Chemical Geology 623 (2023) 121395

Contents lists available at ScienceDirect

Chemical Geology

journal homepage: www.elsevier.com/locate/chemgeo

Portable x-ray fluorescence calibrations: Workflow and guidelines for optimizing the analysis of geological samples

A.C. Da Silva [a,*,1], A. Triantafyllou [a,b,1], N. Delmelle [a]

[a] Department of Geology, University of Liège, Sart Tilman, Liège, Belgium
[b] Lyon Geology Laboratory - Earth, Planets and Environment (LGL-TPE), Université de Lyon, Université Lyon 1, ENS de Lyon, CNRS, UMR 5276, F-69622 Villeurbanne, France

ARTICLE INFO

Editor: Karen Johannesson

Keywords:
Portable X-Ray Fluorescence
Calibration model
pXRF
Limit of Quantification
Sedimentary and igneous rocks geochemistry

ABSTRACT

The use of portable Energy Dispersive X-Ray Fluorescence (pXRF) spectrometers has exploded within the last decade for e.g., environmental, geological, and archaeological field research. Producing reliable geochemical data using a pXRF is a challenging process that requires the development of dedicated and matrix-matched empirical calibrations. Although a large field of applications, several research papers are still published using unquantified pXRF data, using manufacturers' built-in calibrations unsuitable for targeted rocks, or interpreting chemical concentrations without checking the limit of quantification of the pXRF spectrometer and its calibration model. In this paper, we present a general workflow for building an empirical calibration model for a pXRF spectrometer, from choosing reference materials to validating calibration's efficiency and quantifying its limit of quantification. We also compare the geochemical results on a sedimentary and an igneous rocks suite obtained from thirty-two calibration models that were built using different software available, including a simple homemade spreadsheet, the EasyCal calibration software, a free and open-source calibration app named CloudCal, and the built-in GeoMining and GeoExplorer calibrations provided by Bruker manufacturer on the Tracer 5 pXRF. This decoupled approach allows (i) to explore the pros and cons of each calibration model and determine which empirical calibration model brings the most accurate chemical concentration for a sedimentary and an igneous rock suite, (ii) to weigh the impact of each calibration parameter (type of regression line, spectral interferences, normalization, influence coefficients, multiple conditions of excitation and acquisition) on the accuracy of geochemical concentrations on rock samples and (iii) to bring innovative tools to assess the limits of quantification of a newly built calibration model.

1. Introduction

Portable Energy Dispersive X-Ray Fluorescence (pXRF) instruments have become common tools for chemical analysis of major, minor and some trace elements in the last few years. These spectrometers are based on a versatile, relatively cheap, and stable technique, allowing fast measurement of a large number of samples in the field or the laboratory. It is also a non-destructive technique, preventing tedious sample preparation and widening its field of application to fragile or archaeological materials. With the development of lightweight and tough models of pXRF apparatus, its use exploded for environmental, geological and archaeological field research, including, for example, the monitoring of polluted sediments (Frahm et al., 2016), paleoenvironmental, paleoclimatic and sedimentary geochemistry (Sinnesael et al., 2018; Da Silva et al., 2019), stratigraphy and cyclostratigraphy (De Vleeschouwer et al., 2017; Ibáñez-Insa et al., 2017; Da Silva et al., 2020; Saker-Clark et al., 2019), sedimentology and sequence stratigraphy (Stewart and Mauk, 2017; Turner et al., 2016), paleontology (Cohen et al., 2017), provenance studies in geoarchaeology (Kasztovszky et al., 2018; Triantafyllou et al., 2021), but also base and precious metal mineral exploration (Ahmed et al., 2019, Cao et al., 2018) and quantitative analysis of igneous rocks and magmatic processes (Ryan et al., 2017; McNulty et al., 2018; Martí et al., 2020). Although this large field of applications, too many research papers are still published using unquantified pXRF data or using manufacturer's built-in calibrations that are not suitable for the targeted rocks, or (over-) interpreting

* Corresponding author.
E-mail address: ac.dasilva@uliege.be (A.C. Da Silva).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.chemgeo.2023.121395
Received 30 August 2022; Received in revised form 23 December 2022; Accepted 16 February 2023
Available online 20 February 2023

chemical concentrations without checking the limits of quantification of the instrument and its calibration model (see Shackley, 2012 and references therein). Amongst misleading interpretations, this prevents the direct comparison of pXRF chemical analyses gathered from different devices and research groups, which goes against the new era of building large databases to perform statistical geochemistry.

Several research groups showed that concentrations of minor and major chemical elements could be accurately measured using portable XRF spectrometers (Hall et al., 2014; Quye-Sawyer et al., 2015; Mejía-Piña et al., 2016; de Winter et al., 2017; Arenas-Islas et al., 2019). However, this requires building a dedicated calibration model that converts raw peak intensity to the mass concentration of an element in a given material. pXRF raw results are typically provided as counts per second (cps), for which the intensity is directly related to the influence of a chemical element into the sample (Weltje et al., 2015). However, the raw intensity of a given element peak also depends on several additional factors, including e.g., the voltage of the X-ray source, its current, the measurement duration, the type of spectrum deconvolution, the instrumental blank, the nature of the analysed material (Gregory et al., 2019; Röhl and Abrams, 2000; Tjallingii et al., 2007; Lemière, 2018). All these parameters affect the correlation between peak intensity and the actual chemical concentration of a given element and must be carefully corrected in the calibration scheme. There are two main approaches to calibrating raw pXRF signal: the "Fundamental Parameter" approach (Rousseau, 1984) is based on theoretical mathematical models for the excitation of atoms and relaxation processes, the attenuation and excitation of radiation within the sample by certain matrix atoms. This type of calibration is employed by numerous pXRF manufacturers and does not require external reference materials. It is an excellent approach for detecting elements and semi-quantifying an unknown material. However, several research papers showed that it does not exploit full instrument performance, it lacks reproducibility between different pXRF instruments, and does not provide optimal accuracy for several elements (e.g., Ti, Mn, Fe, Zn and Sr; Sheppard et al., 2011; Goodale et al., 20122; Hall et al., 2014). The second approach consists of building an empirical calibration that relies on analysing materials of known concentration (called reference materials) and on complex multilinear regression that accounts for inter-elemental effects. This approach is more challenging to implement but provides more accurate and reliable quantitative measurements (Hall et al., 2014; Mejía-Piña et al., 2016; Rouillon and Taylor, 2016; Sinnesael et al., 2018; Steiner et al., 2017; Al-Musawi and Kaczmarek, 2020) for a given type of material. However, only a few papers provide guidelines to build such sophisticated empirical calibrations and details on the analytical protocol to acquire precise, accurate and reproducible geochemical data (e.g., Adams et al., 2020; Ravansari et al., 2020).

In this perspective, the aim of this paper is: (1) to provide a workflow and guidelines for building an empirical calibration model with a pXRF spectrometer, from the selection of geochemical reference materials, the optimization of calibration corrections, to the statistical validation of a calibration model; (2) to compare the accuracy of chemical results provided by thirty-two empirical calibrations models built using different software packages available on the market. This work allows determining which empirical calibration model brings the most accurate chemical concentration for a sedimentary and an igneous rock suite; (3) to weigh the impact of each calibration parameter on the accuracy of geochemical concentrations on rock samples, and (4) to bring innovative tools to assess the limits of quantification of a newly built calibration model.

2. Workflow and methods for building a pXRF empirical calibration

This section presents a general workflow and the basic terminology for building an empirical calibration model using a pXRF device. An empirical calibration relies on reference materials (materials of known concentration) used to calculate element concentrations of an analyte via a complex multilinear regression that accounts for inter-elemental effects (Johnson et al., 2021). This multi-step process involves the choice of reference materials, the optimization of the excitation source (energy and current), acquisition conditions (time) and the correction of the raw pXRF signal leading to chemical quantification. The pXRF spectrometer used in this study is a Bruker Tracer 5 g from Liège University (Belgium). It is equipped with an X-Ray Rhodium (Rh) source with 4 W power, a maximum current of 200 μA and a voltage of 50 kV coupled to a 20 mm^2 Silicon Drift Detector (SDD) with a 1 μm graphene window and a spectral resolution under 140 eV at 250000 cps (on the Mn Kα spectral line). All measurements were conducted in dry air conditions (i.e. normal conditions without specific gas flow) and using an 8 mm collimator. We aim to quantify the following elements: MgO, Al_2O_3, SiO_2, P_2O_5, K_2O, CaO, TiO_2, MnO, Fe_2O_3, Ni, Cu, Zn, Y, Rb, Sr and Zr, that are commonly used as identification and petrogenetic proxies in igneous and sedimentary petrology. Chemical elements are expressed as oxides compounds (in wt%) or pure elements (in ppm), as they are generally used in geoscientific research papers.

2.1. Building an empirical calibration starts with the choice of reference materials

Building a calibration model in spectroscopy starts with measuring a set of reference materials of known chemical concentration. The latter should ideally consist of matrix-match analytical materials to reduce interferences of other constituents from the sample matrix (Gallhofer and Lottermoser, 2018; Lemière, 2018) and should cover the analyte concentration ranges of interest. It can be synthetic, as several Bruker built-in calibrations are developed using pure silica matrix reference materials, but it can be unsuitable for other types of geological matrices, such as carbonate lithologies. Alternatively, reference materials can be made of well-characterized natural rock powders with a chemical matrix as close as possible to the targeted samples (e.g., Kenna et al., 2011; Rouillon and Taylor, 2016; Young et al., 2016). The ideal approach is to use geochemical reference materials (GRMs), which meet rigorous homogeneity standards and are quantified by multiple reference international laboratories (Govindaraju, 1998). Some research groups also develop in-house reference material (synthetic or natural rocks), which can better agree with the studied samples. The chemical composition of these in-house reference materials requires cross-validation between multiple laboratories using reliable analytical techniques (e.g., Lezzerini et al., 2014).

In this paper, we use a set of GRMs made of powdered natural rocks to build the different calibration models in section 3 (GRMs names, their reference concentration, and associated reference paper can be found in Table 1 in the supplementary material). First, press-pellets of GRMs fine powders were prepared, mixing approximately 5 g of rock powder with 2 mL of methyl methacrylate resin (Elvacite®) as a binding agent. The mixture was then pressed (200 kN/cm^2) for about 60 s. For the sedimentary rock calibration, seventeen GRMs were used, including carbonates, dolomites, sandstones, siltstones, shales, and cherts. Twelve GRMs were selected to build the calibration models for the igneous rock calibration, including basalt, andesite, diorite, granodiorite, and granite.

2.2. Optimizing the excitation and acquisition parameters

The excitation and acquisition conditions include the tube's voltage and current (expressed in kV and μA, respectively) and the analysis duration (in seconds). The size of the beam collimator can be adapted depending on the desired spot size. A fixed large 8 mm collimator was used for this study to cover the largest analysis area on the sample. Most modern pXRF spectrometers are equipped with Rh anode X-ray sources covering 6 up to 50 kV (with a maximum 4 W output). Low voltage conditions (such as 15 kV) enhance the detection of light elements (atomic number $Z < 30$; i.e. Mg, Al, Si, P, K, Ti, Mn, Fe, Cu, Zn), while

higher voltage conditions (30 up to 50 kV) allow the detection of heavier elements (Z > 30; Beckhoff et al., 2006). Light elements are harder to quantify through X-Ray fluorescence than heavier elements because they tend to reabsorb the energy of the emitted photon at their outer electron, leading to weak signal-to-noise ratios (Kikongi et al., 2017). All the fundamentals of portable XRF spectrometry and the physical principle between excitation photons, electrons shells and fluorescence and energy dispersive analysis are beyond the scope of this contribution but can be found in several recent papers and books (see e.g., Knight et al., 2021; Drake and MacDonald, 2023). Most modern pXRF spectrometers allow combining multiple excitation conditions of their X-ray source based on which chemical element can be quantified within its best sensitivity window. For example, the Tracer 5 g pXRF and its X-ray Rh source can produce up to three successive excitation conditions. Two of the built-in analysis modes - named GeoMining and GeoExplorer - includes for example, three successive excitation conditions (called phases): one at 30 kV used as an exploration phase to focus on transition elements between K and Ga, the second phase at 50 kV dedicated for heavy elements detection, between As and U and the third phase at 15 kV for the detection of light elements between Mg and Ca.

Increasing the acquisition duration enhances the peak-to-background ratio, hence improving the reproducibility of quantified results. de Winter et al. (2017) stated that the precision of the chemical analysis does not improve largely beyond 60 s of acquisition time using a Bruker Tracer IV pXRF spectrometer on carbonate rock samples. Similar results were found for the analysis of geological standards and obsidian archaeological materials (Newlander et al., 2015), showing that short acquisition times (40–60 s) were commonly as accurate as pXRF data obtained at longer acquisition times (~120 to 180 s). Therefore, acquisition times in this range (below 90 s for the total analysis) were used in this study (see Table 1). The same excitation conditions (voltage and current) should be used for the analysis of the reference material when building the empirical calibration than for the samples' analysis. Excitation conditions, the number of excitation conditions and acquisition time for each calibration built in this paper are summarized in section 2.4 and Table 1.

2.3. Correction parameters to build an empirical calibration model

In pXRF spectroscopy, building an empirical calibration sensu stricto consists of establishing a multilinear regression model that allows the conversion from raw peak intensity to the mass concentration of an element using a set of reference materials. pXRF raw results are typically provided as counts per second, for which the measured net intensity of an element energy peak (also called spectral line) is directly related to its influence into the sample (Fig. 1a). The calibration step can be a real challenge as several sources of error have to be managed and corrected to efficiently quantify the chemical composition of samples (Beckhoff et al., 2006). There is a large and exhaustive literature on analytical protocols and different correction methods, which are exhaustively summarized in e.g., Hall et al., 2014; Ravansari et al., 2020 and references therein. Therefore, we succinctly summarize in this section the main correction parameters and associated terminology used in this paper.

2.3.1. Influence coefficients and matrix correction

The correction for the chemical elements present in addition to the analyte is referred to as "matrix correction" or "correction for the matrix effect." The challenge is that the measured intensity of a characteristic X-ray of an element can be strongly influenced by the presence of other chemical elements in the sample matrix (Lachance, 1993). The latter can scatter, enhance (i.e. secondary excitation) or absorb the incoming radiation from the X-ray source and/or the emitted fluorescent X-ray on its path to the pXRF detector. The correction of the matrix effect has been the focus of much attention in the last decades, for which several correction algorithms have been developed (see e.g., Beckhoff et al., 2006; Molchanova et al., 2011; Drake and MacDonald, 2023 for a review). The most commonly used these days are variants of the Lachance and Traill (1966) and Lucas-Tooth and Pyne (1964) algorithms (see Rousseau, 2006, for a didactic review of them). Both equations use a mathematical means called influence coefficients (generally noted as 'alphas'). They consist of numerical coefficients that correct for the effect of each matrix element on the analyte concentration.

2.3.2. Spectral interferences

Energy dispersive detectors used in pXRF spectrometers are likely affected by potential interferences from overlapping spectral peaks. Such spectral interferences can be associated with overlapping energy peaks between elements with close atomic numbers or with Compton scattered peaks and Si escape peaks (see Gallhofer and Lottermoser, 2018, for more details on each type of interference). This is commonly observed in pXRF spectra where the K_β peak of an element overlaps and

Table 1
Characteristics of the calibration models used in this paper, including their respective abbreviation, the excitation source conditions (voltage and current), the acquisition duration, and the list of applied corrections (e.g., matrix, spectral interference). Several energy/current/time settings are defined for calibration with multiple excitation conditions (also referred to as phases in the manuscript).

Calibration	Abbreviations	Excitation source		Acquisition time	Applied corrections
		Energy (kV)	Current (μA)	Time (s)	
Manufacturer GeoExploration	GeoExplorer-3phases	15/30/50	13/15/27	25/25/25	Unknown
	GeoExplorer-2phases	15/50	13/27	25/25	
Manufacturer GeoMining	GeoMining-3phases	15/30/50	13/15/27	25/25/25	Unknown
	GeoMining-2phases	15/50	13/27	25/25	
Simple Spreadsheet calibration	SSC-line-no0	40	20	75	Linear, no intercept through 0
	SSC-line-0	40	20	75	Linear, intercept through 0
	SSC-polynomial-no0	40	20	75	Polynomial, no intercept through 0
	SSC-polynomial-0	40	20	75	Polynomial, intercept through 0
CloudCal	CloudCal-40 kV	40	20	75	Normalization on Region of Interest (ROI), linear or polynomial, spectra interferences, matrix effect (Lucas-Tooth-Pyne).
	CloudCal-15 kV	15	13	40	
	CloudCal-50 kV	50	27	40	
EasyCal	Easycal-basic	40	20	75	Compton normalization, line or Polynomial, intercept at 0 or not
	Easycal-basic-matrix				Easycal-basic + Matrix effect corrections (Tertian, 1986)
	Easycal-basic-interferences				Easycal-basic + interference corrections
	Easycal-basic-matrix-interferences				Easycal-basic + Matrix effect and interference corrections
	Easycal-Final				"Final" combination

