Project Title: Statistical Model to Predict the Compressibility of Florida's Soils

(BDV24 TWO 977-24)

Boo Hyun Nam (PI)
Depart. of Civil, Environmental, and Construction Eng.
University of Central Florida

Petros Xanthopoulos (Co-PI)
Stetson University

Presented at
2017 FDOT GRIP Meeting

August 18, 2017

Presentation outline

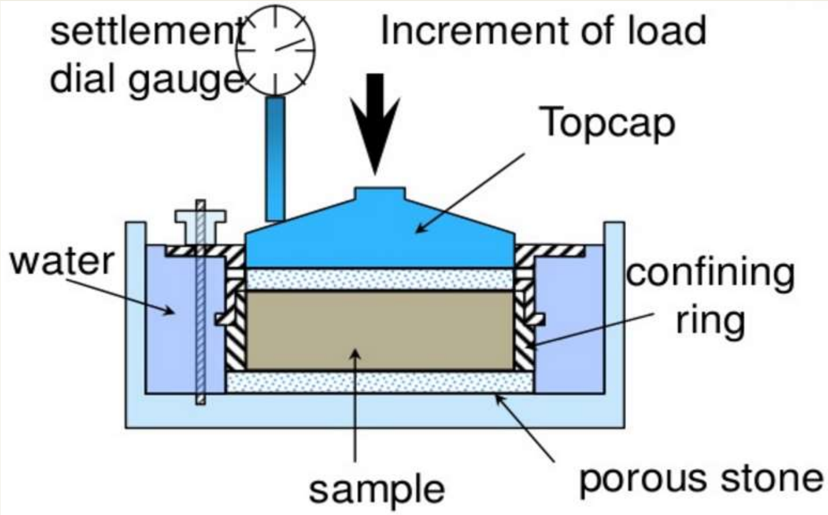
- **Background**
- **Project Objectives**
- **Work Plan**
  - **Task 1:** Identify the existing Cc, Cr, and Cv models
  - **Task 2:** Compare the accuracy with existing models
  - **Task 3:** Collect data and create comprehensive Florida's geotechnical database
  - **Task 4:** Evaluate the correlations of key affecting parameters and soil compressibility
  - **Task 5:** Develop a methodology to construct the statistical Cc, Cr, and Cv models
  - **Task 6:** Develop the soil compressibility prediction models for specific soil types
  - **Task 7:** Evaluate the relationship with field tests
- **Future Work Plan**

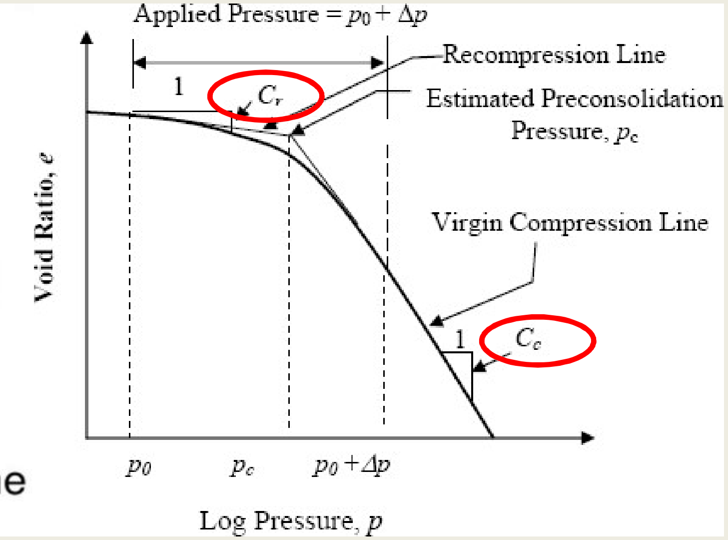
Current works (Tasks 1–5)

Project Objectives

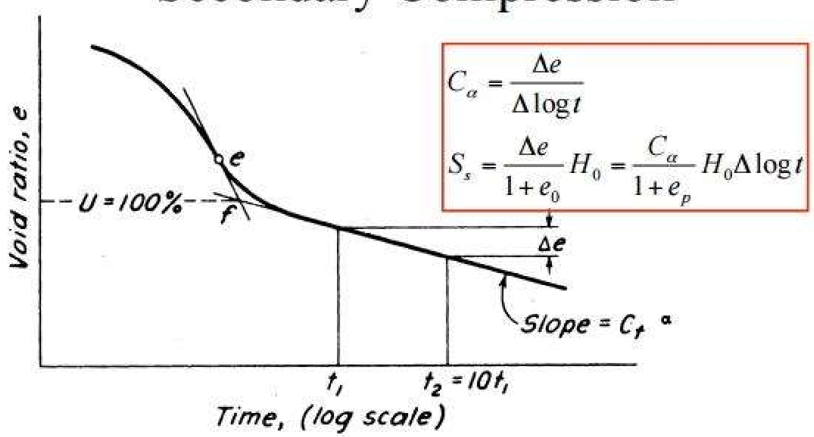
- To identify statistically significant affecting variables on Cc and Cr and to evaluate their correlations
- To identify the most accurate model of soil compressibility (from statistical perspective) for Florida's soils.
- To develop the best performing statistical models to predict Cc, Cr, and Cv for Florida's soils
  - State-of-the-art statistical techniques will be used
  - Models will be developed for specific soil types

Background





Secondary Compression



- Deformation of soil under constant load
- No excess pore pressure
- Affects properties of the clay
- Secondary compression index = C_α

$$c_v = \frac{T H^2_{D_{50}}}{t}$$

where:

T = a dimensionless time factor: for method 12.3.1 use 50 % consolidation with T = T_{50} = 0.197, for method 12.3.2 use 90 % consolidation with T = T_{90} = 0.848,

t = time corresponding to the particular degree of consolidation, s or min; for method 12.3.1 use $t = t_{50}$, for method 12.3.2 use $t = t_{90}$, and

$H_{D_{50}}$ = length of the drainage path at 50 % consolidation,

Cv = coefficient of consolidation

C_α = secondary compression index

Background (cont.)

