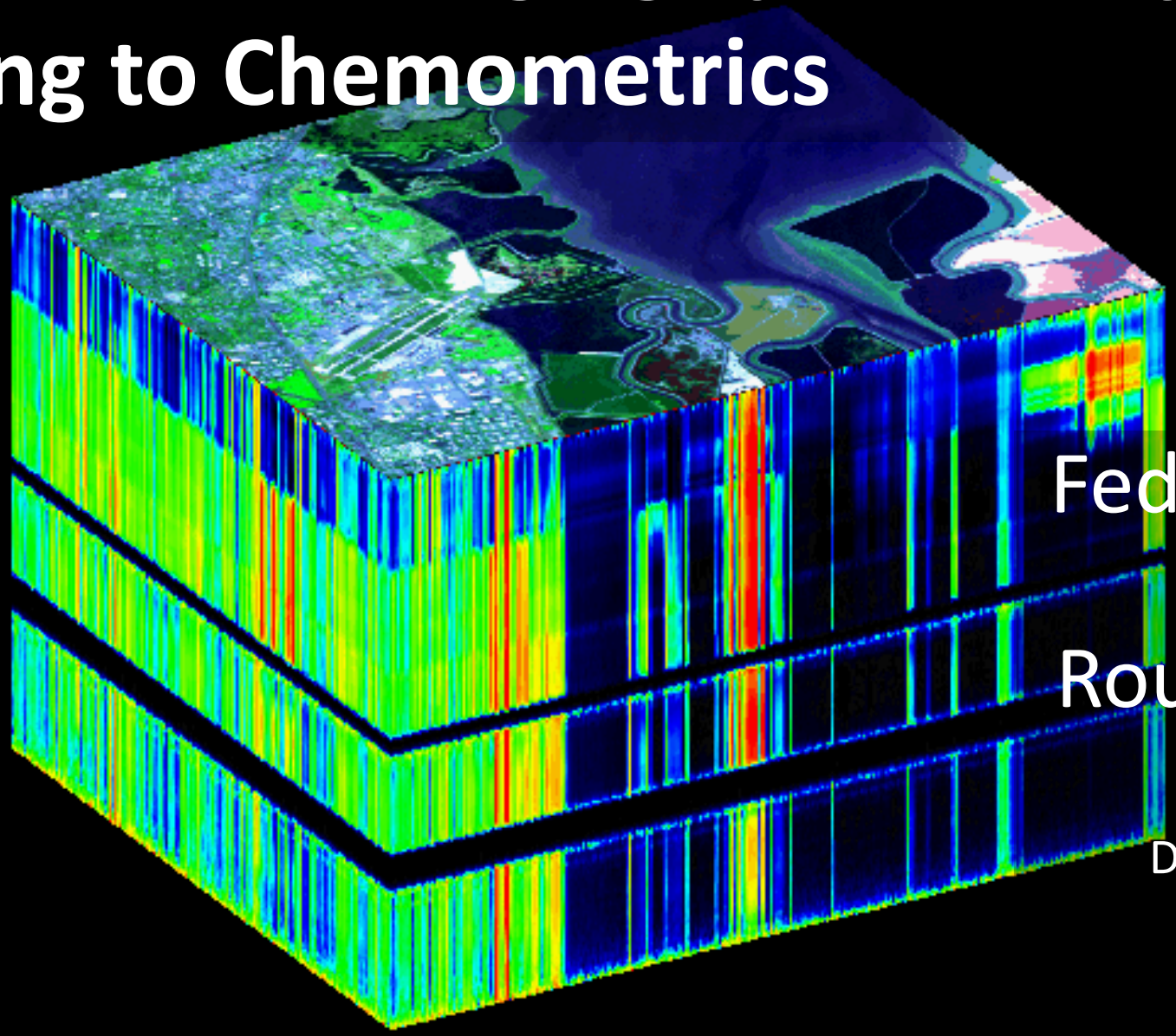
Application of Imaging Spectroscopy and Machine Learning to Chemometrics



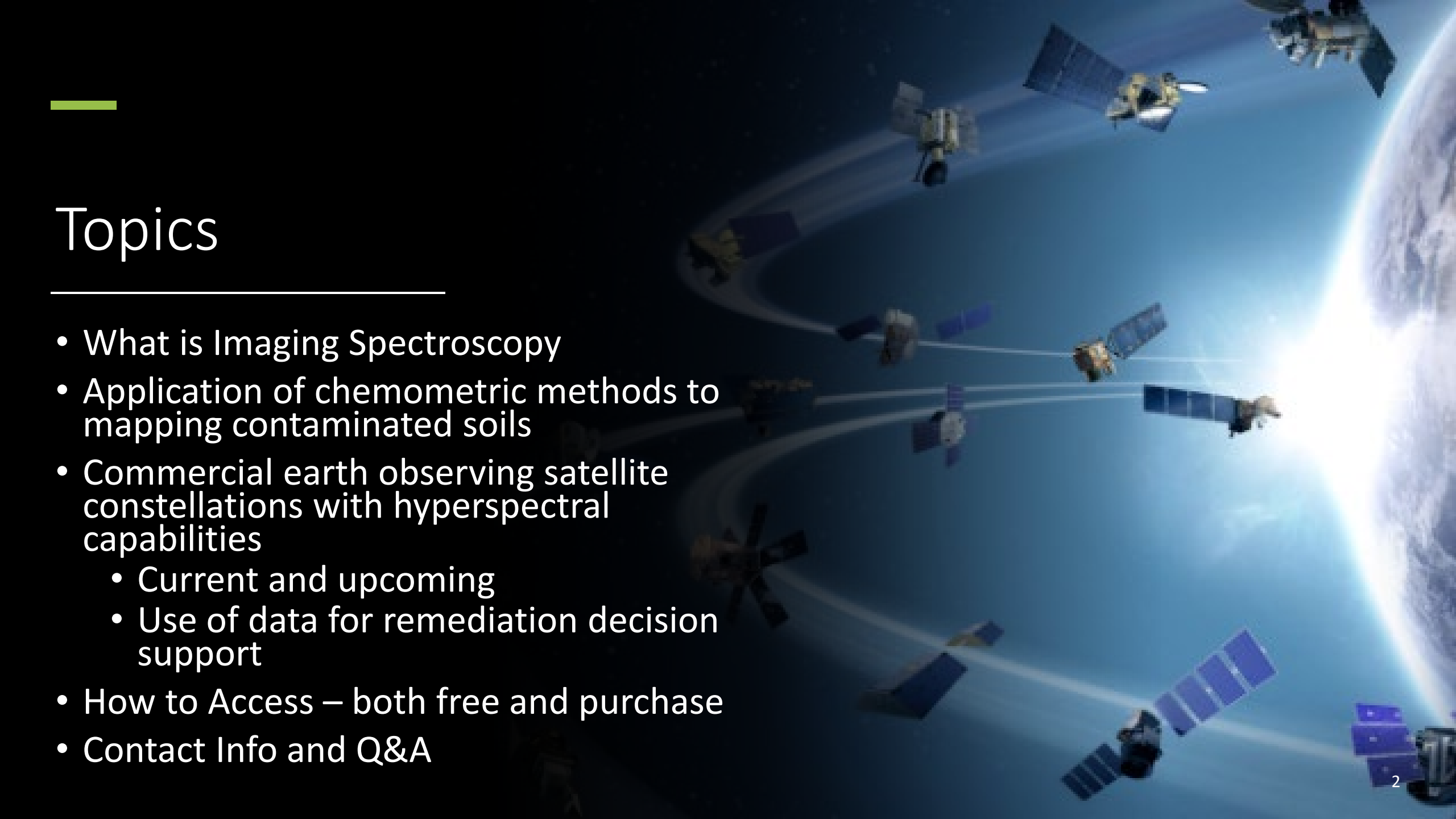
Federal Remediation Technologies Roundtable meeting

