



UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

## TRABAJO FIN DE MÁSTER

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN

EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

**Procesamiento espectral para la pseudocuantificación de especies minerales en espectros del instrumento *Raman Laser Spectrometer* de la misión ExoMars de la ESA**

---

**Spectral processing for the pseudo-quantification of mineral species in spectra of the Raman Laser Spectrometer instrument of ESA ExoMars mission**

Autor:

**D. Guillermo López Reyes**

Tutor:

**Dr. D. Fernando Rull Pérez**

**Dr. D. Ignacio de Miguel Jiménez**

Valladolid, 6 de Septiembre de 2012

---

**TÍTULO:      **Procesamiento espectral para la pseudo-  
cuantificación de especies minerales en espectros  
del instrumento *Raman Laser Spectrometer* de la  
misión ExoMars de la ESA****

**Spectral processing for the pseudo-quantification  
of mineral species in spectra of the Raman Laser  
Spectrometer instrument of ESA ExoMars mission**

**AUTOR:           D. Guillermo López Reyes**  
**TUTOR:           Dr. D. Fernando Rull Pérez**  
**Dr. D. Ignacio de Miguel Jiménez**

DEPARTAMENTO:

---

**TRIBUNAL**

---

**PRESIDENTE:    Dr. D.**  
**VOCAL:           Dr. D.**  
**SECRETARIO     Dr. D.**

---

**FECHA:           6 de Septiembre de 2012**  
**CALIFICACIÓN:**

---

**Resumen del TFM**

El espectrómetro Raman Laser Spectrometer (RLS) forma parte del conjunto de instrumentos incluidos en el futuro *rover* que la misión ExoMars de la ESA y Roscosmos enviarán a Marte en 2018. Como instrumento portátil para exploración planetaria, la operación del instrumento deberá ser totalmente autónoma, sin intervención humana. Por tanto, varios paquetes de software y algoritmos deberán ser definidos y optimizados para proporcionar autonomía al instrumento.

Por otro lado, debido a la simplicidad y robustez requerida en instrumentos de vuelo, así como la interacción necesaria con otros instrumentos dentro del *rover*, su modo de operación tiene algunas características que pueden considerarse anómalas para un espectrómetro Raman típico. Por ejemplo, este instrumento analizará un mínimo de

20 puntos de una muestra en polvo homogéneamente distribuida, en lugar del habitual análisis en materiales masivos. Por tanto, para obtener del instrumento el máximo retorno científico posible, también es necesario desarrollar técnicas analíticas que se adapten a las características de operación del instrumento.

En este trabajo, tanto los aspectos de automatización a bordo como de postprocesado de datos han sido tenidos en cuenta: Se ha propuesto, evaluado y optimizado un algoritmo para la identificación de muestras relevantes a nivel astrobiológico, así como un método estadístico para la pseudo-cuantificación de fases minerales a partir de espectros Raman, y que ha sido probada con muestras reales.

## **Palabras clave**

Espectroscopía Raman, ExoMars, RLS, modo operación, algoritmia, autonomía, procesamiento de datos

## **Abstract**

The Raman Laser Spectrometer (RLS) instrument is a Raman spectrometer which forms part of the analytical laboratory drawer (ALD) of the future ExoMars rover which will be sent to Mars by ESA and Roscosmos in 2018. As a portable instrument for planetary exploration, the instrument operation will have to be autonomous, with no human intervention. Thus, several sets of algorithms will have to be defined and optimized to provide the instrument with autonomy.

Besides, due to the simplicity and robustness expected from flying models, as well as the necessary interaction with other instruments of the ALD, its operation mode is bound to have some atypical characteristics for a Raman spectrometer. For example, this instrument will analyze at least 20 points of a homogeneously powdered sample, instead of the usual analysis on bulk materials. Thus, to obtain the maximum scientific return of the instrument, some analytical techniques adapted to the characteristics of the operation mode of the instrument will have to be developed.

In this project, both the on-board automation and post-processing aspects have been addressed: An algorithm for the identification of astrobiological relevant samples has been proposed, evaluated and optimized, as well as a statistical method for the pseudo-quantification of mineral phases only based on Raman spectra, which has also been tested on real samples.

## **Keywords**

Raman spectroscopy, ExoMars, RLS, operation mode, algorithm, autonomy, data processing

## **Acknowledgements**

To the members of '*505 - Despacho Dioses de la Ingeniería e Innovación*', Alberto, Alejandro and Isaac, for the discussions and brainstormings during the search of new ideas. A special mention to Isaac, for his invaluable help with the programming in this project and his willingness, always present. To Aurelio, without whose wisdom and help it wouldn't have been possible. To the colleagues at the University of Jena (Germany), who provided the original idea for the Simple Baseline algorithm.

I kindly thank my family for their always present and unconditional help, and, of course, to Eva, always with me even in the distance.

## Table of Contents

Capítulo 1. Resumen: Introducción (English follows).....	7
Chapter 1. Introduction and background.....	11
1.1. Motivation and objectives .....	11
1.2. The Aurora Programme of the European Space Agency (ESA).....	11
1.3. The ExoMars programme 2016-2018 .....	12
1.4. The ExoMars rover instrument suite .....	14
1.5. The RLS instrument .....	16
1.5.1. How it works .....	16
1.5.2. Sample preparation and operation modes .....	18
1.6. The operation mode definition: ExoMars RLS Simulator .....	19
1.7. Problem description.....	21
1.7.1. Detection of traces of astrobiological interest.....	23
1.7.2. Pseudo-quantification of mineral species from Raman spectra .....	24
Capítulo 2. Resumen: Detección de bandas espectrales de interés astrobiológico (English follows).....	26
Chapter 2. Detection of spectral bands of astrobiological interest .....	32
2.1. Concept and methodology .....	32
2.1.1. Algorithm development .....	32
2.1.2. Regions of astrobiological interest.....	33
2.1.3. General methodology.....	33
2.2. Baseline removal.....	34
2.2.1. Introduction.....	34
2.2.2. Algorithm description.....	35
2.2.3. Methodology .....	36
2.2.4. Results Discussion.....	38
2.3. Detection of the spectrum peaks .....	40
2.3.1. Algorithm description.....	40
2.3.2. Methodology .....	43
2.3.3. Results and discussion .....	44
Capítulo 3. Resumen: Pseudo-cuantificación de especies minerales a partir de espectros Raman del instrumento RLS (English follows) .....	45
Chapter 3. Pseudo-quantification of mineral species from the RLS instrument Raman spectra.....	46
3.1. Introduction .....	46
3.2. Experiment description.....	46

3.3. Statistical analysis .....	47
3.4. Results discussion .....	48
Capítulo 4. Resumen: Compendio de resultados y conclusiones (English follows) .....	52
Chapter 4. Wrap-up.....	54
4.1. Final considerations and conclusions .....	54
4.1.1. Contribution to the instrument operation mode.....	54
4.1.2. Contribution to data analysis .....	55
4.2. Future work.....	56
4.3. Conference papers related to this work.....	56
References .....	57

## Capítulo 1. Resumen: Introducción (English follows)

### 1.1. Motivación y objetivos

La futura misión ExoMars de la Agencia Espacial Europea enviará en 2018 a Marte, entre los instrumentos de su carga de pago, el espectrómetro Raman RLS (*Raman Laser Spectrometer*). La espectroscopía Raman ha sido siempre considerada una técnica de laboratorio debido a restricciones técnicas, y mediante la operación de un técnico de laboratorio. Ahora que se plantea la exploración planetaria con esta técnica, es necesario desarrollar una algoritmia que provea al instrumento con la capacidad de actuar de manera autónoma.

Además, el instrumento RLS operará de manera diferente a la habitual en un laboratorio, ya que analizará un conjunto de al menos 20 puntos en una muestra molida en polvo homogéneamente distribuido, en lugar de muestras masivas en determinados puntos seleccionados por el operador. Esta manera de realizar los análisis abre la puerta a métodos de análisis estadísticos que pueden llevar no solo a la identificación de minerales, sino también a cuantificar su presencia en mezclas de minerales.

Este documento presenta los resultados de la investigación realizada en el aspecto de la automatización, así como el de post-procesado de los datos. El Capítulo 1 introduce el problema y establece las bases para el estudio. El Capítulo 2 estudia y propone algoritmos para la identificación a bordo del *rover* de muestras de interés astrobiológico. El Capítulo 3 discute un método para la pseudo-cuantificación de especies minerales. Finalmente, el Capítulo 4 agrupa los resultados, establece las conclusiones y pauta líneas futuras para la continuación de la investigación.

### 1.2. El programa Aurora de la Agencia Espacial Europea

El programa Aurora (Figura 1) fue creado en el año 2001 con el objetivo de crear un programa a largo plazo para la exploración robótica y humana del sistema solar (Marte, la luna y asteroides principalmente). También se pretende buscar vida más allá de la tierra.

El programa Aurora es un programa paso a paso cuyas misiones irán creciendo en complejidad, culminando, si todo va bien, en una expedición humana a Marte en la década de 2030. Los pasos a seguir incluyen análisis remotos del ambiente marciano, exploración robótica de la superficie del planeta rojo, misiones de retorno de muestras y bases extraterrestres permanentes, y pasan por la exploración de la luna.

### 1.3. El programa ExoMars 2016-2018

La búsqueda de vida presente o pasada en Marte es una de las cuestiones científicas de mayor relevancia hoy en día. El programa ExoMars de la ESA tiene como finalidad investigar el entorno marciano y demostrar nuevas tecnologías, estableciendo el camino para futuras misiones de retorno de muestras en 2020. La Figura 2 muestra la actividad humana presente y futura en el planeta rojo. Las misiones actuales (MER, MSL) tienen como finalidad dilucidar si existen en Marte condiciones que puedan

indicar si la vida fue y/o es posible en Marte. Las futuras misiones buscarán evidencias de vida presente o pasada.

El programa ExoMars contempla dos misiones. La primera, en 2016, consiste en un orbitador con un detector de gases orbital (TGO) y un módulo demostrador de entrada, descenso y aterrizaje (EDM). La segunda, en 2018, enviará un *rover* a la superficie en colaboración con la Agencia Espacial Rusa Roscosmos, que tendrá movilidad en la superficie, así como la habilidad de analizar muestras extraídas bajo la superficie con un taladro (Figura 3).

Además de la finalidad de desarrollo tecnológico, la misión pretende responder a una serie de cuestiones científicas como la búsqueda de vida presente o pasada, entender el ciclo del agua y la geoquímica de Marte, e investigar los gases atmosféricos y sus fuentes.

La capacidad del *rover* de ExoMars de analizar muestras extraídas bajo la superficie es importante para detectar posibles trazas biológicas, ya que la profundidad protegería de la radiación ultravioleta y del consiguiente proceso de degradación.

#### 1.4. El conjunto de instrumentos del *rover* de ExoMars

La Figura 4 muestra los instrumentos que forman parte de la carga de pago de la misión, llamada Pasteur. El *rover* dispondrá de capacidad de visión con la cámara panorámica (PanCam), un espectrómetro de masas capaz de detectar orgánicos y si son de origen biótico o abiótico (MOMA), un espectrómetro de infrarrojo (MicrOmega), un radar de subsuperficie (WISDOM), el instrumento Ma\_MISS, que analizará las muestras durante el taladrado y el espectrómetro *Raman Laser Spectrometer* (RLS), que es una poderosa herramienta para la identificación y caracterización de minerales y biomarcadores.

#### 1.5. El instrumento RLS

El *Raman Laser Spectrometer* (RLS) es un espectrómetro Raman que permite la identificación de compuestos a partir de vibraciones moleculares excitadas mediante un láser, y captadas mediante un espectrómetro. La espectroscopía Raman se considera una técnica no destructiva, ya que analiza el material sin producir cambios de ningún tipo en su estructura. El funcionamiento del instrumento en Marte será totalmente autónomo (enfoque del sistema óptico, algoritmos de optimización de parámetros de adquisición, guardado y compresión, y finalmente envío a tierra).

Para cumplir con las políticas de protección planetaria (para evitar contaminación de cuerpos extraterrestres con material biológico terrestre), la configuración del instrumento es la que se muestra en la Figura 5, con tres unidades aisladas y conectadas únicamente por cableado óptico o eléctrico. Las unidades son: el espectrómetro (SPU), la unidad de control y excitación (ICEU) y el cabezal óptico (iOH). De esta forma no existe contacto directo con la muestra, que está en la zona ultra limpia del *rover*, evitando así posibles contaminaciones.

El láser de excitación de la muestra es continuo y con una longitud de onda de 532 nm, con un *spot* en la superficie de la muestra de unas 50 micras, y una irradiancia entre



0.6 y 1.2 kW/cm<sup>2</sup>, niveles que se mantienen por debajo del límite de daño térmico de minerales termolábiles. El rango cubre desde 150 a 3800 cm<sup>-1</sup> de desplazamiento Raman (números de onda con respecto a la longitud de onda de excitación), con una resolución de entre 6 y 8 cm<sup>-1</sup>. La Figura 6 muestra una imagen de un prototipo de laboratorio, así como el esquemático del modelo de vuelo del espectrómetro.

El instrumento puede funcionar según dos modos de operación: El primero, escaneo “inteligente”, basado en la información proporcionada por otros instrumentos como MicrOmega, que pueden indicar dónde se encuentran potenciales puntos de interés en la muestra. El segundo método, por defecto, se trata de un escaneo automático de la muestra en determinadas posiciones prefijadas de antemano. La Figura 7 muestra el sistema de preparación y distribución de muestras (SPDS), que aplanar la superficie del polvo antes de colocar la muestra bajo los instrumentos.

## 1.6. Definición del modo de operación: Simulador ExoMars RLS

Se entiende por modo de operación a la serie de operaciones y algoritmia que el instrumento ha de realizar de manera autónoma para la obtención de un espectro Raman de calidad. No ha de confundirse el modo de operación con el desarrollo de la algoritmia necesaria para que el instrumento se convierta en un instrumento de vuelo (telemetría, comunicación con el rover, etc). La definición del modo de operación es una tarea puramente científica, que traslada las actuaciones que el operador humano realiza con el espectrómetro (enfoque del láser en la muestra, ajuste del tiempo de adquisición del espectro, eliminación de la fluorescencia, etc.) a un sistema capaz de decidir autónomamente cuál es el punto de enfoque, detectar si la muestra bajo estudio es fluorescente y por tanto requiere un tratamiento especial, cuál será el tiempo de integración óptimo, etc.

La definición de toda esta algoritmia está basada en experimentos científicos que ayuden a la definición de parámetros objetivos detectables por el instrumento que le permitan realizar aquellas tareas que el operador humano realiza de manera manual en base a su experiencia.

Para realizar esta tarea, en la Unidad Asociada UVa-CSIC-CAB hemos desarrollado el llamado Simulador ExoMars (Figuras 8 a 10), que nos ha permitido realizar experimentos bajo condiciones similares a las proporcionadas por el *rover* de ExoMars, así como desarrollar algoritmos que permiten al instrumento funcionar de manera autónoma, como por ejemplo, optimizar los parámetros de adquisición en cada punto de la muestra. Además, el uso de este sistema ha permitido la sistematización de la adquisición, permitiendo obtener grandes cantidades de datos, y por tanto abriendo la puerta al desarrollo de técnicas de análisis estadístico de los datos en muestras en polvo.

Sin embargo, todavía queda mucho por hacer en lo que a definición del modo de operación se refiere. En particular, el problema atacado en este trabajo se refiere a la detección automática por parte del instrumento de muestras de relevancia astrobiológica, que permitirán priorizar los datos que el instrumento envíe a tierra.

## 1.7. Descripción del problema

Como se ha dicho, la espectroscopía Raman, a pesar de su expansión durante los últimos 25 años, ha sido siempre considerada una técnica de laboratorio, sobre todo debido a sus limitaciones tecnológicas. Sin embargo, recientemente la tecnología ha permitido miniaturizar los equipos hasta el punto de considerar la técnica para su uso en exploración planetaria.

Sin embargo, se necesita desarrollar la algoritmia capaz de hacer que el instrumento funcione de manera autónoma (ya que habitualmente el instrumento es operado por un técnico). Por ejemplo, un algoritmo de la mayor importancia es el ya mencionado de ajuste de los parámetros de adquisición, que se calculan en cada punto que el instrumento analice. Otros algoritmos que han de desarrollarse incluyen la detección y eliminación de fluorescencia o rayos cósmicos, o la detección de trazas astrobiológicamente interesantes en el espectro adquirido.

Este trabajo se ha centrado en este último algoritmo de detección de trazas astrobiológicamente relevantes. La finalidad es poder enviar a tierra los resultados de la adquisición del instrumento en atención a su relevancia, de forma que el equipo científico en tierra pueda tomar decisiones tácticas (que en ocasiones deberán tomarse antes de poder haber analizado todos los datos) basándose en los datos más relevantes.

La Figura 11 muestra un típico espectro Raman. El eje de abscisas se corresponde con el desplazamiento Raman, y el de ordenadas se da en unidades arbitrarias, pero representa la intensidad en cada longitud de onda. La intensidad de un pico se mide con respecto al nivel de la línea de base o *baseline*. Esta línea de base no aporta información útil al espectro Raman. El algoritmo de detección de picos en regiones de interés lo que hará en primer lugar será eliminar la línea de base, y a continuación, determinará las posiciones de los picos. Para ello se presentan dos algoritmos diferentes para el cálculo de la línea de base, y se evalúa un tercer algoritmo para la detección de los picos (ver Capítulo 2).

Para la maximización del retorno científico del instrumento, es interesante analizar si un método estadístico de tratamiento de los datos sería capaz de obtener información cuantitativa de la abundancia de minerales en mezclas. En el Capítulo 3 de este documento se recogen algunos aspectos sobre el particular.

## Chapter 1. Introduction and background

### 1.1. Motivation and objectives

ESA's future ExoMars mission will send a rover to Mars in 2018 with a payload of scientific instruments among which the Raman spectrometer RLS (Raman Laser Spectrometer) is included. Due to technology restrictions, Raman spectroscopy has never been used in planetary missions, and has always been considered as a laboratory technique more than an in-situ one, controlled by a human operator. In the planetary exploration scenario, the instrument will have to perform autonomously, and thus, a set of algorithms needs to be defined, optimized and developed to provide the instrument with the necessary autonomy.

In addition, the mode in which the instrument will operate differs in some aspects from the usual Raman laboratory operation, to meet the flying robustness requirements as well as to comply with the restrictions imposed by the interaction with other instruments of the payload, which share a common Sample Preparation and Distribution System (SPDS). This instrument, instead of analyzing a bulk material in some points of interest as a spectroscopist would do, will analyze a predefined number of points (no less than 20) of a homogeneously powdered sample. This opens the door for statistical analytical methods which can lead not only to the identification of minerals, but also to the quantification of their presence in mineral mixtures.

