BOLETÍN
DE
CIENCIAS DE LA TIERRA

https://revistas.unal.edu.co/index.php/rbct

Evaluation of normalization methods applied to Short-Wavelength Infrared (SWIR) spectroscopy mineral databases from multiple instruments and for vectoring analysis exploration

Juan Camilo Paredes, Yan Carlos Trigos & Camilo Uribe-Mogollon

Eafit University, Medellín, Colombia. jcparedesa@eafit.edu.co, ytrigos@eafit.edu.co, uribemogollon@gmail.com

Received: March 12th 2024. Received in revised from: November 6th 2024. Accepted: November 8th 2024.

Abstract
Over the past decade, short-wave infrared (SWIR) spectroscopy has made significant advances in detecting geochemical variations in minerals like white mica, alunite, and chlorite for exploring hydrothermal ore deposits. These variations provide valuable clues, indicating changes in temperature, pH, and fluid oxidation state towards the mineralized center. However, small calibration differences among devices challenge data integration. This study evaluates the 2200 nm Al-OH absorption feature in four white mica SWIR spectroscopy databases collected by TerraSpec™ and OreXpress™ from samples at the Grasshopper porphyry prospect. It evaluates three normalization methodologies: rescaling, mean normalization, and Z-score, yielding p-values for successful data merging of up to 0.75. Findings suggest effective normalization methods across devices, reducing biases from uncalibrated spectrometers. This research offers a methodology to correct SWIR database biases, facilitating accurate data integration across instruments for vectoring analysis.

Keywords: Reflectance spectroscopy; SWIR; White mica; database normalization.

Evaluación de métodos de normalización aplicados a bases de datos minerales de Espectrografía de Infrarrojo Cercano (SWIR) provenientes de múltiples instrumentos y para análisis de vectores de exploración

Resumen
Durante la última década, la espectroscopía de infrarrojo de onda corta (SWIR) ha experimentado avances significativos en la detección de variaciones geoquímicas en minerales como la mica blanca, la alunita y la clorita para explorar depósitos de minerales hidrotermales. Estas variaciones proporcionan pistas valiosas, indicando cambios en la temperatura, el pH y el estado de oxidación del fluido hacia el centro mineralizado. Sin embargo, las pequeñas diferencias de calibración entre dispositivos representan un desafío para la integración de datos. Este estudio evalúa la característica de absorción del Al-OH a 2200 nm en cuatro bases de datos de espectroscopía SWIR de mica blanca recopiladas por TerraSpec™ y OreXpress™ a partir de muestras en el prospecto de pórfido Grasshopper. Se analizan tres metodologías de normalización: reescalado, normalización de la media y variable centrada reducida, obteniendo valores de p para la fusión exitosa de datos de hasta 0.75. Los hallazgos sugieren métodos de normalización efectivos entre dispositivos, reduciendo sesgos de espectrómetros no calibrados. Esta investigación ofrece una metodología para corregir sesgos de la base de datos SWIR, facilitando la integración precisa de datos entre instrumentos para análisis de vectores.

Palabras clave: Espectroscopía de reflectancia; SWIR; Mica blanca; normalización de bases de datos.

1 Introduction

The use of shortwave infrared (SWIR) spectroscopy to identify changes in the geochemistry of alteration minerals that can be used as markers for the search for hydrothermal ore deposits has made major strides in the last ten years [1,2]. SWIR spectroscopy is a technique that collects reflectance spectra in the range of 1300-2500 nm caused by vibrational process of molecular bonds such as OH, H_2O, NH_4, CO_3, Al- OH, Mg-OH, and Fe-OH [3]. These bonds have a distinctive absorption feature and are usually present in the structure of alteration minerals including

How to cite: Paredes, J.C., Trigos, Y.C., and Uribe-Mogollon, C., (2024). Evaluation of normalization methods applied to Short-Wavelength Infrared (SWIR) spectroscopy mineral databases from multiple instruments and for vectoring analysis exploration. BOLETÍN DE CIENCIAS DE LA TIERRA. 56, pp. 47-53. DOI:https://doi.org/10.15446/rbct.n56.113445