June 13, 2022

David J. Williams, EPA/ORD

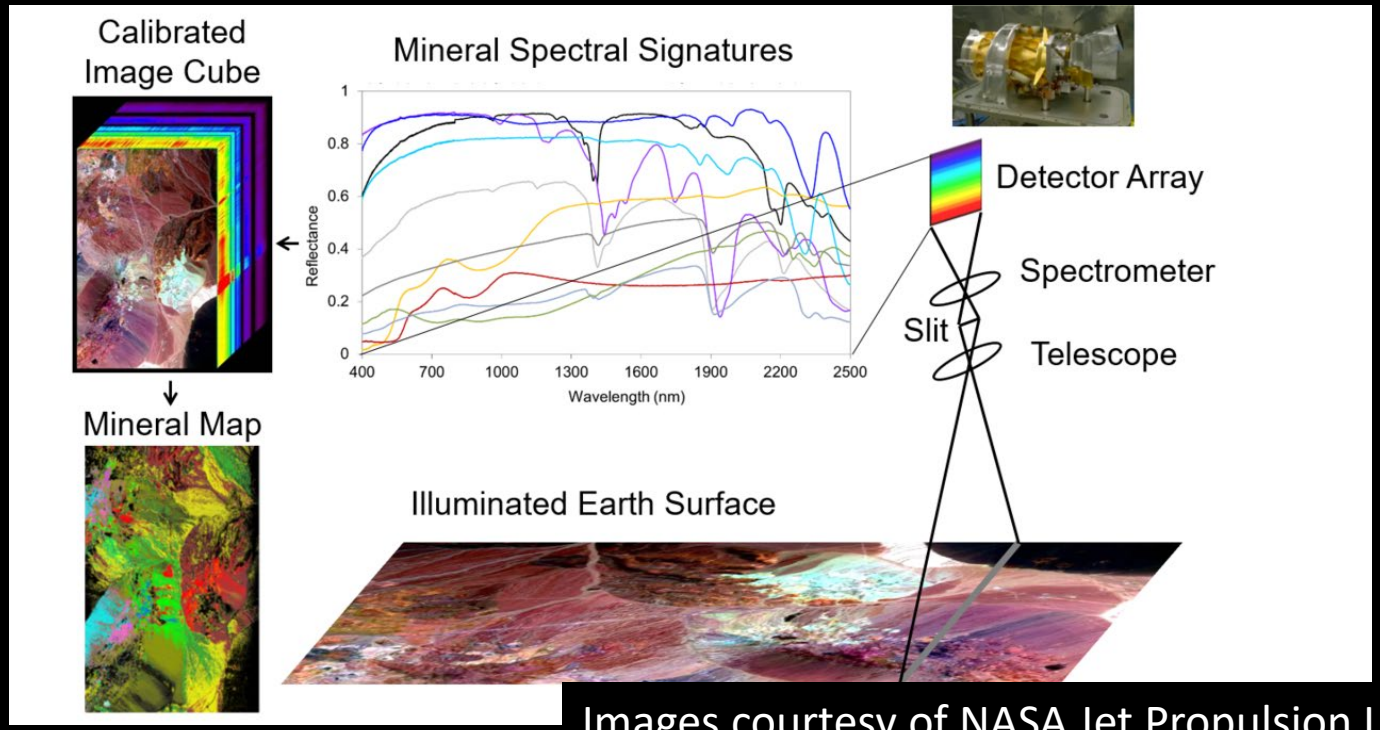
Topics

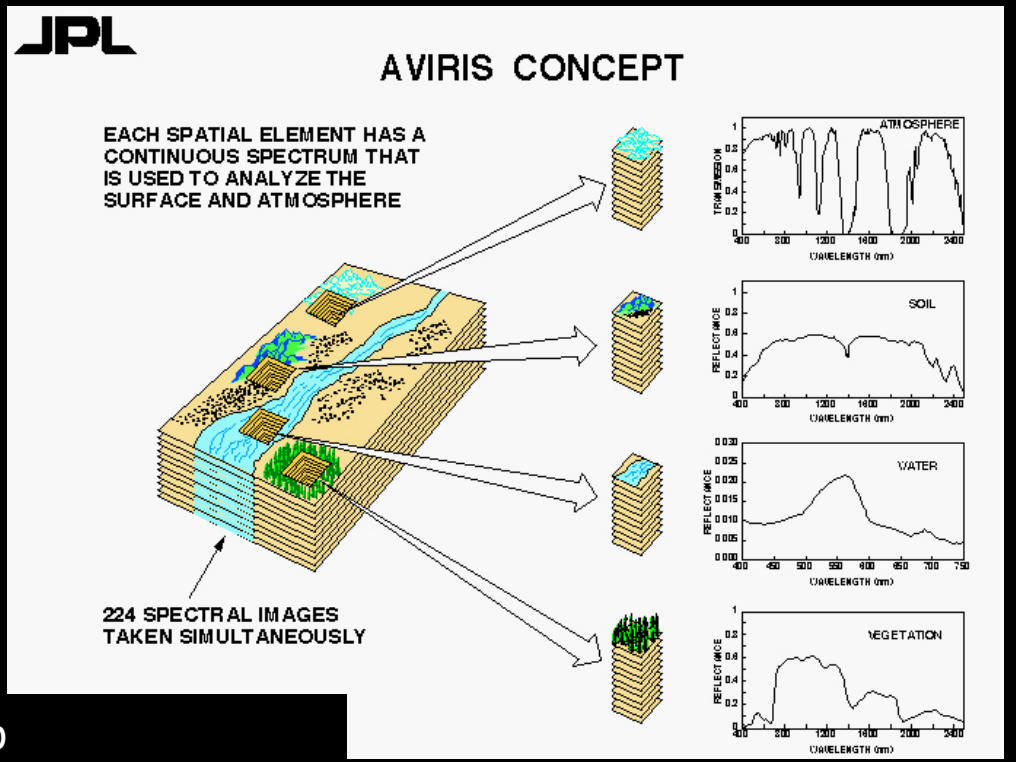
- What is Imaging Spectroscopy
- Application of chemometric methods to mapping contaminated soils
- Commercial earth observing satellite constellations with hyperspectral capabilities
  - Current and upcoming
  - Use of data for remediation decision support
- How to Access – both free and purchase
- Contact Info and Q&A



