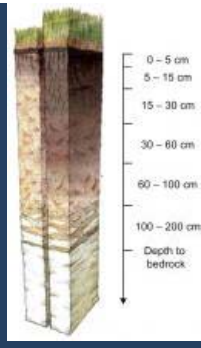
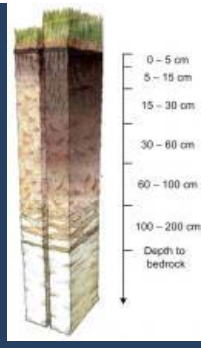


Converting pH 1:1 H_2O (pH1:1$_W$) and/or 1:2 $CaCl_2$ (pH 1:2$_{Ca}$) to 1:5 H_2O (pH1:5$_W$) for the North America node of GlobalSoilMap.net project

Zamir Libohova[1]; Nathan Odgers[2]; Jim Thomson[2]; Rick Nesser[1]; Rich Ferguson[1]; Jon Hempel[1]; Larry West [1]

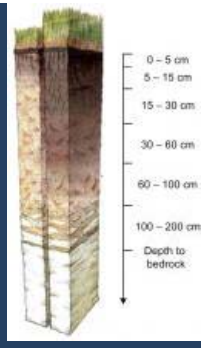
1 - USDA-NRCS-National Soil Survey Center
2 - West Virginia University-Geospatial Research Unit (USDA-NRCS-NSSC)

Rationale and Objectives



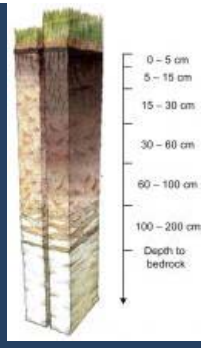
- Rationale:
  - The GlobalSoilMap.net initiative requires the generation of continuous maps for soil properties, including pH in 1:5W;
  - The US Soil Survey employs a pH $1:1_W$ and a pH1:2 $CaCl_2$ (0.01M) (pH $1:2_{Ca}$) for routine pH and more than a 100,000 soil samples have been analyzed with these methods (Soil Survey Staff, 2009).
- Objectives:
  - Establish inter-method relationships representing any soil order, mineralogy, texture, depth, genetic horizons to reliably convert between methods;
  - Understand and define the influence of soil properties on any mutual bias between data sets that might be observed.

Materials and Methods

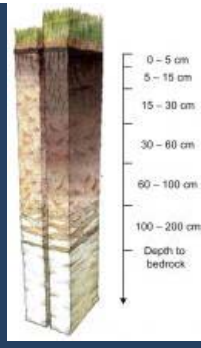


- A subset of soil samples from the National Cooperative Soil Survey Characterization Database that were previously analyzed for pH $1{:}1_W$ and pH $1{:}2_{Ca}$ were selected for pH determination by $1{:}5_W$ and 1:5 $CaCl_2$ (0.01 M) (pH$1{:}5_{Ca}$);
- The samples represent 11 soil orders, 3 mineralogy classes, organic vs. mineral, and 5 texture classes grouped;
- For each category, between 25 and 30 samples were selected to represent a statistically sizable sample.

Soil Sample Locations



Results and Discussions



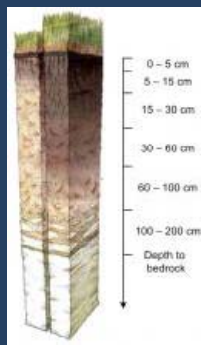
- Linear Models:
  - pH1:5$_W$ vs. pH1:1$_W$
  - pH1:5$_W$ vs. pH1:2$_{Ca}$
- Other Predictive Models.
- Influence of Variables on the prediction:
  - Continuous (pH 1:1$_W$)
  - Categorical (Soil Order, etc.,) (Leverage plots);
- Error Predictions and Error Distribution:
  - pH1:5$_W$ (Measured) vs. pH1:5$_W$ (Predicted):
  - Simple Model (w/o Classes)
  - Complex Model (w Classes)
- Other Error Sources.

Mean pH value comparisons by method and soil conditions

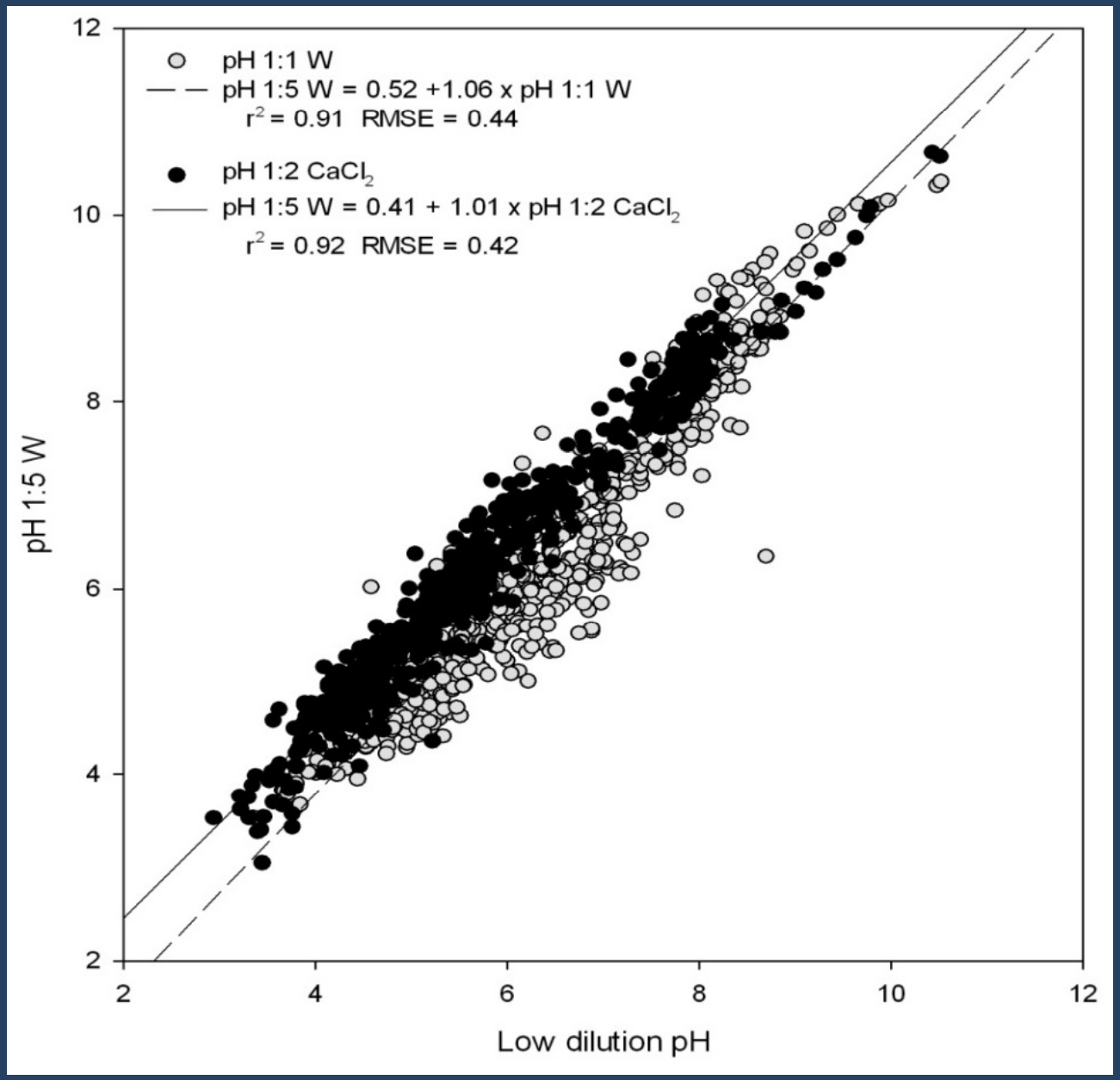
Soil Condition	Classes	No of Obs.	pH Method: pH $1:5_W$ Mean	Std. Err.†	Sig.‡	pH $1:1_W$ Mean	Std. Err.†	Sig.‡	pH $1:2_{Ca}$ Mean	Std. Err.†	Sig.‡
Soil Order	Alfosols	73	5.84	0.08	cd	5.87	0.13	de	5.22	0.13	de
	Andisols	46	5.78	0.10	d	6.29	0.16	cd	5.61	0.17	cd
	Aridisols	24	8.70	0.10	a	8.48	0.22	a	8.07	0.23	a
	Entisols	59	7.14	0.08	b	7.02	0.14	b	6.45	0.15	b
	Gelisols	50	6.60	0.08	bc	6.77	0.15	bc	6.18	0.16	bc
	Histosols	27	4.83	0.12	de	5.69	0.21	def	5.20	0.22	de
	Inceptisols	50	6.06	0.09	cd	6.04	0.15	d	5.47	0.16	cd
	Mollisols	93	6.96	0.07	b	6.99	0.11	b	6.40	0.12	b
	Spodosols	39	4.51	0.10	e	4.86	0.17	f	4.27	0.18	f
	Ultisols	56	5.01	0.10	e	5.27	0.14	ef	4.61	0.15	ef
	Vertisols	40	7.19	0.09	b	6.88	0.17	bc	6.43	0.18	b
Mineralogy	Amorphic	20	6.09	0.09	bc	6.52	0.27	bc	5.93	0.28	bc
	Glassy	11	5.46	0.20	bc	5.71	0.36	bcd	5.30	0.38	bcd
	Glassy over amorphic	11	5.42	0.15	bc	6.15	0.36	bcd	5.43	0.38	bcd
	Halloysitic	13	5.72	0.20	bc	5.76	0.33	bcd	5.37	0.35	bcd
	Isotic	34	5.41	0.08	c	5.78	0.21	cd	5.08	0.22	cd
	Kaolinitic	87	5.12	0.07	c	5.31	0.13	d	4.72	0.14	d
	Mixed	259	6.58	0.02	b	6.59	0.07	b	6.02	0.08	b
	Smectitic	62	7.81	0.06	a	7.66	0.15	a	7.19	0.16	a
Family Texture	Clayey	89	6.91	0.05	a	6.77	0.14	a	6.29	0.15	a
	Coarse Loamy	228	6.12	0.03	b	6.31	0.09	b	5.68	0.09	b
	Fine Loamy	60	6.15	0.06	b	6.17	0.17	ab	5.57	0.18	b
	Fine Silty	54	6.39	0.06	ab	6.54	0.18	ab	6.02	0.19	ab
	Sandy	132	5.94	0.04	b	6.12	0.12	b	5.54	0.12	b

