



Garnet major-element composition as an indicator of host-rock type: a machine learning approach using the random forest classifier

Jan Schönig¹ · Hilmar von Eynatten¹ · Raimon Tolosana-Delgado² · Guido Meinhold^{1,3,4}

Received: 2 July 2021 / Accepted: 13 October 2021
© The Author(s) 2021

Abstract

The major-element chemical composition of garnet provides valuable petrogenetic information, particularly in metamorphic rocks. When facing detrital garnet, information about the bulk-rock composition and mineral paragenesis of the initial garnet-bearing host-rock is absent. This prevents the application of chemical thermo-barometric techniques and calls for quantitative empirical approaches. Here we present a garnet host-rock discrimination scheme that is based on a random forest machine-learning algorithm trained on a large dataset of 13,615 chemical analyses of garnet that covers a wide variety of garnet-bearing lithologies. Considering the out-of-bag error, the scheme correctly predicts the original garnet host-rock in (i) > 95% concerning the setting, that is either mantle, metamorphic, igneous, or metasomatic; (ii) > 84% concerning the metamorphic facies, that is either blueschist/greenschist, amphibolite, granulite, or eclogite/ultrahigh-pressure; and (iii) > 93% concerning the host-rock bulk composition, that is either intermediate–felsic/metasedimentary, mafic, ultramafic, alkaline, or calc–silicate. The wide coverage of potential host rocks, the detailed prediction classes, the high discrimination rates, and the successfully tested real-case applications demonstrate that the introduced scheme overcomes many issues related to previous schemes. This highlights the potential of transferring the applied discrimination strategy to the broad range of detrital minerals beyond garnet. For easy and quick usage, a freely accessible web app is provided that guides the user in five steps from garnet composition to prediction results including data visualization.

Keywords Garnet major-element composition · Database · Host-rock discrimination · Machine-learning · Provenance · Web app

Introduction

Garnet is one of the most useful minerals in Earth sciences providing key information on mantle, metamorphic, metasomatic, and igneous processes (e.g., Baxter et al. 2013).

Due to its widespread occurrence and its wide compositional range that mainly depends on pressure, temperature, and host-rock composition, garnet major-element chemistry became a well-established tool in sedimentary provenance analysis (e.g., Krippner et al. 2014) and economic exploration campaigns (e.g., Hardman et al. 2018). Apart from discriminating different source regions, the potential capability of extracting petrogenetic host-rock information in terms of geological setting, metamorphic conditions, and composition by solely considering garnet single grains is of particular significance in interdisciplinary research linking sedimentary, metamorphic, tectonic, and geodynamic processes (e.g., Schönig et al. 2018a).

For instance, our knowledge about the evolution of metamorphism and plate tectonic regimes through time is mainly based on the preserved crystalline rock record (e.g., Brown and Johnson 2018; Holder et al. 2019), which is increasingly incomplete with increasing age (e.g., Goodwin 1996). To overcome this issue, detrital zircon is commonly used

Communicated by Daniela Rubatto.

✉ Jan Schönig
jan.schoenig@uni-goettingen.de

¹ Geoscience Centre Göttingen, University of Göttingen, Goldschmidtstraße 3, 37077 Göttingen, Germany

² Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resources Technology, Chemnitz Straße 40, 09599 Freiberg, Germany

³ School of Geography, Geology and the Environment, Keele University, Keele, Staffordshire ST5 5BG, UK

⁴ Institute for Geology, TU Bergakademie Freiberg, Bernhard-von-Cotta-Straße 2, 09599 Freiberg, Germany

and has been shown to provide important information on eroded crustal rocks, enabling a more reliable global-scale reconstruction of continental growth and igneous suites throughout Earth history (e.g., Dhuime et al. 2017, and references therein). Similar approaches to reconstruct metamorphic conditions and/or plate tectonic regimes via the detrital record are absent, but garnet major-element chemistry provides a potential tool to tackle this issue. However, to link detrital garnet composition with petrogenetic host-rock information, a robust statistical model with reliable prediction success is required.

Since the first application of detrital garnet major-element chemistry to constrain different source regions (Morton 1985), several petrogenetic schemes have been developed to discriminate specific garnet host rocks (Teraoka et al. 1998; Schulze 2003; Mange and Morton 2007; Aubrecht et al. 2009; Suggate and Hall 2014; Hardman et al. 2018; Tolosana-Delgado et al. 2018). Until 2018, all schemes used strict and subjective compositional fields to discriminate garnet source rocks by considering a subset of available variables. Krippner et al. (2014) and Suggate and Hall (2014) demonstrated that this approach leads to high misclassification rates due to large compositional overlaps of garnets from various source rocks. Notably, although garnet growth stages are often preserved by compositional zoning (e.g., Tracy et al. 1976; Tracy 1982), the composition does (i) not always pinpoint unique growth conditions (e.g., Lanari et al. 2017), (ii) may be influenced by progressive fractionation of the reactive bulk-rock composition (e.g., Lanari and Engi 2017), and (iii) may be modified by post-growth diffusion at high temperatures (e.g., Caddick et al. 2010).

Recently, the usage of robust multivariate statistics has been shown to significantly improve classification results (Hardman et al. 2018; Tolosana-Delgado et al. 2018). While Hardman et al. (2018) focus on the discrimination of mantle-versus-crustal garnet, Tolosana-Delgado et al. (2018) consider five major garnet host-rock types including amphibolite-, granulite-, and eclogite-facies metamorphic rocks, igneous rocks, and ultramafic rocks. The capabilities of extracting petrogenetic information from the detrital garnet record have further been improved by considering trace elements (Čopjaková et al. 2005; Hong et al. 2020), mineral inclusions (e.g., Schönig et al. 2019, 2020; Baldwin et al. 2021), U–Pb geochronology (e.g., Seman et al. 2017; Millonig et al. 2020), as well as Sm–Nd geochronology (Maneiro et al. 2019). However, all these novel techniques require a wealth of experience, caution, equipment, and effort. In contrast, major-element chemistry is routinely applied and enables to efficiently screen a statistically significant number of grains. Although multivariate schemes are already highly advanced, these may be improved by (i) enlarging the database, (ii) including previously unconsidered host-rock types, (iii) considering host-rock composition,

(iv) refining prediction classes, and (v) applying statistical classification methods with enough flexibility to disentangle the strongly overlapping compositional signatures of garnet types. Particularly machine-learning algorithms are suitable to tackle multivariate discrimination tasks as demonstrated by prediction models that are based on bulk-rock chemistry (Petrelli and Perugini 2016; Petrelli et al. 2017; Ren et al. 2019), glass composition (Bolton et al. 2020), as well as single-grain chemistry (Han et al. 2019; Itano et al. 2020). In addition, random forest regression led to an improvement of barometric predictions for majoritic garnet inclusions in diamond that crystallised at $P > 5$ GPa (Thomson et al. 2021), implying that the algorithm may prove suitable for developing a model to discriminate detrital garnet sourced from a potentially broad range of host rocks.

Here we present a new garnet major-element discrimination scheme for host-rock prediction developed by a machine learning approach using the random forest algorithm in classification mode (Breiman 2001). It is based on a large database compiled from the literature ($n = 13,615$), which was used to train and test the discrimination model simultaneously. The new scheme enables garnet discrimination into host-rock setting (mantle versus metamorphic versus igneous versus metasomatic), metamorphic facies (blueschist/greenschist versus amphibolite versus granulite versus eclogite/ultrahigh-pressure), and composition (intermediate–felsic/metasedimentary versus mafic versus ultramafic versus alkaline versus calc–silicates). Besides providing a more detailed classification than any previous scheme, we highlight the usefulness of our scheme by demonstrating the much higher discrimination success and interpretability of results compared to others. We introduce a freely accessible web application, allowing users to easily apply the discrimination scheme to their own data without the need for programming expertise. Finally, application examples emphasize the potential to reflect (i) varying pressure–temperature conditions during garnet growth, (ii) the geological framework of catchments, and (iii) provenance shifts through time.

Database

The compiled chemical garnet database (Supplementary Information 1, <https://rodare.hzdr.de/record/1220>) includes 13,615 observations of eight oxides commonly analysed in lab routines: SiO₂, TiO₂, Al₂O₃, Cr₂O₃, FeO_{total}, MnO, MgO, and CaO (in wt%). Data compilation results from a comprehensive literature survey and benefitted from several published databases, most importantly to mention are Grütter et al. (2004), Krippner et al. (2014), Suggate and Hall (2014), and Hardman et al. (2018). Original references have been cross-checked to promote database quality. If any of the

eight oxides is not included on any particular observation, that value is handled as not available (NA). As most standard routines use a detection limit between 0.02 and 0.03 wt%, we chose 0.03 wt% as the threshold and all values below are handled as below detection limit (BDL).

Garnets have been subdivided according to host-rock type (main group), metamorphic facies (group), and composition (subgroup, Table 1). Host-rock types include mantle (MA), metamorphic (MM), igneous (IG), and metasomatic rocks (MS). For metamorphic facies, estimated pressure–temperature (P – T) conditions in the original references have been used instead of facies-specific mineral assemblages. The assignment follows the facies scheme after Bucher and Frey (2002) for protoliths of mid-oceanic ridge basalt composition. This comprises greenschist facies (GS), blueschist facies (BS), amphibolite facies (AM), granulite facies (GR), eclogite facies (EC), and ultrahigh-pressure metamorphism (UHP). In case geo-thermo-barometry was not applied in the original literature, datasets were only used when rock descriptions and reported facies are conclusive. Host-rock compositions were simplified into ultramafic (UM), mafic (M), intermediate–felsic/metasedimentary (IF/S, mafic meta-sediments omitted), calc–silicate (CS), and alkaline (A).

Model development

For developing the garnet discrimination scheme, the random forest machine-learning algorithm (Breiman 2001) is applied in classification mode. A description of the principle of creating a random forest classification model is given in Supplementary Information 2. For a more detailed and mathematically based explanation, the reader is referred to the original work of Breiman (2001) and reviews treating this topic (Boulesteix et al. 2012; Ziegler and König 2014; Belgiu and Drăgut 2016; Biau and Scornet 2016).

Data processing, calculations, and plotting were performed using the statistic software R (R Core Team 2020). Used packages include ‘compositions’ (van den Boogart and Tolosana-Delgado 2008) for compositional data analysis, ‘dplyr’ (Wickham et al. 2020) for data wrangling, ‘ggtern’ (Hamilton and Ferry 2018) for display, ‘magittr’ (Bache and Wickham 2014) for readability of complex code, and ‘randomForest’ (Liaw and Wiener 2002) for the calculations of the forest itself.

Two models are developed called ‘setting and metamorphic facies’ and ‘composition’. For the ‘setting and metamorphic facies’ model, main groups and groups of Table 1 have partially been merged into seven classes, namely garnet of (i) mantle rocks (class MA, includes main group MA), (ii) blueschist-/greenschist-facies metamorphic rocks (class BS/GS, includes groups BS and GS), (iii) amphibolite-facies

Table 1 Number of observations for individual garnet host-rock types (main group), metamorphic facies (group), and composition (subgroup) included in the database

Main group	Observations (main group)
Group	Observations (group)
Subgroup	Observations (subgroup)
MA: mantle rocks	3439
<i>M: mafic</i>	1662
<i>UM: ultramafic</i>	1777
MM: metamorphic rocks	8276
GS: greenschist facies	786
<i>M: mafic</i>	352
<i>F/S: intermediate–felsic/metasedimentary</i>	434
BS: blueschist facies	418
<i>M: mafic</i>	200
<i>IF/S: intermediate–felsic/metasedimentary</i>	213
<i>CS: calc–silicate</i>	5
AM: amphibolite facies	2605
<i>M: mafic</i>	824
<i>IF/S: intermediate–felsic/metasedimentary</i>	1767
<i>S: calc–silicate</i>	14
GR: granulite facies	1011
<i>M: mafic</i>	180
<i>IF/S: intermediate–felsic/metasedimentary</i>	672
<i>CS: calc–silicate</i>	159
EC: eclogite facies	2095
<i>M: mafic</i>	1813
<i>IF/S: intermediate–felsic/metasedimentary</i>	274
<i>CS: calc–silicate</i>	8
UHP: ultrahigh-pressure metamorphism	1361
<i>M: mafic</i>	882
<i>IF/S: intermediate–felsic/metasedimentary</i>	295
<i>CS: calc–silicate</i>	184
IG: igneous rocks	1074
<i>M: mafic</i>	66
<i>IF: intermediate–felsic</i>	919
<i>A: alkaline</i>	89
MS: metasomatic rocks	826
<i>CS: calc–silicate</i>	826

metamorphic rocks (class AM, includes group AM), (iv) granulite-facies metamorphic rocks (class GR, includes group GR), (v) eclogite-/ultrahigh-pressure-facies metamorphic rocks (class EC/UHP, includes groups EC and UHP),

(vi) igneous rocks (class IG, includes main group IG), and (vii) metasomatic rocks (class MS, includes main group MS) (Table 2). For the ‘composition’ model, subgroups of Table 1 have partially been merged into five classes, namely garnet of (i) ultramafic rocks (class UM, includes subgroup MA-UM), (ii) mafic rocks (class M, includes subgroups MA-M, GS-M, BS-M, AM-M, GR-M, EC-M, UHP-M, and IG-M), (iii) intermediate–felsic/metasedimentary rocks (class IF/S, includes subgroups GS-IF/S, BS-IF/S, AM-IF/S, GR-IF/S, EC-IF/S, UHP-IF/S, and IG-IF), (iv) alkaline rocks (class A, includes subgroup IG-A), and (v) calc–silicates (class CS, includes subgroups BS-CS, AM-CS, GR-CS, EC-CS, UHP-CS, and MS-CS) (Table 2).

