Interactive Gaussian Graphical Models for Discovering Depth Trends in ChemCam Data

Diane Oyen, Caner Komurlu, Nina Lanza

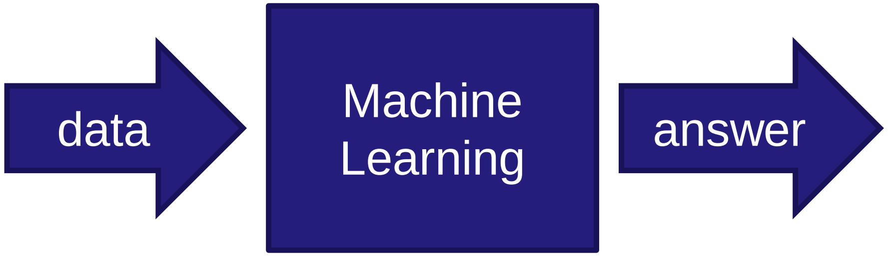
Planetary Science Informatics and Data Analytics

4/26/18

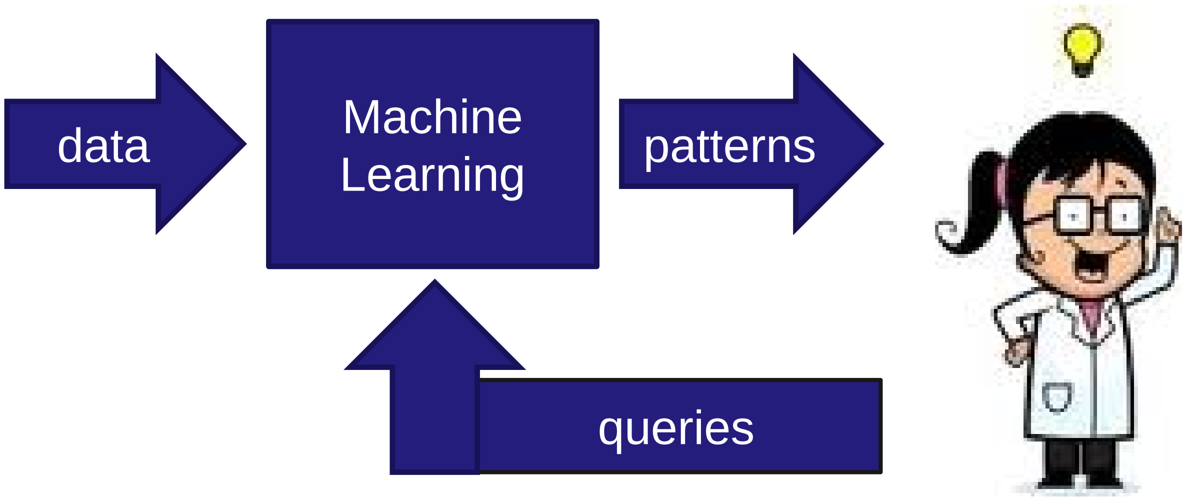
Operated by Los Alamos National Security, LLC for the U.S. Department of Energy's NNSA

Machine Learning for Scientific Discovery

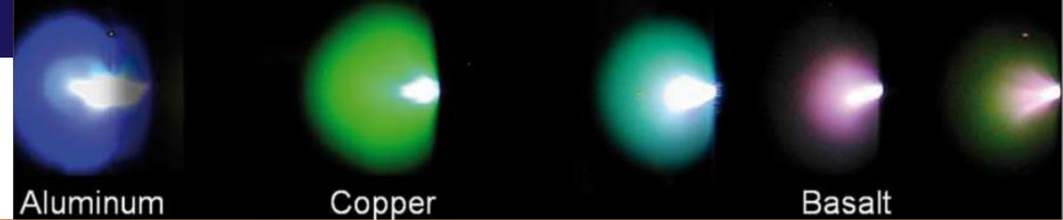
- **Scientific questions may not be answerable through automated data processing**



- **Machine learning should accelerate the speed with which scientists explore answers to their questions**

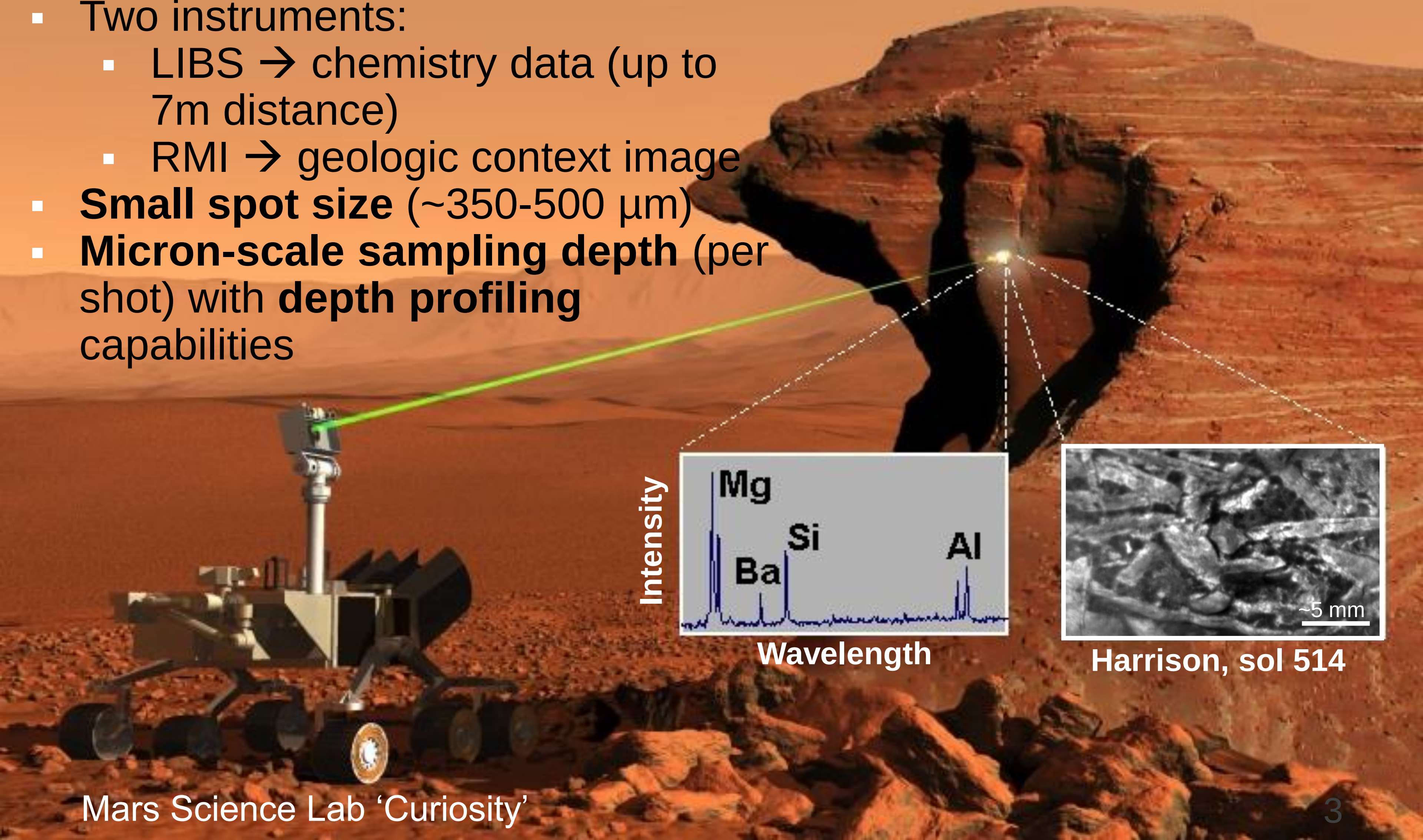


ChemCam: Chemistry and Camera



Above: LIBS plasmas (from Sirven et al., 2007)

