

---

# **pyrolite Documentation**

***Release 0.3.5.post0+2.gd3f5b51***

**Morgan Williams**

**Feb 28, 2024**



## GETTING STARTED



## INTRODUCTION

pyrolite is a set of tools for making the most of your geochemical data.

The python package includes functions to work with compositional data, to transform geochemical variables (e.g. elements to oxides), functions for common plotting tasks (e.g. spiderplots, ternary diagrams, bivariate and ternary density diagrams), and numerous auxiliary utilities.

- On this site you can browse the [API](#), or look through some of the [usage examples](#).
- There's also a quick [installation guide](#), a list of [recent changes](#) and some notes on where the project is heading in the near [future](#).
- If you're interested in [contributing to the project](#), there are many potential avenues, whether you're experienced with python or not.

---

**Note:** pyrolite has been [published in the Journal of Open Source Software](#), and a recent publication focusing on using *lambdas* and tetrads to parameterise Rare Earth Element patterns has been [published in Mathematical Geosciences](#)!

---



## WHY PYROLITE?

The name *pyrolite* is an opportunistic repurposing of a term used to describe an early model mantle composition proposed by Ringwood [?], comprised principally of **pyr**-oxene & **ol**-ivine. While the model certainly hasn't stood the test of time, the approach optimises the aphorism "All models are wrong, but some are useful" [?]. It is with this mindset that pyrolite is built, to better enable you to make use of your geochemical data to build and test geological models.





## GETTING STARTED

### 3.1 Installation

pyrolite is available on [PyPi](#), and can be downloaded with pip:

```
pip install --user pyrolite
```

---

**Note:** With pip, a user-level installation is generally recommended (i.e. using the `--user` flag), especially on systems which have a system-level installation of Python (including *\*.nix*, WSL and MacOS systems); this can avoid some permissions issues and other conflicts. For detailed information on other pip options, see the relevant [docs](#). pyrolite is not yet packaged for Anaconda, and as such `conda install pyrolite` will not work.

---

#### 3.1.1 Upgrading pyrolite

New versions of pyrolite are released frequently. You can upgrade to the latest edition on [PyPi](#) using the `--upgrade` flag:

```
pip install --upgrade pyrolite
```

**See also:**

[Development Installation](#)

#### 3.1.2 Optional Dependencies

Optional dependencies (*dev*, *skl*, *spatial*, *db*, *stats*, *docs*) can be specified during *pip* installation. For example:

```
pip install --user pyrolite[stats]
```

```
pip install --user pyrolite[dev,docs]
```

## 3.2 Getting Started

---

**Note:** This page is under construction. Feel free to send through suggestions or questions.

---

### 3.2.1 Getting Set Up

Before you can get up and running with pyrolite, you'll need to have a distribution of Python installed. If you're joining the scientific Python world, chances are that the [Anaconda distributions](#) are a good match for you<sup>0</sup>. Check PyPI for the most up to date information regarding pyrolite compatibility, but as of writing this guide, pyrolite is known to work well with Python 3.5, 3.6 and 3.7.

When it comes to how you edit files and interact with Python, there are now many different choices (especially for editors), and choosing something which allows you to work the way you wish to work is the key aspect. While you can certainly edit python files (.py) in any text editor, choosing an editor designed for the task makes life easier. Choosing one is subjective - know that many exist, and perhaps try a few. Integrated Development Environments (IDEs) often allow you to quickly edit and run code within the same window (e.g. [Spyder](#), which is typically included in the default Anaconda distribution). Through notebooks and related ideas the [Jupyter](#) ecosystem has broadened how people are interacting with code across multiple languages, including Python. For reference, pyrolite has been developed principally in [Atom](#), leveraging the [Hydrogen](#) package to provide an interactive coding environment using Jupyter.

Finally, consider getting up to speed with simple Git practises for your projects and code such that you can keep versioned histories of your analyses, and have a look at hosted repository services (e.g. [GitHub](#), [GitLab](#)). These hosted repositories together with integrated services are often worth taking advantage of (e.g. hosting material and analyses from papers, posters or presentation, and linking this through to [Zenodo](#) to get an archived version with a DOI).

### 3.2.2 Installing pyrolite

There's a separate page dedicated to [pyrolite installations](#), but for most purposes, the best way to install pyrolite is through opening a terminal (an Anaconda terminal, if that's the distribution you're using) and type:

```
pip install pyrolite
```

To keep pyrolite up to date (new versions are released often), periodically run the following to update your local version:

```
pip install --upgrade pyrolite
```

### 3.2.3 Writing Some Code

If you're new to Python<sup>†0</sup>, or just new to pyrolite, checking out some of the examples is a good first step. You should be able to copy-paste or download and run each of these examples as written on your own system, or alternatively you can run them interactively in your browser thanks to [Binder](#) and [sphinx-gallery](#) (check for the links towards the bottom of each page). Have a play with these, and adapt them to your own purposes.

---

<sup>0</sup> If you're strapped for space, or are bloat-averse, you could also consider using [Anaconda's miniconda distributions](#).

<sup>†</sup> If you're completely new to Python, check out some of the many free online courses to get up to scratch with basic Python concepts, data structures and get in a bit of practice writing code (e.g. the basic Python course on [Codecademy](#)). Knowing your way around some of these things before you dive into applying them can help make it a much more surmountable challenge. Remember that the pyrolite community is also around to help out if you get stuck, and we all started from a similar place! There are no 'stupid questions', so feel free to ping us on [Gitter](#) with any questions or aspects that are proving particularly challenging.

### 3.2.4 Importing Data

A large part of the pyrolite API is based around [pandas](#) DataFrames. One of the first hurdles for new users is importing their own data tables. To make this as simple as possible, it's best to organise - or 'tidy' - your data tables<sup>0</sup>. Minimise unnecessary whitespace, and where possible make sure your table columns are the first row of your table. In most cases, where these data are in the form of text or Excel files, the typical steps for data import are similar. A few simple examples are given below.

To import a table from a .csv file:

```
from pathlib import Path
import pandas as pd

filepath = Path('./mydata.csv')
df = pd.read_csv(filepath)
```

In the case of an excel table:

```
filepath = Path('./mydata.xlsx')
df = pd.read_excel(filepath)
```

There is also a pyrolite function which abstracts away these differences by making a few assumptions, and enables you to import the table from either a csv or excel file:

```
from pyrolite.util.pd import read_table
df = read_table(filepath)
```

### 3.2.5 Gitter Community

A Gitter Community has been set up to serve as a landing page for conversations, questions and support regarding the pyrolite python package and related activities. Here we hope to capture questions and bugs from the community such that these can be addressed quickly, and we can ensure pyrolite and its documentation are as useful as possible to the community. Please feel free to use this space to:

- Ask questions or seek help about getting started with pyrolite or particular pyrolite features
- Get tips for troubleshooting
- Discuss the general development of pyrolite
- Ask about contributing to pyrolite

Items which are related to specific aspects of pyrolite development (requesting a feature, or reporting an identified bug) are best coordinated through [GitHub](#), but feel free to touch base here first. See below and the [contribution](#) and [development](#) guides for more information.

Note that users and contributors in online spaces (including Gitter) are expected to adhere to the [Code of Conduct](#) of this project (and any other guidelines of the relevant platform).

---

<sup>0</sup> Where each variable is a column, and each observation is a row. If you're unfamiliar with the 'Tidy Data' concept, check out [?].

### 3.2.6 Bugs, Debugging & Bug Reporting

This section provides some guidance for what to do when things don't work as expected, and some tips for debugging issues and common errors associated with pyrolite. Note that the scope of these suggestions is necessarily limited, and specific issues which relate more to `matplotlib.pyplot`, `pandas`, and `numpy` are often well documented elsewhere online.

- Checked the documentation, had a look for FAQ and examples here and still stuck?

The maintainers are happy to answer questions and help you solve small bugs. It's useful to know where people get stuck so we can modify things where useful, and this is an easy way to contribute to the project. Consider posting a question on [Gitter](#), and if you think it's something others may run into or could be a problem related to use of another package, perhaps also consider posting a question on [stackoverflow](#) for visibility.

- Think it's a bug or problem with pyrolite specifically?

Some guidelines for reporting issues and bugs are given in the [contributing guide](#).

See also:

[Examples](#), [API](#), [Changelog](#), [Code of Conduct](#)

## 3.3 Examples

This example gallery includes a variety of examples for using pyrolite which you can copy, download and alter, or run on Binder.

### 3.3.1 Plotting Examples

pyrolite provides some functionality for basic plotting of geochemical data in the form of spidergrams (`pyrolite.plot.spider`), ternary diagrams (`pyrolite.plot.tern`) and density diagrams (i.e. 2D histograms, `pyrolite.plot.density`).

#### Stem Plots

Stem plots are commonly used to visualise discrete distributions of data, and are useful to highlight discrete observations where the precision of values along one axis is high (e.g. an independent spatial measure like depth) and the other is less so (such that the sampling frequency along this axis is important, which is not emphasised by a scatter plot).

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

from pyrolite.plot import pyroplot
from pyrolite.plot.stem import stem
```

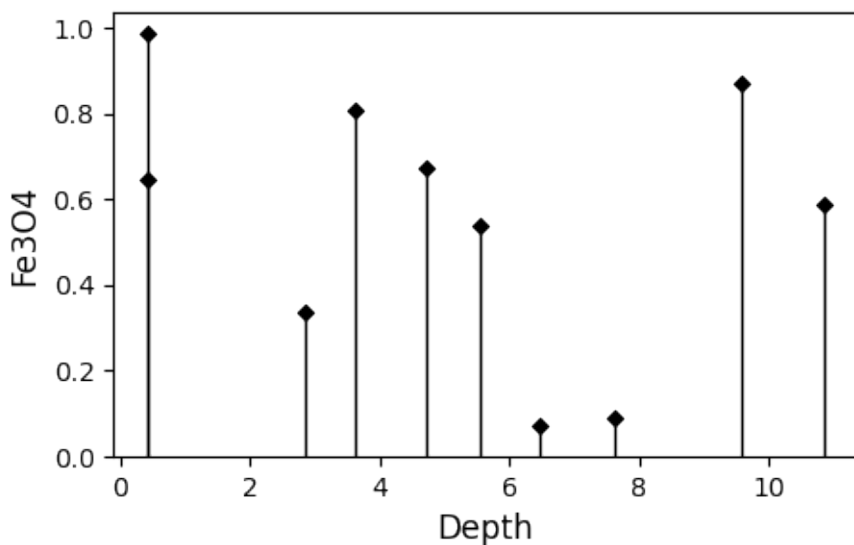
```
np.random.seed(82)
```

First let's create some example data:

```
x = np.linspace(0, 10, 10) + np.random.randn(10) / 2.0
y = np.random.rand(10)
df = pd.DataFrame(np.vstack([x, y]).T, columns=["Depth", "Fe304"])
```

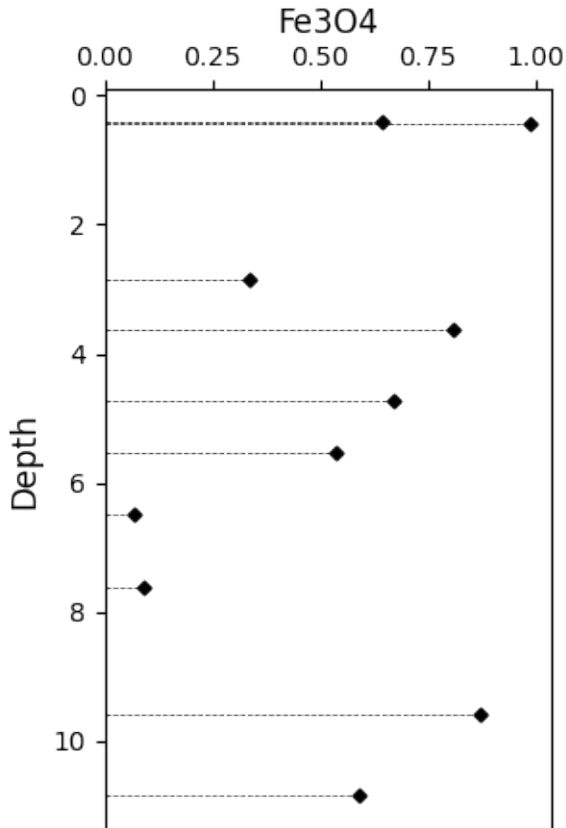
A minimal stem plot can be constructed as follows:

```
ax = df.pyroplot.stem(color="k", figsize=(5, 3))
```



Stem plots can also be used in a vertical orientation, such as for visualising discrete observations down a drill hole:

```
ax = df.pyroplot.stem(
    orientation="vertical",
    s=12,
    linestyle="--",
    linewidth=0.5,
    color="k",
    figsize=(3, 5),
)
# the yaxes can then be inverted using:
ax.invert_yaxis()
# and if you'd like the xaxis to be labeled at the top:
ax.xaxis.set_ticks_position("top")
ax.xaxis.set_label_position("top")
```



Total running time of the script: (0 minutes 0.497 seconds)

## Ternary Plots

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

from pyrolite.plot import pyroplot

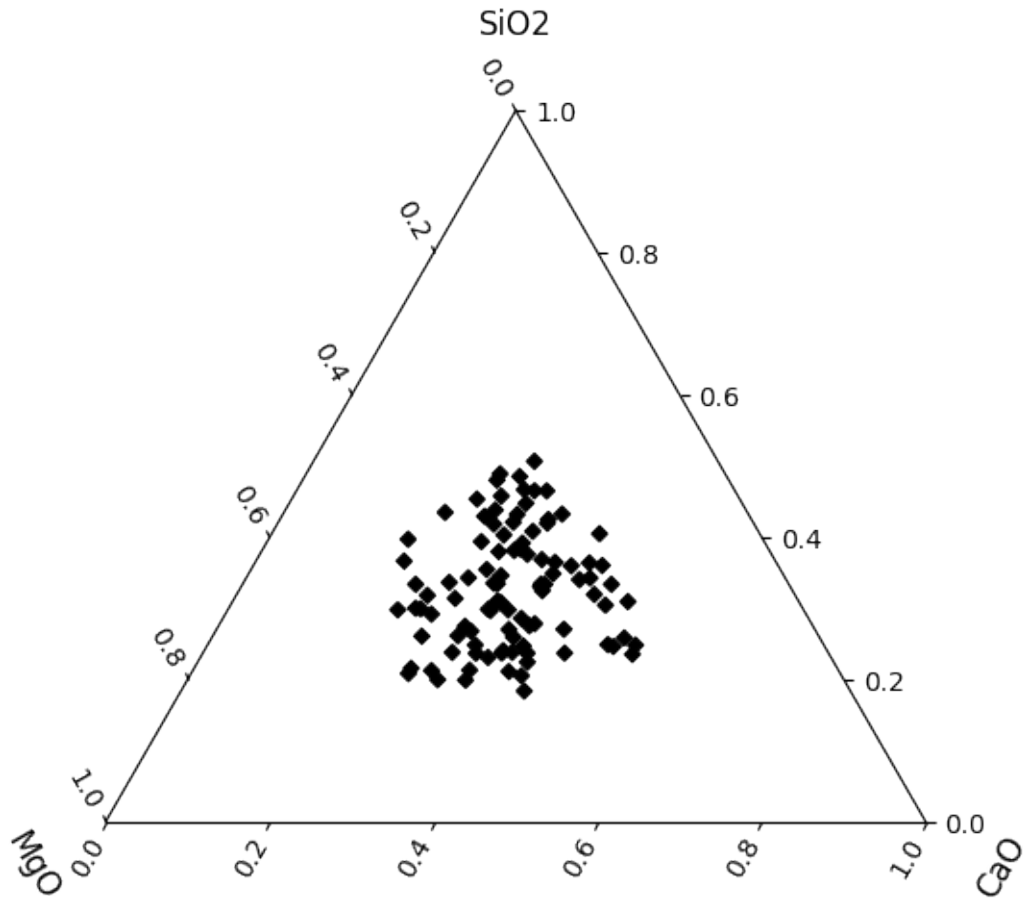
np.random.seed(82)
```

Let's first create some example data:

```
df = pd.DataFrame(data=np.exp(np.random.rand(100, 3)), columns=["SiO2", "MgO", "CaO"])
df.loc[:, ["SiO2", "MgO", "CaO"]].head()
```

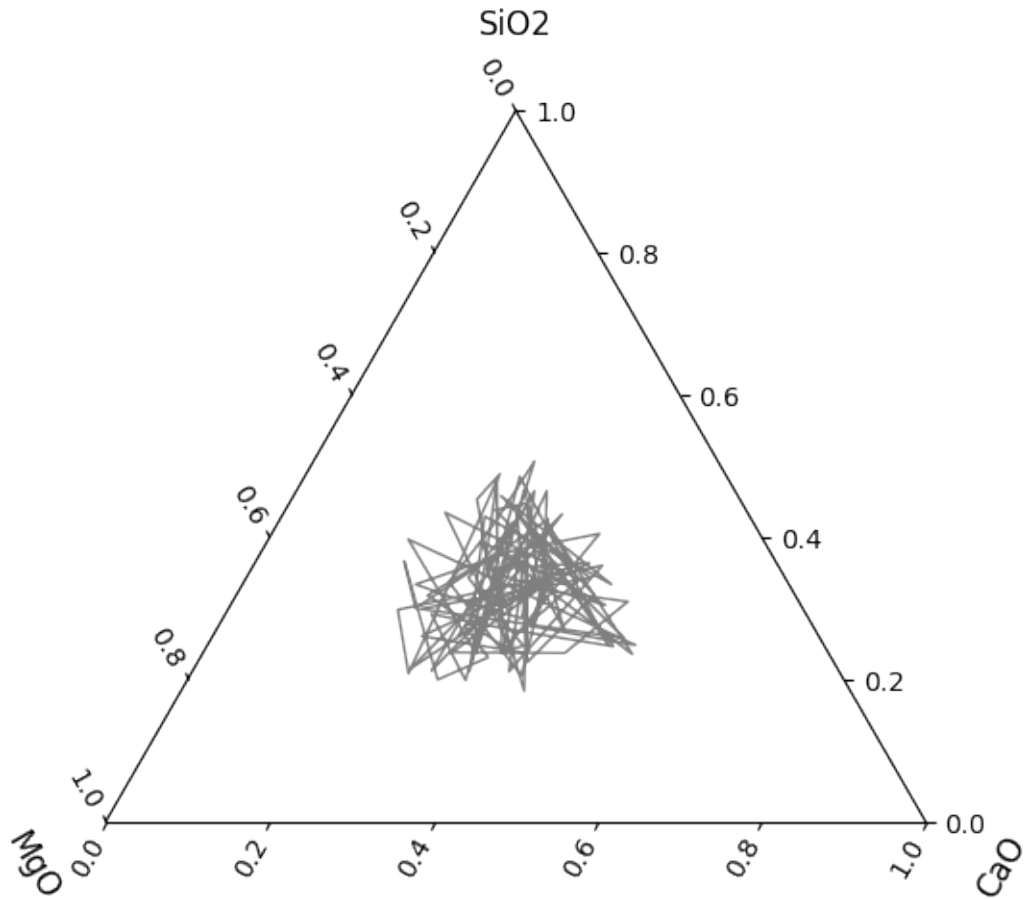
Now we can create a simple scatter plot:

```
ax = df.loc[:, ["SiO2", "MgO", "CaO"]].pyroplot.scatter(c="k")
plt.show()
```



If the data represent some continuous series, you could also plot them as lines:

```
ax = df.loc[:, ["SiO2", "MgO", "CaO"]].pyroplot.plot(color="k", alpha=0.5)
plt.show()
```



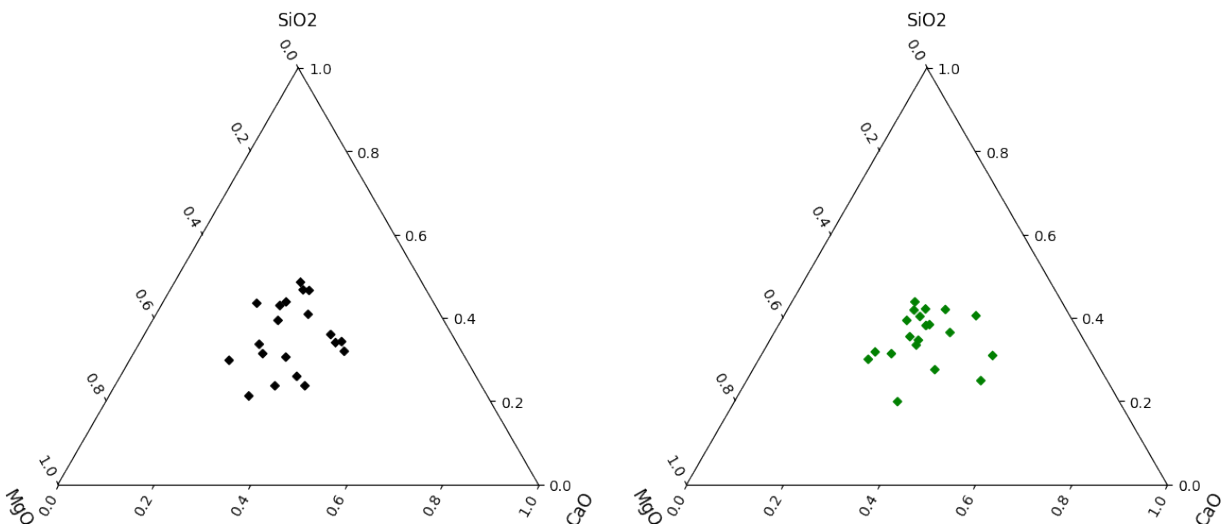
The plotting axis can be specified to use existing axes:

```
fig, ax = plt.subplots(1, 2, sharex=True, sharey=True, figsize=(12, 5))

df.loc[:, ["SiO2", "MgO", "CaO"]].sample(20).pyroplot.scatter(ax=ax[0], c="k")
df.loc[:, ["SiO2", "MgO", "CaO"]].sample(20).pyroplot.scatter(ax=ax[1], c="g")

ax = fig.orderedaxes # creating scatter plots reorders axes, this is the correct order
plt.tight_layout()
```