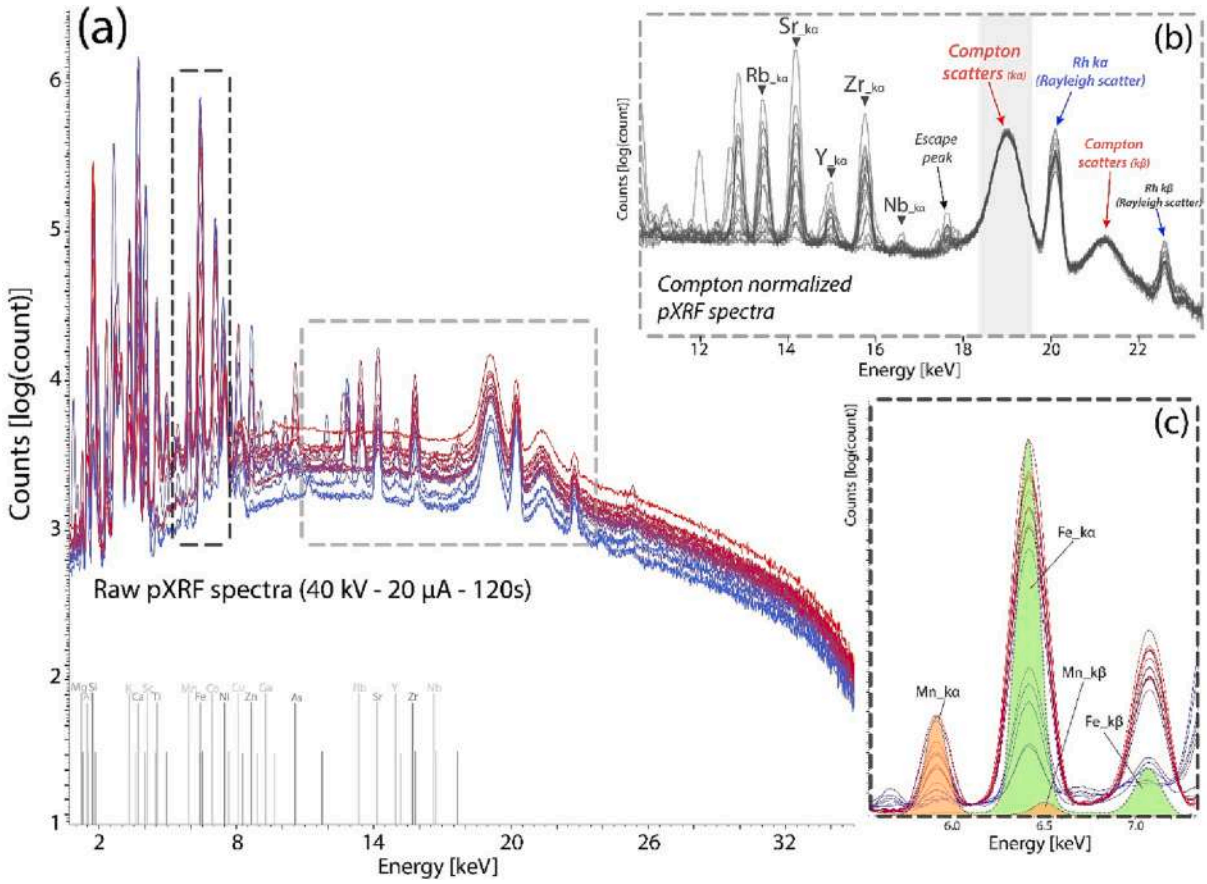

Fig. 1. Portable XRF spectra of all sedimentary geochemical reference materials used in this paper, acquired at 40 kV, 20 µA and during 120 s (see the list of GRM in supplementary material-Table 1). (a) Raw pXRF spectra with the position of k-alpha.

interferes with the K_α peak of the element with one - or two - higher atomic numbers (see Fig. 1c showing the Mn K_β and Fe K_α peak-on-peak overlap). Their combination can result in incorrect detection and hence, inaccurate concentration (over- or under-estimation) of an element in the analysed sample. Such interference can be minimized by adapting the energy integration range of the measured peak or selecting an alternative spectral line that is free from interference. For example, the Fe K_β spectral line could be used to quantify Fe concentration as far as there is no Co in the targeted sample (Fig. 1c). Another approach consists of applying a peak-fitting mathematical routine which can subtract the calculated interference from the measured peak (Conrey et al., 2014); the latter is generally proprietary to the software developer.

2.3.3. Normalization of the energy spectrum

The physical parameters of the pXRF instrument and the environmental conditions during the analysis are not considered in the above corrections. These analysis conditions can be structural due to small angle variations between the pXRF nose and the sample or to sample surfaces rough enough to cause air attenuation or secondary matrix attenuation. These environmental perturbations can also be superficial due to the sample surface's moisture, which can also cause attenuation of the incident and the fluorescent X-rays, instrumental fluctuations due to variations in ambient temperature or drift due to Rh source aging, etc. (Forster et al., 2011). For these reasons, it is critical to normalize the pXRF spectra to an internal standard to allow inter-spectrum comparison and make quantitative results reproducible. Several authors proposed to use the known scatters of the X-ray Rh source as a normalization reference and present in the spectrum of every sample. Rayleigh and Compton scatter peaks are commonly used (Johnson et al., 2021). They are caused by the elastic and inelastic scatter radiation from the Rh excitation source, respectively (Fig. 1b). The main Compton peak occurs between 18.5 and 19.5 keV at 40 kV excitation (Fig. 1b).

Further explanations on the physical background and mathematical equations behind these different correction parameters are beyond the scope of this paper but can be found in several seminal papers and book reviews, e.g., Beckhoff et al., 2006; Brouwer, 2018; Markowicz, 2008; Molchanova et al., 2011; Rousseau et al., 1996; Steiner et al., 2017; Weltje and Tjallingii, 2008; Conrey et al., 2014; Lee Drake et al., 2022 and references therein.

2.4. Different empirical calibration models and software

Several software exist on the market and allows to account for spectrum normalization (e.g., Compton, total energy), matrix effect and spectral interference corrections (based on several algorithms, e.g., Lachance and Traill, 1966; Lucas-Tooth and Price, 1961; Lucas-Tooth and Pyne, 1964), equation of the regression model (e.g., linear, quadratic, polynomial). We compare in this paper four types of calibration: basic calibration built using a spreadsheet (SSC), more sophisticated ones using the EasyCal software, the CloudCal app, and the built-in GeoMining and Geo-Explorer calibrations provided by the Bruker manufacturer. As mentioned before, the Tracer 5 g spectrometer includes a built-in calibration using multiple conditions of excitation (voltages and current).

2.4.1. Simple spreadsheet calibration (SSC)

This calibration is built through in-house spreadsheets using simple linear or quadratic regression models, including minimal signal corrections, applied on the igneous and sedimentary suites. GRMs

measurements were conducted using a single excitation condition, with an energy of 40 kV, during 75 s acquisition time (similar to the CloudCal and EasyCal calibrations below). We tested four different configurations, comparing a linear or quadratic regression curve with and without anchoring the regression line to its origin.

2.4.2. CloudCal calibration (CC)

This calibration is built using the free and open-source CloudCal package (Drake, 2018) on the R platform. We performed three types of calibration on both igneous and sedimentary suites. One calibration model was built using a unique excitation condition with an energy of 40 kV during 75 s acquisition time (similar to SSC and EasyCal calibrations). The second and third calibrations are based on two excitation conditions during the same measurement: one phase is set at 15 kV and the second at 50 kV, each during 40 s of acquisition time (similar configuration as the GeoExploration and GeoMining Calibrations provided by Bruker). This setting requires building one calibration for each excitation condition (i.e. each element is calibrated for each excitation phase), for which quantified results can be combined into a single calibration model. The CloudCal calibration tool allows the selection of the type of calibration line (linear or quadratic) and to apply different types of peak normalization. Corrections for the matrix effect can also be applied using influence factors (through the Lucas-Tooth and Pyne, 1964 correction, applied here), and spectral interferences can also be considered.

2.4.3. EasyCal calibration (EC)

The Spectra EDX software provided by Bruker includes the EasyCal calibration tool with the Tracer 5 g. The EasyCal software allows building a homemade empirical calibration using a single excitation condition. We applied an excitation voltage of 40 kV for 75 s on both igneous and sedimentary suites (similar to SSC and CloudCal). The EasyCal calibration tool allows one to choose the type of regression line (linear or quadratic), the origin anchoring of the regression line, the use of Compton peaks normalization and the correction of peak interferences. Corrections for the matrix effect can also be applied using influence factors (named alphas in the software); main interferences are highlighted and ordered by the importance of influence. This type of calibration can be uploaded in the instrument allowing direct quantifications in the field.

2.4.4. Built-in Bruker calibrations named GeoExplorer and GeoMining

These consist of factory calibrations on the Bruker Tracer 5 g. They include three successive acquisition conditions: phase 1 at 30 kV (exploration phase allowing to have a rough estimate and to focus on transition elements between K and Ga), phase 2 at 50 kV (phase dedicated for heavy elements detection, between As and U) and phase 3 at 15 kV (phase for the light elements between Mg and Ca). These calibrations were built using synthetic reference materials made of a single chemical element with a range of concentrations in a pure SiO_2 matrix (Bruker, 2017). The correction factors included in this type of calibration are proprietary and not detailed in the literature. We tested a configuration with three successive phases 1, 2 and 3, each during 25 s for GeoExplorer and GeoMining (75 s of acquisition time in total) and another configuration only using phases 2 and 3 (50 s of acquisition time) on the two rock suites.

2.5. Validating an empirical calibration model

A crucial step of this workflow is to test the performance of the calibration model. Each calibration model will be used in this study to quantify a set of validation samples. Those validation samples consist of pressed powder pellets of which we know the reference chemical composition and should be similar to the material that the pXRF user wants to study. The validation set should ideally consist of the same matrix and range of analyte concentration as the calibration set (Magnusson and Örnemark, 2014) and should be analysed using the same excitation and similar acquisition conditions as for the calibration. For the igneous rocks, the validation set is made of fifteen GRMs, different from those used for the calibration. Their reference compositions are provided in Govindaraju (1994) and can be found in Table 2 in the supplementary material. For the sedimentary rocks, the validation set includes seven GRMs (not used for the calibration) and eighteen in-house reference sedimentary samples. The latter consist of Devonian sedimentary rocks from Belgium (previously published in Da Silva et al., 2013), including limestones, marls, shales, and carbonated sandstones, with their reference compositions, are provided respectively in Govindaraju (1994) and via laboratory Wavelength Dispersive-XRF measurements (X-ray fluorescence spectrometer ARL PERFORM-X 4200 with results and analytical methods described in Da Silva et al., 2013).

2.6. Defining the limits of detection (LOD) and limits of quantification (LOQ)

The limit of detection (LOD) is defined as the lowest concentration of an analyte that can be reliably detected with a given analytical method. The limit of quantification (LOQ) of a given element is defined as the lowest concentration that can be quantitatively detected (or quantified) with a stated accuracy and precision for this element. The LOQ is generally higher than the LOD and will depend on several parameters, including the type and voltage of the excitation source, on the pXRF detector but mostly on the quality of the empirical calibration. Both LOD and LOQ can be calculated through two different approaches:

2.6.1. Via a statistical calculation of LOD and LOQ

One of the standard estimates of the LOD corresponds to the concentration equivalent to three times the standard counting errors of a set of measurements of the background intensity (LOD = 3 x standard deviation of a blank sample analysis). In this case, the background's signal and standard deviation are estimated through independent blank measurement. An ideal blank sample should be of a similar matrix to the geochemical reference materials and contain none of the targeted analytes (Belter et al., 2014). We selected a few GRMs used for our calibrations with none or very low concentrations of the targeted elements (Table 3 in the supplementary material). In order to estimate the statistical LOQ, blank samples should be analysed more than ten times according to the Analytical Methods Committe (1987). For each element and its calibration, we extracted the 30 measurements from its respective GRM and then calculated three times the standard deviation of these measurements to obtain the statistical LOD (named LOD_Stat hereafter). The statistical LOQ is calculated as ten times the standard deviation of the blank measurements (or 10/3 x LOD_Stat).

2.6.2. Via an empirical estimation of the LOD and the LOQ

In most of the literature, the statistical LOD or LOQ is considered as underestimated (e.g., Fiamegos and de la Calle Guntiñas, 2018; Rousseau, 2001). The empirical estimation of LOD and LOQ is a more indicative way to calculate the actual low range of detection and quantification. The empirical LOQ is calculated through the analysis of a set of validation samples by comparing the difference between the measured and the reference values for each chemical element with its concentration in the sample (i.e. the systematic error, noted %Err = 100 x [[pXRF concentration – reference concentration] / reference concentration] expressed in percent; see Fiamegos and de la Calle Guntiñas, 2018). This relationship is typically defined as a power function where very low amounts of a given element led to a drastic increase in %Err. In this scheme, defining the empirical LOQ consists of choosing the threshold %Err value at which the quantified concentration is considered reliable/sufficient. Defining this threshold is commonly performed by considering a fixed %Err threshold value applied to all chemical elements (chosen by the user and the analytical need of a study) or by considering different %Err ranges for different elements (see a detailed

method in Triantafyllou et al., 2021). In this paper, we provide a new technique to estimate the LOQ and we compare the LOQ obtained through different techniques in section 4.4.

3. Results: comparing calibration models for sedimentary and igneous rock suites

To determine the accuracy of calibration models, each of them is tested on a set of validation samples. The quantified results of each validation sample are compared with their reference compositions and displayed as validation lines (as shown in Figs. 3, 5, 7, 9). Based on these validation lines, three statistical parameters were used to estimate the performance of each calibration model: (i) The Percentage of Error (% Err, displayed in Figs. 2, 4, 6, 8) is calculated for each validation sample. It reflects the discrepancy between measured and reference concentration values (i.e. 100 x [[pXRF concentration – reference concentration] / reference concentration] expressed in percent). This %Err brings information on the accuracy of specific chemical elements at given concentration ranges. To compare different calibration models, we calculate and use the median %Err value (noted m%Err) for all quantified elements of each calibration model. The higher the m%Err value is, the larger the general discrepancy of the calibration model. This aspect will depend on each pXRF user to decide what type of %Err is acceptable for the applications they want to develop. (ii) One of the most indicative parameters when comparing chemical results from a pXRF with laboratory reference methods is the coefficient of determination (denoted R^2). It quantifies the ability of a model to predict an outcome in the linear regression setting ($R^2 = 1.00$ - the residual sum of squares / the total sum of squares; Pedregosa et al., 2011). We calculated R^2 values for each validation line, with an R^2-value close to 1.0, meaning that the reference values show a close match to the measured ones. iii) The slope value of the validation line is expected to be close to the 1.0 value (on the 1:1 trend line) and allows to visualize if your calibration is providing over or under-estimated concentrations (with respectively a slope lower than 1.0 or higher than 1.0). Validation lines are displayed in Figs. 3, 5, 7 and 9, CloudCal and EasyCal full calibration procedures are included in Supplementary Material Table 4 and the validation lines statistics can be found in the table 5 of the Supplementary Materials.

3.1. Simple Spreadsheet Calibrations

Concerning the Simple Spreadsheet Calibrations (SSC), a very limited number of corrections have been applied through four calibrations: the regression line calibration not forced to pass through the origin (SSC-line-no0), the regression line calibration going through the origin (SSC_line_0), the polynomial regression line not forced to pass through the origin (SSC-polynomial-no0) and the polynomial regression anchored to the origin (SSC-polynomial-0).

By comparing the different SSC calibrations, the calibrations without anchoring the regression line to origins (SSC-line-no0 and SSC-polynomial-no0) provide for most of the elements better results than the other calibration models (Figs. 2-3). For example, for the sedimentary suite by removing the anchoring of the calibration to the origin, the m%Err for Al_2O_3 decreases from 57.1 to 18.5% and R^2 goes from 0.89 to 0.97, for MgO m%Err goes from 126.2 to 20.9% and R^2 0.57 to 0.96 and for K_2O m%Err goes from 135.4 to 93.71% and R^2 from 0.71 to 0.75%. In total, on the 64 validation lines for the two rock suites, 42 were statistically enhanced or equivalent when the calibration line was not anchored to the origin (22 got worst when anchored). Interestingly, polynomial and linear regression models give similar quantified results (28 of the 64 R^2 of the validation lines were improved by changing from linear to polynomial calibration model, six did not change significantly, and 30 became worst). For the sedimentary rock suite for all SSC calibrations, the quantifications for P_2O_5 (R^2 lower than 0.3), Cu (R^2 lower than 0.64) and K_2O (R^2 around 0.75) are mostly inaccurate with m%Err values larger than 50%, while m%Err is lower than 50% for SiO_2, Rb, Ca, MgO, Sr, Al_2O_3 and Fe_2O_3 (R^2 higher than 0.9). For the igneous rock suite, calibrations provided for MgO, Ni, Cu, Zr, Rb, Sr and Y show m% Err higher than 50% and R^2 lower than 0.4, while m%Err are lower than 50% for TiO_2 (R^2 between 0.8 and 0.9), MnO (R^2 close to 0.7), CaO (R^2 close to 0.8) and Fe_2O_3 (R^2 higher than 0.9). The validation lines usually follow the 1:1 line for the sedimentary rock suite, meaning that there is

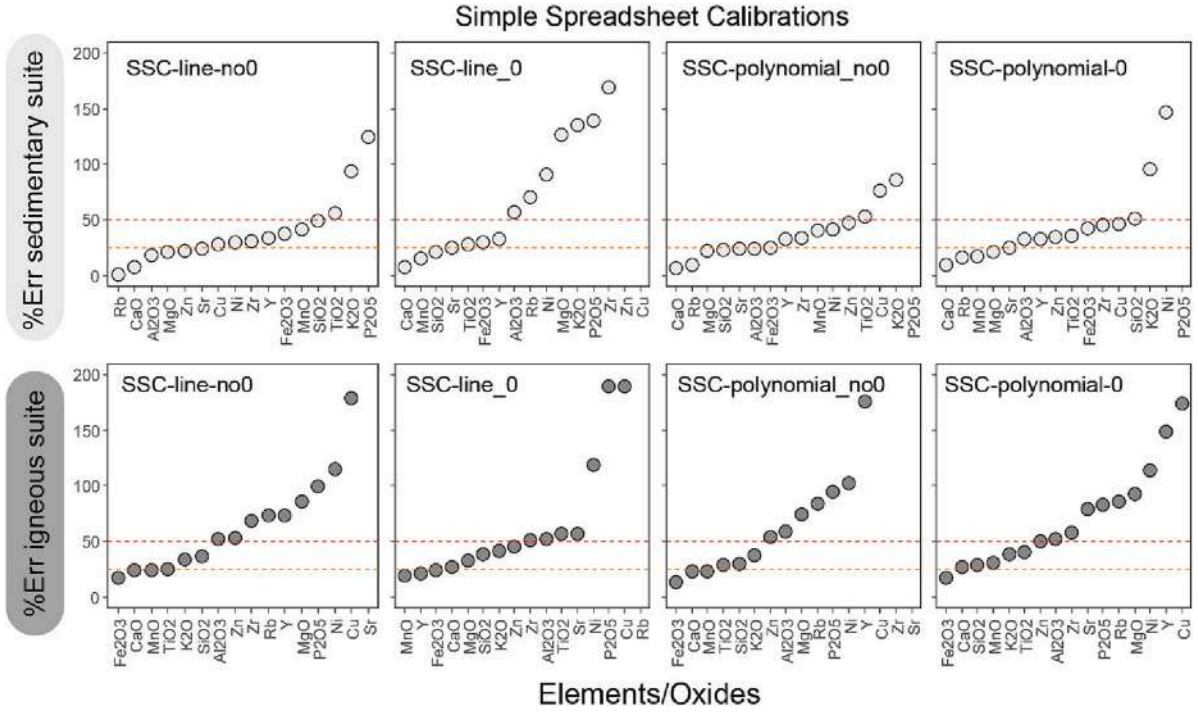

Fig. 2. Median percentage error (m%Err, see main text for definition) for the Simple Spreadsheet Calibration models for the sedimentary suite at the top and the igneous suite at the bottom. Chemical elements on the X axis are ordered by order of increasing m%Err. The horizontal dotted lines mark 25% and 50% of m%Err. The absence of a black dot means the calibration does not quantify the element. Abbreviations for each calibration model are provided in Table 1.

