OPEN SOURCE HARDWARE AND SOFTWARE IN AGRICULTURE:

AN AUTONOMOUS SAP FLOW MEASUREMENT WIRELESS NETWORK &

A USER FRIENDLY MANAGEMENT ZONE DELINEATION TOOL

by

CAMDEN LOWRANCE

(Under the Direction of George Vellidis)

ABSTRACT

The use of wireless sensor networks are vital to Precision Agriculture. Often these technologies are expensive or overly complex to producers. An open source microcontroller platform, Arduino, was used to develop a system to measure sap flow in grapevines (*Vitis vinifera*) in a vineyard in Greece, and to seamlessly transfer this data to a web server for storage. The performance of the system in the summer of 2014 is discussed. Management zones are a pillar of Precision Agriculture research. Spatial variability is apparent in all fields, and understanding this variability by delineating management zones can lead to better management decisions. A custom suite of applications called EZZone was developed for management zone delineation and is available for public use at ezzone.pythonanywhere.com. A demonstration and analysis of the tool is presented.

INDEX WORDS: Precision agriculture, Open source, Arduino, Sap flow, Wireless sensor networks, Agricultural management zones, Management zone delineation, GIS

OPEN SOURCE HARDWARE AND SOFTWARE IN AGRICULTURE:

AN AUTONOMOUS SAP FLOW MEASUREMENT WIRELESS NETWORK &

A USER FRIENDLY MANAGEMENT ZONE DELINEATION TOOL

by

CAMDEN LOWRANCE

BS, University of Georgia, 2012

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2014

OPEN SOURCE HARDWARE AND SOFTWARE IN AGRICULTURE:

AN AUTONOMOUS SAP FLOW MEASUREMENT WIRELESS NETWORK &

A USER FRIENDLY MANAGEMENT ZONE DELINEATION TOOL

by

CAMDEN LOWRANCE

Major Professor:	George Vellidis
Committee:	Miguel Cabrera
	Lan Mu

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

ACKNOWLEDGEMENTS

I would like to thank my major professor George Vellidis for a very unique and rewarding opportunity to live and work in Greece. I want to thank committee members Miguel Cabrera and Lan Mu for their guidance and Marc van Iersel for his backing and guidance of my sap flow project. For the organizers of my time in Greece, my thanks to Spyros Fountas and Theofanis Gemtos. I want to thank Thanos Balafoutis without whom I would likely still be trying to install my sensor network. To my friends in Greece: Matina, Vangelis, Thanos and Nikos, I will not forget your kindness and hospitality anytime soon. Finally I want to thank my parents Richard and Julie who have inspired me all these years to continue to challenge myself and my girlfriend Larissa for her support and devotion.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 **Precision Agriculture's importance, drawbacks, and application in this research**

Precision agriculture (PA) is the application of advanced technologies such as automated sensor networks and GPS-controlled vehicles that enhance the ability of producers to make spatially-aware management decisions. The goal of using these smart technologies is to increase profitability by maximizing output (yield) while optimizing inputs (water, fertilizer, labor etc.) by treating the field as a continually varying surface and adapting unique management to these varying zones of the field [1]. The economic benefits of precision farming are well-documented and these techniques have the potential to lessen the environmental impacts of agriculture while increasing profitability [2]. However, the ability of small-scale farmers to adopt these technologies is limited because of the high costs associated with the technology [3] and also because the development of new technology is designed for large scale production [1]. Adoption of PA by small scale producers can be accomplished by focusing on technologies that are less expensive and cognizant of the different management strategies between producers.

One major drawback of current technology is its usability by farmers [4]. Learning new software is time consuming and often not seen as worthwhile for farmers. The time needed to learn a new software or learn how to operate new machinery directly competes with essential

tasks such as planting. It is known that the complexity of a PA technology is negatively associated with its adoption, farmers are less likely to invest time and money into technology that is overly complex [4]. Thus new software should be automated in as many methods as possible to save valuable time and should be simple to learn and use. New technological designs should focus on providing farmers with valuable information at minimal time investment and maximal usability.

A serious limitation of existing FMIS is their lack of support for Geographic Information Systems (GIS) data [5]. GIS data is so essential to PA decision making that any software dealing with PA should store data in spatial formats. Another limitation of existing software is that it is too feature rich. Most existing software have a dizzying number of tools that often go unused. The latest innovations in technological integration have the potential to transfer Farm Management Information Systems (FMIS) onto web architectures so that they are accessible across the various devices found in a home. A web based technology focused on one specific task that could easily integrate with a wide range of software and technology would greatly increase the use of PA. Thus any new technology developed for management purposes should be web based and accessible across all common devices and it should readily incorporate with other GIS software and common interchangeable data formats.

1.2 **Importance of water management in vineyards**

Precision viticulture, PA in vineyards, is known to be a valuable management strategy under vineyard conditions where spatial variability of yield is relatively stable temporally (year to year) and where the causes of yield or quality variability can be identified and measured easily [6]. Water availability in vineyards is known to be the main factor determining yield and

quality of the grapes [7]. Many researchers consider water stress to be the primary cause of limited plant growth in dry regions of the world [8]. Soil water availability directly affects the uptake of water by grapevines (*Vitis vinifera*) [9]. We can integrate both plant and soil water data to model the water use of plants throughout the season. A system that can accurately model seasonal water use and that is affordable for small scale farmers would greatly increase the use of PA technology.

Like many commercial crops, water availability directly affects grape fruit yield and quality [7]. The climate of Greece is characterized by hot, dry summers with little rainfall. Yield and quality of non-irrigated plots in Greece are affected by environmental conditions [10]. It has been shown that water deficits can lead to grapes that have a better quality. Water deficits leads to an increase in sugar accumulation in the fruit and also the aromatic qualities of the grape [10]. Another important consideration is that different varieties of grapevines respond differently to water stress, indicating the importance of site-specific water monitoring and management [11]. High fruit quality is largely seen as more desirable than high yield, and the precision viticulture industry focuses on applying regulated deficit irrigation (RDI) to ensure that high quality is achieved [7].

The practice of RDI is based on measuring vine water status to ensure that a plant is water stressed at a level that promotes grape quality without jeopardizing yield [12]. It has been shown that RDI can reduce the need for water by as much as 40%, without altering yield and quality [12]. RDI can be a useful technique in dry areas where irrigation is limited because of the water saving potential [13]. The key factor for determining when to irrigate is the water status of the plant.

1.3 **Sap flow research on grapevines**

Sap flow is a common measurement to determine transpiration of grapevines. Sap flow is a synonym for the measurement of the sap flux density ($m^3\ m^{-2}\ s^{-1}$) which is the cubic meter of sap flow per square meter of stem cross section per second. Sap flow measurements are a direct method of measuring whole plant transpiration without disturbing the leaf canopy [14]. In a multi-year study on potted grapevines, it was shown that the correlation of total daily sap flow to total daily water consumption was essentially 1:1 (r^2= .98) and that the instantaneous leaf photosynthetic rate was also correlated strongly with instantaneous sap flow (r^2= .78) [14]. The authors compared the daily and hourly sap flow values with gravimetric measurements of plant water loss. An important aspect of the study is that the strong correlation of sap flow and gravimetric water loss was found over a range of different soil water contents, indicating that changes in soil water availability directly affect sap flow. The authors suggest the decrease in sap flow at decreased soil water content levels is a direct result of decreased stomatal conductance (*g)* of the plant, which was measured indirectly as leaf photosynthetic rate. Because of sap flow's correlation with daily leaf photosynthetic rate and with daily trunk growth (r^2= .91 and r^2= .67, respectively), it can also be used as an indicator of overall plant growth.

Another study relating sap flow to evapotranspiration (ET) found that there was a linear relationship between sap flow and solar radiation but sap flow was related non-linearly with Vapor Pressure Deficit (VPD) [15]. Thus it is important to measure field VPD to explain variability in sap flow measurements. The authors found that sap flow measurements were suitable for estimating total ET at the vineyard level. A study that took place in a Greek

vineyard, which we can assume had similar geographic characteristics as this proposed study, found that a critical daily sap flow value can be determined for irrigation scheduling [8]. In other words, if the daily sap flow of the plant does not exceed a certain value, then the plant is water stressed and should be irrigated. We should note that this study did not actually schedule irrigation from these values, but introduces a method based on minimum daily flow that will be applicable to this research.

There are notable examples of sap flow measurements being used directly for irrigation scheduling [16]–[18]. The study with the most importance to this proposal is the one of Braun & Schmidt [16]. The authors used Granier's Thermal Dissipation Probe (TDP) Method [19] to measure sap flow and schedule irrigation directly based on these measurements. This is the same sap flow method that will be used in this research. The researchers found strong correlation between weight loss of a potted plant and sap flow (r^2= .99). The study also found an apparent lag in sap flow measurements and suggests that the data should be integrated as hourly means. The correlation coefficients for 10, 30 and 60 minute integrations of sap flow and weight loss were .87, .92 and .94, respectively. This indicates that sap flow data are more suited for long-term water use measurements, although is still highly correlated at short intervals.

1.4 **Sap flow theory**

The TDP method works by measuring the heat dissipated from a thermocouple wrapped in a heating element and placed in the xylem of the plant. The amount of heat dissipated by the sap flow through the plant's xylem can be related to the water use of the plant. The system is comprised of two copper-constantan thermocouple probes inserted radially into the stem of

a woody plant. A thermocouple is a junction of two different metals. This junction produces a voltage that is proportional to the temperature at the junction of the two metals and the temperature can be directly calculated by the Seebeck effect [20]. The thermocouple is then entered into a steel needle to form the probe. One of these probes is wrapped with resistance heating wire and is constantly heated. This probe is then inserted into the xylem of the plant. The other probe is positioned upstream (below) the heated probe, and is used as a reference temperature. This reference probe automatically accounts for changes in the xylem temperature. The temperature difference (ΔT) between the two probes is measured to calculate the rate of sap flow. The difference in temperature is associated with the heat dissipation from the heated probe by upward movement of water through the xylem or sapwood. In other words, high sap flow = low differential temperature. Figure 1.1 presents an illustrative diagram of the sap flow probes.

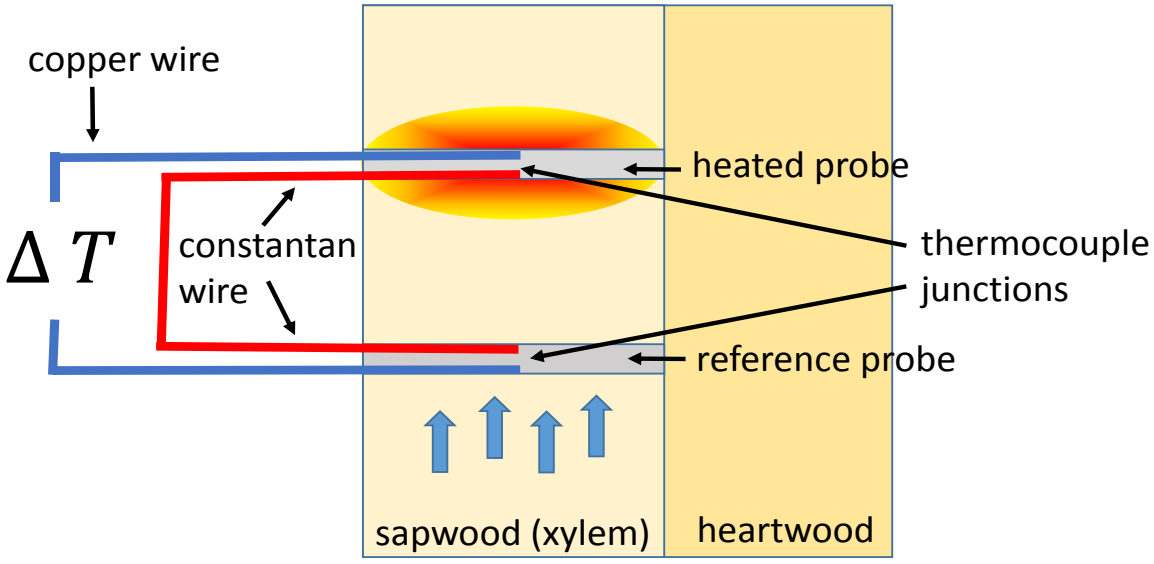


Figure 1.1: Sap flow sensor diagram

When the heated thermocouple probe and the tree are in thermal equilibrium and the sap flux density is constant, the heat input is equal to the amount of heat dissipated at the wall of the probe [19], [21]:

$$h\,S\,(T - T_f) = R\,I^2$$

Equation 1.1

h = coefficient of heat exchange (watts per square meter, per °C [$W \cdot m^{-2} \cdot °C^{-1}$])

S = exchange surface area (m^2)

T = temperature of the heated probe (°C)

T_f = temperature of the wood in absence of heating; reference probe (°C)

R = electrical resistance of the heating element (ohms [Ω])

I = current passed through heating element (amperes [A])

*note: $R \cdot I^2$ is a derivation of Ohm's law which is equal to power (watts [P])

The coefficient of heat exchange (h) was found to be related to the sap flux density (also stated as sap flow) by the following equation:

$$h = h_0\,(1 + \alpha F_d{}^{\beta})$$

Equation 1.2

h_0 = coefficient of heat exchange when sap flux = 0 (F_d = 0)

F_d = sap flux density (cubic meters, per square meter, per second [$m^3 \cdot m^{-2} \cdot s$])

α and β are regression coefficients from Granier, see equation 6 (1985)

We can calculate the coefficient of heat exchange at zero sap flux (h_0) as:

$$h_0 = \frac{P}{S\,(T_{max} - T_f)}$$

Equation 1.3

h_0 = coefficient of heat exchange when sap flux = 0 (F_d = 0)

S = exchange surface area (m^2)

T_{max} = temperature at zero flux (F_d = 0)

T_f = temperature of the wood in absence of heating; reference probe (°C)

$P = R * I^2$

When T_{max} = 0, we can assume that only conductive heat loss is occurring because there is no sap flow. When F_d is not equal to zero and is constant, the sap flux density (F_d) can be calculated as:

$$F_d = \left[\frac{1}{\alpha} \times \frac{h \times h_0}{h}\right]^{1/\beta}$$

Equation 1.4

We can then combine equations 1, 3 and 4 to calculate the sap flux density:

$$F_d = \left[\frac{1}{\alpha} \times \frac{(T_{max} \times T_f) - (T \times T_f)}{T \times T_f}\right]^{1/\beta} = \left[\frac{1}{\alpha} \times \frac{\Delta T_{max} - \Delta T}{\Delta T}\right]^{1/\beta}$$

Equation 1.5

$\frac{\Delta T_{max} - \Delta T}{\Delta T}$ can be called K, flow index (no units)

The original calibration of these sensors were carried out by Granier (1985) on three different tree species and found high correlation between K and F_d. This correlation was found to be independent of tree type and sapwood type. A nonlinear regression model was found to fit the relationship between these two measurements. The equation proposed by Granier (1985) to calculate the sap flux density (F_d) from the flow index (K) is:

$$F_d = 0.000119 \times K^{1.231}$$

Equation 1.6

The above measurement only calculates sap flow in a cylinder encompassing the probe. To calculate the sap flow for the entire stem, we need to know the amount of stem containing xylem flow. We will follow the assumption that the entire stem of the grapevine has some sort of flow as proved by a previous study on the Granier TDP method in grapevines [16]. Thus our equation to calculate the flow rate across the entire stem cross section is:

$$F = F_d \times A_{sw}$$

Equation 1.7

F = whole tree sap flow ($m^3\ s^{-1}$)

A_{sw} = sapwood area (m^2)

This original calibration equation fixed the heating power at 0.2 W. This value was found to not heat the reference temperature probe while providing sufficient sensitivity for the heated probe. Several modifications of the Granier design have been developed, including commercial varieties. A critical review of these designs and the original design indicate that any changes to heat field or geometry of the probes will need new calibration methods [21].

1.5 **Considerations and limitations of sap flow measurements**

The most critical component of sap flow calculation is accurate estimation of the maximum voltage (temperature) difference recorded. Practical applications of the TDP method indicate that the zero-flow state can be limited by weather conditions and vegetative growth in the plant [21]. Thus basic measurements of weather conditions such as temperature and humidity will be implemented and sap flow readings will be assessed for validity. Vapor Pressure Deficit has a strong relationship with sap flow, so this environmental measure will be also be used to validate sap flow measurements. Installation of the probes can also affect the sensor readings, and loss of thermal equilibrium can cause the values to drift over time [21]. Thus a determination of ΔT_{max} should be made over a 7 – 10 day period and should also be determined separately for each probe [22].

Probes are typically installed on the North side of the tree or vine and insulated with reflective insulation to limit solar radiation interference. Morning peaks in sap flow measurements have been observed when probes are exposed to direct sunlight [21]. Another reason to keep the direction constant is that sap flow is not uniform within a plant's xylem. Therefore we should seek to control sap flow probe's insertion parameters to limit probe location discrepancies. There are correction techniques proposed to deal with extensive drought periods and other complications that will be performed if necessary [21]. The Granier TDP method only measures sap flux density within a small cylinder. There are extrapolation methods to determine total sap flow of the tree (*F*) given the measurement area within the xylem of active flow. A study applying the Granier TDP technique to vineyards found that the active conductive xylem completely encompassed the probes and that even 20 year old

grapevines did not develop heartwood [16]. This fact discovered by Braun & Schmidt will also be beneficial in determining the sapwood area. Since we can assume that there is no heartwood in grapevines, the area is simply the area of the stem minus the area of the bark. Given the complexity of the technique, care will be taken to ensure accurate measurements are obtained.

1.6 **Management zone use in precision agriculture**

A management zone is simply a portion of a field that is more homogenous based on a certain measureable characteristic (e.g. soil type, crop yield) than the surrounding area [23]. The number and size of zones are determined by the algorithm producing the zones.

Management zones (MZs) in PA have been used to increase crop productivity, optimize inputs, and reduce environmental costs. MZs are particularly important in Variable Rate Technology (VRT) fertilizer applications [24]. Limitations in grid soil sampling have in turn led to an increased use of management zones to identify areas for heterogeneous fertilizer application. The zones can be integrated with fertilizer application equipment to actively control the amount of a nutrient dispersed. Variable Rate Irrigation (VRI) can incorporate defined management zones to determine irrigation amounts in VRI center pivot technology [25]. When integrated with real time field monitoring, MZs allow farmers to precisely control irrigation amounts throughout their fields.

There are a variety of factors in the field affecting crop yield that can be assessed at detailed spatial resolution with modern technology. It has been shown that soil nutrient levels have a marked effect on crop yield [24]. Because of the large variability in soil qualities, the usage of management zones for Nitrogen application has shown to reduce Nitrogen usage

while not affecting yields [26]. Crop placement can also affect yield, with more eroded areas experiencing less crop yield [27]. While a variety of factors affect crop yield, not all should be included in a management zone delineation tool.

It is important to choose proper characteristics to use in MZs classification. The majority of field characteristics used for determination are temporally stable and correlated with yield. Soil measurements such as electrical conductivity (EC) and fertility are common input parameters. Since EC variability also represents variability of other soil properties, it has become a common measure for zone definition [28]. The measurement is easy to obtain and is often correlated with crop yield, giving it a decided advantage over other measurements such as soil chemical and physical analysis and landscape slope, which require time consuming data acquisition [29]. Because of these advantages, soil EC data will be used as the basis for demonstration of the Management Zone Delineation (MZD) tool presented here.

As resources such as water and fertilizer continue to increase in scarcity, it is important to no longer treat the field as a homogenous surface but to address the spatial variability by creating zones of heterogeneous management. The use of management zones is a promising approach to mitigating agricultural inputs while still obtaining profitability [23]. The process of creating management zones can be daunting for anyone unfamiliar with spatial relationships and analysis. There exists currently one commonly used free tool for management zone delineation, a Windows based software known as Management Zone Analyst [30]. This software has drawbacks that will be directly addressed by this research project. The tool produced by this research will be user friendly and as automated as possible. It will also easily incorporate with other GIS software by outputting the zones in common GIS file formats. By

reducing the barriers to these advanced management tools, this research will allow PA to benefit society on a larger scale.

1.7 **Management zone delineation techniques**

The basis of all management zone delineation (MZD) techniques is grouping of data values into similar classes so that the members of the class are closer in relation than to members of other classes [31]. Although a variety of techniques have been used or proposed, there is no standard procedure and the success of a procedure is largely determined by the input dataset [32]. One common clustering algorithm is the iterative self-organizing data analysis technique (ISODATA). It is an unsupervised technique that clusters data points by minimizing the Euclidean distance between the points [33]. The requirement of variables with Gaussian distributions makes this technique undesirable for a widely used approach such as presented here.

Fuzzy *c*-means (or *k*-means) is a common technique for clustering data that does not require a Gaussian distribution for input data [34]. The technique seeks to minimize the sum of squared distances from data points within a cluster domain. This is an iterative approach in that the cluster criterion, the sum of squared values within the cluster, are recalculated each time a new data point is added to a cluster. The technique is fuzzy in the sense that there are no defined group boundaries, and data points can belong to more than one group, in varying strength of membership. This process has been widely used in PA, and a software known as Management Zone Analyst has been created to simplify use of the technique [30].

1.8 **Management Zone Analyst**

It is worth discussing the Management Zone Analyst (MZA) software because it is one of the only known free tools for MZD [30]. The software runs on Windows systems and has a custom Graphical User Interface. It accepts text files as input data sources, which can be multivariate. There are a few drawbacks of this software. The first is that it was developed to run with Window NT and not been updated since. This limits its ability to be used on modern platforms. Another disadvantage is that it requires skilled knowledge of statistics to understand and choose the correct model parameters for MZD. The third and most important drawback of this technique is that it produces no GIS data file that readily incorporates with other software. A user of MZA needs to interpret a variance-covariance matrix and convert a text file to point data, then to a usable form such as a raster or polygon. The conversion of a point data set to a better representative of a field such as a polygon or raster is no simple matter. This often involves an interpolation algorithm to determine values between points based on the known values and the distance from known values. Because the output of MZA is an ordinal classification number i.e. zone 1, zone 2, zone 3, the interpolation to field data is limited to a few number of techniques. Because of this limitation, point data must first be gridded to a field surface before input into the MZA software [35], [36]. Thus a simplified technique that automated the process of gridding and aligning the point data set to the field that produces zones in common GIS file formats would likely foster much wider use of MZs.

1.9 **Overview and inspiration of proposed tool**

The desire to simplify MZD was obtained after spending many hours in commercial GIS software attempting to devise a useful automated algorithm. The well-known software is very robust in its analysis capabilities, but unfortunately quite expensive. I am of the opinion that publicly funded research should be as committed to free and open access as possible. Thus the abandonment of commercial software and adoption of open source programming languages for a custom tool. A MZD tool should take into account the desired number of zones from users, i.e., it is useless to create a complex zone map if only two regions are desired. The tool should also conform to the spatial characteristics of the field, such as row spacing, row orientation, and planter width. It is pointless to create zones that are smaller than the resolution of a farmer's VRT, especially considering that farmers adjust zones that are unnecessarily complex to conform with logistical constraints [4]. The tool should also work with the type of field of the user, in that the algorithm is different according to field shape and irrigation controller. It should be noted that by emphasizing usability over statistical accuracy, we do create less "perfect" zones, but at the benefit of creating a tool that is accessible to a much wider audience than the research community.

The MZD delineation tool is developed using the Python programming language (Python.org). Python has a rich history of use in scientific fields, especially in GIS analysis. The backbone of the scientific computation in Python is in its *Scipy* and *Numpy* modules [37]. These two modules allow for complex mathematical modeling and analysis. *Numpy* provides support for multidimensional arrays and has object oriented functions to operate on these arrays. *Scipy* provides scientific computing functions for Python such as interpolation of point data to a

defined grid. There are also numerous GIS modules for the language, with *Shapely* being the most relevant to this project (toblerity.org/shapely/). Focusing on GIS integration allows the tool's result to seamlessly transfer to a Farm Management Information System and eliminates the extra expertise needed to generate GIS zones using third-party software. Python also has modules for custom web app development such as *Flask* (flask.pocoo.org). By putting the tool on the web, we ensure open access across all platforms, including mobile browsers, and allow for user input into the MZD. The user will be able to modify field and zone boundaries, input desired spacing based on technology constraints, and draw lines indicating the row orientation. Creating a semi supervised MZD tool ensures that the user feels involved in the process and that the results are unique to both the user's management philosophy and the technological and physical field constraints.

1.10 **Research goals**

The main goals of this thesis:

1. To develop a system that measures water use by sap flow in vineyards that uses low cost, open source hardware and software;
2. To model seasonal water intake accurately and to correlate sap flow measurements with other common environmental and plant parameters; and
3. To create a public access management zone classification tool that automatically interprets field variability and can be used for a variety of management tasks.

One broad goal of this research is to help infuse the use of open source technology into the field researcher's toolkit. Open source means that it is free to use, modify, and share with anyone. While certainly not the first example of open source systems in scientific research, the methods proposed adapt a homemade sensor to an open source control, logging and data transfer system. This research demonstrates the advantages of do-it-yourself field research and thus encourage others to replicate and expand on my results. Another goal is to decrease the barriers to individuals deploying custom wireless sensor networks for agronomic research. Detailed pictures and tutorials have been created for the installation process, and the system prototyping and testing is documented online for public access. Also, written software code is published online for open access on the GitHub repositories *camdenl/sapFlow* and *camdenl/EZZone*. These systems are optimized for use on small farms and are designed to be simple and effective solutions to otherwise expensive management strategies.

I have modeled seasonal water uptake by measuring sap flow of vines that are representative of field variability. I have also assessed field variability in soil moisture, NDVI, midday stem water potential and grape yield and quality. Combining all these measurements, an analysis of their relation to plant water use via sap flow has been obtained. I expect to identify spatial variability with these parameters and discuss the zone specific management practices that could be developed from these results. I have identify primary factors affecting this variability, and discuss implications for future research given these results.