See also:

[Heatscatter Plots](#), [Density Plots](#), [Spider Density Diagrams](#), [Ternary Color Mapping](#)

**Total running time of the script:** (0 minutes 14.158 seconds)

### REE Radii Plots

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

from pyrolite.plot import pyroplot
```

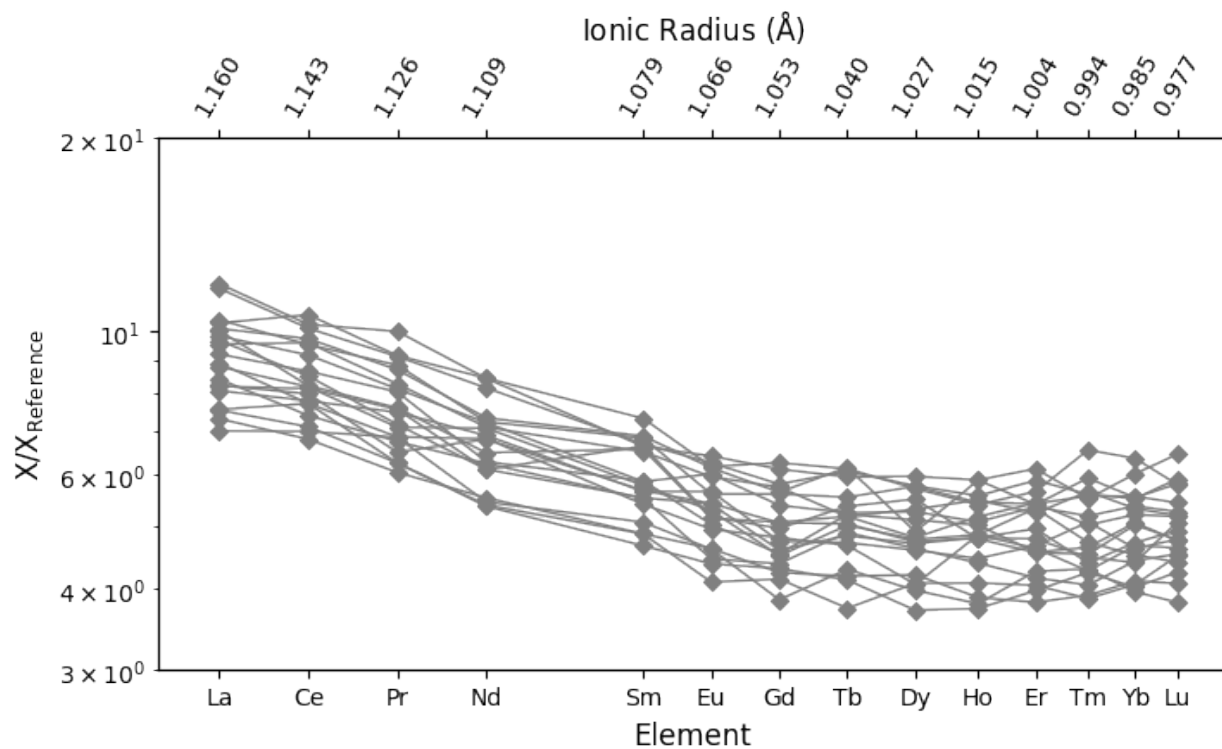
Here we generate some example data, using the `example_spider_data()` function (based on EMORB, here normalised to Primitive Mantle);

```
from pyrolite.util.synthetic import example_spider_data

df = example_spider_data(noise_level=0.1, size=20)
```

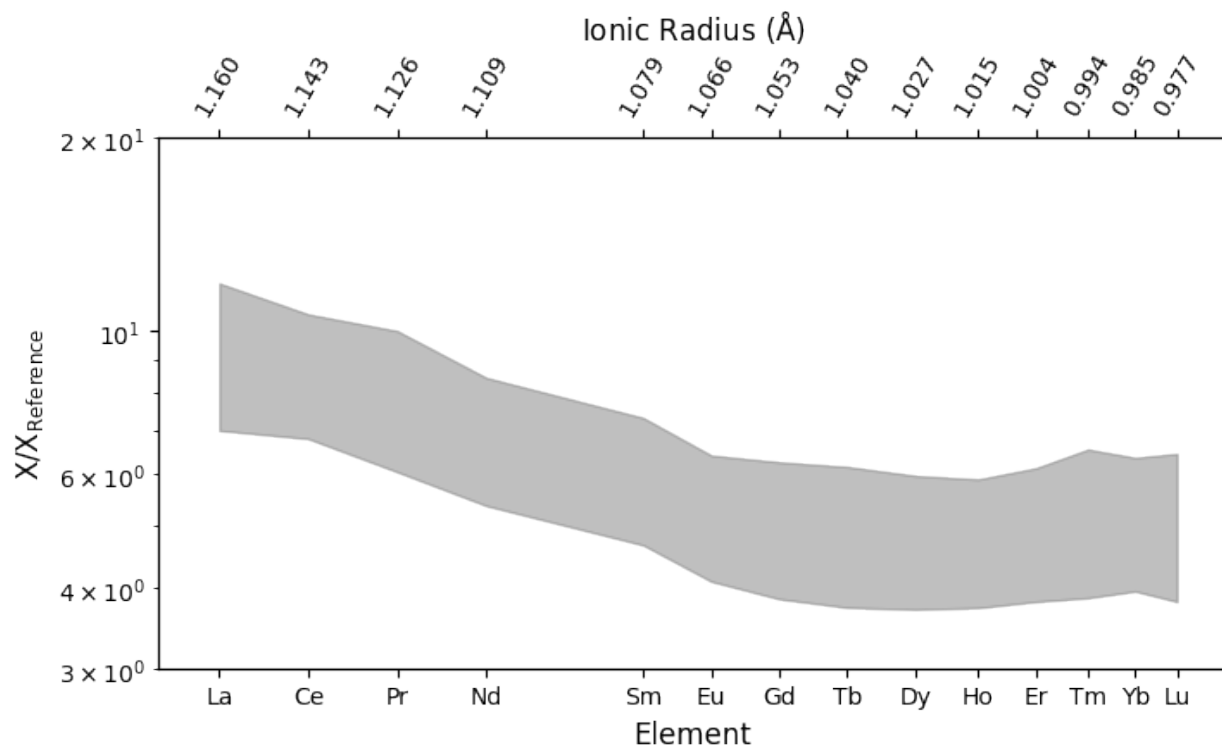
Where data is specified, the default plot is a line-based spiderplot:

```
ax = df.pyroplot.REE(color="0.5", figsize=(8, 4))
plt.show()
```



This behaviour can be modified (see spiderplot docs) to provide e.g. filled ranges:

```
df.pyroplot.REE(mode="fill", color="0.5", alpha=0.5, figsize=(8, 4))
plt.show()
```

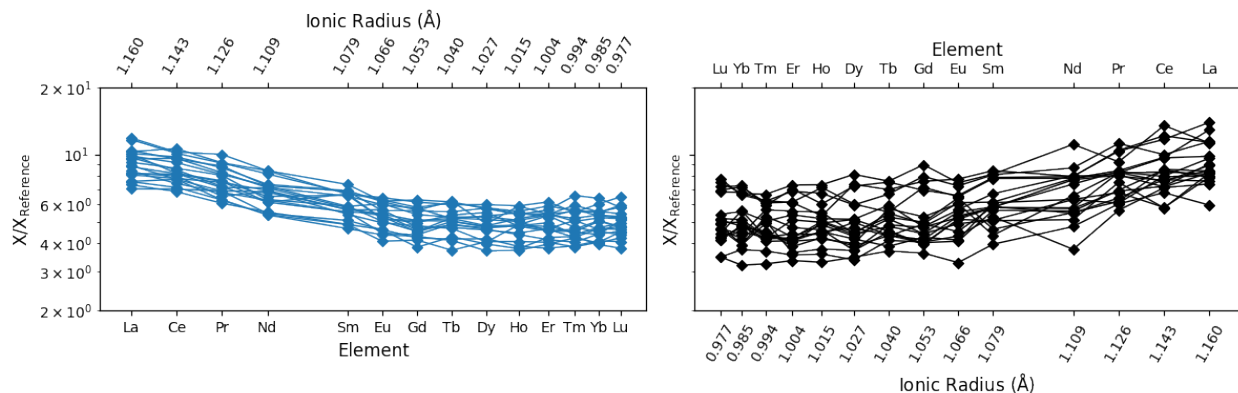


The plotting axis can be specified to use existing axes:

```
fig, ax = plt.subplots(1, 2, sharey=True, figsize=(12, 4))

df.pyroplot.REE(ax=ax[0])
# we can also change the index of the second axes
another_df = example_spider_data(noise_level=0.2, size=20) # some 'nosier' data
another_df.pyroplot.REE(ax=ax[1], color="k", index="radii")

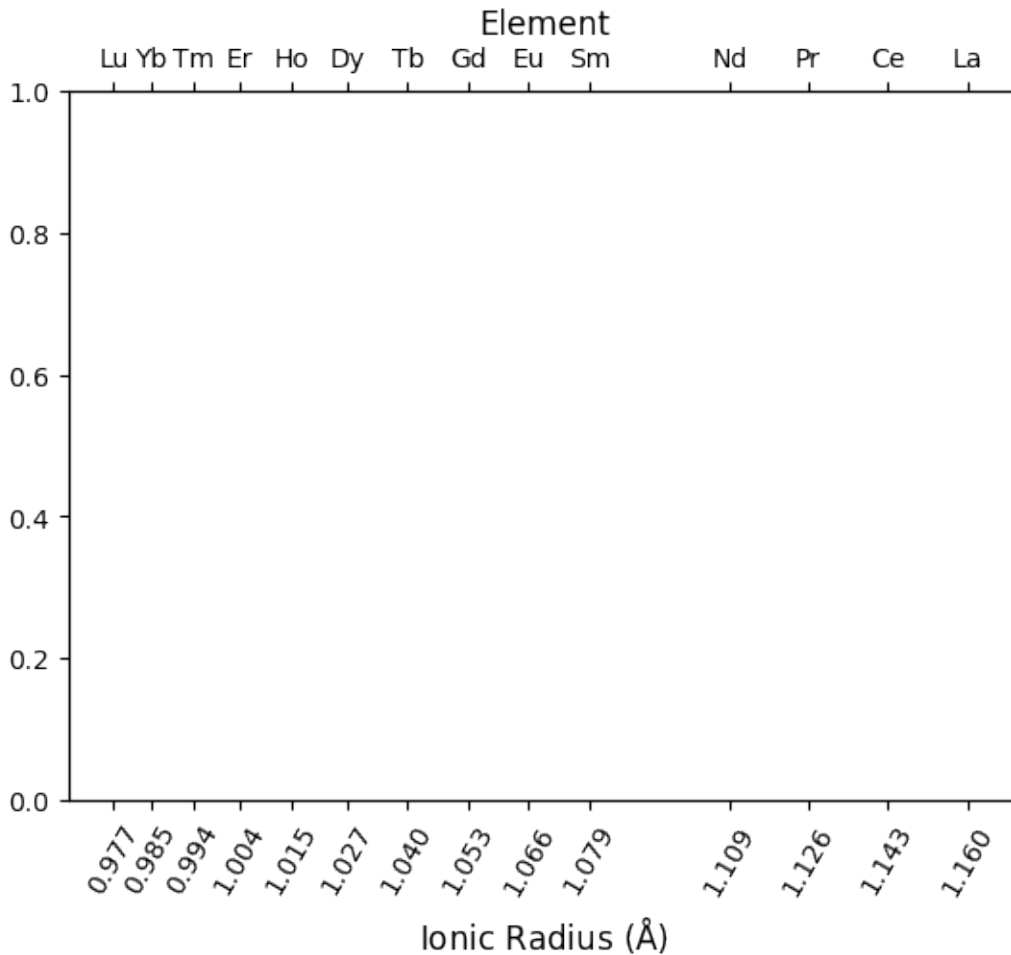
plt.tight_layout()
plt.show()
```



If you're just after a plotting template, you can use `REE_v_radii()` to get a formatted axis which can be used for subsequent plotting:

```
from pyrolite.plot.spider import REE_v_radii

ax = REE_v_radii(index="radii") # radii mode will put ionic radii on the x axis
plt.show()
```



See also:

**Examples:**

[Ionic Radii, Spider Diagrams, lambdas: Parameterising REE Profiles](#)

**Functions:**

`get_ionic_radii()`, `pyrolite.plot.pyroplot.REE()`, `pyrolite.plot.pyroplot.spider()`,  
`lambda_lnREE()`

**Total running time of the script:** (0 minutes 4.714 seconds)

## Ternary Color Systems

pyrolite includes two methods for coloring data points and polygons in a ternary system, `ternary_color()` and `color_ternary_polygons_by_centroid()` which work well with some of the plot templates (`pyrolite.plot.templates`) and associated classifiers (`pyrolite.util.classification`).

```
import numpy as np
import matplotlib.pyplot as plt
```

## Colors by Ternary Position

The `ternary_color()` function serves to generate the color which corresponds to the mixing of three colours in proportion to the components in a ternary system. By default these colours are red, green and blue (corresponding to the top, left, and right components in a ternary diagram). The function returns colours in the form of an RGBA array:

```
from pyrolite.util.plot.style import ternary_color
from pyrolite.util.synthetic import normal_frame

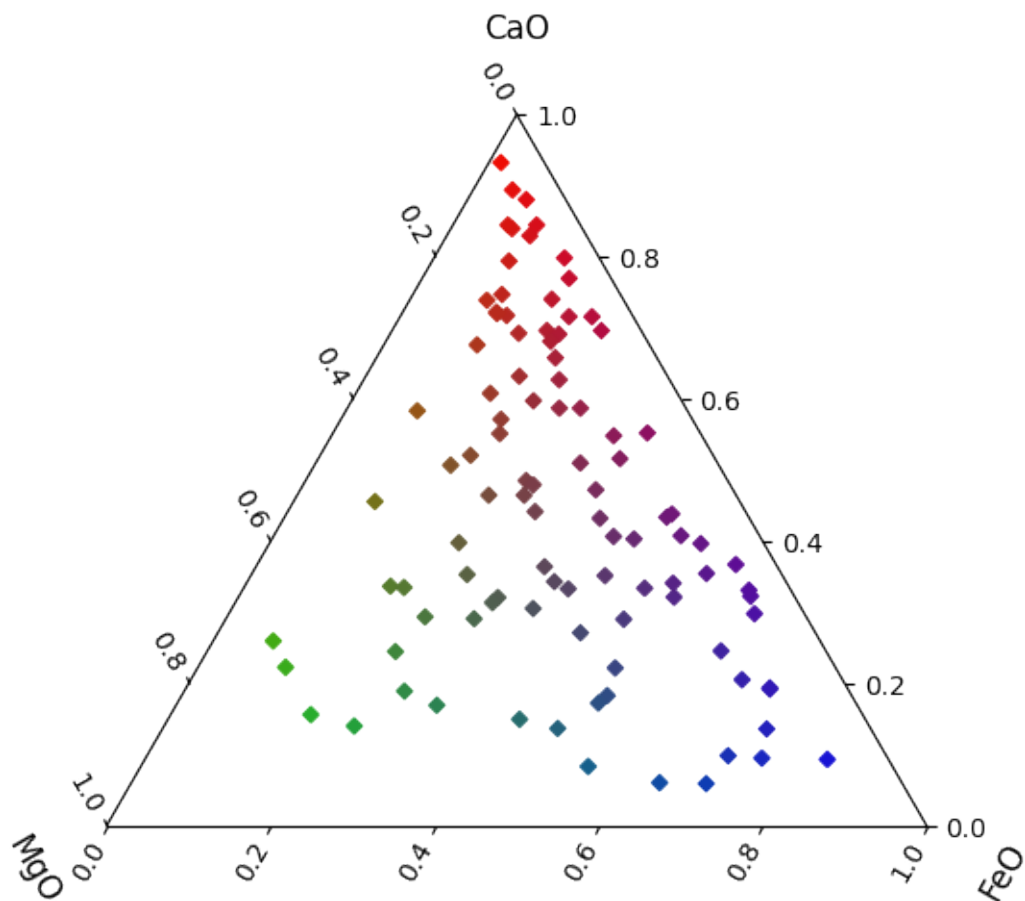
# generate a synthetic dataset we can use for the colouring example
df = normal_frame(
    columns=["CaO", "MgO", "FeO"],
    size=100,
    seed=42,
    cov=np.array([[0.8, 0.3], [0.3, 0.8]]),
)

colors = ternary_color(df)
colors[:3]
```

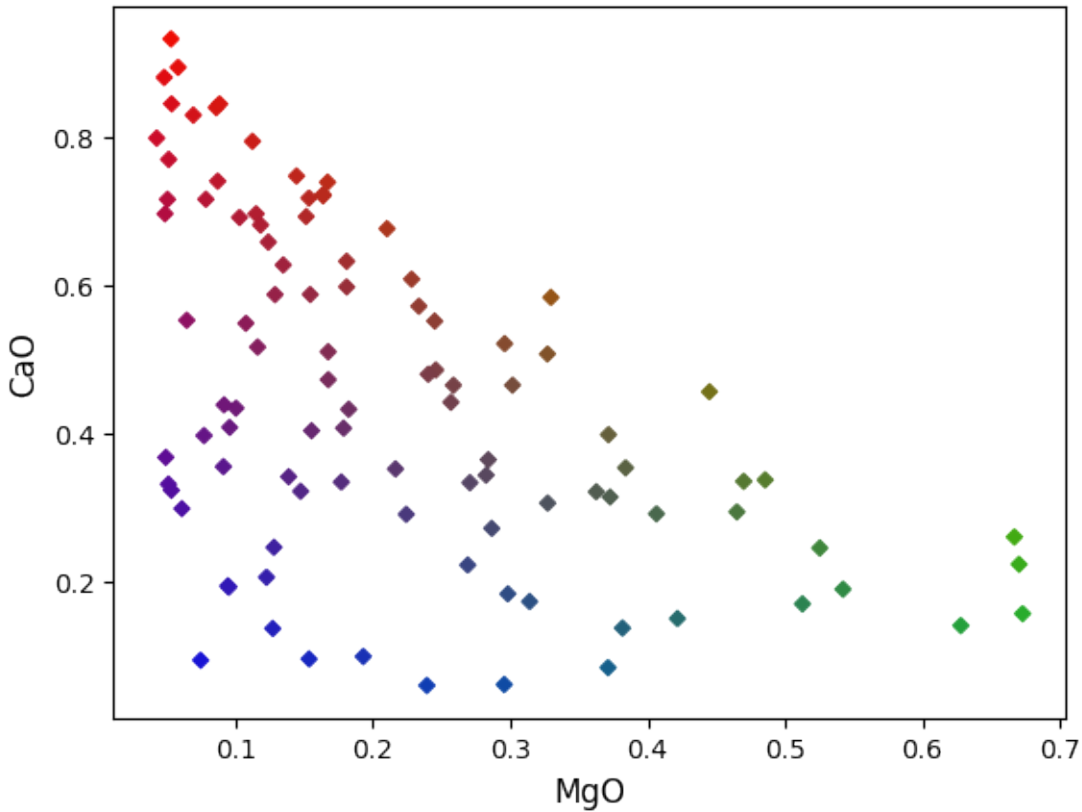
```
array([[0.18902861, 0.54222697, 0.26874443, 0.999999  ],
       [0.5097609 , 0.16739226, 0.32284684, 0.999999  ],
       [0.08310658, 0.37115312, 0.54574029, 0.999999  ]])
```

These can then be readily used in a ternary diagram (or elsewhere):

```
ax = df.pyroplot.scatter(c=colors)
plt.show()
```

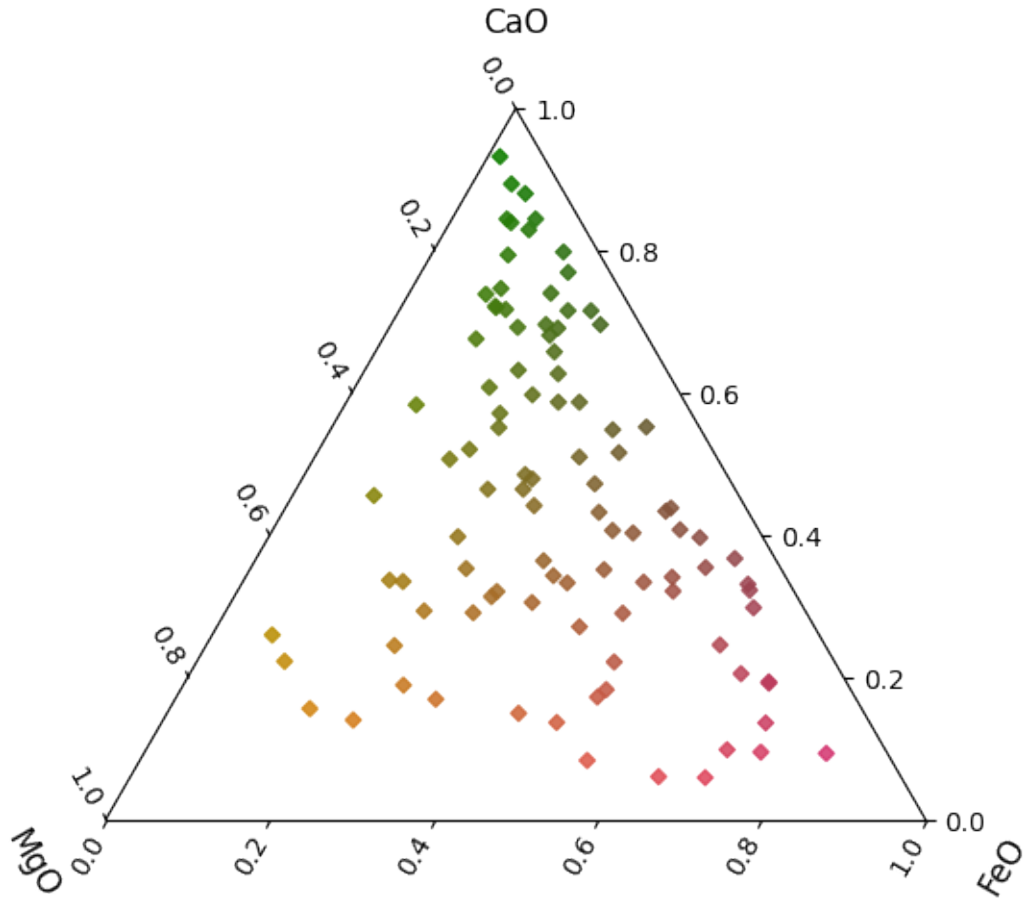


