Intercomparison of XRF Core Scanning Results From Seven Labs and Approaches to Practical Calibration

Ann G. Dunlea1, Richard W. Murray1, Ryuji Tada2, Carlos A. Alvarez-Zarikian1, Chloe H. Anderson4, Adrian Gilli4, Liviu Giosan1, Thomas Gorgas1,6, Rick Hennekam7, Tomohisa Irino8, Masafumi Murayama9, Larry C. Peterson10, Gert-Jan Reichart7, Arisa Seki11,12,13, Hongbo Zheng12,13, and Martin Ziegler5,14, Carlos A. Alvarez

1Woods Hole Oceanographic Institution, Woods Hole, MA, USA, 2Department of Earth and Planetary Science, University of Tokyo, Tokyo, Japan, 3International Ocean Discovery Program, Texas A&M University, College Station, TX, USA, 4MARUM Center for Marine Environmental Sciences, Bremen, Germany, 5Department of Earth Sciences, ETH Zurich, Zurich, Switzerland, 6Now at Fugro, Germany Land GmbH, Berlin, Germany, 7Department of Ocean Systems, Royal Netherlands Institute for Sea Research, Utrecht University, Texel, The Netherlands, 8Department of Environmental Earth Science, Hokkaido University, Sapporo, Japan, 9Center for Advanced Marine Core Research, Kochi University, Kochi, Japan, 10Department of Marine Geosciences, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, USA, 11Faculty of Science, Shinshu University, Matsumoto, Japan, 12College of Science, Nanjing Normal University, Nanjing, China, 13Now at Research Center for Earth System Science, Yunnan University, Kunming, China, 14Now at Earth Science Department, Utrecht University, Utrecht, Netherlands

Abstract

X-ray fluorescence (XRF) scanning of marine sediment has the potential to yield near-continuous and high-resolution records of elemental abundances, which are often interpreted as proxies for paleoceanographic processes over different time scales. However, many other variables also affect scanning XRF measurements and convolute the quantitative calibrations of element abundances and comparisons of data from different labs. Extensive interlab comparisons of XRF scanning results and calibrations are essential to resolve ambiguities and to understand the best way to interpret the data produced. For this study, we sent a set of seven marine sediment sections (1.5 m each) to be scanned by seven XRF facilities around the world to compare the outcomes amidst a myriad of factors influencing the results. Results of raw element counts per second (cps) were different between labs, but element ratios were more comparable. Four of the labs also scanned a set of homogenized sediment pellets with compositions determined by inductively coupled plasma-optical emission spectrometry (ICP-OES) and ICP-mass spectrometry (MS) to convert the raw XRF element cps to concentrations in two ways: a linear calibration and a log-ratio calibration. Although both calibration curves are well fit, the results show that the log-ratio calibrated data are significantly more comparable between labs than the linearly calibrated data. Smaller-scale (higher-resolution) features are often not reproducible between the different scans and should be interpreted with caution. Along with guidance on practical calibrations, our study recommends best practices to increase the quality of information that can be derived from scanning XRF to benefit the field of paleoceanography.

1. Introduction

The invention of nondestructive X-ray fluorescence (XRF) spectrometric scanning of split sediment cores revolutionized the field of paleoceanography (e.g., Croudace et al., 2006; Jansen et al., 1998; Rothwell & Croudace, 2015). High-energy X-rays irradiate elements in the sediment core, exciting electrons that release surplus energy in a characteristic spectral pattern. Detector systems measure this fluorescing energy spectrum and mathematical deconvolutions of the spectrum estimate the individual element intensities, which are estimates of each element’s abundance in the sediment. XRF scanning can measure many major and minor elements simultaneously while preserving the sediment intact (e.g., Croudace et al., 2006; Haschke, 2006; Haschke et al., 2002; Kido et al., 2006; Koshikawa et al., 2003; Richter et al., 2006; Wien et al., 2005). With measurements able to be acquired as close as every 100 μm of core, the near-continuous element records can then be related to past changes in Earth and ocean processes that affect the composition of the sediment (e.g., Hennekam et al., 2019; Peterson et al., 2000; Seki et al., 2019; Ziegler et al., 2013).
In addition to element abundances, however, the physical properties of the core and instrument-specific settings affect the element counts measured by scanning XRF as well. Instrument-specific settings including the type of X-ray excitation source, applied voltage of tube power supply, aging of the X-ray source, measurement time, and the spectrum deconvolution and background estimates also affect the magnitude of intensities measured for each element (e.g., Jarvis et al., 2015). Unlike conventional analyses of dried and homogenized samples, element intensities obtained by XRF scanning are also influenced by variable moisture content, diverse grain sizes, and irregularities of the core surface (Böning et al., 2007; Chen et al., 2016; Ge et al., 2004; Hennekam & de Lange, 2012; Kido et al., 2006; MacLachlan et al., 2015; Ramsey et al., 1995; Tjallingii et al., 2007). Although these effects are known, scanning XRF data are commonly reported as element intensities or counts per second (cps) that are interpreted as records indicating the variations of the element abundance (e.g., Croudace et al., 2006; Richter et al., 2006; Rothwell & Rack, 2006). The easiest and most convenient way to minimize some of these effects (e.g., heterogeneities or physical properties) is by using element ratios, rather than raw intensities (e.g., Calvert & Pedersen, 2007; Croudace et al., 2006; Jansen et al., 1998; Richter et al., 2006; Rothwell et al., 2006; Ziegler et al., 2013). However, if quantitative results were available, XRF scans could be compared between labs and with other geological compositions, making scanning XRF an even more powerful tool.

