

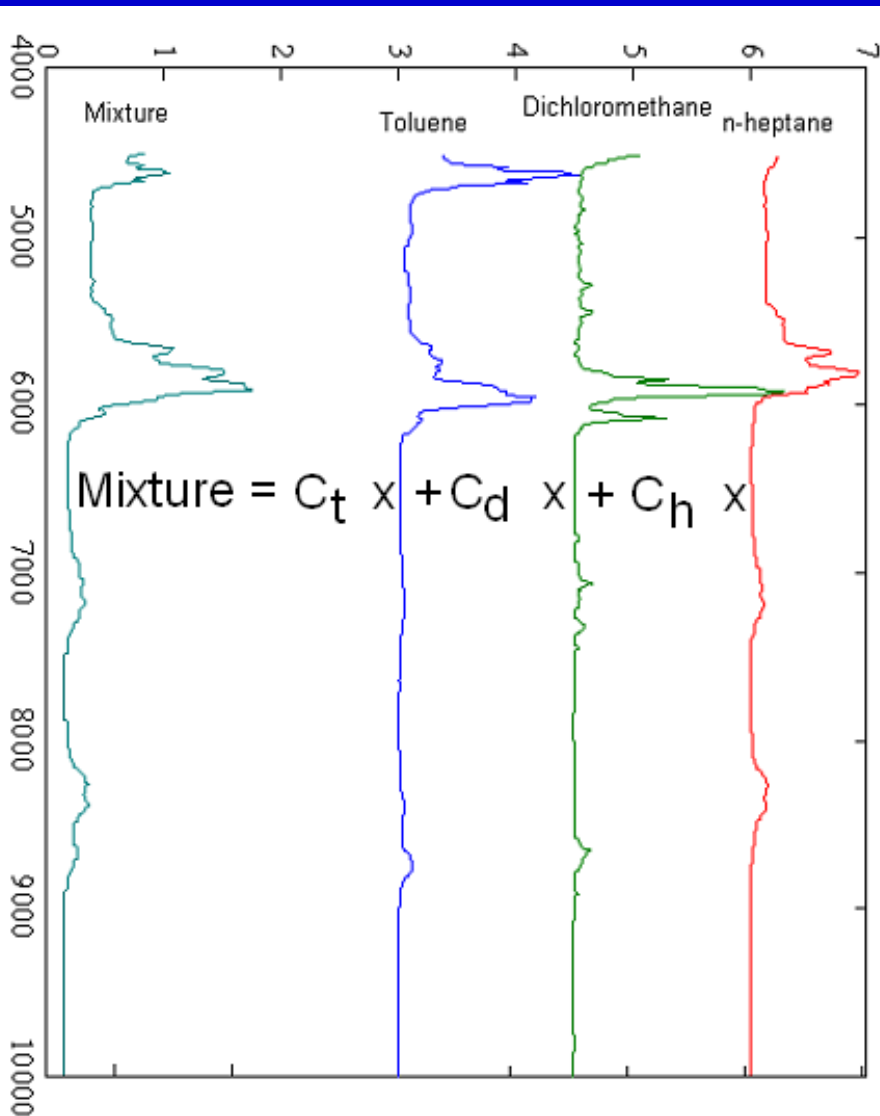
# VOLUME Fractions in Chemometric Calibration

By

**Howard Mark**  
Ron Rubinovitz  
Donald Dahm  
Kevin Dahm  
David Heaps  
Paul Gemeperline

Presented at Pittcon 2011  
Atlanta, Georgia

# Basis of CLS



$$\text{Mixture} = c_t \times \text{Toluene} + c_d \times \text{Dichlor} + c_h \times \text{n-heptane}$$

$$0.4347 = c_t \times 0.4360 + c_d \times 0.5962 + c_h \times 0.3051$$

$$0.4126 = c_t \times 0.4319 + c_d \times 0.5456 + c_h \times 0.2975$$

$$0.3939 = c_t \times 0.4262 + c_d \times 0.5058 + c_h \times 0.2906$$

$$0.3783 = c_t \times 0.4204 + c_d \times 0.4744 + c_h \times 0.2844$$

$$0.3655 = c_t \times 0.4164 + c_d \times 0.4492 + c_h \times 0.2789$$

$$0.3554 = c_t \times 0.4164 + c_d \times 0.4274 + c_h \times 0.2741$$

$$0.3475 = c_t \times 0.4218 + c_d \times 0.4064 + c_h \times 0.2698$$

$$0.3410 = c_t \times 0.4321 + c_d \times 0.3841 + c_h \times 0.2660$$

$$0.3349 = c_t \times 0.4452 + c_d \times 0.3595 + c_h \times 0.2625$$

$$0.3286 = c_t \times 0.4592 + c_d \times 0.3332 + c_h \times 0.2593$$

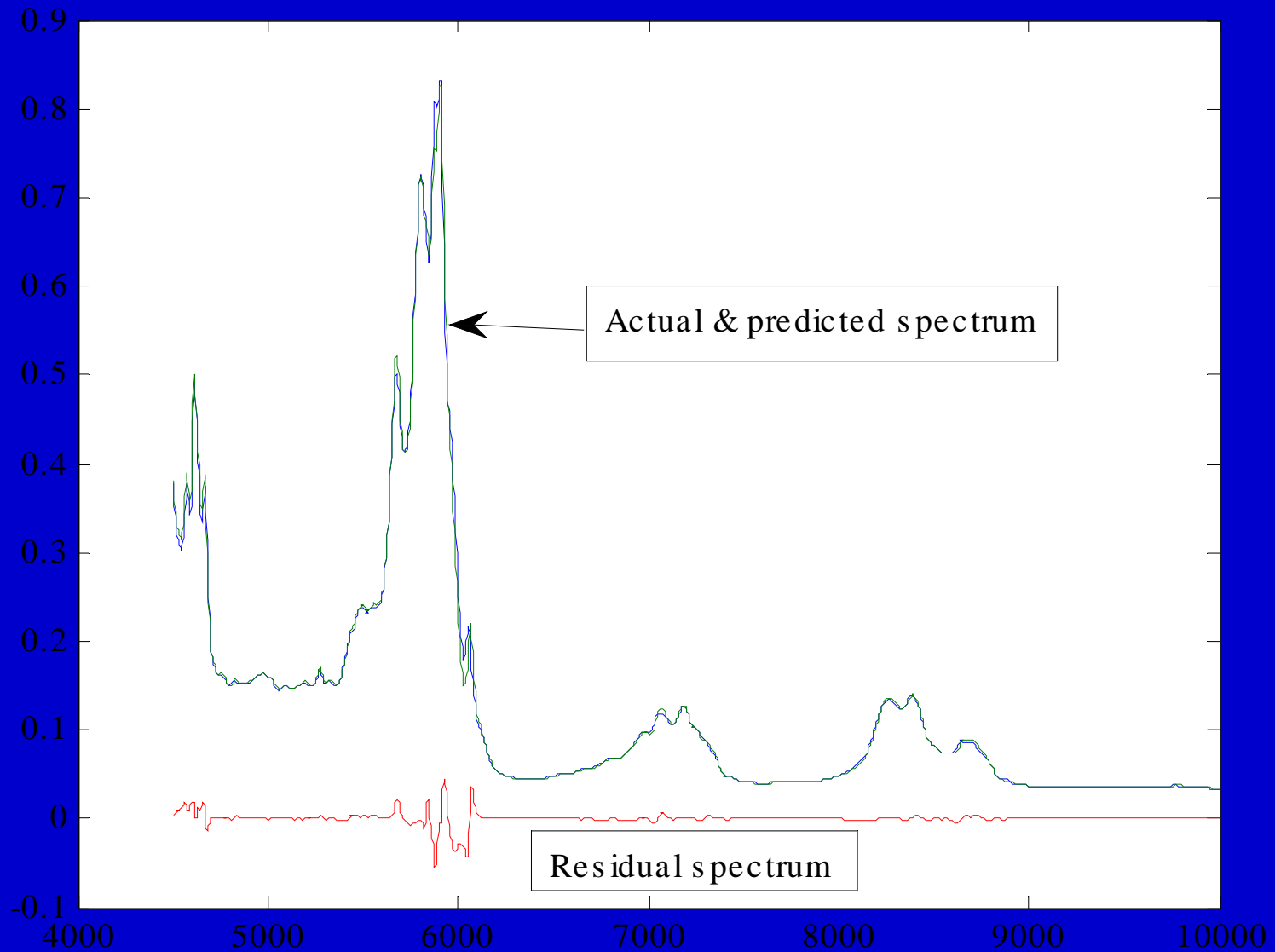
- - - - -

The values for C are coefficients that tell us the fractional contribution of each pure component spectrum to the mixture spectrum.

