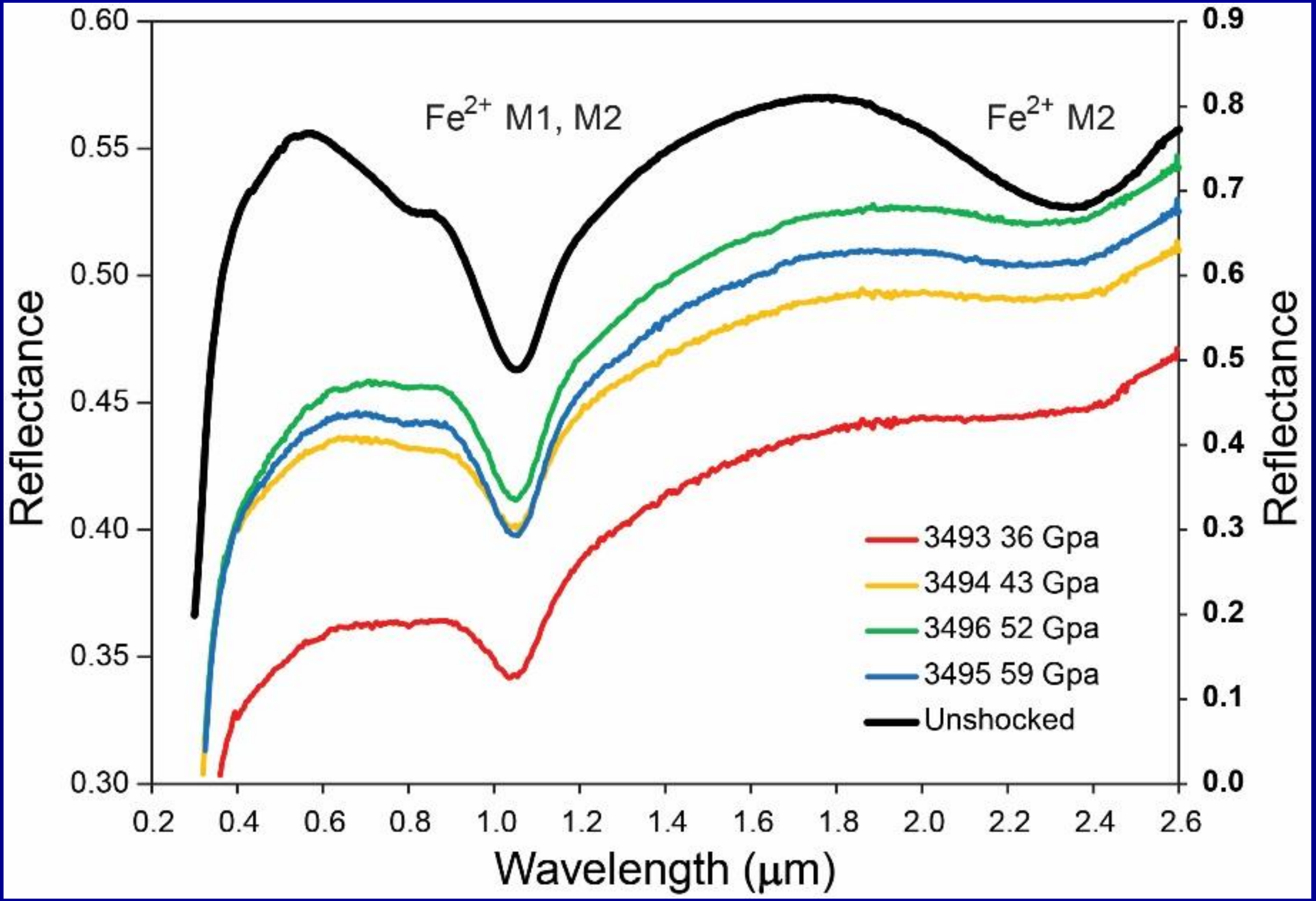
The Future of Spectroscopy

M. Darby Dyar
Dept. of Astronomy
Mount Holyoke College
Senior Scientist
Planetary Science Institute

The Holy Grail: Accurate Mineralogy Derived from Spectroscopy

http://www.movie-roulette.com/photos_big/monty-python-and-the-holy-grail-4-1.jpeg


CAUTION
OFF



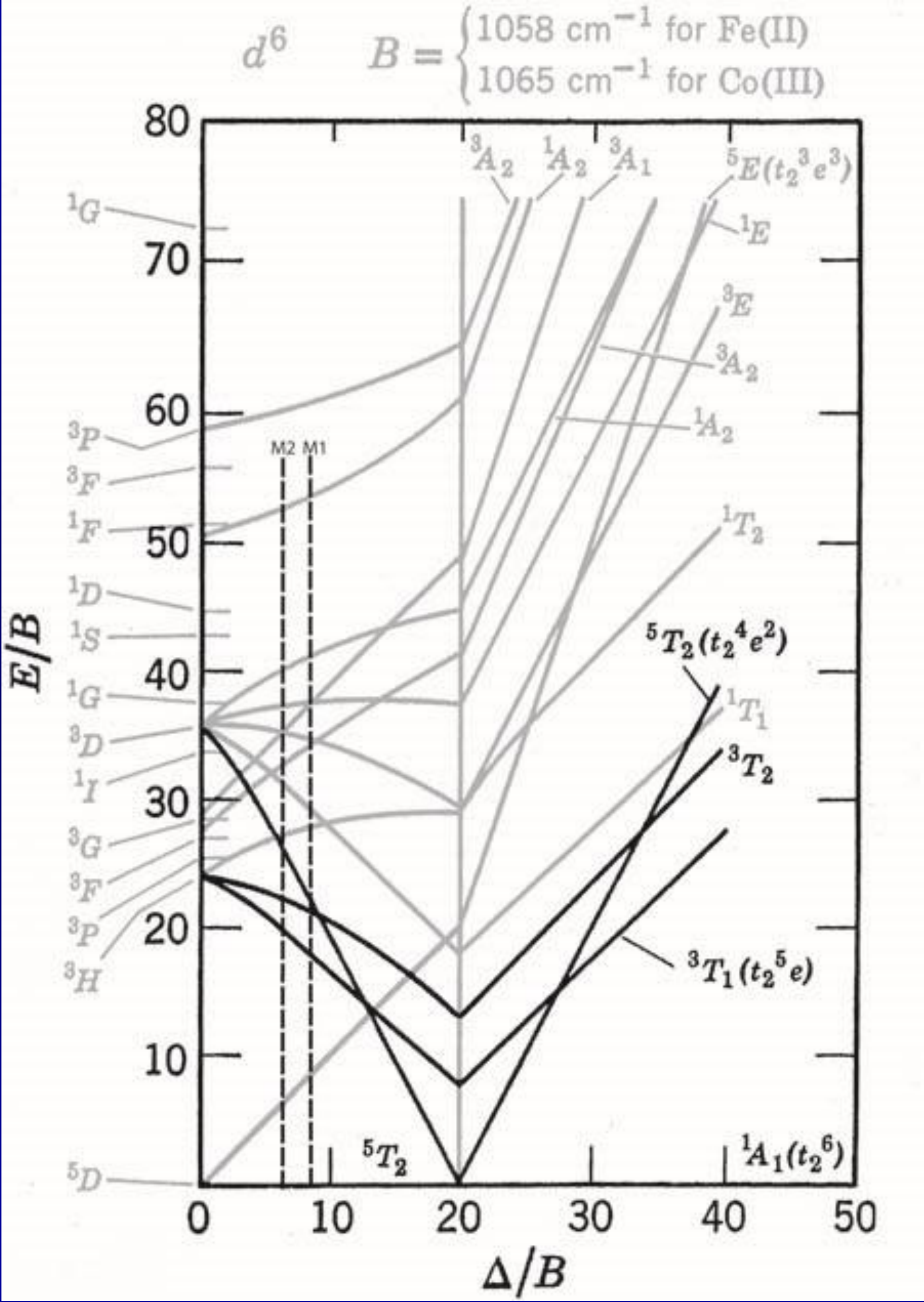
McCanta et al. (2016) Icarus, submitted

Table 1. CRISM Spectral Parameter Summary Products[a]

Name	Parameter	Formulation[b]	Rationale
		Surface Parameters [c]	
R770	0.77 μm reflectance	R770	rock/dust
RBR	red/blue ratio	R770/R440	rock/dust
BD530	0.53 μm band depth	1 − (R530/(a*R648 + b*R440))	crystalline ferric minerals
SH600	0.60 μm shoulder height	R600/(a*R530 + b*R680)	select ferric minerals
BD640	0.64 μm band depth	1 − (R648/(a*R600 + b*R680))	select ferric minerals
BD860	0.86 μm band depth	1 − (R860/(a*R800 + b*R920))	select ferric minerals
RPEAK1	reflectance peak 1	wavelength where 1st derivative = 0 of 5th order polynomial fit to R600, R648, R680, R710, R740, R770, R800, R830	Fe mineralogy
BDI1000VIS	1 μm integrated band depth; VIS wavelengths	divide R830, R860, R890, R920 by RPEAK1 then integrate over (1 − normalized reflectances)	Fe mineralogy
BDI1000IR	1 μm integrated band depth; IR wavelengths	divide R950, R980, R1020, R1050, R1080, R1150 by linear fit from peak R between 1.3–1.87 μm to R2530 extrapolated backward to remove continuum, then integrate over (1 − continuum-corrected reflectances)	Fe mineralogy
IRA	1.3 μm reflectance	R1330	IR albedo
OLINDEX	Olivine index	(R1695/(0.1*R1050 + 0.1*R1210 + 0.4*R1330 + 0.4*R1470)) − 1	olivine will be strongly positive; based on fayalite
LCPINDEX	pyroxene index	((R1330-R1050)/(R1330 + R1050)) * ((R1330-R1815)/(R1330 + R1815)	pyroxene will be strongly positive; favors LCP
HCPXINDEX	pyroxene index	((R1470-R1050)/(R1470 + R1050)) * ((R1470-R2067)/(R1470 + R2067)	pyroxene will be strongly positive; favors HCP
VAR	spectral variance	variance of observed data from a line fit from 1.0–2.3 μm	olivine and pyroxene will have high values
ISLOPE1	−1 * spectral slope1	(R1815-R2530)/(2530-1815)	ferric coating on dark rock
BD1435	1.435 μm band depth	1 − (R1430/(a*R1370 + b*R1470))	CO_2 ice

Pelkey, S. M., et al. (2007), CRISM multispectral summary products: Parameterizing mineral diversity on Mars from reflectance, J. Geophys. Res., 112, E08S14, doi:10.1029/2006JE002831.

Tanabe-Sugano diagram for pyroxene



Klima et al. (2007) MAPS, 42, 235-253

Spectroscopy

+

Machine

Learning

=

Better

Spectroscopy

http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291099-128X

Chemometrics is an interdisciplinary field combining experimental design, physical-chemical measurements, multivariate statistical analysis, mathematical modeling, and information technology for extracting useful information from data.

…Journal of Chemometrics

Chemometric Approaches to:

A. Multivariate analysis

B. X-ray absorption spectroscopy

C. Laser-induced breakdown spectroscopy

D. Baseline removal

E. Calibration transfer

F. Data preprocessing

Most Basic Technique for Multivariate Analysis
Partial Least Squares (PLS)

- Shrink regression equation by creating hybrid channels that are linear combinations of all previous channels.
- Correlate two matrices described by **Y** = **X b** :
  - Spectra (**X**) (***p*** samples × ***N*** channels)
  - Variable(s) of interest (**Y**)
- **This removes co-linearity** because directions in that new vector space are ortho-normal, avoiding the problem that inhibits ordinary least-squares regression.
- PLS analysis thus produces *b*-coefficients for each channel that represent the correlation implicit in **b**.