Management zone classification can be accomplished in a variety of ways. The most common approach is to cluster a point data set of field characteristics based on the measured value of a parameter such as soil moisture. This technique is used in this research, but in a

novel and more GIS-centric manner. The entire process will be focused from a spatial analysis perspective so that the outcome is spatially explicit, meaning it can be loaded into GPS controlled farm machinery with little effort. I treat the field as a continuously varying surface, or raster, and cluster zones based on the whole field value, not just at specific data points. The tool also allows farmers to input parameters such as row or planter spacing, so the zones will readily integrate with GPS technologies and also align with existing field dimensions.

This tool is placed on a web server and uses a Python based web framework known as Flask and a Javascript based web mapping package known as OpenLayers (openlayers.org). Python is an object oriented programming language that has a wide use across many fields. JavaScript is a dynamic programming language that is primarily used for interactive website design in a web client environment. Combining these two technologies allows me to create a user friendly web front end while having all the power of Python for the actual management zone computation. Both the users' field and the management zones created are displayed on the web application. The user will then be able to download the produced zones in commonly used GIS data formats. Based on my literature reviews, this is a novel tool and will be a valuable solution for spreading advanced management strategies to small scale producers, consultants, and other agricultural practitioners.

CHAPTER 2

AN AUTONOMOUS SAP FLOW MEASUREMENT WIRELESS NETWORK[1]

[1]Lowrance, C. and Vellidis, G. To be submitted to *Precision Agriculture*

Abstract

The use of advanced technologies in agriculture is becoming essential to making smart management decisions. Often these technologies are expensive and inaccessible to small-scale producers. The advent of open source hardware and software has made the accessibility of advanced technologies greater than ever before. Open source simply means that access and modification is free and open to the public. An open source microcontroller platform, Arduino, and its analogues, have been used successfully in many industries, including agriculture. The presented research used Arduino-based technology as a vehicle for monitoring the water status of plants with wireless sensor networks. Specifically, a system was developed to measure sap flow in grapevines (Vitis vinifera) in a vineyard in Greece, and to seamlessly transfer this data to a web server for storage. The goal of this research was to increase the ability of small-scale farmers to more accurately measure their plant's water use by developing a low-cost monitoring system from easily obtainable materials.

The system of 17 nodes and a base station were installed in a vineyard outside of Athens, Greece in summer 2014. The system measured whole tree sap flow rates (*F*) for two months and during this time Normalized Difference Vegetation Index (NDVI) and Midday Stem Water Potential (MSWP) values were recorded for each node. Vapor Pressure Deficit (VPD) values were calculated for the whole field. The system had technical issues limiting the amount of sap flow data recorded and six nodes' *F* values were deemed valid for analysis. Field variability of *F* was determined to be significantly different at p=.01 by a Kruskall-Wallis test. The values exhibited moderate to strong correlation with NDVI, MSWP and VPD. Linear regression results for maximum daily *F* and VPD for each node were found significant at p=.05

for four nodes. A linear regression result for maximum daily *F* and for NDVI for all nodes was significant at p=.05. Suggested improvements to the system are presented and uses for the system are discussed.

2.1 **Introduction**

2.1.1 Precision Agriculture's importance, drawbacks, and application in this research

Precision agriculture (PA) is a catch-all term for techniques, technologies, and management strategies aimed at addressing within-field variability of parameters that affect crop growth. These parameters may include soil type, soil organic matter, plant nutrient levels, topography, water availability, weed pressure, insect pressure, etc. The application of advanced technologies such as automated sensor networks and GPS-controlled vehicles enhance the ability of producers to make spatially-aware management decisions. The goal of using these smart technologies is to increase profitability by maximizing output (yield) while optimizing inputs (water, fertilizer, labor etc.) by treating the field as a continually varying surface and adapting unique management to these varying zones of the field [1]. The economic benefits of precision farming are well-documented and these techniques have the potential to lessen the environmental impacts of agriculture while increasing profitability [2]. However, the ability of small-scale farmers to adopt these technologies is limited because of the high costs associated with the technology [3] and also because the development of new technology is designed for large scale production [1]. Adoption of PA by small scale producers can be accomplished by focusing on technologies that are less expensive and cognizant of the different management strategies between producers.

One major drawback of current technology is its usability by farmers [4]. Learning how to use new software or hardware is time consuming and often not seen as worthwhile for farmers. The time needed to learn new software or learn how to operate new technology directly competes with essential tasks such as planting. It is known that the complexity of a PA

technology is negatively associated with its adoption; farmers are less likely to invest time and money into technology that is overly complex [4]. Thus new software should be automated in as many methods as possible to save valuable time and should be simple to learn and use. New technology should focus on providing farmers with valuable information at minimal time investment and maximal usability.

2.1.2 Water management and precision agriculture

Irrigation has become essential to crop production in many agricultural areas of the world. As a result, the competition for available fresh water supplies is increasing. If irrigated agriculture is to survive in this competitive environment, we must use irrigation water efficiently especially in semi-arid and arid regions. A large number of techniques and tools have been developed to assist irrigation system users (irrigators) and especially producers to estimate when and how much water to apply to crops. Yet, despite the availability of these techniques and tools, the vast majority of irrigators still rely either on a fixed schedule or on visual cues of plant stress such as wilting. And typically, irrigators will apply a standard amount (for example 2.5 cm) at each irrigation event. As a result, both the timing and depths of irrigation may be inappropriate and may lead to yield, nutrient, and soil losses. Vories et al. (2006) found that improper timing of irrigation on cotton can result in yield losses of between USD 370/ha to USD 1850/ha. Precision agriculture applied to irrigation, also known as precision irrigation, offers the potential for improving both irrigation efficiency and yields.

2.1.3 Importance of water management in vineyards

Precision viticulture, PA in vineyards, is known to be a valuable management strategy under vineyard conditions where spatial variability of yield is relatively stable temporally (year

to year) and where the causes of yield or quality variability can be identified and measured easily [6]. Water availability in vineyards is known to be the main factor determining yield and quality of the grapes [7]. Many researchers consider water stress to be the primary cause of limited plant growth in dry regions of the world [8]. Soil water availability directly affects the uptake of water by grapevines (*Vitis vinifera*) [9]. We can integrate both plant and soil water data to model the water use of plants throughout the season. A system that can accurately model seasonal water use and that is affordable for small scale farmers would greatly increase the use of PA technology.

Like many commercial crops, water availability directly affects grape fruit yield and quality [7]. The climate of Greece is characterized by hot, dry summers with little rainfall. Yield and quality of non-irrigated plots in Greece are affected by environmental conditions [10]. It has been shown that water deficits can lead to grapes that have a better quality. Water deficits leads to an increase in sugar accumulation in the fruit and also the aromatic qualities of the grape [10]. Another important consideration is that different varieties of grapevines respond differently to water stress, indicating the importance of site-specific water monitoring and management [11]. High fruit quality is largely seen as more desirable than high yield, and the precision viticulture industry focuses on applying regulated deficit irrigation (RDI) to ensure that high quality is achieved [7].

The practice of RDI is based on measuring vine water status to ensure that a plant is water stressed at a level that promotes grape quality without jeopardizing yield [12]. It has been shown that RDI can reduce the need for water by as much as 40%, without altering yield and quality [12]. RDI can be a useful technique in dry areas where irrigation is limited because

of the water saving potential [13]. The key factor for determining when to irrigate is the water status of the plant.

Precision irrigation technology in vineyards is primarily drip irrigation systems that can be controlled at varying degrees of resolution, up to vine-specific amounts. Irrigation studies generally divide the field into zones and determine the irrigation amount for the whole zone based on a parameter measured in the zone [7], [38]. In the studies, irrigation was supplied to the plants when they crossed a water stress threshold and the quality for each irrigation zone were compared to a well-irrigated control group at the end of the season. The results showed it is possible to increase quality of the grapes by controlling irrigation based on the water status of the plant at different levels throughout the field. Spatial variability in vineyard water use can be assessed by a sensor network measuring water use of plants throughout the field. By designing a low cost sensor network, we can assess spatial variability of water use at a satisfactory resolution to manage the vineyard variably. Variable irrigation can lead to an overall better quality of the grape crop by actively controlling irrigation based on the variable water status of the field.

2.1.4 Sap flow research on grapevines

Sap flow is a common measurement to determine transpiration of grapevines. Sap flow is a synonym for the measurement of the sap flux density ($m^3\ m^{-2}\ s^{-1}$) which is the cubic meter of sap flow per square meter of stem cross section per second. Sap flow measurements are a direct method of measuring whole plant transpiration without disturbing the leaf canopy [14]. In a multi-year study on potted grapevines, it was shown that the correlation of total daily sap flow to total daily water consumption was essentially 1:1 (R^2= .98) and that the instantaneous

leaf photosynthetic rate was also correlated strongly with instantaneous sap flow (R^2= .78) [14]. The authors compared the daily and hourly sap flow values with gravimetric measurements of plant water loss. An important aspect of the study is that the strong correlation of sap flow and gravimetric water loss was found over a range of different soil water contents, indicating that changes in soil water availability directly affect sap flow. The authors suggest the decrease in sap flow at decreased soil water content levels is a direct result of decreased stomatal conductance (*g)* of the plant, which was measured indirectly as leaf photosynthetic rate. Because of sap flow's correlation with daily leaf photosynthetic rate and with daily trunk growth (R^2= .91 and R^2= .67, respectively), it can also be used as an indicator of overall plant growth.

Another study relating sap flow to evapotranspiration (ET) found that there was a linear relationship between sap flow and solar radiation but sap flow was related non-linearly with Vapor Pressure Deficit (VPD) [15]. Thus it is important to measure field VPD to explain variability in sap flow measurements. The authors found that sap flow measurements were suitable for estimating total ET at the vineyard level. A study that took place in a Greek vineyard, which we can assume had similar geographic characteristics as this proposed study, found that a critical daily sap flow value can be determined for irrigation scheduling [8]. In other words, if the daily sap flow of the plant does not exceed a certain value, then the plant is water stressed and should be irrigated. We should note that this study did not actually schedule irrigation from these values, but introduces a method based on minimum daily flow that will be applicable to this research.

There are notable examples of sap flow measurements being used directly for irrigation scheduling [16]–[18]. The study with the most importance to this proposal is the one of Braun & Schmidt [16]. The authors used Granier's Thermal Dissipation Probe (TDP) Method [19] to measure sap flow and schedule irrigation directly based on these measurements. The TDP method is based on the difference in temperature between a heated and reference probe that are inserted into the plant's xylem. This is the same sap flow method used in this research. The researchers found strong correlation between weight loss of a potted plant and sap flow (R^2= .99). The study also found an apparent lag in sap flow measurements and suggests that the data should be integrated as hourly means. The correlation coefficients for 10, 30 and 60 minute integrations of sap flow and weight loss were .87, .92 and .94, respectively. This indicates that sap flow data are more suited for long-term water use measurements, although is still correlated at short intervals.

2.1.5 Sap flow theory

The TDP method works by measuring the heat dissipated from a thermocouple wrapped in a heating element and placed in the xylem of the plant. The amount of heat dissipated by the sap flow through the plant's xylem can be related to the water use of the plant. The system is comprised of two copper-constantan thermocouple probes inserted radially into the stem of a woody plant. A thermocouple is a junction of two different metals. This junction produces a voltage that is proportional to the temperature at the junction of the two metals and the temperature can be directly calculated by the Seebeck effect [20]. The thermocouple is then entered into a steel needle to form the probe. One of these probes is wrapped with resistance heating wire and is constantly heated. This probe is then inserted into the xylem of the plant.

The other probe is positioned upstream (below) the heated probe, and is used as a reference temperature. This reference probe automatically accounts for changes in the xylem temperature. The temperature difference (ΔT) between the two probes is measured to calculate the rate of sap flow. The difference in temperature is associated with the heat dissipation from the heated probe by upward movement of water through the xylem or sapwood. In other words, high sap flow = low differential temperature. Figure 1.1 presents an illustrative diagram of the sap flow probes.

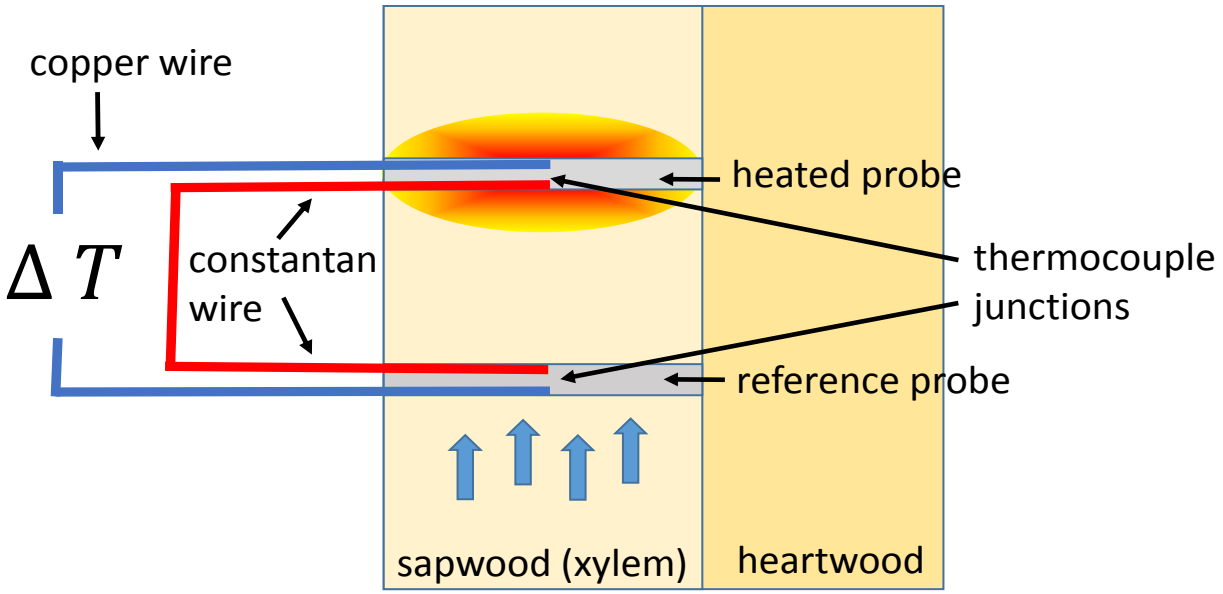


Figure 2.1: Sap flow sensor diagram

When the heated thermocouple probe and the tree are in thermal equilibrium and the sap flux density is constant, the heat input is equal to the amount of heat dissipated at the wall of the probe [19], [21]:

$$h\,S\,(T - T_f) = R\,I^2$$

Equation 2.1

h = coefficient of heat exchange (watts per square meter, per °C [$W \cdot m^{-2} \cdot °C^{-1}$])

S = exchange surface area (m^2)

T = temperature of the heated probe (°C)

T_f = temperature of the wood in absence of heating; reference probe (°C)

R = electrical resistance of the heating element (ohms [Ω])

I = current passed through heating element (amperes [A])

*note: $R \cdot I^2$ is a derivation of Ohm's law which is equal to power (watts [P])

The coefficient of heat exchange (h) was found to be related to the sap flux density (also stated as sap flow) by the following equation:

$$h = h_0 (1 + \alpha F_d{}^{\beta})$$

Equation 2.2

h_0 = coefficient of heat exchange when sap flux = 0 (F_d = 0)

F_d = sap flux density (cubic meters, per square meter, per second [$m^3 \cdot m^{-2} \cdot s$])

α and β are regression coefficients from Granier, see equation 6 (1985)

We can calculate the coefficient of heat exchange at zero sap flux (h_0) as:

$$h_0 = \frac{P}{S (T_{max} - T_f)}$$

Equation 2.3

h_0 = coefficient of heat exchange when sap flux = 0 (F_d = 0)

S = exchange surface area (m^2)

T_{max} = temperature at zero flux (F_d = 0)

T_f = temperature of the wood in absence of heating; reference probe (°C)

$P = R * I^2$

When T_{max} = 0, we can assume that only conductive heat loss is occurring because there is no sap flow. When F_d is not equal to zero and is constant, the sap flux density (F_d) can be calculated as:

$$F_d = \left[\frac{1}{\alpha} \times \frac{h \times h_0}{h}\right]^{1/\beta}$$

Equation 2.4

We can then combine equations 1, 3 and 4 to calculate the sap flux density:

$$F_d = \left[\frac{1}{\alpha} \times \frac{\left(T_{max} \times T_f\right) - \left(T \times T_f\right)}{T \times T_f}\right]^{1/\beta}$$

$$= \left[\frac{1}{\alpha} \times \frac{\Delta T_{max} - \Delta T}{\Delta T}\right]^{1/\beta}$$

Equation 2.5

$\frac{\Delta T_{max} - \Delta T}{\Delta T}$ can be called *K*, flow index (no units)

The original calibration of these sensors were carried out by Granier (1985) on three different tree species and found high correlation between *K* and F_d. This correlation was found

to be independent of tree type and sapwood type. A nonlinear regression model was found to fit the relationship between these two measurements. The equation proposed by Granier (1985) to calculate the sap flux density (F_d) from the flow index (K) is:

$$F_d = 0.000119 \times K^{1.231}$$

Equation 2.6

The above measurement only calculates sap flow in a cylinder encompassing the probe. To calculate the sap flow for the entire stem, we need to know the amount of stem containing xylem flow. We will follow the assumption that the entire stem of the grapevine has some sort of flow as proved by a previous study on the Granier TDP method in grapevines [16]. Thus our equation to calculate the flow rate across the entire stem cross section is:

$$F = F_d \times A_{sw}$$

Equation 2.7

F = whole tree sap flow ($m^3\ s^{-1}$)

A_{sw} = sapwood area (m^2)

This original calibration equation fixed the heating power at 0.2 W. This value was found to not heat the reference temperature probe while providing sufficient sensitivity for the heated probe. Several modifications of the Granier design have been developed, including commercial varieties. A critical review of these designs and the original design indicate that any changes to heat field or geometry of the probes will need new calibration methods [21].

2.1.6 Considerations and limitations of the TDP sap flow method

The most critical component of sap flow calculation is accurate estimation of the maximum voltage (temperature) difference recorded, T_{max}. Practical applications of the TDP method indicate that the zero-flow state can be limited by weather conditions and vegetative growth in the plant [21]. Vapor Pressure Deficit has a strong relationship with sap flow, so this environmental measure is often used to validate sap flow measurements. Installation of the probes can also affect the sensor readings, and loss of thermal equilibrium can cause the values to drift over time [21]. Thus a determination of T_{max} should be made over a 7 – 10 day period and should also be determined separately for each probe [22].

Probes are typically installed on the North side of the tree or vine and insulated with reflective insulation to limit solar radiation interference. Morning peaks in sap flow measurements have been observed when probes are exposed to direct sunlight [21]. Another reason to keep the direction constant is that sap flow is not uniform within a plant's xylem. Therefore one should seek to control sap flow probe's insertion parameters to limit probe location discrepancies. There are correction techniques proposed to deal with extensive drought periods and other complications [21]. The Granier TDP method only measures sap flux density within a small cylinder. There are extrapolation methods to determine total sap flow of the tree (*F*) given the measurement area within the xylem of active flow. A study applying the Granier TDP technique to vineyards found that the active conductive xylem completely encompassed the probes and that even 20 year old grapevines did not develop heartwood [16]. This fact discovered by Braun & Schmidt is also beneficial in determining the sapwood area.

Since there is no heartwood in grapevines, the sapwood area is simply the area of the stem minus the area of the bark.

2.1.7 Justification for a homemade sap flow system

There are many commercial sap flow sensors available with well-documented usage in the literature [39], [40]. The primary reason for constructing homemade probes is cost. The homemade probes can be made for less than $10 per set of sap flow measurement probes [41]. A similar commercial system can be as much as $670 per probe, including the constant heater voltage regulator [41]. It has been found that homemade probes can provide valuable measurements compared with well-known commercial systems, especially considering the possibility that the commercial systems themselves consistently underestimate sap flow [21], [41].

The point at which this research diverges from other published sap flow methodologies is the use of control and logging system for sap flow measurements. Currently there are only a few conference abstracts which describe sap flow measurements using open source microcontroller systems that are easily integrated with sensors, computers and other electronic components, but none sufficiently describe the system, its benefits, and its drawbacks. The described sap flow systems typically use expensive, wired data logging systems which limit the ability of a researcher to collect data that represent field spatial variability. To interface multiple sap flow probes requires a multiplexer and a data logger and one would still be range limited by the length of wire attached to the probes. Accurately measuring small voltages through long sections of wire is another complication of these designs. Thus the advantages of

an Arduino-based system is that it can measure at the same accuracy of a commercial data logging system over a much wider range of the field and for substantially less cost.

The system presented here costs around $1,425. This includes 20 sap flow sensor measurement devices, a base station to collect the data, a cellular modem with service to transmit data to a web server, and all power supplies for the system. A comparable commercial system from Dynamax (Houston, TX) would cost over $12,000. Some could rightly argue that the ease of setting up a commercial data logger with a commercial sap flow sensor far outweigh the price savings from a DIY system, but as a graduate level researcher I feel that the absolute best way to understand the workings of any system is to attempt to emulate how the system functions through your own methods. Following this reasoning, or "learning by doing" I feel I understand much more about the actual methods to determine sap flow than if I were to simply buy a commercial system. As a researcher, it is necessary to be able to critique a system for validity. A DIY system is much easier to assess for validity and fix potential errors than a commercial system with components and measurement techniques that are often guarded to protect intellectual property.

2.1.8 Research goals

One broad goal of this research was to help infuse the use of open source technology into the field researcher's toolkit. Open source means that it is free to use, modify, and share with anyone. While certainly not the first example of open source systems in scientific research, the methods proposed adapt a homemade sensor to an open source control, logging and data transfer system. This research demonstrates the advantages of field research using do-it-yourself instruments and thus encourage others to replicate and expand on my results.

Another broad goal was to decrease the barriers to individuals deploying custom wireless sensor networks for agronomic research.

This research project was associated with the TransAtlantic Precision Agriculture Consortium's (www.vellidis.org/TAPAC) graduate student exchange program between the University of Georgia and the Agricultural University of Athens, Greece. As a result, the field component of this project was conducted in and around Athens, Greece. The project's specific objectives were designed to meet the project's overall goals but to also take advantage of the opportunities and constraints offered by working in Greece. The project's specific objectives were:

1. To develop a system that measures water use by sap flow in vineyards using low cost, open source hardware and software;
2. To model seasonal water intake accurately and to correlate sap flow measurements with other common environmental and plant parameters; and
3. To assess the system's performance and suggest improvements for future system designs.

2.2 **Materials and methods**

2.2.1 Sensor network development

2.2.1.1 Homemade sap flow sensor construction

The Granier TDP (Figure 2.2) construction was explained in detail by Davis et al. (2012b) and also by a publication from the University of California – Los Angeles Center for Embedded Network Sensing [CENS], (2008). The system employed here incorporated aspects of both of these designs as well as new components. The thermocouple used was Type T, copper-constantan, 36 AWG, manufactured by OMEGA Engineering (Stamford, Connecticut, USA). Approximately 12cm of wire was used per probe. To construct the probe, the insulation was stripped from both ends of the wire, and on one end the two conductors were soldered together and trimmed to around 2mm. An 18-gauge standard medical grade needle was used to hold the probe. A notch was filed into the needle 1cm from the base and the needle shaft was trimmed to 2.1cm. The soldered thermocouple junction was then inserted into the needle until it was visible in the notched hole. The junction was then glued into place directly at the notched hole using cyanoacrylate glue. This comprised the construction for a standard (reference) temperature probe. To construct a heated temperature probe, 85 cm of insulated resistance wire with a resistance of 147.4 ohms per meter was wrapped around a probe and attached to a current source. The resistance wire, which was used as heating coil, is called Kanthal D, manufactured by Kanthal Corporation (Hallstahammar, Sweden). The two ends of heating element wire were attached to a low resistance extension wire and then to the constant power source. The constantan (red) wire from the two probes was soldered together

and covered with heat shrink, and the two copper wires attached to the differential inputs on an analog to digital controller.

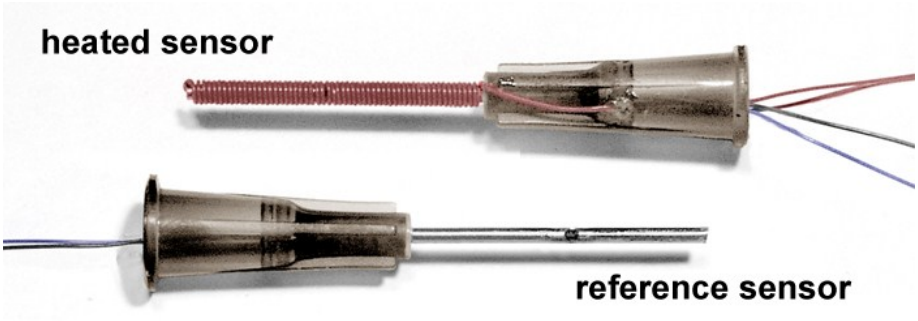


Figure 2.2: The two probe types (CENS, 2008)

Other forms of sap flow measurements which offer both advantages and disadvantages compared to the method just described have been developed. .The Heat Ratio Method (HRM), first proposed by Burgess et al. (2001), works on a heat pulse technique as opposed to a heat balance technique. With this method, a heat pulse is sent out only at certain intervals, therefore a constant power source is not required. The use of both methods in wireless sensor networks was compared in detail by Davis et al. (2012). The important findings of their study was that a 10 second sampling interval was required by the HRM to match the error of the Granier method sampled every 15 minutes. So although the pulsing required less energy than constant heating, pulsing heat and sending data every 10 seconds consumed substantially more energy than a constant heat source with a reduced data sampling rate. In the end both methods gave similar absolute errors.

2.2.1.2 The Arduino platform and the analogue used in this research

The Arduino platform (Arduino, Italy) is an open source microcontroller system that is designed to be easily integrated with sensors, computers and many electronic components. Both the hardware and software of the Arduino platform are open source licensed. The Arduino software language is a variant on the C++ language known as Wiring. An Integrated Development Environment (IDE) is used to write programs that are then compiled and executed on an embedded microcontroller. Many variations and "clones" of the Arduino system have been developed to address specific tasks and to expand compatibility of the platform.