This document presents the results of research performed on both the on-board automation aspect, as well as the post-processing one. Chapter 1 introduces the problem and sets the background for the study. In Chapter 2 a proposal, evaluation and optimization of an algorithm for the identification of astrobiological relevant samples is described. Chapter 3 discusses a statistical method for the pseudo-quantification of mineral phases, which has been tested on real samples. Finally Chapter 4 wraps up the results, discusses the conclusions and proposes future lines of research.

### 1.2. The Aurora Programme of the European Space Agency (ESA)

Aurora (Figure 1) is part of Europe's strategy for space, endorsed by the European Union Council of Research and the ESA Council in 2001. This strategy calls for Europe to:

- Explore the Solar System and the Universe.
- Stimulate new technology.
- Inspire the young people of Europe to take a greater interest in science and technology.

As a result of this challenge, in 2001 ESA set up the Aurora Program. The primary objective of Aurora is to create, and then implement, a European long-term plan for the robotic and human exploration of the solar system, with Mars, the Moon and the asteroids as the most likely targets. A second objective is to search for life beyond the Earth. Future missions under the program will carry sophisticated exobiology payloads



Figure 1. Aurora logo

to investigate the possibility of life forms existing on other worlds within the solar system.

It is clear from these objectives that the interdependence of exploration and technology is the basis of the Aurora Program. On the one hand the desire to explore provides the stimulus to develop new technology while on the other, it is the introduction of innovative technology that will make exploration possible [1].

### Aurora roadmap [1]

Aurora's step-by-step approach means that missions will increase in complexity over time, culminating - if all goes well - in a human expedition to Mars by the year 2030. Steps on the way to Mars will probably include exploration of the Moon as well as:

- Remote sensing of the Martian environment
- Robotic exploration and surface analysis
- Mars sample return missions
- A robotic outpost

Not all these steps towards the ultimate goal of sending humans to Mars will necessarily be part of the Aurora Program. As the result of international cooperation, various collaborating agencies will make a contribution to those missions that best meet their particular requirements and areas of expertise.

## 1.3. The ExoMars programme 2016-2018

Establishing if life ever existed on Mars is one of the outstanding scientific questions of our time. To address this important goal, the European Space Agency (ESA) has established the ExoMars program to investigate the Martian environment and to demonstrate new technologies paving the way for a future Mars sample return mission in the 2020's [2]. This program fits into the international roadmap for the exploration of Mars, shown in Figure 2. Currently, the red planet has several satellites and orbiting instruments, as well as ground vehicles. The Mars Exploration Rovers (MERs) were sent to Mars to look for evidences of water. The recently landed Mars Science Laboratory (MSL) will look for past or present evidences of habitability, i.e., sites in which life could exist or have existed. Future missions will look for evidences of past or present life.

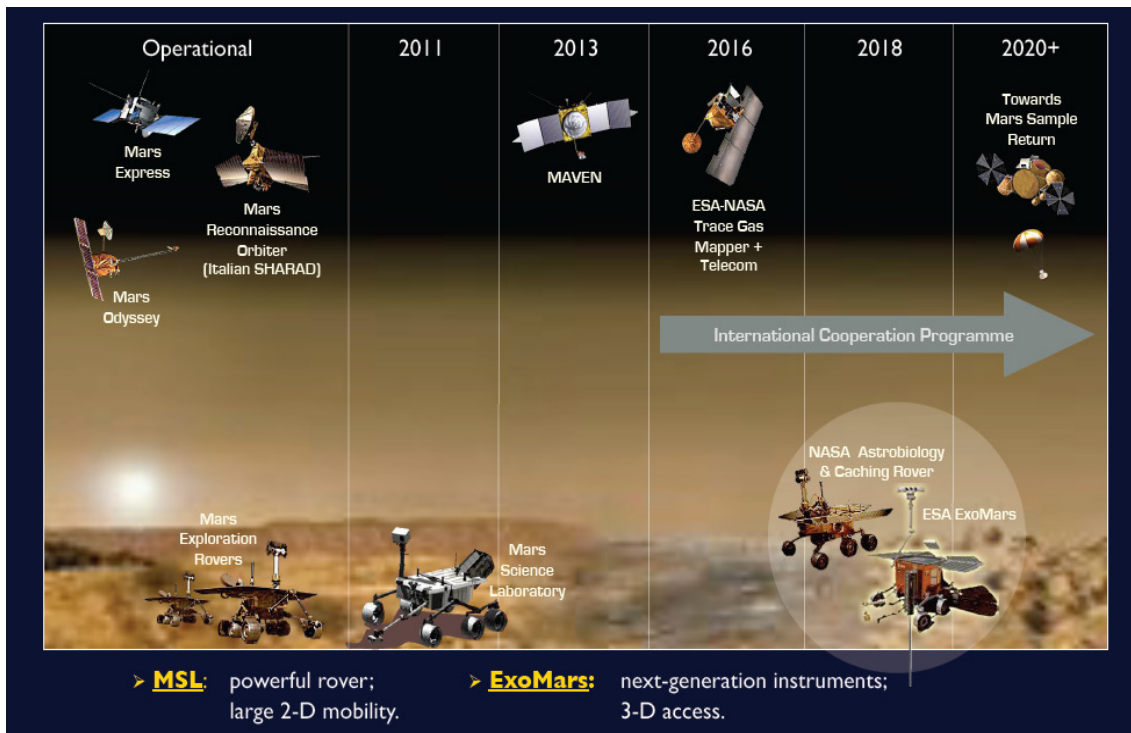


Figure 2. Past, present and planned missions to the red planet.

Two missions are foreseen within the ExoMars program: one consisting of an Orbiter plus an Entry, Descent and Landing Demonstrator Module, to be launched in 2016, and the other, featuring a rover, with a launch date of 2018. Both missions will be carried out in cooperation with the Russian space agency Roscosmos.

The objectives of this program can be subdivided in technological and scientific aspects [3]. In the technological plane, the ExoMars program will demonstrate a number of essential flight and in-situ enabling technologies that are necessary for future exploration missions, such as an international Mars Sample Return mission. These include [2]:

- Entry, descent and landing (EDL) of a payload on the surface of Mars;
- Surface mobility with a rover;
- Access to the subsurface to acquire samples; and
- Sample acquisition, preparation, distribution and analysis.

At the same time a number of important scientific investigations will be carried out, for example [2]:

- Search for signs of past and present life on Mars;
- Investigate how the water and geochemical environment varies;
- Investigate Martian atmospheric trace gases and their sources.

The 2016 mission includes a Trace Gas Orbiter (TGO) and an Entry, Descent and Landing Demonstrator Module (EDM). The Orbiter will carry scientific instruments to detect and study atmospheric trace gases, such as methane. The EDM will contain sensors to evaluate the lander performance as it descends, and additional sensors to study the environment at the landing site.

For the 2018 mission, ESA is developing a Mars rover carrying a comprehensive suite of analytical instruments dedicated to exobiology and geochemistry research named after Pasteur.

The Rover will travel searching for traces of past and present signs of life, collecting and analysing samples from within rocky outcrops and from the subsurface, down to a depth of 2 m with the help of a drill (Figure 3).

An effective chemical identification of biomarkers requires access to well-preserved organic molecules. Because the Martian atmosphere is more tenuous than Earth's, three important physical agents reach the surface of Mars with adverse effects for the long-term preservation of biomarkers [3]:

- The ultraviolet (UV) radiation dose is higher than on our planet and will quickly damage potential exposed organisms or biomolecules.
- UV-induced photochemistry is responsible for the production of reactive oxidant species that, when activated, can also destroy biomarkers; the diffusion of oxidants into the subsurface is not well characterized and constitutes an important measurement that the mission must perform.
- Ionizing radiation penetrates into the uppermost meters of the planet's subsurface.



Figure 3. Artist concept of the ExoMars rover on the Mars surface.

This causes a slow degradation process that, over many millions of years, can alter organic molecules beyond the detection sensitivity of analytical instruments.

Thus, the rover's surface mobility and the 2-m vertical reach of the drill are both crucial for the scientific success of the mission and for the preparation of the Mars Sample Return (MSR) mission, as it will be able to reach potentially non-radiated samples.

#### 1.4. The ExoMars rover instrument suite

Figure 4 shows the instruments forming part of the current configuration of the Pasteur Payload (PPL), name given to the instrument suite that will be included in the ExoMars 2018 mission rover.



Instrument Name	Description	Mass (kg) including maturity margin
PanCam (WAC + HRC)	Panoramic camera system	1.560
MOMA	LD-MS + Pyr GC-MS for organic molecule characterisation	6.100
MicrOmega IR	IR imaging spectrometer	0.960
Raman (internal)	Raman spectrometer	2.260
WISDOM	Shallow ground-penetrating radar	1.380
Ma_Miss included in 2.0-m drill	IR borehole spectrometer	0.650

Figure 4. Current configuration of the Pasteur Payload [4]

Below, a short description of the instruments is provided [2, 5].

The **PanCam** - panoramic camera - has been designed to search for textural information on rocks (for example, laminations or pitting) that can be related to the presence of organisms on Mars. PanCam will also capture other information that will help in revealing the geological characteristics of the Martian environment.

This instrument will provide stereo and 3D imagery of the terrain around the Rover, for the benefit of the mission as a whole.

**MOMA** is the largest instrument in the ExoMars Rover, and the one directly targeting biomarkers. MOMA will answer questions related to the potential origin, evolution and distribution of life on Mars. This instrument will detect organic molecules, even at very low concentrations, and try to establish their biotic or abiotic origin by molecular identification in terms of chirality. In addition to studying the samples collected by the drill, MOMA will also analyse gases in the Martian atmosphere.

**MicrOmega** is a micro-imaging system designed to identify, at grain scale, the mineralogical and the molecular composition of the Martian samples collected by the ExoMars drill.

The Raman Laser Spectrometer (**RLS**) is a Raman spectrometer which provides a powerful tool for the definitive identification and characterization of minerals and biomarkers. Raman spectroscopy is sensitive to the composition and structure of any mineral or organic compound. This capability provides direct information of potential organic compounds that can be related with present or past signatures of life on Mars as well as general mineralogical information for igneous, metamorphous, and sedimentary processes, especially water-related geo-processes.

**WISDOM** (Water Ice Subsurface Deposit Observation on Mars) is radar that will provide a detailed view of the Red Planet's shallow subsurface structure by sounding the upper layers of its crust. It will study the nature of the subsurface remotely, using radar pulses from a UHF ground penetrating radar, covering the frequency range from 500 MHz to 3 GHz, to map the subterranean layers.

Located inside the ExoMars Rover's drill, **Ma\_MISS**, the Mars Multispectral Imager for Subsurface Studies, will be the instrument in closest contact with the Martian subsurface. Ma\_MISS will image the walls of the borehole created by the drill to study Martian mineralogy and rock formation. This will provide valuable information for the study of subsurface soil and rock layers (i.e., stratigraphy), the distribution and state of water-related minerals, and will help to characterize the geophysical Martian environment.

## 1.5. The RLS instrument

As stated, ExoMars is a planetary mission to Mars led by ESA and proposed for launch in 2018, with the Pasteur Payload in which a Raman spectrometer is included: the Raman Laser Spectrometer (RLS).

RLS can identify the composition and structure of minerals and rocks at the mineral grain scale. This capability provides a definitive characterization of a target material. Hence, RLS can be used as a 'rapid target evaluation' for tactical decisions during the mission. Indeed, the Raman spectrometer will be used to identify organic compounds and search for signatures of life, to identify the mineral products and indicators of biological activities, to characterize mineral phases produced by water-related processes and to characterize igneous minerals and their products resulting from alteration processes (e.g. oxidation) [2].

The RLS instrument will also support the scientific measurements by correlating its spectral information with other spectroscopic and imaging instruments such as the MicrOmega Infrared Spectrometer. Furthermore, the Raman instrument is capable of measuring the sample at a fast-pace (within minutes) in order to release the selected sample for further analysis by other ExoMars instruments (e.g. the MOMA instrument).

### 1.5.1. How it works

The Raman instrument will collect and analyze the scattered light emitted by a laser on a sample (e.g. Martian rock). The spectral information received by the spectrometer (number of peaks, position and relative intensities) is determined by the molecular structure and composition of a compound, enabling the identification and characterization of the Martian molecules, minerals and rocks. This method allows analyzing the sample in a non-destructive way, which means that the sample remains intact for further analysis by other instruments.

Due to implementation issues (thermal or planetary protection issues) within the rover analytical lab, the RLS instrument consists of three independent units connected via electrical and optical harnesses [6]. The laser is in the Instrument Control and



Excitation Unit (ICEU). Its collimated beam is conveyed to the Internal Optical Head (iOH), where it is focused on the target. This unit must incorporate a mechanism to adjust the optical focus because of irregularities of the sample surface. The Raman signal is collected through the same front optics, the excitation wavelength is filtered out, and the signal is transmitted to the Spectrometer Unit (SPU) for dispersion and capture on a two-dimensional Charged Coupled Device (CCD).

Figure 5 shows how the sub-units are implemented in the rover Analytical Laboratory Drawer (ALD). This configuration is necessary because the target sample has to be kept within an ultra-clean zone in the rover to comply with the planetary protection policies (aimed at avoiding biological contamination of extraterrestrial bodies), and thus it is not in physical contact with RLS. These subunits are further described below [6].

- Spectrometer unit (SPU). The dispersion of Raman light is obtained by transmission using a holographic grating. The Raman spectrum is registered on a 2048x512 pixel CCD, which is kept at -10°C by means of a thermo-electric cooling device.
- Control and excitation unit (ICEU). It includes de DC/DC power converters and the data processing capability (micro-controller, RAM memory, clock and CANBus). Its role is also to capture the RLS health parameters and to control the thermal management. To support other functionalities of the instrument, it comprises the laser with two redundant excitation outputs, the CCD front-end electronics and the autofocus driver to focus the laser beam on the surface of the sample.
- Optical head (iOH). The range of focus is 2mm simultaneously for the excitation and collection fibers, as it is a confocal instrument.

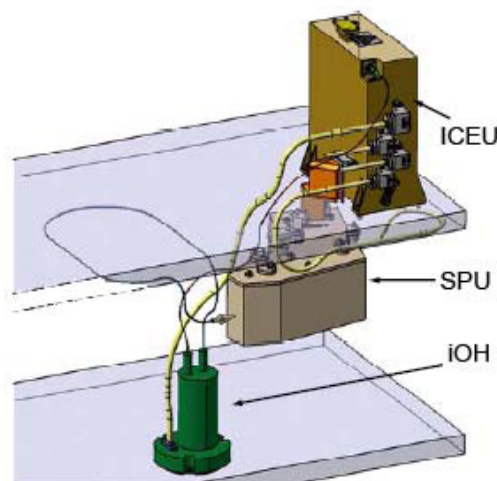


Figure 5 Schematics of the implementation of the RLS instrument in the ExoMars rover

The laser source is continuous (non-pulsed) at a wavelength of 532 nm. This laser will be able to provide a total of 400 individual sets of measurement cycles (10 minutes maximum each). The laser spot size will be around 50 microns on the surface of the sample, to be compatible with the other analytical tools of the mission. This spot size

procures an irradiance level on the surface of the sample between 0.6 and 1.2 kW/cm<sup>2</sup>. The upper limit is fixed to remain below the threshold of powder grain thermal damage, mainly in oxides and hydroxides. The goal of the spectrometer is to cover the spectral shift from around 150 to 3800 cm<sup>-1</sup>, to capture the range of fundamental vibration of rock forming minerals, oxyanionic anions, functional groups of organic species, water and OH vibrations in hydrates. The Raman spectral resolution goal is around 6 cm<sup>-1</sup> in the fingerprint spectral region below 2000 cm<sup>-1</sup> and around 8 cm<sup>-1</sup> above this limit [7]. Figure 6 shows the schematics and a laboratory prototype of the RLS instrument SPU.

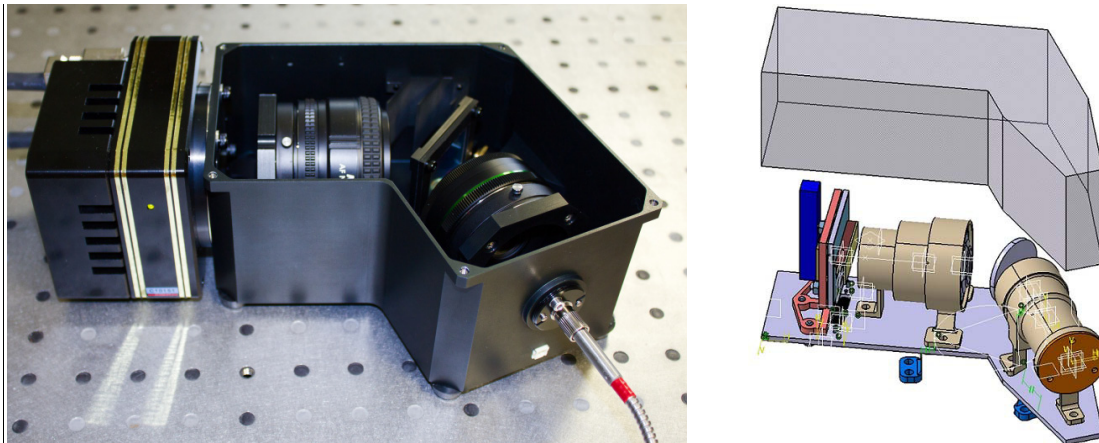


Figure 6. Laboratory prototype and schematics of the RLS spectrometer unit (SPU).

### 1.5.2. Sample preparation and operation modes

The RLS instrument will operate on samples collected from the surface and the subsurface down to 2 meters, by means of the drill included in the ExoMars rover. Then, the samples will be crushed into a fine powder and by means of a dosing station the powder will then be presented to RLS and other instruments [7, 8].

There shall be two operational modes [6]: an automatic scanning and a smart scanning.

The mode selection will be performed by spacecraft commands from ground. During automatic scanning, the rover shall place the target following a preconfigured sequence of movements in front of RLS optical head. RLS shall take at least 20 shots per sample at regularly spaced spots on the target. During smart scanning, MicroOmega IR images shall be processed by the rover to determine if there is any target of interest which can be reached by the RLS instrument. The rover shall place the target of interest below the RLS optical head. In case no target of interest is found, the preconfigured sequence of movements shall be performed.

Figure 7 shows the sample preparation and distribution system of the rover. The powdered sample is provided to the analytical laboratory drawer (ALD) and flattened to provide the analytical instruments with a flattened surface to analyze.