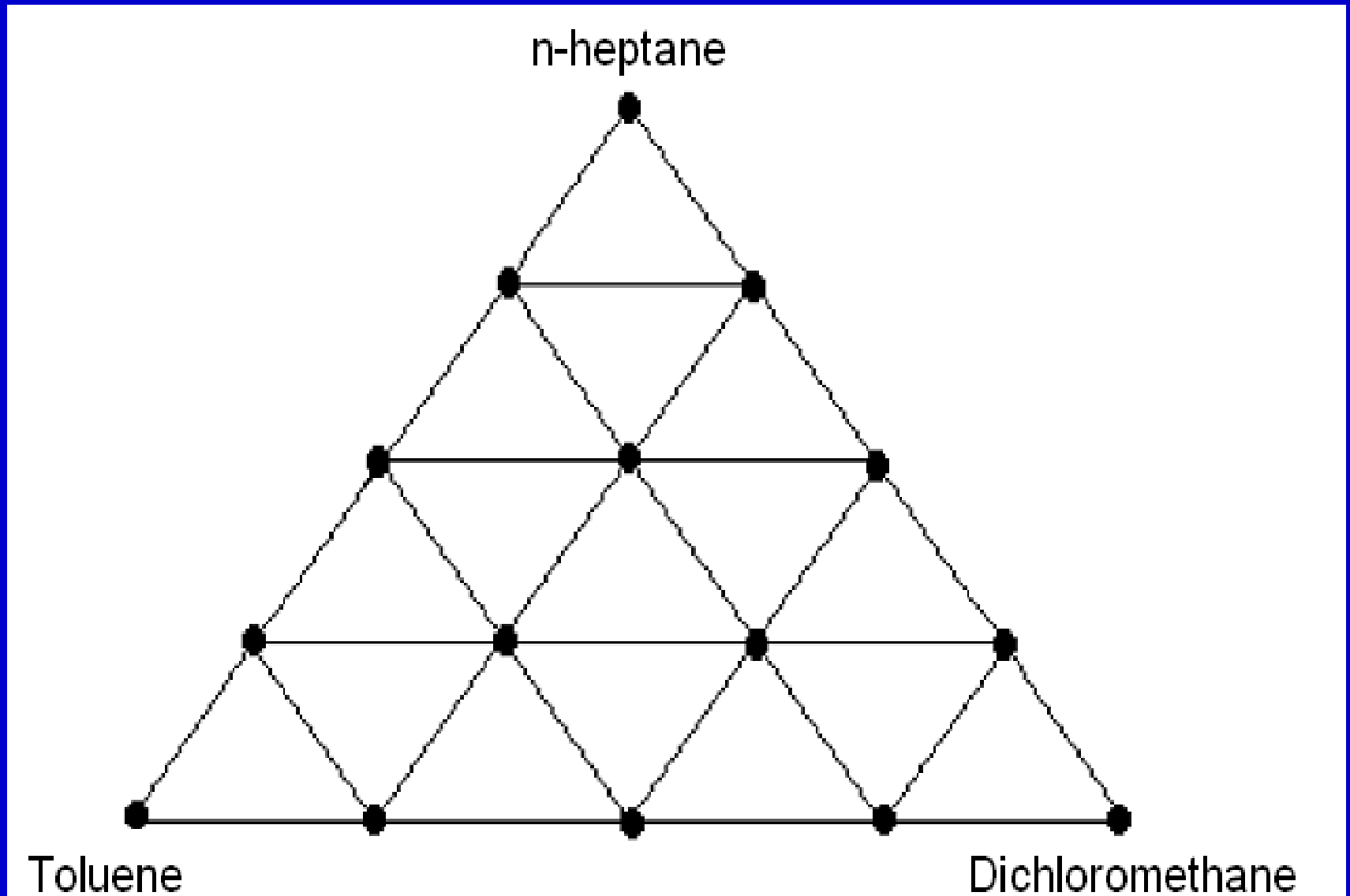
# Properties of CLS

- 1) The values for C are coefficients that tell us the fractional contribution of each pure component spectrum to the mixture spectrum.
- 2) It's an “absolute” method that does not require a conventional “calibration” step to provide analytical answers.
- 3) The coefficients only relate spectra to spectra, but do not in themselves provide a link to the physical characteristics of samples.

# Confirmation of CLS



# 3-Component Mixture Experimental Design



# Comparison with Gravimetric Results

	Toluene		Dichloromethane		n-heptane	
	Grav %	Spec %	Grav %	Spec %	Grav %	Spec %
1	100	100	0	0	0	0
2	76.4	83.06	23.6	14.39	0	1.98
3	74.1	70.45	0	0.72	25.9	29.21
4	50.3	61.04	49.7	35.35	0	3.03
5	48.9	50.76	25.11	15.84	25.99	33.79
6	49.94	45.57	0	0.76	50.06	54.33
7	25.3	34.78	74.77	62.43	0	2.68
8	25.33	28.82	49.65	35.17	25.02	36.63
9	23.86	23.67	26.45	15.54	49.68	61.3
10	25.19	21.98	0	0.47	74.81	77.87
11	0	0	100	100	0	0
12	0	-0.47	75.04	61.19	24.96	40.87
13	0	0.5	49.54	34.76	50.46	67.21
14	0	0.33	24.34	14.72	75.66	86.47
15	0	0	0	0	100	100

# Comparison with Mole % Results

	Toluene		Dichloromethane		n-heptane	
	Mole %	Spec %	Mole %	Spec %	Mole %	Spec %
1	100	100	0	0	0	0
2	77.83	83.06	22.16	14.39	0	1.98
3	72.45	70.45	0	0.72	27.54	29.21
4	52.32	61.04	47.67	35.35	0	3.03
5	48.74	50.76	23.07	15.84	28.17	33.79
6	47.84	45.57	0	0.76	52.15	54.33
7	26.84	34.78	73.15	62.43	0	2.68
8	25.76	28.82	46.55	35.17	27.67	36.63
9	23.32	23.67	23.84	15.54	52.82	61.3
10	23.64	21.98	0	0.47	76.35	77.87
11	0	0	100	100	0	0
12	0	-0.47	71.81	61.19	28.18	40.87
13	0	0.5	45.42	34.76	54.57	67.21
14	0	0.33	21.42	14.72	78.57	86.47
15	0	0	0	0	100	100

# Comparison with Volumetric Values

	Toluene		Dichloromethane		n-heptane	
	Vol. %	Spec %	Vol. %	Spec %	Vol. %	Spec %
1	100	100	0	0	0	0
2	83.3	83.06	16.69	14.39	0	1.98
3	69.29	70.45	0	0.72	30.7	29.21
4	60.93	61.04	39.06	35.35	0	3.03
5	49.82	50.76	16.6	15.84	33.57	33.79
6	44.03	45.57	0	0.76	55.96	54.33
7	34.27	34.78	65.72	62.43	0	2.68
8	28.37	28.82	36.08	35.17	35.53	36.63
9	22.94	23.67	16.5	15.54	60.56	61.3
10	20.98	21.98	0	0.47	79.01	77.87
11	0	0	100	100	0	0
12	0	-0.47	60.6	61.19	39.39	40.87
13	0	0.5	33.44	34.76	66.55	67.21
14	0	0.33	14.43	14.72	85.66	86.47
15	0	0	0	0	100	100

# Fundamental Causes

- “It isn’t what we don’t know that gives us trouble, it’s what we know that ain’t so” –  
(Variously attributed to Will Rogers or to Mark Twain)

# Fundamental Causes

- The first thing we “know” that ain’t so is that Beer’s Law is the same as CLS.
- Beer’s Law:  $A = abc$
- CLS algorithm:  $A = c_1A_1 + c_2A_2 + \dots$
- The confusion comes from the use of the symbol “c” in both formulas.
- In Beer’s Law, c stands for concentration
- In CLS, c stands for coefficient
- But CLS only relates spectra to spectra, There is no connection in CLS between the spectra and the concentrations.

# Fundamental Causes

- The second thing we “know” that ain’t so is that different measures of concentration are equivalent except for a possible scaling factor.

# Fundamental Causes

- BUT

- Conversion between weight % & volume % is not unique
- Conversion between weight % & volume % is not linear
- THIS HAS NOTHING TO DO WITH THE SPECTROSCOPY