Previous studies have calibrated scanning XRF results to quantitative concentrations using a variety of approaches and with varying degrees of success. Some studies calibrated XRF scanning results on an element-by-element basis assuming a linear relationship with known compositions for the calibration curve (e.g., Hunt et al., 2015; Wirth et al., 2013). Sometimes a correction for moisture effects is performed (e.g., Tjallingii et al., 2007) or a normalized-median weight percent calculated (e.g., Chen et al., 2016; Lyle et al., 2012; Lyle & Backman, 2013; Shackford et al., 2014) prior to a linear calibration with known compositions. Combining principles of compositional data analysis with XRF-spectrometry theory, other studies recommend using ratios or log-ratios during calibration (e.g., Bloemsma & Weltje, 2015; Tachikawa et al., 2011; Weltje et al., 2015; Weltje & Tjallingii, 2008; Ziegler et al., 2013). While calibration points are typically taken from the core being scanned to matrix match the sediment and standards, the log-ratio technique claims to provide a model that can calibrate XRF scans with sediment of a similar lithology, not necessarily from the core being scanned. However, the extent to which these approaches will allow records to be compared and/or combined between labs has not been well constrained.

In this study, we sent seven 1.5 m long sections of marine sediment cores to seven XRF core scanning laboratories around the world. We compare the XRF scanning results as raw element cps, ratios (cps/cps), and calibrated concentration ratios. Using Fe and Ca as representative elements, we provide guidance as to how to perform a practical calibration of XRF scanning results. The results of the study inform our recommendations for the best practices to calibrate and compare XRF scanning data between labs and other independently determined compositional data.

2. Methods
2.1. Interlab Comparison Experiment Design
2.1.1. XRF Scanning of Core Sediment Sections
The seven 1.5 m core sections of marine sediment core used in this study were drilled during Integrated Ocean Drilling Program (IODP) Expedition 346 at Site U1424 in the Japan Basin (40°11.39′N, 138°13.90′E, 2,808 m water depth) and Site U1425 on the Yamato Rise (39°29.43′N, 134°26.55′E, 1,909 m water depth; Tada et al., 2015). The sections selected (Hole U1424C Sections 1H4, 2H5, and 3H5 and Hole U1425C Sections 2H3, 2H4, and 2H6 and 3H6) cover a range of sediment compositions. U-channels extracted continuous marine sediment approximately 1 cm thick from the center of each split core section. Over the course of four years (2014 to 2017), the set of seven u-channels was shipped around the world to seven labs with XRF scanners including, in no particular order, the Kochi Core Center at Kochi University (Japan), IODP Core Repository at Texas A&M University (USA), Nanjing Normal University (China), Rosenstiel School of Marine and Atmospheric Science at the University of Miami (USA), ETH Zurich (Switzerland), Woods Hole Oceanographic Institution (USA), and the Royal Netherlands Institute of Sea Research (The Netherlands).

The lead investigators overseeing the XRF scanning in these labs were shipboard participants on IODP Expedition 346 (Tada et al., 2015) and are among the authors of this paper. The only instructions to each
Scanners, two ITRAX Core Scanners, and one Geotek Core Scanning Logger) with different X-ray tubes. The labs used various types and different generations of XRF scanning instruments (four Avaatech Core Scanners, two ITRAX Core Scanners, and one Geotek Core Scanning Logger) with different X-ray sources (Rhodium, Molybdenum; Table 1). Three of the labs scanned the cores at two or three excitation energies (e.g., 10, 30, and 50 kV). Each lab reported a different suite of elements, but all included Ca, Fe, K, Mn, Si, Sr, Ti, and Zr. Six labs also reported Al, Br, Cr, Cu, Ni, Pb, Rb, S, and Zn and five labs reported Ba, Cl, Ga, Mo, V, and Y (supporting information Table S1).

As the u-channels were shipped around the world for 4 years, the sediment dried, cracked, and shifted slightly within the container even though they were wrapped with protective films. Especially toward the end of the study, shrinkage cracks were abundant and the surface of the sediment was irregular and lower than the edge of the u-channel from being scraped so many times. One lab scanned sections from different holes at the same sites (U1424A, U1425B, and U1425D) that were stratigraphically aligned with the sections in this study (Irino et al., 2018). Therefore, the effects imposed by core-aging and interhole variability are included in this study in addition to differences in instrument type and settings.

Our goal was to assess the extent to which the records can be compared among all of these variables. We do not comment on the quality of a given scan and intentionally do not identify which lab generated which scans, as many of the variables (e.g., X-ray tube aging, detector aging, and/or dehydration of the core material) could affect any instrument at various times or be exacerbated during the transit between labs. The complete data set of XRF scanning results is provided in supporting information Data Set S1 and six major elements are plotted in supporting information Figures S1–S7. For the main text, we focus on the XRF scanning results of Fe and Ca from Section U1425C-2H3 (or U1425D-2H2 63.5–150 cm and U1425D-2H3 0–63.5 cm). Every lab scanned Fe and Ca as they are two of the most common elements used in paleoceanographic interpretations (Croudace et al., 2006; Rothwell & Croudace, 2015). Each XRF scan should reflect the sediment composition, which did not change among the labs, and this study seeks to find the best way to compare that common signal.