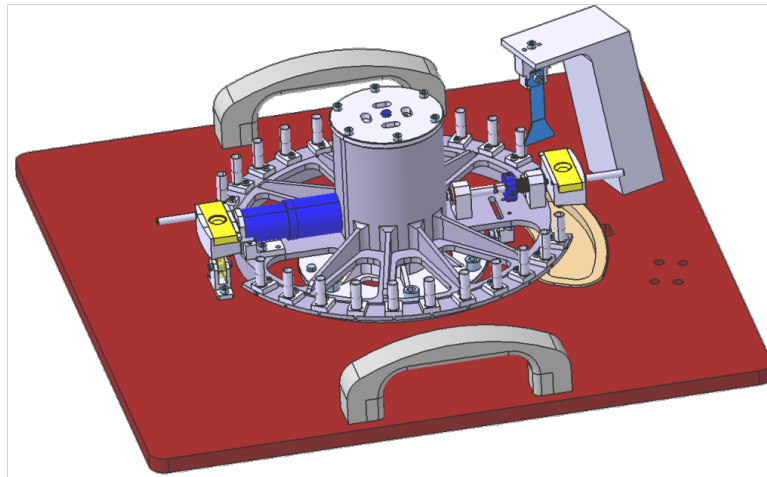


Figure 7. Schematics of the rover Sample Preparation and Distribution System

## 1.6. The operation mode definition: ExoMars RLS Simulator

The potential of Raman spectroscopy as a non-invasive or destructive procedure applied to all kinds of materials defines it as a very important technique for all types of materials, minerals or biological analyses. This spectroscopic technique has remained, however, unprecedented in planetary exploration, so it has been traditionally carried out with the intervention of a human operator.

The operation mode of the RLS instrument is understood as the series of operations and algorithms that the instrument is to perform autonomously for obtaining a Raman spectrum of the highest quality. This must not be confused with the development of the algorithms necessary for the instrument to become a flight instrument (telemetry, communication with the rover, etc). The definition of the operation mode is a purely scientific task, which shifts the actions that the human operator performs with the spectrometer (laser focusing on the sample, adjustment of the spectrum's acquisition time, elimination of fluorescence, etc) to a system able to autonomously decide which is the point of focus, to detect whether the sample under study is fluorescent or not and therefore it requires a special treatment, what will the optimal integration time be, etc.

The definition of all these algorithms is based on scientific experiments that help to define objective parameters which are measurable by the instrument, to enable it to autonomously carry out those tasks that the human operator manually performs based on his experience.

For the development and optimization of these necessary algorithmic tools for the instrument, we needed to develop an adequate setup. The idea is to provide the scientist with a platform in which realize the experiments in a rover-like scenario, taking into account how the RLS instrument will interact with the rover, setting the basis for the development of the operation mode.

The so-called ExoMars Simulator that we have developed at UVa-CSIC-CAB Associated Unit [6, 9] (Figures 8, 9 and 10) has enabled to perform many scientific experiments

under conditions similar to those provided by the Analytic Laboratory Drawer (ALD) of the ExoMars rover. Thanks to it, it has been possible to define how to take spectra autonomously, without the presence of a human operator. So far, some sophisticated commercial Raman equipment were able to work in a pseudo-automatic way [10], which performs a series of measurements at different points in a sample, treating all points equally. However, the use of the ExoMars RLS Simulator has allowed the definition of an automatic acquisition system that optimizes the parameters in each one of the points, thereby providing better results in less time.

On the other hand, the use of this system has allowed performing experiments that have helped understand the possibilities offered within the framework of planetary exploration, in order to maximize the scientific return of the resulting products from the operation of the instrument on Mars. In particular, the development of the ExoMars simulator has paved the way for the systematic study of the operation on powdered samples.

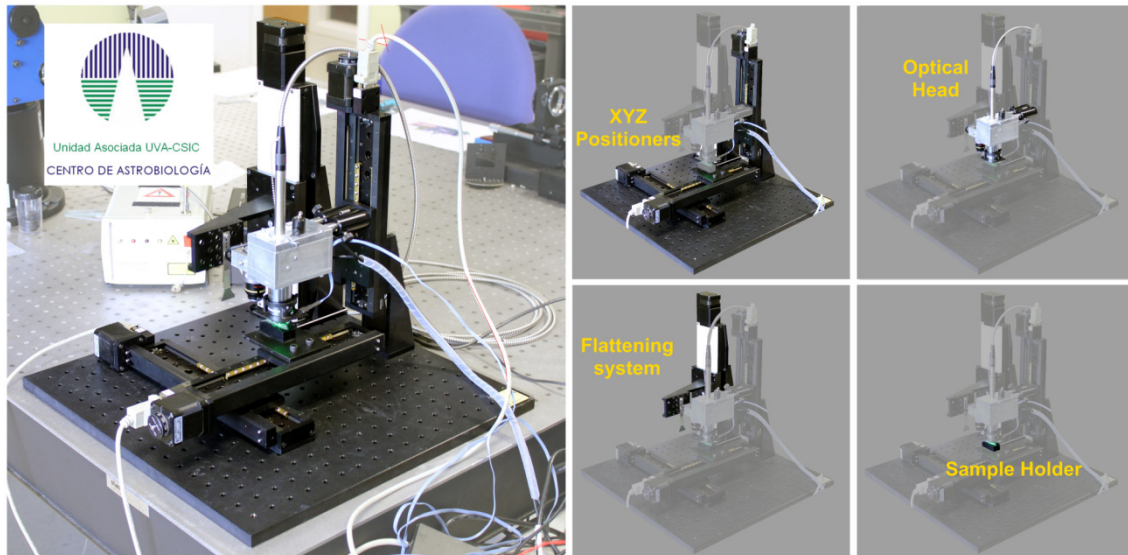


Figure 8. ExoMars RLS Simulator prototype and schematics

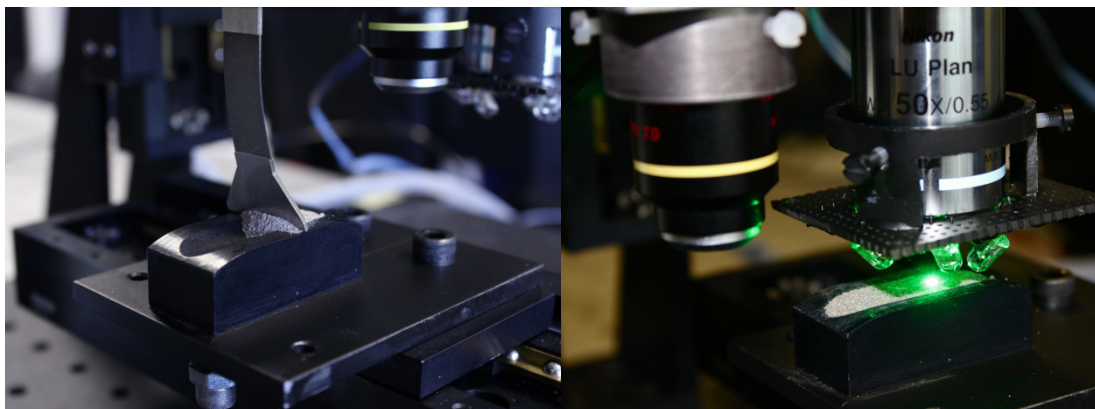


Figure 9. Detail of the sample preparation and subsequent analysis



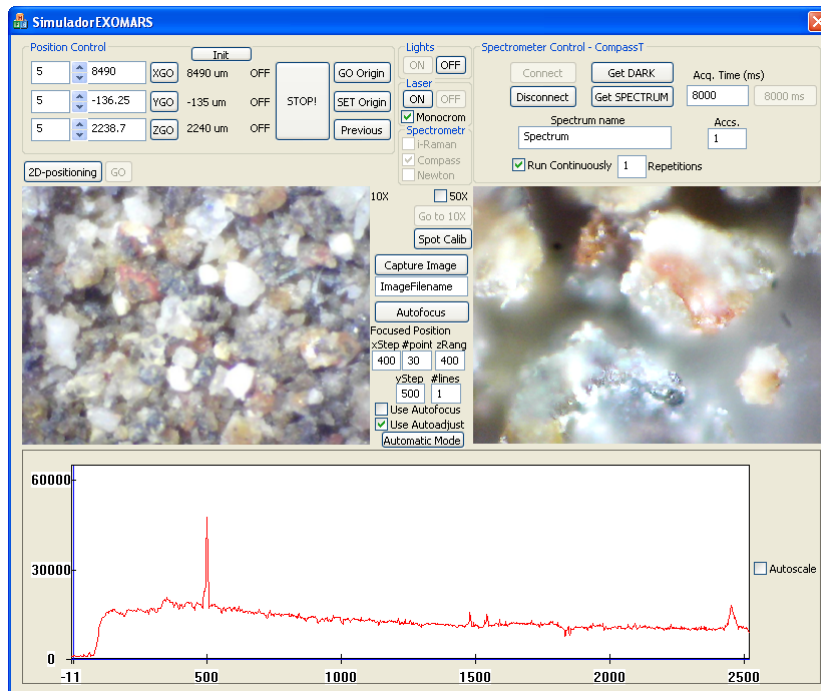


Figure 10. Software developed for controlling the Simulator

Despite the level of development achieved thanks to the ExoMars Simulator, the definition of the operation mode is far from completion. The desire to obtain the best possible results highlights the need to develop new algorithms that provide the instrument of a greater intelligence. These new algorithms include the special treatment of points at which materials of special scientific interest are detected (eg. biogenic material), intelligent minimization of fluorescence levels depending on the specific spectrum, detection and treatment of thermolabile samples or the collaborative science with other instruments, so that information obtained by other instruments could be used to improve the performance of the RLS instrument and vice versa. Many of these aspects have been defined and explained in various documents [11-14]. In the present one, the definition of an algorithm for the detection of astrobiological relevant samples is addressed.

### *Cell for Martian Environment Simulation*

So far, simulations of the operation mode have been conducted in laboratory conditions. With the addition of a Martian Environment Simulation Cell [15] to the equipment, the capabilities of the former get extended and a new and more realistic scientific approach to the autonomous simulation of the operation mode of the mission is reached.

## **1.7. Problem description**

Raman spectroscopy is well recognized as a powerful tool for the chemical and structural identification of materials in the solid, liquid or gas state. Its analytical capabilities, without the need to perturb a sample, have made this technique unique for many applications where the materials are scarce or very valuable and rare. As

consequence of these capabilities a great expansion in applications has been seen in the last 25 years across many fields ranging from chemistry to mineralogy and geology, art and archaeology, forensic sciences, polymers and biomaterials [16].

In spite of this expansion and the capabilities of Raman spectroscopy, it has been, and to a great extent still is, considered as a laboratory technique. Nevertheless, in recent years several technological developments in laser sources, optical elements, spectrometers and detectors have led to the possibility of developing robust, compact and miniaturized spectrometers, which have many of the spectral capabilities of laboratory-based systems.

These developments, thus, also paved the way for the potential use of Raman spectroscopy in planetary exploration as part of rover or lander instrumentation, in particular for the exploration of Mars. Furthermore, at present, Raman spectroscopy is considered as a fundamental, next-generation instrument for the characterization of mineralogical and organic material during the exploration of Mars [16].

However, even if the technological development of the technique has reached a high-enough level to be considered for planetary exploration, there are still many other aspects related to the operation of the instrument and the processing of the results that need to be addressed.

In effect, being Raman spectroscopy a laboratory technique, the presence of a human operator has always been a constant in the operation of a Raman spectrometer, as well as for the analysis of the resulting spectra. However, possible planetary missions with this technique will have to operate in an autonomous way, without the presence of a human operator. This means that brand new algorithms have to be developed to provide the system with parameters based on which the system can take autonomous decisions which lead to maximizing the scientific return from the instrument.

For example, the acquisition parameters issue. An individual Raman spectrum is typically acquired based on two parameters: Integration time ( $T_i$ ) and Number of Accumulations ( $N_a$ ). The integration time is the time during which the spectrometer CCD is acquiring photons from the collection fiber of the spectrometer, and the number of accumulations defines how many acquisitions with  $T_i$  are averaged, to increase the SNR of the final spectrum. These two parameters have different optimum values from one spectrum to another, even in the same sample (except if both are taken in the very same spot), so they have to be adjusted at each new acquisition.

The integration time should be set to a value which maximizes the number of photons received at the spectrometer (the higher  $T_i$ , the higher the number of received photons) but which do not saturate the CCD. On the other hand, the number of accumulations reduces the noise in the final spectrum by averaging several of them. By averaging, the random noise generated during the acquisition will be somewhat cancelled. However, the noise decrement will be smaller with each new accumulation, and thus a trade-off between time and reachable SNR has to be reached.

Typically, the standard Raman operation would comprise a human operator which, looking at reference spectra acquisitions with 'random'  $T_i$  and  $N_a$ , optimizes the acquisition parameters based on his experience, to obtain the best possible final

spectrum. Some factors that the operator can take into account can be: maximum and minimum intensity of the spectra, background shape, random noise level, etc.

Providing the instrument with capabilities to realize this operation is of utmost importance if it is going to operate autonomously, and thus, a huge effort has to be made to develop a set of algorithms which can respond to all the different needs of the Raman instrument operation in an optimum way to achieve the best possible science from it.

Following, a list of some possible capabilities that the instrument has to be provided with is shown:

- Fluorescence detection and removal. Some samples have a fluorescent response that can conceal the Raman response due to their higher intensity with time. However, this fluorescence decays with time, with different rates depending on the sample. The instrument will have to decide for how long it will wait before acquiring the final spectrum.
- Cosmic ray detection and removal. Detect and erase random peaks which can appear during acquisition.
- Acquisition parameters adjustment. It needs, among other things, to evaluate the SNR of the spectrum (background subtraction, noise evaluation and peak intensity calculation).
- Detection of astrobiologically interesting traces in the final spectrum.

In addition to providing the instrument with all the capabilities to perform autonomous and optimized spectral acquisitions, it is also important to develop algorithms to extract the maximum information from the instrument configuration. That is, given that the RLS instrument will acquire several points along a line on a homogeneously mixed powdered sample, it is important to investigate if there is potential information that can be extracted from a statistical treatment of all the acquired spectra on a sample.

### **1.7.1. On-board Raman spectrum processing: Detection of traces of astrobiological interest**

From all the algorithms that have to be developed, the present project has focused on the detection of astrobiological traces of interest. When an especially important feature is detected in a spectrum, it will be marked as high priority spectrum to be sent to earth in the first place, allowing the scientists to take tactical decisions based on the most relevant results first.

A typical Raman spectrum and its characteristics can be seen in Figure 11. The x axis represents the Raman shift (wavenumber of the Raman vibration minus the laser one). The units for the intensity of the peaks do not directly correspond to the number of photons received by the CCD due to inefficiencies of the devices as well as the mediation of an ADC (Analog to Digital converter), and this is why the counts are measured in a. u. (arbitrary units). The laser excitation is always positioned at  $0\text{ cm}^{-1}$ , and filtered by a notch filter which weakens the light received by the spectrometer in

several orders of magnitude. Usually, Raman signals are often swamped or obscured by the background or baseline, due to fluorescence from organic molecules and contaminations, and do not provide any useful information for the Raman spectrum. Thus, the peak intensity of the Raman peaks is defined as the number of counts at the position of the peak, minus the counts of the baseline or background. In order to simplify the interpretations of the spectra, determine the peak positions, estimate the peak width, and measure the peak intensity accurately, the elimination of the background is necessary [17].

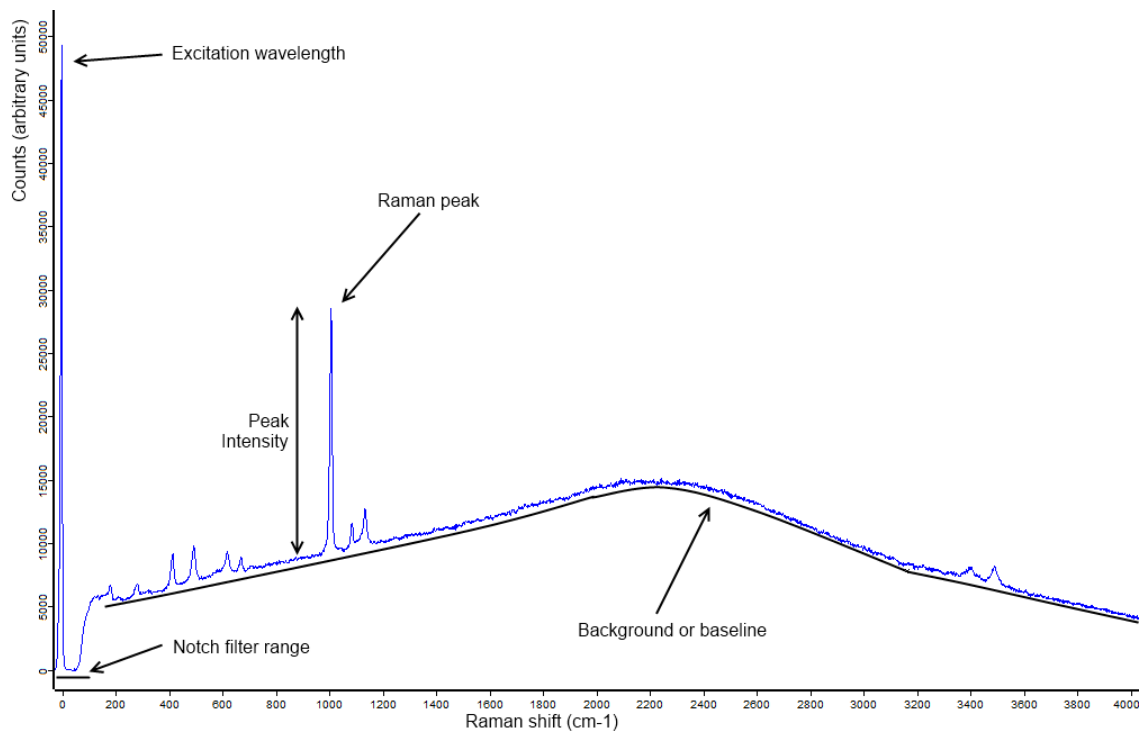


Figure 11. Typical Raman spectrum

The detection of traces of interest algorithm requires that the instrument is capable of subtracting the background of the spectrum, as well as detecting its peaks, with the minimum number of operations and memory usage. By comparing the detected peaks to several important astrobiological indicators, the acquired spectrum can be marked as high priority.

The research here presented evaluates two different algorithms for the detection of the baseline, optimizing the parameters for all of them, and comparing their performances. For the peak detection another algorithm is proposed, evaluated and optimized, and the performance of both the baseline removal and peak detection algorithms are evaluated jointly. This issue is presented in Chapter 2.

