A Functional Anova Approach to Detecting Changes in Soil Moisture and Temperature

Manju M. Johny
Major Professor: Petruţa Caragea

May 10, 2017

Outline

Motivation

- Climate change poses significant challenges to soil ecosystem.
- Extensive research on this topic has been done by Dr. Diane Debinski's lab.
    - Montane Meadows in Grand Teton National Park, WY
    - Early snow melt simulated through snow removal
    - Warmer temperatures simulated through passive heating chambers
    - Collected soil moisture and temperature at 25 cm and 5 cm depths respectively
- Interested in studying effect of simulated climate change on soil moisture and temperature.

Outline

Data

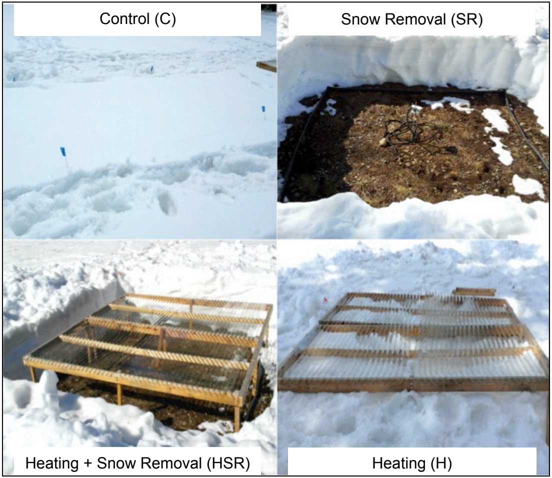
- Data collected by Dr. Diane Debinski's lab
- 4 treatments
  1. Control (C)
  2. Heating (H)
  3. Heating + Snow Removal (HSR)
  4. Snow Removal (SR)
- 3 replicates (east, center, west) established in plots for each treatment.
- Soil moisture and temperature measurements obtained hourly from May 27, 2011 to September 27, 2011

Data

- Data collected by Dr. Diane Debinski's lab
- 4 treatments
  1. Control (C)
  2. Heating (H)
  3. Heating + Snow Removal (HSR)
  4. Snow Removal (SR)
- 3 replicates (east, center, west) established in plots for each treatment.
- Soil moisture and temperature measurements obtained hourly from May 27, 2011 to September 27, 2011
- **Question: Are the treatments different?**

