Performance-Based Methods and In-House Method Validation

CCME PHC's Validation Issues

CWS CCME PHC Method contains essential elements for establishing method equivalence

Appendix 2 – Method Validation

- Calls for analysis of four reference samples in at least triplicate using the exact CWS PHC reference method to compare.

- Data for both methods must (be) within 20% for all samples

Current equivalence protocols could be improved.

Potential Issues with Validation/Equivalence Protocols

Appendix 2 – Method Validation

- Calls for analysis of four reference samples in at least triplicate using the exact CWS PHC reference method to compare.

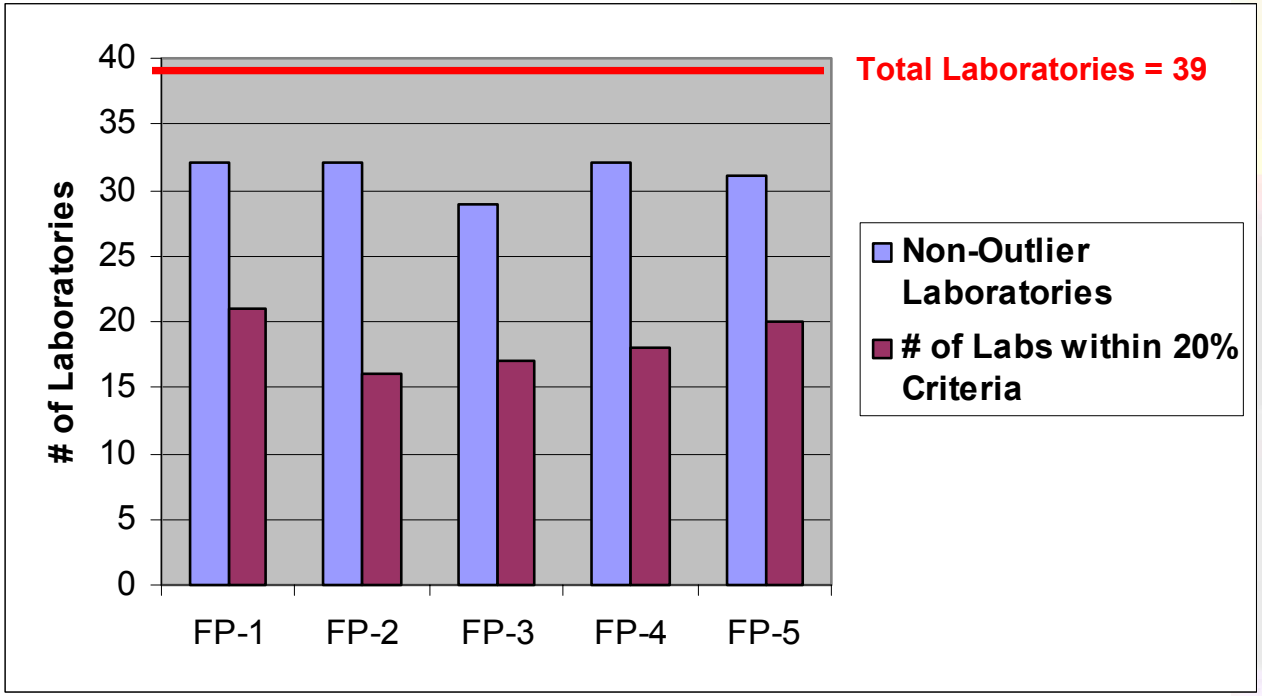
  Most likely very large variability in reference soils. NOC/humic matter, % clay, hydrocarbon characteristics, contaminated soil samples versus spikes etc.

- Data for both methods must (be) within 20% for ***all*** samples.

  Arbitrary limits that may not reflect the variance of the reference method.

  Only 5/39 (2/soxhlet, 3/non-soxhlet) of Participating Laboratories had *all* samples within 20% of the reference value for F3. Using the current protocol only 13% of the labs would actually pass the criteria of having no samples fail.

Less than 50% of the laboratories participating in the recent Inter- laboratory study had results within 20% of the Reference Value for the F3 Fraction



Potential Issues with Validation/Equivalence Protocols

What does "Within 20% for all samples" Mean

- Do the averages of each *reference sample* or *all individual samples* have to be within 20% ?