Most Basic Technique for Multivariate Analysis Partial Least Squares (PLS)

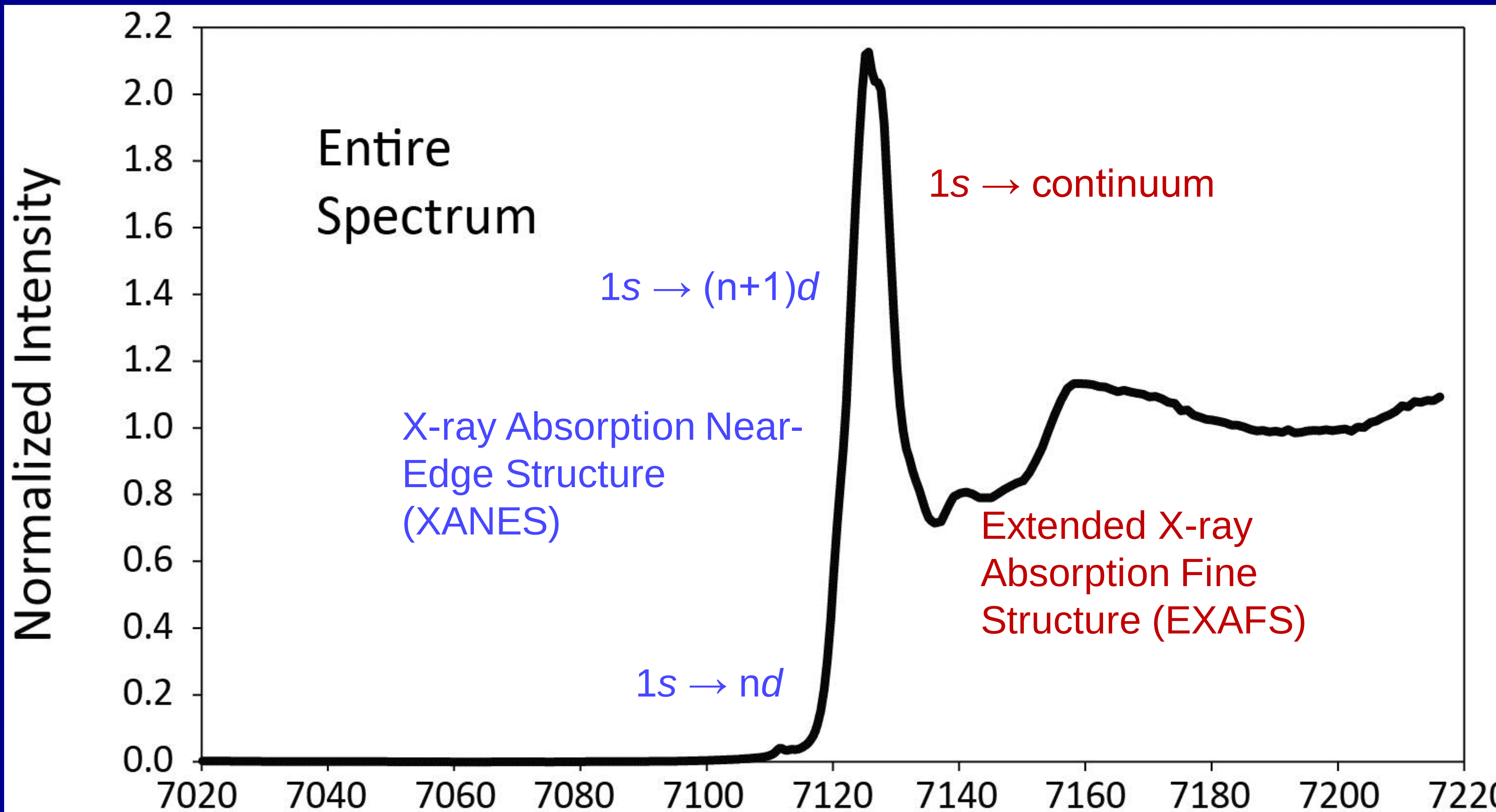
Each Spectral Channel is an Independent Variable

Prediction Quantities (elements, mineralogy, %Fe^{3+}) are the Dependent Variable(s)

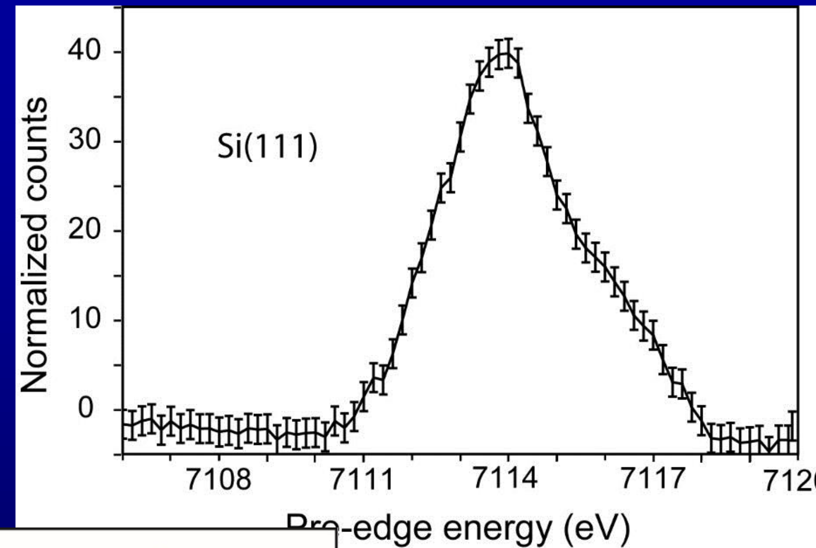
	Fe^{3+}	λ1 (keV)	λ2 (keV)	λ3 (keV)	λ4 (keV)	λ5 (keV)	λ6 (keV)	λ7 (keV)	λ8 (keV)	λ9 (keV)	λ10 (keV)	……	λn (keV)
Sample 1													
Sample 2													
Sample 3													
Sample 4													
Sample 5													
…													
Sample n													

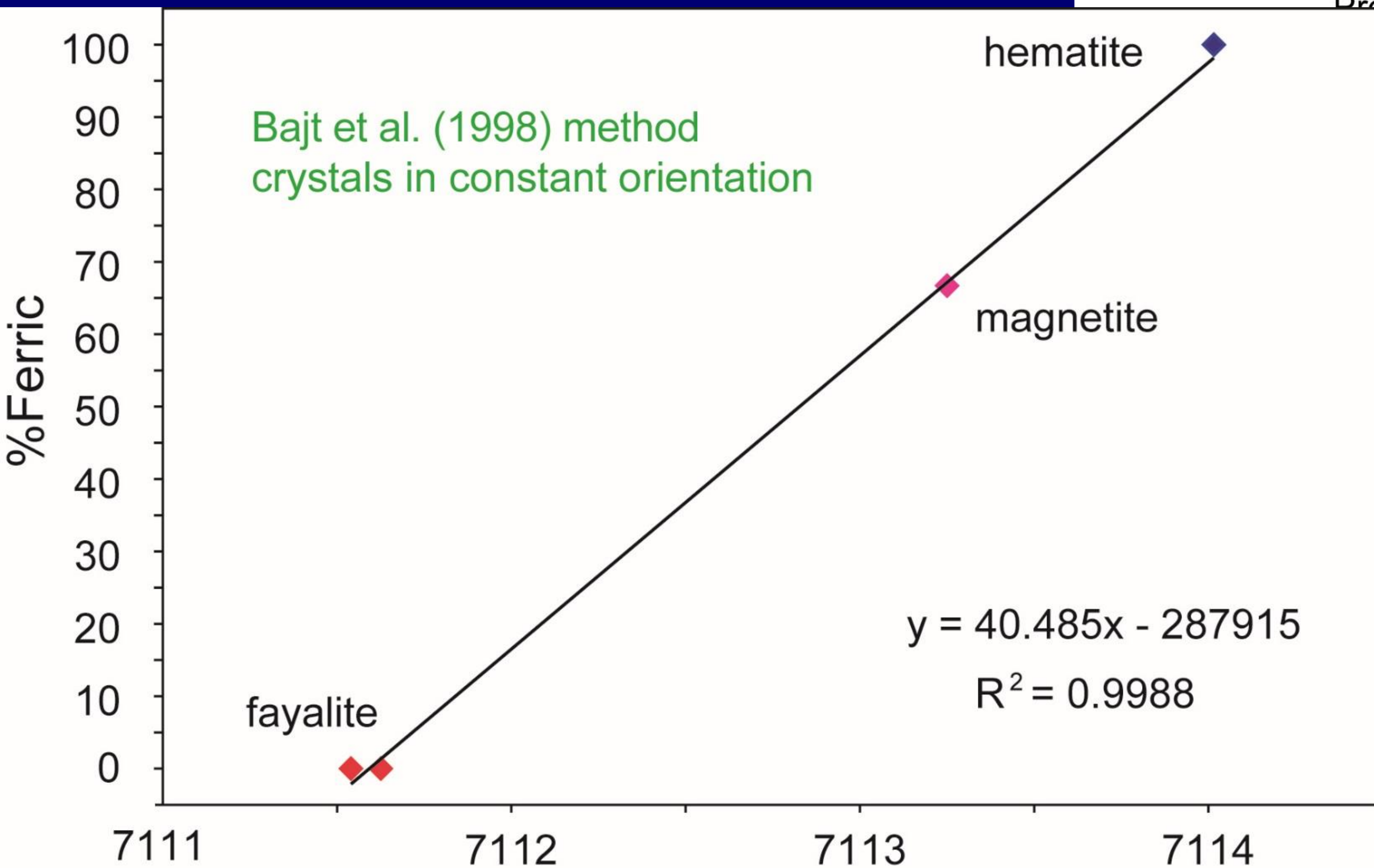
Example #1:

X-ray Absorption Spectroscopy



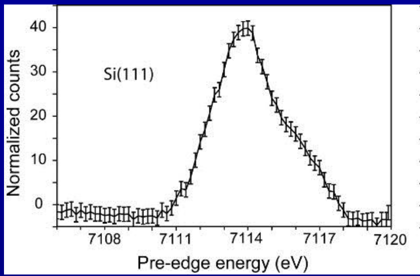
Using XAS to measure redox state of Fe

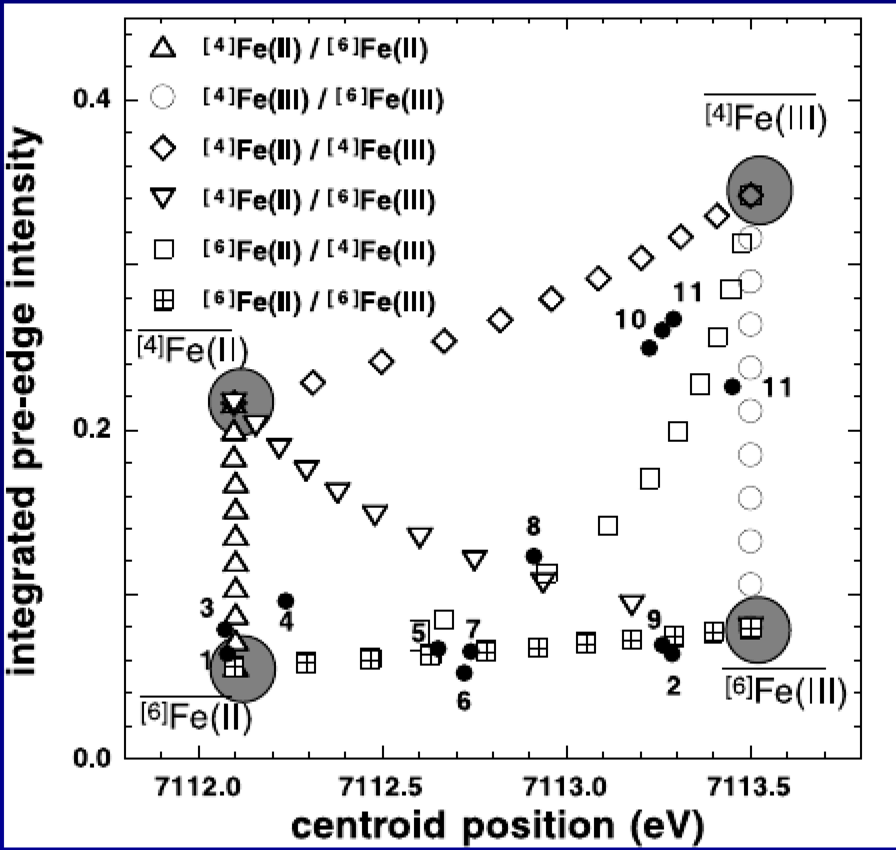


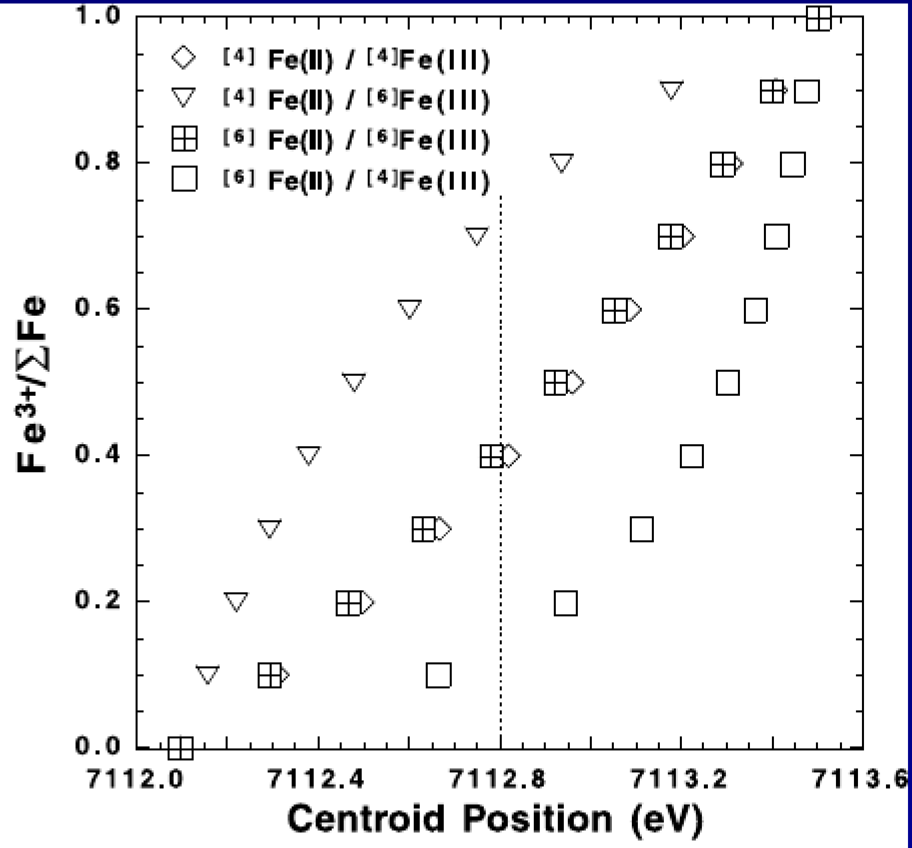


Bajt et al. (1994): Centroid of pre-edge moves with valence state

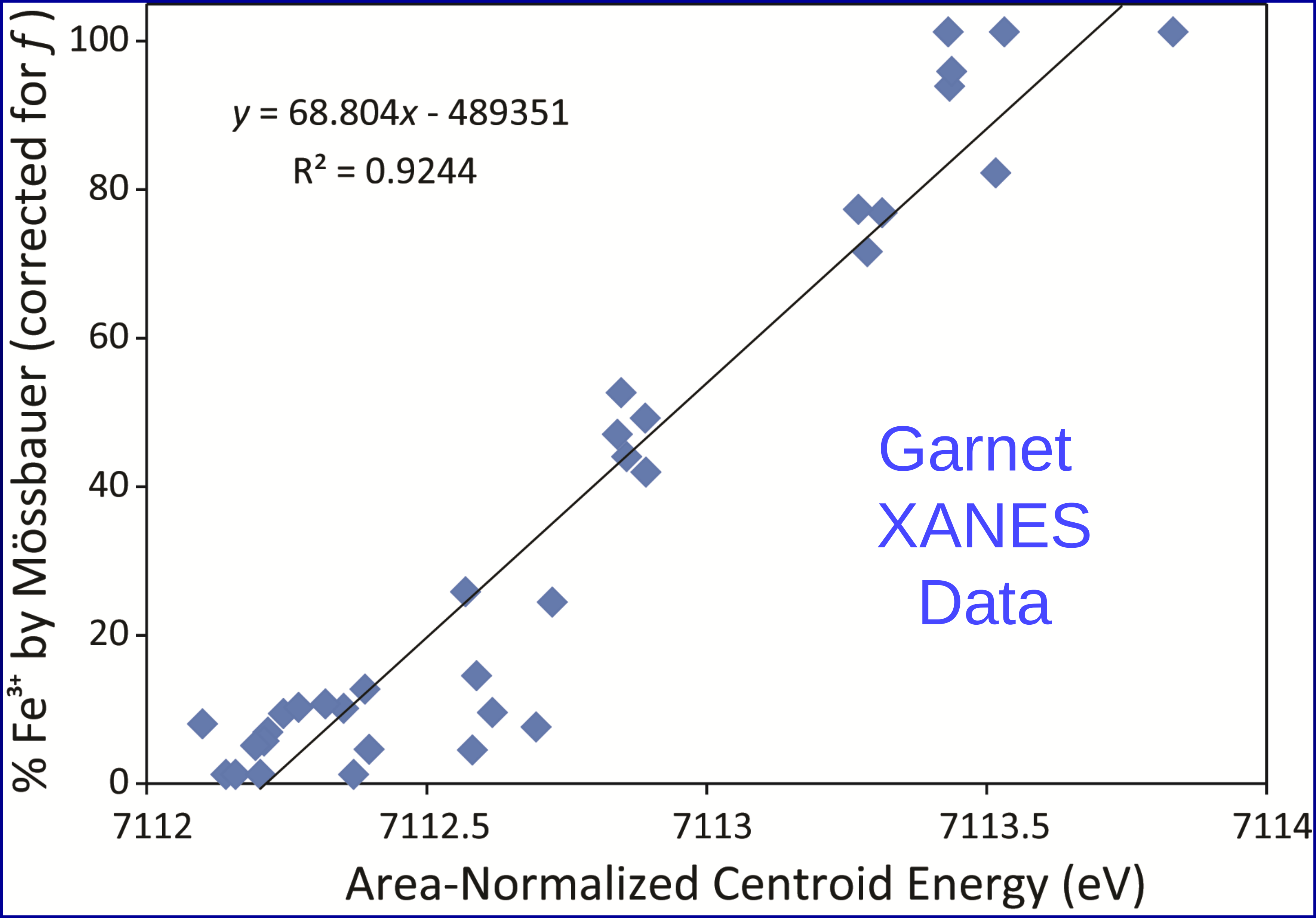
Fe XANES Data used to Predict %Fe^{3+} in Powders using XANES Pre-Edge



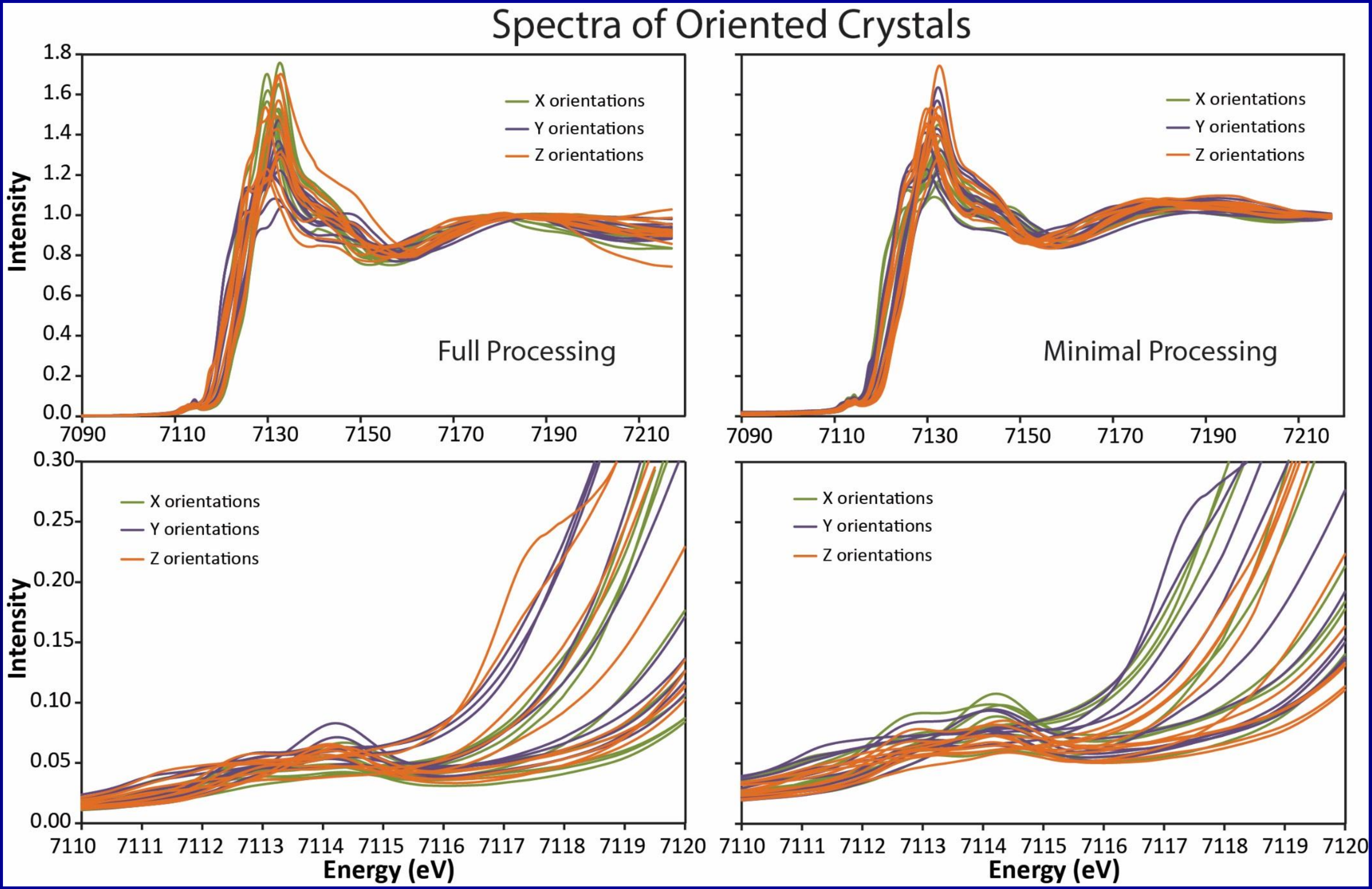




Wilke et al. (2001) *Am. Min.* 86, 714-730, calibration for powders



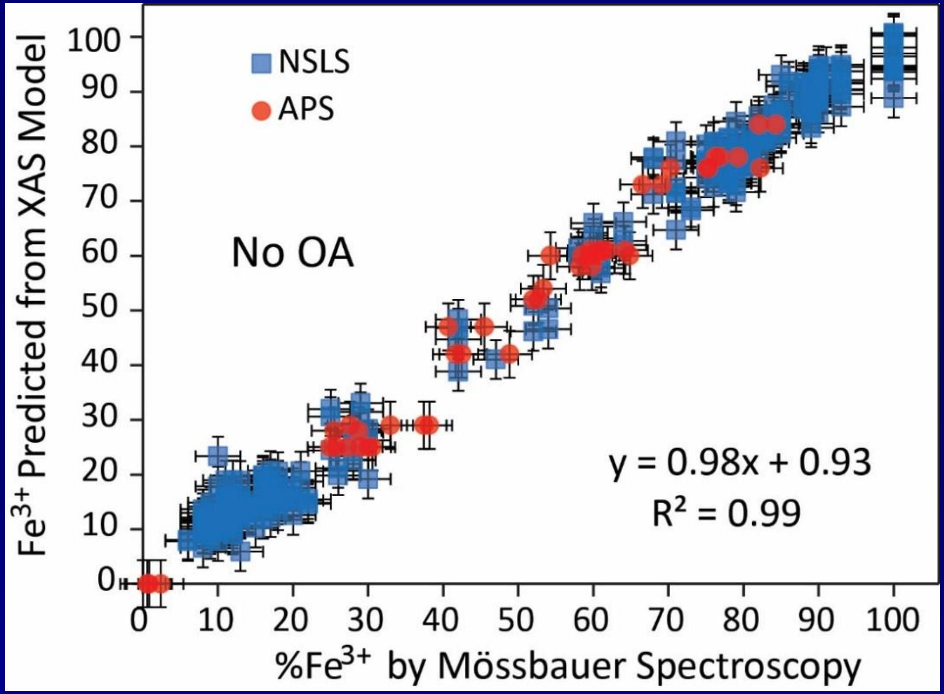
% Fe^{3+} by Mössbauer (corrected for f)
100
80
60
40
20
0
$y = 68.804x - 489351$
$R^2 = 0.9244$
Garnet XANES Data
7112
7112.5
7113
7113.5
7114
Area-Normalized Centroid Energy (eV)