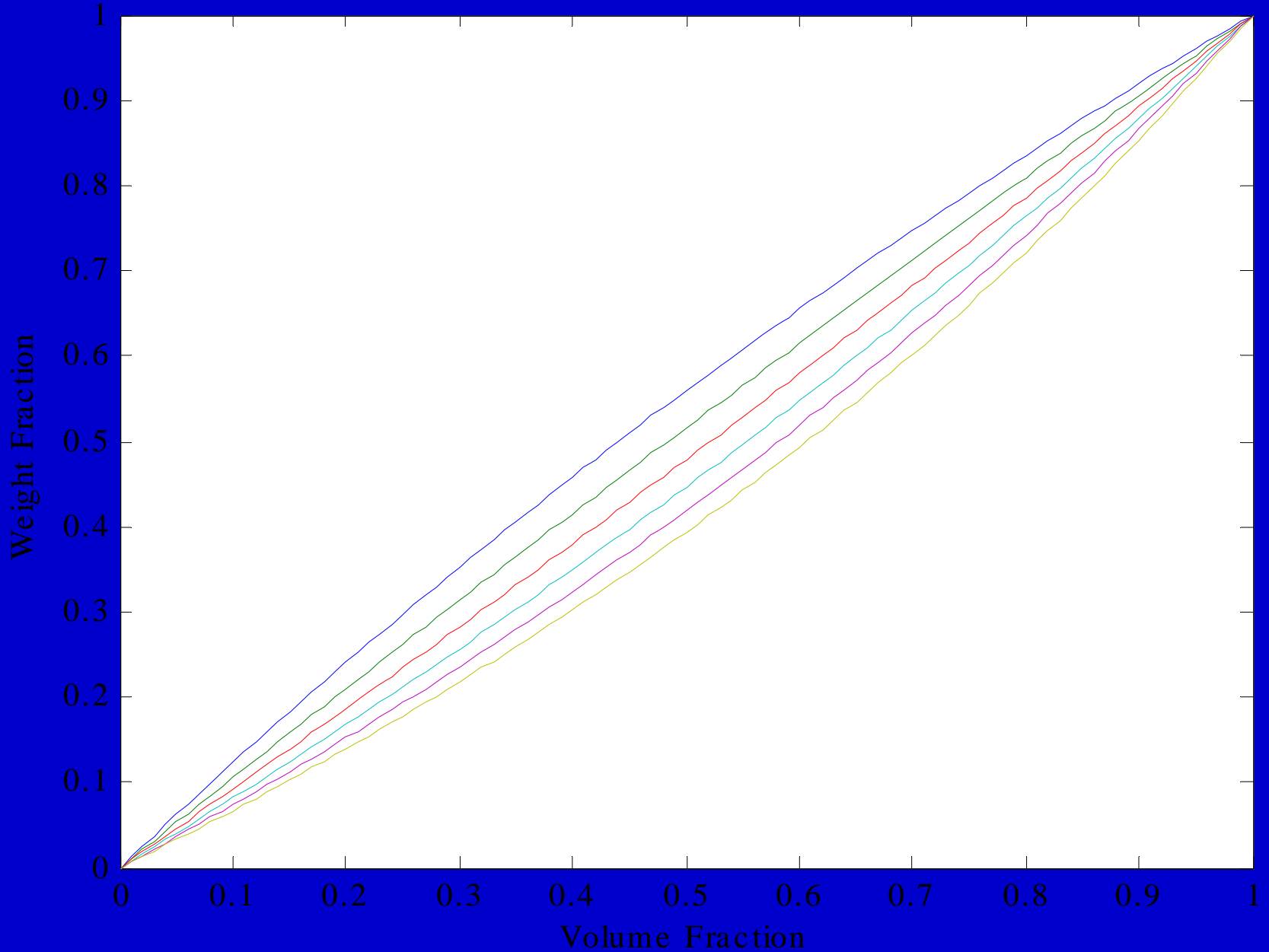
# Conversion between weight % and volume % is not unique

Weight %			Volume %		
Tol	Dichl	Hep	Tol	Dichl	Hep
100	0	0	100	0	0
76.40	23.60	0	83.30	16.69	0
74.10	0	25.90	69.29	0	30.70
50.30	49.70	0	60.93	39.06	0
48.90	25.11	25.99	49.82	16.60	33.79
49.94	0	50.06	44.03	0	55.96
25.30	74.77	0	34.27	65.72	0
25.33	49.65	25.02	28.37	36.08	35.53
23.86	26.45	49.68	22.93	16.50	60.56
25.19	0	74.81	20.98	0	79.01
0	100	0	0	100	0
0	75.04	24.96	0	60.60	39.39
0	49.54	50.46	0	33.44	66.55
0	24.34	75.66	0	14.13	85.86
0	0	100	0	0	100

Note the discrepancies are much larger than any lab error

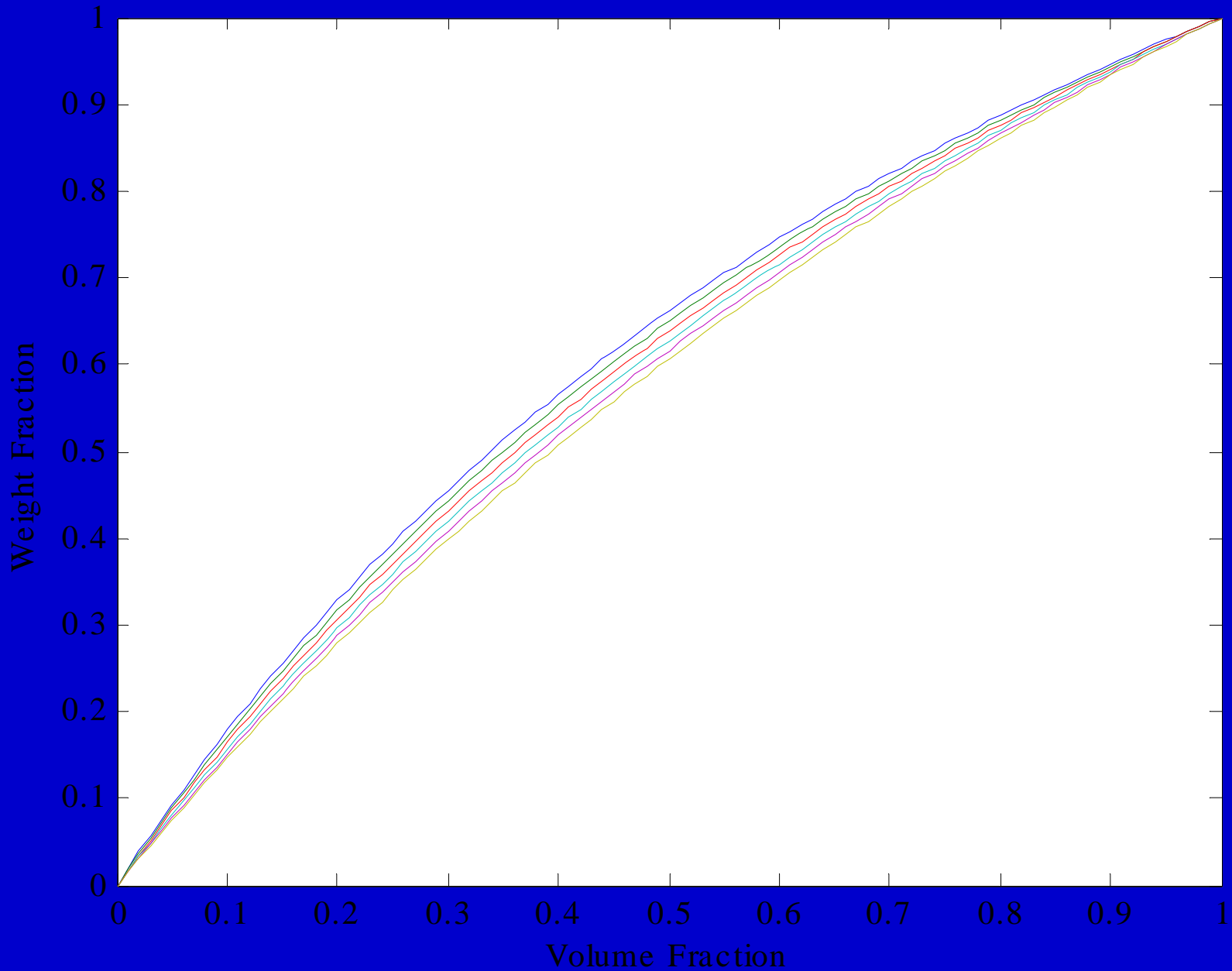
# Non-linear Relation: weight % and volume % (Toluene)