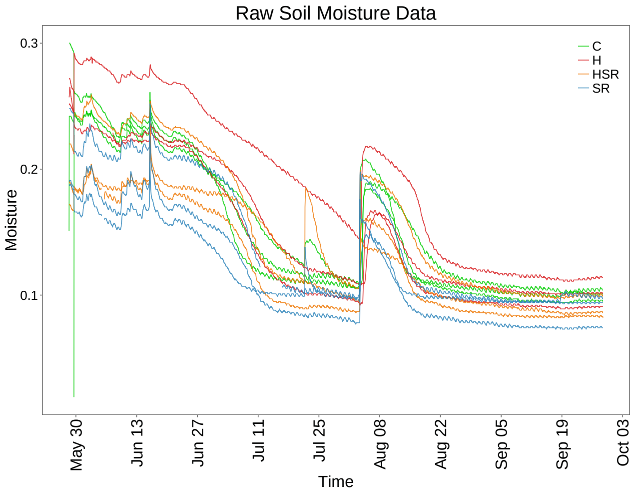
Treatments

Figure 1: Experimental Design



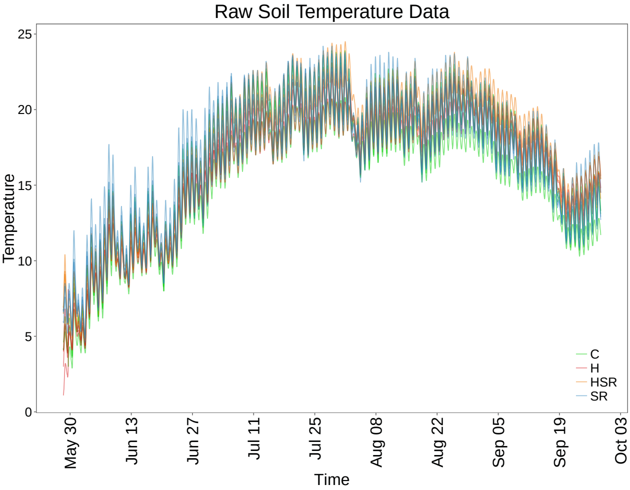
Data: Soil Moisture

Figure 2: Time series measurements of soil moisture for all treatments



Data: Soil Temperature

Figure 3: Time series measurements of soil temperature for all treatments



Considerations for Analysis

1. Data has temporal dependence, and possible correlation among series
2. Classical one-way anova inadequate
3. Previous work: Non-parametric permutation test (Sherwood et al.)
    - Utilizes features of the entire sequence
    - p-value limited for small number of series

Considerations for Analysis

1. Data has temporal dependence, and possible correlation among series
2. Classical one-way anova inadequate
3. Previous work: Non-parametric permutation test (Sherwood et al.)
   - Utilizes features of the entire sequence
   - p-value limited for small number of series
4. Our Approach: Functional Anova
   - Considers each series as a continuous function
   - Extends the classical one-way anova approach to functional data
   - Bootstrap procedure allows for visualization of uncertainty in differences
   - p-value obtained through bootstrap and Monte Carlo

Outline

Data Pre-Processing – Smoothing

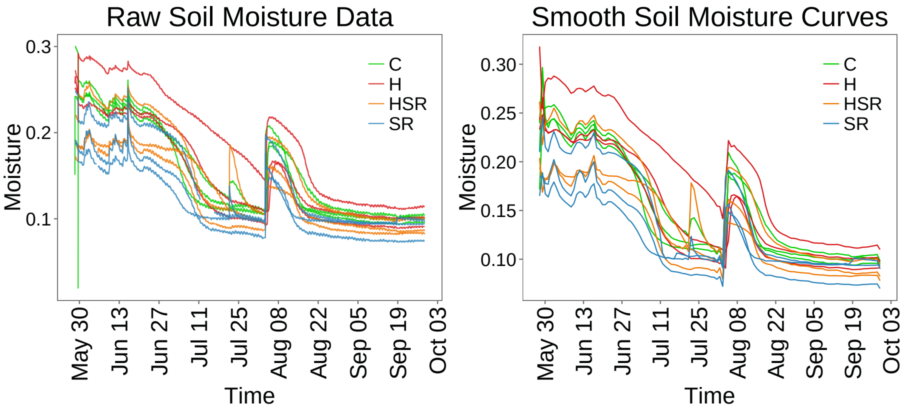
- Goal: Turn discrete data into functions
- Smooth data using basis functions

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t) \tag{1}$$

  where ϕ_k is the k^{th} basis function
- Fourier bases used for cyclical data (temperature)
- B-spline bases used for non-cyclical data (moisture)
- Optimal number of basis, K, obtained by minimizing GCV criterion

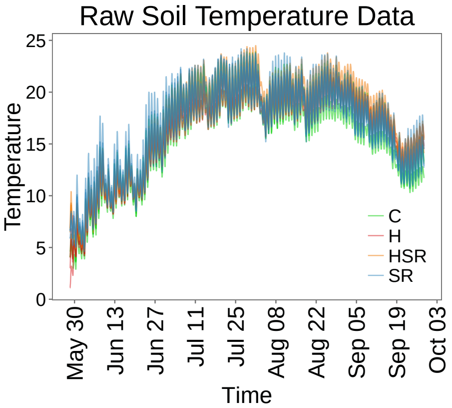
Smooth Moisture Curves

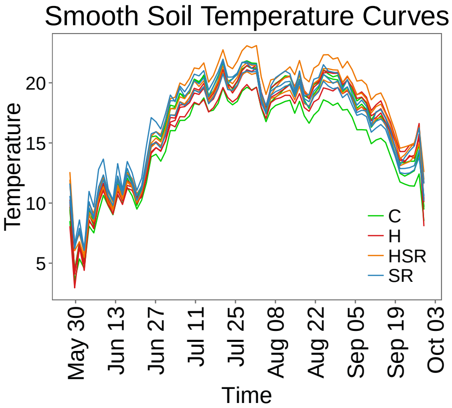
Figure 4: Moisture data smoothed using 120 B-spline bases functions of 3^{rd} order