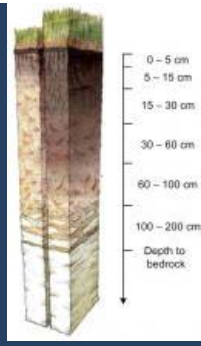
† Std. Err., Standard Error.

‡ Sig, Significance. Letters indicate significance within categories.



Linear Models

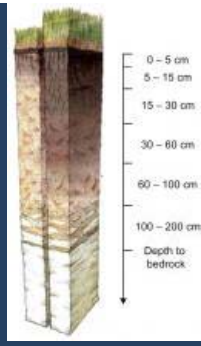




Other Predictive Models

	Regression Equation	Model	
Model	**pH 1:5$_W$ vs. pH 1:2$_{Ca}$**	**R^2**	**RMSE**
Linear	**0.40 + 1.01*pH 1:2$_{Ca}$**	**0.92**	**0.42**
Polynomial (2nd)	**0.44 + 1.00*pH 1:2$_{Ca}$ + 0.01*(pH 1:2$_{Ca}$-5.76)^2**	**0.92**	**0.42**
Polynomial (3rd)	**0.04 + 1.06*pH 1:2$_{Ca}$ + 0.04*(pH 1:2$_{Ca}$-5.76)^2 - 0.01*(pH 1:2$_{Ca}$-5.76)^3**	**0.92**	**0.41**
Smoothing Spline		**0.93**	**0.40**
	pH 1:5$_W$ vs. pH 1:1$_W$		
Linear	**- 0.51 + 1.05*pH 1:1$_W$**	**0.91**	**0.44**
Polynomial (2nd)	**- 0.45 + 1.04*pH 1:1$_W$ + 0.05*(pH 1:1$_W$-6.34)^2**	**0.92**	**0.43**
Polynomial (3rd)	**- 0.74 + 1.08*pH 1:1$_W$ + 0.06*(pH$_W$-6.34)^2 - 0.01*(pH 1:1$_W$-6.34)^3**	**0.92**	**0.42**
Smoothing Spline		**0.93**	**0.41**

Influence of Variables on the prediction

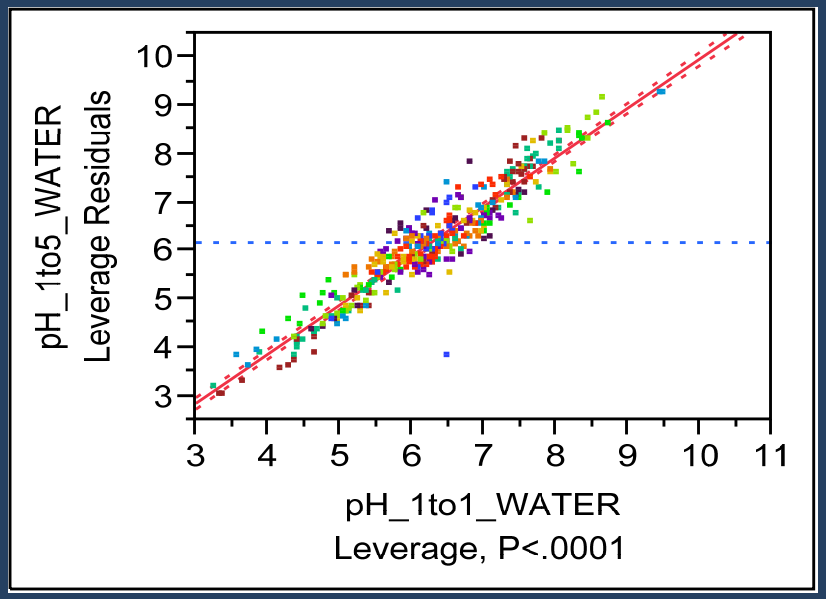


Model Parameters	R^2	RMSE	p value	AICc	BIC
pH $1:1_W$	0.91	0.44	<0.0001	676	689
pH $1:2_{Ca}$	0.92	0.42	<0.0001	621	534
pH $1:1_W$ & $1:2_{Ca}$	0.93	0.40	<0.0001	570	592
Soil Order	0.41	1.16	<0.0001	1760	1811
pH $1:1_W$ & Soil Order	0.94	0.37	<0.0001	489	587
pH $1:2_{Ca}$ & Soil Order	0.95	0.35	<0.0001	441	538
pH $1:1_W$ & $1:2_{Ca}$ & Soil Order	0.96	0.33	<0.0001	383	570
Soil Mineralogy	0.29	1.30	<0.0001	1681	1719
pH $1:1_W$ & Mineralogy Class	0.94	0.39	<0.0001	484	555
pH $1:2_{Ca}$ & Soil Mineralogy	0.94	0.38	<0.0001	473	543
pH $1:1_W$ & $1:2_{Ca}$ & Mineralogy Class	0.95	0.36	<0.0001	433	567
Soil Texture	0.05	1.46	<0.0001	2033	2059
pH $1:1_W$ & Family Textural Class	0.92	0.43	<0.0001	650	697
pH $1:2_{Ca}$ & Family Textural Class	0.93	0.41	<0.0001	596	643
pH $1:1_W$ & $1:2_{Ca}$ & Family Textural class	0.94	0.38	<0.0001	526	615
Genetic Master Horizons	0.003	1.51	0.6035	1972	1994
pH $1:1_W$ & Genetic Master Horizons	0.92	0.44	<0.0001	655	693
pH $1:2_{Ca}$ & Genetic Master Horizons	0.93	0.42	<0.0001	590	628
pH $1:1_W$ & $1:2_{Ca}$ & Genetic Master Horizons	0.94	0.38	<0.0001	546	618
Soil Depth Intervals	0.006	1.50	0.6302	2060	2094
pH $1:1_W$ & Soil Depth Intervals	0.92	0.44	<0.0001	685	7445
pH $1:1_{Ca}$ & Soil Depth Intervals	0.92	0.41	<0.0001	620	680
pH $1:1_W$ & $1:2_{Ca}$ & Soil Depth Intervals	0.93	0.40	<0.0001	572	682

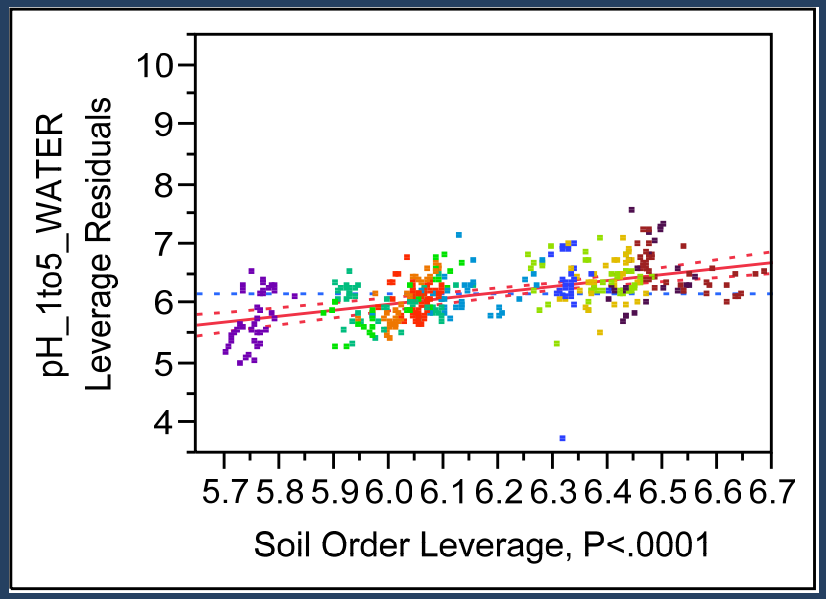
Entering texture class to the prediction increases the R^2 and BIC, however the increase, though significant, is small and has little influence on the prediction and yields a complex prediction equation.