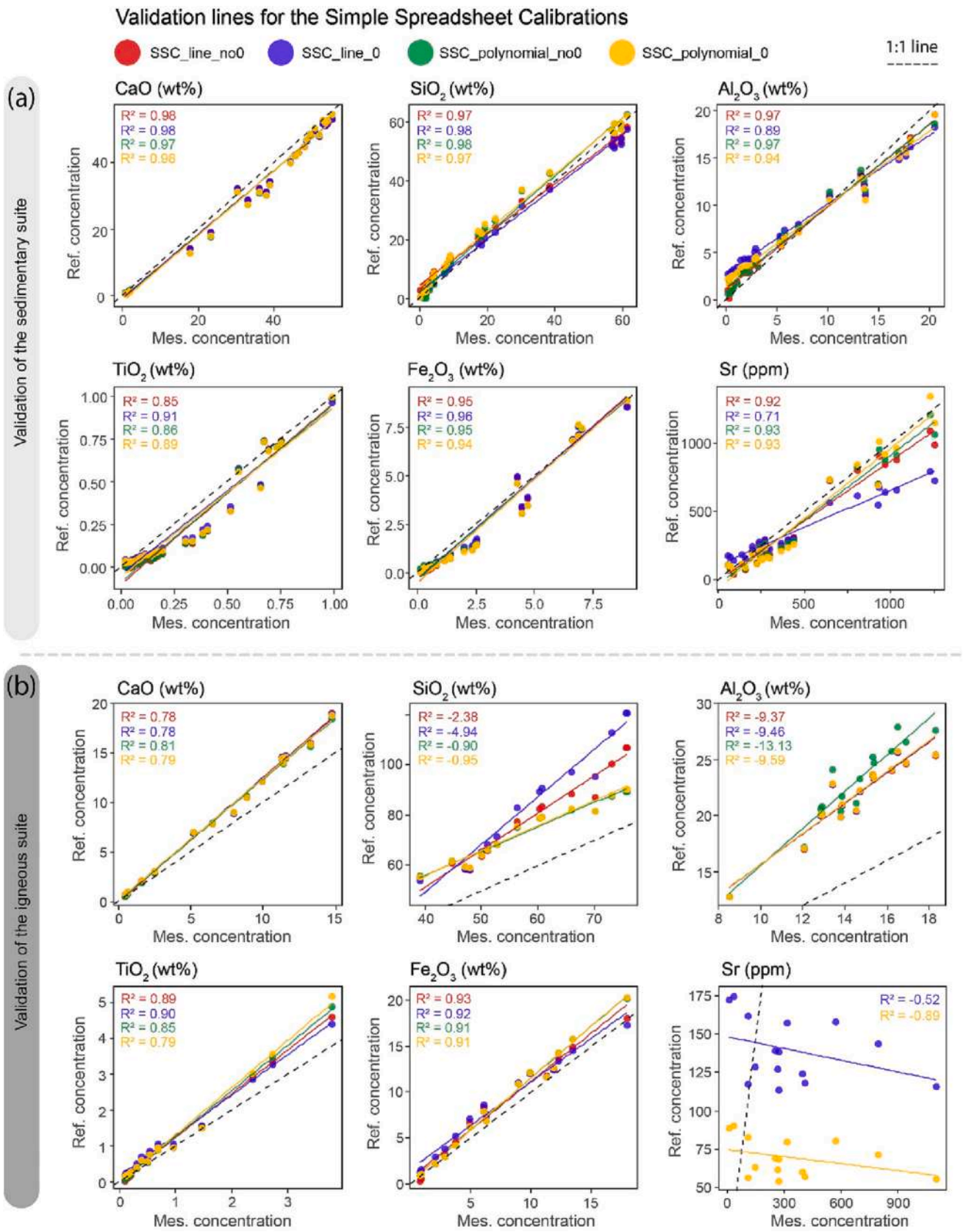

Fig. 3. Validation lines for the Simple Spreadsheet Calibration models of CaO, SiO_2, Al_2O_3, TiO_2, Fe_2O_3 and Sr for the sedimentary suite at the top and the igneous suite at the bottom. These plots compare reference concentrations ('Ref. conc.' on the X axis) with measured concentrations ('Mes. Conc.' on the Y axis) for each calibration model. A linear trend is fitted for each validation set, and its correlation coefficient (R) and its slope (S) are calculated and displayed on each plot. The ideal 1:1 line (black dashed) is also displayed for comparison. Abbreviations of calibration names are provided in Table 1. (Statistics of validation lines of all chemical elements/oxides are given in Suppl. Material-Table 5).

no significant over- or under-estimation of the expected composition, with the best validation lines classically displaying slope values ranging between 0.98 and 1.10 (except for Zr and Al, best validation lines slope respectively of 0.90 and 0.86). The R^2 value of the best validation lines for the sedimentary suite typically ranges between 0.97 and 0.99, except for TiO_2 (0.91), Sr (0.93) and Zr (0.89). For the igneous suite, there is typically an offset with the 1:1 line (i.e. slopes range between 0.85 and 0.95 and 1.1 and 1.5), and some validation lines are marked by a low to a negative coefficient of determination (Fig. 3 with $R^2 = -0.08$ for Rb, $R^2 = -0.89$ for Sr, $R^2 = -0.64$ for Zr).

3.2. CloudCal calibrations

Three CloudCal calibrations were built, including one with an excitation source condition at 40 kV and an analysis duration of 75 s (named CloudCal_40kV), and two calibrations operating respectively at 15 and 50 kV (named CloudCal_15kV and CloudCal_50kV) for 40 s each. These two last calibrations are then combined to provide better quantification of all targeted chemical elements. They will be considered separately hereafter to identify which transitional element is better analysed by which configuration.

Considering the lighter oxides (from Mg to Ti), the CloudCal_15kV and the CloudCal_40kV calibrations give a much better peak-to-noise ratio, hence better accuracy than the CloudCal_50kV calibration. For example, for the sedimentary rock suite, the m%Err of MgO with CloudCal_50kV is 136.9% (R^2 = -0.35), while it is 13.3% for CloudCal_40kV ($R^2 = 0.82$) and 13.5% for CloudCal_15kV ($R^2 = 0.93$). The m%Err for Al_2O_3 are for 50 kV of 30.6, for 40 kV of 10.4 and 15 kV of 14.5% (R^2 evolving from 0.69, 0.99, 0.98), for K_2O 163.9, 61.5 and 54.8% ($R^2 = 0.1$, 0.88 and 0.91) (Fig. 4). Concerning the transition elements, the three calibration models all give relatively similar results for MnO quantification with m%Err of 16.0% ($R^2 = 0.99$) for CloudCal_40kV, 21.0% ($R^2 = 0.99$) for CloudCal_15kV and 18.5% (0.98) for CloudCal_50kV; while for Fe and heavier elements, the CloudCal_50kV with the highest voltage offers better quality calibration, while CloudCal_15kV become less accurate (m%Err of Fe_2O_3 for CloudCal_40kV is 5.1% ($R^2 = 0.99$), CloudCal_50kV is 3.9% ($R^2 = 1.00$), and CloudCal_15kV is 31.5% ($R^2 = 0.93$); of Ni is respectively 34.0 ($R^2 = 0.22$), 27.3 ($R^2 = 0.63$) and 366.2% (R^2 = -194.15); while Rb, Sr, Y and Zr are not quantified with the CloudCal_15kV calibration). Similar results are observed for the igneous suite (Figs. 4-5).

For the sedimentary suite, the CloudCal_40kV calibration is marked by m%Err lower than 50% and R^2 higher than 0.8 for MgO, Al_2O_3, SiO_2, K_2O, CaO, TiO_2, MnO, Fe_2O_3, Zn and Sr (Figs. 4-5) while calibrations for Rb, Ni, Y P_2O_5, Cu and Zr are highly inaccurate (m%Err higher than 50% or R^2 lower than 0.4; see Figs. 4-5). For the combination of CloudCal_15kV (MgO to MnO) and CloudCal_50kV (Fe_2O_3 to Zr) calibrations, m%Err are lower than 50%, and R^2 is higher than 0.6 for MgO, Al_2O_3, SiO_2, K_2O, CaO, TiO_2, MnO, Fe_2O_3, Ni, Zn, Y, Rb, Sr, Zr, while m%Err is higher than 50% or R^2 is lower than 0.4 for Cu and P_2O_5 only (Figs. 4-5). The best calibrations for all elements of the sedimentary suite provide validation lines with an R^2 value of 0.99–1.00 and a slope of 1.0, except for TiO_2 ($R^2 = 0.96$, S = 1.03), Rb ($R^2 = 0.61$, S = 0.65) and Zr ($R^2 = 0.96$, S = 0.88; Fig. 5).

For the igneous suite, the CloudCal_40kV calibration provides m%Err lower than 50% and R^2 higher than 0.7 for all oxides/elements except for Al2O3 ($R^2 = 0.19$), Cu ($R^2 = 0.47$) and Ni ($R^2 = 0.31$). By combining the two phases' calibrations (15 kV and 50 kV), m%Err are lower than 50%, and R^2 is higher than 0.7 for all oxides/elements except Zr (R^2 = -0.69) (Fig. 4), but the accuracy of the calibrations is slightly lower than for the sedimentary suite. Indeed, most of the calibration lines provided by CloudCal for the igneous rock suites are in good agreement with the 1:1 line (S = 0.98–1), with high R^2 values (0.97–0.99), pointing to a relatively good accuracy (Fig. 5), except for SiO_2 ($R^2 = 0.93$, S = 0.92); Al_2O_3 ($R^2 = 0.77$; S = 1.1); MnO ($R^2 = 0.95$; S = 0.93); Rb ($R^2 = 0.87$; S = 0.94) and Zr ($R^2 = 0.87$; S = 1.1). All the results of the statistics of validation lines can be found in the supplementary materials (Table 5).

3.3. EasyCal calibrations

Five EasyCal calibrations were built by adding successive correction factors to the calibration model, going from a relatively simple calibration model to a more sophisticated one. This approach allows assessing the role of each correction factor and its impact on the accuracy of quantified results. The first EasyCal calibration model (named EasyCal_basic) is built using a polynomial regression line not anchored

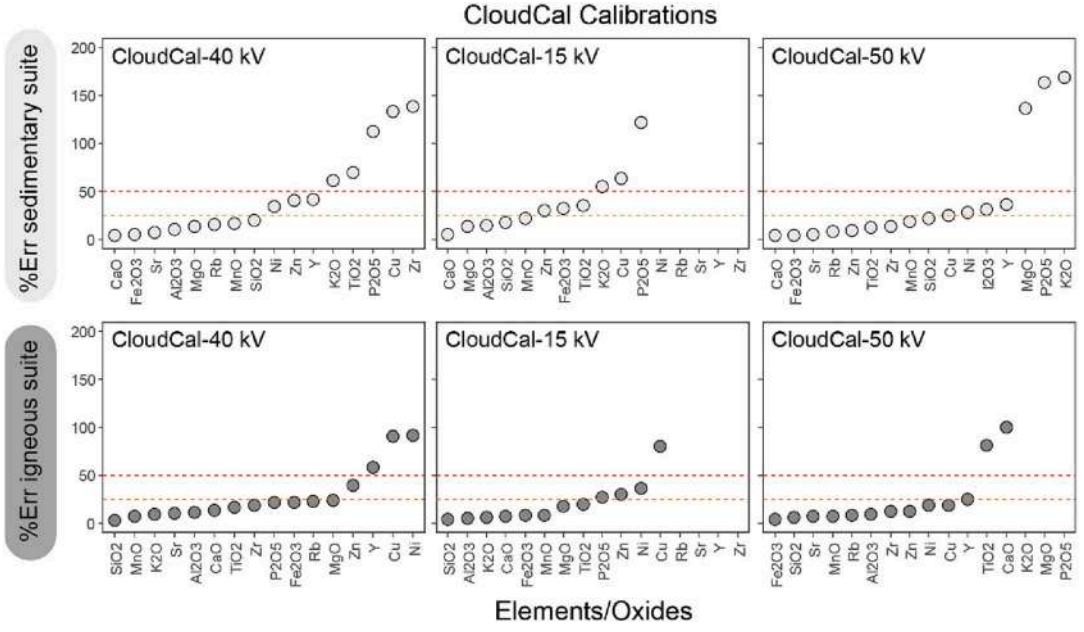

Fig. 4. Median percentage error (m%Err, see main text for definition) for the CloudCal calibration models for the sedimentary suite at the top and the igneous suite at the bottom. Chemical elements on the X axis are ordered by order of increasing m%Err. The horizontal dotted lines mark 25% and 50% of m%Err. The absence of a black dot means the calibration does not quantify the element. Abbreviations for each calibration model are provided in Table 1.

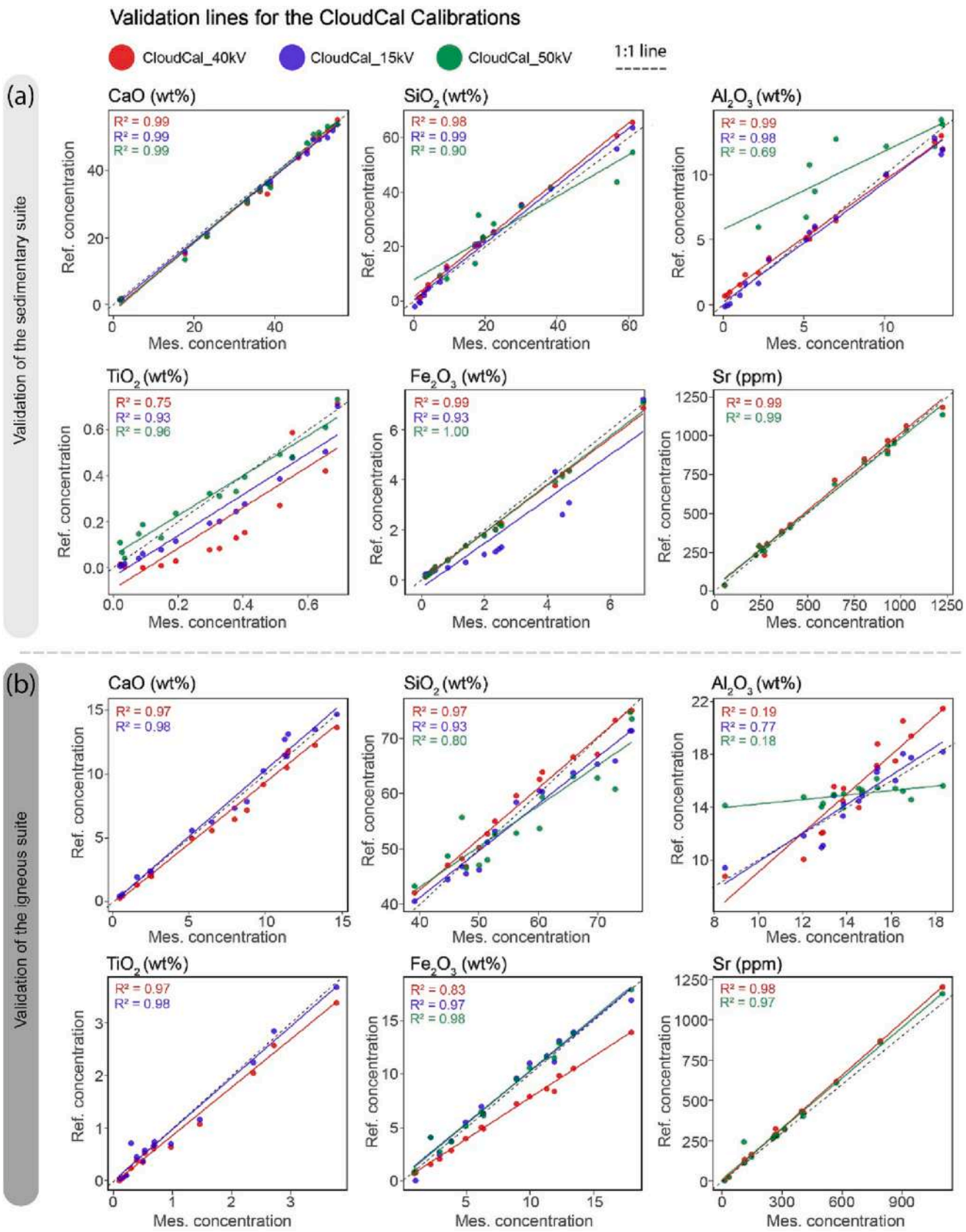

Fig. 5. Validation lines for the CloudCal calibration models of CaO, SiO_2, Al_2O_3, TiO_2, Fe_2O_3 and Sr for the sedimentary suite at the top and the igneous suite at the bottom. These plots compare reference concentrations ('Ref. conc.' on the X axis) with measured concentrations ('Mes. Conc.' on the Y axis) for each calibration model. A linear trend is fitted for each validation set, and its correlation coefficient (R) and its slope (S) are calculated and displayed on each plot. The ideal 1:1 line (black dashed) is also displayed for comparison. Abbreviations of calibration names are provided in Table 1. (Statistics of validation lines of all chemical elements/oxides are given in Suppl. Material-Table 5).

to the origin and by normalizing each spectrum to the Compton peak (see section 2.3). The other calibrations models include all the correction factors of the EasyCal_basic calibration but also the use of influence coefficients to correct for the effects of the matrix composition on the analyte intensity (EasyCal_basic-matrix) or the correction for spectra interferences (EasyCal-basic-interferences), or both (EasyCal-basic-matrix-interferences). Ultimately, we provide a final calibration model that combines the optimal corrections from each EasyCal calibration model for an optimal calibration model (EasyCal-Final).