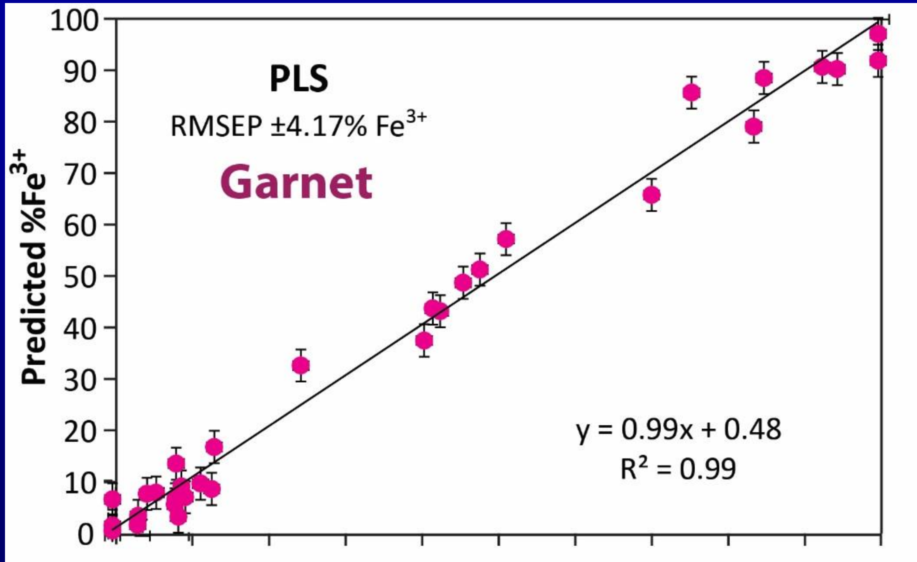


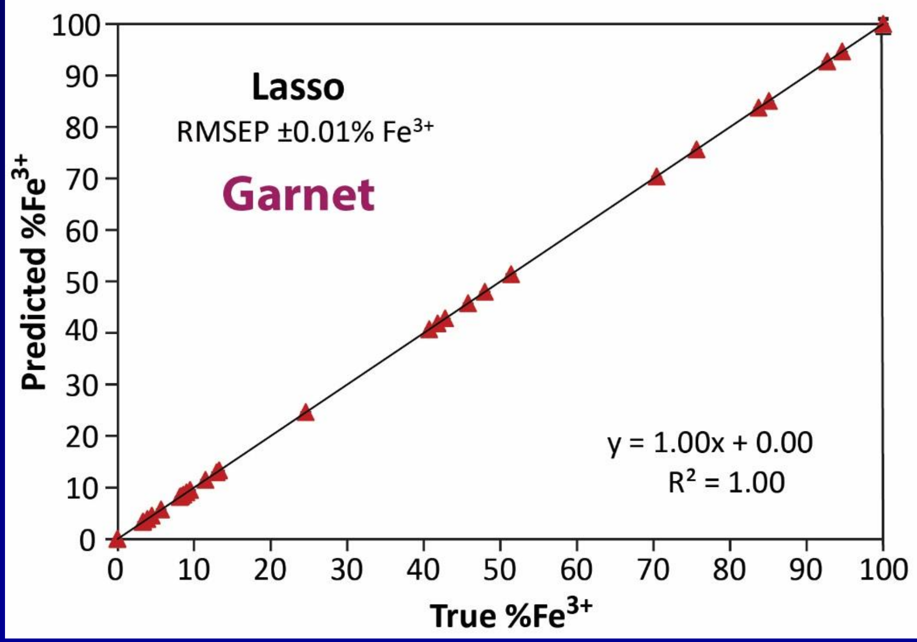
B. XAS Example


WORLD
NASA

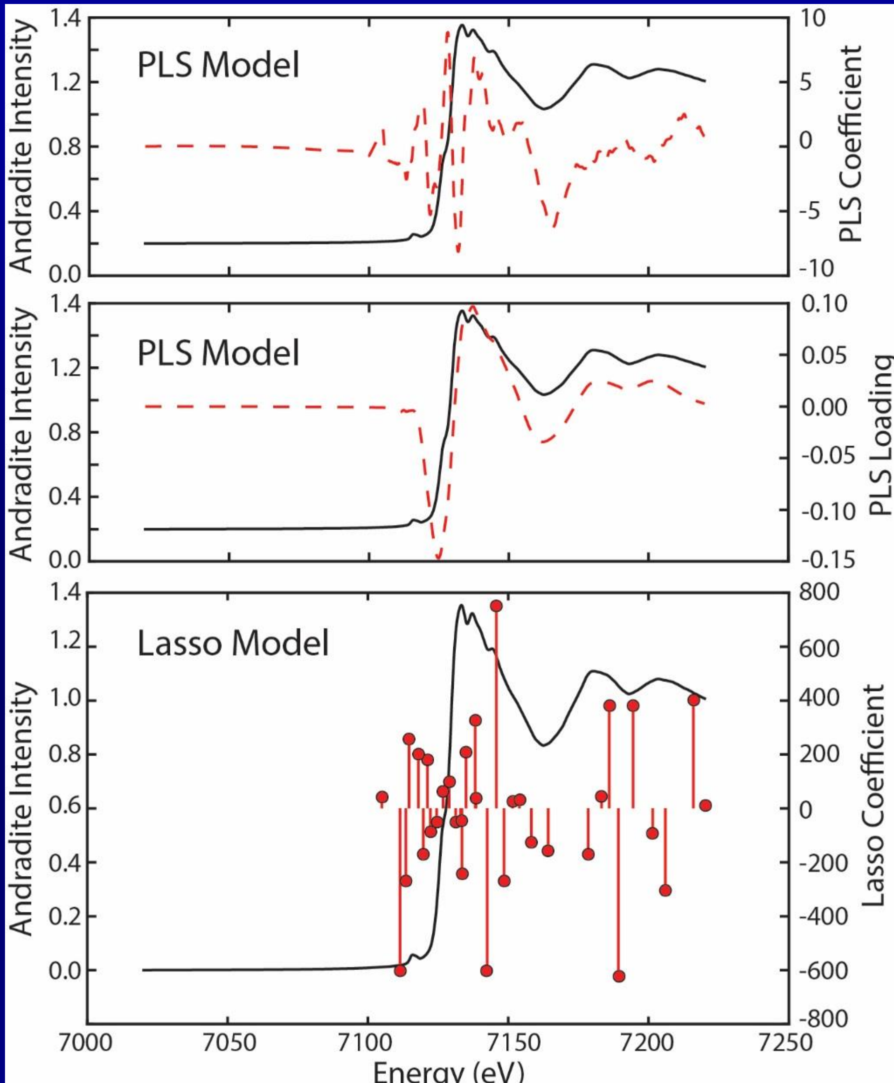
Multivariate Predictions of Fe^{3+} in Silicate Glass and Garnet