Each observation in the database includes eight variables in form of oxide wt% (see “Database” section). These observations have been acquired by many operators since major-element analysis has become a standard analytical tool. Thus, observations in the database include a wide range of used systems, analysed oxides, calibrations, operating conditions, and data processing techniques. This mainly results in discrepancies in the total wt% of the eight oxides considered for discrimination, making the amount of the whole (or total sum by sample) a non-informative quantity. In addition, all chemical components are rarely analysed, but rather the amount of each component is limited to that whole. Thus, only relative changes are relevant (e.g., van den Boogart and Tolosana-Delgado 2008). To tackle this issue and to get rid of spurious anti-correlations (Chayes 1960), the natural logarithms of ratios between all pairs of oxides are used as variables instead of the single values (Aitchison 1986). The usage of log-ratios is a mathematically elegant transformation that enables the use of standard unconstrained multivariate statistics (Aitchison and Egozcue 2005). This approach increases the total number of variables from eight oxides measured to 28 pairwise log-ratios. Besides the advantages, this introduces a difficulty when handling values BDL, that is < 0.03 wt%. The log-ratios with values BDL in the numerator and/or denominator can potentially span a wide range of values. Considering a log-ratio that

contains a value BDL in the numerator with a detection limit DL, the log-ratio of numerator and a denominator with value x is always $< \ln(DL \times x^{-1})$. Conversely, a value BDL in the denominator with a detection limit DL leads to a log-ratio $> \ln(x \times DL^{-1})$ with x the value of the numerator. Thus, log-ratios with a value BDL in the numerator or denominator have been replaced by the minimum of that log-ratio in the database minus one or the maximum plus one, respectively (see pair-wise log-ratio function in Supplementary Information 3). This approach ensures that values BDL are treated in the same way than fully observed values, while maintaining and making use of this information. Those log-ratios with values BDL both in the numerator and denominator are replaced by NA. All missing values (those that involve an NA and those that involve two BDLs) are treated by the ‘na.roughfix’ function, which first replaces missing values by the median of not-missing values, trains a random forest model, computes the proximity matrix between samples, and refines the missing values by replacing it with the weighted median of non-missing values (using the proximity values as weights).

The method parameters of the random forest were chosen in a double procedure. The ‘sampsiz’ parameter was set by a discretionary approach. This parameter controls the number of observations taken from each class for each tree by random sampling. As random forest is optimized for creating discrimination models that have the highest overall classification success, classes that include more observations are often better classified compared to those where fewer observations are available (Supplementary Information 2). The aim of the garnet discrimination model was to balance the classification success rates for the individual classes as well as possible. Therefore, the ‘sampsiz’ for classes containing a higher number of observations was reduced (Table 2). This results in a slightly lower classification success for the entire database, but more balanced success rates for the individual classes (Chen et al. 2004). The other parameters were set by formal exhaustive cross-validation. Both models have been computed by five iterations with all combinations of ‘ntree’ between 200 and 6,000 (step size = 200), ‘mtry’ between 1

Table 2 Classes and parameters of the random forest garnet discrimination models

‘Setting and metamorphic facies’ (ntree = 3400; mtry = 5; nodesize = 1)							
Classes	MA	BS/GS	AM	GR	EC/UHP	IG	MS
Observations	3439	1204	2605	1011	3456	1074	826
Sampsiz	1011	1011	1200	1011	1100	1011	826
‘Composition’ (ntree = 3200; mtry = 6; nodesize = 1)							
Classes	UM	M	IF/S	A	CS		
Observations	1777	5979	4574	89	1196		
Sampsiz	1777	2600	2100	89	1196		

and 12 (step size = 1), and ‘nodesize’ between 1 and 3 (step size = 1). Those parameter values giving on average the lowest OOB (out-of-bag) error have been chosen for the final models (Table 2).

Model performance

Performance of the ‘setting and metamorphic facies’ model

The ‘setting and metamorphic facies’ model predicts the correct class out of seven classes for > 88% of all observations included in the database based on the OOB error (Table 3). Notably, the predicted classes are more detailed and useful in terms of petrogenetic information compared to the most frequently used scheme after Mange and Morton (2007), which includes prediction classes that do not point to a specific host-rock type (Supplementary Information 2).

To take full and quick advantage of the classification regarding provenance, the voting results are shown in two separate graphical schemes for ‘setting’ and ‘metamorphic facies’, each representing the four ternary sides of a tetrahedron (Fig. 1). The ‘setting’ scheme discriminates garnet sourced from MA, IG, MS, and metamorphic rocks (MM) based on the votes for each class. MM is represented by the maximum vote of the four metamorphic classes, that is BS/GS, AM, GR, and EC/UHP. The decision of taking the maximum vote instead of the sum of votes is based on two major points. First, taking the sum of votes artificially introduces classification results that are based on a much larger ‘sampsizе’ for training the random forest model for MM (1011 + 1200 + 1011 + 1100 = 4322) compared to MA (1011), IG (1011), and MS (826, see Table 2). Consequently, the balancing introduced by making use of the ‘sampsizе’ argument (see “Model development” section) is getting out of balance, resulting in higher classification success rates for MM at the expense of the three other classes. Second, taking the sum of votes entangles the two different questions

to be answered by a single classification model, which is impermissible.

Votes of the ‘setting’ scheme are shown as kernel density maps in four ternary plots to represent each setting class (Fig. 1a). The voting result of each individual observation from the database is solely plotted in one of the four ternary plots, that is, the plot representing the three highest votes. The vast majority of the votes plots close to the apexes of the corresponding correct class. Only very minor overlaps occur for MM versus MA, MM versus IG, and MM versus MS. This is obvious in the corresponding majority vote bar plot showing that on average > 95% are assigned to their correct class (Fig. 1a).

Considering solely the discrimination of MA versus crustal garnet (MM, IG, and MS), MA are correctly classified in 96% and crustal garnet in 99%, giving a class average of 97%. Thus, the prediction success excels the graphical mantle-versus-crustal garnet discrimination after Hardman et al. (2018), that is 95% based on the presented database (Supplementary Information 2). In the developed model, MA-M is the only mantle subgroup exceeding 2% of the observations misclassified as MM, and EC-M as well as UHP-M are the only crustal subgroups exceeding 2% of the observations misclassified as MA (Fig. 1b). However, even these challenging groups show high success rates: 93% for MA-M, 97% for EC-M, and 96% for UHP-M.

Garnets of class MS are correctly identified in > 97% and only minor amounts (< 3%) are misclassified as MM (Fig. 1a, b). Contrary, metamorphic calc–silicates that formed at high temperature, which are subgroups AM-CS and GR-CS, are often misclassified as MS. Besides their chemical similarity, this is caused by the underrepresentation of observations from metamorphic calc–silicates compared to mafic and intermediate–felsic/metasedimentary host-rock compositions, in particular for AM (Table 1). In addition to calc–silicates, some of the alkaline igneous garnets (7%) are misclassified as MS.

IG garnet is correctly classified in > 93% (Fig. 1a). The highest success rates are given for subgroups IG-IF and IG-A with 94% (6% misclassified as MM) and 93% (7%

Table 3 Classification rates and confusion matrix of the ‘setting and metamorphic facies’ random forest model based on the out-of-bag estimate

		Correct classification: 88.4%						
		True class						
		MA	BS/GS	AM	GR	EC/UHP	IG	MS
Predicted class	MA	3292	0	0	3	90	0	0
	BS/GS	0	1047	139	5	147	13	3
	AM	1	86	2166	29	165	28	0
	GR	4	4	91	839	136	13	13
	EC/UHP	142	46	122	100	2890	13	4
	IG	0	12	74	10	24	1001	4
	MS	0	9	13	25	4	6	802
Correct classification		95.7%	87.0%	83.1%	83.0%	83.6%	93.2%	97.1%

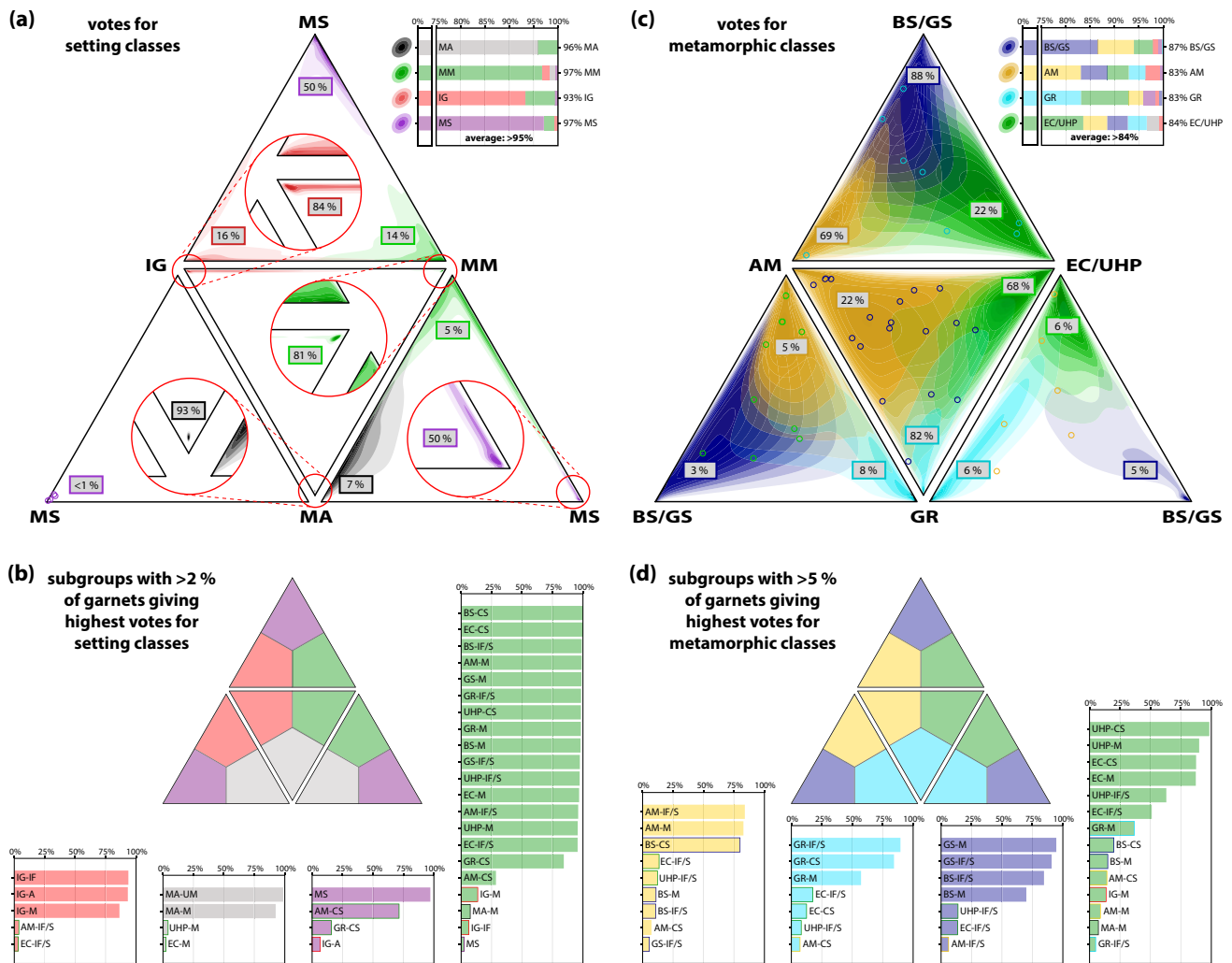


Fig. 1 Performance of the ‘setting and metamorphic facies’ model based on the out-of-bag error. The distribution of votes for ‘setting’ classes (a) and ‘metamorphic facies’ classes (c) are shown as kernel density estimate maps in four ternary diagrams that represent the sides of a tetrahedron: *MS* metasomatic rocks, *BS/GS* blueschist/green-schist facies, *AM* amphibolite facies, *GR* granulite facies, and *EC/UHP* eclogite/ultrahigh-pressure facies. Individual observations are included only in the ternary diagram that represents the three highest votes. Note that *MM* represents the maximum vote for *BS/GS*, *AM*, *GR*, or *EC/UHP*. Kernel density bandwidth calculated after Venables

and Ripley (2002). Corresponding classification success rates based on the majority vote are shown as barplots (note break in scale at 75%). Proportions of observations from each class occurring in the individual ternary diagrams are given in small rectangular boxes within the ternary diagrams of (a) and (c). The proportions of observations from individual subgroups assigned to the prediction classes are shown as barplots for the ‘setting’ (b) and ‘metamorphic facies’ (d) scheme (see Table 1 for abbreviations). Bars representing misclassifying votes are highlighted by frames that are colour coded according to the true class

misclassified as *MS*), respectively. *IG-M* shows a slightly lower success with 86%, which still represents a well-discriminated subgroup. Noteworthy, very low amounts of garnet from other subgroups are misclassified as *IG*, with *AM-IF/S* and *EC-IF/S* being the only subgroups exceeding 2% misclassification as *IG* (Fig. 1b).

Garnets of the setting class *MM* are correctly classified in >96% (Fig. 1a), with 15 of the 17 metamorphic subgroups listed in Table 1 being correctly assigned in >95%.

Subgroups showing lower success rates are restricted to calc-silicates (*GR-CS* and *AM-CS*; Fig. 1b).

Votes of the ‘metamorphic facies’ scheme are as well shown as kernel density maps in four ternary plots to represent each metamorphic class (Fig. 1c). The plot includes all observations giving the highest votes for *MM* in the ‘setting’ scheme. The voting result of each individual observation is solely plotted in the ternary diagram representing the three highest votes. Compared to the ‘setting’ scheme, the spread

in votes and overlaps are more pronounced but the maxima are clearly located close to the apexes of the corresponding class. The barplot in Fig. 1c shows that classification success rates of the individual classes range from 83 to 87%, giving an average of > 84% correctly classified garnets. Beyond the additional and separate prediction of garnet sourced from BS/GS, classification clearly improved compared to the scheme of Tolosana-Delgado et al. (2018), which shows an average classification success rate of 65% for metamorphic classes (Supplementary Information 2).