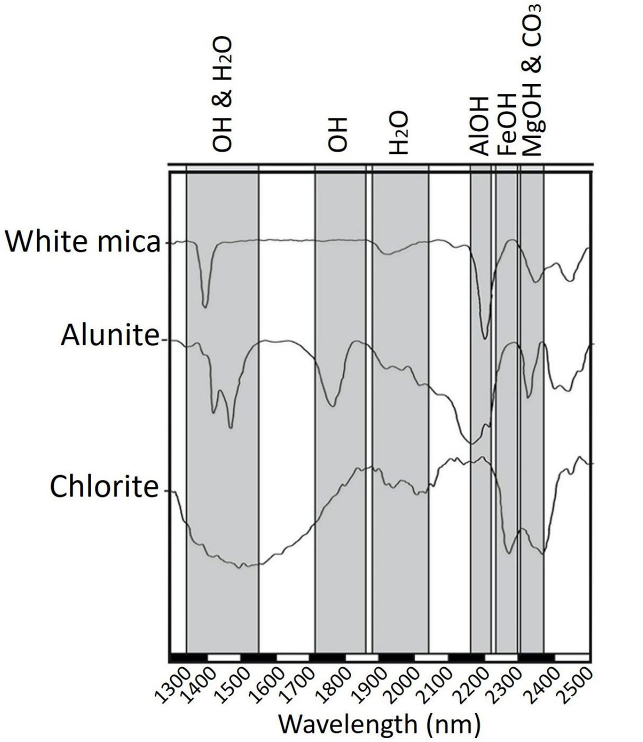

Figure 1. This image presents behavior plots of reflectance spectra for white mica, alunite, and chlorite, corrected using hull quotient. White mica is characterized by a distinctive Al-OH absorption feature at 2200 nm. Alunite exhibits a significant OH and H_2O absorption feature at 1480 nm. In contrast, chlorite shows distinct dips for the Fe-OH absorption feature at 2250 nm and the Mg-OH feature at 2340 nm. These plots are valuable for identifying and distinguishing these minerals in geological studies.
Source: authors.

phyllosilicates, hydroxylated silicates, sulphates, and carbonates (Fig. 1) [4]. Studies of the variation of the spectral absorption features indicate changes in the mineral chemistry as a function of physicochemical conditions under which they were formed [5].

In the context of porphyry deposits, the study of the spectral variations in hydrothermal minerals such as white mica, alunite, and chlorite has provided relevant vectoring information towards the mineralized center produced by changes in temperature, pH, and/or oxidation state of the fluids [6-8]. For example, at the Copper Cliff porphyry Cu deposit in Montana, the wavelength position of the Al-OH spectral absorption feature at ~2200 nm in white micas proved effectiveness in the identification of two distinct phyllic alteration events: 1) an early green-colored expression associated with Fe-bearing micas and high grades of hypogene copper mineralization characterized by longer Al-OH absorption wavelengths (2206-2210 nm) at the deposit center, and 2) a later barren expression characterized by Fe-poor bearing white-colored micas and shorter Al-OH absorption wavelengths (2197-2206 nm). The difference in the location of the white mica Al-OH absorption feature between both phyllic alteration styles is principally controlled by chemical variations in the octahedral site within the white mica structure and attributed to redox changes in the system, where the early event was formed by oxidized magmatic fluids and the latest by more reduced fluids [8].

Likewise, for alunite the wavelength position of the OH absorption feature at ~1480 nm has reported systematic trends in the range of 1478-1482 nm, with the longest wavelengths towards the intrusive center at the Lepanto lithocap in Philippines [6]. This increase in the wavelength position of the OH absorption feature in alunite is associated with a higher content of Na and lower K, following the Na/(Na+K) relationship which is related to a higher formation temperature (Chang et al., 2011). For chlorite, the Fe-OH absorption feature around 2250 nm and the Mg-OH located near 2340 nm showed systematic decreases from 2254 nm to 2249 nm and from 2343 nm to 2332 nm towards the center of the Batu Hijau Cu-Au porphyry system in Indonesia. The shifts in the wavelength position of the Fe-OH and Mg-OH absorption features correlate with variations in the content of Fe^{2+} and Mg^{2+} in octahedral site of the chlorite and linked to fluid temperature [9].

The successful application of SWIR spectroscopy as a vectoring tool is based on the instrument precision. However, during data collection most exploration projects may use multiple SWIR instruments with different calibration settings. As a result, in cases when no inter-instrument calibration is performed, variability of the value of a spectral feature among SWIR instruments has been identified and the application of data for vector analysis is often difficult [5,10-11]. This issue opens the possibility to develop a methodology to compare, quantify and correct the SWIR databases biases for an accurate integration of data from multiple instruments, which is the focus of the present work.

2 Materials and methods

2.1 Study case

To evaluate the different normalization approaches mentioned above, we have selected four SWIR databases taken from the work conducted by Uribe-Mogollon and Maher (2020) [11] at the Grasshoper porphyry prospect in Montana (Fig. 2). The spectral data correspond to white micas from a suite of eighty-five rock samples presenting distinctive phyllic alteration events and collected with two SWIR devices (1) OreXpress^{TM}, manufactured by Spectral Evolution, and (2) TerraSpec^{TM} 4 Hi-Res Mineral Spectrometer, manufactured by Analytical Spectral Devices. For ensuring data reproducibility, it was employed in this research a measurement methodology for OreXpress™ and TerraSpec™ devices with a 20-mm probe window, conducting 50 repeat measurements on six samples under identical conditions. Additionally, the remaining samples underwent 10 repeat measurements, contributing to the overall reliability of the study. Detailed information regarding the methodology and context of the samples can be found in the cited source.

The four databases in this study are: 1) OreXpress^{TM} 13MAR2018 (number of samples=85), 2) TerraSpec^{TM} 14FEB2018 (number of samples =35), 3) TerraSpec^{TM} 25FEB2018 (number of samples =85), and 4) TerraSpec^{TM} 13MAR2018 (number of samples =85).