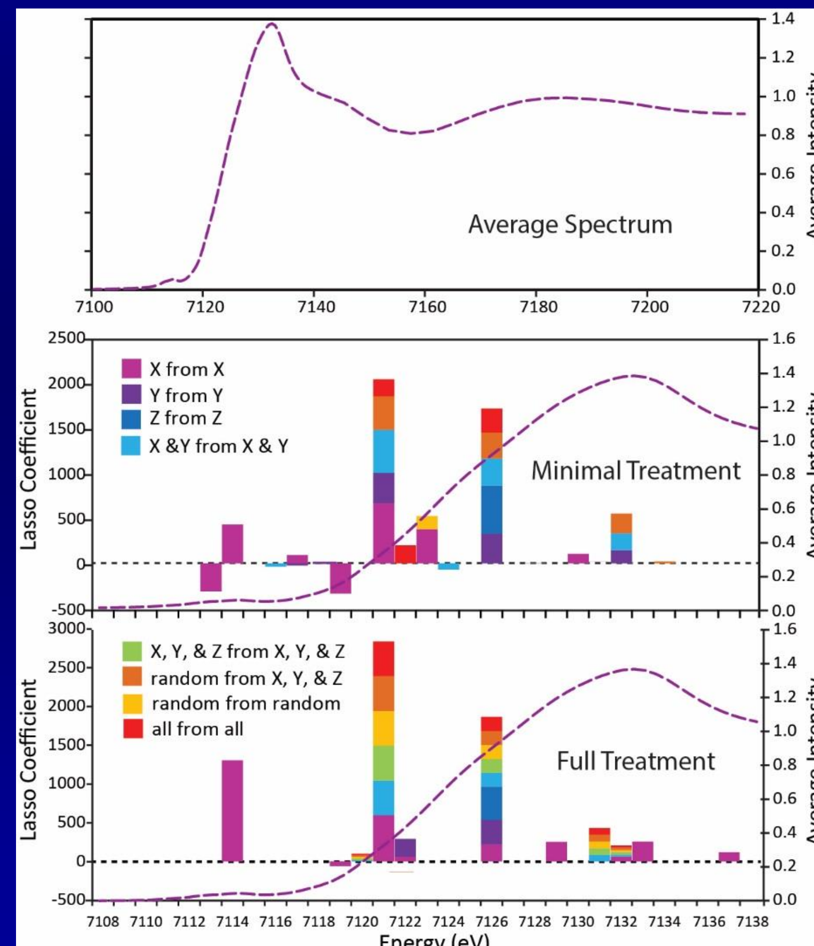


Identification of Key Predictive Channels



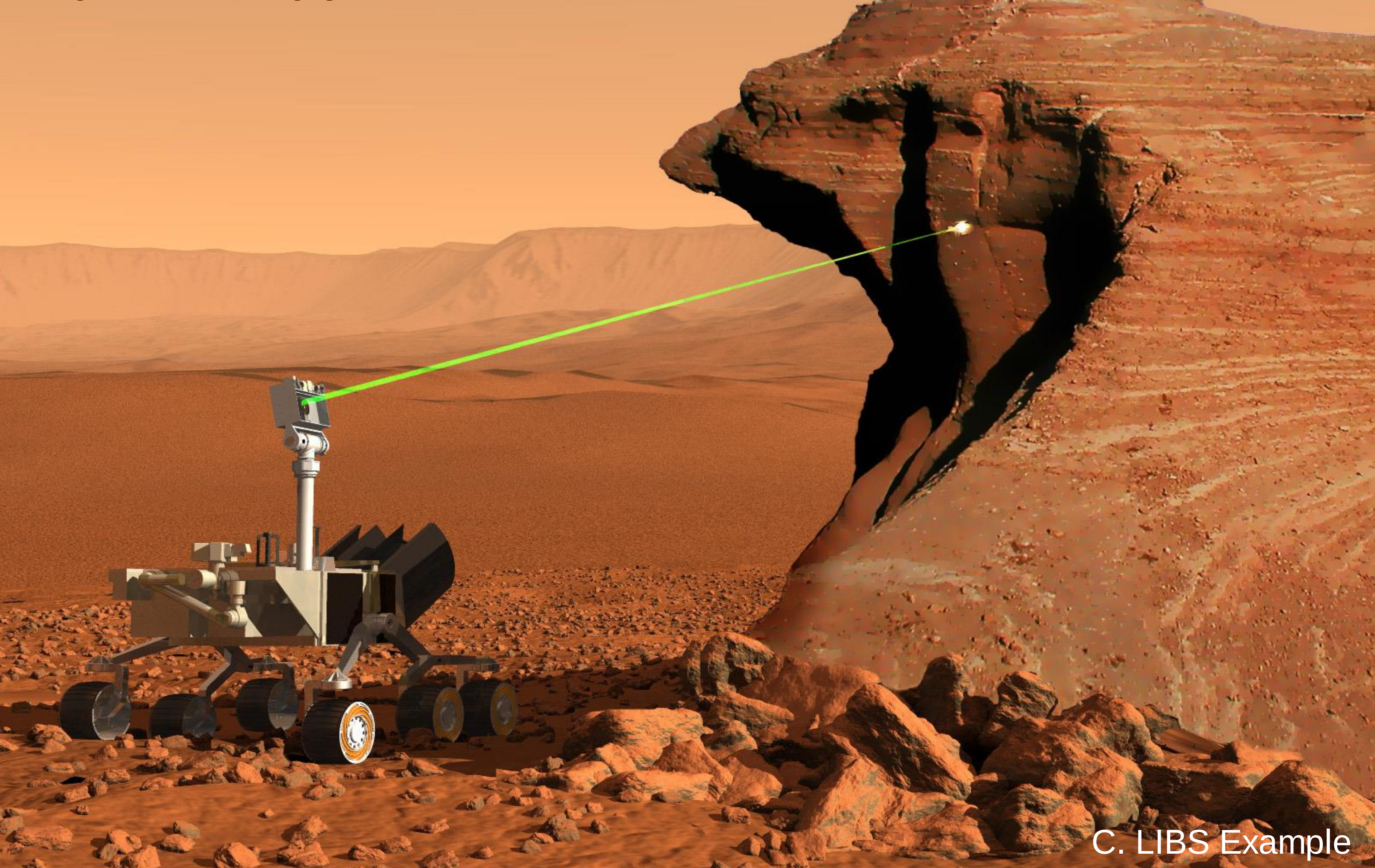
Garnet XAS Data

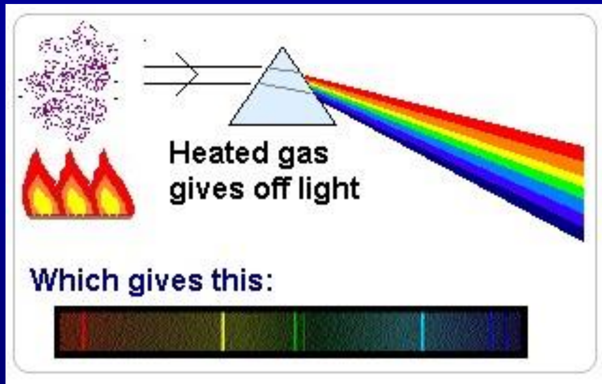
Amphibole XAS Data

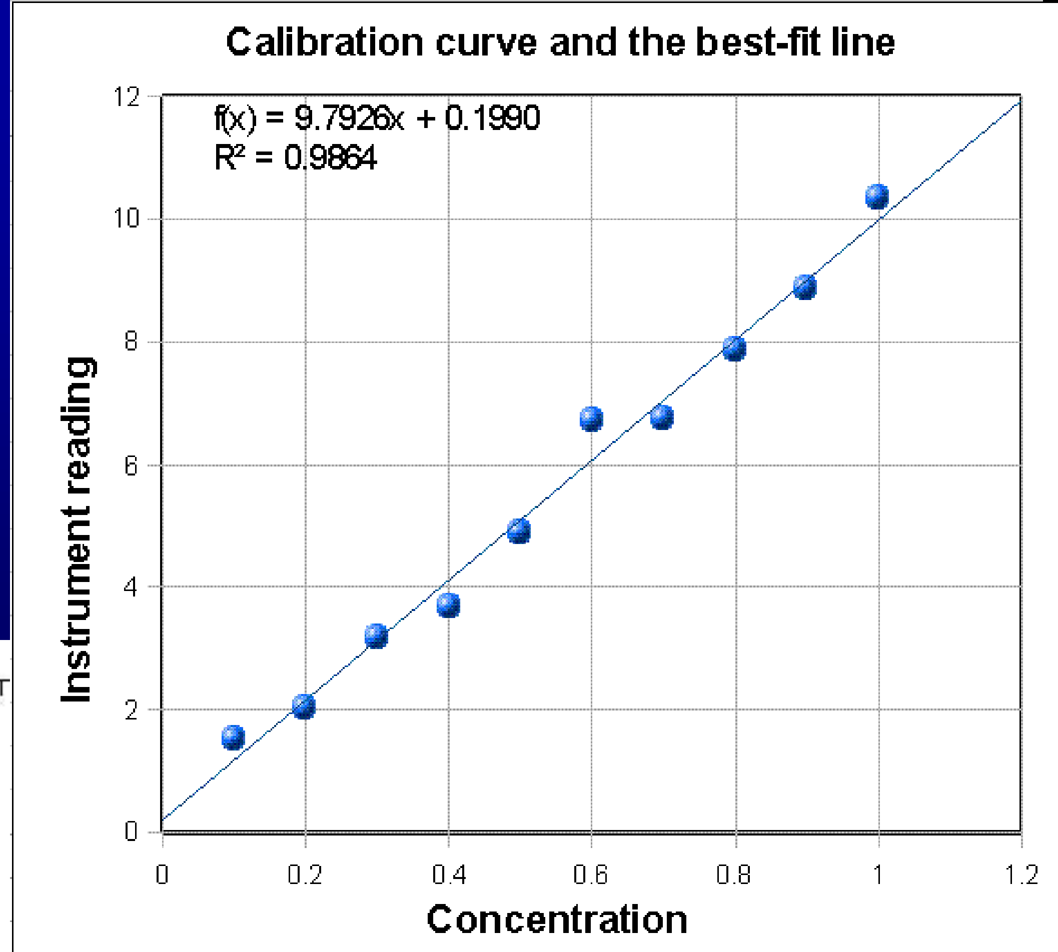


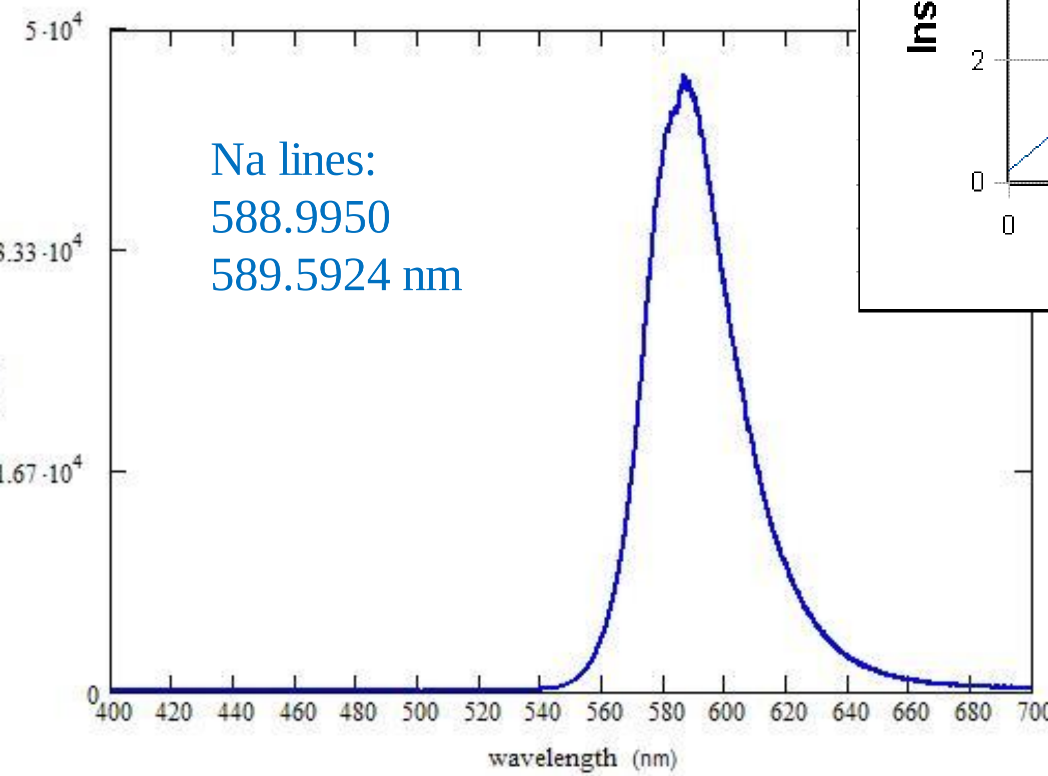
B. XAS Example

Laser-Induced Breakdown Spectroscopy: LIBS



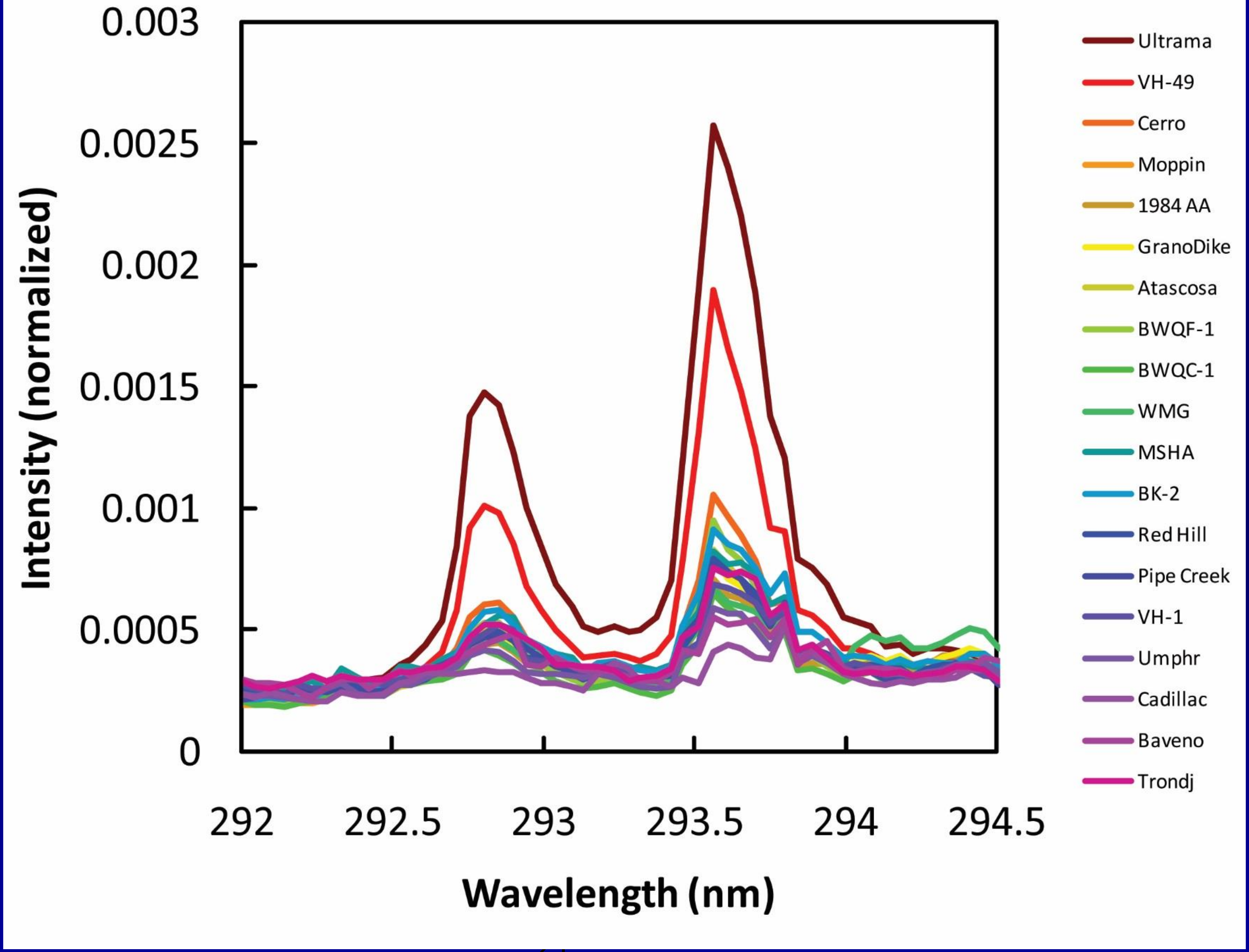