### 1.7.2. On-earth data analysis: Pseudo-quantification of mineral species from Raman spectra

To maximize the scientific return of the instrument, it is important to adapt the data analysis techniques to the concrete operation mode of the flying instrument jointly



with the rover. In this case, as the acquisition is being performed in several points of a homogeneously mixed powdered sample, a study has been realized to evaluate the possibility of obtaining data from a set of spectra on the same sample, rather than on one spectrum alone.

Typically, Raman spectroscopy is used to identify the mineral phases that can be found in a sample. However, the intensity ratios between the peaks belonging to one mineral phase and the others do not usually provide information about the abundance of each of them, especially due to the fact that the Raman scattering efficiency can be very different (several orders of magnitude) between different mineral phases. By statistically analyzing all the several spectra obtained during the instrument operation, the question to solve is if a pseudo-quantification of the abundances of the mineral species would be possible. Chapter 3 gathers some results to this respect.

## Capítulo 2. Resumen: Detección de bandas espectrales de interés astrobiológico (English follows)

### 2.1. Concepto y metodología

Es sabido que el tiempo de transmisión entre la tierra y marte es de 14 minutos. Debido a esta restricción, unida a la presencia intermitente de los satélites que orbitan marte para su uso como repetidores, y las posiciones relativas de los planetas, la capacidad de ponerse en contacto directo con el *rover* queda muy disminuida. Por tanto, se necesitará que el vehículo y el instrumento sean capaces de actuar de manera autónoma y, además, que las ventanas abiertas de transmisión se aprovechadas al máximo, enviando en primer lugar aquellos datos que se consideran más relevantes.

El algoritmo de detección de bandas espectrales de interés pretende dotar al instrumento RLS con la habilidad de detectar puntos de especial interés, a los que se otorgará mayor prioridad y por tanto preferencia a la hora de enviarse a tierra.

El **desarrollo del algoritmo** sigue la aproximación estándar de eliminación de la línea de base en un primer paso, y detección de los picos y regiones de interés en un segundo paso.

Los algoritmos estudiados en este trabajo se eligieron entre otros muchos debido a su simplicidad teórica y buen comportamiento en términos de número de operaciones y uso de memoria.

#### Regiones de interés astrobiológico

Estas regiones se definen teniendo en cuenta la frecuencia con la que vibrarán las emisiones Raman de orgánicos o materiales astrobiológicamente relevantes. El criterio seguido para la asignación de regiones de interés se ha basado en las vibraciones Raman de los enlaces Carbono-Carbono (C=C), Carbono-Hidrógeno (C-H), Carbono-Oxígeno (C=O) y bandas de hidratación (O-H libre y H<sub>2</sub>O). Estas cuatro regiones no se solapan entre sí, como puede verse en la Figura 12.

La **metodología general** para el análisis y evaluación de los diferentes algoritmos propuestos, se utilizó un conjunto de 113 espectros extraídos de la base de datos de la Unidad Asociada UVa-CSIC-CAB. Estos espectros fueron procesados manualmente en primera instancia para extraer su línea de base y regiones de interés, y de esta forma los resultados obtenidos por los algoritmos pueden compararse a algo.

Todos los algoritmos propuestos se optimizaron para obtener los mejores resultados posibles para el conjunto de los 113 espectros, y se describen en las siguientes secciones.

## 2.2. Eliminación de la línea de base

### 2.2.1. Introducción

Varios métodos para la corrección automática o semi-automática de la línea de base han sido encontrados en la literatura, con diversas aproximaciones, desde análisis en el dominio de la frecuencia con transformadas de Fourier o *wavelet* para eliminar componentes de baja frecuencia, uso de derivadas para distinguir entre línea de base y bandas espectrales, el uso de análisis estadístico o técnicas multivariantes, o ajustando la línea de base por una función polinómica.

Sin embargo, la mayoría de los algoritmos que se encuentran en la bibliografía no prestan una especial atención a la complejidad computacional del algoritmo o su uso de memoria, puesto que estas soluciones suelen plantearse para el análisis de datos múltiple con equipos de laboratorio. Por tanto, estas aproximaciones no son aplicables en general en el marco de la exploración planetaria, especialmente dadas las limitadísimas capacidades del microcontrolador del instrumento RLS.

### 2.2.2. Descripción de los algoritmos

#### *Algoritmo Simple baseline (Alemanes)*

Este algoritmo, cuya idea nos fue propuesta por espectroscopistas de la Universidad de Jena (Alemania), originalmente establece la línea de base como el propio espectro, e iterativamente reduce los puntos de la misma basándose en los *delta* siguientes puntos, promediando el valor actual de línea de base con los valores en dichos puntos. El parámetro *delta* debe establecerse con un valor tal que cubra la banda Raman más ancha que uno espere encontrarse.

Uno de los principales problemas de las aproximaciones basadas en ventana (como ésta) es que el parámetro *delta* necesita ajustarse dependiendo del tipo de espectro bajo análisis (de la anchura de sus picos). Para intentar minimizar esta dependencia, proponemos una mejora que consiste en utilizar dos ventanas (parámetros *delta1* y *delta2*) en lugar de una sola. Esto se justifica por el hecho de que, típicamente, los picos Raman más anchos se encuentran en las regiones espectrales más elevadas (parámetro *delta2*), mientras que los estrechos se encuentran en las regiones más bajas (donde se usará el parámetro *delta1*, con  $\text{delta2} > \text{delta1}$ ). Para definir la posición espectral en la que el valor del parámetro de ventana cambia de *delta1* a *delta2*, se evaluó el algoritmo con diferentes combinaciones de parámetros, y teniendo en cuenta que los picos anchos que se encuentran en posiciones más bajas del espectro (uno de los picos del hematite y también la banda D del grafito) no van por debajo de  $1300 \text{ cm}^{-1}$ . Con todo esto, el valor óptimo se calculó en  $1100 \text{ cm}^{-1}$ .

Las Figuras 13 y 14 ejemplifican el problema. La Figura 13 muestra un resultado con el algoritmo optimizado para un solo parámetro, mientras que la Figura 14 muestra cómo mejora la estimación de la línea de base introduciendo un segundo parámetro de ventana.

Este algoritmo tendrá por tanto que optimizar los dos parámetros *delta1* y *delta2*.

### Algoritmo Weakley

Este algoritmo propuesto en la bibliografía localiza mínimos en regiones equiespaciadas del espectro, tomando la primera derivada del espectro de mínimos, y generando un espectro diferencia de segundo orden. Para decidir si un mínimo pertenece o no a la línea de base se define una función umbral adaptativa basada en análisis estadístico. A diferencia de otras propuestas, no hace asunciones sobre el tipo de espectro bajo análisis, lo cual lo convierte en un posible candidato para el instrumento RLS.

Este algoritmo también fue modificado con el criterio de doble ventana, dejando un total de tres parámetros para optimizar: *window1* (ventana en el rango espectral bajo), *window2* (ventana en el rango espectral alto), y *tolerance* (parámetro de la función umbral).

### 2.2.3. Metodología

La evaluación de los algoritmos pretende conocer la bondad de los mismos así como su robustez.

Para evaluar la **bondad**, los parámetros del algoritmo han de ser optimizados para minimizar el error global de los resultados. El criterio para la optimización de los parámetros consiste en minimizar el MSE (*Mean Square Error*) de la diferencia entre la línea de base calculada por el algoritmo y la línea de base ideal u objetivo. La línea de base objetivo fue calculada por un espectroscopista.

El procedimiento de optimización consiste en ejecutar el algoritmo con diferentes combinaciones de parámetros, calculando el MSE en cada caso. Por tanto, para un algoritmo con dos parámetros, se calculará una matriz de tamaño  $m \times n$ , con  $m$  y  $n$  el número de valores aplicados a los parámetros barridos. En el caso de hacer un barrido sobre tres parámetros se obtendrá un cubo de datos tridimensional. La Figura 15 muestra una representación de una matriz de error bidimensional, donde la configuración óptima se observa en la zona azul oscura.

Se entiende por **robustez** la habilidad del algoritmo para comportarse correctamente con nuevos espectros no utilizados para la optimización de los parámetros. Para evaluarla, y a pesar de que un entrenamiento basado en 113 espectros podría considerarse estadísticamente válido, se aplicó un proceso de validación cruzada *Leave-One-Out* (LOOCV). De esta forma, se ejecutan 113 iteraciones de entrenamiento con 112 espectros, y validación con el sobrante. De esta forma se puede hacer una estimación grosera de la robustez de los algoritmos.

### 2.2.4. Discusión de resultados

En primer lugar, para obtener una referencia de valores de MSE para cada algoritmo, estos se optimizaron para cada espectro por separado, obteniendo los resultados de la Tabla 1. Claramente puede observarse como el algoritmo Alemanes tiene un comportamiento superior utilizando dos parámetros de ventana con respecto a la propuesta original con uno solo. En el caso del algoritmo Weakley, el comportamiento es equivalente, por lo que cabría decir que no hay mejora al utilizar dos ventanas en

lugar de una sola. Además, comparando los algoritmos entre sí, se puede ver que el algoritmo Weakley funciona mejor que el Alemanes.

Sin embargo, los parámetros no deben ser ajustados para cada espectro, sino que hay que buscar la optimización global de los mismos (Tablas 2 y 3), promediando las matrices MSE obtenidas para todos los espectros. Los errores en este caso se pueden observar en la Tabla 4. La conclusión es que, globalmente, ambos algoritmos tienen un comportamiento muy similar. También se concluye que la modificación de doble ventana supone una mejora sustancial con respecto a la propuesta con una sola ventana (aproximadamente un 20% y 13% más de error en el caso de una sola ventana, para Alemanes y Weakley, con parámetros óptimos 290 y 50, respectivamente).

En segundo lugar, los resultados obtenidos al evaluar el MSE de los espectros de test del procedimiento LOOCV se encuentran en las Tablas 5 y 6, donde puede verse que el algoritmo Alemanes obtiene valores muy cercanos a los obtenidos anteriormente, mientras que el algoritmo Weakley presenta un mayor error. Además, la desviación de los parámetros óptimos presenta para este segundo algoritmo valores más elevados, con excursiones máximas también bastante elevadas del valor de los parámetros (hasta un 50%).

## 2.3. Detección de los picos del espectro

### 2.3.1. Descripción del algoritmo

La detección de picos se realiza sobre el espectro al que se le ha eliminado la línea de base previamente, luego este algoritmo se ejecutará a continuación del de eliminación de la línea de base.

#### *Restricciones del algoritmo*

Una de las características del algoritmo de detección de picos más importante es que el funcionamiento tiene que ser autónomo, o no supervisado ya que habitualmente se necesita la interacción con el usuario para ajustar algún parámetro umbral, que variará dependiendo del espectro bajo estudio. En un escenario de exploración planetaria no es posible considerar la interacción con el usuario. Por tanto, nuestra propuesta es calcular dicho umbral basándose en características espectrales de cada espectro. En concreto, dos parámetros espectrales que utilizamos son la intensidad del pico máximo y el ruido del espectro.

Otra condición importante en este caso es que no se produzcan falsos positivos, es decir, no se detecten picos en regiones donde no los hay. De esta forma se minimiza la probabilidad de que las muestras potencialmente interesantes no se traten de errores de software. Además, es deseable que el número de falsos negativos sea el menor posible, siempre que no se hayan producido falsos positivos.

### *Propuesta de algoritmo*

Una propuesta típica basada en cálculos matemáticos es aquella basada en métodos de segunda derivada para detectar los picos, que funcionan muy bien con espectros con elevada SNR. Sin embargo, el ruido afecta de manera importantísima, tendiendo a dar falsos positivos.

Otra típica propuesta para la detección de los picos está basada en la definición de un umbral por encima del cual todos los máximos locales detectados se consideran como picos. El problema principal consiste en la definición de dicho umbral, que será diferente para cada espectro.

En el caso bajo estudio, para cumplir con los requisitos propuestos al algoritmo, proponemos aplicar la aproximación estándar con umbrales dependientes de características espectrales del propio algoritmo (intensidad del pico máximo y ruido del espectro). Además, esta propuesta contempla la definición de dos tipos de parámetros. Por un lado, la intensidad mínima de cada pico, definida como el valor mayor de entre un determinado porcentaje del pico máximo y un valor dependiente del ruido del espectro. Con esta condición se reduce la probabilidad de que picos secundarios débiles puedan dar falsos positivos (han de ser mayores que un cierto porcentaje del pico máximo), y también se evita detectar falsos positivos en espectros muy ruidosos (han de ser mayores que un valor dependiente del ruido). Un ejemplo de ambas situaciones puede verse en la Figura 17, con umbral (línea verde) basado en la intensidad del pico máximo y 18, con umbral (línea azul) basado en el ruido del espectro. En este último caso se ve cómo la condición impide al algoritmo detectar ningún pico, ya que el espectro es muy ruidoso y no habría suficientes garantías.

Por otro lado, se define un segundo parámetro que servirá para incrementar iterativamente el umbral por encima del cual se detectan los picos. Este incremento se calcula como el máximo de los dos valores siguientes: un factor de la desviación estándar de la intensidad del ruido (para evitar detectar fluctuaciones de ruido como si fueran picos), y un 5% del pico máximo (para limitar el número de iteraciones a 20). La finalidad de esta iteración es poder detectar dobles y triples picos, ya que el algoritmo solamente marca como pico el máximo de todos los valores adyacentes que se encuentran por encima del umbral. La Figura 19 muestra un ejemplo de detección de un doblete: hasta que no se itera y se define el umbral como la línea rosa, el pico de  $1360\text{ cm}^{-1}$  pasará inadvertido, ya que no es el máximo de todos los puntos que se encuentran por encima del umbral verde.

#### **2.3.2. Metodología**

El criterio para optimizar los dos parámetros (porcentaje del máximo pico y factor que multiplica a la desviación estándar de ruido) es que no se produzcan falsos positivos, e intentar minimizar el número de falsos negativos. La Figura 20 muestra las matrices de falsos positivos y de falsos negativos para las distintas combinaciones de parámetros. Como era de esperar, valores muy altos de los parámetros elevarán los umbrales, reduciendo el número de falsos positivos pero aumentando el falsos negativos, mientras que valores muy bajos provocarán el efecto contrario.

Al aplicar el criterio tomando el valor que proporciona cero falsos positivos y el mínimo número de falsos negativos, se está eligiendo una configuración que no es la más robusta. Es más, se ejecutó un procedimiento de validación cruzada LOOCV y se vio que, para algunos conjuntos de espectros, los parámetros calculados hacían que el algoritmo diera falsos positivos para los espectros no utilizados en el entrenamiento, lo cual es inaceptable.

Para resolver la situación, los parámetros se relajan (incrementan) para aumentar la probabilidad de que nuevos espectros no resulten en falsos positivos.

Las condiciones que definen las regiones espectrales a detectar se refieren a los rangos espectrales con diferentes tipos de enlaces químicos: C=C, C=O, C-H e hidratación (H<sub>2</sub>O u OH) (ver Figura 12).

Para obtener los resultados y compararlos, se ejecuta el algoritmo sobre los espectros con corrección manual de línea de base, así como sobre los espectros corregidos automáticamente por los algoritmos Alemanes y Weakley propuestos anteriormente.

### **2.3.3. Resultados y discusión**

Los resultados se muestran en la Tabla 7. Puede verse cómo los resultados con los espectros óptimos (corregidos manualmente) el número de regiones de interés no detectadas asciende al 27%. Es un valor relativamente elevado debido al compromiso adquirido entre robustez y número de falsos negativos. Esto también afecta a los resultados obtenidos con los algoritmos de línea de base, donde el empeoramiento del error del 6% es el doble del que se obtendría con los parámetros óptimos.

## Chapter 2. Detection of spectral bands of astrobiological interest

### 2.1. Concept and methodology

The RLS instrument, in its current operation mode, will analyze a minimum number of 20 points of the same homogenized powdered sample, and then, at some point, send the results back to Earth. However, it is well known that the communication with the red planet is not real-time, but there is a travelling time for signals of 14 minutes. Furthermore, mission and operation restrictions, as well as blackouts due to the relative positions between the Earth, Mars and the data relays orbiting the red planet, will make that the information transfer between the control center and the rover be slow and gapped. This will usually provide the rover commanding team with very short time windows to take tactical decisions.

So, to ensure that the tactical decisions are best chosen, it is of utmost importance that the most relevant information reaches the scientific team in the first place, so they can take decisions for commanding the rover based on that information, when it is not possible to receive all the information in advance. For example, if a very important sample has been found on a determined outcrop, it would be a tactical error to send the rover away instead of performing more experiments at the same site.

The standard operation mode, by itself, lacks the capability of establishing differences among the different points of the sample, and thus it gives all of them the same priority and importance.

The algorithm to detect spectral bands of interest is intended at providing the RLS instrument with the ability to detect potential points of interest along the line. This way the instrument will have the ability to establish higher priorities to those spectra which are considered to be astrobiologically interesting, commanding then the rover to upload them to Earth first.

#### 2.1.1. Algorithm development

To develop this kind of algorithm, the best approach is to divide it in two steps: First, the spectrum baseline is estimated and subtracted, and second, the peaks are detected on the spectrum without baseline [18].

During the development of the RLS ExoMars Simulator, a series of algorithms were implemented, including some baseline calculations and basic peak detection, to estimate the SNR of a spectrum, providing the system with means to calculate the acquisition parameters. However, those algorithms were created only to verify if the automation of the system was possible. They will only work under some restrictions (they need very noisy spectra, for instance), and, furthermore, those algorithms were not optimized in terms of operations or memory usage.

The algorithms studied for this project were chosen among many others due to their theoretical simplicity and good performance, which means low number of operations and memory usage. These will not only be applicable to the detection of traces of



interest, as, for example, the baseline correction algorithms can also be used for the estimation of the SNR during the acquisition parameters adjustment.

### 2.1.2. Regions of astrobiological interest

The parameters which make a Raman spectrum interesting or not are, in this case, related to the possibilities of finding a sample related to extant or past life, as well as with water related mineral activity. Hence, the importance of a spectrum is based on its spectral characteristics, i.e., having peaks at certain spectral regions, due to the fact that some determined molecular bonds will vibrate at a determined frequency, in which no other vibrations are found (Figure 12).

For instance, vibrations at interesting regions would be any of those forming part of organic material, as the so-called  $\nu$  band of the Carbon-Carbon bond (C=C) (around  $1600\text{ cm}^{-1}$  for organic material, and both from  $1335$  to  $1380\text{ cm}^{-1}$  and from  $1550$  to  $1586\text{ cm}^{-1}$  for graphitic carbon), the Carbon-Hydrogen bond (C-H) ( $2800 - 3100\text{ cm}^{-1}$ ) or the Carbon-Oxygen bond (C=O) ( $1680 - 1820\text{ cm}^{-1}$ ). Another important vibration is the Oxygen-Hydrogen one (O-H), which might denote the presence of hydrated minerals ( $3100 - 3650\text{ cm}^{-1}$ ) [19]. The carbonate ( $1080-1100\text{ cm}^{-1}$ ) and sulfate ( $980-1000\text{ cm}^{-1}$ ) spectral regions are also somewhat interesting, as they can be indicators of past hydrothermal activity, but do not necessarily have a close relation with organic material, and thus have not been included as interesting regions to analyze.