All databases present the same number of samples (85 samples), except for TerraSpec^{TM} 14FEB2018 where only a sub-suite of 35 samples was measured.

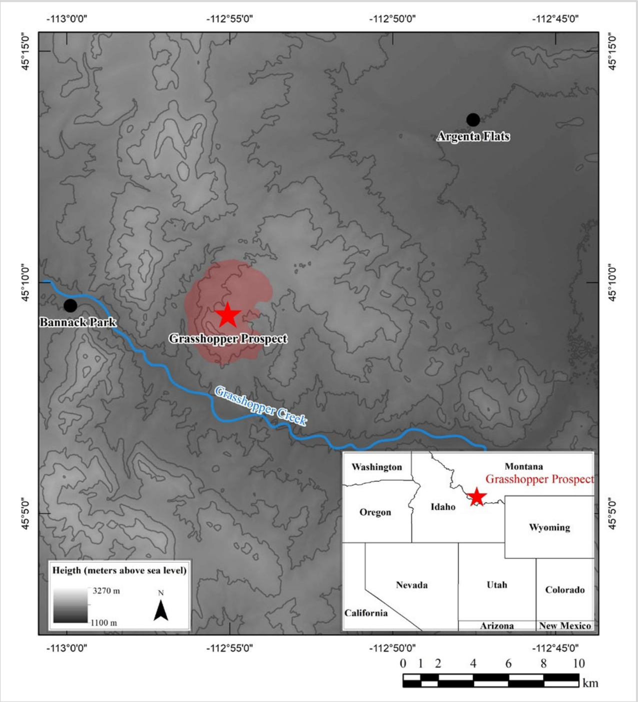

Figure 2. Location of the Grasshopper Porphyry prospect in Montana. In color red is signaled the boundaries of the porphyr.
Source: adapted from Uribe-Mogollon and Maher (2020).

2.2 Phyton script: data arrangement, visualization plots and statistical tables

The spectral data of each sample consists of a CSV file containing the reflectance percentage of each wavelength in the range of 1300 to 2500 nm. Since the wavelength position of the white mica Al-OH spectral absorption feature is at ~2200 nm, a Python script was built to extract it in each sample in all databases. This would be the wavelength position where the reflectance percentage is closer to 0%. In addition to the extraction of the spectral feature, the Python script creates frequency tables, histograms, and Kernel density estimation (KDE) plots to visualize the distribution of the data in each database, and a statistical table with parameters such as the mean, median, standard deviations, quantiles, minimum, and maximum values. All the above products are essential to the selection of the database of reference and the normalization method to be applied.

It is noteworthy that KDE is preferred for a better representation of the distribution of the data [13-15], and it is used in this study to compare between the reference database and the normalized databases. The advantage of KDE is that, contrary to histograms, the shape of the data is not lost by the placement of bins, as kernels are centered on each data point. For more information about the KDE calculation and bandwidth selection please refer to Appendix A.

2.3 Application of statistical methods

Statistical normalization refers to data transformation processes in which databases on different value ranges are adjusted to one in reference, or changes in the probability of distribution are made to align the databases to one in reference. The following are the statistical normalization methods used in this study.

2.3.1. Rescaling normalization

Also known as min-max normalization, this method consists in the range normalization of independent variables, adjusting the minimum and maximum values in a database to one in reference. It is expressed as the eq. (1) [12]:

$$x(normalized) = \frac{(x-y)*(b-a)}{(z-y)} + a, \tag{1}$$

where the value of a data point in the initial database is represented by (x), the mini-mum and maximum values from the reference database are represented by (a) and (b) respectively, and the initial database minimum and maximum value are correspondingly represented by (y) and (z).

2.3.2. Mean normalization

The mean normalization method allows the transformation and fit of a database to one in reference by using a conjunction of its means. It is expressed in the eq. (2) [12]:

$$x(normalized) = \mu_R + (X - \mu_i), \tag{2}$$

where the value of a data point in the initial database is represented by (x), the mean value from the reference database is represented by (μR), and the initial database mean value is represented by (μi).

2.3.3. Z-score normalization

Also known as a standard score, this normalization technique allows the standardization of the standard deviation in relation to the mean of the reference database. It is expressed in the eq. (3) [12]:

$$x(normalized) = \mu_R + (X - \mu_i) * \sigma_R/\sigma_i, \tag{3}$$

where the value of a data point in the initial database is represented by (x), the mean value from the reference database is represented by (μR), the initial database mean value is represented by (μi), and the standard deviation from the reference database and the initial database is represented by (σR) and (σi) respectively.

Table 1.
Descriptive statistics for the SWIR databases from Grasshopper.

