Acta Astronautica 67 (2010) 1356–1360

Contents lists available at ScienceDirect

Acta Astronautica

journal homepage: www.elsevier.com/locate/actaastro

Detection of biomarkers using infra-red spectroscopy

S. Dumas *, Y. Dutil, G. Joncas

Dépt. de physique, de génie physique et d'optique et, Centre de Recherche en Astrophysique du Québec, Canada G1V 0A6

ARTICLE INFO

Article history:
Received 8 February 2009
Received in revised form
26 November 2009
Accepted 18 January 2010

Keywords:
Astrobiology
Biomarkers
Infrared
Spectroscopy

ABSTRACT

In the near future, more missions of exploration will be sent to our neighbour planets and moons to detect signs of life. The automatisation of such search is not an easy task and selecting the right samples for the analysis is even less easy. This paper proposes a technique using infra-red spectroscopy coupled with statistical and multivariate algorithms to preselect targets for detailed search for biomarkers.

1. Background

Future exploration missions to the planets and moons will bring more and more sophisticated equipment to detect life. They will be unmanned mission and would require complex artificial intelligence to cope with the surrounding. Obviously they would not have the time nor the resource to study a huge number of the samples. A way to filter the samples to process only those with a high probability of harbouring life would be a better way than searching in all the rock samples.

This paper reports on an experiment to detect biomarkers in rock using infra-red spectroscopy and the use of multivariant analysis to isolate the bio-signatures.

Several techniques [1] of detecting life forms in rock, in different environments, exist but they all destroy the samples. The proposed method will not destroy nor affect those life forms. Furthermore, it is not requiring complicated procedure to prepare the samples.

2. Biomarkers

Biomarkers [1,2] are molecules that are the by-product of the organic matter (via metabolisation) or their nutrient (i.e. carbonate). It may be difficult to detect directly the micro-organisms but not so difficult to detect their trace in the environment. The biomarkers are relatively simpler molecules than the micro-organisms.

Such biomarkers are, for example: calcium oxalate, alkanes, alkenes, some FeO-type of molecules and carbolic acid. Their IR spectra are well known.

3. Method and instrumentation

The use of infra-red spectroscopy is dictated by the fact that the organic matter shows absorption bands in this region of the light spectrum. The region of middle infra-red (2.5–15.37 μm, 4000–650 cm^{-1}) is especially important. The measurements of all rock samples were performed using a IR spectrometer from Nicolet (MAGNA series) and a diffuse reflectance kit. This device uses mirrors to redirect the IR beam on specific locations on the rock.

* Corresponding author.
E-mail addresses: stephane_dumas@sympatico.ca (S. Dumas), yvan.dutil@sympatico.ca (Y. Dutil).

0094-5765/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.actaastro.2010.01.018

4. The samples

The samples used during this study come from two distinct places in Canada. They were collected in relation to other bioastronomy projects and we obtained a few specimens.

The first group [3] of rocks comes from near the station of Eureka, Nunavut, Canada. The specimens are composed mainly of sandstone and quartz. Sixteen samples were produced from this group.

The second group [4] comes from near the town of Guelph, Ontario, Canada. They are composed mainly of dolomite, calcite and sphalerite. Twenty samples were produced from this group.

In each group, there are several rocks exhibiting visual traces of endolithes (red and green patches). Also further analyses [3,4] of both the groups report the presence of organic matter.

For each sample, several spectra were taken in different locations on the rock (directly on the endolithes, on regions beside and on region completely without endolithes).

5. Reference spectra

In the process of analysis of the Nunavut and Guelph spectra, we used a series of reference spectra for comparison. Those spectra were compared to the rock sample spectra through multivariant analysis.

The mineral spectra come from the JPL-ASTER [5] database and are: olivine, sandstone, calcite, hydroxide, sulphate, carbonate, sulphide, dolomite and quartz. Those are the main constituents of the rock (aside endolithes).

