Estimating Soil Saturation from Temperature

Non-Linear Curvefitting, Solving the Heat Equation, Correlation of Water Saturation to Heat Dissipation

James Tyler, Damian Anderson, Jake Murphy

Theory of Predictive Modeling

CS 580/PHSCS 513R

Background:

Agriculture is a major economic component. Watering costs affect agricultural development. We look to reduce the amount of water needed to grow crops by associating water saturation and temperature.
One problem is water saturation sensors are very expensive, manpower intensive, and relatively fragile. Thus the cost of monitoring water saturation outstrips any possible financial benefit.
Another problem is that most equipment used to water fields doesn't allow for varying irrigation across the field.
One possible solution to both of these problems might be using temperature sensors, instead of water saturation sensors, to determine where in the field more water is needed. Unlike water saturation sensors, temperature sensors can be miniaturized cost effectively, and then spread along with seeds in the soil. Such sensors are already in production here at BYU.
We have both water saturation and soil temperature datasets for several center pivot irrigation systems in Idaho. Our goal is to run a feasibility analysis using the data, to see if a strong relative correlation between temperature and water saturation can be found.

Figure 1: A center pivot irrigation system. This one does not allow for variable irrigation by regions of the field.

Problem Statement:

Our goal is to study the relationship between soil temperature and soil water saturation, to see if thermometers can be used in place of water saturation sensors in determining where in a field to water.
The purpose of this project was not to establish a direct method for doing so, but to test the feasibility using water saturation and temperature data already collected.
Ultimately, we had 3 problem statements that were well posed.
1 - We wanted to find a function to represent our day data-sets that minimized the residuals to our actual observed data.
2 - We wanted to solve the heat equation for differing dissipation coefficients that would minimize the predicted and observed 18 inch depth temperatures.
3 - We wanted to study the correlation between dissipation coefficients and the observed water saturation.