Influence of variables on pH $1{:}5_W$ Prediction (Soil Order)

pH $1{:}1_W$ Influence (Leverage)

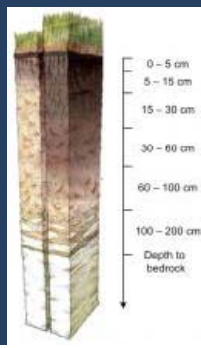


Soil Order Influence (Leverage)



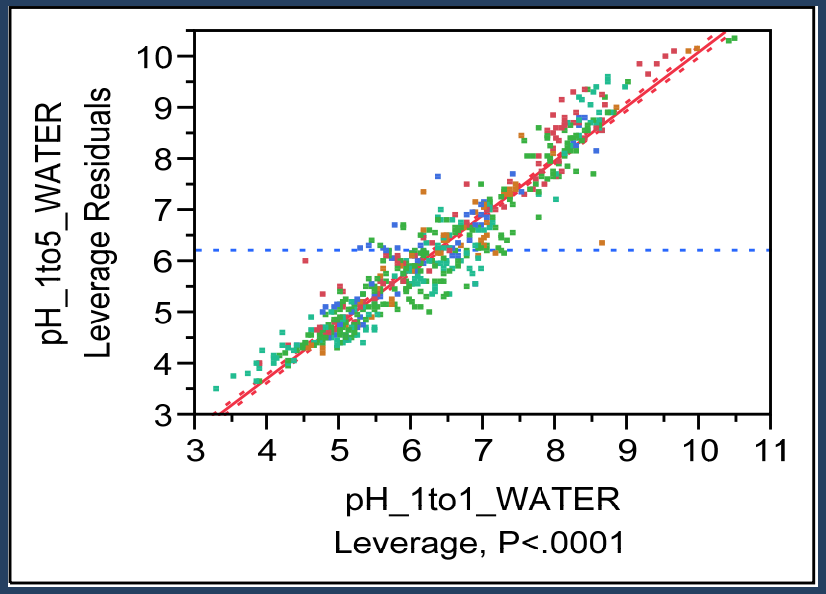
Source	Sum of Squares	F Ratio	Prob > F
Soil Order	23	16.6	<.0001*
pH $1{:}1_W$	463	3286.6	<.0001*

The leverage plot indicates that the continuous variable (pH$1{:}1_W$) has the highest influence on the (pH$1{:}5_W$) compared to categorical variable (Soil Order).

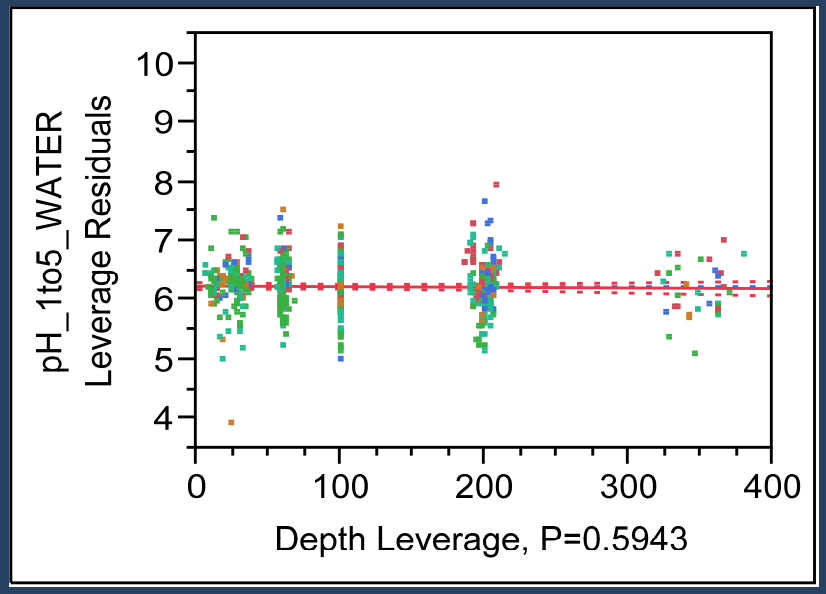


Influence of variables on pH $1{:}5_W$ Prediction (Soil Depth Intervals)

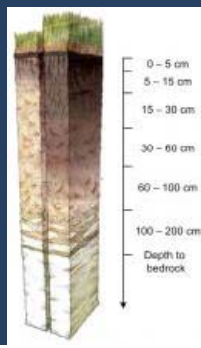
pH $1{:}1_W$ Influence (Leverage)



Soil Depth Influence (Leverage)

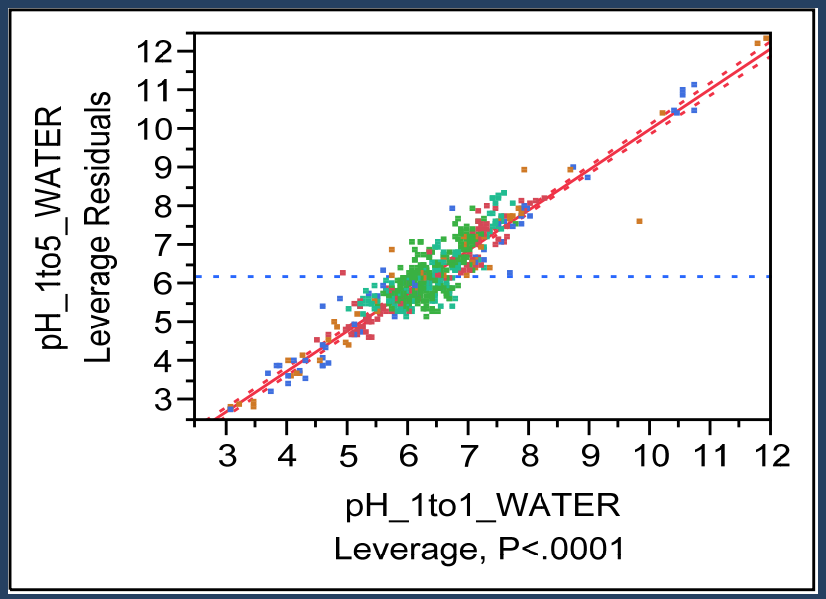


The leverage plot indicates that the continuous variable ($pH1{:}1_W$) has the highest influence on the ($pH1{:}5_W$) compared to variable (Soil Depth Intervals).

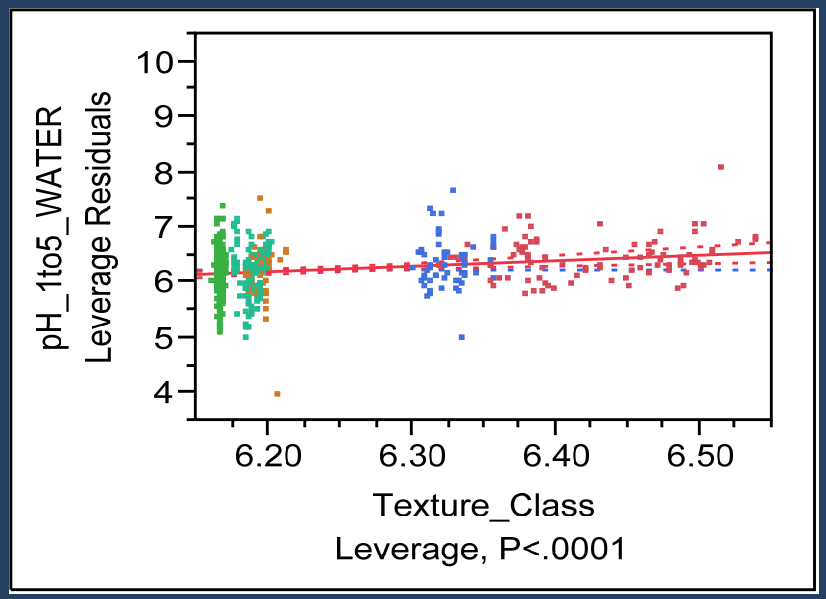


Influence of variables on pH $1{:}5_W$ Prediction (Texture Class)

pH $1{:}1_W$ Influence (Leverage)

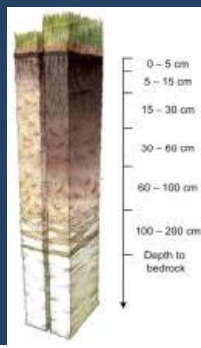


Texture Class Influence (Leverage)

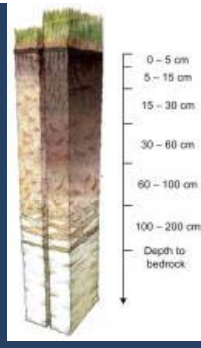


Source	Sum of Squares	F Ratio	Prob > F
pH $1{:}1_W$	692.0	3808.5	<.0001*
Texture Class	5.6	7.7	<.0001*
Texture_Class*pH $1{:}1_W$	1.0	1.4	0.2226

The leverage plot indicates that the continuous variable (pH$1{:}1_W$) has the highest influence on the (pH$1{:}5_W$) compared to categorical variable (Texture Class).