The Arduino clone used in this research is known as the Moteino (Figure 2.3), and is manufactured by LowPowerLabs (LowPowerLabs.com). The benefit of using the Moteino is that it also includes a high-powered 434 Megahertz (MHz) radio transceiver that will be used to wirelessly transmit and receive packets of information containing sap flow measurements and programming instructions. The specific radio transceiver utilized by the Moteino is the RFM69HW, manufactured by HopeRF (Shenzhen, China). This radio frequency was chosen because it will travel farther through obstacles than the higher frequency radios available. Range tests I have performed indicate that the signals reach over 250 meters in a dense forest using the default transmission settings.

The Arduino community is an interconnected development community through web forums and social media. One helpful aspect of an interconnected development community is the collaboration and creation of libraries. Libraries in the Arduino world are a set of functions that can be called in a program to operate specific pieces of hardware or to control aspects of the microcontroller's functionality. The benefits of libraries are numerous, but the primary goal

is to ease program construction and readability. For instance, a function called **blinkLight** is much easier to interpret than the actual commands to turn on and off the light.

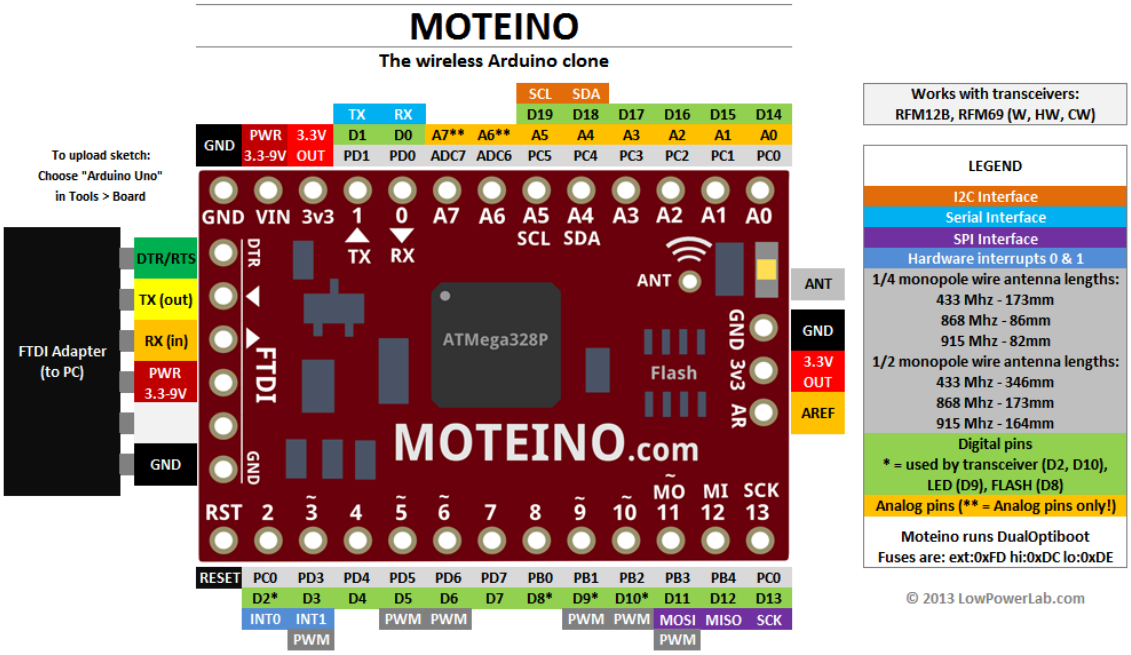


Figure 2.3: Moteino diagram (LowPowerLabs, 2013)

2.2.1.3 Accurate sensing of thermocouple voltages

Much consideration was given to the sensing of the low voltages created by the sap flow thermocouple junctions. An Analog to Digital Converter (ADC) is used to convert an analog signal, in this case a voltage measurement, to a digital representation in the microcontroller system. The standard Analog to Digital Converter (ADC) in an Arduino has a 10 bit resolution, meaning the analog signal can be represented in $2^{10} = 1024$ digital signals (expressed as a positive integer). This resolution is not high enough to discern voltage signals from thermocouples accurately. The Type T thermocouple used in this research varies with 0.041 millivolts (mV) per degree Celsius. The MCP3422 manufactured by Microchip (Chandler,

Arizona, USA) is a low-noise, high accuracy ADC with 2 differential inputs and up to 18 bits (2^{18} = 262144) of resolution. It also has a precision voltage reference of 2.048V and an 8x precision gain amplifier. With this resolution and voltage reference we can obtain divisions as low as 1 microvolts. At this point the accuracy of the sap flow thermocouple system becomes much more limiting than the resolution of the ADC.

The accuracy of the Type T thermocouple wire is +/-1 Degree C. I programmed the system to take numerous measurements and average the values to obtain a more accurate measurement. The MCP3422 chip can be controlled using the Inter-Integrated Circuit (I^2C) bus on the Moteino (pins A4 and A5). The I^2C bus is a method of controlling electronic devices by serial commands, or in the software. The MCP3422 is used in this system because an Arduino library was created and released to the public for free use. One drawback of this chip is that it is surface mount technology. An adapter manufactured by SchmartBoard (Freemont, California) is used to convert the pins to the Moteino's size.

2.2.1.4 Data acquisition and transfer station

The Moteinos used to measure sap flow do not store the data, instead data are sent wirelessly to a base station that contains a SD card to write the data to files and also a GSM cellular module so data can be transferred to a web server. This base station is a Raspberry Pi, a miniature computer that runs a variant of the Debian Linux Operating System known as Raspbian. The Raspberry Pi is connected via the serial terminal to a Moteino that acts as the gateway for all the nodes in the field. The Moteino transfers the incoming transmissions to the Raspberry Pi by simply printing a line of text to the serial terminal. A Python application on the Raspberry Pi listens for the incoming transmissions and writes them to a text file. The

Raspberry Pi contains a software clock, so that the time of day can be attached to the incoming sap flow measurements. The GSM module used in this project connects to the internet using a SIM card to access the cellular networks on either 2G or 3G bands. The Pi accesses the GSM module through a Wi-Fi connection and uploads the text files to a web server to be stored in a database. This ensures that the files are safely secured in at least two locations and allows for real time display of the measurements. Because the base station was constantly acquiring, recording, and transferring sap flow data, its power requirements are far greater than for the Moteino-based sap flow system.

2.2.1.5 System power supply

Each Moteino sap flow measurement system was powered by 3, 1.2 V Nickel Metal Hydride batteries wired in series for a total voltage of 3.6 V and a total amperage of 2.5 A. The batteries used are manufactured by Tenergy (Fremont, California, USA). The capacity of the battery supply was roughly 2500 MilliAmp hours (mAh). This means that if we have a system that draws 2500 mA of current, the batteries will theoretically be depleted in an hour. The solar panels used to recharge the batteries were 1.5 W panels capable of producing 270 mA. The solar panels were purchased from SeeedStudio (Shenzen, China). Considering a reasonable estimate of 250 mA charging the batteries, approximately 10 hours were needed to completely recharge the batteries. The solar panel and battery system were designed cover daily power requirements of the system, while still keeping the system economical. Also, the panel voltage, battery voltage, and system load matched well enough to use just a simple blocking diode to regulate backflow from the panel. This method of charging is known as trickle charging and should ensure the batteries were not supplied with too much current.

The power supply for the base station incorporated a high capacity 12V battery charged by a 16-18V solar panel. A commercial Morningstar (Newtown, Pennsylvania, USA) charge controller was used to control the battery voltage. The Arduino board's voltage regulator does not perform efficiently at 12V input, so a generic step-down 5 V regulator was used to provide a more efficient voltage. The reason for a large capacity battery for the base station was primarily that the GSM module can consume up to 0.5 A of current in heavy use. The Raspberry Pi was also running continuously, another reason for a larger capacity battery.

2.2.1.6 Constant voltage heat source

The Granier TDP requires a coiled heating element that is constantly on. There are variations to this technique that have been successfully adapted, but they require less frequent measurement intervals (Lu et al., 2004). The Granier TDP method requires a heat field of 0.2 W and deviation from this field necessitates recalibration. The system I designed modifies the original components used for the heating probe and also for the constant heating supply, but the heat field itself is not changed. A constant voltage breakout board was used to ensure the heat field remains constant. A 5 V step-up/step-down voltage regulator S7V7F5 manufactured by Pololu (Las Vegas, Nevada, USA) was employed to output 5 V (+5%/-3%) at different input voltage levels. Since this system incorporated rechargeable batteries and solar power supplies, it was necessary to regulate a fluctuating input voltage that can be above and below the 5 V supply. From Ohm's Law we can calculate the current required for the heat source as:

$$V = I R:$$

$$5\,V = .04\,A \times 125\,\Omega$$

Equation 2.8

40 mA of current was theoretically required hourly to power the source. The voltage regulator has an average of 90% efficiency, so assuming that our current requirement was 50mA, this brought our estimated total daily consumption to:

$$50\ mA \times 24\ h\ =\ 1200\ mAh$$

Equation 2.9

This was the largest current sink in the system. Compared to this, the power requirements of the Moteino were negligible. Ensuring that the Moteino was in a sleep mode most of the time, during which it consumed as little as 20 µA per hour, resulted in an estimated daily power consumption of 4.8 mAh. Given a conservative supply estimate of 250 mA per hour from the solar panels, the daily power consumption was theoretically replenished in around 5 hours of optimal daylight. The completed node schematic can be seen in Figure 2.4.

2.2.1.7 Project costs

Table 2.1 presents a detailed breakdown of the essential items and their costs. A few items were already in possession by the University of Georgia and were not purchased specifically for this project, however equivalent versions are included in the table.

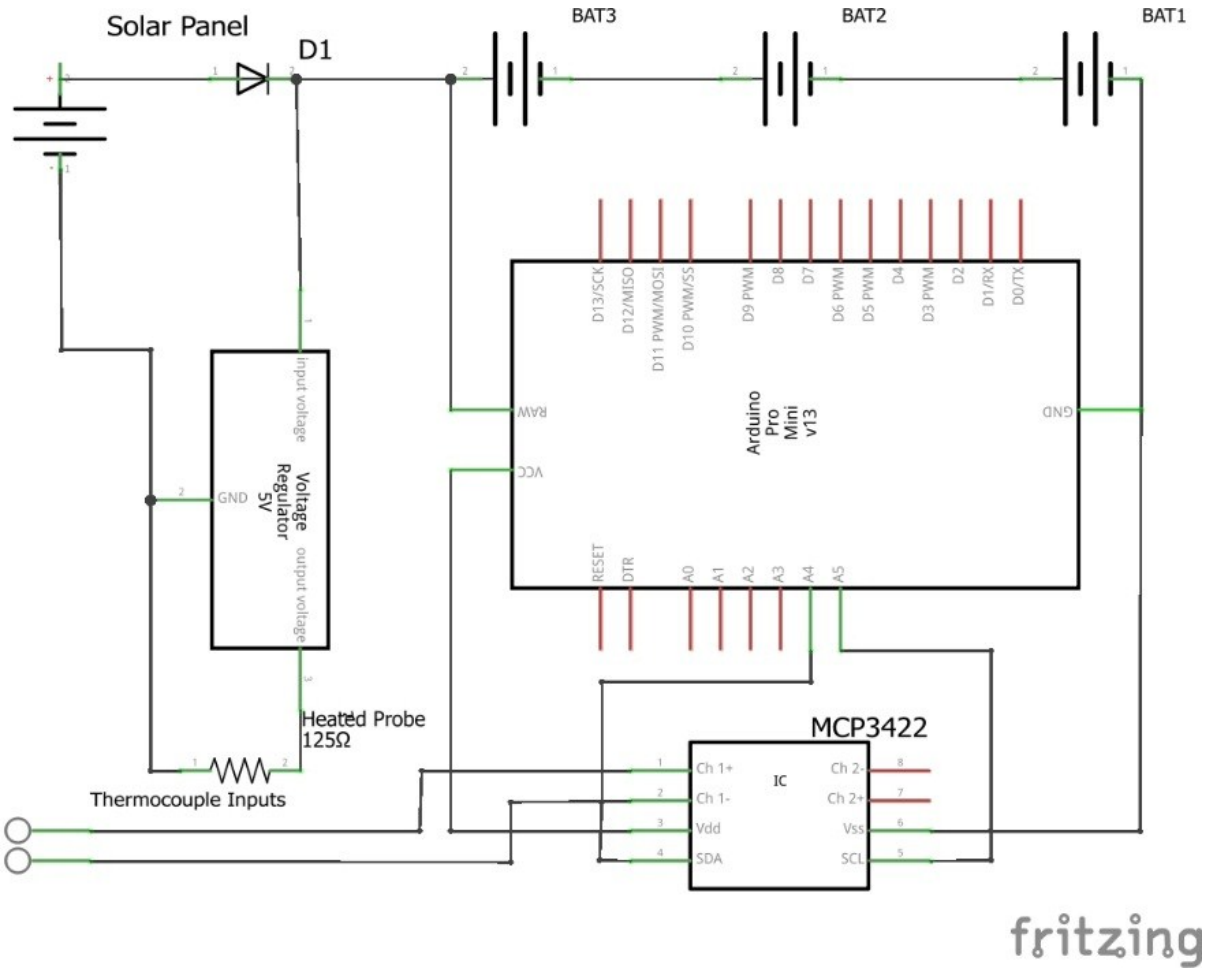


Figure 2.4: Node schematic

Table 2.1: Sap flow network essential parts list

Item	Vendor	Part Number	Price per part	Quantity	Total price
Moteino	LowPowerLabs.com	R4 RFM69HW – 433Mhz	$20.95	20	$419.00
FTDI adapter	LowPowerLabs.com	FTDI-Adapter	$14.50	1	$14.50
MCP3422	Mouser Electronics	579-MCP3422A0-E/SN	$3.28	20	$65.60
SchmartBoard SMT to DIP adapter	Mouser Electronics	872-204-0004-01	$6.85	10	$68.50
3 AA battery holder	Mouser Electronics	534-2487	$1.49	20	$29.80
Type T thermocouple wire, 36 gauge	OMEGA Engineering	TT-T-36	$34.00	1	$34.00
18 gauge steel needle	McMaster-Carr	75165A75A	$11.75	1	$11.75
Kanthal D wire	Ebay, seller kingorchild	251354584134	$4.25	1	$4.25
Step up/down voltage regulator	Pololu	S7V7F5	$4.25	20	$85
1.5 W solar panel	SeeedStudio	POW92152O	$6.99	20	$139.80
Blocking diode	Mouser	512-1N4007	$0.14	20	$2.80
60 AA rechargeable NiMH battery, Tenergy	Amazon	B003YW8IKU	$78.69	1	$78.69
Huawei E5331 21 Mbps Mobile WiFi	Amazon	B00871NPP6	$59.54	1	$59.54
Raspberry Pi Model B+	Amazon	B00LPESRUK	$39.20	1	$39.20
Aleko 30 W Solar Panel	Amazon	B00G4JCV1W	$65.00	1	$65.00
Docooler 10 A Solar Charge Controller	Amazon	B00JKDZVRU	$9.99	1	$9.99
Keedox 12 V to 5 V Step Down Converter	Amazon	B00A71CMDU	$8.99	1	$8.99
Total Cost					**$1136.41**

2.2.2 Field installation and data acquisition

The network was installed in a rainfed vineyard near Athens, Greece in the month of May and June, 2014. The approximate vineyard location is 37° 59' 31.2" N, 23° 51' 19.4" E and is .694 ha. The vineyard exhibits visible, marked soil differences from the high to low elevations although the elevation change is less than 20 m. The vineyard has a north-facing slope. The plants are untrellised, nonirrigated, and experience significant water stress in the summer. The sap flow measurement devices were installed on the North side of the vines and wrapped with thermal insulation to minimize environmental heating effects, Figure 1.1 presents the installed sap flow sensor schematic and Figure 2.4 presents the entire node schematic. The probes were vertically spaced 10 to 15 cm apart. This distance is a compromise between being too close as to alter the temperature of the reference probe and too far as to allow external thermal gradients to overwhelm the measurements [21]. The two probes were inserted radially into the trunk, the base station was installed at a central location so that the wireless transceiver range is roughly equal across the field. The station and sap flow measurement devices were encased in protective PVC housing to ensure they are properly secured from the elements (Figure 2.5). Data were measured at 15 minute intervals from each probe. The probes were programmed to send data at different times, so that the base station only receives one packet of information at a time. The Moteino sends a packet of information until it receives a confirmation of reception from the base station. If collisions occur, the program was structured to send until the packet is received by causing the Moteino to sleep for a random number of seconds and then resend the packet. The chances of collision were small, considering that a packet of sap flow data can be represented in a few bytes and that the default data transfer

rate was around 40 kilobytes per second. The remote cellular connections also presented a possible connection error, so a protocol was developed to ensure files are confirmed once sent. A map illustrating the node placement can be seen in Figure 2.7.

Figure 2.5: Base station (left) and Node (right)

The base station battery, charge controller, Raspberry Pi and GSM module were placed inside a plastic box for protection from the elements. The solar panel outputs were plugged into the solar section of the charge controller, the battery into the battery section, and the input for the 12 V to 5 V step down converter were placed into the load section. The output of the converter was connected to the Raspberry Pi's micro USB power adapter. The Moteino receiving transmissions was powered from a USB port of the Raspberry Pi. The solar panel was placed on the lid of the box and the box lid was secured with rocks. The Moteino was place on

a pole above the lid to ensure signal reception. Figure 2.6 shows a simple schematic for the base station components.

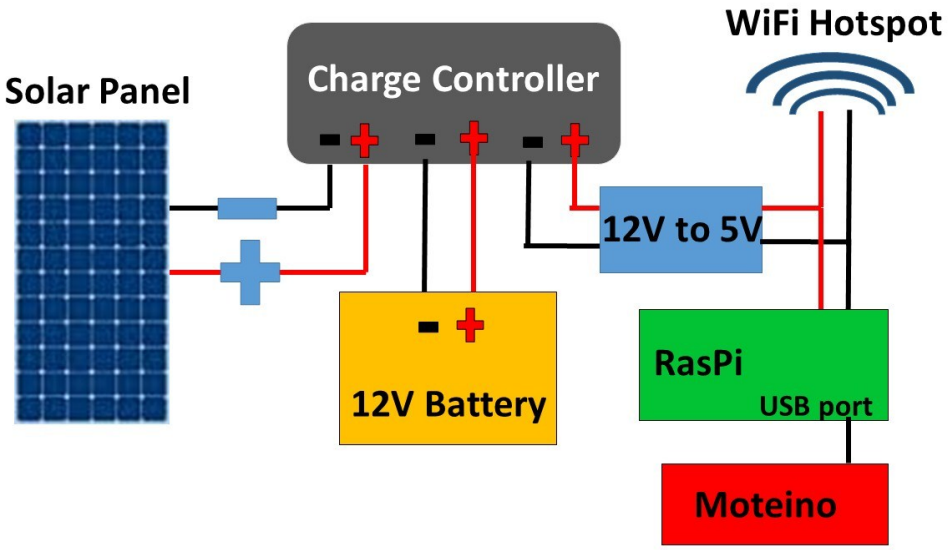


Figure 2.6: Base station schematic

2.2.3 Network performance throughout the season

By design, data were stored only at the base station and not the individual sensors. Sensor network performance was remotely assessed by viewing uploaded data files from the base station. This proved to be a critical component of the system, as the base station was stolen in late July and unable to be replaced. This left over a month of missing data, as the harvest was in early September. In addition, there were multiple issues with the network that we attempted to remedy during the season but with little success. The first and most important issue was the lack of power for the nodes. Despite the careful design calculations, the power system was undersized. In addition, many solar panels were coated with bird excrement, eliminating power for days at a time. Batteries were recharged by a commercial charger in the laboratory and node batteries were replaced and cycled in this manner. Many

batteries also failed, probably due to the high stress of completely discharging and then recharging from an unstable power source (the solar panels). The second problem directly related to the first was the lack of brown out detection. Brown out detection is a safety feature that turns off the microcontroller if the input voltage is below a certain value. This prevents data corruption from insufficient voltages. This was originally turned off to lower power consumption of the microcontroller, but turned out to be a necessary function. Without the brownout detection, nodes corrupted and were unable to perform their tasks. The boards were reprogrammed in the field to allow for brownout detection. Once the brownout detection was restored and the boards with robust batteries were left, the system performed adequately from mid to late June until the base station was stolen in late July. The transceivers performed adequately throughout the season, although a recurring issue at night consumed more battery than necessary. The nodes were programmed to send a packet and wait for confirmation from the base station that their packet was received, this is known as requesting an ack (acknowledgement). It seemed that during the night the nodes either waited an insufficient amount of time for an ack, or the base station Moteino was unable to send an ack, because the message would be resent by the node multiple times, usually exhausting the number of retries programmed. Interestingly the base station had no problem receiving these messages because they are recorded in the dataset, but seemed to have issues responding to acks during the night and for certain nodes. It seems that nodes farther from the base station had the most trouble receiving acks. The antennas used in the project were a simple monopole antenna, or just a piece of copper wire trimmed to a certain length. I have no explanation for why the problems occurred at night.

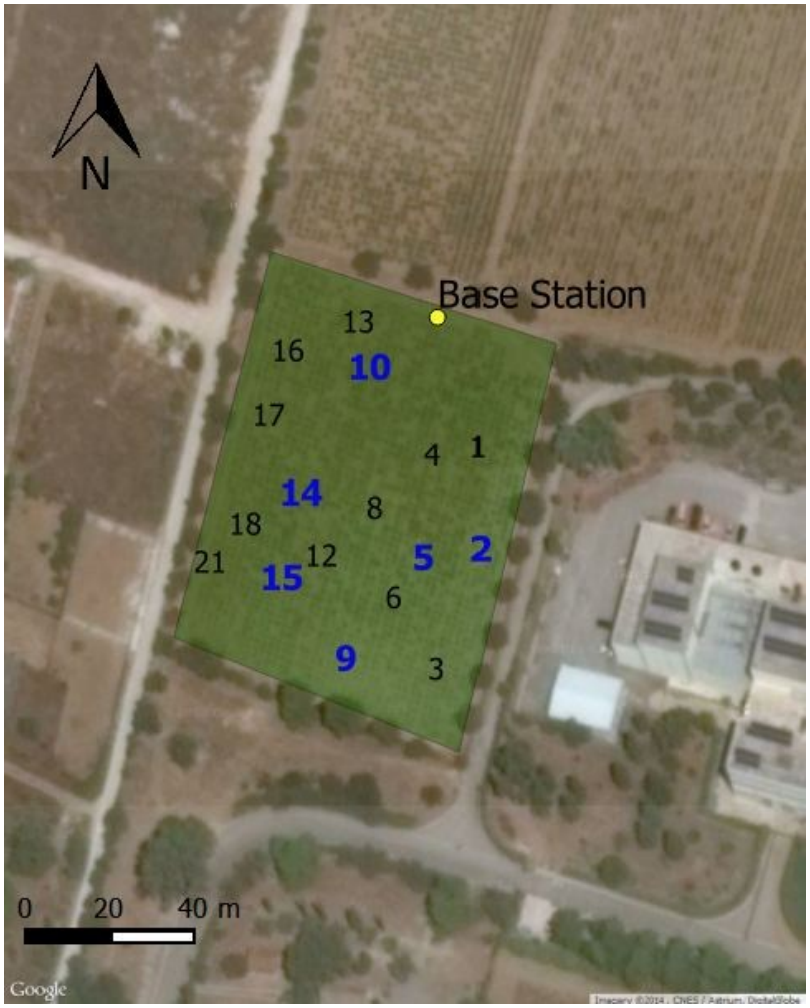


Figure 2.7: Sensor network arrangement, blue denotes nodes used for analysis

2.2.4 Ancillary data collection

I used several other types of data to assess sap flow measurements for validity. The Normalized Difference Vegetation Index (NDVI) is a ratio of the reflectance of two isolated spectrums from a target. The ratio is indicative of the amount of biomass on the plant. NDVI measurements were obtained from each vine on which sensors were installed using the RapidScan portable NDVI meter (Holland Scientific, Lincoln, Nebraska). The RapidScan measures reflectance values at 670 nm, 730 nm and 780 nm. The 670nm and 730nm bands were used for the NDVI calculations. Vines were scanned from approximately 170 cm above

the soil on June 22nd, July 2nd, and July 13th, 2014. Percent Soil Volumetric Water Content (VWC) were obtained at 10 sap flow sensor locations using the Decagon 10HS sensor (Pullman, Washington). The 10HS sensor was powered through one of the Moteino's digital pins. The sensor outputs a voltage that was measured using an analog input pin. The VWC value was calculated from the measured voltage using factory supplied equations. The measurements were taken at the same frequency as the sap flow measurements as described in 0. Midday Stem Water Potential in MPa (MSWP) measurements were obtained on June 25th , July 2nd, and July 13th, 2014 during the growing season using the methods described by Williams and Araujo, 2002. Two leaves of moderate size and shaded by the canopy were selected and covered in black plastic and aluminum foil for roughly one hour. Covering the leaf stops transpiration and it becomes equilibrated with stem water potential. The leaves are then removed and inserted into a pressure chamber and pressurized. The pressure, in MPa, at the moment at which water capsizes the stem column is recorded. Weather data temperature and humidity were acquired from Weather Underground's API (wundergroud.com) from a weather station approximately 500 m from the vineyard. The station number is IATTICAK2. Vapor Pressure Deficit (VPD) was calculated according to FAO guidelines [43].