TOLUENE Conversion: Weight % to Volume %



# Non-linear Relation: weight % and volume % (Dichl.)

DICHLOROMETHANE Conversion: Weight % to Volume %



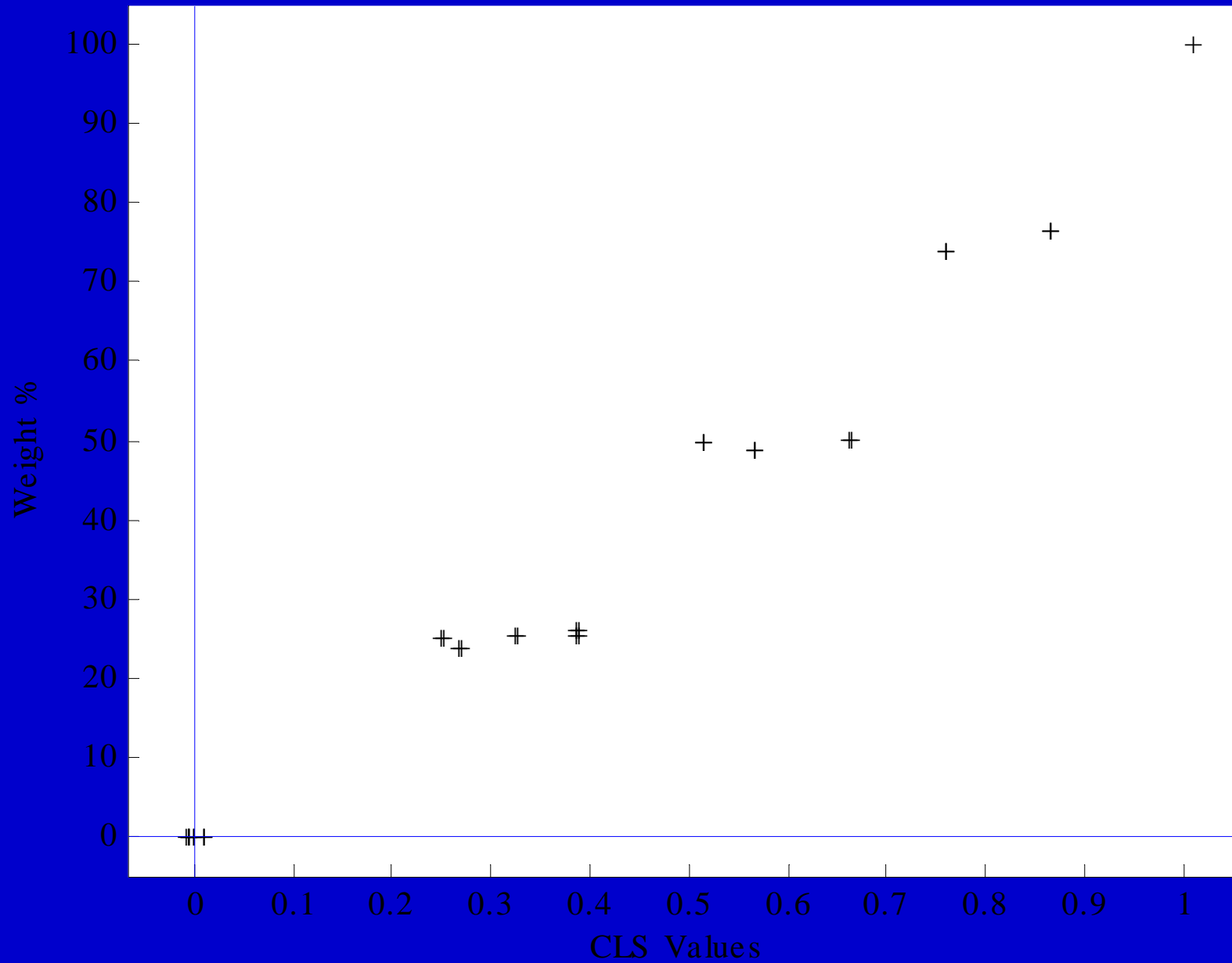


# Conclusions about relationship between weight % and volume %

- 1) The relationship doesn't depend on the spectroscopy, therefore nothing you do to the spectra will change it
- 2) Anything linearly related to volume % must have the same NON-linearity with respect to weight %
- 3) Note in the above graphs that the different lines are not equally spaced.

