Multi-Element Analysis of Minerals using Laser Ablation Inductively Coupled Plasma Time Of Flight Mass Spectrometry and Geochemical Data Visualization using t-Distributed Stochastic Neighbor Embedding: Case Study on Emeralds

Hao A. O. Wang,[1,*] Michael S. Krzemnicki[1,2]

1 Swiss Gemmological Institute SSEF, CH-4051, Basel, Switzerland

2 Department of Environmental Sciences, Mineralogy and Petrology, University of Basel, CH-4056, Basel, Switzerland

hao.wang@ssef.ch; michael.krzemnicki@ssef.ch;

Supplementary Table S-1. Number of emerald samples from various deposits analyzed in this study.

Geological Provenance	**Number of Samples**
Afghanistan (Classic Material)	18
Afghanistan (Colombia-Like Material)	9
Brazil	24
Colombia	62
Ethiopia	18
Tanzania	10
Zambia	27
Total Number of Samples	**168**

Supplementary Table S-2. LA-ICP-TOF-MS instrument operating parameter.

Parameters	**LA-ICP-Time-Of-Flight-MS**
Laser model and type	NWR 193 UC, Elemental Scientific Lasers
Laser wavelength (nm)	193
Fluence ($J{\cdot}cm^{-2}$)	5.6
Repetition rate (Hz)	20
Ablation duration (s)	30
Spot diameter (μm)	75 for NIST SRMs and 100 for emeralds
Carrier gas flow *(Helium, L·min-1)*	0.8-0.9
Mass spectrometer *model and type*	*icp*TOF, TOFWERK (modified from iCAP Qc, Thermo Fisher Scientific)
RF power (W)	1250-1400
Make-up gas flow *(Argon, $L{\cdot}min^{-1}$)*	0.65 - 0.68 (tuned daily)

Mass-to-charge ratio measured (unit for mass-to-charge ratio is Thomson, Th)	full mass spectrum (m/Q = 7 to 255, $^{7}Li^{+}$ to $^{238}U^{16}OH^{+}$)
Total averaging time per output data point (s)	0.151 (averaging 5000 mass spectra)
ThO^{+}/Th^{+}	< 0.6%
$^{238}U^{+} / {}^{232}Th^{+}$	~1
$^{138}Ba^{2+}{}_{calculated} / {}^{69}Ga^{+}$	<3-4%
Notch filter setting at m/Qs (Th)	18, 28, 41

Supplementary Table S-3. Twenty elements used for t-SNE calculation in Figure 4(e).

