Analysis of Interactions Between Soil and Water

James S. Tyler

A project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Blake Barker, Chair
Emily Evans
Jared Whitehead

Department of Mathematics

Brigham Young University

ABSTRACT

Analysis of Interactions Between Soil and Water

James S. Tyler
Department of Mathematics, BYU
Master of Science

In this project, we discuss different methods for analyzing relationships between soil and water. We carry out two studies, the first analyzing the flow of nutrients through an ecosystem as measured by analyzing stream data, and the second developing a model to predict soil water content from soil temperature. We use both traditional "mechanistic" approaches and more modern "characteristic" approaches on both data sets in order to form predictive models. We develop a novel Non-Uniform Label Smoothing (NULS) [3] technique in a Deep Neural Network (DNN) environment to create a predictive model which performs well on the second study, and yet fails to do so on the first. This new technique helps with uncertainty quantification of the predictive model [1] in a way that is typically quite difficult to do in a regression context. We discuss the pros and cons of mechanistic and characteristic approaches for these two studies. We relate our work to the role of the Data Analyst in modern America as "big data" becomes more abundant and DNNs become cheaper to implement and more accessible then ever [27].

Acknowledgements

I would like to give special thanks to the Brigham Young University Math Department for the research opportunities I was given.

I would like to thank Dr. Neil Hansen for his willingness to share the data, which he spent a great deal of trouble obtaining. I would also like to thank the collaborators on the irrigation team, Damian Anderson, Sophia Anderson, Blake Barker, Joseph Engel, Neil Hansen, Jane Housley, Jake Murphy, and Brynn Woolley.

I would like to thank Haley Moon for her patient teaching. I would like to thank the collaborators on the Wildfire research team, Ben Abbott, Kallin Austin, Addison Bliss, Haley Moon, Jake Murphy, Adam Norris, Robby Sainsbury, and Daniel Smith.

I would like to thank the committee members for their time with this project. I would like to thank Dr. Evans for her work with the Math F.I.R.E. group, and Dr. Whitehead for first igniting my interest in machine learning. I would like to thank my advisor and mentor, Dr. Blake Barker, for his work with the Smart Irrigation. Without his mentoring, I would not have had the high quality education I did. His classes first taught me humility, and then taught me that I could do hard things. I would like to thank Lonette Stoddard for her dedication to the many students here at BYU.

Most of all, I would like to thank my lovely wife, Rebekah Tyler. Though I'm the one receiving honors, without her loving support none of this would have been possible. Her patience with long cryptic mathematical rants is unending, and there are not words enough to express my gratitude for her work as a wife and mother, though I perhaps hope that this can start. Thank you.

Contents

List of Figures

Chapter 1. Mathematical Background

1.1 Research Introduction

The efficient use of soil and water is essential to sustain and maintain the human population on a planet with limited resources [32]. As such, understanding soil and water interactions is key to a flourishing society. A better understanding allows for more efficient preservation of current ecosystems as well as better use of the agricultural resources at our disposal.

In this paper, we explore methods for analyzing the relationship between soil and water in multiple contexts. We do so in order to better understand nutrient flow through an ecosystem and to improve agricultural appropriation of resources. We also illustrate novel data analysis techniques utilizing deep neural networks in an effort to compare their effectiveness to more traditional methods of data analysis.

Our first study will be an experimental analysis of water samples recovered as runoff from areas affected recently by wildfire. Levels of Biodegradable Organic Material (hereafter BDOM) are an important measurement of ecological health [31], and experiments are conducted on samples to better understand the precise nature of a recovering environment. Analysis was done by the Brigham Young University (hereafter BYU) Math F.I.R.E. group.

Our second study is an observational study on the relationship between soil temperature and volumetric water content (hereafter VWC). The ultimate goal of this study is the application of thermometers to more precisely understand the dynamics

of VWC throughout a field. The application of this analysis could increase efficiency of irrigation systems worldwide [32].

Our methods can be broadly summarized into two categories. In the first general method, we use first principle models to understand the relationships at play between soil and water. We then use traditional methods of data analysis to explore correlations and make a predictive model. Hereafter, this will be referred to broadly as the Mechanistic approach.

The second general method is to utilize Deep Neural Networks (hereafter DNN) with a variety of inputs to generate a model, wherein the underlying relationships are not explored, but emphasis is put into making accurate and precise predictions on the data. Hereafter, this will be referred to broadly as the Characteristic approach.

We illustrate the relative pros and cons of the two approaches and utilize both, to the extent possible, on each study. Results for the two studies are then discussed along with potential ongoing research topics.

1.2 Characteristic versus Mechanistic Methodologies

The two research datasets we studied had many similarities, but they lent themselves mathematically to very distinct problem sets. This made them ideal for highlighting differences between traditional and modern analysis methods. We discuss the pros and cons of each method further below.

Figure 1.1 contrasts the methodologies in a diagram. An example of what a Mechanistic Approach diagram might look like is given. Inputs are known and problem formulation concerns their direct relationship/affect on the output. Then an ex-

ample of a Characteristic Approach diagram is given. The underlying relationships governed by the hidden layer weights are not known, which relates to the output in a non-linear manner.

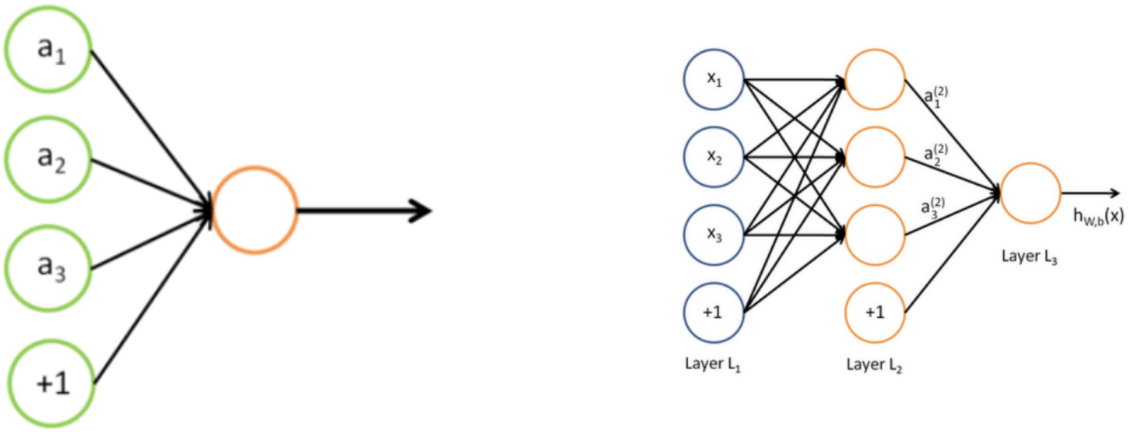

Figure 1.1: On the left is the mechanistic approach, on the right is the characteristic approach. Figure courtesy of Afroz Chakure, DEV.TO.

In determining whether or not to take the mechanistic or characteristic approach, many different factors are considered. Below are three:

(i) The dataset available [11]. A large factor in constructing a DNN is the size/-type of data available. Most DNN's require large datasets, especially relative to the questions that can be answered using a Multiple Linear Regression (hereafter MLR) approach. Additionally, in a DNN caution has to be taken to keep from utilizing unbalanced data [12] or overfitting of data [17]. As a tradeoff, less worry is given to factors such as independence between variables, which a DNN can account for naturally, unlike a MLR model .

(ii) The problem formulation. For many problems, the characteristic approach answers the question naturally using methods that are well understood and easily generalizable. Given the novelty of machine learning techniques and how the mathematical foundations of these are understood, often the characteristic approach may be adequate. If the problem of study is the relationship between variables, then a DNN may not be well-suited. However, in predictive problem formulations, a DNN can often out perform a traditional MLR model as it can handle non-linearities with relative ease and isn't hampered by a need to understand the underlying relationships or account for many confounding factors.

(iii) The generalizability of the study. Many times a DNN can be constructed on a particular dataset and yet found to work very poorly on similar datasets [15]. Because of a lack of understanding on the underlying relationships, DNN's may not generalize well to other scenarios. However, if the relationships between variables is discovered in the traditional sense, then often we can account for this in making a predictive model. Traditional methods work better in extrapolating outside of initial datasets, whereas DNN's often fail to find anything meaningful on data even marginally different from that in the training set. Research into the underlying relationships governing a DNN is ongoing [21].