```
ax = df[["MgO", "CaO"]].pyroplot.scatter(c=colors)
plt.show()
```



You can use different colors for each of the vertices if you so wish, and mix and match named colors with RGB/RGBA representations (note that the alpha will be scaled, if it is passed as a keyword argument to `ternary_color()`):

```
colors = ternary_color(df, alpha=0.9, colors=["green", "orange", [0.9, 0.1, 0.5, 0.9]])
ax = df.pyroplot.scatter(c=colors)
plt.show()
```



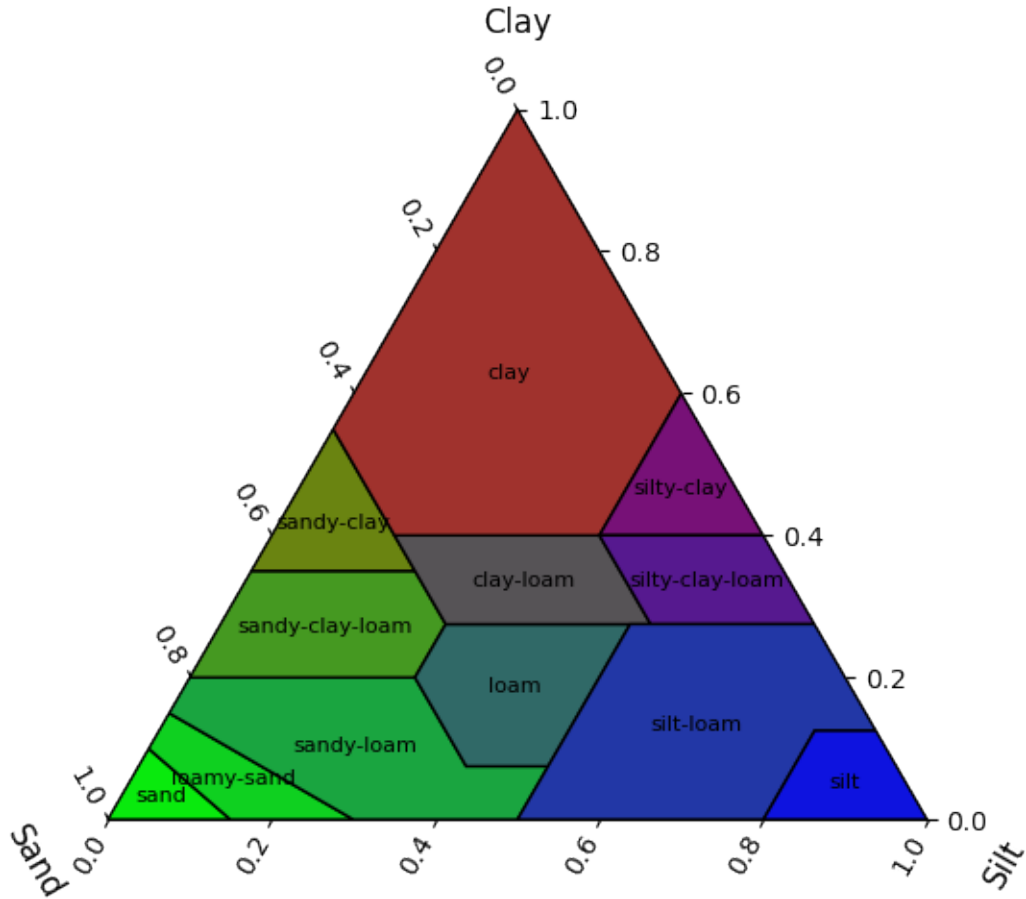
### Colors by Centroid Position

We can also colour polygons within one of these templates by the ternary combination of colours (defaulting to red, green and blue) at the polygon centroid:

```
from pyrolite.util.classification import USDAsoilTexture
from pyrolite.util.plot.style import color_ternary_polygons_by_centroid

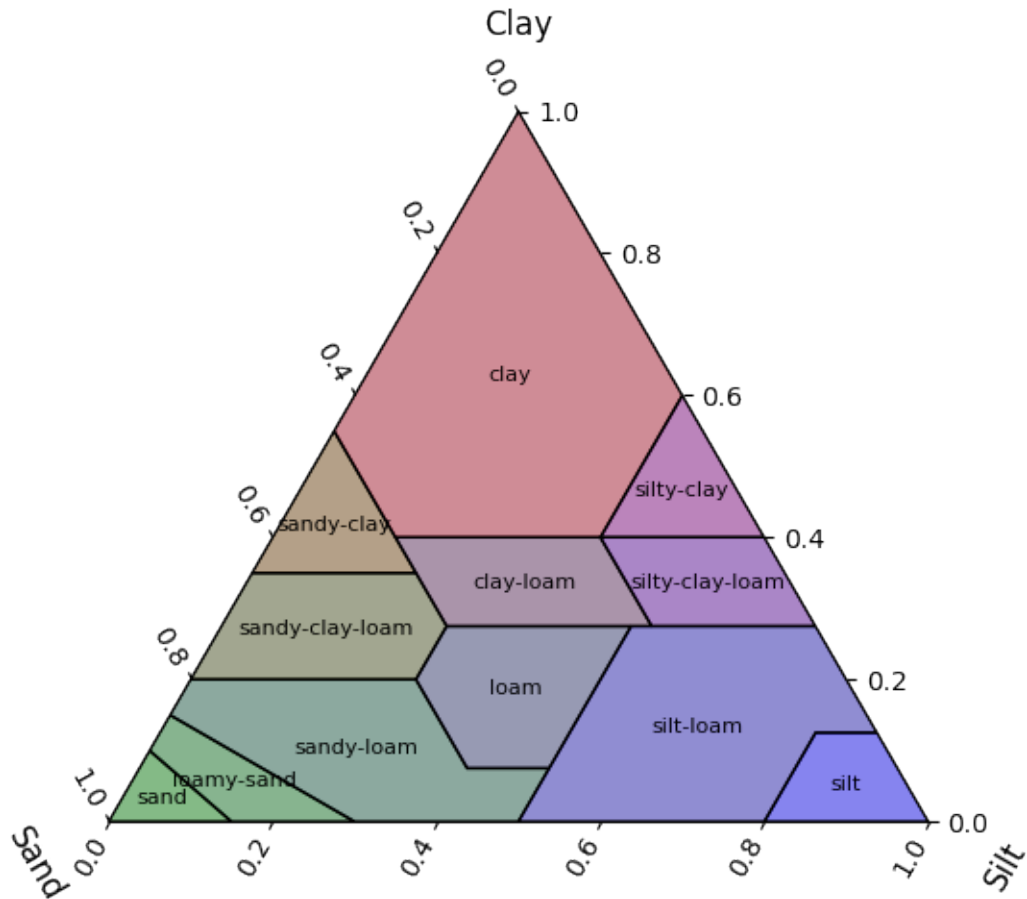
clf = USDAsoilTexture()
ax = clf.add_to_axes(ax=None, add_labels=True, figsize=(8, 8))
color_ternary_polygons_by_centroid(ax)
plt.show()
```





There are a range of options you can pass to this function to control the ternary colors (as above), change the scaling coefficients for ternary components and change the opacity of the colors:

```
color_ternary_polygons_by_centroid(
    ax, colors=("red", "green", "blue"), coefficients=(1, 1, 1), alpha=0.5
)
plt.show()
```



See also:

Examples:

[Ternary Diagrams](#), [Plot Templates](#)

**Total running time of the script:** (0 minutes 12.989 seconds)

## Parallel Coordinate Plots

Parallel coordinate plots are one way to visualise data relationships and clusters in higher dimensional data. pyrolite now includes an implementation of this which allows a handy quick exploratory visualisation.

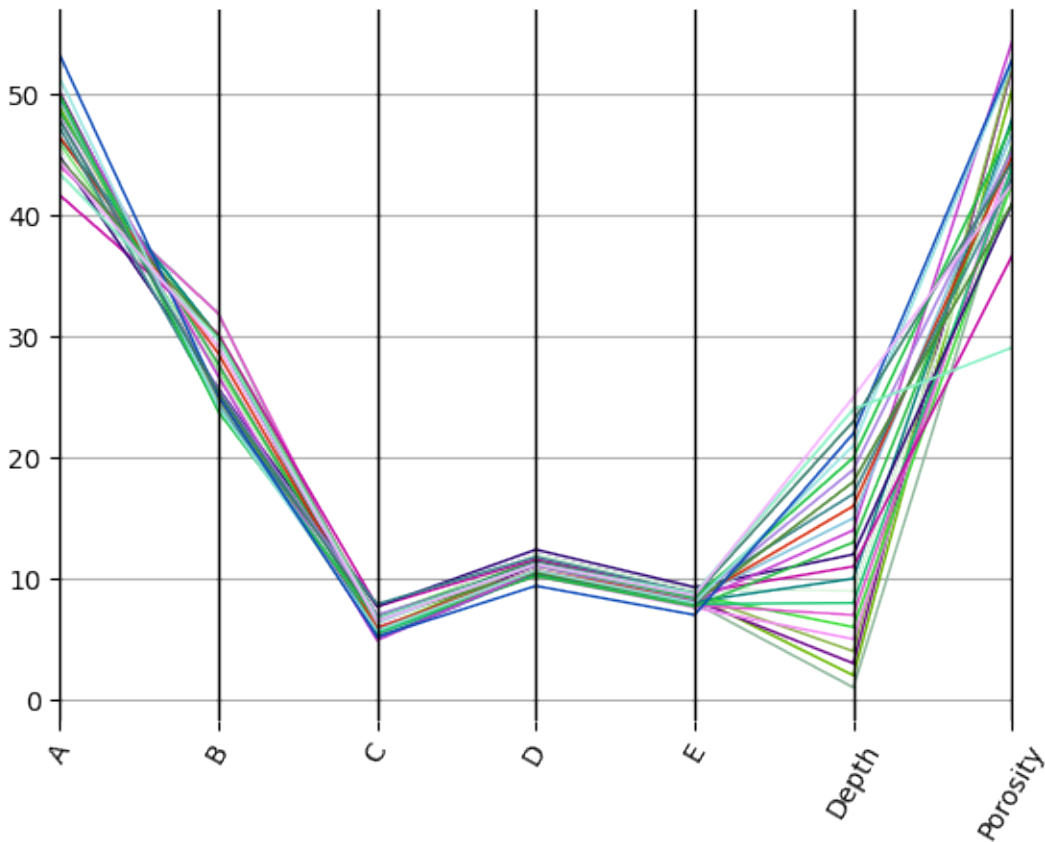
```
import matplotlib.axes
import matplotlib.pyplot as plt
from mpl_toolkits.axes_grid1 import make_axes_locatable
import numpy as np
import pandas as pd

import pyrolite.data.Aitchison
import pyrolite.plot
```

To start, let's load up an example dataset from Aitchison

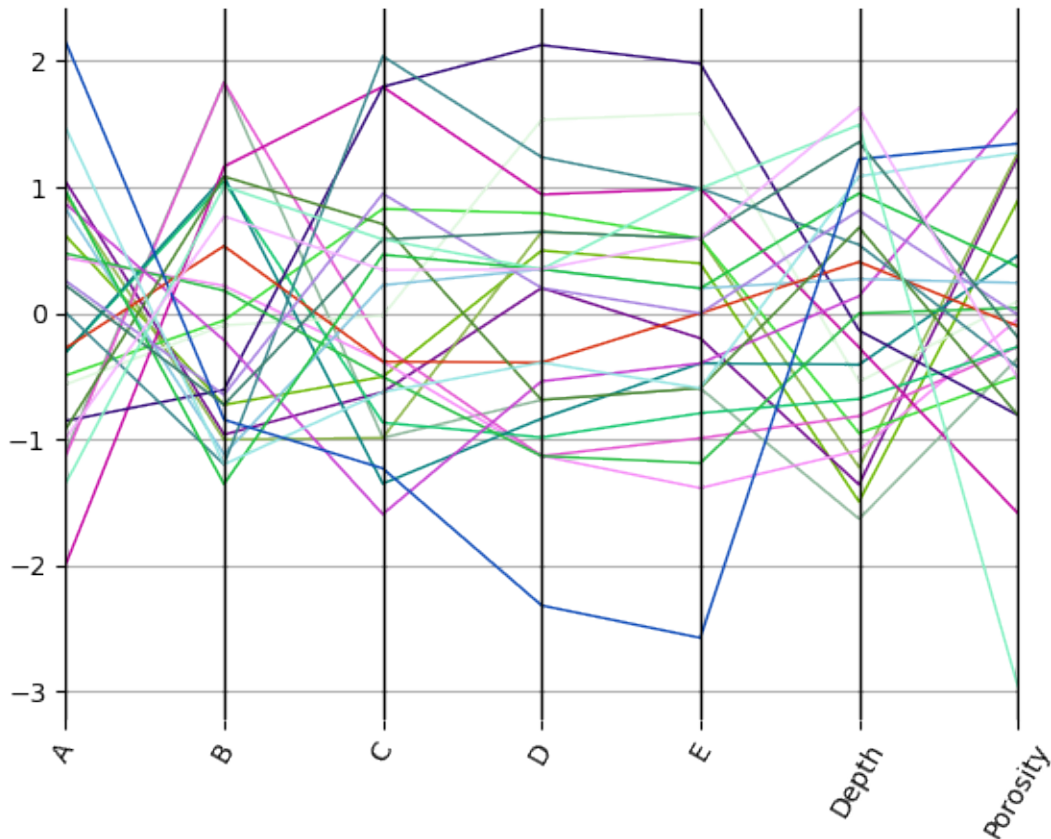
```
df = pyrolite.data.Aitchison.load_coxite()
comp = [
    i for i in df.columns if i not in ["Depth", "Porosity"]
] # compositional data variables
```

```
ax = df.pyroplot.parallel()
plt.show()
```



By rescaling this using the mean and standard deviation, we can account for scale differences between variables:

```
ax = df.pyroplot.parallel(rescale=True)
plt.show()
```

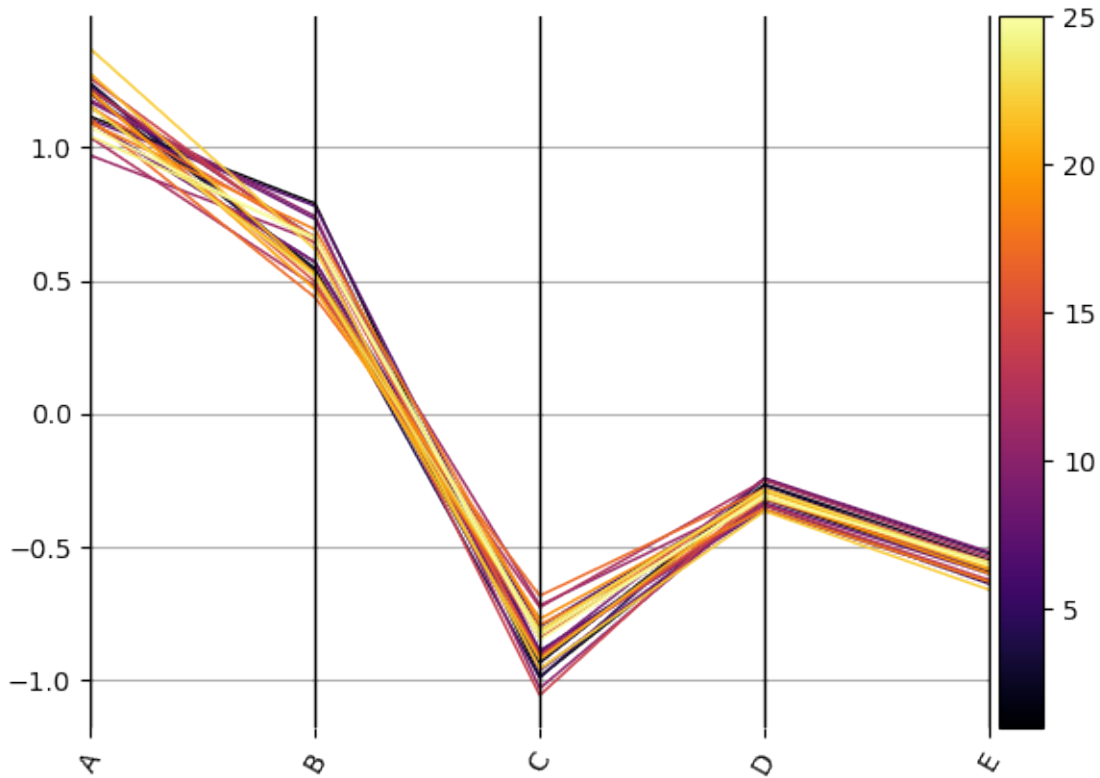


We can also use a centred-log transform for compositional data to reduce the effects of spurious correlation:

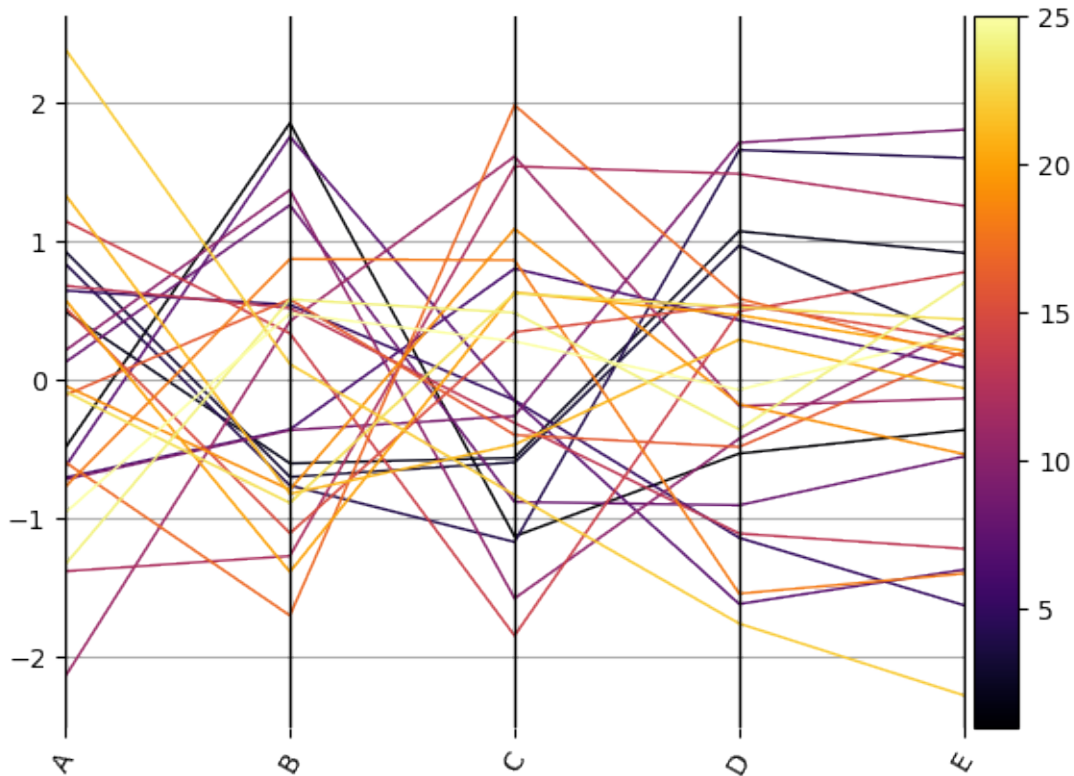
```
from pyrolite.util.skl.transform import CLRTransform

cmap = "inferno"
compdata = df.copy()
compdata[comp] = CLRTransform().transform(compdata[comp])
ax = compdata.loc[:, comp].pyroplot.parallel(color=compdata.Depth.values, cmap=cmap)
divider = make_axes_locatable(ax)
cax = divider.append_axes('right', size='5%', pad=0.05)

# we can add a meaningful colorbar to indicate one variable also, here Depth
sm = plt.cm.ScalarMappable(cmap=cmap)
sm.set_array(df.Depth)
plt.colorbar(sm, cax=cax, orientation='vertical')
plt.show()
```



```
ax = compdata.loc[:, comp].pyroplot.parallel(
    rescale=True, color=compdata.Depth.values, cmap=cmap
)
divider = make_axes_locatable(ax)
cax = divider.append_axes('right', size='5%', pad=0.05)
plt.colorbar(sm, cax=cax, orientation='vertical')
plt.show()
```



**Total running time of the script:** (0 minutes 1.170 seconds)

## Heatscatter Plots

While `density()` plots are useful summary visualizations for large datasets, scatterplots are more precise and retain all spatial information (although they can get crowded).

A scatter plot where individual points are coloured by data density in some respects represents the best of both worlds. A version inspired by similar existing visualisations is implemented with `heatscatter()`.

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

from pyrolite.plot import pyroplot

np.random.seed(12)
```

First we'll create some example data

```
from pyrolite.util.synthetic import normal_frame, random_cov_matrix

df = normal_frame(
    size=1000,
    cov=random_cov_matrix(sigmas=np.random.rand(4) * 2, dim=4, seed=12),
    seed=12,
)
```

We can compare a minimal `heatscatter()` plot to other visualisations for the same data:

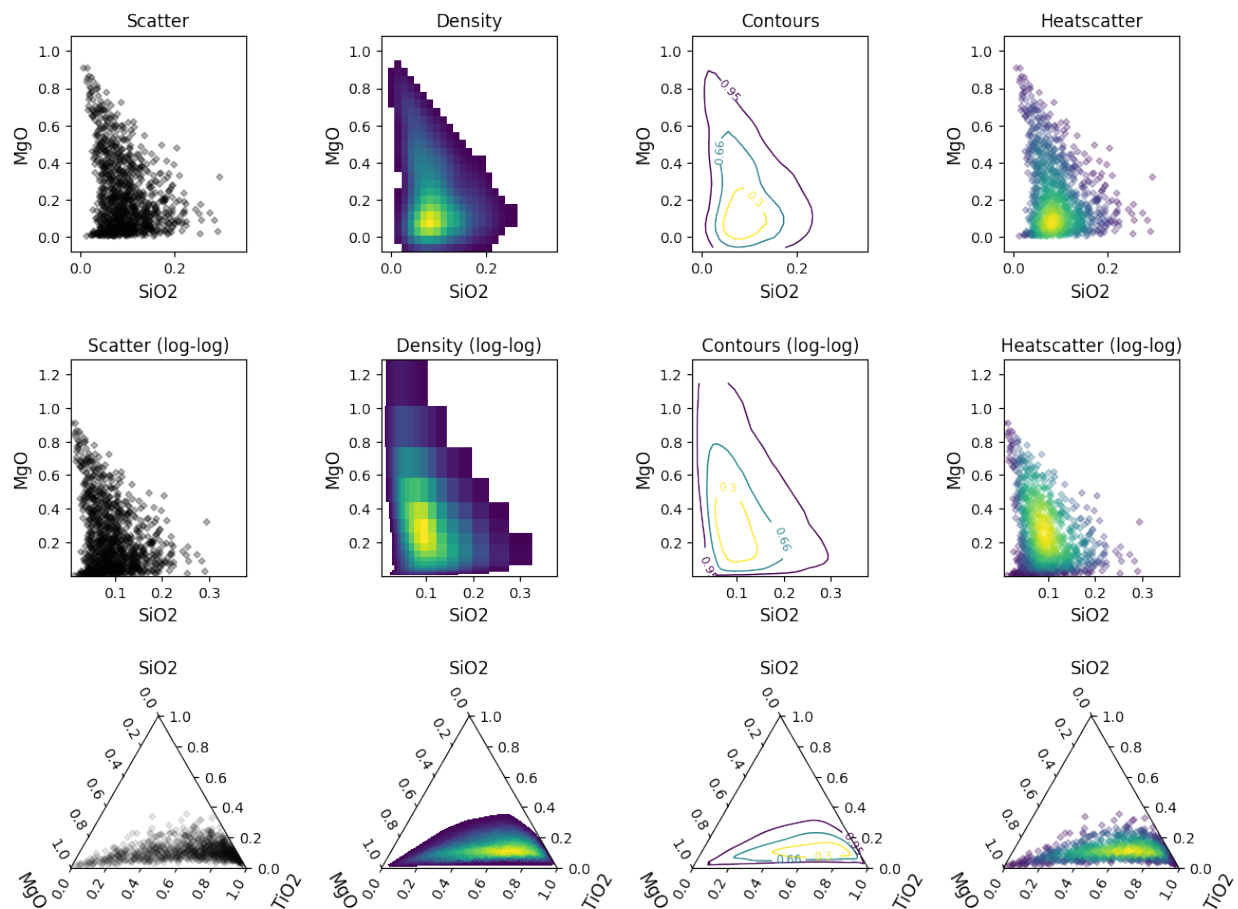
```
from pyrolite.util.plot.axes import share_axes

fig, ax = plt.subplots(3, 4, figsize=(12, 9))

ax = ax.flat
share_axes(ax[:4], which="xy")
share_axes(ax[4:8], which="xy")
share_axes(ax[8:], which="xy")

contours = [0.95, 0.66, 0.3]
bivar = ["SiO2", "MgO"]
trivar = ["SiO2", "MgO", "TiO2"]
# linear-scaled comparison
df.loc[:, bivar].pyroplot.scatter(ax=ax[0], c="k", s=10, alpha=0.3)
df.loc[:, bivar].pyroplot.density(ax=ax[1])
df.loc[:, bivar].pyroplot.density(ax=ax[2], contours=contours)
df.loc[:, bivar].pyroplot.heatscatter(ax=ax[3], s=10, alpha=0.3)
# log-log plots
df.loc[:, bivar].pyroplot.scatter(ax=ax[4], c="k", s=10, alpha=0.3)
df.loc[:, bivar].pyroplot.density(ax=ax[5], logx=True, logy=True)
df.loc[:, bivar].pyroplot.density(ax=ax[6], contours=contours, logx=True, logy=True)
df.loc[:, bivar].pyroplot.heatscatter(ax=ax[7], s=10, alpha=0.3, logx=True, logy=True)
# ternary plots
df.loc[:, trivar].pyroplot.scatter(ax=ax[8], c="k", s=10, alpha=0.1)
df.loc[:, trivar].pyroplot.density(ax=ax[9], bins=100)
df.loc[:, trivar].pyroplot.density(ax=ax[10], contours=contours, bins=100)
df.loc[:, trivar].pyroplot.heatscatter(ax=ax[11], s=10, alpha=0.3, renorm=True)
fig.subplots_adjust(hspace=0.4, wspace=0.4)

titles = ["Scatter", "Density", "Contours", "Heatscatter"]
for t, a in zip(titles + [i + " (log-log)" for i in titles], ax):
    a.set_title(t)
plt.tight_layout()
```



See also:

[Ternary Plots](#), [Density Plots](#), [Spider Density Diagrams](#)

**Total running time of the script:** (0 minutes 17.892 seconds)

## Plot Templates

pyrolite includes some ready-made templates for well-known plots. These can be used to create new plots, or add a template to an existing `matplotlib.axes.Axes`.

```
import matplotlib.pyplot as plt
```

## Bivariate Templates

First let's build a simple total-alkali vs silica ( `TAS()` ) diagram:

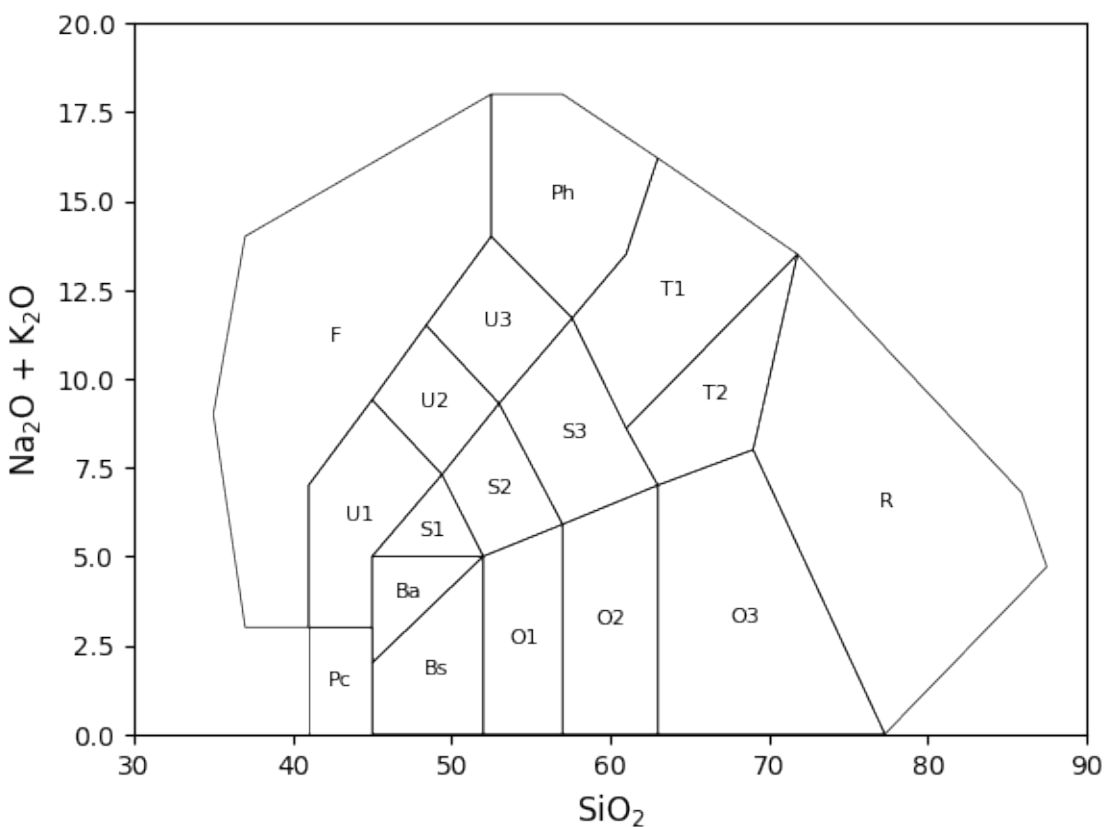
```
from pyrolite.plot.templates import TAS, SpinelFeBivariate
from pyrolite.util.plot.axes import share_axes
```

(continues on next page)



(continued from previous page)

```
ax = TAS(linewidth=0.5, add_labels=True)
plt.show()
```

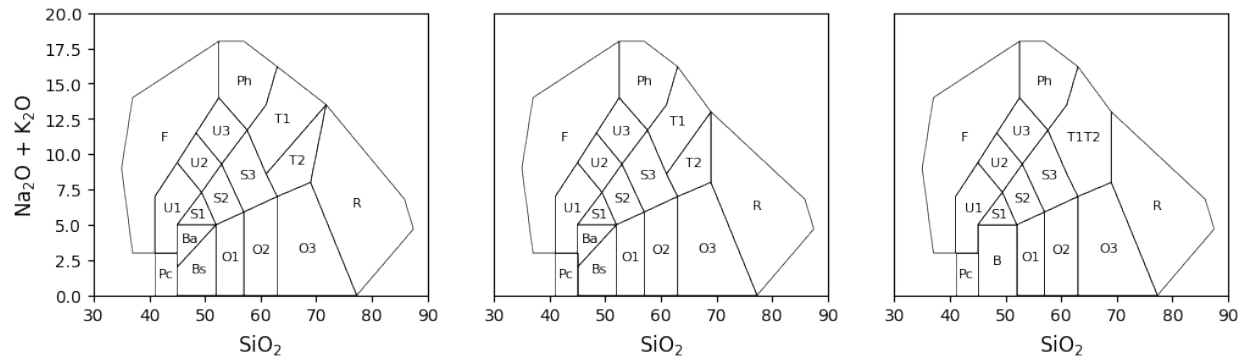


A few different variants are now available, with slightly different positioning of field boundaries, and with some fields combined:

```
fig, ax = plt.subplots(1, 3, figsize=(12, 3))

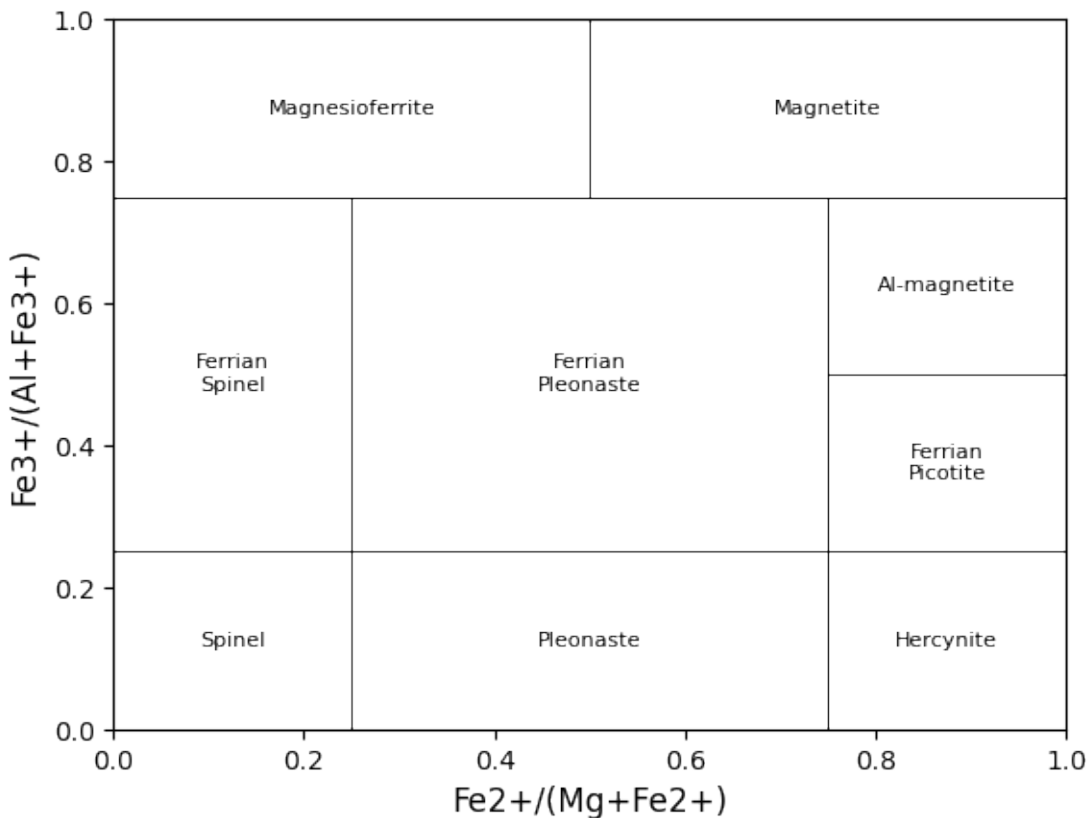
TAS(ax=ax[0], linewidth=0.5, add_labels=True, which_model=None) # Middlemost's TAS
TAS(ax=ax[1], linewidth=0.5, add_labels=True, which_model="LeMaitre") # LeMaitre's TAS
TAS(ax=ax[2], linewidth=0.5, add_labels=True, which_model="LeMaitreCombined")

for a in ax[1:]:
    a.set(yticks=[], ylabel=None)
```



For distinguishing Fe-rich variants of spinel phases, the bivariate spinel diagram can be useful:

```
ax = SpinelFeBivariate(linewidth=0.5, add_labels=True)
plt.show()
```



```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
util/plot/axes.py:219: RuntimeWarning: More than 20 figures have been opened. Figures
created through the pyplot interface (`matplotlib.pyplot.figure`) are retained until
explicitly closed and may consume too much memory. (To control this warning, see the
rcParam `figure.max_open_warning`). Consider using `matplotlib.pyplot.close()`.
fig, ax = plt.subplots(1, **subkwargs(kwargs, plt.subplots, plt.figure))
```

pyrolite contains templates for the Pearce diagrams, used to discriminate mafic rocks (and particularly basalts) based on their whole-rock geochemistry. Two templates are included: *pearceThNbYb()* and *pearceTiNbYb()*. We can

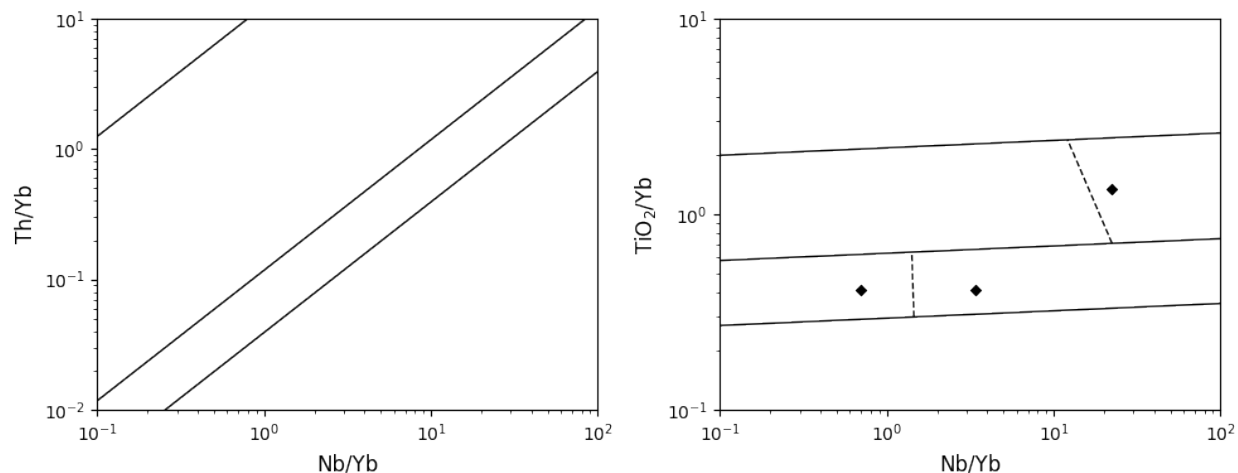
create some axes and add these templates to them:

```
from pyrolite.plot.templates import pearceThNbYb, pearceTiNbYb

fig, ax = plt.subplots(1, 2, figsize=(10, 4))
share_axes(ax, which="x") # these diagrams have the same x axis

pearceThNbYb(ax=ax[0])
pearceTiNbYb(ax=ax[1])

plt.tight_layout() # nicer spacing for axis labels
```



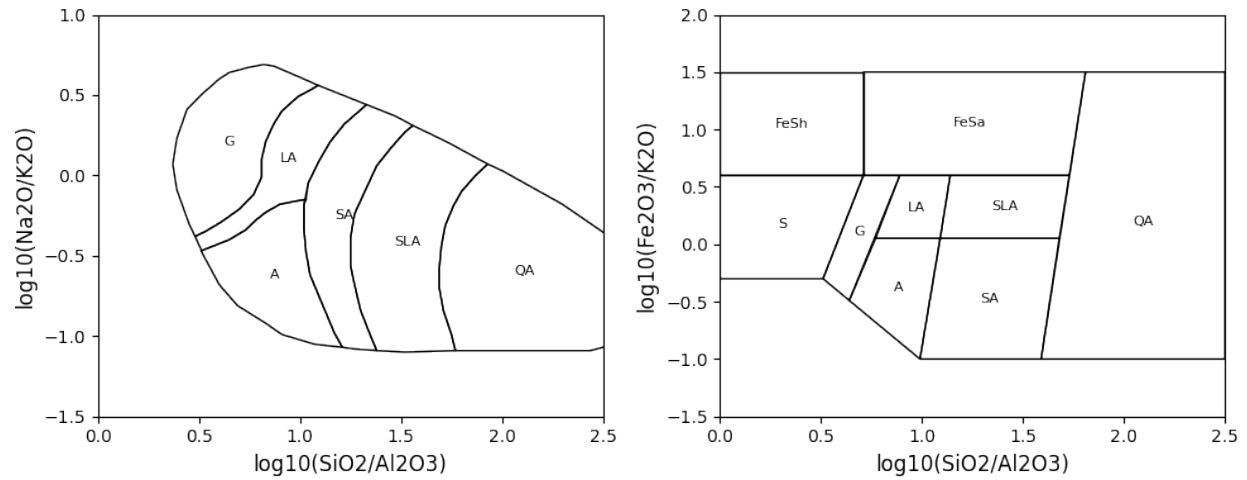
pyrolite also now includes some diagram templates for discrimination of sandstones based on their whole-rock geochemistry (*Pettijohn()*, *Herron()*):

```
from pyrolite.plot.templates import Herron, Pettijohn

fig, ax = plt.subplots(1, 2, figsize=(10, 4))
share_axes(ax, which="x") # these diagrams have the same x axis

Pettijohn(ax=ax[0], add_labels=True)
Herron(ax=ax[1], add_labels=True)

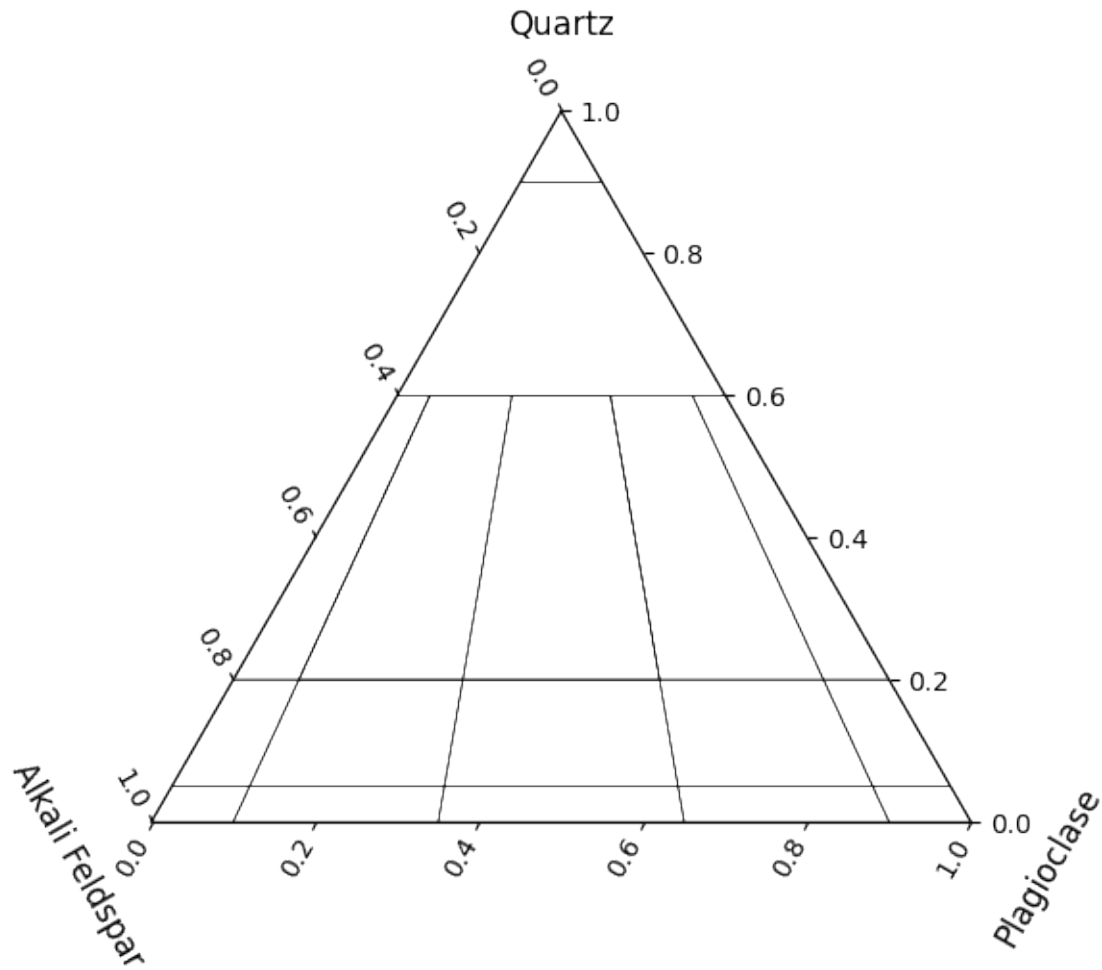
plt.tight_layout()
```