- Is "Within 20%" mean RPD or % difference versus soxhlet ?

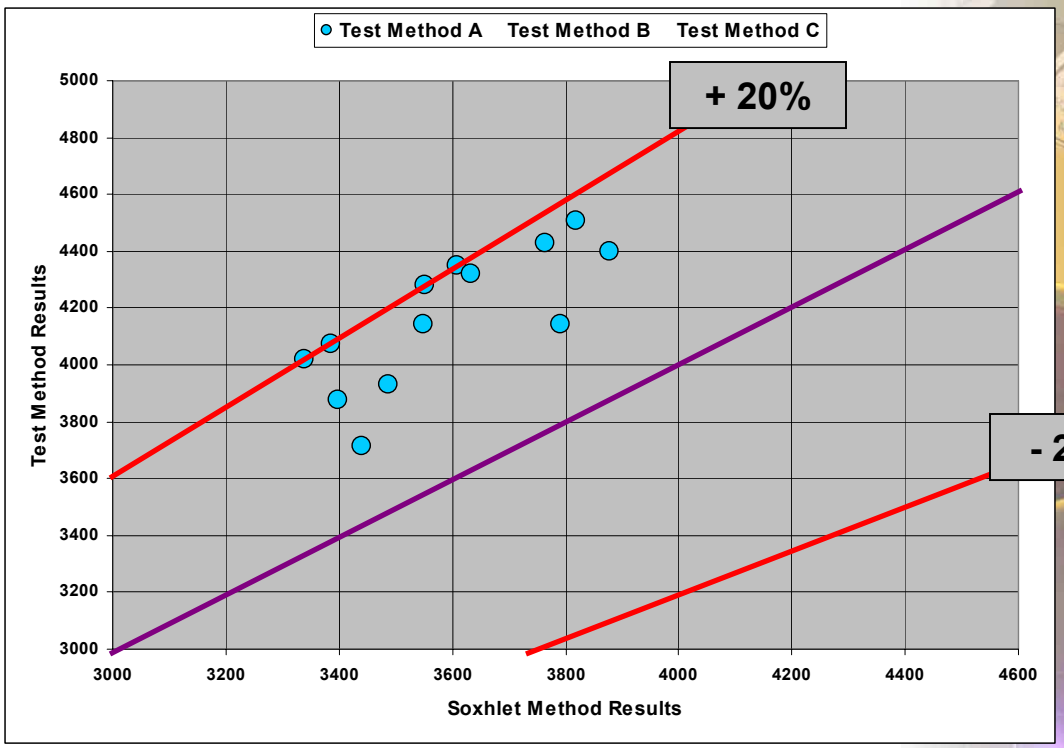
Soxhlet A	Test A	Dif	RPD%	Test/ Soxhlet
3550	4141	-591	15	117
3551	4280	-729	19	121
3440	3715	-275	8	108
3609	4348	-739	19	120
3385	4073	-688	18	120
3764	4429	-665	16	118
3791	4141	-350	9	109
3340	4022	-682	19	120
3818	4505	-687	17	118
3397	3875	-478	13	114
3487	3933	-446	12	113
3879	4401	-522	13	113
3634	4320	-686	17	119