Smooth Temperature Curves

Figure 5: Temperature data smoothed using 75 Fourier bases functions





Outline

Functional Anova Framework

For $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$,

- $x_{ij}(t)$ is the j^{th} sample curve from i^{th} group
- x_{ij} generated from an L_2-process X_i having mean μ_i and covariance $K_i(s,t) = Cov(X_i(s), X_i(t))$.
- $\bar{x}_{i.}(t)$ is sample mean curve of i^{th} group
- $\hat{K}_i(s,t)$ is sample covariance of i^{th} group

To test equality of mean curves between m independent groups

$$H_0 : \mu_1(t) = \mu_2(t) = \ldots = \mu_m(t) \tag{2}$$

FANOVA Procedure

Test Statistic

$$T_n = \sum_{i<i'}^{m} \|\bar{x}_{i.}(t) - \bar{x}_{i'.}(t)\|^2 \tag{3}$$

where $\|.\|$ represents the L_2-norm.

- $\bar{x}_{i.}$ are rediscretized into vectors of length K
- $\|.\|$ approximated by euclidean distance

FANOVA Procedure

Reference Distribution of T_n under H_0

For each i^{th} group and N=2000 simulations,

1. Generate n_i curves from a Gaussian distribution with mean 0 and covariance $\hat{K}_i(s,t)$.
2. Obtain the mean curve from the previous step. This is a resample curve, Z_i^*.
3. For $l = 1, \ldots, N$, simulate Z_{il}^*
4. For $l = 1, \ldots, N$, $T_l = \sum_{i<i'}^{m} \|Z_{il}^* - Z_{i'l}^*\|^2$
5. p-value $= \frac{1}{N}\sum_{l=1}^{N} I(T_l > T_n)$

Outline

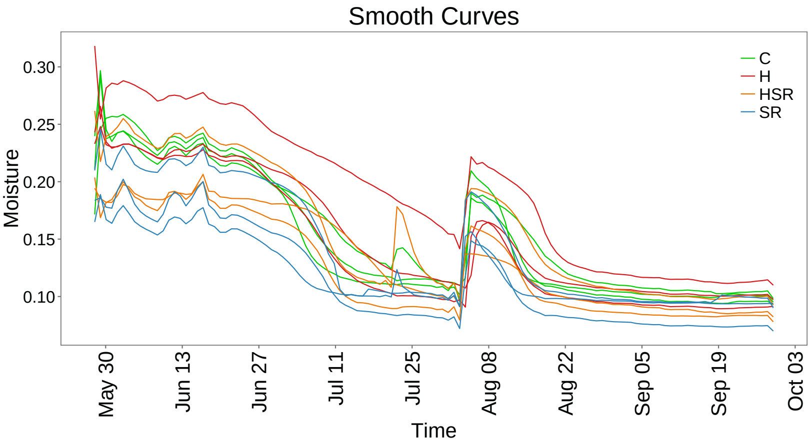
Results

Conduct 3 fanova tests on moisture and temperature data

1. Test for difference between the treatments (4 groups)
   1. C
   2. H
   3. HSR
   4. S
2. Test for difference due to snow removal (2 groups)
   1. $\tilde{\mathrm{S}}$ = 6 curves from SR and HSR
   2. $\mathrm{N}\tilde{\mathrm{S}}$ = 6 curves from C and H
3. Test for difference due to heating (2 groups)
   1. $\tilde{\mathrm{H}}$ = 6 curves from H and HSR
   2. $\mathrm{N}\tilde{\mathrm{H}}$ = 6 curves from C and SR