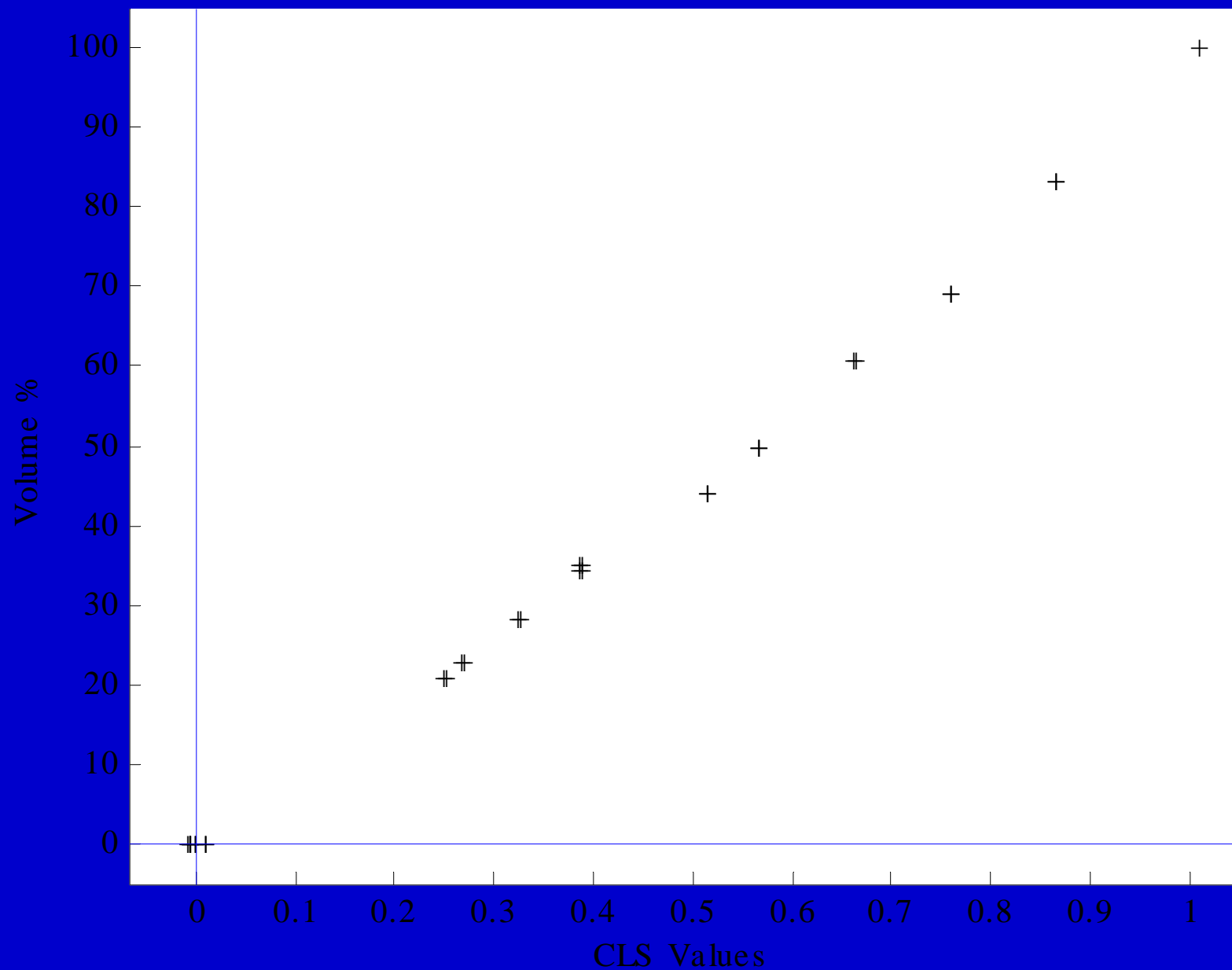
# Toluene: CLS versus weight %

TOLUENE: CLS vs Weight %



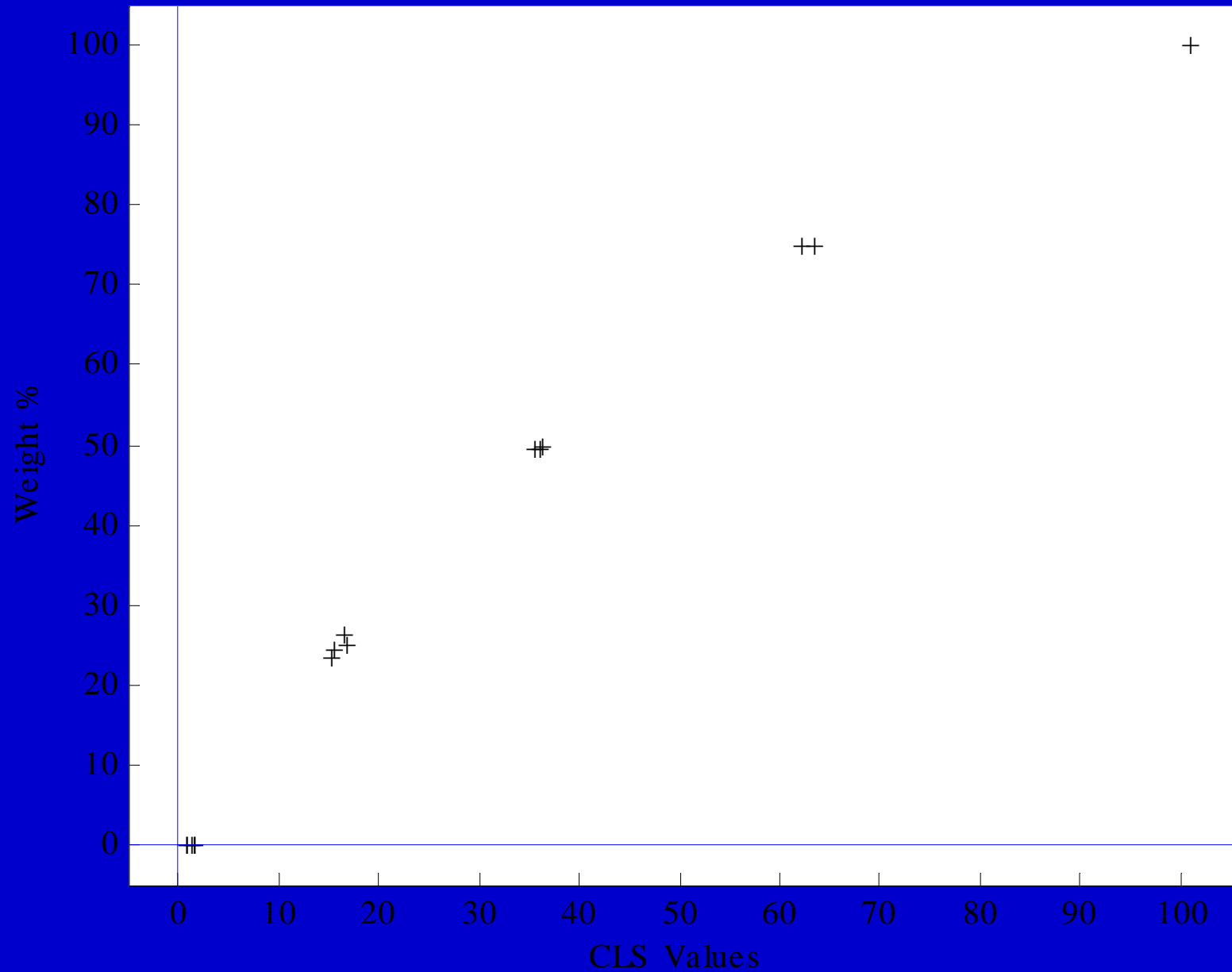
# Toluene: CLS versus volume %

TOLUENE: CLS vs Volume %



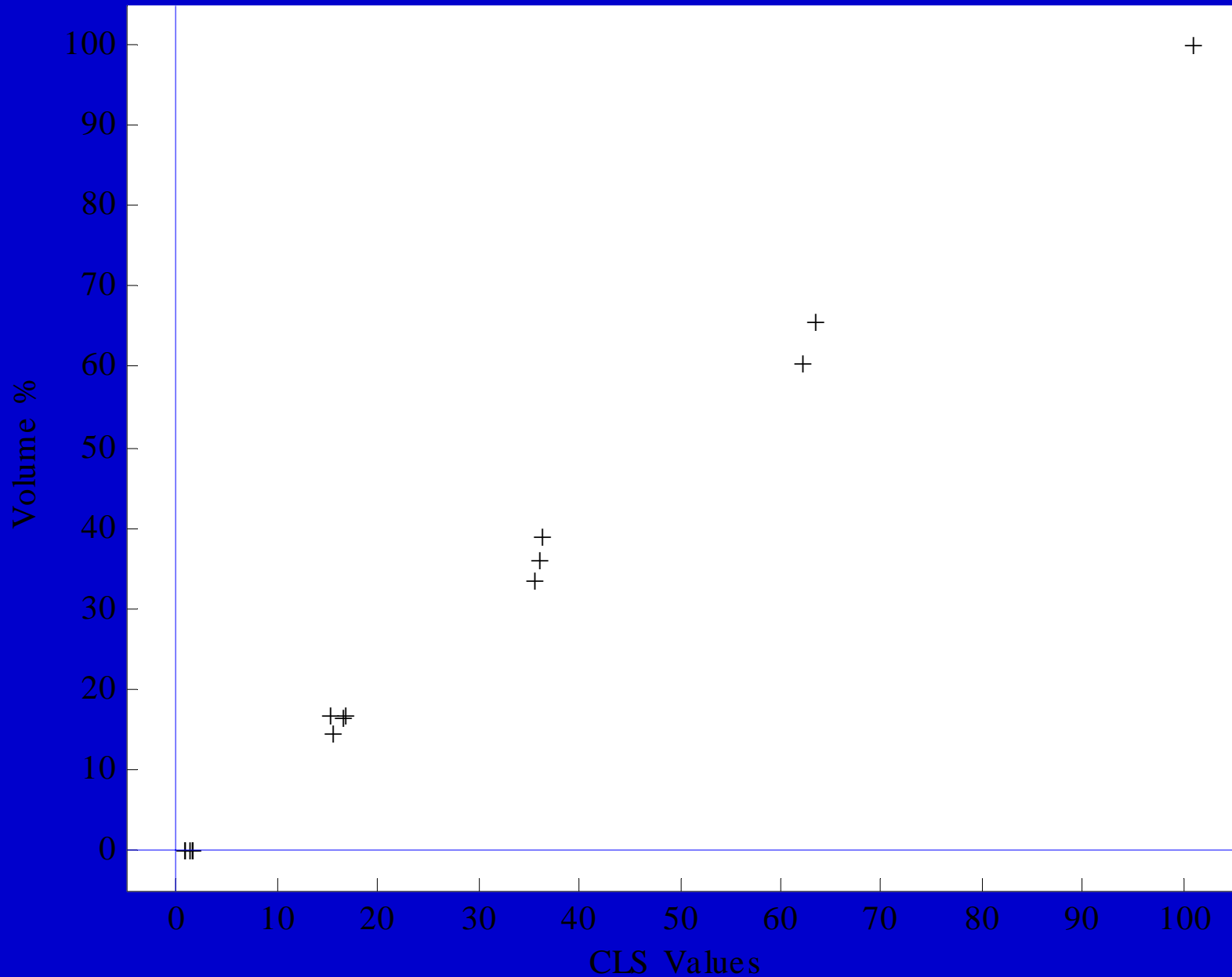
# Dichloromethane: CLS versus Weight %

DICHLOROMETHANE: CLS vs Weight %



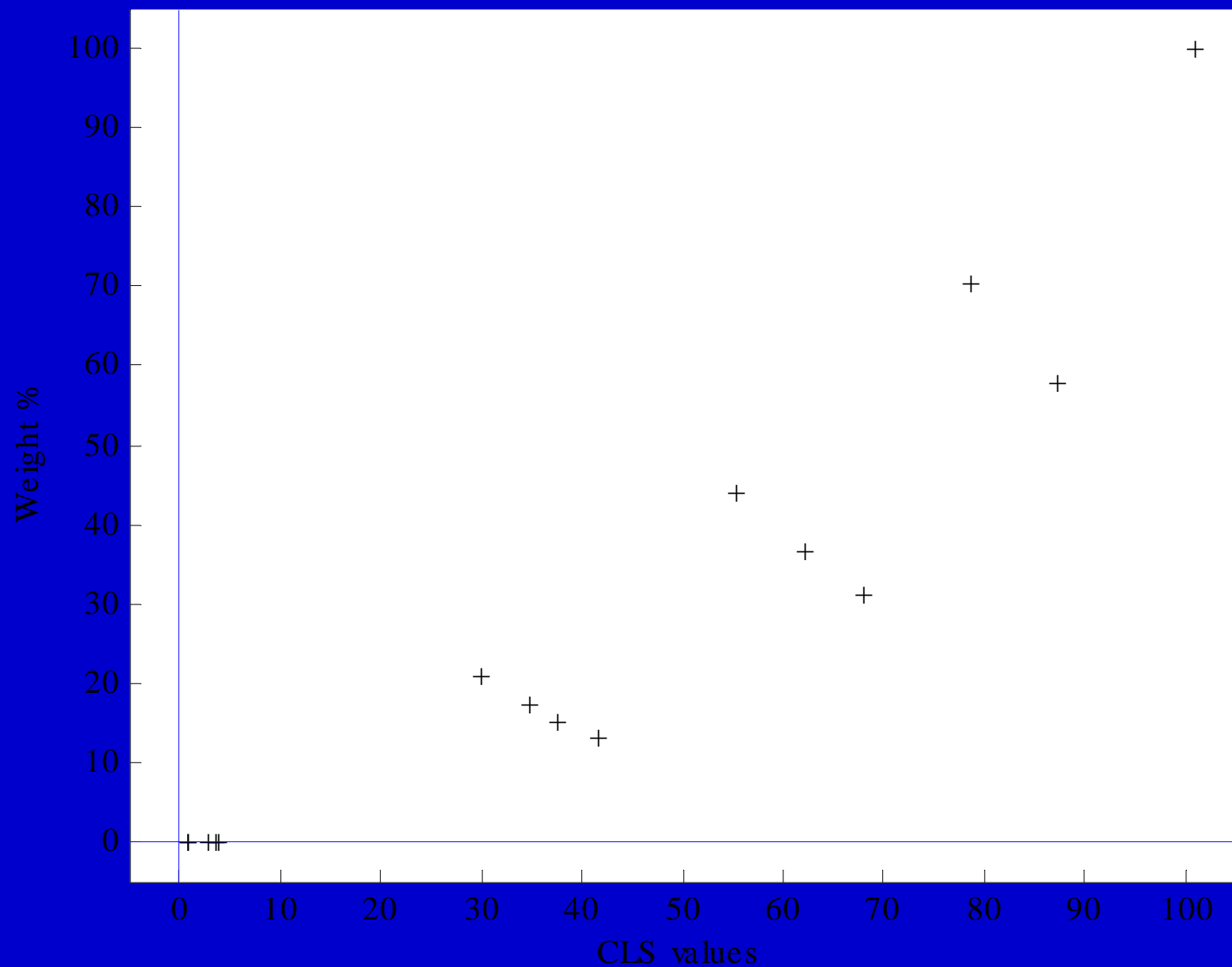
# Dichloromethane: CLS versus Volume %

DICHLOROMETHANE: CLS vs Volume %



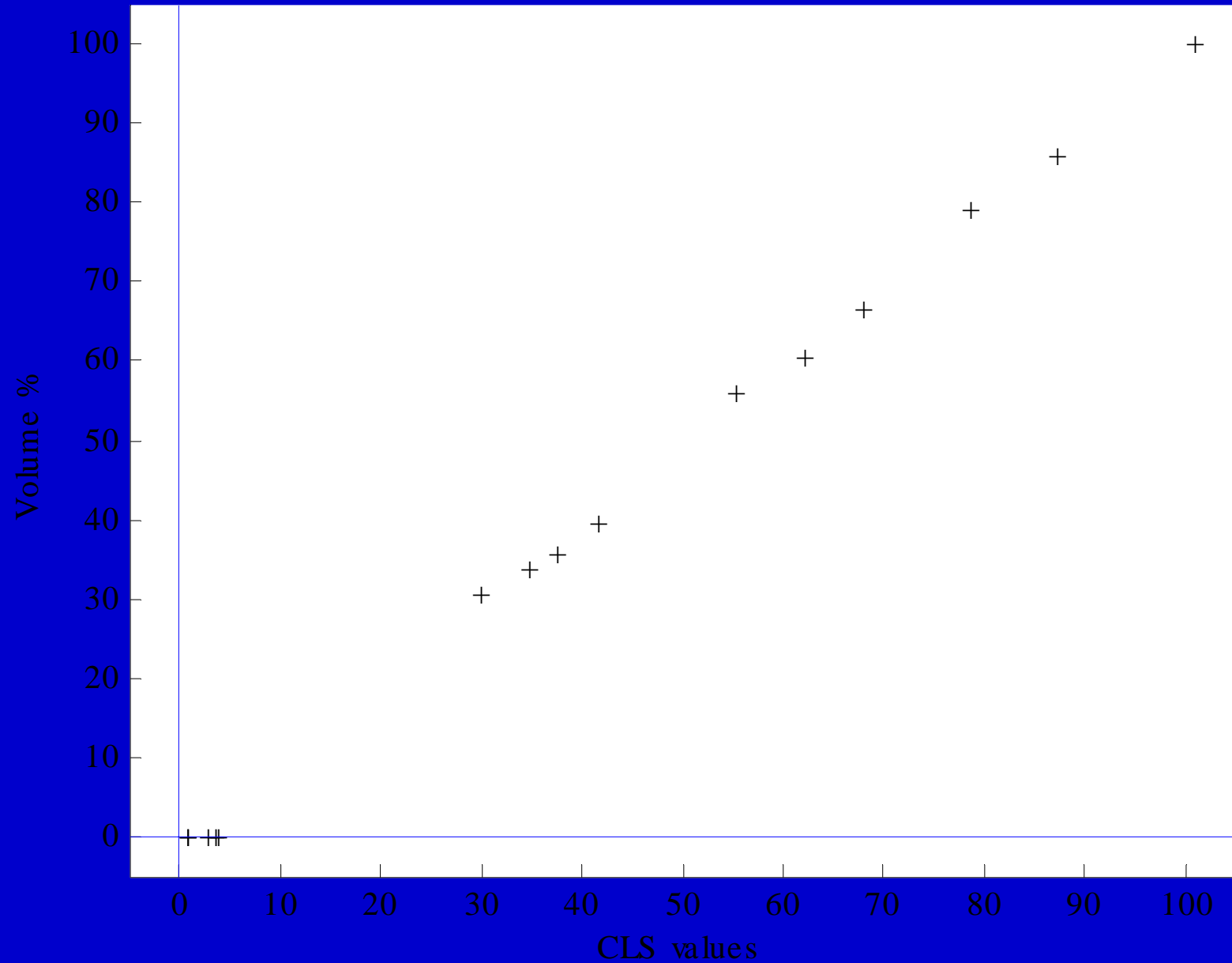
# Heptane: CLS versus weight %

HEPTANE: CLS vs Weight %



# Heptane: CLS versus volume %

HEPTANE: CLS vs Volume %



# Conclusions (1)

- 1) Computed CLS coefficients relate spectra to spectra, but have no link to the physical properties of the samples.
- 2) Because of this, spectra predict spectra using CLS, but whether those spectra predict concentrations, depends on how the concentrations are expressed.
- 3) The work performed shows that the Spectroscopy is sensitive to volume percent (or fraction) of the constituents.
- 4) A measurement that is linear with respect to volume fraction is inherently nonlinear with respect to weight fraction.
- 5) Since the relationship between volume % and other units does not depend on the spectroscopy, nothing you do to the spectra can affect it.

# Conclusions (2)

- 6) Since weight fraction is the unit almost invariably used, the discrepancies caused by the conversion non-linearities due to incorrect units for concentration have plagued chemometric calibrations since the beginning of modern NIR.
- 7) This is exacerbated by the fact that a given volume percent of an analyte can correspond to different weight percents, and vice versa.
- 8) The discrepancy between the values of concentrations in different units is the largest source of error in quantitative spectroscopic analysis, and this error source has been completely ignored by both the spectroscopic and chemometric communities.