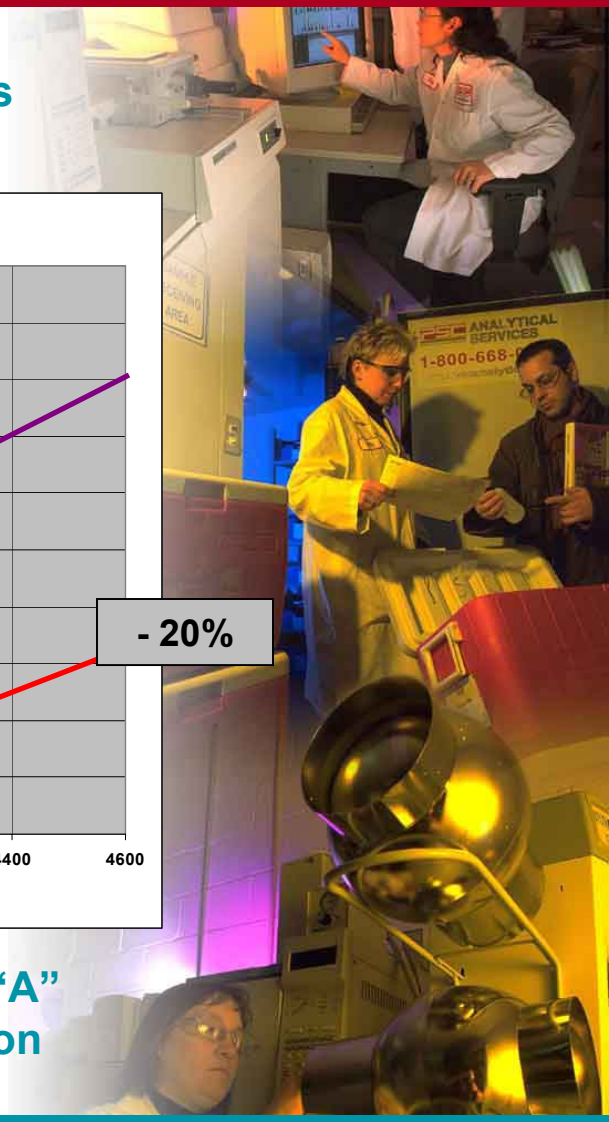
Is there a need for a standard reference material for CWS users to demonstrate equivalence?

Method “A” Results – All samples within 20% Criteria but is the method equivalent to soxhlet?

Results

Soxhlet A	Test A	Dif	RPD
3550	4141	-591	15
3551	4280	-729	19
3440	3715	-275	8
3609	4348	-739	19
3385	4073	-688	18
3764	4429	-665	16
3791	4141	-350	9
3340	4022	-682	19
3818	4505	-687	17
3397	3875	-478	13
3487	3933	-446	12
3879	4401	-522	13
3634	4320	-686	17





Although samples pass 20% criteria Test Method “A” produces biased high results with similar precision

Method "A" Results – Differences between the two test populations are statistically different than zero.

	Soxhlet A	Test A	Dif	RPD%	Test/ Soxhlet
	3550	4141	-591	15	117
	3551	4280	-729	19	121
	3440	3715	-275	8	108
	3609	4348	-739	19	120
	3385	4073	-688	18	120
	3764	4429	-665	16	118
	3791	4141	-350	9	109
	3340	4022	-682	19	120
	3818	4505	-687	17	118
	3397	3875	-478	13	114
	3487	3933	-446	12	113
	3879	4401	-522	13	113
	3634	4320	-686	17	119
Average	3588	4168		15	
StDev	179	238		1.3	
Avg-Differences			-579.8		
StDev of Differences			152.0		
alpha (two-sided)=			0.05		
# of paired observations=			13		
degrees of freedom=			12		
t(calc)=			-13.755		
t(critical)=			2.201		
reject null hypotheses=			yes		