Simple vs. Complex Model



Simple

pH 1:5$_W$ = -0.52 + 1.06* **pH 1:1$_W$**

<u>Without</u> Texture Class

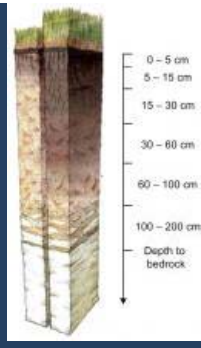
Complex

pH 1:5$_W$ = -0.42 + Match(Texture_Class) + 1.05 * **pH 1:1$_W$** *

"Clayey"	⇒ 0.20
"Coarse Loamy"	⇒ -0.10
"Fine Loamy"	⇒ 0.06
"Fine Silty"	⇒ -0.08
"Sandy"	⇒ -0.08
else	⇒ .

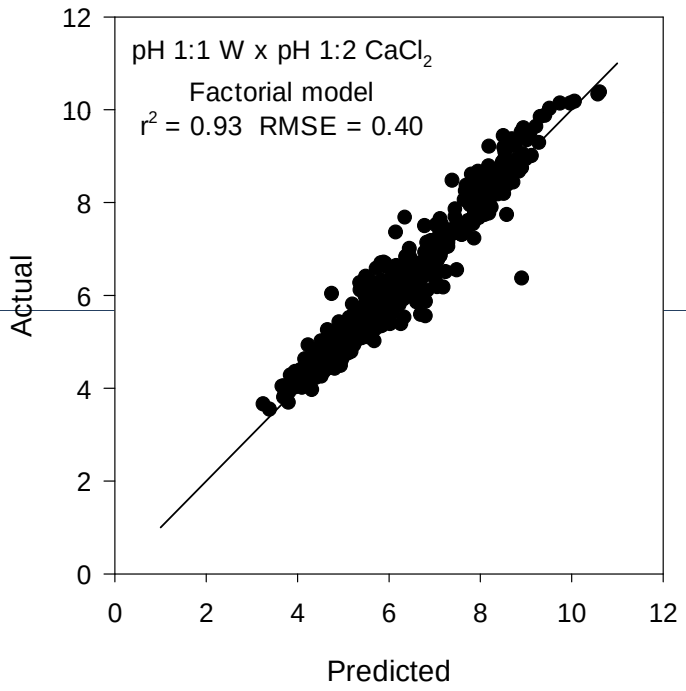
<u>With</u> Texture Class

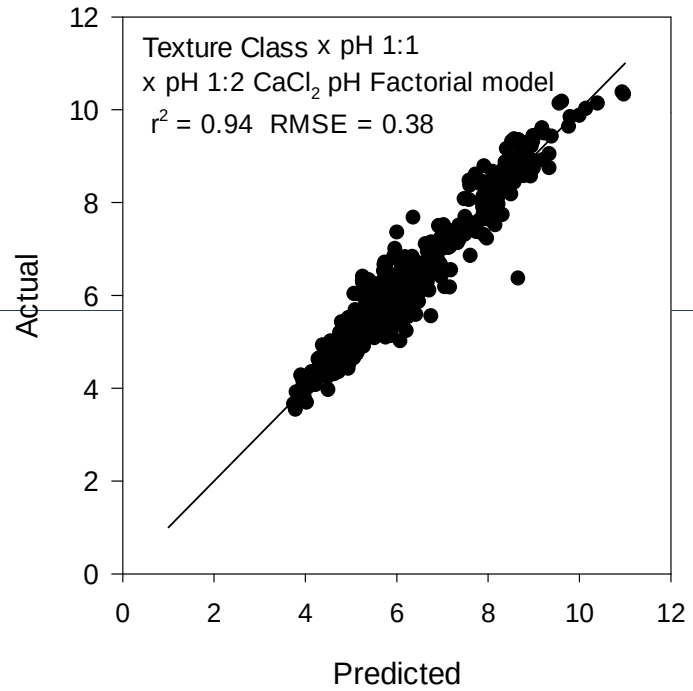
Probability of information loss

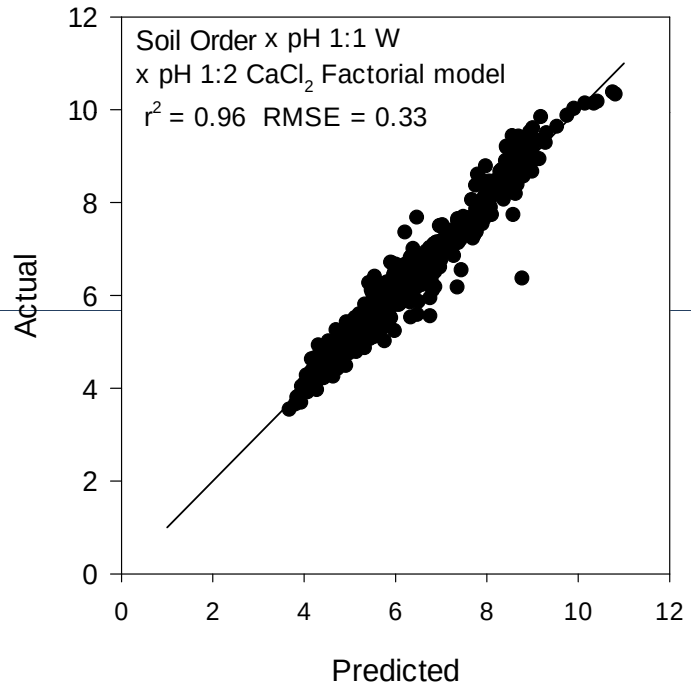


Model Parameter	Prob_AIC
$pH1{:}2_{Ca}$ and $pH1{:}1_{w}$	
$pH1{:}2_{Ca}$ vs. $pH1{:}1_{w}$	9.05762E-13
$pH1{:}2_{Ca}$ + $pH1{:}1_{w}$ vs. $pH1{:}1_{w}$	8.60241E-24
$pH1{:}2_{Ca}$ + $pH1{:}1_{w}$ vs. $pH1{:}2_{Ca}$	9.49743E-12
$pH1{:}2_{Ca}$ and $pH1{:}1_{w}$ and Soil Order (SO)	
$pH1{:}1_{w}$ + SO vs. $pH1{:}1_{w}$	2.43754E-41
$pH1{:}2_{Ca}$ + SO vs. $pH1{:}2_{Ca}$	7.27486E-52
$pH1{:}2_{Ca}$ + $pH1{:}1_{w}$ + SO vs. $pH1{:}1_{w}$	2.65235E-64
$pH1{:}2_{Ca}$ + $pH1{:}1_{w}$ + SO vs. $pH1{:}2_{Ca}$	2.92831E-52
$pH1{:}2_{Ca}$ and $pH1{:}1_{w}$ and Soil Mineralogy (SM)	
$pH1{:}1_{w}$ + SM vs. $pH1{:}1_{w}$	2.01088E-42
$pH1{:}2_{Ca}$ + SM vs. $pH1{:}2_{Ca}$	7.88773E-33
$pH1{:}2_{Ca}$ + $pH1{:}1_{w}$ + SM vs. $pH1{:}1_{w}$	1.18177E-53
$pH1{:}2_{Ca}$ + $pH1{:}1_{w}$ + SM vs. $pH1{:}2_{Ca}$	1.30472E-41
$pH1{:}2_{Ca}$ and $pH1{:}1_{w}$ and Soil Texture (ST)	
$pH1{:}1_{w}$ + ST vs. $pH1{:}1_{w}$	1.93578E-06
$pH1{:}2_{Ca}$ + ST vs. $pH1{:}2_{Ca}$	3.9571E-06
$pH1{:}1_{w}$ + $pH1{:}2_{Ca}$ + ST vs. $pH1{:}1_{w}$	2.28258E-33
$pH1{:}1_{w}$ + $pH1{:}2_{Ca}$ + ST vs. $pH1{:}2_{Ca}$	2.52007E-21
$pH1{:}2_{Ca}$ and $pH1{:}1_{w}$ and Genetic Master Horizons (GMH)	
$pH1{:}1_{w}$ + GMH vs. $pH1{:}1_{w}$	1.95996E-05
$pH1{:}2_{Ca}$+ GMH vs. $pH1{:}2_{Ca}$	2.3352E-07
$pH1{:}1_{w}$ + $pH1{:}2_{Ca}$ + GMH vs. $pH1{:}1_{w}$	6.07978E-29
$pH1{:}1_{w}$ + $pH1{:}2_{Ca}$ + GMH vs. $pH1{:}2_{Ca}$	6.71233E-17
$pH1{:}2_{Ca}$ and $pH1{:}1_{w}$ and Soil Depth Intervals (SDI)	
$pH1{:}1_{w}$ + SDI vs. $pH1{:}1_{w}$	0.013167547
$pH1{:}2_{Ca}$ + SDI vs. $pH1{:}2_{Ca}$	0.568360147
$pH1{:}1_{w}$ + $pH1{:}2_{Ca}$ + SDI vs. $pH1{:}1_{w}$	2.3501E-23
$pH1{:}1_{w}$ + $pH1{:}2_{Ca}$ + SDI vs. $pH1{:}2_{Ca}$	2.59461E-11