Imaging spectroscopy

- Also called hyperspectral imaging: 100s of continuous spectral bands
- Collects data that is identical to laboratory spectrometer but in an image format. Used to map materials including soils and wastes





Images courtesy of NASA Jet Propulsion Lab

Example: EPA's TEROS instrument

- 10.4 lbs weight with lens
- 4 x 8.5 x 8 inches (WxHxL)
- power requirements: 12 VDC at 3 W

- Headwall Photonics
- High Efficiency Hyperspec® VNIR-E Series
  - Spectral range: 400-1,000 nm
  - High Efficiency grating
  - 923 spectral channels
  - 2.5 nm spectral resolution
  - 1600 spatial channels
  - 100 fps frame rate
  - sCMOS FPA, TE cooled
  - Very high SNR

- 8.5 lbs weight with lens
- 4 x 8.5 x 8 inches (WxHxL)
- power requirements: 12 VDC at 3 W

- Headwall Photonics
- Hyperspec SWIR
  - Spectral range: 950-2,500 nm
  - High Sensitivity SWIR MCT camera; 320 x 256 pixels; 14 bits
  - 320 spatial bands x 200 spectral bands

TEROS Configurations

TEROS can fly on commercial Cessna aircraft
Cost to fly ~$10k/week (pilot + aircraft)

Pod upside down to show instruments

These instruments are integrated into the pod.

- VNIR & SWIR imaging spectrometers
- Hi-Res camera for “LiDAR-like” data
- Thermal camera
- GPS/INS

TEROS can fly on NASA aircraft:

- King Air B-200 aircraft for large projects.
- Cost: $2,500/hour

- NASA Cessna 206H for local projects on the east coast
  Cost: $400/hour
- Flown using wing strut
- or pod

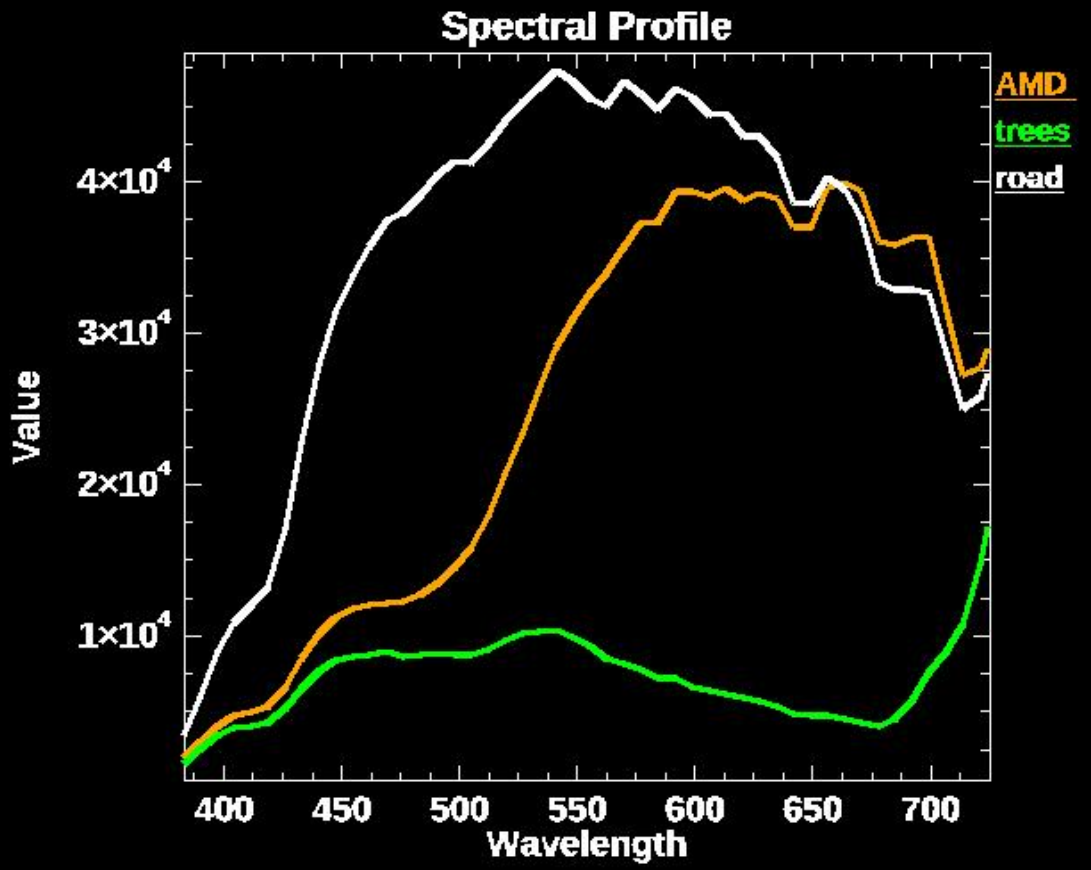
TEROS instruments in wing strut

VNIR spectrometer

Uncorrected image

Orthorectified image

VNIR data example



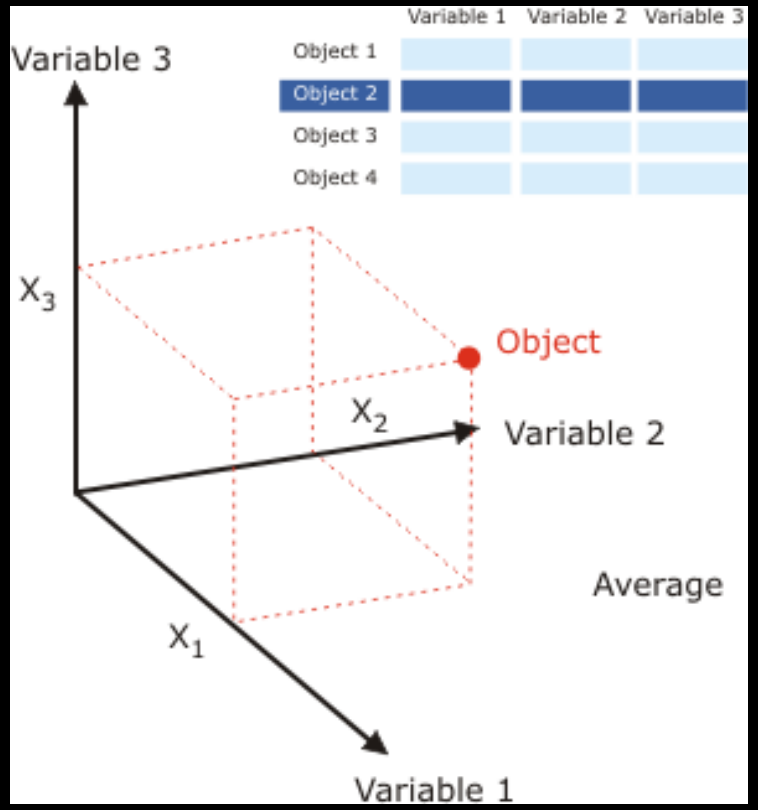
Chemometrics and imaging spectroscopy

- Hyperspectral imaging is great for mineral mapping, material ID, vegetation monitoring and water quality assessment. But contaminated soils are a special case due to the complexity of the metal and matrix
- For example, metals and metalloids (arsenic) can be adsorbed to iron oxides, clays, organic matter depending on soil media, chemical reactions and other factors
- Chemometrics methods leverages machine learning to use spectral information from sample data to predict contaminate levels
  - Machine learning uses training data with known soil metal concentrations

