AUTOMATING ASTEROID SURFACE COMPOSITION IDENTIFICATION FROM REFLECTANCE SPECTRA. K. A. Farry[1], J. S. Graham[2], F. Vilas[3], and K. S. Jarvis[4]. [1]National Research Council (NRC) Research Associate, NASA/JSC/SN3, Houston, TX 77058, farry@farry.com. [2]9311 Tree Branch, Houston, TX 77064. [3]NASA/JSC/SN3, Houston, TX 77058. [4]LMSMSS, Houston, TX 77058.

Introduction. We are applying genetic programming, an evolutionary programming technique,[1] to identifying the minerals in spectra of asteroids from telescopes. We have done a basic feasibility test of this new identifier concept using US Geological Survey (USGS) spectra[2] of three terrestrial minerals likely to be present in low-albedo asteroid regoliths: Antigorite, Hematite, and Jarosite. Initial results are very promising. Functions produced by genetic programming correctly identify 96% of 140 spectra corrupted by measurement noise, scale uncertainty, and background continua removal uncertainty.

Background on Genetic Programming. The genetic programming process[1] mimics the "survival of the fittest" evolution of species in producing a solution. It "evolves" or optimizes computer programs to meet a fitness criteria. For a data reduction and identification problem such as asteroid composition, we define this fitness criteria in terms of maximizing correct identification of various minerals or maximizing a norm of (or distance between) spectral features of different minerals.

A scenario using genetic programming to evolve a mineral identifier begins with a pool of data items and functions that might be useful in relating the data items to each other. A set or population of candidate mineral identification programs is randomly created from this pool. The genetic programming system evaluates the fitness of each of these identification programs by testing them on known mineral spectral signatures and computing their identification success. It then uses the best to create the next generation of programs, by a combination of "mutation" (randomly changing portions of the programs), "crossover" (joining portions of two programs to make a new program), and "reproduction" (simply copying a program from the old generation into the new). The fitness evaluation and creation process repeats for many generations; the best program across all generations becomes our mineral identifier.

Genetic programming is especially well-suited for identifying minerals from reflectance spectra because it can:

- develop identifiers for problems where the relationships between data and the phenomena generating the data are complex, subtle, or poorly understood.
- assist the scientist in determining which data items reveal the most about the underlying phenomena.
- overcome many errors in the posing of the identification problem.
- incorporate additional "known cases" into the identifier from the specimen database automatically.
- incorporate the outputs of many different instruments into the identification process.
- produce identification programs that can run on small embedded processors in real time in field instruments or spacecraft.
- modify identification programs rapidly to accommodate instrument failures, possibly salvaging a scientific mission.

Unlike (for example) neural networks, where the structure of the solution is set and only weights on portions of the structure can adapt, genetic programming allows the structure of the solution to adapt. Thus, it has proven more effective in identifying patterns in certain biological signals (achieving 100% accuracy in some cases)[3] than neural network solutions (achieving only 93% accuracy on the same data).[4] Like neural networks, however, genetic programming requires many spectra of known minerals to guide the evolution of the identifier (equivalent to "training" in neural networks).

Approach. First, we create a database of known spectra to guide evolution. Since measuring good spectra is tedious and expensive, we use all spectra of the minerals of interest in the USGS digital spectral library.[2] We make 10 copies of each and then add simulated measurement noise. We also randomly vary the average amplitude and the overall slope of each spectrum to simulate uncertainties in scaling and removing background continua typical in telescope spectrograph data reduction.[5] The result is a large group of noisy spectra.

We can apply genetic programming directly to the raw spectral measurements; however, the evolution computational resources (time and memory) are a function of the input data items. Consequently, we started with a strategy of reducing the measured spectrum (consisting of magnitudes measured at 470 wavelengths from 0.4 to 3.0 μm) into a few features that capture each spectrum's uniqueness. Our first feature set consists of:

- mean frequency (MF) (centroid of area under spectrum after normalizing its area to one);
- spectral line length (SLL);
- number of slope sign changes (SC) (a total number of maxima, minima, and inflection points);
- number of minima (NMin);
- number of maxima (NMax);
- frequency of highest peak (FHP); and
- frequency of lowest absorption minima (FLA).

The background continua uncertainty affects mean frequency the most, and highest peak and lowest absorption minima frequency to varying degrees. Noise affects the other features more than background continua removal uncertainty.

We filter (using a 5-point boxcar averaging filter) and then compute these features for all spectra. We use 50% of the database to guide the evolution of an identifier. We reserve the remaining 50% for testing the identifier that performed the best on the evolution data set. This gives us an independent confirmation of the identifier's performance.