The two projects researched below provide comparison for the two methodologies. In the Wildfire research, we found that mechanistic methods quickly and clearly answered all of our questions and provided an understanding of the variables at play, and yet the characteristic model based off the dataset never provided a good predictive model due to unreliable training and imbalanced data [12]. Alternatively, in the Irrigation research, we found that the underlying relationships of the physical mechanisms were not understood well enough to model with a mechanistic approach. While research in that is ongoing, we did find the data was well suited to a characteristic predictive model. This suited the needs of the research group, as our primary interest was in relative spatial predictive modeling of VWC. Research into the underlying relationships is still ongoing. It's important to note that in both scenarios, methods from one category may help guide another. This is illustrated in the Irrigation research section below.

In all characteristic models, datasets are split into training and testing sets for purposes of evaluating the model and minimizing loss. This keeps the DNN from learning on its own testing set, introducing error into validation. [33]

1.3 Non-Uniform Label Smoothing

In regression DNN models, the objective is to create an architecture with a single output layer. Its been found that utilizing large DNN models with many hidden

layers makes a model susceptible to overfitting and may make it generalize poorly [6]. Calibration of a DNN model is analogous to uncertainty quantification in a traditional data analysis approach. Model predictions need a probability distribution placed around the predicted value in order to be more generally applicable [18]. Calibration in a DNN model is fine-tuning so that the predicted value is not only accurate but consistently within a specific known range. This is best seen in classification models where often a soft-max activation function will generate a probability vector as an output. Calibration here can be easily measured as probability assigned (or confidence) and can be leveraged towards uncertainty quantification. However, in regression DNN models, activation functions on the final layer give only one output. This means that "confidence" in the prediction is not measured.

One common remedy to this is a pragmatic approach of measuring the spread of the data and arbitrarily drawing 95% confidence intervals based off a spread that captures 95% of the predictions. However, in a time-series model, this is not possible due to lack of replication.

The method that we propose is to translate the regression problem into a discrete classification problem with ordinal categorical variables. Then using Non-Uniform Label Smoothing (hereafter NULS) along with the cross-entropy loss function, we appropriately incentivize the model to learn the underlying relationships [3], and make predictions in such a way as to control the distribution of the outputs. A brief explanation of the above methodology is given below. For details about the mathematics of machine learning see [8]. For further information on the mathematics of label smoothing, see [35]. For another instance of Non-Uniform Label Smoothing

classification, see [3].

A DNN consists broadly of an input layer, hidden layers, and an output layer. The output layer is determined by the task which the DNN tries to accomplish. In regression, the task of a DNN is to predict a single continuous variable as output. In our research studies described below, this is BDOM or VWC.

In classification problems, a DNN will give a softmax vector output which represents a probability distribution among the different discrete categories. In order for the model output to be a probability distribution, all probabilities must sum to one and be in the range of zero to one. The softmax activation function preserves the outputs of the DNN while giving it these qualities, associating larger input components with larger probabilities by utilizing the formula below on the output vector of the DNN:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, \text{ for } 1 \leq i \leq n, \ z = (z_1, z_2, ..., z_n) \in \mathbb{R}^n.$$

In this way, if z is the input vector, then $\hat{z} = (\sigma(z_1), \sigma(z_2), ..., \sigma(z_n))$ is the softmax output vector. If we have an n-dimensional output vector from our DNN, the softmax will act with the above formula on every element of our output vector to produce a probability distribution vector of dimension n.

In a classification problem, the softmax output is compared with a "One-Hot Encoded" target vector: a vector of all zeros besides the correct classifier. These classifiers are unrelated to each other and the accuracy of the DNN is measured by its ability to predict the correct classification. These are compared most often with Cross Entropy [36]. In a uniformly label smoothed target vector, the vector is scaled so that it also represents a probability distribution, but the incorrect classifiers are

given a weight corresponding to a fine-tuned hyper-parameter. This means that a cross-entropy comparison will now penalize models for having large predictions in incorrect classifiers even if overall the model predicts correctly. This keeps models from becoming overconfident as a means of calibration.

Non-uniform label smoothing establishes a relationship between categories of classification. In the label smoothing process, instead of giving a uniform small weight to the incorrect classifiers of the target vector, another method for distributing the weights is utilized to keep the target vector as a probability distribution. The method that we utilize is a normal distribution, as per the formula:

$$\phi(x_i : x_j, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_i - x_j}{\sigma}\right)^2\right), \text{ for } 1 \leq i \leq n,\ x = (x_1, ..., x_n) \in \mathbb{R}^n,$$

where x_j is the correct classifier, σ is the fine-tuned standard deviation of the model, and $\hat{x} = (\phi(x_1 : x_j, \sigma), \phi(x_2 : x_j, \sigma), ..., \phi(x_n : x_j, \sigma)) \in \mathbb{R}^n$ is the Non-Uniformly Label Smoothed target vector. This creates a vector which assigns a relationship between close ordinally related classifiers.

The most common loss function in state of the art classification DNN's is cross entropy [36]. The cross entropy could be thought of as measuring the similarity between two probability distributions. Below is an explanation of how this loss function behaves on the one-hot encoded, Uniformly Label Smoothed, and Non-Uniformly Label Smoothed target vectors. The formula for cross entropy is given below as a negative-log likelihood, where x is the treated target vector and z is the output vector of our DNN, with $x, z \in \mathbb{R}^n$:

$$l(x, z) = -\sum_{i=1}^{n} x_i \times \log(z_i).$$

Note from the formula above, if x is a target vector when x is one-hot encoded, only one element of the summation is non-zero. This means that only one element of the probability distribution z is taken into account in minimizing the loss function in the corresponding DNN when the target vectors are one-hot encoded.

When the target vector is uniformly label smoothed, the entire probability distribution x is now taken into account. However, only in order to ensure that the model doesn't assign weights disproportionately large to secondary guesses. That is, the model is penalized for uncertainty between two guesses.

When the target vector x is normally-distributed into a NULS target vector, its clear that this loss function will incentivize the model to reflect that close association. The probability distribution z is not only rewarded for a correct classification, but for putting similar confidence in closely associated ordinal classifiers.

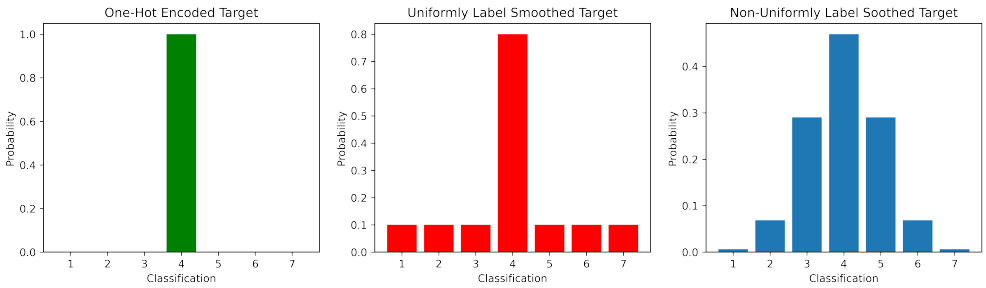

Figure 1.2: Left, One-Hot Encoded. Middle, Uniformly Label Smoothed. Right, Normally Distributed Non-Uniform Label Smoothed

Figure 1.2 shows an illustration of these three different classes of target vectors side by side. In the example figure, class "4" is the correct label. All of these reflect that, but it is clear that only the Normally Distributed NULS target vector establishes association between closely related classifiers.

In theory, and as results show below, this allows training a DNN which utilizes a classifier based architecture and yet mimics a regression DNN [3]. Utilizing a normally-distributed NULS calibrates the accuracy versus precision of the model and the probabilistic output can be used to measure directly the confidence with which the model asserts a prediction. The standard deviation of the normal distribution used in this smoothing is a hyper-parameter that is tuned until a reasonable trade-off between precision and accuracy is reached.

Chapter 2. Wildfire Research

2.1 Experimental Background

In 2018, the Spanish Fork River Watershed experienced a mega-fire, burning approximately 610 km^2. Scientists fear that such events become more catastrophic as humanity alters the ecosystem [5, 10]. Fire suppression may lead to buildup of underbrush which could possibly lead to an increased frequency of mega-fires [19]. Our study examines how mega-fires affect a watershed ecosystem, specifically nutrient flow through the water.

Haley Moon headed a research project from the Abbott Lab at BYU to collect water samples at 69 sites with differing frequencies. These sites are representative of the differing water shed locations from sections of the larger Spanish Fork River watershed, which were either highly affected by the mega-fire, moderately affected by the mega-fire, or else not affected by the fire but in a similar ecosystem to the fire range.