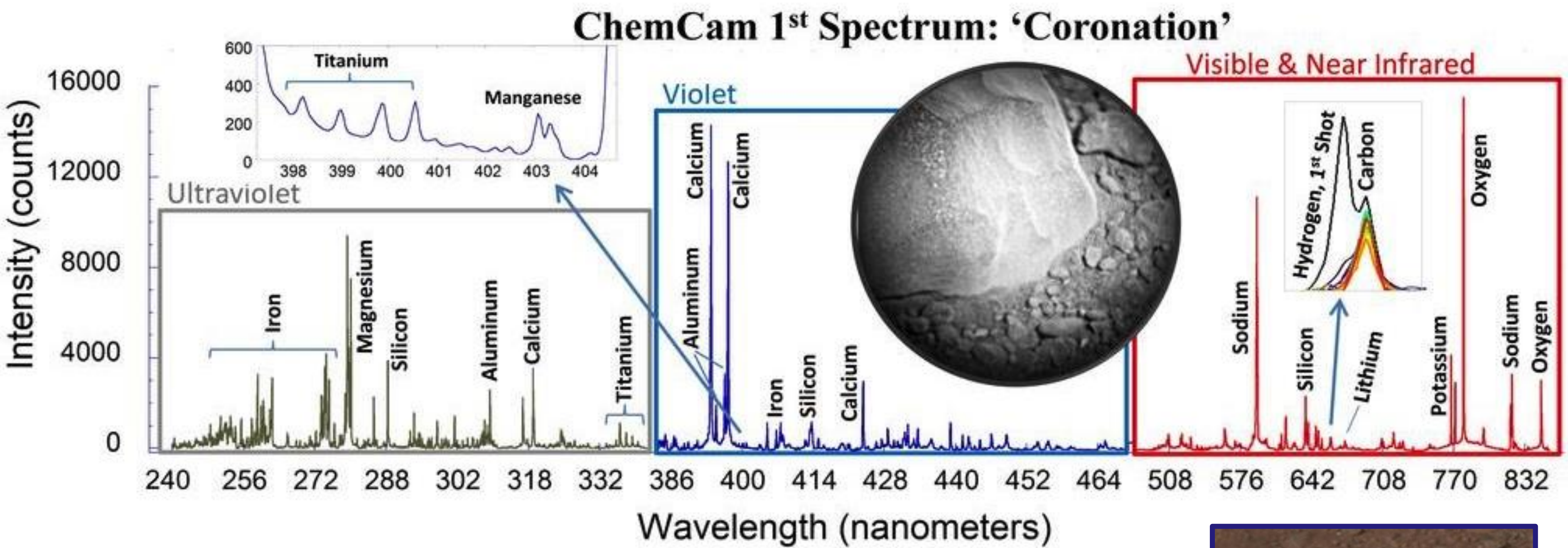
- Two instruments:
  - LIBS → chemistry data (up to 7m distance)
  - RMI → geologic context image
- **Small spot size** (~350-500 µm)
- **Micron-scale sampling depth** (per shot) with **depth profiling** capabilities



Harrison, sol 514

Mars Science Lab 'Curiosity'

Example LIBS spectrum
First ChemCam target, Coronation (sol 14)



- This is data averaged over 50 shots in one location
- Quantification of element abundance is difficult

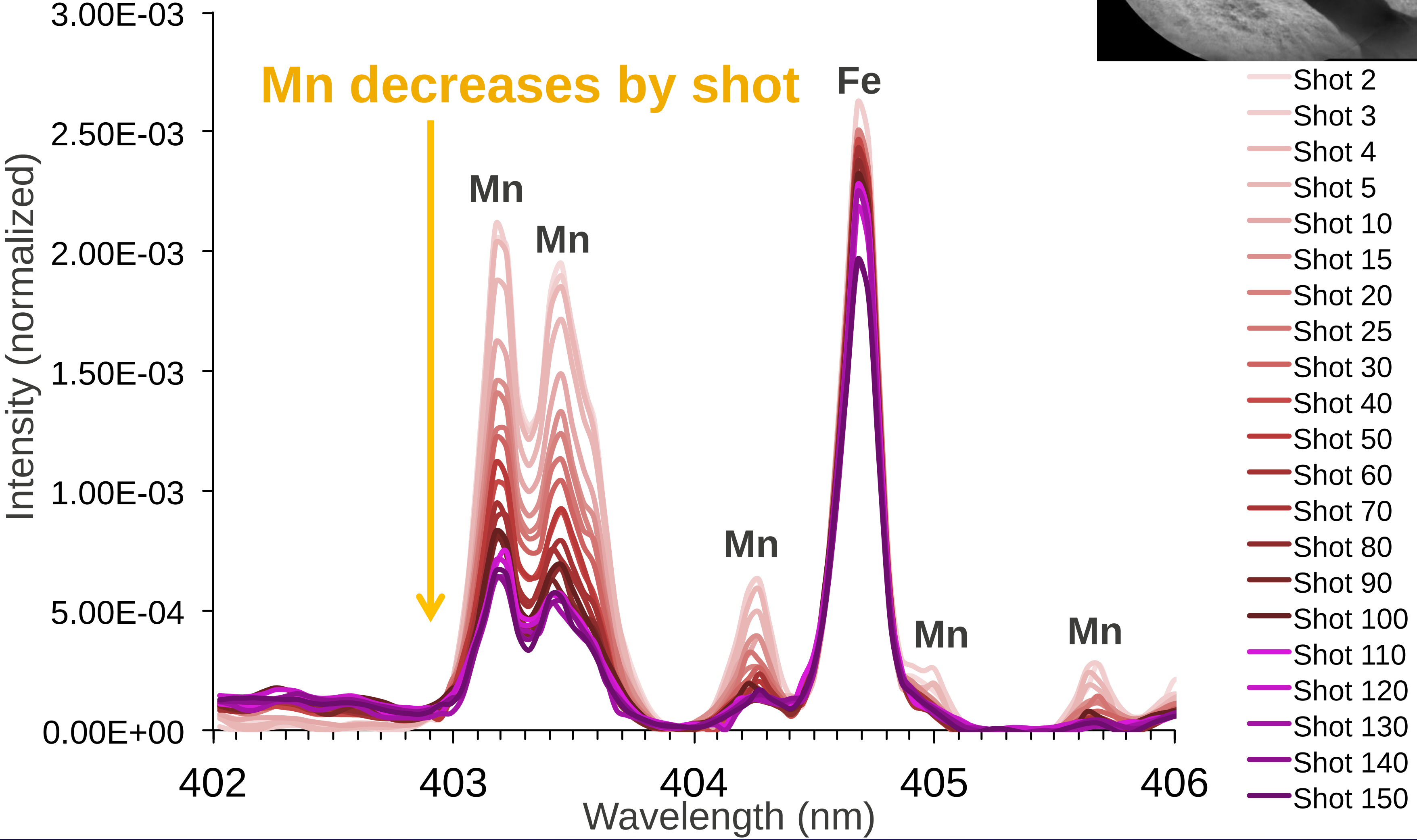
Machine Learning for Discovering Patterns in ChemCam Spectral Data

- **ChemCam spectroscopy**
  - Typically 30 – 150 shots in one location
  - Several shots in same location give depth profile
  - Multiple locations (1-15) per target
  - Relative elemental abundance is more precise than the accuracy of absolute abundance

- **Surface coatings, veins, stratigraphic layers**
  - Relationships among shots give clues about dynamics of chemical compositions
  - Different approach than trying to quantify elements first, followed by looking for dynamics