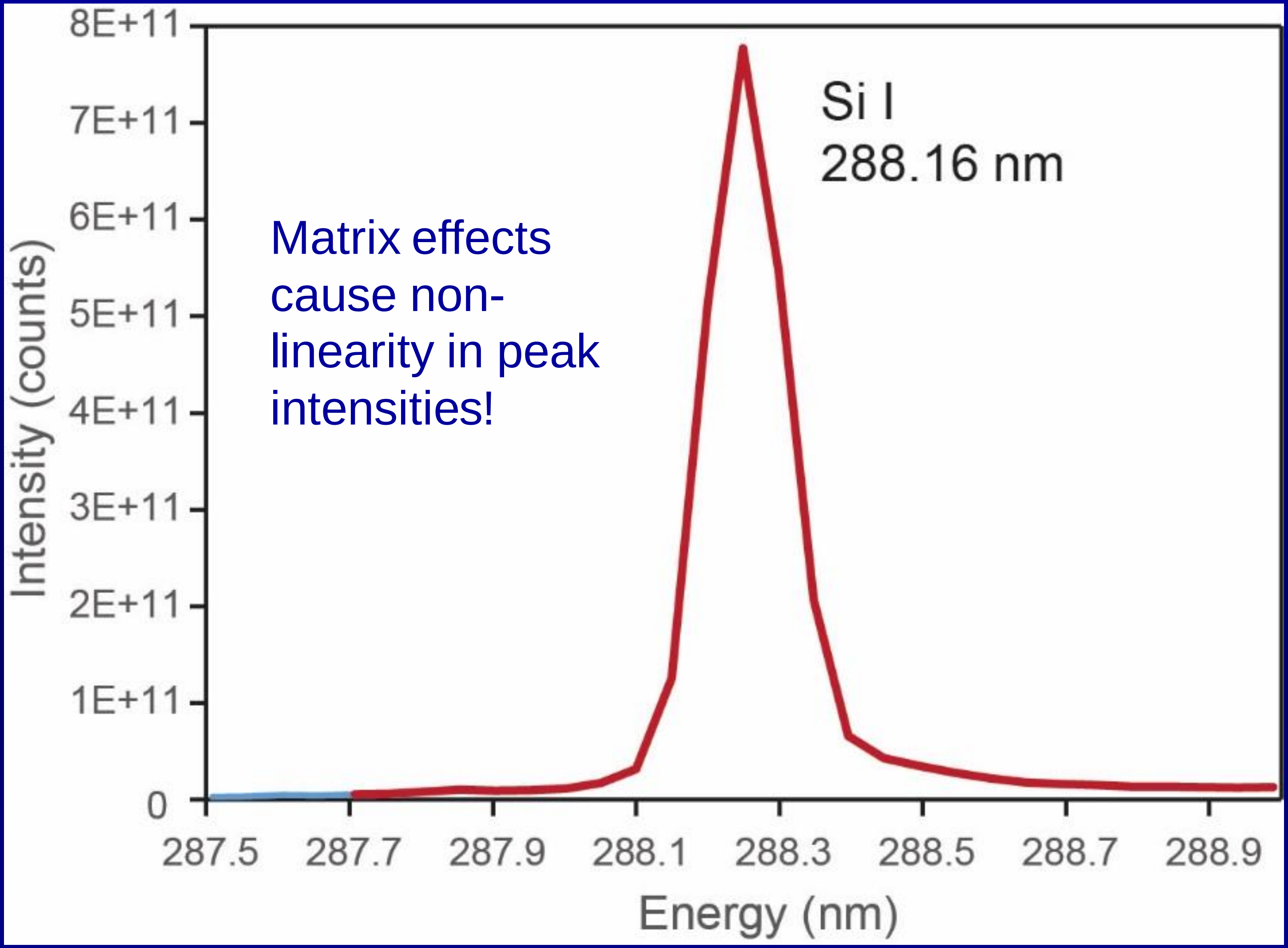
C. LIBS Example

LIBS Challenge for Geological Samples: Matrix Effects!



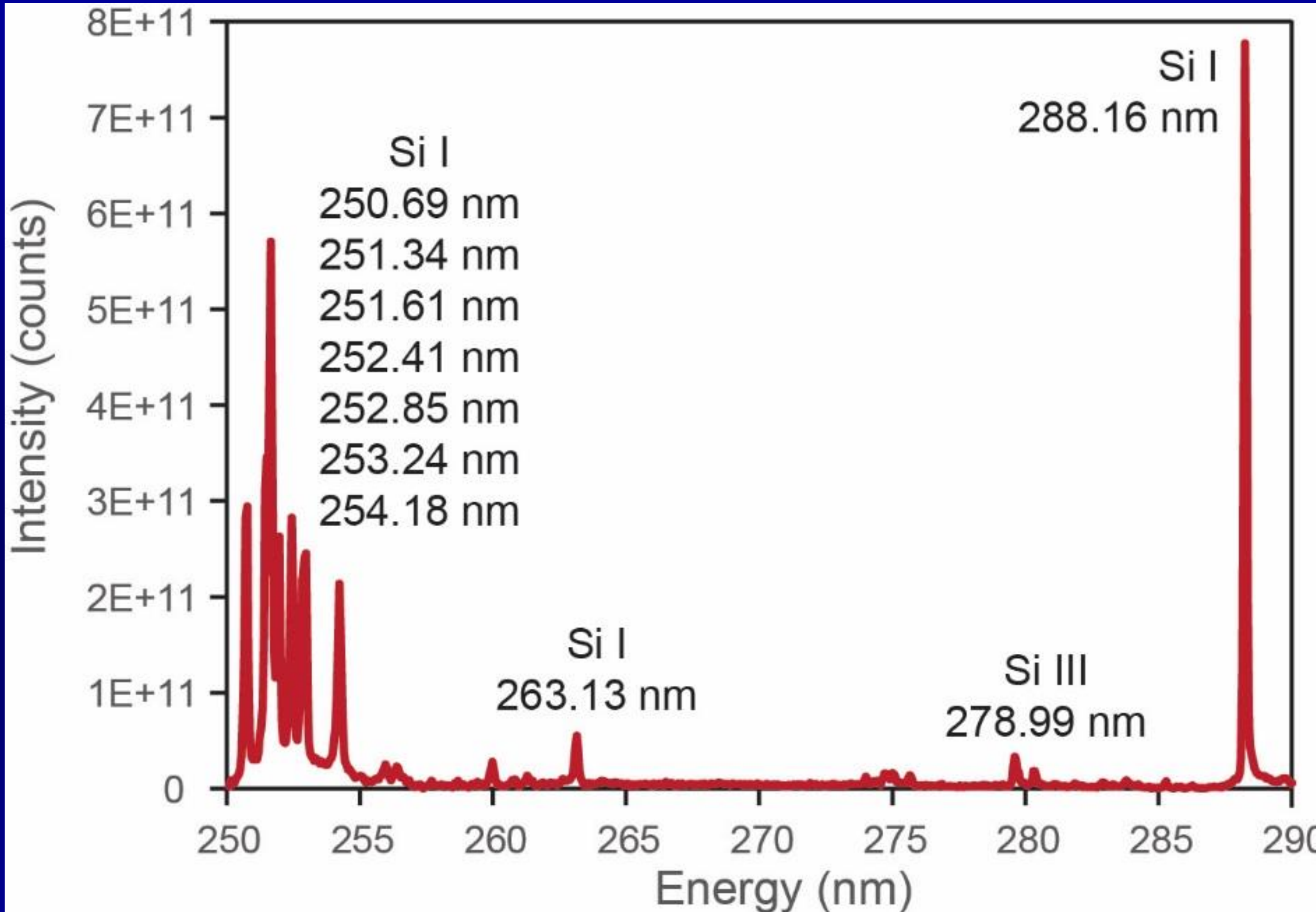
High Ni

Low Ni

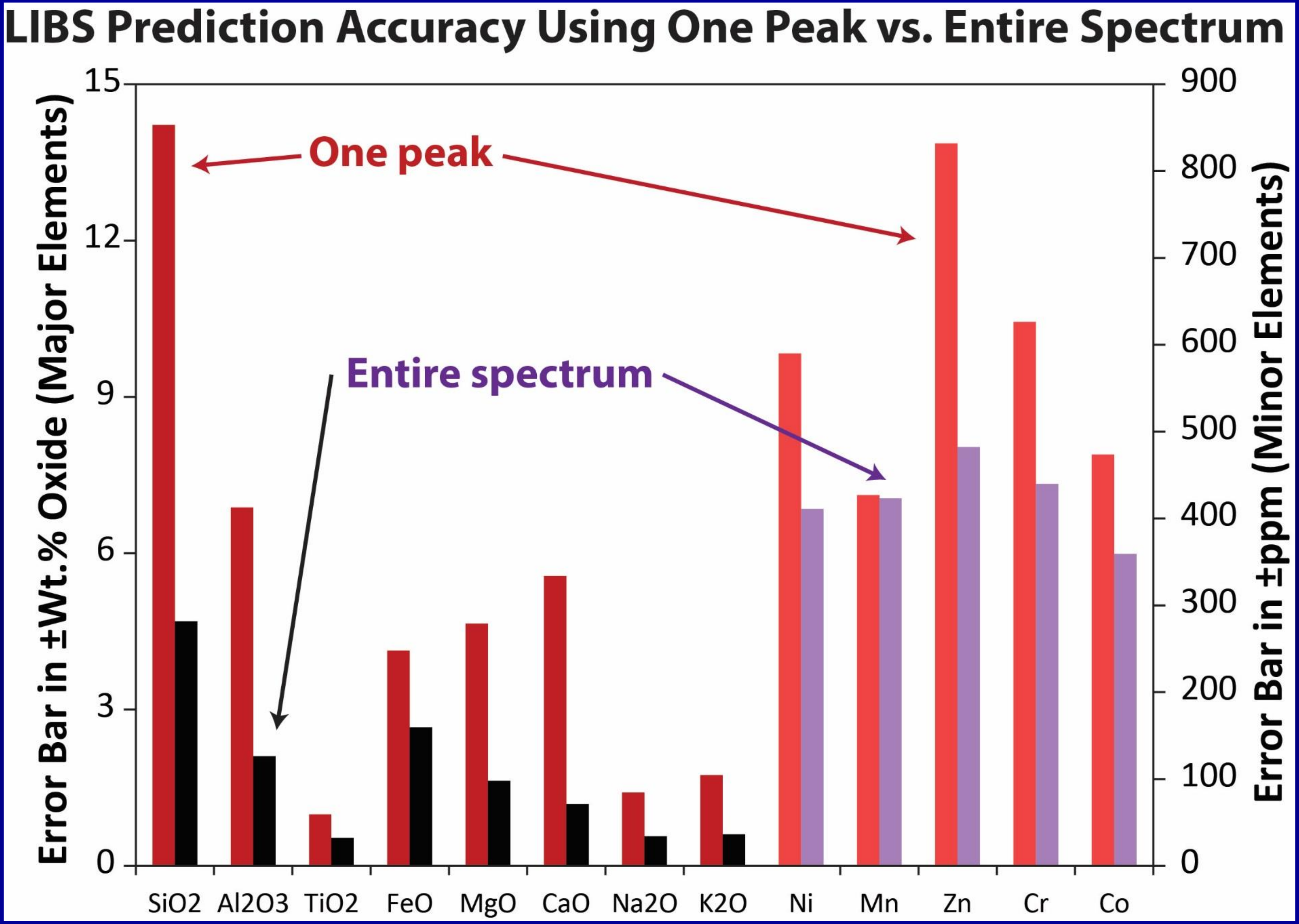


Prediction of SiO_2 contents of LIBS standards (1354 samples, 3 plasma temperatures, Mars conditions) using this single Si I emission line is:

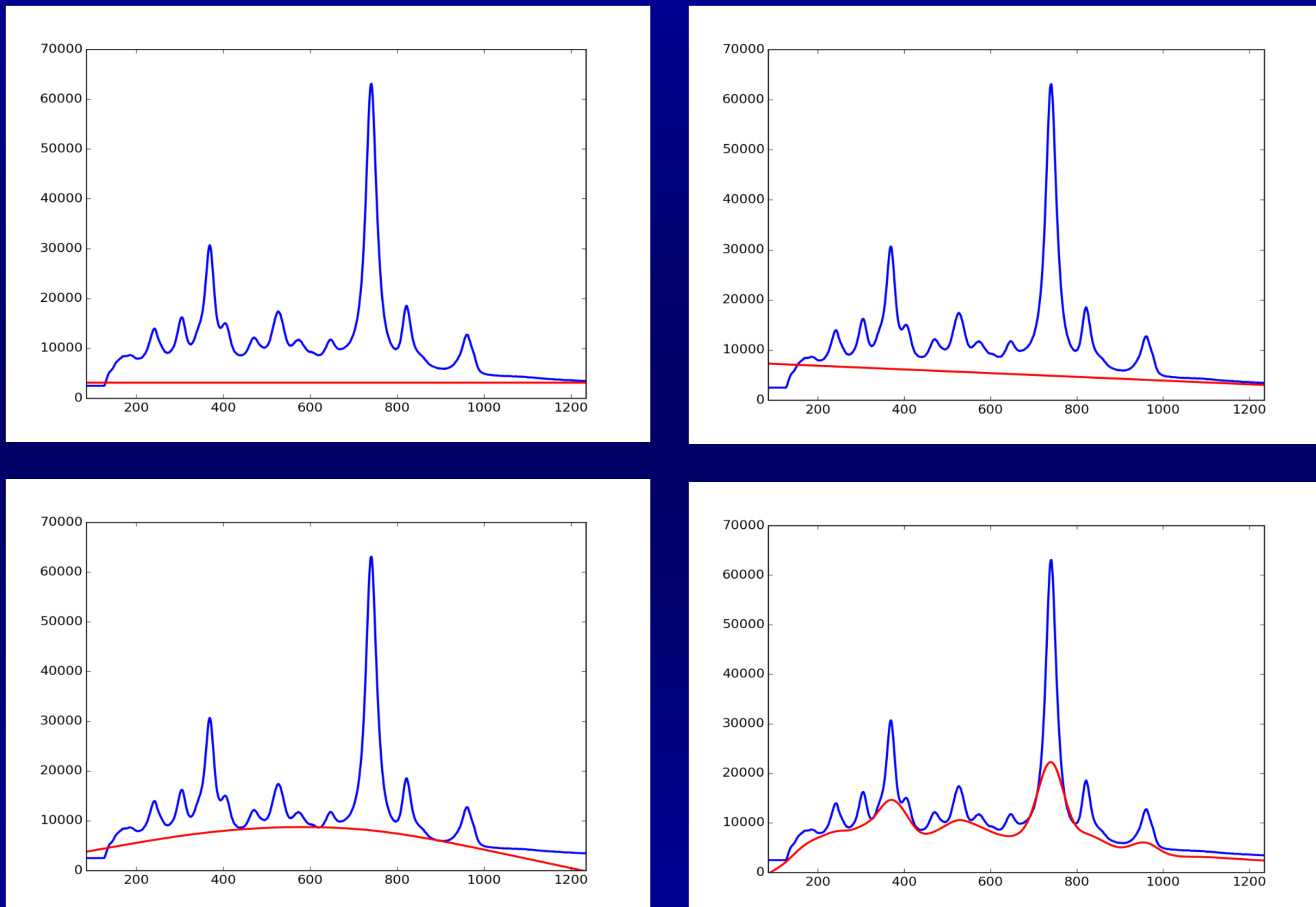
±13.76 wt.% SiO_2



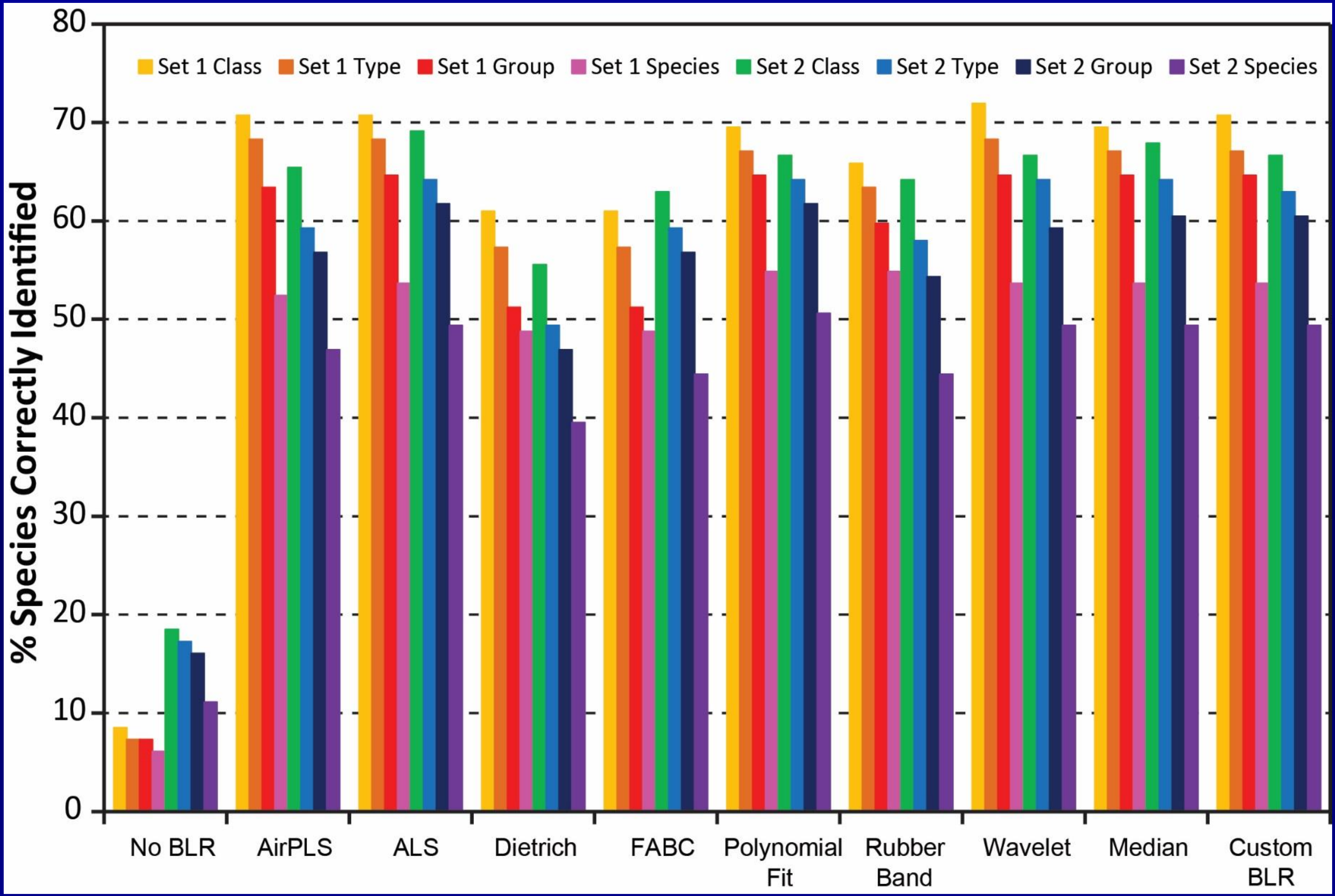
Prediction of SiO_2 using all channels of this spectrum is: ±4.69 wt.% SiO_2


LIBS Prediction Accuracy Using One Peak vs. Entire Spectrum
Error Bar in ±Wt.% Oxide (Major Elements)
Error Bar in ±ppm (Minor Elements)
One peak
Entire spectrum
0
3
6
9
12
15
0
100
200
300
400
500
600
700
800
900
SiO2
Al2O3
TiO2
FeO
MgO
CaO
Na2O
K2O
Ni
Mn
Zn
Cr
Co

Chemometrics (Machine Learning) gives us answers to vexing problems…

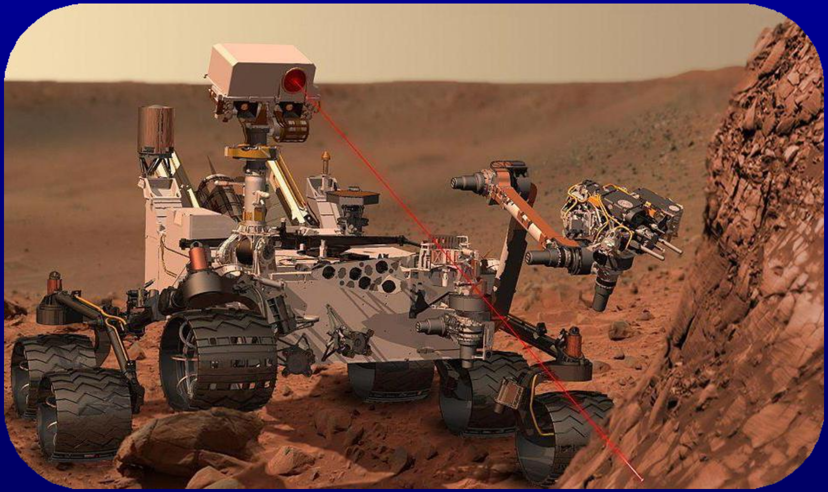


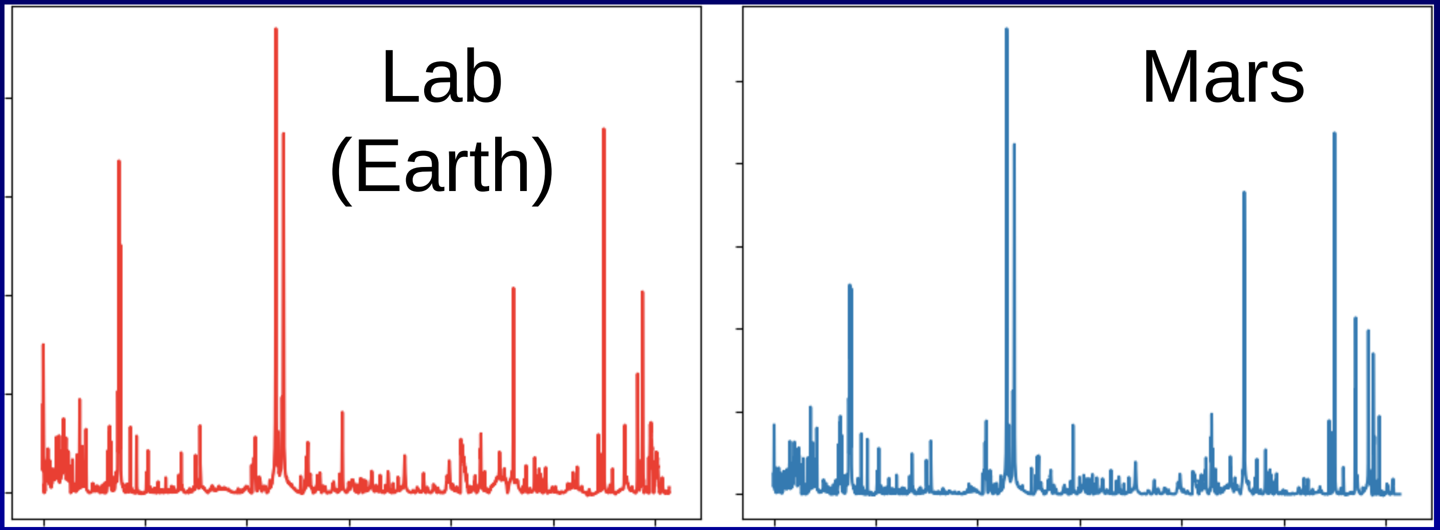
The baseline removal conundrum



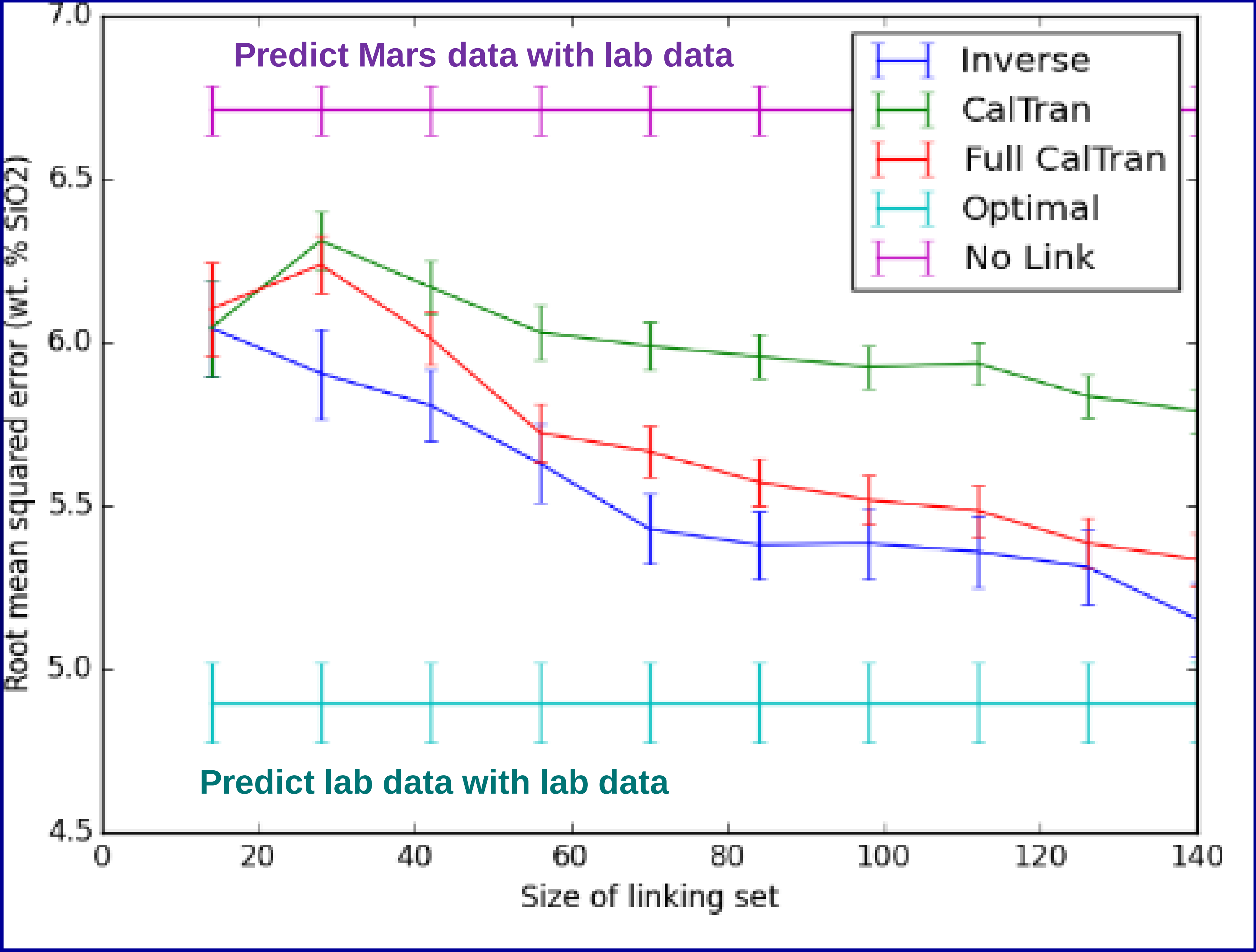
Species matching success for Raman spectroscopy comparing optimized baseline removal methods to no baseline removal (far left) and Custom BLR (far right) by taxonomic rank (Dana classification number).

