doi:10.1017/S1431927622002793
Microsc. Microanal. 28 (Suppl 1), 2022
© Microscopy Society of America 2022

A Comparison of Quartz Source Determination Using Supervised Machine Learning verses a Traditional Ratio Approach

Angela Halfpenny[1*], Meaghan Wetherell[2], and Craig Scrivner[1].

1. Department of Geological Sciences, Central Washington University (CWU), Ellensburg, WA, USA.
2. School of Information, University of Arizona, Tucson, Arizona, USA.
* Corresponding author: angela.halfpenny@cwu.edu

Quartz is an extremely common mineral that can display a wide range of colours and textures and is important for industry as it is used to make oscillators for watches, televisions and GPS equipment. Quartz growth can also be associated with ore minerals like gold, silver and copper. For each colour of quartz there are multiple locations, or sources, around the world where they grow. Sometimes a particular texture is characteristic of a specific source location, but most of the time this is not the case. Archaeological samples, such as obsidian arrowheads, have been measured using portable x-ray fluorescence spectrometry (pXRF) to determine the elemental composition and various trace element ratios have been used to determine the source location of the rock source the arrowheads have been made from [1].

This project aims to continue determining the characteristic trace element fingerprint exhibited by blue quartz gemstones from different source locations [2]. This research uses a non-destructive, Bruker Tracer 5i, portable X-ray fluorescence spectrometer (pXRF) to measure the trace element signature of blue quartz samples [3]. As most of the samples used for this research were loaned from various jewellery companies, the Kittitas County Museum, the Central Washington University (CWU) geology department collection and from private individuals, no sample preparation was performed before pXRF analysis. Each sample was first photographed, then measured using calipers and then the best side of the sample (flattest, showing least amount of post-mineralization flaws) was analysed using the pXRF for major and trace elements using different operating conditions for the instrument (Fig.1).

After measuring hundreds of blue quartz samples with known origins, like the Ellensburg Blues and Blue Candy samples from Washington [4], Holley Blues from Oregon, USA, from Mexico, Indonesia and Africa, the results show a similar trace element signature with the concentrations of the trace elements varying between locations. The results from the pXRF analysis were used to develop diagnostic trace element ratios that are associated with each source of blue quartz and we wrote software to apply those ratios to determine sources for new samples. With this traditional approach, we had to manually decide which trace elemental ratios are of geological significance to determine one source vs another (Fig. 2).

Another approach to analyzing pXRF data is make the computer do the work for you through the process of supervised machine learning. We built a RandomForest model in R using the most diagnostic elements in the dataset (e.g. Zr, Mn, Zn, and Co). Machine learning algorithms are particularly effective for use on data with overlapping composition, as the algorithm can detect and apply finer-scale differences than human eyes [5]. Currently the model has over 90% accuracy in identification of the four groups with sufficient sample sizes (>10), with most mistakes being the misclassification of Blue Candy specimens as Ellensburg Blues (Fig. 3). Both models require further testing and refinement and then the models can be used to provide source identification for samples from unknown locations. A future is to invite people to get their blue quartz samples analysed to test the software’s ability to determine the samples source location and compare the machine learning vs ratio approaches for performing the source determination.

https://doi.org/10.1017/S1431927622002793 Published online by Cambridge University Press

References
[1] A. B.M. De Francesco, G.M. Crisci, in “X-ray fluorescence spectrometry (XRF) in Geoarchaeology”, ed. M.S. Shackley, (Springer) p. 81-108.
[2] A. Halfpenny, Microscopy and Microanalysis. 25 (S2) (2019) p. 2480-2481.
[3] R.H. Tykot, Applied Spectroscopy, 70 (1) (2016) p. 42-56.
[4] J.P. Thomson, “Ellensburg Blue” (Kittitas County Museum) p. 1-22
[5] J.L. McGuire, Journal of Mammalogy, 92(6) (2011) p. 1383-1394

