UCLouvain

École polytechnique de Louvain

Data fusion of Raman hyperspectral imaging data using multi-block chemometric methods for the characterization of copper preservation

Author : Benjamin Wasserman
Supervisors : Faten Ammari, Arnaud Delcorte, Delphine Neff
Readers : Christine Dupont, Benoît Hackens
Academic year 2025–2026
Master [120] in Chemical and Materials Engineering

Abstract

The copper corrosion patina is an intricate chemical matrix consisting primarily of copper, cuprite (Cu_2O), and brochantite ($Cu_4SO_4(OH)_6$); along with several other minority copper salt constituents. The outermost layer, brochantite, is the distinctive blue-green patina associated with copper corrosion [1–3]. Contemporary preservation research is focused on non-toxic, eco-friendly, sol-gel coatings for copper corrosion protection. The coatings for cultural heritage artifacts present a unique challenge, it must preserve the underlying metal, the current corrosion layers, and prevent modification of the aesthetics [2]. Specifically, alkoxysilane sol-gels doped with carboxylic acid compounds have been shown to inhibit the corrosion process. This study intends to evaluate the copper corrosion inhibition treatment of tetraethyl orthosilicate (TEOS) doped with octanoic acid applied to the copper patina surface. Treated copper cross-sections were analyzed by Raman spectroscopy using two different excitation wavelengths (473 nm and 532 nm) to analyze the penetration depth and interactions of the inhibitor within the corrosion layers. The corrosion products and inhibitor produce distinct responses from the two wavelengths to create unique, but complementary "blocks" of data. Machine learning and multi-block techniques exploit this collective relationship to enhance the identification of the corrosion and coating-related phases [4]. The multi-block fusion techniques revealed that the octanoic acid penetrated the corrosion layer and reacted with the brochantite to form copper octanoate. But, this reaction was concentrated on the outer boundary of the layer rather than distributed uniformly throughout the brochantite layer. Comparing the different multi-block methods showed that they produced largely convergent results for this two-block dataset. These findings support the use of dual-excitation Raman imaging combined with multi-block fusion as an accessible, non-destructive strategy for evaluating corrosion inhibitor coatings on cultural heritage copper artifacts.

1. Introduction

Cultural heritage objects are fundamental to the cultural identity of communities, nations, and humanity, serving as tangible anchors that connect past generations to contemporary society [5]. Along with being hallmarks of cultural identity, cultural heritage artifacts are drivers to economic and social welfare. Unfortunately, these objects are subject to degradation from environmental factors, bio-contamination, natural or anthropogenic disasters, and incorrect restoration interventions; degradation often occurs both on macro- and nano-scale [5,6].

Copper has been utilized for thousands of years in both artistic and pragmatic applications, and it remains a crucial material for society [1,4]. The earliest archaeological records utilizing native copper deposits traces back to at least 10,000 BCE to the Epipaleolithic period or the early Neolithic period; primitive copper metallurgy traces back to at least 5,000 BCE to Anatolia [1,7,8]. This vast history has yielded a diverse tapestry of copper objects forming a critical component of the archaeological record and humanity's shared cultural heritage [9,10]. Contributing to this diversity are the eclectic burial and atmospheric environments in which these artifacts are discovered in [11]. Consequently, it is important to understand the physical and mechanical properties of copper corrosion to better preserve the material.

Copper Corrosion

Fundamentally, copper corrosion is a form of wet corrosion, an electrochemical redox reaction that occurs in the presence of oxygenated water [10]. Although visible water may not always be present on the surface of an artifact, a microscopically thin layer of water is almost always present due to ambient atmospheric humidity [10,12]. But, the local environmental conditions heavily influence this process; parameters such as relative humidity, wet/dry cycling, rainfall frequency, and ambient pollutants critically dictate the overall corrosion rate .[10] As a redox reaction, it involves two half-reactions,

the oxidation of the metal at the surface (the electron donating anode, Equation 1) and the reduction of an environmental chemical species (the electron accepting cathode, Equation 2) [10,11,13].

The resulting corrosion layer on the copper is a multilayered tripartite system that forms in successive stages over time. The initial step is the oxidation of copper to form the mineral, cuprite (Cu_2O) (Equation 3), shifting the surface color from a bright salmon-pink color to the dark brown-black color commonly associated with copper [1,3,10,12,14]. In moist, aerated environments this step occurs instantaneously, the copper moves to +1 oxidation state, via the chemisorption and reduction of oxygen [3,10]. The cuprite layer forms a continuous, polycrystalline layer with excellent adherence to the underlying metal [10]. The cuprite prevents water from further reaching the copper's surface [10].

$$4Cu + 2H_2O \rightarrow 2Cu_2O + 4H^+ + 4e^-$$

Equation 1: Anode reaction for copper corrosion [15,16]

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$$

Equation 2: Cathode reaction for copper corrosion [15,16]

$$4Cu + O_2 \rightarrow 2Cu_2O$$

Equation 3: Complete reaction for cuprite during the copper corrosion process [15,16].

However, the degradation process continues with the cuprite, the surface copper ions are subsequently oxidized with the trace atmospheric impurities present, and the copper ion transitions from Cu(I) oxidation state to Cu(II) [3,10,14,16]. The presence of molecular atmospheric pollutants dictate the final Cu(II) species on the corrosion layer [10]. This additional corrosion layer is the blue-green patina associated with copper corrosion caused by the Cu(II) species [2,3,14,17].

The corrosion products produced on the surface of copper can vary depending on time and environmental conditions. In sulfate-rich environments posjnakite ($Cu_4SO_4(OH)_6 \cdot H_2O$), brochantite ($Cu_4SO_4(OH)_6$), chalcantite ($CuSO_4 \cdot 5H_2O$) and antlerite ($Cu_3(SO_4)(OH)_4$) are the frequently occurring compounds, see Equation 4 [1,14,16,18]. Atacamite ($Cu_2Cl(OH)_3$), paratacamite ($Cu_2Cl(OH)_3$) and its derivatives are commonly found in marine chloride-rich environments, see Equation 5 [1–3,14,16]. Other environmental conditions can form other types of copper oxide minerals such as exposure to bird feces; it can also form unique copper oxide products such as cornetite ($Cu_3PO_4(OH)_4$), and moolooite ($Cu(C_2O_4) \cdot n(H_2O)$) [16,19,20]. Gerhardite ($Cu_2(NO_3)(OH)_3$) can be found in presence of nitrous oxides, and malachite in the presence of carbonates ($Cu_2(CO_3)(OH)_2$) [16,19].

$$Cu_2O + SO_4^{-2} + H_2O \rightarrow Cu_x(SO_4)(OH)_y * nH_2O$$

Equation 4: The reaction to form the copper sulfate compounds: posnjakite, brochantite, antlerite [16]

$$Cu_2O + Cl^- + H_2O \rightarrow Cu_2Cl(OH)_3$$

Equation 5: The reaction to form the copper chloride compounds: atacamite, paratacamite [16]

Brochantite constitutes the primary component of the long-term corrosion in the Cu(II) layer in most environments, except for marine environments [1–3,14,16]. Brochantite is characterized by high porosity and the presence of numerous cracks [8]. Typically, posjnakite is the first copper oxide to develop, and brochantite develops later often taking years to form. In marine environments, the major constituents being atacamite and paratacamite with smaller amounts of brochantite [1–3,19]. For example, the Statue of Liberty in New York City harbor still contained brochantite in its copper patina despite being surrounded by salt seawater [3,21]. The ever-presence of brochantite in all copper patina is due to its stability; copper hydroxysulphate phases are 100 times more stable than the equivalent chloride and carbonate compounds, and 10,000 times more stable than the equivalent nitrate compounds [3].

But the other copper oxides can still exist in this corrosion layer depending on exposure time, quantity of pollutants, and metastable intermediate compounds that lead to brochantite [3,8]. A scheme of the tripartite copper corrosion layer system is demonstrated in Figure 1. Overall, the copper corrosion layer is complex matrix of different copper compounds.

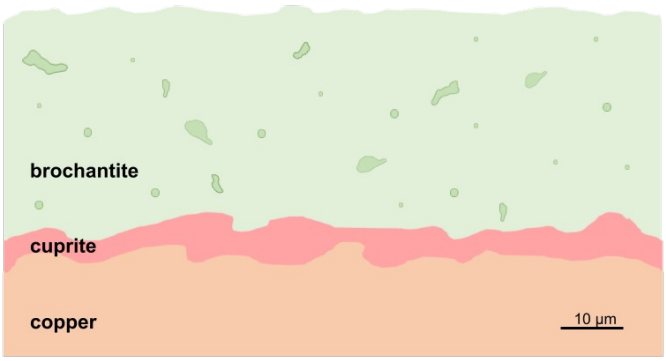


Figure 1: Diagram of the copper corrosion layers. Copper is at the bottom, cuprite is the middle, and brochantite is on top. Brochantite is the hydrophilic, blue-green, porous Cu(II) layer.

Corrosion inhibition and Coatings

Broadly defined, anti-corrosion inhibitors are any chemical substances that are not traditionally present in the corrosion process, but suppresses the corrosion rate and the chemical processes [8]. The inhibitor commonly bonds and modify the metal's surface to out-compete the absorption of corrosion inducing molecules, and creates a thin film over the metal's surface [8]. Anti-corrosion coatings can be classified by their mechanism of inhibition or type of material utilized [4,8,20]. The corrosion inhibitor could use either electrochemical inhibition consisting of either anode, cathode, or a combination of the two to inhibit the corrosion reaction process [2,4,8].

The main material types used specifically for cultural heritage materials are waxes, acrylic resins, nitrocellulose lacquers, carboxylates, fluorinated polymers, and organo-silane compounds [21]. Among these, microcrystalline waxes are preferred among conservators in France [4]. These inhibitors primarily provide a physical barrier, such as a hydrophobic molecular thin film that stops reactants from reaching the metal's surface [4,8]. Recently, there has been a growing awareness of the toxicity of anti-corrosion inhibitors and developing environmentally friendly solutions to replace toxic inhibitors [8,22].

Waxes are the traditional protective treatment applied to metal surfaces as barriers against moisture and oxygen, as well as topcoats over varnishes to reduce gloss and improve appearance. Conservators appreciate waxes for providing a natural finish, though they can darken porous patinas and collect dust easily. Overall, waxes offer relatively poor long-term corrosion protection and

weathering resistance compared to synthetic varnishes, requiring frequent renewal. Under certain humid or chloride-rich conditions, the porous nature of wax layers can even trap moisture and accelerate metal corrosion. Modern heritage conservation primarily uses microcrystalline waxes and polyethylene waxes, while natural waxes are generally avoided due to the formation of damaging organic acids upon aging [23].

Acrylic resins were first introduced to heritage conservation in the 1960s and 1970s; they are thermoplastic polymers valued for their transparency, strong adhesion, mechanical strength, and chemical stability over time. The Paraloid product line dominates conservation field, with Paraloid B-72 being widely favored across Europe for protecting indoor silver, iron, and copper artifacts. For outdoor applications, *Incralac,* a formulation consisting of Paraloid B-44, benzotriazole (BTA) corrosion inhibitor, and UV stabilizers is the industry standard for bronze sculptures and monuments, typically providing a protective service life of 3 to 5 years. Other specialized acrylic variants include Paraloid B-48N, which offers enhanced hardness and abrasion resistance [23].

Nitrocellulose lacquers are polynitrate esters of natural cellulose marketed under trade names such as Agateen, Frigilene, and Ercolene. Historically, they have been primarily deployed as thin protective films on indoor silver objects to prevent tarnishing in atmospheres containing hydrogen sulfide. While effective at preventing tarnish, nitrocellulose lacquers suffer from notable long-term degradation issues, including yellowing, embrittlement, and a loss of solubility in their original solvents over time. Although less common on base metals, additions of organic corrosion inhibitors have been shown to improve their performance on patinated bronzes and iron [23].

Fluorinated polymers (fluoropolymers) feature carbon-fluorine bonds that impart high chemical inertness, hydrophobic properties, and superior resistance to UV radiation and aggressive environmental weathering. Polyvinylidene fluoride (PVDF), such as *Kynar*, is frequently utilized in metallic heritage, often blended with acrylic modifiers or used as a topcoat over acrylic basecoats to overcome its inherent lack of adhesion to bare metal surfaces. Other evaluated systems include fluoroethylene vinyl ether (FEVE) copolymers (*Lumiflon*) and fluorinated acrylic copolymers (*Funcosil AG*). Despite their outstanding weather resistance and high corrosion inhibition, potential drawbacks include susceptibility to water-whitening and difficult reversibility without specialized chemical modifications [23].

Alkoxysilanes have gained considerable traction in recent years as anti-corrosion coatings due to the development of the sol-gel coating process [4]. They are more environmentally friendly compared to traditional chemical treatment processes. Alkoxysilanes can form excellent metallic-siloxane bonds to strength their adhesion to the metallic substrate [4]. Silane sol-gel coatings offer very interesting characteristics such as possibility of application by different techniques, good adhesion, and good penetration [21]. The sol-gel process comprises of four stages; hydrolysis of precursor molecules, polycondensation of monomers to form chains and/or particles, polymer particle growth, agglomeration of the polymer particles to form networks extending through the liquid medium resulting in thickening to eventually form a gel [23]. Typically, it applied to object prior to the gelation stage [23]. The most common application techniques are electrochemical deposition, spray-coating, dip-coating, or spin-coating. The sol-gel process is user-friendly for deposition on substrates with large areas with simple, inexpensive equipment [23].

Recent research has been conducted on alkoxysilanes doped with carboxylic acid compounds due to their non-toxicity and anti-corrosion properties [4]. They are fatty acid derivatives extracted from vegetable oils. The current application technique for carboxylic acids requires their dissolution in a solvent and the submersion of the metallic artifact; it is time-consuming and not well-suited for outdoor applications, the sol-gel coating process eliminates these issues [4].

The carboxylic acid in the coating penetrates the pores of the brochantite and reacts with the copper ions in the corrosion layer to form copper carboxylates, as illustrated in Figure 2. The hydrophilic carboxylic groups bonds to the copper species with the non-polar hydrophobic alkyl projecting outwards. It forms an insoluble, hydrophobic layer on the corrosion layer that decreases wettability, prevent contact with water, aggressive species (Cl^-), inhibiting the cathodic reaction, and subsequently hindering the corrosion process. The quantity, alkyl chain length, penetration depth of the carboxylic acids into the corrosion layer is crucial parameters to be an effective corrosion inhibitor [4,24]. The increased length of the alkyl groups, and the density of the carboxylic acid improves the corrosion inhibition. Recent research in Lob et al. 2024, demonstrated tetramethyl orthosilicate (TMOS) doped with three types of carboxylic acids (heptanoic acid, octanoic acid & decanoic acid) indicated that TMOS doped with carboxylic acids can be an effective corrosion inhibitor with high hydrophobicity [4].

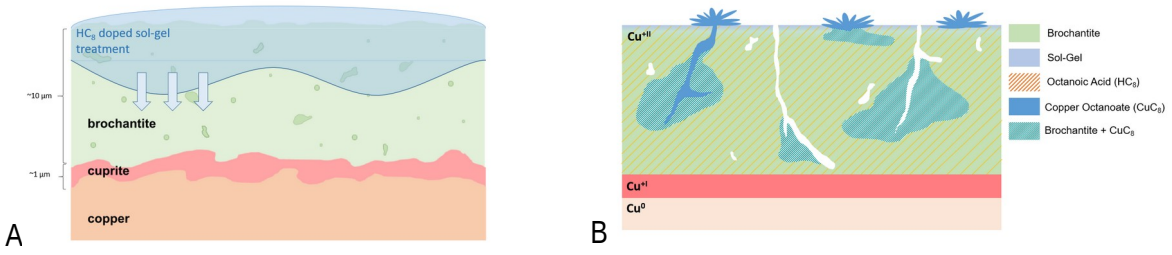


Figure 2: Diagram (A) of the sol-gel treatment penetrating the brochantite layer. Diagram (B) of the hypothetical distribution of corrosion inhibitor treatment in the brochantite layer.

Raman Hyperspectral Imaging

Hyperspectral imaging (HSI) technology was initially developed to be used in astrophysics and geospatial remote sensing applications [27–31]. The technology has since been applied to a wide range of other applications such as pharmaceuticals, biochemistry, food science, and semiconductor manufacturing [27,31]. It is a non-invasive and non-destructive technology making it a perfect analysis technology for cultural heritage science [27]. HSI system captures electromagnetic radiation across hundreds of contiguous spectral bands, compared to traditional RGB (red-blue-green) imaging which discretizes light into three broad, non-contiguous bands [30]. The resulting output is a high-dimensionality, three-dimensional dataset known as a hypercube. The dimensions of a hypercube are represented as (M x N x λ), the M and N represent x,y spatial coordinates, while λ represents the spectral dimension, see Figure 3 [31]. Each pixel within the hypercube contains the high-resolution spectrum reflecting the unique reflectance, transmittance, or absorbance properties of the material at that specific spatial location [28].

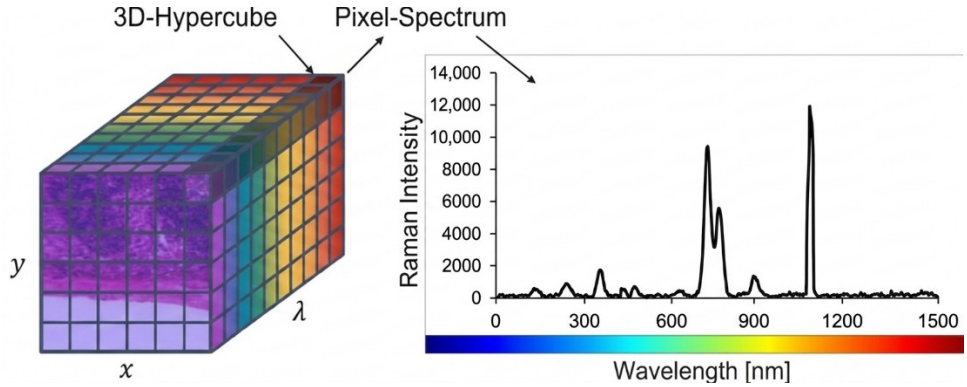


Figure 3:Diagram representation of a hypercube with spatial dimensions (x,y) and the spectral dimension (λ). The hypercube consists of a full spectrum for each spatial pixel. The right panel displays the spectrum of the top-right pixel, illustrating its measured Raman intensity spectrum. [32]

Raman spectroscopy is a non-destructive analytical technique used to identify molecular structures by measuring the vibrational, rotational, and low-frequency modes of chemical bonds. It illuminates a sample with a monochromatic laser source, the technique exploits the inelastic scattering of light, wherein incident photons exchange energy with the molecular bonds of the target material. All the possible vibrational modes of a molecule are a series of independent, fundamental vibrations, named normal vibrations, which correspond to its vibrational frequency (ν) [34–38].

When monochromatic electromagnetic radiation interacts with a molecule, the vast majority of photons undergo elastic Rayleigh scattering with no net energy loss (an absorption-emission phenomenon). However, a smaller amount of photons experiences inelastic scattering by momentarily perturbing the electron cloud to an unstable virtual energy state. This causes a net loss or gain in energy between the molecule and the photons, providing the basis for the Raman effect [38–40].