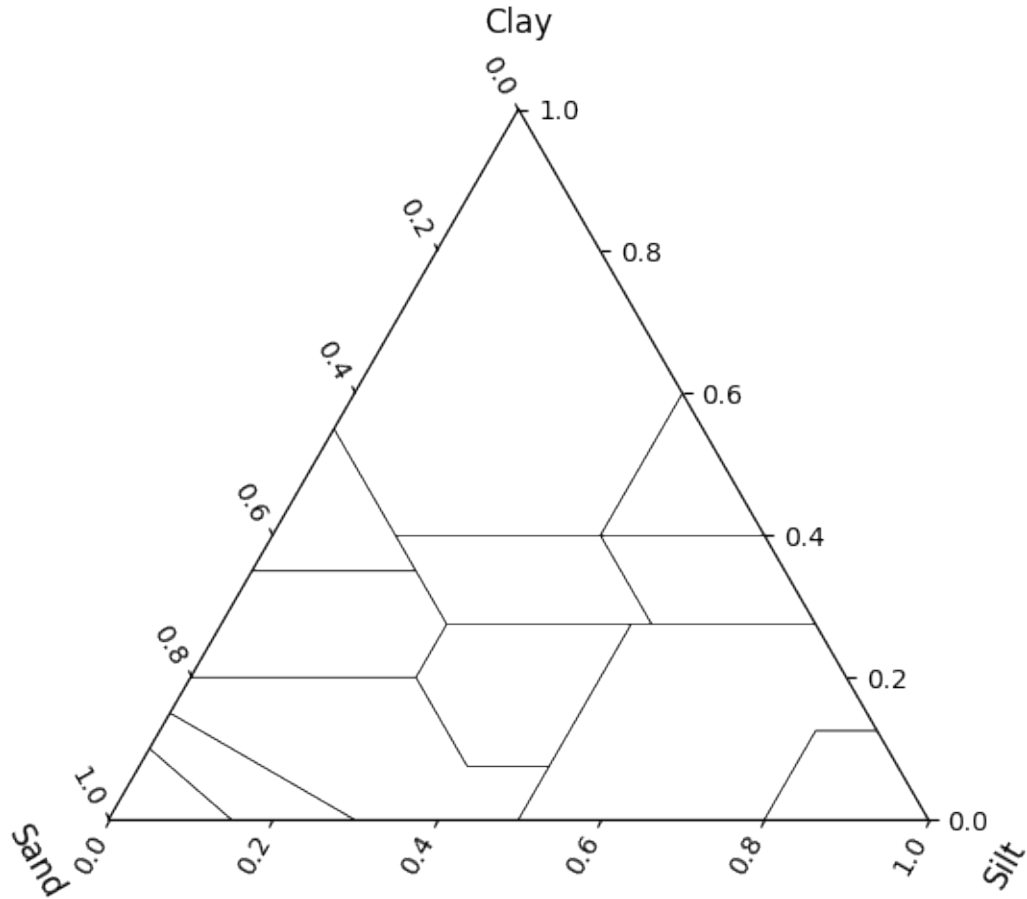
## Ternary Templates

pyrolite now also includes ternary classification diagrams including the *QAP()* and *USDAsoilTexture()* diagrams:

```
from pyrolite.plot.templates import (  
    QAP,  
    FeldsparTernary,  
    JensenPlot,  
    SpinelTrivalentTernary,  
    USDAsoilTexture,  
)  
  
ax = QAP(linewidth=0.4)  
plt.show()
```

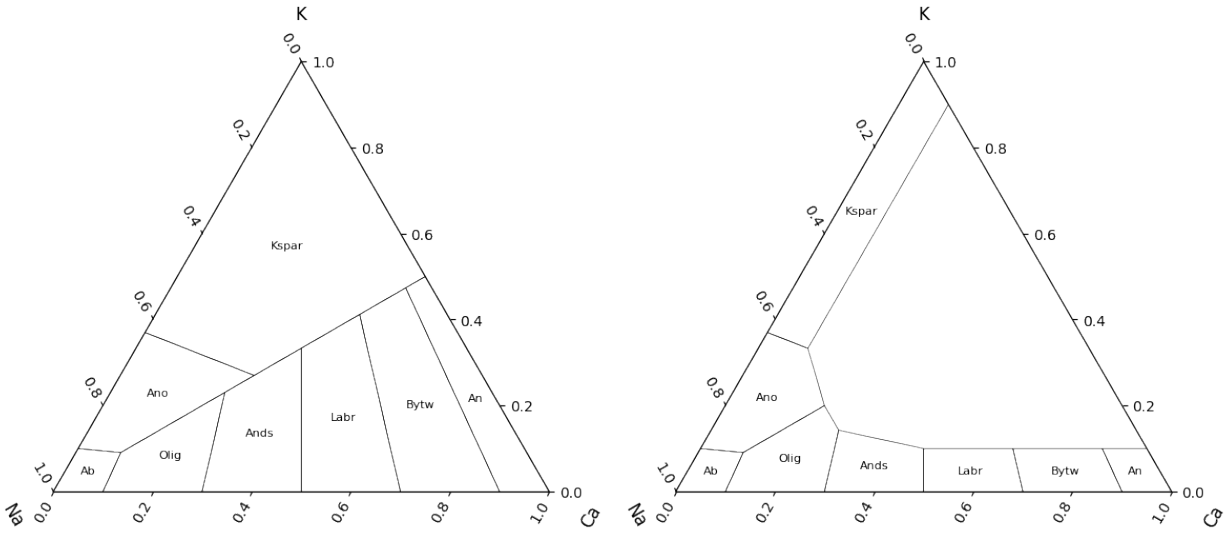


```
ax = USDAsoilTexture(linewidth=0.4)
plt.show()
```



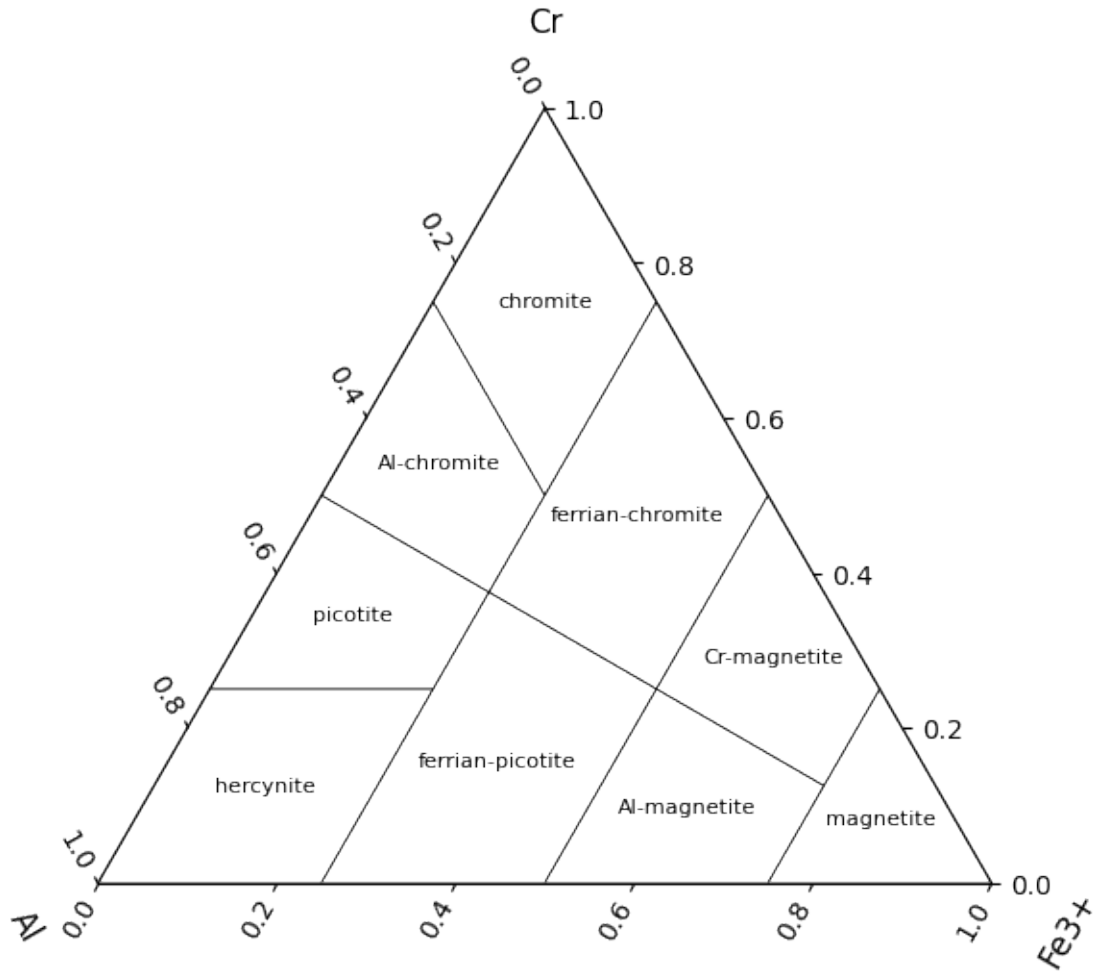
For the feldspar ternary diagram, which is complicated by a miscibility gap, there are two modes: *'default'* and *'miscibility-gap'*. The second of these provides a simplified approximation of the miscibility gap between k-feldspar and plagioclase, whereas *'default'* ignores this aspect (which itself is complicated by temperature):

```
fig, ax = plt.subplots(1, 2, figsize=(12, 6))
FeldsparTernary(ax=ax[0], linewidth=0.4, add_labels=True, mode="default")
FeldsparTernary(ax=ax[1], linewidth=0.4, add_labels=True, mode="miscibility-gap")
plt.tight_layout()
plt.show()
```



For general spinel phase discrimination, a ternary classification diagram can be used to give labels based on trivalent cationic content ( $\text{Cr}^{3+}$ ,  $\text{Al}^{3+}$ ,  $\text{Fe}^{3+}$ ):

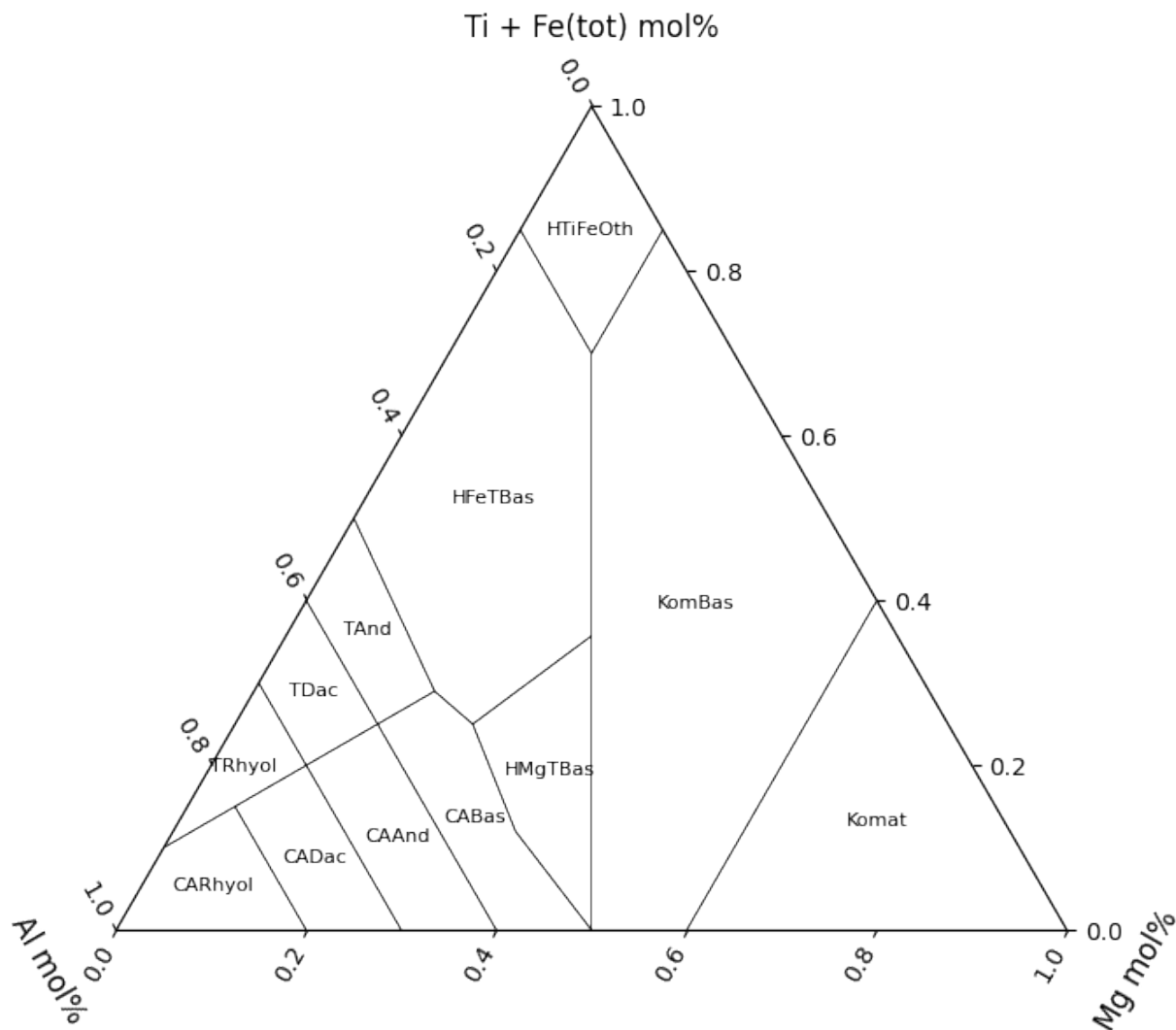
```
SpinelTrivalentTernary(linewidth=0.4, add_labels=True, figsize=(6, 6))
plt.show()
```



The Jensen plot is another cationic ternary discrimination diagram (Jensen, 1976), for subalkaline volcanic rocks:

```
JensenPlot(linewidth=0.4, add_labels=True, figsize=(7, 7))  
plt.show()
```





References and other notes for diagram templates can be found within the docstrings and within the pyrolite documentation:

```
help(TAS)
```

Help on function TAS in module pyrolite.plot.templates.TAS:

```
TAS(ax=None, add_labels=False, which_labels='ID', relim=True, color='k', which_
  ↳model=None, **kwargs)
    Adds the TAS diagram to an axes. Diagram from Middlemost (1994) [#pyrolite.plot.
  ↳templates.TAS.TAS_1]_,
    a closed-polygon variant after Le Bas et al (1992) [#pyrolite.plot.templates.TAS.TAS_
  ↳2]_.
```

Parameters

-----

```
ax : :class:`matplotlib.axes.Axes`
    Axes to add the template on to.
```

(continues on next page)

(continued from previous page)

```

add_labels : :class:`bool`
    Whether to add labels at polygon centroids.
which_labels : :class:`str`
    Which labels to add to the polygons (e.g. for TAS, 'volcanic', 'intrusive'
    or the field 'ID').
relim : :class:`bool`
    Whether to relimit axes to fit the built in ranges for this diagram.
color : :class:`str`
    Line color for the diagram.
which_model : :class:`str`
    The name of the model variant to use, if not Middlemost.

Returns
-----
ax : :class:`matplotlib.axes.Axes`

References
-----
.. [#pyrolite.plot.templates.TAS.TAS_1] Middlemost, E. A. K. (1994).
    Naming materials in the magma/igneous rock system.
    Earth-Science Reviews, 37(3), 215-224.
    doi: `10.1016/0012-8252(94)90029-9 <https://dx.doi.org/10.1016/0012-
↪8252(94)90029-9>`__

.. [#pyrolite.plot.templates.TAS.TAS_2] Le Bas, M.J., Le Maitre, R.W., Woolley, A.R. ↪
↪(1992).
    The construction of the Total Alkali-Silica chemical
    classification of volcanic rocks.
    Mineralogy and Petrology 46, 1-22.
    doi: `10.1007/BF01160698 <https://dx.doi.org/10.1007/BF01160698>`__

```

**See also:****Examples:**

TAS Classifier, Ternary Colour Mapping

**Modules:***pyrolite.util.classification***Classes:***TAS, QAP, USDAsoilTexture***Total running time of the script:** (0 minutes 22.646 seconds)

## Density and Contour Plots

While individual point data are useful, we commonly want to understand the the distribution of our data within a particular subspace, and compare that to a reference or other dataset. Pyrolite includes a few functions for visualising data density, most based on Gaussian kernel density estimation and evaluation over a grid. The below examples highlight some of the currently implemented features.

```

import matplotlib.pyplot as plt
import numpy as np

```

(continues on next page)

(continued from previous page)

```
import pandas as pd

from pyrolite.comp.codata import close
from pyrolite.plot import pyroplot
from pyrolite.plot.density import density

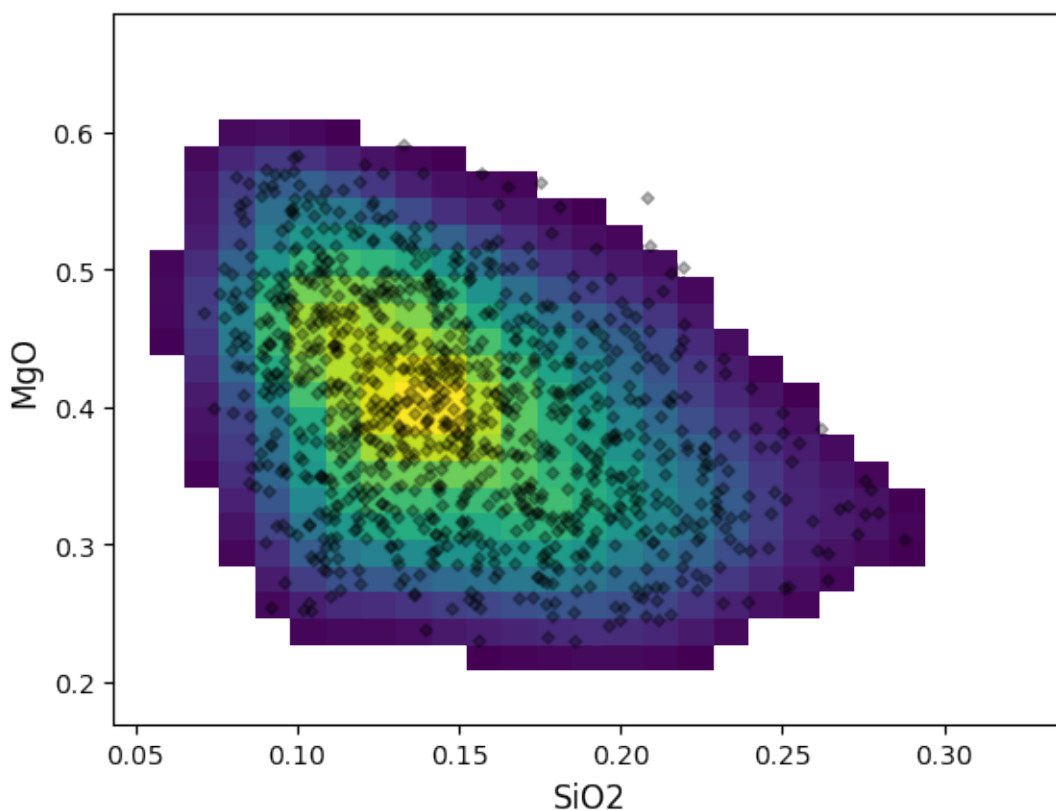
np.random.seed(82)
```

First we create some example data :

```
oxs = ["SiO2", "CaO", "MgO", "Na2O"]
ys = np.random.rand(1000, len(oxs))
ys[:, 1] += 0.7
ys[:, 2] += 1.0
df = pd.DataFrame(data=close(np.exp(ys)), columns=oxs)
```