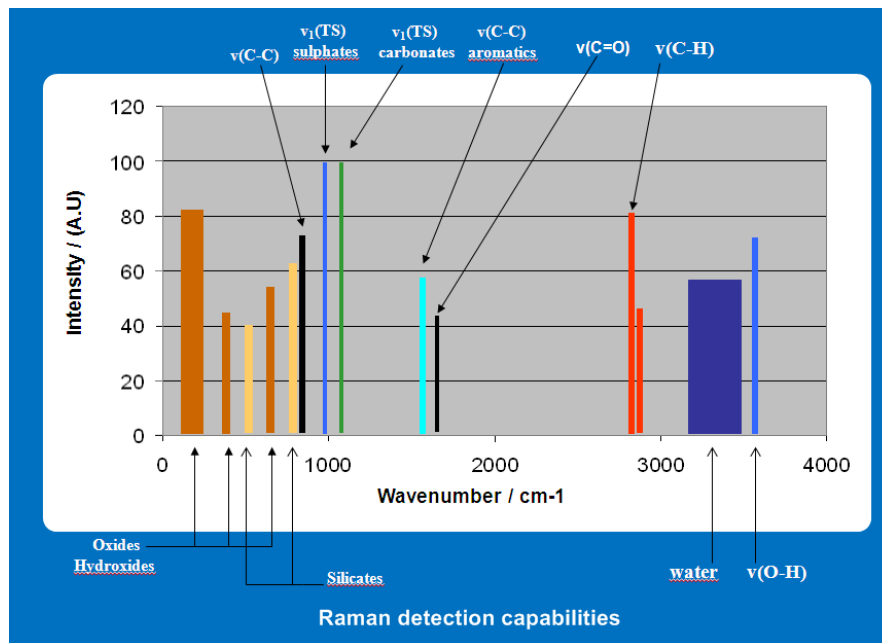


Figure 12. Raman spectral ranges by functional group

### 2.1.3. General methodology

For the analysis and evaluation of the different proposed algorithms, a set of 113 spectra extracted from the Unidad Asociada UVA-CSIC-CAB database [20] were used. All these spectra were manually processed to extract their baseline and regions of

interest, and their results were compared to those obtained by the different algorithms.

Several algorithms for detecting the baseline were proposed, modified or implemented, and another one proposed for the identification of the regions of interest. All of them were then optimized to obtain the best possible results, comparing with the objective results. These algorithms are described in the following section.

## **2.2. Baseline removal**

### **2.2.1. Introduction.**

The characteristically narrow bands observed in Raman and infrared spectra are typically superimposed on broad, low frequency components referred to as baseline or background [21].

Methods of semi-automatic and automatic baseline correction have been reported and vary in complexity and computational rigor [22]. Methods derived from the field of digital signal processing employ such techniques as Fourier or wavelet transformations to remove low-frequency components [23, 24]. Other methods include applying first and/or higher order derivatives to identify and distinguish between baseline and spectral bands [25, 26], the use of statistical and probabilistic approaches such as principal components analysis or maximum entropy algorithms [27-29], or assuming that the baseline has a continuous functional form, usually a low-order polynomial, in which case a routine is employed that gradually “strips” the baseline from the raw spectrum [30, 31]. In general, methods tend to employ a combination of the aforementioned techniques depending on application and need [32, 33].

However, none of the previous studies pay attention to the algorithm complexity in terms of the processor load and memory usage, due to the fact that they are usually focused on the baseline analysis in automated environments for laboratory equipment, where such restrictions are not found. The approach that the RLS instrument implement has to take this issue into account. Unhappily, many of the algorithms found in the literature are processor consuming by nature (Fast Fourier Transforms, for example, will not be suitable for the instrument microprocessor) or because they perform heavy iterative approaches to the final result [34, 35], thus drastically increasing the number of operations. This kind of algorithms cannot be used in a planetary exploration framework, due to the very limited capabilities of the RLS instrument microprocessor.

This study proposes a low number of operations and low memory consumption approach for the baseline calculation, and also proposes a modification and evaluation of the derivative based proposal by [22].

## 2.2.2. Algorithm description

### *Simple baseline algorithm (Alemans)*

The Simple baseline algorithm (to which we usually refer as Alemans), is a very simple window based algorithm proposed to us by some colleagues at the University of Jena (Germany), which originally establishes the baseline as the spectrum itself, and then iteratively reduces the baseline based on the *delta* following values, by averaging the current baseline value with the one in those points. The *delta* parameter should be defined as the number of points needed to cover the broadest Raman band one expects to measure.

One of the main problems with all the window based approaches is that the *delta* parameter needs, a priori, to be adjusted depending on the type of spectrum under analysis (mainly, depending on the width of peaks). To this respect, an important improvement has been proposed and applied to all the algorithms under study. On one hand, the baseline calculation algorithms which work correctly for thin peaks do not work well for the wide ones, because they take the wide peak as baseline, and sometimes the contrary also happens, when the algorithm detects an only baseline region as a peak. On the other hand, we observed that the wide Raman peaks are located in higher parts of the spectrum than the thin ones. So our contribution is to add a threshold in the spectrum in which the *delta* parameter changes. This way, instead of having a fixed *delta* value for the whole spectrum, the algorithm will use a *delta1* value for the low part of the spectra, and a *delta2* for the higher part. By evaluating the algorithm with several combinations of the Raman shift limit between regions, and taking into account that the lowest wide peaks (belonging to the graphite D band or one of the hematite peaks) are placed around  $1300\text{ cm}^{-1}$ , the optimal Raman Shift limit was found to be at  $1100\text{ cm}^{-1}$ .

This issue is represented graphically in Figures 13 and 14, and quantitatively in section 2.2.4. Figure 13 shows how the optimized algorithm for one parameter deviates largely from the objective baseline, while in Figure 14, with two different window parameters, is much closer to the desired baseline.

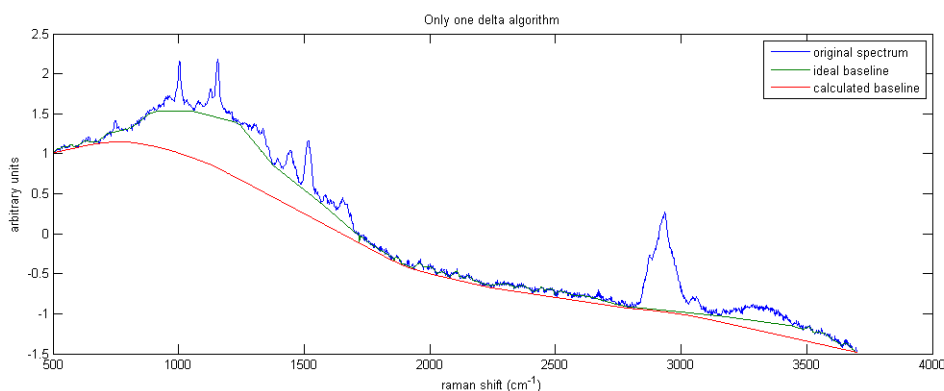


Figure 13. Simple Baseline detection with only one delta parameter for the whole range

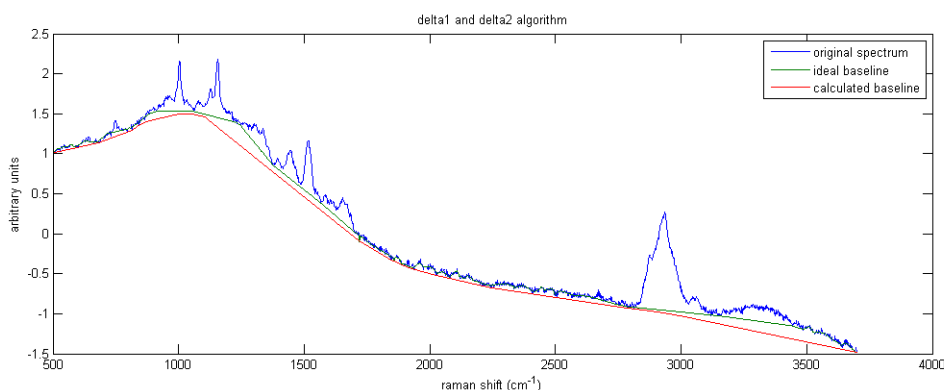


Figure 14. Simple Baseline detection optimized for a  $\delta_1 + \delta_2$  algorithm

Being this criterion theory-based, it is possible to apply it to any other window based algorithm as the Weakley one, with the certainty that it will improve the results.

So, this algorithm proposal has both parameters  $\delta_1$  and  $\delta_2$  to be optimized.

### **Weakley algorithm** [22]

This algorithm proceeds by locating minima in equally spaced regions of the spectrum, taking the first derivative of the intermediate minima-only spectrum, generating a second-order difference spectrum, and then developing a statistically based, adaptive thresholding scheme to reject or accept minima points as part of the baseline. Once the baseline points are located, simple linear interpolation is employed to remove the baseline from the raw spectrum. Unlike most derivative methods, this technique makes few assumptions about the positivity, smoothness, band shape, or noise level in the raw spectrum. In fact this method is “model free” insofar as it assumes that the perturbing baseline has no global functional form but contains relatively smooth local variations. So, this algorithm can be considered as a possible candidate for the RLS instrument.

This algorithm was also modified to include the two-region spectral range as described in the previous section. The parameters to optimize in this case are three: *window1* and *window2* (corresponding to the window width in the different spectral regions) and the *tolerance* parameter (threshold parameter to decide whether a minimum is baseline or not).

### **2.2.3. Methodology**

#### **Algorithm goodness**

The methodology followed for the evaluation of the implemented algorithms consists in the optimization of the algorithm parameters to provide the best possible results compared to those considered as objective.

The criterion for the minimization of the error is to minimize the Mean Squared Error (MSE) of the difference between the calculated baseline and the objective one, for the 113 spectra in the training set. The objective or ideal baseline was manually calculated

by a human operator, and downloaded from the spectra database. The algorithm parameters are swept with different values, and the MSE is calculated for each combination of parameters. This way, when sweeping two parameters 1 and 2, an  $m \times n$  MSE matrix is obtained, with  $m$  and  $n$  the number of different values applied to parameter 1 and 2, respectively. By searching for the minimum MSE value in the matrix, the optimum combination of parameter 1 and 2 can be found. In the case that three parameters are swept, a three-dimensional array of MSE values is calculated. Figure 15 shows a graphic representation of the MSE values in a two-dimensional matrix (two parameters). The minimum value is somewhere around the dark blue area.

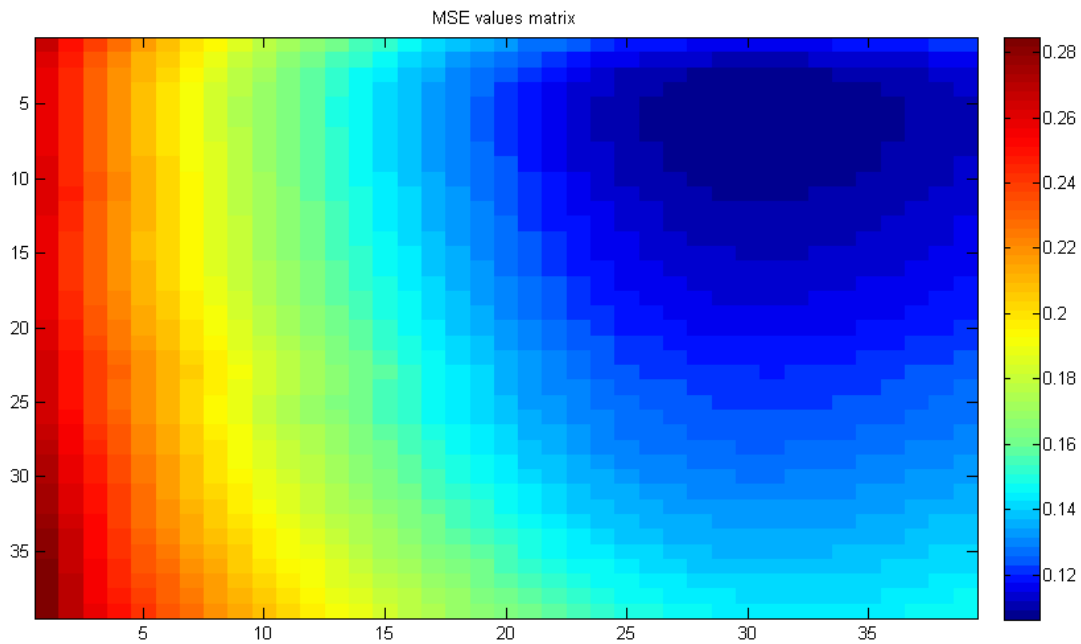


Figure 15. Example of an MSE values matrix, for two swept parameters.

### *Algorithm robustness*

The robustness of an algorithm is understood as its ability to behave correctly when a new input that has not been used to optimize the algorithm parameters is used.

Even if the training test is formed by 113 spectra, which probably is a statistically high enough number to avoid overfitting, to ensure the accuracy of the optimization and that the optimized parameters perform correctly for non-training spectra, a Leave One Out Cross-Validation (LOOCV) procedure is used. This way, one of the 113 spectra is left out, and the rest are used for the optimization of the parameters. The average MSE of the 112 training spectra provides a set of optimum parameters with which the algorithm is executed for the single spectrum, and its MSE is calculated. By leaving out each of the 113 spectra, an array of 113 MSE values, as well as 113 sets of parameters are obtained. Using this method, a gross estimation of the robustness of the algorithms can also be obtained, in addition to their goodness.

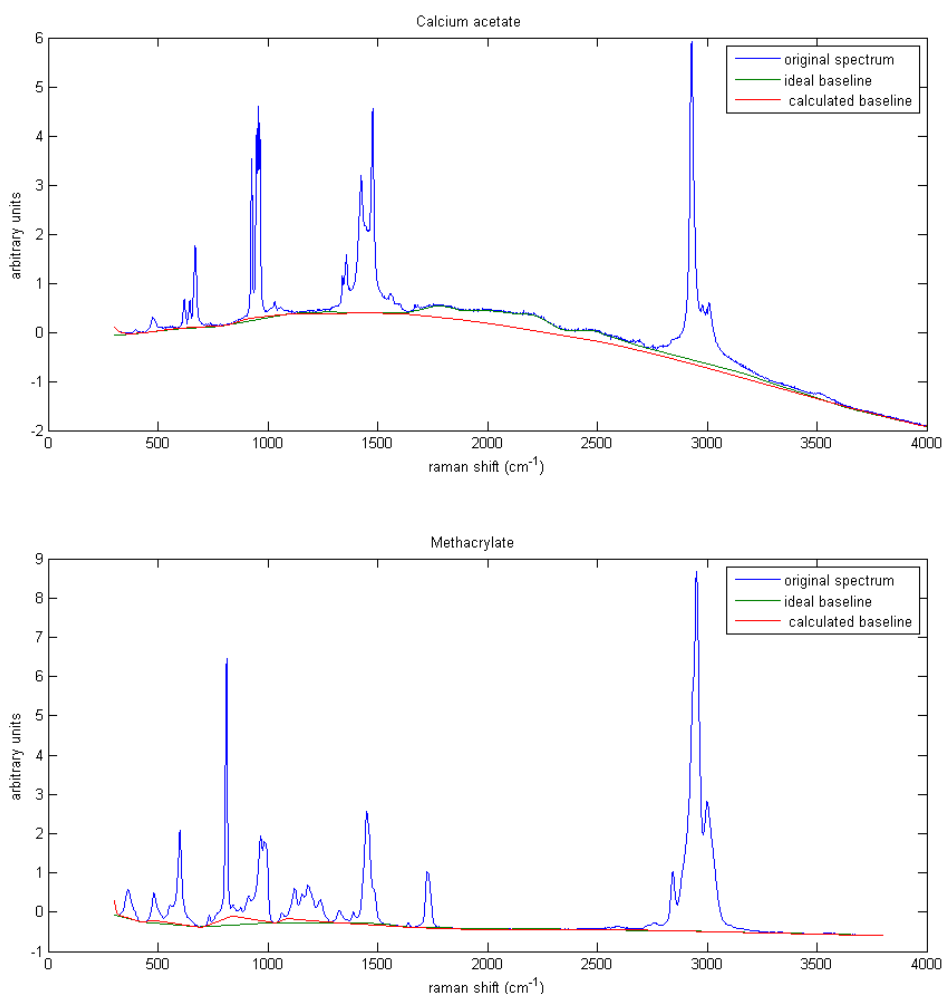


Figure 16. Examples of baseline detection for the optimized parameters

Figure 16 shows a couple of examples of baseline detection with the optimized parameters for the Alemanes algorithm.

## 2.2.4. Results Discussion

### Algorithms goodness

To provide a reference magnitude to which compare the obtained MSE values, the algorithm parameters were optimized for each spectrum individually to obtain the best possible result for each of them. Table 1 shows the average and standard deviation of the obtained MSE values for the original algorithms, and also for the double window versions. The Alemanes algorithm clearly shows a better performance with two parameters, while for the Weakley one, the behavior is exactly the same (this means *window1* and *window2* have the same value). With this result we could be tempted to say that the one parameter Weakley algorithm is better than the two parameter one. Furthermore, in the sight of these results, it seems that the modified Weakley algorithm can be more accurate than the Simple Baseline one when optimizing the algorithm parameters for the spectrum under study.

However, the parameters cannot be adjusted for each spectrum, but they should be fix. To decide these values, the MSE matrices of the 113 spectra were averaged,

obtaining the optimum parameters globally. These values are shown in Table 2 and Table 3, and minimize the overall error of the 113 spectra. With these parameters, the resulting MSE obtained is shown in Table 4. The conclusion is that, globally, both algorithms have a very similar performance, with very similar MSE averages and standard deviations, and also that the double window modification provides overall better results than the original proposals (around 20% and 13% more error for the original proposals, with the Alemanes and Weakley algorithms, respectively).

For the Alemanes one window algorithm, *delta* was 290, and for the modified Weakley, *window* was set to 50.