Statistical parameter	Terraspec^{TM} 14FEB2018	Terraspec^{TM} 25FEB2018	Terraspec^{TM} 13MAR2018	OreXpress^{TM}
Count	35.0	85.0	85.0	85.0
Ave	2202.1	2201.6	2201.5	2203.6
Median	2202.0	2201.0	2201.0	2203.0
Min	2197.0	2194.0	2194.0	2197.0
Max	2207.0	2208.0	2208.0	2210.0
StdDev	2.9	3.9	4.0	3.7
Q1	2199.0	2198.0	2198.0	2200.0
Q3	2205.0	2205.0	2205.0	2207.0

Source: the authors.

3 Results

3.1 Descriptive statistics

Table 1 shows the descriptive statistics for each of the SWIR databases from the Grasshopper prospect. TerraSpec^{TM} 25FEB2018 and TerraSpec^{TM} 13MAR2018 present almost identical statistical parameters. However, the average and median values of these are 2 nm lower than OreXpress^{TM}. The minimum and maximum values for the TerraSpec^{TM} 25FEB2018 and TerraSpec^{TM} 13MAR2018 are 2194 and 2208 nm respectively, whereas for the OreXpress^{TM} these are 2197 and 2210 nm.

TerraSpec^{TM} 14FEB2018 has an average and median values of 2202 nm, which is 1 nm longer than TerraSpec^{TM} 25FEB2018 and TerraSpec^{TM} 13MAR2018, and 1 nm shorter than OreXpress^{TM}. The minimum value for TerraSpec^{TM} 14FEB2018 and OreXpress^{TM} is 2197 nm. However, the maximum value for the TerraSpec^{TM} 14FEB2018 is 2207, which is 3 nm shorter than OreXpress^{TM}.

In terms of the standard deviation, it is observed that TerraSpec^{TM} 14FEB2018 has the lowest value (2.9) among all SWIR databases, but it also has the lowest number of samples (n=35). The best standard deviation between the databases presenting equal number of samples (n=85) is found in the OreXpress^{TM} (3.7). Therefore, this has been selected as the database of reference to use in the normalization methods. Fig. 3 shows the histograms of all TerraSpec^{TM} databases plotted against the OreXpress^{TM}, and Fig. 4 shows all databases plotted as KDE density functions.

3.2 Normalization methods

Fig. 5 shows each TerraSpec^{TM} database normalized to the reference OreXpress^{TM} database using, (A) rescaling normalization, (B) mean normalization, and (C) z-score normalization. The descriptive statistical parameters of the normalized database are shown in Tables 2, 3, and 4. After the application of the rescaling normalization method, Fig. 5-A and Table 2 present the TerraSpec^{TM} databases with the respectively normalized 2197 nm and 2210 nm minimum and maximum values of the OreX-press^{TM} database. For the mean normalization method, it can be observed in Fig. 5-B and Table 3 that all the TerraSpec^{TM} databases means coincide with the OreXpress^{TM} (2203.6 nm) mean. Finally, Fig. 5-C and Table 4 show the z-score normalization where all TerraSpec^{TM} databases present a standard deviation of 3.7 and a mean value of 2203.6 nm like the OreXpress^{TM} database.

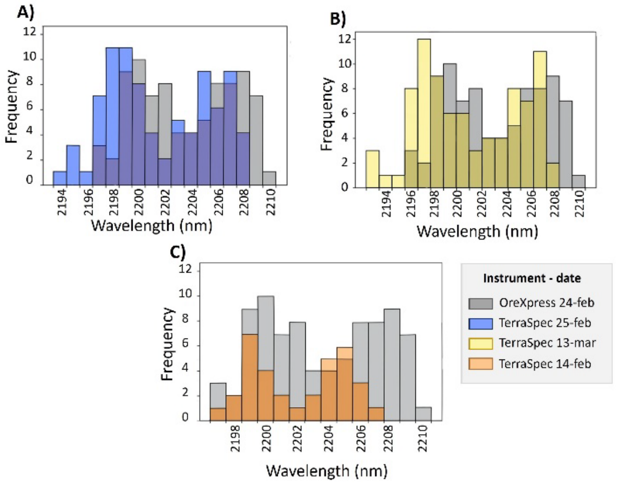

Figure 3. This Figure displays histograms illustrating the minimum wavelength positions for each TerraSpec™ sampling dataset on specific dates, comparing them to the minimums from the OreXpress™ dataset. The Figure contrasts the OreXpress™ database (shown in grey) with three TerraSpec™ datasets: (A) TerraSpec™ 25FEB2018 (depicted in blue), (B) TerraSpec™ 13MAR2018 (in yellow), and (C) TerraSpec™ 14FEB2018 (in orange). Notably, there is a consistent 2 nm difference observed between the two devices.
Source: authors.