The use of the influence coefficient to correct for the matrix effect (EasyCal-basic-matrix) strongly improves the accuracy of the validation lines by decreasing the m%Err for most chemical elements (e.g., for the sedimentary suite: Al_2O_3 m%Err decreases from 71.4 to 9.2% (R^2 from 0.8 to 0.96); for SiO_2 from 40.6 to 12.0% (R^2 from 0.93 to 0.99); for K_2O from 153.4 to 29.0% (R^2 from 0.22 to 0.93); for Rb from 51.1 to 14.7% (R^2 from 0.23 to 0.6), for Sr from 9.8 to 5.4% (R^2 from 0.99 to 0.99), and for Zr from 296.9 to 67.5% (R^2–1.57 to −0.37; see Fig. 6). However, the use of influence coefficients makes the calibration slightly less accurate for CaO (m%Err increases from 3.3 to 8.4%, R^2 from 1 to 0.98), Fe_2O_3 (from 5.4 to 8.7%, R^2 1.0 to 1.0) and MnO (from 16.7 to 17.2% R^2 from 0.89 to 0.89) for the sedimentary suite, and for Fe_2O_3 (increasing from 3.9 to 5.1%, R^2 from 0.99 to 0.99), and Sr (from 3.9 to 5.7% R^2 from 1.0 to 0.99) for the igneous rock suite (see Fig. 6).

Applying the spectra interference correction (EasyCal-basic-interferences) improves the calibration by decreasing their m%Err, for a few oxides such as P_2O_5 (R^2 no data to 0.78) in the sedimentary suite and from 47.4 to 14.4% and R^2 0.19 to 0.96 in the igneous rocks suite, TiO_2 (m%Err from 19.1 to 10.9% and R^2 0.95 to 0.99 in the sedimentary rock suite), SiO_2 (m%Err from 11.6 to 7.5% and R^2 0.14 to 0.65 in the igneous suite), Fe_2O_3 (m%Err 5.4 to 5.3% and R^2 stable at 1.00 in the sedimentary rock suite) and Zr (m%Err from 296.9 to 18.4% and − 1.57 to 0.96 in the sedimentary rock suite and from 28.3 to 15.6% and R^2 0.72 to 0.91 in the igneous rock suite). However, in some cases, the correction for spectral interference makes the calibration less accurate for Al_2O_3 (m%Err 71.4 to 80.9% and R^2 from 0.8 to 0.79) and SiO_2 (m%Err 40.6 to 62.6% and R^2 from 0.93 to 0.87) in the sedimentary suite and TiO_2 (m%Err from 9.0 to 9.6% and R^2 1 to 0.97) and Y (m%Err 22.7 to 62.9% and R^2 from 0.75 to 0.56) in the igneous suite.

Combining influence coefficient correction with spectral interference corrections (EasyCal-basic-matrix-interferences) does not improve the calibration obtained either through the correction of spectral interference or via the use of influence coefficient only, except for P_2O_5. Furthermore, combining the two types of corrections in some cases leads to higher m%Err (for MgO and Al_2O_3 in the two rock suites) than the calibration obtained with only one correction, showing how over-correction can lead to a decrease in calibration accuracy.

For the sedimentary suite, the validations lines are in good agreement with the 1:1 line (Fig. 7), with the R^2 value of the validation lines ranging between 0.96 and 1.00, except for Rb ($R^2 = 0.52$), for Y ($R^2 = 0.88$), for MnO ($R^2 = 0.89$), with slope values close to the 1:1 trend, except for Al_2O_3 (S = 0.90), MnO (S = 1.3), Rb (S = 0.92) and Zr (S = 0.96). For the igneous suite, the calibrations bring slightly less accurate results for SiO_2 ($R^2 = 0.88$, S = 1.1), Al_2O_3 ($R^2 = 0.80$, S = 1.1), MnO ($R^2 = 0.86$, S = 0.8); and Zr ($R^2 = 0.59$, S = 0.97).

To sum up, the optimal EasyCal calibration models are made using matrix effect correction only (for MgO, MnO, Ni, Cu, Zn, Rb, and Sr for both rock suites) or applying spectra interference correction only (for CaO, TiO_2, Zr for the sedimentary suite, P_2O_5 and Fe_2O_3 for the igneous suite) or both of them (for Al_2O_3, SiO_2, P_2O_5, Zr for the sedimentary suite). As a result, the optimal calibration settings (EasyCal-Final) provide m%Err below 50% or R^2 higher than 0.5 for all the oxides/elements in both suites, except for Ni and Cu in the sedimentary suite ($R^2 = 0.16$ for Ni and $R^2 = 0.4$ for Cu; Figs. 6 and 7) and will be discussed in section 4. All the results of the statistics of validation lines can be found in Table 4 of the supplementary materials.

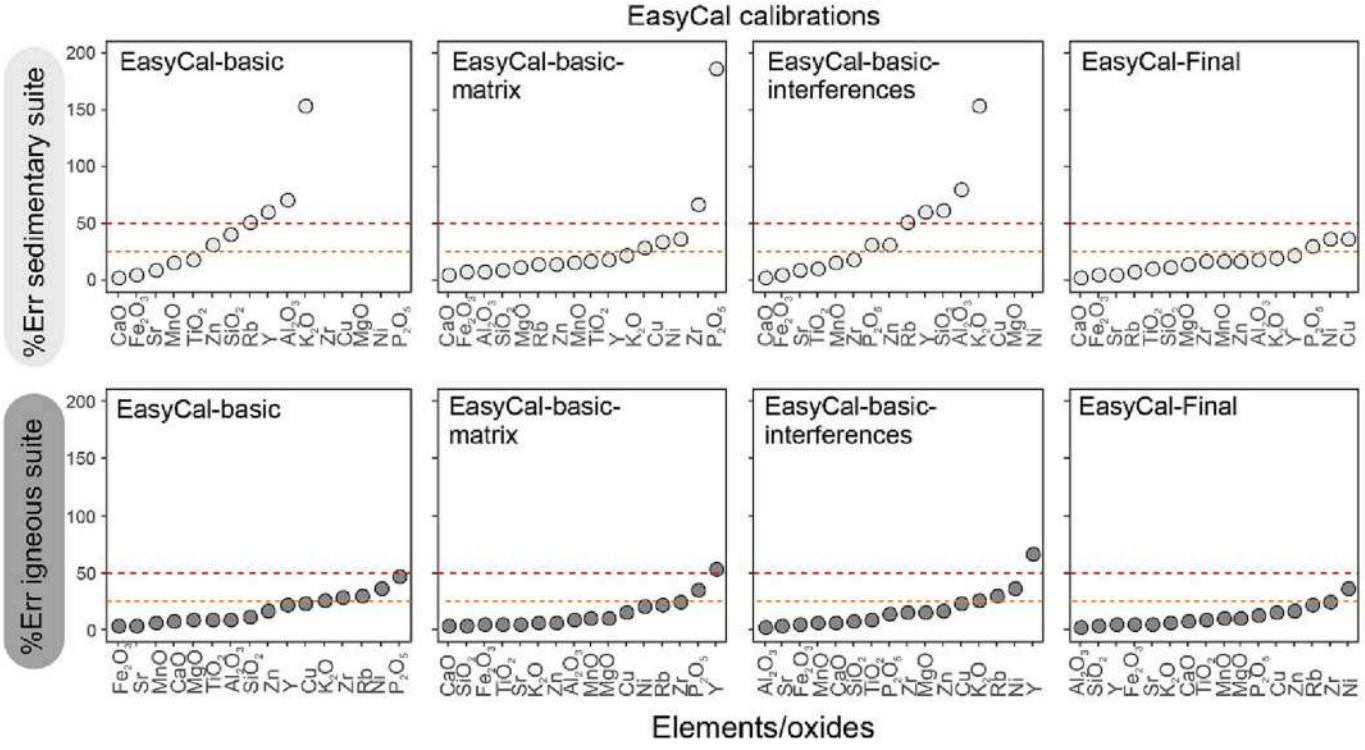

Fig. 6. Median percentage error (m%Err, see main text for definition) for the EasyCal calibration models for the sedimentary suite at the top and the igneous suite at the bottom. Chemical elements on the X axis are ordered by order of increasing m%Err. The horizontal dotted lines mark 25% and 50% of m%Err. The absence of a black dot means the calibration does not quantify the element. Abbreviations for each calibration model are provided in Table 1.

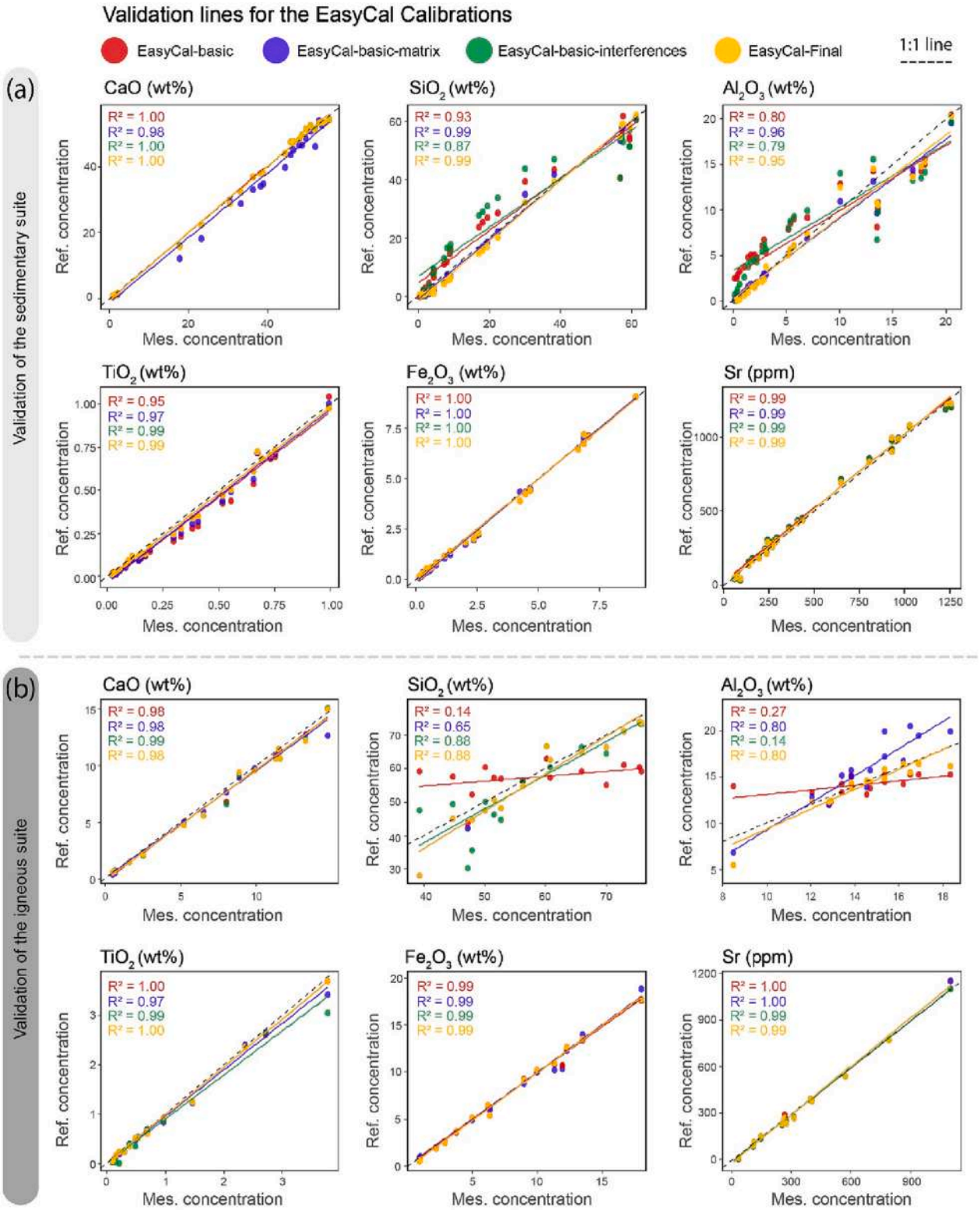

Fig. 7. Validation lines for the EasyCal calibration models of CaO, SiO_2, Al_2O_3, TiO_2, Fe_2O_3 and Sr for the sedimentary suite at the top and the igneous suite at the bottom. These plots compare reference concentrations ('Ref. conc.' on the X axis) with measured concentrations ('Mes. Conc.' on the Y axis) for each calibration model. A linear trend is fitted for each validation set, and its correlation coefficient (R) and its slope (S) are calculated and displayed on each plot. The ideal 1:1 line (black dashed) is also displayed for comparison. Abbreviations of calibration names are provided in Table 1. (Statistics of validation lines of all chemical elements/oxides are given in Suppl. Material-Table 5).

3.4. GeoMining and GeoExplorer calibrations

The GeoMining and GeoExplorer calibrations differ in the oxides/elements added to the SiO_2 matrix of their in-house reference materials (Bruker, 2017). For GeoMining and GeoExplorer, a calibration with only two phases and one with three phases of source excitation were compared.

Four calibration models are proposed hereafter: GeoMining and GeoExplorer with three phases (named GeoMining-3phases or GeoExplorer-3phases with three phases of 25 s each, with a total analysis duration of 75 s), GeoMining and GeoExplorer with two phases (named GeoMining-2phases or GeoExplorer-2phases with two phases, each operating for 25 s, for a total analysis duration of 50 s).

No significant difference was observed for the quantification of light and heavy elements between the two or three phases' calibrations (Figs. 8-9) for both sedimentary and igneous suites. In turn, using three excitation source phases improves the quantification of transition elements such as Ni, Cu and Zn. The GeoExplorer calibration provides slightly better results for the sedimentary rock suite, while GeoMining provides slightly more accurate calibrations for the igneous rock suite (see Figs. 8 and 9). Concerning Ni and P_2O_5, both calibrations models provide m%Err higher than 50% and R^2 negative, while for MgO, Al_2O_3, SiO_2, K_2O, CaO, TiO_2, MnO, Fe_2O_3, Rb, and Zr, m%Err values show values lower than 50% (Fig. 8). For the sedimentary suite, the validation sets (Fig. 9) show coefficient of determination of 0.93–0.99, except for Rb ($R^2 = 0.62$) and Zr ($R^2 = 0.77$), and for the igneous suite, except for Al_2O_3 ($R^2 = -0.53$), and Zr ($R^2 = 0.60$). However, for most chemical elements, the validation lines show a poor fit with the 1:1 line. For the sedimentary suite, the validation lines have a slope of 0.98 or 1.0–1.1 for SiO_2, Al_2O_3, TiO_2 and MnO (Fig. 9), while for the other elements, it ranges between 0.53 and 0.96. For the igneous suite, R^2 values range between 0.91 and 0.99 for most oxides/elements, except for Al_2O_3, MnO and Zr, while their slope values range between 0.80 and 0.96 and 1.30, suggesting that quantified results provided by GeoMining or GeoExplorer calibrations are either over- or under-estimated. All the results of the statistics of validation lines can be found in the supplementary materials (Table 5).

4. Discussion and perspectives

4.1. Statistics on validation lines is the way to quantify the accuracy of a pXRF calibration

The accuracy of a calibration model can be quantified: (i) by calculating statistics directly on the calibration line which compare peak intensity and reference concentrations (using e.g., the coefficient of correlation of the calibration line, denoted "r", the Cook distance analysis, QQ plots; see examples of this approach in Al-Musawi and Kaczmarek, 2020; Mejía-Piña et al., 2016; Rouillon and Taylor, 2016) or (ii) by testing a calibration model on a set of validation samples and by analysing the statistics on the *validation lines* which compare measured and known concentrations (using e.g., the coefficient of determination R^2, the slope value of the linear regression line (S), %Err, m%Err; see details in section 2 and examples of this approach in VanCott et al., 1999; Rouillon and Taylor, 2016; Declercq et al., 2019; Triantafyllou et al., 2021). The first approach has been extensively used in several research papers but can be misleading as the regression line in a calibration model may show an excellent coefficient of determination but an important mismatch in expected concentrations when applied on a reference material for validation. We compared the correlation coefficient (r value) of the calibration lines with the median value of the percentage of error (m%Err; Fig. 10) of our set of validation samples for each quantified chemical element. These parameters are expected to be negatively correlated (i.e. lowest m%Err of the validation set corresponding to the highest r value of the calibration line). However, the correlation between r of the calibration line and m%Err is very low to non-existent. Most of the data are distributed as a plateau defining high correlation coefficients ($r > 0.97$) for the calibration line corresponding to increasing m%Err values (i.e. inaccurate calibration model). Some specific chemical elements and oxides (Rb, MgO, K_2O, P_2O_5) show off trends with high m%Err (>70%) meaning they are poorly quantified. However, these elements still show high r values for their calibration

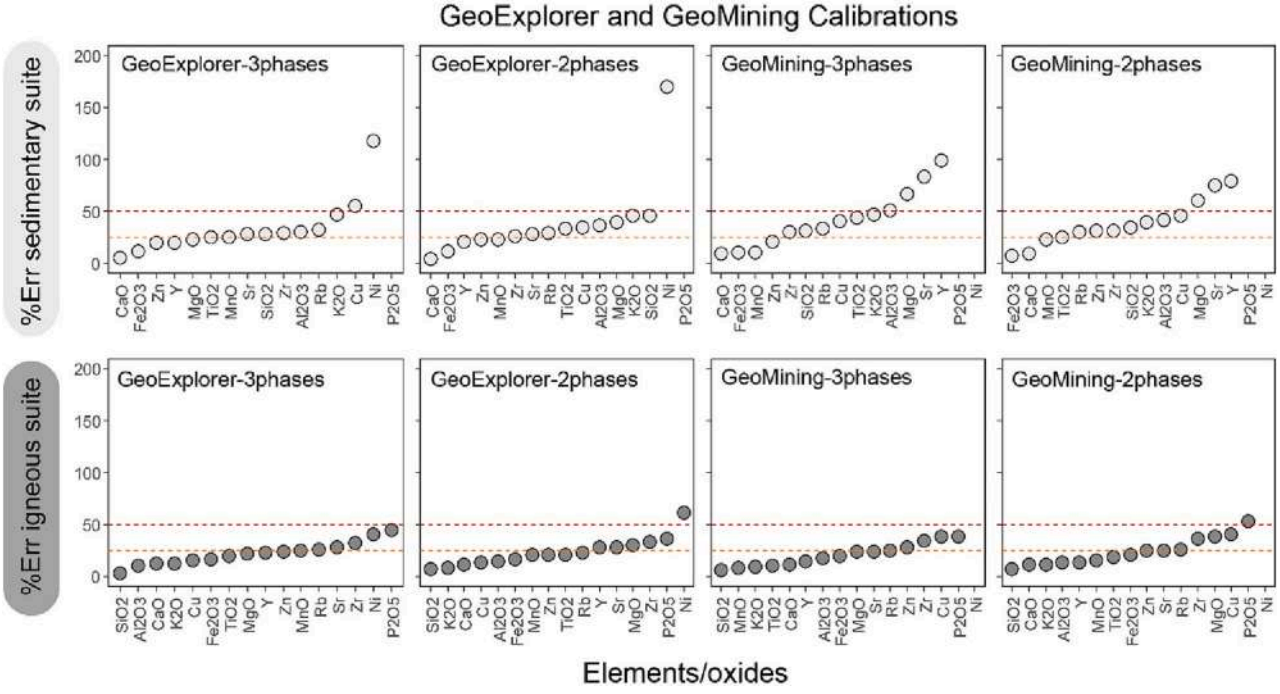

Fig. 8. Median percentage error (m%Err, see main text for definition) for the GeoMining and GeoExplorer calibrations for the sedimentary suite at the top and the igneous suite at the bottom. Chemical elements on the X axis are ordered by order of increasing m%Err. The horizontal dotted lines mark 25% and 50% of m%Err. The absence of a black dot means the calibration does not quantify the element. Abbreviations for each calibration model are provided in Table 1.