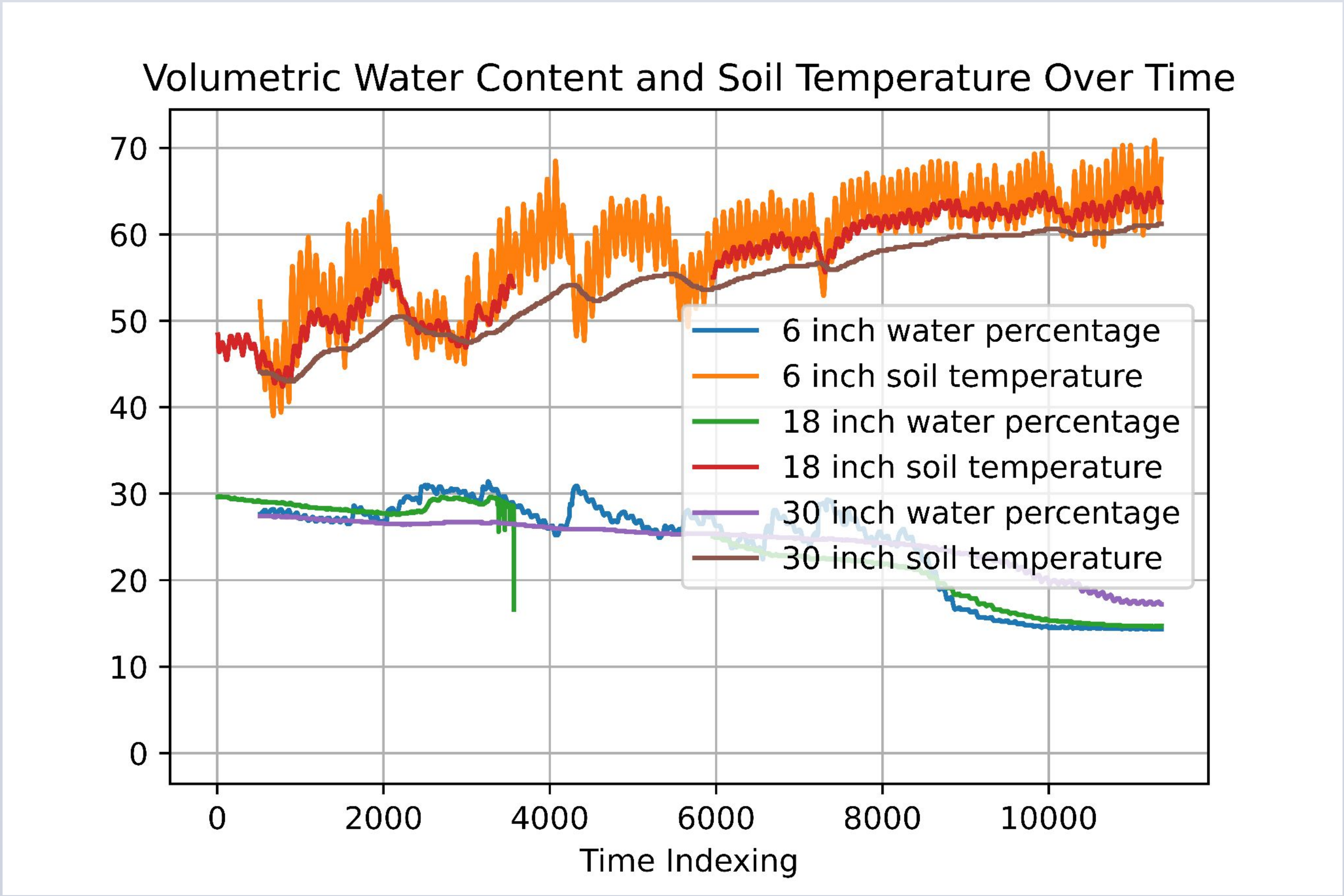


Figure 2: Water Saturation versus Soil Temperature over time of one location for our field in Idaho. Taken over a 3 month period.

Assumptions:

1 - Temperature is not directly related to water saturation, but rather water saturation affects the heat capacity of the soil. This implies that water saturation is related to the change in temperature. From research into first principles, the heat equation was appropriate:

$$\frac{dT}{dt} = \kappa \frac{d^2T}{dx^2}$$

where we assume that the water saturation is related to the dissipation coefficient, x is the depth of the soil sample, t is the time, and T is the temperature.

2 - We assumed that the water saturation didn't vary with respect to time or depth. These are both very clearly false. However, the water saturation varies much less with respect to time than temperature does, making it a safe assumption for our purposes. However, the water saturation does vary more with respect to depth. We made the assumption that it doesn't vary with respect to depth, as it greatly decreased the computational cost of minimization.

3 - We assumed that the dissipation coefficient was linearly related to the water saturation, and most of our methods are attempting to back up this assumption. This was suggested by an agricultural domain expert.

4 - We assumed uniform quality of soil throughout the data measurement. This is a bad assumption, as soil erosion and weathering, as well as natural growing processes, can change the type of soil being measured, which in turn changes the amount of water that the soil can hold, as well as the change in water saturation of the soil.