2.1.2. Sediment Pellets for Quantitative Calibration

In addition to the seven core sediment sections, we freeze-dried and powdered four discrete samples that were pressed into disk-shaped pellets about 2 cm in diameter from nearby Core MD01-2407 on the Oki Ridge (37°04′N, 134°42′E, 932 m water depth; Kido et al., 2007). The four samples have a similar matrix to the seven sediment sections scanned in this study. The four samples from Core MD01-2407 covered a range of sediment types (calcareous, siliceous, light, and dark colored; Kido et al., 2007) that span the dynamic range of the Fe and Ca element cps scanned for this study. A set of four pellets were sent to four of the seven labs (one ITRAX and three Avaatech) involved in the study to be scanned using the same instrument parameters they used on the sediment sections. Three labs used the same instrument and parameters used for the sediment section, but the fourth lab replaced the X-ray tube in between scanning the pellets and sediment sections (supporting information Data Set S2).

We also analyzed the four sample powders for major and trace element concentrations using inductively coupled plasma (ICP)-optical emission spectrometry (OES) and ICP-mass spectrometry (MS) in the Analytical Geochemistry Facilities at Boston University (e.g., Anderson et al., 2018, 2019; Dunlea et al., 2015). Together, these two instruments analyzed concentrations of Al, Ba, Ca, Cr, Cu, Fe, K, Mn, Mo, Ni, Pb, Rb, Si, Sr, Ti, V, Y, and Zr in the four homogenized sediment powders (supporting information Data Set S2). For major elements, samples were digested using a flux fusion method and analyzed on ICP-OES (e.g., Murray et al., 2000). For trace elements, a heated acid cocktail was used to dissolve 20 ± 2 mg of sample powders under clean-lab conditions (HNO₃, HCl, and HF, with later additions of HNO₃ and H₂O₂ after samples were dried down), which were analyzed with ICP-MS (e.g., Dunlea et al., 2015). Triplicate digests of an in-house matrix-matched standard were analyzed along with the samples to quantify precision (3 times the standard deviation/average), which was consistently ~2%. Standard reference material BHVO-2 was analyzed as an unknown during the same run, and the results were accurate within precision. Given the
accuracy and precision of the ICP measurements, we take the ICP measurements as the “known” values of the pellets, which we used to calibrate the XRF scanning results.

2.2. Data Processing

2.2.1. Compiling Data for Comparison

The shifts in the sediment within the u-channel during transport, along with the depth where scanning was started, caused minor misalignments in the downcore profiles of the same section generated in different labs. No major offsets were observed and the overall patterns and magnitudes of the results should nonetheless be comparable. The depths of the record from Sections U1425D-2H2 (63.5–150 cm) and U1425D-2H3 (0–63.5 cm) were stratigraphically aligned with Section U1425C-2H3 using the very precise offsets reported in Irino et al. (2018). The u-channels included 0.1 m of padding and structural support at the top of the section, so an extra 0.1 m was added to the stratigraphically aligned sections from Hole U1425D for a more precise match.

The element intensities were normalized by the measurement time (in seconds) to determine cps, which are commonly used in paleoceanographic and paleoenvironmental studies. We divided the data from Avaatech instruments by “Real Time” and used “Live Time” to normalize data from ITRAX and Geotek instruments. This is an imperfect normalization because ITRAX instruments keep the “Live Time” for counting for each measurement consistent, while Avaatech instruments account for “Dead Time” to keep the “Real Time” consistent for each measurement. We report and use the cps rather than total counts because some labs measured the calibration pellets for a different amount of time than the sediment sections. Because the count times are the same for each measurement within a scan, the adjustment from counts to cps scales all of the data but does not impose additional variability within the data. To better correct for differences in count time on the ITRAX, element cps could be normalized to the sum of several common elements in each sample (e.g., Jansen et al., 1998; Weltje & Tjallingii, 2008).

Background corrections may have been included when modeling the XRF energy spectra, but no additional background correction was performed after the element cps values were reported. Scanning XRF records that showed outliers with anomalously low values for nearly every element corresponded with anomalously high values of Ar or S. Possibly caused by cracks, air in between the sample and detector, or an instrument glitch, the outliers were identified by high cps of Ar or S and eliminated from the record.