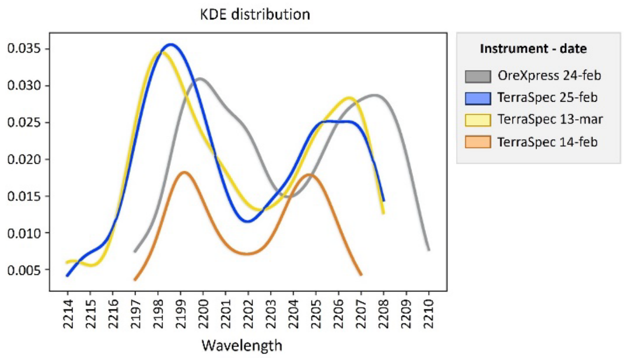

Figure 4. Graph presenting the KDE distribution for each database from the Grasshopper porphyric prospect. In the Figure is compared the OreXpress™ database (grey) against TerraSpec™ 25FEB2018 (blue), TerraSpec™13MAR2018 (yellow), and TerraSpec™ 14FEB2018 (orange). It can also be observed the generalized 2 nm discrepancy between both devices.
Source: authors.

3.3 Method viability determination

Results of the K-S two sample test are presented in Table 5 as a comparison between each TerraSpec^{TM} database versus OreXpress^{TM} database. In this table is presented the null and alternative hypothesis validity between each normalization method applied to each sampling date from the TerraSpec^{TM} databases referenced to the OreXpress^{TM} database. The acceptance of these hypotheses is established by the (p) value, which is an equivalence for the maximum difference between the cumulative distributions and the sample size. With the establishment of a 0.05 acceptance value (α) (this value is commonly used in the scientific community for this type of test), its determined that both databases have similar probability of distribution when the (p) value is equal to or larger than alpha (α) (H0: p≥a).

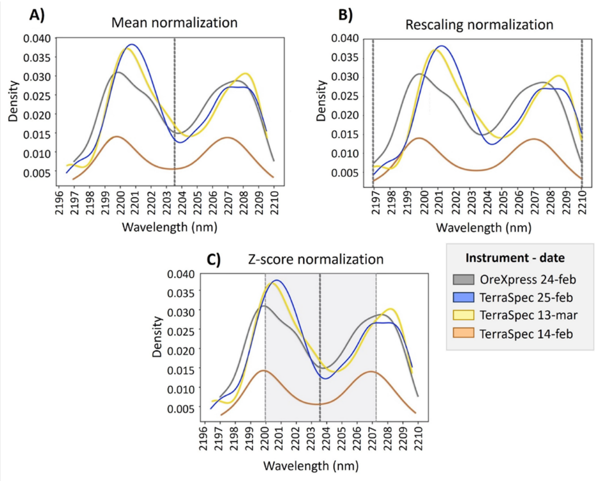

Figure 5. KDE plots of all TerraSpec™ databases from the Grasshopper porphyric prospect normalized to the reference OreXpress™ database by: (A) mean normalization with a dotted line representing the normalized mean value; (B) rescaling normalization with dotted lines representing the minimum and maximum values selected as normalization reference; and (C) z-score normalization with a grey area representing the normalized standard variation value.
Source: authors.

Table 2.
Statistical parameters after rescaling normalization.

Statistical parameter	Terraspec^{TM} 14FEB2018 rescaling	Terraspec^{TM} 25FEB2018 rescaling	Terraspec^{TM} 13MAR2018 rescaling	OreXpress^{TM} rescaling
Count	35.0	85.0	85.0	85.0
Ave	2203.6	2204.1	2204.0	2203.6
Median	2203.5	2203.5	2203.5	2203.0
Min	2197.0	2197.0	2197.0	2197.0
Max	2210.0	2210.0	2210.0	2210.0
StdDev	3.8	3.6	3.7	3.7
Q1	2199.6	2200.7	2200.7	2200.0
Q3	2207.4	2207.2	2207.2	2207.0

Source: the authors.

Table 3.
Statistical parameters after mean normalization.

Statistical parameter	Terraspec^{TM} 14FEB2018 mean norm.	Terraspec^{TM} 25FEB2018 mean norm.	Terraspec^{TM} 13MAR2018 mean norm.	OreXpress^{TM} mean norm.
Count	35.0	85.0	85.0	85.0
Ave	2203.6	2203.6	2203.6	2203.6
Median	2203.5	2203.0	2203.1	2203.0
Min	2197.0	2196.5	2196.6	2197.0
Max	2210.0	2209.5	2209.6	2210.0
StdDev	3.8	3.6	3.7	3.7
Q1	2199.6	2200.2	2200.3	2200.0
Q3	2207.4	2206.7	2206.8	2207.0

Source: the authors.

Table 4.
Statistical parameters after z-score normalization.

Statistical parameter	Terraspec^{TM} 14FEB2018 z-score norm.	Terraspec^{TM} 25FEB2018 z-score norm.	Terraspec^{TM} 13MAR2018 z-score norm.	OreXpress^{TM} z-score norm.
Count	35.0	85.0	85.0	85.0
Ave	2203.6	2203.6	2203.6	2203.6
Median	2203.5	2203.0	2203.1	2203.0
Min	2197.1	2196.4	2196.5	2197.0
Max	2209.8	2209.6	2209.6	2210.0
StdDev	3.7	3.7	3.7	3.7
Q1	2199.6	2200.2	2200.2	2200.0
Q3	2207.3	2206.8	2206.8	2207.0

Source: the authors.