The 48 different organics references were taken from the book “Structure Determination of Organic Compounds” [6]. Here is a short list: anhydride acid, alcohols, aldehydes, alkanes, ammonium, aromatic hydrocarbons, furans, primary amides and carboxylic acids.

We obtained a few grams of JCS-1 [7] which is a Martian soil simulant. Its behaviour in the IR domain is very similar to the Martian soil. It was used to verify if the soil could mask the organics signatures we were looking for on Mars.

6. Correlation test

We performed a series of correlation tests on the spectra. This is a way to evaluate the correlation between two sets of data. Each spectrum contained in our database is a vector of intensity values for each wavenumber in a particular interval (i.e. middle infra-red).

The first series of tests evaluated the correlation between each pair of spectrum. A value between −1 and +1 indicates the relation between those two spectra. A positive value indicates that both spectra are similar while a negative value shows different spectra.

This process was done for each pair of spectra from Nunavut, Guelph and the group of references organics spectra. The matrix (Fig. 1) resulting from this operation indicates similarities amongst Nunavut spectra and amongst Guelph spectra. But not much similarity between both groups. The similarities amongst each group are explained by their mineral composition which has a stronger spectral signature than the organics.

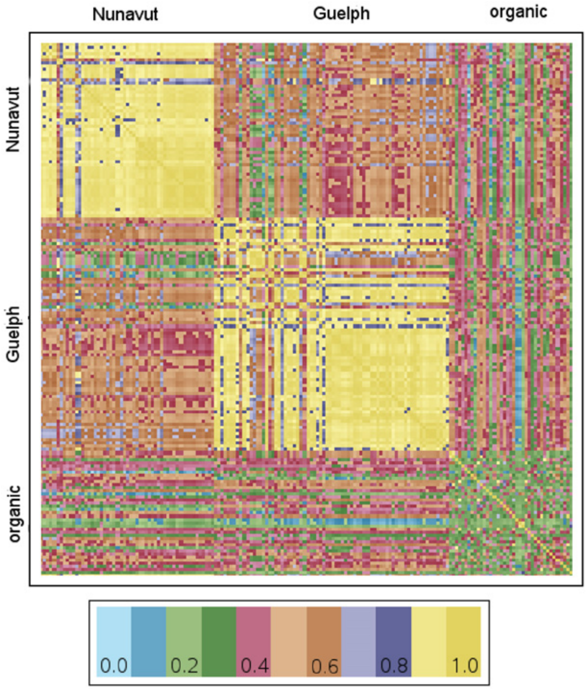

Fig. 1. Correlation matrix using spectra as vector data. The colour coding is from blue-green (negative) to yellow (positive) while red-brown indicate neutral. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The matrix also indicates very little connection with the organic spectra (represented by the green square at the bottom right).

The whole process was also performed using vectors of wavenumber instead of spectra. In the first scenario, the vectors being compared were the spectra themselves. Meaning that each element of the vectors was a wavenumber.

In the second scenario the correlation matrix took vectors containing the value of the same wavenumber across the spectra. Therefore, comparing the importance of each wavenumber in the whole database.

The resulting matrix (Fig. 2) indicates a region where the information is concentrated, where the correlation is lower. This region is the “fingerprint” region (600–1800 cm^{-1}) and is known to contain all the signatures of biological molecules.

Another way to see the results of the second matrix is related to the redundancy of information in the spectra. The matrix shows that only the fingerprint region may be used since most of the information is contained in that region.

The correlation test is not enough to identify the organics in the samples. Fig. 2 shows interesting features that will be used in the next step of the analysis.