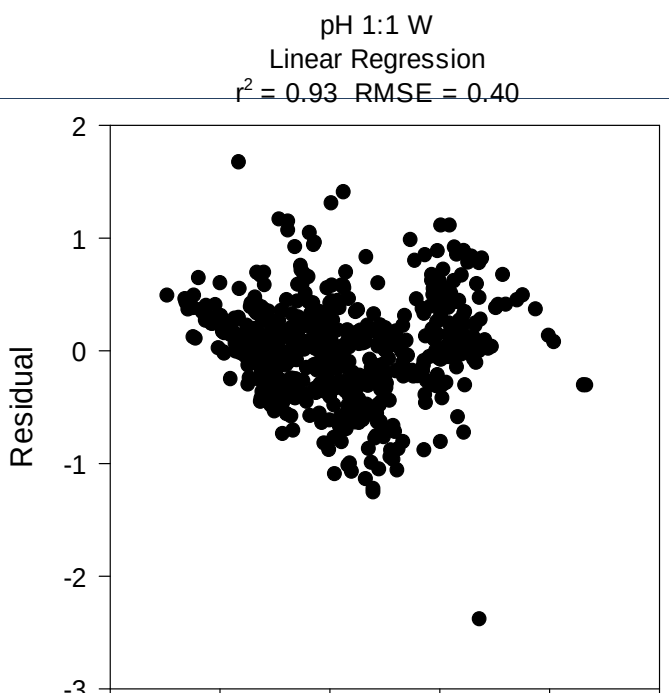
Linear Regression Models

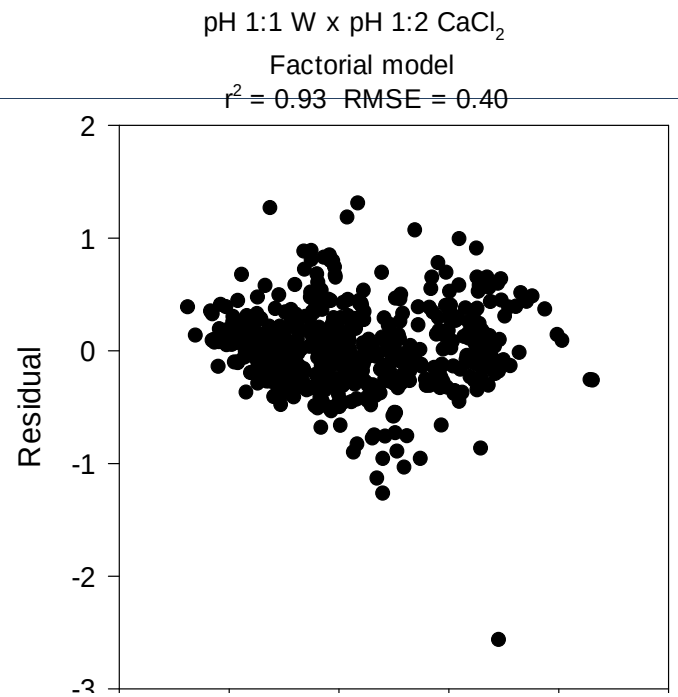


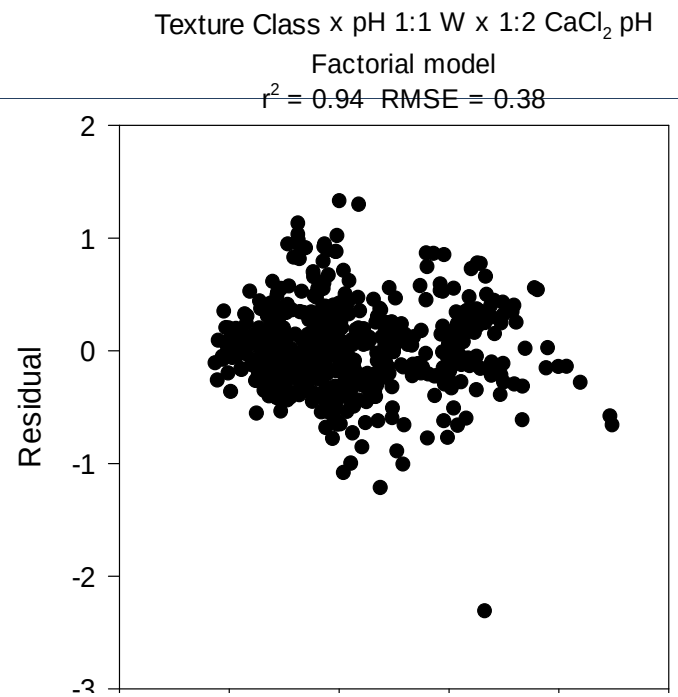




Residuals of Measured pH (Actual) versus Predicted pH by different factorial models

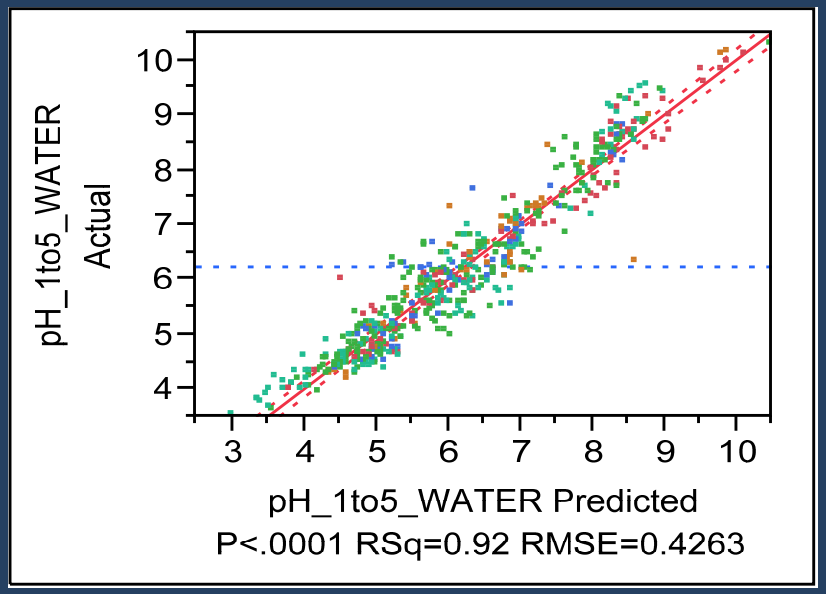




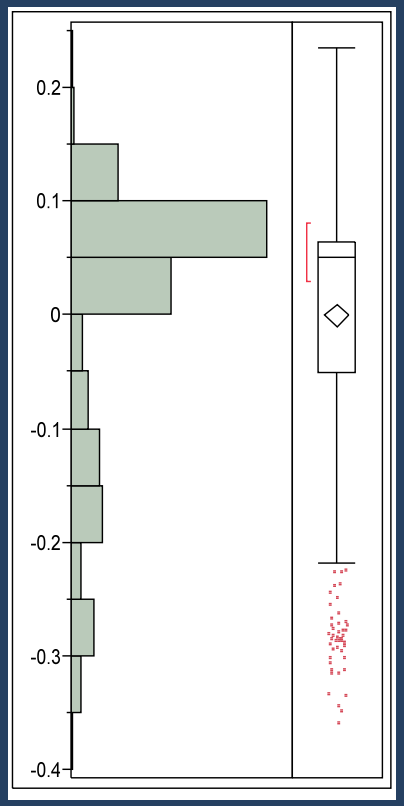


Overall Prediction Error

Actual vs. Predicted Plot



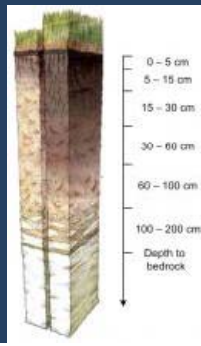
Distribution of pH1:5$_W$ predicted by pH 1:1$_W$ only – pH1:5$_W$ predicted by pH 1:1$_W$ **plus** texture class.



Quintiles

100.0%	Maximum	0.23462
99.5%		0.19457
97.5%		0.13105
90.0%		0.1032
75.0%	Quartile	0.06436
50.0%	**Median**	**0.0508**
25.0%	Quartile	-0.0504
10.0%		-0.1891
2.5%		-0.2944
0.5%		-0.3451
0.0%	Minimum	-0.359

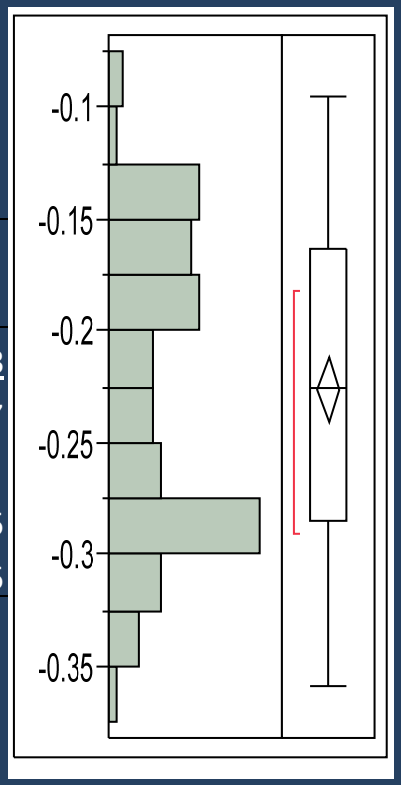
Overall the median difference is only 0.05 small and the majority fall within this range.



“Clayey” Prediction Error

Distribution of pH1:5$_W$ predicted by pH 1:1$_W$ only – pH1:5$_W$ predicted by pH 1:1$_W$ **plus** “Clayey” texture class.