Table 5.
Comparison between the statistical results from the application of the K-S two simple test to each TerraSpec^{TM} sampling date database versus OreXpress^{TM} database.

P-Value	OreXpress ^{TM} vs Terraspec^{TM} 14FEB2018	OreXpress ^{TM} vs Terraspec^{TM} 25FEB2018	OreXpress ^{TM} vs Terraspec^{TM} 13MAR2018
Non-normalized data	0.038	0.017	0.017
Rescaling norm.	0.705	0.366	0.366
Mean norm.	0.705	0.366	0.477
Z-score norm.	0.705	0.366	0.477

Source: the authors.

The pre-normalization TerraSpecTM 25FEB and 13MAR database dates present (p) values results of 0.017 in the comparison against OreXpressTM database, and for TerraSpecTM 14FEB this result is 0.038 in comparison to OreXpressTM database. After the application of the normalization methods the test presents a (p) value of 0.366 in all the normalization methods for TerraSpecTM 25FEB vs OreXpressTM, and in the TerraSpecTM 13MAR for the rescaling method. The z-score and mean normalization methods result in a (p) value of 0.477 in the TerraSpecTM 13MAR compared to the OreXpressTM database. In all the normalization methods the TerraSpecTM 14FEB obtained (p) values of 0.705.

4 Discussion

Fig. 6 presents a graphical summary of the transformation methods in the form of boxplots. In these plots, the general statistical behavior can be observed by looking at the minimum, maximum, median, and mean values (shown as white dots). In Fig. 6-A, it is possible to observe the non-normalized data, where the disparity is evidenced in the mean values. For example, the OreXpressTM database mean deviates from the TerraSpecTM 14FEB database by 1.5 nm, and from the TerraSpecTM 25FEB database by 2.0 nm.

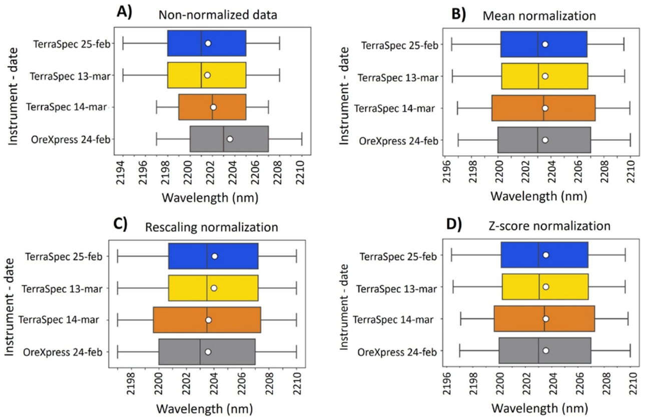

Figure 6. Boxplots from the Grashopper porphyric prospect presenting: (A) non-normalized data, (B) data after the mean normalization method, (C) data after rescaling normalization method, and (D) data after z-score normalization method. In blue TerraSpec™ 25FEB, yellow TerraSpec™ 13MAR, orange TerraSpec™ 14FEB, and grey OreXpress™.
Source: authors.

In Fig. 6-B, it is presented the TerraSpecTM databases transformation by the rescaling normalization method. In this case, we used the OreXpressTM 2197 nm minimum and 2210 nm maximum values as reference. TerraSpecTM 25FEB and TerraSpecTM 13MAR initially had 2194 nm as minimum value and 2208 nm as maximum value. In the same way, for TerraSpecTM 14FEB these values initially were 2197 nm and 2207 nm, respectively. By adjusting the range of the databases with respect to the OreXpressTM, it was observed that the mean values of the TerraSpecTM databases changed. The mean value adjustment is approximately 0.5 nm or 67% for the TerraSpecTM 25FEB and TerraSpecTM 13MAR databases in comparison to the OreXpressTM database. For TerraSpecTM 14FEB, the mean value changed from 2201 nm to 2203.6 nm, by extending the maximum value to 2210 nm.

The mean normalization method is shown in Fig. 6-C, where all the TerraSpecTM databases mean, values were set to the OreXpressTM database 2203 nm mean value. By transforming the means, it is observed a displacement in the minimum and maximum values. For example, the initial TerraSpecTM 25FEB which had a 2194 nm minimum and a 2208 nm maximum values were converted to 2196.5 and 2209.5 nm, respectively. The positive shift corresponds to ~2 nm of difference between the original and reference mean values. This same shift is observed in the other databases.

The z-score normalization is observed in Fig. 6-D. This method places the OreXpressTM 2203.6 nm mean value and 3.7 nm standard deviation value for reference in all the TerraSpecTM databases. As a consequence, there is an approximate difference of 0.1 nm in the minimum and maximum values for all normalized TerraSpecTM databases. In Fig. 5-C, the darker gray area that goes between 2199.9 and 2207.3 nm represents one standard deviation from the mean value, equivalent to 63% of the TerraSpecTM datapoints.