2.2.2. Quantitative Calibrations

For this study, we applied two calculations to convert the cps measured by scanning XRF to quantitative element concentrations (e.g., ppm or wt%). First, we tested a simple linear calibration of scanning XRF data on an element-by-element basis. A simple least squares linear regression was fit to the element cps measured by scanning XRF and the element concentrations measured by ICP for the four homogenized pellet samples. The XRF cps were regarded as the dependent variable and the ICP concentrations as the independent variable, so the best fit line minimized the deviations between the XRF data and the predicted values while the ICP values were fixed. A Pearson correlation coefficient ($r^2$) assessed the goodness of fit and the equation of the linear fit (e.g., $\text{Fe}_{\text{XRF(cps)}} = m \cdot (\text{Fe}\text{ICP(wt%)})) + b$, where $m$ and $b$ are the slope and y intercept, respectively) was used to convert the scanning XRF cps to concentrations.
The second calculation used a log-ratio calibration that followed the approach described in Weltje and Tjallingii (2008) and Weltje et al. (2015). For each XRF measurement, we divided an element (e.g., FeXRF(cps)) by a common denominator (e.g., CaXRF(cps)) that was measured simultaneously in the XRF spectra and took the natural log of that ratio [e.g., ln (FeXRF(cps)/CaXRF(cps))]. The natural logarithm (ln) of the ratio for the ICP data was also calculated using the same denominator (e.g., CaICP(wt%)) from the same sample powder [e.g., ln (FeICP(wt%)/CaICP(wt%))].

Figure 1. Results of the scanning XRF interlab comparison. Each row has results from a different lab that XRF scanned the marine sediment from Section U1425C-2H3. XRF#7 scanned the stratigraphically aligned segment of sediment from Hole U1425D. The x axis of every plot is depth in meters from the top of the section. The columns are (a) Fe (cps × 10^3), (b) Ca (cps × 10^3), and (c) Fe (cps)/Ca (cps) of the original data reported from each lab, not yet calibrated to be concentrations. The XRFs are numbered in the chronological order in which they were scanned.
We selected Ca as the denominator to follow the results of Weltje and Tjallingii (2008), although the best denominators may be different for separate studies. The log-ratios of XRF and ICP were linearly regressed against each other for each element [e.g., ln (FeXRF(cps)/CaXRF(cps)) = m*ln (FeICP(wt%)/CaICP(wt%)) + b, where m and b are the slope and y intercept, respectively]. An r² was calculated to estimate the goodness of fit. This regression deviates from the method of Weltje and Tjallingii (2008) that relates the log-ratios with a singular value decomposition to include error in both the x and y axes during the fit. Given the well-constrained precision and accuracy of the ICP data, we assume that the uncertainty exclusively resides in the scanning XRF data. We used the equation of the regression of the log-ratios to convert the XRF log-ratios to quantitative log-ratios and finally to quantitative ratios [e^(log-ratio) = ratio].

At this stage, Weltje and Tjallingii (2008) calculated proportions of the seven major elements (Si, Al, Ti, Fe, Mn, Ca, and K) and Cl in their study sum to 100 wt%, in order to convert their ratios back to semiquantitative elemental concentrations. This approach does not provide fully quantified results as it only regards a subset of elements and does not include some major elements (Mg, Na, and P), organic matter, carbonates, and oxygen (e.g., MnO vs. Mn) that contribute to the 100 wt% of a total analysis of sediment composition. As is typical for XRF scanning results, our study also did not include all the major elements nor did we want to assume that the elements measured summed to a constant value, so we did not take this final step. A promising approach to constrain quantitative “absolute” element concentrations from scanning XRF data has been developed using a multivariate log-ratio calibration, which includes the unmeasured elements as part of the model (e.g., Hennekam et al., 2019; Weltje et al., 2015). However, the four standards used to calibrate data in this study are insufficient to effectively apply the multivariate technique. Thus, we left the data as quantitative ratios (units: weight%/weight%) rather than as putative “absolute” abundances.

3. Results
3.1. Direct Comparison: Raw cps and Ratios

The complete data set of the seven sections scanned by seven labs for multiple elements is reported in supporting information Data Set S1. For the main text, we focus on one representative sediment section (Section U1425C-2H3) and the elements Fe and Ca, which highlight some of the key complexities in comparing XRF scans from different labs. The absolute magnitudes of raw cps will vary between labs with different instrument settings, so in this section we focus mainly on the relative changes in each profile. Overall, most of the distinct variations are reproduced in every profile of raw cps of Fe in the XRF scans of Section U1425C-2H3 generated by the different labs (Figure 1a). However, the magnitudes of the local minima relative to the local maxima are not the same, such as the local minimum of Fe around 1.3 m below the top of the section (Figure 1a). Additionally, the Fe (cps) from XRF#1 increases from 0.5 to 1.4 m (Figure 1a), while other profiles remain flat (e.g., Figure 1a, XRF#3 and XRF#7) or decrease (Figure 1a, XRF#6). Similarly, for the Ca profiles (Figure 1b), the two main local maxima centered on ~0.75 and 1.3–1.5 m below the section top are reproduced in each profile, but their shape and relative magnitude vary among the different labs. The magnitude of the local maximum in Ca (cps) at 0.75 m is relatively lower (Figure 1b, XRF#1 and XRF#3), higher (Figure 1b, XRF#4, XRF#5, and XRF#6), or the same as (Figure 1b, XRF#2 and XRF#7) the local maximum at 1.3–1.5 m, depending on which profile is examined. In some of the profiles the local maxima in Ca (cps) at 0.75 m may be interpreted as dual peaks (e.g., Figure 1b, XRF#4, XRF#5, XRF#6, and XRF#7), whereas other profiles suggest it is one sustained local maxima (e.g., Figure 1b, XRF#1, XRF#2, and XRF#3). Similar to Fe, XRF#1 shows overall increasing Ca in the deeper half of the section (Figure 1b), while the other profiles remain low surrounding the two local maxima. The discrepancies between profiles cannot be ascribed to one particular factor (e.g., type of XRF, or if the section was scanned early or late in the shipping-around-the-world schedule).