MSE (individual best parameters)	Alemanes (original)	Alemanes ( <i>delta1, delta2</i> )	Weakley (original)	Weakley ( <i>window1, window2</i> )
Average	0.0740	0.0640	0.0513	0.0513
Standard deviation	0.1890	0.0775	0.0674	0.0674

Table 1. MSE value with the optimum parameters for each spectrum

Global best parameters (Alemanes)	
Delta 1	90
Delta 2	340

Table 2. Simple Baseline algorithm best parameters

Global best parameters (Weakley)	
Window 1	50
Window 2	180
Tolerance	0.9

Table 3. Weakley algorithm best parameters

MSE (individual best parameters)	Alemanes (original)	Alemanes ( <i>delta1, delta2</i> )	Weakley (original)	Weakley ( <i>window1, window2</i> )
Average	0.1260	0.1063	0.1138	0.1013
Standard deviation	0.1086	0.0948	0.1230	0.1093

Table 4. MSE value with the optimum parameters for all the spectra jointly

### Algorithms robustness

To evaluate the robustness of the algorithms, a Leave-One-Out procedure was followed, as explained in the previous section. The MSE values calculated for the outsider spectra are shown in Table 5 and Table 6.

It is plain that the Simple Baseline algorithm obtains values which are very close to the ones obtained previously, while the modified Weakley one has a steeper deviation. Furthermore, the deviation from the optimum parameters in the Simple baseline algorithm is below 2%, with also low values of maximum excursion among the 113 iterations. However, for the Weakley one the deviation is higher and the maximum excursion goes up to 50% for some of the parameters.

These results are good indicators that the modified Weakley algorithm is less robust and will probably obtain worse results when a non-training spectrum is used than the

Simple Baseline (Alemanes) one. This is also indicated somewhat by the difference between the MSE average in Table 1 (parameters optimized for each spectrum individually) and Table 4 (parameters optimized for all the spectra globally), and confirmed by the LOOCV results.

Alemanes LOOCV results	MSE	Delta 1	Delta 2
Average	0.1096	89.7	338.8
Standard deviation	0.0977	1.9%	1%
Max. Excursion	--	16%	3%

Table 5. LOOCV results for the Simple Baseline (Alemanes) algorithm

Weakley LOOCV results	MSE	Window 1	Window 2	Tolerance
Average	0.1210	50.3	179.5	0.87
Standard deviation	0.1380	3.3%	4.5%	10.7%
Max. Excursion	--	20%	50%	46%

Table 6. LOOCV results for the modified Weakley algorithm

## 2.3. Detection of the spectrum peaks

### 2.3.1. Algorithm description

Once the spectrum baseline has been calculated and subtracted, it is necessary to detect the peaks in the spectrum. For an SNR calculation algorithm, it would only be necessary to search for the spectrum maximum to obtain the maximum signal of the spectrum. In this case, however, it would be desirable to find all the peaks of the spectrum, or at least as many as possible. In this section a proposal for such an algorithm is described.

#### *Restrictions of the algorithm*

##### Unsupervised

Usually, the peak detection algorithms need that some threshold parameters be adjusted in order to obtain the best possible results. This is a good approach as long as interacting with the user is possible. For planetary exploration purposes, this cannot be met.

Thus, for the automatic and unsupervised detection of the spectral regions of interest, we propose that this threshold is calculated for each spectrum, as a function of its spectral characteristics. This way the unsupervised aspect of the decision can be met. Namely, the spectral characteristics refer to the maximum peak intensity, on one hand, and a parameter related to the spectrum noise, on the other.



### No false positives

A false positive is understood as a positive result from the algorithm in a spectral region in which no peak is present. It is mandatory that the algorithm do not detect false positives, i.e., that the algorithm is very robust in this sense. So, this is a very important restriction for the algorithm, to be sure that the potentially astrobiological interesting samples are not falsely detected.

### Minimum number of false negatives

A false negative happens when an extant peak in the spectrum is not detected by the algorithm. It is important that the algorithm is able to detect as many peaks as possible, but this condition is always subject to the previous one.

### *Algorithm proposal*

A typical theory-based approach to the detection of spectral peaks is based in the use of second derivative methods. These methods have proved to perform extremely well with very high SNR spectra. Unhappily, they are extremely sensitive to the spectrum noise [26], and thus are not valid for the RLS instrument purposes, as they tend to provide false positives.

Another typical approach for the automatic detection of peaks will be based on the definition of a threshold over which all the relative maxima detected are considered as peaks. The main problem of this approach is to define that threshold, as it will be different for different spectra. Typically, this kind of algorithm is used in scenarios in which some sort of interaction with the user is possible, as for example, for database recognition or in user oriented spectroscopy software as OPUS (see OPUS online Help in OPUS 2.06 software), where this kind of approach is referred to as standard.

To meet the “Unsupervised” and the “No false positives” restrictions, our proposal consists on applying the standard approach with spectrum dependent parameters, i. e. the thresholds of the algorithm will be defined as a function of the intensity of the maximum peak, as well as the spectrum noise. This way, there is no need to interact with the user to define the optimum parameter. In addition, our proposal defines two types of thresholds. On one hand, the minimum intensity of any peak. This threshold is defined as the highest of two values: a determined percentage of the maximum peak and a value dependent on the spectrum noise. Both values are optimized for the training spectra set. With this two restrictions it is ensured that weak secondary peaks which can appear in some spectral regions of interest without really being so are excluded (peaks have to be higher than a percentage of the maximum peak), and also that in very noisy spectra (in which noise peaks are comparable to Raman peaks) no noise peak is considered as Raman peak. For example, Figure 17 represents a threshold determined as a fraction of the maximum intensity peak (in green). This way, no peak is detected below this limit, concealing potential undesirable secondary peaks. Figure 18 is a good example of the contrary situation. Noise is so high that the threshold (cyan line) is set to a value which prevents the algorithm from detecting any peak at all.

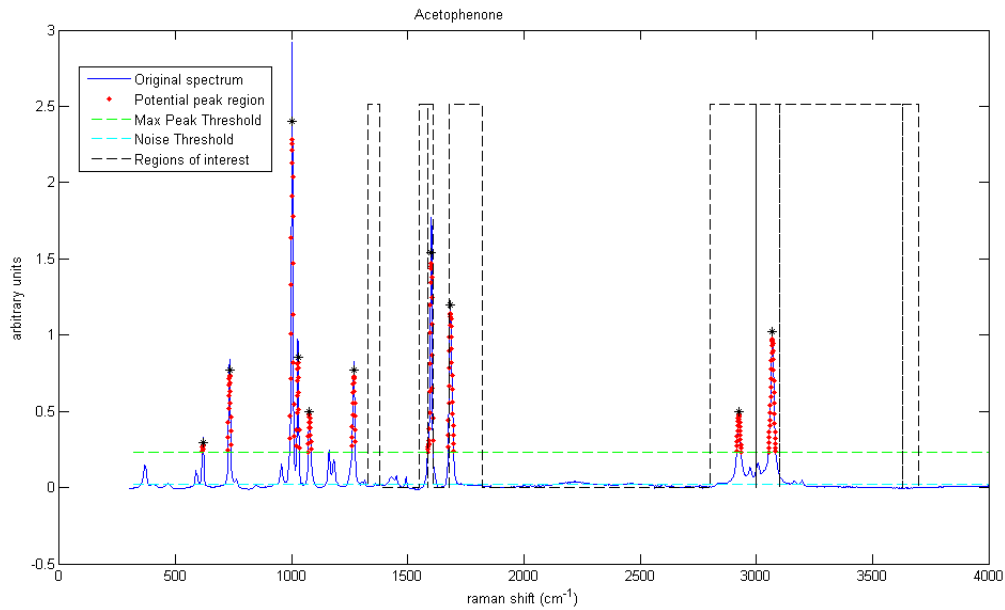


Figure 17. Example of the peak detection algorithm with a maximum intensity threshold definition

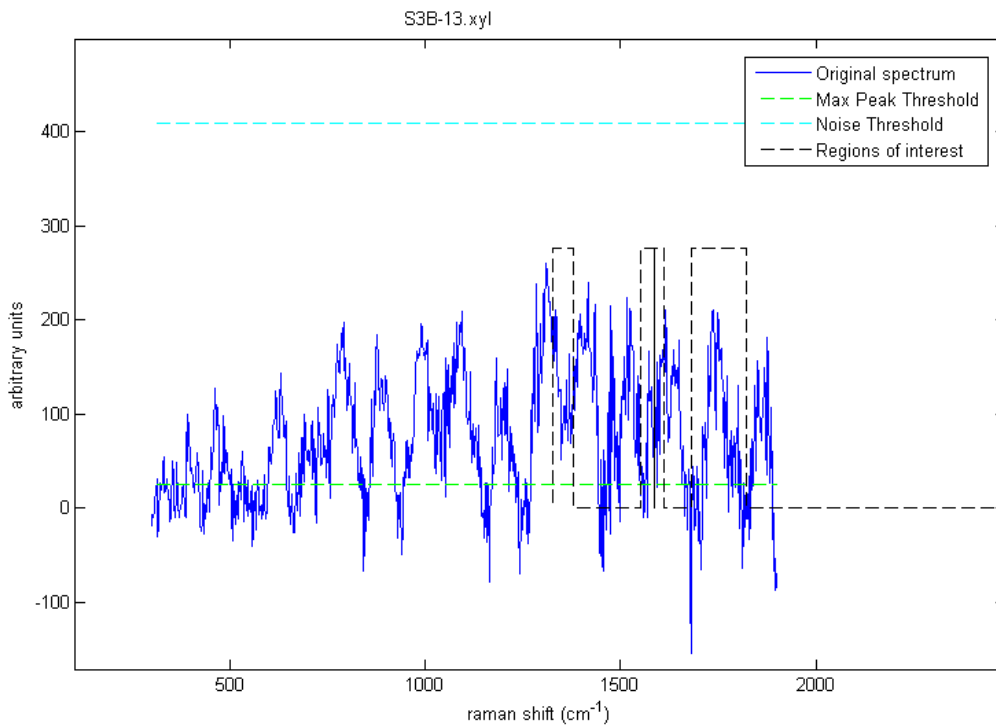


Figure 18. Example of the peak detection algorithm with a noise threshold definition. No peaks are detected

On the other hand, a threshold increment is defined as the maximum of an adjustable factor of the non-peak points standard deviation (to avoid detecting noise fluctuations as peaks), and a 5% of the maximum peak (to limit the number of iterations to 20). This value is defined to iteratively increase the threshold, to provide the algorithm with the ability of detecting doublets (double peaks) or triplets (triple peaks). Figure 19 shows an example of doublet detection. The detected peaks are considered to be the maximum value of all the values in a row which are found over the threshold. This allows the first threshold (green line) to detect the quartz peak at  $465 \text{ cm}^{-1}$ , as well as the graphite one at  $1600 \text{ cm}^{-1}$ . The following threshold (pink line), however, will find

both graphite peaks ( $1360$  and  $1600\text{ cm}^{-1}$ ) in separated regions, and thus be able to detect both of them.

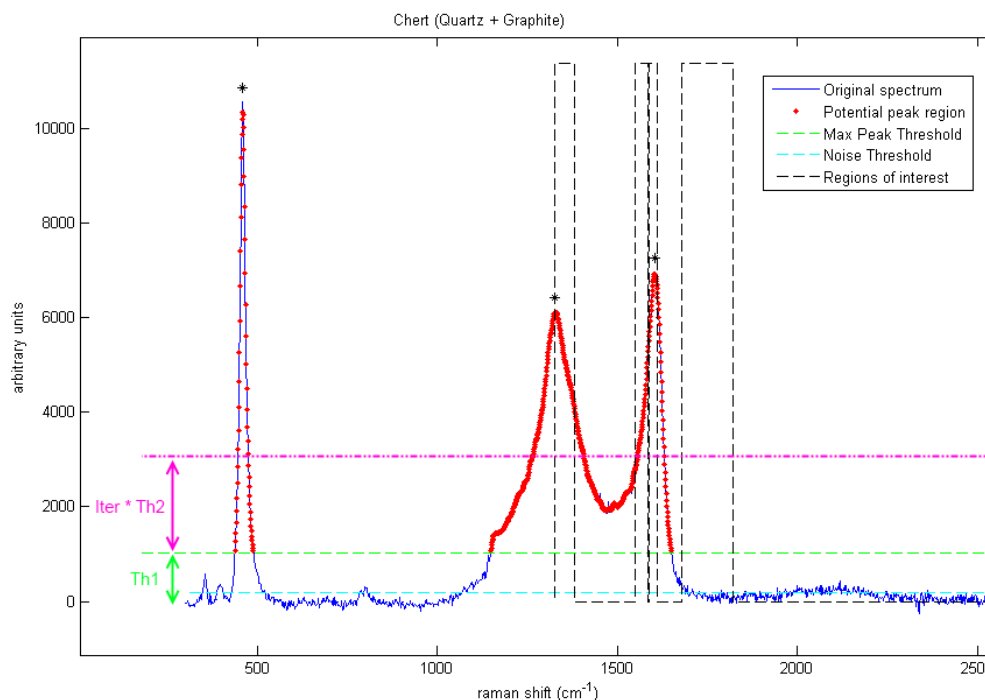


Figure 19. Example of the peak detection algorithm with a doublet spectrum.

### 2.3.2. Methodology

This algorithm has two parameters to define its thresholds: the percentage of the maximum intensity peak, and the factor by which the noise standard deviation is multiplied. The criterion for optimizing the parameters consists on applying the restrictions described before: no false positives have to be detected, and the minimum number of false negatives should be returned. These values are calculated by comparing the calculated regions of interest with the manually calculated (objective) results. As there are two parameters to configure, two bi-dimensional matrixes are obtained: a two dimensional matrix with the number of false positives for each combination of parameters, and the equivalent one with the number of false negatives. Figure 20 shows the representation of the false positives and false negatives matrixes. It can be seen that, as expected, the lower the thresholds, the lower the number of false negatives, but the higher the number of false positives.

The definition of the criterion could be, for example, to minimize the total number of errors. However, in the case we are considering, the criterion definition is to minimize the number of false negatives from among all those combinations in which no false positives are obtained.

After applying this criterion for the optimization of the parameters, we observed that it did not provide the most robust configuration. Furthermore, by executing a LOOCV procedure, it was found that, for some sets of spectra, the calculated parameters provided false positives for the left-out spectra (this means over-fitting), which was unacceptable.

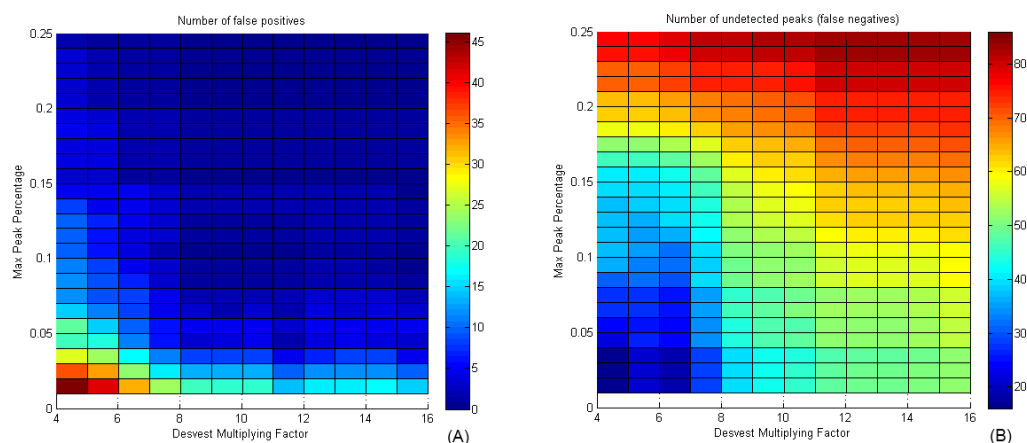


Figure 20. (A) False positive matrix and (B) False negatives matrix

To solve this situation, instead of taking the absolute minimum number of false negatives, the parameters were relaxed (incremented) to increase the probability that new spectra do not result in false positives. This was verified with the LOOCV results, and it results in higher rates of false negatives, as a tradeoff for robustness.

The conditions of the spectral regions to be found were defined to distinguish among different bonding types: C=C, C=O, C-H and hydration (H<sub>2</sub>O or OH), instead of also trying to distinguish between different frequencies of the same bonding types (e.g. water from free OH). The regions were defined with the wavenumbers described in Figure 12.

The algorithm was run on the manually corrected spectra for reference, and then again with the spectra corrected with the baseline removal algorithms Simple Baseline and the modified Weakley.

### 2.3.3. Results and discussion

The results obtained with the previous considerations can be found in Table 7. As it can be seen, the corrected spectra results, which provide the reference and, in principle, best possible results, have an error rate of 27% undetected regions of interest. This is quite a high value, due to the fact that the false positives have to be 0, and also because the chosen parameters do not correspond to the optimized ones, to address for more robustness.

On the other hand, when using the automatic baseline correction algorithms, the results worsen a 6%. However, this figure is double than what could be obtained by taking the optimal parameters, as a consequence of the tradeoff between accuracy and robustness.

	Manually Corrected spectra (objective)	Simple Baseline (Alemanes)	Modified Weakley
Max Peak Percentage	9	8	9
Noise factor	9	10	9
Undetected peaks (%)	27%	33%	33%
Undetected peaks (wrt corrected spectra)	--	6%	6%

Table 7. Performance results for the peak detection algorithm.

### **Capítulo 3. Resumen: Pseudo-cuantificación de especies minerales a partir de espectros Raman del instrumento RLS (English follows)**

El modo de operación del instrumento define que se obtienen espectros en varios puntos a lo largo de una muestra en polvo en teoría homogéneamente distribuida. Lo que se pretende es realizar un análisis de forma que se evalúe si es posible la identificación de todas las fases minerales presentes, y además ver si, a partir del análisis de varios puntos, es posible cuantificar la proporción de las especies minerales presentes.

En este capítulo se definen un método para la pseudo-cuantificación de especies minerales a partir del análisis de un relativamente pequeño número (30) de espectros de mezclas sintéticas con diferentes materiales y proporciones. Todas las mezclas son de dos elementos. Para el análisis de los datos, se desarrolló una herramienta software para automatizar el proceso y poder tratar los datos de manera eficiente.

Para cada mezcla de materiales, se analizan con el Simulador ExoMars 5 líneas de 30 puntos cada una, en un conjunto de 9 mezclas de diferente proporción, entre el 1% y el 99%. En total, para cada pareja de materiales, se generan 1350 espectros. Las Figuras 21 y 22 muestran los espectros de los materiales que se utilizaron.

El análisis estadístico concreto consiste en calcular la intensidad Raman integrada de todos los picos Raman correspondientes a cada uno de los materiales, y calcular un parámetro de cuantificación como el promedio del ratio de intensidades de cada uno de los materiales con respecto al total. Para calcular la intensidad integrada se usa el método de la Figura 23, aprovechando que los picos a analizar son conocidos de antemano. La Figura 24 muestra un ejemplo de integración de las intensidades de los diferentes materiales.