This inelastic exchange manifests in two distinct forms depending on the starting energy of the molecule. In Stokes scattering, the incident photon transfers energy to the molecule, resulting in scattered light of lower frequency and longer wavelength (lower energy than the incident beam), which produces a positive Raman shift. Conversely, anti-Stokes scattering occurs when a photon absorbs vibrational energy from an already excited molecule, yielding scattered light of higher frequency (higher energy than the incident beam) and a negative Raman shift. The energy difference between the incident photons and the scattered photons corresponds to the energy difference between the vibrational ground state and the first excited vibrational state of the molecule. This energy difference is quantified as the Raman shift, defined as the difference in frequency between the incident laser and the inelastically scattered light, and is typically expressed in units of wavenumbers (cm^{-1}) [38–41].

A unique, characteristic spectral fingerprint is generated by plotting the scattered photon intensity as a function of the Raman shift. Each peak in the spectrum corresponds to a specific normal vibration that induces a change in the molecular polarizability, enabling precise identification of functional groups, chemical bond environments, and material phase changes. Furthermore, when combined with HSI, the spectral dimension (λ) of the hypercube maps the intensity of these inelastic scattering events across a continuum of frequencies. This allows the system to generate high-resolution chemical distribution, molecular orientation, and solid-state phase transitions within complex, heterogeneous samples [32,38,39].

In the last few decades, Raman spectroscopy has become an important analytical technique in archaeometry and cultural heritage preservation, valued for its non-destructive and non-invasive approach, a critical aspect when handling irreplaceable art and artifacts [38,39,42,43]. The technique offers exceptional versatility, enabling researchers to identify a vast array of both inorganic and organic materials, including mineral pigments, ceramics, gemstones, metallic corrosion layers, and complex organic binders. A major catalyst for its widespread adoption has been the development of portable,

miniaturized instrumentation; these handheld devices allow for in-situ analysis directly within museum galleries, outdoor excavation sites, or on architectural monuments, completely circumventing the need for physical sampling or risky transportation. Raman analysis plays an important role in determining artifact provenance, authentication, and ancient manufacturing techniques. Furthermore, it offers diagnostic information regarding the deterioration of materials, guiding and improving contemporary approaches to preservation and restoration [38,43–45].

Multi-block fusion

Multi-block analysis is a data fusion technique used to analyze complex datasets obtained from multiple sources or distinct analytical techniques simultaneously. High-dimensional data generated from the different sources originating from the sample have become increasingly popular, especially in the domain of chemometrics. While traditional single block methods are restricted to processing data from a single source, multi-block analysis are methods to merge multiple data matrices generated from the same samples, such as combining different spectroscopic techniques or combining different chemical measurements. By performing this integrated analysis, it is possible to extract both the shared common and unique distinct information across all datasets. Ultimately, this leads to improved data visualization, enhanced predictive performance in modeling, and a much deeper comprehensive understanding of the process or sample being studied [46,47].

Methodologically, multi-block data fusion strategies are categorized into three distinct levels, namely between measurement level fusion, feature level fusion, and decision level fusion commonly referred to as low-level, mid-level, and high-level data fusion [47–49]. Low-level data fusion is the simplest method, it is concatenating the data matrices into one single matrix [47–51]. It does not require the application of any feature extraction or variable reduction technique [50]. Depending on the commonality of the data blocks, it is concatenated by directly placing the datasets next to the other or coupling them across the variables [48]. Pre-processing is crucial for low-level data fusion to prevent data blocks with different magnitudes from dominating one another [48].

Mid-level data fusion involves first extracting the relevant features, then recombining the extracted features into a new single matrix [47–51]. Instead of merging high-dimensional raw datasets directly, dimensionality reduction or variable selection techniques such as Principal Component Analysis (PCA), Partial Least Squares (PLS), or other feature selection algorithms are applied independently to each data block. The most informative features, such as score vectors or selected key variables, are then concatenated into a streamlined, lower-dimensional fused matrix. This strategy effectively eliminates uninformative noise, mitigates the curse of dimensionality, and balances the relative weight of distinct analytical blocks regardless of their original variable sizes. However, careful optimization is required during the feature extraction stage to ensure that subtle yet critical information is not inadvertently discarded prior to fusion [47,48,50,51].

High-level fusion, also called decision level fusion, integrates the regression or classification outcomes generated by separate, independent models built for each data block [47–51]. It is a technique that is infrequently utilized in chemometrics field compared to low and mid-level data fusion methods [48]. Rather than combining raw variables or extracted features, individual predictive models are first trained on their respective datasets to produce preliminary decisions values. These individual outputs are then fused into a final consolidated decision using combination strategies such as weighted averaging, majority voting, or Bayesian consensus. This approach is particularly advantageous when dealing with highly heterogeneous data sources or distinct analytical techniques that require specialized modeling architectures. Furthermore, high-level fusion offers strong robustness against noise, as a poor-quality or failing data block can be easily down-weighted or excluded without redesigning the overall workflow. However, a primary trade-off of this strategy is that it cannot capture synergistic, cross-block variable interactions that might exist prior to the final decision stage [47–51].

In this study, a specific alkoxysilane formulation was investigated composed of tetraethyl orthosilicate (TEOS) used as the silica precursor doped with octanoic acid as the corrosion inhibitor. The primary objective was to map the spatial distribution and penetration depth of the corrosion inhibition treatment within the copper corrosion layer. Ideally, the TEOS would remain on the surface while the octanoic acid penetrates deeper into the brochantite layer, reacting to form a uniformly distributed copper octanoate phase. To assess this mechanism, the corrosion products and protective treatment were analyzed using Raman spectroscopy with 473 nm and 532 nm laser sources. This dual-wavelength excitation strategy leverages distinct spectral responses to enhance compound identification, yielding a comprehensive, spatially resolved map of coating integration. Multi-block data analysis methods were utilized to fuse both excitation wavelengths datasets, offering a holistic model of the coating system to spatially resolve corrosion inhibition treatment.

2.1 Materials

2.1.1 Copper sample

The copper coupon sample was taken from the roof of Saint-Martin Church in Metz, France with natural corrosion from around 100 years. This same sample was previously described in Lob *et al.* 2024, Apchain *et al.* 2021, and Donnici *et al.* 2021 [2,4,24]. The corrosion layer structure features consist of an approximately 10 µm inner cuprite layer beneath a thicker (40–50 µm), porous outer brochantite layer.

2.1.2 Corrosion Inhibition Formulation

The treatment synthesized in this study was a solution of tetraethyl orthosilicate (TEOS), $Si(OCH_3)_4$), octanoic acid ($C_8H_{12}O_2$) as the corrosion inhibitor with water and isopropanol as solvent. The TEOS/HC8/IprOH/H2O solution was prepared in a 1:4:12:4 molar ratio. Octanoic acid at 99% purity and the TEOS at 98% purity used as silicon precursor were purchased from Sigma Aldrich. The isopropanol at 99.9% purity was purchased from Merck.

2.1.3 Reference Compounds

Pure reference compounds of synthetic brochantite and copper octanoate were prepared for Raman spectroscopy analysis. The synthesis followed the processes described by Kratschmer, Wallinder & Leygraf 2002 and the Robinet method [12,25].

2.2 Software

Both Python and R were utilized for this study. In Python, data manipulation was performed using Pandas and NumPy, while data visualization used Seaborn, Matplotlib, and Plotly. For data analysis, Scikit-learn and MVLearn were used in Python, alongside the R packages of Multiblock and MBAnalysis [52,53].

2.3 Sample Preparation

The sol-gel formulation was prepared in a 50 mL Eppendorf tube. At 50°C and with stirring at 800 rpm, 14 mL of octanoic acid was solubilized in 18 mL of isopropanol. Then, 5 mL of TEOS is added to the mixture before adding 1.6 mL of water. The solution is stirred at 800 rpm at 30°C for 30 minutes after the addition of water.

The copper coupon treatment was carried out by dip-coating. This technique allows control of the immersion and withdrawal speeds of the treated substrates in order to obtain homogeneous and reproducible films for viscous substances. The copper coupon was immersed in the sol-gel solution for one minute. After soaking the sample in the sol-gel formulation, the sample was immediately placed in a desiccator under controlled conditions, 30°C temperature and 60% relative humidity. The sample was kept in these conditions for one week in order to dry.

Tin foil was used to cover the coupon and prevent resin penetration into the brochantite porosity. Then, the sample was embedded in resin and sectioned to expose the cross-section. Then mechanically polished with silicon carbide to reveal the corrosion product layers while minimizing damage to them. The samples are subsequently placed in an ultrasonic water bath for two minutes and dried in an oven at 30°C for 24 hours.

2.4 Raman analysis

Micro-Raman spectroscopy analyses were conducted on the treated copper cross-sections using a Renishaw InVia spectrometer equipped with a Leica DMLM optical microscope. Spectral acquisitions were performed using Renishaw Wire 3.4 software. Two successive mappings were carried out over a 124 x 72 µm area, using a 4 µm step size to collect 1080 spectra per map. To ensure exact spatial alignment between the datasets, the sample's position remained unchanged throughout both runs. While both mappings used a x50 objective lens and an 1800 lines/mm diffraction grating, they differed in their laser wavelength and specific acquisition parameters.

The first mapping utilized a 473 nm (blue) laser with a 10 s acquisition time. To prevent phase transformations, a filter was applied to transmit only 5% of the light, keeping the laser power on the sample below 120 µW. The second mapping was conducted using a 532 nm (green) laser with a 50 s acquisition time. Here, the power was kept below 100 µW using a 1% transmission filter to similarly avoid phase alterations. The reference spectra were acquired under the same conditions. The Raman spectroscopy measurements of the copper coupon used two different laser excitations yielded two three-dimensional spectral hypercubes. The dimensions of hypercube are expressed as H = (M x N x λ) the M and N represent x,y spatial coordinates, while λ represents the spectral dimension [32].

2.5 Data Pre-processing

Data pre-processing is broken up into two different stages. The first stage, consisting of the intra-block pre-processing stage, operates independently on each individual data block prior to fusion and encompasses Level I and Level II operations. Level I focuses on signal-level pre-processing such as denoising, baseline drift correction, and scatter corrections remove physical or instrumental artifacts specific to each measurement modality. Level II addresses intra-block scaling and centering to ensure that variables within the same block contribute equitably regardless of their underlying scale. The second stage consists of inter-block pre-processing (Level III), which operates across the integrated matrices to equalize systemic differences between blocks. Level III employs strategies such as block variance scaling, block rank scaling, or block normalization to adjust for disparities in variable count, block variance, and pseudo-rank, thereby preventing high-dimensional or large-variance blocks from dominating the final data fusion model [46,54].

2.5.1 Intra-block pre-processing

Raw Raman spectra often contain unwanted noise, background signals, and artifacts that should be removed to obtain useful information [26]. One type of unwanted sharp signals are cosmic spikes caused by high-energy cosmic rays from outer space hitting the detector [27]. Software provided by Renishaw enabled the removal of the cosmic spikes.

Then, background correction was performed to extract the true Raman peak intensities required for subsequent data processing. Asymmetric Least Squares (ALS) method was used for the baseline estimation, leveraging the distinct frequency separation between positive Raman spectral features and the underlying background. The ALS method creates a smooth baseline under experimental data. This equation works by balancing two factors; penalizing sharp changes via the second derivative to keep the curve smooth, and using asymmetric weights to properly adjust for data deviations. Following background correction, Total peak area normalization was utilized due to its ability to correct global intensity variations while preserving the relative intensities across all spectral peaks [26].

Total peak area normalization calculates the percentage of each peak area relative to the total area of all peaks [28].

2.5.2 Inter-block pre-processing

Pre-treatment of inter-block pre-processing of the data is aimed at equalizing differences across different data blocks, such as variations in scales, pseudo-ranks, and the number of variables. The x,y dimensions were equalized across both hypercube datasets. This step is critical because many multi-block analysis methods naturally favor blocks with larger overall variations, which can lead to model bias. In this study, six scaling techniques described in Campos & Reis 2020 (See Table 1) were analyzed to prevent one block from dominating the model, ultimately improving model interpretation. The different scaling methods produced a scaling factor K_b, that was divided by the blocks during pre-processing to produce the scaled blocks. [46,54]. The first category of scaling methods is simply known as block-scaling methods that consist of soft-block scaling (SBS), hard-block scaling (HBS) and super hard-block scaling (SHBS). These approaches uses the number of variables to calculate the scaling factor. The second category of scaling uses the block standard deviation to calculate the scaling factor [54,55].

Scaling Method	**Scaling Factor, K_b**
Soft-Block Scaling (SBS)	$\frac{1}{\sqrt[4]{n_{\text{block } j}}}$
Hard-Block Scaling (HBS)	$\frac{1}{\sqrt{n_{\text{block } j}}}$
Super Hard-Block Scaling (SHBS)	$\frac{1}{n_{\text{block } i}}$
Soft-Block Variance Scaling (SBVS)	$\frac{\sqrt[4]{n_{\text{block } j}}}{\sqrt{\sum \sigma_i^2}}$
Hard-BlockVariance Scaling (HBVS)	$\frac{1}{\sqrt{\sum \sigma_i^2}}$
Super Hard-Block Varaince Scaling (SHBVS)	$\frac{1}{\sum \sigma_i}$

Table 1: Inter-block scaling techniques used in this study described in Campos & Reis 2020 [54]

n_{block} : number of variables in the block *j* (j = 1… n).

σ_i : standard deviation of each block column (variable) (I = 1… n).

2.6 Data Processing: Individual blocks

2.6.1 Individual Blocks: Maximum Spectral Intensity Heatmap

Firstly, the individual hypercubes (473 nm and 532 nm) were condensed into two-dimensional spatial grids. This dimensionality reduction was achieved by extracting the maximum intensity value at each spatial coordinate along the spectral dimension (λ). This step collapsed the spectral, z-direction

dimension from hundreds of features down to a single feature (the maximum intensity value) and this was performed separately before any unsupervised chemometric analysis [32,56].

2.6.2 Individual Blocks: Principal Component Analysis

Secondly, Principal Component Analysis (PCA) was then conducted on the individual hypercube, it is an unsupervised linear transformation technique used for dimensionality reduction. It projects high-dimensional data onto a lower-dimensional subspace of orthogonal axes, known as principal components, to maximize variance and minimize information loss [57–59].

Since PCA is conventionally applied to a two-dimensional matrix. Therefore, the three-dimensional 473 nm and 532 nm hyperspectral hypercubes, $H \in R^{M \times N \times \lambda}$, were first flattened into two-dimensional matrices. Specifically, the spatial coordinates K = M x N were reshaped into a single array, while the spectral dimension (λ) was preserved. This unfolding process yielded a two-dimensional data matrix, $X \in R^{K \times \lambda}$ [58,60,61].

Next, the flattened matrix was decomposed via Singular Value Decomposition (SVD), that factorizes the data as $X = U\Sigma V^T$. Where $U \in R^{K \times K}$ represents the left-singular vectors, describing the spatial coordinates in the orthogonalized space, $\Sigma \in R^{K \times \lambda}$ contains the singular values, and $V \in R^{\lambda \times \lambda}$ contains the orthonormal right-singular vectors representing the spectral loadings. Dimensionality reduction to *k* principal components (PCs) is achieved by extracting the first *k* columns of **V** to define the spectral loading matrix, $P \in R^{\lambda \times k}$, ensuring that the primary axes capture the maximum variance expressed across the datasets. The spectral data was then projected onto this *k*-dimensional subspace, yielding the score matrix, $T \in R^{K \times k}$ computed as $T = XP$. Finally, spatial context was restored by reshaping the *K* rows of the score matrix back into the original *M* x *N* spatial grid. This reconstruction process yields a dimensionally reduced hypercube of size M x N x *k*, reducing the number of features in the z-direction down to *k* [46–48].

2.7 Data Processing: Multi-block Fusion Methods

Exploratory multi-block data analysis methods can be classified into two categories based on how methods handle common and distinct information in the blocks. The first category comprises methods that identify only the shared information among the different blocks, and then explore the contribution of each block to these shared components. The second category comprises methods that simultaneously extract both shared information and information specific to individual blocks [62].

Shared Components Analysis Only

2.7.1 Concatenated PCA (SUM PCA)

The most straightforward method for multi-block data fusion is concatenated PCA, commonly referred to as SUM-PCA. In this method, individual flattened data blocks $X_1 \in R^{K \times \lambda_1}$ (473 nm block) and $X_2 \in R^{K \times \lambda_2}$ (532 nm block) are concatenated horizontally along their shared spatial dimension *K* to construct a unified fused matrix $X_{\text{fused}} = [X_1, X_2] \in R^{K \times (\lambda_1 + \lambda_2)}$, as shown in Figure 4. SVD is subsequently applied to X_{fused}, following the procedure described in Section 2.6 for individual block processing. This decomposition yields a global score matrix $T_{fused} \in R^{K \times k}$, which provides a spatial map representing shared chemical variations across both excitation wavelengths, and a global loading matrix, $P_{\text{fused}} \in R^{(\lambda_1 + \lambda_2) \times k}$. Finally, the global loading matrix is partitioned into block-specific sub-matrices $P_1 \in R^{\lambda_1 \times k}$ and $P_2 \in R^{\lambda_2 \times k}$ to evaluate spectral contributions and variable loadings from each respective excitation wavelengths [31,37–40,49,50].

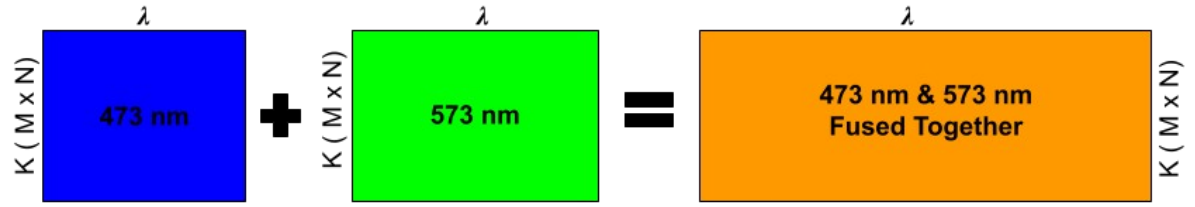


Figure 4: Diagram illustrating SUM-PCA. The two data matrices (473 nm & 573 nm) are concatenated together to form a new concatenated matrix. The matrices are fused together in the spatial dimension, elongating the spectral dimension.

2.7.2 Common Components and Specific Weight Analysis (ComDim / CCSWA)

ComDim, also known as Common Components and Specific Weight Analysis (CCSWA), is a multi-block unsupervised method designed to identify shared features across multiple datasets. It operates by extracting global common components from the combined data space while simultaneously calculating block-specific weights (saliences) for each individual data block. These weights quantify the relative contribution or importance of each data source to each extracted dimension, allowing for a clear evaluation of shared versus unique variance. In this study, ComDim is applied to identify common spatial trends (common scores) shared across the 473 nm and 532 nm spectral blocks, while simultaneously determining the specific weight contributed by each spectral block and the individual wavelength contributions via block loadings [51–53].