The barplots in Fig. 1d show that (i) the correct subgroups constitute the vast majority of observations assigned to the individual metamorphic classes, except for several of the CS subgroups; (ii) garnet from IF/S host-rock composition is better classified for GR, AM, and BS/GS subgroups than for EC/UHP subgroups and vice versa; and (iii) misclassifications are mainly restricted to adjacent classes in P – T space, that is BS/GS and GR with AM and EC/UHP, respectively, while AM and EC/UHP share borders with all other classes (see also barplot in Fig. 1c). Points (i) and (ii) can be related to the distribution of observations from the individual subgroups (Table 1), but point (iii) implies a rather continuous change in garnet composition with changing P – T conditions, which is well known from metamorphic petrology and reflected by the votes (Supplementary Information 2).

Performance of the ‘composition’ model

The ‘composition’ model predicts the correct class out of five for > 92% of all observations included in the database based on the OOB error estimate (Table 4). It discriminates garnet from intermediate–felsic/metasedimentary (IF/S), mafic (M), ultramafic (UM), calc–silicates (CS), and alkaline (A) host rocks.

Votes of the ‘composition’ scheme are shown as kernel density maps in four ternary plots to represent four of the five classes (Fig. 2a). Because alkaline garnet has the lowest number of observations, is rarely misclassified, and misclassifications are restricted to class CS (Table 4), class A is excluded from the plots. Thus, plots include all observations giving not the highest votes for class A. As in the preceding plots, voting results for each individual observation from the

database is solely plotted in one of the four ternary plots, which is the plot representing the three highest votes. The vast majority of the votes plots close to the apexes of the corresponding correct class. However, overlaps occur for M versus IF/S and M versus UM. This is highlighted in the corresponding majority vote bar plot showing that on average > 93% are assigned to their correct class (Fig. 2a).

The barplots in Fig. 2b show that the correct subgroups constitute the vast majority of observations assigned to the individual composition classes. In particular, garnet classified as CS and UM are well represented and only scarcely garnet of subgroups IG-A and MA-M receive the highest votes for CS or UM, respectively. Garnets classified as IF/S and M are dominated by the correct subgroups, too. However, up to almost 16% of the M subgroups are assigned to IF/S and IF/S subgroups to M. Conspicuously, EC-IF/S (50%), BS-M (50%), and BS-CS (100%, 60% in M, 40% in IF/S) show high misclassifications.

Understanding the models

The two presented garnet discrimination models consist of 3400 and 3200 trees (Table 2), each based on a different bootstrapped random sample and deeply grown without pruning leading to between 1619 and 1949 decision nodes for each tree of the ‘setting and metamorphic facies’ model and between 1033 and 1269 nodes for the ‘composition’ model. Thus, following the decision process in detail is not feasible, giving random forest models a kind of ‘black box’ character.

To understand the basic discrimination decisions performed by the models, the importance of variables is explored in three ways: (i) Considering the mean decrease in Gini impurity (Supplementary Information 2) for individual variables, that is the weighted average of the decrease in Gini impurity between parent and child nodes in all trees of the trained forest when the values of an individual variable are permuted; (ii) considering the mean decrease in accuracy for individual variables, that is the decrease of prediction accuracy when an individual variable is removed from the OOB test set; and (iii) considering the increase in

Table 4 Classification rates and confusion matrix of the ‘composition’ random forest model based on the out-of-bag estimate

		Correct classification: 92.3%				
		True class				
		IF/S	M	UM	CS	A
Predicted class	IF/S	4209	340	1	14	0
	M	353	5480	145	10	0
	UM	1	145	1626	0	0
	CS	11	14	5	1167	6
	A	0	0	0	5	83
Correct classification		92.0%	91.7%	91.5%	97.6%	93.3%

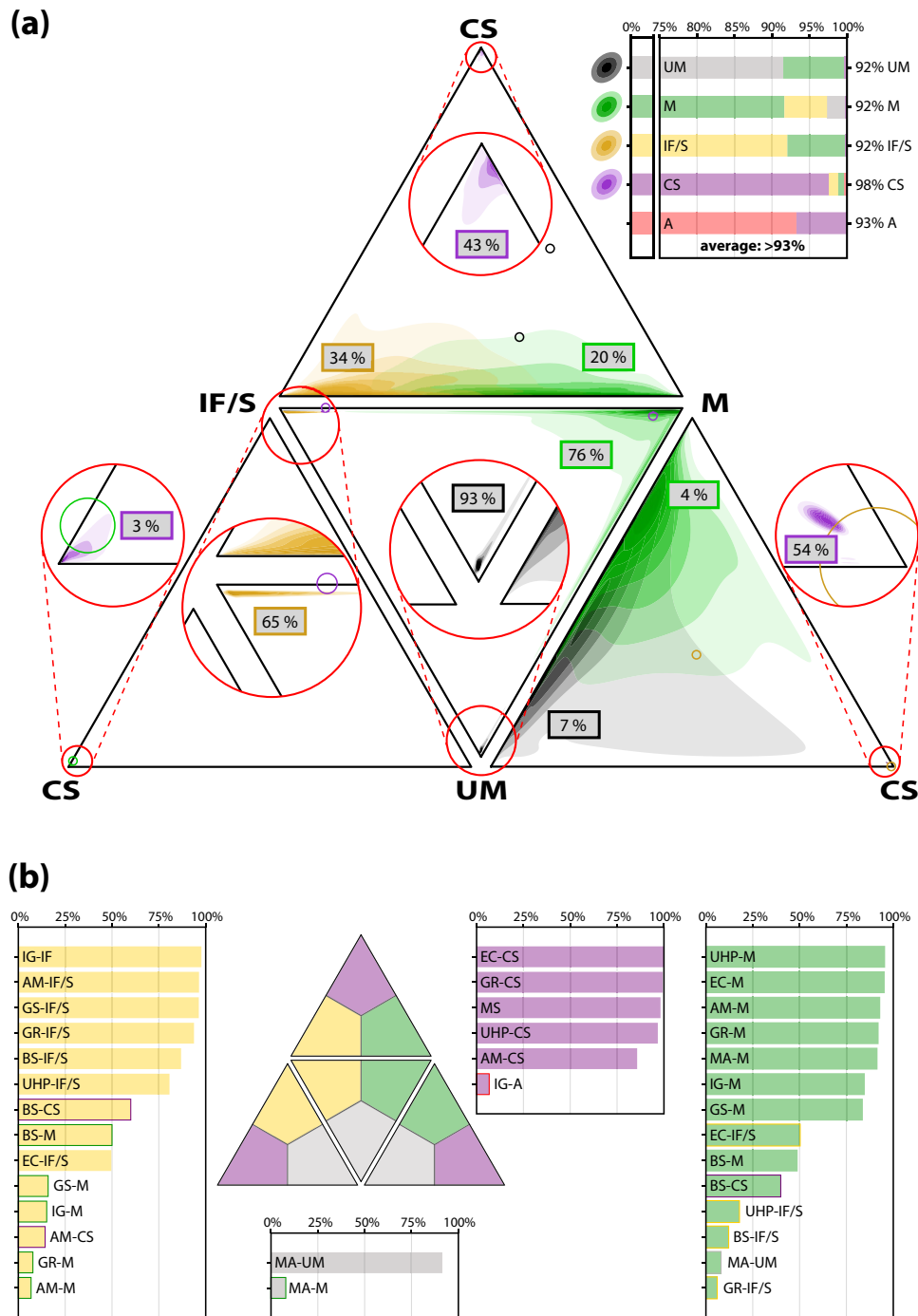


Fig. 2 Performance of the ‘composition’ model based on the out-of-bag error. **a** Distribution of votes shown as kernel density estimate maps in four ternary diagrams that represent the sites of a tetrahedron: *CS* calc–silicates, *IF/S* intermediate–felsic/metasedimentary, *UM* ultramafic, and *M* mafic. Votes for alkaline (*A*) composition are not shown. Individual observations are included only in the ternary diagram that represents the three highest votes. Kernel density bandwidth calculated after Venables and Ripley (2002). Corresponding

classification success rates based on the majority vote are shown as barplots (note break in scale at 75%). Proportions of observations from each class occurring in the individual ternary diagrams are given in small rectangular boxes. **b** Proportions of observations from individual subgroups assigned to the prediction classes shown as barplots (see Table 1 for abbreviations). Bars that represent misclassifying votes are highlighted by colour coded frames according to the true class

misclassification rates for individual classes when individual oxides are removed from the database for training the forest. Note that the exclusion of each oxide means that all seven log-ratios that include this oxide are removed. Finally, the most important variables to discriminate individual classes and subgroups are retraced and their origin is discussed.

Variable importance of the ‘setting and metamorphic facies’ model

The highest variable importance based on the mean decrease in Gini impurity and accuracy is given by log-ratios involving MgO. This particularly includes $\ln(\text{FeO}_{\text{total}} \times \text{MgO}^{-1})$, followed by $\ln(\text{SiO}_2 \times \text{MgO}^{-1})$ and $\ln(\text{Al}_2\text{O}_3 \times \text{MgO}^{-1})$ (Fig. 3a). All three show a high spread for observations of the dataset with distinct ranges for the individual classes (Fig. 3b). Notably, values of these ratios decrease from metasomatic garnet (MS), over low-temperature metamorphic (BS/GS and AM) and igneous garnet (IG), to high-grade metamorphic garnet (GR and EC/UHP), and mantle garnet (MA) shows the lowest values.

The next important set of log-ratios includes CaO versus Al_2O_3 , SiO_2 , MgO, MnO, and $\text{FeO}_{\text{total}}$ (Fig. 3a). This is dominated by the importance of discriminating class MS which shows higher CaO values than the other classes (Fig. 3b). Another major difference between class MS and all other classes is the high range of values for $\ln(\text{SiO}_2 \times \text{Al}_2\text{O}_3^{-1})$ for MS compared to the very tight range of the other classes, which is also visible in the values for $\ln(\text{SiO}_2 \times \text{FeO}_{\text{total}}^{-1})$ and $\ln(\text{Al}_2\text{O}_3 \times \text{FeO}_{\text{total}}^{-1})$. In contrast to class MS, garnets of class GR and to a lesser extent IG show high values for log-ratios including CaO in the denominator, even higher than those for the other metamorphic classes, making log-ratios including CaO useful to separate garnet of classes GR and IG. Besides MS and GR, $\ln(\text{MgO} \times \text{CaO}^{-1})$ is particularly important for the discrimination of IG versus BS/GS, which are in most other respects compositionally similar (Fig. 3b).

Behind the log-ratios mainly dictated by MgO and CaO along with $\text{FeO}_{\text{total}}$, a set of log-ratios including MnO becomes next most important (Fig. 3a). Remarkably, values for log-ratios with MnO in the denominator versus MgO, Al_2O_3 , SiO_2 , and $\text{FeO}_{\text{total}}$ show a reverse trend to log-ratios with MgO in the denominator (Fig. 3b). Values successively increase from lower-temperature metamorphic and igneous garnet, over higher-temperature metamorphic garnet, to high-grade metamorphic garnet, and finally mantle garnet. Distinct is the behaviour of metasomatic garnet showing higher similarities with low-grade metamorphic garnet for $\ln(\text{MnO} \times \text{MgO}^{-1})$ and $\ln(\text{Al}_2\text{O}_3 \times \text{MnO}^{-1})$, and higher similarities with high-grade metamorphic garnet for $\ln(\text{SiO}_2 \times \text{MnO}^{-1})$ and $\ln(\text{FeO}_{\text{total}} \times \text{MnO}^{-1})$.

Considering the mean decrease in Gini impurity and accuracy of the entire tree ensemble, the individual

log-ratios including TiO_2 or Cr_2O_3 are placed in a subordinate role for creating high-purity splits. However, some specific log-ratios are important for splitting individual classes. In particular, $\ln(\text{TiO}_2 \times \text{CaO}^{-1})$ is essential for purifying the discrimination of class IG versus metamorphic classes. The variables $\ln(\text{Cr}_2\text{O}_3 \times \text{FeO}_{\text{total}}^{-1})$, $\ln(\text{Cr}_2\text{O}_3 \times \text{MnO}^{-1})$, $\ln(\text{Cr}_2\text{O}_3 \times \text{CaO}^{-1})$, $\ln(\text{TiO}_2 \times \text{FeO}_{\text{total}}^{-1})$ and $\ln(\text{TiO}_2 \times \text{MnO}^{-1})$ are important for discriminating MA versus EC/UHP (Fig. 3).

In addition to the exploration of variable importance based on the decrease in Gini impurity and accuracy, Fig. 4 shows the implications for misclassification rates when individual oxides are excluded for the development of a random forest ‘setting and metamorphic facies’ discrimination model. Misclassification rates for all classes show the highest increase when MnO is excluded, followed by TiO_2 (strongly influencing the correct classification of IG), MgO (with notable contribution to the quality of BS/GS and MS reclassification), CaO, Al_2O_3 , $\text{FeO}_{\text{total}}$, SiO_2 , and Cr_2O_3 in decreasing order of global importance (Fig. 4, bold grey line). The higher importance of MnO and TiO_2 compared to MgO, CaO, and $\text{FeO}_{\text{total}}$ is contrary to the importance order based on the mean decrease in Gini impurity (Fig. 3). This indicates that MnO and TiO_2 either gain their importance by considering many log-ratios including MnO and TiO_2 or by being specifically important for the discrimination of individual classes, or a combination of both. A detailed description of consequences for individual classes by excluding individual oxides is provided in Supplementary Information 2.

Variable importance of the ‘composition’ model

The highest variable importance based on the mean decrease in Gini impurity and accuracy is given by log-ratios with CaO versus Al_2O_3 , SiO_2 , $\text{FeO}_{\text{total}}$, MnO, and MgO (Fig. 5a). Calc-silicate (CS) and alkaline (A) garnet show low values for all of these ratios and thus can be discriminated from all other groups (Fig. 5b). Mafic garnet (M) shows higher values than CS and A, but lower values than intermediate-felsic/metasedimentary (IF/S), enabling to create high-purity splits. In addition, the ratios with CaO versus $\text{FeO}_{\text{total}}$, as well as MgO, are useful to separate ultramafic (UM) and M garnet.