Problem Statement:

- Consolidation testing for the Cc, Cr, and Cv (coefficient of consolidation) is not simple and easy. Those indexes are difficult to quantify and have large uncertainty.
- Two ways to determine Cc, Cr, and Cv: (1) direct measurement via lab test and (2) correlation to other soil data determined from lab tests.
- Many previous studies on prediction models of soil compressibility such as Cc, Cr, and Cv. However, those models may not be accurate enough for Florida's soil conditions because the models are constructed based on local soils and most models are based on a simple linear regression model.

Research team

- **Geotechnical Engineers**
  - Boo Hyun Nam (Associate Prof)

    Dept. of Civil, Environmental, and Construction Eng., UCF
  - Yongje Kim (PhD student)

    Dept. of Civil, Environmental, and Construction Eng., UCF
- **Data Scientists**
  - Petros Xanthropoulos (Assist. Prof)

    Dept. of Decision and Info Sci., Stetson Univ.
  - Orestis Panagopoulos (Assist. Prof)

    College of Business, California State Univ.




CURRENT WORKS

Task 1: Identify the existing Cc, Cr, and Cv models

- Literature review to identify the existing “mathematical” models of soil compressibility.
- Review on the guidance or standard methods of federal and state agencies
- Review on previous studies of Florida (including “database” if any)
  - Informal survey to Districts and consultants

Existing prediction models

Ind. Variable	Dep. Variable	Equation	Reference	Notes
Cc	w	$C_c = 0.01w - 0.05$	Azzouz (1976)	All soils
		$C_c = 0.01w$	Koppula (1981)	Clays
		$C_c = 0.01w - 0.075$	Herrero (1983)	Clays
		$C_c = 0.013w - 0.115$	Park, Lee (2011)	Clays
		$C_c = 0.0075w$	Miyakawa (1960)	Peat
		$C_c = 0.011w$	Cook (1956)	Peat
	e	$C_c = 0.54e - 0.19$	Nishida (1956)	Clays
		$C_c = 0.43e - 0.11$	Cozzolino (1961)	Clays
		$C_c = 0.75e - 0.38$	Sowers (1970)	Clays
		$C_c = 0.49e - 0.11$	Park, Lee (2011)	Clays
		$C_c = 0.4(e-0.25)$	Azzouz (1976)	All soils
		$C_c = 0.15e + 0.01077$	Bowles (1989)	Clays
		$C_c = 0.287e - 0.015$	Ahadiyan (2008)	Clays
		$C_c = 0.6e$	Sowers (1970)	Peat
		$C_c = 0.3(e-0.27)$	Hough (1957)	Clays
	LL	$C_c = 0.006(LL-9)$	Azzouz (1976)	Clays
		$C_c = (LL-13)/109$	Mayne (1980)	Clays
		$C_c = 0.009(LL-10)$	Terzaghi, Peck (1967)	Clays

		$C_c = 0.014LL-0.168$	Park, Lee (2011)	Clays
		$C_c = 0.0046(LL-9)$	Bowles (1989)	Clays
		$C_c = 0.011(LL-16)$	McClelland (1967)	Clays
	w, LL	$C_c = 0.009w + 0.005LL$	Koppula (1981)	Clays
		$C_c = 0.009w + 0.002LL - 0.01$	Azzouz (1976)	Clays
	e, w	$C_c = 0.4(e + 0.001w - 0.25)$	Azzouz (1976)	All soils
	e, LL	$C_c = -0.156 + 0.411e - 0.00058LL$	Al-Khafaji, Andersland (1992)	Clays
		$C_c = -0.023 + 0.271e + 0.001LL$	Ahadiyan (2008)	Clays
	e, w, LL	$C_c = 0.37(e + 0.003LL +).0004w - 0.34)$	Azzouz (1976)	Clays
		$C_c = -0.404 + 0.341e + 0.006w + 0.004LL$	Yoon, Kim (2008)	Clays
	w, LL, e, γ_{dry}	$C_c = 0.1597(w^{-0.0187})(1 + e)^{1.592}(LL^{-0.0638})(\gamma_{dry}{}^{-0.8276})$	Ozer (2008)	Clays
		$C_c = 0.151 + 0.001225w + 0.193e - 0.000258LL - 0.0699\gamma_{dry}$	Ozer (2008)	Clays
Cr	e	$C_r = 0.156e + 0.0107$	Elnaggar, Krizek (1971)	Clays
		$C_r = 0.208e + 0.0083$	Peck, Reed (1954)	Clays
		$C_r = 0.14(e+0.007)$	Azzouz (1976)	All soils
	w	$C_r = 0.003(w + 7)$	Azzouz (1976)	All soils
	LL	$C_r = 0.002(LL + 9)$	Azzouz (1976)	All soils
	e, w	$C_r = 0.142(e - 0.009w + 0.006)$	Azzouz (1976)	All soils
	w, LL	$C_r = 0.003w + 0.0006LL + 0.004$	Azzouz (1976)	All soils
	e, LL	$C_r = 0.126(e + 0.003LL-0.06)$	Azzouz (1976)	All soils
	e, w, LL	$C_r = 0.135(e + 0.1LL-0.002w - 0.06)$	Azzouz (1976)	All soils