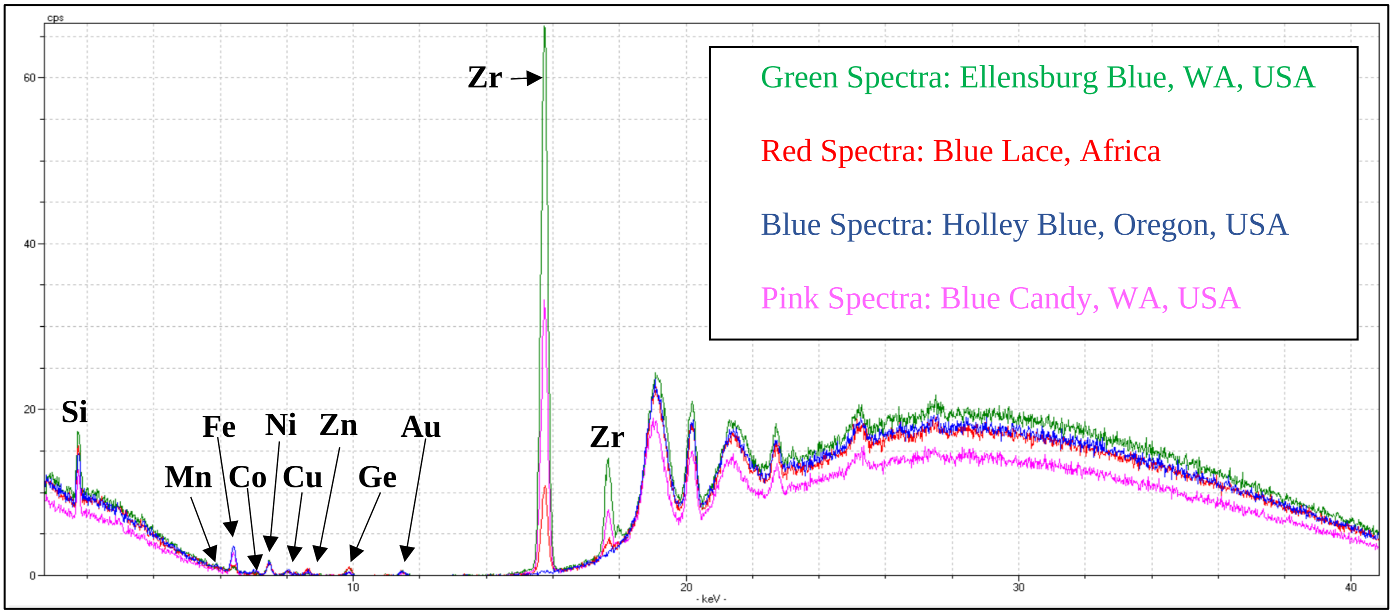

Figure 1: Collected x-ray fluorescence spectra for blue quartz samples from known sources. The spectra exhibit peaks of similar trace elements but the concentration varies between samples. Zr is the most important element followed by Mn, Co, Ni, Cu, Zn ,Ge, Au and Fe (Fe is often increased by post-mineralisation affects).

<table>
<tr><th>Analysis Number</th><th>pXRF Method</th><th>Source Identifications</th></tr>
<tr><td>E001a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E010a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue</td></tr>
<tr><td>E011a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue</td></tr>
<tr><td>E013a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue | Holley Blue, Oregon, USA</td></tr>
<tr><td>E014a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue | Holley Blue, Oregon, USA</td></tr>
<tr><td>E015a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue</td></tr>
<tr><td>E016a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue | Holley Blue, Oregon, USA | Red Top Blue, Kittitas County, Washington, USA</td></tr>
<tr><td>E017a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue | Holley Blue, Oregon, USA | Red Top Blue, Kittitas County, Washington, USA</td></tr>
<tr><td>E018a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue | Holley Blue, Oregon, USA | Red Top Blue, Kittitas County, Washington, USA</td></tr>
<tr><td>E019a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E021a</td><td>Trace</td><td>Ellensburg Blue | Holley Blue, Oregon, USA</td></tr>
<tr><td>E022a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E023a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E024a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E025a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E026a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E027a</td><td>Trace</td><td>Blue Candy, Skagit County, Washington, USA | Ellensburg Blue | Holley Blue, Oregon, USA</td></tr>
<tr><td>E028a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E029a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E030a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E031a</td><td>Trace</td><td>Ellensburg Blue | Holley Blue, Oregon, USA | Red Top Blue, Kittitas County, Washington, USA</td></tr>
<tr><td>E033a</td><td>Trace</td><td>Blue Lace, Africa | Ellensburg Blue | Holley Blue, Oregon, USA</td></tr>
<tr><td>E034a</td><td>Trace</td><td>Ellensburg Blue | Holley Blue, Oregon, USA</td></tr>
<tr><td>E035a</td><td>Trace</td><td></td></tr>
<tr><td>E036a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E037a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E039a</td><td>Trace</td><td>Ellensburg Blue | Holley Blue, Oregon, USA</td></tr>
<tr><td>E040a</td><td>Trace</td><td>Ellensburg Blue</td></tr>
<tr><td>E041a</td><td>Trace</td><td>Ellensburg Blue | Holley Blue, Oregon, USA</td></tr>
</table>

Figure 2: Results from ratio based source determination. Showing issues with multiple source possibilities for the same sample, with further testing and refinement of the code this misindexing should be reduced. As more samples from known sources are measured this will decrease the number of unknown samples that the code does not find a source match for like E035a

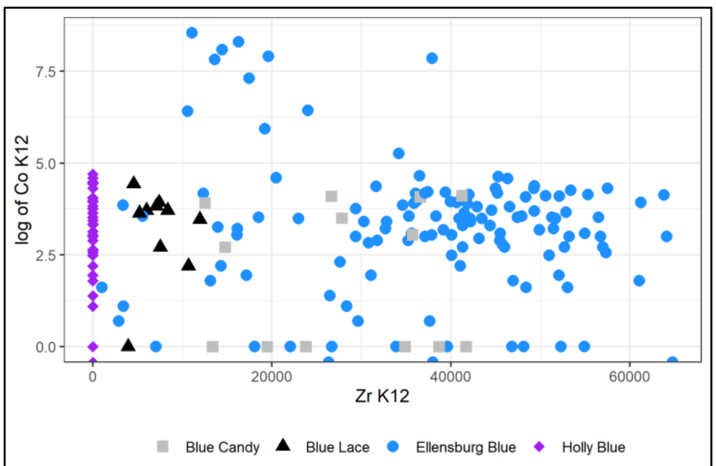

Figure 3: Scatterplot of two important elements showing separation within the four largest sample categories. Oregon Holley Blues are easy to separate just using Zirconium (Zr), but Blue Candy and Blue Lace are harder to discriminate visually. Machine Learning models can take large overlapping datasets like this and use them to separate and sort samples into defined categories.

https://doi.org/10.1017/S1431927622002793 Published online by Cambridge University Press