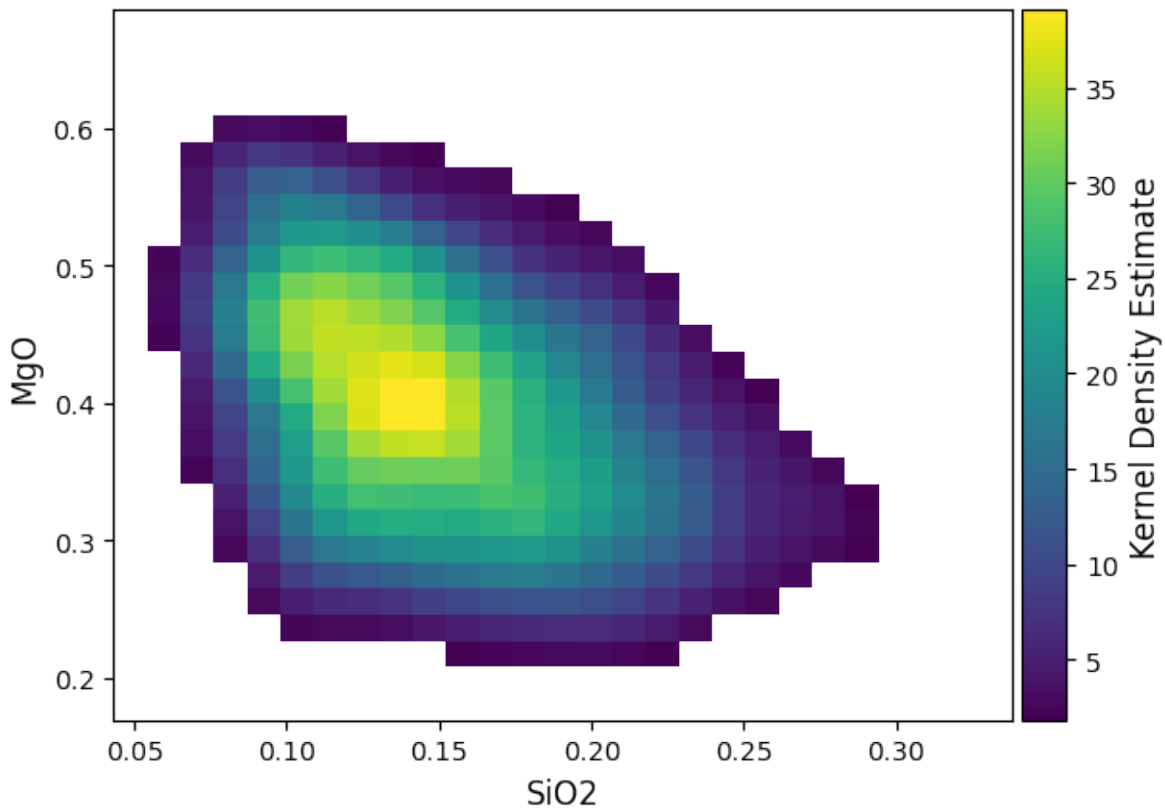
A minimal density plot can be constructed as follows:

```
ax = df.loc[:, ["SiO2", "MgO"]].pyroplot.density()
df.loc[:, ["SiO2", "MgO"]].pyroplot.scatter(ax=ax, s=10, alpha=0.3, c="k", zorder=2)
plt.show()
```



A colorbar linked to the KDE estimate colormap can be added using the *colorbar* boolean switch:

```
ax = df.loc[:, ["SiO2", "MgO"]].pyroplot.density(colorbar=True)
plt.show()
```

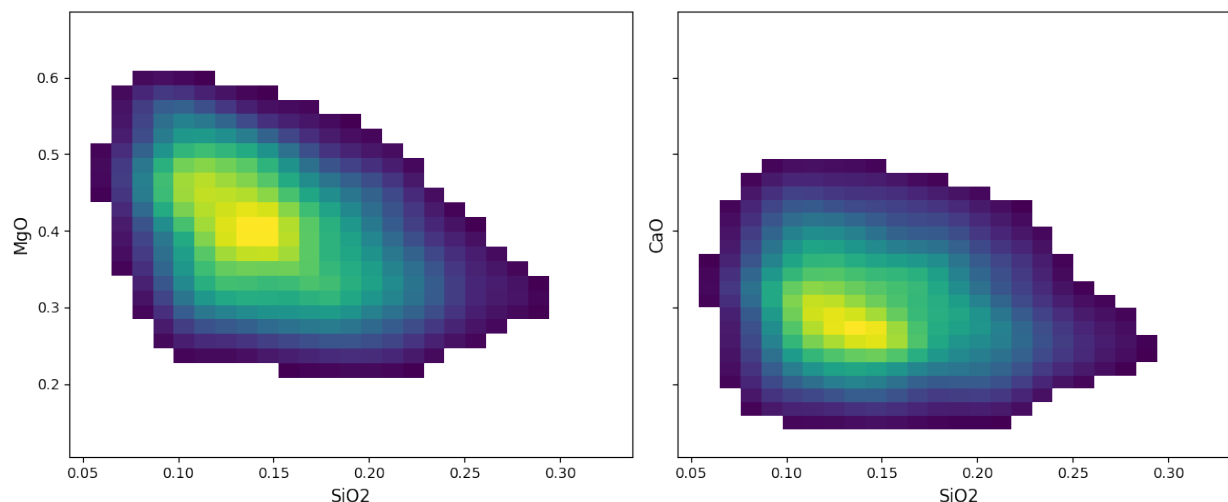


*density* by default will create a new axis, but can also be plotted over an existing axis for more control:

```
fig, ax = plt.subplots(1, 2, sharex=True, sharey=True, figsize=(12, 5))

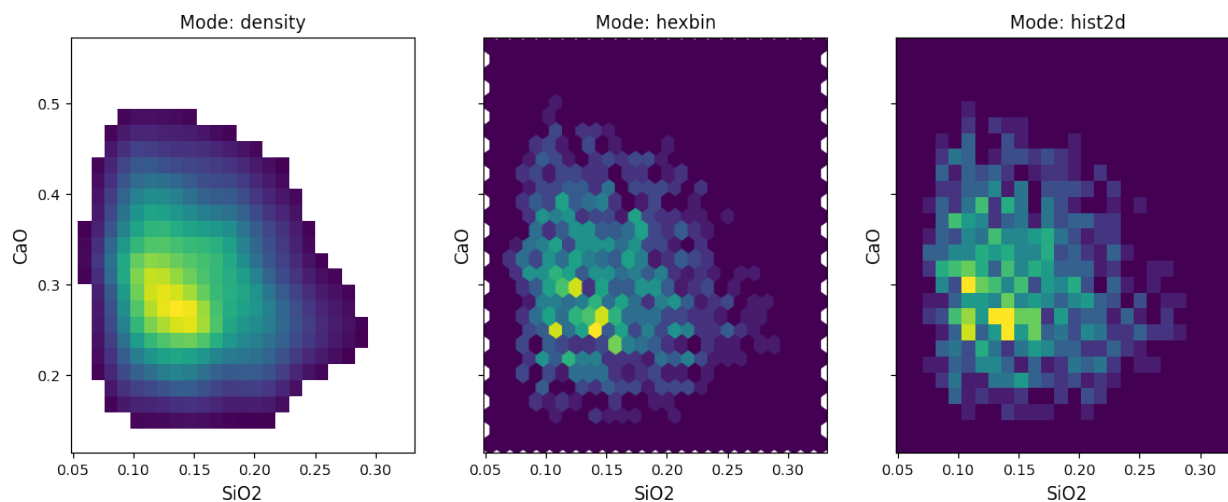
df.loc[:, ["SiO2", "MgO"]].pyroplot.density(ax=ax[0])
df.loc[:, ["SiO2", "CaO"]].pyroplot.density(ax=ax[1])

plt.tight_layout()
plt.show()
```



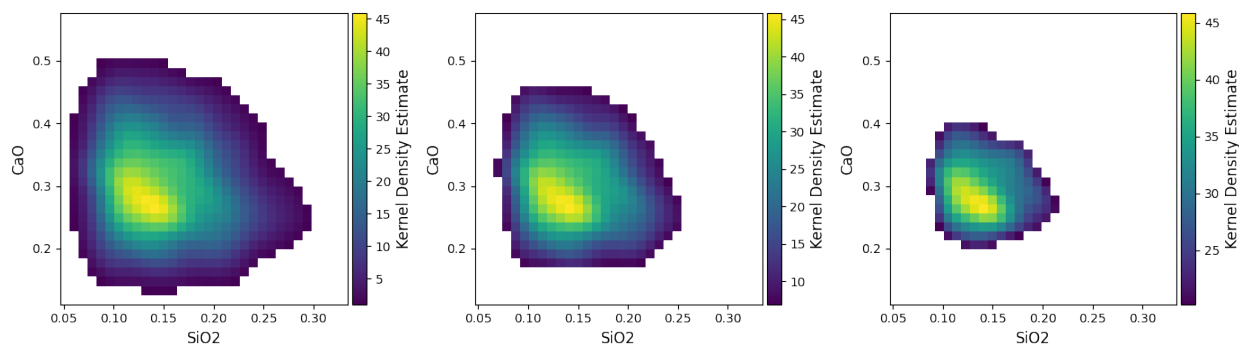
There are two other implemented modes beyond the default *density*: *hist2d* and *hexbin*, which parallel their equivalents in matplotlib. Contouring is not enabled for these histogram methods.

```
fig, ax = plt.subplots(1, 3, sharex=True, sharey=True, figsize=(14, 5))
for a, mode in zip(ax, ["density", "hexbin", "hist2d"]):
    df.loc[:, ["SiO2", "CaO"]].pyroplot.density(ax=a, mode=mode)
    a.set_title("Mode: {}".format(mode))
plt.show()
```



For the density mode, a *vmin* parameter is used to choose the lower threshold, and by default is the 99th percentile (*vmin=0.01*), but can be adjusted. This is useful where there are a number of outliers, or where you wish to reduce the overall complexity/colour intensity of a figure (also good for printing!).

```
fig, ax = plt.subplots(1, 3, figsize=(14, 4))
for a, vmin in zip(ax, [0.01, 0.1, 0.4]):
    df.loc[:, ["SiO2", "CaO"]].pyroplot.density(ax=a, bins=30, vmin=vmin, colorbar=True)
plt.tight_layout()
plt.show()
```

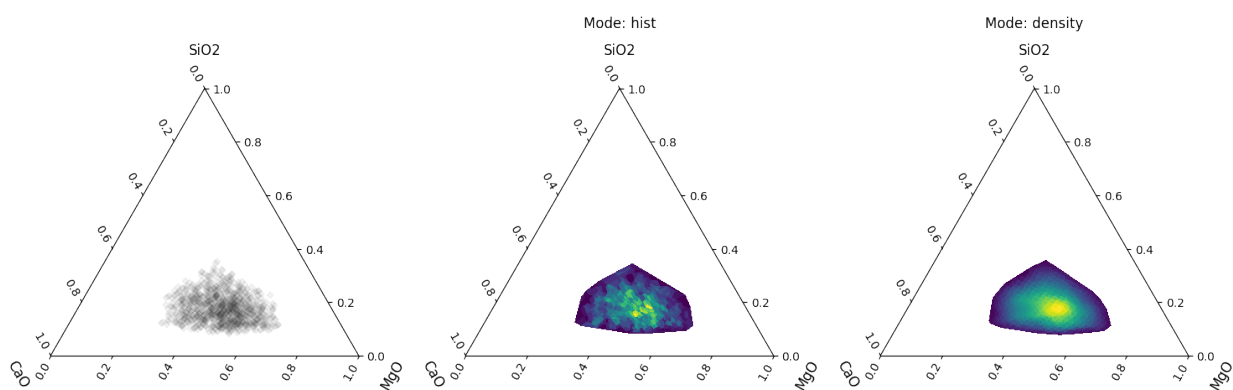


```
plt.close("all") # let's save some memory..
```

Density plots can also be used for ternary diagrams, where more than two components are specified:

```
fig, ax = plt.subplots(
    1,
    3,
    sharex=True,
    sharey=True,
    figsize=(15, 5),
    subplot_kw=dict(projection="ternary"),
)
df.loc[:, ["SiO2", "CaO", "MgO"]].pyroplot.scatter(ax=ax[0], alpha=0.05, c="k")
for a, mode in zip(ax[1:], ["hist", "density"]):
    df.loc[:, ["SiO2", "CaO", "MgO"]].pyroplot.density(ax=a, mode=mode)
    a.set_title("Mode: {}".format(mode), y=1.2)

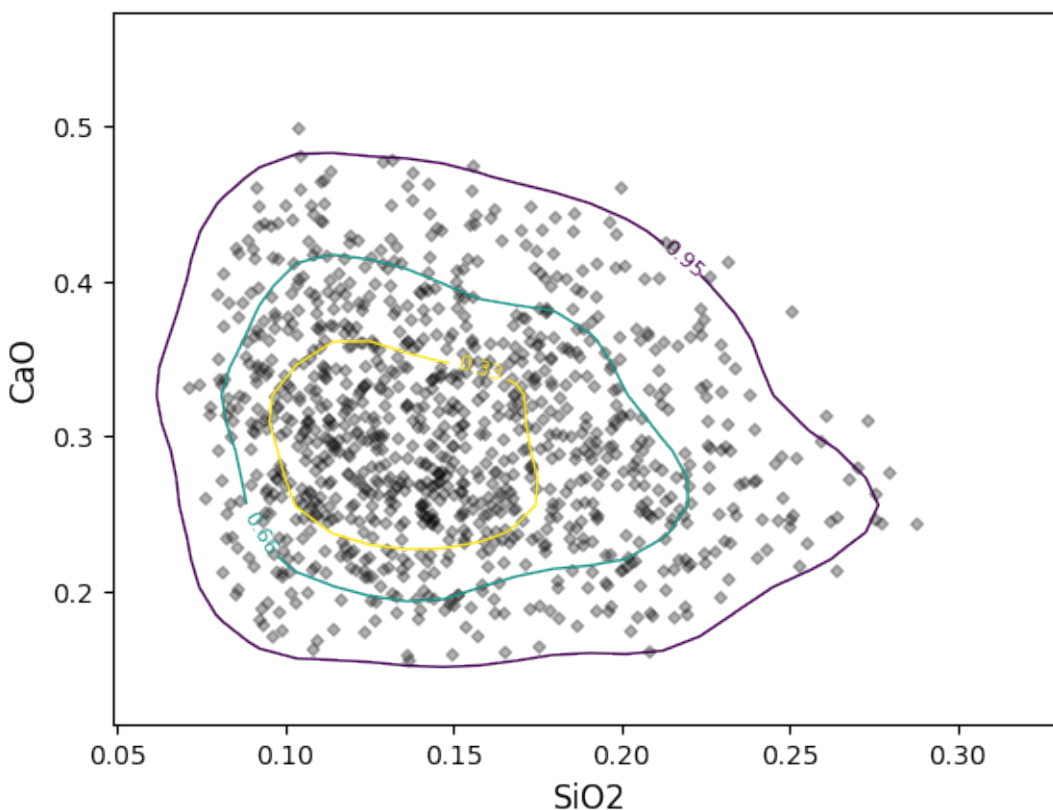
plt.tight_layout()
plt.show()
```



Contours are also easily created, which by default are percentile estimates of the underlying kernel density estimate.

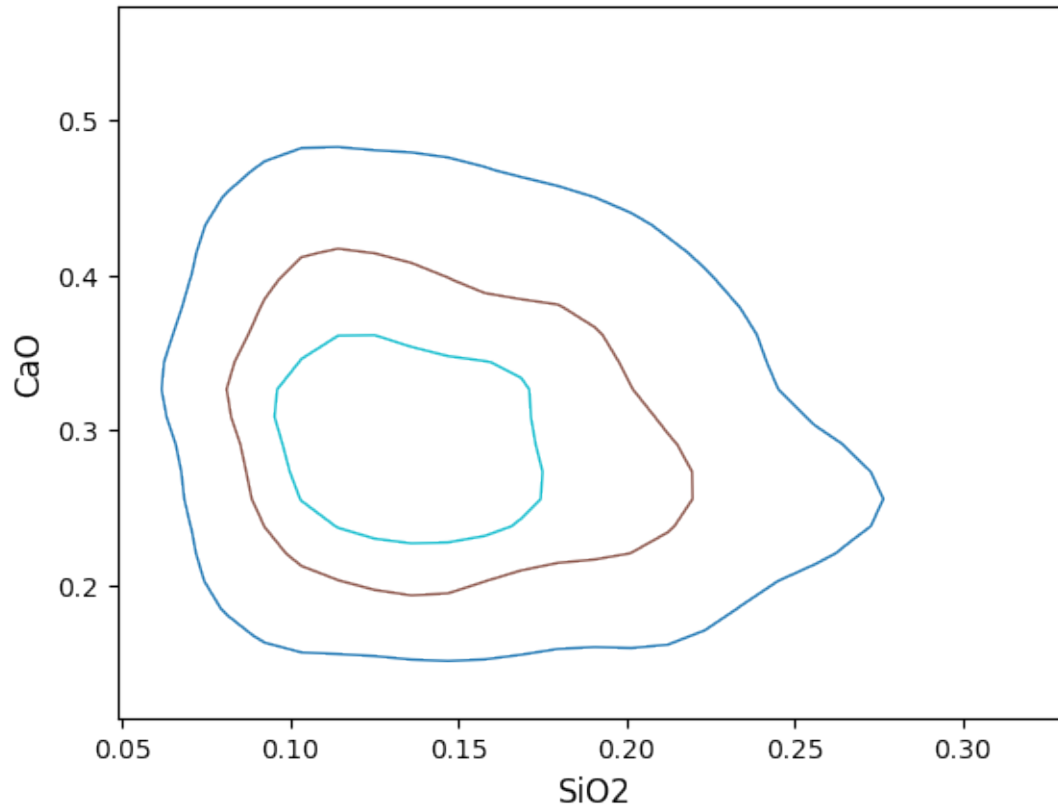
**Note:** As the contours are generated from kernel density estimates, assumptions around e.g. 95% of points lying within a 95% contour won't necessarily be valid for non-normally distributed data (instead, this represents the approximate 95% percentile on the kernel density estimate). Note that contours are currently only generated; for `mode="density"`; future updates may allow the use of a histogram basis, which would give results closer to 95% data percentiles.

```
ax = df.loc[:, ["SiO2", "CaO"]].pyroplot.scatter(s=10, alpha=0.3, c="k", zorder=2)
df.loc[:, ["SiO2", "CaO"]].pyroplot.density(ax=ax, contours=[0.95, 0.66, 0.33])
plt.show()
```



You can readily change the styling of these contours by passing *cmap* (for different colormaps), *colors* (for bulk or individually specified colors), *linewidths* and *linestyles* (for bulk or individually specified linestyles). You can also use 'label\_contours=False' to omit the contour labels. For example, to change the colormap and omit the contour labels:

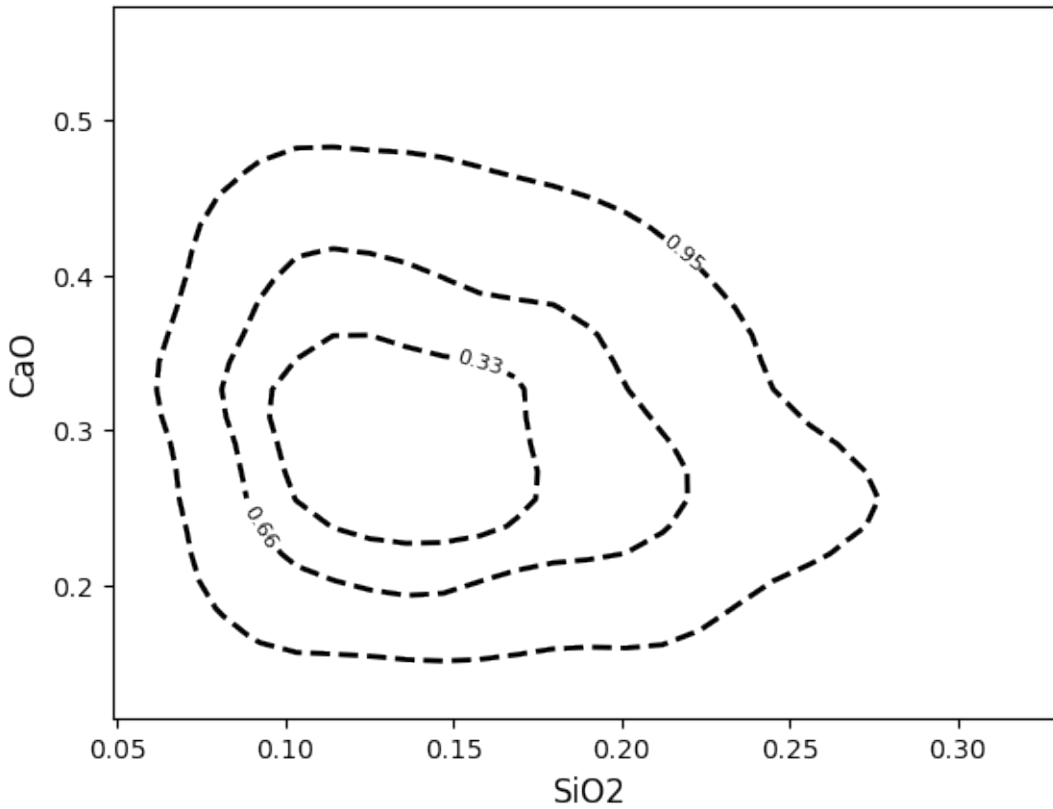
```
ax = df.loc[:, ["SiO2", "CaO"]].pyroplot.density(
    contours=[0.95, 0.66, 0.33], cmap="tab10", label_contours=False
)
plt.show()
```



You can specify one styling value for all contours:

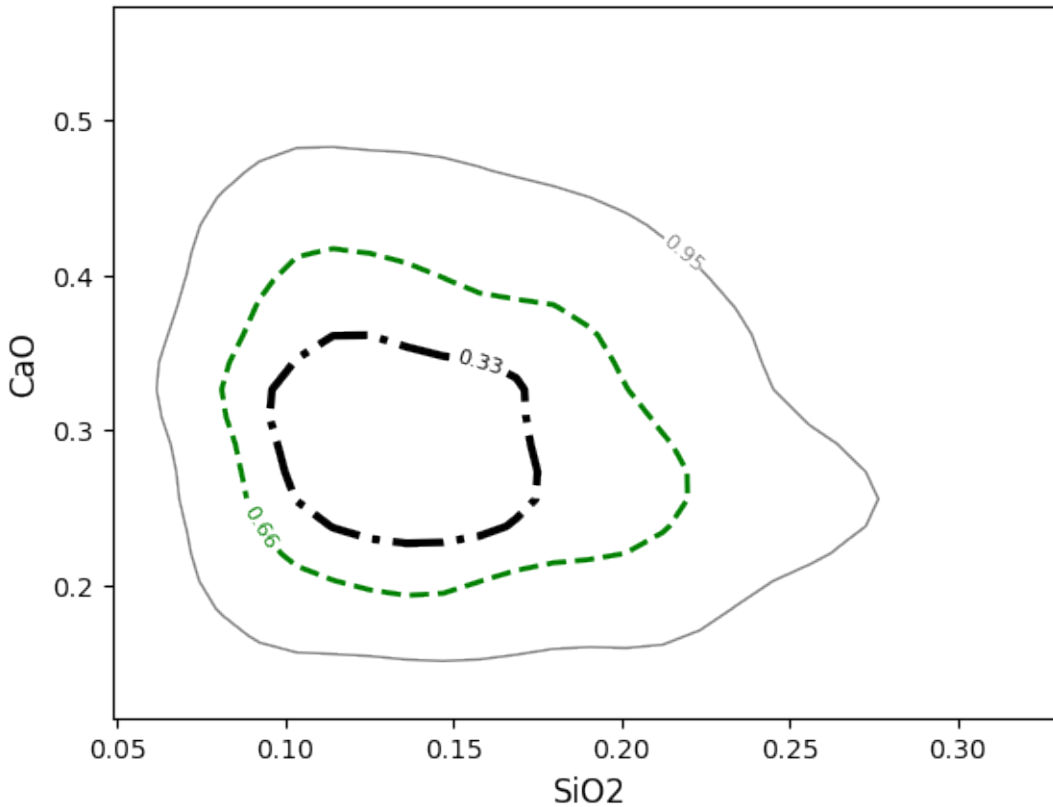
```
ax = df.loc[:, ["SiO2", "CaO"]].pyroplot.density(  
    contours=[0.95, 0.66, 0.33], linewidths=2, linestyle="--", colors="k"  
)  
plt.show()
```