Task 2: Compare the accuracy with existing models

- Compare the accuracy of the prediction models identified in Task 1.
  - With respect to root mean square error (RMSE) and R2
- Identify good performing models for Florida's soils

Equation	Reference	Notes	R^2	RMSE
$C_c = 0.01w - 0.05$	Azzouz (1976)	All soils	0.7448	0.8359
$C_c = 0.01w$	Koppula (1981)	Clays	0.5202	0.4191
$C_c = 0.01w - 0.075$	Herrero (1983)	Clays	0.5189	0.4336
$C_c = 0.013w-0.115$	Park, Lee (2011)	Clays	0.6729	0.3953
$C_c = 0.0075w$	Miyakawa (1960)	Peat	0.5784	1.5194
$C_c = 0.011w$	Cook (1956)	Peat	0.6611	1.9601
$C_c = 0.54e - 0.19$	Nishida (1956)	Clays	0.7236	0.3945
$C_c = 0.43e - 0.11$	Cozzolino (1961)	Clays	0.6120	0.4046
$C_c = 0.75e - 0.38$	Sowers (1970)	Clays	0.7362	0.5552
$C_c = 0.49e - 0.11$	Park, Lee (2011)	Clays	0.6847	0.3924
$C_c = 0.4(e - 0.25)$	Azzouz (1976)	All soils	0.5676	0.7501
$C_c = 0.15e + 0.01077$	Bowles (1989)	Clays	0.3157	0.7536
$C_c = 0.287e - 0.015$	Ahadiyan (2008)	Clays	0.3847	0.7692
$C_c = 0.6e$	Sowers (1970)	Peat	0.6715	1.7876
$C_c = 0.3(e-0.27)$	Hough (1957)	Clays	0.4081	0.5425
$C_c = 0.006(LL - 9)$	Azzouz (1976)	Clays	0.2857	0.6213
$C_c = (LL-13)/109$	Mayne (1980)	Clays	0.4323	0.5638
$C_c = 0.009(LL -10)$	Terzaghi, Peck (1967)	Clays	0.4236	0.5641
$C_c = 0.014LL - 0.168$	Park, Lee (2011)	Clays	0.5569	0.7921

Compare the accuracy (cont.)

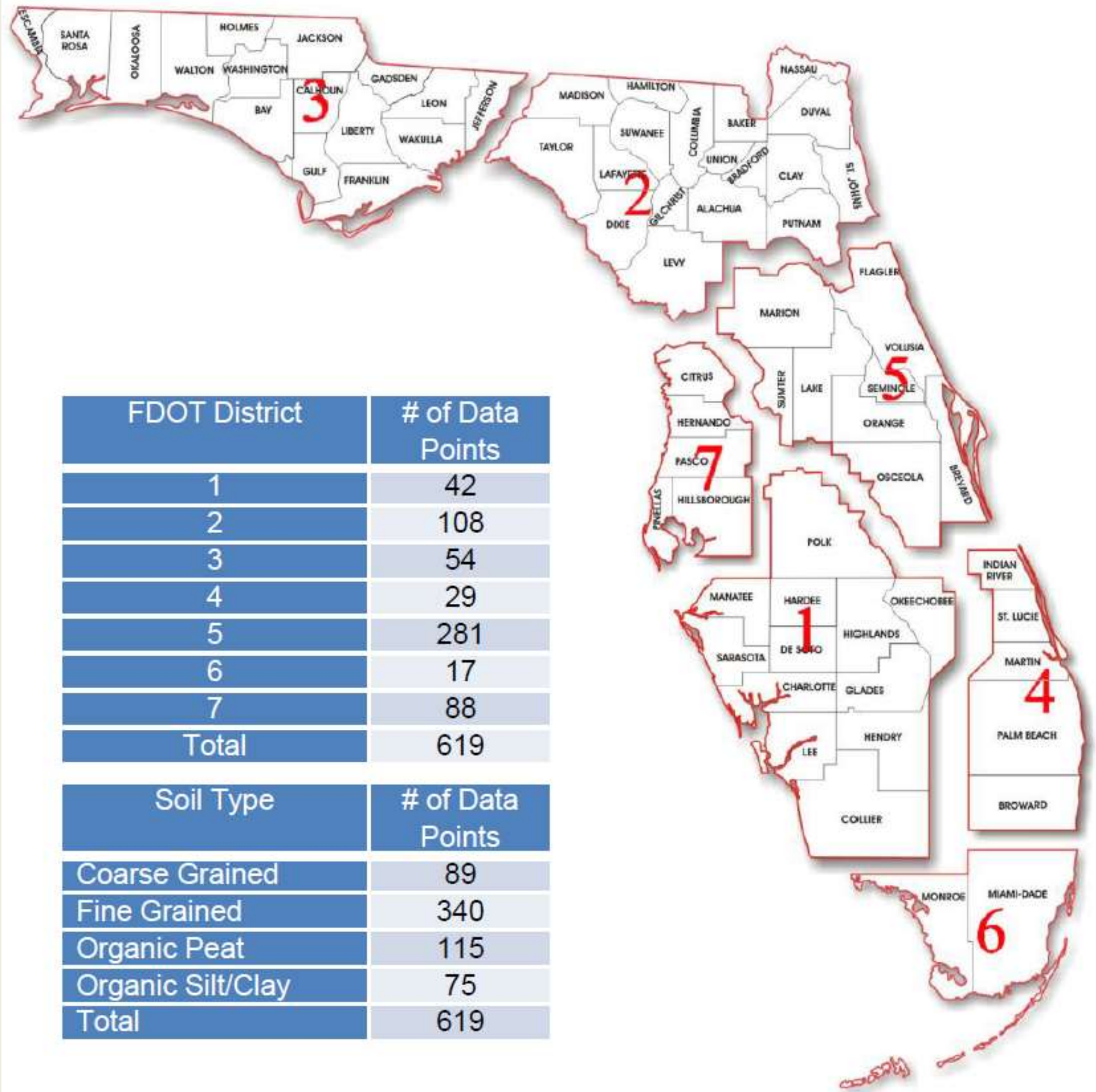
Equation	Reference	Notes	R^2	RMSE
$C_c = 0.0046(LL-9)$	Bowles (1989)	Clays	0.2780	0.6989
$C_c = 0.011(LL-16)$	McClelland (1967)	Clays	0.5094	0.5991
$C_c = 0.009w + 0.005LL$	Koppula (1981)	Clays	0.5701	0.5518
$C_c = 0.009w + 0.002LL – 0.01$	Azzouz (1976)	Clays	0.5866	0.4875
$C_c = 0.4(e + 0.001w – 0.25)$	Azzouz (1976)	All soils	0.7057	0.7414
$C_c = -0.156 + 0.411e – 0.00058LL$	Al-Khafaji, Andersland (1992)	Clays	0.5276	0.3881
$C_c = -0.023 + 0.271e + 0.001LL$	Ahadiyan (2008)	Clays	0.3400	0.4597
$C_c = 0.37(e + 0.003LL +).0004w – 0.34)$	Azzouz (1976)	Clays	0.5014	0.3888
$C_c = -0.404 + 0.341e + 0.006w + 0.004LL$	Yoon, Kim (2006)	Clays	0.6805	0.4991
$C_c = 0.1597(w^{-0.0187})(1 + e)^{1.592}(LL^{-0.0638})(\gamma_{dry}^{-0.8276})$	Ozer (2008)	Clays	0.6824	0.5886
$C_c = 0.151 + 0.001225w + 0.193e – 0.000258LL – 0.0699\gamma_{dry}$	Ozer (2008)	Clays	0.3006	0.5204
$C_r = 0.156e + 0.0107$	Elnaggar, Krizek (1971)	Clays	0.5330	0.2536
$C_r = 0.208e + 0.0083$	Peck, Reed (1954)	Clays	0.5419	0.3643
$C_r = 0.14(e+0.007)$	Azzouz (1976)	All soils	0.6016	0.3369
$C_r = 0.003(w + 7)$	Azzouz (1976)	All soils	0.5780	0.4415
$C_r = 0.002(LL + 9)$	Azzouz (1976)	All soils	0.5485	0.1682
$C_r = 0.142(e – 0.009w + 0.006)$	Azzouz (1976)	All soils	0.6089	0.1802
$C_r = 0.003w + 0.0006LL + 0.004$	Azzouz (1976)	All soils	0.5674	0.2344
$C_r = 0.126(e + 0.003LL-0.06)$	Azzouz (1976)	All soils	0.5808	0.2109
$C_r = 0.135(e + 0.1LL-0.002w – 0.06)$	Azzouz (1976)	All soils	0.5548	0.3131