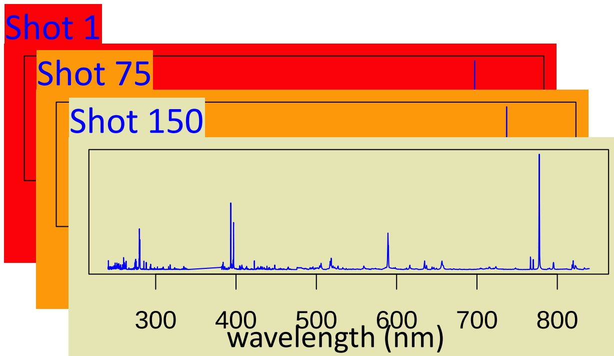
Trends with depth on Mars

Stephen sol 630 DP, location 4



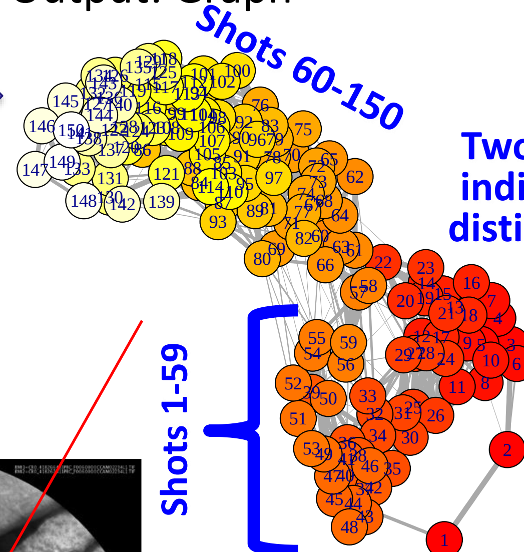
Depth Trends in ChemCam Spectra Data

Input: Chemistry spectra

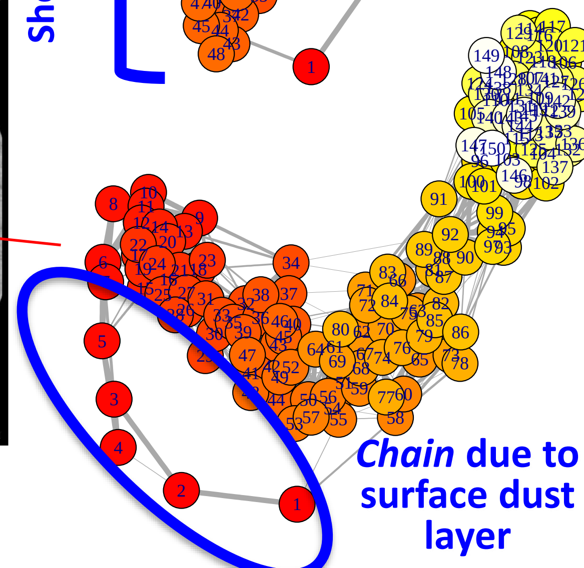


Interact: Which elements (wavelengths) affect graph structure?

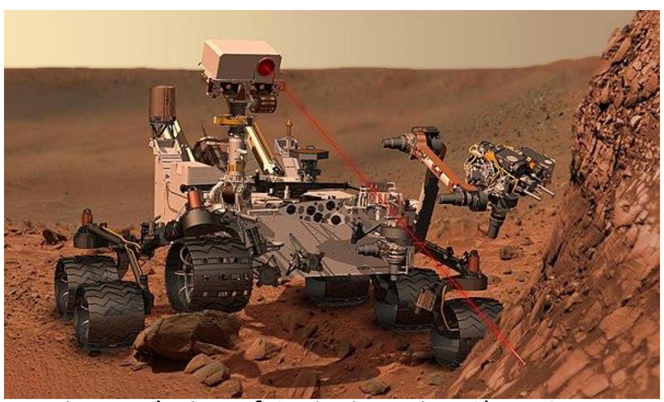
Output: Graph



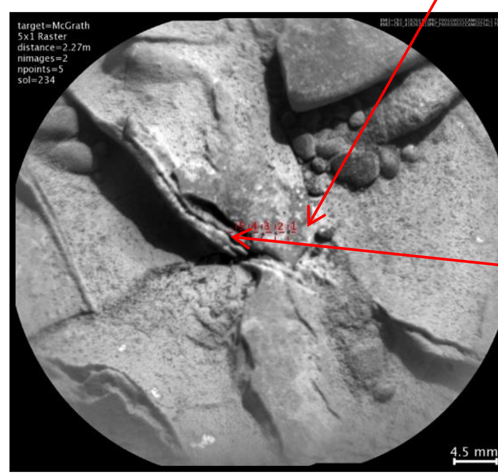
Two *clusters* indicate two distinct layers



***Chain* due to surface dust layer**



Artist rendering of Curiosity using ChemCam on Mars [NASA/JPL]

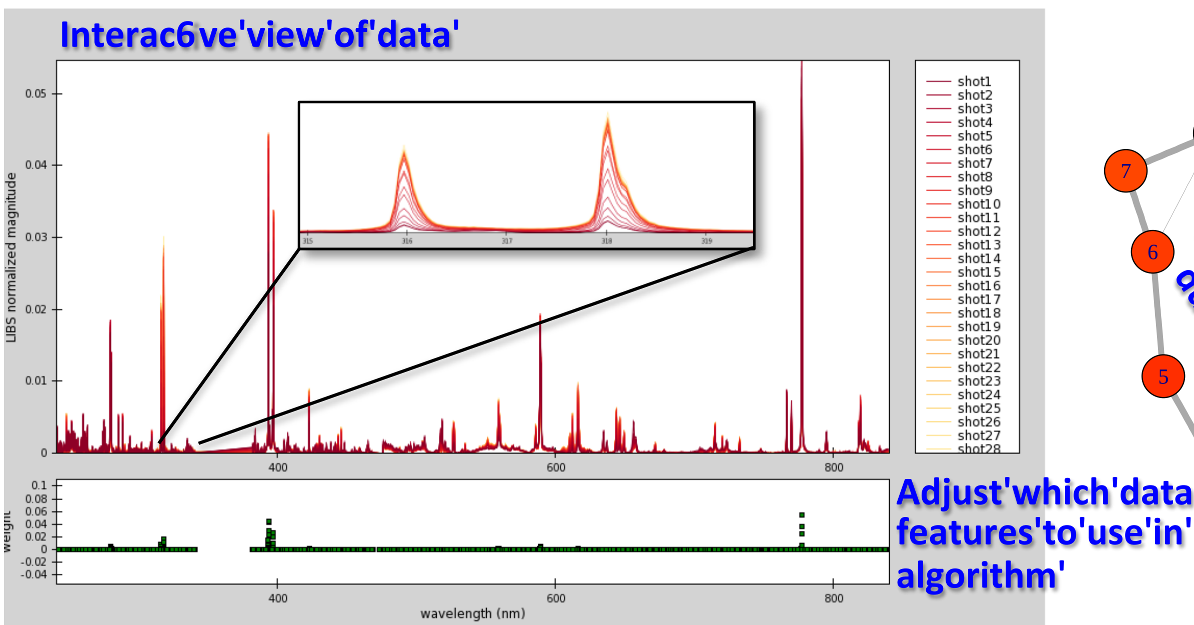


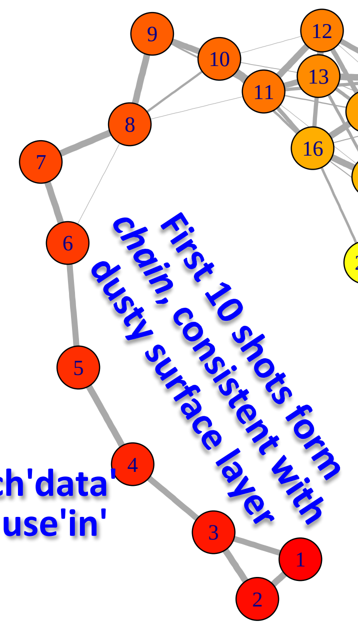
McGrath, sol 234

Interactive Learning of Probabilistic Graphical Models from ChemCam Spectra

- Probabilistic graphical models show relationships among variables
- Quickly identify surface features
  - Dust, surface coatings, veins
- Explore chemistry through interaction

- Nodes colored by shot number
- Edge weights are averaged over models of varying sparsity
- Drawn using a spring layout