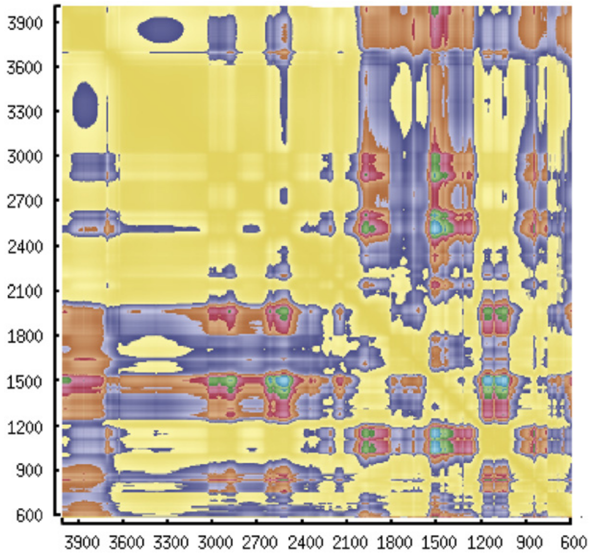

Fig. 2. Correlation matrix using wavenumber as vector data.

7. Principal component analysis

The principal component analysis [8,9,10] (PCA) is a vector space transform often used to reduce multidimensional data sets to lower dimension for analysis. It is also known as Karjunen–Loève transform (KLT), Hotelling transform or the proper orthogonal decomposition.

The PCA will transform the spectra from the original wavenumber space to the eigenspace. It will also provide information about the minimum number of dimensions required to represent the data in the new space. This can be seen as a compression technique but it will not be used in that way here. Nevertheless, that information will be useful to reject the unnecessary dimensions and reduce the size of the data set. While each element of the original spectrum is the amplitude for a given wavenumber, each element in the eigenvector is an amplitude for per principal component (or PC). The PC vectors have the same length as the wavenumber spectra.

The PCA can be summarised by Eq. (1), where each part of the equation is matrices. The matrix A (of dimension $m\times n$) contains the original data. In our case, it contains m IR spectra of n wavenumber each. U contains the m new vectors expressed in the eigenspace. W is a diagonal matrix with the eigenvalues and V is a matrix with the eigenvector. Both W and V are $n\times n$ matrices.

$$A=UWV^{\mathrm{T}} \tag{1}$$

Several numerical algorithms [8] exist to compute this result and we will not explain them here. The algorithm we used (from the mathematical library Numerical Recipes [11]) computes and returns U, W and V given the input matrix A.

Fig. 3 illustrates the process for 2D vectors. Each vector S is defined by a pair of (X,Y). Those vectors are then assembled into a matrix A. The resultant vectors S' are defined in the eigenspace (X′–Y′). S' are the rows of U.

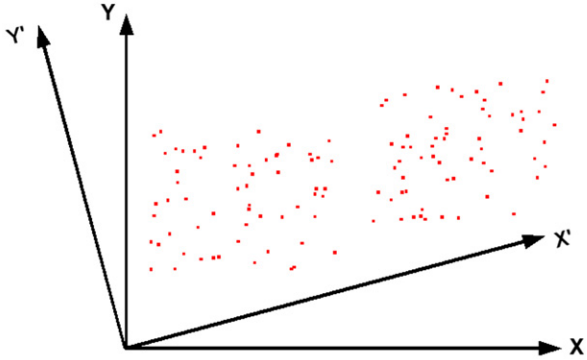

Fig. 3. Example of PCA in 2D.

It can also be seen as a rotation from the X–Y to the X′–Y′ system. This rotation being defined by WV^{T}. So, once the W and V matrices have been computed, they can be used as a single rotation matrix to bring unknown spectrum into the eigenspace and identify them using reference (i.e. known) spectra.

It is useful to represent the spectra by their corresponding point in a N-dimensions PCA space. Therefore, each spectrum can be seen as a single point, in the PCA space, composed of N components.

8. Application of PCA

In order to use the PCA as described in the previous section, the values of each spectrum must be slightly processed by removing the mean value of each wavenumber. The method works best when using only variations around the mean.

The resulting vectors are then put in a matrix called A. It is then processed using a PCA algorithm [11] from a software library. Three matrices (U, W and V) are produced as output.

Using the values of W, it is possible to determine how many principal components (PC) to use to describe each vector of U. In our experiment, the first three components contained enough information to represent the spectra in the PCA space. We have used those components in pairs (pc1–pc2 and pc2–pc3) to isolate the organic from the mineral.