# IMPLICATIONS AND RAMIFICATIONS

# CALIBRATION TRANSFER

# Calibration transfer

- 1) Spectroscopy; 22(6); p.20-26 (2007)
- 2) Using data from IDRC 2002 “Shootout”
- 3) Two instruments measuring the same samples gave the same predictions whether the “constituent” was an actual constituent or a set of random numbers

# Calibration transfer

- 4) At the time there was no explanation
- 5) Now we know what ain't so:
  - Calibration transfer is not limited by instrument differences
  - Different sample sets bring with them different sets of non-linear relations between what the instrument sees and what we tell it.

# Calibration transfer caveat

- 6) This does not mean that inter-instrument differences won't also prevent calibration transfer.
- 7) The problem of calibration transfer is different than the problem of validating a calibration
- 8) "A chain is only as strong as its weakest link"
- 9) We have found a new, previously unsuspected "weak link".

# IMPLICATIONS FOR CONVENTIONAL CALIBRATIONS

# Implications for Conventional Calibrations (1)

- 1) It took CLS to bring it out but the physical effects are independent of the algorithm, therefore MLR and PCR and PLS are just as affected.
- 2) This finding explains many of the “mysteries” we’ve been fighting all these years:
- 3) Need for too many factors
- 4) Error measures do not go to the noise floor when # factors = # variables
- 5) Reason why spectral analysis considered good only for “dilute solutions”

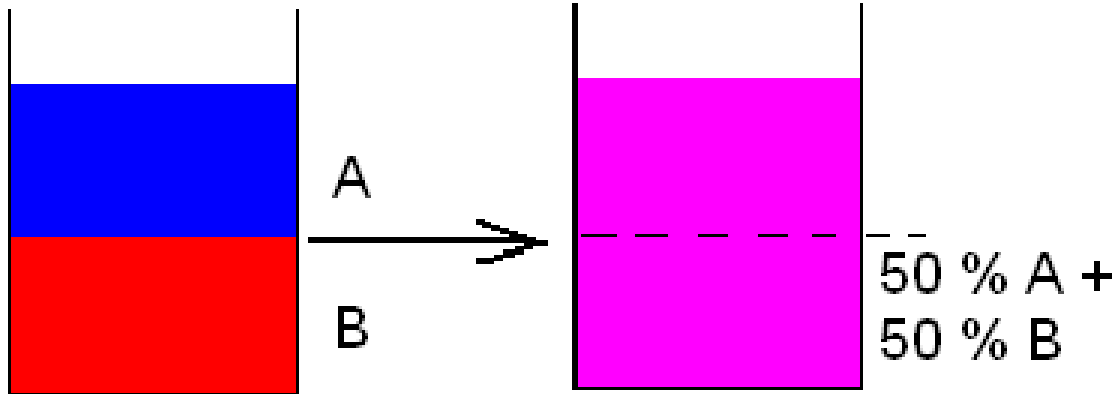
# Implications for Conventional Calibrations (2)

- 6) Reason why calibrations have not been extrapolatable
- 7) Reason calibration transfer is so difficult.
- 8) You can multiply volume percent by sample properties (e.g., density, MW) so that conventional algorithms can use other measures of concentration, that contain a volume in the denominator (e.g., weight / volume, molarity), which will then be linear with respect to the spectroscopy.