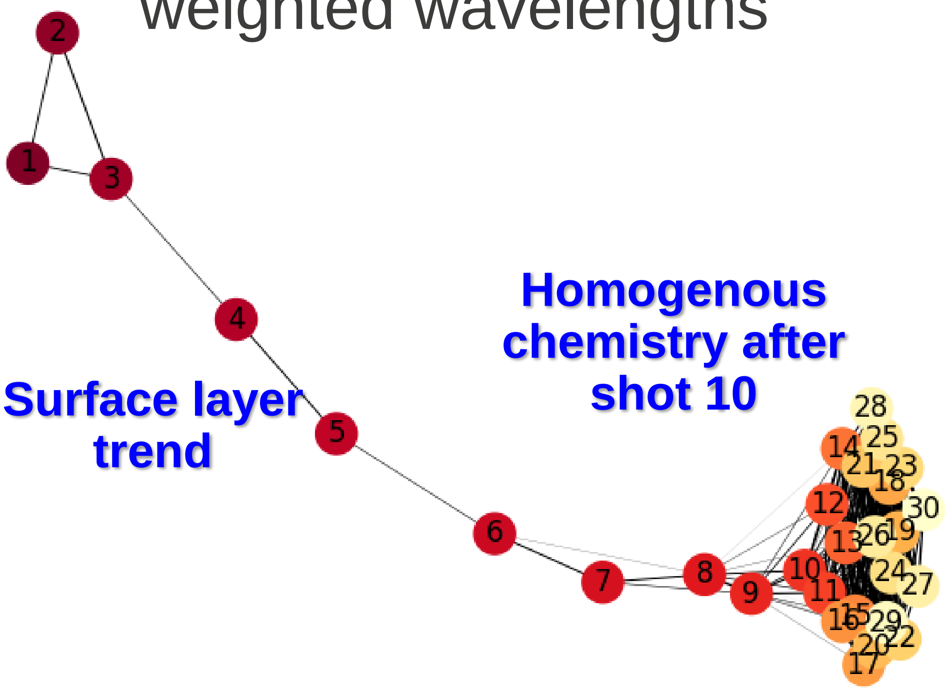


iGGM: Interactive Gaussian Graphical Model

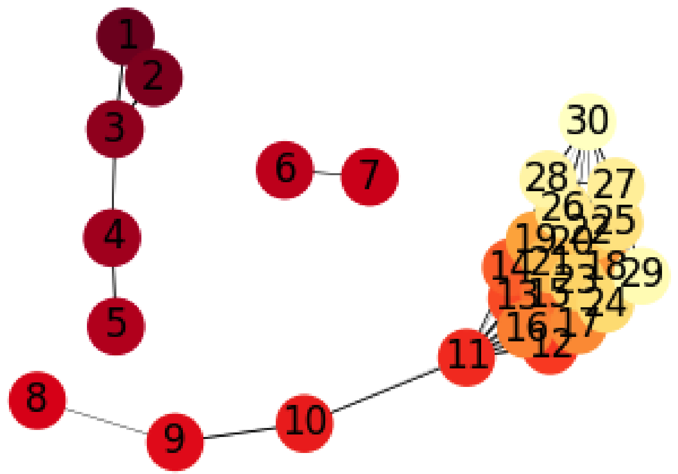
- **Explain chemistry behind the observed trends in the GGM**
  - Which wavelengths are responsible for the GGM structure?
- **Human-initiated interaction**
  - Mask or weight wavelengths and re-learn graph structure from weighted samples
  - Examples: Mask out major elements or down-weight “bright” elements
- **Machine-initiated interaction**
  - Interactive machine learning algorithm selects a subset of wavelengths that if masked would most change the GGM structure
  - Calculate gradient of covariance matrix with respect to sample weights

Human-initiated interaction for Bell Island

- Graph with equally-weighted wavelengths



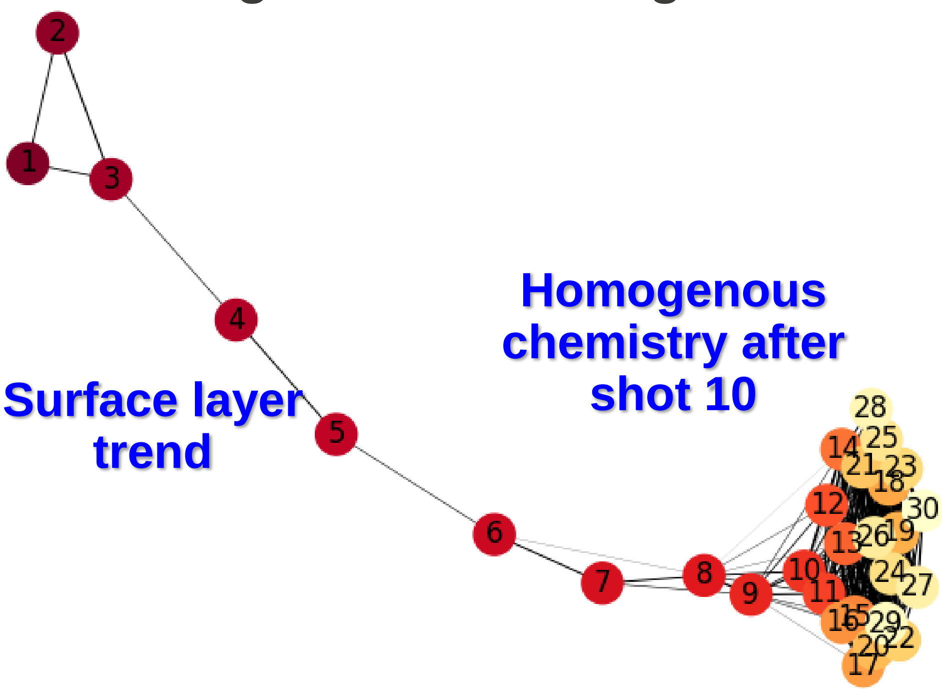
- Graph with only wavelengths of trace elements (lithium, chromium, manganese, rubidium, strontium, and barium)



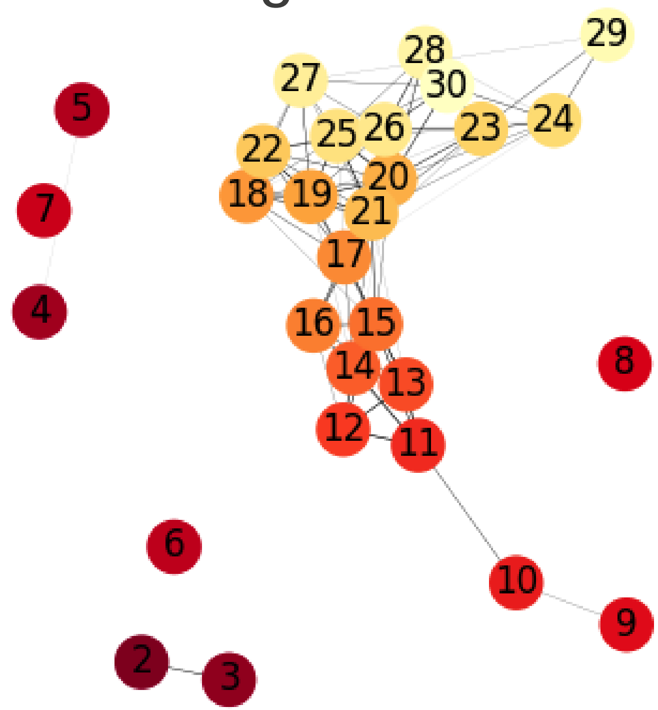
Masking out major elements breaks some shot-to-shot correlations in surface layer

