4.2. Spline-Halves Method with Various Parameters

(all errors in μM)

Name	Observed	Est.	Threshold
SplineHalves(25 pieces, degree 3)	132.03	131.74	149.65
SplineHalves(35 pieces, degree 4)	233.20	232.82	271.50
SplineHalves(10 pieces, degree 4)	294.58	291.75	375.96
SplineHalves(10 pieces, degree 3)	462.13	457.92	573.61
SplineHalves(15 pieces, degree 4)	600.60	598.03	718.32
SplineHalves(10 pieces, degree 5)	609.30	602.02	786.17
SplineHalves(15 pieces, degree 3)	877.82	868.89	1098.33
SplineHalves(20 pieces, degree 4)	922.54	911.86	1179.14
SplineHalves(15 pieces, degree 5)	1043.77	1027.25	1372.55
SplineHalves(35 pieces, degree 3)	1169.12	1157.33	1502.30
SplineHalves(25 pieces, degree 4)	1326.68	1310.24	1713.72
SplineHalves(30 pieces, degree 5)	1372.59	1347.87	1845.31
SplineHalves(25 pieces, degree 5)	1387.31	1360.27	1863.12
SplineHalves(20 pieces, degree 3)	1467.86	1456.14	1823.47
SplineHalves(20 pieces, degree 5)	1497.71	1471.24	2014.38
SplineHalves(30 pieces, degree 4)	1536.41	1500.71	2115.35
SplineHalves(35 pieces, degree 5)	1551.32	1522.18	2082.39
SplineHalves(30 pieces, degree 3)	1844.13	1824.58	2327.90

running time: 14.18 min.

4.1.2   Round One

Table 4.4 shows the results for the round one testing. Since the pre-testing to choose parameters was done under round one conditions, the rows of this table are essentially repeats of rows found in tables 4.1, 4.2, and 4.3.

A somewhat arbitrary choice was made to include the top three from each of the pre-test tables, as well as the single best setting for each MDA variation that was not in the top three.

In terms of observed error, all three versions of spline-slopes were the best, and all MDA methods except 203-D resampled FFT beat all versions of spline-halves. The inconsistency of the top two MDA methods stood out. 200-D FFT in particular is the only method in table 4.4 with a higher estimated error than observed error and has a threshold error higher than four methods which are inferior in terms of observed error.[2]

Six of the methods tested in round one had observed and estimated errors less than 107.5 μM. (See Section 3.2.3.) Four had threshold errors less than that amount.

---

[2]One way to think of estimated and threshold error is that the estimated error is less affected by outliers than the observed error, while the threshold error is more affected by high outliers and less affected by low outliers.

Table 4.3. MDA Methods with Various Parameters

(all errors in μM)

Name	Observed	Est.	Threshold
Fft(200-D)	55.60	80.65	217.18
DifferencesFft(202-D)	87.92	87.83	103.92
Fft(201-D)	91.24	91.20	99.09
Resampled(199-D)	113.66	113.57	122.67
Resampled(200-D)	113.66	113.59	122.61
Resampled(197-D)	113.66	113.59	122.55
Resampled(198-D)	113.66	113.64	122.34
DifferencesFft(204-D)	134.46	132.00	190.66
DifferencesFft(201-D)	171.84	164.00	279.81
Fft(202-D)	180.94	185.02	296.89
DifferencesFft(200-D)	186.37	176.44	341.13
Fft(198-D)	197.09	194.71	259.43
ResampledFft(203-D)	246.55	244.47	308.94
ResampledFft(204-D)	246.55	243.66	308.58
ResampledFft(201-D)	246.55	244.23	309.34
ResampledFft(202-D)	246.55	244.02	308.98
DifferencesFft(203-D)	315.13	311.33	411.48
ResampledFft(200-D)	535.98	529.03	704.06
Fft(199-D)	650.84	641.11	862.18
Resampled(201-D)	835.67	829.33	1015.54

running time: 51.75 min.

Despite being based on a method already in use, the spline-halves method was generally the worst in round one.