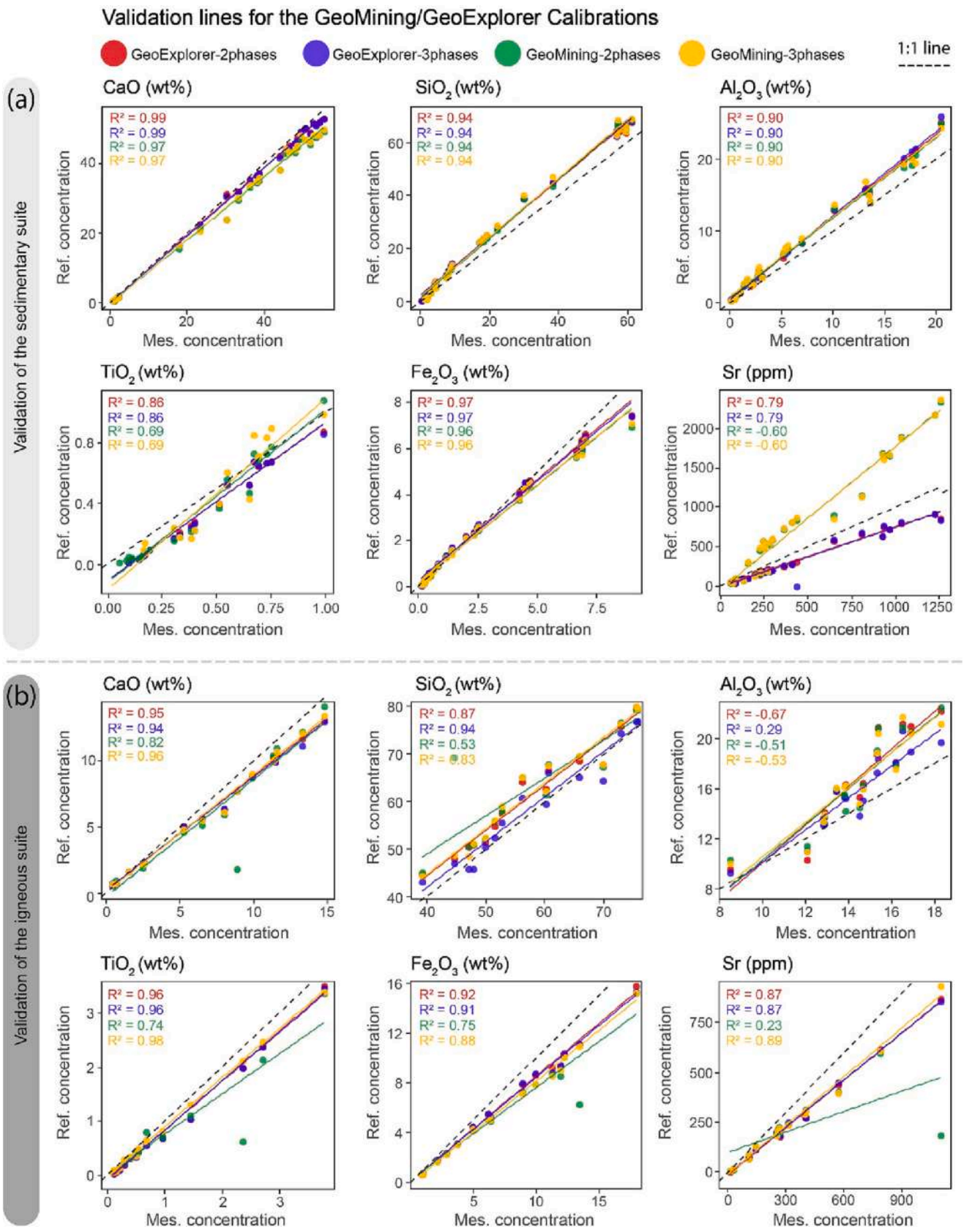

Fig. 9. Validation lines for the GeoMining and GeoExplorer calibrations of CaO, SiO_2, Al_2O_3, TiO_2, Fe_2O_3 and Sr for the sedimentary suite at the top and the igneous suite at the bottom. These plots compare reference concentrations ('Ref. conc.' on the X axis) with measured concentrations ('Mes. Conc.' on the Y axis) for each calibration model. A linear trend is fitted for each validation set, and its correlation coefficient (R) and its slope (S) are calculated and displayed on each plot. The ideal 1:1 line (black dashed) is also displayed for comparison. Abbreviations of calibration names are provided in Table 1. (Statistics of validation lines of all chemical elements/oxides are given in Suppl. Material-Table 5).

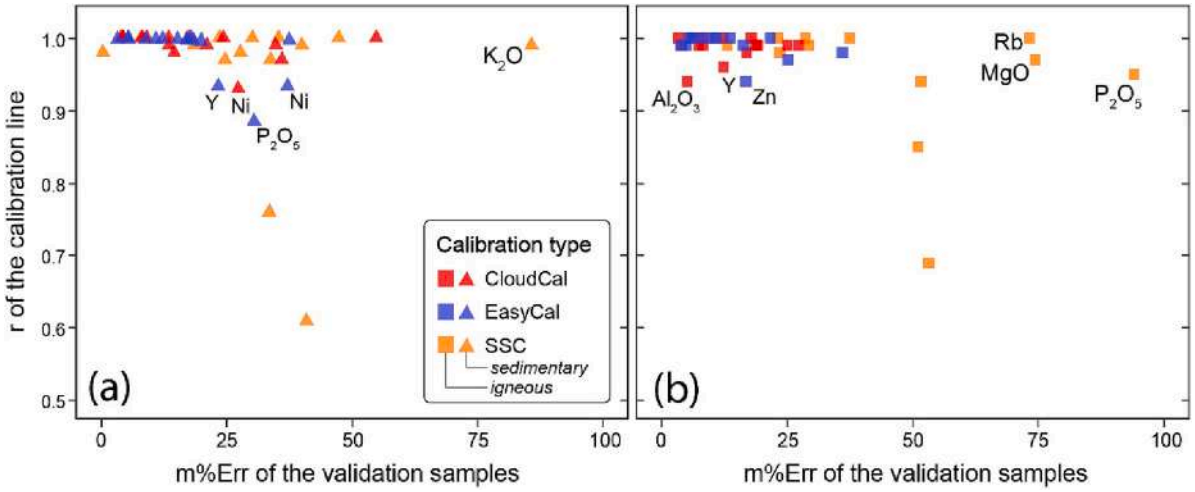

Fig. 10. Calibration line correlation coefficient (R) versus median percentage of error (m%Err) of the validation samples (each data point corresponding to an element), which should be negatively correlated. Bruker does not provide the coefficient of determination of the calibration lines from GeoMining and GeoExplorer.

lines ($r > 0.95$; Fig. 10). If we take K_2O for the sedimentary suite as an example, when we look at the calibration line (Fig. 11a), it show an excellent alignment and r values close to one, although their validation line displays a significant offset with the line of equality (i.e. 1:1 line in Fig. 11b). Another example comes from Rb in the igneous suite (Fig. 11c-d), in the SSC calibration line, for which the coefficient of correlation of the calibration line is artificially high ($r = 1$), because of the non-uniform spread of GRMs composition, leading to poor results on the

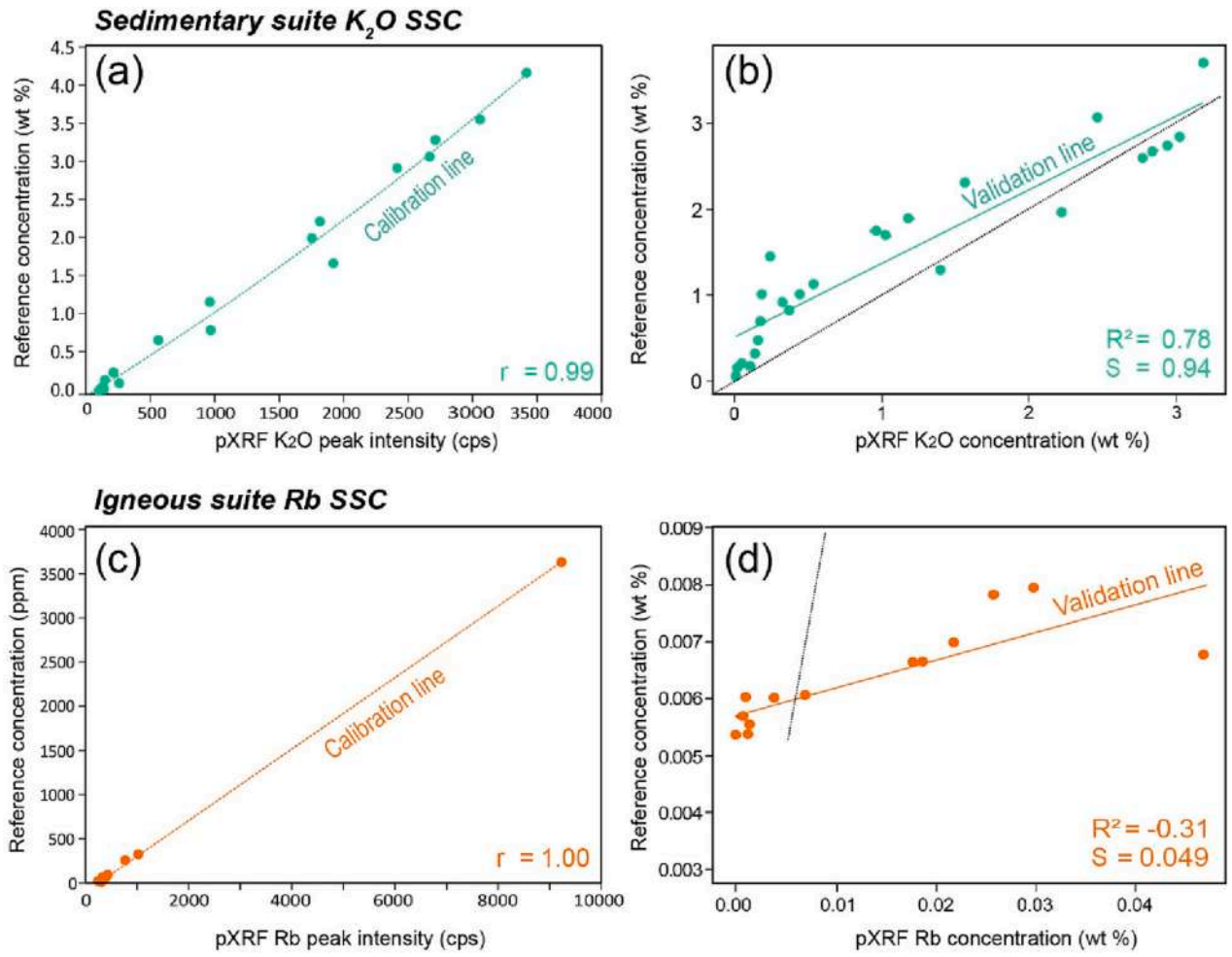

Fig. 11. Calibration and validation lines with the Simple Spreadsheet for K_2O for the sedimentary suite and Rb for the igneous suite. (a-b) Calibration and validation lines for K_2O for Simple Spreadsheet calibration Polynomial not going through 0. (c-d) Calibration and validation line for Rb for Simple Spreadsheet calibration Line not going through 0. r = correlation coefficient, R^2 = coefficient of determination, S = Slope. The validation line can be compared with an ideal 1:1 dotted black line.

validation line (S = 0.005, R^2 negatives). Furthermore, some elements show r value of their calibration lines below 0.95 (including Al_2O_3 and Y in the igneous rock suite in the CloudCal calibration, Zn in the igneous suite calibrated using the EasyCal software). However, still leading to m %Err below 25%, indicative of an accurate quantification (see Fig. 10).

Using validation lines is the only approach to test a "black box" calibration provided by a pXRF manufacturer. For example, the GeoExplorer and GeoMining built-in calibrations come without the details on their calibration models, their statistics and used reference materials. It appears that their performances are not ideal on our set of validation samples, showing a clear offset for most of the oxides/elements on the two rock suites (Fig. 9), which are over- or under-estimated. A similar offset on a validation set of Si-rich igneous rocks was already identified on the Tracer IV, using the Soil calibration (Triantafyllou et al., 2021). We suggest that such discrepancy could be due to the composition of their reference materials which might not be close enough to igneous and/or sedimentary rock matrices (e.g., Conrey et al., 2014; Steiner et al., 2017). Considering these observations, we strongly recommend future pXRF users to test the efficiency of their calibration model by gathering a set of well characterized validation samples and by building and analysing the statistics on validation lines. This would validate calibration models provided by manufacturers, or build in-house and it would allow to estimate the impact of unknowns' matrix effect and/or field conditions on quantified results.

4.2. *Which empirical calibration model brings the most accurate chemical concentration?*

For each calibration approach (i.e. Simple Spreadsheet, CloudCal app, EasyCal software or built-in GeoExplorer or GeoMining calibrations), one calibration model was selected bringing the most accurate chemical concentrations on the set of validation samples (see section 3). For the SSC, only basic corrections could be applied, and the most accurate validation lines were selected for each chemical oxides or element, combining linear and quadratic regression curves (Table 2, section 3.1.). For the GeoExplorer and GeoMining calibrations, the three-phases calibration was selected, with successive source voltage at 15, 30 and 50 kV, during 25 s acquisition time for each phase. The GeoExplorer calibration model appears slightly more accurate for the sedimentary rock suite and the GeoMining calibration model for the igneous suite (Section 3.4.). For the CloudCal calibration, we could see that the calibration at low voltage performed better for light elements than the calibration at higher voltage and that the calibration at high voltage performed better for heavier elements (Section 3.2.). We thus selected a combination of two phases calibration (at 15 kV and 50 kV each of them lasting 40 s). As for the CloudCal calibration, the GeoExplorer and GeoMining modes, multiple phase calibration provides better quantification than single phase calibration. For the EasyCal calibration, we modified the calibration parameters aiming at the best correlation coefficient for our calibration and we checked the performance of our calibration through the values obtained for the statistic parameters obtained on the validation lines (%Err, R^2 and slope). We then came back to the calibration, modified some parameters, and checked how it behaved on the validation set. Through this iterative process, we aimed at reaching the optimal calibration model (EC-Final), applying different correction procedures, with the correction of the matrix effect being key for most of the oxides /elements (see Section 3.3.).

All the configurations of these final calibrations are available in supplementary material Table 4; while all the statistics of the validation lines of these final calibrations are available in supplementary material 5 We calculate the %Err of each oxides/element for each calibration based on their validation set of samples (Fig. 12); the lowest %Err corresponding to the closest correspondence with the reference concentration. For both suites (Fig. 12), globally, we can see that the spread of the %Err values in the SSC (orange) and GeoExplorer (green) calibrations is stronger and that the %Err for these calibrations tends to be higher, pointing to lower accuracy calibrations (even if it's not always the case).

For the igneous rock suite, the %Err for the SSC range between 13 and 94%; for the GeoExplorer calibration between 2 and 44%, for EasyCal between 3 and 36% and for CloudCal between 3 and 27%. For the sedimentary rock suite, these %Err for the SSC range between 8 and 85%; for the GeoExplorer calibration between 4 and 117%, for EasyCal between 3 and 37% and for CloudCal between 4 and 55% (Fig. 12). These results show that the CloudCal and the EasyCal calibrations provide better accuracy on obtained chemical results than both Simple Spreadsheet and GeoExplorer calibrations. The high %Err for the SSC and GeoExplorer calibrations, are also attested by the poor fitting of their validation lines with the 1:1 line (see Figs. 3 and 9) for both rock suites.

For the sedimentary rocks suite, all the calibrations provided relatively low-quality results for K_2O, P_2O_5, Ni and Y. Concerning the igneous rock suite, the poorest quantification results are for MgO (except through EasyCal), Ni and P_2O_5, with %Err largely bigger than 100% (Fig. 12). P_2O_5 is systematically poorly quantified, which has been also observed in other studies (Hall et al., 2014; Kenna et al., 2011; Rowe et al., 2012). Hunt and Speakman (2015) and Hall et al. (2014) suggest that it could due to, the low concentration in the measured material; the low energy of the characteristic X-rays of P, the interference of the P Kα line with Ca escape peak and its neighbourhood to Si Kα and Kβ lines. Other studies have also documented poor pXRF performances in quantifying Mg as a light element with a low peak-to-noise ratio which makes it more difficult to quantify without using gas flux (e.g., Brouwer, 2018; Hall et al., 2014; Rowe et al., 2012). Ni was also poorly quantified by Kenna et al. (2011), while Hall et al. (2014) noted excellent quantification results for samples relatively rich in Ni (above 250 ppm) which is not the case for our set of samples. K_2O is poorly quantified for the sedimentary rock suite (m%Err is 85% for SSC, 20% for EasyCal, 55% for CloudCal and 42% for GeoExplorer), but its accuracy is better for the igneous rock suite (m%Err is 34% for SSC, 6% for EasyCal and CloudCal and 12% for GeoMining). This could be caused by the spectral interference with the Ca Kα line - Ca being more abundant in most of the sedimentary rock suite - but also to the low K content for most of the analysed rocks (< 3 wt% K_2O) in this study.

Table 2
Best calibration model for each oxide/element and rock suite using the Simple Spreadsheet calibration type. Line-no0 = calibration line not going through 0; Line-0 = calibration line going through 0; poly-0 = polynomial calibration going through 0; poly-no0 polynomial calibration not going through 0. The hyphens indicate that the calibration model could not provide sufficient correlation between the reference and quantified concentrations (see Fig. 3).