2.2.5 Description of statistical tests implemented

The Pearson correlation coefficient, R, was used to assess the relationship between sap flow and the ancillary data variables. R is a measure of the strength of the linear relationship between two variables, and is between -1 and 1, inclusive. It is calculated by dividing the covariance between two variables by the product of their standard deviation. The coefficient is used as a precursor to linear regression. Ordinary least squares or linear regression was used to

predict observed values of sap flow from the ancillary data collected. The ancillary measurements were our independent variable and sap flow were our dependent variable. Linear regression is a technique used to predict the best fit linear line through the input points. The best fit line is determined by minimizing the error between predicted and actual observations for the dependent variable. We assume that the relationship between sap flow and our ancillary values is linear. For univariate linear regression, our R^2 value is simply the Pearson correlation coefficient squared. This is an indicator of the amount of variance in the dependent variable that is explained by our model, or by our independent variable. The higher an R^2 value, the stronger the relationship between sap flow and the respective ancillary measurement. To assess the differences between nodes of sap flow measurements, a Kruskal-Wallis test was performed. This test has very few assumptions of the input data set and is a method of determining if the distribution of data is different between groups. We tested for difference between nodes because if the distribution per node is different it means it is likely that there is spatial variability throughout the field in the water use of the vines.

2.3 **Results and discussion**

2.3.1 Description of sap flow data

Nodes 2, 5, 9, 10, 14 and 15 (Figure 2.7) performed well enough for their data to be considered for further analysis. These nodes reported data for at least a 2 week period and throughout the night. Sap flow data was converted into hourly averages corresponding with the beginning of each hour. Because of the variability of sap flow measurements due in part to the insufficient power supply, it was necessary to manually examine the datasets and remove obvious outliers that were the result of ADC errors and times of insufficient power. Points that

drastically differed from the general trend of the dataset were eliminated. Because of reported time lags in sap flow data [16], [21], the interval was moved to the beginning of the hour, meaning that all measurements occurring up to 59 minutes after the start of an hour were averaged together. Values of sap flux density, F_d, recorded in this study were substantially less than published values of grapevine sap flow [16] although those values were not measured in field experiments. The values were collected from a potted plant with insulated soil to simulate using a cover crop to trap moisture in the soil. The test area in this study is not irrigated and the hot and dry summers of Greece put the plants into significant water stress which may account for the lower values we measured. Whole tree sap flow, *F*, was calculated by multiplying the instantaneous flow rate by the sap wood area (Equation 1.7), which was assumed to be the entire stem as found by Braun & Schmidt, 1999. The summary statistics for *F* and F_d are presented in Table 2.2 including summary statistics for all nodes and for each node. 4939 hourly averages were observed overall. The variance between nodes in the count of observed values is from different installation dates of the sensors and also from node failure due to power loss. The data showed marked differences between nodes of observed *F* and F_d rates. The relatively large standard deviation for each node indicates the data are not centered on the mean. The large skewness and kurtosis also confirms this. A histogram plot of *F* for each node included in Figure 2.8 shows that the data exhibits a positive skew, this also evident in Table 2.2. The hourly averaged values of *F* for each node are presented in Figure 2.10. All nodes exhibit the expected diurnal trend with sap flow peaking during the day. Because of the skewed data distribution, maximum daily *F* rates, F_{max} were extracted for each node for comparison with the ancillary data collected as described in 2.2.4. Table 2.3 presents the

summary statistics for F_{max}. We can see that the skewness is much less severe than with the hourly data values. The histograms for each node in Figure 2.9 also confirm this.

Table 2.2: Sap flow summary statistics, in $F_d * 10^{-6}$ and $F * 10^{-3}$

$\mathbf{F_d * 10^{-6}}$							
Node ID	**Count**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**
All	4939	.10	607.9	29.2	54.3	4.3	23.7
2	447	.15	26.1	4.2	4.7	1.7	2.8
5	646	.54	131.9	25.9	21.3	1.4	2.4
9	1271	.11	26.6	9.0	7.0	.57	-1.0
10	839	.10	110.1	13.4	13.3	2.5	9.8
14	1374	.26	607.9	71.7	87.2	2.2	5.8
15	362	.10	51	16.9	12.0	.65	-.53
$\boldsymbol{F * 10^{-3}}$							
All	4939	.03	167.3	10.0	15.3	3.8	20.0
2	447	.08	14.0	2.3	2.5	1.7	2.8
5	646	.28	67.4	13.2	10.9	1.4	2.4
9	1271	.05	12.1	4.1	3.2	.57	-1.0
10	839	.05	55.2	6.7	6.7	2.5	9.8
14	1374	.07	167.3	19.7	24.0	2.2	5.8
15	362	.03	15.0	4.9	3.5	.65	-.53

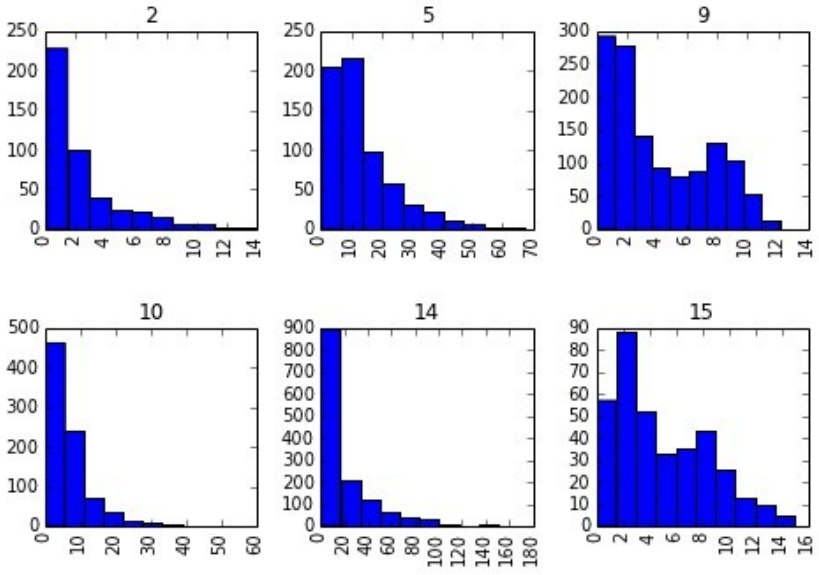


Figure 2.8: Histogram of F for each node

Table 2.3: Summary statistics for F_{max}

Node ID	Count	Min	Max	Mean	Std. Dev.	Skewness	Kurtosis
All	255	0.35	167.31	21.72	24.18	2.69	9.01
2	24	0.35	13.98	7.85	3.27	-0.41	-0.15
5	32	1.51	67.35	31.52	15.55	0.34	-0.40
9	68	0.41	12.11	8.53	2.81	-1.68	2.37
10	38	0.35	55.19	19.89	10.79	0.90	1.72
14	70	0.92	167.31	40.19	35.70	1.40	1.73
15	23	0.45	14.95	8.37	4.20	-0.19	-0.86

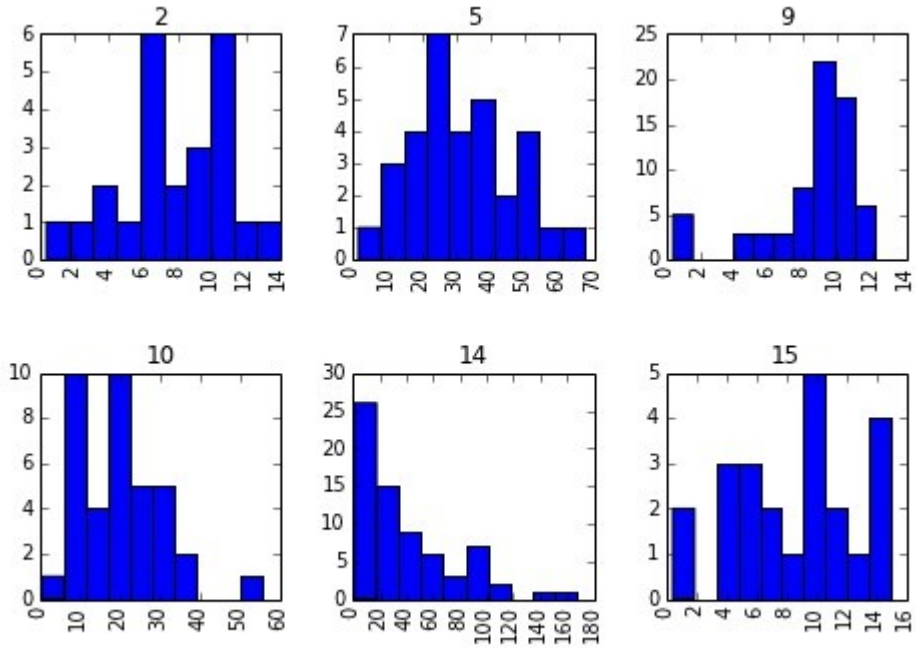


Figure 2.9: Histogram of F_{max} for each node

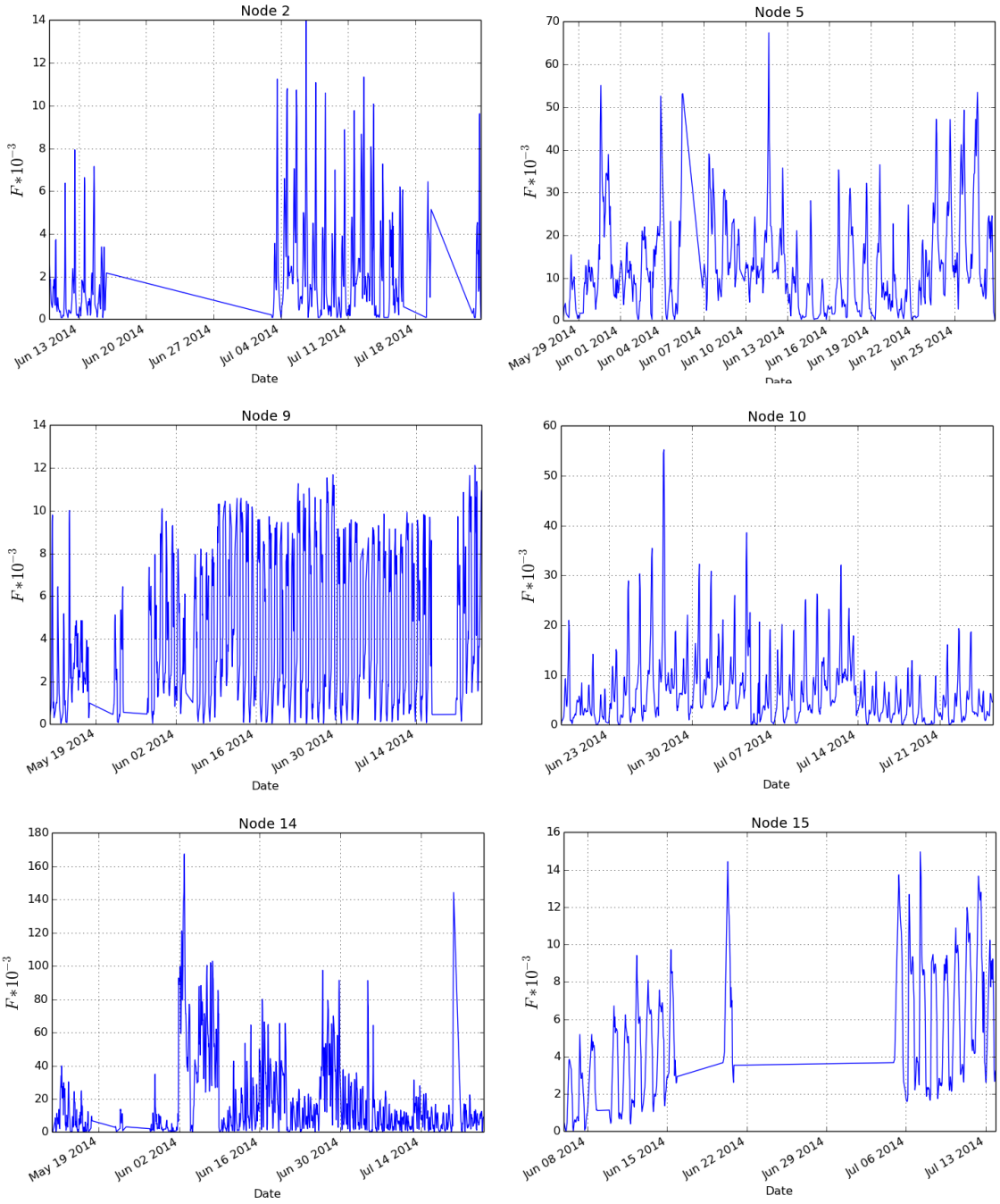


Figure 2.10: Hourly values of F for all nodes

2.3.2 Sap flow data validation

2.3.2.1 Differences between nodes in F rates

A Kruskal-Wallis test was performed on the *F* values for the nodes. The test showed significance at a *p* value of .01. This means that it is highly likely the distribution of *F* is different between nodes throughout the vineyard. This is important because a difference in *F* values is an indicator of spatial variability in sap flow rates, and thus water use throughout the vineyard.

2.3.2.2 Vapor Pressure Deficit

The Pearson correlation coefficients, R, for daily maximum VPD, $VPD_{max,}$ and F_{max}, and for VPD and *F* in Table 2.4. It was also necessary to analyze the dataset for selection of an optimal T_{max} based on the correlation of the selection with VPD. Five nodes (2, 5, 9, 10 and 15) exhibited a moderate positive relationship between VPD and *F* while one node (14) exhibited a weak positive relationship. The daily R values presented in Table 2.4 for certain nodes 2, 10 and 15 are higher than the hourly averages. This confirms prior findings that sap flow values are more suitable for long term water use measurements [16]. A simple linear regression was run on the results to see if maximum daily VPD predicted maximum daily *F*. The regression model has the form of:

$$F_{max} = \alpha + \beta * VPD_{max}$$

Equation 2.10

This produced R^2 values of .32, .27, .52 and .42 for nodes 2, 9, 10 and 15, respectively but nodes 5 and 14 had less than optimal results (Figure 2.12). The regression for nodes 2, 9, 10 and 15 was significant at p = .05. One explanation of the poor results from nodes 5, and 14 is that the sap flow data these aforementioned nodes are too noisy. Interestingly these nodes

were installed nearly a month before the rest of the nodes, it is possible that the stressed vines caused the plant tissue to shrink and the probes lost their thermal contact with the plant stem. Another possibility is that the lack of available water in the soil limited the plant's transpiration.

Table 2.4: Comparison of R values of VPD and F, F_{max}

Node ID	*R* of *F* and VPD	*R* of F_{max} and VPD_{max}
2	.48	.56
5	.50	.31
9	.40	.52
10	.50	.72
14	.19	.06
15	.56	.64

2.3.2.3 Volumetric Water Content

Unfortunately the installed soil moisture probes exhibited strange results, forming a saw-tooth shaped response curve (Figure 2.11). It is most likely an issue with the resolution of the built in ADC of the Moteino or arises from powering the sensors. There was also very little change in the soil moisture readings. The most likely cause of this is that the area where the soil moistures were installed did not contain a significant number of active plant roots. Installation in a more active root zone would lead to an apparent change in soil moisture from plant uptake.

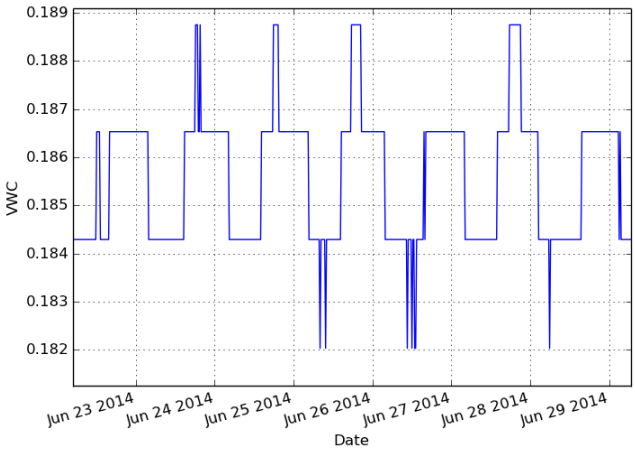


Figure 2.11: Decagon soil moisture sensor response example

2.3.2.4 Midday Stem Water Potential

Midday Stem Water Potential (MSWP) measurements were compared to F_{max}. Values were matched by day the MSWP measurements were made. There were 12 matching records for all nodes, the minimum was -1.45 MPa, the maximum -0.75 MPa and the mean -1.07 MPa. The relationship between Daily maximum *F* and MSWP for all nodes has a weak negative correlation of -.33. We would expect a negative correlation because smaller values indicate more water stressed plants [38]. We should expect a stronger correlation than observed however. The pressure chamber used in this study was a homemade machine that took some experience in correctly reading. The observation of when exactly the stem column breaks is subjective and varies according to the observer. I did all the readings myself, but still feel that the machine takes a great deal of skill to operate correctly. Nevertheless, the negative correlation observed proves that we are obtaining valid measurements for sap flow.

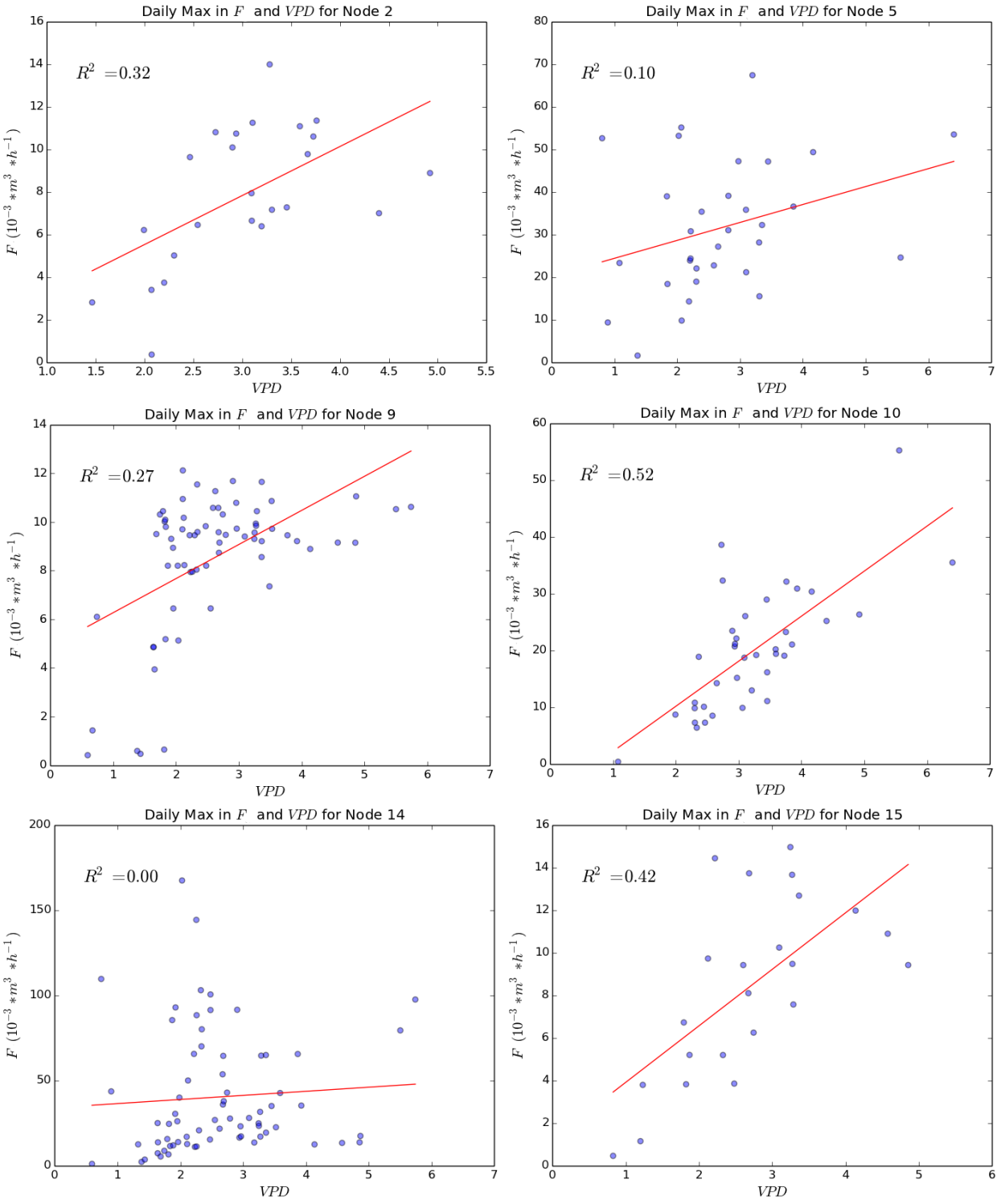


Figure 2.12: Linear regression results for VPD_{max} and F_{max}

2.3.2.5 Normalized Difference Vegetation Index

NDVI measurements were compared to F_{max}. There were 12 matching records for all nodes, the minimum was .67, the maximum was .86 and the mean was .77. The NDVI measurements exhibited a strong positive correlation with $F_{max,}$ with a correlation coefficient of .58. This is a good indicator of sap flow measurement validity because we would expect that plants with greater NDVI values would have a higher biomass and thus have higher transpiration rates. A linear regression for NDVI predicting sap flow was run in the form of:

$$F_{max} = \alpha + \beta * NDVI$$

Equation 2.11

The regression results in indicate that 34% of the variability in *F* is explained by the model as shown in Figure 2.13, the results were significant at p = .05.

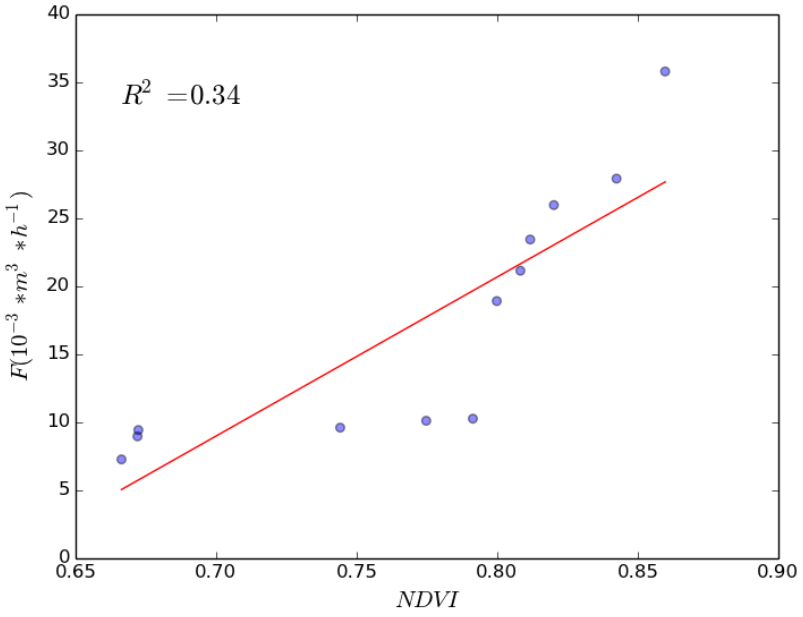


Figure 2.13: Linear regression results of F and NDVI

2.3.3 Suggested improvements to sensor network

The first and foremost improvement to the network is the increasing the power supply of the nodes. The battery capacity should be doubled to 5000 mAh and a 5 W solar panel should be used. Another addition should be a voltage regulator to charge the batteries at a fixed rate. This will help to protect the batteries from voltage spikes from the panel. A revised schematic including a low dropout (LDO) voltage regulator and expanded battery sources can be seen in Figure 2.14. The LDO regulator is an adjustable voltage regulator that is fixed to 4.4 V. The charged NiMH cell should reach 1.4 V, three in series should reach 4.2 V. The next improvement also related to power is to monitor the battery voltage with the Moteino microcontroller. This can be accomplished by adding a voltage divider to the battery positive terminal and measuring the output of the divider at an analog pin on the Moteino. We must divide the voltage because the Moteino cannot handle analog voltages above 3.3 V. These two additions should decrease the chances of power failure and also provide a mechanism to monitor power availability throughout the season. Table 2.5 is a revised essential parts list including the aforementioned suggested improvements.

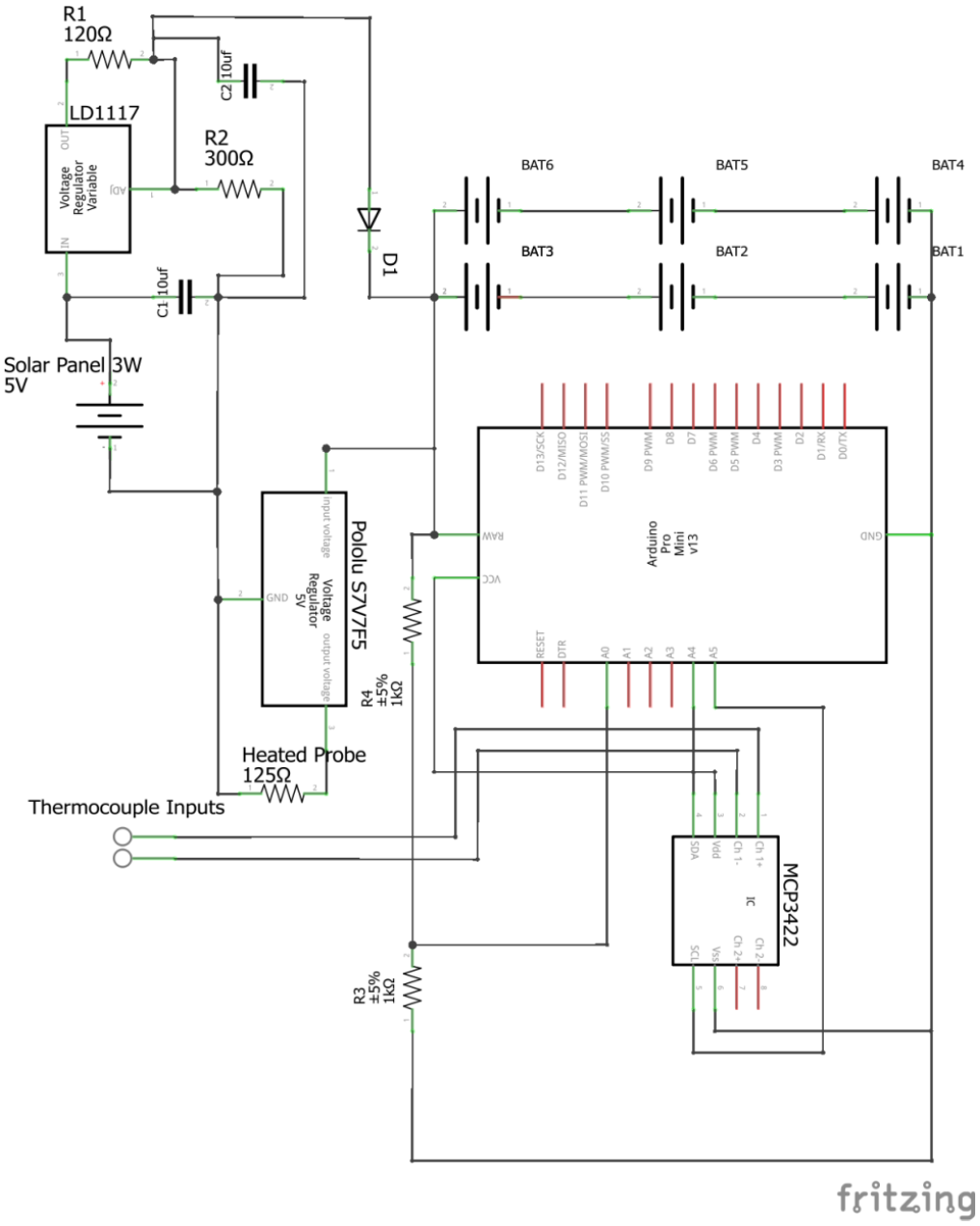


Figure 2.14: Revised node schematic

The Granier TDP is usually used in wooded forest crops. This generally means that the plant would provide some shade over the probe insertion point. Shading the probes limits the effect of solar radiation on the measurements. The probes in this study were shaded with reflective aluminum insulation. The insulation was insufficient because morning peaks in sap flow rates as described elsewhere [21] were still present in the measurements Figure 2.15. It is possible that the bare soil is causing adverse effects on the measurements because it is exposed to solar radiation at greater amounts than the plant column. A future installation should experiment with shading techniques when bare soil is present to minimize morning sap flow peaks.