Figure 2.1 shows the differing watersheds of the Spanish Fork River along with the collection sites. Opacity relates to the overlap of watershed areas, and orange indicates areas affected by the mega-fire.

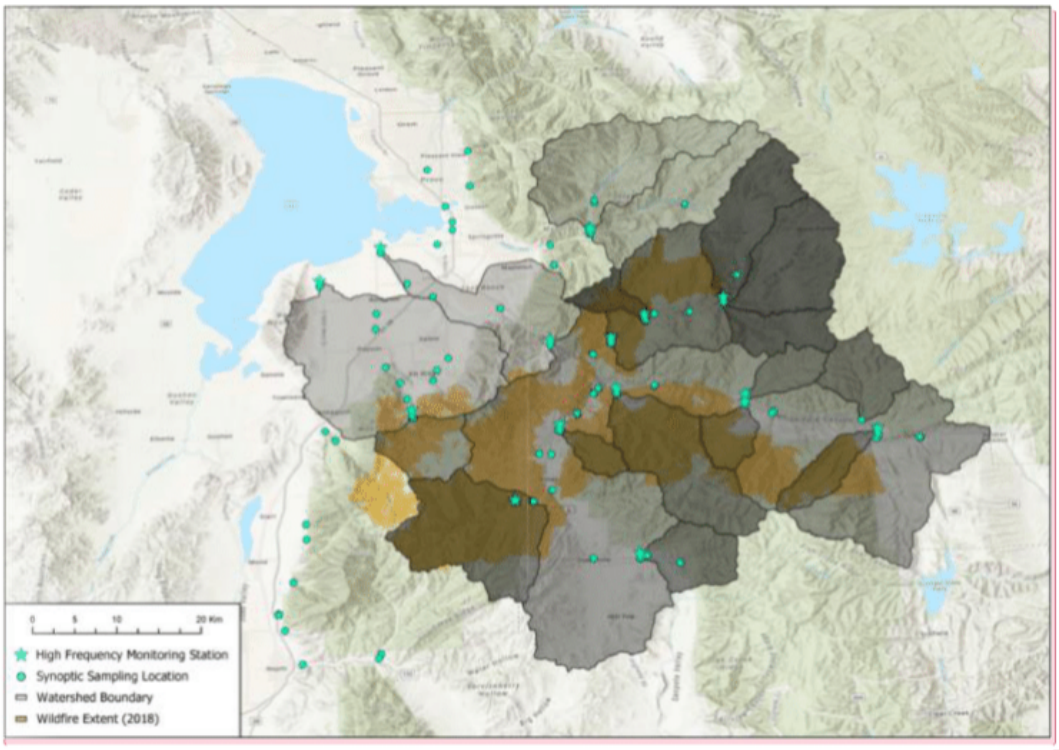

Figure 2.1: The Wildfire research watersheds and sample site locations. Figure courtesy of the Haley Moon research team.

Multiple samples were taken from each location for replication, but analysis focused on samples taken for comparison in the Spring and Summer. The samples underwent different experimental treatments. Some were given HCl to suspend microbial activity to be treated as a control. Others were given nutrients, and all were given a microbial inoculum. They were then incubated for 28 days in either a light environment or a dark environment. All samples, including the controls, were eval-

uated at the end of the 28 day incubation period. This was done in order to remove machine calibration as a potential confounding factor.

Figure 2.2 shows how the experiment was designed. Observations were split categorically along Spring/Summer, as well as experimentally along light or dark environments for incubation, as well as solution mixture.

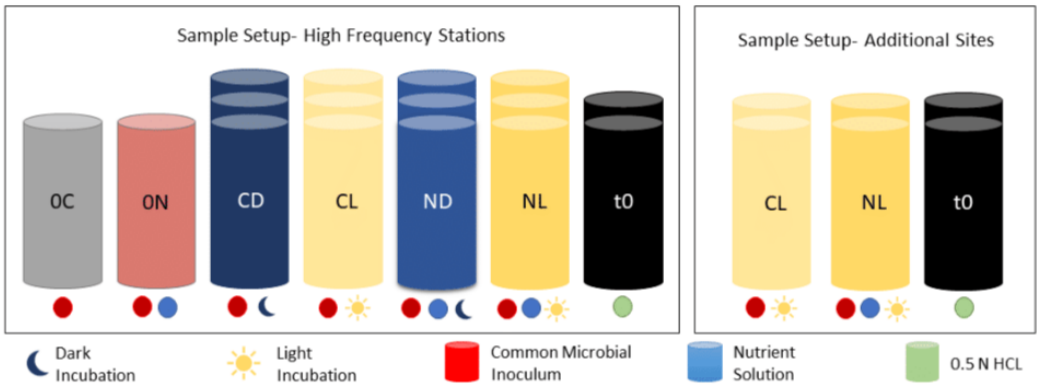

Figure 2.2: The Wildfire experimental setup illustration. Figure courtesy of the Haley Moon research team.

After this period, a large number of optical measurements were taken from each of the samples. Other measurements for each sample were known prior to the experiment, such as the properties of the watershed area that the sample was taken from, as well as the burn level of the same.

The goal of these tests was to evaluate the amount of Biodegradeable Organic Material (hereafter BDOM) in the water. This was measured as a change in organic material over the incubation period. Properties of the water sample were measured in order to determine the health of the ecosystem from which the sample was taken.

Data analysis revolved not just around establishing correlation between different

factors, but in determining if measurements after incubation could have been predicted before incubation. For this reason, numerous characteristic and predictive models were created for determining BDOM, among other factors, based off treatment. Measurements that could have reasonably been known before the experiment was conducted were included as input parameters.

Motivation for constructing predictive models is potentially knowing how land will be affected by a mega-fire. Predictive modeling could inform biologists of potential damage before field measurements have been completed, in order to know where mitigating actions need to be taken to preserve the ecosystem following a fire [10]. It is important for ecologists to understand nutrient availability, as many of Utah's water bodies, including Utah Lake in particular, already suffer from nutrient problems.

To this end, the main result that was desired of our research group was on the effect of the experimental treatment on the differing samples. We also considered a BDOM predictive model to determine if the effects of the experiment could be predicted before the incubation period. This model became the problem formulation for the methodologies below.

For additional details about the biological experimental background, see the following articles on dissolved organic material [5, 10].

2.2 Mechanistic Methodology

Traditional Analysis of Variance (hereafter ANOVA) methods were adequate for answering the questions of interest to the researchers. The following ANOVA table of

comparisons was found. Our choice of significance level was the standard $\alpha = 0.05$, as neither error type seemed catastrophic. The null-hypothesis was that the treatments produced no significant difference in the BDOM, and the alternative hypothesis was that the treatments produced significantly different BDOM measurements. "Reject" would then indicate that the null-hypothesis is rejected.

Table 2.1 shows an ANOVA table comparing treatments pairwise. CL stands for "Control Light", CD for "Control Dark", NL for "Nutrient Light", and ND for "Nutrient Dark".

Treatments Compared	p-value	Reject/Accept
CL-CD	0.0337	Reject
ND-CD	0.8510	Accept
NL-CD	9.79e-14	Reject
ND-CL	0.0370	Reject
NL-CL	2.27e-8	Reject
NL-ND	2.14e-13	Reject

Table 2.1: An ANOVA table comparing treatments.

Similar t-tests found no significant differences between Spring and Summer trials for the CD, CL, and ND trials, but there was a significant difference between the NL trials for summer and spring with a p-value of $1.39e-5$.

Figure 2.3 shows the observed change in BDOM by both trial versus treatment. This is a visualization of the observed quartile ranges, with circles representing outliers. While there were statistically significant differences between trials and treatments, it is clear that the most stark contrast was in a Nutrient Light environment, particularly in the summer.

This answered the main questions of interest to the researchers, but the further

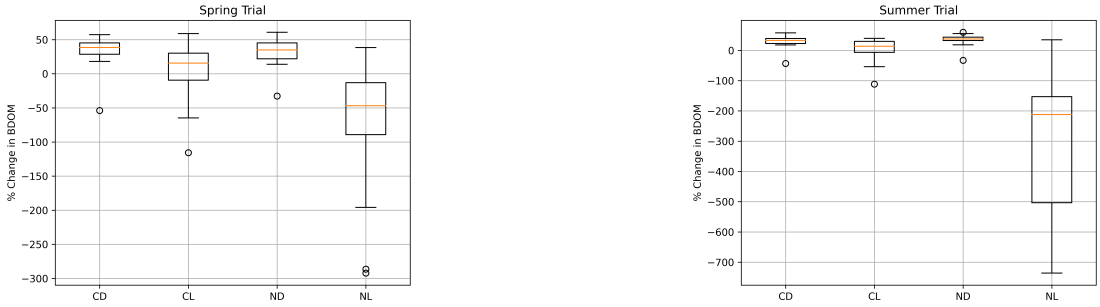

Figure 2.3: The Wildfire experiment BDOM observations.

hope was that perhaps a good predictive model could be made for predicting the change in BDOM based off factors that were known before incubation. For our purposes, the input parameters of our model included time of year, treatment, photographic properties, and properties of the watershed the sample was taken from, such as elevation, slope, etc. Further details on variables can be found in the related works [5, 10]. The output to our model was then the change in measured dissolved organic carbon, or BDOM.