The PCA was applied to the whole database of spectra using different parts of the IR spectrum. The first attempt was performed using the complete range of middle infrared. Then other filters were applied in order to optimise the process (e.g. fingerprint region, functional region (1800–3300 cm^{-1}) and a combination of some sub regions (i.e. CH and NN bonds)).

9. Results

The best results were obtained using the fingerprint region (Figs. 4 and 5). The points (in the PCA space) representing the organics spectra formed a group, or compact structure. Several points (representing the spectra taken on the rocks) were close and others were very far.

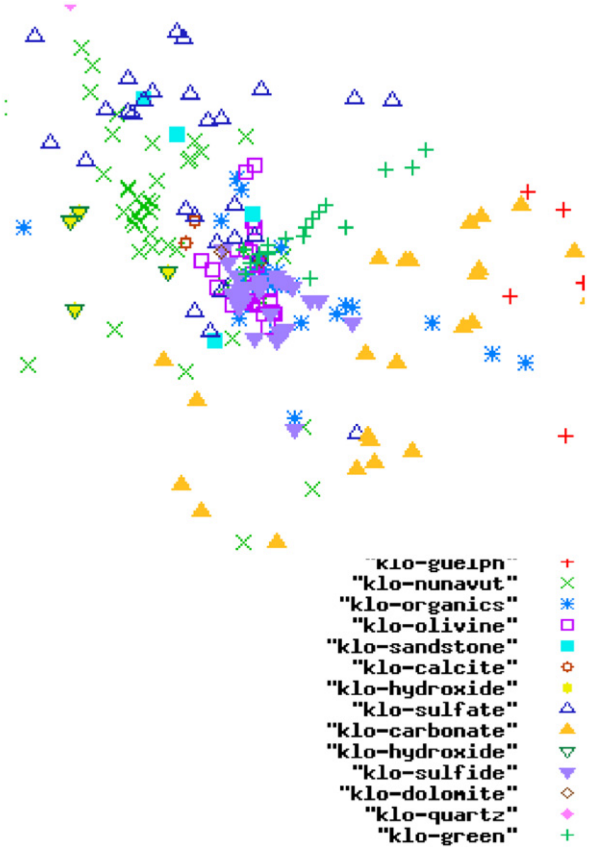

Fig. 4. Results of PCA using the fingerprint region of IR. This is a portion of the graphic (pc2–pc3). The blue stars represent the organic spectra used as reference. The samples associated to the spectra from Nunavut (green X) and Guelph (red +) near those blue stars contain traces of endolithes (visible green and red patches on the samples). The other points represent mineral taken from the JPL-ASTER database. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The points near the group of points representing organics spectra were identified as representing spectra containing trace of organic matter. Samples linking to those points show visual markers such as red and green patches.

When using the whole middle IR range, the grouping of points was dictated by the mineral contains of the samples rather then the organic composition. The organic group was isolated in a corner of the PCA space. The rock spectra were distributed among the mineral spectra. It was relatively easy to identify the mineral composition of each sample.

10. Other samples

We obtained some IR spectra from Antarctica [9] samples and submitted them to our PCA technique.

The series of points generated by the new spectra fall near the organic cloud previously obtained. Not all the new spectra were close to the cloud but those that were presented an interesting validation [8] which confirmed the fact that some IR signatures harboured traces of biomarkers.