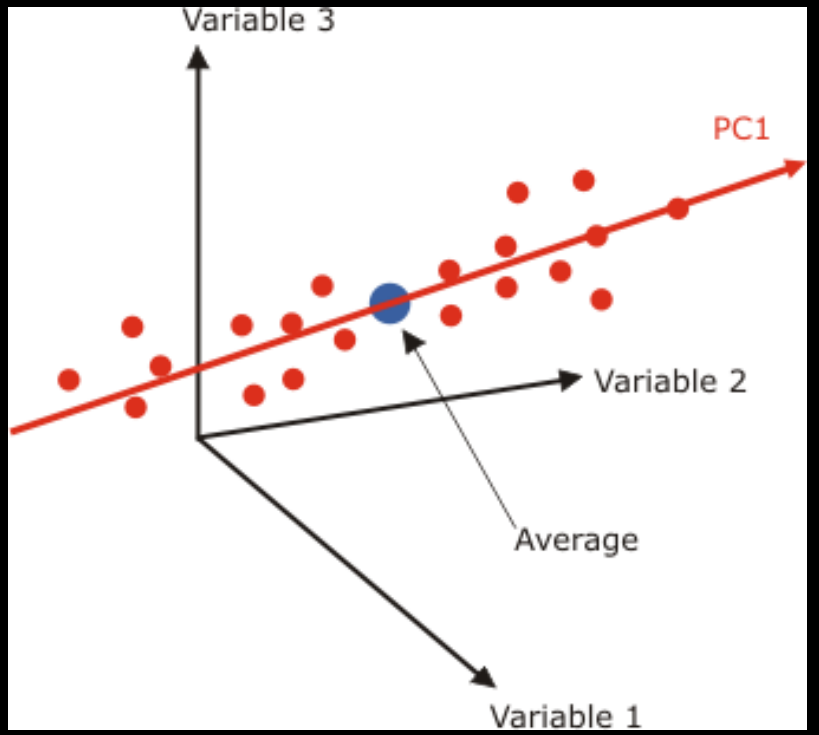
Basics of Chemometrics using PLSR

- Partial Least Squares Regression (PLSR) is a method for relating two data matrices, X and Y, by a linear multivariate model, but goes beyond traditional regression in that it models also the structure of X and Y.
  - In this case, X is spectra and Y is XRF
- The method creates a statistical model that is calibrated using the data set. A subset of the data is used to validate the model.
- PLSR uses an approach similar to principle components analysis (PCA)

PLSR, like PCA, reduces the spectral space to linear combinations



Spectral space: (Variables 1,2,3 are spectral bands like red, blue, green)

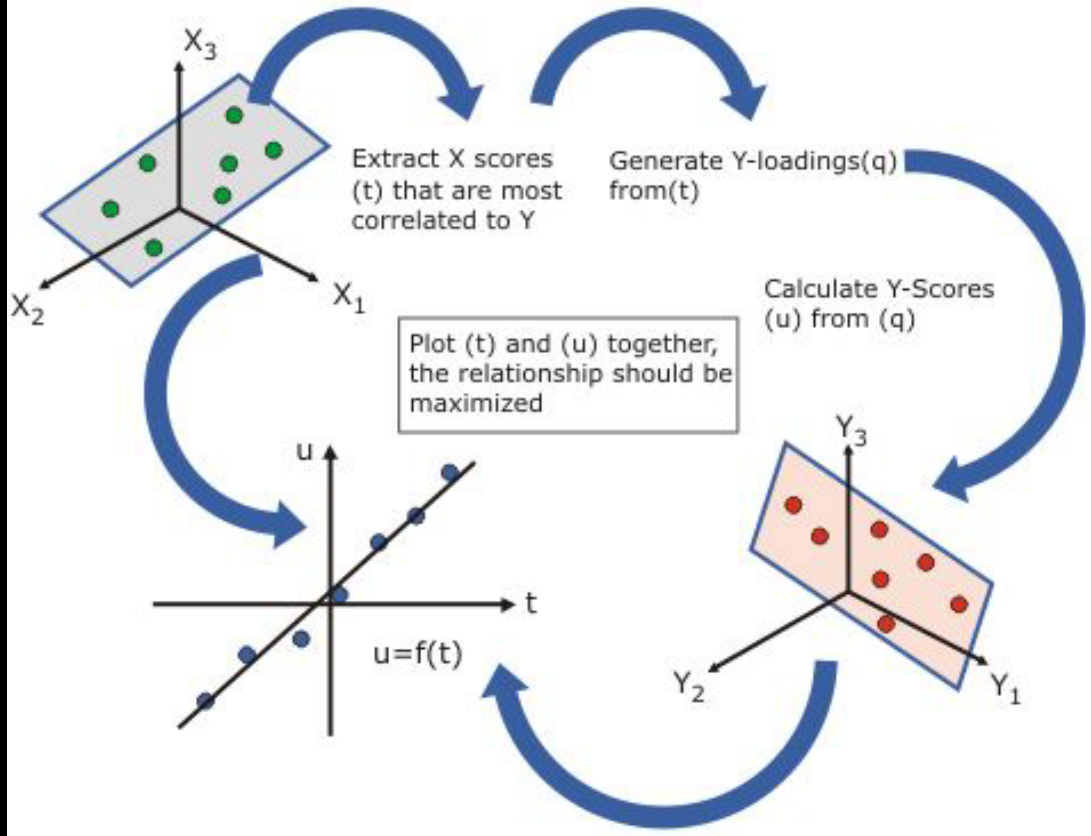


PC space: (PC1 is an uncorrelated linear combinations of the input bands)