The ratio of Fe (cps)/Ca (cps) is more comparable between labs than the raw cps. The increasing trend of XRF#1 from 0.5 to 1.4 m is flattened to be more similar to the profiles from the other labs (Figure 1c). The relative magnitudes of the two local minima of Fe (cps)/Ca (cps) are equal in the profiles from each lab (Figure 1c). Compared to the total cps of each element individually, the total magnitude of the Fe (cps)/Ca (cps) is also more consistent among the labs, except for XRF#2 and XRF#7, which are substantially higher than the other labs.
3.2. Calibrated Data

3.2.1. Linear Calibration

The linear regressions demonstrate a good fit between the ICP (wt%) data and the XRF (cps) data scanned by four labs for Fe ($r^2 \approx 0.85$ to 0.96; Figure 2a) and Ca ($r^2 = 0.99$; Figure 2b) in the four pellets. Only one pellet has a high concentration of Ca while the other three are relatively low, rendering it effectively a two-point calibration curve (Figure 2b). If the pellet with high Ca is excluded, the linear regression of the lower three calibration points yields poor fits with primarily negative Ca concentrations and thus is not considered an option. The XRF data calibrated with the linear regression produced Fe and Ca concentration profiles that are rescaled but otherwise identical in shape to the original, uncalibrated XRF data (cps) profiles (Figures 3a and 3b). The linearly calibrated Fe (wt%) and Ca (wt%) produce a ratio [Fe (wt%)/Ca (wt%)] that shows overall similar trends to the uncalibrated Fe (cps)/Ca (cps) (Figures 3c and 4a and 4b). However, the

Figure 2. The four-point calibration curves. Each row has results from a different lab that XRF scanned four pellets of homogenized sediment with compositions that span the dynamic range of Fe (cps) and Ca (cps) scanned by XRF in this study. The y axes of each plot are the measurements of the four pellets from scanning XRF. The x axes are the element concentrations of the four pellets measured by ICP. The columns are (a) the linear calibration of Fe, (b) the linear calibration of Ca with the inset plots zoomed in on the lower range of the axes, and (c) the calibration of the natural log of Fe/Ca. The $r^2$ values are reported within each plot.
linearly calibrated Fe (wt%)/Ca (wt%) increases in the deeper half of the profile for XRF#1, similar to the uncalibrated Fe (cps) and Ca (cps) profiles.

3.2.2. Log Ratio Calibration
The calibration curves of the log-ratios [ln (Fe/Ca)] of XRF and ICP data are well fit ($r^2 = 0.96$ to 0.98; Figure 2c). The scanning XRF element concentration ratios [Fe (wt%)/Ca (wt%)] that were calibrated with log-ratios are similar to the uncalibrated ratio plots [Fe (cps)/Ca (cps)] but are rescaled and the relative minima and maxima are of different magnitudes likely due to the use of natural logs in the calibration. The increase in Fe (wt%)/Ca (wt%) from 0.5 to 1.4 m in the profile from XRF#1 observed in the linearly calibrated results is not present in the log-ratio calibrated data. The absolute magnitudes of the log-ratio calibrated profiles of Fe (wt%)/Ca (wt%) are also more comparable than they were as uncalibrated Fe (cps)/Ca (cps) and linearly calibrated Fe (wt%)/Ca (wt%) (Figures 3d and 4a–4c). The absolute and relative variability of smaller-scale features (e.g., within a 20 cm interval) are also more similar with the log-ratio calibration than with the linear calibration, but discrepancies are still observed (Figures 4d–4f).

4. Discussion
4.1. “Best-Practice” for Comparing and Calibrating Scanning XRF Data
Aside from element abundances, many factors depending on the physical properties of the core or instrument settings affect the magnitude of cps measured by scanning XRF (e.g., Böning et al., 2007; Chen et al., 2016; Ge et al., 2004; Henneke & de Lange, 2012; Jarvis et al., 2015; Kido et al., 2006; MacLachlan et al., 2015; Ramsey et al., 1995; Tjallingii et al., 2007). Our study includes these “real-world” factors that are inherent in XRF scanning practices for paleoceanographic work. For example, some investigators may be scanning a core that has been recently acquired, while other investigators (perhaps stimulated by the initial work) may scan the same core years later and/or with a different type of scanner (e.g., ITRAX or AvaaTech). Additionally, individual users may have unique habits to prepare core, and human error can introduce an extra source of uncertainty. A combination of these core and instrument effects likely explains the range of magnitude in cps for the element abundances as well as the differences in smaller features measured by the various labs. For example, XRF#1 shows an overall trend unique from the other profiles: low Fe (cps) and Ca (cps) at 0.5 m that increases until 1.4 m below the top of the section (Figure 1a, XRF#1, and Figure 1b, XRF#1). This background trend could be due to instrumental drift or inhomogeneous moisture content that did not affect the other instruments that were measuring a drier sediment section with different settings. Alternatively, improper mounting of the core may cause the sensor to reach the surface incorrectly and impose a downhole trend. The XRF scans are numbered chronologically in the order they were scanned, so one would expect that the core surface would be of the highest core quality at the beginning of the study (XRF#1) when the cores were freshest. If the XRF results are to represent the true element concentrations, the analyses need to account for these other variables affecting the measurements.