Task 3: Collect data and create comprehensive Florida's geotechnical database

- Data collection has included:
  - Natural moisture
  - Wet density
  - Fines (passing No. 200)
  - Initial void ratio
  - Automatic hammer SPT N value
  - Effective overburden pressure
  - Atterberg limits (liquid limit (LL), Plasticity index (PI))
  - Organic content
  - CPT data (if possible)
  - Etc.

Data collection

A total of <u>619</u> consolidation test data so far. Each consolidation test has an accompanying SPT boring to provide a description of the soil's stiffness. The vast majority of the data collected is from the FDOT District 5 which includes the counties of Volusia, Seminole, Orange, Osceola, Brevard, Lake, Marion, Sumter, and Flagler.



FDOT District	# of Data Points
1	42
2	108
3	54
4	29
5	281
6	17
7	88
Total	619

Soil Type	# of Data Points
Coarse Grained	89
Fine Grained	340
Organic Peat	115
Organic Silt/Clay	75
Total	619

Task 4: Evaluate the correlations of key affecting parameters and soil compressibility

- Evaluate the correlation between key index parameters and soil compressibility (Cc and Cr) and Cv
  - Specific index parameters can have dominant influence in compressibility of specific soil types.

  - High plasticity clays (Plasticity Index > 70)
  - Low plasticity clays (Plasticity Index <= 70)
  - Silts
  - High organic soils (Natural Moisture >= 160)
  - Low organic soils (Natural Moisture < 160)

*Coefficient of Determination (R^2) and Root Mean Square Error (RMSE) values were calculated to quantify the performance level of the key index parameters and appear on the Excel file.

Top 3 Index Parameters - Positive Correlation with Cc

Silts	Pearson's Correlation Coefficient
Natural Moisture (%)	0.7388
Liquid Limit (LL)	0.7469
Plasticity Index (PI)	0.7347

Pearson's correlation between – 1 and 1

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Low plasticity clays	Pearson's Correlation Coefficient
Effective Overburden Pressure (ksf)	0.1803
Natural Moisture (%)	0.6168
Initial Void Ratio (e)	0.6454

Low organic Soils	Pearson's Correlation Coefficient
Natural Moisture (%)	0.3455
Automatic Hammer Blow Count	0.4187
Initial Void Ratio (e)	0.8130

High plasticity clays	Pearson's Correlation Coefficient
Natural Moisture (%)	0.8149
Liquid Limit (LL)	0.4601
Initial Void Ratio (e)	0.8286

High organic Soils	Pearson's Correlation Coefficient
Natural Moisture (%)	0.6420
Organic Content (%)	0.3701
Initial Void Ratio (e)	0.6480