Oxides/Elements	MgO	Al_2O_3	SiO_2	P_2O_5	K_2O	CaO	TiO_2	MnO	Fe_2O_3
Sedimentary	Line-no0	Line-no0	Poly-no0	–	Poly-no0	Poly-no0	Lines-0	Poly-no0	Line-0
Igneous	Poly-no0	Line-no0	Poly-0	Poly-no0	Line_no0	Poly-no0	Line_no0	Line_no0	Poly-no0

Oxides/Elements	Ni	Cu	Zn	Ga	Rb	Sr	Y	Zr
Sedimentary	–	Line-no0	Poly-no0	–	Lines_no0	Poly-0	Line_no0	Poly-no0
Igneous	–	–	–	–	–	–	–	Line-0

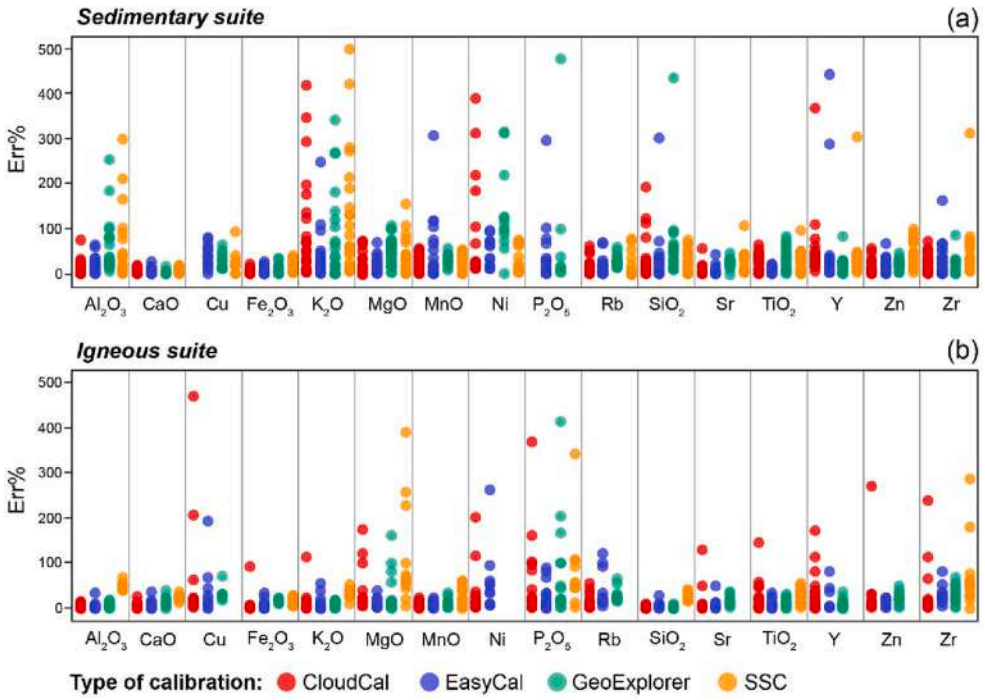

Fig. 12. Percentage of Error (Err %) comparison of the different calibrations for the sedimentary suite at the top and the igneous suite at the bottom for a selection of chemical elements and oxides on the X-axis. The highest %Err corresponds to less accurate quantification results. The SSC and GeoExplorer calibration models show the highest %Err for most chemical elements and rock suites.

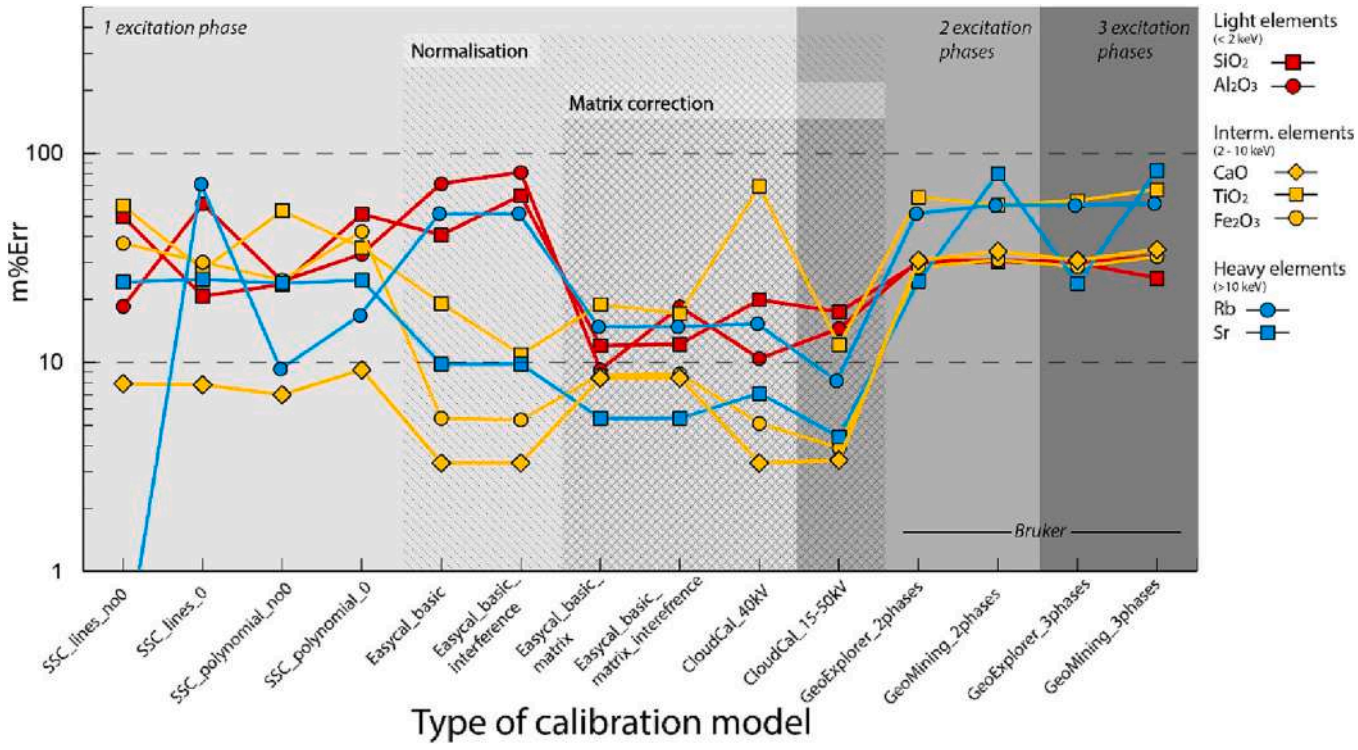

Fig. 13. Line plot comparing the median percentage of Error (m%Err) of all the calibration models for the sedimentary suite and seven chemical elements: light (SiO_2 and Al_2O_3, < 2 keV), intermediate (CaO, TiO_2 and Fe_2O_3, from 2 to 10 keV) and heavy elements (Rb and Sr, > 10 keV). This plot allows to weigh the impact of the different correction parameters and acquisition conditions discussed in section 2. Calibration names are explained in Table 1. 'Bruker' stands for built-in calibrations provided for the Tracer 5 g.

4.3. Weighting the impact of calibration parameters on quantification accuracy

Several correction parameters can be applied by building an empirical calibration model, including the mode of spectra normalization, the type and geometry of the regression line or the correction for the matrix effect. It is challenging to weigh the effect of each of these correction parameters. This is shown in Fig. 13, displaying the variation in m%Err for light (SiO_2, Al_2O_3; < 2 keV), intermediate (CaO, TiO_2, Fe_2O_3; from 2 to 10 keV) and heavy elements (Rb, Sr; > 10 keV in the raw pXRF spectrum) for each calibration model. This set of elements allows to track each calibration's efficiency at different parts of the energy spectrum and for different ranges of concentrations (major, minor and trace elements).

Concerning the geometry of the regression line, comparing SSC calibrations shows that not anchoring the calibration line at the origin can be slightly more efficient than when it is anchored, mostly for the polynomial line. Indeed, in Fig. 13, the %Err tends to increase (lower efficiency) from SSC-polynomial-no0 (no anchoring) to SSC-polynomial-0 (anchoring). Using a polynomial in the SSC calibration instead of a line usually does not significantly improve the calibration (Fig. 13 and section 3.1.). However, using a polynomial regression function to build calibration curves instead of a straight linear regression can help to counterbalance the matrix effect (Kido et al., 2006), while not passing through the origin can help to account for spectral interferences (Conrey et al., 2014).

For most of the elements, except for light elements, the normalization to the Compton or Rayleigh peak of the raw pXRF spectra considerably improves the accuracy of the chemical results. This correction can be evaluated by comparing the SSC calibration (no normalization) with the EasyCal-basic calibration on Fig. 13. The transition is marked by a significant drop in m%Err from e.g., 23.9 to 9.8% for Sr, from 24.7 to 5.4% for Fe_2O_3 and from 27.7 to 19.1% for TiO_2. However, the impact of the normalization on the Compton peak, at 18.5–19.5 keV, decreases for lighter elements, and we can see that the Compton normalization did not improve the quantification of Al_2O_3 and SiO_2 (1.55 to 71.4% for Al_2O_3 and 1.48 to 40.6 for SiO_2). The influence of peak overlap (or interference) correction strongly depends on the element considered and the rock matrix. It is crucial for element peaks close to a highly abundant element in the analysed rock. The interference correction is only efficient for a few oxides/elements such as CaO, Fe_2O_3, MnO, TiO_2, Zr and P_2O_5 (see Fig. 13 and section 3.3.). Considering a single-phase analysis, comparing the EasyCal calibrations with each other (Figs. 6-7), the use of matrix effect corrections appears to be the most efficient correction parameter for all elements (except for TiO_2), lowering their m%Err values (Fig. 13 and details in section 3.3.). It has a major impact on light elements like SiO_2 and Al_2O_3, for which median relative errors are lower from 40.6 to 12.0% and 71.4 to 9.2%, respectively (Fig. 13).

The calibration procedure using the EasyCal software allows to compare the reference concentration with the calculated concentration directly and thus, to visualize the improvement of the calibration line on the fly while applying correction parameters. The calibration line of Al_2O_3 provided by EasyCal (Fig. 14a) with the corrections from the normalization on the Compton peak only and a polynomial regression line being considered a basic configuration leads to a correlation coefficient r of 0.976 for the calibration line. In Fig. 14b, we apply additional corrections for the overlap of Mg kα1, K kα1 and Si kα1 peaks, which provides a calibration line with an r value of 0.990. In Fig. 14c, the basic calibration also includes the influence coefficient for Mg, Fe, Ca, K and Si to correct for matrix effects. This correction increases the coefficient of correlation of the calibration line to $r = 0.998$, highlighting the very large impact of the matrix correction on the calibration model. In Fig. 14d, the calibration line is built by including all corrections factors combined from the previous calibrations, showing the highest coefficient of determination ($r = 1.00$). The influence of matrix effect correction is important for elements such as Zr, Sr and Zn, which are less well calibrated with SSC calibrations than the CloudCal and EasyCal calibration models. Those elements are in relatively low amounts in the

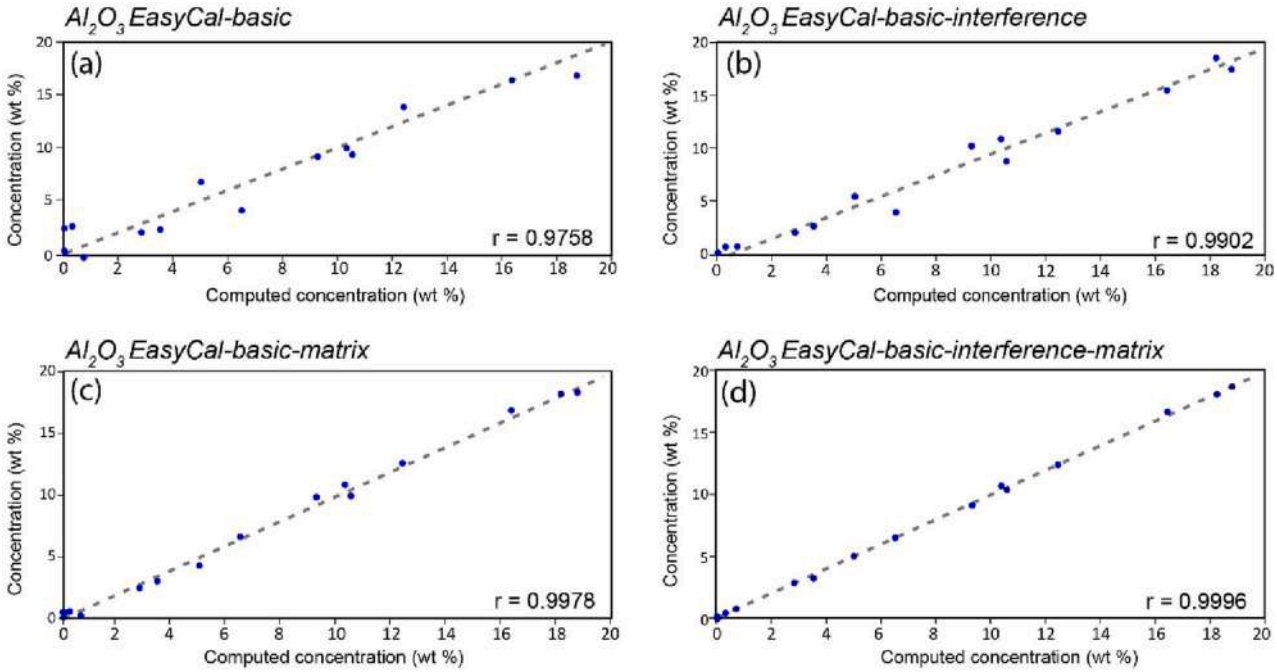

Fig. 14. EasyCal Calibration lines (dashed black lines) and their coefficient of correlation (r) for Al_2O_3 on the sedimentary suite (each blue dot represents the analysis of a sedimentary GRM). The reference chemical concentration (wt%) is shown on the Y axis, and the computed concentration on the X axis. (a) Calibration using a polynomial regression line with the Compton normalization (named 'EasyCal-basic'). (b) EasyCal-basic calibration with additional correction for spectral interference (named 'EasyCal-basic-interferences'). (c) EasyCal-basic calibration with additional matrix corrections using influence coefficients (named 'EasyCal-basic-matrix') and (d) EasyCal-basic calibration combining the correction for spectral interferences and matrix effect (named 'EasyCal-basic-interferences-matrix'). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sedimentary and igneous suites and strongly sensitive to the matrix effect of elements such as Si and Ca in high abundance. Hence, the lack of matrix correction in SSC calibration is visible, leading to weaker calibration efficiencies. Oxides such as CaO, Fe_2O_3 and TiO_2 have good calibrations with SSC and those oxides, as shown in EasyCal calibration steps, have their calibrations weakly impacted by the matrix effect (indeed in Fig. 13, at the transition between EasyCal-basic and EasyCal-matrix the m%Err are decreasing for all light and heavy elements in red and blue and increasing for the intermediate elements CaO, Fe_2O_3 and TiO_2 in yellow).

The impact of using single versus multiple excitation phases analysis can be weighted by comparing quantification results from CloudCal calibrations (see also section 3.2.). One is built using a single (40 kV) and the other one using two excitation phases (15 kV and 50 kV) acquisition. Using a multiple phases acquisition for most elements allows a more accurate quantification by calibrating each element with the highest signal in a given energy window (Fig. 13, the transition between CloudCal-40 kV to CloudCal-15-50 kV leads to a general decrease of m% Err). This lowers the median relative % error value on the sedimentary suite: for SiO_2 from 19.9 to 17.4%, for Fe_2O_3 from 5.1 to 3.9% and Sr from 15.2 to 8.1% (Fig. 4, Fig. 13). Although GeoExplorer and GeoMining calibrations provided by Bruker are using a three-phases (15 kV, 30 kV, 50 kV) excitation mode, their quantification is weaker than for most single-phase calibrations using CloudCal and EasyCal (Fig. 12). Generally, the manufacturer's built-in calibrations show m%Err higher than 20, as most of our in-house and matrix-matched calibrations display m%Err lower than 20%. Most quantified elements reach m%Err lower than 10% for sophisticated calibration models (using the EasyCal software and/or the CloudCal app).

Our study shows that normalizing spectra positively impacts the accuracy of the calibration models for intermediate and heavy elements (> 20 keV in the energy spectrum). Furthermore, the matrix correction is a crucial parameter for the accuracy of most of the calibration models. Finally, combining multiple phases of acquisition is recommended as it optimizes the detection and quantification of elements on the entire energy spectrum from light to heavy elements. However, this is not mandatory, as shown by the m%Err, which are equivalent between the single-phase EasyCal calibrations and the two-phase CloudCal calibrations (Fig. 12).

4.4. *Assessing the lower and upper limits of quantification?*

The range of elemental concentration on which a calibration model is applicable can be assessed through the LOQs, combined with the upper limit of calibration.