The next important set of log-ratios includes MgO versus SiO_2 , Al_2O_3 , $\text{FeO}_{\text{total}}$, and MnO (Fig. 5a). This is not solely related to the creation of high-purity splits but also to highly increase the model accuracy. Like log-ratios with CaO, the ratios with MgO allow the separation of most CS and A garnet (high values) from M and IF/S garnet (intermediate values), and UM garnet (low values) (Fig. 5b).

Besides $\ln(\text{FeO}_{\text{total}} \times \text{CaO}^{-1})$ and $\ln(\text{FeO}_{\text{total}} \times \text{MgO}^{-1})$, other ratios that include $\text{FeO}_{\text{total}}$ like $\ln(\text{SiO}_2 \times \text{FeO}_{\text{total}}^{-1})$,

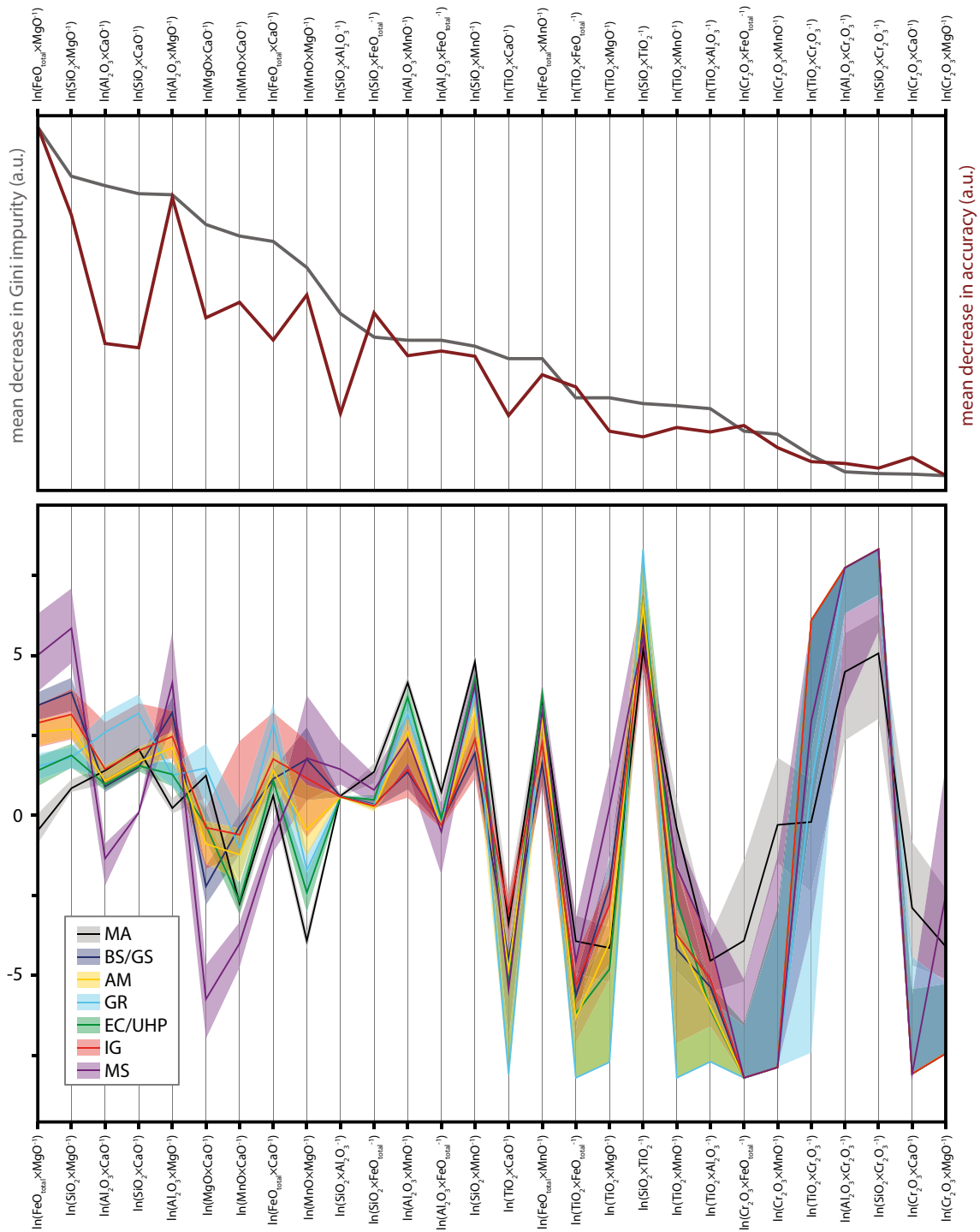


Fig. 3 Variable importance for the ‘setting and metamorphic facies’ model. Upper diagram shows the mean decrease in Gini impurity and mean decrease in accuracy for individual variables ordered from the highest decrease in Gini impurity (left) to lowest (right). Lower diagram shows corresponding variable values for individual classes as

25% quantile, median, and 75% quantile (see Table 1 for abbreviations). For some variables, nominator or denominator are often below detection limit, resulting in similar values for median and quantiles (lines overlap)

$\ln(\text{Al}_2\text{O}_3 \times \text{FeO}_{\text{total}}^{-1})$, and $\ln(\text{FeO}_{\text{total}} \times \text{MnO}^{-1})$ show only intermediate importance for the decrease in Gini impurity. Nevertheless, they are highly important for the accuracy

of the model (Fig. 5a). By contrast, log-ratios that include Cr_2O_3 also show intermediate importance for the purity of splits, but much lower importance for the model accuracy.

However, these ratios are obviously useful to separate UM garnet (Fig. 5b).

The log-ratios including TiO_2 are placed in a subordinate role for the mean decrease in Gini impurity and accuracy (Fig. 5a). By contrast, excluding TiO_2 when developing the ‘composition’ discrimination model leads to the highest increase in misclassification rates (owing to its dramatic influence in the correct reclassification of alkaline garnets), followed by CaO, MgO, MnO, Cr_2O_3 , $\text{FeO}_{\text{total}}$, SiO_2 , and Al_2O_3 in decreasing order of importance (Fig. 6). A detailed description of consequences for individual classes by excluding individual oxides is provided in Supplementary Information 2.

Origin of main discriminators

The discrimination of classes in the ‘setting and metamorphic facies’ as well as the ‘composition’ model is complex, and for sufficient separation, individual subgroups have to be considered individually. A detailed exploration of main discriminators for individual subgroups is provided in Supplementary Information 2. Here we focus on the most important class discriminators.

Garnet of class MS in the ‘setting and metamorphic facies’ model shows the highest classification success (Table 3), reflecting its distinct chemical composition. MS garnet is CaO rich, MgO poor, and shows a broad range in Al_2O_3 contents with an average lower than the other classes (Fig. 7a). The distinct composition is strongly related to the formation environment. Although skarn garnet can form in a wide range of settings and different protolith lithologies, by far the most are associated with igneous activity that leads to contact metamorphism of carbonates by heat supply and infiltrating metasomatic fluids at depth < 12 km (e.g., Meinert 1992). Garnet mainly forms at the prograde anhydrous stage together with clinopyroxene, both being high in Ca^{2+} due to the availability given by the chemical composition of the protolith. Furthermore, high oxygen fugacity enables the stabilization of andradite (end member composition: $\text{Ca}_3\text{Fe}_2\text{Si}_3\text{O}_{12}$), that involves the substitution of Al^{3+} by Fe^{3+} compared to grossular ($\text{Ca}_3\text{Al}_2\text{Si}_3\text{O}_{12}$), resulting in a garnet solid solution rich in grossular–andradite (e.g., Zhang and Saxena 1991). In contrast, clinopyroxene crystallizes mainly as solid solutions between diopside ($\text{CaMgSi}_2\text{O}_6$) and hedenbergite ($\text{CaFeSi}_2\text{O}_6$), and thus much of the available Mg^{2+} and Fe^{2+} are incorporated in clinopyroxene (Bin and Barton 1988). An exceptional case includes skarns that formed under low oxygen fugacity conditions like those associated with tungsten and tin mineralization, where some garnet populations are rich in Al^{3+} (e.g., Zhang and Saxena 1991; Meinert 1992). However, considering recently published compositions of garnet from these reduced environments (Duan et al. 2020; Im et al. 2020), > 87% are correctly

classified as MS, and the remaining are mainly classified as GR garnet of CS composition.

Garnet of class MA shows the second highest classification success in the ‘setting and metamorphic facies’ model (Table 3). In particular, the comparatively high content of MgO and Cr_2O_3 and the low content of $\text{FeO}_{\text{total}}$ and MnO in mantle rocks compared to crustal rocks represent the most important difference (Fig. 7b). In addition, a high importance of TiO_2 is observed for the discrimination of some subgroups (Supplementary Information 2). The enrichment of Cr in MA garnet is mainly related to the lithophile behaviour of Cr, resulting in the accumulation of Cr in mantle mineral phases like chromium spinel during partial melting of the upper mantle (e.g., Matrosova et al. 2020), while spinel is replaced by garnet at greater depths (e.g., Klemme et al. 2009). The higher TiO_2 content of MA garnet is probably related to the higher formation temperatures, resulting in increasing solubility of Ti in garnet (Aulbach 2020). Notably, caution should be taken for some rare UHP-IF/S garnets that can have extremely high MgO contents (Chopin 1984) leading to a misclassification as MA garnet.

Compared to MS and MA garnet, the discrimination of IG versus MM garnet is more challenging. Very distinct are only IG garnets of alkaline composition (Supplementary Information 2). Otherwise, the higher content of MnO and lower content of MgO as well as CaO represent the main discriminators (Fig. 7c). This is related to the much higher abundance of garnet in felsic igneous rocks compared to alkaline or mafic igneous rocks. High MnO contents reflect the abundant crystallization from highly fractionated Al- and Mn-rich magmas (e.g., Dahlquist et al. 2007), enabling garnet growth at pressures as low as 3 kbar (e.g., Green 1977). The low-pressure formation conditions also agree with low CaO contents, and Ca-rich garnet in igneous rocks only occurs in deeply emplaced intrusions (e.g., Anderson et al. 2008). In addition, at crystallization temperatures of felsic melts, Mg is partitioning into the melt, resulting in low-Mg garnet (Green 1977). However, this superordinate trend is only sufficient to separate parts of class IG and subpopulations have to be considered individually. Particularly notable is the classification improvement given by the higher content of TiO_2 in IG garnet compared to MM. Besides the relevance for garnet of alkaline igneous rocks, which are very rich in Ti (e.g., Huggins et al. 1977), TiO_2 is highly important for garnet from intermediate–felsic igneous rocks (Supplementary Information 2). Understanding the partitioning of Ti in garnet of igneous versus metamorphic systems is not straightforward, but the observed preferred incorporation of Ti into igneous biotite (Samadi et al. 2021) may also apply to garnet.

With regard to the discrimination of individual MM classes, the most important variables to separate BS/GS from other metamorphic classes include MgO and MnO

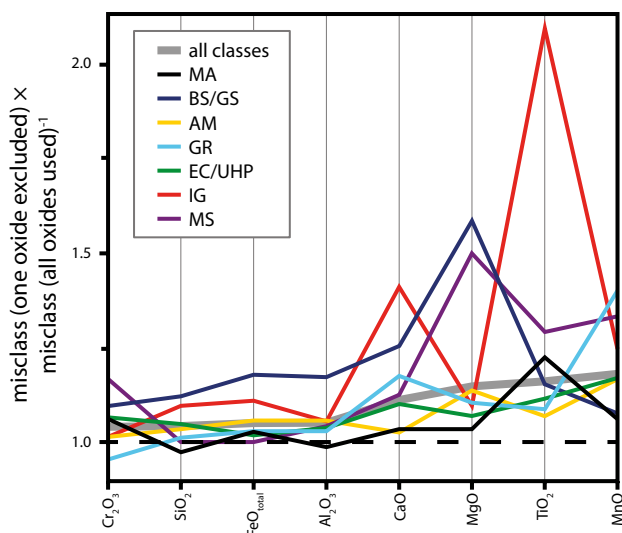


Fig. 4 Misclassification rates of the ‘setting and metamorphic facies’ model for individual oxides excluded compared to the full model with all oxides used (see Table 1 for abbreviations)

(Fig. 7d). The low content of MgO agrees with the low-temperature formation conditions and the many exchange thermometers that imply increasing $\text{Fe} \times (\text{Fe} + \text{Mg})^{-1}$ with decreasing temperature (Reverdatto et al. 2019, and references therein). High MnO contents are consistent with the increasing stability field of garnet to lower P – T conditions with increasing MnO content of the protolith. Thus, the earliest grown garnet typically shows the highest MnO contents (e.g., Carlson 1989) and removes MnO from the effective bulk composition leading to an up-temperature shift of garnet stability and bell-shaped Mn-zoning patterns (Evans 2004).

For the discrimination of GR from AM and EC/UHP garnet, the combination of high MgO and low CaO is most important (Fig. 7e). This highlights the fundamentals of Fe–Mg exchange thermometry (decreasing $\text{Fe} \times (\text{Fe} + \text{Mg})^{-1}$ with increasing temperature) and the garnet–aluminosilicate–plagioclase–quartz geo-thermo-barometer based on the higher stability of anorthite at high-temperature/low-pressure (low Ca in garnet) compared to the higher stability of grossular + aluminosilicate + quartz at low-temperature/high-pressure (e.g., Ghent 1976; Koziol and Newton 1988). Similarly, the best separation of AM and EC/UHP is observed by considering the higher values of $\text{FeO}_{\text{total}}$ versus MgO (lower temperature) and MnO versus CaO (lower pressure) for AM (Fig. 7f).