Top 3 Index Parameters – Negative Correlation with Cc

Silts	Pearson's Correlation Coefficient
Dry Density (pcf)	-0.6111
Fines (-200) (%)	-0.6704
Specific Gravity	-0.7154

Low plasticity clays	Pearson's Correlation Coefficient
Wet Density (pcf)	-0.4888
Dry Density (pcf)	-0.5513
Automatic Hammer Blow Count	-0.2066

Low organic Soils	Pearson's Correlation Coefficient
Wet Density (pcf)	-0.5159
Dry Density (pcf)	-0.4111
Fines (-200) (%)	-0.2751

High plasticity clays	Pearson's Correlation Coefficient
Wet Density (pcf)	-0.7025
Dry Density (pcf)	-0.7885
Fines (-200) (%)	-0.3365

High organic Soils	Pearson's Correlation Coefficient
Wet Density (pcf)	-0.2536
Dry Density (pcf)	-0.6138
Effective Overburden Pressure (ksf)	-0.2496

Top 3 Index Parameters - Positive Correlation with Cr

Silts	Pearson's Correlation Coefficient
Natural Moisture (%)	0.7176
Liquid Limit (LL)	0.6159
Plasticity Index (PI)	0.6336

Low plasticity clays	Pearson's Correlation Coefficient
Liquid Limit (LL)	0.2207
Natural Moisture (%)	0.3812
Initial Void Ratio (e)	0.3743

Low organic Soils	Pearson's Correlation Coefficient
Effective Overburden Pressure (ksf)	0.1673
Natural Moisture (%)	0.3084
Initial Void Ratio (e)	0.7481

High plasticity clays	Pearson's Correlation Coefficient
Natural Moisture (%)	0.4541
Liquid Limit (LL)	0.4394
Initial Void Ratio (e)	0.4355

High organic Soils	Pearson's Correlation Coefficient
Natural Moisture (%)	0.4795
Organic Content (%)	0.1715
Initial Void Ratio (e)	0.8125

Top 3 Index Parameters – Negative Correlation with Cr

Silts	Pearson's Correlation Coefficient
Dry Density (pcf)	-0.5752
Fines (-200) (%)	-0.6351
Specific Gravity	-0.7426

Low plasticity clays	Pearson's Correlation Coefficient
Wet Density (pcf)	-0.2332
Dry Density (pcf)	-0.3242
Automatic Hammer Blow Count	-0.0894

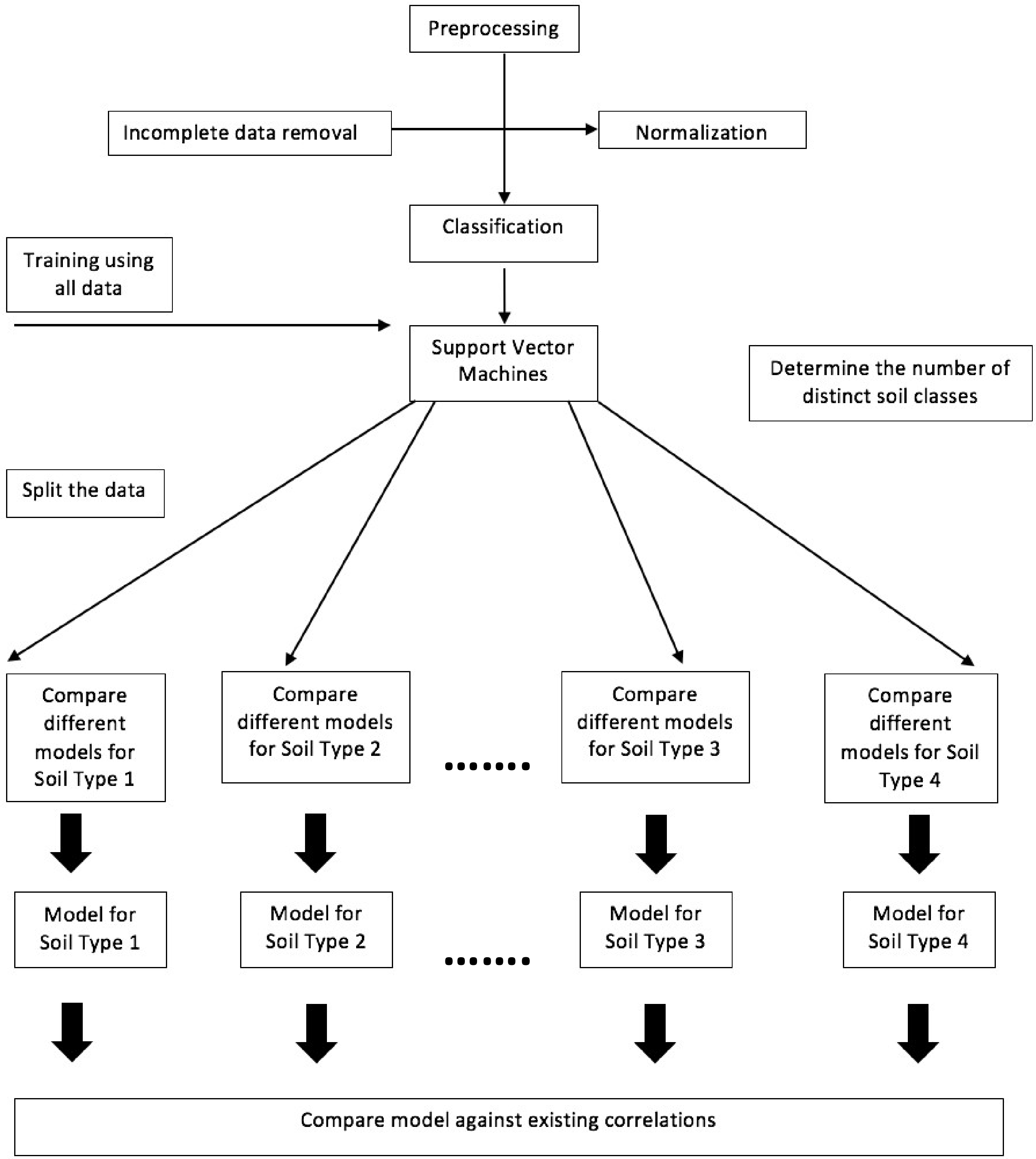
Low organic Soils	Pearson's Correlation Coefficient
Wet Density (pcf)	-0.4484
Dry Density (pcf)	-0.3324
Fines (-200) (%)	-0.2098