A multiple linear regression model was trained utilizing the Scikit-learn python package. We mainly used 95% confidence intervals as the main predictor. As the results show, we constructed a model that confidently predicted when a "catastrophic" change in BDOM was going to take place. While the model could not accurately predict BDOM, it was capable of indicating when a significant change was going to take place.

2.3 Characteristic Methodology

The Wildfire research was an example of when the mechanistic method better answered the questions of interest than the characteristic method. Because the questions of interest related to significance of treatment, and the relationships studied are relatively straightforward, traditional analysis methods better answered the researchers main questions.

When it comes to predictive modeling, a multiple linear regression model was found to be adequate, but we still constructed a DNN predictive model on the data. After fine-tuning, the result performed inferior to the MLR model for reasons which we discuss below.

The fine-tuned model consisted of 18 inputs discussed in the experimental background. Each layer utilized a Rectified Linear Unit activation function with parameters numbering 128, 64, 16, and 1 for layers 1, 2, 3, and the output layer, respectively. We utilized ADAM for the optimizer with a learning rate fine-tuned at 0.0001. The loss function was the Mean Square Error, and we utilized a batch size of 20 data points training over 30,000 epochs. Training for this number of epochs overfit the model, but the point at which the model overfit was inconsistent, and so validation of the model was implemented on the weights of the model before overfitting as determined by which epoch minimized the loss function.

Figure 2.4 illustrates some of the trouble the DNN had. The Wildfire DNN loss function is on the left. Fine-tuning showed that it converged inconsistently with the number of training epochs before overfitting fluctuated. Its predictions are on the right, and the loss never achieved anything comparable to the mechanistic method.

Multiple iterations of training the DNN, with different fine-tuning of parameters, did not noticeably affect the model.

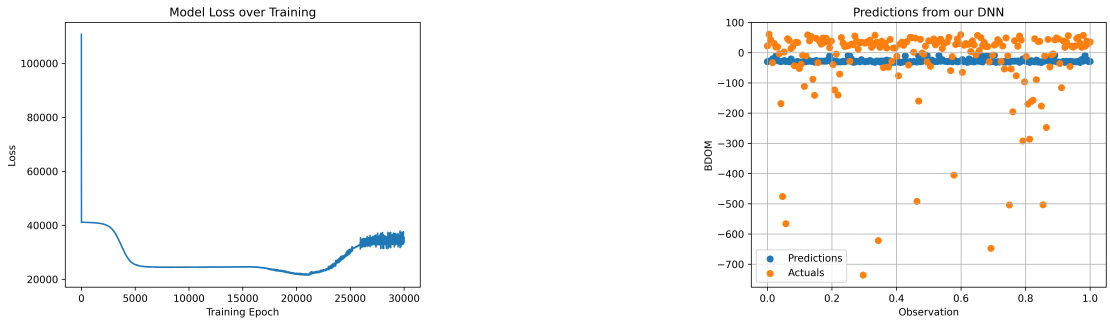

Figure 2.4: The Wildfire DNN loss function is on the left, visualizing loss versus training epoch. The predictions for the most accurate model predictions are on the right.

We consider now some theories as to why this DNN performed so poorly.

The first problem we ran into was quality of data, specifically its imbalance [12] and size [29]. Typically in a machine learning setting, models train on thousands of examples to trillions [16], however we had less than 200 data points. The amount of data required for machine learning depends on the task to perform [29], but it was clear that we did not have enough data for a DNN to explain the underlying relationships between the variables.

Additionally, the data was unbalanced [12]. This is more noticeable with small data sets, but only a quarter of the dataset was in the Nutrient-Light treatment, which experienced the most significant change in BDOM. Since many of these samples were themselves still within range of other treatments, it led to a dataset that was not equally balanced between differing outcomes. This could be remedied with a larger dataset to pull from, but unfortunately no larger dataset is available.

Another problem that was discovered in the DNN architecture was that the loss function never seemed well suited to machine learning optimization. The DNN was not able to reliably arrive at a minimum when training, and so as it trained over epochs, it was continuously fluctuating.

It is possible that there is a scenario wherein a DNN could run a good predictive model on BDOM changes in Utah Lake, but the data we obtained posed significant problems. Potentially, analysis could be conducted in the future utilizing machine learning techniques other than DNNs, or else utilizing differing loss functions.

Chapter 3. Irrigation Research

3.1 Physical Problem Formulation

Many farmers utilize a central pivot irrigation system for watering their fields [23]. Historically, these irrigation systems have utilized a uniform watering distribution, watering all sections of the fields evenly. While this is adequate, the obvious question is if some sections of the field are being watered more than is necessary, and if water could be saved by utilizing a non-uniform watering distribution. Recent technological advancements have made Variable Rate Irrigation (hereafter VRI) systems more common and affordable, increasing the desire for research on how to allocate resources [23].

Unfortunately, VWC sensors remain quite expensive, are relatively fragile, and man-power intensive to operate. This makes the prospect of monitoring soil VWC and watering where appropriate infeasible. However, temperature sensors do not

struggle from any of these drawbacks. BYU, in conjunction with Dr. Neil Hansen, is in the process of synthesizing temperature sensors which are cheap, relatively durable, miniaturized, and which can transmit data wirelessly to a collection center. This motivates research to develop a predictive model for soil VWC from soil temperature.

Figure 3.1 shows what the data initially looked like. The VWC and Temperatures for Site 1 are on the left, plotted against the time in the growing season. On the right, a scatter plot of VWC against Temperature. This plot obviously hints, as first principles affirm, that the relationship between the two is non-linear and indirect, but present [9].

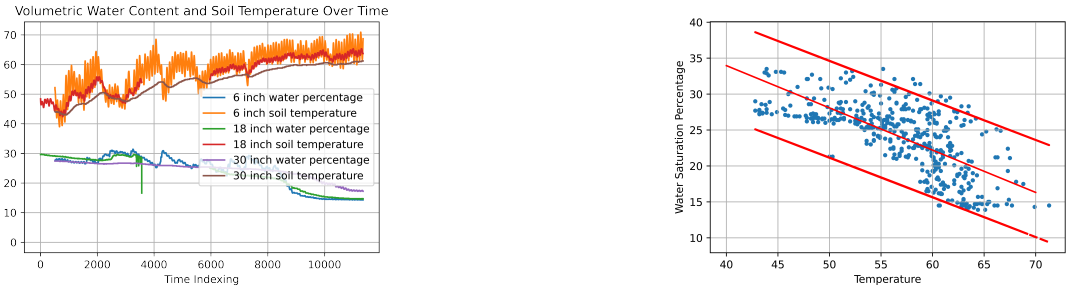

Figure 3.1: The VWC and Temperatures for Site 1 are on the left, indexed chronologically by sample, and a scatter plot of VWC against Temperature on the right. The non-linearity invalidates the assumptions for a 95% confidence interval.

It was presumed going into the experiment that multiple factors would be confounding in a direct statistical analysis [6, 9]. Many of these, however, could be controlled for. For example, while having VWC sensors all throughout a field is infeasible, having a single centralized sensor for relative data comparison is feasible. The ultimate goal is to remove VWC sensors completely. Other confounding factors include soil type, atmospheric temperature and pressure, relative humidity and

precipitation, solar radiation, wind speeds, and others.

Data was collected for use by Dr. Neil Hansen's lab from the BYU Department of Plant and Wildlife Sciences. Research has been done on the physical foundation of this problem [9]. Numerous studies have defined the relationship between soil temperature and VWC as is discussed in the mechanistic methodology below. However, creating a predictive model between the two poses a significant problem.

We formulated the problem from a predictive standpoint of first finding models that would relatively accurately estimate VWC from soil temperature, and then applying these to predict relative spatial VWC distribution with a centralized sensor for catching confounding factors.