Calibration Transfer



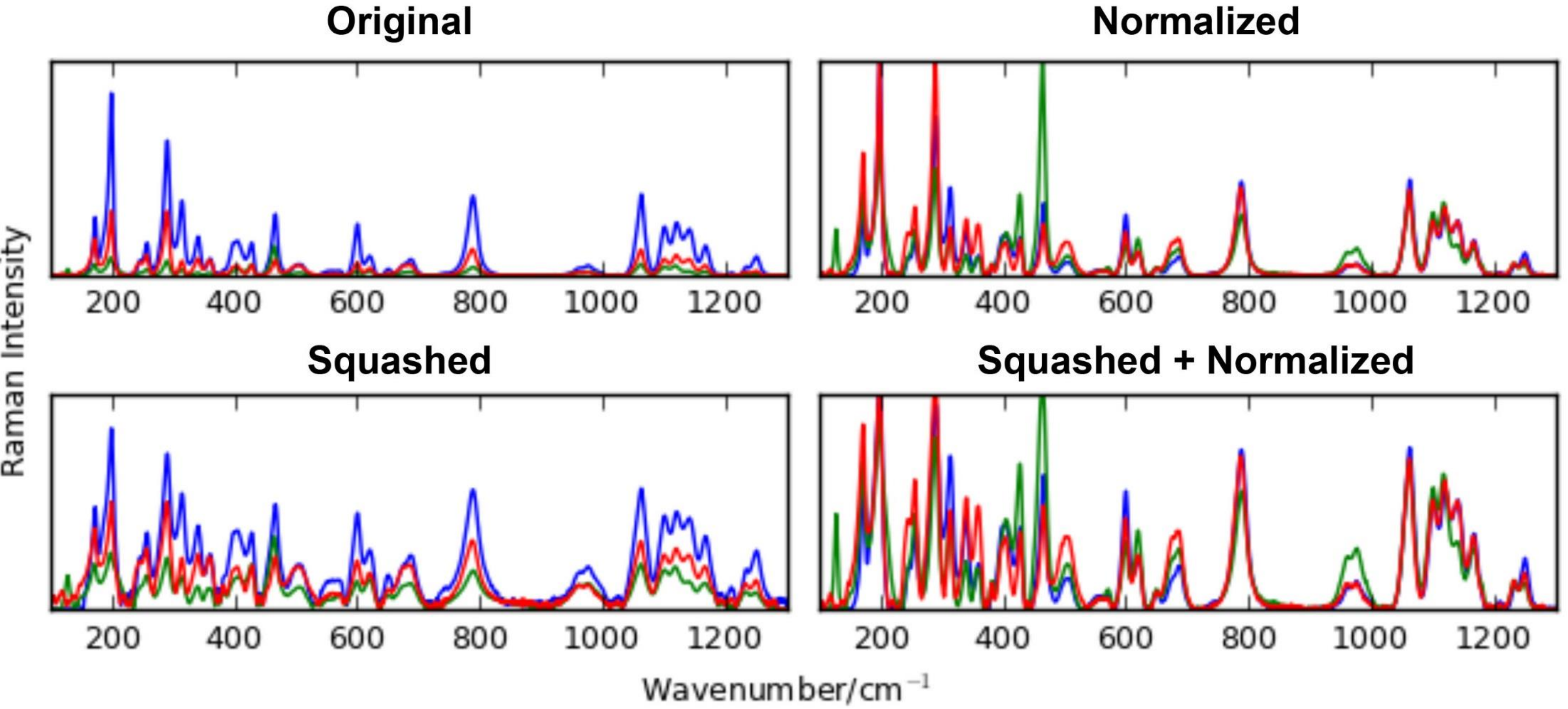


E. Cal Trans


Predict Mars data with lab data
Predict lab data with lab data
Inverse
CalTran
Full CalTran
Optimal
No Link
Root mean squared error (wt. % SiO2)
Size of linking set
7.0
6.5
6.0
5.5
5.0
4.5
0
20
40
60
80
100
120
140

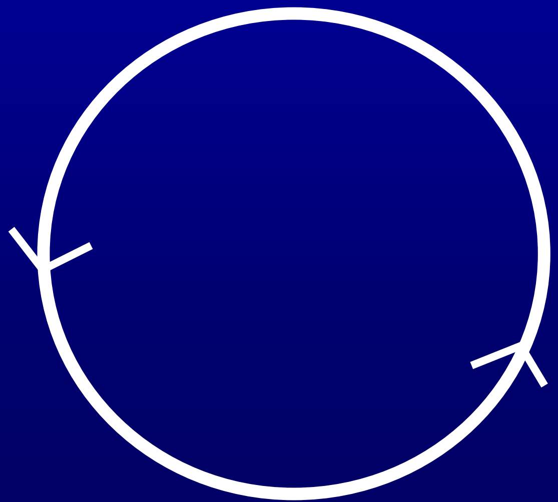
E. Cal Trans

Visualization of Spectral Preprocessing Steps



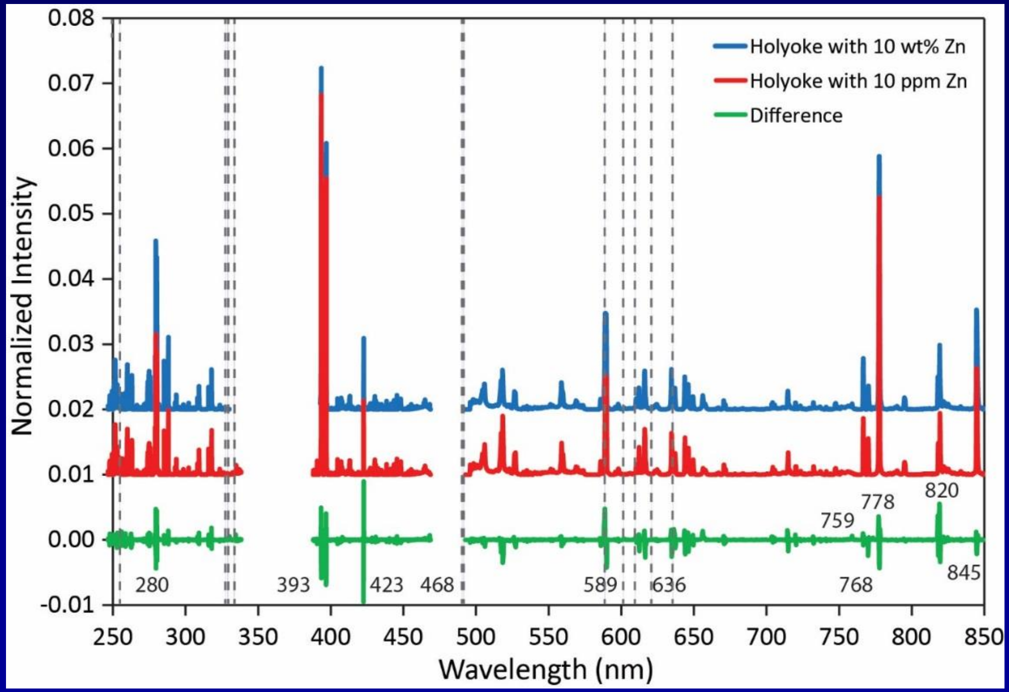
Raman Data

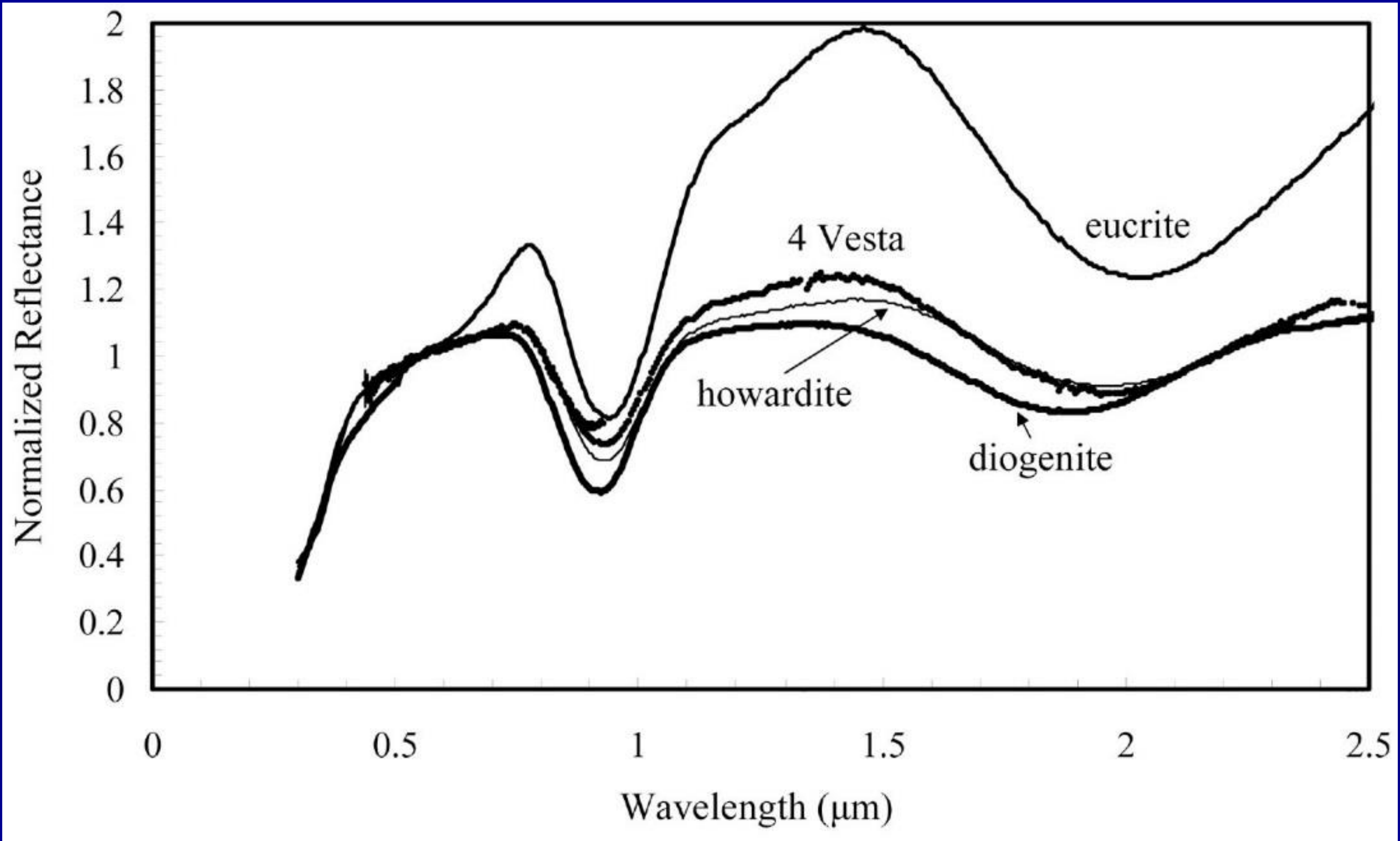
Protocols Based on Individual Peaks and Underlying Physical Principles



Insights from Machine Learning

Machine Learning can enable fundamental Insights into spectra





Burbine et al. (2009)

Barriers to Using Machine Learning

1. Too little overlap between planetary and computer science communities
2. Steep learning curve to understand new methods
3. Reluctance to move on from fundamentals-based approaches
4. Inadequate and silo-ed spectral databases
5. Ignorance of instrumental differences

Benefits of Using Machine Learning

1. Utilize & evaluate all channels of spectral data using automated (objective) feature selection
2. Quantifiable error bars for conclusions based on spectral data
3. Improved instrument design for planetary exploration
4. Calibration transfer between data sets
5. Ability to integrate data from multiple types of spectroscopy in a single model
6. Gain new insights into fundamental physical processes


KEEP
CALM