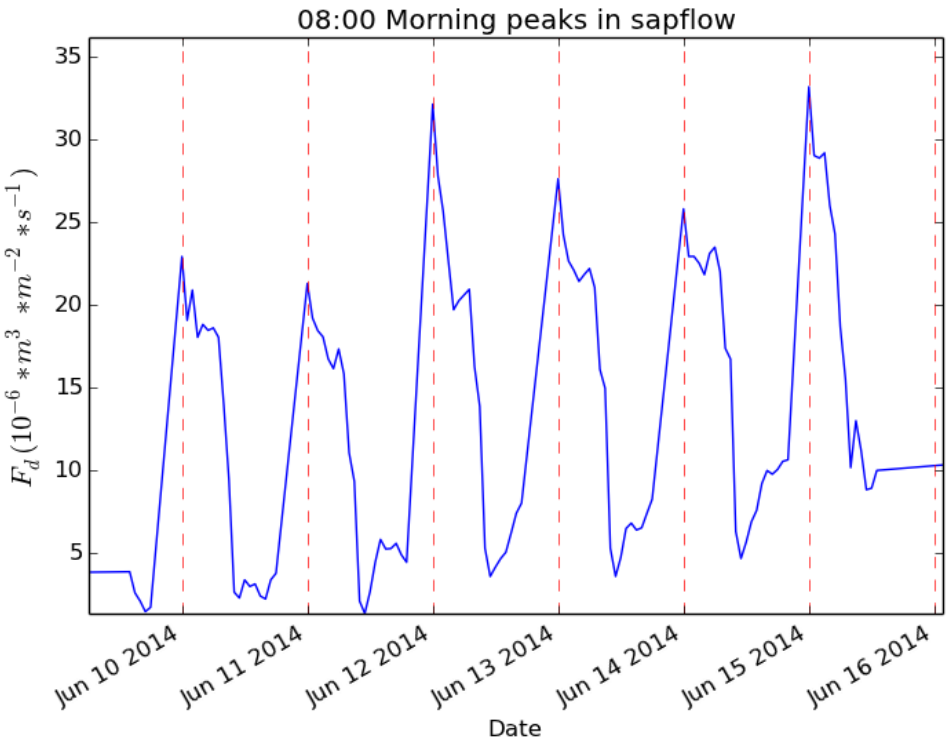


Figure 2.15: Morning peaks in sap flow for node 15, dotted red line indicates 08:00 hours

Table 2.5: Revised essential parts list

Item	Vendor	Part Number	Price per part	Quantity	Total price
Moteino	LowPowerLabs.com	R4 RFM69HW 433 MHz	$20.95	20	$419.00
FTDI adapter	LowPowerLabs.com	FTDI-Adapter	$14.50	1	$14.50
MCP3422	Mouser Electronics	579-MCP3422A0-E/SN	$3.28	20	$65.60
SchmartBoard SMT to DIP adapter	Mouser Electronics	872-204-0004-01	$6.85	10	$68.50
3 AA battery holder	Mouser Electronics	534-2487	$1.49	40	$59.60
Type T thermocouple wire, 36 gauge	OMEGA Engineering	TT-T-36	$34.00	1	$34.00
18 gauge steel needle	McMaster-Carr	75165A75A	$11.75	1	$11.75
Kanthal D wire	Ebay, seller kingorchild	251354584134	$4.25	1	$4.25
Step up/down voltage regulator	Pololu	S7V7F5	$4.25	20	$85
3 W solar panel	SeeedStudio	POW92136O	$11.95	20	$239.00
Blocking diode	Mouser	512-1N4007	$0.14	20	$2.80
60 AA rechargeable NiMH, Tenergy	Amazon	B003YW8IKU	$78.69	2	$157.38
Huawei E5331 21 Mbps Mobile WiFi	Amazon	B00871NPP6	$59.54	1	$59.54
Raspberry Pi Model B+	Amazon	B00LPESRUK	$39.20	1	$39.20
Aleko 30 W Solar Panel	Amazon	B00G4JCV1W	$65.00	1	$65.00
Docooler 10 A Solar Charge Controller	Amazon	B00JKDZVRU	$9.99	1	$9.99

Item	Vendor	Part Number	Price per part	Quantity	Total price
Sealed Lead Acid Battery, 12 V, 35 Ah	Amazon	D5722	$64.99	1	$64.99
Keedox 12 V to 5 V Step Down Converter	Amazon	B00A71CMDU	$8.99	1	$8.99
LD1117 Voltage Regulator	Mouser	511-LD1117V	$0.43	20	$8.60
180 Ω Resistor	Mouser	660-CF1/4C181J	$0.10	20	$2.00
300 Ω Resistor	Mouser	291-300-RC	$0.10	20	$2.00
10 uf Capacitor	Mouser	140-REA100M1EBK0511P	$0.10	40	$4.00
Total Cost					**$1425.69**

2.3.4 Proposed uses for sap flow system

Because even the relatively small vineyard (.694 ha) used in this study exhibited significant differences between *F* throughout the field, the system is ideal for scheduling VRI in different zones. Capturing spatial variability of water use is essential for VRI technology. Given that there is a relationship between *F* and NDVI, zones in the field could be established from NDVI data of the entire field. We can assume that *F* of probes within zones will have less variability than probes between zones and calculate the *F* per zone for irrigation application. The low cost of this system allows users to capture a much higher spatial resolution than with similar commercial systems.

2.4 **Conclusions**

The goal of this system was to obtain accurate sap flow measurements. Although I encountered problems with the instrumentation and the field experiment, I was able to show that a home-made system can be developed inexpensively and produce good results of plant water status in a non-irrigated vineyard. We have also shown that there is variability in these measurements throughout the field. We have observed the expected relationships between sap flow and other indicators of plant water status. Although we expect these measurements to have been more accurate in a more intensely managed vineyard where the vines are healthier and where irrigation is applied, there are many improvements to be made to increase the accuracy of the measurements. This system is important because it reduces the cost of acquiring field scale variability of plant water status. As water resources continue to become a major global issue, it is imperative that PA produce water management technologies that are cost effective and useful to users. This has been accomplished here by adapting an open source

microcontroller platform to monitor the water status of plants in real time and can be easily integrated with existing FMIS. All of the software is documented online so that others can easily replicate the system here. I hope this work will inspire others to share new technologies with the global research community and to realize the importance of open source technology to the future of agriculture.

CHAPTER 3

A USER FRIENDLY MANAGEMENT ZONE DELINEATION TOOL, EZZONE[1]

[1]Lowrance, C. Vellidis, G and Fountas, S.. To be submitted to *Computer and Electronics in Agriculture*.

Abstract

Management zones are a pillar of Precision Agriculture research. Spatial variability is apparent in all fields, and assessing this variability through measurement devices can lead to better management decisions. The use of Geographic Information Systems for agricultural management is common, especially with management zones. Although many algorithms have been produced in research settings, no online software for management zone delineation exists. This research used a common grouping technique based on minimizing the sum of squares between groups to create an open source tool called EZZone for management zone delineation. The tool is accessible by anyone at ezzone.pythonanywhere.com, and allows users to upload their data for delineation. The tool is designed to be user friendly and easily integrate with other GIS software. EZZone was applied and evaluated on data collected from small to large fields, as well as fields incorporating Variable Rate Technology. Five fields were chosen from Georgia, USA, the UK and Greece. The zones were constructed in these fields until the Goodness of Variance fit was above 0.8. The results for each field are discussed and the strengths and weaknesses of the algorithm are examined.

3.1 **Introduction**

3.1.1 Precision Agriculture's importance, drawbacks, and application in this research

Precision agriculture (PA) is a catch-all term for techniques, technologies, and management strategies aimed at addressing within-field variability of parameters that affect crop growth. These parameters may include soil type, soil organic matter, plant nutrient levels, topography, water availability, weed pressure, insect pressure, etc. The application of advanced technologies such as automated sensor networks and GPS-controlled vehicles enhance the ability of producers to make spatially-aware management decisions. The goal of using these smart technologies is to increase profitability by maximizing output (yield) while optimizing inputs (water, fertilizer, labor etc.) by treating the field as a continually varying surface and adapting unique management to these varying zones of the field [1]. The economic benefits of precision farming are well-documented and these techniques have the potential to lessen the environmental impacts of agriculture while increasing profitability [2]. However, the ability of small-scale farmers to adopt these technologies is limited because of the high costs associated with the technology [3] and also because the development of new technology is designed for large scale production [1]. Adoption of PA by small scale producers can be accomplished by focusing on technologies that are less expensive and cognizant of the different management strategies between producers.

One major drawback of current technology is its usability by farmers [4]. Learning how to use new software or hardware is time consuming and often not seen as worthwhile for farmers. The time needed to learn new software or learn how to operate new technology directly competes with essential tasks such as planting. It is known that the complexity of a PA

technology is negatively associated with its adoption; farmers are less likely to invest time and money into technology that is overly complex [4]. Thus new software should be automated in as many methods as possible to save valuable time and should be simple to learn and use. New technology should focus on providing farmers with valuable information at minimal time investment and maximal usability.

A serious limitation of many existing Farm Management Information Systems (FMIS) is their lack of support for geospatially referenced data [5]. Geospatially referenced data are so essential to PA decision making that any software dealing with PA should store data in spatial formats. Another limitation of existing software is that it is too feature rich. Most existing software has a dizzying number of tools that often go unused. The latest innovations in technological integration have the potential to transfer FMIS onto web architectures so that they are accessible across the various devices such as smartphones and tablets which are now commonly used by many farmers and farm advisors. Web based PA software focused on a specific task that can easily integrate with a wide range of software and technology would greatly increase its potential for use by PA practitioners. Thus any new technology developed for management purposes should be web based and accessible across all common devices and it should readily incorporate with other Geographic Information System (GIS) software and common interchangeable data formats.

3.1.2 Management zone use in precision agriculture

A management zone is a portion of a field that is more homogenous than the overall field based on a certain measureable characteristic (e.g. soil type, crop yield) [23]. Management zones (MZs) in PA have been used to increase crop productivity, optimize inputs, and reduce environmental costs. MZs are particularly important when Variable Rate Technology (VRT) is used to vary the rate of agricultural inputs such as fertilizer, lime, or water [24]. The zones can be integrated with application equipment to actively control the amount of an input dispersed. Variable Rate Irrigation (VRI) can incorporate defined management zones to determine irrigation amounts in VRI center pivot technology [25]. When integrated with real time field monitoring, MZs allow farmers to precisely control irrigation amounts throughout their fields.

There are a variety of factors in the field affecting crop yield that can be assessed at detailed spatial resolution with modern technology. It has been shown that soil nutrient levels have a marked effect on crop yield [24]. Because of the large variability in soil qualities, the usage of management zones for nitrogen application has shown to reduce nitrogen usage while not affecting yields [26]. Crop placement can also affect yield, with more eroded areas experiencing less crop yield [27]. While a variety of factors affect crop yield, not all should be included in a management zone delineation tool. The number and size of MZs is currently determined by a wide array of techniques some of which are qualitative while others are based on geostatistics.

It is important to choose proper characteristics to use in MZs classification. The majority of field characteristics used for determination are temporally stable and correlated with yield.

Soil measurements such as electrical conductivity (EC) and fertility are common input parameters. Since soil EC variability also represents variability of other soil properties, it has become a common measure for zone definition [28]. The measurement is easy to obtain and is often correlated with crop yield, giving it a decided advantage over other measurements such as soil chemical and physical analysis and landscape slope, which require time consuming data acquisition [29].

As resources such as water and fertilizer continue to increase in scarcity, it is important to no longer treat the field as a homogenous surface but to address the spatial variability by creating zones of heterogeneous management. The use of management zones is a promising approach to mitigating agricultural inputs while still obtaining profitability [23]. The process of creating management zones can be daunting for anyone unfamiliar with spatial relationships and analysis. There exists currently one commonly used free tool for management zone delineation, a Windows based software known as Management Zone Analyst [30]. This software has drawbacks that will be directly addressed by this research project. The tool produced by this research will be user friendly and as automated as possible. It will also easily incorporate with other GIS software by outputting the zones in common GIS file formats. By reducing the barriers to these advanced management tools, this research will allow PA to benefit society on a larger scale.

3.1.3 Management zone delineation techniques

The basis of all management zone delineation (MZD) techniques is grouping of data values into similar classes so that the members of the class are closer in relation than to members of other classes [31]. Although a variety of techniques have been used or proposed, there is no standard procedure and the success of a procedure is largely determined by the input dataset [32]. One common clustering algorithm is the iterative self-organizing data analysis technique (ISODATA). It is an unsupervised technique that clusters data points by minimizing the Euclidean distance between the points [33]. The requirement of variables with Gaussian distributions makes this technique undesirable for any commonly used approach.

Fuzzy *c*-means (or *k*-means) is a common technique for clustering data that does not require input data to have a Gaussian distribution [34]. The technique seeks to minimize the sum of squared distances from data points within a cluster domain. This is an iterative approach in that the cluster criterion, the sum of squared values within the cluster, is recalculated each time a new data point is added to a cluster. The technique is fuzzy in the sense that there are no defined group boundaries, and data points can belong to more than one group, in varying strength of membership. This process has been widely used in PA, and Management Zone Analyst software mentioned earlier uses this technique [30].

3.1.4 Management Zone Analyst

Management Zone Analyst (MZA) was created by United States Department of Agriculture scientists and the software is therefore in the public domain and available to users at no cost for MZD [30]. The software runs on Windows systems and has a custom Graphical User Interface. It accepts text files as input data sources which can be multivariate. There are a few drawbacks to this software. The first is that it was developed to run with Window NT and not been updated since. This limits its ability to be used on modern platforms. Another disadvantage is that it requires skilled knowledge of statistics to understand and choose the correct model parameters for MZD. The third and most important drawback of this technique is that it does not produce a GIS data file that can be readily used by other software. A user of MZA needs to interpret a variance-covariance matrix and convert a text file to point data, then to a usable form such as a raster or polygon. The conversion of a point data set to a better representative of a field such as a polygon or raster is no simple matter. This often involves an interpolation algorithm to determine values between points based on the known values and the distance from known values. Because the output of MZA is an ordinal classification number i.e. zone 1, zone 2, zone 3, the interpolation to field data is limited to a few techniques. Because of this limitation, point data must first be gridded to a field surface before input into the MZA software [35], [36]. Thus a simplified technique that automated the process of gridding and aligning the point data set to the field that produces zones in common GIS file formats would likely foster much wider use of MZs.

3.1.5 Overview and inspiration of proposed tool

The desire to simplify MZD was obtained after spending many hours in commercial GIS software attempting to devise a useful automated algorithm. The well-known software is very robust in its analysis capabilities, but unfortunately quite expensive. I am of the opinion that publicly funded research should be as committed to free and open access as possible -- thus the abandonment of commercial software and adoption of open source programming languages for a custom tool.

A MZD tool should take into account the desired number of zones from users, i.e., it is useless to create a complex zone map if only two regions are desired. The tool should also conform to the spatial characteristics of the field, such as row spacing, row orientation, and planter width. It is fruitless to create zones that are smaller than the resolution of a farmer's VRT, especially considering that farmers adjust zones that are unnecessarily complex to conform with logistical constraints [4]. The tool should also take into account field shape and in the case of creating irrigation management zones, it should take into account the configuration of the VRI system. It should be noted that by emphasizing usability over statistical accuracy, we do create less "perfect" zones, but at the benefit of creating a tool that is accessible to a much wider audience than the research community.

The MZD delineation tool was developed using the Python programming language (Python.org). Python has a rich history of use in scientific fields, especially in GIS analysis. The backbone of the scientific computation in Python is in its *SciPy* and *NumPy* modules [44]. These two modules allow for complex mathematical modeling and analysis. *NumPy* provides support for multidimensional arrays and has object oriented functions to operate on these arrays. *SciPy*

provides scientific computing functions for Python such as interpolation of point data to a defined grid. There are also numerous GIS modules for the language, with *Shapely* being the most relevant to this project. *Shapely* is a library for geometric operations on objects in the Cartesian coordinate system. Focusing on GIS integration allows the tool's result to seamlessly transfer to a Farm Management Information System and eliminates the extra expertise needed to generate zones using third-party software. Python also has a module for custom web app development called *Flask*. By putting the tool on the web, we ensure open access across all platforms, including mobile browsers, and allow for user input into the MZD. The web GIS client portions of the tool were developed using JavaScript, specifically the OpenLayers (openlayers.org) and jQuery libraries (jQuery.com). The user is able to modify field and zone boundaries, input desired spacing based on technology constraints, and draw lines indicating the row orientation. Creating a semi supervised MZD tool ensures that the user feels involved in the process and that the results are unique to both the user's management philosophy and the technological and physical field constraints. A diagram of the programs used to create EZZone can be seen in *Figure 3.1*.

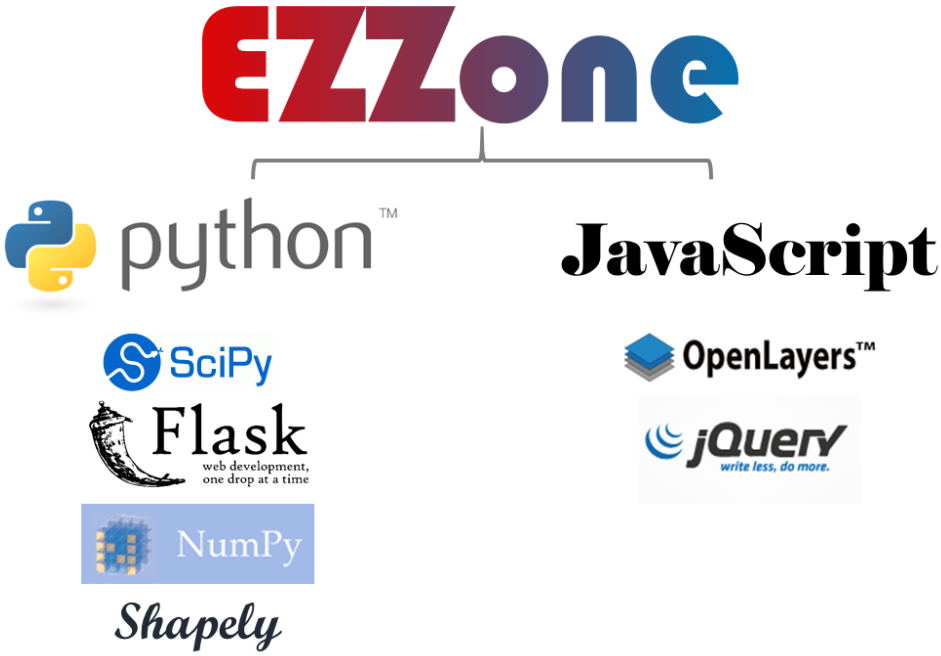


Figure 3.1: The different packages used to create EZZone.

3.1.6 Research goals

The primary goal of this research was to develop an algorithm for MZD that is easy to use, efficient, and useful. The completed software, EZZone, is free to use and available at http://ezzone.pythonanywhere.com. The MZD algorithm used in EZZone was evaluated for its accuracy in classifying field data on 5 separate fields and those results are presented here. I hope this tool will be used by farmers to better understand the spatial relationships in their field and to use their resources more effectively by incorporating the management zones created by the tool into their management practices.

3.2 **Materials and methods**

3.2.1 Overview of MZD algorithm

The MZD algorithm has four major phases. The first phase is the user input screen. The user uploads their *.txt or *.csv comma delimited file that contains the location in longitude and latitude and value they would like to zone (*Figure 3.2*, A). At this phase the user inputs characteristics about their field according to the field type. The next phase takes the user to a screen with a computed boundary of their field based on the input dataset (*Figure 3.2*, B). The user adjusts this boundary and the boundary is used in the later phases of the algorithm. Depending on the field type, the user adds attributes to the map that represent characteristics of the field. After the boundary layer is created, the point data are interpolated to create a surface of the entire field (*Figure 3.2*, C). A polygon grid is created based on the field type and input parameters from the user (*Figure 3.2*, D). The polygon grid values are calculated from the created raster surface (*Figure 3.2*, E). The grids are then classified based on their values and an optimal number of classification zones is determined based on the Goodness of Variance Fit. The third phase shows the user a graphic of the recommended zones. The user selects the number of zones they want to use and the geometries are merged according to class (*Figure 3.2*, F). The completed classified polygons are shown to the user in the fourth phase. The user has the opportunity to adjust the zones based on their knowledge of the field. The user can then download the finished zones in both GeoJSON and ESRI Shapefile formats.

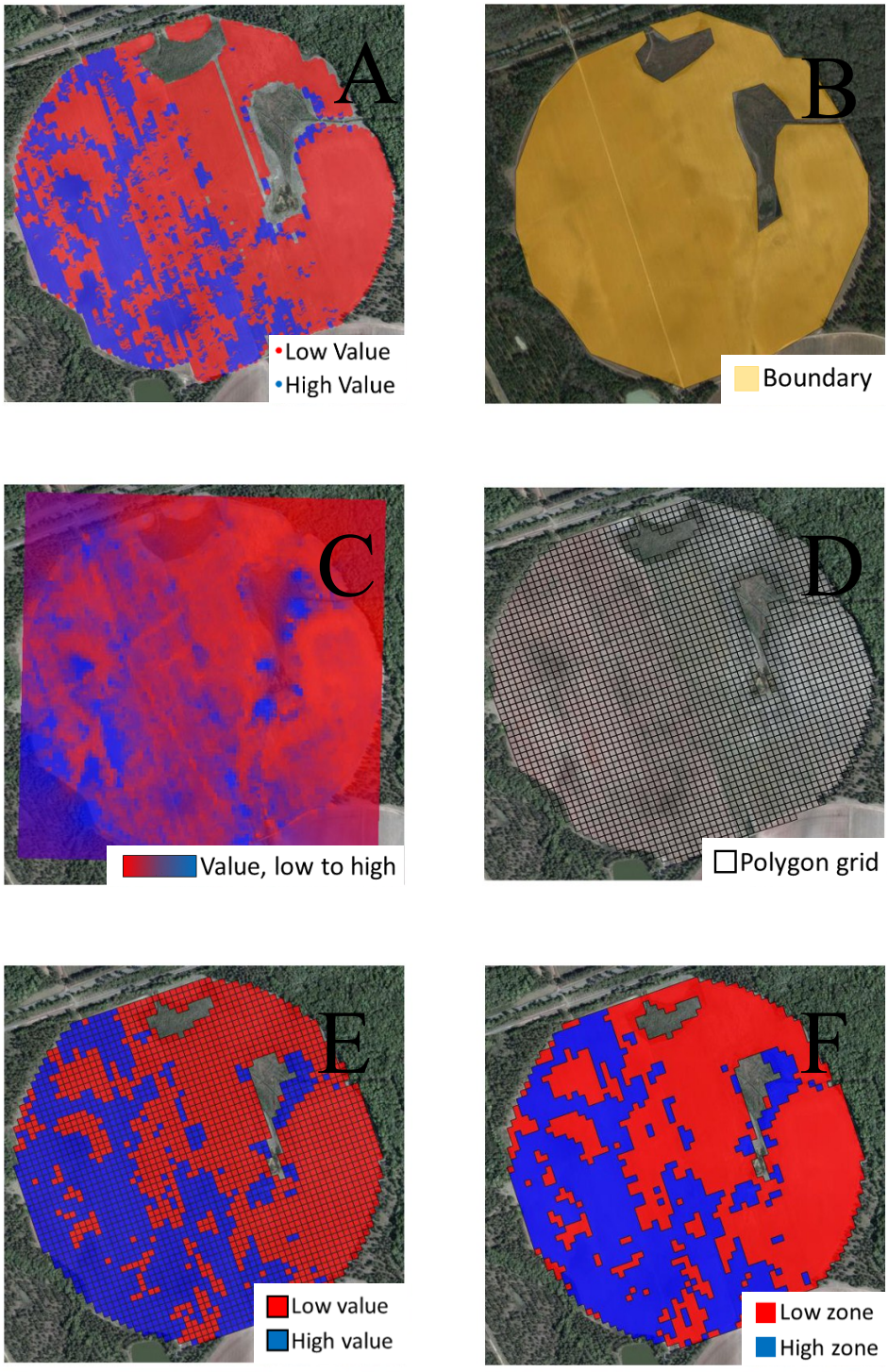


Figure 3.2: Illustrating the MZD algorithm process, the colors red and blue represent the two different zones. A: input dataset, B: field boundary, C: raster created from input, D: polygons created from field dimensions, E: polygon values extracted from raster, F: classified polygons and merged geometries.