The upper limit of quantification is directly related to the range of the GRMs concentrations used to build the calibration. Therefore, quantified concentrations obtained by extrapolating the range of quantification of a calibration model will result in important bias, specifically in the upper range of concentrations (Duée et al., 2019). However, Bruker (2017)

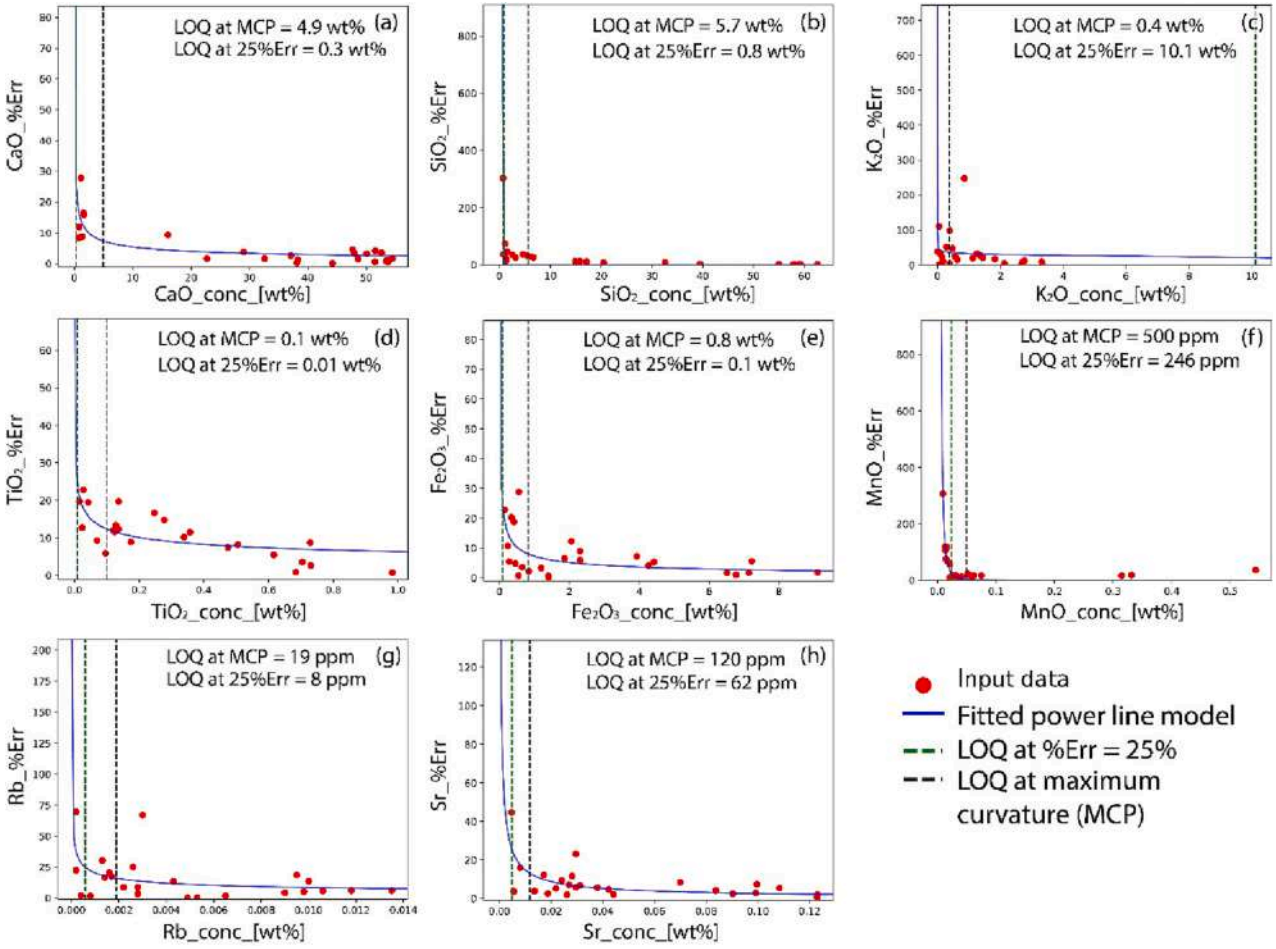

Fig. 15. Estimation of the empirical Limit of Quantification (LOQ) for a selection of elements (Empirical LOQs for all the chemical elements are given in the suppl. Material-Table 7). The empirical LOQ is estimated by comparing the chemical concentration against the percentage of Error (%Err) for a given element. This relationship is defined by a power trendline (blue line). The Empirical LOQ concentration is calculated either at the intercept of %Err = 25% (LOQ at.%Err 25, or LOQ_Emp_25; green dashed line); or at the maximum curvature of the power trendline (LOQ_Emp_MC; black dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

proposes that a calibration model can be extrapolated up to 10% over the highest value.

As defined in section 2.6, a Lower Limit of Detection (LOD) can be calculated through statistics (three times the standard deviation of a blank sample), and a Lower Limit of Quantification is classically defined as 10/3 x statistical LOD. In most of the literature, the statistical LOQ is considered as underestimated (e.g., Rousseau, 2001; Fiamegos and de la Calle Guntiñas, 2018). The empirical estimation of LOQ is a more indicative way to calculate the actual low range of quantification. We used the pXRF analyses from our validation set and compared their reference concentration with their %Err (x-axis and y-axis, respectively in Fig. 15) for each calibration and each chemical element (see a detailed method in Triantafyllou et al., 2021). These patterns generally define a power trendline, with an increase in the %Err value for low concentrations. An equation of the power function was fitted to the data, and the concentration at 25% of %Err was calculated. This approach allows estimating the empirical LOQ at a fixed %Err value of 25% for each oxide/element and each calibration mode. We also use this power function to propose a new methodology to estimate the LOQ at maximum curvature (see section 4.4.1.). In section 4.4.2. we compare these LOQs and use them in conjunction with the upper limit of calibration to assess the limit of quantification of our calibrations.

4.4.1. *Empirical lower limit of quantification at maximum curvature*

The major limitation of defining empirical LOQs is that users have to define the efficiency of the lower limits of quantification manually (using the %Err value for example, see in our example below empirical LOQ at %Err 25%) depending on the project's deliverables and specific geochemical results targets. This is why we propose a novel approach to automatically estimate the empirical LOQ by calculating the point of maximum curvature of the power function (named hereafter as LOQ_Emp_MC; see Fig. 15). We thus developed an in-house python code (available here: https://github.com/antoinetri/loq_max_curv) which calculates the equation of a power function with the best fit to the data and the point of maximum curvature on the modeled curve. The point of maximum curvature typically corresponds to the drastic increase in % Err (Fig. 15). The concentration at the point of maximum curvature was calculated and referred to as LOQ_Emp_MC. The main advantage of this method is that it precludes the use of an arbitrary %Err value by the pXRF user, which is one major drawback of defining empirical LOQ. The

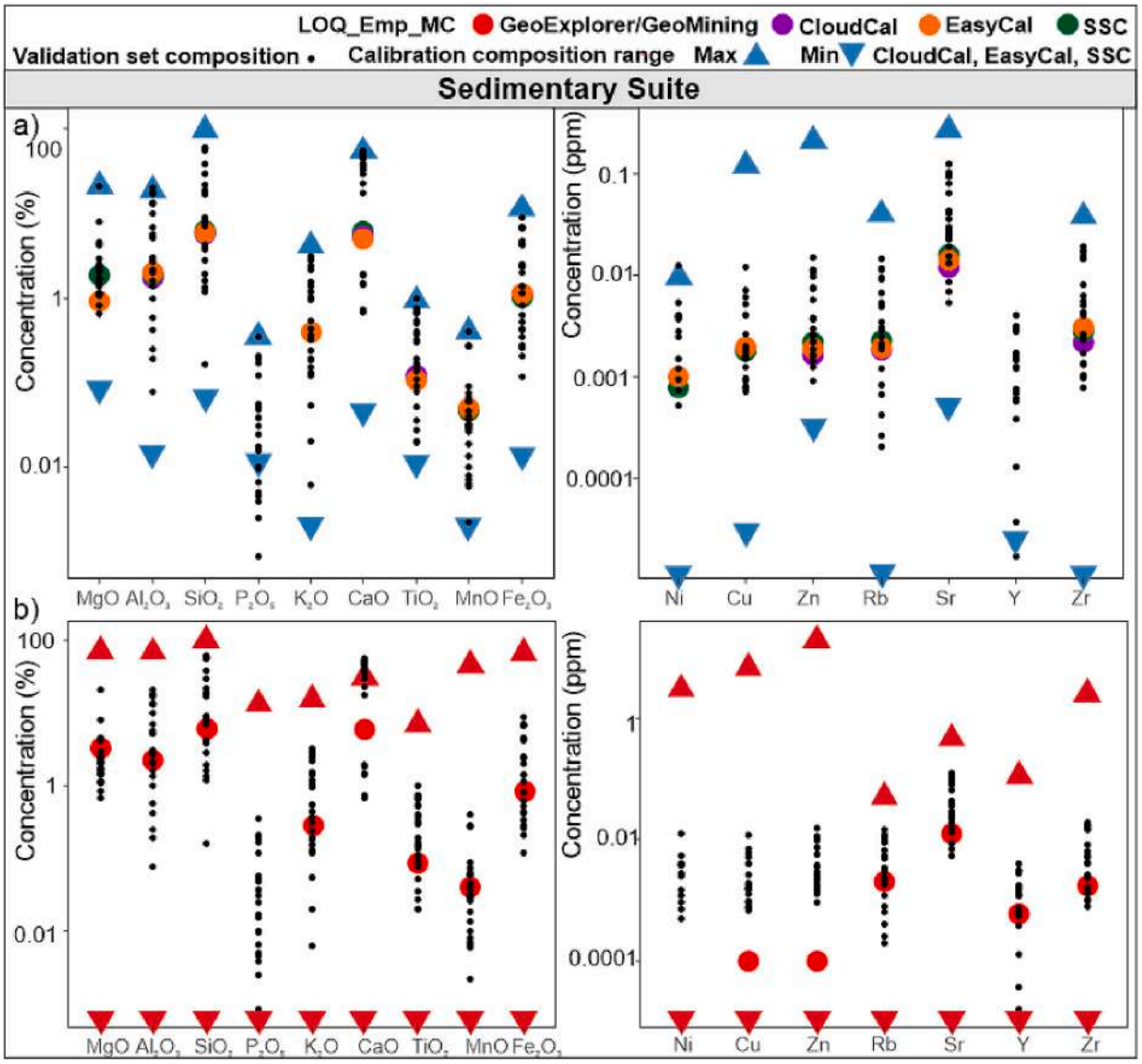

Fig. 16. Comparison of minimal (min) and maximal (max) elemental concentration of the GRMs used for the different calibrations, with the Empirical LOQ obtained through Maximal Curvature (LOQ_Emp_MC) and with the elemental concentration of the samples from the validation sets. a) CloudCal, EasyCal, and SSC calibrations on the sedimentary suite; b) GeoExplorer calibration on the sedimentary suite.

main issue is that in some cases, the maximum curvature could be reached at very high % Err, leading to LOQs values still associated with poorly quantified data. Considering the large range of measurements required to estimate the LOQ, we only provide those for the sedimentary suite (selected because it provides a larger range of matrices than the igneous suite). So, in the following section, only the sedimentary suite will be considered.

4.4.2. Elemental range for quantification

When comparing the LOQ obtained from different methods, the LOQs developed through empirical methods are usually higher than the statistical LOQ, and the LOQ_Emp_MC are higher than the LOQ_Emp_25, specifically for the lighter elements. The empirical LOQ is generally ten times higher than the statistical LOQ on average, confirming the view that statistical LOQ is classically underestimated (e.g., Fiamegos and de la Calle Guntiñas, 2018); Rousseau, 2001). The LOQs provided by Bruker (2017) for the GeoExplorer calibration are usually lower than all the other LOQ calculations, except for Mg and Zn (See Fig. 16 and supplementary material Table 6). We consider that the empirical LOQ at maximum curvature is less optimistic and should be used to assess the range of elemental concentration on which a calibration is applicable, combined with the upper limit of calibration, directly related to the range of the GRMs concentrations used to build the calibration.

Figure 16 includes the minimal and maximal elemental concentration of the GRMs used for the different calibrations, the Empirical LOQ at maximum curvature, compared with the elemental concentration of the samples from the validation (the equivalent of any geological object studied by the pXRF user) set for the sedimentary suite. It allows us to check which type of measurement can be considered relatively valid, or should be considered carefully, only for semi-quantitative measurement or just for identification purposes.

For the sedimentary suite and CloudCal, EasyCal and SSC calibrations (Fig. 16a), MgO, Al_2O_3, SiO_2, K_2O, CaO, TiO_2, MnO, Fe_2O_3, Cu, Zn, Rb, Sr and Zr composition of the validation set are all included between the minimal and maximal concentration of the GRM. For P_2O_5, the concentration of numerous samples is below the minimal concentration, so those samples should be considered carefully only for semi-quantitative or relative measurements. For most of the oxides/elements, some samples of the validation sets have concentrations below Empirical LOQ at maximum curvature, and so those measurements should be considered even more carefully, only for semi-quantitative or identification considerations. Only a few concentrations are higher than the maximum concentration of the GRM used for the calibration. For Al_2O_3 and Ni, one sample has a composition higher than the maximal concentration. For Al_2O_3, this composition is within 10% of this maximal value and Bruker (2017) proposes that a calibration can be extrapolated up to 10% over the highest value. Indeed, if we check the corresponding % Err for this point, it is 4% for SSC and CloudCal and only 0.2% for EasyCal, pointing indeed to a still efficient calibration. For Ni, we also have a value higher than the maximal value from the GRM, but this time it is higher than 10% from the highest value of the GRM, and the corresponding %Err for this sample is near 75% for SSC and EasyCal, so the calibration is less efficient.

The CaO composition of the validation set for thesedimentary suite includes some samples with compositions richer than the reference material used to build the GeoExplorer calibration. This does not seem to affect the %Err of the calibration of CaO, which are all in the same range as those obtained through CloudCal and EasyCal (Fig. 11) and lower than those obtained through SSC. However, the low proportion of CaO in the reference material used for the GeoExplorer calibration may have led to an under-evaluation of the matrix effect and an over or under-estimation of multiple elements.

5. Concluding remarks

This paper aims to share experience in the challenging but critical journey of building empirical calibrations for portable XRF spectrometers. Our initial motivation was to ensure the quality of our results and chemical analyses, to define analytical limitations and to optimize the capabilities of the Bruker Tracer 5 g pXRF. We realized that these challenges require the development of lab-made, dedicated, matrix-match empirical calibration models for which a tremendous literature exists as well as a specific and complex analytical terminology, both of which may act as a brake to further calibration development. Through the building and the comparison of 32 calibration models, we present a general workflow and numerous guidelines for the demanding and rigorous geoscientific community. These are listed hereafter:

1. Calibrations provided by the manufacturer can be inadequate for geological samples because the standard used to build those calibrations can have different matrices and “unrealistic” elemental composition for sedimentary or igneous rocks.
2. To test the efficiency of a newly built or built-in pXRF calibration model, the coefficient factors of the calibration lines are not sufficient and may be misleading. Therefore, users should gather a set of chemically well-characterized samples called “validation samples” (equivalent to the geological objects pXRF users intend to measure). Measurements on these validation samples will allow testing of the calibration model’s efficiency by building and analysing the statistics on validation lines.
3. Sophisticated calibration models allow the application of several correction parameters to the original raw spectrum. However, more corrections do not necessarily mean better quantification. We were able to estimate the effect of the different correction procedures on the quantified results. We showed spectra normalization and matrix correction coefficients are the most important parameters to improve the quantified results.
4. As the impact of the matrix effect corrections on a calibration are essential for most elements, it means that using a non-matrix matched calibration set or simple calibration lines (SSC) without matrix correction possibilities is strongly less efficient.
5. The LODs are generally used to estimate the capabilities of a pXRF spectrometer for low concentrations material. However, we show that using LOQ is more reliable when users aim to produce quantified results. We propose an original method to estimate the LOQ empirically without choosing a threshold value of %Err and its corresponding concentration. Instead, this approach fits a power function to the %Err versus concentration graph and calculates the LOQ at the point of maximum curvature. This approach is the best way to test a LOQ of a pXRF spectrometer, a given calibration, and a sample matrix combined.
6. We showed that the empirical LOQs for most chemical elements could be ten times higher than the statistical LOQ or those provided by pXRF manufacturers. The empirical LOQ combined with the upper elemental concentration for the calibration lines allow the user to estimate the application range of their newly built calibration.

We believe this paper will act as a roadmap for geoscientists who want to develop their empirical calibrations and validate them, ultimately bringing a better knowledge of their instrument and its capabilities.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

A.C.D.S. acknowledges the "*Conseil Universitaire de la recherche et la valorisation*", as well as « *Subside Fédéral de la recherche* » for financial support for the acquisition of the Portable XRF Bruker Tracer 5 g, as well as 'Fond National de la Recherche Scientifique' grant (FRS, T.0051.19 and J.0037.21). The program IGCP-652 on improving the geologic time scale is also acknowledged. A.T. acknowledges support from the Université Claude Bernard de Lyon with the "*BQR Accueil EC 2022*" and the *"AAP LYON 1 - Équipement de Recherche 2022"* granted to the LGLTPE. The Bruker Support Service and Brandon Lee Drake (CloudCal app) are also thanked for constructive discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemgeo.2023.121395.

References

Adams, C., Brand, C., Dentith, M., Fiorentini, M., Caruso, S., Mehta, M., 2020. The use of pXRF for light element geochemical analysis: a review of hardware design limitations and an empirical investigation of air, vacuum, helium flush and detector window technologies. Geochem. Explor. Environ. Anal. 20, 366–380. https://doi.org/10.1144/geochem2019-076.

Ahmed, A., Crawford, A.J., Leslie, C., Phillips, J., Wells, T., Garay, A., Hood, S.B., Cooke, D.R., 2019. Assessing copper fertility of intrusive rocks using field portable X-ray fluorescence (pXRF) data. Geochem. Explor. Environ. Anal. 20, 81–97.

Al-Musawi, M., Kaczmarek, S., 2020. A new carbonate-specific quantification procedure for determining elemental concentrations from portable energy-dispersive X-ray fluorescence (PXRF) data. Appl. Geochem. 113, 104491 https://doi.org/10.1016/j.apgeochem.2019.104491.

Analytical Methods Committe, 1987. Recommendations for the definition, estimation and use of the detection limit. Analyst 112, 199–204. https://doi.org/10.1039/AN9871200199.

Arenas-Islas, D., Huerta-Diaz, M.A., Norzagaray-López, C.O., Mejia-Piña, K.G., Valdivieso-Ojeda, J.A., Otero, X.L., Arcega-Cabrera, F., 2019. Calibration of portable X-ray fluorescence equipment for the geochemical analysis of carbonate matrices. Sediment. Geol. 391, 105517. https://doi.org/10.1016/J.SEDGEO.2019.105517.

Beckhoff, B., Kanngießer, Habil, B., Langhoff, N., Wedell, R., Wolff, H., 2006. Handbook of Practical X-Ray Fluorescence Analysis, Springer. ed. Springer.

Belter, M., Sajnóg, A., Barałkiewicz, D., 2014. Over a century of detection and quantification capabilities in analytical chemistry – Historical overview and trends. Talanta 129, 606–616. https://doi.org/10.1016/j.talanta.2014.05.018.

Brouwer, P., 2018. Theory of XRF. Getting Acquainted with the Principles, 5th edition. Malvern Panalytical, Almelo, The Netherlands.

Bruker, 2017. Bruker - Tracer 5i: Geo Exploration Calibration (P/N: 730.0187) and Geo Mining Calibration (P/N: 730.0203).

Cao, Y., Linnen, R., Good, D., Samson, I., 2018. Applications of the combined portable XRF-benchtop SEM methodology to PGE exploration. Ore Geol. Rev. 101, 32–53. https://doi.org/10.1016/j.oregeorev.2018.07.006.

Cohen, D.R., Cohen, E.J., Graham, I.T., Soares, G.G., Hand, S.J., Archer, M., 2017. Geochemical exploration for vertebrate fossils using field portable XRF. J. Geochemical Explor. 181, 1–9. https://doi.org/10.1016/J.GEXPLO.2017.06.012, 12p.

Conrey, R.M.M., Goodman-Elgar, M., Bettencourt, N., Seyfarth, A., Van Hoose, A., Wolff, J.A.A., 2014. Calibration of a portable X-ray fluorescence spectrometer in the analysis of archaeological samples using influence coefficients. Geochem. Explor. Environ. Anal. 14, 291–301. https://doi.org/10.1144/geochem2013-198.