Ratios are a simple yet powerful way to reduce the variability imposed by the inhomogeneity of the sediment core, device-specific settings, and instrument aging or drift (e.g., Calvert & Pedersen, 2007; Lüwemark et al., 2011). Ratios cancel out the effects that equally influence the element intensities of a measurement such as imperfect sample preparation and count times. The results of this study show that the overall shape of the Fe/Ca profiles are more similar than Fe (cps) and Ca (cps) as single element profiles among labs (Figures 1 and 4a). The increase in Ca (cps) and Fe (cps) unique to XRF#1 measurements is no longer observed in the Fe (cps)/Ca (cps). The overall shapes and relative magnitude of local minima in the profiles of Fe (cps)/Ca (cps) matches between all the labs better than the raw Fe (cps) and Ca (cps) (Figure 1).

Despite a significant improvement, the ratios from different labs are still only relative composition variations and there are significant differences in the values of Fe (cps)/Ca (cps) (Figure 4a). The Fe (cps)/Ca (cps) of XRF#2 and XRF#7 is nearly an order of magnitude higher than the scans from other labs, likely due to the instrument type and the X-ray sources used (supporting information Table S1). For better unity between the profiles from different labs, a calibration of the XRF data is required.

When a linear calibration is applied to XRF data on an element-by-element basis, the calibration curve is subjected to the uncertainties imposed by inhomogeneities in the core and device-specific effects that modify the element cps. The linear calibration of XRF data in this study rescales the element cps but does not account for the nonelement effects nor improve the match between the shape of the profiles from each lab.
The profiles of Fe (wt%)/Ca (wt%) calculated from the linearly calibrated Fe (wt%) and Ca (wt%) show some of the same inconsistencies as the uncalibrated Fe (cps) and Ca (cps) (Figures 1 and 4b). The linearly calibrated Fe (wt%)/Ca (wt%) profile from XRF#1 shows the unique increasing background trend observed in the uncalibrated cps data but is removed in the uncalibrated Fe (cps)/Ca (cps). Thus, the linear calibration on an element-by-element basis carried over some of the effects of the drift or inhomogeneity that affected only XRF#1.

Using element ratios in the calibration curve minimizes many of the inhomogeneity and instrument-specific differences prior to the calibration (Weltje & Tjallingii, 2008). Furthermore, log-ratios account for the inherent nonlinear relationship between element concentrations and instrument response imposed by matrix effects (Weltje & Tjallingii, 2008). The log-ratio calibration of XRF data in this study made the profiles from the different labs very close both in absolute values and relative magnitude of variability (Figures 3d and 4c). Smaller-scale variations are more comparable too but still show differences that should be considered when interpreting the calibrated data (Figures 4d–4f). The amount of noise varies between the records too but can only be quantified if replicate measurements are available (Löwemark et al., 2019; Weltje et al., 2015). Overall, the similarities between profiles suggest that a log-ratio calibration curve is significantly better than linear calibrations to directly compare data between labs (Figure 4).

Our study demonstrates that external standards may be effective for calibrating and comparing XRF scans from different labs when a log-ratio calibration is used, as Weltje and Tjallingii (2008) and Weltje et al. (2015) proposed. Most previous studies constructed calibration curves with discrete samples taken directly from the core being scanned to ensure a perfect matrix match. We used marine sediment samples from a nearby site to construct a calibration curve, and these samples are thus a close, yet not perfect, matrix match. Regardless,
the results of our study (Figure 4) indicate an effective calibration may be achieved with external standards that have a matrix similar, but not identical, to the core being scanned. A common set of calibration standards routinely scanned within each lab would help the XRF core scanning results between labs be more comparable.

In summary, we recommend reporting XRF scanning results as ratios (cps/cps) rather than raw cps to minimize device-specific and core effects. In line with Weltje and Tjallingii (2008) and Weltje et al. (2015), log-ratio calibrations are significantly better than linear calibrations in making XRF scanning results comparable between labs. To enhance future comparability among labs, we recommend that each scanning XRF lab is provided a common set of many external standards that present options to match the matrix being scanned as closely as possible. While our simple four-point calibration appears sufficient to demonstrate the effectiveness of this approach, we recommend using additional points for an even more accurate calibration curve (e.g., >30 calibration points; Weltje et al., 2015). All together these recommendations would help advance the field toward making XRF scanning an even more powerful and reliable tool.
4.2. Paleoceanographic Implications