Most of the mechanistic methodology below focuses on predicting VWC from soil temperature as our data suited this problem best. Characteristic analysis was conducted on predicting relative spatial VWC distribution as that was the practical application of the research. It is important to note that our dataset was not adequate to accurately solve either problem completely, but we still found many statistically significant results and correlations.

For details about observational agricultural background, see the following articles on predictive methods of VWC for utilizing VRI more effectively [7, 22, 32], and new methods for utilizing satellite imagery in application [30].

3.2 Mechanistic Methodology

Initially we studied the underlying relationships governing soil temperature and VWC. Much work has been done on this in agricultural sciences [9], and lots of

research guided our analysis. Ultimately this did not succeed in producing a reliable predictive model, however we did highlight a lot of interesting correlations, which further research will utilize in determining underlying relationships.

It was assumed from first principles models that temperature is not linearly correlated to VWC, however they are indirectly connected by the heat equation [9]. Water trapped in soil changes its heat capacity and thermal conductivity, which in turn changes its ability to diffuse heat through the soil. This leads to different temperatures for soil samples of different VWC. Utilizing the heat equation for mapping, we let κ be the diffusivity coefficient, U_t be the derivative of the temperature with respect to time, U_{xx} be the second derivative of the temperature with respect to depth of the soil, C be the thermal conductivity, ρ be the density of the soil, and k be the heat capacity. Then assuming the thermal diffusivity constant κ is constant, the equation which we tried to solve was given below:

$$U_t = \kappa U_{xx}.$$

We measured temperature, and tried to identify the diffusivity coefficient and heat capacity. We anticipated a strong correlation between the heat capacity and VWC.

Figure 3.2 illustrates our Data Assimilation generalized method, which is discussed more below. On the left is an example of Vandermonde Polynomial Interpolation to a single day from our dataset. We ultimately found that an 8th-degree interpolation best refined the data, without observing artificial fluctuation, as each day reflected 98 observations. Interpolating the data was necessary for utilizing it as Dirichlet boundary conditions to the heat equation. On the right can be seen a

solved heat equation, giving a temperature map.

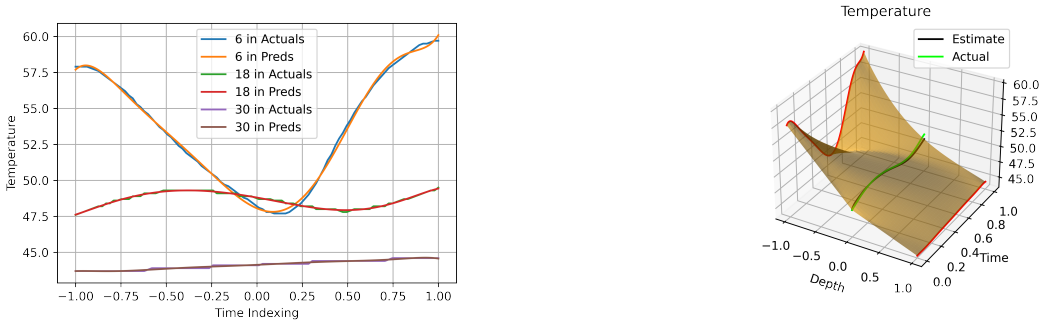

Figure 3.2: On the left is single day polynomial interpolation, on the right we utilize Crank-Nicholson to solve the heat equation.

The first problem was that the data was not precise nor frequent enough for solving the heat equation to the required accuracy. In order to account for this, we used polynomial interpolation to fit a curve to the observed temperatures, and to synthesize data that fit the curve to the frequency that we needed. When utilizing polynomial interpolation in a predictive model, care needs to be taken to avoid extrapolation beyond the dataset [20]. However, we found that the temperature was a continuous non-chaotic variable, and so the synthesized data was reasonable for our purposes, as we did not intend to make extrapolations far outside of the observed data range.

With the synthesized temperature data, we blocked the observations into 24-hour periods. The assumption was that VWC changed relatively slowly in comparison to temperature, and so we assumed a constant diffusivity coefficient over the 24-hour period and utilized the synthesized data as Dirichlet boundary conditions for solving the heat equation. We used a sophisticated implementation of the Crank-Nicholson

[34] method to solve the heat equation rapidly. For each solution to the heat equation, we obtained a diffusivity coefficient κ, and used that to predict intermediary temperature values of the soil. We utilized the Euclidean norm of the difference between the prediction and the actuals as our loss function. We used Scipy's minimize function, with the Levenberg-Marquardt Algorithm [25], to minimize the loss function and obtain the diffusivity coefficient of best fit.

The ultimate goal was feature engineering in hopes that the diffusivity coefficient would correlate to VWC strongly. It did not ever correlate above $r^2 = 0.1$, being statistically insignificant. Possible reasons are the assumptions that the diffusivity coefficient was inversely proportional to VWC or that the diffusivity coefficient was relatively constant with respect to the 24-hour periods of observation.

We did find that the loss function was easily minimized, with no worry that the diffusivity coefficient was difficult to obtain. We also found that the best fit coefficient did not vary chaotically over time, but consistently.

The goal for ongoing research is to be able to isolate the thermal conductivity of a soil sample over a related 24-hour period in order to better get an estimate of the heat capacity in relation to the diffusivity coefficient. In principle, this should improve the estimated relationship.

From the heat equation, we believed that the rate of change of temperature would be more highly related to VWC than temperature itself would. This was also motivated by soil temperatures' dependence on atmospheric temperature, something which is not related to the VWC. We attempted a few different methods to control for this, but did not find any statistically significant correlations.

One common practice in this area of research is to measure the fluctuation of soil temperature over every 24-hour period [6]. This can be thought of as a feature related somewhat qualitatively to the soil heat capacity. This did show significant correlation and we ultimately used this variable in our characteristic approach.

Another common practice is to account for time-serial correlation in the data by filtering out the high frequency change of the temperature. This was again another type of feature engineering. We did this by taking a 24-hour rolling average of each temperature measurement. This did indeed increase correlation significantly over temperature measurement and ultimately this feature did end up in our characteristic approach.

The results of this approach were less than satisfactory. Constructing a multiple linear regression model utilizing any reliably attainable factors did not lead to a predictive model that gave the kind of precision desired. While the model was often comparably accurate, the variability of the differing factors when their relationships were not entirely known led to a high variance in the output of the predictive model. This in turn led to large fluctuating predictions which were often quite far from the actual data. This is what we expected of a MLR model with multiple co-linearities [28].

In a multiple linear regression model, it is common practice to remove variables which are not statistically significant to your model [14], as well as trying to account for co-linearities which confound the model assumptions [24]. However, due to a lack of knowledge of how variables relate directly to the prediction, beyond simple correlation, this is not plausible.

It is important to note that research into the underlying relationships did help to analyze which variables should be utilized in the approach below. Additional research into the underlying relationships of this mechanistic model is still ongoing.

3.3 Characteristic Methodology

For the characteristic method, the motivation was to create a deep neural network that would predict well the VWC from a list of variables which would be likely attainable in application. This meant that temperatures were going to be the main contributor to the neural network, but that other easily attainable factors would be included as well.

It was assumed that, in practice, a central VWC sensor could be placed in a field for controlling confounding factors and calibrating, as could an atmospheric sensor. A model that utilized this central VWC sensor was called an anchored model.

In much of the research into similar DNN models, common factors utilized were mean soil temperature, minimum and maximum soil temperatures, and soil temperature fluctuation over 24 hour periods [6, 2, 4, 7]. Other research often included mean air temperature, precipitation, humidity, atmospheric pressure, and other variables. Most of these are attainable on atmospheric sensors, which can be easily maintained. It was assumed then in application that such factors could be present in the DNN.

The DNN models utilized 18 parameters as inputs, with 3 hidden layers, containing 128, 64, and 32 parameters respectively. All three hidden layers utilized ReLU activation functions. For the regression models, the output layer had one parameter, and for the NULS models, the number of outputs was a fine-tuned parameter

representing the range of possible VWC predictions. All models utilized 0.0001 as the learning rate with the NULS model using a cross entropy loss function and all regression models using the mean square error loss function. Training took place in batches of size 32, over $10,000$ epochs. The largest difference between models was the method of treating the target vector, as discussed above. Predictions of these models were VWC values.

In general, the DNN approach made a model that fluctuated less than both the MLR and the data. It often was marked by a small change when the actual VWC fluctuated more dramatically, and captured well small changes in the VWC. This was contrasted with the MLR model, which would make very dramatic regular fluctuations.