Mathematically, for each data block matrix X_i an inter-pixel cross-product matrix is defined as $W_i = X_i X_i^T \in R^{K \times K}$, capturing the inherent variance and similarity structure within that block. To extract a global common component, an iterative optimization procedure is performed. The block weights w_i are initially set to equal values, and a weighted global matrix W_G is constructed, $W_G = \sum_{i=1}^{B} w_i W_i$. SVD is performed on W_G to extract its dominant eigenvector, establishing the global common score vector, $q \in R^{K \times 1}$. The specific block weights are then updated by projecting **q** back onto each block's cross-product matrix, $\left(w_i = q^T W_i q\right)$ and normalizing them such that $\sum_{i=1}^{B} w_i^2 = 1$. The updated weights are substituted back to reconstruct W_G, and this cycle repeats iteratively until the scores and weights reach convergence. Upon convergence, the block loadings $\mathbf{p_i}$ each block are calculated as, $p_i = \left(X_i^T q\right) / \left(q^T q\right)$. Once convergence is reached for a component, the global scores, block loadings, and block weights are stored. To calculate subsequent dimensions, each data block matrix X_i is deflated by subtracting the variance explained by the current common component. The entire iterative optimization process is then repeated on the deflated matrices until the designated number of common components has been extracted [51–53].

2.7.3 Hierarchical PCA (H-PCA)

Hierarchical Principal Component Analysis (H-PCA) is an exploratory multi-block technique designed to reveal covariant patterns both within and across multiple multivariate data matrices. Similar to ComDim/CCSWA, H-PCA models the 473 nm and 532 nm spectral blocks by capturing global trend information while preserving block-specific contributions. However, unlike standard concatenated PCA or cross-product based methods, H-PCA constructs a hierarchical score-loading architecture using an iterative procedure derived from the Nonlinear Iterative Partial Least Squares (NIPALS) framework.

In the first stage, model parameters for a single component including block weights, block scores, global weights, and global scores are computed iteratively. The algorithm initializes with an

arbitrary normalized global score vector $t_{\text{globe}} \in R^{K\times 1}$. For each data block, individual block weight vectors $w_i \in R^{\lambda_i \times 1}$ are calculated by regressing the block features onto **t**, followed by normalization to unit length. Individual block score vectors are subsequently derived as $t_i = X_i w_i$. These block scores are concatenated horizontally to form a super-score matrix, $T_{\text{sup}} = \left[t_1 t_2 \ldots t_B\right] \in R^{(K\times B)}$. A super-weight vector $q \in R^{B\times 1}$ is calculated to determine the relative importance of each block, and the global score vector is updated as $t = T_{\text{sup}} q$. This cycle is iterated until numerical convergence of **t** is achieved. Upon convergence of a global component, block loading vectors $p_i \in R^{\lambda_i \times 1}$ are calculated by regressing each block X_i onto the converged global score vector **t** via $p_i = X_i^T t / \left(t^T t\right)$. In the second stage, higher-order latent components are extracted through block-wise deflation. The variance explained by the current global component is removed from each data block using the rank update. The iterative NIPALS algorithm is then repeated on the resulting residual matrices to extract subsequent hierarchical principal components, providing a structured decomposition of common and block-specific variance across the excitation wavelengths [54–56].

2.7.4 Multi-Block PCA (MB-PCA) / Consensus PCA (C-PCA)

Consensus PCA (C-PCA), often designated interchangeably with Multi-Block PCA (MB-PCA), captures global spatial consensus while accounting for differences in total variance across blocks. Operating under a NIPALS framework similar to H-PCA, C-PCA applies explicit block-scaling constraints to ensure equal block representation. Given an initial global consensus score vector $\boldsymbol{t}_{\textbf{globe}}$, normalized block loadings, $p_i = X_i^T t_{\text{glob}} / \left\| X_i^T t_{\text{glob}} \right\|$, and block scores, $t_i = X_i p_i$, are calculated for each block. To prevent high-variance blocks from dominating, block scores are normalized prior to constructing the super-score matrix, $T_{\text{sup}} = \left[t_1^c t_2^c \ldots t_B^c\right]$. The consensus weight vector $\mathbf{w_g}$ is calculated, and the global score vector is updated as $t_{\text{glob}} = T_{\text{sup}} w_g$ until convergence. Block deflation is performed to sequentially extract orthogonal consensus dimensions [65,66].

2.7.5 Multiple Co-Inertia Analysis (MCOA)

Multiple Co-Inertia Analysis (MCOA) provides an exploratory multiblock integration framework that maximizes co-inertia (covariance) across diverse data tables measured on identical samples. The MCOA algorithm proceeds through two stages. First, single-table ordination via independent standard PCA is performed on each preprocessed block $\mathbf{X_i}$. Second, MCOA constructs a common synthetic reference vector $q \in R^{K\times 1}$ by maximizing the sum of squared co-inertias between **q** and the individual block score projections $t_i = X_i p_i$. Unlike MB-PCA and H-PCA, MCOA performs deflation on individual block loadings rather than on the global score space, preserving complementary and orthogonal block information across higher dimensions [65,67].

2.7.6 Multi-Factor Analysis (MFA)

Multiple Factor Analysis (MFA) is tailored to integrate multiple data blocks measured on identical observations by balancing the influence of each block. In the first stage of MFA, individual standard PCAs are performed independently on each centered data block $X_i \in R^{K\times \lambda_i}$ to determine its first singular value $\sigma_{i,1}$, which normalizes the maximum directional variance of each block to 1 and prevents larger spectral blocks from dominating the model. In the second stage, all normalized blocks are concatenated horizontally into a grand matrix, $X_{\text{MFA}} = \left[X_1^* X_2^* \ldots X_B^*\right]$ and subjected to a global unweighted PCA. MFA extracts global compromise factor scores $\mathbf{T_{MFA}}$ and partial factor scores,

$T_i = B \cdot X_i^* P_{i,\mathrm{MFA}}$ by projecting individual normalized blocks onto the compromise space, enabling direct evaluation of commonalities and discrepancies [64,66,68].

Shared and Distinct Component Analysis

2.7.7 Distinct and Common Simultaneous Component Analysis (DISCO-SCA)

Distinct and Common Simultaneous Component Analysis (DISCO-SCA) explicitly separates shared variations across all blocks from distinct variations specific to individual blocks. In the first step, Simultaneous Component Analysis (SCA) is performed on the concatenated data matrix $X_{\mathrm{fused}} = \left[X_1 X_2 \ldots X_B\right] = T P_{\mathrm{fused}}^T + E$, yielding unrotated global score (**T**) and loading ($\mathbf{P}_{\mathrm{fused}}$) matrices. Because standard SCA components contain mixed common and unique signals, the second step applies an orthogonal target rotation matrix $R \in R^{k \times k}$. The rotated loadings $P_{\mathrm{rot}} = P_{\mathrm{fused}} R$ are rotated toward a predefined target pattern $\mathbf{P}_{\mathrm{target}}$ containing fixed zero constraints for distinct blocks and non-zero entries for common blocks. The corresponding scores are updated as $T_{\mathrm{rot}} = TR$, providing an interpretable structural separation of joint and block-unique chemical signals [47,69,70].

2.7.8 Angle-based Joint Individual and Variance Explained

Angle-based Joint and Individual Variation Explained (AJIVE) is an analysis technique designed to separate shared (joint) variations across multiple datasets from the unique (individual) variations inherent to each specific data block. Unlike standard PCA which evaluates data blocks in isolation, AJIVE explicitly decomposes each input data X_k into three additive components. A joint signal matrix J_k, a block individual signal matrix I_k, and a residual noise matrix E_k. For the dual-excitation Raman imaging analysis involving flattened matrices from the 473 nm X_1 and 532 nm X_2 laser excitations, this decomposition is expressed as:

$$X_1 = J_1 + I_1 + E_1 \qquad \text{and} \qquad X_2 = J_2 + I_2 + E_2$$

Where $X_k \in R^{K \times \lambda}$ represents the dataset for block *k*, consisting of K common spatial pixels and λ_k spectral variables. In this framework, $J_k \in R^{K \times \lambda}$, captures common spatial-spectral patterns shared across both excitation wavelengths, $I_k \in R^{K \times \lambda}$ captures unique spectral behavior specific to a single wavelength, and $E_k \in R^{K \times \lambda}$ accounts unmodeled residual noise.

To achieve this decomposition efficiently, AJIVE employs a two-stage spectral approach based on the geometric principal angles between the spatial (row score) subspaces of the data matrices. First, SVD is applied independently to each column-centered and scaled block. Based on initial rank selection criteria, low-rank approximations with ranks r_1 and r_2 are extracted to yield the dominant spatial score matrices $U_1 \in R^{K \times r_1}$ and $U_2 \in R^{K \times r_2}$. To evaluate spatial alignment, these score matrices are concatenated horizontally to form a combined matrix, $M = \left[U_1 U_2\right] \in R^{K \times (r_1 + r_2)}$. A second SVD is then performed on M^T**,** yielding singular values d_i that directly relate to the principal angles θ_i = across (d_i) between the row spaces. Principal angles close to 0° indicate shared spatial patterns (joint signals), whereas angles approaching 90° represent orthogonal patterns unique to an individual block, enabling a robust statistical separation of joint and individual variation [71–73].

2.8 Summary of Methods & Procedures

Overall, the procedure of the methods used in this study can be summarized in Figure 5, describing the process that the analyses were conducted. In summary, the copper coupons were prepared, and analyzed using Renishaw InVia Raman spectrometer. Then, the raw Raman data underwent intra-block pre-processing. Either the two data blocks underwent inter-block data pre-processing for multi-

block exploratory techniques or PCA was applied to the individual blocks separately as the final step. After inter-block data pre-processing was applied the different multi-blocks methods (ex: SUM-PCA, ComDim, AJIVE etc.) were applied to the Raman datasets as the final step in this procedure.

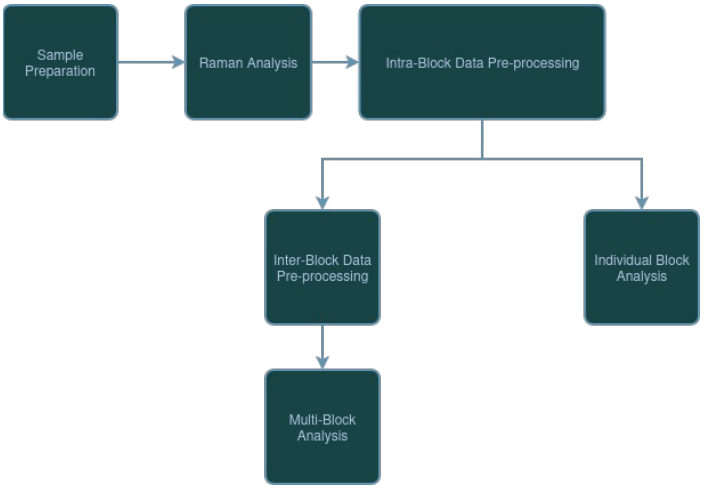


Figure 5: Flowchart describing the process which the proceeding methods were conducted in this study.

3. Results

3.1 Results of the Individual Blocks

3.1.1 Reference Spectrum

The reference spectra of brochantite, copper octanoate, cuprite, octanoic acid and TEOS from the 473 nm laser are shown in Figure 6. Notably, brochantite spectrum has the intense, sharp peak at 973 cm^{-1} that corresponds to previously studied Raman spectra of brochantite in Bouchard & Smith 2003, Fan et al. 2020 and Valadas et al. 2015 [26–29]. This peak at 973 cm^{-1} is indicative of the symmetric stretching vibration of the sulfate present in brochantite [33]. The brochantite spectrum also has a dense cluster of smaller peaks at 194, 242, 318, 391, 481, 596, 608, 1076, 1100, and 1128 cm^{-1}. The peaks at 1076, 1100, and 1128 cm^{-1} are characterized by the anti-symmetric stretching mode of the sulfate in brochantite [29]. The cluster of smaller peaks from ~100cm^{-1} to ~600cm^{-1} is the distinct fingerprint region associate with Raman spectroscopy of brochantite, but are difficult to assign to specific structural features [34,35]. The most intense peak for copper octanoate is a broad peak intensity at 2884 cm^{-1} this corresponds to the stretching vibration mode of -CH_2 or -CH_3 groups in the alkyl chain of the copper octanoate [36]. There is also minor, narrow peak at 1448 cm^{-1} corresponding to the scissoring and twisting mode of -CH_2 or -CH_3 groups in the alkyl chain of the copper octanoate [37].

The cuprite spectrum has peaks at 212, 512, 642 cm^{-1} corresponding to cuprite previously described literature values of 218, 495, and 638 cm^{-1} [32,38,39]. Previous Raman studies of cuprite reported the 212 cm^{-1} to be the most intense, but in our study the 642 cm^{-1} was the most intense peak, and second more intense peak being at 212 cm^{-1}. This could be caused by the differences in the experimental conditions during Raman analysis [32,38,39].

The reference spectrum of the octanoic acid shows a broad, intense peak at 2932 cm^{-1} this corresponds to the stretching vibration mode of -CH_2 or -CH_3 groups in the alkyl chain similar to the copper octanoate. There are minor peaks at 1302 and 1443 cm^{-1}, these peaks correspond to scissoring and twisting mode of -CH_2 or -CH_3 groups in the alkyl chain [37]. The TEOS spectrum was noisy, and the distinct peak from this spectrum corresponds to 2924 cm^{-1}. This peak can be assigned to the stretching vibrations of the -CH bonds present, and Gnyba et al. 2003 showed the presence of this peak in their study [37,40,41].

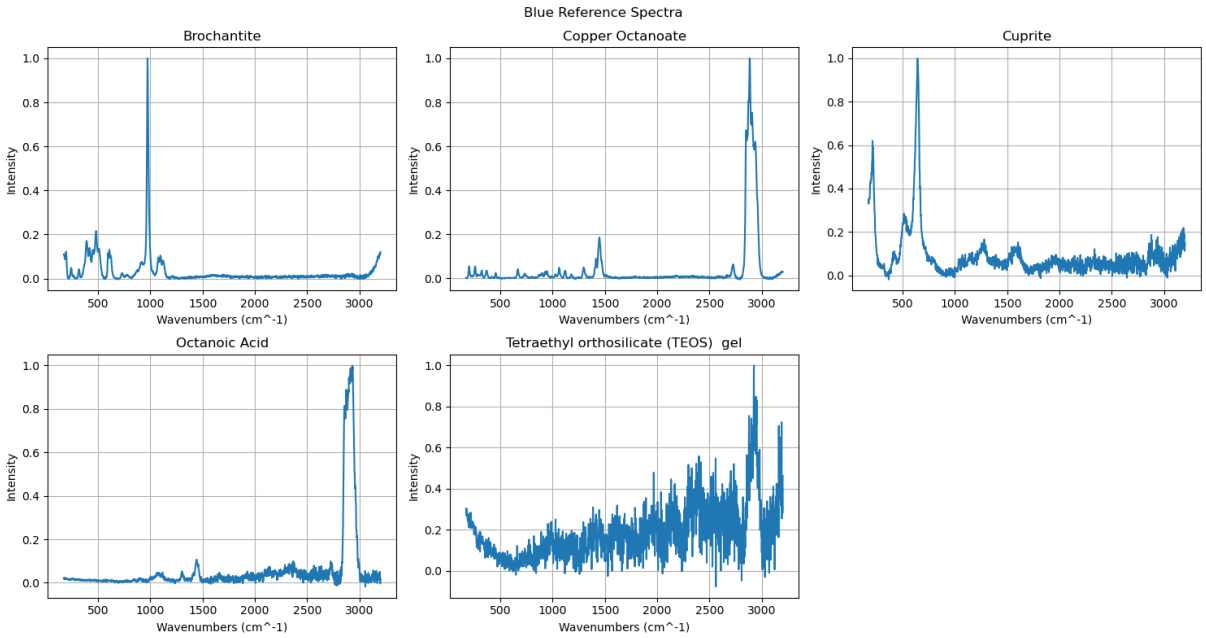


Figure 6: The reference spectra for brochantite, copper octanoate, cuprite, octanoic acid and TEOS from the 473 nm ("blue") laser.

The reference spectra of brochantite, copper octanoate, cuprite, and TEOS from the 532 nm laser are shown in Figure 7. The brochantite spectrum for the green laser is very similar for the blue laser; there is a sharp, intense peak at 974 cm^{-1} with minor peaks in the brochantite fingerprint region at 140, 200, 392, 486, 597, and 611 cm^{-1}. The copper octanoate spectrum has broad peak intensity at 2884 cm^{-1} the same the blue reference spectrum. Copper octanoate has minor, moderate peaks at 111, 202, 260, 324, 374, 458, 670, 1065, 1117, 1299, 1450 cm^{-1}; these minor peaks from ~0 to ~1500 cm^{-1} are caused by the alkyl hydrocarbon chain present on the copper octanoate [37]. The cuprite green reference spectrum is noisy. Cuprite has a distinct peak at 111, 152, 219, 558 and 656 cm^{-1} corresponding to previously described literatures values [32,38]. It is important to note the unique peak compositional layout for cuprite, it has a strong signal at the beginning of the spectrum at ~100, that is followed by peaks at 558 and 656 cm^{-1}, this peak composition is similar to the 473 nm reference spectrum.

The reference spectrum for the octanoic acid shows a broad, intense peak at 2932 cm^{-1} similar to the 473 nm reference spectrum. There are other minor peaks present at 558, 1084, 1117, 1307, 1441, 2454 cm^{-1} from the other vibrational stretching modes on the alkyl chain [37]. The reference spectrum TEOS was noisy; there are distinct peaks at 484, 983, 2933 cm^{-1}. The peaks 484 and 983 cm^{-1} correspond to the stretching of the Si-O bonds, and the peak at 2933 cm^{-1} corresponds to the -CH bonds present [37,40,41]. Based off the data from the two-reference spectrum, it appears the 532 nm laser produces stronger responses from the alkyl chain than the 473 nm laser based off the data from reference spectrum.