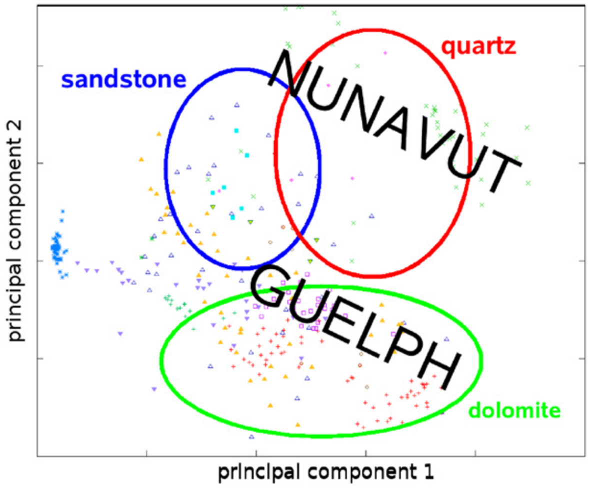

Fig. 5. Results of PCA when applied using the whole middle IR spectrum. It is easy to identify the mineral composition of each sample. The little blue cluster of points at the far left of the graphic is the organic spectra. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We had also accessed to the Mars Exploration Rovers (MER) spectra and submitted them to the PCA test. The MER spectra did not cover all the IR regions needed and part of them have a lot of noise rendering the interpretation of the results very difficult. We could not extract meaningful information from the results regarding the organic composition. Some mineral spectral signatures (i.e. sulphate and sulphide) were similar to those found on the surface [12,13] of Mars.

11. Conclusion

It is possible to use infra-red spectroscopy to detect biomarkers in rocks. The principal component analysis can be used to isolate the organics from the mineral. However, while the classification performed by the PCA can highlight the presence of organic matters, it cannot identify them.

The method must be carefully calibrated using known references spectra. Then the composition of any unknown spectrum may be determined by the process providing that the references database contains the proper spectra.

It is not necessary to recompute the whole PCA each time we add a new spectrum. Once the matrices *W* and *V* have been computed, they can be used as a simple rotation matrix to bring any spectrum to the PCA space.

Furthermore, the JCS-1 did not mask nor has influence on the IR signatures we were interested in. Therefore, the Martian soil will not mask organics matter.

References

[1] C. Ascaso, J. Wierzchos, New approaches to the study of Antarctic lithobiotic microorganisms and their inorganic traces, and their application in the detection of life in Martian rocks, International Microbiology 5 (2002) 215–222.

[2] J. Toporski, A. Steele, The relevance of bacterial biomarkers in astrobiological research, in: Proceedings of the Second European Workshop of Exobiology, 2002, pp. 239–242.

[3] C.R. Omelon, W.H. Pollard, F.G. Ferris, Environmental controls on microbial colonization of high arctic cryptoendolithic habitats, Polar Biology 160 (2005) 329–338.

[4] U. Matthes, S.J. Turner, D.W. Larson, Light attenuation by limestone rock and its constraint on the depth distribution of endolithic algae and cyanobacteria, International Journal of Plant Science 162 (2001) 263–270.

[5] E. Pretsch, P. Buhlmann, C. Affolter, Structure Determination of Organic Compounds, Springer, Berlin, 2000.

[6] M.D. Lane, M.D. Dyar, J.L. Bishop, Spectroscopic evidence for hydrous iron sulfate in the martien soil, Geophysical Research Letters 31 (2004) L19702.

[7] D. McKay, J. Carter, W. Boles, JSC-1: a new lunar regolith simulant, Lunar and Planetary Science 24 (1993) 963–964.

[8] S. Ronen, Principal component analysis of synthetic galaxy spectra, Monthly Notices of the Royal Astronomical Society 303 (1999) 284–296.

[9] A. Broersen, R. van Liere, Transfer functions for imaging spectroscopy data using principal component analysis, in: Eurographics—IEEE VGTC Symposium on Visualization, 2005.

[10] M.E. Wall, A practical approach to microarray data analysis, D.P. Berrar, W. Dubitsky, M. Granzow (Eds.), 2003 (Chapter 5).

[11] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes—The Art of Scientific Computing, Cambridge, 1992.

[12] K.P. Hand, R.W. Carlson, M. Anderson, Utilizing active mid-infrared microspectrometry for in-situ analysis of cryptoendolithic microbial communities of battleship promontory, dry valleys, Antarctica, Astrobiology and Planetary Missions 5906 (2005) 302–310.

[13] A. Aubrey, H.J. Cleaves, J.H. Chalmers, Sulfate minerals and organic compounds on mars, Geological Society of America 34 (2006) 357–360.