Concerning composition, class CS shows the highest discrimination success, followed by A (Table 4). CS and A garnet are well separated from all other classes by their higher CaO content (Fig. 8a), reflecting the high host-rock CaO content. The high TiO_2 content of A garnet allows the discrimination to CS (Fig. 8b). Furthermore, UM garnet

separates from IF/S and M garnet by the higher MgO and Cr_2O_3 content in combination with the lower $\text{FeO}_{\text{total}}$ content (Fig. 8c), in line with the element availability defined by the host-rock composition as well as the high-temperature formation conditions of mantle rocks. For the discrimination of IF/S and M garnet, $\ln(\text{FeO}_{\text{total}} \times \text{MgO}^{-1})$ and $\ln(\text{MnO} \times \text{CaO}^{-1})$ are the most important variables, both being higher for IF/S compared to M garnet (Fig. 8d). This agrees with the often higher CaO content of M host rocks and the stabilization effect of garnet at lower temperatures by increased host-rock MnO contents that are mainly observed from IF/S protoliths.

Sensitivity and applicability

One of the major aims in sedimentary provenance analysis is the reconstruction of source rock assemblages of sediments and the climatic–physiographic conditions under which they formed (von Eynatten and Dunkl 2012). The main potential of garnet single-grain analysis lies in reflecting variations in metamorphic conditions of rocks located in the source region, in particular when other characteristic minerals are lost due to their lower mechanical and/or chemical stability (Morton 1985). Thus, to represent a robust tool, predictions of the introduced garnet discrimination scheme should (i) be sensitive to changes in P – T conditions, (ii) reflect catchment specific host-rock characteristics, and (iii) be able to identify provenance shifts occurring in sedimentary successions or between samples.

Sensitivity to P – T changes

The sensitivity to variations in P – T conditions during garnet growth is tested by comparing the proposed discrimination scheme predictions for different garnet growth zones and/or different garnet populations from samples that record multistage garnet growth based on geo-thermo-barometric investigations.

Schantl et al. (2019) studied garnet-bearing granulites from the Moldanubian Zone in Lower Austria by a combination of rutile thermometry, biotite breakdown reactions, Ti content of biotite, garnet–aluminosilicate–plagioclase–quartz and amphibole–plagioclase thermo-barometry as well as thermodynamic modelling. For a sample set of five felsic granulites, the authors deduced 810–820 °C and 16–25 kbar for garnet core formation (eclogite facies), after which the core composition has been entirely reset by diffusion during decompression under increasing temperature to 1000–1050 °C and 15–17 kbar (eclogite–granulite-facies transition). Subsequent fast decompression cooling to 770 °C and 8 kbar (granulite–amphibolite facies transition)

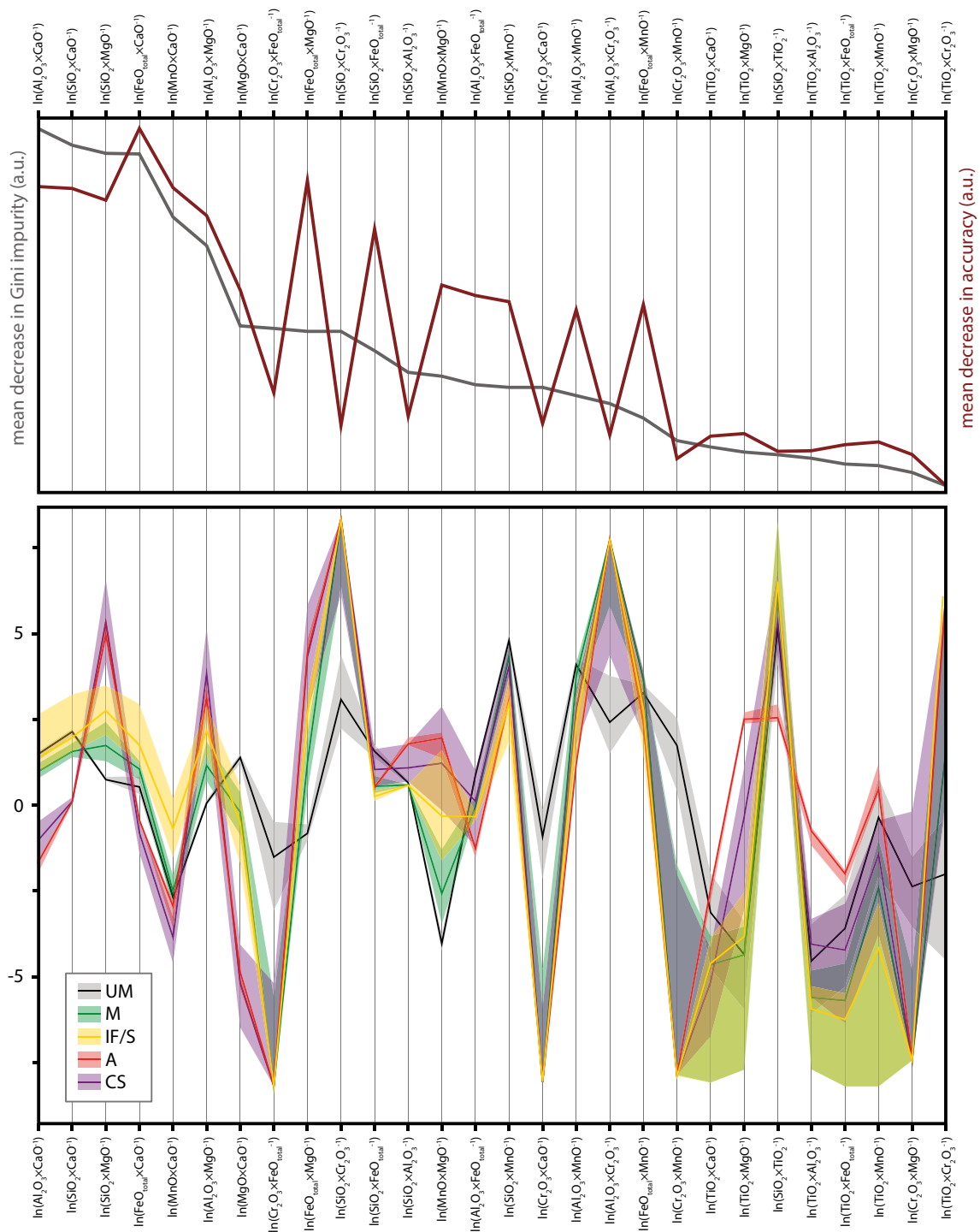


Fig. 5 Variable importance for the ‘composition’ model. Upper diagram shows the mean decrease in Gini impurity and mean decrease in accuracy for individual variables ordered from the highest decrease

Gini impurity (left) to lowest (right). Lower diagram shows corresponding variable values for individual classes as 25% quantile, median, and 75% quantile (see Table 1 for abbreviations)

lead to diffusional modification of the rim composition and minor modification of the core composition, as well as the formation of a second garnet generation (Fig. 9a, red path in P – T facies diagram for MORB composition). This P – T

evolution is well reflected by the arithmetic mean votes for garnet core ($n = 5$), garnet rim ($n = 5$), and retrograde garnet ($n = 2$, only observed in one sample) compositions (Fig. 9a, red path in ternary vote plots). Core compositions received

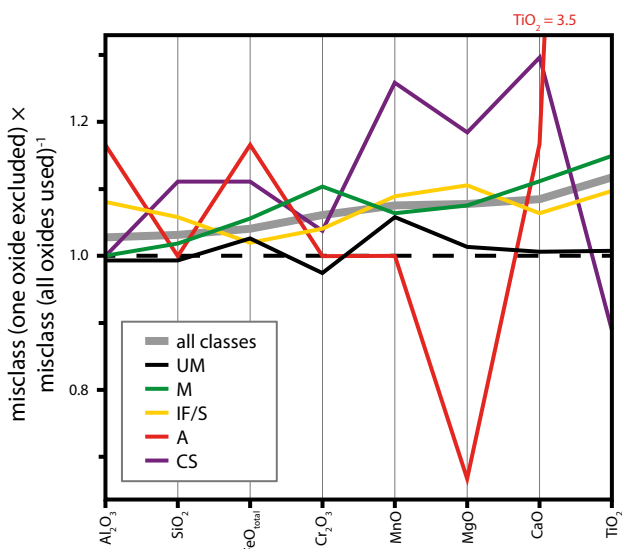


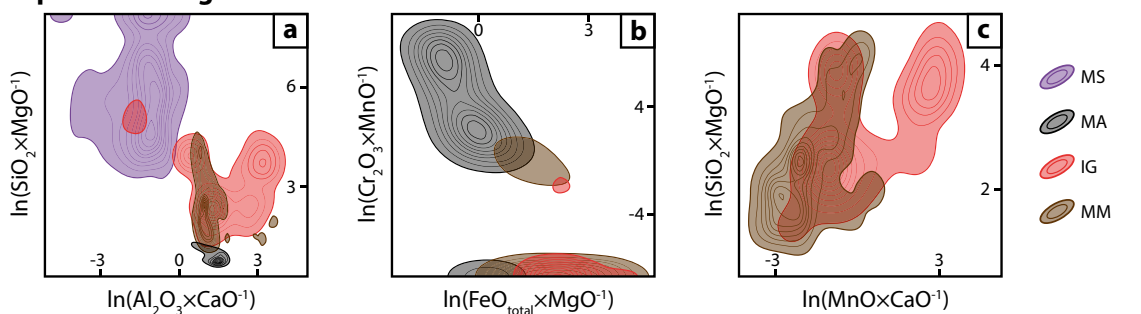
Fig. 6 Misclassification rates of the ‘composition’ model for individual oxides excluded compared to the full model with all oxides used (see Table 1 for abbreviations)

the highest votes for EC/UHP being five times higher than for AM and GR, rim compositions give similar votes for EC/UHP and GR being three times higher than AM, and retrograde garnet compositions show the highest votes for GR

followed by slightly lower votes for AM, both being more than two times higher than for EC/UHP.

Giuntoli et al. (2018) investigated complexly zoned garnets from micaschists of the Sesia Zone in the Western Italian Alps. Thermodynamic modelling of garnet growth zones records two orogenic cycles, most completely preserved by sample ‘1249’. The pre-alpine metamorphic event gives 730 °C and 6 kbar for garnet core formation (high-temperature amphibolite facies), followed by isobaric cooling to 620 °C and 6 kbar recorded by the first rim (amphibolite facies) (Fig. 9a, pale-blue path in *P–T* facies diagram). Alpine metamorphism at 620 °C and 16 kbar led to diffusional re-equilibration of the outermost rim and in proximity to fractures, denoted as the second rim (eclogite facies). Growth of a third rim resulted from subsequent temperature increase to 660 °C at almost isobaric conditions of 15 kbar (eclogite facies). This complex path is nicely reflected by the discrimination scheme votes (Fig. 9a, pale-blue path in ternary vote plots). For the high-temperature amphibolite-facies core formation, predictions are highest for class AM followed by GR with very minor votes for BS/GS and EC/UHP. Votes for the composition of the first rim that formed under isobaric cooling within the amphibolite facies stay highest for AM, but the second highest votes switched to BS/GS reflecting the colder conditions. The strong increase in pressure recorded by the second and third rim results

separate setting classes



separate metamorphic facies classes

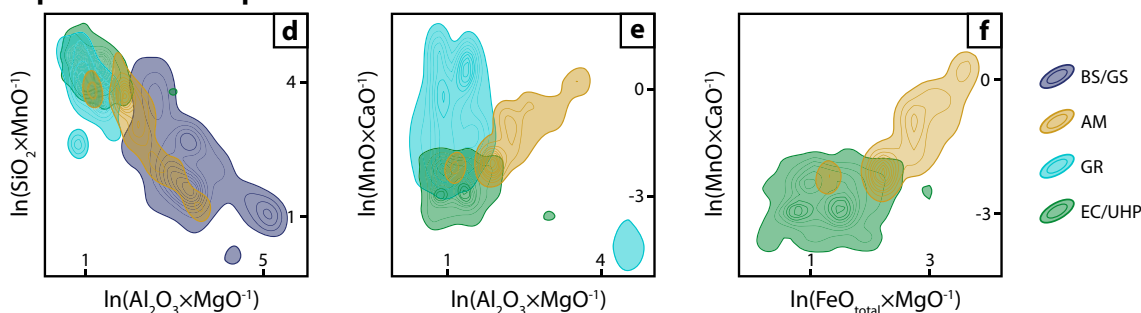


Fig. 7 Main discriminators to separate setting (a–c) and metamorphic facies classes (d–f). **a** MS versus MA, IG, and MM; **b** MA versus IG and MM; **c** IG versus MM; **d** BS/GS versus GR, AM, and EC/UHP; **e**

GR versus AM and EC/UHP; **f** AM versus EC/UHP (see Table 1 for abbreviations). Data shown as kernel density estimate maps

separate composition classes

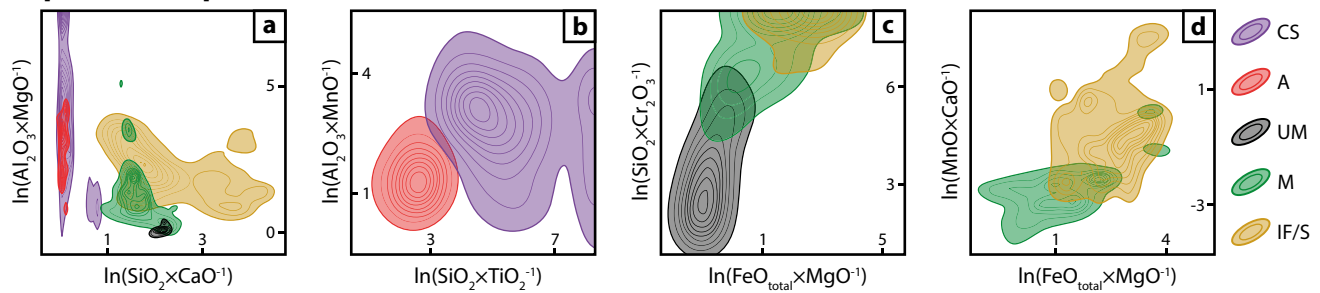


Fig. 8 Main discriminators to separate composition classes. **a** CS and A versus UM, M, and IF/S; **b** CS versus A; **c** UM versus M and IF/S; **d** M versus IF/S (see Table 1 for abbreviations). Data shown as kernel density estimate maps

in dominant votes for EC/UHP. Such increase in votes for EC/UHP is also observed for the increased pressure conditions from the innermost garnet core (545 °C and 15 kbar, blueschist–eclogite-facies transition) to the outermost core (560 °C and 21 kbar, eclogite facies) of a chloritoid-bearing micaschist studied by Negulescu et al. (2018) (Fig. 9a, yellow path). It is important to note that in all three P – T paths tested, the host-rock composition for stages outside the eclogite facies is correctly predicted as IF/S, but growth stages within the eclogite facies are misclassified as M. This highlights the difficulty of assigning the correct composition class to subgroup EC-IF/S (cf. Figure 2b).