El análisis de los datos para una mezcla de calcita y yeso dio lugar a los resultados numéricos mostrados en la Tabla 8. Pueden observarse dos cosas a partir de estos datos. Por un lado, los componentes minoritarios son detectados sin problema por el instrumento, incluso en aquellas proporciones en las que se encuentran al 1%. Por tanto, puede decirse que el instrumento presenta niveles de detección de hasta un 1%. Por otro lado, puede observarse cómo el parámetro de cuantificación se encuentra relativamente próximo al valor real de la proporción de la mezcla, por lo que parece que puede ser un buen método para la pseudo-cuantificación de los materiales. La Figura 25 muestra la evolución de la intensidad integrada para cada material en cada uno de los 30 puntos de una línea, con diferentes proporciones, donde se puede apreciar cómo varía la intensidad integrada de los espectros cuando se varían las proporciones entre los materiales. La Figura 26 representa la curva de calibración obtenida para esta mezcla, con una pendiente bastante lineal.

También se probó con otras mezclas como Oxalato de sodio y Nitrato potásico, permitiendo nuevamente la pseudo-cuantificación de la abundancia de los materiales presentes en la mezcla basándose en este método.

## **Chapter 3. Pseudo-quantification of mineral species from the RLS instrument Raman spectra**

### **3.1. Introduction**

The qualitative detection capabilities of the ExoMars RLS Raman spectrometer have been shown in several studies of natural samples [36, 37]. The next logical step, taking into account that the RLS instrument will acquire at least 20 samples of each sample, is to try to obtain some kind of quantitative information out of these multiple spectra. However, to do so, large quantities of data have to be analyzed, and thus, this task can be tedious.

The aim of this part of the project was to develop a software tool to automate the analytical process, to provide a fast and relatively unsupervised means for the processing of the generated data of the pseudo-quantitative analysis.

In this section, the methodology is described, but more attention is paid to the scientific results that can be obtained based on the use of the software and RLS simulator. Here, it is shown that a pseudo-quantitative analysis of the mineral phases on the sample might be performed with this instrument. This conclusion is based upon the results obtained from a relatively small number of points (30) on powdered synthetic admixtures with different materials and proportions. Samples are binary admixtures with known proportions.

### **3.2. Experiment description**

For each admixture, a series of 5 lines of 30 points were analyzed for a set of up to 9 proportions ranging from 1% to 99% in weight, a total of 1350 points were analyzed (focus at each point, calculation of the acquisition parameters and final spectrum acquisition). This was feasible thanks to the RLS ExoMars Simulator described in previous sections, which allowed for the automation of the process.

The samples grain size distribution was 250 microns and smaller, and the spot size of the Raman Optical Head on the sample was 50 microns. The materials used for the samples were chosen with the only restriction of not having common spectral bands in order to facilitate the analysis. In case that the spectra have overlapping peaks, a deconvolution process to identify the individual peaks would be necessary. Sodium Oxalate+Potassium Nitrate and Calcite+Gypsum are two of the analyzed admixtures, and their individual spectra can be seen in Figures 21 and 22, respectively.

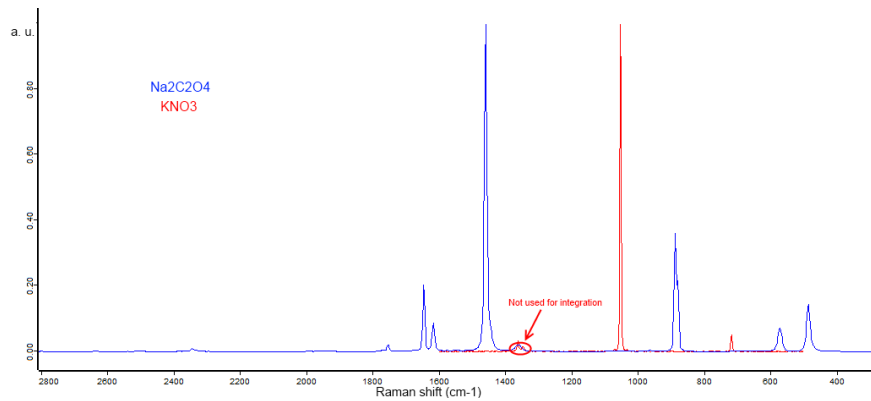


Figure 21. Potassium Nitrate (KNO<sub>3</sub>) and Sodium Oxalate (Na<sub>2</sub>C<sub>2</sub>O<sub>4</sub>) spectra

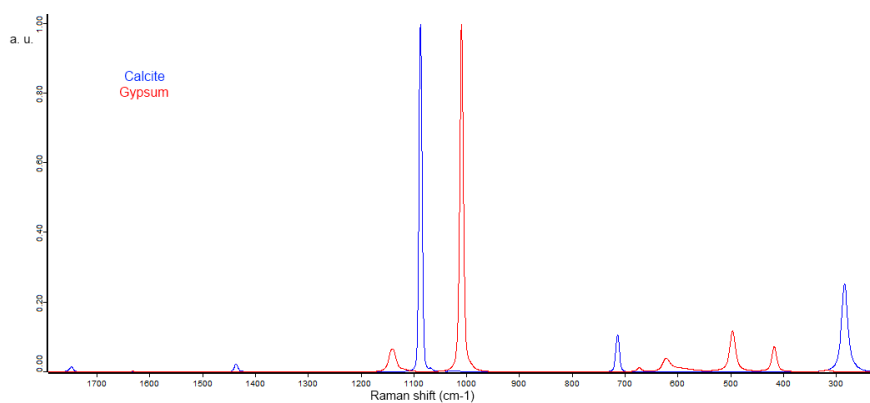


Figure 22. Calcite and gypsum spectra

### 3.3. Statistical analysis

The proposed procedure for analyzing the spectra is divided in three steps:

- Correct the spectra baselines and calculate the Raman integrated intensity at each point for each material found in the spectrum. To do so, the area under the peaks belonging to a material is integrated.
- Calculate the ratio  $r_{point}$  between the integrated intensity of each material ( $I_1$  and  $I_2$ ) and the total intensity of all the materials ( $I_1 + I_2$ ).
- Calculate the quantification parameter  $r_{material}$  as the average of the intensity ratio for each material of the admixture (average of  $r_{point\_material}$ ).

$$r_{point_1} = \frac{I_1}{I_1 + I_2}$$

$$r_{point_2} = \frac{I_2}{I_1 + I_2}$$

All this procedures were developed into a software platform to perform all these tasks pseudo-automatically.

The baseline correction in this case is simpler than for the processes described in the previous sections, because, in this case the materials under study are previously known. To integrate the peaks intensity, the algorithm is input with the peak frequency (or wavenumber) limits, as shown in Figure 23. This way, the baseline is defined by linear interpolation between the two indicated limits, and the intensity easily calculated by a trapezoidal integration method, as shown in Figure 24.

The only inputs the software needs are the wavenumber limits for all the peaks of the material that wants to be calculated. Then the software calculates the total intensity for a material, for all the spectra set (in this case, for 9 executions of 150 spectra each). The results are saved into an excel file from where they are easily treated.

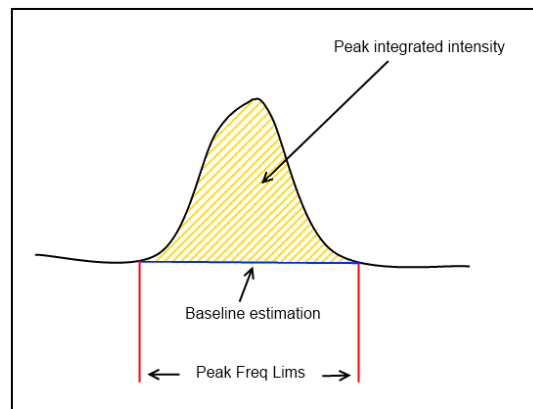


Figure 23. Peak intensity integration method

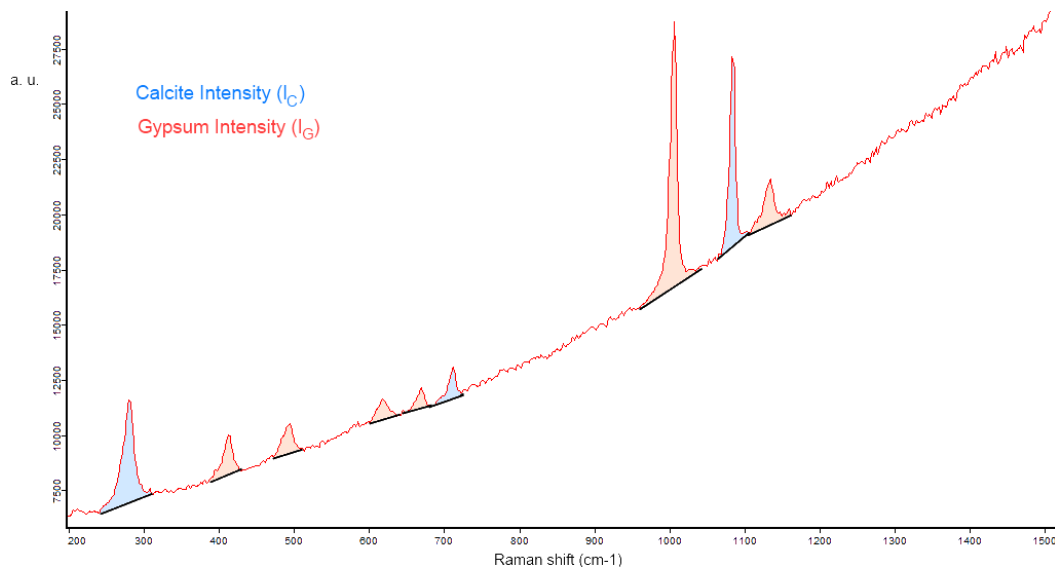


Figure 24. Example of a mixture spectrum. The intensity of each material is the integration of all its peaks

### 3.4. Results discussion

By means of the developed software, it was possible to analyze in a practical way a huge amount of data. This kind of software will be included in a wider package for the mission EGSE (Electronic Ground Support Equipment).



For the calcite and gypsum admixture, the quantification parameters for calcite  $r_C$  and gypsum  $r_G$  were calculated as the average along a line of the ratios  $I_C/(I_C+I_G)$  and  $I_G/(I_C+I_G)$  obtained at each point of the line, where  $I_C$  and  $I_G$  represent the spectral intensity of calcite and gypsum in a given spectrum, respectively, as described before.

Five lines were analyzed for each sample (for each mixture proportion). For this admixture, the  $r_{\text{point}_C}$  and  $r_{\text{point}_G}$  values for each of the 30 points in a line were calculated. The graphical representation of these values, for each of the 30 points in the line can be seen in Figure 25.

The quantification parameter  $r_{\text{material}}$  was calculated for each line and material, as the average of the  $r_{\text{point}_C}$  and  $r_{\text{point}_G}$  values in the line. The numerical results for the calcite + gypsum admixture can be seen in Table 8.

Calcite Intensity / Total Intensity					
Calcite proportion	Line 0	Line 1	Line 2	Line 3	Line 4
1%	4%	1%	1%	1%	0.4%
10%	11%	13%	13%	9%	13%
25%	27%	26%	22%	16%	34%
50%	58%	50%	45%	50%	52%
75%	75%	71%	78%	76%	79%
90%	91%	91%	91%	88%	93%
99%	97%	98%	96%	97%	96%

Gypsum Intensity / Total Intensity					
Gypsum proportion	Line 0	Line 1	Line 2	Line 3	Line 4
99%	96%	99%	99%	99%	99.6%
90%	89%	87%	87%	91%	87%
75%	73%	74%	78%	84%	66%
50%	42%	50%	55%	50%	48%
25%	25%	29%	22%	24%	21%
10%	9%	9%	9%	12%	7%
1%	3%	2%	4%	3%	4%

**Table 8. Intensity ratios for the 5 different lines vs the real proportion in the Calcium+Gypsum admixture**

By taking a look at the obtained results we can see that all the materials were found in every sample, even the minor ones in the 1%-99% proportions. This would mean that, with a 30 point line, the analyzed materials have a detection threshold down to 1% with this instrument. Due to the <250 microns grain size, in most of the cases there were points in which both materials were found simultaneously, thus lowering the detection limit for these samples. Of course, the Raman efficiency of the materials plays an important role in the experiment: For instance, a very poor Raman scatterer in a very low concentration could be concealed by the presence of a Raman efficient material.

The results also showed that a semi-quantitative analysis can be performed based on the results obtained by RLS, as the quantification parameter is relatively close to the actual values of the proportion (see Table 8).

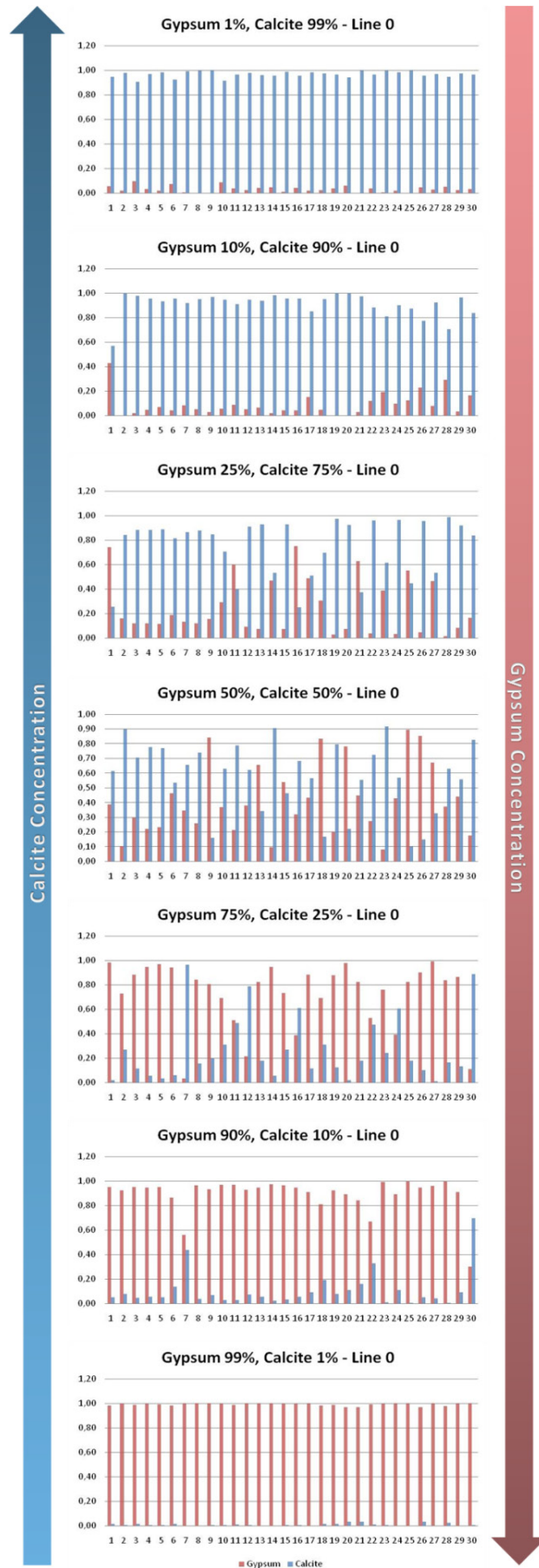


Figure 25. Intensity evolution along a 30 point line for a set of calcite + gypsum mixture

It is important to note that, in this case, the differences in the scattering efficiencies of the materials are somehow compensated by the fact that the integration times for each spectrum are automatically calculated by the Simulator software at each point. This way, the integration time is somewhat proportional to the concentration of the less efficient Raman scatterer, thus providing a quite linear calibration curve (see Figure 26), which will be used to predict the relative concentrations of future unknown mixtures.

This same procedure was applied to the Sodium Oxalate + Potassium Nitrate mixtures, and the final results were similar to those here presented, but obtaining not so linear calibration curves. Anyhow, this calibration curves still allow the pseudo-quantification of unknown mixtures based only on their Raman spectra.

Future experiments will need to address the overlapping of the peaks and the subsequent necessary deconvolution process.

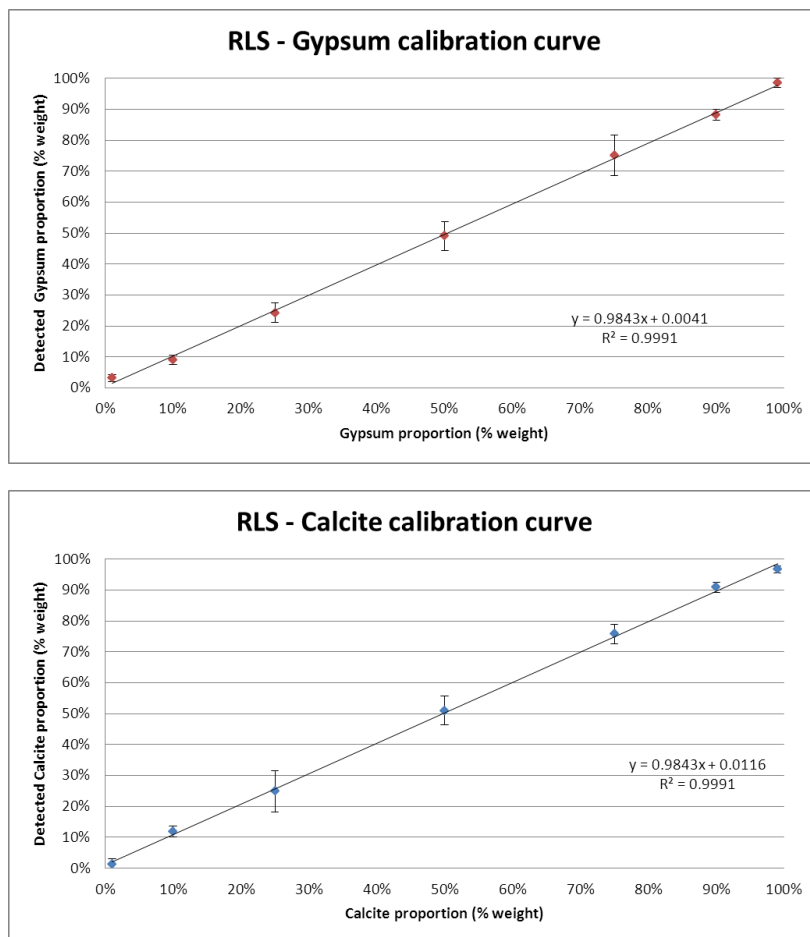


Figure 26. Calibration curves of the calcite+gypsum mixture

## Capítulo 4. Resumen: Compendio de resultados y conclusiones (English follows)

Este trabajo ha presentado una serie de propuestas en dos direcciones complementarias del desarrollo del instrumento RLS: por un lado se han propuesto y estudiado una serie de algoritmos para dotar al instrumento de la capacidad de detectar regiones espectrales de potencial interés astrobiológico, de forma que dicha información pueda ser priorizada durante las comunicaciones con la tierra. Por otro lado, se ha propuesto también un método para el análisis ulterior de los espectros del instrumento, abriendo la puerta a la posibilidad de pseudo-cuantificar la abundancia de especies minerales en mezclas minerales.