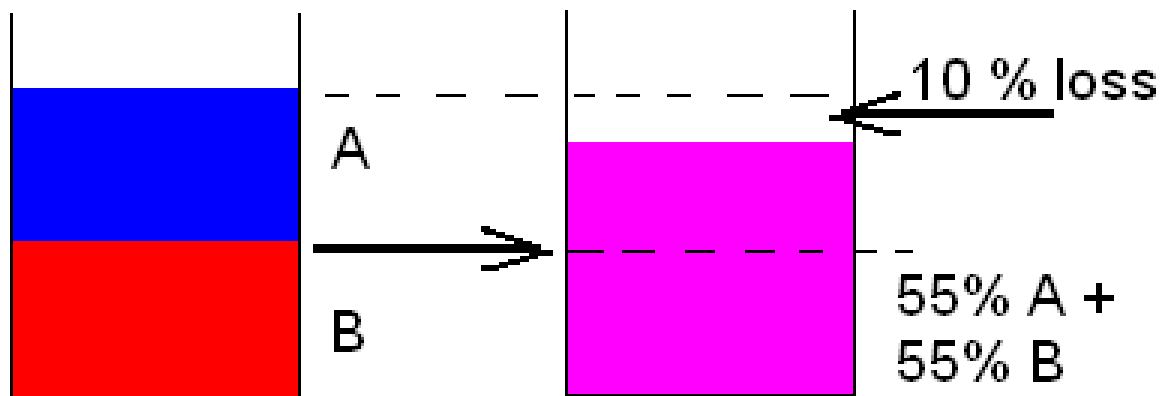
# PARTIAL MOLAL VOLUMES

# Partial Molal Volumes

IDEAL



REAL



# Partial Molal Volumes

- 1) In an ideal equi-volume mixture, there would be 50% of each component by volume
- 2) CLS coefficients would be each be 0.5 and add to 1.0
- 3) PMV effect would reduce the volume of mixture, so that a given volume would contain  $> 50\%$  of each pure component.
- 4) CLS coefficients would each be  $> 0.5$  and add to  $> 1.0$

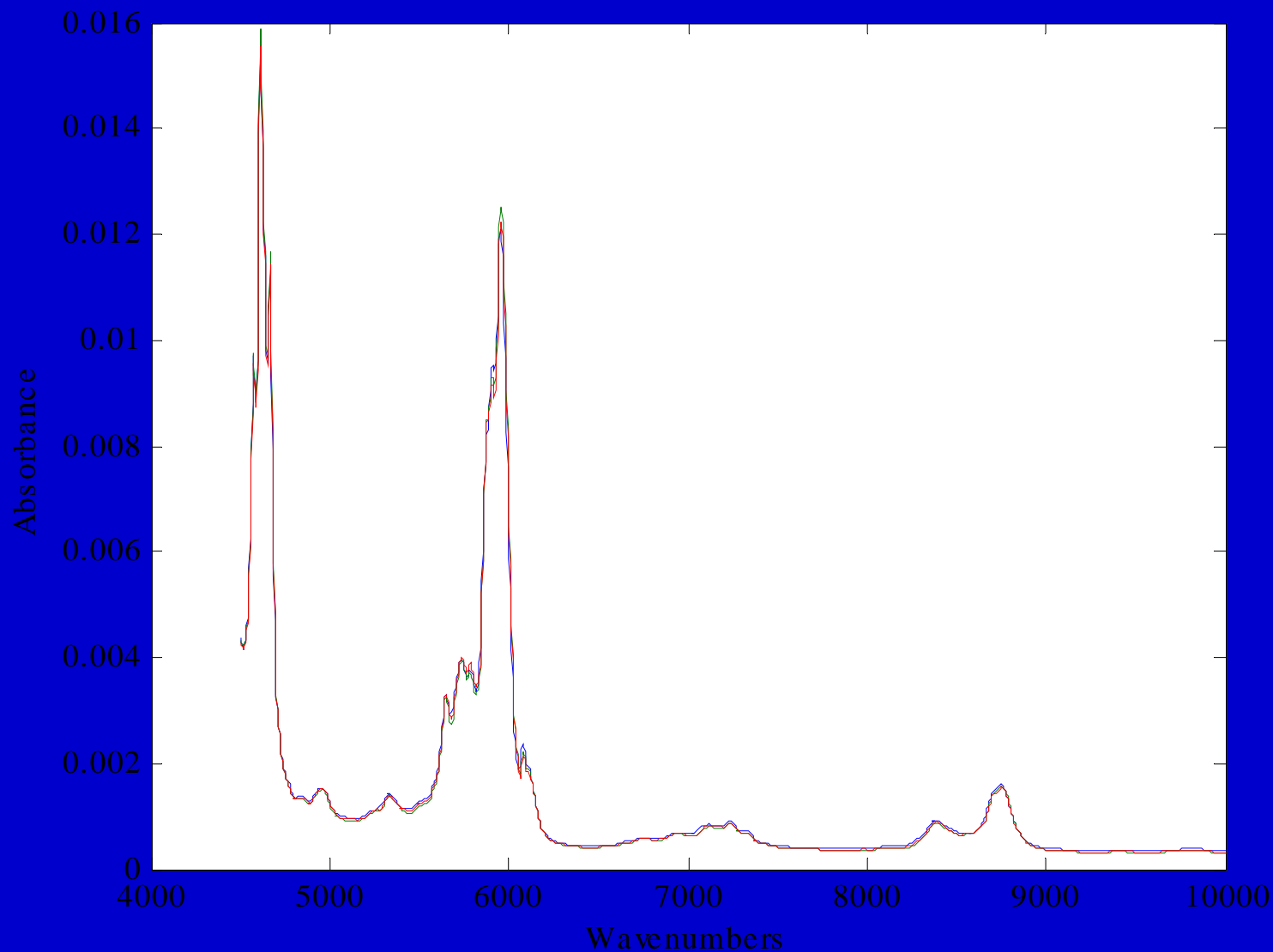
# Partial Molal Volumes for Water-Methanol-Acetic Acid Mixtures

Water-Methanol			Water-Acetic acid			Methanol-Acetic Acid		
Water	Methanol	CLS Sum	Water	Acetic acid	CLS Sum	Methanol	Acetic acid	CLS Sum
100	0	1.000	100	0	1.000	100	0	1.000
75	25	1.003	75	25	1.074	75	25	0.994
50	50	1.012	50	50	1.110	50	50	1.030
25	75	1.011	25	75	1.085	25	75	1.030
0	100	1.000	0	100	1.000	0	100	1.000

EFFECT ON K-MATRIX  
CALCULATION OF PURE-  
COMPONENT SPECTRA FROM  
WEIGHT % vs VOLUME %

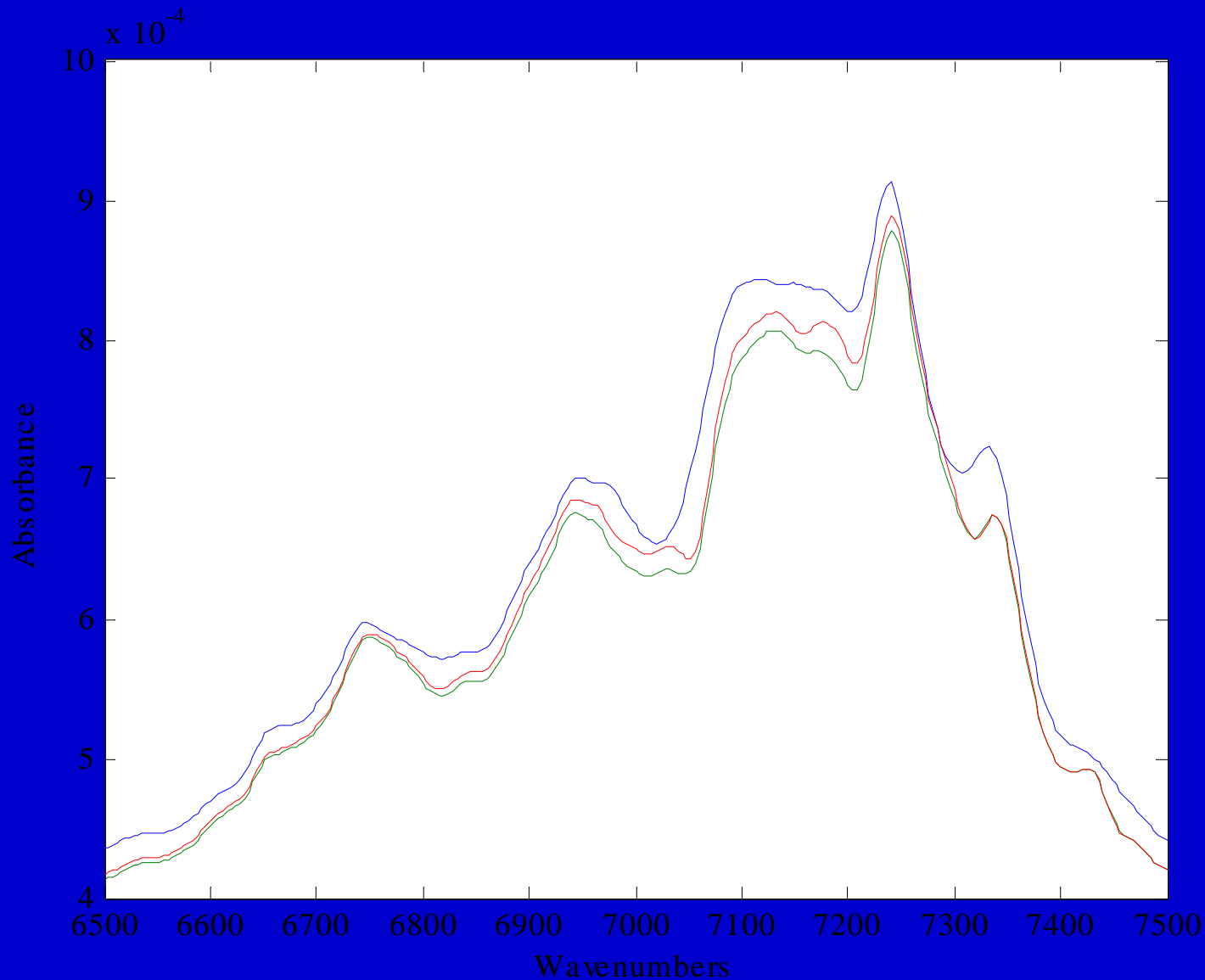
# Toluene calculation, Full Spectral Range