Li et al. (2021) studied mafic granulite lenses from the South Altyn Orogen in West China by thermodynamic modelling including garnet and plagioclase isopleth thermobarometry. The authors observed three garnet generations, whereby conditions of 600–655 °C and 15.8–19.2 kbar (eclogite facies) are recorded by the core of garnet one, followed by garnet one mantle growth at 920 °C and 36.2 kbar (ultrahigh-pressure eclogite facies), formation of garnet two from 820 °C and 17.8 kbar (eclogite facies) to 826 °C and 11.5 kbar (high-pressure granulite facies), and rim formation of garnet three at 826–735 °C and 11.5–8.7 kbar (high-pressure granulite–amphibolite facies transition) (Fig. 9a, dark blue path and white points in P – T facies diagram). While the formation of garnet one is correctly predicted as EC/UHP by the discrimination scheme, decompression is not well reflected, and although votes for GR increase, they stay highest for EC/UHP (Fig. 9a, dark blue path and white points in ternary vote plots). This agrees with the issue of the metamorphic class prediction for garnet of subgroup GR-M (cf. Figure 1d).

Reflecting catchment specific garnet host-rock characteristics

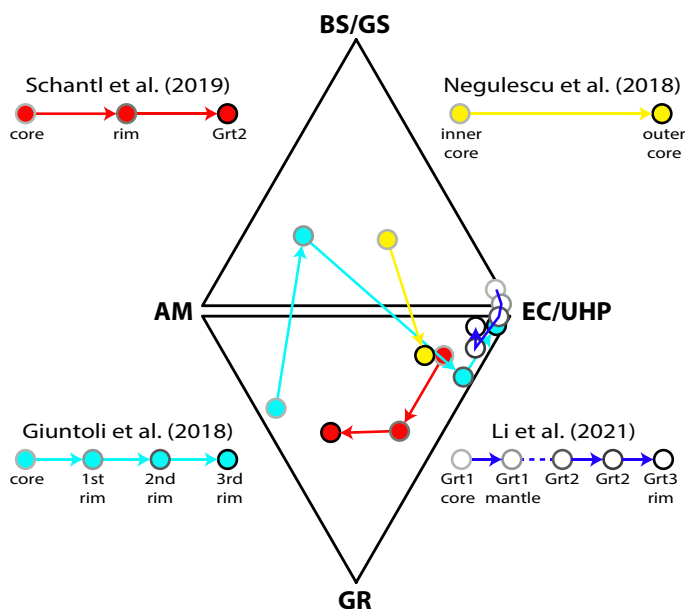
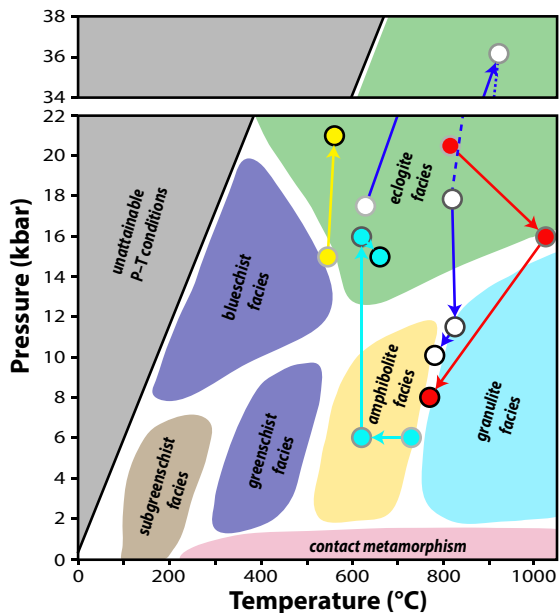
To test the sensitivity of the discrimination scheme to reflect changes in the relative abundance of different garnet-bearing

host rocks contributing erosional material to sediments, two modern sand case studies are chosen with a well-known geological framework of the catchments. Thereby, we focus on discrimination issues that are not able to be tackled by previous schemes, that is the discrimination between blueschist/greenschist and eclogite facies sources as well as the discrimination between crustal and mantle eclogites.

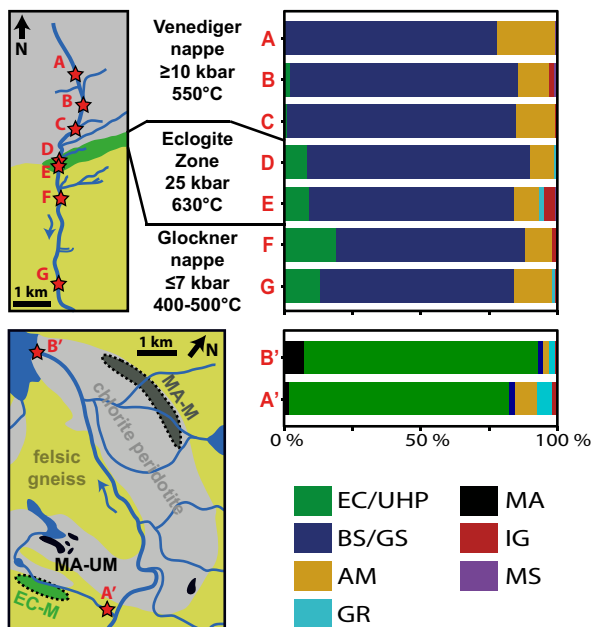
Krippner et al. (2015) studied detrital garnets from tributaries draining the western Hohe Tauern window in the central Eastern Alps, Austria. The Dorfertal stream section was sampled by seven modern sand samples, here renamed as A–G (Fig. 9b). Along the profile, the stream drains (i) the Venediger nappe that preserves few pre-Alpine eclogites (e.g., von Quadt et al. 1997) and underwent alpine peak metamorphic conditions of ~550 °C and ≥ 10 kbar mainly corresponding to the blueschist facies (Selverstone 1993), (ii) the ‘Eclogite Zone’ that records peak conditions of ~630 °C and ~25 kbar corresponding to the eclogite facies (Hoschek 2007), and (iii) the Glockner nappe that underwent metamorphism at 400–500 °C and ≤ 7 kbar mainly corresponding to the greenschist facies (Selverstone 1993). This drainage path is well represented by the discrimination scheme predictions, showing that (i) garnets of samples A–C collected within the Venediger nappe are dominantly assigned to a BS/GS origin, (ii) an increasing amount of EC/UHP assigned garnet is observed once the ‘Eclogite Zone’ is entered (samples D and E) which is becoming highest directly downstream of the eclogite zone (sample F), and (iii) the proportion of EC/UHP assigned garnet dilutes further downstream in sample G of the Glockner nappe (Fig. 9b, barplot).

The second case study relates to modern sands of tributaries draining the Almklovdalen peridotite in the Western Gneiss Region of Norway (Krippner et al. 2016). This area represents a key region for the long-lasting debate about the crustal (in-situ) versus mantle (exotic) origin of eclogites, whereby it finally turned out that both interpretations are feasible depending on the type of eclogite one is looking at (Brueckner 2018). A breakthrough was made by Medaris

(a) pressure-temperature variations



(b) catchment specific characteristics



(c) provenance shifts

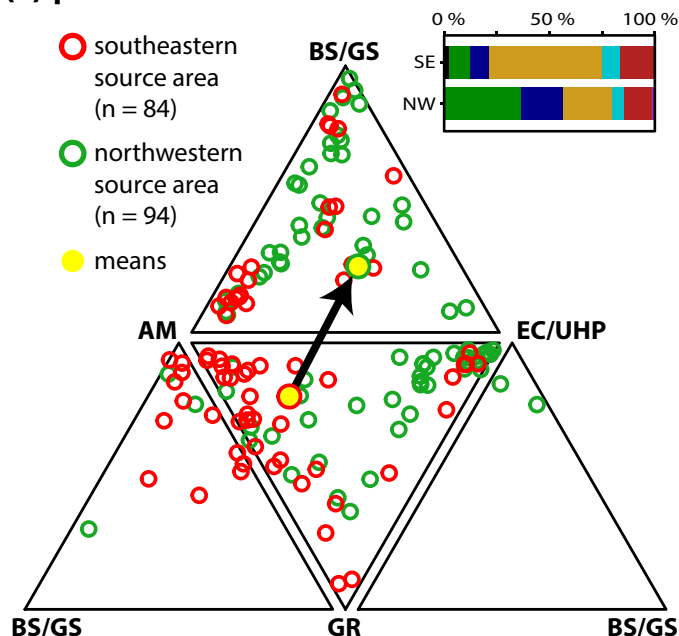


Fig. 9 Sensitivity and applicability of the introduced garnet discrimination scheme. **a** Sensitivity to changes in pressure and temperature during garnet growth tested by crystalline rock examples. Tested examples are from Giuntoli et al. (2018), Negulescu et al. (2018), Schantl et al. (2019), and Li et al. (2021). Conditions of individual growth zones are shown in a pressure–temperature diagram with metamorphic facies for MORB composition after Bucher and Frey (2002), modified from Schönig et al. (2018b). Vote results for individual growth zones by the introduced discrimination scheme are shown in ternary diagrams. **b** Sensitivity to reflect catchment specific characteristics tested on example case studies of Krippner

et al. (2015) and Krippner et al. (2016). Simplified geological maps show sampling locations in the western Hohe Tauern (upper map) and Almkloddalen area (lower map). Class assignments of garnets from individual samples by the introduced discrimination scheme are shown as barplots. **c** Sensitivity to identify shifts in provenance tested on an example case study by von Eynatten and Gaupp (1999). Individual votes as well as arithmetic mean votes for garnet grains of sediments sourced from southeast and northwest are shown in the ternary ‘metamorphic facies’ plot. Barplot shows the corresponding class assignment (see Table 1 for abbreviations)

(1980), showing that so-called ‘external’ or ‘in-situ’ eclogites that occur within felsic host gneisses are of crustal origin and underwent a prograde path during subduction, while ‘internal’ or ‘exotic’ eclogites (or other garnet-bearing assemblages) within peridotites are of mantle origin, were emplaced in the subducted crust, and underwent a retrograde path during exhumation.

By investigating the composition of very fine-sand-sized detrital garnets of four sediment samples in comparison to the garnet composition of very fine- to coarse-sand-sized garnets of one sample, Krippner et al. (2016) observed that very pyrope-rich grains that imply a mantle origin only occur in the coarsest fraction. Re-evaluation of the fine-sand garnet data using the new discrimination scheme reveals a mantle origin for a significant proportion. Two sampling locations of Krippner et al. (2016) are marked on the simplified geological, renamed as A’ and B’ (Fig. 9b). In the small catchment of sample A’, garnet-free chlorite peridotite, felsic gneiss, and frequent lenses of crustal eclogites (EC-M) occur. In contrast, in the lower stream section of sample B’, frequent mantle eclogite lenses (MA-M) shed material to the sedimentary system. This is well represented by class prediction, showing that garnet of mantle origin is enriched in sample B’ (Fig. 9b), and a mafic host-rock composition is assigned to these mantle garnets (not shown).

Detecting provenance shifts

To test the capability of the new garnet discrimination scheme regarding the identification of shifts in provenance, a multi-method provenance study of Cretaceous sedimentary rocks from the Northern Calcareous Alps is chosen (von Eynatten and Gaupp 1999). Based on framework petrography, heavy mineral analysis, and the chemical composition of amphiboles, white mica, tourmaline, garnet, and chloritoid, the authors distinguished two contrasting source regions. Particularly the occurrence of glaucophane and phengite indicate the contribution of erosional material from low-temperature/high-pressure metamorphic rocks located in a north-western source area, which are absent in sediments sourced from a south-western source.

Re-evaluation of the detrital garnet data by the introduced discrimination scheme strongly supports this observation. Garnet grains from the south-eastern source are dominantly classified as AM (~54%), while class assignments to BS/GS, GR, and EC/UHP are each < 10% (Fig. 9c). By contrast, garnet grains sourced from northwest are dominantly assigned to EC/UHP (~36%) with significant but much lower proportions of AM (~23%), and much higher proportions of BS/GS (~20%). This is further highlighted by the arithmetic mean votes that shift from the AM–GR–EC/UHP diagram (37% AM, 13% GR, 14% EC/UHP, 11% BS/GS) for the south-eastern source, to

the AM–BS/GS–EC/UHP diagram (25% AM, 9% GR, 31% EC/UHP, 18% BS/GS) for the north-western source (Fig. 9c, black arrow).

The ‘garnetRF’ web application

To enable a user-friendly application of the garnet random forest discrimination scheme without the need for installing software and having programming expertise, we developed a freely and worldwide accessible web application called ‘garnetRF v1.0’. The app was developed using the ‘shiny’ R package of Chang et al. (2021) and the ‘shinydashboard’ package of Chang and Ribeiro (2018). Used packages further include those mentioned in the “Model development” section as well as ‘readxl’ (Wickham and Bryan 2019) and ‘openxlsx’ (Schauberger and Walker 2020) for importing spreadsheet files, as well as ‘colourpicker’ (Attali 2020) and ‘cowplot’ (Wilke 2020) for visualisation.