Similar Precision

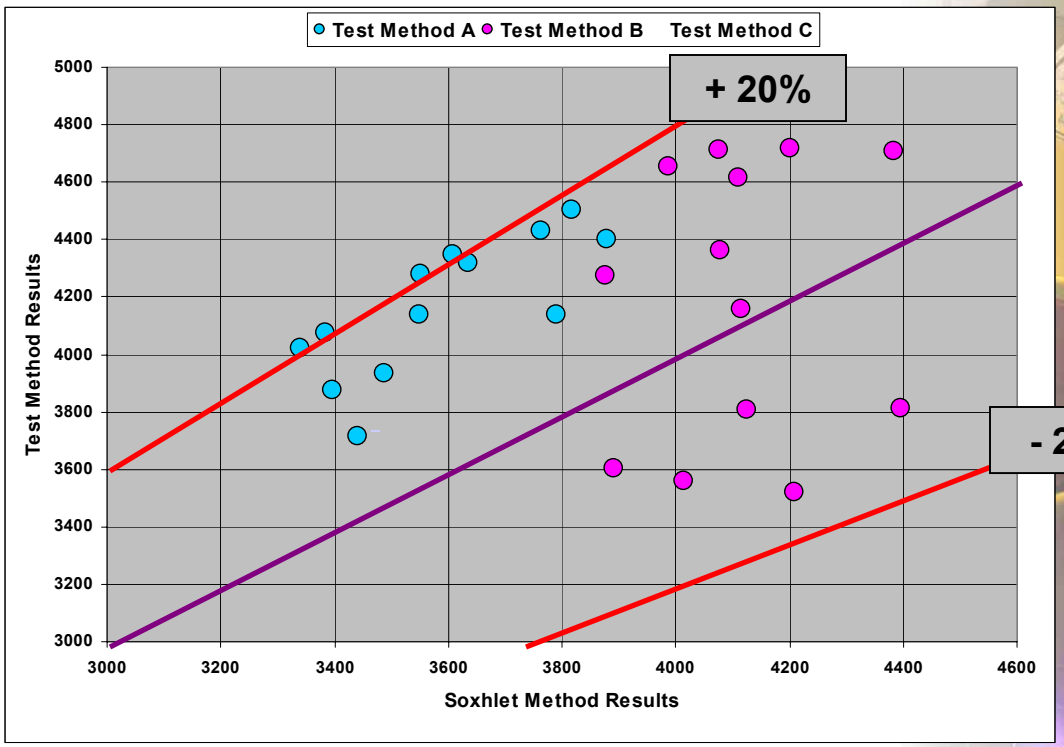
The Null Hypothesis can be rejected so the difference of the means of the between the two populations may be different than zero.

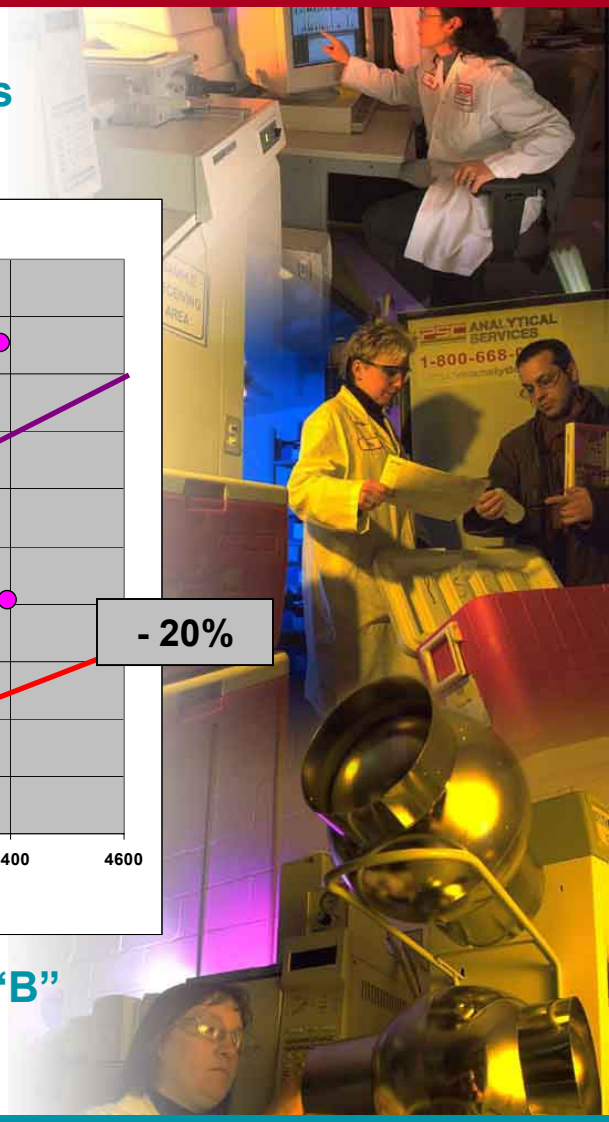
Statistically speaking the methods are different but they would still pass the CWS criteria.

Method "B" Results – All samples within 20% Criteria but is the method equivalent to Reference Method?