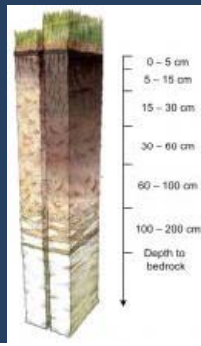
	Measured pH 1:5 W		Predicted pH 1:5 $_W$		
Texture Class	Mean	Std Dev	Mean	Std Dev	Diff.
Clayey	**6.91**	**1.68**	**6.69**	**1.58**	**-0.23**
Coarse Loamy	6.12	1.40	6.19	1.38	-0.07
Fine Loamy	6.15	1.15	6.05	1.13	0.10
Fine Silty	6.39	1.32	6.44	1.29	-0.05
Sandy	5.94	1.58	5.99	1.52	-0.05

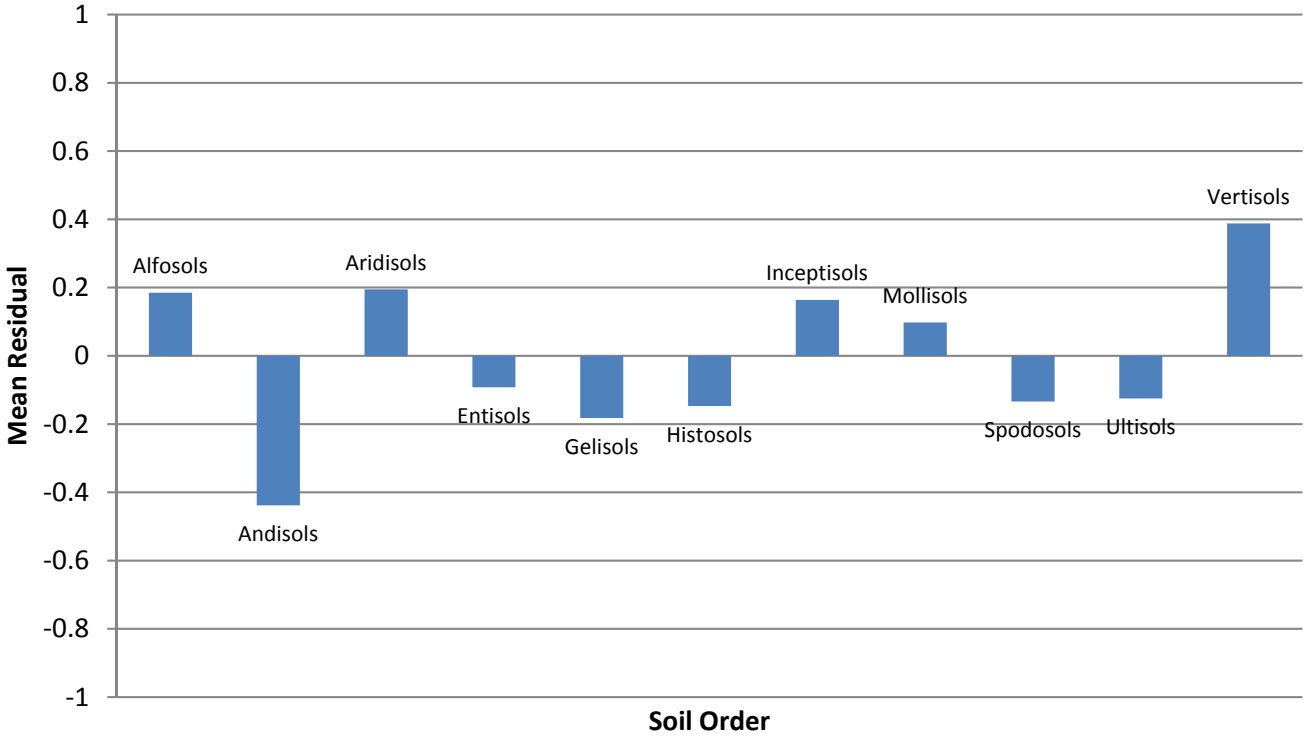


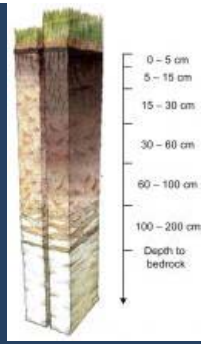
Quintiles

100.0%	Maximum	-0.0956
99.5%		-0.0956
97.5%		-0.1006
90.0%		-0.1402
75.0%	Quartile	-0.1639
50.0%	**Median**	**-0.2257**
25.0%	Quartile	-0.2845
10.0%		-0.3121
2.5%		-0.3477
0.5%		-0.359
0.0%	Minimum	-0.359

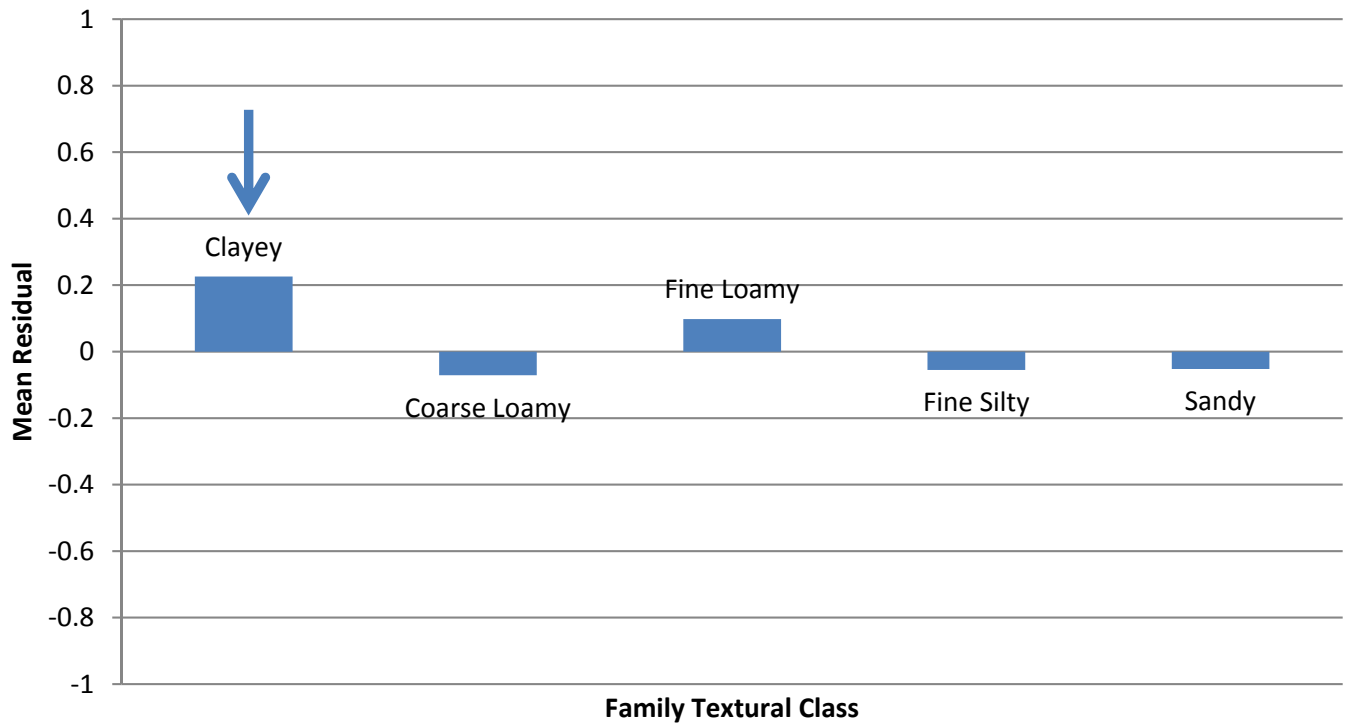
“Clayey” seems to be the most influential .