4.1.3 Round Two

Table 4.5 shows the results for round two. Three methods had all three measures of error under 150 μM. (See Section 3.2.3.) The best had an error of roughly 100 μM, and the second- and third- best had errors of roughly 140 μM.

Two methods improved their errors relative to round one. Spline-halves with 25 pieces and degree 3 improved by all three measures and jumped from eighth place to first place in terms of observed error. Spline-halves with 10 pieces and degree 4 also improved in all three measures and moved from last place to third.

200-D FFT and 202-D differences FFT both appeared more consistent in round two than in round one, in the sense that estimated and threshold errors were closer to the observed error. The

Table 4.4. Round One (Calibration Scans Only) Results

(all errors in μM)

Name	Observed	Est.	Threshold
SplineSlopes(65 pieces, degree 3)	36.93	36.92	39.89
SplineSlopes(80 pieces, degree 3)	38.52	38.49	41.44
SplineSlopes(70 pieces, degree 5)	38.73	38.69	41.65
Fft(200-D)	55.60	79.83	214.31
DifferencesFft(202-D)	87.92	87.61	103.66
Fft(201-D)	91.24	91.24	99.18
Resampled(199-D)	113.66	113.57	122.66
SplineHalves(25 pieces, degree 3)	132.03	131.80	149.41
SplineHalves(35 pieces, degree 4)	233.20	232.43	270.59
ResampledFft(203-D)	246.55	244.31	309.41
SplineHalves(10 pieces, degree 4)	294.58	292.21	376.67

running time: 16.82 min.

difference is that 200-D FFT was consistently bad, while 202-D differences FFT came in second place.

The spline-slopes versions were close together, as they were in round one, but performed near the bottom in round two even though they were at the top in round one.

4.2 Analysis

Which round should be believed? Round one says that spline-slopes is the best method and spline-halves the worst. Round two has it almost the other way around. To answer this question, it may be helpful to consider what, in general, were the differences between the two rounds.

Some of the differences involve the nature of calibration scans and field scans. On one hand, the true chemical concentrations of the calibration scans are known, while those of the field scans are only inferred. On the other hand, analyzing field scans is more realistic than analyzing calibration scans.

Field scans have more noise and more extraneous information than calibration scans. That could explain why spline-halves was the worst in round one and the best in round two, since it is inspired by a noise-resistant, chemical-specific method used by Moore, et al. [20] It would not explain why spline-halves improved numerically from round one to round two.

Table 4.5. Round Two (Field Scans) Results

(all errors in μM)

Name	Observed	Est.	Threshold
SplineHalves(25 pieces, degree 3)	99.55	99.55	100.49
DifferencesFft(202-D)	136.99	136.97	141.63
SplineHalves(10 pieces, degree 4)	138.61	138.61	139.90
Fft(201-D)	295.48	295.48	298.66
SplineSlopes(80 pieces, degree 3)	400.86	400.87	403.74
SplineSlopes(65 pieces, degree 3)	401.55	401.58	404.42
SplineSlopes(70 pieces, degree 5)	422.98	422.98	426.03
SplineHalves(35 pieces, degree 4)	423.78	423.77	430.28
Resampled(199-D)	468.20	468.22	472.92
Fft(200-D)	507.60	507.64	516.06
ResampledFft(203-D)	3912.58	3912.66	3959.87

running time: 188.26 min.

A possible explanation for lower error in round two is that the "true" values used in round two are wrong. The largest improvement in observed error was roughly 156 μM. It is possible, though it would be unfortunate, that the expert-provided concentrations have that much error.

Another possible explanation is that round two gave a better error estimate because it used more data. There were over one hundred times as many field scans as calibration scans in the sample data, and that could have helped round two.

It is likely that all the factors discussed above played a role, and it's not clear, on balance, which round is more reliable.

4.3   Possible Future Work

Scientists doing *in situ* cyclic voltammetry could simply begin using some of the tested methods and decide which ones they like, but there are several other opportunities for future work:

Automatic Output Cleaning Domain experts typically smooth and otherwise adjust the output of existing methods. For example, the expert-provided concentration values for the field scans used in round two were averaged in groups of five, cleaned of outliers, and in some cases coerced to zero. (All this was done by the domain expert in the normal course of using the data. See Section 3.1.3.) To give a clearer picture of accuracy, the outputs of the methods being tested were not adjusted in any way. In practice, smoothing and certain other types of cleaning (such as setting negative values to zero) could easily be made automatic.