Results

Soxhlet B	Test B	Dif	RPD
4396	3815	581	14
4114	4158	-44	1
3892	3605	287	8
4209	3520	689	18
4013	3559	454	12
4126	3809	317	8
4076	4715	-639	15
4078	4363	-285	7
3876	4274	-398	10
4109	4615	-506	12
3987	4655	-668	15
4200	4716	-516	12
4384	4706	-322	7





Although samples pass 20% criteria Test Method "B" produces results with poorer precision

Method "B" Results – Differences between the two test populations are not statistically different than zero.

	Soxhlet B	Test B	Dif	RPD
	4396	3815	581	14
	4114	4158	-44	1
	3892	3605	287	8
	4209	3520	689	18
	4013	3559	454	12
	4126	3809	317	8
	4076	4715	-639	15
	4078	4363	-285	7
	3876	4274	-398	10
	4109	4615	-506	12
	3987	4655	-668	15
	4200	4716	-516	12
	4384	4706	-322	7
	4112	**4193**	**-81**	**11**
	159	**477**		**3.0**
	Avg-Differences		-80.8	
	StDev of Differences		486.4	
	alpha (two-sided)=		0.05	
# of paired observations=			13	#
	degrees of freedom=		12	
		t(calc)=	-0.599	
		t(critical)=	2.201	
	reject null hypotheses=		no	

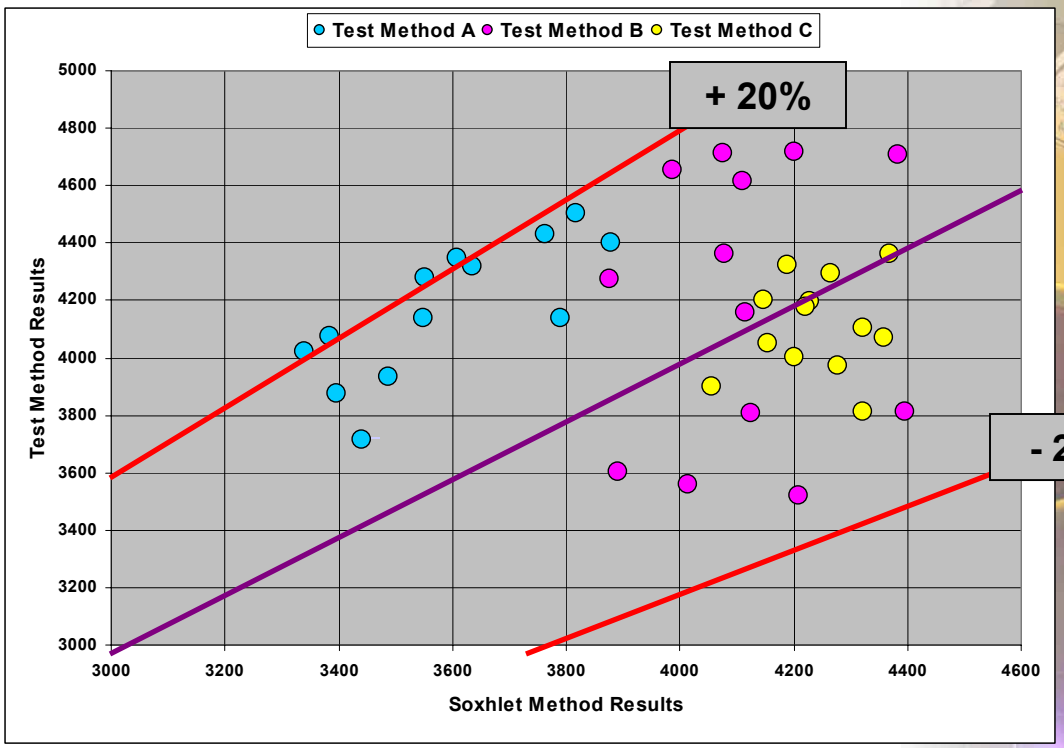
Method "B" data has a SD approximately 3x the Reference method. Is this acceptable?