Machine-initiated interaction for Bell Island

- Graph with equally-weighted wavelengths



- Graph with auto-selected wavelengths masked

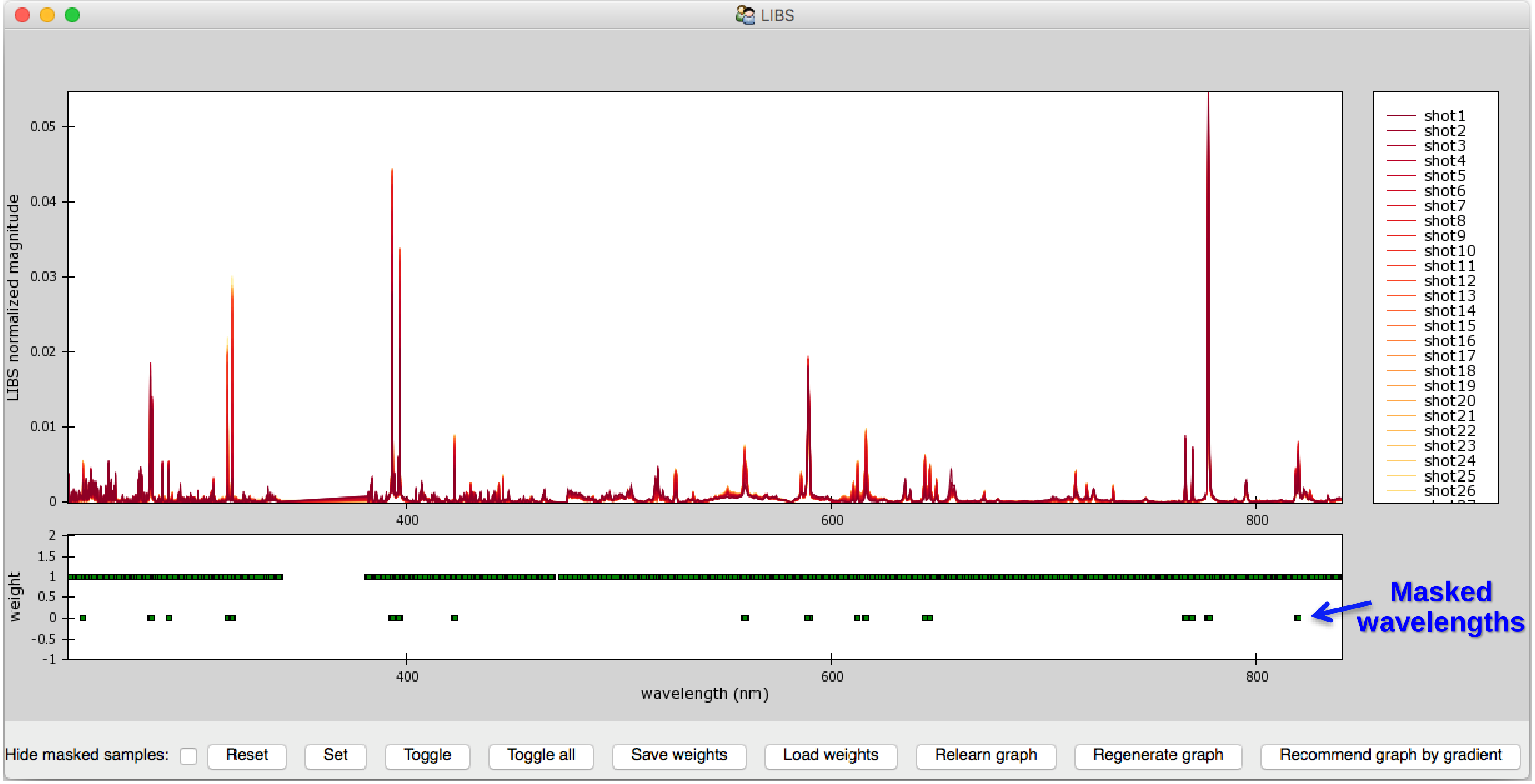


We now know which wavelengths create the prominent surface trend.

Also see that interior shots may have a layer change at shot 17.

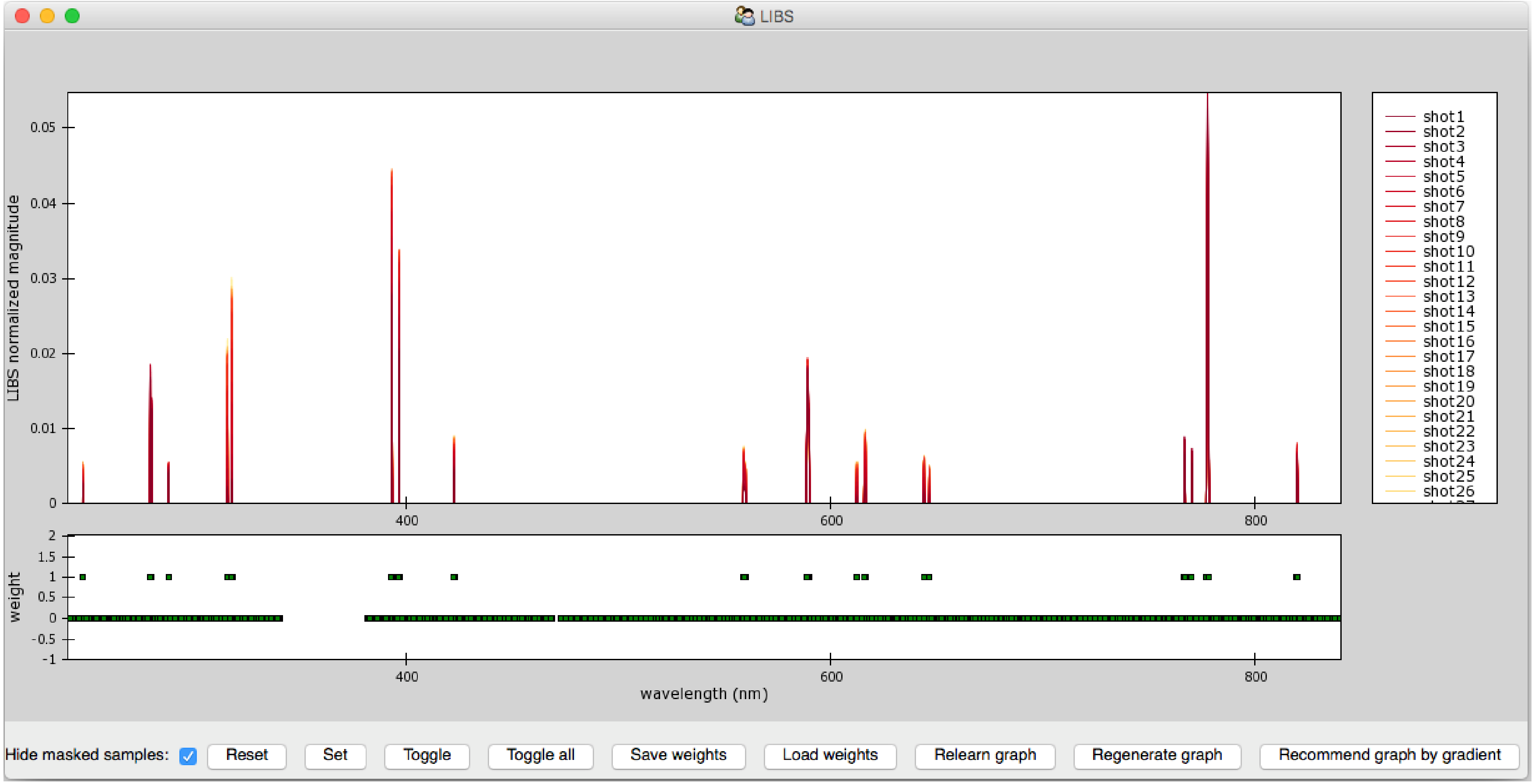
Bell Island, sol 113, 30 shots in one location

- **95 wavelengths auto-selected by iGGM that most change the structure of the GGM**