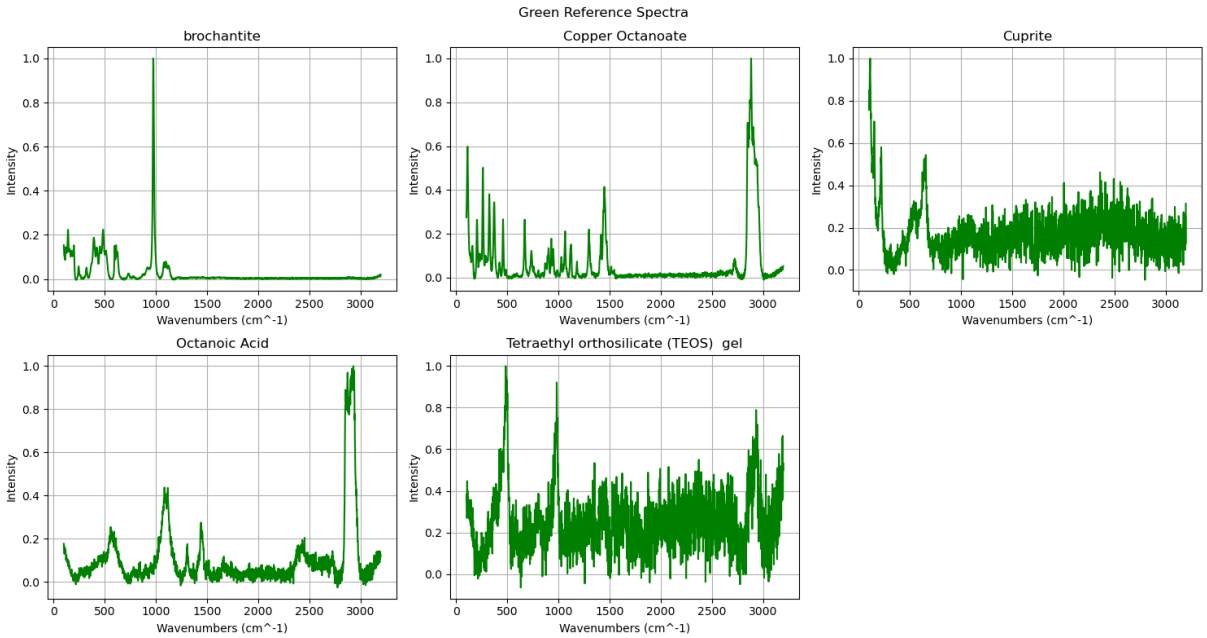


Figure 7: The reference spectra for brochantite, copper octanoate, cuprite, octanoic acid, and TEOS from the 532 nm ("green") laser.

3.1.2 Hypercube structure individual blocks

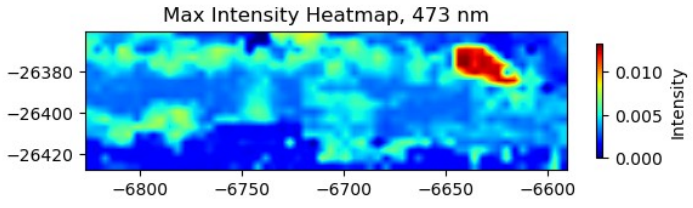
The dimensions for the 473 nm block are 1052 x 1052 x 1433, and the dimensions for the 532 nm block are 1052 x 1052 x 1944.

3.1.3 Results of the maximum intensity of the individual blocks

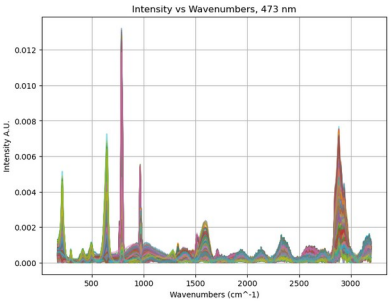
The maximum spectra intensity maps of the individual blocks were analyzed. Figure 8 shows the maximum-intensity hyperspectral heatmaps and the corresponding Raman spectra intensity graphs for the individual blocks. The maximum-intensity hyperspectral heatmaps were constructed by taking the maximum intensity present in the λ dimension corresponding to the spatial coordinates, (M x N).

Some structural information of the corrosion layer can be determined in the 473 nm heatmap, while very little structural information can be determined in the 532 nm heatmap as shown in Figure 8A and Figure 8D. The 473 nm heatmap in Figure 8A shows minor intensity difference throughout the heatmap. But it is difficult to determine the precise compounds from this heatmap due to the overlapping, superimposed spectra in the corresponding Raman spectra plot. The 532 nm heatmap shown in Figure 8C is mostly homogenous for intensity, showing only a little intense area in the corrosion layer.

A



B



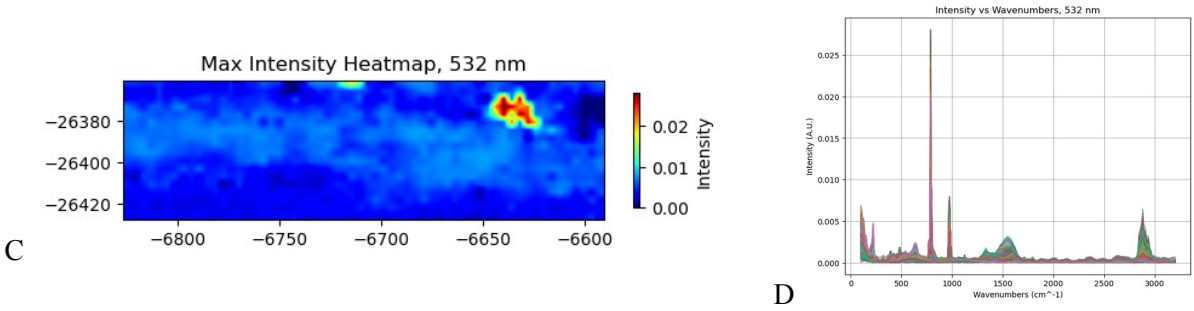


Figure 8: The maximum-intensity heatmaps and the corresponding Raman spectra intensity graphs for the two laser wavelengths. The top-left (A) image is the heatmap of the 473 nm laser and the top-right (B) image is Raman spectra intensity graph for the 473 nm laser. The bottom-left (C) image is the hyperspectral heatmap graph for the 532 nm laser and the bottom-right (D) image is Raman spectra intensity graph for the 532 nm laser. Figure 8B and Figure 8D are the spectra for each pixel superimposed upon one another. The intensity scale is the relative intensity for each heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

In both hyperspectral heatmaps there is an area of high intensity as highlighted in Figure 9. This area is believed to be contamination of silicon carbide from the sample preparation polishing process, not a component of the corrosion layers or the coating. This contamination is present in all subsequent analyzes conducted in this study.

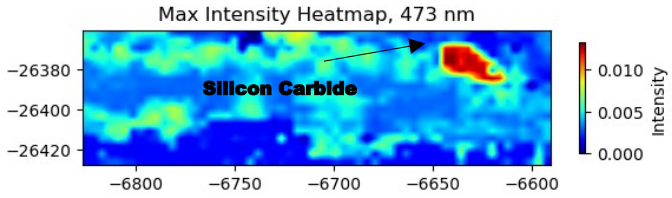


Figure 9: The contamination present in the sample, highlighted on the hyperspectral heatmap from the 473 nm block heatmap. The contamination is from the polishing procedure during sample preparation, and not a component of the corrosion layers or the coating.

3.1.4 Results of PCA applied to individual blocks

SVD was conducted to evaluate the optimal number of PCs for data processing, the corresponding scree plots for both blocks are shown in Figure 10. PCA was subsequently applied to each individual block to identify orthogonal directions of maximum variance within the data. Crucially, since principal components are linear combinations of the original variables rather than chemical components, PCA does not perform spectral unmixing. Consequently, a single PC loading can contain Raman bands from multiple phases, particularly when the variances of those phases are correlated. For the 473 nm dataset, the first six PCs captured 97.61% of the explained variance, whereas for the 532 nm laser block, six PCs captured 95.83%. The 473 nm Raman laser dataset retained a higher percentage of explained variance within the first six PCs compared to the 532 nm dataset. Notably, for both datasets, the second PC captures the variance associated with silicon carbide (SiC) contamination introduced during the sample preparation process, as highlighted in Figure 9.

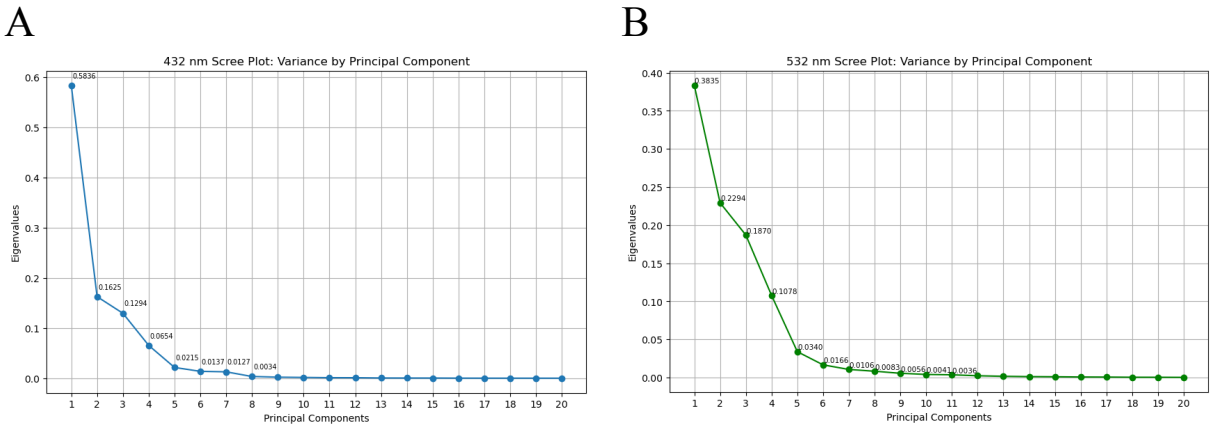


Figure 10: Scree plots for both individual blocks. On the left (A) is the scree corresponding to the blue data set, and on the right (B) is the scree plot for the green data set. The y-axis is the eigenvalues ("the explained variance") and the x-axis is the principal components amount. The first eight principal components are labeled with their corresponding values. The first 20 PC were calculated to highlight the convergence in the scree plots.

Results of PCA applied to the 473 nm block

PC1 for this block is shown in Figure 11 with the heatmap and the corresponding loadings spectrum graph. This PC represents for 58.36% of the explained variance in the data set. A broad, intense peak is observed at 2885 cm^{-1}, which corresponds to the octanoic acid that is present in the sample as compared to the reference spectra in section, 3.1.1 Reference Spectrum. The areas in the with highest intensity expressed in the heatmap for Figure 11A would correspond to this peak from spectrum loadings graph for these chemical constituents. Also, there is a minor, sharp peak of brochantite at 974 cm^{-1}. The areas of intermediate intensity displayed in the Figure 11A heatmap correspond to the brochantite peak identified in the spectral loadings graph.

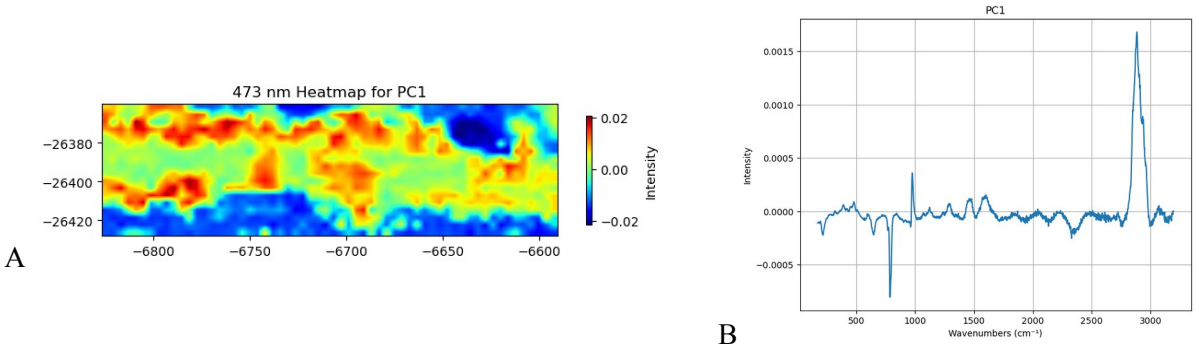


Figure 11: PC1 for the 473 nm "blue" block with the heatmap on the left and the loading spectrum (B) on the right. The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC2 is illustrated with its heatmap and the corresponding loadings spectrum graph. This PC represents 16.25% of the explained variance. This PC represents the silicon carbide (SiC) as shown in Figure 9 from the mechanical polishing during sample preparation. There are two distinct peaks at 789 cm^{-1} and 961 cm^{-1}, these peaks correspond to SiC values described in Nakashima et al. 2003, and Nakashima and Harima 1997 [43–45].

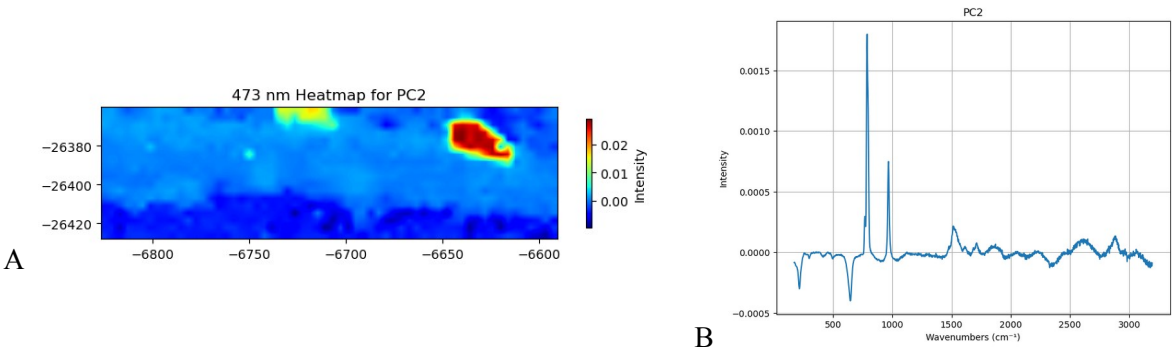


Figure 12: PC2 for the 473 nm "blue" block with the heatmap (A) and the loading spectrum graph (B). This PC display the containment of silicon carbide present in the sample, in the area of greatest intensity on the heatmap. The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC3 is shown in Figure 13 with the heatmap and the corresponding loadings intensity graph. This PC represents the 12.94% of the explained variance in this data set. Sharp peak spikes appear at 218, 645, 786 cm^{-1}. These features represent cuprite in the sample. However, while the peak locations are similar. Its distribution on the heatmap aligns with the expected location of cuprite. As shown in Appendix Figure 30, the bottom portion of the image represents the copper substrate, proceeding upward through cuprite, brochantite, and finally the mounting resin.

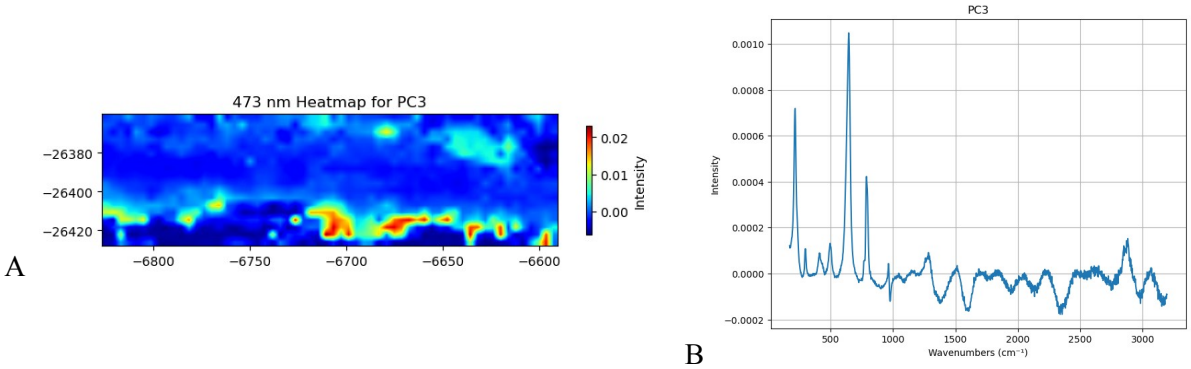


Figure 13: PC3 for the 473 nm "blue" block with the heatmap (A) and the loading spectrum graph (B). The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC4 is shown in Figure 14 and represents 6.54% of the explained variance. This PC is characterized by a sharp peak at 985 cm^{-1} that corresponds to the profile of brochantite. On the heatmap, the areas of highest intensity correspond directly to the distribution of brochantite. Additionally, there are also peak intensities at 1079, 1099, 1129 cm^{-1}, that forms a triplet structure that matches the reference brochantite spectrum immediately following the sharp 985 cm^{-1} peak [29].

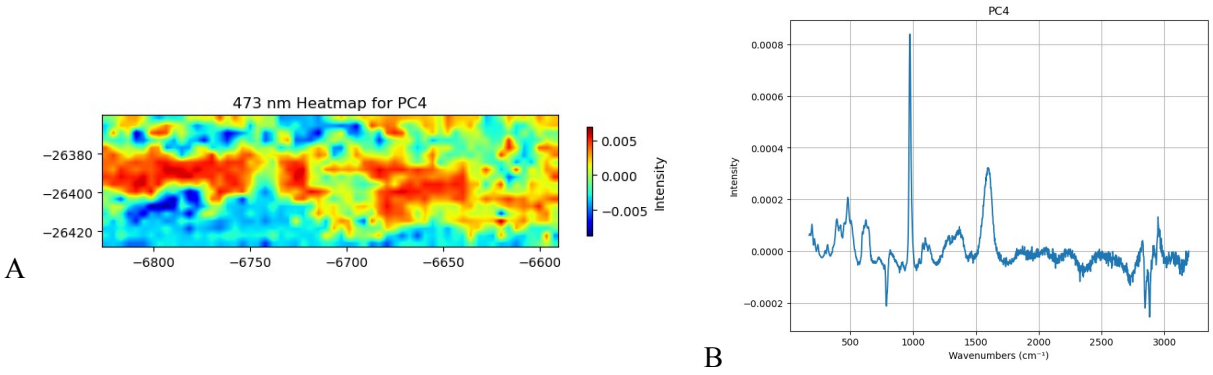


Figure 14: PC4 for the 473 nm "blue" block with the hyperspectral heatmap (A) and the loading spectrum graph (B). The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC5 is shown in Figure 15 with the heatmap and the corresponding loadings spectrum graph. This PC represents the 2.15% of the explained variance in this data set. The loadings spectrum graph for this PC is noisy and difficult to extract peaks as the baseline breakdown. Some structural components can be identified in the areas of increased intensity, but it is difficult to identify these areas due to the noise of the intensity graph. Subsequent PCs were not evaluated due to diminishing results.

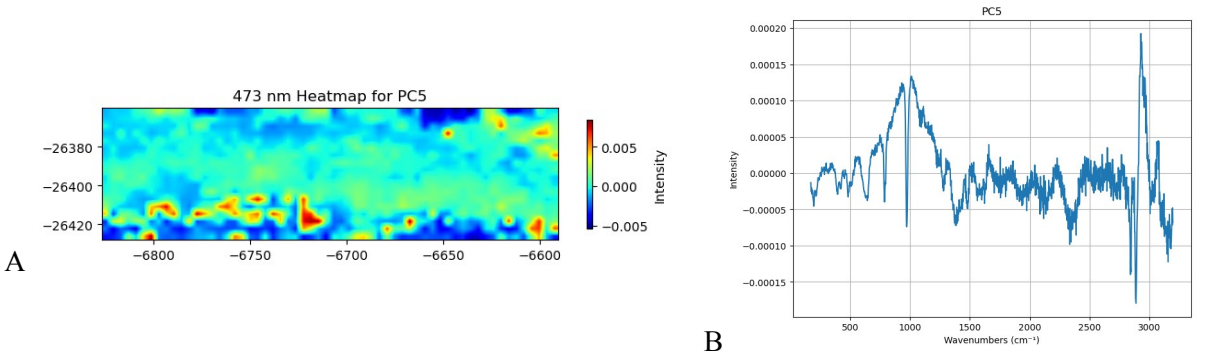


Figure 15: PC5 for the 473 nm "blue" block with the heatmap (A) and the loading spectrum graph (B). The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

Results of PCA applied to the 532 nm block

PC1 for this shown in Figure 16 with the heatmap and the corresponding loading spectrum graph. The loadings spectrum shows a strong, sharp peak intensity at 973 cm^{-1} that is indicative of brochantite, alongside a moderate, broad peak at 2883 cm^{-1} indicative of copper octanoate. The heatmap shown in Figure 16A shows a thick layer of brochantite, along the octanoate acid/copper octanoate present in the sample this corresponds visually to the micrograph in appendix Figure 30.