Figure 3.3 illustrates the loss and predictions of the different DNN Models. On the left is the first model DNN Loss function, trained over epochs. The monotonic nature of this curve is what indicates a well-trained DNN [13] and should be compared to the Wildfire DNN in section 3.3. It indicates that the model has achieved a local minimum [13]. On the right is a sample DNN set of predictions on the VWC over time compared to actual observations. After this training, we addressed the question of how the model did in spatial relative VWC comparisons. To rephrase, did it rank the sampling sites from wettest to driest in the correct order a statistically significant portion of the time?

With 4 sites, yielding 24 possible permutations of sites, we would expect a model that behaved poorly in this aspect to yield approximately $\frac{100}{24} \approx 4.1\overline{6}\%$ accuracy. The model obtained an observed accuracy of 43%, which is statistically significant

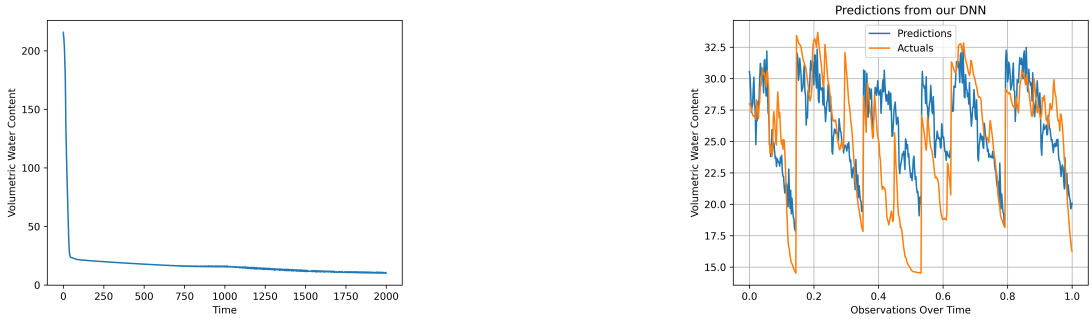

Figure 3.3: On the left is the first model DNN Loss function. On the right is a sample DNN set of predictions on the VWC over time, compared to actual observations.

and shows that the model did indeed rank the data from wettest to driest, to an extent. This is a simple indication study: specifically, this method of spatial relativity ranking does not take into account the margins of inaccuracy, making results indicative, but not conclusive.

Fine tuning was conducted on three distinct machine learning models, the initial regression DNN, the anchored DNN, and the NULS DNN. The initial regression DNN and anchored DNN differed only in that the anchored DNN was given the temperature of one site as one of its input values, mimicking a central atmospheric sensor. The NULS DNN utilized the Non-Uniform Label Smoothed target vectors, with a classification architecture. It is important to note that in a DNN, unlike in a MLR model, uncorrelated factors in your model are often less detrimental as the neurons connecting to them will simply die off [26].

Model 1 : The input was 24 hour sections of the dataset. The variables were max temperature over those periods, min temperature, mean temperature, fluctuation (max minus min), air temperature, humidity, precipitation, atmospheric pres-

sure, wind speed, and date. An MLR was run on these data points first, in order to determine the correlation the model could reasonably attain, measured from its r^2 value. The MLR model had $r^2 = 0.653$. Date was used as a stand in for amount of vegetation over the growing period. All of the temperature values were motivated from the analysis done by others in similar works [2, 6]. Humidity, precipitation, atmospheric pressure, and wind speed were all motivated by both prior works, and first principles.

Model 2 : For the second model, we utilized site 2 as an anchor point. We wanted to test the assumption that a farmer would have a single central atmospheric and VWC sensor, which could be used to help calibrate the DNN. For this DNN, we passed in the maximum, minimum, mean, and fluctuation temperatures of site 2 as well. Site 2 was utilized as it had the most consistent dataset, allowing it to make a good "anchor". These were the only differences between models 1 and 2.

Model 3 : For the third model, we utilized a NULS DNN on the same variables as in model 1. We had to fine-tune as hyper-parameters the length of the categorical variables vectors, as well as the number of categories in that vector, and the standard deviation of the normal distribution utilized in the NULS (as it was always assumed the observation was going to be the mean). Note that the

NULS model did not take site 2 as an anchor.

In order to compare the outputs of these models, we couldn't compare the losses directly, due to differing loss functions. We used the model to predict actual values of the VWC at given dates for the 5 sites they had in common, sites 1, 3, 4, 5 and 6, and then found the 2-norm of the differences between the predicted and the observed VWC values. This allowed for comparison between the models in a way that would be quantitatively analogous.

This was still not a perfect comparison method. The MLR model, for example, did get results for predictions comparable to the NULS DNN as far as error of predictions goes. However, the MLR model had extremely high variation, with wild fluctuation between accurate and inaccurate predictions and many outliers.

The first model performed well and anchoring the model with the site 2 VWC information did improve it, but only marginally. The largest improvement to the DNN came however with the NULS DNN, with the loss being orders of magnitude lower.

Figure 3.4 compares the losses of the differing models. Note the log scale on the y axis. On the right is a comparison of the predictions of the models on site 1, compared to its actual observed value.

Figure 3.5 compares a little more clearly the model predictions to their respective observations. On the left we compare the anchored DNN predictions of VWC with the observations. This did not visibly outperform the initial DNN, much to our

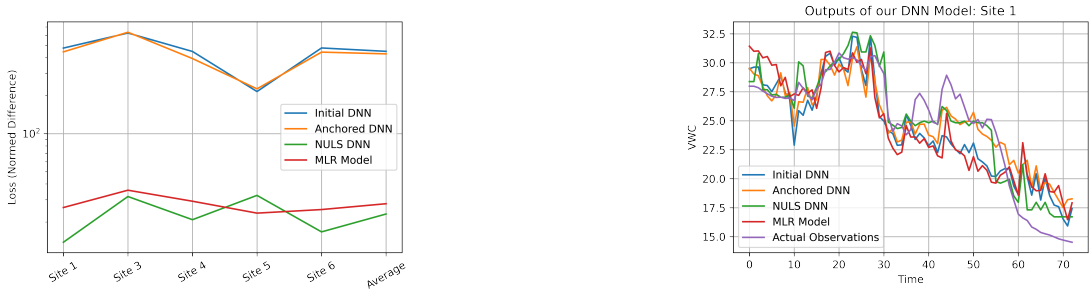

Figure 3.4: On the left is a comparison plot of the loss of the differing models. On the right we compare the model predictions to their observed values.

surprise. On the right we have the NULS DNN. Note that while the NULS is most accurate, it did seem more likely to have outliers in the predictions. This is likely an artifact of unbalanced data [12], and the NULS DNN architecture.

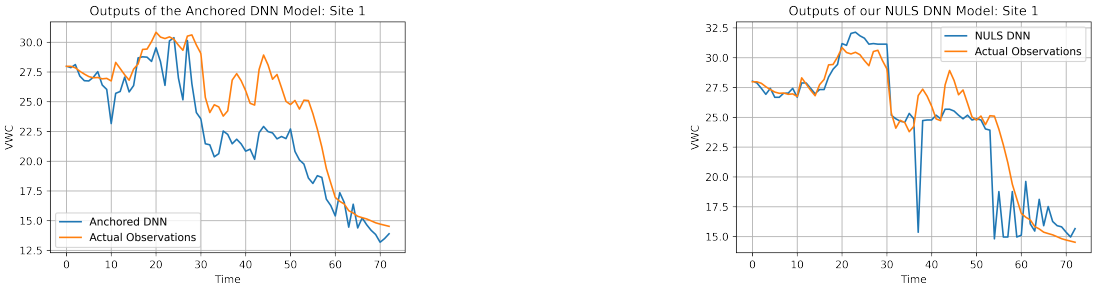

Figure 3.5: On the left we compare the anchored DNN predictions of VWC with the observations, and on the right only the NULS DNN.

The loss is best viewed on a log-scale. The NULS DNN was still sub-optimal. Fine-tuning the hyper parameters helped significantly, but its clear that the DNN behaved as a classification DNN, and not as a regression DNN. Classification DNN's need a much more balanced data set than a regression model, that is, you'd like to have equal amounts of training data for each category you're classifying into. That was not possible for our application which leaves open the possibility that, should

a larger dataset become available, further improvements to the NULS DNN could still be made. This would likely help with some errant classifications seen as large fluctuations in the predictions of the NULS DNN.

The results were, we found it feasible that temperature can be used to predict VWC in soil samples, though more data would likely increase the capability to do so. Generalizing these results to other fields is not statistically validated, given all of the data was from one field. Extrapolation in a characteristic model is even less likely, and so if the results were to be replicated in other environments, it is likely much more data would need to be acquired but that extrapolations could then be made [2].