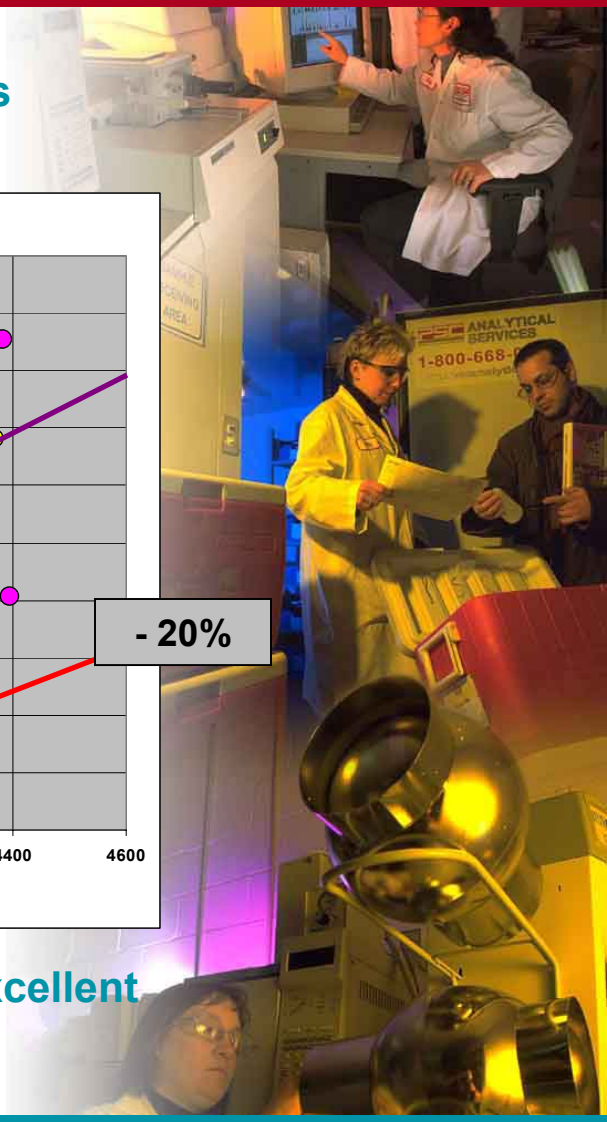
The Null Hypothesis cannot be rejected so the difference of the means of the between the two populations may not be different than zero. Statistically speaking the methods may be equivalent.

Method "C" Results – All samples within 20% Criteria but is the method equivalent to Reference Method?

Results

Soxhlet C	Test C	Dif	RPD
4229	4195	34	1
4368	4363	5	0
4321	4107	214	5
4266	4295	-29	1
4147	4200	-53	1
4189	4326	-137	3
4200	4001	199	5
4278	3974	304	7
4321	3812	509	13
4154	4050	104	3
4056	3901	155	4
4221	4180	41	1
4358	4070	288	7
4239	**4113**	**126**	**4**
91	**167**		**1.8**





All Method "C" samples pass 20% criteria and have excellent precision versus Reference Method

Method "C" Results – Differences between the two test populations are not statistically different than zero.

Soxhlet C	Test C	Dif	RPD
4229	4195	34	1
4368	4363	5	0
4321	4107	214	5
4266	4295	-29	1
4147	4200	-53	1
4189	4326	-137	3
4200	4001	199	5
4278	3974	304	7
4321	3812	509	13
4154	4050	104	3
4056	3901	155	4
4221	4180	41	1
4358	4070	288	7
4239	**4113**	**126**	**4**
91	**167**		**1.8**
Avg-Differences		125.7	
v of Differences		176.7	
pha (two-sided)=		0.05	
l observations=		13	
es of freedom=		11	
t(calc)=		2.565	
t(critical)=		2.201	
ull hypotheses=		no	

Method "C" data has a SD < 2x the Reference method.

The Null Hypothesis cannot be rejected so the difference of the means of the between the two populations may not be different than zero. Statistically speaking the methods may be equivalent.

For the labs using soxhlet and hexane/acetone the average relative standard deviation for the five F3 samples was ~ 25% (outliers excluded).

Summary and Recommendations

- The current protocol doesn't work, thus we need a new one
- It must be practical, 30 data points = infinity
- Guidelines
  - Acceptable PE performance (CAEAL?) and audit
  - Precision equal or better than reference method
  - Accuracy (recovery) equal or somewhat better (20%) than ref.

Further Considerations

- The agreed deviation validation protocol will become the de facto standard for all performance based methods
- We need something simple and doable that is at the same time technically and legally defensible