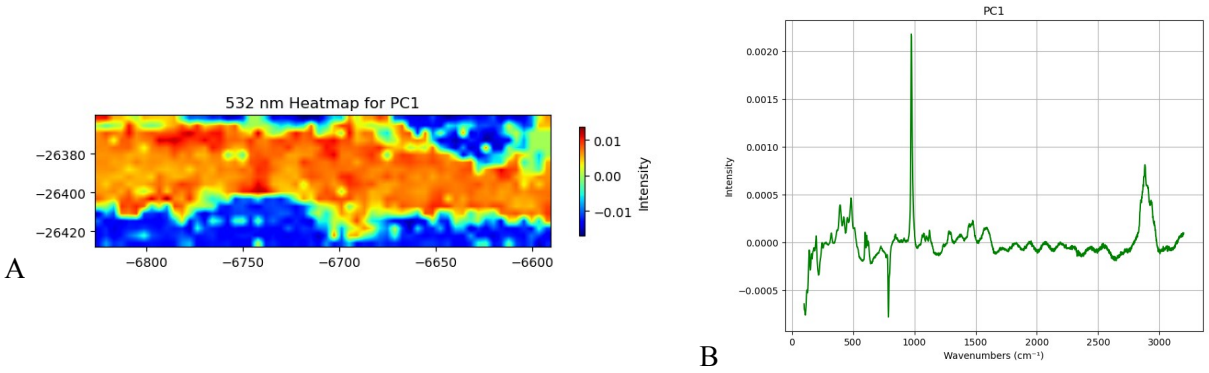


Figure 16: PC1 for the 532 nm block with the hyperspectral heatmap on the left and the loadings intensity graph on the right.The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC2 is illustrated in Figure 17, this PC accounts for 22.94% of the explained variance. It represents the SiC contamination from mechanical polishing. Sharp, intense peaks at 791 and 970 cm^{-1} correspond to the previously described SiC features seen in Figure 9 and Figure 12.

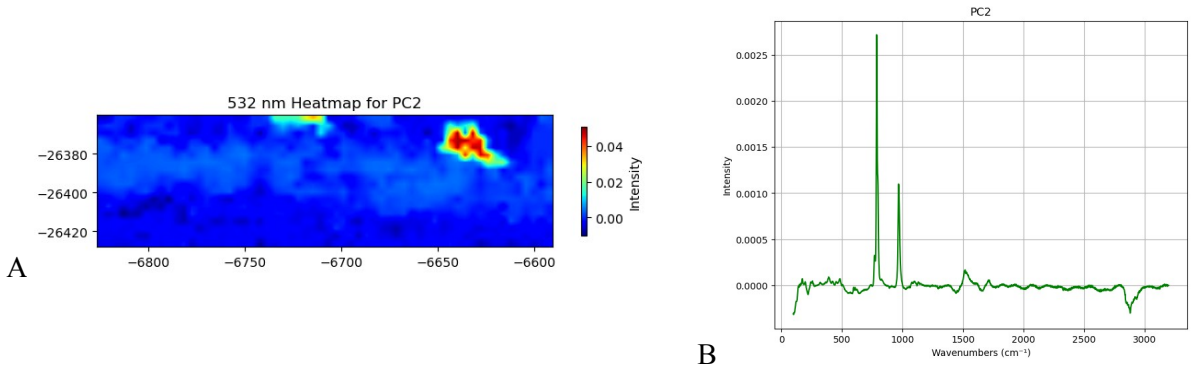


Figure 17: PC2 for the 532 nm "green" block with the heatmap (A) and the loading spectrum graph (B). This PC display the containment of silicon carbide present in the sample, in the area of greatest intensity on the heatmap. The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC3 is shown in Figure 18 with the heatmap and the corresponding loadings intensity graph, this PC represents 18.70% of the explained variance. There is a strong, sharp peak at 973 cm^{-1} that matches reference data for brochantite. The spatial distribution in the heatmap aligns with expected location of brochantite, closely aligned with the results of PC1 in Figure 16, and matching the micrograph in appendix Figure 30.

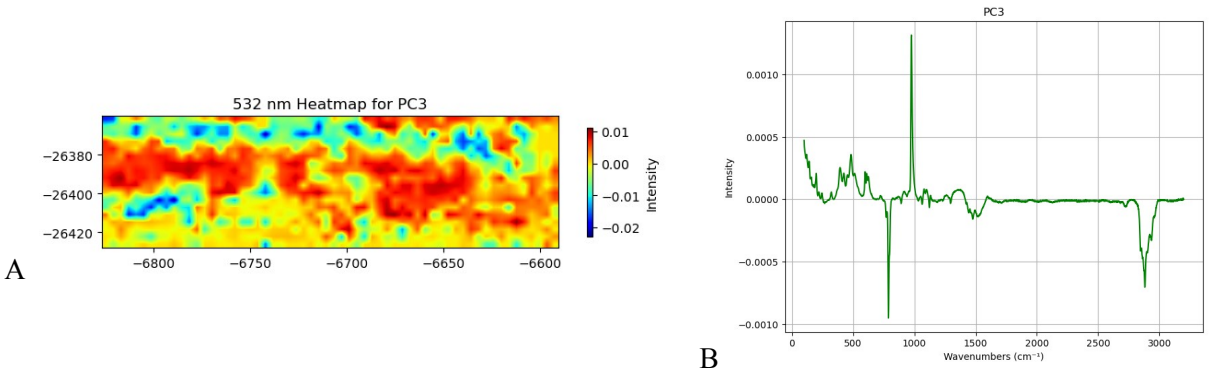


Figure 18: : PC3 for the 532 nm "green" block with the heatmap (A) and the loadings intensity graph (B). The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC4 is illustrated in Figure 19 with the heatmap and the corresponding loadings spectrum graph, and representing 10.78% of the explained variance. A broad peak at 2888 cm^{-1} represents the alkyl chains of the copper octanoate from the corrosion inhibition treatment. While sharp peaks at 221 and 790 cm^{-1} potentially represent alternative vibrational modes of these alkyl chains from the corrosion inhibition treatment. A prominent feature is the intense signal at the start of the spectrum (~99 cm^{-1}), followed by a cluster of smaller peaks at 417, 478, 539, and 639 cm^{-1}. This composition closely corresponds with the 532 nm reference spectrum for cuprite, suggesting its presence in the sample, but its signal intensity is mixed with the signal from the corrosion inhibition treatment making its precise identification difficult.

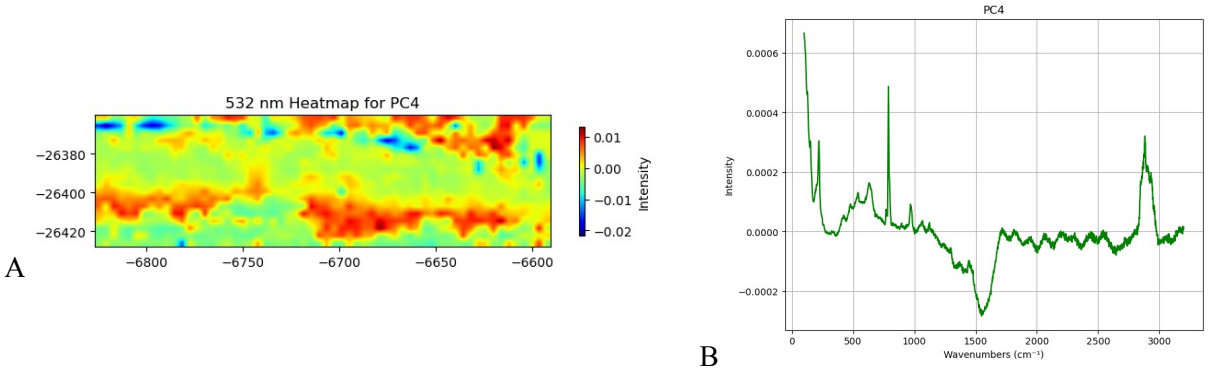


Figure 19: PC4 for the 532 nm "green" block with the heatmap (A) and the loadings intensity graph (B). The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC5 is shown in Figure 20 with the heatmap and the corresponding loadings spectrum graph, representing 3.40% of the explained variance. There is a broad signal peak at 1536 cm^{-1} representing amorphous carbon, and corresponds to other described literature for amorphous carbon [44,45]. The baseline does breakdown in this PC, and there is noise present, preventing further structural information from being obtained, besides the presence of amorphous carbon. But this baseline breakdown and noise does hinder the presence and signal of the amorphous carbon.

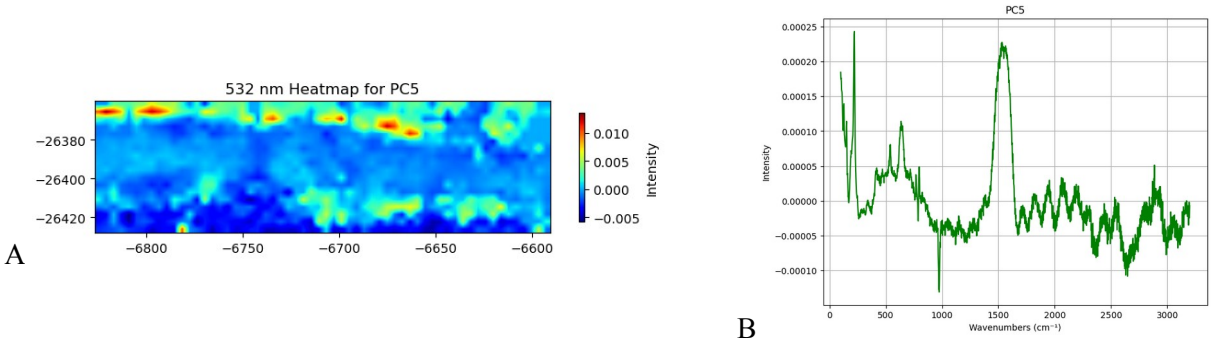


Figure 20: PC5 for the 532 nm "green" block with the heatmap (A) and the loadings intensity graph (B). The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

PC6 is shown in Figure 21 with the heatmap and the corresponding loadings spectrum graph, and representing 1.66% of the explained variance. There are major, sharp intense peaks at 225 cm and 646 cm. While the loadings spectrum graph in Figure 21B preserves the baseline and remains mostly unaffected by noise, little structural information can be obtained from this PC. The peak present in Figure 21B do not correspond to the 532 nm reference spectra. This makes chemical identification in corrosion layers difficult. It could be extrapolated that the high-intensity areas correspond to cuprite

based on their approximate locations; however, the observed peak intensities and composition are inconsistent with standard cuprite profiles.

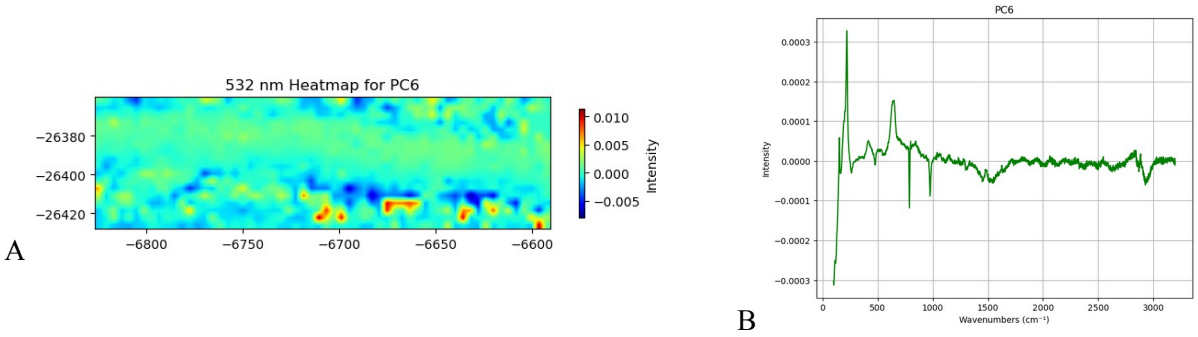


Figure 21: PC6 for the 532 nm "green" block with the heatmap (A) and the loadings spectrum graph (B). The intensity scale (A) is the relative intensity for the heatmap, and x,y axis corresponds to the internal positional coordinates of the spectrometer.

3.2 Results of the multi-block fusion

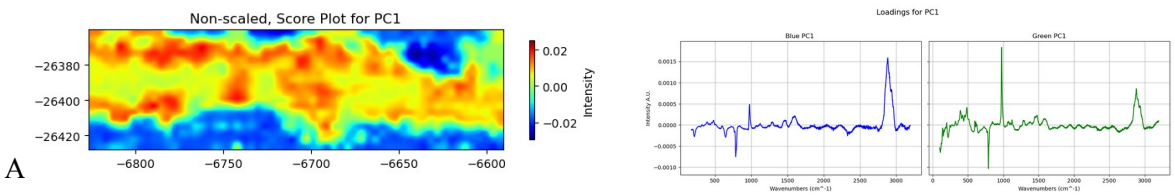
3.2.1 Results of the inter-block scaling

Inter-block scaling techniques between the two datasets were explored based on methods described by Campos & Reis [48]. These scaling techniques were evaluated using the SUM-PCA multi-block method, and compared against non-scaled data. Table 2 shows the total explained variance for the first PC1-6, along with their combined average value and its standard deviation. The remarkably low standard deviation is 0.00108 indicating that the choice of scaling technique does not significantly alter the multi-block results. Based off our comparative analysis scaling techniques for SUM-PCA, the other subsequent multi-block techniques were conducted using non-scaled data, SBS, and SBSV scaled data

Table 2: Comparison of the non-scaled versus scaling techniques using the first six principal components.

Principal Component #	No Scaling	SBS	HBS	SHBS	SBSV	HBVS	SHBVS	Average	Standard Deviation
PC1-6	0.9146	0.9148	0.9151	0.9170	0.9147	0.9149	0.9170	0.9154	0.00108

The heatmaps for PC1 are illustrated in Figure 18, comparing the non-scaled data against the SHBS and SBS scaling methodologies. Upon visual inspection, the PC1 profiles across all three preprocessing methods exhibit identical characteristics along their corresponding loadings spectra. This indicates that the choice of scaling method does not influence the model for this study and multi-block method, further validating the results from Table 2.



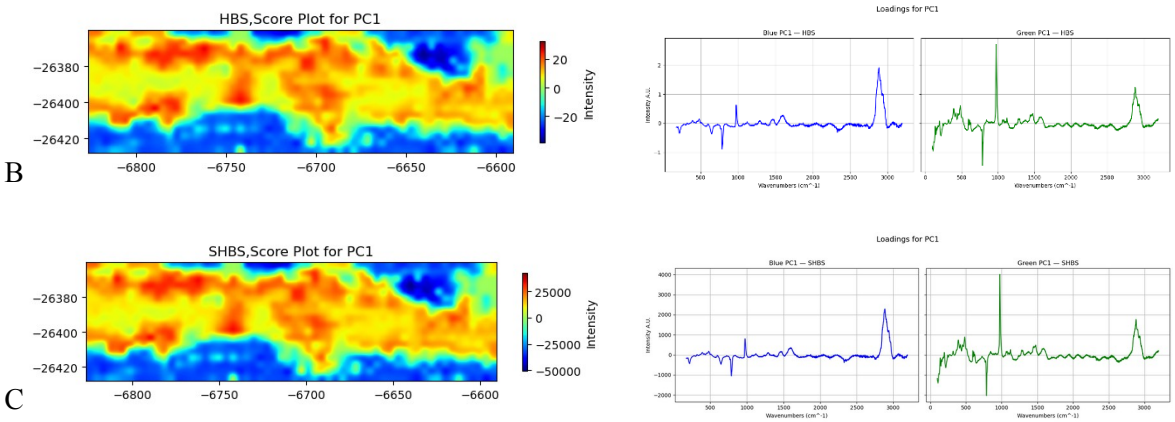


Figure 22: PC1 heatmaps for non-scaled (A), HBS (B), SHBS (C) scaled data. These are together to highlight that different scaling techniques produces similar results for the heatmaps and loading spectra indicating that scaling does not affect the results of these data sets.

3.2.2 Results of SUM-PCA

The SUM-PCA method only considers the shared components within each block, the corresponding scree plot is shown in Figure 23. While the first six PCs account for 91.46% of the cumulative explained variance, only four PC were usable for the objectives of this study. As previously discussed, and shown in Table 2 scaling does not greatly affect the results of this method. The explained variance is lower for SUM-PCA, 91.46%, when compared in the individual 473 nm, 97.61%, block and the 532 nm block, 95.83% as shown in Figure 10. This decrease in the explained variance stems from the mathematical constraints imposed by forcing both data blocks into a shared subspace. Which prevents the algorithm from perfectly optimizing for each block's each individual variance. Since SUM-PCA is only optimizing for the shared, global components, it represents a compromise between both datasets; any unshared block variance deflates the cumulative explained variance.

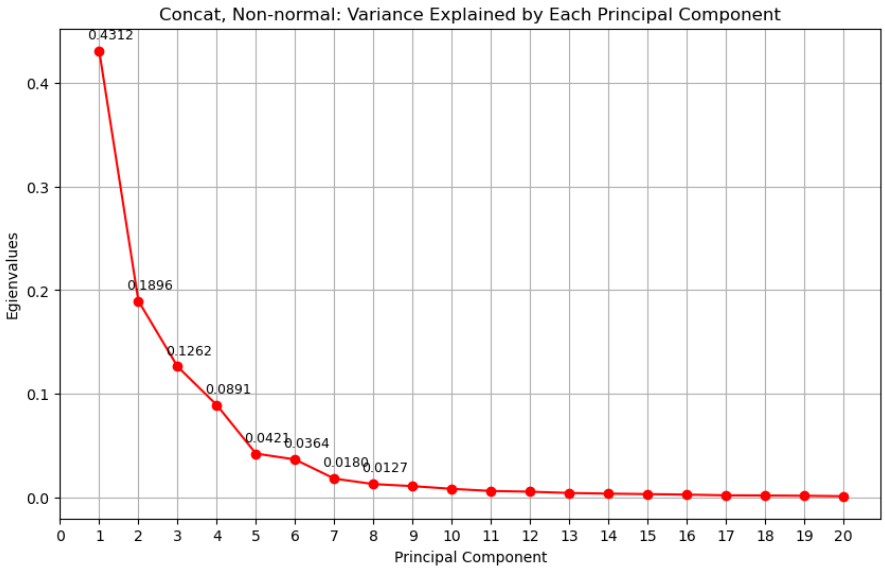


Figure 23: SUM-PCA scree plot for non-scaled data.