Chapter 4. Results

4.1 Wildfire Research

We were able to answer all of the researchers' questions of interest using mechanistic methods. Many of the questions on relationships between variables, and so characteristic predictive models would not have been sufficient there regardless.

We found that the MLR predictive models produced a better prediction on the change in BDOM than the DNN models. This is likely due to a number of factors, which are discussed more in section 3.3. However, it is clear that the Wildfire Research did benefit most from the traditional methods of analysis.

Figure 4.1 shows some predictions of how BDOM would change from our MLR predictive model. Its important to note that our goal was not actually to predict

accurately BDOM, but to get a solid indicator for when BDOM would change significantly.

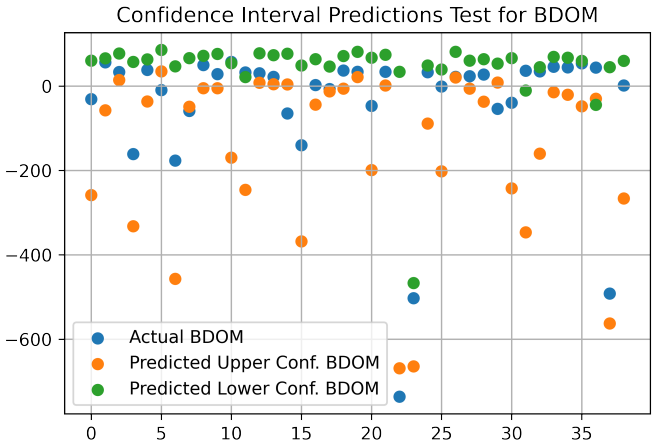

Figure 4.1: The BDOM above is predicted with a 95% confidence interval.

We found that burn did significantly affect BDOM. If we think of BDOM as a measure of change in nutrients through an ecosystem over time, this means that fires do drastically affect BDOM.

The ANOVA table in section 3.2 also indicates that different treatments of a watershed do affect how nutrients flow through the system. The only treatments wherein we failed to reject this null-hypothesis was in the treatment of ND and CD, as they were both incubated in the dark. The treatment that had the most significant impact on BDOM was the NL treatment, as the ANOVA table shows most significant p-values in pairwise comparisons of NL, and as is visually indicated by the box plots of the BDOM by treatment.

Our ability to predict BDOM is significant scientifically as it can indicate where

a drastic change in a watershed environment may occur, to an extent. Its important to note that the samples were not chosen at random, and so the results of this study to not generalize out directly to other watersheds, but the results are certainly indicative of a broader natural phenomenon.

4.2 Irrigation Research

In the Irrigation research, we were not trying to determine underlying relationships between the variables. Most of those have been, to an extent, fleshed out in past studies. We were instead simply trying to build a predictive model from available data, and see how well it could be applied to solve our problem formulation.

We found that the characteristic approach was very capable to solve this problem. In the mechanistic approach, we had hoped to engineer a feature vector (diffusivity coefficients) which would be inversely proportional to the VWC, as first principles models would suggest. However, confounding factors have kept us from getting a strong correlation of these variables yet.

Fortunately, our data was well suited for a characteristic approach. We were able to get solid predictions of VWC from temperature, as shown in section 4.3. Further improvements are highly likely, and while our models are not generalizable to other environments, they indicate that it is likely, given enough data, such a model could be constructed.

Our NULS DNN performed the best on this data set, often getting quite accurate predictions. However, it is clear that our data was not as well conditioned for a classification DNN as it was for a regression DNN. This is something that again

could be remedied with a larger data set.

The NULS DNN did require fine-tuning of the standard deviation as a hyper-parameter, in order to achieve the best results on accuracy versus precision. Accuracy for us was measured as the norm of the difference between the observations and the predictions, and precision was measured as the percentage of predictions inside the 95% confidence interval, as would be indicated by said standard deviation. A numerical analysis of this is given in the results on NULS below.

4.3 Characteristic versus Mechanistic Datasets

The above results indicate that there are clearly times when Characteristic or Mechanistic methods are preferred. The largest factors to take into account in predictive modeling is the type of data available. If the question of interest is on the underlying relationship between variables, than the mechanistic method will be superior. However, in constructing a predictive model, other factors must be taken into account.

Size of the data set is an important factor to consider. We found that the size of the data set was impactful on the quality of the predictions in a characteristic approach, but that it was relatively unimportant in the mechanistic approach. While we did not conduct analysis on any datasets besides those discussed above, outside research seems to validate the claim that generalizability in a DNN is a factor of the size of the data set.

We also found that the dataset was very impactful on the type of DNN that could be constructed. The balance of data in a DNN is important to make sure that the loss function is not "gaming the system", minimizing by skewing relationships, for

example. Data sets that were also very specific do not generalize well, and in a DNN regression model, this may cause high variation, as seen in the NULS DNN.

Additionally, more statistical inference can be made from a mechanistic model than from a characteristic one. Because mechanistic models have been more heavily researched, there are often inferences on the coefficients of these models, with corresponding confidence intervals, p-values, etc. These are of course lacking in a DNN, where such coefficients are often either unknown, or represent non-linear abstract relationships that we do not wish to study.

4.4 Non-Uniform Label Smoothing and its Drawbacks

We were very excited about the results of the NULS DNN. Non-Uniform Label Smoothing has never, to our knowledge, been used in a regression context such as this before, and the fact that the results were not only comparable, but exceeded those of a traditional DNN motivate further research.

This is only further motivated by the fact that typical classification DNN's, such as a NULS DNN, are overly susceptible to unbalanced data, such as those we received. Its possible then that our NULS DNN could be further improved if a data set where a balance of those ordinal categorical responses we're interested in were available.

The tradeoff for such a balance would be "throwing out" data, in order to train your model, balancing the data often requires randomly selecting samples from sets, to ensure comparable size. This is not something however that can be tested on our data, due to size constraints.

It is further interesting to note that our original reason for attempting the NULS

DNN was to obtain some sort of uncertainty quantification for our predictions. This was validated after training the model. It is clear from the results that there were times when the model simply underperformed, and for those sites the uncertainty quantification was inaccurate. However, in 3 of the 6 sites, the 95% confidence interval was found to be accurate.

Figure 4.2 shows most clearly the results of our differing NULS models. We ran the NULS DNN for a range of standard deviation hyper parameters, as part of the fine-tuning process. We compare on the left the different accuracies of each model (as measured by the 2-norm of the difference between the predictions and observations), and on the right we compare the differing precisions of each model (as measured by a percentage of observations inside the predicted 95% confidence interval).

Figure 4.2: On the left we measure accuracy, and on the right we measure precision.

4.5 Discussion

It is the assertion of this project that the modern Data Analyst will need to be able to utilize both traditional mechanistic and modern characteristic tools to handle today's dynamic problems.

An ever increasing abundance of data, combined with novel research into Deep Neural Network architectures, means that more and more research will be done with exciting predictive models. However, its clear from the results above that there will always be a time and a place for the mechanistic approach.

As for the novel Non-Uniform Label-Smoothing DNN applied to a regression problem, we don't propose to have completely researched this model. We only applied this to our datasets, and found it to be superior. However, we did observe strange variations in the data which suggest that, if given a larger dataset, this model could be improved.

Its also important to note that we did not study the generalizability of such a method. We would like to believe that our method could create a model which, given the right dataset, could give a generalizable method for analysis of soil versus water with a specific uncertainty quantification, but we really only breached the research on this to prove feasibility.

This motivates further research into NULS DNN's, to see if truly well balanced large datasets could be used for extremely accurate regression solutions, and to see if they really could be the answer to highly efficient Variable Rate Irrigation systems.

Bibliography

[1] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul W. Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges, arXiv:2011:06225. *CoRR*, abs/2011.06225, 2020.

[2] Umesh Acharya, Aaron L. M. Daigh, and Peter G. Oduor. Machine learning for predicting field soil moisture using soil, crop, and nearby weather station data in the red river valley of the north, Journal of Soil Systems, https://www.mdpi.com/2571-8789/5/4/57. *Soil Systems*, 5(4), 2021.

[3] Riadh Kobi José Dolz Hervé Lombaert Ismail ben Ayed Adrian Galdran, Jihed Chelbi and Hadi Chakor. Non-uniform label smoothing for diabetic retinopathy grading from retinal fundus images with deep neural networks, NIH, https://pubmed.ncbi.nlm.nih.gov/32832207/. 2020.