High plasticity clays	Pearson's Correlation Coefficient
Wet Density (pcf)	-0.4228
Dry Density (pcf)	-0.4673
Automatic Hammer Blow Count	-0.1667

High organic Soils	Pearson's Correlation Coefficient
Effective Overburden Pressure (ksf)	-0.2307
Dry Density (pcf)	-0.4709
Fines (-200) (%)	-0.2100

Task 5: Develop a methodology to construct the statistical Cc, Cr, and Cv models

- Step 1: Data Preprocessing (consisting of normalization and outlier detection).
- Step 2: Identification of distinct soil types through supervised learning (Support Vector Machines).
- Step 3: Cc, Cr, and Cv model development for each soil type identified from Step 2. Development of appropriate regression models and investigation of potential interaction effects.
- Step 4: Models validation. Investigation of the ability of models to predict on data that haven't been used for model development.
- Step 5: Quantification of uncertainties and estimation of confidence intervals that will quantify the predictive accuracy of the proposed models.


Preprocessing
Incomplete data removal
Normalization
Classification
Training using all data
Support Vector Machines
Determine the number of distinct soil classes
Split the data
Compare different models for Soil Type 1
Compare different models for Soil Type 2
Compare different models for Soil Type 3
Compare different models for Soil Type 4
Model for Soil Type 1
Model for Soil Type 2
Model for Soil Type 3
Model for Soil Type 4
Compare model against existing correlations

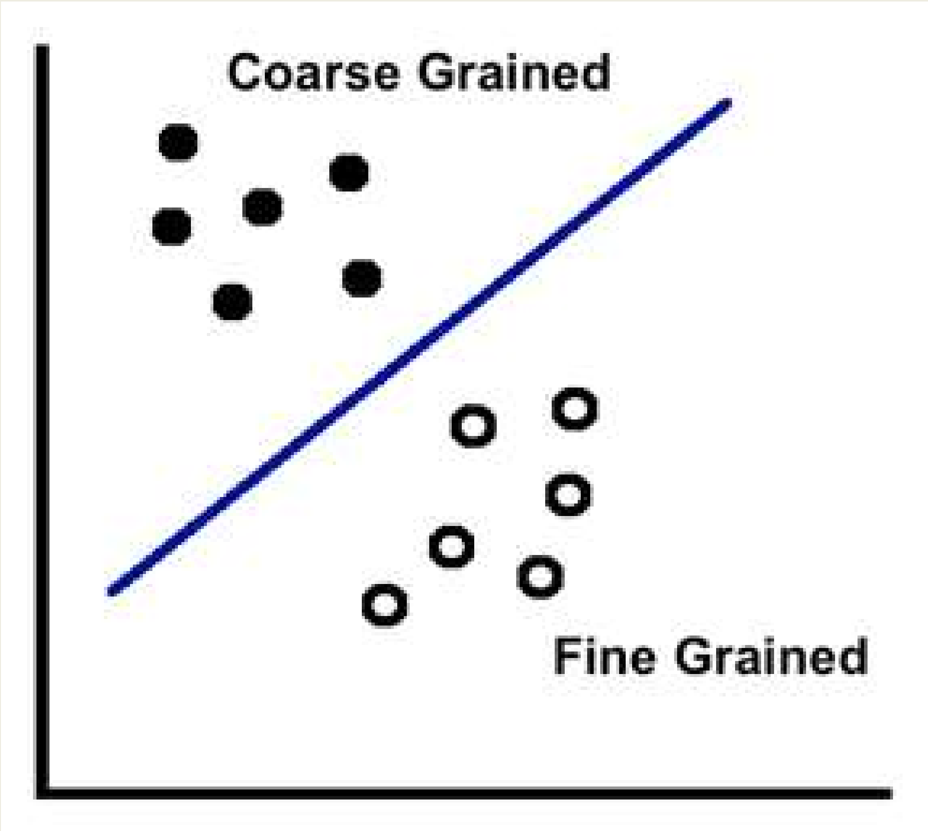
Preprocessing

- Full data sets were segregated from non-full data sets. Full data sets include the following parameters: moisture content, initial void ratio, dry unit weight, wet unit weight, automatic hammer blow count, overburden stress, and fines content.
- For simplification purposes, and abundance of full data sets, the non-full data sets were not included as part of this study.
- Data is normalized through z-score normalization, which offers a way to compare observations that are measured on different scale. Each soil classification has a unique dimensionality due to a varying number of full data sets.

Classification

- A classification model is developed that assists in determining the number of distinct soil groups that exists.
- The goal is to confirm or reject the hypothesis that each soil type requires a different statistical model.
- The data is comprised of two sets – training and testing. Training data is used to teach the algorithm and testing data is used to evaluate the accuracy and predictability of the model.
- The confusion matrix illustrates how well the testing data was filtered to the class that the training data predicted it would fall into.
- As can be observed from the table, the assumed classifications were confirmed by the testing data, with the exception of Organic Silt/Clay. The testing data predicted that this classification behaved more like a Fine Grained soil.