Da Silva, A.C., De Vleeschouwer, D., Boulvain, F., Claeys, P., Fagel, N., Humblet, M., Mabille, C., Michel, J., Sardar Abadi, M., Pas, D., Dekkers, M.J.J., 2013. Magnetic susceptibility as a high-resolution correlation tool and as a climatic proxy in Paleozoic rocks – Merits and pitfalls: examples from the Devonian in Belgium. Mar. Pet. Geol. 46, 173–189. https://doi.org/10.1016/j.marpetgeo.2013.06.012.

Da Silva, A.C., Dekkers, M.J., De Vleeschouwer, D., Hladil, J., Chadimova, L., Slavík, L., Hilgen, F.J., 2019. Millennial-scale climate changes manifest Milankovitch combination tones and Hallstatt solar cycles in the Devonian greenhouse world. Geology 47. https://doi.org/10.1130/G45511.1.

Da Silva, A.C., Sinnesael, M., Claeys, P., Davies, J.H.F., de Winter, N.J., Percival, L., Schaltegger, U., Vleeschouwer, D., 2020. Anchoring the Late Devonian mass extinction in absolute time by integrating climatic controls and radio-isotopic dating. Sci. Rep. 10, 12940, 12p.

De Vleeschouwer, D., Königshof, P., Claeys, P., 2017. Reading time and paleoenvironmental change in the Emsian–Eifelian boundary GSSP section (Wetteldorf, Germany): A combination of cyclostratigraphy and facies analysis. Newsl. Stratigr. https://doi.org/10.1127/nos/2017/0397.

Declercq, Y., Delbecque, N., De Grave, J., De Smedt, P., Finke, P., Mouazen, A.M., Nawar, S., Vandenberghe, D., Van Meirvenne, M., Verdoodt, A., 2019. A comprehensive study of three different portable XRF scanners to assess the soil geochemistry of an extensive sample dataset. Remote Sens. 11 https://doi.org/10.3390/rs11212490.

Drake, B.L., 2018. CloudCal v3.0. https://doi.org/10.5281/zenodo.2596154.

Drake, B.L., MacDonald, B.L., 2023. Advances in portable X-ray fluorescence spectrometry. Intrumental, application and interpretation. Royal Society of Chemistry, p. 547p.

Duée, C., Orberger, B., Maubec, N., Laperche, V., Capar, L., Bourguignon, A., Bourrat, X., El Mendili, Y., Chateigner, D., Gascoin, S., Nolte, H., Koert, P., 2019. Impact of heterogeneities and surface roughness on pXRF, pIR, XRD and Raman analyses: challenges for on-line, real-time combined mineralogical and chemical analyses on drill cores and implication for "high speed" Ni-laterite exploration. J. Geochem. Explor. 198, 1–17.

Fiamegos, Y., de la Calle Guntiñas, M.B., 2018. Validation strategy for an ed-xrf method to determine trace elements in a wide range of organic and inorganic matrices based on fulfilment of performance criteria. Spectrochim. Acta - Part B At. Spectrosc. 150 https://doi.org/10.1016/j.sab.2018.10.009.

Forster, N., Grave, P., Vickery, N., Kealhofer, L., 2011. Non-destructive analysis using PXRF: methodology and application to archaeological ceramics. X-Ray Spectrom 40, 389–398. https://doi.org/10.1002/xrs.1360.

Frahm, E., Monnier, G.F., Jelinski, N.A., Fleming, E.P., Barber, B.L., Lambon, J.B., 2016. Chemical soil surveys at the Bremer Site (Dakota county, Minnesota, USA): measuring phosphorous content of sediment by portable XRF and ICP-OES. J. Archaeol. Sci. 75, 115–138. https://doi.org/10.1016/j.jas.2016.10.004.

Gallhofer, D., Lottermoser, B.G., 2018. The influence of spectral interferences on critical element determination with portable X-ray fluorescence (pXRF). Minerals 8. https://doi.org/10.3390/min8080320.

Goodale, N., Bailey, D.G., Jones, G.T., Prescott, C., Scholz, E., Stagliano, N., Lewis, C., 2012. PXRF: A study of inter-instrument performance. J. Archaeol. Sci. 39, 875–883. https://doi.org/10.1016/j.jas.2011.10.014.

Govindaraju, K., 1994. Compilation of working values and sample description for 383 geostandards. Geostand. Newslett. 18, 1–158. https://doi.org/10.1046/j.1365-2494.1998.53202081.x-i1.

Govindaraju, K., 1998. Geochemical reference materials. In: Geochemistry. Encyclopedia of Earth Science. Springer, Dordrecht. https://doi.org/10.1007/1-4020-4496-8_138.

Gregory, B.R.B., Patterson, R.T., Reinhardt, E.G., Galloway, J.M., Roe, H.M., 2019. An evaluation of methodologies for calibrating Itrax X-ray fluorescence counts with ICP-MS concentration data for discrete sediment samples. Chem. Geol. 521, 12–27. https://doi.org/10.1016/j.chemgeo.2019.05.008.

Hall, G.E.M., Bonham-Carter, G.F., Buchar, A., 2014. Evaluation of portable X-ray fluorescence (pXRF) in exploration and mining: phase 1, control reference materials. Geochem. Explor. Environ. Anal. 14, 99–123. https://doi.org/10.1144/geochem2013-241.

Hunt, A.M.W., Speakman, R.J., 2015. Portable XRF analysis of archaeological sediments and ceramics. J. Archaeol. Sci. 53, 626–638. https://doi.org/10.1016/j.jas.2014.11.031.

Ibáñez-Insa, J., Pérez-Cano, J., Fondevilla, V., Oms, O., Rejas, M., Fernández-Turiel, J.L., Anadón, P., 2017. Portable X-ray fluorescence identification of the Cretaceous–Paleogene boundary: application to the Agost and Caravaca sections, SE Spain. Cretac. Res. 78, 139–148. https://doi.org/10.1016/j.cretres.2017.06.004.

Johnson, L.R.M., Ferguson, J.R., Freund, K.P., Drake, L., Duke, D., 2021. Evaluating obsidian calibration sets with portable X-Ray fluorescence (ED-XRF) instruments. J. Archaeol. Sci. Rep. 39, 103126.

Kasztovszky, Z., Maróti, B., Harsányi, I., Párkányi, D., Szilágyi, V., 2018. A comparative study of PGAA and portable XRF used for non-destructive provenancing archaeological obsidian. Quat. Int. 468, 179–189. https://doi.org/10.1016/j.quaint.2017.08.004.

Kenna, T.C.T.C., Nitsche, F.O.F.O., Herron, M.M.M., Mailloux, B.J.B.J., Peteet, D., Sritrairat, S., Sands, E., Baumgarten, J., 2011. Evaluation and calibration of a Field Portable X-Ray Fluorescence spectrometer for quantitative analysis of siliciclastic soils and sediments. J. Anal. At. Spectrom. 26, 395–405. https://doi.org/10.1039/C0JA00133C.

Kido, Y., Koshikawa, T., Tada, R., 2006. Rapid and quantitative major element analysis method for wet fine-grained sediments using an XRF microscanner. Mar. Geol. 229, 209–225. https://doi.org/10.1016/j.margeo.2006.03.002.

Kikongi, P., Salvas, J., Gosselin, R., 2017. Curve-fitting regression: improving light element quantification with XRF. X-Ray Spectrom. 46, 347–355. https://doi.org/10.1002/xrs.2760.

Knight, R.D., Kjarsgaard, B.A., Russell, H.A.J., 2021. An analytical protocol for determining the elemental chemistry of Quaternary sediments using a portable X-ray fluorescence spectrometer. Appl. Geochem. 131, 105026. https://doi.org/10.1016/j.apgeochem.2021.105026.

Lachance, G.R., 1993. Correction procedures using influence coefficients in X-ray fluorescence spectrometry. Spectrochim. Acta Part B At. Spectrosc. 48, 343–357. https://doi.org/10.1016/0584-8547(93)80040-2.

Lachance, G., Traill, R., 1966. A practical solution to the matrix problem in X-ray analysis. Can. J. S. 11, 43–48.

Lemière, B., 2018. A review of pXRF (field portable X-ray fluorescence) applications for applied geochemistry. J. Geochem. Explor. 188, 350–363. https://doi.org/10.1016/j.gexplo.2018.02.006.

Lucas-Tooth, H., Pyne, C., 1964. The accurate determination of major constituents by X-ray fluorescent analysis in the presence of large interelement effects. Adv. X-ray Anal. 7, 523–541.

Lezzerini, M., Tamponi, M., Bertoli, M., 2014. Calibration of XRF data on silicate rocks using chemicals as in-house standards. Atti Soc. Tosc. Sci. Nat. Mem. Series A 121, 65–70, 10.24.24/ASTSN.M.2014.16.

Lucas-Tooth, H.J., Price, B.J., 1961. A mathematical method for the investigation of interelement effects in X-Ray fluorescence analysis. Metallurgia 64, 149–152.
Magnusson, B., Örnemark, U., 2014. The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics, Eurachem Guide.
Markowicz, A.A., 2008. Chapter 2: Quantification and correction procedures. In: Potts, P. J., West, M. (Eds.), Portable X-Ray Fluorescence Spectrometry - Capabilities for In Situ Analysis, pp. 13–38.
Martí, J., Zafrilla, S., Andújar, J., Jiménez-Mejías, M., Scaillet, B., Pedrazzi, D., Doronzo, D., Scaillet, S., 2020. Controls of magma chamber zonation on eruption dynamics and deposits stratigraphy: the case of El Palomar fallout succession (Tenerife, Canary Islands). J. Volcanol. Geotherm. Res. 399, 106908 https://doi.org/10.1016/j.jvolgeores.2020.106908.
McNulty, B.A.B.A., Fox, N., Berry, R.F.R.F., Gemmell, J.B.B., 2018. Lithological discrimination of altered volcanic rocks based on systematic portable X-ray fluorescence analysis of drill core at the Myra Falls VHMS deposit, Canada. J. Geochem. Explor. 193, 1–21. https://doi.org/10.1016/j.gexplo.2018.06.005.
Mejía-Piña, K.G., Huerta-Diaz, M.A., González-Yajimovich, O., 2016. Calibration of handheld X-ray fluorescence (XRF) equipment for optimum determination of elemental concentrations in sediment samples. Talanta 161, 359–367. https://doi.org/10.1016/j.talanta.2016.08.066.
Molchanova, E.I., Smagunova, A.N., Shcherbakov, I.V., 2011. Specific features of matrix correction in the X Ray fluorescence analysis of samples of widely varied composition. J. Anal. Chem. 66, 824–830.
Newlander, K., Goodale, N., Jones, G.T., Bailey, D.G., 2015. Empirical study of the effect of count time on the precision and accuracy of pXRF data. J. Archaeol. Sci. Rep. 3, 534–548. https://doi.org/10.1016/j.jasrep.2015.07.007.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html.
Quye-Sawyer, J., Vandeginste, V., Johnston, K.J.K.J., 2015. Application of handheld energy-dispersive X-ray fluorescence spectrometry to carbonate studies: opportunities and challenges. J. Anal. At. Spectrom. 30, 1490–1499. https://doi.org/10.1039/C5JA00114E.
Ravansari, R., Wilson, S.C., Tighe, M., 2020. Portable X-ray fluorescence for environmental assessment of soils: Not just a point and shoot method. Environ. Int. 134, 105250. https://doi.org/10.1016/j.envint.2019.105250.
Röhl, U., Abrams, L.J., 2000. High-Resolution, downhole and nondestructivre core measurements from sites 999 and 1001 in the Caribbean. In: Leckie, R.M., Sigurdsson, H., Acton, G.D., Draper, G. (Eds.), Proceedings of the Ocean Drilling Program, Scientific Results, 165, pp. 191–203.
Rouillon, M., Taylor, M.P., 2016. Can field portable X-ray fluorescence (pXRF) produce high quality data for application in environmental contamination research? Environ. Pollut. 214, 255–264. https://doi.org/10.1016/j.envpol.2016.03.055.
Rousseau, R.M., 1984. Fundamental algorithm between concentration and intensity in XRF analysis 1—theory. X-Ray Spectrom. 13, 115–120. https://doi.org/10.1002/xrs.1300130306.
Rousseau, R.M., 2001. Detection limit and estimate of uncertainty of analytical XRF results. Rigaku J. 18, 33–47.
Rousseau, R.M., 2006. Corrections for matrix effects in X-ray fluorescence analysis—a tutorial. Spectrochim. Acta Part B At. Spectrosc. 61, 759–777. https://doi.org/10.1016/j.sab.2006.06.014.
Rousseau, R.M., Willis, J.P., Duncan, A.R., 1996. Practical XRF Calibration Procedures for Major and Trace elements. X-Ray Spectrom. 25, 179–189. https://doi.org/10.1002/(SICI)1097-4539(199607)25:4<179::AID-XRS162>3.0.CO;2-Y.
Rowe, H., Hughes, N., Robinson, K., 2012. The quantification and application of handheld energy-dispersive x-ray fluorescence (ED-XRF) in mudrock chemostratigraphy and geochemistry. Chem. Geol. 324–325, 122–131.
Ryan, J.G.G., Shervais, J.W.W., Li, Y., Reagan, M.K.K., Li, H.Y.Y., Heaton, D., Godard, M., Kirchenbaur, M., Whattam, S.A.A., Pearce, J.A.A., Chapman, T., Nelson, W., Prytulak, J., Shimizu, K., Petronotis, K., Chapman, T., Nelson, W., Prytulak, J., Shimizu, K., Petronotis, K., 2017. Application of a handheld X-ray fluorescence spectrometer for real-time, high-density quantitative analysis of drilled igneous rocks and sediments during IODP Expedition 352. Chem. Geol. 451, 55–66. https://doi.org/10.1016/j.chemgeo.2017.01.007.
Saker-Clark, M., Kemp, D.B., Coe, A.L., 2019. Portable X-Ray fluorescence spectroscopy as a tool for cyclostratigraphy. Geochem. Geophys. Geosyst. 20, 2531–2541. https://doi.org/10.1029/2018GC007582.
Shackley, M.S., 2012. Portable X-ray fluorescence spectrometry (pXRF): the good, the bad, and the ugly. Archaeol. Southwest Mag. 26, 1–8.
Sheppard, P.J., Irwin, G.J., Lin, S.C., McCaffrey, C.P., 2011. Characterization of New Zealand obsidian using PXRF. J. Archaeol. Sci. 38, 45–56. https://doi.org/10.1016/j.jas.2010.08.007.
Sinnesael, M., de Winter, N.J.N.J., Snoeck, C., Montanari, A., Claeys, P., 2018. An integrated pelagic carbonate multi-proxy study using portable X-ray fluorescence (pXRF): Maastrichtian strata from the Bottaccione Gorge, Gubbio, Italy. Cretac. Res. 91, 20–32. https://doi.org/10.1016/j.cretres.2018.04.010.
Steiner, A.E.E., Conrey, R.M.M., Wolff, J.A.A., 2017. PXRF calibrations for volcanic rocks and the application of in-field analysis to the geosciences. Chem. Geol. 453, 35–54. https://doi.org/10.1016/j.chemgeo.2017.01.023.
Stewart, E.K., Mauk, J.L., 2017. Sedimentology, sequence-stratigraphy, and geochemical variations in the Mesoproterozoic Nonesuch Formation, northern Wisconsin, USA. Precambrian Res. 294, 111–132. https://doi.org/10.1016/j.precamres.2017.03.023.
Tertian, R., 1986. Mathematical matrix correction procedures for x-ray fluorescence analysis. A critical survey. X-Ray Spectrom. 15, 177–190. https://doi.org/10.1002/xrs.1300150307.
Tjallingii, R., Röhl, U., Kölling, M., Bickert, T., 2007. Influence of the water content on X-ray fluorescence core-scanning measurements in soft marine sediments. Geochem. Geophys. Geosyst. 8 https://doi.org/10.1029/2006GC001393.
Triantafyllou, A., Mattielli, N., Clerbois, S., Da Silva, A.C., Kaskes, P., Claeys, P., Devleeschouwer, X., Brkojewitsch, G., 2021. Optimizing multiple non-invasive techniques (PXRF, pMS, IA) to characterize coarse-grained igneous rocks used as building stones. J. Archaeol. Sci. 129, 105376 https://doi.org/10.1016/j.jas.2021.105376.
Turner, B.W., Tréanton, J.A., Slatt, R.M., 2016. The use of chemostratigraphy to refine ambiguous sequence stratigraphic correlations in marine mudrocks. An example from the Woodford Shale, Oklahoma, USA. J. Geol. Soc. Lond. 173, 854–868. https://doi.org/10.1144/jgs2015-125.
VanCott, R.J., McDonald, B.J., Seelos, A.G., 1999. Standard soil sample preparation error and comparison of portable XRF to laboratory AA analytical results. Nucl. Inst. Methods Phys. Res. Sect. A Accel. Spectro. Detect. Assoc. Equip. 422, 801–804. https://doi.org/10.1016/S0168-9002(98)01000-6.
Weltje, G., Bloemsma, M., Tjallingii, R., Heslop, D., Röhl, U., Croudace, I., 2015. Prediction of Geochemical Composition from XRF Core Scanner Data: A New Multivariate Approach Including Automatic selection of Calibration Samples and Quantification of Uncertainties. In: Croudace, I., R.R (Eds.), Of Sediment Cores. Developments in Paleoenvironmental Research, vol. 17, pp. 507–534. https://doi.org/10.1007/978-94-017-9849-5_21. Dordrecht.
Weltje, G.J., Tjallingii, R., 2008. Calibration of XRF core scanners for quantitative geochemical logging of sediment cores: Theory and application. Earth Planet. Sci. Lett. 274, 423–438. https://doi.org/10.1016/j.epsl.2008.07.054.
de Winter, Niels J., Sinnesael, M., Makarona, C., Vansteenberge, S., Claeys, P., 2017. Trace element analyses of carbonates using portable and micro-X-ray fluorescence: performance and optimization of measurement parameters and strategies. J. Anal. At. Spectrom. 32, 1211–1223. https://doi.org/10.1039/C6JA00361C.
Young, K.E., Evans, C.A., Hodges, K.V., Bleacher, J.E., Graff, T.G., 2016. A review of the handheld X-ray fluorescence spectrometer as a tool for field geologic investigations on Earth and in planetary surface exploration. Appl. Geochem. 72, 77–87. https://doi.org/10.1016/j.apgeochem.2016.07.003.