Li	Be	Na	Mg	Al	K	Ca	Sc
Ti	V	Cr	Mn	Fe	Ni	Zn	Ga
Ge	Rb	Sn	Cs				

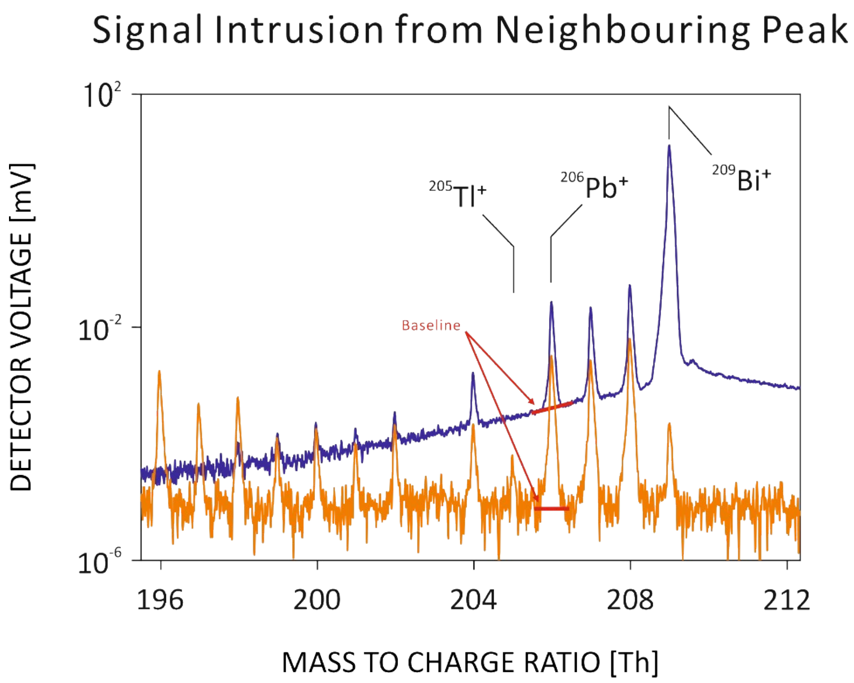

Figure S-1. Signal intrusion from neighboring peak is illustrated in two tourmaline samples, one with high Bi content (blue spectrum) and one with low Bi content (orange spectrum). Baseline of mass spectrum on trace elements such as Pb and Tl can be affected, therefore a baseline correction is necessary. Baselines on $^{206}Pb^{+}$ are demonstrated in both mass spectra in red lines. Detailed description of the baseline correction can be found in the main text.

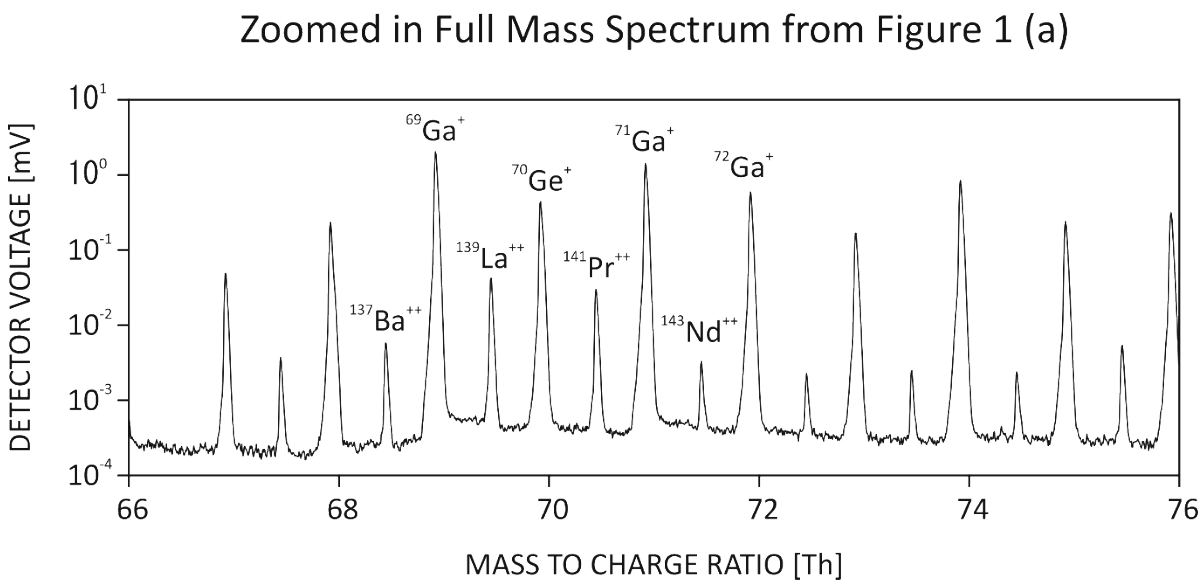

Figure S-2. Zoomed-in mass spectrum from Figure 1(a) focused on m/Q between 66Th and 76Th. Due to ~2000 mass resolving power (MRP) at the middle mass range, doubly charged Ba and lanthanide isotopes at half masses can be resolved. These doubly charged isotopes at half masses are used for mathematical correction of interferences, e.g. on $^{69}Ga^{+}$. A summed mass spectrum is acquired from NIST SRM 610.