Basics

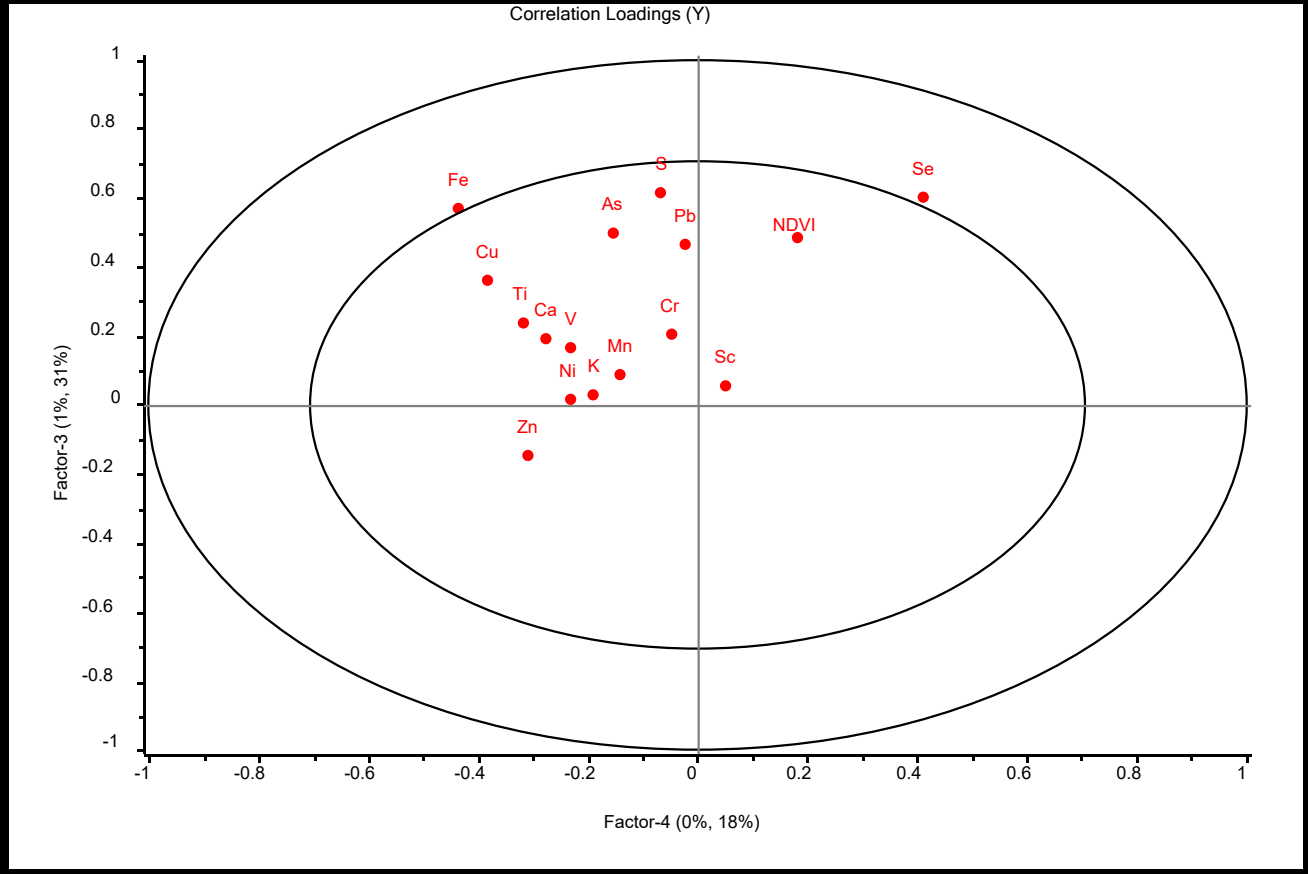
PLSR maximizes the covariance between X and Y. In this case, convergence of the system to a minimum residual error is often achieved in fewer factors than using PCR. This is in contrast to PCR, which first performs Principal Component Analysis (PCA) on X and then regresses the scores (T) vs. the Y data. A conceptual illustration for PLSR is shown graphically below.

PLSR Procedure



PLSR may be carried out with one or more Y variables, meaning that multiple Y responses can be used during regression modeling.

PLSR analysis – loading factors



Factors are linear combinations that best explain the data. The loadings indicate how the factors are related to the data. The inner circle is 50% explanation, outer is 100%. In this example, Fe and Se explain 50% of the correlation, As about 30%.

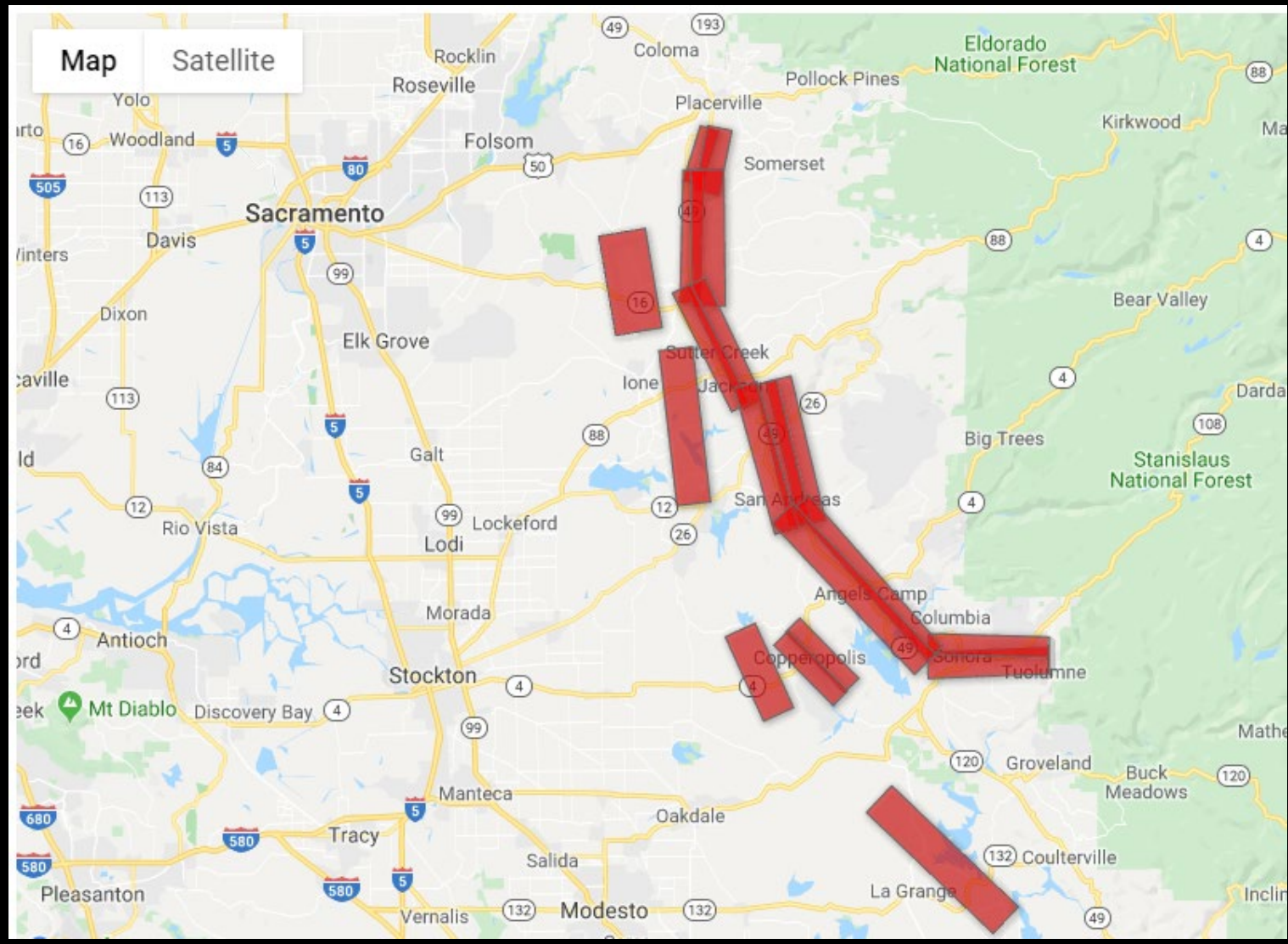
Example: Mapping As contaminated soils in the Mother Load gold belt in California