3.2.2 Server and client operations

EZZone uses the Python web framework *Flask* to serve a collection of functions that together form the application. Nearly all of the computation is done server-side, or in other words behind the scenes in a web browser. The primary client side application is known as *OpenLayers.* It is a JavaScript library used for GIS interaction and allows for easy manipulation of GIS data and can render data on satellite overlays of the Earth. *OpenLayers* is used to gather input from the user and to render the user's data, Python libraries *SciPy, NumPy* and *Shapely* are used to perform the MZD and *Flask* is used to serve the results of the computations back to the client. *Flask* can be thought of as the bridge that the data crosses between the user and the MZD algorithms.

3.2.3 Inputting data

When navigating to the site's home page, the user is greeted with a description about the project, instructions on how to use the tool and a form to upload their data to and fill out their data's characteristics (*Figure 3.3*). The first step is to upload data for MZD. This data must be either a text file or a comma separate value file in X, Y, Z format. X is longitude, Y is latitude and Z is the value to zone. The next choice is field type. Three field types are supported by EZZone, *Rectangular*, *Circular with Variable Rate Irrigation (VRI)*, and *Irregular and Circular without VRI*. The user can choose the number of zones desired, from 2 to 9, or can have the zones optimized by the algorithm. When a field type of *Rectangular* or *Irregular and Circular without VRI* is selected, the user fills out the *Row Spacing in Meters* and has the option of creating smoother output by adjusting the *Smoothing Factor*. When *Circular with VRI* is

selected, the user fills out the *Spacing of Span in Meters*, which is the length of the VRI span that is individually controllable; the user also fills out *Length of Pivot in Meters* parameter.

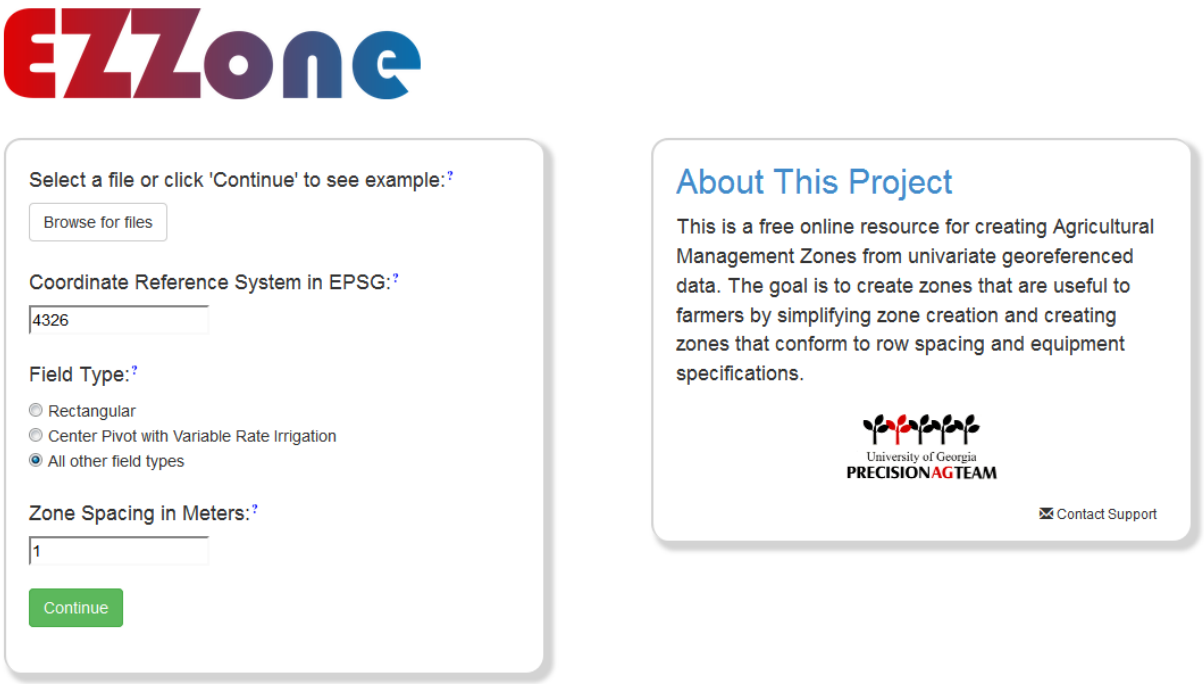


Figure 3.3: The EZZone index page.

3.2.4 Creating a field boundary and adding spatial parameters

The next step in the procedure is to take the user's input data and create a polygon representing the field boundary from the points (*Figure 3.2*, B). This is accomplished by creating a *Shapely* MultiPoint feature from all of the points and then using the convex hull of the MultiPoint feature as coordinates for a new polygon. The polygon is rendered to the user and the user can adjust the vertices of the polygon to their liking. For rectangular fields, no further input is necessary; the user simply adjusts the 4 rectangle vertices. For circular fields with VRI, the user places a point representing the location of the center pivot. For irregular and circular without VRI fields, the user draws a line indicating the row orientation of their field if

they wish the zones to conform to the row orientation. *Figure 3.4* illustrates the user input options.

Figure 3.4: Illustrating adding a line (left) and a point (right) so that the created zones conform to field type and technology.

3.2.5 Raster creation

A raster is a two dimensional array of equally spaced dimensions that represents a variable value. The cell size of a raster is the same across all cells; therefore we can determine the coordinates from the structure of the array holding the values. The resolution of a raster is the width of its cell. A raster with a resolution of 2 m or 0.5 m is created from the user's data depending on the field size. To create the raster we need to know the minimum and maximum X and Y dimensions of the field. This is derived from the boundary of the field created before. A grid is created by iterating through the X and Y dimensions and adding points at the resolution interval. Once our grid is created, we can then interpolate our Z values onto the grid at the predefined grid points. There are two main steps in the interpolation procedure. The

first is to take our X and Y data file points and to perform a Delaunay triangulation. Delaunay triangulation for a set of points constructs triangles so that no point falls into another's circle circumference (*Figure 3.5*).

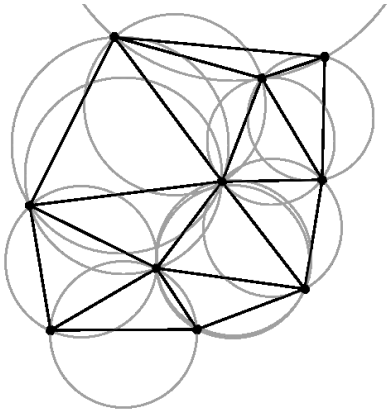


Figure 3.5: A visualization of the Delaunay triangulation (from Wikipedia).

The algorithm then performs linear barycentric interpolation on the Delaunay triangulations to calculate the value of Z at each grid point. The linear barycentric interpolation method is a weighted average approach to determining a value at a given point. From our Delaunay triangulation, we have three vertexes of a triangle and three values at each vertex (x1, x2, x3). If we want to calculate the value of a point (p) that falls within the triangle, i.e. a point in our grid, we simply take the weighted average from the area of the three vertexes (A1, A2, A3) based on the position of the point in the triangle. See *Figure 3.6* for an illustration. We can then extrapolate these values outside the range of points by using the nearest neighbor algorithm, which simply returns the known value closest to the grid cell point in question.

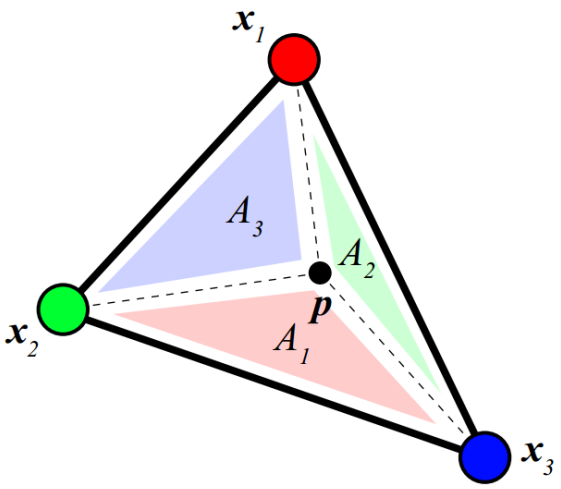


Figure 3.6: A visualization of the linear barycentric interpolation method (from UCSC).

3.2.6 Polygon grid creation

3.2.6.1 Rectangular field

Polygons are created that cover the boundary of the field and are based on the field type and dimensions. For rectangular field types, polygons are created based on the length, width and angle of the field boundary and the row spacing of the field. The grid is always created from north to south, west to east. First our 4 polygon vertices are ordered based on their Y coordinates. Next the top and left side line is determined by finding the difference in X (ΔX) of the maximum Y coordinate pair and the second to maximum Y coordinate pair. Given ΔX we can determine the three points that make up the top line and the left line of the polygon. We calculate the number of polygon cells by dividing the top line and left line length by the user's row spacing. The angle of rotation for the polygon grid points is calculated from the angle of the field. The first vertex of the first gridded polygon is the top left point. We then iterate through the field dimensions and use the computed values to create the rest of the polygons. *Figure 3.7* illustrates the process and the result.

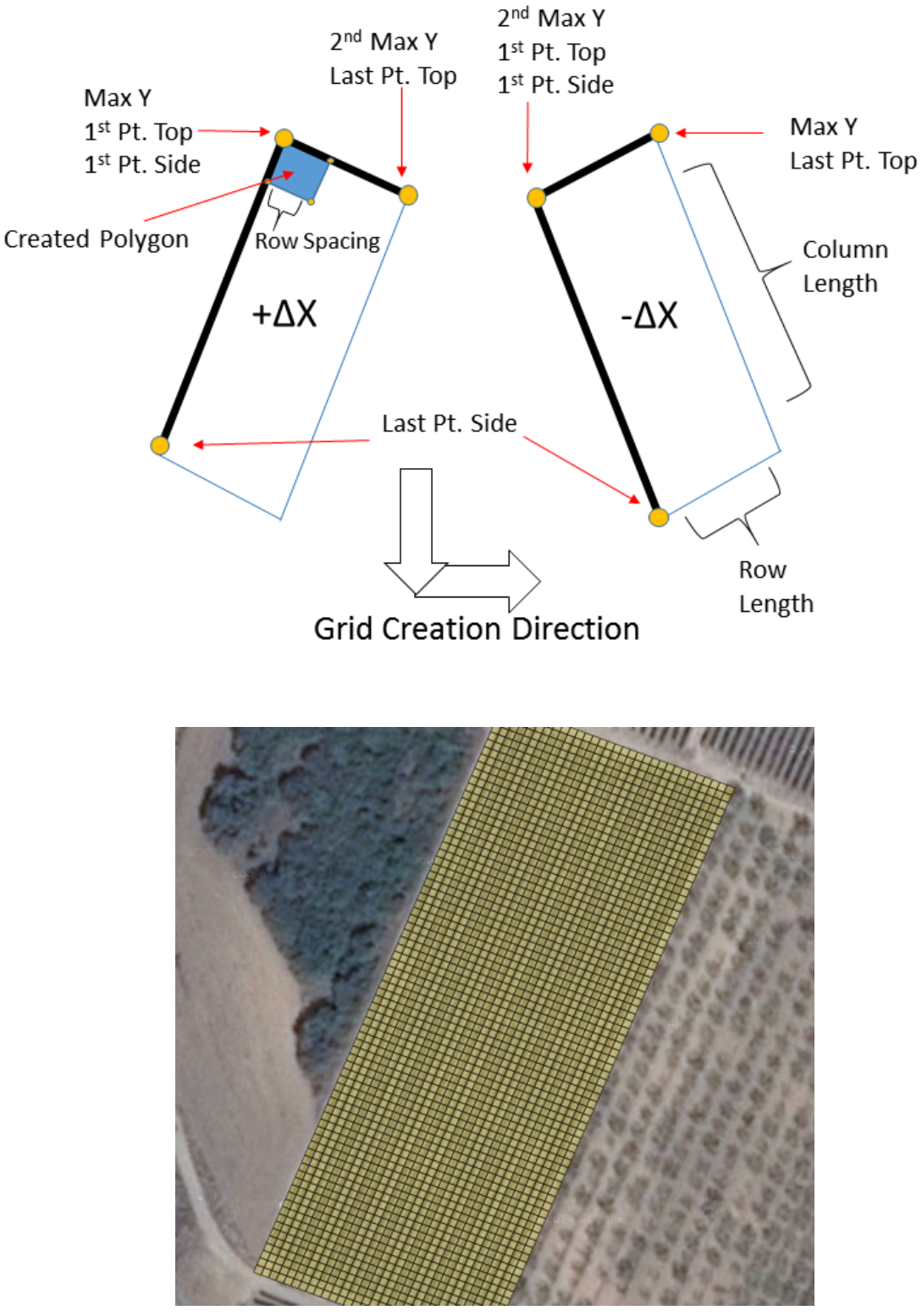


Figure 3.7: The process of rectangular grid creation (top) and result (bottom).

3.2.6.2 Circular field with VRI

For circular fields with VRI, triangular polygons that look like pie slices are created that together form a circular shape corresponding with the VRI sprinkler control zone for the field. Starting from the center, polygon vertices are created going towards the edge in a clockwise orientation using the polar coordinate system. The vertices are determined by the sin and cosine of the x and y angles, respectively, and the distance between each vertex. There are 90 pie slices whose divisions are determined by the spacing of the VRI sections and whose length is determined by the length of the pivot. Once the edge is reached, the process repeats itself but the pie slice is rotated 4° clockwise. *Figure 3.8* illustrates this process and result.

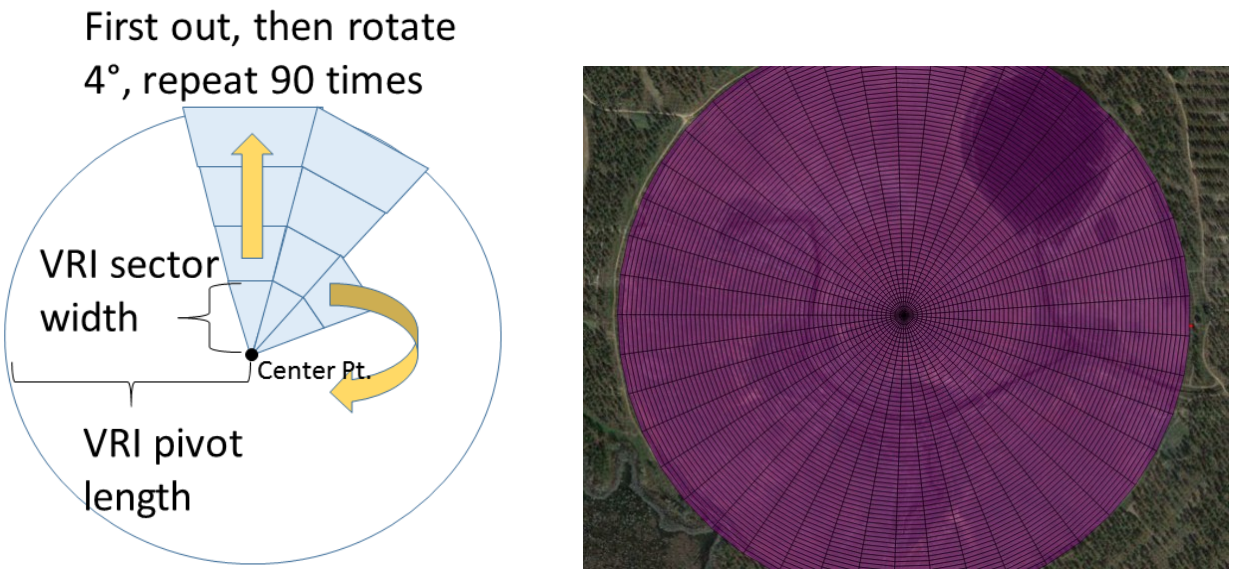


Figure 3.8: The process of circular grid creation (left) and result (right).

3.2.6.3 Irregular field and Circular field without VRI

For irregular fields, polygons are created similarly to a rectangular field except the point of rotation for the grid is based on the centroid of the field and the angle of rotation is based on

the row orientation. If there is no row line drawn then the grid is not rotated. The grid is also expanded +/- 200 m of the field boundary bounding box. *Figure 3.9* illustrates the process and result.

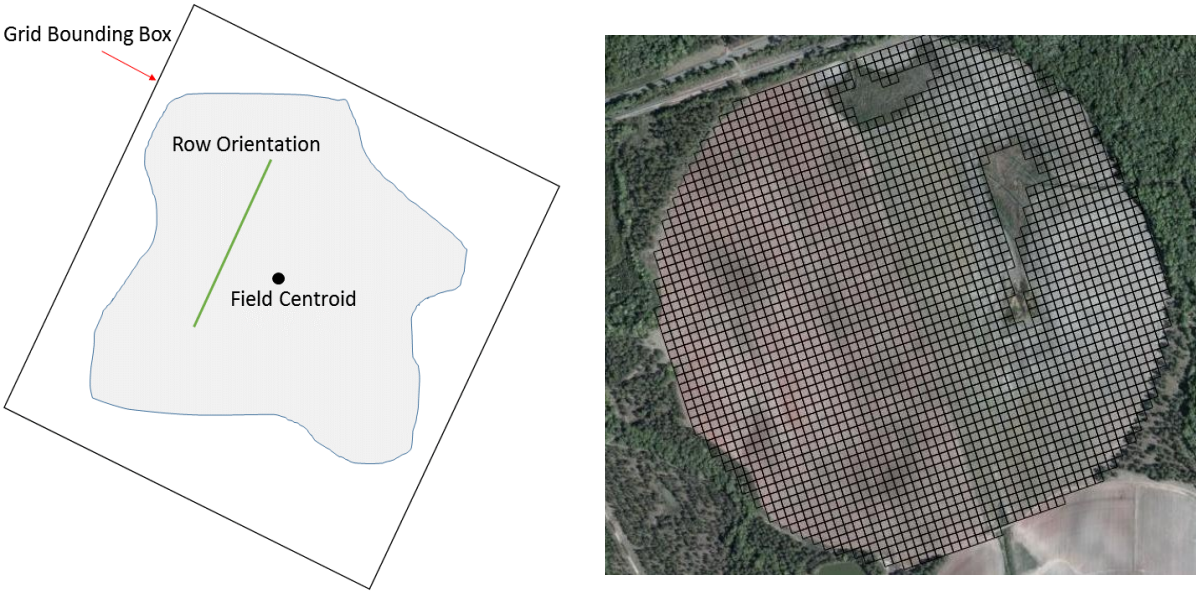


Figure 3.9: The process of irregular grid creation (left) and result (right).

3.2.7 Clustering of polygon attributes

The clustering method employed here is a univariate method of grouping based on minimizing the member deviation from the class mean while maximizing class mean deviation from other classes. It is commonly referred to as Jenks natural breaks optimization and is a well-known technique for statistical mapping (Jenks, 1967). It can also be thought of as a univariate *k*-means, the aforementioned clustering technique used in MZA (Dent et al., 2008). The data is first divided into arbitrary groups based on the number of classes desired and the sum of squared deviations between classes (SDBC) is calculated.

Recall that the sum of squared deviations (SSD) is:

$$SSD = \sum (y - \bar{y})^2$$

Equation 12: Sum of squared deviations (SSD)

Where:

y = array value

$\bar{y}$ = array mean

SSD = sum of squared deviations

The SDBC is the sum of the SSD for each class. Next the sum of squared deviations for the array mean (SDAM), i.e. the input array, is calculated by taking the SSD of the array. Note that this value will not change throughout the iterations. Subtracting SDAM by the SDBC for a class gives us the sum of squared deviations for the class mean (SDCM). Then an item from the class with the highest SDCM is moved to the lowest SDCM, and the SDBC are recalculated. This process is repeated until the Goodness of Variance Fit (GVF) is maximized. The GVF is the difference between SDAM and SDCM, divided by SDAM.

The Jenks natural breaks method is performed on the polygon values extracted from the raster surface. The values are extracted from the raster using the *rasterstats* module in python. The final step is to merge polygons that share the same management zone class, simplifying our geometries. The results of EZZone produce zones ranging from low to high values of the input dataset. Results presented here are always color coded blue to red, with blue indicating low values and red indicating high values.

3.2.8 Smoothing zones for usability

Oftentimes the results of MZD are valid statistically but not applicable to management situations because of their over resolution. EZZone has an option to smooth the zones created so that any created polygons with an area below a threshold are joined to the next group. This threshold is determined by the squared area of the row spacing multiplied by the smoothing parameter. The logic for the procedure can be seen in *Figure 3.10*. The smoothing parameter has a selected value of 1 to 10. *Figure 3.11* presents a smoothing factor of 5 for the zones presented in *Figure 3.2*. Currently there is no method of assessing the GVF of the smoothed zones, although it is assumed that the value decreases.

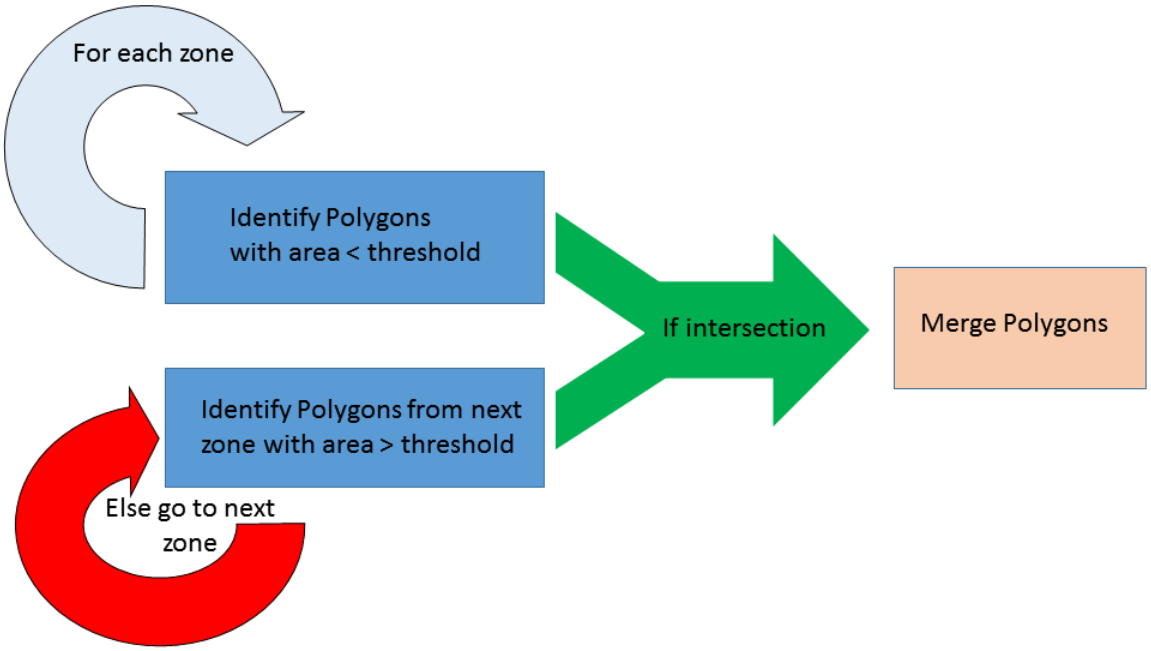


Figure 3.10: A flowchart depicting the smoothing logic

Figure 3.11: Smoothed zones, using a smoothing parameter of 5.

3.2.9 Description of test datasets

My intent was to use spatially explicit data like soil electrical conductivity from the vineyard near Athens, Greece described in Chapter 2 as well as other fields to evaluate EZZone. Spatially explicit datasets from the vineyard were problematic so all testing was done with data contributed by partner researchers. EZZone was tested on all the aforementioned field types and on two different data types -- yield and EC. The field referencing is based on the variable type, either *ec* or *yield*, and abbreviated field type, *Rect, Circ,* and Irreg for Rectangular, Circular with VRI and Irregular fields. Because there are two irregular fields with yield data zoned, the reference name for one of these fields has a 2 added to the end. The VRI field zoning procedure was tested on the EC values of a field in Calhoun county in southwestern Georgia

located at approximately -84.55669, 31.46651 (EPSG:4326). This field is referenced as *ecCirc* (*Figure 3.12*, A). The rectangular field zoning procedure was tested on EC values of a vineyard outside of Volos, Greece located at approximately 22.73895, 39.26553 and referenced as *ecRect* (*Figure 3.12*, B). The irregular zoning procedure was tested on three different fields. The first field is located in Miller County in southwestern Georgia at approximately -84.75147, 31.17223 and EC values were used for zoning; this field is referenced as *ecIrreg* (*Figure 3.12*, C). The next field is located in Tift county in southwestern Georgia at approximately -83.55373, 31.51142 and cotton yield values were used; the field is referenced as *yieldIrreg* (*Figure 3.12*, D). The final field is located in the United Kingdom at approximately -0.45902, 52.07069. Yield data was used for this field and it is referenced as *yieldIrreg2* (*Figure 3.12*, E). The summary statistics for the fields' Z values and area are presented in Table 3.1.