PC1 is shown in Figure 24Figure 24 for SUM-PCA, represents 43.12% of the explained variance. In the 473 nm component, the most prominent feature is a broad peak at 2885 cm^{-1} with a secondary sharp peak at 974 cm^{-1}. The most prominent feature in the 532 nm component is dominated by the sharp 974 cm^{-1} peak, with the broad 2882 cm^{-1} peak acting as a secondary feature. The 974 cm^{-1} peak represents brochantite, while the peaks at 2885 and 2882 cm^{-1} represent from the alkyl

chains of copper octanoate from the corrosion inhibition treatment. The co-existence of these signals is captured in the heatmap (Figure 24A), where regions of maximum intensity highlight the interaction zones between the corrosion inhibition treatment and the brochantite substrate.

The two signals from each block (473 nm & 532 nm) component are reflected in the corresponding heatmap shown in Figure 24A. Since the most intense peaks are in this PC are the brochantite and copper octanoate, the areas with the greatest intensity represent the combined contributions of both the copper octanoate and brochantite. Since, during the treatment of the copper coupon, the TEOS gel remains on the surface while the octanoic acid penetrates and reacts with brochantite to form copper octanoate. When looking at the heatmap in Figure 24A, the TEOS coating is on the outer most section of the brochantite forming copper octanoate. It appears the octanoic acid is penetrating deeper into the brochantite, and reacting with the brochantite to form copper octanoate. It can be extrapolated that the corrosion inhibition treatment is not uniformly present on the outermost section of the brochantite. Since, the areas of greatest intensity are not uniformly present, and these areas represent the combined contributions of both the copper octanoate/octanoate acid and brochantite. This is not ideal; the ideal scenario would be a uniform presence of the corrosion inhibition treatment throughout brochantite layer. The heatmap does show some deeper, uniform presence of the corrosion inhibition treatment coating into brochantite layer.

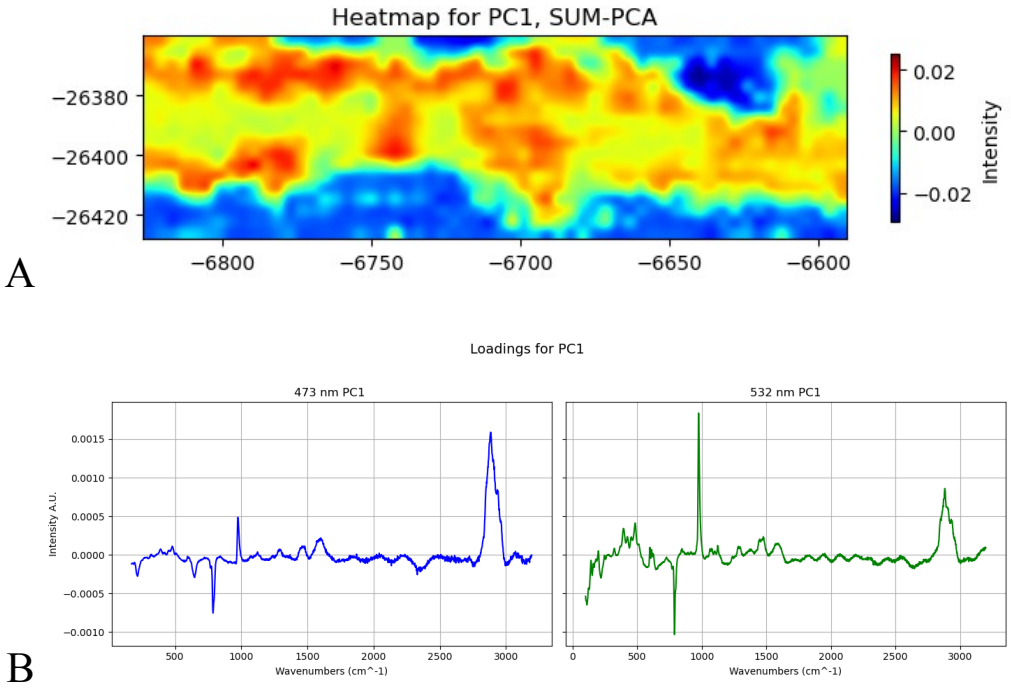


Figure 24: PC1 heatmap (A) and the loadings intensity graph (B) for SUM-PCA. Non-scaled data

The excitation responses, as shown in Figure 24B, of identical chemical exhibit distinct variations when comparing the 473 nm and 532 nm lasers. Specifically, excitation at 473 nm yields a significantly more pronounced response to the alkyl functional groups present in TEOS, copper octanoate, and octanoic acid than does the 532 nm source. Conversely, the 532 nm laser demonstrates a stronger excitation response to brochantite compared to the 473 nm laser.

PC3 for SUM-PCA is illustrated in Figure 25, accounts for 18.96% of the explained variance. Strong, sharp peaks appear at 972 cm^{-1} (473 nm block) and 974 cm^{-1} (532 nm block), confirming a robust brochantite response that is visibly stronger under 532 nm excitation. The corresponding heatmap maps the distribution of this brochantite phase. Negative features at 2881 and 2891 cm^{-1} correspond to the alkyl groups of the copper octanoate, complemented by overlapping negative bands

at 777 and 787 cm^{-1} that likely represent alternative alkyl vibrational modes. Consequently, the areas of lowest intensity in this heatmap effectively track the other locations of the gel coating, and copper octanoate.

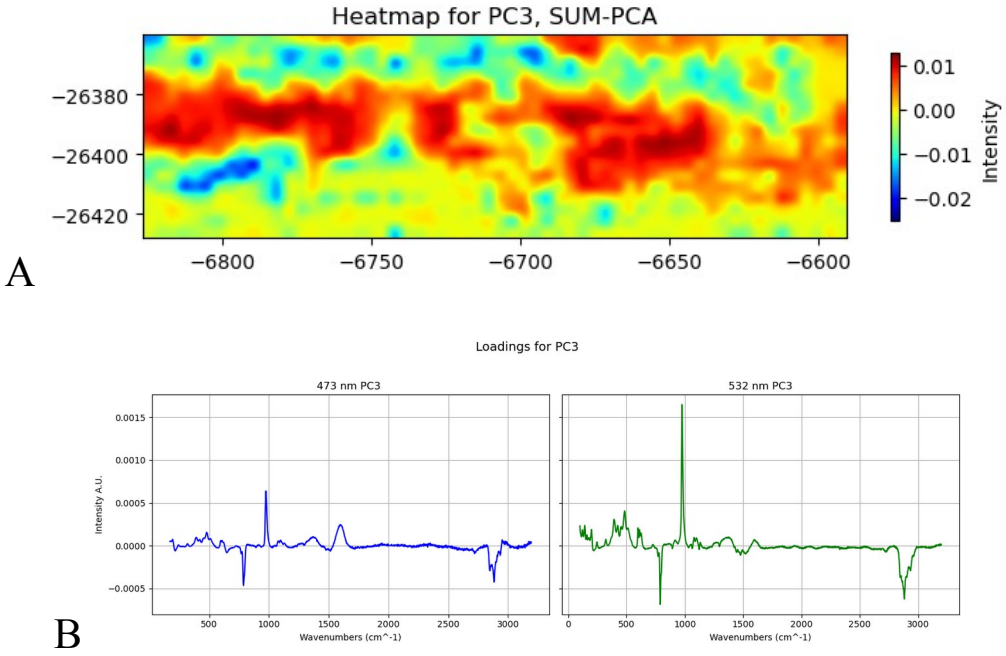


Figure 25: PC3 heatmap (A) and the loadings intensity graph (B) for SUM-PCA. Non-scaled data

PC4 is shown in Figure 26 for SUM-PCA representing 8.91% of the explained variance, and provides insight regarding the presence of cuprite distribution. The 473 nm spectrum shows sharp peaks at 210 and 650 cm^{-1}, while the 532 nm spectrum exhibits a characteristic low-wavenumber feature near 100 cm^{-1}, followed by peaks at 223, 630, and 785 cm^{-1}. The areas of maximum intensity match where cuprite (Figure 26A) is expected; however, the data does not show a continuous, well-defined layer of cuprite beneath the brochantite, but rather isolated, fragmented pockets. Notably, the most intense peak for 473 nm spectrum is at 650 cm^{-1} and the most intense peak for the 532 nm spectrum was at ~100 cm^{-1}.

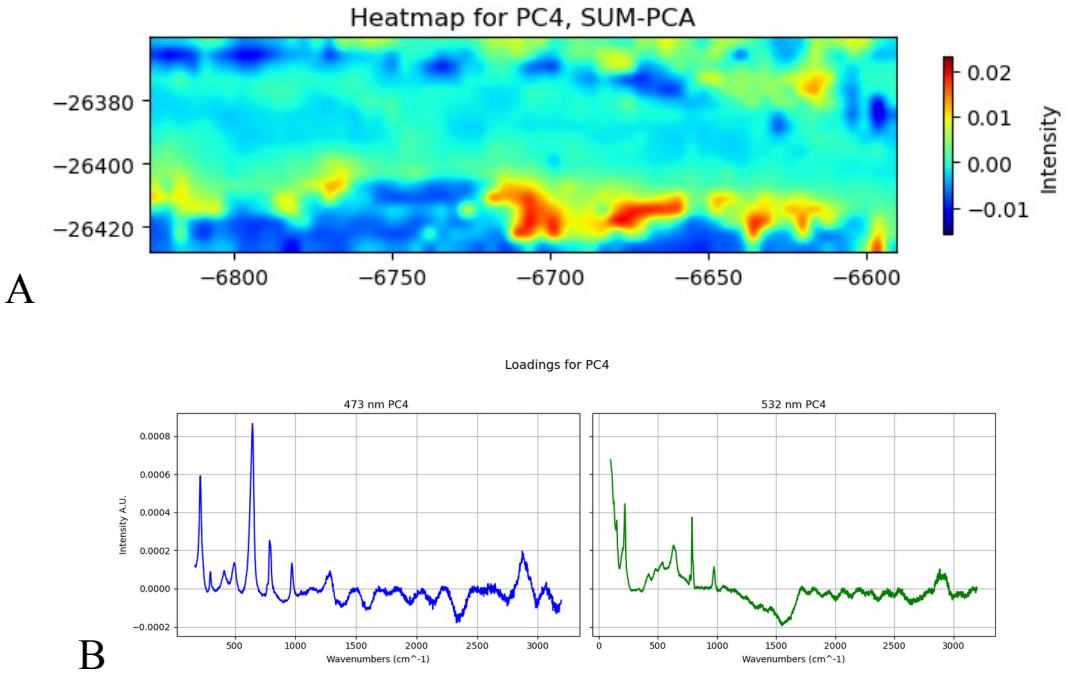


Figure 26: PC4 heatmap (A) and the loadings intensity graph (B) for SUM-PCA. Non-scaled data.

PC5 did not yield any useful information regarding the chemical distribution. PC6 accounts for 3.64% of the explained variance and could represent the location of cuprite, (Figure 27). The 473 nm spectrum in Figure 27B aligns with the cuprite reference spectrum (Figure 6). The 532 nm spectrum has a broad peak intensity at 1558 cm^{-1} corresponding to amorphous carbon present. High intensity areas in this heatmap (Figure 27A) could denote either cuprite or amorphous carbon.

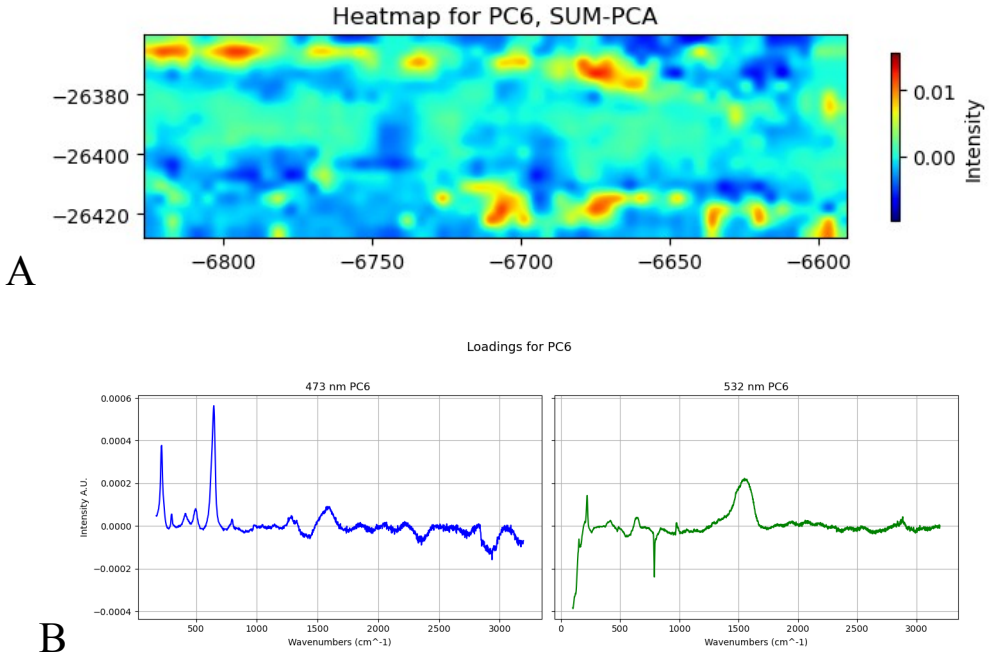


Figure 27: PC6 heatmap (A) and the loadings intensity graph (B) for SUM-PCA. Non-scaled data.

3.2.3 Results for AJIVE

The AJIVE method segregates the data into shared, joint and individual, block-specific variations. These unique features are expressed in heatmaps corresponding to their original data block. The two data blocks were found to have a joint score of zero, indicating the datasets within each block

do not share any common features. This is further supported by canonical correlation analysis (CCA), the calculated coefficient of determination, R^2, had a value of zero; indicating the two blocks do not have common features. Notably, AJIVE does not rank components strictly by variance in a scree plot because it structurally decomposes the data into independent shared and individual spaces, unlike traditional PCA which relies on a single variance metric [44].

The first AJIVE component is similar to PC1 of SUM-PCA and individual block models. For the 473 nm component, the primary feature remains the broad alkyl chain peak at 2885 cm^{-1} possibly from the copper octanoate. For the 532 nm component, the brochantite peak at 974 cm^{-1} dominates. Similar for SUM-PCA results in Figure 24B, the results in Figure 28C compared to the 532 nm laser, the 473 nm excitation laser elicits a stronger response from chemical components containing alkyl functional groups. Conversely, the 532 nm laser elicits a stronger response toward brochantite than the 473 nm laser.

By comparing these two distinct heatmaps (Figure 28A & Figure 28B) provides a clear picture of how the corrosion inhibition treatment is distributed within the brochantite layer. Since the areas of greatest intensity in the 473 nm component represent the inhibition treatment, and the areas of greatest intensity for the 532 nm component represent the brochantite. It can be extrapolated the estimated location that the corrosion inhibition treatment is located in the sample. The individual heatmaps show that the corrosion inhibition treatment is located the outermost boundary of the brochantite layer. While there is some penetration of the octanoic acid into the brochantite layer forming copper octanoate, but suffers from a non-uniform distribution and limited penetration into the brochantite. The non-uniform distribution of the protective treatment within the brochantite might likely arise from the porous structure of the brochantite, which unevenly channels the diffusing octanoic acid as it reacts to yield copper octanoate. This non-uniform distribution could also be caused by polarities differences between the brochantite and octanoic acid.

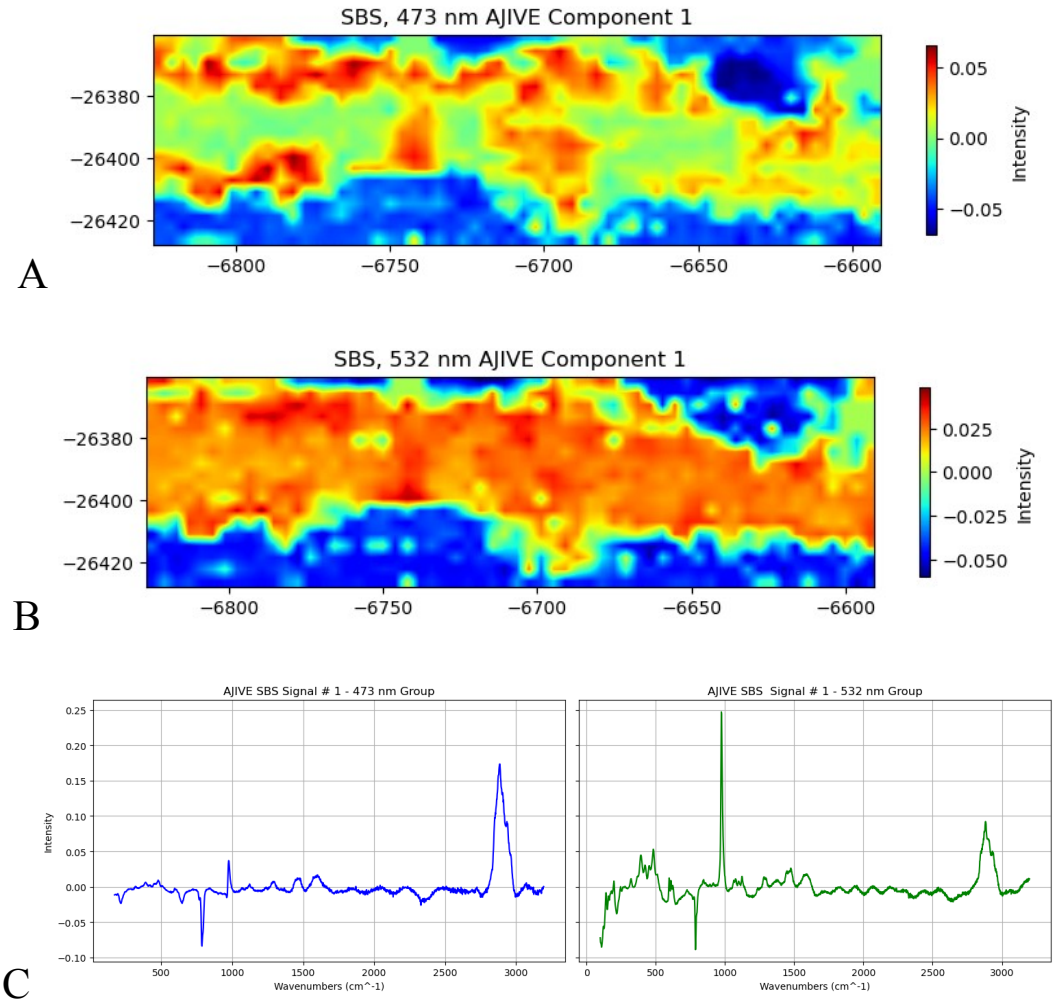


Figure 28: The first component for AJIVE for the 473 nm component heatmap (A), the 532 nm component, and the intensity spectra graphs (C).