The ‘garnetRF’ app is standalone and users are guided by an ‘instructions’ tab. By five quick steps, users get from their garnet major-element data to discrimination scheme results and corresponding plots:

1. Visit the ‘garnetRF’ app via web browser at <http://134.76.17.86:443/garnetRF/>, hosted at the servers of GWDG, the data and IT service company of the University of Göttingen and the Max Planck Institute in Göttingen.
2. Give your project a title. This is optional but facilitates data handling as any file output includes the project title.
3. Download the .xlsx template via ‘Data Preparation’ and copy your garnet chemical data to the template. Note that all analyses from each individual sample should have the same sample name for adequate data visualization and calculation of means. Individual grain or spot names can optionally be assigned in the ‘grain_no’ column. Ensure the right order of oxides, give values in wt%, use a point as decimal separator, give a numeric value < 0.03 for oxide contents below detection limit, and leave cells empty for not-analysed oxides. Empty cells are only allowed for Cr₂O₃ and TiO₂ (Note that this reduces prediction accuracy: cf. Figs. 4 and 6).
4. Upload your prepared data. Once upload and calculations are complete, results can be found and downloaded in the ‘Data Tables’ tab. Individual results for each grain, the arithmetic mean for each sample, and class assignments (majority vote) for each sample are given.
5. Click ‘Create Plots’ to visualize individual and mean votes as ternary plots for each scheme as well as barplots

for class assignments. Marker size, shape, and colour can be modified and plots can be downloaded as.pdf, and thus easily be modified via graphical vector programs.

Conclusion

A large database including 13,615 observations of chemical garnet major-element analyses is compiled from the literature. Observations are subdivided into 23 petrogenetic subgroups with regard to host-rock type, metamorphic facies, and composition. With this database, a new discrimination scheme is developed that aims at predicting the original host-rock of detrital garnet grains. To focus on the most substantial information in terms of sedimentary provenance, database subgroups are merged into seven classes to predict the ‘setting and metamorphic facies’ of the garnet-bearing host-rock as well as five classes to predict the ‘composition’. For both prediction issues, the random forest classification machine-learning algorithm is applied.

The final discrimination scheme is easily applicable via a provided web app. Considering the out-of-bag error, the scheme is able to correctly predict (i) the host-rock setting with a class average of > 95%, (ii) the host-rock metamorphic facies with a class average of > 84%, and (iii) the host-rock composition with a class average of > 93%.

Success rates for individual classes included in setting, metamorphic facies, and composition predictions differ $\leq 5\%$ from the class average, emphasizing substantial discrimination balance. However, balancing between individual subgroups merged in the prediction classes is much less pronounced. Particularly notable are higher misclassification rates in (i) setting prediction for metamorphic calc–silicate garnet (misclassified as metasomatic) and mafic igneous garnet (misclassified as metamorphic), (ii) metamorphic facies prediction for intermediate–felsic/metasedimentary eclogite/ultrahigh-pressure facies garnet (misclassified as granulite, blueschist/greenschist, and amphibolite facies) and mafic granulite-facies garnet (misclassified as eclogite/ultrahigh-pressure facies), and (iii) composition prediction for mafic blueschist/greenschist facies as well as mafic igneous garnet (misclassified as intermediate–felsic/metasedimentary) and intermediate–felsic/metasedimentary eclogite/ultrahigh-pressure garnet (misclassified as mafic).

Detailed exploration of the discrimination models reveals that decisions mainly follow characteristic partitioning trends that are the building blocks of geo-thermo-barometry. The strength of the developed model is the step-wise consideration of many of those element ratios in a randomized way that leads to a significant increase in classification success without compromising generalization. This procedure also uses individual decision lines for subpopulations

occurring within the prediction classes, which seems to be the main reason for high classification success rates for the host-rock metamorphic facies without knowledge of the host-rock composition and vice versa. In addition, model exploration uncovers the high potential of minor variations in TiO_2 to strongly increase prediction accuracy, underlining the importance to include TiO_2 in the standard protocol of in situ chemical analysis of garnet.

Application examples to crystalline rocks that record multiple stages of garnet growth, to modern sand samples from tributaries draining variable lithologies, and to sedimentary successions that received erosional material from different sources demonstrate that the discrimination scheme is sensitive to identify (i) variations in P – T conditions, (ii) catchment specific host-rock characteristics, and (iii) shifts in provenance. Although the scheme is designed for provenance applications and subsists on a statistically significant number of grains analysed, it may also be useful in crystalline rock petrology to quickly narrow down sample sets to the most interesting samples to study these in more detail.

Outlook and call on the community

Although garnet host-rock predictions clearly enhanced compared to previous discrimination schemes, there is still potential for improvement, in particular with regard to robustness and balance within individual classes. One issue is the underrepresentation of some individual groups/subgroups, especially BS, GS, EC/UHP-IF/S, IG-M, and low-temperature MM-CS. These are also those groups/subgroups that show the lowest classification success. Another issue is the separate treatment of ‘setting and metamorphic facies’ and ‘composition’.

We suggest that a combined model with a prediction of detailed classes and derivation of the most important information is the best future direction. For example, class GR could be split into GR-IF/S, GR-M, and GR-CS, and after prediction, the maximum vote of the three classes defines the GR prediction, similar to the approach for MM in the presented setting scheme. In the same way, the maximum vote of all mafic classes, that would be MA-M, IG-M, BS/GS-M, AM-M, GR-M, and EC/UHP-M, defines the prediction for M. This also allows to balance all subgroups during the training phase of model development.

To reach this aim, the dataset has to be extended, in particular for underrepresented groups/subgroups mentioned. In many works, solely the most representative garnet compositions are reported, but it is clear that a much higher amount of data was acquired during the project. Thus, we want to encourage the community to contribute their supplementary as well as newly published garnet data with known host-rock type, metamorphic grade, and composition to the dataset.

We also appreciate hints on references that are so far not included in the dataset. In addition, we welcome any kind of criticism or detected problems. Based on feedback and new data acquisition, we plan to submit an updated scheme within the next couple of years.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00410-021-01854-w>.

Acknowledgements We thank Pierre Lanari and an anonymous reviewer for detailed and valuable comments and Daniela Rubatto for the editorial work.

Author contributions JS: conceptualization, software, formal analysis, investigation, data curation, visualization, writing—original draft; HvE: conceptualization, supervision, funding acquisition, writing—review and editing; RT-D: conceptualization, software, formal analysis, writing—review and editing; GM: funding acquisition, writing—review and editing.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Research Foundation [DFG grant EY 23/27-1].

Availability of data and material All data used are included in the main manuscript or available via the Supplementary Information.

Code availability A bare-bone version of the code for reproducing the classification models is provided in Supplementary Information 3. The ‘garnetRF v1.0’ web application is available via web browser at <http://134.76.17.86:443/garnetRF/>, hosted at the servers of GWDG, the data and IT service company of the University of Göttingen and the Max Planck Institute in Göttingen.

Declarations

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aitchison J (1986) The statistical analysis of compositional data. Chapman and Hall, London
- Aitchison J, Egozcue JJ (2005) Compositional data analysis: where are we and where should we be heading? *Math Geol* 37:829–850
- Anderson JL, Barth AP, Wooden JL, Mazdab F (2008) Thermometers and thermobarometers in granitic systems. *Rev Mineral Geochem* 69:121–142. <https://doi.org/10.2138/rmg.2008.69.4>
- Attali D (2020) Colourpicker: a colour picker tool for shiny and for selecting colours in plots. R package version 1.1.0. <https://CRAN.R-project.org/package=colourpicker>. Accessed 14 Sept 2020
- Aubrecht R, Méres S, Sýkora M, Mikuš T (2009) Provenance of the detrital garnets and spinels from the Albian sediments of the Czorsztyn Unit (Pieniny Klippen Belt, Western Carpathians, Slovakia). *Geol Carpath* 60:463–483
- Aulbach S (2020) Temperature-dependent rutile solubility in garnet and clinopyroxene from mantle eclogite: Implications for continental crust formation and V-based oxybarometry. *J Petrol* 61:egaa065. <https://doi.org/10.1093/ptrology/egaa065>
- Bache SM, Wickham H (2014) magrittr: a forward-pipe operator for R. R package version 1.5. <https://CRAN.R-project.org/package=magrittr>. Accessed 22 Nov 2014
- Baldwin SL, Schönig J, Gonzalez JP, Davies H, von Eynatten H (2021) Garnet sand reveals rock recycling processes in the youngest exhumed high- and ultrahigh-pressure terrane on Earth. *Proc Natl Acad Sci* 118:e2017231118. <https://doi.org/10.1073/pnas.2017231118>
- Baxter EF, Caddick MJ, Ague JJ (2013) Garnet: Common mineral, uncommonly useful. *Elements* 9:415–419. <https://doi.org/10.2113/gselements.9.6.415>
- Belgiu M, Drăguț L (2016) Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* 114:24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Biau G, Scornet E (2016) A random forest guided tour. *TEST* 25:197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Bin Z, Barton MD (1988) Compositional characteristics of garnets and pyroxenes in contact-metamorphic skarn deposits and their relationship with metallization. *Chin J Geochem* 7:329–335. <https://doi.org/10.1007/BF02842337>
- Bolton MS, Jensen BJ, Wallace K, Praet N, Fortin D, Kaufman D, De Batist M (2020) Machine learning classifiers for attributing tephra to source volcanoes: an evaluation of methods for Alaska tephra. *J Quat Sci* 35:81–92. <https://doi.org/10.1002/jqs.3170>
- Boulesteix AL, Janitza S, Kruppa J, König IR (2012) Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip Rev Data Min Knowl Discov* 2:493–507. <https://doi.org/10.1002/widm.1072>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Brown M, Johnson TE (2018) Secular change in metamorphism and the onset of global plate tectonics. *Am Miner* 103:181–196. <https://doi.org/10.2138/am-2018-6166>
- Brueckner HK (2018) The great eclogite debate of the Western Gneiss Region, Norwegian Caledonides: The in situ crustal v. exotic mantle origin controversy. *J Metamorph Geol* 36:517–527. <https://doi.org/10.1111/jmg.12314>
- Bucher K, Frey M (2002) Petrogenesis of metamorphic rocks. Springer-Verlag, Berlin-Heidelberg
- Caddick MJ, Konopásek J, Thompson AB (2010) Preservation of garnet growth zoning and the duration of prograde metamorphism. *J Petrol* 51:2327–2347. <https://doi.org/10.1093/ptrology/egq059>
- Carlson WD (1989) The significance of intergranular diffusion to the mechanisms and kinetics of porphyroblast crystallization. *Contrib Miner Petrol* 103:1–24. <https://doi.org/10.1007/BF00371361>
- Chang W, Ribeiro BB (2018) shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1. <https://CRAN.R-project.org/package=shinydashboard>. Accessed 17 Oct 2018
- Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2021) shiny: Web Application