Table 3.1: Summary statistics for test fields

Field	*Count*	*Mean*	*Std Dev*	*Min*	*Max*	*Field Area (ha)*
ecCirc	22385	9.9	5.4	0.3	56.6	85.3
ecRect	12676	73.3	20.0	8.8	114.0	1.0
ecIrreg	6298	7.8	4.2	0.4	31.3	28.8
yieldIrreg	8602	2.6	1.3	0.0	8.3	9.2
yieldIrreg2	5052	8.2	0.9	4.2	11.2	16.3

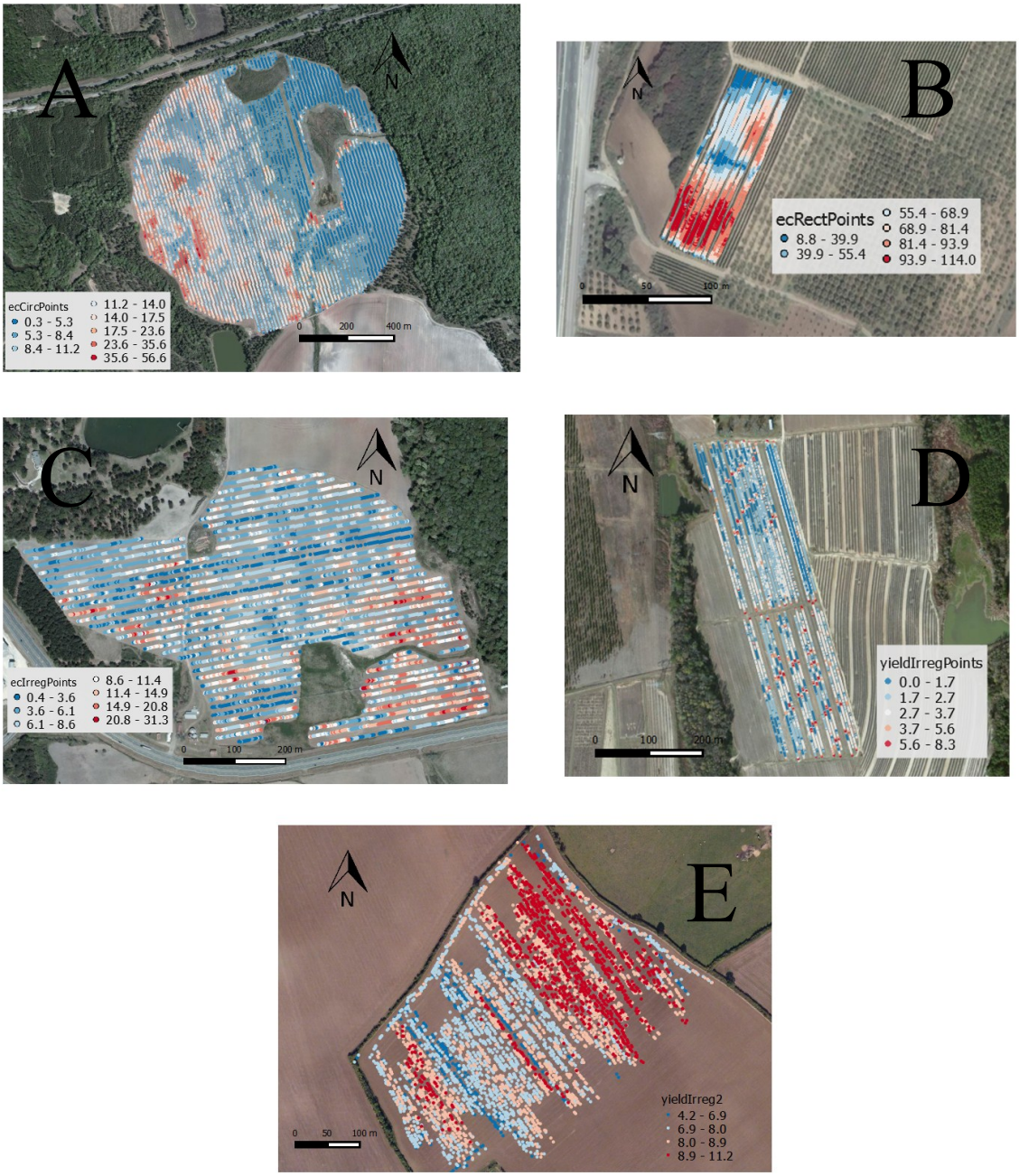


Figure 3.12: Test fields' data points. A: ecCirc, B: ecRect, C: ecIrreg, D: yieldIrreg, E: yieldIrreg2

3.3 **Results and discussion**

3.3.1 Description of test procedure

EZZone was tested on each of the 5 test fields by classifying the fields until the GVF was greater than or equal to 0.8. The GVF is measured between 0 and 1 unless the classification procedure is inaccurate, in which case it will be a negative number. Because the GVF is a measurement of the homogeneity within the classes, it represents the accuracy of our classification [45]. It should be noted that 0.8 is an arbitrary cut off point but I feel it represents a compromise between classification optimization and a usable number of zones. EZZone allows the user to select the number of zones and recommends a number of zones based on the GVF (*Figure 3.13*). The zones presented here are always represented in a blue to red color scheme, blue representing low values and red representing high values.

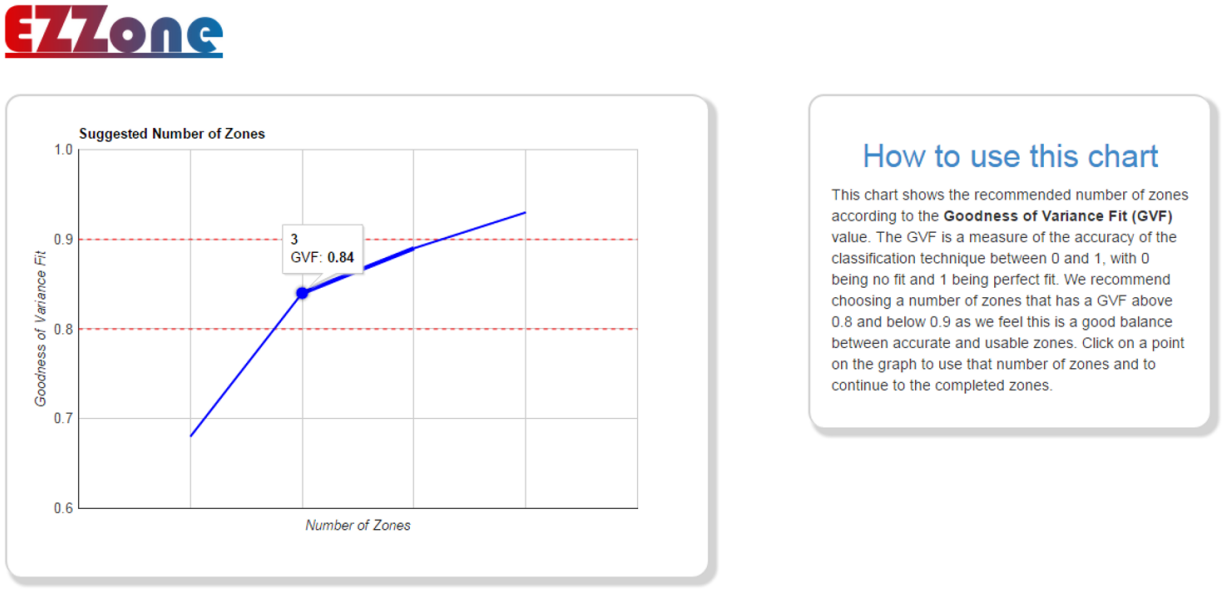
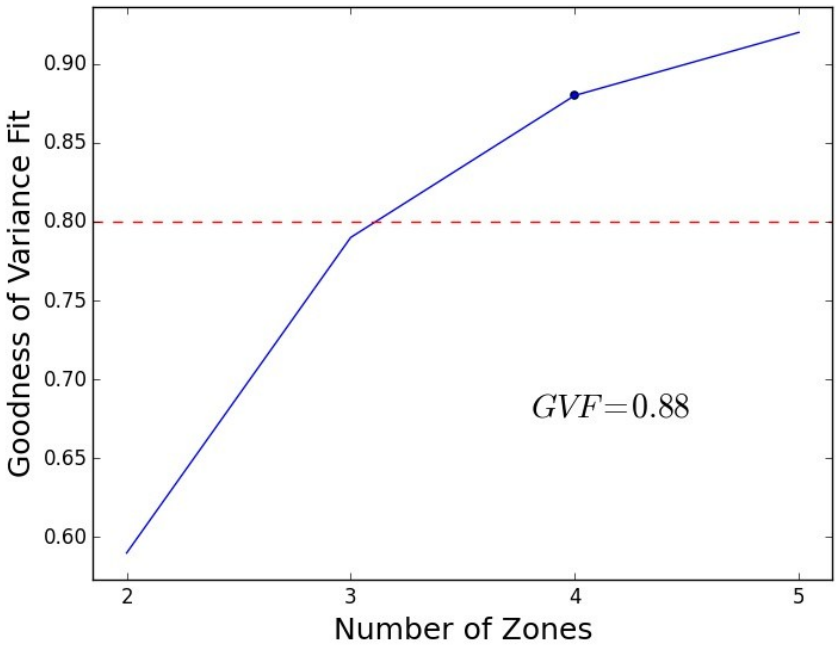


Figure 3.13: The GVF selection screen of EZZone.

3.3.2 ecCirc classification

The ecCirc field is a central pivot field that was classified using the algorithm for VRI fields. The field exhibits marked homogeneity in the Eastern half but marked heterogeneity in the Eastern half, as evident in *Figure 3.12*, A. The field boundary was adjusted to account for large areas of the pivot span that are not planted with crops. The GVF was optimized at 4 zones with a value of .88. The resultant zones and GVF optimization can been seen in *Figure 3.14*. We can see that the zones are also largely homogeneous in the Eastern half and heterogeneous in the Western half. This indicates that the spatial variability of EC throughout the field is not static and the Western half of the field likely has very different soil attributes over a small distance. We can observe there is a general trend of low to high EC from East to West. A recommendation for the usability of these zones would be to establish a relationship between EC and the soil types of the field and then base the irrigation amount on the water holding capacity of the soil.



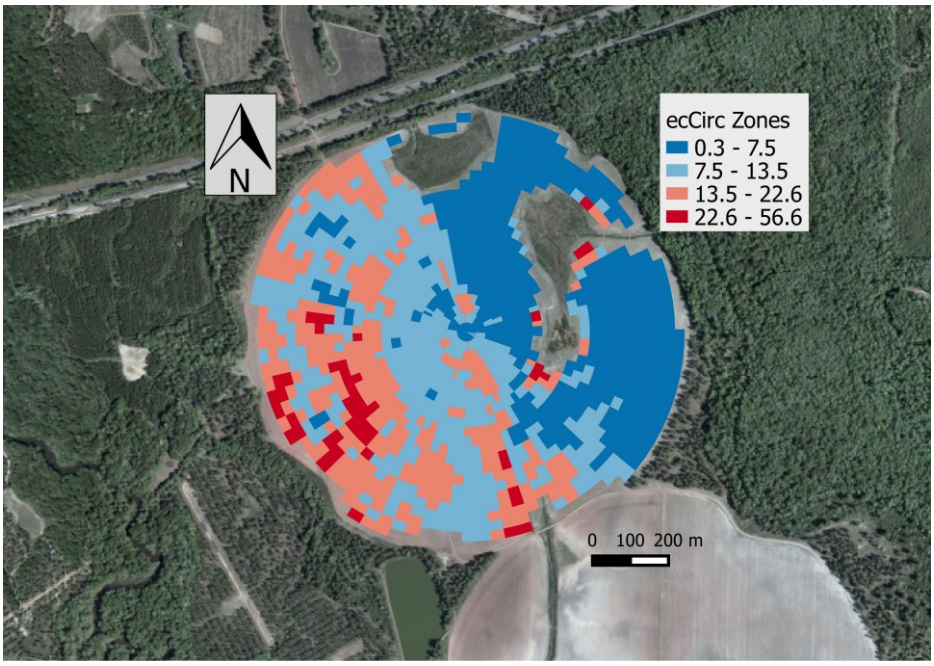


Figure 3.14: The ecCirc GVF (top) and zones (bottom).

3.3.3 ecRect classification

This field was classified using the rectangular polygon procedure, with a row spacing of 2 m. The input data for the field also seems to exhibit GPS coordinate errors because the point data does not cover the entire field (*Figure 3.15*). In cases like this, it is best to zone areas for which the points represent, even if the zone map doesn't cover the entire field. While only 1 ha in size, this field had the largest standard deviation of all the fields. We can see a distinct trend in the EC values of the field from northwest to southeast. Because of this strong trend over a relatively small area, it took 3 zones to optimize the GVF and the resulting zones are very homogenous (*Figure 3.16*). This field is a good candidate for application of VRT because it is small enough that different techniques could be applied relatively easily.

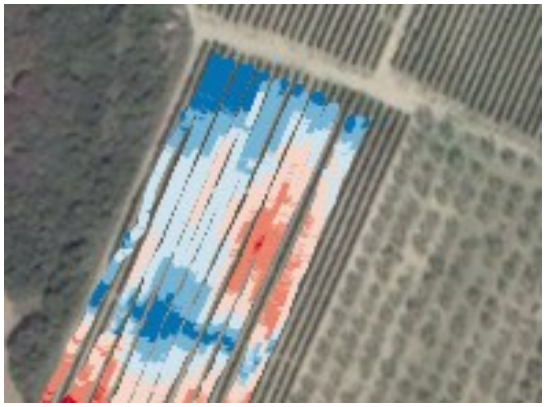
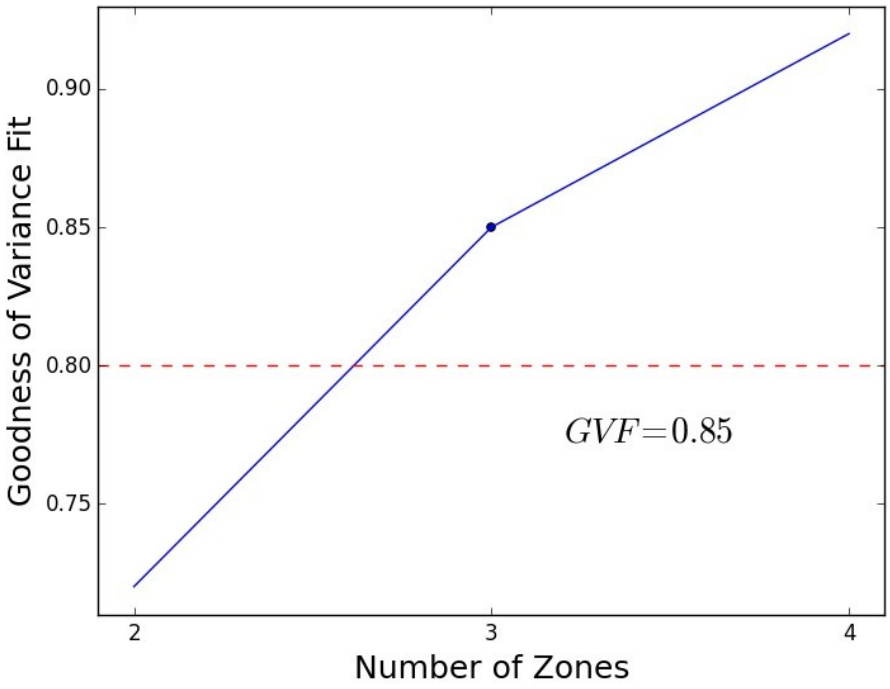


Figure 3.15: An example of inaccurate georeferencing in ecRect, note how the points do not cover the eastern part of the field.



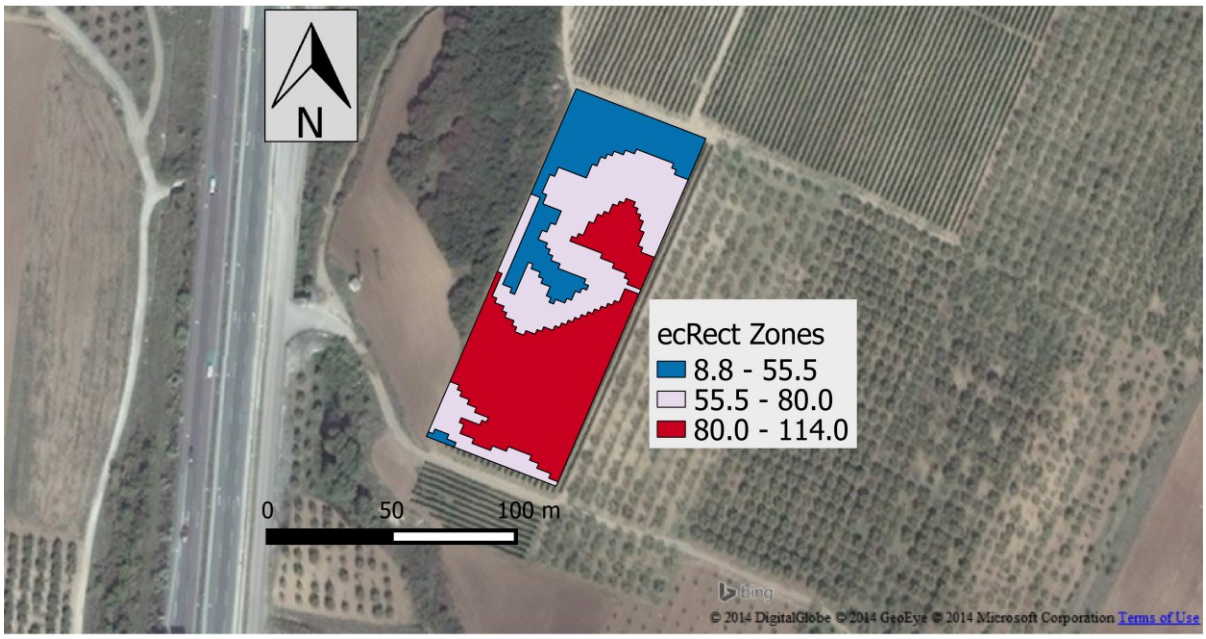


Figure 3.16: The ecRect GVF (top) and zones (bottom).

3.3.4 ecIrreg classification

This field has a uniquely irregular shape and there are lowland areas that are unplanted. There is little spatial homogeneity present in the field and the EC values seem to be irregularly distributed. This results in very heterogeneous zones of the field and the GVF optimized at 3 zones using a spacing of 2.75 m (*Figure 3.17*). This highlights one of the drawbacks of MZD in general. In this case to obtain a valid statistical result, we sacrifice some readability or usability of the map. A feature unique to EZZone is the ability to smooth the completed zones via a smoothing parameter. This allows us to take zone polygons and merge them with larger polygons of different zones. At this point we do sacrifice some validity for the benefit of usability. The results of the GVF optimized zones and the smoothed zones using a smoothing parameter of 5 are seen in *Figure 3.18.*

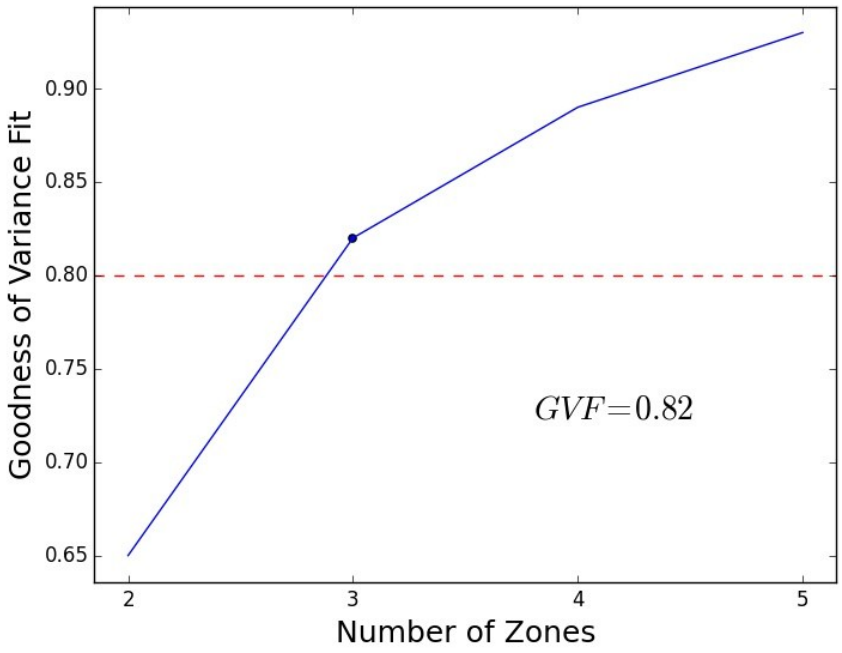


Figure 3.17: Illustrating the ecIrreg GVF optimization.

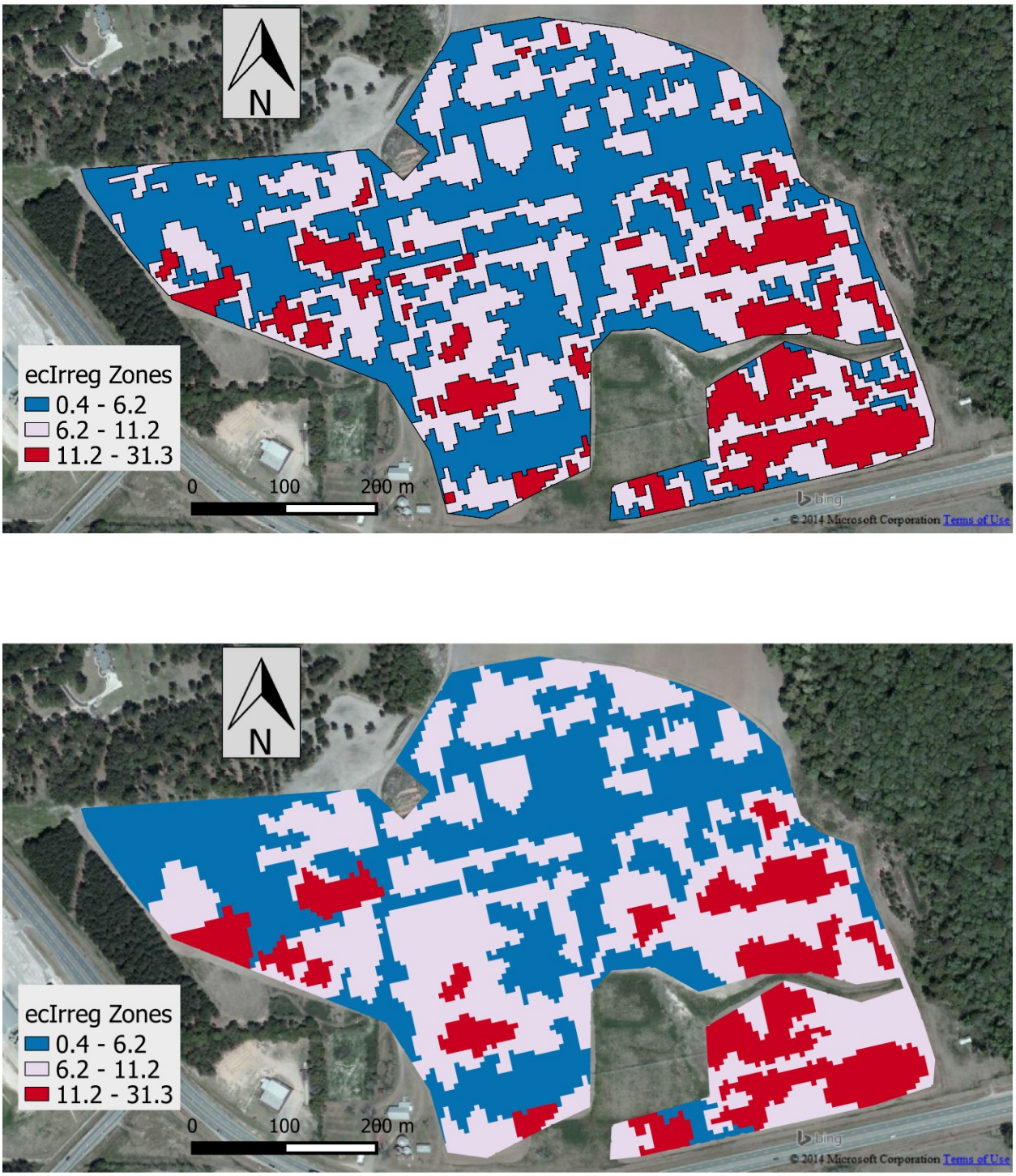


Figure 3.18: The ecIrreg GVF optimized zones (top) and smoothed zones (bottom), using a smoothing parameter of 5.

3.3.5 yieldIrreg classification

The yield for a cotton field was zoned with the irregular field method although this field is somewhat rectangular (*Figure 3.12*, D). The reasoning behind this is that there were a collection of points that were completely removed from the rest of the points in the field (*Figure 3.19*, A). These are data collected by the yield monitor during calibration. EZZone's strength lies in its visualization of the MZD process. In this case, we recognized that the produced boundary did not match the boundary of the field and adjusted the created boundary to match the actual boundary. In the process we are instructing EZZone to ignore those points outside of our created boundary. The GVF for yieldIrreg optimized at 4 zones and a smoothed result using a parameter of 5 is presented in *Figure 3.20*. In the middle of the field we notice that the yield is higher and exhibits a break from the trend above and below the zone (*Figure 3.19*, B). This area is actually the edge of the row and these zones are the results of edge effects in the field, where the yield is abnormally high due to increased solar radiation or sensor error. This field had no visible trend and this is reflected in the created zones.

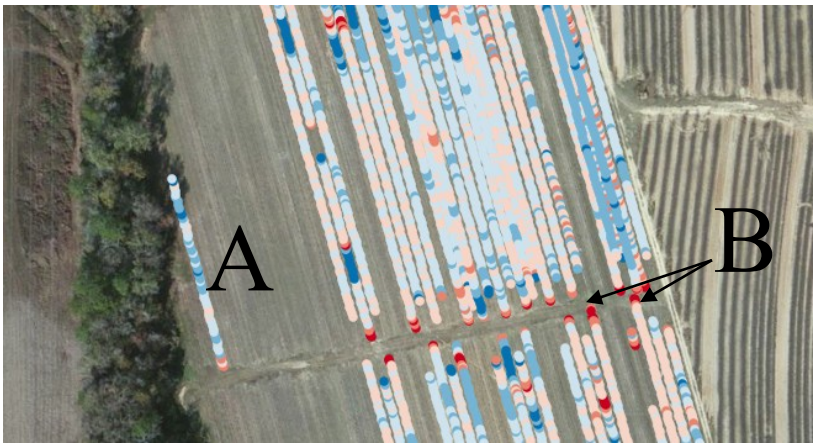


Figure 3.19: The yieldIrreg calibration strip (A) and edge effects location (B).