Environmental Corrections The domain expert reported making no corrections for factors such as temperature or pH (See Section 2.2.1.) and so no corrections were made during testing. Such corrections should be tested to see how they affect the accuracy of the methods and should be included for users who want them.

Better Data The data used here was simply partial old data the domain expert had from previous work. A dataset created for testing might give a better picture of how the methods perform. Ideally, the dataset would contain field scans with known true concentrations. That could be done by taking a sample of seawater back to the lab and analyzing it using both cyclic voltammetry and other methods.

Other Methods Automatic testing of automatic methods has previously been absent from the analysis of *in situ* cyclic voltammetry data. The testing approach used here can be applied independently of the actual methods, and could potentially justify methods that would otherwise be too difficult to investigate.

4.4 Conclusion

Several of the methods tested were improvements over the existing prototype aevum. Aside from aevum, they are the first fully automatic methods for multi-chemical analysis of *in situ* cyclic voltammetry data, and their accuracies have been measured numerically.

Which single method among those tested is best depends on how the test results are interpreted. Further testing with better data should be done to help choose amongst the methods and improve on them. The difference, for testing purposes, between calibration and field scans may be larger than expected[3].

Users will ultimately decide whether any method meets their requirements.

---

[3]While working on this project, the domain experts and I generally assumed that analyzing a calibration scan rather than a field scan was an easier version of the same problem. The differences between round one and round two results suggest that different principles might apply.

Bibliography

[1] Christian Amatore and Bruno Fosset. Equivalence between microelectrodes of different shapes: Between myth and reality. *Analytical Chemistry*, 68(24):4377–4388, Dec 1996. doi: 10.1021/ac960421s.

[2] Paul J. Brendel and George W. III Luther. Development of a Gold Amalgam Voltammetric Microelectrode for the Determination of Dissolved Fe, Mn, O2, and S(-II) in Porewaters of Marine and Freshwater Sediments. *Environmental Science & Technology*, 29(3):751–761, 1995.

[3] Gwendolyn Bristow and Martial Taillefert. VOLTINT: A Matlab-based program for semi-automated processing of geochemical data acquired by voltammetry. *Computers & Geosciences*, 34(2):153–162, 2008.

[4] Trevor J. Davies and Richard G. Compton. The cyclic and linear sweep voltammetry of regular and random arrays of microdisc electrodes: Theory. *Journal of Electroanalytical Chemistry*, 585(1):63–82, 2005.

[5] C. de Boor. *A practical guide to splines*. Springer, 1978.

[6] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley & Sons, 2nd edition, 2001.

[7] Bradley Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1993.

[8] D. J. Gavaghan. An exponentially expanding mesh ideally suited to the fast and efficient simulation of diffusion processes at microdisc electrodes. 1. derivation of the mesh. *Journal of Electroanalytical Chemistry*, 456(1-2):1–12, 1998.

[9] D. J. Gavaghan. An exponentially expanding mesh ideally suited to the fast and efficient simulation of diffusion processes at microdisc electrodes. 2. application to chronoamperometry. *Journal of Electroanalytical Chemistry*, 456(1-2):13–23, 1998.

[10] D. J. Gavaghan. An exponentially expanding mesh ideally suited to the fast and efficient simulation of diffusion processes at microdisc electrodes. 3. application to voltammetry. *Journal of Electroanalytical Chemistry*, 456(1-2):25–35, 1998.

[11] Brian T. Glazer, Adam G. Marsh, Kevin Stierhoff, and George W. Luther. The dynamic response of optical oxygen sensors and voltammetric electrodes to temporal changes in dissolved oxygen concentrations. *Analytica Chimica Acta*, 518(1-2):93–100, 2004.

[12] Peter Hall and Susan R. Wilson. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47(2):757–762, 1991.