The third AJIVE component (Figure 29) reflects the individual block PCA results for PC3 (Figure 13 and Figure 18) . The 473 nm spectrum displays sharp cuprite markers at 215, 645, and 786 cm^{-1}. Notably, the heatmap in Figure 29A closely resembles the individual 473 nm PC3 heatmap in Figure 13A, though their intensities are inverted; the AJIVE 473 nm component exhibits negative spectral loading peaks, whereas Figure 13B displays positive loading peaks. While the 532 nm spectrum characterizes the brochantite at 975 cm^{-1}, and moderate negative peak at 2883 cm^{-1} that represents alkyl groups on from copper octanoate. But this signal overlaps with other negative in the 532 nm that prevents good identification of the corrosion inhibition treatment present.

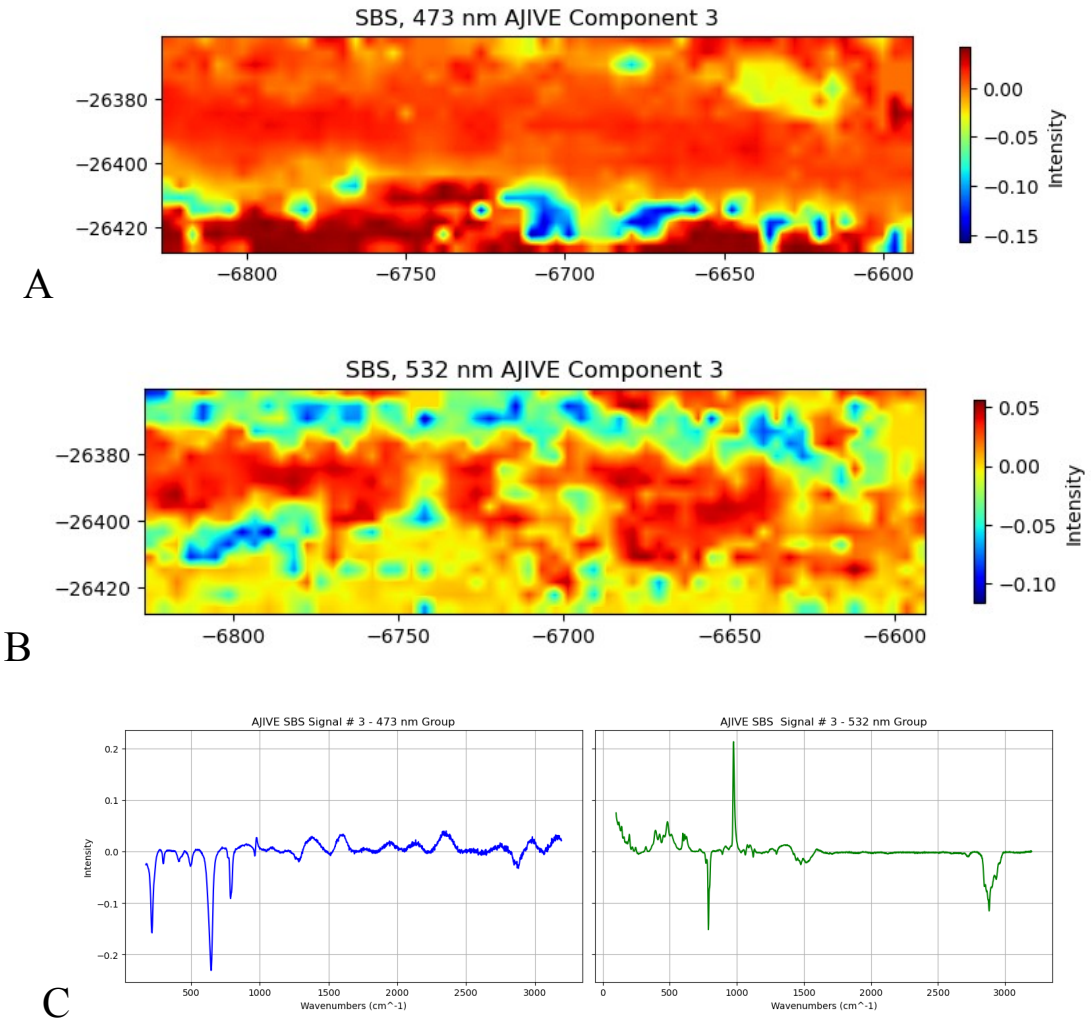


Figure 29: The third component for AJIVE for the 473 nm component heatmap (A), the 532 nm component, and the intensity spectra graphs (C)

Overall, the AJIVE method results are similar with the individual block PCA analyses discussed in Section 3.1.4. There are slight differences in the spectrum from the spectra of AJIVE and individual block PCA such as the inverse spectrum shown in Figure 29A compared to Figure 13B.

Table 3 is a summary of the different multi-block techniques evaluated in this study. The methods are differentiated by whether they utilize only shared features or both shared and unique features. The output type indicates whether the method generates joint heatmaps or individual block heatmaps. Out of six calculated components, the number of meaningful components is listed alongside the specific chemical constituents identified by each technique. Overall, the majority of the methods yielded similar results to SUM-PCA methods expect for the AJIVE method. Notably, both the 473 nm and 532 nm reference spectra for both copper octanoate and octanoic acid had broad peaks present at 2884, 2932 cm^{-1}, 2884, and 2925 cm^{-1} respectively. This similar peak location and structure can make identification between the two compounds, possibly the copper octanoate identified could be identified as octanoic acid due the two spectrum being very similar.

Multi-block Method	Shared Components Analysis Only	Shared & Distinct Component Analysis	Joint or Individual Heatmap	The number of components useful out of six components	Chemical constituents Identified
SUM-PCA	√		Joint Heatmap	4	brochantite, copper octanoate, cuprite, octanoic acid,
ComDim	√		Joint Heatmap	4	brochantite, copper octanoate, cuprite, octanoic acid
H-PCA	√		Joint Heatmap	4	brochantite, copper octanoate, cuprite, octanoic acid
CPCA	√		Joint Heatmap	4	brochantite, copper octanoate, cuprite, octanoic acid
MOCA	√		Joint Heatmap	4	brochantite, copper octanoate, cuprite, octanoic acid
MFA	√		Joint Heatmap	4	brochantite, copper octanoate, cuprite, octanoic acid
DISCO-SCA		√	Joint Heatmap	4	brochantite, copper octanoate, cuprite, octanoic acid
AJIVE		√	Individual Heatmap	2	brochantite, copper octanoate, cuprite, octanoic acid

Table 3: Table outlining the different multi-block techniques used in this study. The different methods are differentiated if it utilizes only shared feature or shared & unique features in its calculation. The methods are listed if the results produce a joint, overlaid heatmap of the two blocks or individual heatmaps of the two blocks. Six components were calculated for each method, the number of useful components from each method is displayed in the table. The different chemical components that were able to be identified are listed in the table.

4. Discussion

Overall, this study demonstrated that the carboxylic acid is present from the corrosion inhibition treatment; it reacted on the surface of the corrosion product layer of the copper coupon. In the cross-section of the corrosion layer indicated that the octanoic acid successfully penetrated the corrosion layers, and one part reacted with the brochantite by forming copper octanoate. Corrosion protection provided by copper carboxylate is explained by its hydrophobic character, which decreases wettability, prevent contact with water and aggressive species (Cl^-), inhibiting the cathodic reaction. The remaining octanoic acids, present in the corrosion product layer (CPL) could inhibit the anodic reaction by complexing copper cation, produced during the corrosion process. While the coating was present on the outermost boundary of the brochantite layer, and there was deeper penetration of the coating into the brochantite layer. Ideally, the corrosion inhibitor would achieve uniform penetration throughout the entire brochantite layer. However, the present results indicate that complete uniformity was not realized. It is crucial to have deep penetration to achieve a greater resistance to environmental stresses and degradation; it ensures the entire corrosion layer is protected with the copper octanoate, not only the surface portion of the brochantite.

The Raman signature of TEOS gel was difficult to discern during analysis due its poor crystalline structure creating a weak Raman response. Furthermore, the alkyl chain present in the TEOS, copper

octanoate, and octanoic acid exhibit overlapping distinct signals specifically around ~2880 cm^{-1}. Given that the coating formulation is a solution of TEOS and octanoic acid, it could be assumed these alkyl prominent peaks represents the coating formulation's presence, including the octanoic acid, providing corrosion protection. Alternatively, it is possible that the octanoic acid failed to migrate alongside the TEOS gel, thereby failing to provide the intended corrosion protection. Further research and techniques are required to solve this issue, specifically the presence of the octanoic acid, copper octanoate, and TEOS in the sample as their current spectra are still highly mixed together. Potentially, an additional Raman excitation wavelength may offer a viable pathway to resolve these compounds.

In this study, we showed that the 473 nm and 532 nm had different signal intensity responses for the chemical components containing the alkyl groups and the brochantite. The 473 nm laser had a stronger intensity response than the 532 nm laser for the chemical components containing the alkyl groups. For non-resonant Raman scattering, the intensity, I, is proportional to $1/\ \lambda^4$ where λ is the excitation wavelength, this causes bonding in alkyl groups to produce stronger signal responses for shorter wavelengths [90–92].The 532 nm laser had a stronger intensity response for the brochantite when compared to the 473 nm laser. This is due to the Raman spectroscopy resonance effect with the 532 nm laser being close to photon electronic transition of the brochantite. Since brochantite is a greenish color, it reflects green light and 532 nm corresponds to green-light, the vibrational response is amplified producing a stronger peak intensity. The issue differentiating between the parts of the corrosion inhibition treatment could be solved by using a shorter Raman wavelength, since the peak signals are very close together. For non-resonant Raman scattering produces a greater response for shorter wavelengths. Utilizing a longer excitation wavelength may improve the identification of cuprite by approaching its primary reflection band, thereby enhancing the resonance Raman effect and yielding a stronger spectral response.

Deep learning and classical supervised learning techniques are best suited for scenarios with massive amounts of diverse, labeled data. Deep learning can introduce a risk of over-fitting, especially given that labeled datasets are not readily available. Consequently, an exploratory, unsupervised approach is required because the specific chemical constituents of the sample are unknown. The goal of these exploratory techniques is to resolve potential chemical constituents without prior knowledge, accounting for the possibility that reference components may not be present in the sample.

Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) is a popular chemometrics technique used to resolve individual spectra from a mixture, similar to Independent Component Analysis (ICA). However, while MCR-ALS can operate with minimal prior knowledge, it typically assumes or requires initial estimates of the components expected to be in the sample components, while presence cannot be guaranteed here. While autoencoders could serve as an unsupervised alternative, they map data into non-linear spaces. Because the spectral mixtures in this study adhere to a linear space, using a non-linear auto-encoder would distort the physical reality, making it incredibly difficult to accurately match and interpret the resolved spectra.

In section 3.2.1, the results of the inter-block scaling indicated that scaling methods do not influence the results. This is a narrowly specific result of scaling method techniques based off solely SUM-PCA and our two blocks. Future research can be done analyzing if the scaling techniques would affect other multi-block methods or if additional Raman data blocks would affect the scaling results. In Campos & Reis 2020, utilized predictive techniques to evaluate the performance of each scaling technique, and the concluded indecisively which method was the best [54]. Possibly, incorporating prediction techniques in this study could be used an additional metric to evaluate scaling method performance.

Analytical comparison methods, such as the Structural Similarity Index Measure (SSIM), are commonly used to assess quality differences between a processed image and a reference image. SSIM evaluates visual degradation by analyzing three distinct components; luminance, contrast, and

structure. However, when applying multi-block techniques, each method computes intensity values based on its own distinct mathematical framework. Consequently, the relative intensity from a heatmap generated by one multi-block method will not match the intensity scale of another. Even if the heatmaps appear visually identical, these differing intensity scales distort the final index calculations, rendering direct comparisons invalid. Other studies comparing multi-block methods often incorporate predictive techniques to evaluate model performance, this is option was not available. Methodologies can be researched and developed to introduce quantitative framework for comparison to this study.

Although this study focuses on hyperspectral imaging for chemometrics and cultural heritage materials, the methodology is readily applicable to other domains that utilize multi-block techniques or hyperspectral imaging. In other chemometric fields, multi-block techniques are frequently employed to integrate data from disparate analytical platforms through data fusion. While a significant portion of this relatively new multi-block fusion literature targets biological applications such as multi-modal, multi-platform analyses in the food and medical sectors. Hyperspectral imaging historically originated from geospatial satellite imaging; hyperspectral imaging has expanded into chemometrics [45]. Although chemometrics is gaining traction in cultural heritage, its adoption has been slower due to the smaller size of the research community. Only recently have these multi-block fusion techniques been applied to analyze cultural heritage materials, as demonstrated by Li et al. (2024), Lambert et al. (2016), and Pagnin et al. (2022) [46–48]. Furthermore, these multi-block methods could be translated back to their geospatial origins, data can be potentially be fused together from multiple sensors on satellites or drone technology [45].

All the methods that produced a joint, overlaid heatmap (Table 3) yielded similar results. These methods calculated a total of four useful components from the initial six, mainly because they condense the datasets of each block into a global, blended consensus. AJIVE produces different results because it mathematically partitions the datasets into joint and individual structures using subspace angles without ever computing a global consensus. Although DISCO-SCA also isolates distinct features, it establishes a global consensus first before analyzing individual variation. Interestingly, DISCO-SCA required several days to compute compared to the seconds or minutes required by the other multi-block methods. Since these traditional multi-block methods are designed to be used with three or more blocks, adding an extra dataset, such as an additional Raman excitation wavelength, would likely yield a larger variety in the results.

SUM-PCA and AJIVE were implemented in Python, using well-documented and standard machine learning Python packages such as Scikit-learn. While the other methods were implemented in R using less documented and specific multi-block R packages. It is important to consider this difference in programming languages. Since Python is the current standard for chemometrics and machine learning, an easily approachable syntax. R has historically been the favored programming for statistical analysis, but its syntax is more cumbersome, and lacks matured machine learning libraries, that do not integrate well into other contemporary machine learning workflows. Similarly, some of the R packages in this study have been implemented in MATLAB as well. This also presents similar issues as R, due to Python as the current standard for machine learning and proprietary nature of MATLAB software. There is software development needed to translate these packages into Python to be used for contemporary machine learning application for more widespread adoption for multi-block fusion methods [52,96–98].

Given that the different multi-block methods offered comparable results, it is important to determine which method is optimal for future research. Furthermore, while this study focused exclusively on octanoic acid as the corrosion inhibitor, the methodology could potentially be extended to other carboxylic acids. SUM-PCA would be the preferred method, since almost every multi-block yielded similar results. It is computationally and methodologically simple method, it can be

accomplished with well-documented Python packages. While the other multi-block methods are more computationally and methodologically complex and uses more obscure Python or R packages.

Future research could expand upon this study by increasing the number of data blocks analyzed. Because this study was limited to two blocks utilizing Raman excitation wavelengths of 473 nm and 532 nm. The incorporation of additional Raman excitation wavelengths could yield a more robust evaluation of multi-block techniques. These methods are intended for more complex data fusion problems consisting of three or more blocks. Furthermore, expanding the dataset could enhance the resolution of chemical species within the heatmaps, particularly for those containing alkyl groups. Future work should also focus on the underlying algorithms of these methods. For example, AJIVE first described in the literature in 2018, and improves upon the original JIVE method by achieving comparable results with lower computational expense [44]. Additionally, translating the existing R packages into Python would better integrate these techniques into modern machine learning workflows. In conclusion, utilizing the most straightforward and least computationally intensive method is optimal for simpler data fusion scenarios, including future research into anti-corrosion coatings.

5. Conclusion

Overall, this study successfully identified the different constituent components of the copper corrosion layer system and the corrosion inhibition treatment using multi-block data fusion techniques. Further, research is still needed to improve the identification methods as some of the chemical constituents are still difficult to resolve from one another. Multi-block fusion does improve the qualitative analysis of the corrosion layer to provide a better comprehensive understanding how the corrosion treatment is interacting with the corrosion layer. This study showed the different multi-block methods yield very similar results for two-block hyperspectral imaging. It is recommended to perform the computationally less intense and simplistic method to for future research. Over the course of this study, the dual-wavelength Raman hyperspectral approach combined with multi-block data fusion proved to be a viable, non-destructive strategy for characterizing both the corrosion stratigraphy and the octanoic acid-doped TEOS coating, confirming that the inhibitor reacted with the brochantite layer to form copper octanoate at the outer boundary of the corrosion product layer, though complete penetration through the brochantite was not achieved.

These findings carry practical significance for the conservation of copper-based cultural heritage artifacts. By confirming that octanoic acid reacts with brochantite to form a hydrophobic copper octanoate barrier, this work supports the continued development of eco-friendly, sol-gel based corrosion inhibitors as an alternative to more toxic conventional treatments. At the same time, the incomplete penetration of the inhibitor observed here underscores the need for further formulation optimization before such treatments can be reliably deployed on artifacts with thick or heterogeneous corrosion layers. Methodologically, this study demonstrates that multi-block data fusion, and SUM-PCA in particular, offers conservation scientists an accessible and computationally efficient framework for combining multi-wavelength Raman data, one that could be extended to additional spectroscopic modalities or corrosion inhibitor formulations in future work. Taken together, these results advance both the chemical understanding of copper corrosion inhibition and the analytical toolkit available for evaluating protective treatments, laying groundwork for more robust, standardized approaches to cultural heritage conservation.

Appendix

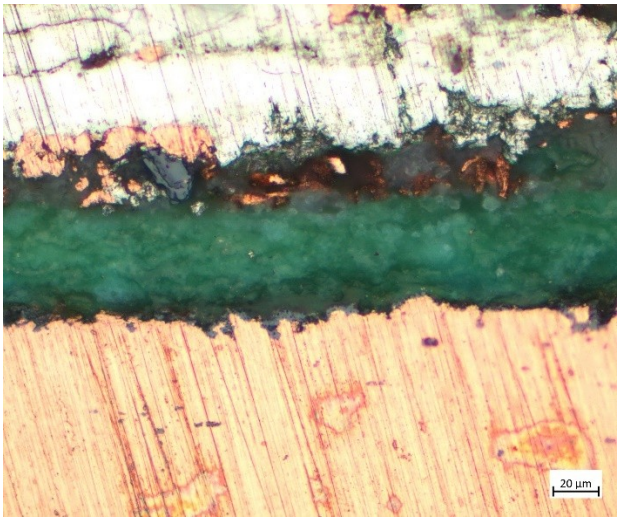


Figure 30: Micrograph cross-section of the copper coupon after polishing with the corrosion layers present from this study. The copper layer is present at the bottom, and the corrosion layers of cuprite (10 µm) beneath a thicker of brochantite (40–50 µm) are present above the copper metal.

References

1. Leygraf, C., Chang, T., Herting, G. & Odnevall Wallinder, I. The origin and evolution of copper patina colour. *Corros. Sci.* **157**, 337–346 (2019).

2. Apchain, E. *et al.* Efficiency and durability of protective treatments on cultural heritage copper corrosion layers. *Corros. Sci.* **183**, 109319 (2021).