Results. The first trial was based on 7 USGS Antigorite, 12 Hematite (across many different grain

sizes), and 9 Jarosite spectra. After duplicating each spectrum 10 times (for a total of 280 spectra), we added white noise with standard deviation of 5 times the measurement uncertainty and randomly varied the overall slopes and scale by up to 5%. In the genetic programming process, we defined fitness as simply the percentage of correct identifications of the minerals producing the spectra. Using 140 (half) of the corrupted spectra, we evolved three identifying functions, one for each mineral of interest:

(1) Antigorite (from 60th generation)

```
if (FHP + FHP) < [if (FHP^NMin) < (1- NMin)
                    then return 1
                    else return NMax]
  then return (FHP/(NMax^MF))
  else return sign(NMax)
```

(2) Hematite (from 32nd generation)

```
if (1/sqrt(FHP)) ≥ (NMax + MF + MF)
  then
    if (FHP - SLL) < SLL
      then return 3
      else return sign(MF)
  else return 1
```

(3) Jarosite (from12th generation)

```
if cos(2) ≥ mod(FHP)
  then return
    exp[if FLA < MF
      then return MF
      else return 2]
  else
    if sign(SC) < (MF * SLL)
      then
        if FLA < MF
          then return MF
          else return 2
      else return 3
```

where mod returns the fractional portion of its argument and sign returns -1 for negative numbers and 1 for zero or positive numbers. These functions output the mineral's identifying number (within a machine epsilon) when they encounter a spectrum from that mineral. Their percent correct identifications on the database spectra were:

Mineral (Identifying Number)	**Evolution Database Performance** (140 spectra total)	**Testing Database Performance** (140 spectra total)
Antigorite (0)	98.57%	95.71%
Hematite (1)	100%	97.86%
Jarosite (2)	96.43%	96.43%
Average	98.33%	96.67%

We reemphasize that none of the 140 testing database spectra was used in the evolution process. Note also that the performance percentage reported here is for each identifier used alone. Using all three together can improve the results as their errors generally do not overlap.

These identifying functions are neither particularly intuitive nor parallel in structure. Nor do they use all the available data. The genetic programming process uses whatever is most effective, without bias.

Future Work. This simple test of genetic programming for mineral identification is only a beginning, albeit a promising one. As we add more minerals to our identification problem and experiment with mixtures, we expect to need more complex features to adequately identify them. We are adding new features to our spectral order reduction, focusing first on:

- 0.43, 0.7, and 1.0 μm fine feature measures; [6]
- wavelength locations of all minima and maxima; and
- modified Gaussian deconvolution for broad feature capture. [7]

Asteroid reflectance spectra sometimes suffer from incomplete atmospheric extinction, so we are also introducing variations on this into our evolution and testing database. Furthermore, actual asteroid spectra have varying resolution, generally much less than that in the USGS and other spectral libraries based on laboratory measurements. We will convolve our evolution database to reduce laboratory spectral resolutions to match actual asteroid survey data. Finally, we will use the identifier evolved from our genetic programming process on actual asteroid reflectance data. We also plan to compare the genetic programming results with those from neural network identifiers.

Acknowledgements. The National Research Council supported Dr. Kristin Farry's work on this project.

[1] Koza, J., *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: The MIT Press, 1992.

[2] R. N. Clark *et al*, *The USGS Survey, Digital Spectral Library*, Version 1: 0.2 to 3.0 microns, US Geological Survey Open File Report 93-592.

[3] Fernandez, J.J., Farry, K.A., and Cheatham, J.B. "Myoelectric Signal Recognition Using Genetic Programming," Genetic Programming 1996 Conference, Stanford, CA, July 28-31, 1996.

[4] Farry, K.A., Walker, I.D., and Baranuik, R., "Myoelectric Teleoperation of a Complex Robotic Hand," *IEEE Transactions on Robotics and Automation*, Vol. 12, No. 5, October 1996, pp. 775-788.

[5] Vilas, F. *The Nature and Origin of Outer Solar System Asteroids from Reflectance Spectrophotometry*, Ph.D. Dissertation, University of Arizona, Tucson, 1984.

[6] Vilas, F. "A Cheaper, Faster, Better Way to Detect Water of Hydration on Solar System Bodies," *Icarus*, 111, 1994, pp. 456-467.

[7] Sunshine, J.M., Pieters, C.M., and Pratt, S.F., "Deconvolution of Mineral Absorption Bands: An Improved Approach," *J. Geophysical Research*, 95, B5, May 10, 1990, pp. 6955-6966.