- Flew NASA JPL's Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) Next Generation (NG)
- Collected data at a ground sample distance (GSD) of 4.5m
- Utilized XRF measurements from previous investigations by EPA Superfund as PLSR training data

Parameter	Value
Wavelength	380 nm to 2510 nm
Spectral Resolution (FWHM, minimum)	5nm ± 0.5 nm
Field of View	36 ± 2 degrees with 600 resolved elements
Instantaneous Field of View	1.0 1.4 mrad ± mrad
Spatial Sampling (maximum observed at resolved elements)	1.0 mrad ± 0.1 mrad
Spectral Distortion (smile)	Uniformity > 97%
Spectral Distortion (keystone)	Unformity >97%
FPA	480 (spectral direction) X 640 (cross track)
Frame Rate	10 - 100 frames per second
Pixel Size	27 microns x 27 microns
Calibration	On-board calibrator
Data Resolution	14 bits
Data Rate	Up to 74 MB/s of throughput
Data Volume	Up to 1.0 TB of raw data before disk swap

Locations of AVIRIS-NG and ASO flights

Instruments flown on Dynamic Aviation's King Air A200 aircraft

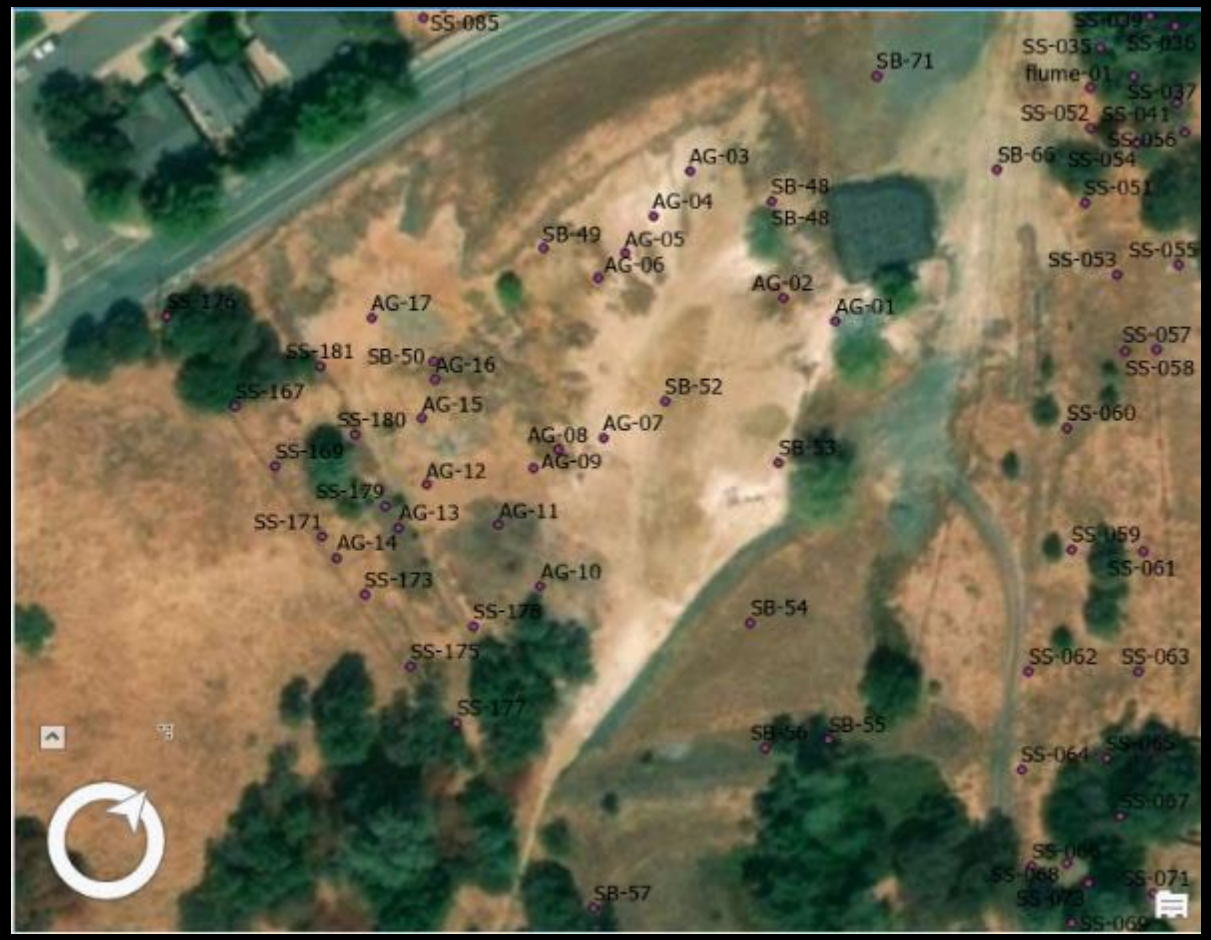


Example of an AVIRIS-NG image

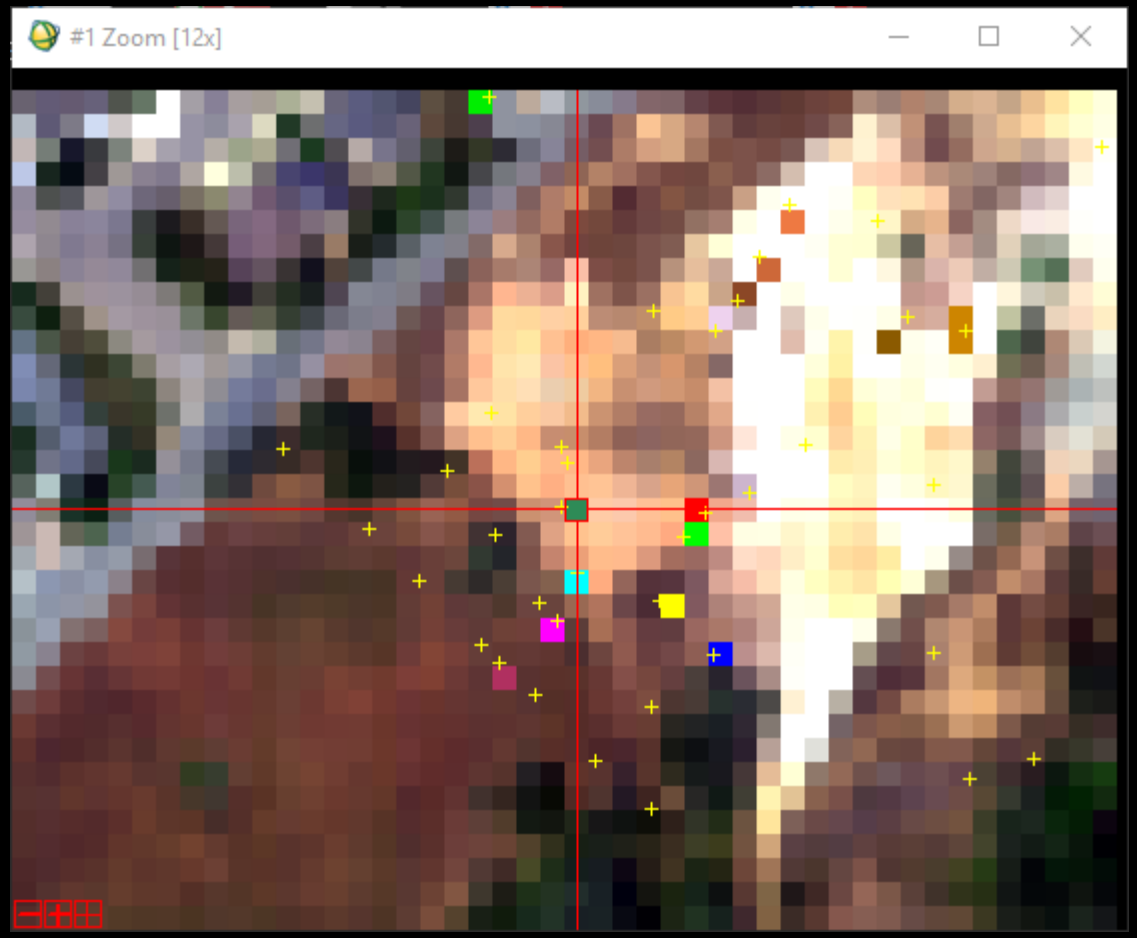
FLIGHT NAME: ang20190623t202347
Date: 6/23/2019 UTC 20:29
Site Name: JacksonSutter_1
Investigator: David J. Williams
Comments: Alt = 16 kft
SOG = 146 kts
Clouds = Clear
Pixel Size: 4.5m
Solar Elevation: 15.48
Solar Azimuth: 195.91
Rotation: 28