3. FitzGerald, K. P., Nairn, J., Skennerton, G. & Atrens, A. Atmospheric corrosion of copper and the colour, structure and composition of natural patinas on copper. *Corros. Sci.* **48**, 2480–2509 (2006).

4. Lob, S., Neff, D., Tran-Thi, T.-H., Richter, M. C. & Rivron, C. Hydrophobic coating using sustainable sol-gel process doped with carboxylic acids to protect heritage copper artefacts. *Prog. Org. Coat.* **186**, 108035 (2024).

5. Blake, J. On Defining the Cultural Heritage. *Int. Comp. Law Q.* **49**, 61–85 (2000).

6. How Science Can Contribute to the Remedial Conservation of Cultural Heritage - Baglioni - 2021 - Chemistry – A European Journal - Wiley Online Library. https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/chem.202100675.

7. Scott, D. A. *Copper and Bronze in Art: Corrosion, Colorants, Conservation*. (Getty Conservation Institute, Los Angeles, 2002).

8. Muşkara, Ü., Bozbaş, S. K., Telli, M. B. & Konak, A. Early copper production by the last hunter-gatherers. *J. Archaeol. Sci. Rep.* **62**, 105051 (2025).

9. Huisman, H. *et al.* Change lost: Corrosion of Roman copper alloy coins in changing and variable burial environments. *J. Archaeol. Sci. Rep.* **47**, 103799 (2023).

10. Ferrari, E. Understanding the inhibition corrosion mechanisms for copper based alloys by a physico-chemical multi-scale approach.

11. Benzarti, Z. *et al.* Copper corrosion mechanisms, influencing factors, and mitigation strategies for water circuits of heat exchangers: critical review and current advances. *Corros. Rev.* **43**, 429–455 (2025).

12. Abdel-Karim, A. M. & El-Shamy, A. M. A Review on Green Corrosion Inhibitors for Protection of Archeological Metal Artifacts. *J. Bio- Tribo-Corros.* **8**, 35 (2022).

13. Vargas, I. T. *et al.* Copper Corrosion and Biocorrosion Events in Premise Plumbing. *Materials* **10**, 1036 (2017).

14. Krätschmer, A., Odnevall Wallinder, I. & Leygraf, C. The evolution of outdoor copper patina. *Corros. Sci.* **44**, 425–450 (2002).

15. Ives, D. J. G. & Rawson, A. E. Copper Corrosion: III . Electrochemical Theory of General Corrosion. *J. Electrochem. Soc.* **109**, 458 (1962).

16. de la Fuente, D., Simancas, J. & Morcillo, M. Morphological study of 16-year patinas formed on copper in a wide range of atmospheric exposures. *Corros. Sci.* **50**, 268–285 (2008).

17. Fitzgerald, K. P., Nairn, J. & Atrens, A. The chemistry of copper patination. *Corros. Sci.* **40**, 2029–2050 (1998).

18. Knotkova, D. & Kreislova, K. Atmospheric corrosion and conservation of copper and bronze. in *WIT Transactions on State of the Art in Science and Engineering* (ed. Moncmanová, A.) vol. 1 107–142 (WIT Press, 2007).

19. Bernardi, E., Bowden, D. J., Brimblecombe, P., Kenneally, H. & Morselli, L. The effect of uric acid on outdoor copper and bronze. *Sci. Total Environ.* **407**, 2383–2389 (2009).

20. Balogh, K., Slížková, Z. & Kreislová, K. Research on effects of bird excrement on metal materials copper and bronze. **2017**, (2017).

21. Graedel, T. E. Copper patinas formed in the atmosphere—II. A qualitative assessment of mechanisms. *Corros. Sci.* **27**, 721–740 (1987).

22. Antonijevic, M. M. & Petrovic, M. B. Copper Corrosion Inhibitors. A review. *Int. J. Electrochem. Sci.* **3**, 1–28 (2008).

23. Molina, M. T., Cano, E. & Ramírez-Barat, B. Protective coatings for metallic heritage conservation: A review. *J. Cult. Herit.* **62**, 99–113 (2023).

24. Chaubey, N., Savita, Qurashi, A., Chauhan, D. S. & Quraishi, M. A. Frontiers and advances in green and sustainable inhibitors for corrosion applications: A critical review. *J. Mol. Liq.* **321**, 114385 (2021).

25. Artesani, A., Di Turo, F., Zucchelli, M. & Traviglia, A. Recent Advances in Protective Coatings for Cultural Heritage–An Overview. *Coatings* **10**, 217 (2020).

26. Cini, D. K. Sol-Gel Coatings for the Protection of Ferrous Heritage Metal.

27. Wysocka, J., Cieslik, M., Krakowiak, S. & Ryl, J. Carboxylic acids as efficient corrosion inhibitors of aluminium alloys in alkaline media. *Electrochimica Acta* **289**, 175–192 (2018).

28. Fischer, C. & Kakoulli, I. Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications. *Stud. Conserv.* **51**, 3–16 (2006).

29. de Juan, A. Hyperspectral image analysis. When space meets Chemistry. *J. Chemom.* **32**, e2985 (2018).

30. Dong, X. *et al.* A review of hyperspectral imaging for nanoscale materials research. *Appl. Spectrosc. Rev.* **54**, 285–305 (2019).

31. Tejasree, G. & Agilandeeswari, L. An extensive review of hyperspectral image classification and prediction: techniques and challenges. *Multimed. Tools Appl.* **83**, 80941–81038 (2024).

32. Amigo, J. M., Babamoradi, H. & Elcoroaristizabal, S. Hyperspectral image analysis. A tutorial. *Anal. Chim. Acta* **896**, 34–51 (2015).

33. Kurz, W. *et al.* Dimensionality reduction in hyperspectral imaging using standard deviation-based band selection for efficient classification. *Sci. Rep.* **15**, 34478 (2025).

34. Jones, R. R., Hooper, D. C., Zhang, L., Wolverson, D. & Valev, V. K. Raman Techniques: Fundamentals and Frontiers. *Nanoscale Res. Lett.* **14**, 231 (2019).

35. The molecular scattering of light.

36. Innocenti, S. *et al.* Historical Pigments and Paint Layers: Raman Spectral Library with 852 nm Excitation Laser. *Minerals* **14**, 557 (2024).

37. Rodriguez, A., Serada, V., Stover, P. & Kurouski, D. Raman on ramen: Nutritional analysis and brand identification. *J. Raman Spectrosc.* **54**, 806–813 (2023).

38. Edwards, H. G. M., Vandenabeele, P. & Colomban, P. *Raman Spectroscopy in Cultural Heritage Preservation*. (Springer International Publishing, Cham, 2023). doi:10.1007/978-3-031-14379-3.

39. Pandey, D. K., Kagdada, H. L., Sanchora, P. & Singh, D. K. Overview of Raman Spectroscopy: Fundamental to Applications. in *Modern Techniques of Spectroscopy: Basics, Instrumentation, and Applications* (eds Singh, D. K., Pradhan, M. & Materny, A.) 145–184 (Springer, Singapore, 2021). doi:10.1007/978-981-33-6084-6_6.

40. Ferraro, J. R. *Introductory Raman Spectroscopy*. (Elsevier, 2003).

41. Chandra, A. *et al.* Unveiling the Molecular Secrets: A Comprehensive Review of Raman Spectroscopy in Biological Research. *ACS Omega* **9**, 50049–50063 (2024).

42. Chiriu, D., Pisu, F. A., Ricci, P. C. & Carbonaro, C. M. Application of Raman Spectroscopy to Ancient Materials: Models and Results from Archaeometric Analyses. *Materials* **13**, 2456 (2020).

43. Yogurtcu, B., Cebi, N., Koçer, A. T. & Erarslan, A. A Review of Non-Destructive Raman Spectroscopy and Chemometric Techniques in the Analysis of Cultural Heritage. *Molecules* **29**, 5324 (2024).

44. Smith, G. D. & Clark, R. J. H. Raman microscopy in archaeological science. *J. Archaeol. Sci.* **31**, 1137–1160 (2004).

45. Coccato, A. & Caggiani, M. C. An overview of Principal Components Analysis approaches in Raman studies of cultural heritage materials. *J. Raman Spectrosc.* **55**, 125–147 (2024).

46. Mishra, P. *et al.* Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC Trends Anal. Chem.* **137**, 116206 (2021).

47. Smilde, A. K., Næs, T. & Liland, K. H. *Multiblock Data Fusion in Statistics and Machine Learning: Applications in the Natural and Life Sciences*. (John Wiley & Sons, 2022).

48. Azcarate, S. M., Ríos-Reina, R., Amigo, J. M. & Goicoechea, H. C. Data handling in data fusion: Methodologies and applications. *TrAC Trends Anal. Chem.* **143**, 116355 (2021).

49. Mafata, M., Brand, J., Medvedovici, A. & Buica, A. Chemometric and sensometric techniques in enological data analysis. *Crit. Rev. Food Sci. Nutr.* **63**, 10995–11009 (2023).

50. Hayes, E., Greene, D., O'Donnell, C., O'Shea, N. & Fenelon, M. A. Spectroscopic technologies and data fusion: Applications for the dairy industry. *Front. Nutr.* **9**, 1074688 (2023).

51. Borràs, E. *et al.* Data fusion methodologies for food and beverage authentication and quality assessment – A review. *Anal. Chim. Acta* **891**, 1–14 (2015).

52. Perry, R. *et al.* mvlearn: Multiview Machine Learning in Python.

53. Mahieu, B., Tchandao Mangamana, E., Vigneau, E. & Cariou, V. MBAnalysis: An R package for multiblock data analysis. *SoftwareX* **35**, 102780 (2026).

54. Campos, M. P. & Reis, M. S. Data preprocessing for multiblock modelling – A systematization with new methods. *Chemom. Intell. Lab. Syst.* **199**, 103959 (2020).

55. Campos, M. P., Sousa, R., Pereira, A. C. & Reis, M. S. Advanced predictive methods for wine age prediction: Part II – A comparison study of multiblock regression approaches. *Talanta* **171**, 132–142 (2017).

56. Ammari, F., Le Penglau, R., Bouhier, M. & Neff, D. Characterization of long-term atmospheric corrosion of iron in indoor and outdoor conditions using Raman spectroscopy coupled to multivariate curve resolution-alternating least squares. *Talanta* **305**, 129569 (2026).

57. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **374**, 20150202 (2016).

58. Farrugia, J., Griffin, S., Valdramidis, V. P., Camilleri, K. & Falzon, O. Principal component analysis of hyperspectral data for early detection of mould in cheeselets. *Curr. Res. Food Sci.* **4**, 18–27 (2021).

59. Luciano, G., Leardi, R. & Letardi, P. Principal component analysis of colour measurements of patinas and coating systems for outdoor bronze monuments. *J. Cult. Herit.* **10**, 331–337 (2009).

60. Moncayo, S. *et al.* Exploration of megapixel hyperspectral LIBS images using principal component analysis. *J. Anal. At. Spectrom.* **33**, 210–220 (2018).

61. Uddin, Md. P., Mamun, Md. A. & Hossain, Md. A. PCA-based Feature Reduction for Hyperspectral Remote Sensing Image Classification. *IETE Tech. Rev.* **38**, 377–396 (2021).

62. Artesani, A., Di Turo, F., Zucchelli, M. & Traviglia, A. Recent Advances in Protective Coatings for Cultural Heritage–An Overview. *Coatings* **10**, 217 (2020).

63. Ríos-Reina, R., Callejón, R. M., Savorani, F., Amigo, J. M. & Cocchi, M. Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars. *Talanta* **198**, 560–572 (2019).

64. Van Deun, K., Smilde, A. K., van der Werf, M. J., Kiers, H. A. & Van Mechelen, I. A structured overview of simultaneous component based data integration. *BMC Bioinformatics* **10**, 246 (2009).

65. Hanafi, M., Kohler, A. & Qannari, E.-M. Connections between multiple co-inertia analysis and consensus principal component analysis. *Chemom. Intell. Lab. Syst.* **106**, 37–40 (2011).

66. Smilde, A. K., Westerhuis, J. A. & de Jong, S. A framework for sequential multiblock component methods. *J. Chemom.* **17**, 323–337 (2003).

67. Hanafi, M., Mazerolles, G., Dufour, E. & Qannari, E. M. Common components and specific weight analysis and multiple co-inertia analysis applied to the coupling of several measurement techniques. *J. Chemom.* **20**, 172–183 (2006).

68. Abdi, H. & Valentin, D. Multiple Factor Analysis (MFA).

69. Måge, I., Smilde, A. K. & van der Kloet, F. M. Performance of methods that separate common and distinct variation in multiple data blocks. *J. Chemom.* **33**, e3085 (2019).

70. Schouteden, M., Van Deun, K., Wilderjans, T. F. & Van Mechelen, I. Performing DISCO-SCA to search for distinctive and common information in linked data. *Behav. Res. Methods* **46**, 576–587 (2014).

71. Feng, Q., Jiang, M., Hannig, J. & Marron, J. S. Angle-based joint and individual variation explained. *J. Multivar. Anal.* **166**, 241–265 (2018).

72. Lock, E. F., Hoadley, K. A., Marron, J. S. & Nobel, A. B. JOINT AND INDIVIDUAL VARIATION EXPLAINED (JIVE) FOR INTEGRATED ANALYSIS OF MULTIPLE DATA TYPES. *Ann. Appl. Stat.* **7**, 523–542 (2013).

73. Skotare, T., Sjögren, R., Surowiec, I., Nilsson, D. & Trygg, J. Visualization of descriptive multiblock analysis. *J. Chemom.* **34**, e3071 (2020).

74. Fan, X., Wang, Q. & Wang, Y. Non-destructive in situ Raman spectroscopic investigation of corrosion products on the bronze dagger-axes from Yujiaba site in Chongqing, China. *Archaeol. Anthropol. Sci.* **12**, 90 (2020).

75. Bouchard, M. & Smith, D. C. Catalogue of 45 reference Raman spectra of minerals concerning research in art history or archaeology, especially on corroded metals and coloured glass. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **59**, 2247–2266 (2003).

76. Martens, W., Frost, R. L., Kloprogge, J. T. & Williams, P. A. Raman spectroscopic study of the basic copper sulphates—implications for copper corrosion and 'bronze disease'. *J. Raman Spectrosc.* **34**, 145–151 (2003).

77. Valadas, S. *et al.* On the Use of the Unusual Green Pigment Brochantite (Cu_4 (SO_4)$(OH)_6$) in the 16th-Century Portuguese-Flemish Paintings Attributed to The Master Frei Carlos Workshop. *Microsc. Microanal.* **21**, 518–525 (2015).

78. Želinská, J., Kopecká, I., Svobodová, E., Milovská, S. & Hurai, V. Stratigraphic EM-EDS, XRF, Raman and FT-IR analysis of multilayer paintings from the Main Altar of the St. James Church in Levoča (Slovakia). *J. Cult. Herit.* **33**, 90–99 (2018).

79. Bresson, S., Bormann, D. & Khelifa, B. Raman studies of the C–H stretching modes in various cholesteryl alkanoates. *Vib. Spectrosc.* **16**, 163–171 (1998).

80. Czamara, K. *et al.* Raman spectroscopy of lipids: a review. *J. Raman Spectrosc.* **46**, 4–20 (2015).

81. Ciupiński, Ł. *et al.* Comparative Laser Spectroscopy Diagnostics for Ancient Metallic Artefacts Exposed to Environmental Pollution. *Sensors* **10**, 4926–4949 (2010).

82. Balkanski, M., Nusimovici, M. A. & Reydellet, J. First order Raman spectrum of Cu2O. *Solid State Commun.* **7**, 815–818 (1969).

83. (PDF) Sol-Gel materials investigation by means of Raman Spectroscopy. *ResearchGate* https://www.researchgate.net/publication/229010847_Sol-Gel_materials_investigation_by_means_of_Raman_Spectroscopy.

84. Sassi, Z., Bureau, J. C. & Bakkali, A. Spectroscopic study of TMOS±TMSM±MMA gels Previously identi®cation of the networks inside the hybrid material.

85. Nakashima, S., Higashihira, M., Maeda, K. & Tanaka, H. Raman Scattering Characterization of Polytype in Silicon Carbide Ceramics: Comparison with X-ray Diffraction. *J. Am. Ceram. Soc.* **86**, 823–829 (2003).

86. Machev, P., O'Bannon, E. F., Bozhilov, K. N., Wang, Q. & Dobrzhinetskaya, L. Not all moissanites are created equal: New constraints on moissanite from metamorphic rocks of Bulgaria. *Earth Planet. Sci. Lett.* **498**, 387–396 (2018).

87. Nakashima, S. & Harima, H. Raman Investigation of SiC Polytypes. *Phys. Status Solidi A* **162**, 39–64 (1997).

88. Yuan, R. *et al.* Raman spectroscopy analysis of disordered and amorphous carbon materials: A review of empirical correlations. *Carbon* **238**, 120214 (2025).

89. (PDF) Raman Spectroscopy of Amorphous Carbon. *ResearchGate* https://doi.org/10.1557/PROC-498-37 doi:10.1557/PROC-498-37.

90. Loppnow, G. R. *et al.* UV Raman Spectroscopy of Hydrocarbons. *Philos. Trans. Math. Phys. Eng. Sci.* **362**, 2461–2476 (2004).

91. Kerr, L. T., Byrne, H. J. & Hennelly, B. M. Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen. *Anal. Methods* **7**, 5041–5052 (2015).

92. Campbell, E. *et al.* Methanol-to-Olefins Studied by UV Raman Spectroscopy as Compared to Visible Wavelength: Capitalization on Resonance Enhancement. *J. Phys. Chem. Lett.* **15**, 6826–6834 (2024).

93. Lambert, D., Muehlethaler, C., Esseiva, P. & Massonnet, G. Combining spectroscopic data in the forensic analysis of paint: Application of a multiblock technique as chemometric tool. *Forensic Sci. Int.* **263**, 39–47 (2016).

94. Pagnin, L., Calvini, R., Sterflinger, K. & Izzo, F. C. Data Fusion Approach to Simultaneously Evaluate the Degradation Process Caused by Ozone and Humidity on Modern Paint Materials. *Polymers* **14**, 1787 (2022).

95. Li, Z. *et al.* Multimodal HSI Combined with Multiblock Data Fusion: A New Tool for the Study of Time-Dependent Alteration Processes in Dyed Textiles. *Anal. Chem.* **96**, 19939–19946 (2024).

96. Ahmed, Z., Kinjol, F. J. & Ananya, I. J. Comparative Analysis of Six Programming Languages Based on Readability, Writability, and Reliability. in *2021 24th International Conference on Computer and Information Technology (ICCIT)* 1–6 (2021). doi:10.1109/ICCIT54785.2021.9689813.

97. Ozgur, C., Colliau, T., Rogers, G. & Hughes, Z. MatLab vs. Python vs. R. *J. Data Sci.* **15**, 355–372 (2022).

98. Sudhaka, K. Python vs. R Programming Language. *Int. J. Manag. IT Eng.* **8**, 70–79 (2018).

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain
Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/epl**