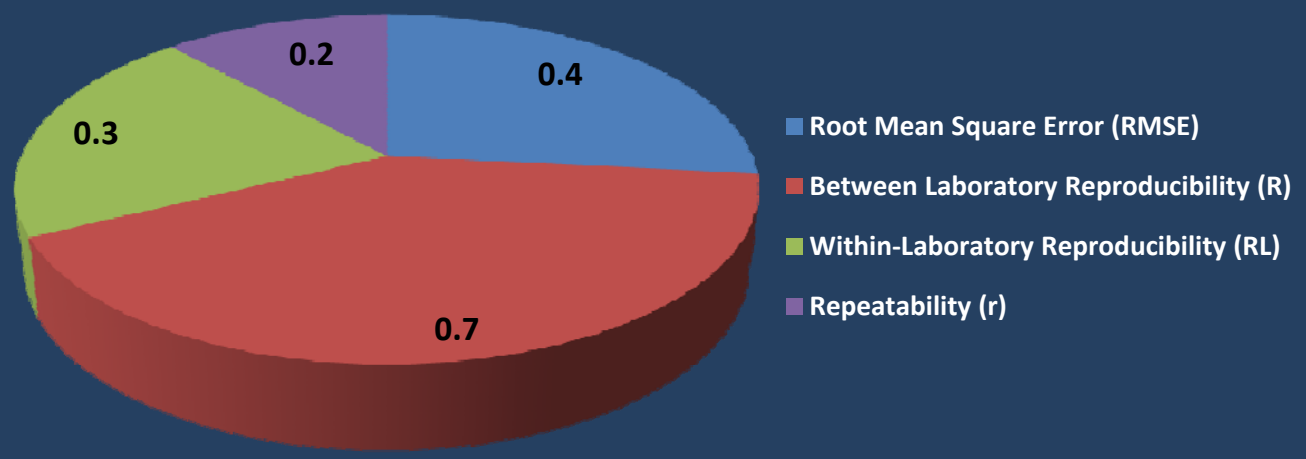
Error Distribution



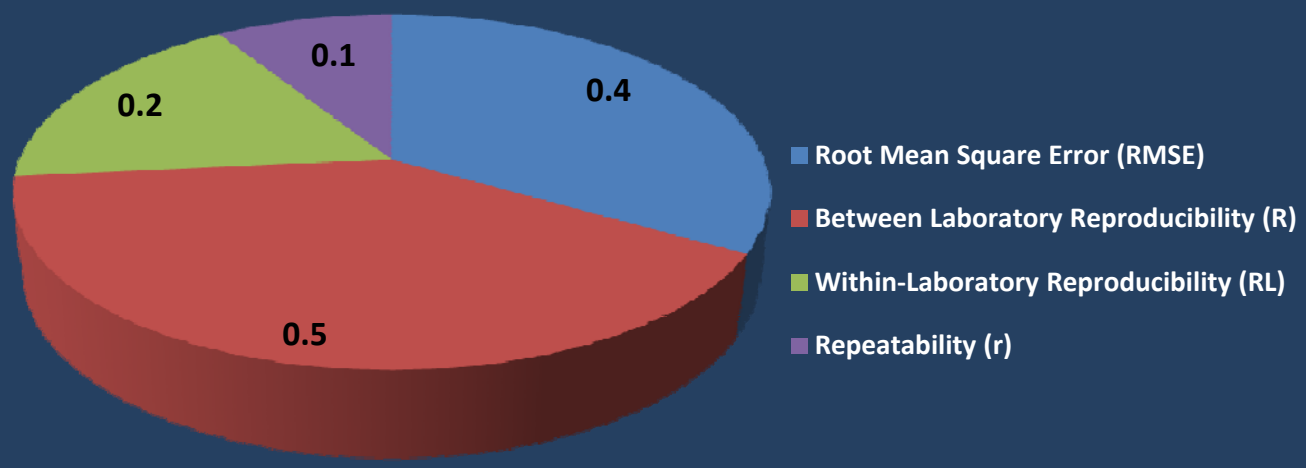
GlobalSoilMap.net

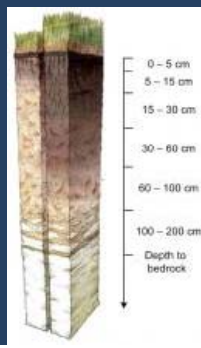
Error source and magnitude associated with pH $1{:}1_W$ and pH $1{:}2_{Ca}$ predictors and analytical pH methods

pH $1{:}1_W$

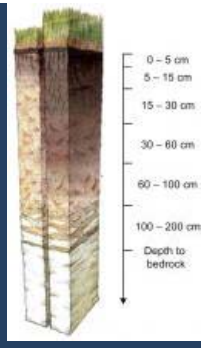


pH $1{:}2_{Ca}$





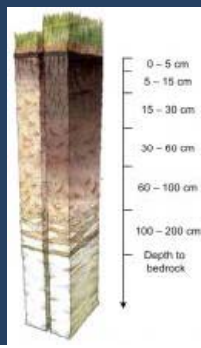
Conclusions



- The simple linear regression model using only pH $1{:}1_W$ or pH $1{:}2_{Ca}$ should be used for predicting pH $1{:}5_W$.
- Soil orders, mineralogy classes and family textural classes explained a small amount of measured pH $1{:}5_W$ variability.
- The errors associated with these predictive models are comparable to those associated with the analytical methods alone.
- The simplicity of the linear model with pH $1{:}1_W$ or pH $1{:}2_{Ca}$ as the only predictor outweighed the benefits of adding more categorical parameters (soil order, mineralogy classes or family textural classes) to the model.
- The lack of significant differences for genetic master horizons and depth intervals indicate that these simple linear regression equations are adequate to generate a predictive soil pH $1{:}5_W$ map for standards depths (0-5; 5-15; 15-30; 30-60; 60-100; and 100-200cm) for GlobalSoilMap.net.

Preliminary/(Pre) mature Conclusions

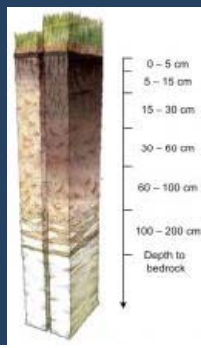
- There is a highly significant correlation between measured pH 1:1 Water/1:2 $CaCl_2$ (0.01M) and predicted pH 1:5 Water;
- Categorical variables (Soil Order and Texture Class) influence the prediction;
- Their influence is small relative to the measured pH 1:1 Water/1:2 $CaCl_2$ (0.01M) and its inclusion in the prediction equation need to be evaluated against increased model complexity;
- The “Clayey” texture class is the most influential in relative terms compared to the other texture classes;
- Mineralogy (not shown); Genetic Horizons (not shown) and Soil Depth Intervals do not significantly influence the prediction;
- Further Statistical Analysis are being conducted.