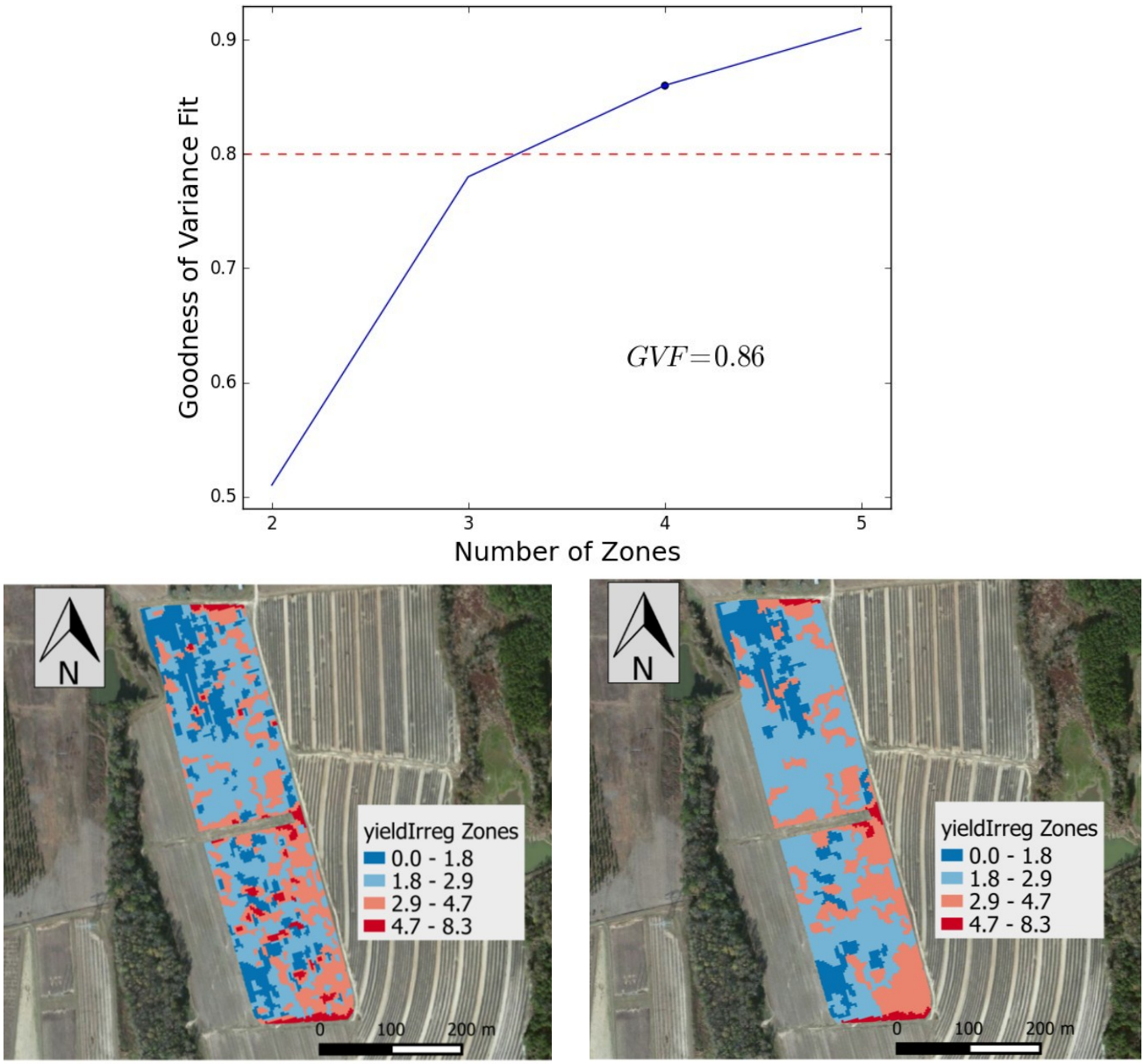


Figure 3.20: The yieldIrreg GVF (top) and raw zones (bottom left) and smoothed zones (bottom right).

3.3.6 yieldIrreg2 classification

The yield for wheat was zoned using the irregular field method and with a row spacing of 1 m. EZZone optimized the GVF in three zones (*Figure 3.21*). The resulting zones can be seen in *Figure 3.22*. In this field we can see large homogenous zones and a marked variability in yield in a distinct southwest to northeast trend. This field is a good candidate for VRT because of the distinct variability in yield as evident in the created zones. Considering that the yield in portions of the field is over two times greater than in other portions, this is a good opportunity to adapt strategies of site-specific management. Yield maps are arguably the primary target of zoning at least for their ability to express to farmers why VRT is important.

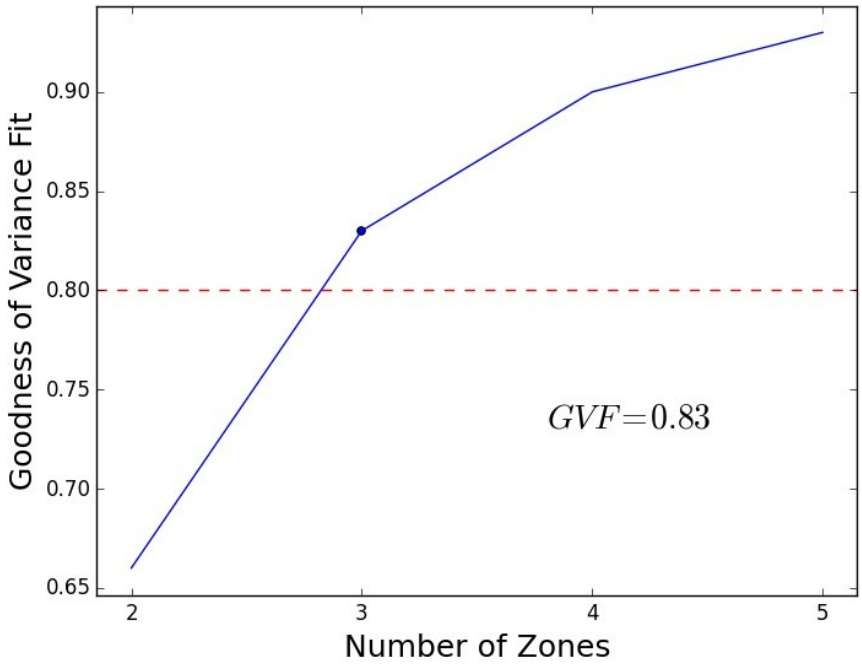


Figure 3.21: yieldIrreg2 GVF

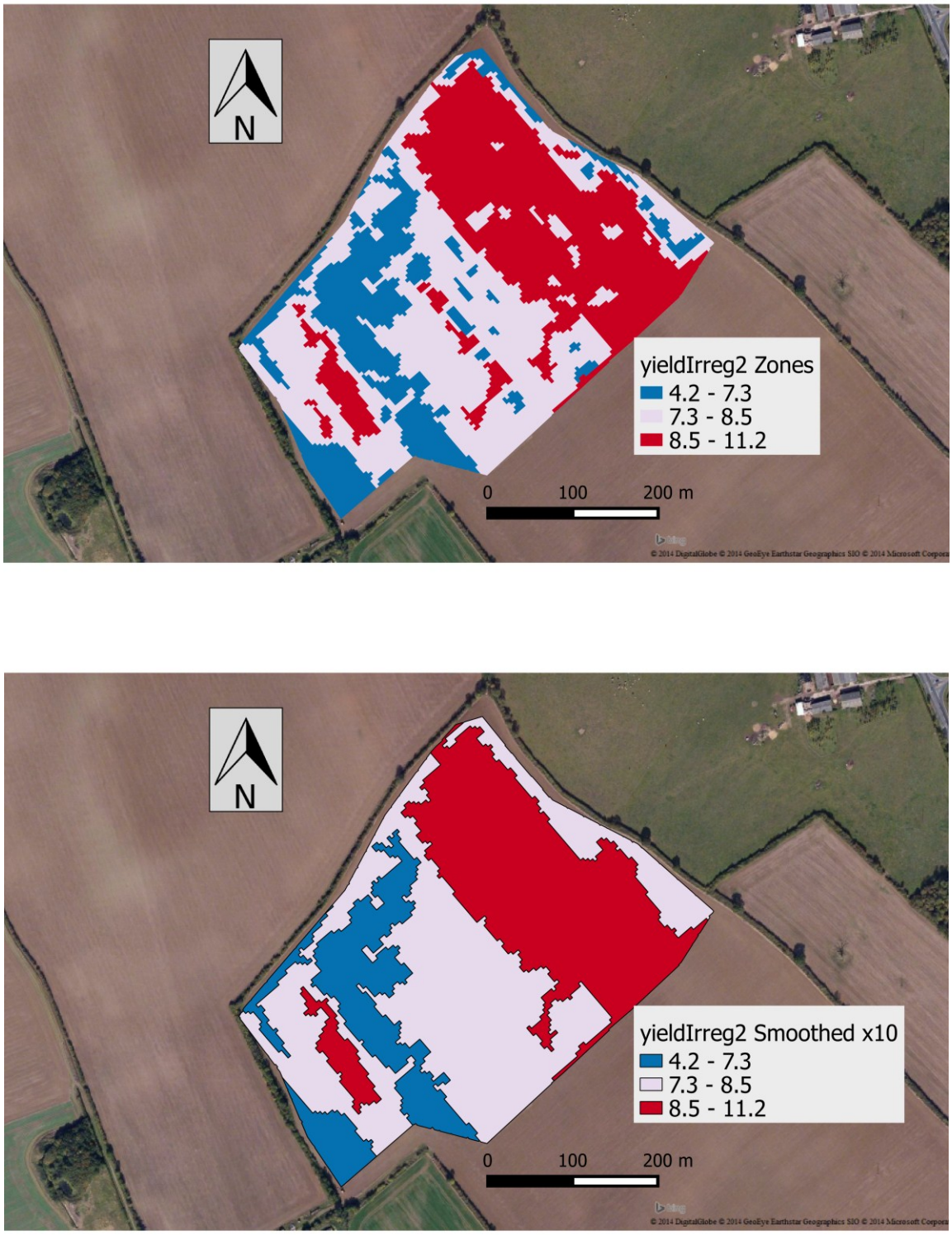


Figure 3.22: The completed yieldIrreg2 zones (top) and smoothed zones (bottom) using a smoothing parameter of 10

3.3.7 Limitations and advantages of proposed technique

There are limitations to the proposed technique that should be explained and evaluated. The first and foremost limitation is this EZZone is a strictly univariate technique. Limiting the technique to one variable requires the user to run the analysis more than once if more than one variable is available and leaves the user with the conundrum of how to integrate potentially different management zones resulting from each run. However, this also creates an advantage in that it encourages the user to input a variable that has the most impact on their management strategy. For instance, a user seeking a map for nitrogen application zones could select a different variable than one seeking Variable Rate Irrigation zones. Because of the ease of use of the system, users could create multiple zone maps for their different inputs and technologies. This is a decided advantage over grouping all variables into one MZD to produce one group of management zones that theoretically account for all variables. It is also much harder to interpret a multivariate zone map than a univariate one. One can easily identify the characteristics of the zones produced by EZZone because they are shown to the user and included in the output. To determine the characteristics of a multivariate zone is more complex and limits their usability. Another disadvantage of the clustering algorithm is that it only considers the attribute and not the location of the attribute in relation to other points, in other words the clustering does not account for the location of the point, only its value. This lack of spatial consideration in the zone classification is undesirable but the existing libraries for spatial clustering in Python are not easily integrated with the data structures used by EZZone.

Another limitation is the extrapolation of data values outside of the input data set. This is a problem with any interpolation technique, the current method of extrapolation data values

in the algorithm is to use a nearest neighbor function, which simply returns the known value closest to the point of interpolation. Of course the interpolation is susceptible to the quality of the input data, so an outlier used as an extrapolated value in the nearest neighbor function could create inaccurate zones. For these reasons, extrapolation should be minimized as much as possible, as demonstrated in 3.3.3. The classification technique has its flaws as well, and the GVF was optimized faster when there was a distinct pattern in the fields as evident in *ecRect* and *yieldIrreg2*. We have seen that in the case of *yieldIrreg* that the zones were very heterogeneous. It would take significant smoothing of the zones to create a map that captures the general trend of the field. The smoothing algorithm has no assessment of the GVF of the smoothed zones and is based on the geometry alone. By smoothing the zones we capture the general trend in the field but lose a fine spatial resolution.

The use of field attributes such as row/planter spacing and row orientation is a unique aspect of this algorithm that allows for easier integration with farm equipment and practices. The goal of any MZD should be to produce zones that are useful to the users. This is certainly the primary goal of EZZone and the user's field type and dimensions are the key variable for the algorithm. A user could take the output of EZZone and integrate it with a VRT prescription map with ease because EZZone produces GIS data that can be read by any agricultural GIS software.

3.4 **Conclusions and future work**

EZZone is available online at ezzone.pythonanywhere.com and does not require a user account or payment to operate. The tool's code is also available at the GitHub repository camdenl/EZZone. I encourage anyone interested in the tool to look at the source code and to

modify or contribute to the development of EZZone. The tool is in a stable release but future expansions for the tool will be planned if there is a positive reaction from the public. Because of the strong GIS foundation of the tool, it could integrate multivariate zoning techniques and spatial clustering algorithms, however this was not a goal of the initial release. I feel this tool is a prime example of the advantages of using open source technology in agriculture and hope that this tool will foster more open collaboration in the field of agricultural research. Because the process of creating a zone map is easy, fast and accessible to anyone with an internet connection and a computer, I think this tool will increase the use of management zones in agriculture and will lead to more efficient and sustainable agricultural management practices.

CHAPTER 4

CONCLUSIONS

The goal of this system was to obtain accurate sap flow measurements. Although I encountered problems with the instrumentation and the field experiment, I was able to show that a home-made system can be developed inexpensively and produce good results of plant water status in a non-irrigated vineyard. We have also shown that there is variability in these measurements throughout the field. We have observed the expected relationships between sap flow and other indicators of plant water status. Although we expect these measurements to have been more accurate in a more intensely managed vineyard where the vines are healthier and where irrigation is applied, there are many improvements to be made to increase the accuracy of the measurements. This system is important because it reduces the cost of acquiring field scale variability of plant water status. As water resources continue to become a major global issue, it is imperative that PA produce water management technologies that are cost effective and useful to users. This has been accomplished here by adapting an open source microcontroller platform to monitor the water status of plants in real time and can be easily integrated with existing FMIS. All of the software is documented online so that others can easily replicate the system here. I hope this work will inspire others to share new technologies with the global research community and to realize the importance of open source technology to the future of agriculture.

EZZone is available online at ezzone.pythonanywhere.com and does not require a user account or payment to operate. The tool's code is also available at the GitHub repository camdenl/EZZone. I encourage anyone interested in the tool to look at the source code and to modify or contribute to the development of EZZone. The tool is in a stable release but future expansions for the tool will be planned if there is a positive reaction from the public. Because of the strong GIS foundation of the tool, it could integrate multivariate zoning techniques and different clustering algorithms, however this was not a goal of the initial release. I feel this tool is a prime example of the advantages of using open source technology in agriculture and hope that this tool will foster more open collaboration in the field of agricultural research. Because the process of creating a zone map is easy, fast and accessible to anyone with an internet connection and a computer, I think this tool will increase the use of management zones in agriculture and will lead to more efficient and sustainable agricultural management practices.

REFERENCES

[1] S. E. Cook and R. G. V. Bramley, “Precision agriculture — opportunities, benefits and pitfalls of site-specific crop management in Australia,” *Aust. J. Exp. Agric.*, vol. 38, no. 7, p. 753, 1998.

[2] T.-G. Katalin, T. Rahoveanu, M. Magdalena, and T. István, “Sustainable New Agricultural Technology – Economic Aspects of Precision Crop Protection,” *Procedia Econ. Financ.*, vol. 8, pp. 729–736, 2014.

[3] D. W. Lamb, P. Frazier, and P. Adams, “Improving pathways to adoption: Putting the right P’s in precision agriculture,” *Comput. Electron. Agric.*, vol. 61, no. 1, pp. 4–9, 2008.

[4] M. J. Robertson, R. S. Llewellyn, R. Mandel, R. Lawes, R. G. V Bramley, L. Swift, N. Metz, and C. O’Callaghan, “Adoption of variable rate fertiliser application in the Australian grains industry: Status, issues and prospects,” *Precis. Agric.*, vol. 13, no. 2, pp. 181–199, 2012.

[5] R. Nikkilä, I. Seilonen, and K. Koskinen, “Software architecture for farm management information systems in precision agriculture,” *Comput. Electron. Agric.*, vol. 70, no. 2, pp. 328–336, Mar. 2010.

[6] R. G. V. Bramley and R. P. Hamilton, “Understanding variability in winegrape production systems,” *Aust. J. Grape Wine Res.*, vol. 10, no. 1, pp. 32–45, Mar. 2008.

[7] C. Acevedo-Opazo, S. Ortega-Farias, and S. Fuentes, “Effects of grapevine (Vitis vinifera L.) water status on water consumption, vegetative growth and grape quality: An irrigation scheduling application to achieve regulated deficit irrigation,” *Agric. Water Manag.*, vol. 97, no. 7, pp. 956–964, Jul. 2010.

[8] A. Patakas, B. Noitsakis, and A. Chouzouri, “Optimization of irrigation water use in grapevines using the relationship between transpiration and plant water status,” *Agric. Ecosyst. Environ.*, vol. 106, no. 2–3 SPEC. ISS., pp. 253–259, 2005.

[9] A. Pellegrino, E. Lebon, M. Voltz, and J. Wery, “Relationships between plant and soil water status in vine (Vitis vinifera L.),” *Plant Soil*, vol. 266, no. 1–2, pp. 129–142, 2005.

[10] S. Koundouras, V. Marinos, A. Gkoulioti, Y. Kotseridis, and C. Van Leeuwen, “Influence of vineyard location and vine water status on fruit maturation of nonirrigated cv.

Agiorgitiko (Vitis vinifera L.). Effects on wine phenolic and aroma components," *J. Agric. Food Chem.*, vol. 54, no. 14, pp. 5077–5086, 2006.

[11] H. R. Schultz, "Water relations and photosynthetic responses of two grapevine cultivars of different geographical origin during water stress," *Strateg. to Optim. Wine Grape Qual. 427*, pp. 251–266, 1995.

[12] M. Greven, S. Green, S. Neal, B. Clothier, M. Neal, G. Dryden, and P. Davidson, "Regulated deficit irrigation (RDI) to save water and improve Sauvignon Blanc quality?," *Water Sci. Technol.*, vol. 51, no. 1, pp. 9–17, Jan. 2005.

[13] L. G. Santesteban, C. Miranda, and J. B. Royo, "Regulated deficit irrigation effects on growth, yield, grape quality and individual anthocyanin composition in Vitis vinifera L. cv. 'Tempranillo,'" *Agric. Water Manag.*, vol. 98, no. 7, pp. 1171–1179, 2011.

[14] J. Escalona and H. Medrano, "Drought effects on water flow, photosynthesis and growth of potted grapevines /," *Vitis.*, vol. 41, no. 2, pp. 57 – 62, 2002.

[15] Y. Zhang, S. Kang, E. J. Ward, R. Ding, X. Zhang, and R. Zheng, "Evapotranspiration components determined by sap flow and microlysimetry techniques of a vineyard in northwest China: Dynamics and influential factors," *Agric. Water Manag.*, vol. 98, no. 8, pp. 1207–1214, May 2011.

[16] P. Braun and J. Schmid, "Sap flow measurements in grapevines (Vitis vinifera L.) 2. Granier measurements," *Plant Soil*, vol. 215, no. 1, pp. 47–55, Sep. 1999.

[17] I. A. M. Yunusa, R. R. Walker, B. R. Loveys, and D. H. Blackmore, "Determination of transpiration in irrigated grapevines: comparison of the heat-pulse technique with gravimetric and micrometeorological methods," *Irrigation Science*, vol. 20, no. 1. pp. 1–8, 2000.

[18] J. E. Fernández, S. R. Green, H. W. Caspari, A. Diaz-Espejo, and M. V. Cuevas, "The use of sap flow measurements for scheduling irrigation in olive, apple and Asian pear trees and in grapevines," *Plant Soil*, vol. 305, no. 1–2, pp. 91–104, Aug. 2007.

[19] A. Granier, "Une nouvelle méthode pour la mesure du flux de sève brute dans le tronc des arbres," *Ann. des Sci. For.*, vol. 42, no. 2, pp. 193–200, 1985.

[20] A. W. Van Herwaarden and P. M. Sarro, "Thermal sensors based on the seebeck effect," *Sensors and Actuators*, vol. 10, no. 3–4. pp. 321–346, 1986.

[21] P. Lu, L. Urban, and P. Zhao, "Granier's Thermal Dissipation Probe (TDP) method for measuring sap flow in trees: Theory and practice," *Acta Bot. Sin.*, vol. 46, no. 6, pp. 631–646, 2004.

[22] A. Granier, “Evaluation of transpiration in a Douglas-fir stand by means of sap flow measurements,” *Tree Physiol.*, vol. 3, no. 4, pp. 309–320, Dec. 1987.

[23] N. Zhang, M. Wang, and N. Wang, “Precision agriculture—a worldwide overview,” *Computers and Electronics in Agriculture*, vol. 36, no. 2–3. pp. 113–132, 2002.

[24] K. L. Fleming, D. G. Westfall, D. W. Wiens, and M. C. Brodahl, “Evaluating Farmer Defined Management Zone Maps for Variable Rate Fertilizer Application,” *Precis. Agric.*, vol. 2, no. 2, pp. 201–215, Oct. 2000.

[25] G. Vellidis, M. Tucker, C. Perry, C. Kvien, and C. Bednarz, “A real-time wireless smart sensor array for scheduling irrigation,” *Comput. Electron. Agric.*, vol. 61, no. 1, pp. 44–50, 2008.

[26] R. Khosla, K. Fleming, J. A. Delgado, T. M. Shaver, and D. G. Westfall, “Use of site-specific management zones to improve nitrogen management for precision agriculture,” *J. Soil Water Conserv.*, vol. 57, no. 6, pp. 513–518, 2002.

[27] A. J. Jones, L. N. Mielke, C. A. Bartles, and C. A. Miller, “Relationship of landscape position and properties to crop production,” *J. Soil Water Conserv.*, vol. 44, no. 4, pp. 328–332, Jul. 1989.

[28] Y. Li, Z. Shi, and F. Li, “Delineation of Site-Specific Management Zones Based on Temporal and Spatial Variability of Soil Electrical Conductivity,” *Pedosphere*, vol. 17, no. 2. pp. 156–164, 2007.

[29] D. L. Corwin and S. M. Lesch, “Apparent soil electrical conductivity measurements in agriculture,” *Comput. Electron. Agric.*, vol. 46, no. 1–3, pp. 11–43, Mar. 2005.

[30] J. J. Fridgen, N. R. Kitchen, K. A. Sudduth, S. T. Drummond, W. J. Wiebold, and C. W. Fraisse, “Management zone analyst (MZA): software for subfield management zone delineation,” *Agronomy journal*, 2004. [Online]. Available: http://naldc.nal.usda.gov/catalog/8380. [Accessed: 22-Sep-2014].

[31] R. M. Lark, “Forming Spatially Coherent Regions by Classification of MultiVariate Data: An Example from the Analysis of Maps of Crop Yield,” *Int. J. Geogr. Inf. Sci.*, vol. 12, no. 1, pp. 83–98, 1998.

[32] F. Guastaferro, A. Castrignanò, D. de Benedetto, D. Sollitto, A. Troccoli, and B. Cafarelli, “A comparison of different algorithms for the delineation of management zones,” *Precis. Agric.*, vol. 11, no. 6, pp. 600–620, 2010.

[33] ESRI, “Grid Commands.” 1994.

[34] B. J. Irvin, S. J. Ventura, and B. K. Slater, “Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin,” in *Geoderma*, 1997, vol. 77, no. 2–4, pp. 137–154.

[35] Y. Li, Z. Shi, F. Li, and H.-Y. Li, “Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land,” *Computers and Electronics in Agriculture*, vol. 56, no. 2. pp. 174–186, 2007.

[36] F. J. Moral, J. M. Terrón, and J. R. M. da Silva, “Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques,” *Soil Tillage Res.*, vol. 106, no. 2, pp. 335–343, 2010.

[37] T. E. Oliphant, “Python for scientific computing,” *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 10–20, 2007.

[38] J. Girona, M. Mata, J. Del Campo, A. Arbonés, E. Bartra, and J. Marsal, “The use of midday leaf water potential for scheduling deficit irrigation in vineyards,” *Irrig. Sci.*, vol. 24, pp. 115–127, 2006.

[39] D. M. Smith and S. J. Allen, “Measurement of sap flow in plant stems,” *J. Exp. Bot.*, vol. 47, no. 305, pp. 1833–1844, 1996.

[40] K. B. Wilson, P. J. Hanson, P. J. Mulholland, D. D. Baldocchi, and S. D. Wullschleger, “A comparison of methods for determining forest evapotranspiration and its components: Sap-flow, soil water budget, eddy covariance and catchment water balance,” *Agric. For. Meteorol.*, vol. 106, no. 2, pp. 153–168, 2001.

[41] T. W. Davis, C.-M. Kuo, X. Liang, and P.-S. Yu, “Sap flow sensors: construction, quality control and comparison.,” *Sensors (Basel).*, vol. 12, no. 1, pp. 954–71, Jan. 2012.

[42] S. S. O. Burgess, M. A. Adams, N. C. Turner, C. R. Beverly, C. K. Ong, A. A. H. Khan, and T. M. Bleby, “An improved heat pulse method to measure low and reverse rates of sap flow in woody plants,” *Tree Physiol.*, vol. 21, no. 9, pp. 589–598, Jun. 2001.

[43] R. G. Allen, L. S. Pereira, D. Raes, M. Smith, and A. B. W, “Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56,” *Irrig. Drain.*, pp. 1–15, 1998.

[44] P.-S. Yu, “SciPy: Open source scientific tools for Python.” 2001.

[45] C. Cauvin, F. Escobar, and A. Serradj, *Cartography and the Impact of the Quantitative Revolution*, vol. 2. John Wiley and Sons, 2013.