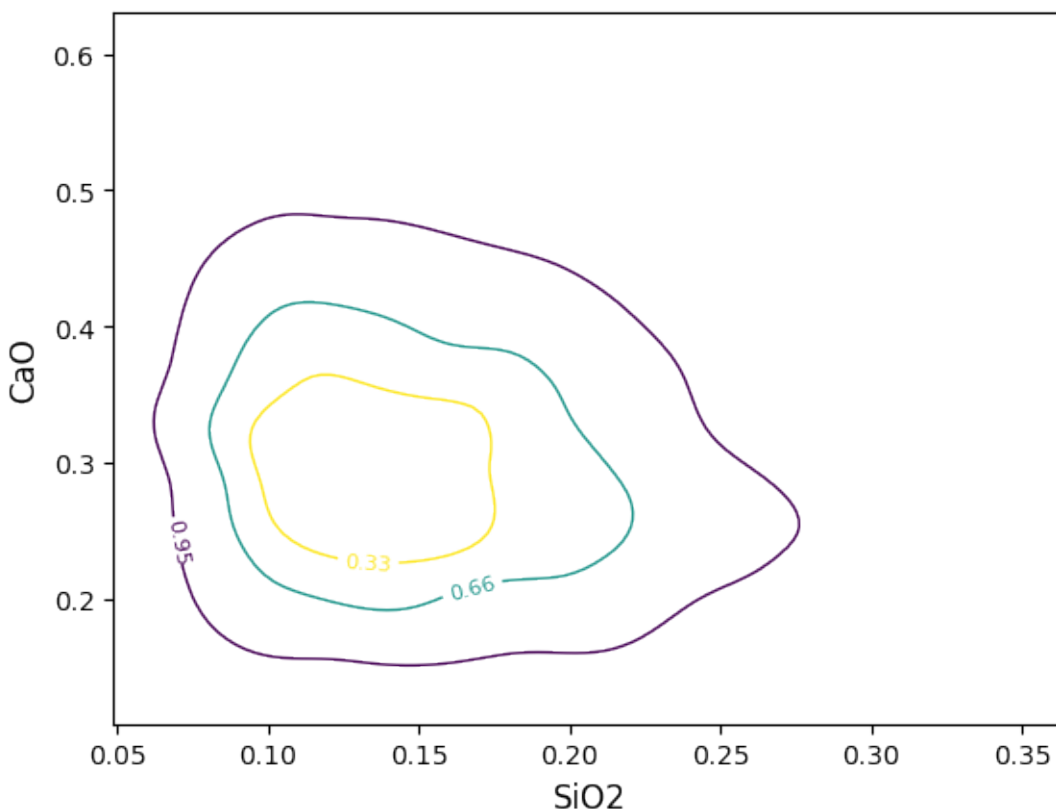
Or specify a list of values which will be applied to the contours in the order given in the *contours* keyword argument:

```
ax = df.loc[:, ["SiO2", "CaO"]].pyroplot.density(  
    contours=[0.95, 0.66, 0.33],  
    linewidths=[1, 2, 3],  
    linestyles=["-", "--", "-."],  
    colors=["0.5", "g", "k"],  
)  
plt.show()
```



If for some reason the density plot or contours are cut off on the edges, it may be that the span of the underlying grid doesn't cut it (this is often the case for just a small number of points). You can either manually adjust the extent ( *extent* = ( *min\_x*, *max\_x*, *min\_y*, *max\_y* ) ) or adjust the coverage scale ( e.g. *coverage\_scale* = 1.5 will increase the buffer from the default 10% to 25%). You can also adjust the number of bins in the underlying grid (either for both axes with e.g. *bins*=100 or individually with *bins*=(*<xbins>*, *<ybins>*)):

```
ax = df.loc[:, ["SiO2", "CaO"]].pyroplot.density(  
    contours=[0.95, 0.66, 0.33], coverage_scale=1.5, bins=100  
)  
plt.show()
```



```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
util/plot/density.py:202: UserWarning: The following kwargs were not used by contour:
'coverage_scale'
cs = contour(
```

Geochemical data is commonly log-normally distributed and is best analysed and visualised after log-transformation. The density estimation can be conducted over logspaced grids (individually for x and y axes using *logx* and *logy* boolean switches). Notably, this makes both the KDE image and contours behave more naturally:

```
# some assymetric data
from scipy import stats

xs = stats.norm.rvs(loc=6, scale=3, size=(200, 1))
ys = stats.norm.rvs(loc=20, scale=3, size=(200, 1)) + 5 * xs + 50
data = np.append(xs, ys, axis=1).T
asym_df = pd.DataFrame(np.exp(np.append(xs, ys, axis=1) / 25.0))
asym_df.columns = ["A", "B"]
grids = ["linxy", "logxy"] * 2 + ["logx", "logy"]
scales = ["linscale"] * 2 + ["logscale"] * 2 + ["semilogx", "semilogy"]
labels = ["{}-{}".format(ls, ps) for (ls, ps) in zip(grids, scales)]
params = list(
    zip(
        [
            (False, False),
            (True, True),
            (False, False),
```

(continues on next page)

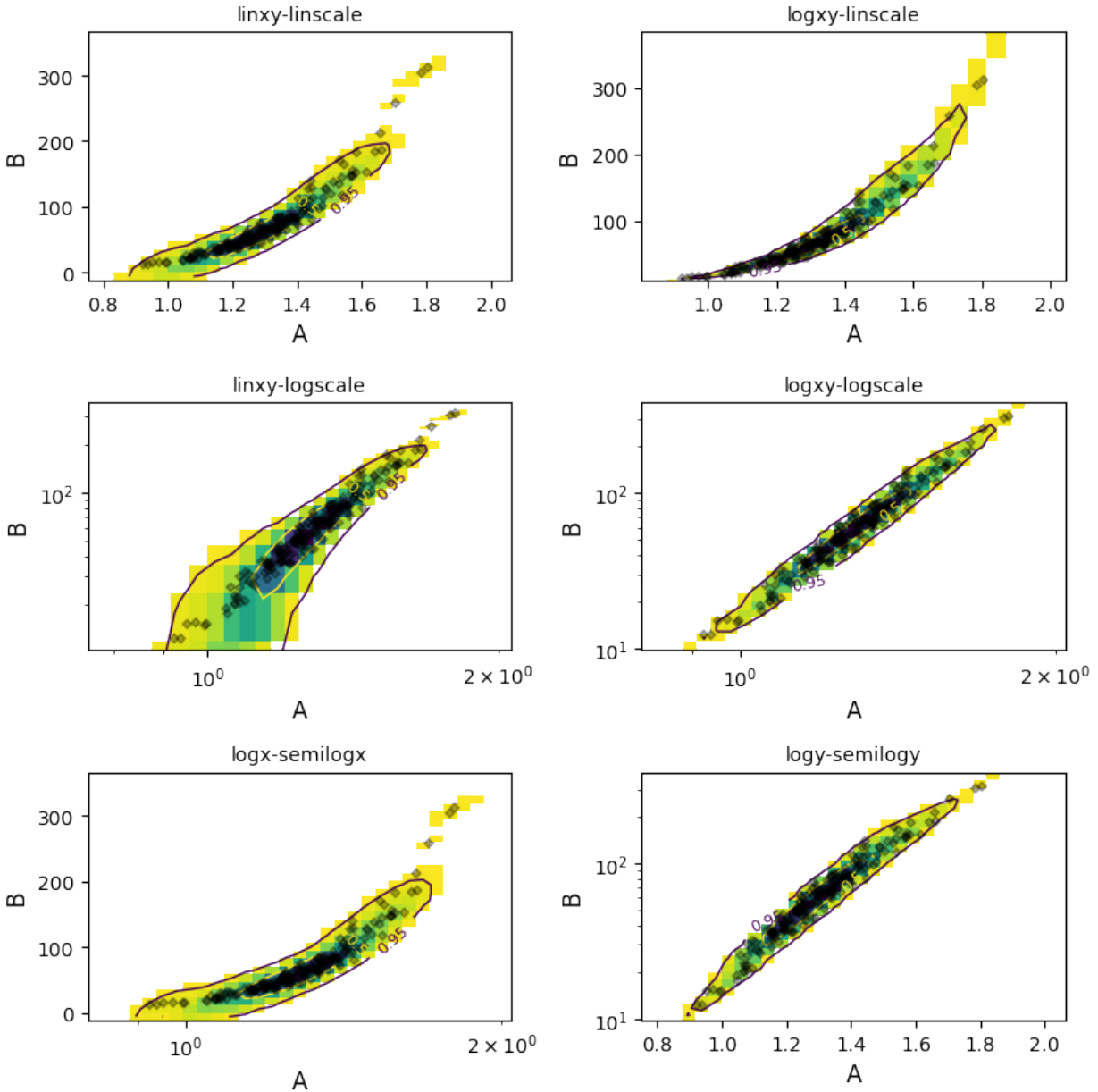
(continued from previous page)

```
        (True, True),
        (True, False),
        (False, True),
    ],
    grids,
    scales,
)
)
```

```
fig, ax = plt.subplots(3, 2, figsize=(8, 8))
ax = ax.flat

for a, (ls, grid, scale) in zip(ax, params):
    lx, ly = ls
    asym_df.pyroplot.density(ax=a, logx=lx, logy=ly, bins=30, cmap="viridis_r")
    asym_df.pyroplot.density(
        ax=a,
        logx=lx,
        logy=ly,
        contours=[0.95, 0.5],
        bins=30,
        cmap="viridis",
        fontsize=10,
    )
    asym_df.pyroplot.scatter(ax=a, s=10, alpha=0.3, c="k", zorder=2)

    a.set_title("{}-{}".format(grid, scale), fontsize=10)
    if scale in ["logscale", "semilogx"]:
        a.set_xscale("log")
    if scale in ["logscale", "semilogy"]:
        a.set_yscale("log")
plt.tight_layout()
plt.show()
```



```
plt.close("all") # let's save some memory..
```

**Note:** Using alpha with the density mode induces a known and old matplotlib bug, where the edges of bins within a pcolormesh image (used for plotting the KDE estimate) are over-emphasized, giving a gridded look.

See also:

[Heatscatter Plots](#), [Ternary Plots](#), [Spider Density Diagrams](#)

**Total running time of the script:** (0 minutes 11.233 seconds)

## Spiderplots & Density Spiderplots

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

Here we'll set up an example which uses EMORB as a starting point. Typically we'll normalise trace element compositions to a reference composition to be able to link the diagram to 'relative enrichment' occurring during geological processes, so here we're normalising to a Primitive Mantle composition first. We're here taking this normalised composition and adding some noise in log-space to generate multiple compositions about this mean (i.e. a compositional distribution). For simplicity, this is handled by `example_spider_data()`:

```
from pyrolite.util.synthetic import example_spider_data

normdf = example_spider_data(start="EMORB_SM89", norm_to="PM_PON")
```

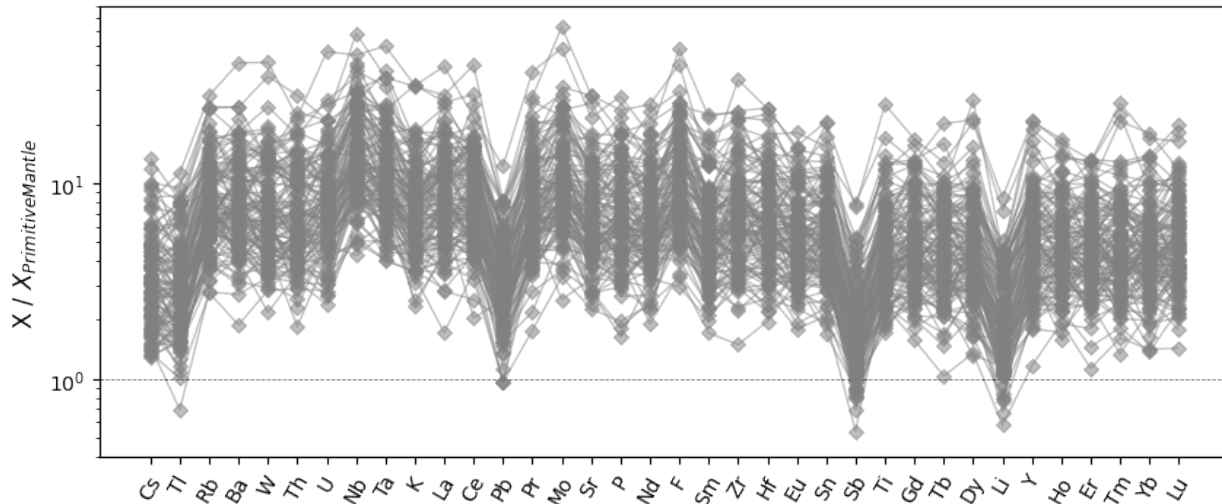
See also:

Normalisation

Basic spider plots are straightforward to produce:

```
import pyrolite.plot

ax = normdf.pyroplot.spider(color="0.5", alpha=0.5, unity_line=True, figsize=(10, 4))
ax.set_ylabel("X / XPrimitiveMantle")
plt.show()
```

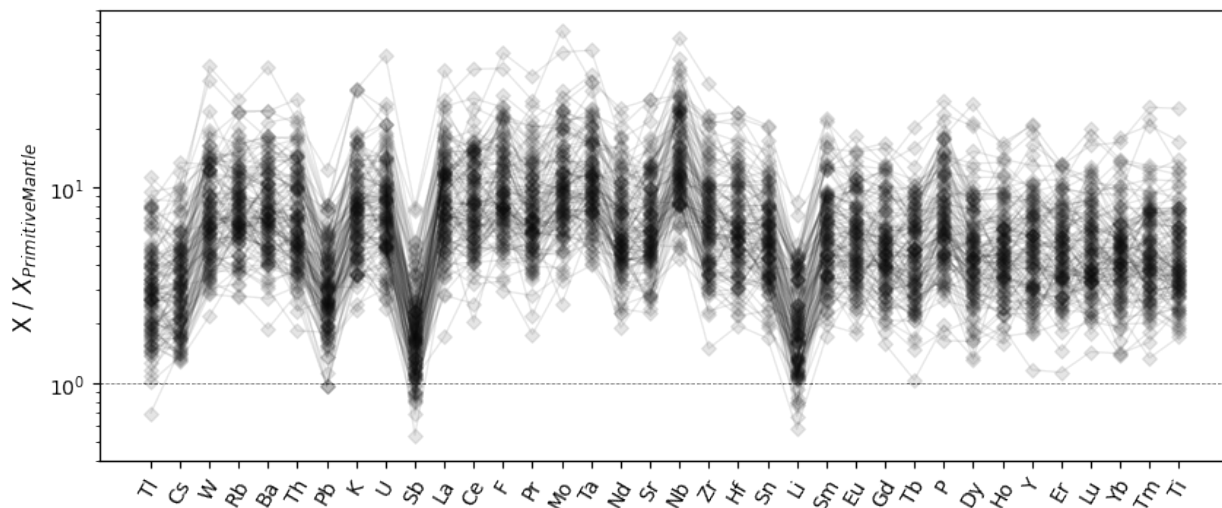


## Index Ordering

The default ordering here follows that of the dataframe columns, but we typically want to reorder these based on some physical ordering. A `index_order` keyword argument can be used to supply a function which will reorder the elements before plotting. Here we order the elements by relative incompatibility (using `pyrolite.geochem.ind.by_incompatibility()` behind the scenes):

```
from pyrolite.geochem.ind import by_incompatibility

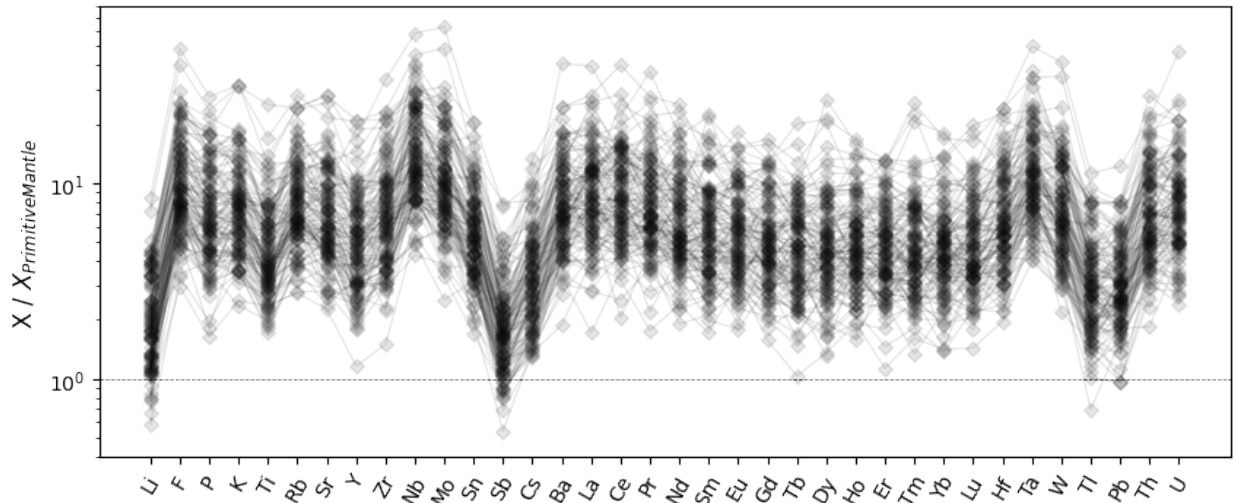
ax = normdf.pyroplot.spider(
    color="k",
    alpha=0.1,
    unity_line=True,
    index_order="incompatibility",
    figsize=(10, 4),
)
ax.set_ylabel("X / X_{Primitive Mantle}")
plt.show()
```



Similarly, you can also rearrange elements to be in order of atomic number:

```
from pyrolite.geochem.ind import by_number

ax = normdf.pyroplot.spider(
    color="k",
    alpha=0.1,
    unity_line=True,
    index_order="number",
    figsize=(10, 4),
)
ax.set_ylabel("X / X_{Primitive Mantle}")
plt.show()
```



## Color Mapping

We can also specify either continuous or categorical values to use for the colors, and even map categorical values to specific colors where useful:

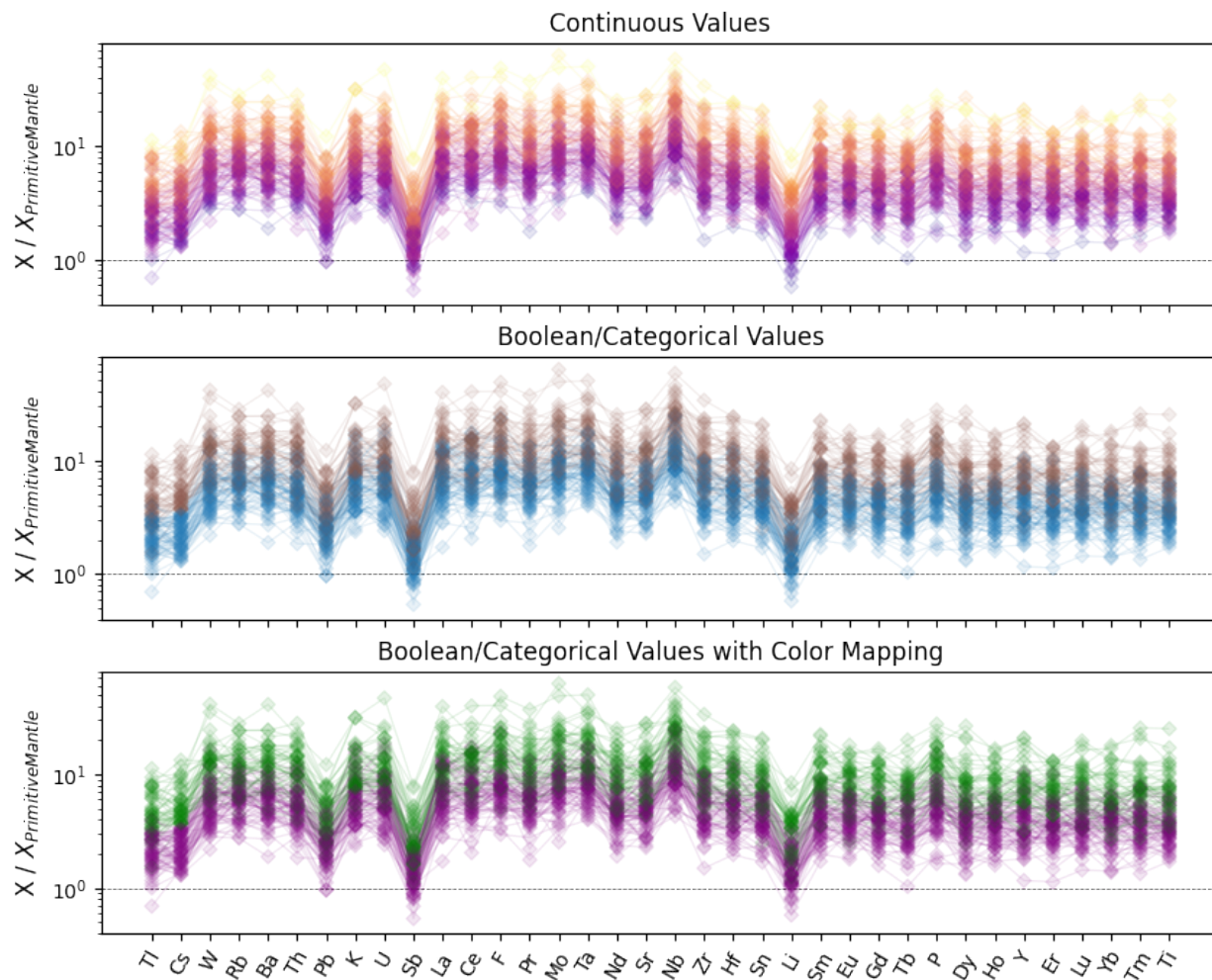
```
fig, ax = plt.subplots(3, 1, sharex=True, sharey=True, figsize=(10, 8))
ax[0].set_title("Continuous Values")
normdf.pyroplot.spider(
    ax=ax[0],
    unity_line=True,
    index_order="incompatibility",
    cmap="plasma",
    alpha=0.1,
    color=np.log(normdf["Li"]), # a range of continous values
)
ax[1].set_title("Boolean/Categorical Values")
normdf.pyroplot.spider(
    ax=ax[1],
    alpha=0.1,
    unity_line=True,
    index_order="incompatibility",
    color=normdf["Cs"] > 3.5, # a boolean/categorical set of values
)
ax[2].set_title("Boolean/Categorical Values with Color Mapping")
normdf.pyroplot.spider(
    ax=ax[2],
    alpha=0.1,
    unity_line=True,
    index_order="incompatibility",
    color=normdf["Cs"] > 3.5, # a boolean/categorical set of values
    color_mappings={ # mapping the boolean values to specific colors
        "color": {True: "green", False: "purple"}
    },
)
[a.set_ylabel("X / X_{Primitive Mantle}") for a in ax]
```