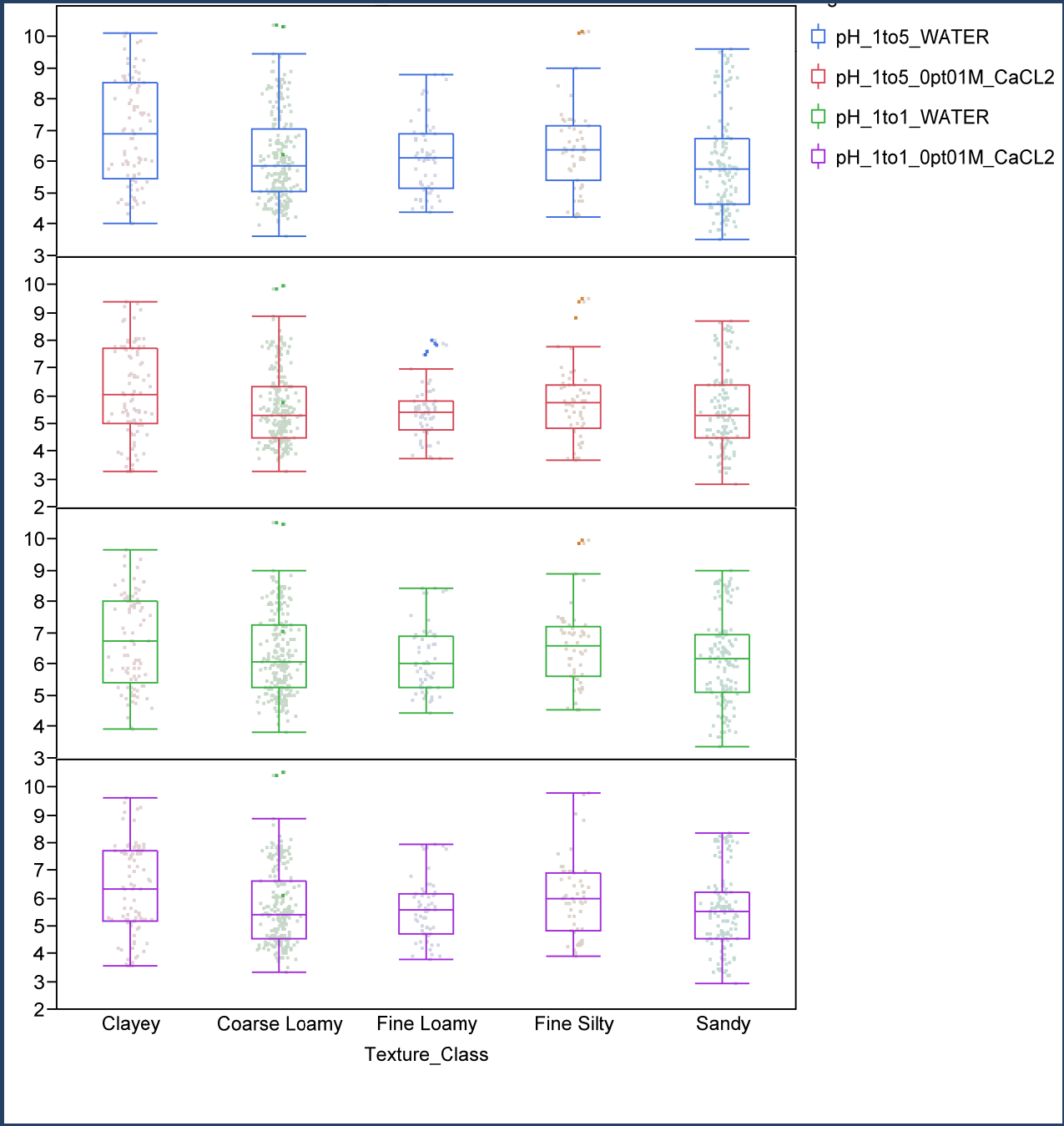


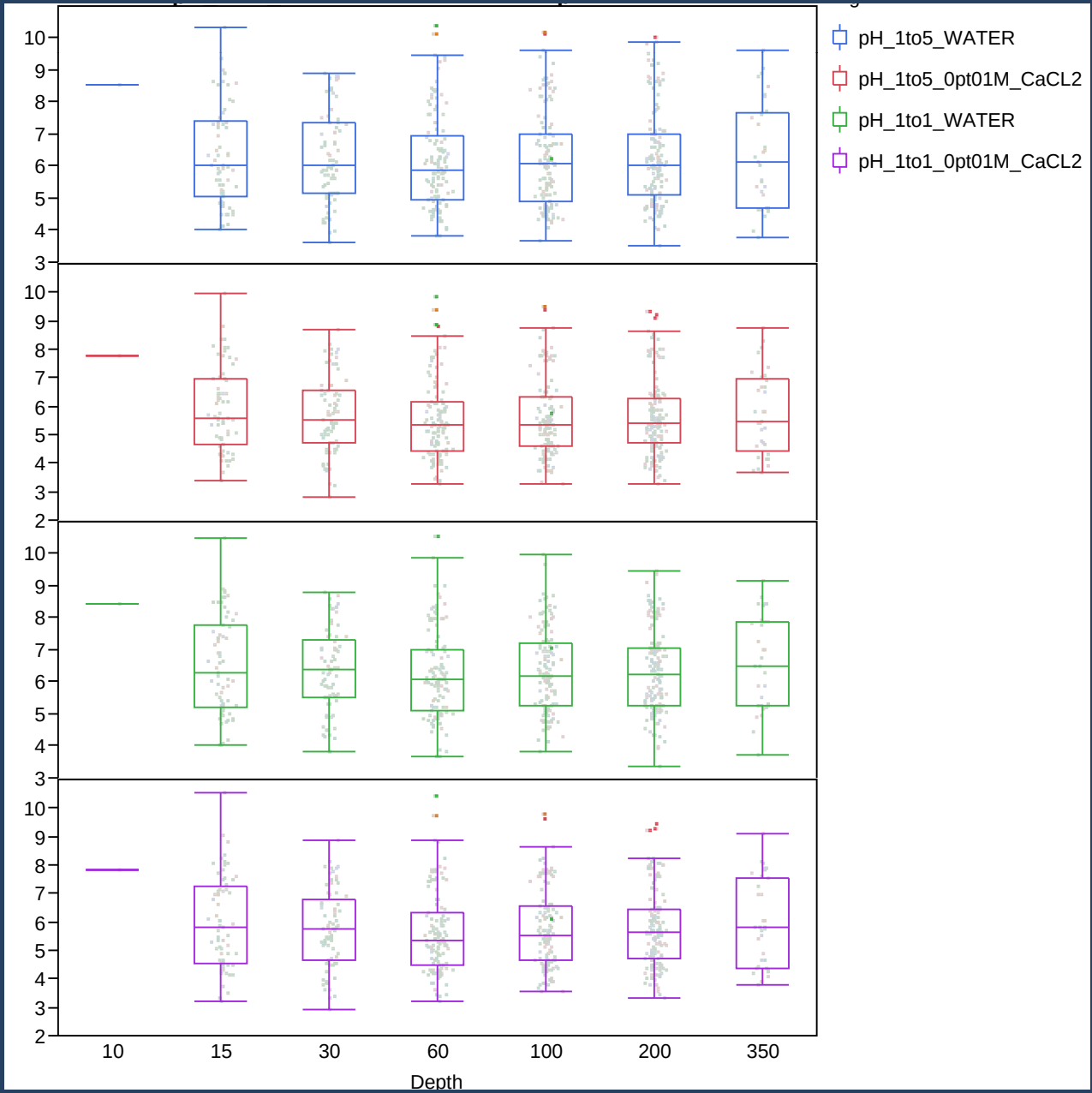
Stepwise Forward selection for Texture Class

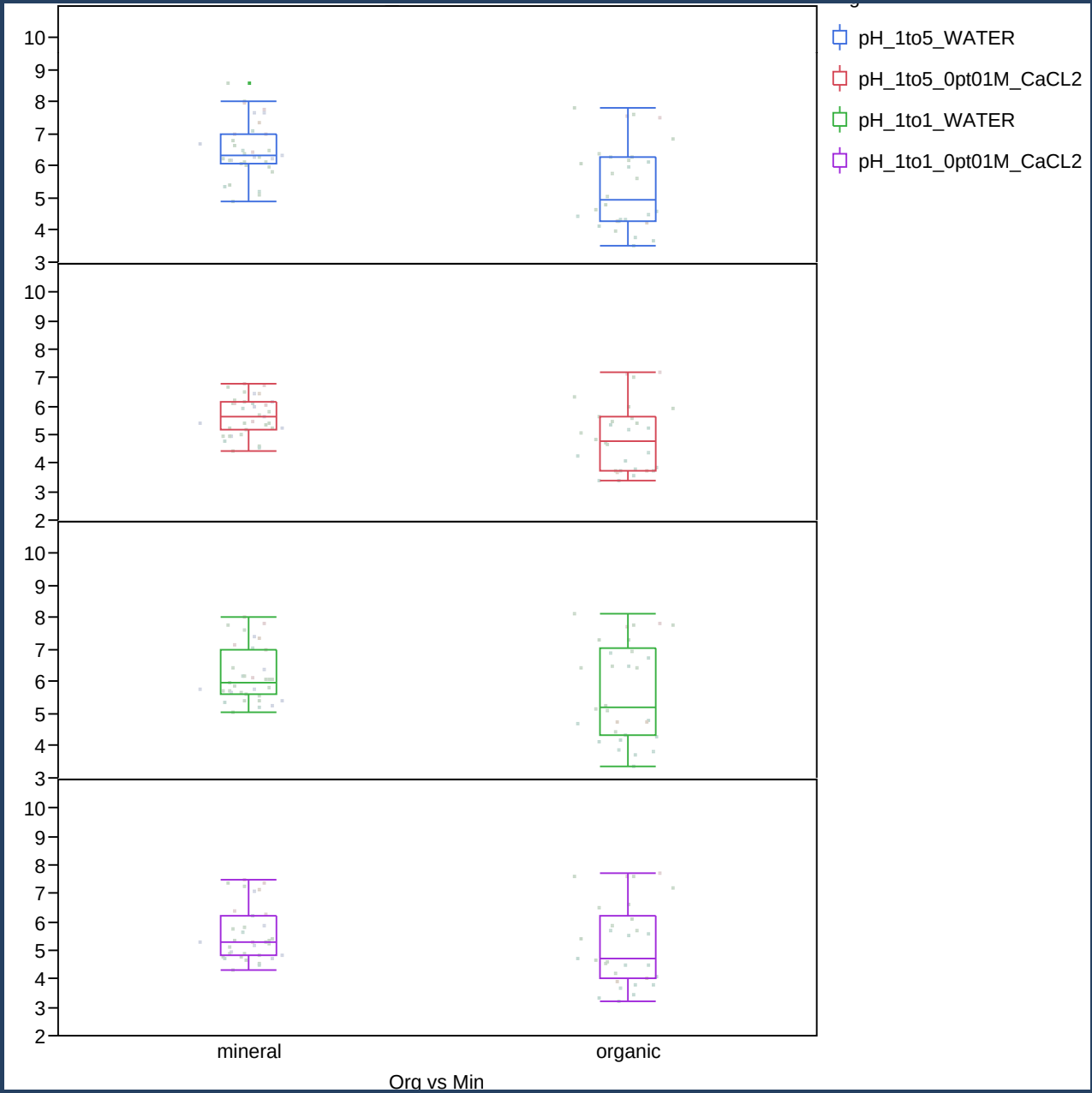
Step	Parameter	Action	Prob.	Seq. SS	R-Square	Cp	AICc	BIC
1	**pH_1to1_WATER**	**Entered**	**0.00**	**1142.9**	**0.913**	**37.6**	**676**	**689**
2	Texture Class {Sandy&Coarse Loamy&Fine Loamy&Fine Silty-Clayey}	Entered	0.00	6.9	0.919	7.7	647	677
3	(pH_1to1_WATER-6.34441)*Texture_Class {Sandy-Coarse Loamy&Fine Loamy&Fine Silty}	Entered	0.22	1.0	0.917	10.0	649	697
4	**Best**	**Specific**			**0.919**	**7.7**	**647**	**677**

Entering texture class to the prediction increases the R^2 and BIC, however the increase, though significant, is small and has little influence on the prediction and yields a complex prediction equation.




pH_1to5_WATER
pH_1to5_0pt01M_CaCL2
pH_1to1_WATER
pH_1to1_0pt01M_CaCL2
Clayey
Coarse Loamy
Fine Loamy
Fine Silty
Sandy
Texture_Class


pH_1to5_WATER
pH_1to5_0pt01M_CaCL2
pH_1to1_WATER
pH_1to1_0pt01M_CaCL2
10
15
30
60
100
200
350
Depth


pH_1to5_WATER
pH_1to5_0pt01M_CaCL2
pH_1to1_WATER
pH_1to1_0pt01M_CaCL2
mineral
organic
Org vs Min