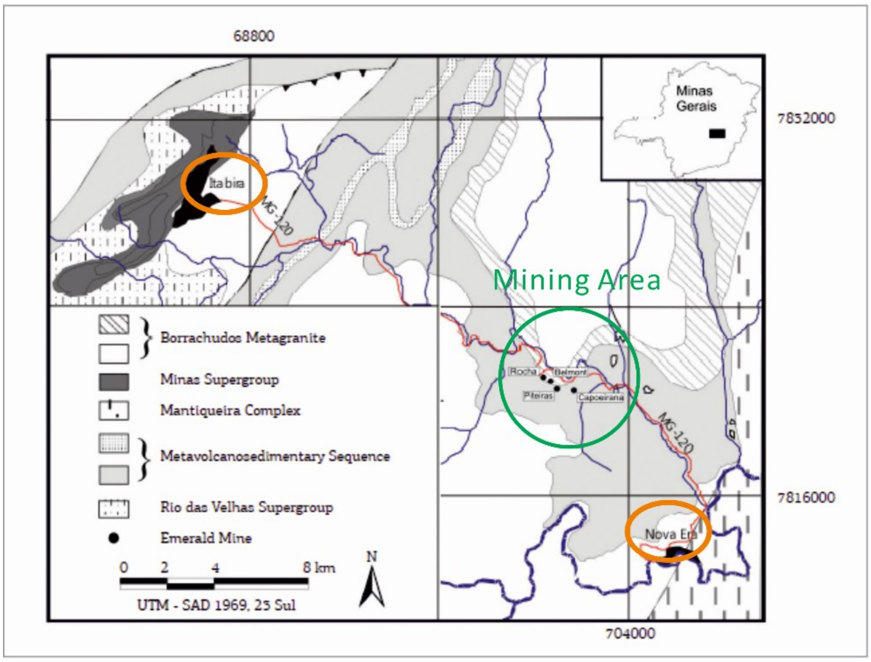

Figure S-3. Geographic provenance of emeralds from Itabira and Nova Era (marked in orange circles) in the state of Minas Gerais, Brazil. These two towns are separated by only 20km. Green circle indicates major mining area, which include Belmont, Capoeirana, Piteiras and Rocha mines. [Adapted from Figure 1 in Ref: Jordt-Evangelista, H.; Lana, C.; Delgado, C.E.R.; Viana, D.J. Age of the emerald mineralization from the Itabira-Nova Era District, Minas Gerais, Brazil, based on LA-ICP-MS geochronology of cogenetic titanite. *Brazilian Journal of Geology* **2016**, 46, 427–437, with permission]. Small mines also spread in this area, however not marked.

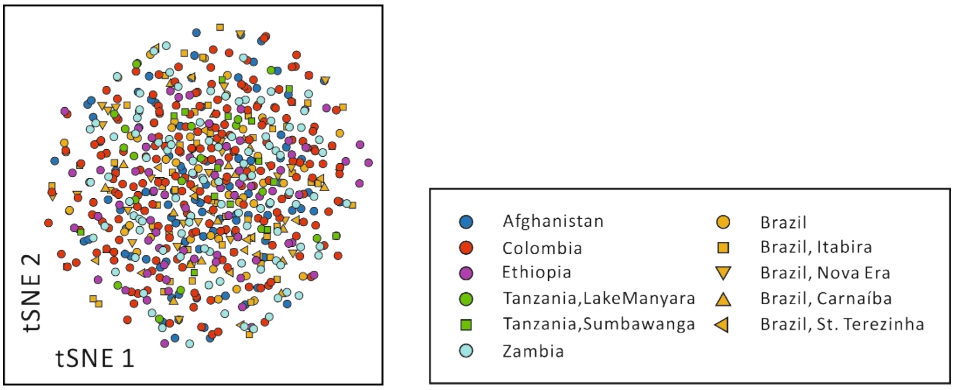

Figure S-4. The t-SNE scatter plot of emerald using all 20-element concentrations. All concentrations are substituted by random values generated from the LOD log-normal histogram of each isotope respectively. The plot has shown no obvious features indicating geographic provenance related clustering. It is to demonstrate that clustering of emeralds shown in Figure 4(e) is due to the similarity of elemental concentrations.

Data Availability Statement

The full datasets and scripts generated during the current study are not available to access as they contain confidential information. A step-to-step data evaluation process is described in detail in the text, and t-SNE calculation packages can be found in various programming languages, such as Python and R.