If the uncalibrated Fe (cps) and Ca (cps) or the linearly calibrated Fe (wt%) and Ca (wt%) were used for paleoceanographic research, the interpretations likely would have been different for the profiles generated by each lab. For example, consider if Fe was being used as a proxy for dust abundance and Ca for marine calcareous biogenic deposition, as they commonly are in paleoceanographic research. The profile from XRF#1 may be interpreted as having dust contributions decrease from 1 to 0.5 m in the section where it begins increasing again to 0 m (Figure 1a, XRF#1). Simultaneously, calcareous biogenic productivity could be interpreted as decreasing up the section with two episodes of high productivity at 1.4 and 0.75 m depths (Figure 1b, XRF#1). In contrast, the uncalibrated XRF#6 profile could be interpreted as having increasing dust from deeper (1.25 m) to shallower (0.5 m) in the section, and four bursts of calcareous productivity at 0.7, 0.8, 1.1, and 1.4 m during the otherwise constant calcareous input over the length of the core (Figure 1a, XRF#6, and Figure 1b, XRF#6). Thus, while the major features of the section are more or less reproduced by each of the XRF scanners even when comparing only raw cps, there are enough inconsistencies to cause major differences in the paleoceanographic narratives if minor features are interpreted.

Additionally, any element concentration (i.e., abundance) is effectively a ratio between that element (e.g., Fe) and the sum of all elements that make up the total bulk sediment. Changes in the other components diluting that element (e.g., organic material or opal) will affect the element concentration (wt% or ppm), the classic “closed sum” problem. Thus, it is important to interpret accumulation rates rather than concentrations, simultaneously examine all components of the sediment for a holistic perspective or interpret element ratios.

In this study, a more consistent paleoceanographic interpretation would be drawn from the uncalibrated Fe (cps)/Ca (cps) and log-ratio calibrated Fe (wt%)/Ca (wt%) profiles. For example, if the Fe/Ca were interpreted as the ratio of dust to calcareous productivity, each plot would be broadly interpreted as having two periods of high productivity relative to dust and/or low dust input relative to productivity. Semiquantitative results cannot determine the total magnitude of variations in the record, which in some cases could lead to poorly constrained interpretations. For example, what appears to be two periods of high productivity in the record may actually reflect subtle changes in aluminosilicate provenance. The quantitative Fe (wt%)/Ca (wt%) calibrated with log-ratios in Section U1425-2H3 ranges from ~0.5 to 6 (Figure 4c), which overlaps the range of Fe (wt%)/Ca (wt%) for common aluminosilicate compositions such as Chinese Loess (CL; 0.5; Jahn et al., 2001), average upper continental crust (UCC; 1.4; Rudnick & Gao, 2014), and post-Archean average Australian shale (PAAS; 4.9; Taylor & McLennan, 1985). The direct, quantitative comparison with geological reference material suggests that the range of Fe/Ca measured by XRF scanning in this section is narrow and may be more sensitive to changes in aluminosilicate provenance rather than productivity. Additionally, subseaﬂoor redox conditions may alter sediment chemistry after deposition (e.g., Fe remobilization) and postrecovery changes to the sediment cores (e.g., pyrite oxidation) may affect the Fe/Ca measured. Thus, quantitative log-ratio calibrated data can be extremely useful in constraining the actual magnitude of variability, interpretation of proxies, and the paleoceanographic information derived from a marine sediment core.

5. Conclusions

Seven international labs XRF scanned seven marine sediment sections using the element menus and instrument parameters they typically use for their independent paleoceanographic research. Despite the diversity of instrument types and settings as well as the aging of the sediment core during transit, the raw scanning XRF (cps) results from each lab broadly reproduced the relative major variations in the element profiles of the sediment section assessed. However, important smaller-scale features and overall increasing/decreasing trends varied among the labs, likely due to inhomogeneity in grain size, moisture content, irregularities in the split core surface, instrumental drift, or other unquantifiable parameters. A linear calibration of XRF Fe (cps) and Ca (cps) with pellet samples with compositions well constrained by ICP-OES and ICP-MS rescaled the scanning XRF profiles to be more comparable between labs, but the inconsistencies in the smaller features and overall trends persisted.
Taking a ratio of the uncalibrated XRF results \([\text{Fe (cps)/Ca (cps)}]\) successfully minimized many of the effects imposed by variable count time, irregular core surfaces, and instrumental drift and made the relative variations and overall trends of the XRF results from different labs more comparable. A log-ratio calibration of the XRF data, as recommended by Weltje and Tjallingii (2008) and Weltje et al. (2015), significantly improved the match between the profiles from different labs in terms of absolute and relative magnitude of variability. These results were achieved for a few elements with only a four-point calibration curve composed of matrix-similar external samples that spanned the dynamic range of the section scanned. Additional points on the calibration curve may improve the results and possibly allow a multivariate log-ratio calibration to convert ratios back into “absolute” concentrations (Weltje et al., 2015). The results of our study suggest that the log-ratio calibration is an effective way to convert XRF scans to quantitative values and unite results from different labs. Even with this approach, however, many smaller scale features are nonreproducible between the labs and should be interpreted with caution. Quantitative XRF data allow direct comparison with other geological composition data, help constrain interpretation of proxies, and quantify the absolute magnitude of variability within the sediment core thereby greatly enhancing the subsequent paleoceanographic research.

### Data Availability Statement

The data generated during and used in this study are included in tables in the supporting information and are publically available on the PANGAEA® Data Publisher (https://doi.org/10.1594/PANGAEA.922043). Any additional data may be obtained from A. G. D. (email: adunlea@whoi.edu).