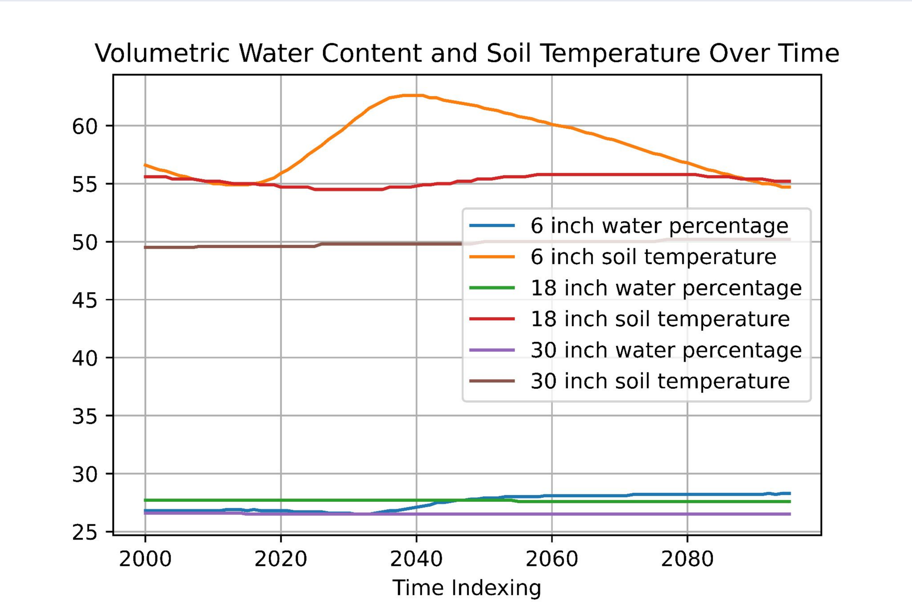


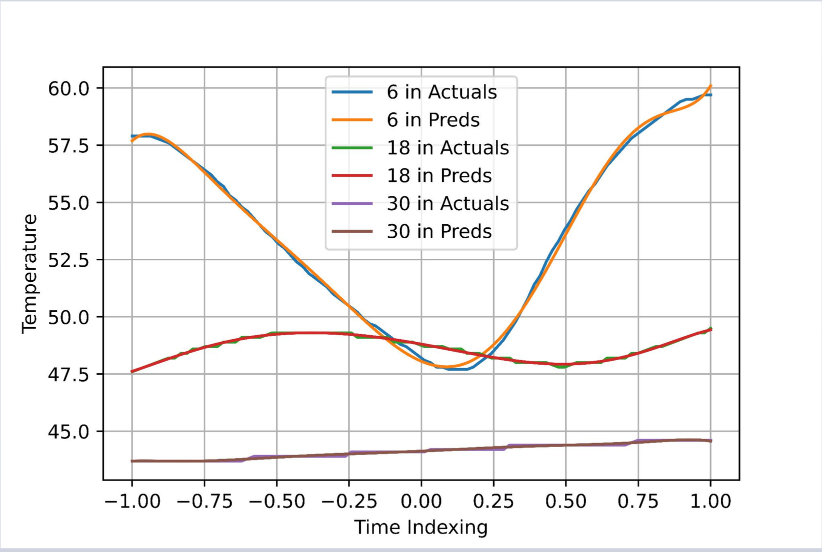
Figure 3: Water Saturation and Temperature over a single day data-set.

Methods:

1 - Data Pre-Processing: The data required relatively little pre-processing. We utilized Pandas in Python for importing the data to a Python Jupyter notebook. We created new dataframes with all the relevant variables, and removed any incomplete data points. We did no other normalization, aside from assuming all time indices were between -1 and 1.
We then broke down the data into 24 hour segments, each containing 96 data points, all being 15 minutes apart.

2 - Interpolative Curve Fitting: We wanted functions for interpolating of our relatively coarse data. The precision of the temperatures was not great enough for our purposes, and so we needed a function to interpolate the data for two reasons. First was that we needed a finer dataset in our solution algorithm for the heat equation. Second was that we wanted a more precise estimate to the instantaneous rate of change of the temperature.
This essentially involved finding a function that fit our data relatively well, and then synthesizing temperature data from that function.

i- We first attempted to fit a non-linear curve-fitting using a combination of the sine and cosine function, minimizing over variables. This was done at the advice of a domain expert in agriculture. However, minimization was not able to reliably able to get an accurate interpolation function. Sympy Optimization Minimization was used.



ii- We next attempted polynomial curve-fitting, with both Monomial and Legendre bases. This was not our preferred attempt, as polynomial curve-fitting often results in overfitting, and very poor extrapolation, due to the behavior of polynomial tails. However, we eventually decided that would not be a problem for this study, as we only wanted our data for interpolation, and we had enough data points for every 24 hour period that we could get an accurate curve while not over-fitting.

Figure 4: Degree 8 Polynomial Curve Fitting to Temperature Over Time

Methods (Cont):

3 - Curve Fitting the Dissipation Constant and Solving the Heat Equation: Once we had refined data sets, we utilized the Crank Nicholson algorithm to solve our differential equation. We utilized the 6 inch and 30 inch depths of our dataset as Dirichlet Boundary Conditions, and then were able to minimize the difference in our predicted 18 inch depth and our observed 18 inch depth, by changing our dissipation coefficient. This gives us an objective function to minimize:

$$\min_{\kappa} ||y - \hat{y}||$$

Where k is our dissipation coefficient, y is our observed middle temperature data, and y hat is the middle temperature data predicted from our heat equation solution.
We mapped the cost for differing dissipation coefficients of a few day data-sets and, noting similarities, decided there seemed to be a unique solution to the problem statement above for each day data-set.
We began the process of finding the dissipation coefficient that solved the problem statement above for every day data-set. Unfortunately, the computational cost of solving the above statement means that there is still a lot of day data-sets that we've not incorporated yet, see ongoing studies.