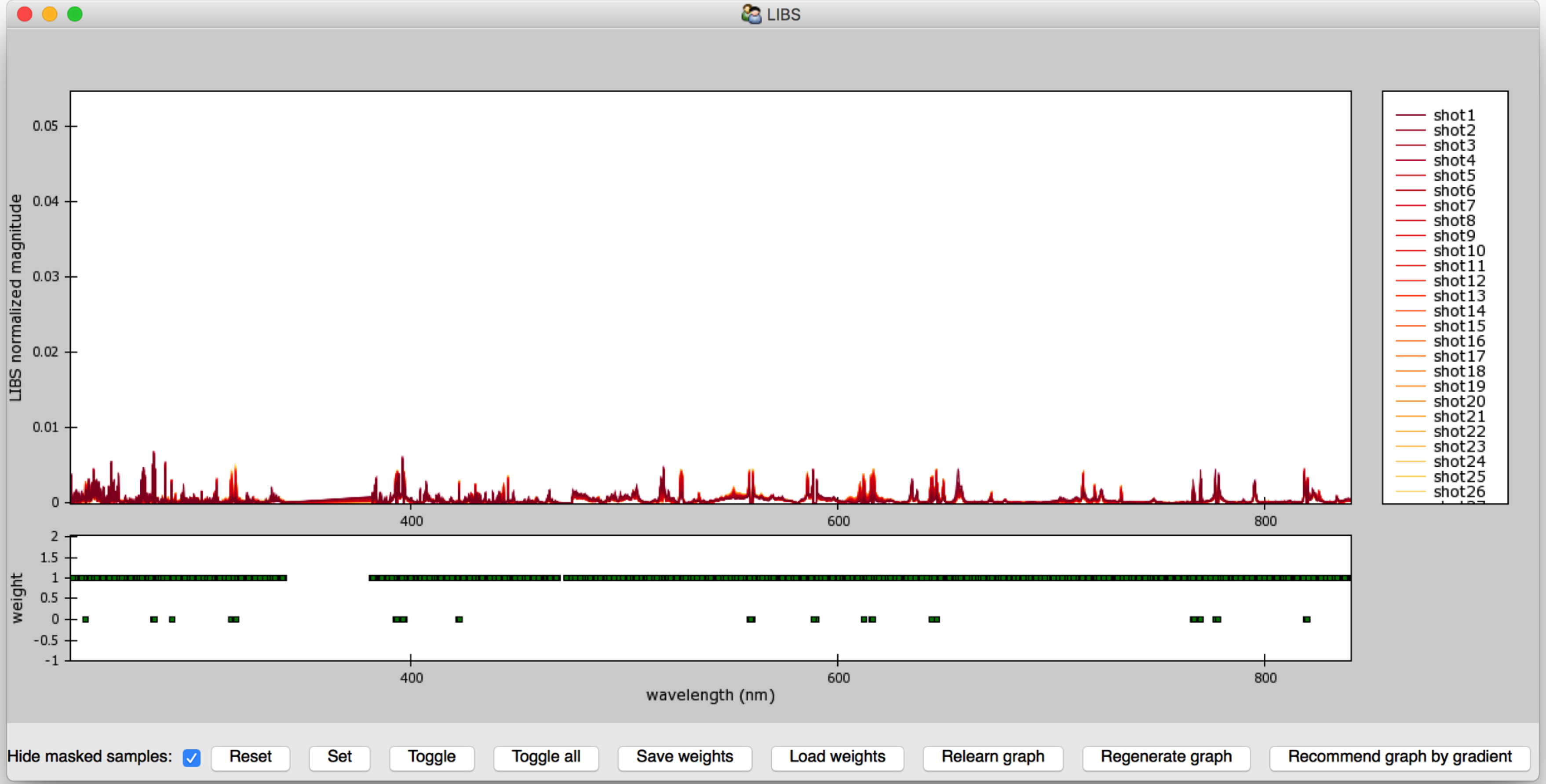
Bell Island, sol 113, 30 shots in one location

- **Spectra of only the 95 masked-out wavelengths auto-selected by iGGM that most change the structure of the GGM**



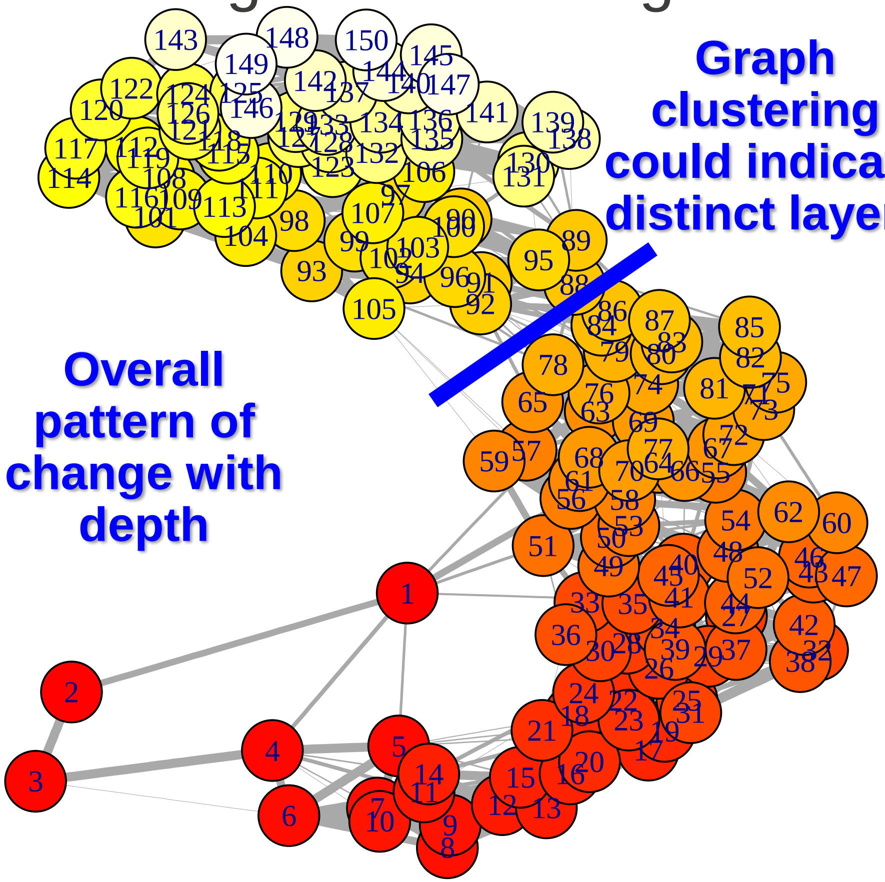
Bell Island, sol 113, 30 shots in one location

- **Spectra of remaining wavelengths not auto-selected by iGGM**



Stephen, sol 619, 150 shots in one location

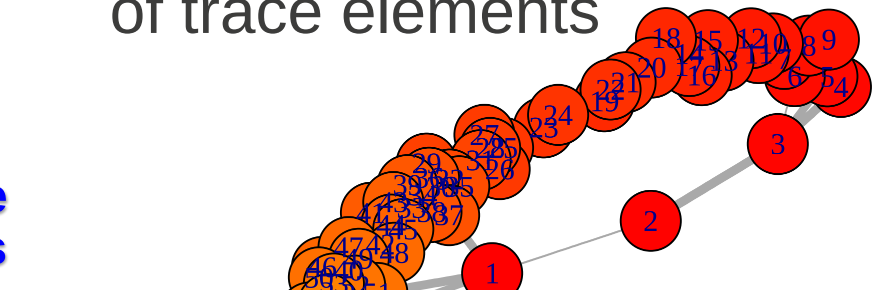
- Graph with equally-weighted wavelengths



Graph clustering could indicate distinct layers

Overall pattern of change with depth

- Graph with only wavelengths of trace elements



Systematic and rapid change in trace elements consistent through most of 150 shots on Stephen