### Acknowledgments

We would like to thank R. Tjallingii and an anonymous reviewer for their thoughtful comments on the manuscript. We would like to thank A. Karasuda, C. C. McKinley, and U. Brüpbacher for their help with the preparation of the pellets and XRF scanning. This research used samples and data provided by the Integrated Ocean Drilling Program (IODP). Funding for this research was provided by the U.S. National Science Foundation to R. W. M. (Grant 1130531). USSSP postcruise support was provided to Expedition 346 shipboard participants A. G. D., R. W. M., L. G., C. A. Z., and L. P. Portions of this material are based upon work supported while R. W. M. was serving at the National Science Foundation.

### References

Anderson, C. H., Murray, R. W., Dunlea, A. G., Giosan, L., Kinsley, C. W., McGee, D., & Tada, R. (2019). Aeolian delivery to Ulleung Basin, Korea (Japan Sea), during development of the East Asian Monsoon through the last 12 Ma. *Geological Magazine*, 157(3), 806–817. https://doi.org/10.1017/S001675681900013X

Bloomstra, M. R., & Weltje, G. J. (2015). Reduced-rank approximations to spectroscopic and compositional data: A universal framework based on log-ratios and counting statistics. *Chemometrics and Intelligent Laboratory Systems*, 142(C), 206–218. https://doi.org/10.1016/j.chemolab.2015.02.001

Böning, P., Baud, E., & Rose, J. (2007). Toward direct, micron-scale XRF elemental maps and quantitative profiles of wet marine sediments. *Geochemistry, Geophysics, Geosystems*, 8, Q05004. https://doi.org/10.1029/2006GC001480

Calvert, S. E., & Pedersen, T. F. (2007). Elemental proxies for palaeoclimatic and palaeoceanographic variability in marine sediments: Interpretation and application, In *Developments in Marine Geology* (Vol. 1, pp. 567–644). Developments in Marine Geology.

Chen, Q., Kiesel, C., Govin, A., Liu, Z., & Xie, X. (2016). Correction of interstitial water changes in calibration methods applied to XRF core scanning major elements in long sediment cores: Case study from the South China Sea. *Geochemistry, Geophysics, Geosystems*, 17, 1925–1934. https://doi.org/10.1002/2016GC006320

Croudace, I. W., Rindby, A., & Rothwell, R. G. (2006). ITRAX: Description and evaluation of a new multifunction X-ray core scanner. In R. G. Rothwell (Ed.), *New Techniques in Sediment Core Analysis*, Special Publications (Vol. 267, pp. 51–63). London: The Geological Society of London.

Dunlea, A. G., Murray, R. W., Sauvage, J., Spivack, A. J., Harris, R. N., & D’Hondt, S. (2015). Dust, volcanic ash, and the evolution of the South Pacific Gyre through the Cenozoic. *Palaeoceanography*, 30, 1078–1099. https://doi.org/10.1002/2015PA002829

Ge, L., Lai, W., & Lín, Y. (2004). Influence of adsorption and correction for moisture in rocks, soils and sediments on in situ XRF analysis. *X-Ray Spectrometry*, 34(1), 28–34. https://doi.org/10.1002/xrs.782

Haschke, M. (2006). The Eagle III BKA system, a novel sediment core X-ray fluorescence analyser with very high spatial resolution. In R. G. Rothwell (Ed.), *New Techniques in Sediment Core Analysis*, Special Publications (Vol. 267, pp. 31–37). London: The Geological Society of London.

Haschke, M., Scholz, W., Theis, U., Nicols, J., Scruggs, B., & Herzeg, L. (2002). Description of a new Micro-XRF spectrometer. *Journal de Physique IV*, 12, 83–95. https://doi.org/10.1051/jpi:20020216

Hennemann, R., & de Lange, G. (2012). X-ray fluorescence core scanning of wet marine sediments: Methods to improve quality and reproducibility of high-resolution palaeoenvironmental records. *Limnology and Oceanography: Methods*, 10(12), 991–1003. https://doi.org/10.4319/lom.2012.10.991

Hennemann, R., Sweere, T., Tjallingii, R., de Lange, G. J., & Reichart, G.-J. (2019). Trace metal analysis of sediment cores using a novel X-ray fluorescence scanning method. *Quaternary International*, 514, 55–67. https://doi.org/10.1016/j.quaint.2018.10.018

Hunt, J. E., Croudace, I. W., & MacLachlan, S. E. (2015). Use of calibrated ITRAX XRF data in determining turbidite geochemistry and provenance in Agadir Basin, northwest African passive margin. In *Micro-XRF Studies of Sediment Cores* (Vol. 17, pp. 127–146). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9649-5_4

Irino, T., Tada, R., Ikahara, K., Sagawa, T., Karasuda, A., Kurokawa, S., et al. (2018). Construction of perfectly continuous records of physical properties for dark-light sediment sequences collected from the Japan Sea during Integrated Ocean Drilling Program Expedition 346 and their potential utilities as palaeoceanographic studies. *Progress in Earth and Planetary Science*, 5(23). https://doi.org/10.1186/s40645-018-0176-7