Blue = measured, black = wt %, red = vol %



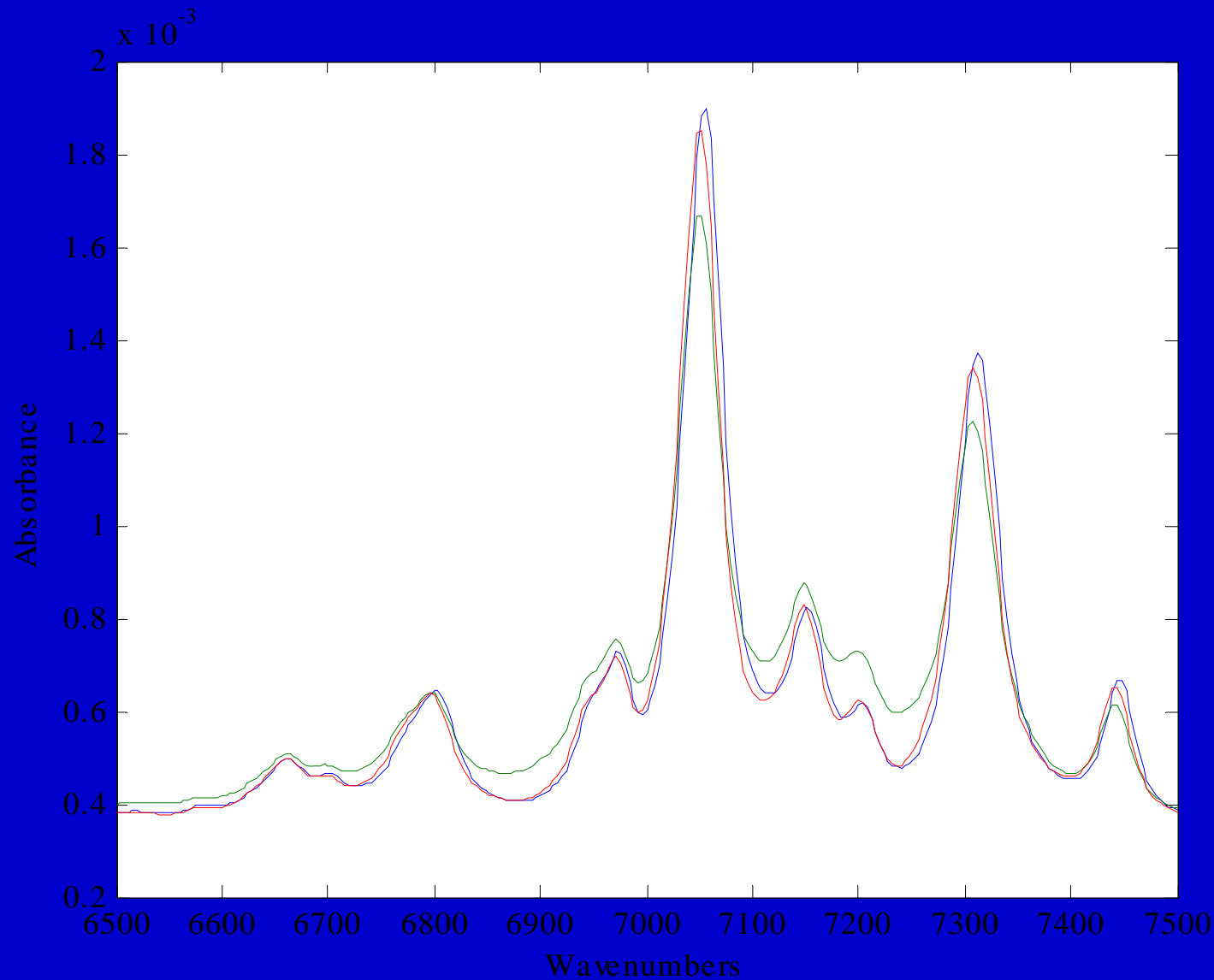
# Toluene calculation, 6500 – 7500 $\text{cm}^{-1}$

Blue = measured, black = wt %, red = vol %



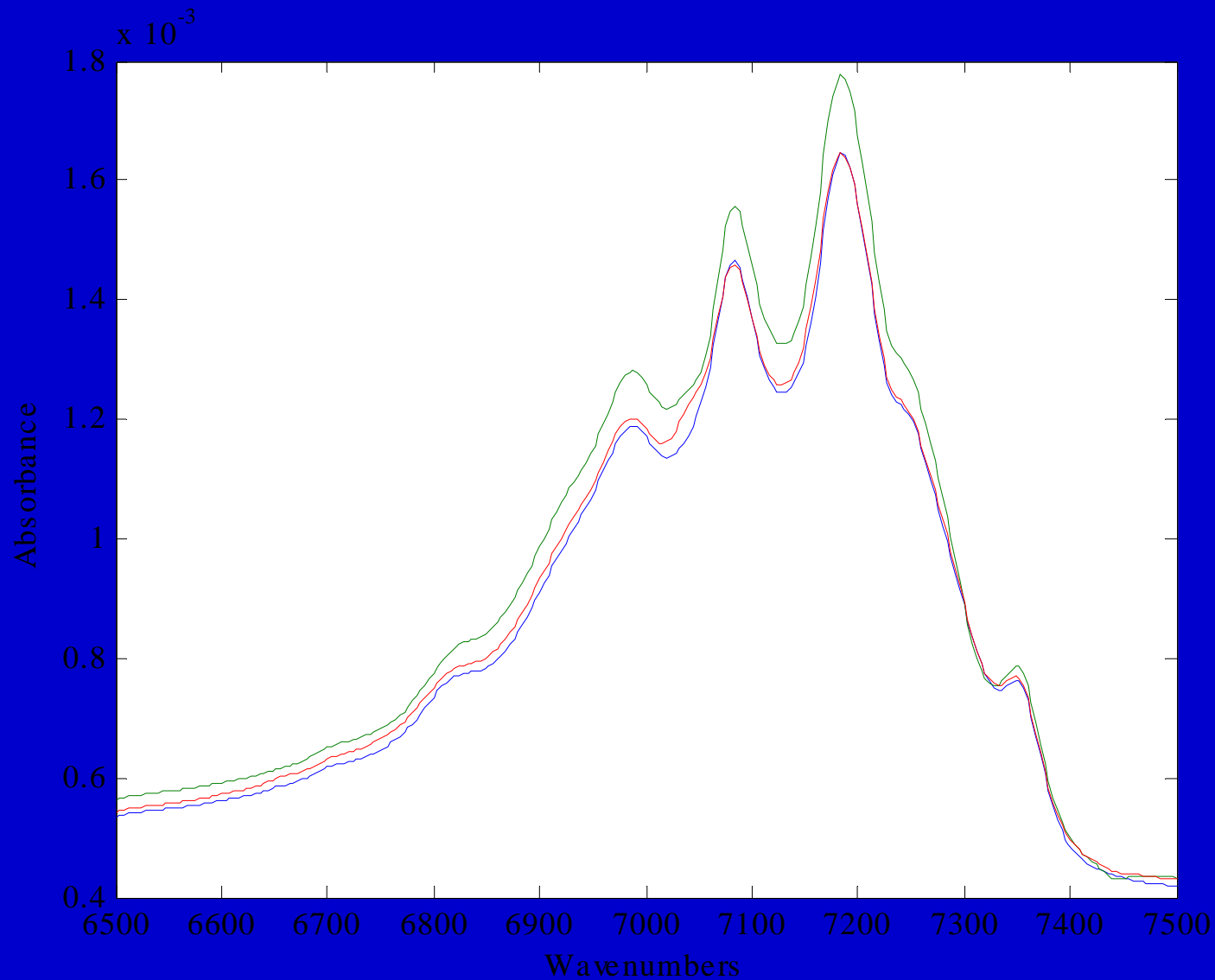
# CH<sub>2</sub>Cl<sub>2</sub> calculation, 6500 – 7500 cm<sup>-1</sup>

Blue = measured, black = wt %, red = vol %



# n-heptane calculation, 6500 – 7500 $\text{cm}^{-1}$

Blue = measured, black = wt %, red = vol %



# OTHER IMPLICATIONS

# Other Implications

- 1) Sum of CLS calculated concentration = inverse of Partial Molal Volume
- 2) The CLS algorithm is the nearest thing we have to an “absolute” analytical method. therefore results are obtained without the need for conventional calibration. (Caveat: Beer’s Law must hold)
- 3) Because of that, Reference Lab error is not an issue
- 4) CLS results are obtained with zero PLS or PCR factors
- 5) Because of that, “Overfitting” is not possible (although underfitting is).

# Published

- Applied Spectroscopy; 64(9), p.995-1006 (2010)
- (note that some results and insights reported here did not make it into the paper)

# CONJECTURES and IMPLICATIONS FOR FUTURE EXAMINATION

- Do these effects also occur in solid samples?  
(And address problem area: how can we determine volume % of components in a solid sample?)
- Do these effects occur in other spectral regions or for other types of spectroscopic analysis (Raman, Fluorescence, Emission)?