La algoritmia de detección de especies minerales de interés se divide en dos fases. La primera es la detección y eliminación de la línea de base, para lo cual se estudiaron dos algoritmos existentes (Alemanes y Weakley) y se propusieron una serie de mejoras, para posteriormente comparar su bondad y robustez. El resultado fue que el algoritmo Alemanes, que en principio parecía funcionar peor cuando se optimizaban los parámetros para cada espectro concreto, sin embargo igualaba en comportamiento al algoritmo Weakley cuando se evaluaba de forma general para un conjunto de 113 espectros. Además, cuando se realizó validación cruzada LOOCV se demostró que el algoritmo Alemanes tiene un comportamiento más robusto que el Weakley modificado. La segunda fase consiste en la detección de los picos sobre el espectro sin línea de base. Para ello se utilizó un algoritmo estándar basado en umbrales ampliamente utilizado en software de análisis espectral y que, sin embargo, requiere la interacción con el usuario para optimizarlo. La contribución que se hace en este trabajo consiste en aplicar un criterio para decidir dichos umbrales en función exclusivamente de parámetros espectrales (intensidad del pico máximo y nivel de ruido del espectro). Los parámetros se optimizaron para minimizar la probabilidad de obtener falsos positivos, y minimizar en la medida de lo posible el número de picos no detectados, de manera que el algoritmo se mantenga con una cierta robustez para nuevos espectros.

En lo referente al análisis de datos, el instrumento realizará análisis sobre un determinado número no menor de 20 espectros en la misma muestra. Esto supone la pérdida de contexto geológico de la muestra, que no podrá analizarse en material masivo. Sin embargo, el análisis de muestras en polvo mediante la metodología explicada en este trabajo ha demostrado ser más eficaz de lo esperado. Por un lado, los límites de detección de este sistema pueden estar entorno al 1%. Por otro, es posible realizar una pseudo-cuantificación de la abundancia de especies minerales en mezclas.

En la sección de trabajo futuro se definen tareas pendientes de realizar. En lo referente al modo de operación del instrumento hay todavía un gran número de funcionalidades de las que dotar al instrumento. Además, en lo referente al algoritmo de detección de trazas de interés, sería interesante, por ejemplo, establecer nuevas condiciones que permitan distinguir, por ejemplo, entre la vibración de los OH libres y las bandas de agua. Una posibilidad sería analizar la forma y anchura de los picos.

En cuanto a la cuantificación de especies minerales, el uso de técnicas multivariantes podría ser interesante, ya que estas técnicas, especialmente la aplicación de redes neuronales artificiales, han demostrado que pueden ser útiles para la identificación de mezclas de minerales, y podrían también serlo para la cuantificación.

En lo referente a la divulgación de los resultados, se está escribiendo un artículo para su publicación en un número especial del European Journal of Mineralogy con motivo del congreso GeoRaman X<sup>th</sup>. Además, se han presentado las siguientes comunicaciones a congresos:

G. Lopez-Reyes, F. Rull, A. Catala, A. Sanz, J. Medina, I. Hermosilla, B. Lafuente (2012). A simple statistical method for the pseudo-quantification of mineral phases within the ExoMars Raman RLS instrument. GeoRaman X<sup>th</sup> 2012, Nancy. [38]

I. Hermosilla, G. Lopez-Reyes, A. Catala, A. Sanz, D. R. Llanos and F. Rull (2012). Raman spectra processing algorithms and database for RLS-ExoMars. European Planetary Science Congress EPSC, Madrid. [39]

## Chapter 4. Wrap-up

### 4.1. Final considerations and conclusions

#### 4.1.1. Contribution to the instrument operation mode

The present project has addressed a number of issues, all of them forming part of a whole entity in the framework of the definition and development of the mode of operation of the ExoMars RLS instrument. The final aim is to obtain the best possible results and scientific return from a portable instrument in a planetary exploration environment, which means that the instrument characteristics will be far from being those of laboratory equipments. Furthermore, the operation of the instrument has to be reinvented, as the Raman spectroscopy technique has always been considered a laboratory technique, and besides, it has always required the intervention of a human operator (which is something that will not be possible in the aforementioned conditions). So the automation of the operation is a key point for the improvement of the instrument capabilities.

To automate the operation procedures, new sets of algorithms have to be studied, defined and developed. Here, the efforts were centered on the development of an algorithm for the on-board automatic detection of spectral bands of astrobiological interest (those spectra presenting vibrations corresponding to C=C, C=O, C-H or OH and H<sub>2</sub>O bondings). The aim is to prioritize these spectra of astrobiological interest over the less relevant ones, so they are sent to earth first. This way, the scientific team will be provided with the most relevant data as soon as possible to take fast and adequate tactical decisions for the rover.

To develop this kind of algorithm, the spectrum baseline has to be subtracted prior to the detection of the peaks.

For the baseline detection algorithm, two algorithms were defined or modified: the Simple Baseline (Alemanes) and the modified Weakley algorithms. Their parameters were optimized to evaluate their goodness, and a Leave-one-out Cross-Validation procedure was executed to evaluate their robustness. The quantification parameter was the MSE of the subtraction of the calculated baseline and the objective one. The results showed that, even if the modified Weakley algorithm outperforms the Simple Baseline one when optimizing the parameters for each spectrum, when these are optimized taking into account all the training spectra, they perform very similarly. However, the LOOCV procedure highlighted that the Simple Baseline algorithm was noticeably more robust than the modified Weakley.

For the detection of the peaks, another algorithm was proposed, based on standard procedures. However, the typical standard algorithms require the user intervention in order to optimize the parameters. In order to remove this need for interaction, we prepared the algorithm to define intensity thresholds, not as a constant value, but as a function of spectral characteristic, namely, the maximum peak intensity and the spectrum noise level. The aim of the algorithm is to detect the maximum quantity of peaks with the certainty that no false positives are detected, as we do not want to prioritize spectra based on non-reliable data. The optimization process included also a

LOOCV procedure to estimate the robustness of the algorithm, and to help decide the optimum parameters as a tradeoff between undetected peaks and robustness. The robustness is understood as the capability of the algorithm to behave properly with spectra that have not been used for the optimization of the algorithm parameters.

The chosen configuration of the algorithm showed that 73% of the peaks were correctly detected when using the spectra without baseline, and 6 percentage points less (67% of correctly detected peaks) when using any of the baseline correction algorithms to subtract the baseline. The behavior of the algorithm can be improved by changing the parameters, but at the expense of robustness.

#### 4.1.2. Contribution to data analysis

The instrument operation mode is not also bound by the need of automation, but also to the simplicity, robustness and interaction with other instruments of the Analytical Laboratory Drawer. These factors also bind the instrument to work under determined conditions, as for example, work on crushed samples obtained from the martian subsurface with a drill, instead of working on bulk materials. The reason for this is that some instruments require crushed material to work with but, typically, Raman spectroscopy is operated on bulk samples. The crushing process, by itself, can induce some changes on the Raman spectrum of the materials and implies a loss of the geological context of the samples [14], which is an important drawback. However, the methodology on the powder has proven to be far more effective than expected, being able to detect all compounds present in the sample, with detection levels even below 1%. Furthermore, by acquiring several spectra on a homogeneously distributed powdered sample, the door to statistical analysis and other innovative techniques opens.

This way, in this project a simple statistical method has been proposed and applied to quantify, only based on raman spectra, the mineral phases present on a sample. In effect, a set of synthetic mixtures with different proportions were prepared and analyzed by means of the ExoMars RLS Simulator, obtaining a total of 1350 spectra per sample type to analyze. The proposed statistical method consists on averaging, for each spectrum, the integrated intensity of all the peaks belonging to a determined mineral phase, divided by the total intensity of all the peaks.

For the calculation of the statistical parameters, a software platform was developed in order to subtract the baseline and integrate the intensity for each peak, for all the spectra. The advantage in this case is that the mineral phases were known in advance, making the baseline subtraction easy by using the method described in Figure 23.

By means of this software, it was possible to see that some kind of pseudo-quantification is possible using only several Raman spectra of mixtures. This kind of software will be included in a wider package for the mission EGSE (Electronic Ground Support Equipment).

## 4.2. Future work

The definition of the operation mode of the ExoMars RLS instrument is still work in progress. Many other algorithms need to be defined and optimized. Relating the algorithm for the detection of astrobiologically interesting traces, probably there are some other proposals which can improve the number of undetected peaks without losing robustness. In addition, some other restrictions could be added to the algorithm to further differentiate the interesting traces, as, for example, differentiate free OH vibrations from H<sub>2</sub>O bands. This could be done by including a condition on the peak shape and width.

With respect to the quantification of mineral phases, the use of multivariate techniques could be interesting. The techniques as Principal Component Analysis (PCA), Partial Least-Squares Regression (PLSR) and, especially, Artificial Neural Networks (ANN) seem to be promising for the detection and identification of mixtures of minerals based on one Raman spectrum [40], and they should also be considered for the quantification of the mineral species.

## 4.3. Conference papers related to this work

The contributions presented in this document have or will be published or presented to conferences. At present, a paper for a special issue related to the GeoRaman X<sup>th</sup> congress of the European Journal of Mineralogy. In addition, the following abstracts have been presented to conferences.

Description of the statistical method for the pseudo-quantification of mineral species:

G. Lopez-Reyes, F. Rull, A. Catala, A. Sanz, J. Medina, I. Hermosilla, B. Lafuente (2012). A simple statistical method for the pseudo-quantification of mineral phases within the ExoMars Raman RLS instrument. GeoRaman X<sup>th</sup> 2012, Nancy. [38]

Description of some instrument on-board algorithms:

I. Hermosilla, G. Lopez-Reyes, A. Catala, A. Sanz, D. R. Llanos and F. Rull (2012). Raman spectra processing algorithms and database for RLS-ExoMars. European Planetary Science Congress EPSC, Madrid. [39]



## References

1. ESA. *Aurora program web page* at [www.esa.int](http://www.esa.int). 2006; Available from: [http://www.esa.int/esaMI/Aurora/SEMZOS39ZAD\\_0.html](http://www.esa.int/esaMI/Aurora/SEMZOS39ZAD_0.html).
2. ESA. *Exomars project web page* at [www.esa.int](http://www.esa.int). 2012; Available from: <http://exploration.esa.int/science-e/www/object/index.cfm?fobjectid=46048>.
3. Catalá-Espí, A., *Operation modes and spectral data treatment in the Raman instrument for ExoMars*, in *Departamento de Astrofísica*. 2010, Universidad de La Laguna: Tenerife.
4. ESA Human spaceflight, M.a.E.p.b., *The ExoMars Payload: confirmation of the payload suite Following restructuring of the mission*. 2009, ESA.
5. A. Pacros, A.H., J. Vago and the ESA ExoMars Project Team. *Overview of the ExoMars Payload*. in *European Planetary Science Congress EPSC*. 2009.
6. F. Rull, S.M., E. Diaz. C. Tato. A. Pacros & the RLS Team. *The Raman Laser Spectrometer (RLS) on the EXOMARS 2018 Rover Mission*. . in *42nd Lunar and Planetary Science Conference LPSC*. 2011.
7. Rull, F. *Raman spectroscopy for planetary exploration*. in *Georaman X 10*. 2012.
8. J. L. Vago, B.G., P. Baglioni, G. Kmínek, D. Gianfiglio and the ExoMars Project Team. *ExoMars: ESA's mission to search for signs of life on the red planet*. in *Lunar and Planetary Science Congress LPSC*. 2006.
9. F. Rull, G.L., A. Catalá, J. Medina, A. Sansano, A. Sanz, F. Sobrón. . *Raman spectroscopy analysis on powdered samples inside the Exomars mission*. . in *ESA Congrex Lisbon 2011*. . 2011.
10. WITec. *WITec Alpha 300R Confocal Raman Microscope. Specifications brochure*. 2012; Available from: . <http://www.witec.de/en/products/raman/alpha300r/>.
11. Alejandro Catalá, G.L., Aurelio Sanz, Alberto Vegas, Fernando Rull, , *Raman focus vs. mean focus measurements on powdered samples*. . 2010.
12. G. Lopez, A.C., A. Vegas, I. Hermosilla, F. Rull., *Considerations and improvements proposed for RLS Mission Operation*. . 2011, CAB INTA.
13. F. Foucher, G.L.R., N. Bost, F. Rull-Perez, P. Rübmann and F. Westall. . *Effect of the crushing process on Raman analyses: consequences for the Mars 2018 mission*. . in *EGU General Assembly 2012*. 2012.
14. F. Foucher, G.L.R., N. Bost, F. Rull-Perez, P. Rübmann and F. Westall. . *Effect of the crushing process on Raman analyses: consequences for the Mars 2018 mission*. . in *GeoRaman 2012*. 2012.
15. Catalá Espí, A.L.R., G.; Rull Pérez, F.; Vegas Hernández, A.; . *Design of a small Martian chamber for simulating the RLS-ExoMars operation mode*. in *European Planetary Science Congress EPSC*. 2011.
16. Rull, F., Martínez-Frías, J. , *Raman Spectroscopy goes to Mars*. . *Spectroscopy Europe*, 2006. **1818**(1).
17. Zhang, Z.-M., et al., *An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy*. *Journal of Raman Spectroscopy*, 2010. **41**(6): p. 659-669.
18. García, L.A., *Analysis and development of signal processing algorithms related to Raman and LIBS spectroscopy to identify chemical compounds*, in *Teoría de la Señal y Comunicaciones e Ingeniería Telemática*. 2011, Universidad de Valladolid: Valladolid.

19. Yvon, H.J., *Raman Spectroscopy for Analysis and Monitoring*.
20. CAB, U.a.U.-C.a.t.d.C.d.A. *Standard Spectra Database*. Available from: <http://84.124.106.189/DB/index.php>.
21. Schulze, G., et al., *Investigation of selected baseline removal techniques as candidates for automated implementation*. *Applied Spectroscopy*, 2005. **59**(5): p. 545-574.
22. Weakley, A.T., P.R. Griffiths, and D.E. Aston, *Automatic Baseline Subtraction of Vibrational Spectra Using Minima Identification and Discrimination via Adaptive, Least-Squares Thresholding*. *Applied Spectroscopy*, 2012. **66**(5): p. 519-529.
23. Shao, L. and P.R. Griffiths, *Automatic baseline correction by wavelet transform for quantitative open-path Fourier transform infrared spectroscopy*. *Environmental Science & Technology*, 2007. **41**(20): p. 7054-7059.
24. Keselbrener, L., M. Keselbrener, and S. Akselrod, *Nonlinear high pass filter for R-wave detection in ECG signal*. *Medical Engineering & Physics*, 1997. **19**(5): p. 481-484.
25. Mosierboss, P.A., S.H. Lieberman, and R. Newbery, *Fluorescence rejection in Raman spectroscopy by shifted-spectra, edge-detection and FFT filtering techniques*. *Applied Spectroscopy*, 1995. **49**(5): p. 630-638.
26. Rowlands, C. and S. Elliott, *Automated algorithm for baseline subtraction in spectra*. *Journal of Raman Spectroscopy*, 2011. **42**(3): p. 363-369.
27. Hasenoehrl, E.J., J.H. Perkins, and P.R. Griffiths, *Rapid functional group characterization of gas-chromatography Fourier-transform infrared spectra by a Principal Components Analysis based expert system*. *Analytical Chemistry*, 1992. **64**(7): p. 705-710.
28. Lewis, D.M. and P.C. Chatwin, *The treatment of atmospheric dispersion data in the presence of noise and baseline drift*. *Boundary-Layer Meteorology*, 1995. **72**(1-2): p. 53-85.
29. Tervo, R.J., T.J. Kennett, and W.V. Prestwich, *An automated background estimation procedure for gamma-ray spectra*. *Nuclear Instruments & Methods in Physics Research*, 1983. **216**(1-2): p. 205-218.
30. Lieber, C.A. and A. Mahadevan-Jansen, *Automated method for subtraction of fluorescence from biological Raman spectra*. *Applied Spectroscopy*, 2003. **57**(11): p. 1363-1367.
31. Goehner, R.P., *Background subtract subroutine for spectral data*. *Analytical Chemistry*, 1978. **50**(8): p. 1223-1225.
32. Iwata, T. and J. Koshoubu, *New method to eliminate the background noise from a line spectrum*. *Applied Spectroscopy*, 1994. **48**(12): p. 1453-1456.
33. Balcerowska, G. and R. Siuda, *Inelastic background subtraction from a set of angle-dependent XPS spectra using PCA and polynomial approximation*. *Vacuum*, 1999. **54**(1-4): p. 195-199.
34. Schulze, H.G., et al., *A Model-Free, Fully Automated Baseline-Removal Method for Raman Spectra*. *Applied Spectroscopy*, 2011. **65**(1): p. 75-84.
35. Schulze, H.G., et al., *A Small-Window Moving Average-Based Fully Automated Baseline Estimation Method for Raman Spectra*. *Applied Spectroscopy*, 2012. **66**(7): p. 757-764.

36. Fernando Rull, A.C., Guillermo López, Antonio Sansano and Aurelio Sanz. . *Capabilities of detection on crushed samples by Raman Spectroscopy: Exomars Raman Instrument.* . in *GeoRaman 2010*. 2010.
37. A. Sansano, G.L., J. Medina, F. Rull and AMASE'10 Team. *Analysis of Arctic Carbonates by Raman Spectroscopy using Exomars's RLS set up* in *European Planetary Science Congress EPSC-DPS*. 2011. Nantes.
38. Lopez-Reyes, G., Rull, F., Catala, A., Sanz, A., Medina, J., Hermosilla, I., Lafuente, B. *A simple statistical method for the pseudo-quantification of mineral phases within the ExoMars Raman RLS instrument.* in *GeoRaman 2012*. 2012. Nancy.
39. I. Hermosilla, G. Lopez-Reyes, A. Catalá, A. Sanz, D. R. Llanos and F. Rull. *Raman spectra processing algorithms and database for RLS-ExoMars.* in *European Planetary Science Congress EPSC*. 2012. Madrid.
40. G. Lopez-Reyes, P.S., B. Lafuente and F. Rull. *Application of multivariate analysis techniques for the identification of sulphates from Raman spectra – Implications for ExoMars.* in *European Planetary Science Congress EPSC*. 2012. Madrid.