Closeup of Argonaut Mine

Creating training data with AVIRIS-NG image

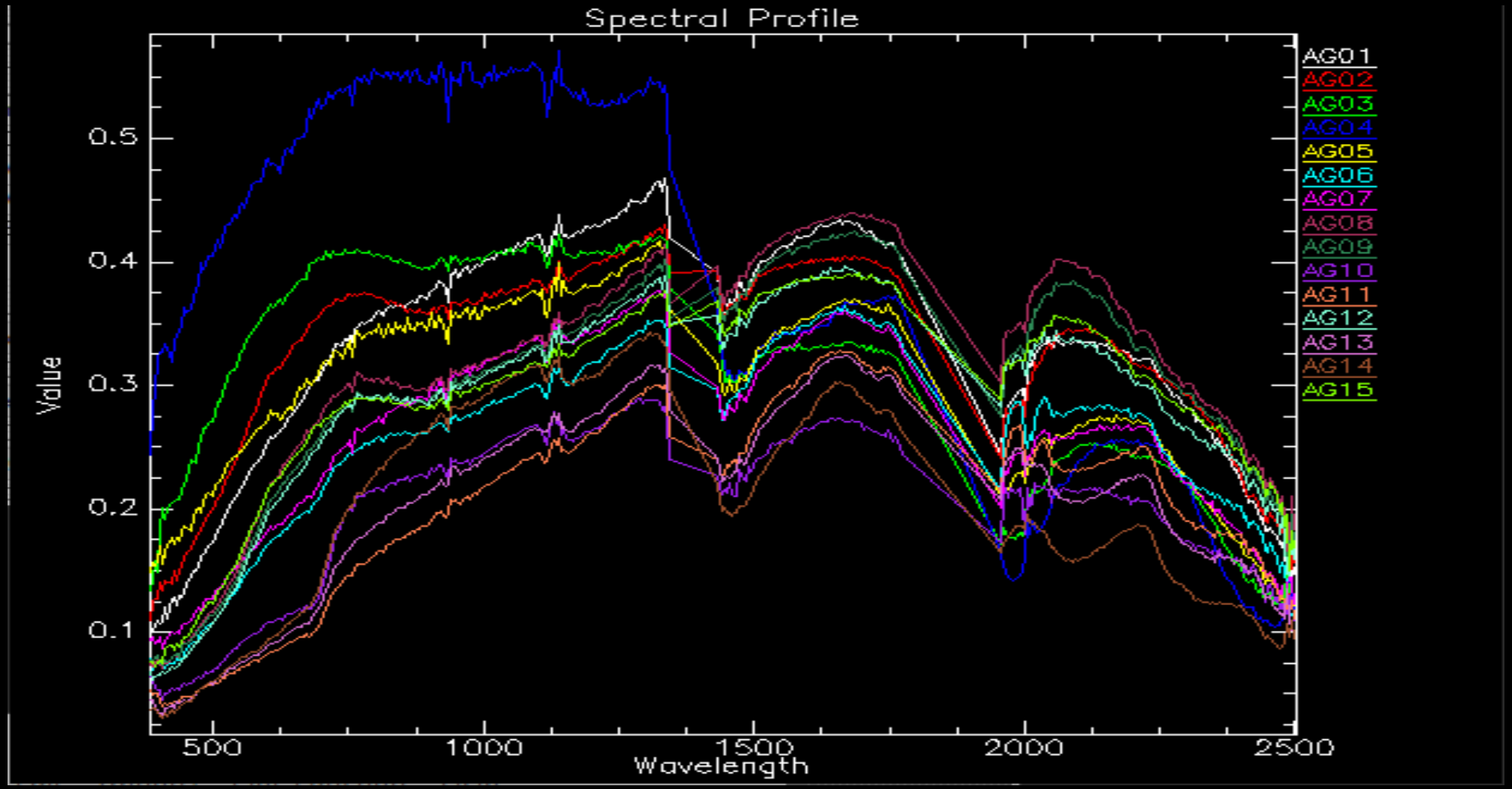


Ground sampling locations on GIS map



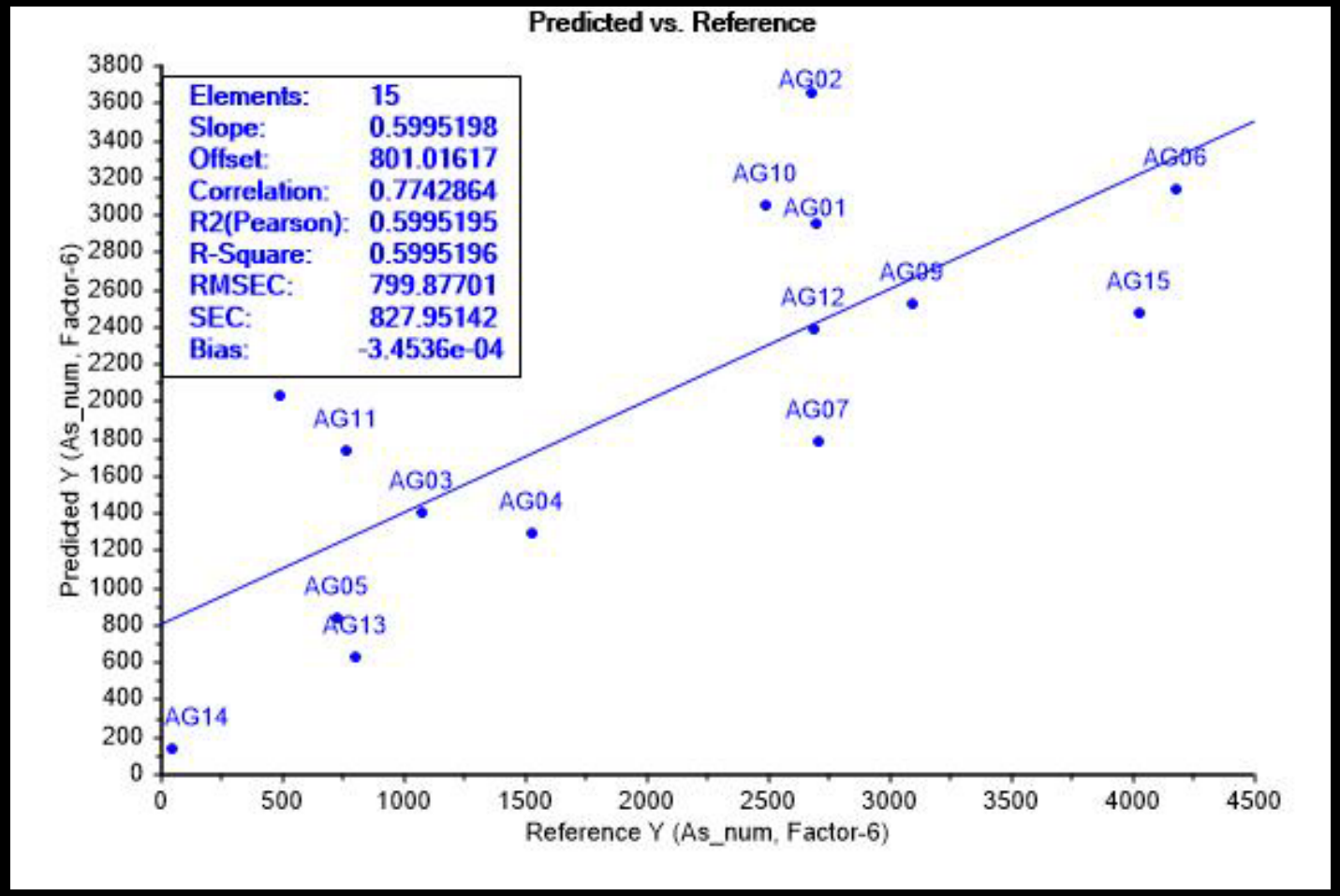
Ground sampling locations, shown by + overlayed on AVIRIS image

Collect spectra in AVIRIS image for each ground sampling location



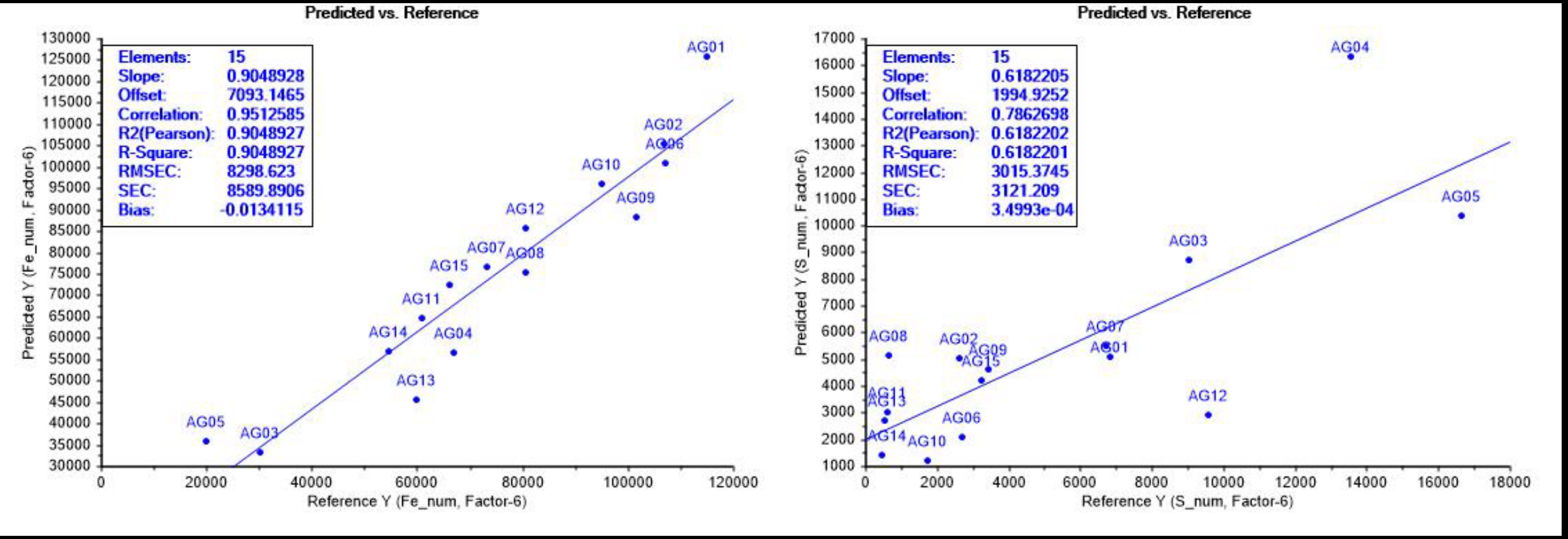
PLSR results

analysis was done with very small subset of data. A larger sample size should improve model



Results for XRF measured Arsenic. Preliminary fit is $r^2 = 0.59$

Iron and Sulfer



The model results will be used to predict metal concentrations from unknown samples (areas without XRF measurements) to locate contaminated soils on and off site of the mine

Commercial UAV, airborne and satellite systems with hyperspectral capabilities

- Airborne & UAS (drone) services – many companies provide regional services.
  - Examples: NV5 (https://www.nv5.com/geospatial/), Precision Hawk (https://www.precisionhawk.com/)
- Satellite imagery providers will soon provide hyperspectral data at a low cost relative to airborne providers
  - Several companies will launch a constellation of imaging satellites that will provide data at 1m – 8m spatial resolution
  - Examples: Orbital Sidekick (https://www.orbitalsidekick.com/)
    - HyspecIQ (https://hyspeciq.com/)
    - Pixxel Space (https://www.pixxel.space/)

How to Access and Order

For Federal Civil agencies and organizations under contract:

- The National Reconnaissance office (NRO) has a new multi-vendor contract called the Electro-Optical Commercial Layer (EOCL)
- Vendors including HyspecIQ and others will be available under the EOCL
- The NRO funds the EOCL and make the data available to fedciv at no cost
- Until the EOCL has these vendors under contact, GSA contracts are available to purchase data
- For EPA, contract OMS Geospatial Support team

Questions and Contact Information

David J Williams

ORD

(919) 889-0632

Williams.davidj@epa.gov

Steve Lennartz

National Geospatial Support Manager

EPA/OMS/OITO/EHD/ADHPCB

RTP MC N211F

(984) 710-4013

lennartz.steve@epa.gov