Actual Class \ Predicted Class	Coarse Grained	Fine Grained	Organic Peat	Organic Silt/Clay
Coarse Grained	11	1	1	-
Fine Grained	1	44	-	1
Organic Peat	-	-	12	1
Organic Silt/Clay	-	7	1	3



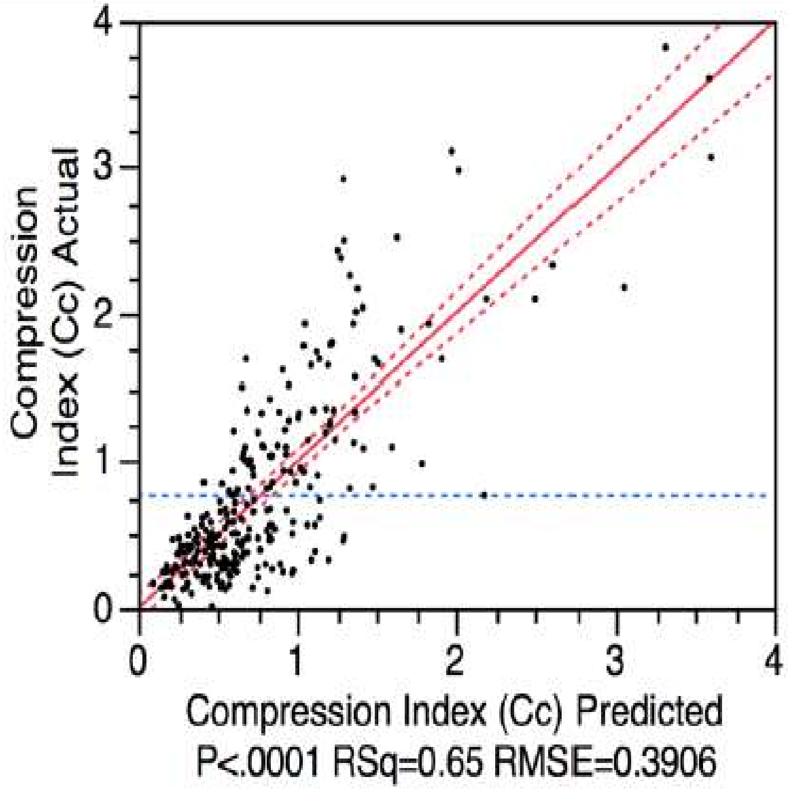
Example of data classification (e.g. classified soils group as coarse and fine grained)

Preliminary Results

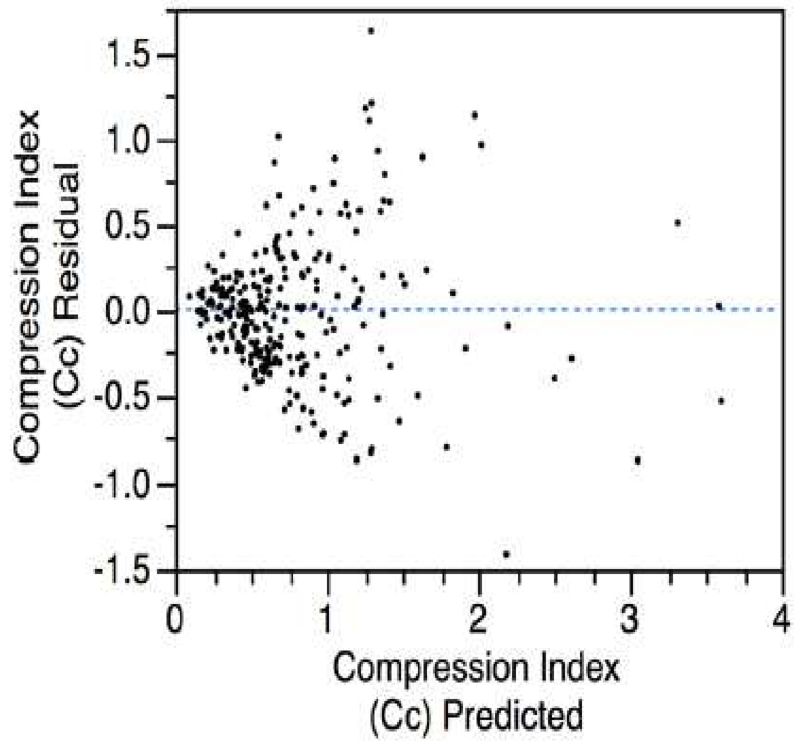
- A regression model was developed with interactions for each distinct group/class (Coarse Grained, Fine Grained, and Organic Peat).
- The table below illustrates the regression models that had reliable predictive capability:

Equation	Notes	R^2	R^2_{adj}	RMSE
$C_c = -0.146 + 0.001 * \gamma_{wet} - 0.003 * \gamma_{dry} + 0.007 * N + 0.005 * Fines + 0.373 * e_0 - 0.0006 * [(\gamma_{wet}\ 115.484) * (N - 6.493)] + 0.001 * [(\gamma_{wet} - 115.484) * (Fines - 31.584)] + 0.032 * [(Fines - 31.584) * (e_0 - 1.028)] + 0.001 * [(\gamma_{wet} - 115.484) * (\gamma_{wet} - 115.484)] - 0.0003 * [(\gamma_{dry} - 86.024) * (\gamma_{dry} - 86.024)] - 0.0005 * [(N - 6.493) * (N - 6.493)]$	Coarse Grained	0.9079	0.8888	0.1108
Reduced Model $C_c = 0.759 + 0.0048 * \gamma_{wet} - 0.012 * \gamma_{dry} - 0.002 * N - 0.0012 * e_0 - 0.0006 * [(\gamma_{wet} - 115.484) * (\gamma_{wet} - 115.484)]$		0.8308	0.8133	0.1436
$C_c = -0.217 + 0.006 * W + 0.287 * e_0$	Fine Grained	0.6487	0.6462	0.3906
$C_c = 1.272 + 0.006 * W - 0.021 * Fines + 0.121 * e_0 - 0.000009 * [(W - 359.133) * (Fines - 65.666)] - 0.000985 * [(W - 359.133) * (e_0 - 5.543)] + 0.0521 * [(e_0 - 5.543) * (e_0 - 5.543)]$	Organic Peat	0.7724	0.7480	1.0904

