

COMPARING LIBS SPECTRA FROM CHEMCAM INSTRUMENT ON BOARD CURIOSITY (MSL) : HOW DIFFERENT CAN THEY BE AND STILL BE “THE SAME” ?. **Éric LEWIN**¹, Jérémie LASUE², Olivier FORNI², Bob TOKAR^{3,4}, Diana BLANEY⁵, Noureddine MELIKECHI⁶, Olivier GASNAULT², Sylvestre MAURICE², Roger WIENS⁴, the ChemCam team and the MSL team, ¹ISTerre (OSUG : uJF-Grenoble & INSU.CNRS), Grenoble.FR (<Eric.LEWIN@uJF-Grenoble.fr>), ²IRAP, Toulouse.FR, ³PSI, Tucson.AZ. USA, ⁴LANL, Los Alamos.NM.USA, ⁵JPL.NASA/CalTech, Pasadena.CA.USA, ⁶Delaware State Univ., Dover.DE. USA.

Introduction: From a metrological point of view, instrumental uncertainty is estimated with experiments of reproducibility defined through a protocol that refer directly or indirectly to standards obtained through highly accurate and highly precise measurements. However for planetary field instruments, operation conditions do barely even approach laboratory ones, and reproducibility as well as accuracy are much harder to assess. In the case of an instrument that is destructive of the measured sample of the target, such as the laser induced breakdown spectroscopy (LIBS) instrument ChemCam [Wiens, 2012; Maurice, 2012], one can resort to homogeneous calibration targets [Fabre, 2011; Vaniman, 2012; Tokar, 2013] such as those on board Curiosity. However, since distance of operation is an important effective parameter [Melikechi, 2013], the specific proximity of the ChemCam calibration targets (~1.5m) relative to normally operated ones (>2.8 m) requires an added modeling step, distance correction, which therefore hampers direct application of reproducibility results to field sample measurements.

Another approach is to look at apparently homogeneous targets, such as fine-grain soils ([Blaney, 2013]) or the so-called omnipresent martian fine dust [Melikechi, 2013]. Statistical dispersion studies on acquired spectra sets for such targets thus permit to analyze their distance induced intensity shifts and to define an upper bound level for the minimal variability in a repeatability "on-field" experiment. Thus, the improved analysis of the LIBS spectra derives principally from our ability to measure in-situ the behavior of specific emission lines with distance, and to predict quantitatively the correction needed. A third approach, slightly different from the previous one is to look at spectra sets of the spot target basis and exhibit those with particularly low dispersion level.

Once the instrumental reproducibility level is quantified, it can be compared with the dispersion level over a set of spectra from different spots on a single geological target or from different rocks or soils. This permits to assess the chemical heterogeneity of a rock, of a geological unit, of a planetary traversed landscape. The statistical tool for processing such comparisons is called ANOVA, stated for "ANalysis Of VAriance".

The multivariate essence of the LIBS spectral data brings some technical difficulty for applying this technique straightforwardly. One of this is that variance is mathematically defined as a square definite positive matrix with 6144 rows and lines (ChemCam spectra are defined over $6144 = 3 \times 2048$ channels encompassing three spectral ranges). Since many LIBS spectral channels are highly correlated, either that they are oversampling a singular spectral line, or because they are sampling different spectral lines of the same chemical element, etc., the true dimensionality of this variance matrix is much smaller. Mathematically, in the case of an optimal instrumental reproducibility test, this dimensionality is maximal. This work will present a first attempt to estimate the minimum admissible boundary value for this dimension.

Another important aspect of statistical comparison of multivariate data is to define a metric of the spectral space in order to permit the definition of a (generalized euclidean) distance (or metric) between two spectra. Classically, two options are considered. The first one, which we can call the photonic distance considers the straightforward quadratic sum of the difference in intensity for the two spectra over the whole channels set. In this case, each channel intensity is defined in term of photons per wavelength unit range (per nanometer ; the wavelength width of ChemCam spectrometers channels are constant inside each of the three subsets, UV, VIS and VNIR, but are somewhat different between these three). Practically these intensities are not used in their absolute values, but are rather normalized to the total intensity over each respective wavelength subrange. The second option for defining the distance between two spectra is the statistical Mahalanobis distance, or at least its more classical much simplified version the normalized euclidean distance, for which the intensity difference is normalized by the respective standard deviation estimated with the whole set of spectra under consideration. Classical multivariate analysis techniques use this option, which therefore yield the same statistical weight for each channel, either it belongs to a spectral line or it is purely noise. Advantage to the former channels, those partially covering a spectral line, comes from correlations

which somehow aggregate these channels and enforce their statistical weight as a group. A third option is to define the multispace metric from metrological or geochemical considerations [Allègre, 1987, EPSL ; Allègre, 1995, EPSL]. In this work, we will use the reproducibility previously defined for such purpose. With this, each channel is given a weight which corresponds to its reproducibility, therefore accounting instrumental uncertainty with a more rigorous approach.