Moisture - All Treatments

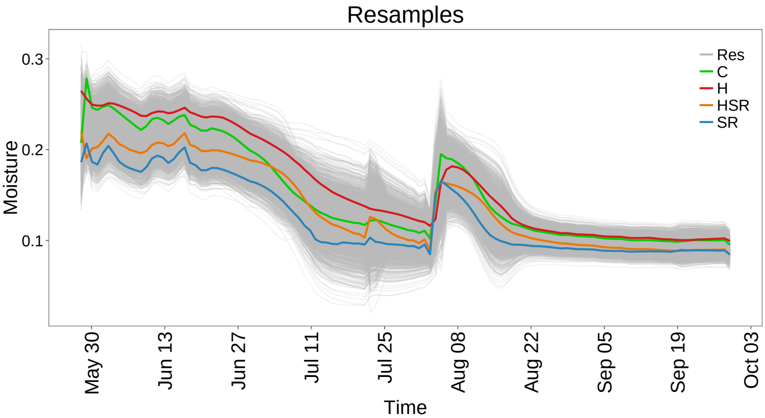
Figure 6: Test for Difference between 4 treatments - Smooth Moisture Curves



Moisture - All Treatments

- 2000 resample curves (Z^*_{il}) simulated under H_0 shown in grey
- Sample mean curves of 4 treatment groups shown in color

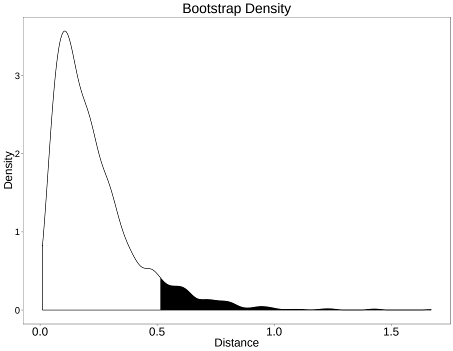
Figure 7: Test for Difference between 4 treatments - Resample Curves



Moisture - All Treatments

- Reference distribution obtained under H_0 shown below.
- Test Statistic $= T_n = 0.513$
- p -value $= 0.074$

Figure 8: Test for Difference between 4 treatments - Bootstrap Density



Moisture – Remaining Tests

Figure 9: Difference Due to Snow Removal

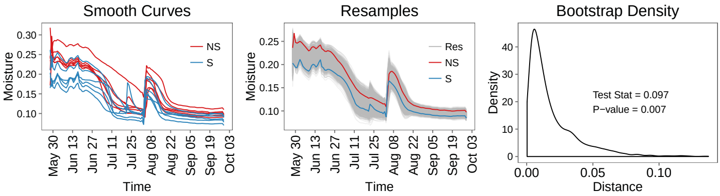
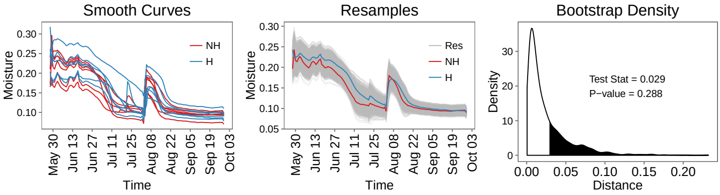


Figure 10: Difference Due to Heating



Summary of Moisture and Temperature Fanova Tests

- Table gives p-values for the following soil moisture and temperature fanova tests:
  - Test for difference between all treatments (C−H−HSR−SR)
  - Test for difference due to snow removal ($\tilde{\text{S}} - \text{N}\tilde{\text{S}}$)
  - Test for difference due to heating ($\tilde{\text{H}} - \text{N}\tilde{\text{H}}$)

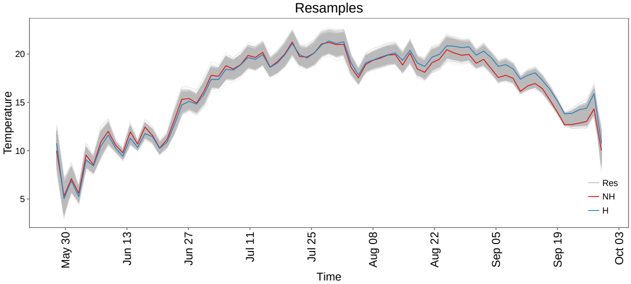
Measurement	C−H−HSR−SR	$\tilde{\text{S}} - \text{N}\tilde{\text{S}}$	$\tilde{\text{H}} - \text{N}\tilde{\text{H}}$
Soil Moisture	0.074	0.007	0.288
Soil Temperature	0.150	0.064	0.168