Image credit: MSSS/JPL/NA

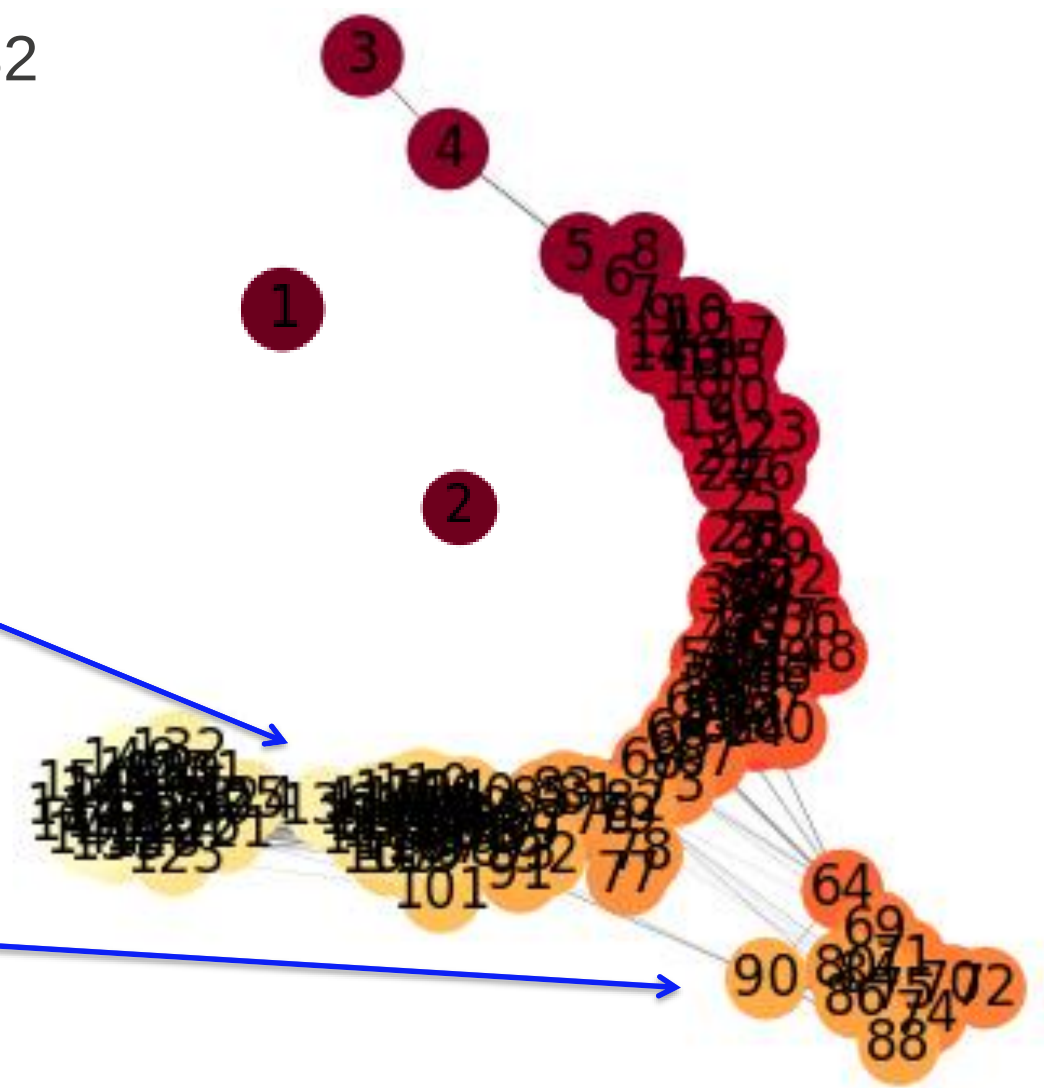
Machine-initiated interaction for Stephen

- iGGM results after 282 wavelengths masked

Masked wavelengths responsible for bottleneck structure at near shot 60.

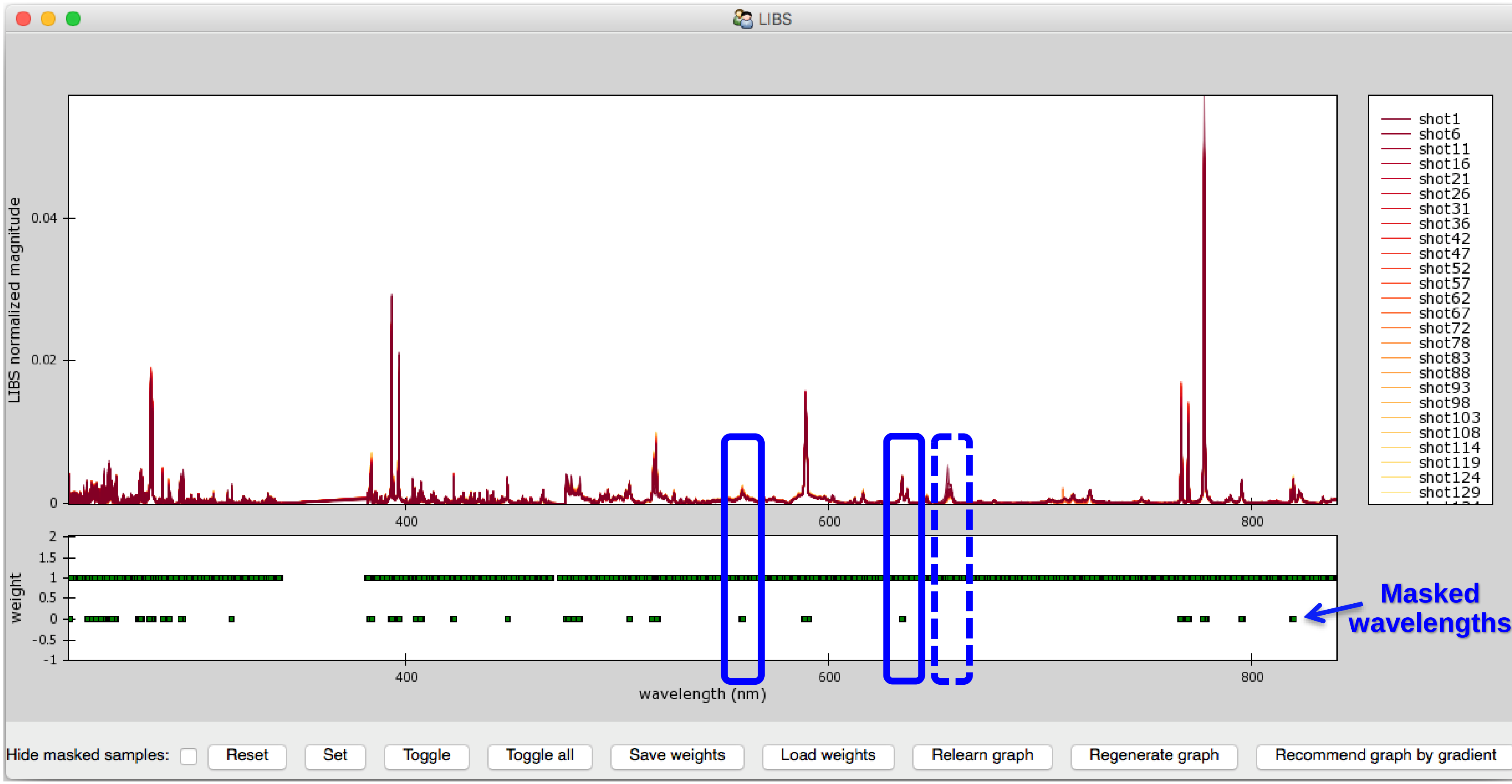
We now see another bottleneck near shot 130.

And an interesting feature around shots 70-90.