In general, similar results are observed in the application of the z-score and mean normalization methods. It can be noted a successful data adaptation after the application of the normalization methods, being in general the highest value for improvement for the TerraSpecTM 14FEB2018 and the lowest for the TerraSpecTM 25FEB2018. However, the rescaling normalization method is less reliable as it presents bigger differences in the distribution of the accumulated database values. This is especially visible in the gap at the “tails” in comparison of the mean normalization method (Fig. 7). In addition, the K-S two sample test supports that the mean and z-score normalization methods are the most reliable (Table 5).

Comparing TerraSpecTM 14FEB2018 database results in the Fig. 6 is noticeable how the database gets excessively deformed consequence of application of these normalization methods in a database which its wavelength range of distribution has a large difference in comparison to the base normalization database (in this case TerraSpecTM 14FEB2018 is 23,1% shorter than OreExpressTM). Thus, as the method reshapes the database to be contained in the base normalization database it is not recommended the application of the normalization method in this type of databases.

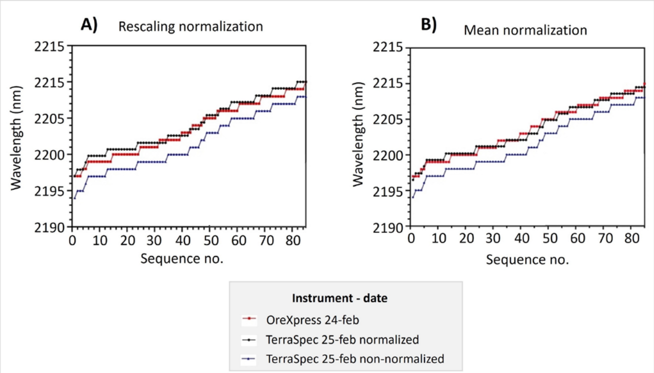

Figure 7. Low-to-high plot of variation comparing the sequence numbers from the wavelength position from the Grashopper porphyric prospect database after: (A) rescaling normalization method, and (B) mean normalization method. Each sample was assigned a sequential number.
Source: authors.

5 Conclusion

The proposed goal of this research of developing a methodology that allows the normalization of databases from different uncalibrated spectrometers was achieved. It was developed a Python scrip that filters and organizes SWIR databases according to the desired absorption feature range and presents the results in visually convenient figures and tables with common statistical parameters. It was applied and discussed the rescaling normalization method, mean normalization method, and the z-score normalization method. In general, the most reliable methods to normalize are the mean normalization and z-score methods, because these unify the mean value and re-solve efficiently the data deviation as presented in the methods discussion. Having both the highest (p) values in the K-S two sample test and presenting the lowest relative deformation in the wavelength position variation comparison between databases. It is expected that the findings of this work will have a great impact in the mineral exploration industry by allowing better processing of SWIR databases and providing a tool that permits the creation of a normalized and more robust composite that is useful during the interpretation process of SWIR data as vectoring analysis. Future works should focus on the practical application of this methodology, in addition to other ab-sorption features and minerals.

References

[1] Cohen, J.F., Compositional variations in hydrothermal white mica and chlorite from wall-rock alteration at the Ann-Mason porphyry copper deposit, Nevada, 2011.

[2] Halley, S., Dilles, J.H., and Tosdal, R.M., Footprints: hydrothermal alteration and geochemical dispersion around porphyry copper deposits. SEG Discovery, (100), pp. 1-17, 2015. DOI: https://doi.org/10.5382/SEGnews.2015-100.fea

[3] Clark, R.N., King, T.V., Klejwa, M., Swayze, G.A., and Vergo, N., High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research: Solid Earth, 95(B8), pp. 12653-12680, 1990. DOI: https://doi.org/10.1029/JB095iB08p12653

[4] Pontual, S., Merry, N., and Gamson, P., Spectral interpretation field manual, G-MEX edition 3. Spectral Analysis guides for mineral exploration. 1997.

[5] Thompson, A.J., Hauff, P.L., and Robitaille, A.J., Alteration mapping in exploration: application of short-wave infrared (SWIR) spectroscopy. SEG

Discovery, (39), pp. 1-27, 1999. DOI: https://doi.org/10.5382/SEGnews.1999-39.fea

[6] Chang, Z., Hedenquist, J.W., White, N.C., Cooke, D.R., Roach, M., Deyell, C.L., ... & Cuison, A.L., Exploration tools for linked porphyry and epithermal deposits: example from the Mankayan intrusion-centered Cu-Au district, Luzon, Philippines. Economic Geology, 106(8), pp. 1365-1398, 2011. DOI: https://doi.org/10.2113/econgeo.106.8.1365

[7] Xiao, B., Chen, H., Hollings, P., Wang, Y., Yang, J., and Wang, F., Element transport and enrichment during propylitic alteration in Paleozoic porphyry Cu mineralization systems: Insights from chlorite chemistry. Ore Geology Reviews, 102, pp. 437-448, 2018. DOI: https://doi.org/10.1016/j.oregeorev.2018.09.020