Obtaining Significance in Fanova Tests

- p-values
  - Provide a one number summary of whether entire sequence of mean curves are significantly different between groups
- Plot of resample curves
  - Provides information that p-values alone can't provide
  - Visualizes uncertainty in differences between mean curves
  - Can identify shifts in behavior of mean curves over time

Temperature – Motivation for splitting the domain

Figure 11: Temperature Resample Curves for Difference Due to Heating (p-value=0.168)

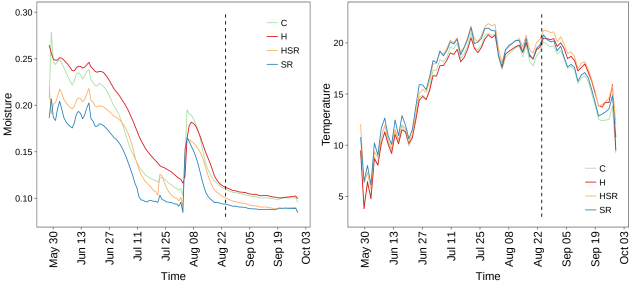


- Shift in behavior of mean curves in August
- Difference becomes more significant over time
- Could signify a change in the process generating the sample curves

Temperature – Motivation for splitting the domain

- Large moisture event in August corresponds to changes in mean temperature curves
- Dotted line (August 24) marks approximate end to large moisture event
- Split domain on August 24, and perform fanova tests separately on each part of domain

Figure 12: Sample mean soil moisture and temperature curves



Summary of Temperature Fanova Tests on Split Domain

- Domain split into 2 parts
  - May 30, 2011 - Aug 24, 2011 (Part 1)
  - Aug 24, 2011 - Sept 29, 2011 (Part 2)
- Table gives p-values for the following soil temperature fanova tests applied separately on the domain:
  - Test for difference between all treatments (C−H−HSR−SR)
  - Test for difference due to snow removal ($\tilde{\mathrm{S}} - \mathrm{N}\tilde{\mathrm{S}}$)
  - Test for difference due to heating ($\tilde{\mathrm{H}} - \mathrm{N}\tilde{\mathrm{H}}$)

Domain Split	C−H−HSR−SR	$\tilde{\mathrm{S}} - \mathrm{N}\tilde{\mathrm{S}}$	$\tilde{\mathrm{H}} - \mathrm{N}\tilde{\mathrm{H}}$
Part 1	0.160	0.024	0.549
Part 2	0.106	0.336	0.018

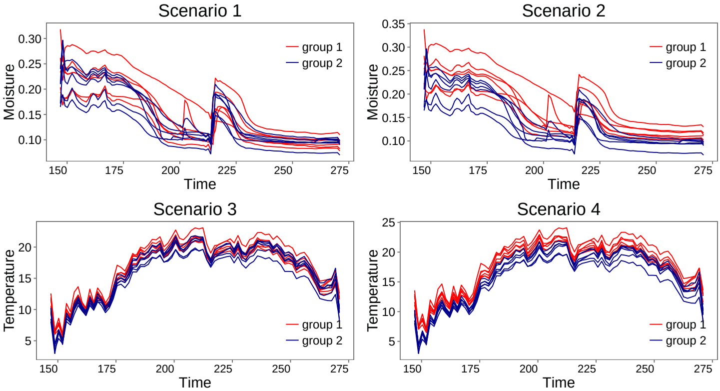
Outline

Simulation Study

- Our study has small number of sample curves per group
- A simulation study investigates effect of number of sample curves on statistical power of fanova test.
- Let $\mu_1(t)$ and $\mu_2(t)$ be the expected values of the L_2-processes generating sample curves in group 1 and 2.
- If $\mu_1(t)$ and $\mu_2(t)$ differ, we're interested in the procedure's ability to correctly reject H_0
- In order to recreate features similar to our data set, $\mu_i(t)$ and $K_i(s,t)$ are constructed from sample data