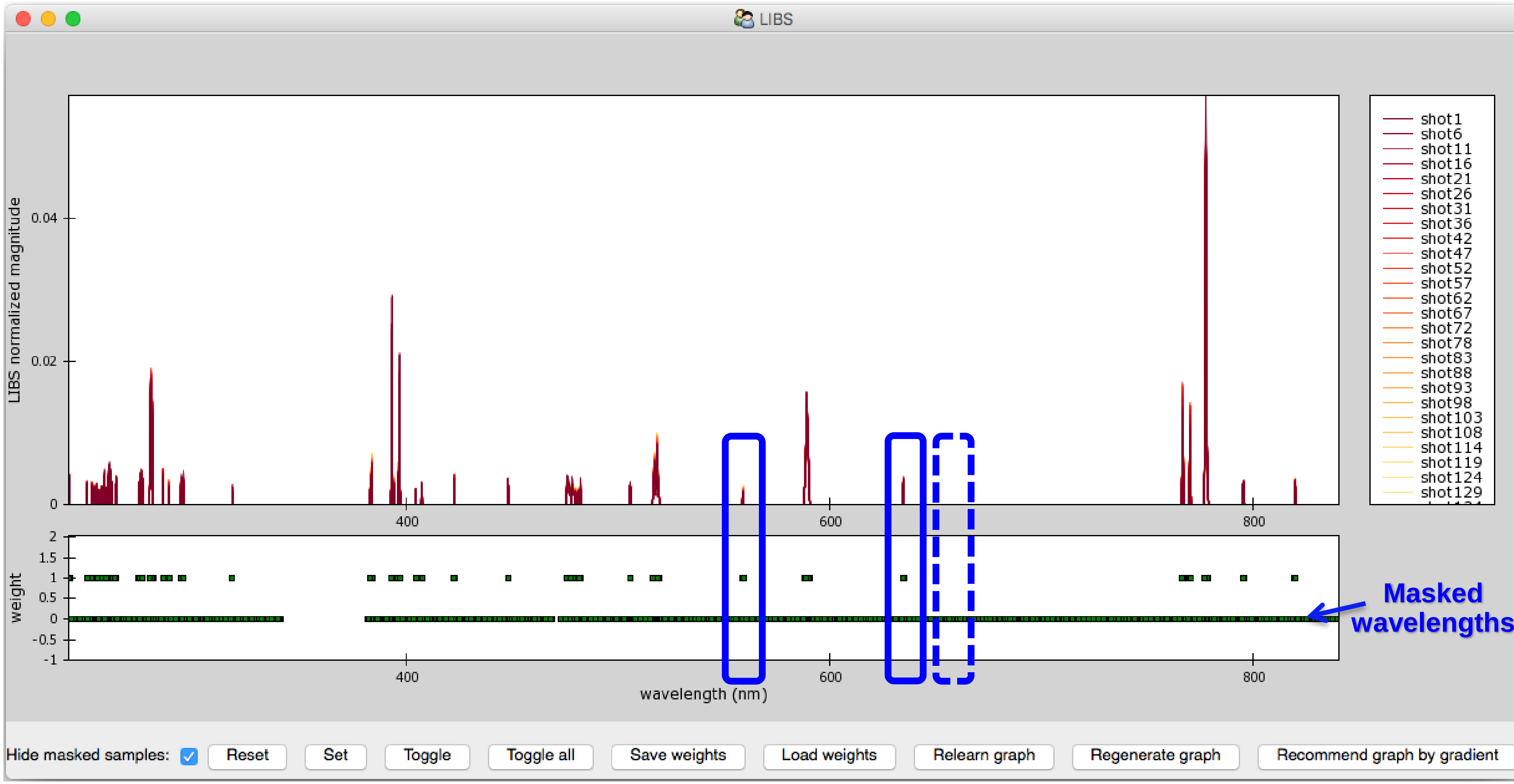
Stephen, sol 619, 150 shots in one location

- **282 wavelengths auto-selected by iGGM that most change the structure of the GGM**



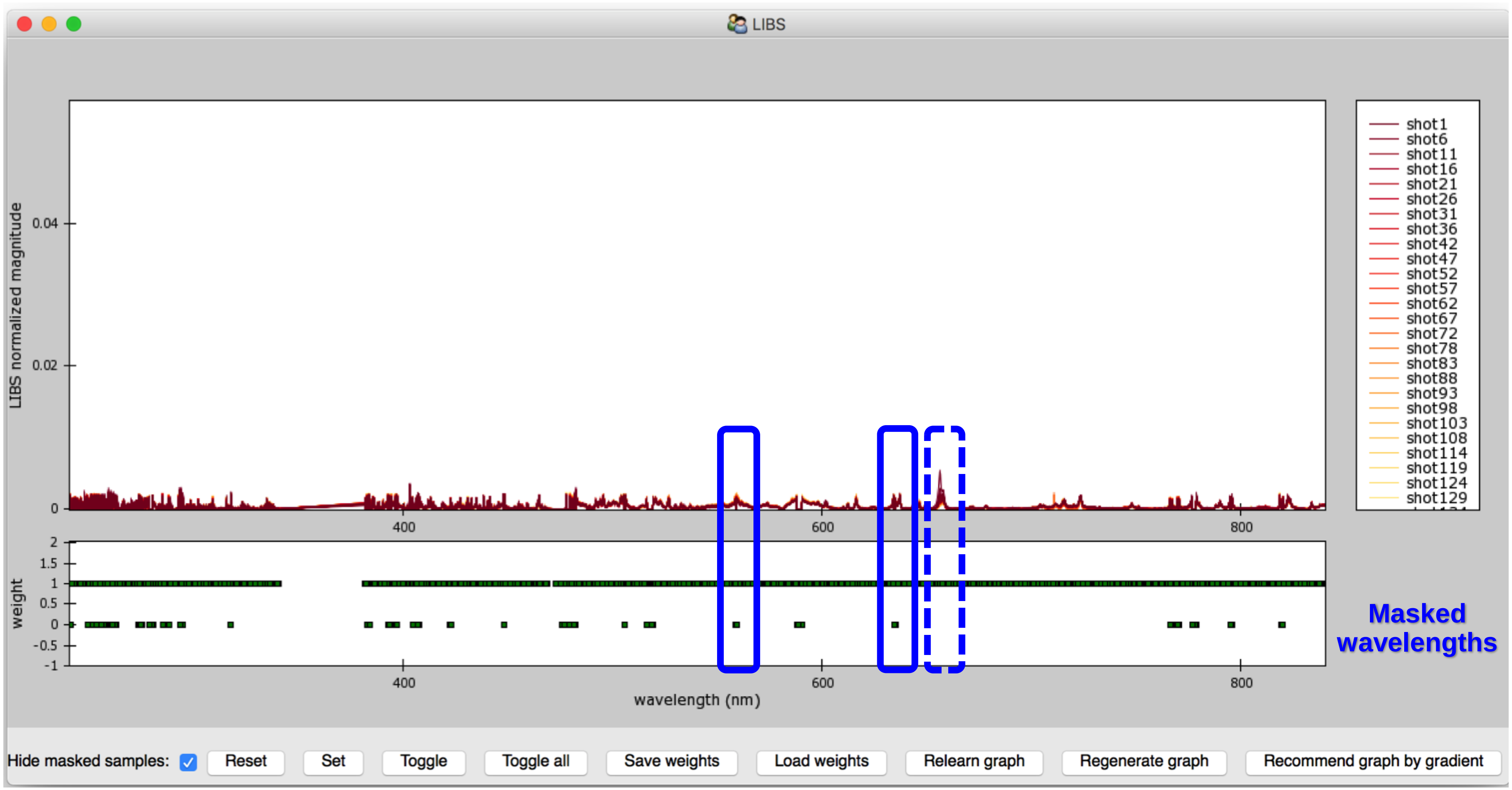
Stephen, sol 619, 150 shots in one location

- **Spectra of only the 282 masked-out wavelengths auto-selected by iGGM that most change the structure of the GGM**



Stephen, sol 619, 150 shots in one location

- **Spectra of remaining wavelengths not auto-selected by iGGM**



Graphs summarize depth trends

- **Shapes of learned Gaussian graphical models reflect compositional trends**
  - Graphs learned from full spectrum vs hand analysis performed on specific elements
  - Quick visual summary of possible chemical trends in targets
- **Interactive machine learning goes beyond giving “only answer”**
  - Typical machine learning gives an *optimal* solution
  - Scientific discovery requires investigating all likely solutions
  - Interactive machine learning allows scientist to vary inputs and/or learning parameters/constraints
- **Machine-initiated results project patterns back into data space**

Questions?

- **Coming soon**

  LibsInteractive Python package on GitHub

  email doyen@lanl.gov if interested