[13] George W. Luther III, Paul J. Brendel, Brent L. Lewis, Bjorn Sundby, Lucie Lefrancois, Norman Silverberg, and Donald B. Nuzzio. Simultaneous Measurement of O2, Mn, Fe, I-, and S(-II) in Marine Pore Waters with a Solid-State Voltammetric Microelectrode. *Limnology and Oceanography*, 43(2):325–333, 1998.

[14] George W. Luther III, Brian T. Glazer, Shufen Ma, Robert E. Trouwborst, Tommy S. Moore, Edouard Metzger, Chareonkwan Kraiya, Tim J. Waite, Gregory Druschel, Bjrn Sundby, Martial Taillefert, Donald B. Nuzzio, Timothy M. Shank, Brent L. Lewis, and Paul J. Brendel. Use of voltammetric solid-state (micro)electrodes for studying biogeochemical processes: Laboratory measurements to real time measurements with an in situ electrochemical analyzer (ISEA). *Marine Chemistry*, 108(3-4):221–235, 2008.

[15] Kenneth S. Johnson, Joseph A. Needoba, Stephen C. Riser, and William J. Showers. Chemical Sensor Networks for the Aquatic Environment. *Chemical Reviews*, 107(2):623–640, Jan 2007. doi: 10.1021/cr050354e.

[16] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[17] Robert Kalvoda, editor. *Electroanalytical Methods in Chemical and Environmental Analysis*. Plenum Press, 1987.

[18] Shufen Ma, George W. Luther III, Robert W. Scarborough, and Michael G. Mensinger. Voltammetry: An In Situ Tool to Monitor the Health of Ecosystems. *Electroanalysis*, 19(19-20):2051–2058, 2007.

[19] Tommy S. Moore, Katherine M. Mullaugh, Rebecca R. Holyoke, Andrew S. Madison, Mustafa Ycel, and George W. Luther. Marine Chemical Technology and Sensors for Marine Waters: Potentials and Limits. *Annual Review of Marine Science*, 1(1):91–115, 2009.

[20] Tommy S. Moore, Donald B. Nuzzio, Timothy W. Deering, Martial Taillefert, and George W. Luther III. Use of Voltammetry to Monitor O2 Using Au/Hg Electrodes and to Control Physical Sensors on an Unattended Observatory in the Delaware Bay. *Electroanalysis*, 19(19-20):2110–2116, 2007.

[21] R. S. Nicholson and Irving. Shain. Theory of stationary electrode polarography. single scan and cyclic methods applied to reversible, irreversible, and kinetic systems. *Analytical Chemistry*, 36(4):706–723, April 1964. doi: 10.1021/ac60210a007.

[22] Mark A. Pitt and In Jae Myung. When a good fit can be bad. *Trends in Cognitive Sciences*, 6(10):421–425, 2002.

[23] Edward J. Rykiel. Testing ecological models: the meaning of validation. *Ecological Modelling*, 90(3):229–244, 1996.

[24] M. Unser. Splines: a perfect fit for signal and image processing. *Signal Processing Magazine, IEEE*, 16(6):22–38, Nov 1999.

[25] Basil H. Vassos and Galen W. Ewing. *Electroanalytic Chemistry*. John Wiley & Sons, 1983.

[26] E. Viollier, C. Rabouille, S. E. Apitz, E. Breuer, G. Chaillou, K. Dedieu, Y. Furukawa, C. Grenz, P. Hall, F. Janssen, J. L. Morford, J. C. Poggiale, S. Roberts, T. Shimmield, M. Taillefert, A. Tengberg, F. Wenzhfer, and U. Witte. Benthic biogeochemistry: state of the art technologies and guidelines for the future of in situ survey. *Journal of Experimental Marine Biology and Ecology*, 285-286:5–31, 2003.

[27] R. Mark Wightman and David O. Wipf. High-speed cyclic voltammetry. *Accounts of Chemical Research*, 23(3):64–70, May 2002. doi: 10.1021/ar00171a002.

[28] David O. Wipf and R. Mark Wightman. Submicrosecond measurements with cyclic voltammetry. *Analytical Chemistry*, 60(22):2460–2464, May 2002. doi: 10.1021/ac00173a005.