[8] Uribe-Mogollon, C., and Maher, K., White mica geochemistry of the Copper Cliff porphyry Cu deposit: Insights from a vectoring tool applied to exploration. Economic Geology, 113(6), pp. 1269-1295, 2018. DOI: https://doi.org/10.5382/econgeo.2018.4591

[9] Neal, L.C., Wilkinson, J.J., Mason, P.J., and Chang, Z., Spectral characteristics of propylitic alteration minerals as a vectoring tool for porphyry copper deposits. Journal of Geochemical Exploration, 184, pp. 179-198, 2018. DOI: https://doi.org/10.1016/j.gexplo.2017.10.019

[10] Chang, Z., and Yang, Z., Evaluation of inter-instrument variations among short wavelength infrared (SWIR) devices. Economic Geology, 107(7), pp. 1479-1488. 2012. DOI: https://doi.org/10.2113/econgeo.107.7.1479

[11] Uribe-Mogollon, C., and Maher, K., White mica geochemistry: Discriminating between barren and mineralized porphyry systems. Economic Geology, 115(2), pp. 325-354, 2020. DOI: https://doi.org/10.5382/econgeo.4706

[12] Silverman, B.W., Density estimation for statistics and data analysis. Routledge, New York, 2018, 176 P. DOI: https://doi.org/10.1201/9781315140919

[13] Danese, M., Lazzari, M., Murgante, B., Integrated geological, geomorphological and geostatistical analysis to study macroseismic effects of 1980 Irpinian Earthquake in urban areas (Southern Italy). In: Gervasi, O., Taniar, D., Murgante, B., Laganà, A., Mun, Y., Gavrilova, M.L., (eds), Computational Science and Its Applications – ICCSA 2009. ICCSA 2009. Lecture Notes in Computer Science, vol 5592. Springer, Berlin, Heidelberg. 2009. DOI: https://doi.org/10.1007/978-3-642-02454-2_4

[14] Barnett, R.M., and Deutsch, C.V., Multivariate imputation of unequally sampled geological variables. Mathematical Geosciences, 47(7), pp. 791-817, 2015. DOI: https://doi.org/10.1007/s11004-014-9580-8

[15] Tian, X., Kong, Y., Gong, Y., Huang, Y., Wang, S., and Du, G., Dynamic geothermal resource assessment: integrating reservoir simulation and Gaussian Kernel density estimation under geological uncertainties. Geothermics, 120(103017), art. 103017, 2024. DOI: https://doi.org/10.1016/j.geothermics.2024.103017

J.C. Paredes, he received the BSc. in Geology in 2022, and a postgraduate degree in Project Management in 2024, both from the Eafit University in Medellín, Colombia. He began his professional career working from 2022 to 2024 with the Calidra Group in mineral exploration and control for calcium mineral deposits. Currently, he is focusing his interests on information interpretation and geological exploration.
ORCID: 0009-0000-8121-4446

Y.C. Trigos, he received the BSc. in Geology in 2022, from the Eafit University in Medellín, Colombia. He began his professional career working with Colas Quebec in Canada in the in the infrastructure sector, asphalt and quarries quality control.
ORCID: 0009-0004-1846-8231

C. Uribe-Mogollon, is an economic geologist with expertise in mineral geochemistry and hydrothermal alteration for mineral exploration. He earned his PhD. in Earth and Environmental Sciences from the New Mexico Institute of Mining and Technology in 2019. His research focused on porphyry vectoring techniques in phyllic-altered rocks, identifying geochemical markers in white micas that serve as indicators of proximity to mineralized ore bodies.
ORCID: 0009-0004-6036-2529

Appendix A

Kernel density estimation method calculation

Kernel is a non-parametric estimation technique which uses a weighting function to produce a continuous setting as a probability density function from a random variable. It produces a function for each datapoint that satisfies eq. (A1-A2):

a) Normalization: $\int_{-\infty}^{\infty} K(u)du = 1;$ (A1)

b) Symmetry: K(-u) = K(u) for all values of u. (A2)

The normalization of the kernel (K) parameters allows the construction of the kernel density estimation (KDE) function, and the symmetry is to ensure that the average of the KDE is the same as the original data sample.

With the weighting obtained for each datapoint, it is calculated the density of the probability with the eq. (A3) [12].

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \quad (A3)$$

Where the density distribution ($\hat{f}$) is the final function, the kernel (K) is a window function, the bandwidth (h) is the smoothing parameter of the function, and the number of datapoints (n) is determined by the data sample.

The bandwidth for the KDE selected is 0.5 as it reduces the mean integrated squared error. It was obtained using the rule-of-thumb [12] bandwidth estimator at the OreXpress™ dataset, resulting as an average of the application in the Gaussian-like distribution shapes produced by the dataset division by its mean and choosing the one with the lowest value.

The result is a function that estimates the probability density of the initial variables, allowing smoothing the data and providing the probability shape of each dataset.