Preliminary Results

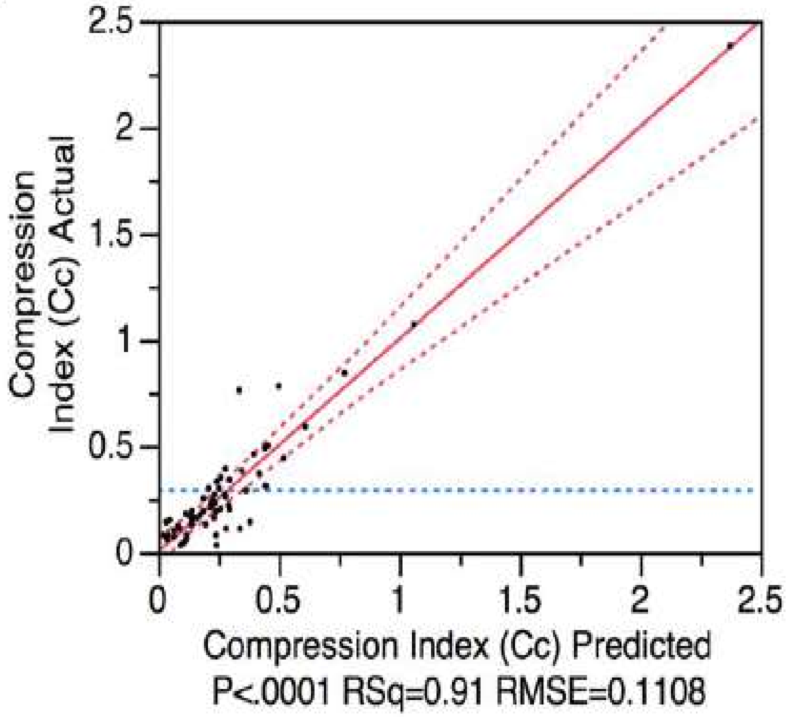


Comparison between the measured C_c and the predicted C_c for Fine Grained including Organic Silt/Clay Class

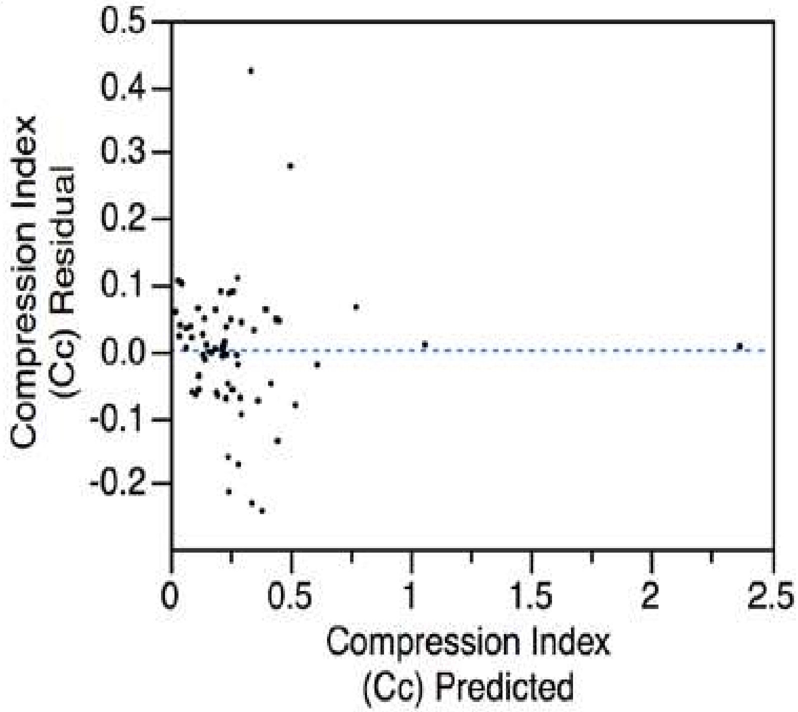


Residual by Predicted Plot for C_c model of Fine Grained including Organic Silt/Clay Class

Preliminary Results



Comparison between the measured C_c and the predicted C_c for Coarse Grained Class



Residual by Predicted Plot for C_c model of Coarse Grained Class

FUTURE WORKS

Task 5: Develop a methodology to construct the statistical Cc, Cr, and Cv models

- Develop the model based on the “complete” data set
- Develop the methodology to account for missing data set
- Quantify the uncertainty level

Task 6: Develop the soil compressibility prediction models for specific soil types

- With the framework developed in Step 5, the models to be developed include:
  - Cc and Cr model
  - Cv model
  - Cα model (if available data)
- Soil types to be considered:
  - High plasticity clays
  - Low plasticity clays
  - Silts
  - High organic soils
  - Low organic soils

Task 7: Evaluate the relationship with field tests

- Evaluate the relationship between field tests (e.g. SPT and CPT if any) and the soil compressibility especially with highly compressible soils.
  - correlation of the compressibility (Cc and Cr) and SPT N values and CPT tip resistance (if any) will be then investigated.

Thank you!

Question?