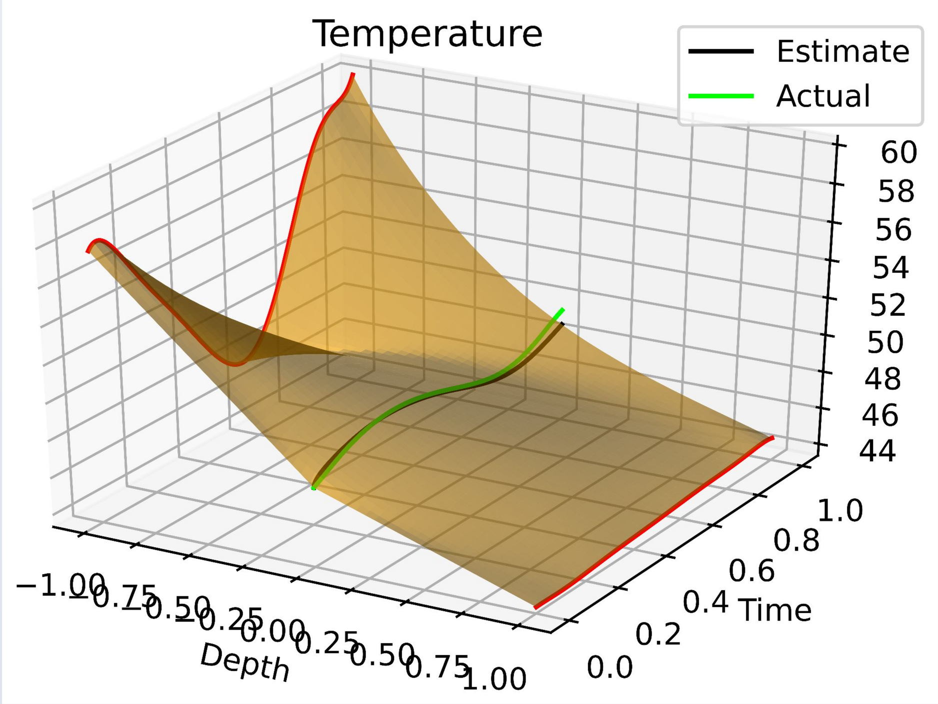
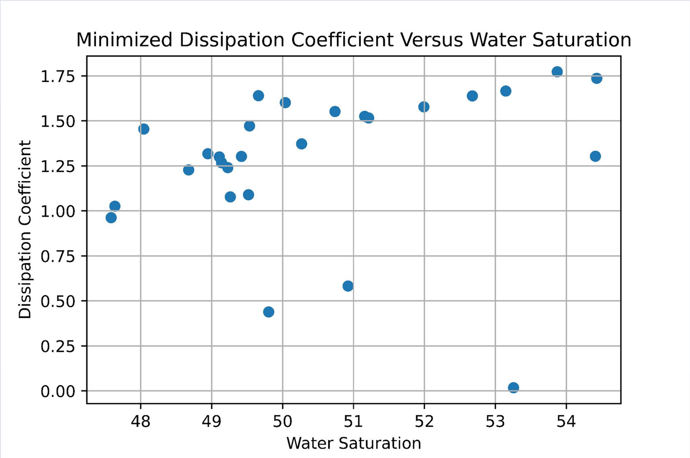


Figure 5: Solved PDE for a minimized dissipation coefficient. The red and green cross sections are observed data, the black cross section was used in the objective function as the predicted solution to the pde.
Depth and time have been normalized.

4 - Correlation with Water Saturation: We compared the observed dissipation coefficients that minimized our problem statement to the average water saturation rates over the corresponding days, checking for any correlation that would imply that water saturation is linearly related to the rate of heat dissipation.

Results:

Polynomial Interpolation to curve fit the temperature worked reliably well for the study.
Solving the PDE, and minimizing the dissipation coefficient, seems to be computationally expensive, but promising.
There does appear to be a correlation between the dissipation coefficient we've minimized and the mean water saturation, when excluding outliers.



Ongoing Study:

This is just a preliminary feasibility study of the data we've been given. Its possible by broadening our assumptions, we could get a better correlation. Specifically, we'd like to account for the water saturation changing with regards to depth. Additionally, we have atmospheric data that could allow us to control for confounding variables, such as atmospheric temperature, precipitation, and irrigation.
Ultimately, our goal is not just to get a relationship between temperature and water saturation with data we have, but to be able to extrapolate on new soil sources water saturation from temperature, despite water saturation sensors never having been used.
Its also possible that we can develop a formula for classification of soil types based off our soil temperature and water saturation, as its assumed that differing soil types are affected differently by water saturation.

Bibliography:

Figure 1: https://theconversation.com/a-water-strategy-for-the-parched-west-have-cities-pay-farmers-to-install-more-efficient-irrigation-systems-185820
Determining the Heat Equation in Soil Samples: https://www.sciencedirect.com/science/article/abs/pii/S0309170815001979
A link to the Google Colab where this work is done: https://colab.research.google.com/drive/1qWeJuGJPwuHgauWKLiRe-jFp3GGiVHlW?usp=sharing

www.PosterPresentations.com