- Framework for R. R package version 1.6.0. <https://CRAN.R-project.org/package=shiny>. Accessed 25 Jan 2021
- Chayes F (1960) On correlation between variables of constant sum. *J Geophys Res* 65:4185–4193. <https://doi.org/10.1029/JZ065i012p04185>
- Chen C, Liaw A, Breiman L (2004) Using random forest to learn imbalanced data. Technical report, # 666, Department of Statistics, University of California, Berkeley, CA. <http://www.stat.berkeley.edu/tech-reports/666.pdf>. Accessed 29 May 2021
- Chopin C (1984) Coesite and pure pyrope in high-grade blueschists of the Western Alps: a first record and some consequences. *Contrib Mineral Petrol* 86:107–118. <https://doi.org/10.1007/BF00381838>
- Čopjaková R, Sulovský P, Paterson BA (2005) Major and trace elements in pyrope–almandine garnets as sediment provenance indicators of the Lower Carboniferous Culm sediments, Drahaný Uplands, Bohemian Massif. *Lithos* 82:51–70. <https://doi.org/10.1016/j.lithos.2004.12.006>
- Dahlquist JA, Galindo C, Pankhurst RJ, Rapela CW, Alasino PH, Saavedra J, Fanning CM (2007) Magmatic evolution of the Peñón Rosado granite: petrogenesis of garnet-bearing granitoids. *Lithos* 95:177–207. <https://doi.org/10.1016/j.lithos.2006.07.010>
- Dhuime B, Hawkesworth CJ, Delavault H, Cawood PA (2017) Continental growth seen through the sedimentary record. *Sed Geol* 357:16–32. <https://doi.org/10.1016/j.sedgeo.2017.06.001>
- Duan XX, Ju YF, Chen B, Wang ZQ (2020) Garnet geochemistry of reduced Skarn system: implications for fluid evolution and Skarn formation of the Zhuxiling W (Mo) deposit, China. *Minerals* 10:1024. <https://doi.org/10.3390/min10111024>
- Evans TP (2004) A method for calculating effective bulk composition modification due to crystal fractionation in garnet-bearing schist: implications for isopleth thermobarometry. *J Metamorph Geol* 22:547–557. <https://doi.org/10.1111/j.1525-1314.2004.00532.x>
- Ghent ED (1976) Plagioclase–garnet– Al_2SiO_5 –quartz: a potential geobarometer–geothermometer. *Am Mineral* 61:710–714
- Giuntoli F, Lanari P, Engi M (2018) Deeply subducted continental fragments: I. Fracturing, dissolution-precipitation and diffusion processes recorded by garnet textures of the central Sesia Zone (Western Italian Alps). *Solid Earth* 9:167–189. <https://doi.org/10.5194/se-2017-87>
- Goodwin AM (1996) Principles of precambrian geology. Academic Press, London
- Green TH (1977) Garnet in silicic liquids and its possible use as a PT indicator. *Contrib Mineral Petrol* 65:59–67
- Grütter HS, Gurney JJ, Menzies AH, Winter F (2004) An updated classification scheme for mantle-derived garnet, for use by diamond explorers. *Lithos* 77:841–857. <https://doi.org/10.1016/j.lithos.2004.04.012>
- Hamilton NE, Ferry M (2018) ggtern: ternary diagrams using ggplot2. *J Stat Softw* 87:1–17. <https://doi.org/10.18637/jss.v087.c03>
- Han S, Li M, Ren Q (2019) Discriminating among tectonic settings of spinel based on multiple machine learning algorithms. *Big Earth Data* 3:67–82. <https://doi.org/10.1080/20964471.2019.1586074>
- Hardman MF, Pearson DG, Stachel T, Sweeney RJ (2018) Statistical approaches to the discrimination of crust-and mantle-derived low-Cr garnet–major-element-based methods and their application in diamond exploration. *J Geochem Explor* 186:24–35. <https://doi.org/10.1016/j.gexplo.2017.11.012>
- Holder RM, Viete DR, Brown M, Johnson TE (2019) Metamorphism and the evolution of plate tectonics. *Nature* 572:378–381. <https://doi.org/10.1038/s41586-019-1462-2>
- Hong D, Jian X, Fu L, Zhang W (2020) Garnet trace element geochemistry as a sediment provenance indicator: an example from the Qaidam basin, northern Tibet. *Mar Pet Geol* 116:104316. <https://doi.org/10.1016/j.marpetgeo.2020.104316>
- Hoschek G (2007) Metamorphic peak conditions of eclogites in the Tauern Window, Eastern Alps, Austria: thermobarometry of the assemblage garnet + omphacite + phengite + kyanite + quartz. *Lithos* 93:1–16. <https://doi.org/10.1016/j.lithos.2006.03.042>
- Huggins FE, Virgo D, Huckenholz HG (1977) Titanium-containing silicate garnet; I, The distribution of Al, Fe^{3+} , and Ti^{4+} between octahedral and tetrahedral sites. *Am Mineral* 62:475–490
- Im H, Jeong JY, Shin D (2020) Genetic environment of W skarn and Pb–Zn vein mineralization associated with the Imog granite in the Taebaeksan Mineralized District, South Korea. *Ore Geol Rev* 126:103721. <https://doi.org/10.1016/j.oregeorev.2020.103721>
- Itano K, Ueki K, Iizuka T, Kuwatani T (2020) Geochemical discrimination of monazite source rock based on machine learning techniques and multinomial logistic regression analysis. *Geosciences* 10:63. <https://doi.org/10.3390/geosciences10020063>
- Klemme S, Ivanic TJ, Connolly JAD, Harte B (2009) Thermodynamic modelling of Cr-bearing garnets with implications for diamond inclusions and peridotite xenoliths. *Lithos* 112:986–991. <https://doi.org/10.1016/j.lithos.2009.05.007>
- Koziol AM, Newton RC (1988) Redetermination of the anorthite breakdown reaction and improvement of the plagioclase–garnet– Al_2SiO_5 –quartz geobarometer. *Am Mineral* 73:216–223
- Krippner A, Meinhold G, Morton AC, von Eynatten H (2014) Evaluation of garnet discrimination diagrams using geochemical data of garnets derived from various host rocks. *Sed Geol* 306:36–52. <https://doi.org/10.1016/j.sedgeo.2014.03.004>
- Krippner A, Meinhold G, Morton AC, Russell E, von Eynatten H (2015) Grain-size dependence of garnet composition revealed by provenance signatures of modern stream sediments from the western Hohe Tauern (Austria). *Sed Geol* 321:25–38. <https://doi.org/10.1016/j.sedgeo.2015.03.002>
- Krippner A, Meinhold G, Morton AC, Schönig J, von Eynatten H (2016) Heavy minerals and garnet geochemistry of stream sediments and bedrocks from the Almkuldalen area, Western Gneiss Region, SW Norway: implications for provenance analysis. *Sed Geol* 336:96–105. <https://doi.org/10.1016/j.sedgeo.2015.09.009>
- Lanari P, Engi M (2017) Local bulk composition effects on metamorphic mineral assemblages. *Rev Mineral Geochem* 83:55–102. <https://doi.org/10.2138/rmg.2017.83.3>
- Lanari P, Giuntoli F, Loury C, Burn M, Engi M (2017) An inverse modelling approach to obtain P–T conditions of metamorphic stages involving garnet growth and resorption. *Eur J Mineral* 29:181–199. <https://doi.org/10.1127/ejm/2017/0029-2597>
- Li X, Liu L, Liao X, Gai Y, Ma T, Geng G, Li T (2021) Metamorphic evolution of mafic granulites in the Kuruksayi area of the South Altyn Orogen, West China: insights from petrography, phase equilibria modelling and geochronology. *J Asian Earth Sci* 213:104766. <https://doi.org/10.1016/j.jseaes.2021.104766>
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2:18–22
- Maneiro KA, Baxter EF, Samson SD, Marschall HR, Hietpas J (2019) Detrital garnet geochronology: Application in tributaries of the French Broad River, Southern Appalachian Mountains, USA. *Geology* 47:1189–1192. <https://doi.org/10.1130/G46840.1>
- Mange MA, Morton AC (2007) Geochemistry of heavy minerals. *Dev Sedimentol* 58:345–391. [https://doi.org/10.1016/S0070-4571\(07\)58013-1](https://doi.org/10.1016/S0070-4571(07)58013-1)
- Matrosova EA, Bobrov AV, Bindi L (2020) Geochemistry of chromium in the earth’s mantle. Springer Nature Switzerland, Cham. <https://doi.org/10.1007/978-3-030-27018-6>
- Medaris LG (1980) Convergent metamorphism of eclogite and garnet-bearing ultramafic rocks at Lien, West Norway. *Nature* 283:470–472. <https://doi.org/10.1038/283470a0>
- Meinert LD (1992) Skarns and Skarn Deposits. *Geoscience Canada* 19:145–162. <https://journals.lib.unb.ca/index.php/GC/article/view/3773>. Accessed 25 Oct 2021
- Millonig LJ, Albert R, Gerdes A, Avigad D, Dietsch C (2020) Exploring laser ablation U–Pb dating of regional metamorphic garnet–The

- Straits Schist, Connecticut, USA. *Earth Planet Sci Lett* 552:116589. <https://doi.org/10.1016/j.epsl.2020.116589>
- Morton AC (1985) A new approach to provenance studies: electron microprobe analysis of detrital garnets from Middle Jurassic sandstones of the northern North Sea. *Sedimentology* 32:553–566. <https://doi.org/10.1111/j.1365-3091.1985.tb00470.x>
- Negulescu E, Săbău G, Massonne H-J (2018) Growth of chloritoid and garnet along a nearly isothermal burial path to 70 km depth: an example from the Bughea Metamorphic Complex, Leaota Massif, South Carpathians. *Mineral Petrol* 112:535–553. <https://doi.org/10.1007/s00710-017-0552-9>
- Petrelli M, Perugini D (2016) Solving petrological problems through machine learning: the study case of tectonic discrimination using geochemical and isotopic data. *Contrib Mineral Petrol* 171:1–15. <https://doi.org/10.1007/s00410-016-1292-2>
- Petrelli M, Bizzarri R, Morgavi D, Baldanza A, Perugini D (2017) Combining machine learning techniques, microanalyses and large geochemical datasets for tephrochronological studies in complex volcanic areas: New age constraints for the Pleistocene magmatism of central Italy. *Quat Geochronol* 40:33–44. <https://doi.org/10.1016/j.quageo.2016.12.003>
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 4.0.2. <https://www.R-project.org/>. Accessed 22 June 2020
- Ren Q, Li M, Han S, Zhang Y, Zhang Q, Shi J (2019) Basalt tectonic discrimination using combined machine learning approach. *Minerals* 9:376. <https://doi.org/10.3390/min9060376>
- Reverdatto VV, Likhonov II, Polyansky OP, Sheplev VS, Kolobov VY (2019) Mineral geothermobarometry. In: Reverdatto VV, Likhonov II, Polyansky OP, Sheplev VS, Kolobov VY (eds) *The nature and models of metamorphism*. Springer Nature Switzerland, Cham, pp 55–82. https://doi.org/10.1007/978-3-030-03029-2_2
- Samadi R, Torabi G, Kawabata H, Miller NR (2021) Biotite as a petrogenetic discriminator: chemical insights from igneous, meta-igneous and meta-sedimentary rocks in Iran. *Lithos* 386:106016. <https://doi.org/10.1016/j.lithos.2021.106016>
- Schantl P, Hauzenberger C, Finger F, Müller T, Linner M (2019) New evidence for the prograde and retrograde PT-path of high-pressure granulites, Moldanubian Zone, Lower Austria, by Zr-in-rutile thermometry and garnet diffusion modelling. *Lithos* 342:420–439. <https://doi.org/10.1016/j.lithos.2019.05.041>
- Schauberger P, Walker A (2020) openxlsx: read, write and edit xlsx files. R package version 4.2.3. <https://CRAN.R-project.org/package=openxlsx>. Accessed 14 Sept 2020
- Schönig J, Meinhold G, von Eynatten H, Lünsdorf NK (2018a) Tracing ultrahigh-pressure metamorphism at the catchment scale. *Sci Rep* 8:2931. <https://doi.org/10.1038/s41598-018-21262-8>
- Schönig J, Meinhold G, von Eynatten H, Lünsdorf NK (2018b) Provenance information recorded by mineral inclusions in detrital garnet. *Sed Geol* 376:32–49. <https://doi.org/10.1016/j.sedgeo.2018.07.009>
- Schönig J, von Eynatten H, Meinhold G, Lünsdorf NK (2019) Diamond and coesite inclusions in detrital garnet of the Saxonian Erzgebirge, Germany. *Geology* 47:715–718. <https://doi.org/10.1130/G46253.1>
- Schönig J, von Eynatten H, Meinhold G, Lünsdorf NK, Willner AP, Schulz B (2020) Deep subduction of felsic rocks hosting UHP lenses in the central Saxonian Erzgebirge: implications for UHP terrane exhumation. *Gondwana Res* 87:320–329. <https://doi.org/10.1016/j.gr.2020.06.020>
- Schulze DJ (2003) A classification scheme for mantle-derived garnets in kimberlite: a tool for investigating the mantle and exploring for diamonds. *Lithos* 71:195–213. [https://doi.org/10.1016/S0024-4937\(03\)00113-0](https://doi.org/10.1016/S0024-4937(03)00113-0)
- Selverstone J (1993) Micro- to macroscale interactions between deformational and metamorphic processes, Tauern window, Eastern Alps. *Schweiz Mineral Petrogr Mitt* 73:229–239
- Seman S, Stockli DF, McLean NM (2017) U-Pb geochronology of grossular-andradite garnet. *Chem Geol* 460:106–116. <https://doi.org/10.1016/j.chemgeo.2017.04.020>
- Suggate SM, Hall R (2014) Using detrital garnet compositions to determine provenance: a new compositional database and procedure. In: Scott RA, Smyth HR, Morton AC, Richardson N (eds) *Sediment provenance studies in hydrocarbon exploration and production*, vol 386. Geological Society London Special Publications, pp 373–393. <https://doi.org/10.1144/SP386.8>
- Teraoka Y, Suzuki M, Kawakami K (1998) Provenance of cretaceous and paleogene sediments in the Median Zone of Southwest Japan. *Bull Geol Soc Japan* 49:395–411 (in Japanese with English abstract)
- Thomson AR, Kohn SC, Prabhu A, Walter MJ (2021) Evaluating the formation pressure of diamond-hosted majoritic garnets: a machine learning majorite barometer. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2020JB020604>
- Tolosana-Delgado R, von Eynatten H, Krippner A, Meinhold G (2018) A multivariate discrimination scheme of detrital garnet chemistry for use in sedimentary provenance analysis. *Sed Geol* 375:14–26. <https://doi.org/10.1016/j.sedgeo.2017.11.003>
- Tracy RJ, Robinson P, Thompson AB (1976) Garnet composition and zoning in the determination of temperature and pressure of metamorphism, central Massachusetts. *American Mineralogist* 61:762–775. <https://doi.org/10.1016/j.cageo.2006.11.017>
- Tracy RJ (1982) Compositional zoning and inclusions in metamorphic minerals. In: Ferry JM (ed) *Characterization of metamorphism through mineral equilibria*. Reviews in Mineralogy and Geochemistry 10:355–398. <https://doi.org/10.1515/9781501508172-013>
- van den Boogaart KG, Tolosana-Delgado R (2008) “Compositions”: a unified R package to analyze compositional data. *Comput Geosci* 34:320–338. <https://doi.org/10.1016/j.cageo.2006.11.017>
- Venables WN, Ripley BD (2002) *Modern applied statistics with S*. Springer-Verlag, New York. <https://doi.org/10.1007/978-0-387-21706-2>
- von Eynatten H, Dunkl I (2012) Assessing the sediment factory: the role of single grain analysis. *Earth Sci Rev* 115:97–120. <https://doi.org/10.1016/j.earscirev.2012.08.001>
- von Eynatten H, Gaupp R (1999) Provenance of Cretaceous synorogenic sandstones in the Eastern Alps: constraints from framework petrography, heavy mineral analysis and mineral chemistry. *Sed Geol* 124:81–111. [https://doi.org/10.1016/S0037-0738\(98\)00122-5](https://doi.org/10.1016/S0037-0738(98)00122-5)
- von Quadt A, Günther D, Frischknecht R (1997) The evolution of pre-Variscan eclogites of the Tauern Window (Eastern Alps): a Sm/Nd-, conventional and Laser ICP-MS zircon U-Pb study. *Schweiz Mineral Petrogr Mitt* 77:265–279
- Wickham H, Bryan J (2019) readxl: Read Excel Files. R package version 1.3.1. <https://CRAN.R-project.org/package=readxl>. Accessed 13 Mar 2019
- Wickham H, François R, Henry L, Müller K (2020) dplyr: A grammar of data manipulation. R package version 1.0.2. <https://CRAN.R-project.org/package=dplyr>. Accessed 12 Aug 2020
- Wilke CO (2020) cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1. <https://CRAN.R-project.org/package=cowplot>. Accessed 15 Dec 2020
- Zhang Z, Saxena SK (1991) Thermodynamic properties of andradite and application to skarn with coexisting andradite and hedenbergite. *Contrib Mineral Petrol* 107:255–263
- Ziegler A, König IR (2014) Mining data with random forests: current options for real-world applications. *Wiley Interdiscip Rev Data Min Knowl Discov* 4:55–63. <https://doi.org/10.1002/widm.1114>