[4] Elliott R.L. Stevens E.W. Altendorf, C.T. and M.L. Stone. Development and validation of a neural network model for soil water content prediction with comparison to regression techniques, ASABE, https://elibrary.asabe.org/abstract.asp??jid=3aid=13230. 1999.

[5] Jeremy B. Jones Jr. William B. Bowden Benjamin W. Abbott, Julia R. Larouche and Andrew W. Balser. Elevated dissolved organic carbon biodegradability from thawing and collapsing permafrost, AGU, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014jg002678. 2014.

[6] Yu Cai, Wengang Zheng, Xin Zhang, Lili Zhangzhong, and Xuzhang Xue. Research on soil moisture prediction model based on deep learning, PLOS ONE, https://doi.org/10.1371/journal.pone.0214508. *PLOS ONE*, 14(4):1–19, 04 2019.

[7] Yu Cai, Wengang Zheng, Xin Zhang, Lili Zhangzhong, and Xuzhang Xue. Research on soil moisture prediction model based on deep learning, PLOS ONE, https://doi.org/10.1371/journal.pone.0214508. *PLOS ONE*, 14(4):1–19, 04 2019.

[8] Pratik Chaudhari. Principles of deep learning, https://pratikac.github.io/. 2022.

[9] Jianzhi Dong, Susan C. Steele-Dunne, Tyson E. Ochsner, and Nick van de Giesen. Determining soil moisture by assimilating soil temperature measurements using the Ensemble Kalman Filter, Advances in Water Resources, https://www.sciencedirect.com/science/article/pii/s0309170815001979. *Advances in Water Resources*, 86:340–353, 2015. Data assimilation for improved predictions of integrated terrestrial systems.

[10] Scott Zolkos Sadie Textor Stephanie Ewing Jane Klassen Robert G. M. Spencer David C. Podgorski Suzanne E. Tank Michelle A. Baker Jonathan A. O'Donnell Kimberly P. Wickland Sydney S. W. Foks Jay P. Zarnetske Joseph Lee-Cullin Futing Liu Yuanhe Yang Pirkko Kortelainen Jaana Kolehmainen Joshua F. Dean Jorien E. Vonk Robert M. Holmes Gilles Pinay Michaela M. Powell Jansen Howe Rebecca J. Frei Samuel P. Bratsman Ethan Wologo, Sarah Shakil and Benjamin W. Abbott. Stream dissolved organic matter in permafrost regions shows surprising compositional similarities but negative priming and nutrient effects, NIH, https://www.ncbi.nlm.nih.gov/pmc/articles/pmc7816262/. 2020.

[11] Shuo Feng, Huiyu Zhou, and Hongbiao Dong. Using deep neural network with small dataset to predict material defects, Science Direct, https://www.sciencedirect.com/science/article/pii/s0264127518308682. *Materials Design*, 162:300–310, 2019.

[12] Yang Feng, Min Zhou, and Xin Tong. Imbalanced classification: a paradigm-based review, arXiv:2002.04592. 2021.

[13] Timur Garipov, Pavel Izmailov, Dmitrii Podoprikhin, Dmitry Vetrov, and Andrew Gordon Wilson. Loss surfaces, mode connectivity, and fast ensembling of dnns, arXiv:1802:10026. 2018.

[14] Karen Grace-Martin. When to leave insignificant effects in a model, The Analysis Factor, https://www.theanalysisfactor.com/insignificant-effects-in-model/.

[15] Shuyue Guan and Murray Loew. Analysis of generalizability of deep neural networks based on the complexity of decision boundary, international conference on machine learning and applications, 10.1109/icmla51294.2020.00025. Dec 2020.

[16] Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George

van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, and Laurent Sifre. Training compute-optimal large language models, arXiv:2203:15556. 2022.

[17] Sai Nikhilesh Kasturi. Underfitting and overfitting in machine learning and how to deal with it, Towards Data Science, https://towardsdatascience.com/underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6fe4a8a49dbf.

[18] Nadja Klein, David J. Nott, and Michael Stanley Smith. Marginally calibrated deep distributional regression, arxiv:1908:09482. *Journal of Computational and Graphical Statistics*, 30(2):467–483, Sep 2020.

[19] J. R. Larouche, B. W. Abbott, W. B. Bowden, and J. B. Jones. The role of watershed characteristics, permafrost thaw, and wildfire on dissolved organic carbon biodegradability and water chemistry in arctic headwater streams, Journal of Biogeosciences, https://bg.copernicus.org/articles/12/4221/2015/. *Biogeosciences*, 12(14):4221–4233, 2015.

[20] Vijaysinh Lendave. An illustrative guide to extrapolation in machine learning, Analaytics India, https://analyticsindiamag.com/an-illustrative-guide-to-extrapolation-in-machine-learning/.

[21] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, arXiv:1705:07874. 2017.

[22] Jasmine Neupane and Wenxuan Guo. Agronomic basis and strategies for precision water management: A review, Agronomy, https://www.mdpi.com/2073-4395/9/2/87. *Agronomy*, 9(2), 2019.

[23] Evett S.R. Colaizzi P.D. Andrade M.A. Marek T.H. Heeren D.M. Lamm F.R. O'Shaughnessy, S.A. and J.L. LaRue. Identifying advantages and disadvantages of variable rate irrigation, Applied Engineering in Agriculture, https://digitalcommons.unl.edu/biosysengfacpub/631/. 2019.

[24] Vishwa Pardeshi. Handling multi-collinearity in ML models, Towards Data Science, https://towardsdatascience.com/handling-multi-collinearity-6579eb99fd81. 2020.

[25] Omead Pooladzandi and Yiming Zhou. Improving Levenberg-Marquardt algorithm for neural networks, arXiv:2212:08769. 2022.

[26] Luthfi Ramadhan. Neural network: The dead neuron, Towards Data Science, https://towardsdatascience.com/neural-network-the-dead-neuron-eaa92e575748.

[27] Murali Rohana. The role of data analytics and machine learning in digital transformation, LinkedIn, https://www.linkedin.com/pulse/role-data-analytics-machine-learning-digital-murali-rohana/. 2023.

[28] Nathan Rosidi. A beginner's guide to collinearity: What it is and how it affects our regression model, Towards Data Science, https://towardsdatascience.com/a-beginners-guide-to-collinearity-what-it-is-and-how-it-affects-our-regression-model-d442b421ff95. 2021.

[29] Neil Slater. What is the approx minimum size of dataset required to build 90 percent correct model?, Stack Exchange, https://datascience.stackexchange.com/questions/63721/what-is-the-approx-minimum-size-of-dataset-required-to-build-90-correct-model. 2019.

[30] Ryan Smith, Lindi Oyler, Colin Campbell, Elisa A. Woolley, Bryan G. Hopkins, Ruth Kerry, and Neil C. Hansen. A new approach for estimating and delineating within-field crop water stress zones with satellite imagery, International Journal of Remote Sensing, https://doi.org/10.1080/01431161.2021.1931536. *International Journal of Remote Sensing*, 42(16):6003–6022, 2021.

[31] Craig N Spencer, Kristin Odney Gabel, and F.Richard Hauer. Wildfire effects on stream food webs and nutrient dynamics in Glacier National Park, USA, Forest Ecology and Management, https://www.sciencedirect.com/science/article/pii/s0378112703000586. *Forest Ecology and Management*, 178(1):141–153, 2003. The Effect of Wildland Fire on Aquatic Ecosystems in the Western USA.

[32] Jeffrey D. Svedin, Ruth Kerry, Neil C. Hansen, and Bryan G. Hopkins. Identifying within-field spatial and temporal crop water stress to conserve irrigation resources with variable-rate irrigation, Agronomy, https://www.mdpi.com/2073-4395/11/7/1377. *Agronomy*, 11(7), 2021.

[33] Jimin Tan, Jianan Yang, Sai Wu, Gang Chen, and Jake Zhao. A critical look at the current train/test split in machine learning, arXiv:2106:04525. 2021.

[34] Georg Walther. The Crank-Nicolson method implemented from scratch in Python, https://georg.io/2013/12/03/crank$_n$*icolson*. 2013.

[35] Chang-Bin Zhang, Peng-Tao Jiang, Qibin Hou, Yunchao Wei, Qi Han, Zhen Li, and Ming-Ming Cheng. Delving deep into label smoothing, arXiv:2011:12562. *CoRR*, abs/2011.12562, 2020.

[36] Zhilu Zhang and Mert R. Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels, arXiv:1805:07836. 2018.