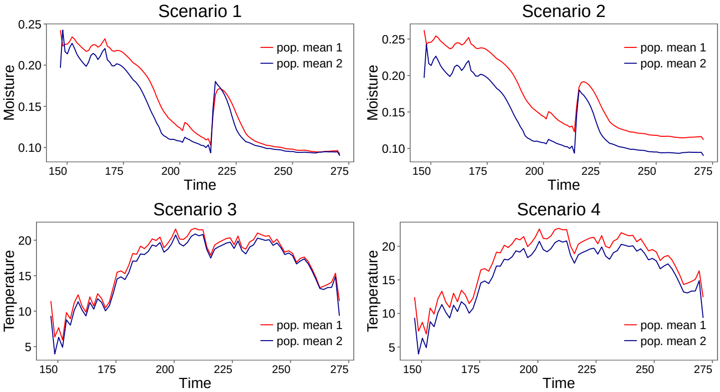
Construction of Simulation Scenarios

Figure 13: $K_1(s,t)$ and $K_2(s,t)$ constructed from sample covariance of group 1 and group 2 for each scenaro. $\mu_1(t)$ and $\mu_2(t)$ constructed from sample mean of group 1 and group 2 – shown on next slide



Construction of Simulation Scenarios

Figure 14: $\mu_1(t)$ and $\mu_2(t)$ from which sample curves will be generated in simulation study



Simulation Procedure

For each simulation scenario and $n_i = 6, 12, 18, 24, 30$,

- n_i sample curves were generated using a Gaussian distribution with mean μ_i and covariance $K_i(s,t)$
- Fanova procedure was applied and repeated for 10,000 simulation
- Proportion of simulations for which p-value of less than 0.05 was obtain is recorded.

Simulation Results

Scenario	6	12	18	24	30
1 (Moist. Small Diff)	0.235	0.385	0.550	0.691	0.817
2 (Moist. Large Diff)	0.724	0.958	0.996	1.000	1.000
3 (Temp. Small Diff)	0.539	0.879	0.984	0.999	1.000
4 (Temp. Large Diff)	0.993	1.000	1.000	1.000	1.000

1. Fanova procedure powerful for small n_i if differences between groups are large
2. Need moderate or large n_i to detect small differences
3. If covariance structure is large, may need even larger n_i to detect small differences

Outline

Conclusion

- Evidence of difference due to snow removal in soil moisture
- Evidence of change in behavior of temperature curves corresponding to large moisture event
  - Prior to large moisture event, differences apparent due to snow removal but not heating
  - After large moisture event, difference apparent due to heating but not snow removal
- Since many data can be considered functional in nature, fanova procedure widely applicable.
  - Useful for obtaining p-values through bootstrap
  - Can visualize uncertainty in differences between mean curves
  - Can identify shifts in behavior of mean curves over time
  - Loss of power when number of sample curves is small

References

A. Cuevas, M. Febrero, and R. Fraiman.
An anova test for functional data.
Computational Statistics and Data Analysis, 47(1):111–122, 2004.

J. A. Sherwood, D. M. Debinski, P. C. Caragea, and M. J. Germino.
Effects of experimentally reduced snowpack and passive warming on montane meadow plant phenology and floral resources.
Ecosphere, 8(3), 2017.

J. Tarrio-Saaverda, S. Naya, M. Francisco-Fernandez, R. Artiaga, and J. Lopez-Beceiro.
Application of functional anova to the study of thermal stability of micronano silica epoxy composites.
Chemometrics and Intelligent Laborotory Systems, 105:114–124, 2011.

G. Wahba and P. Craven.
Smoothing noisy data with spline functions. estimating the correct degree of smoothing by the method of generalized cross-validation.
Numerische Mathematik, 31:377–404, 1978/79.