(continues on next page)



(continued from previous page)

plt.show()



### Legend Proxies for Spiderplots

While it's relatively straightforward to style spider plots as you wish, for the moment can be a bit more involved to create a legend for these styles. Where you're creating styles based on a set of categories or labels, a few of pyrolite's utility functions might come in handy. Below we'll go through such an example, after creating a few labels (here based on a binning of the Cs abundance):

```
labels = pd.cut(
    np.log(normcdf["Cs"]), bins=4, labels=["Low", "Mid. Low", "Mid High", "High"]
)
pd.unique(labels)
```

```
['Mid. Low', 'Mid High', 'Low', 'High']
Categories (4, object): ['Low' < 'Mid. Low' < 'Mid High' < 'High']
```

Below we'll use `process_color()` and `proxy_line()` to construct a set of legend proxies. Note that we need to pass the same configuration to both `spider()` and `process_color()` in order to get the same results, and that the order

of labels in the legend will depend on which labels appear first in your dataframe or series (and hence the ordering of the unique values which are returned).

```
from pyrolite.plot.color import process_color
from pyrolite.util.plot.legend import proxy_line

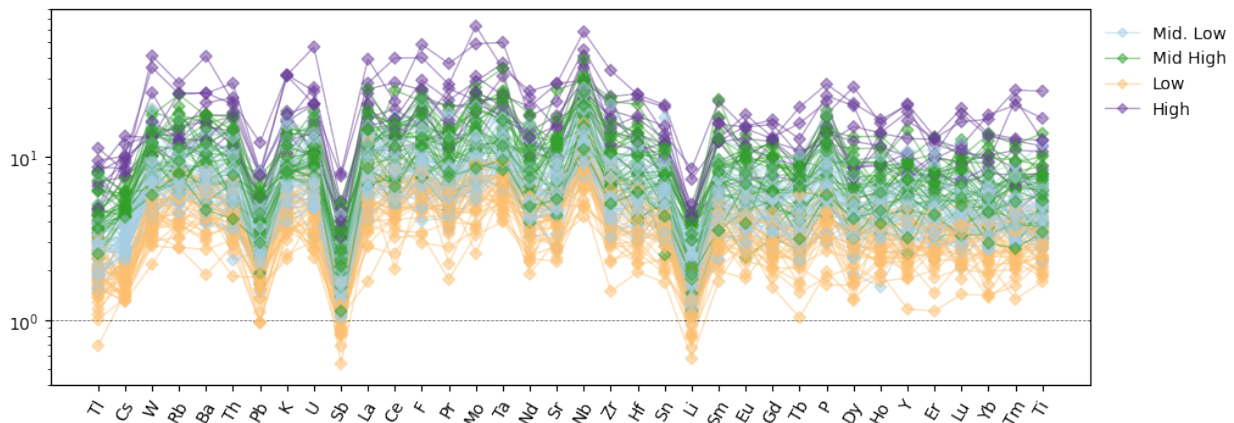
ax = normdf.pyroplot.spider(
    unity_line=True,
    index_order="incompatibility",
    color=labels, # a categorical set of values
    cmap="Paired",
    alpha=0.5,
    figsize=(11, 4),
)

legend_labels = pd.unique(labels) # process_color uses this behind the scenes

proxy_colors = process_color(color=legend_labels, cmap="Paired", alpha=0.5)["c"]

legend_proxies = [proxy_line(color=c, marker="D") for c in proxy_colors]

ax.legend(legend_proxies, legend_labels)
plt.show()
```



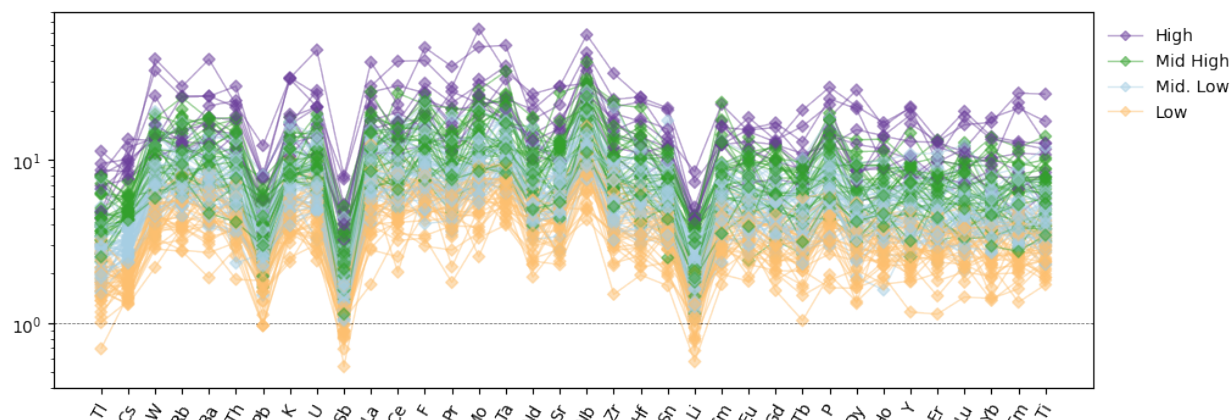
If the specific order of the labels in your legend is important or you only want to include some of the legend entries for some reason, you could use a dictionary to store the key-value pairs and remap the order of the legend entries manually:

```
proxies = {
    label: proxy_line(color=c, marker="D")
    for label, c in zip(legend_labels, proxy_colors)
}

ordered_labels = ["High", "Mid High", "Mid. Low", "Low"]

ax.legend([proxies[l] for l in ordered_labels], ordered_labels)

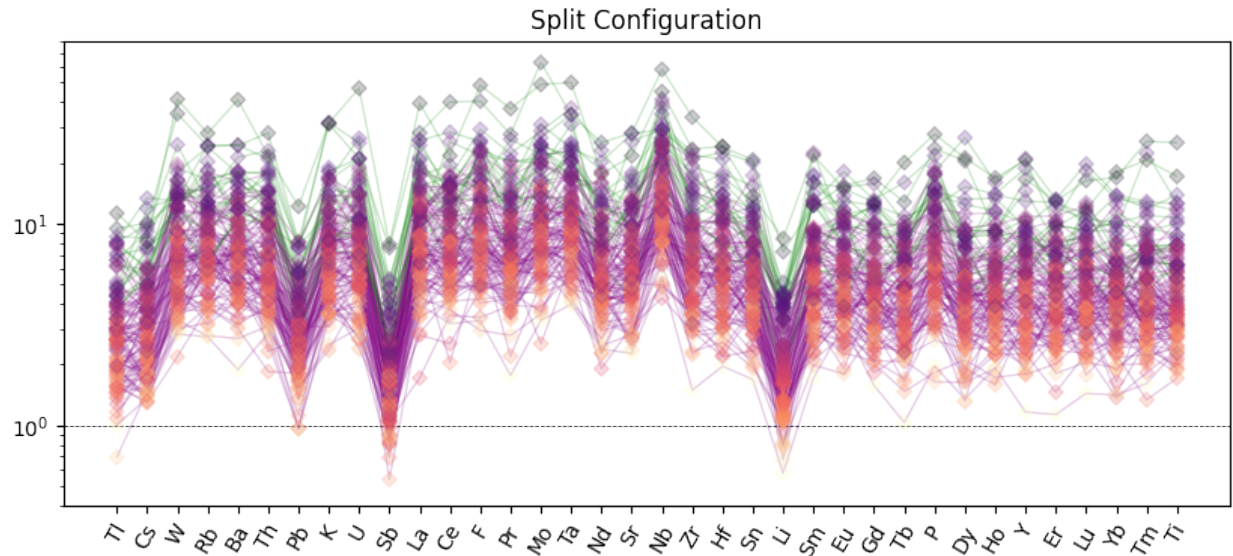
plt.show()
```



### Split Configuration

If you have potential conflicts between desired configurations for the lines and markers of your plots, you can explicitly separate the configuration using the `scatter_kw` and `line_kw` keyword arguments:

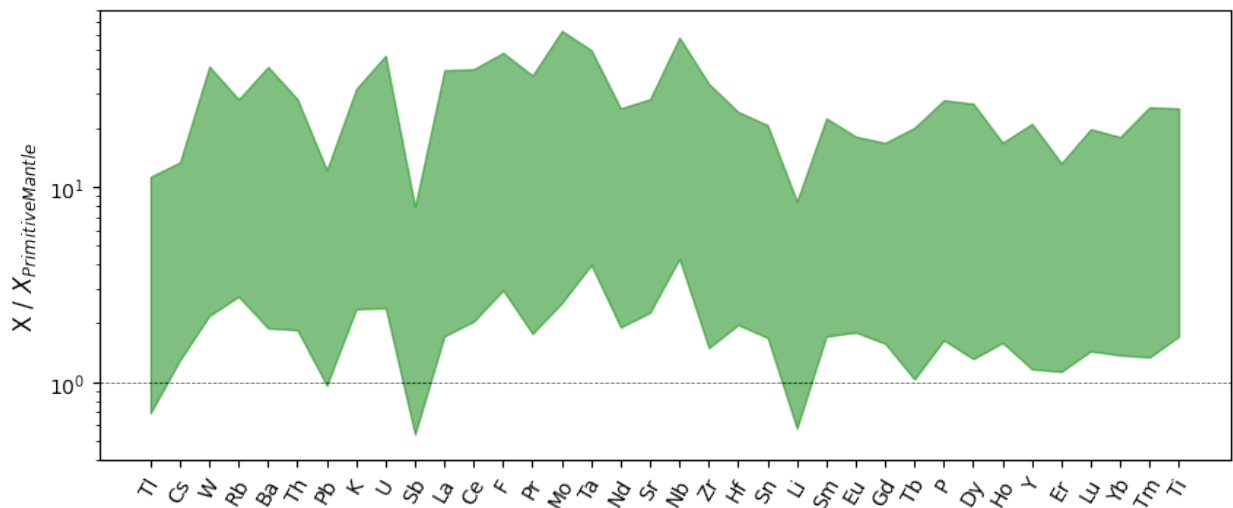
```
fig, ax = plt.subplots(1, 1, sharex=True, sharey=True, figsize=(10, 4))
ax.set_title("Split Configuration")
normdf.pyroplot.spider(
    ax=ax,
    unity_line=True,
    index_order="incompatibility",
    scatter_kw=dict(cmap="magma_r", color=np.log(normdf["Li"])),
    line_kw=dict(
        color=normdf["Cs"] > 5,
        color_mappings={"color": {True: "green", False: "purple"}}),
    alpha=0.2, # common alpha config between lines and markers
    s=25, # argument for scatter which won't be passed to lines
)
plt.show()
```



## Filled Ranges

The spiderplot can be extended to provide visualisations of ranges and density via the various modes. We could now plot the range of compositions as a filled range:

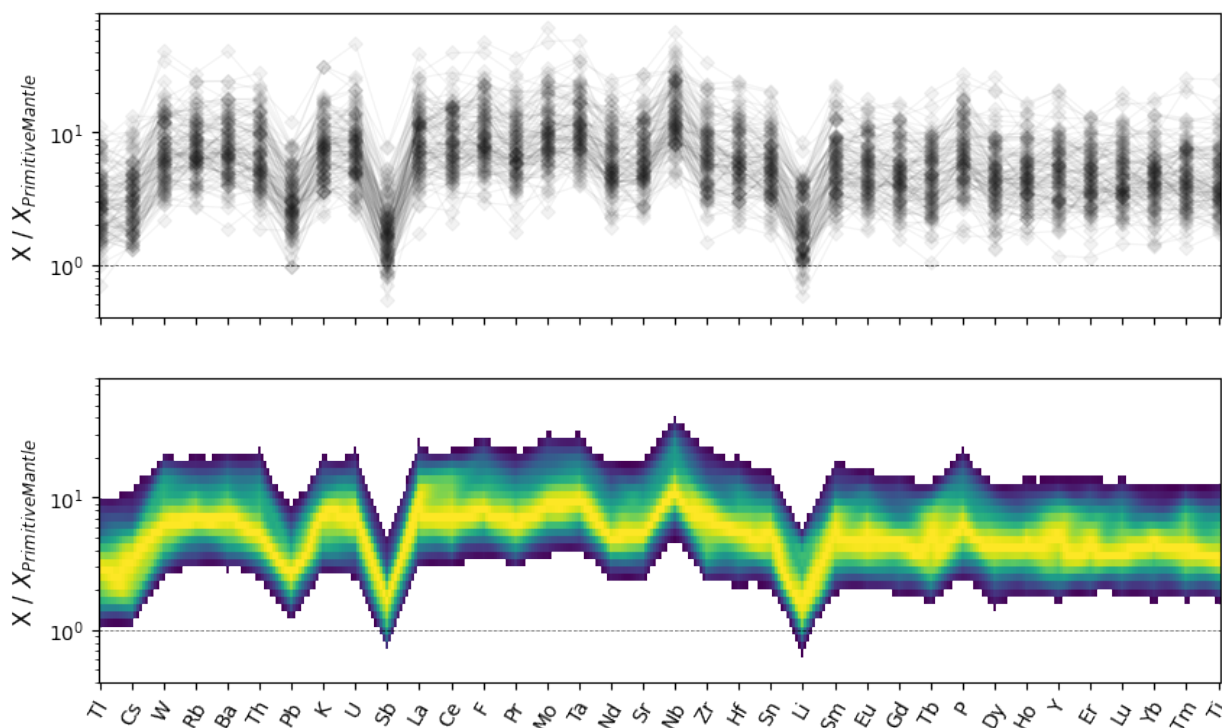
```
ax = normdf.pyroplot.spider(
    mode="fill",
    color="green",
    alpha=0.5,
    unity_line=True,
    index_order="incompatibility",
    figsize=(10, 4),
)
ax.set_ylabel("X /  $X_{\text{Primitive Mantle}}$ ")
plt.show()
```



## Spider Density Plots

Alternatively, we can plot a conditional density spider plot:

```
fig, ax = plt.subplots(2, 1, sharex=True, sharey=True, figsize=(10, 6))
normdf.pyroplot.spider(
    ax=ax[0], color="k", alpha=0.05, unity_line=True, index_order=by_incompatibility
)
normdf.pyroplot.spider(
    ax=ax[1],
    mode="binkde",
    vmin=0.05, # 95th percentile,
    resolution=10,
    unity_line=True,
    index_order="incompatibility",
)
[a.set_ylabel("X / XPrimitiveMantle") for a in ax]
plt.show()
```



We can now assemble a more complete comparison of some of the conditional density modes for spider plots:

```
modes = [
    ("plot", "plot", [], dict(color="k", alpha=0.01)),
    ("fill", "fill", [], dict(color="k", alpha=0.5)),
    ("binkde", "binkde", [], dict(resolution=5)),
    (
        "binkde",
        "binkde contours specified",
        [],
    ),
]
```

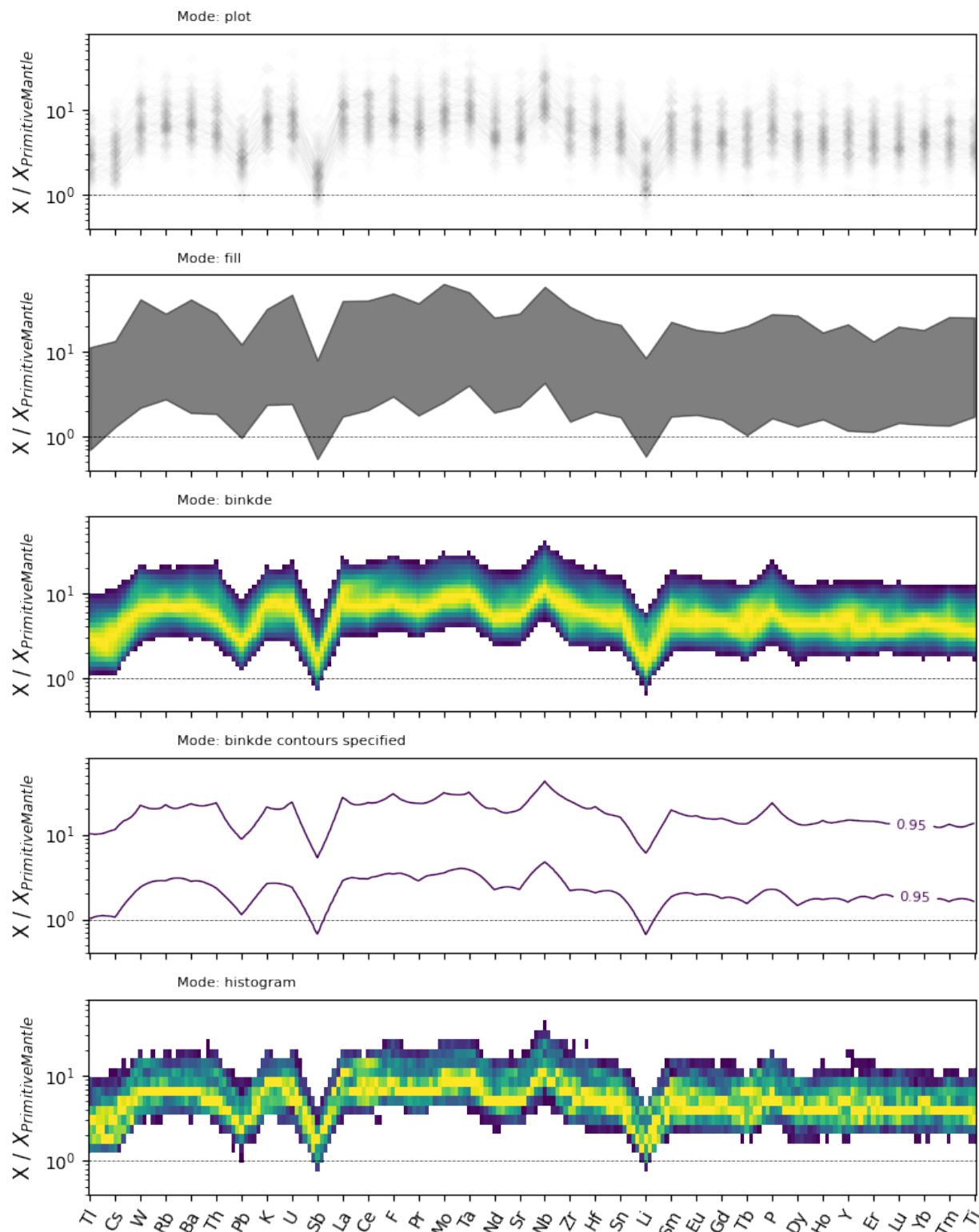
(continues on next page)

(continued from previous page)

```
    dict(contours=[0.95], resolution=5), # 95th percentile contour
),
("histogram", "histogram", [], dict(resolution=5, bins=30)),
]
```

```
down, across = len(modes), 1
fig, ax = plt.subplots(
    down, across, sharey=True, sharex=True, figsize=(across * 8, 2 * down)
)
[a.set_ylabel("X / $X_{\text{Primitive Mantle}}$") for a in ax]
for a, (m, name, args, kwargs) in zip(ax, modes):
    a.annotate( # label the axes rows
        "Mode: {}".format(name),
        xy=(0.1, 1.05),
        xycoords=a.transAxes,
        fontsize=8,
        ha="left",
        va="bottom",
    )
ax = ax.flat
for mix, (m, name, args, kwargs) in enumerate(modes):
    normdf.pyroplot.spider(
        mode=m,
        ax=ax[mix],
        vmin=0.05, # minimum percentile
        fontsize=8,
        unity_line=True,
        index_order="incompatibility",
        *args,
        **kwargs
    )
plt.tight_layout()
```





```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→ plot/spider.py:211: UserWarning: No data for colormapping provided via 'c'. Parameters
→ 'vmin' will be ignored
```

(continues on next page)

(continued from previous page)

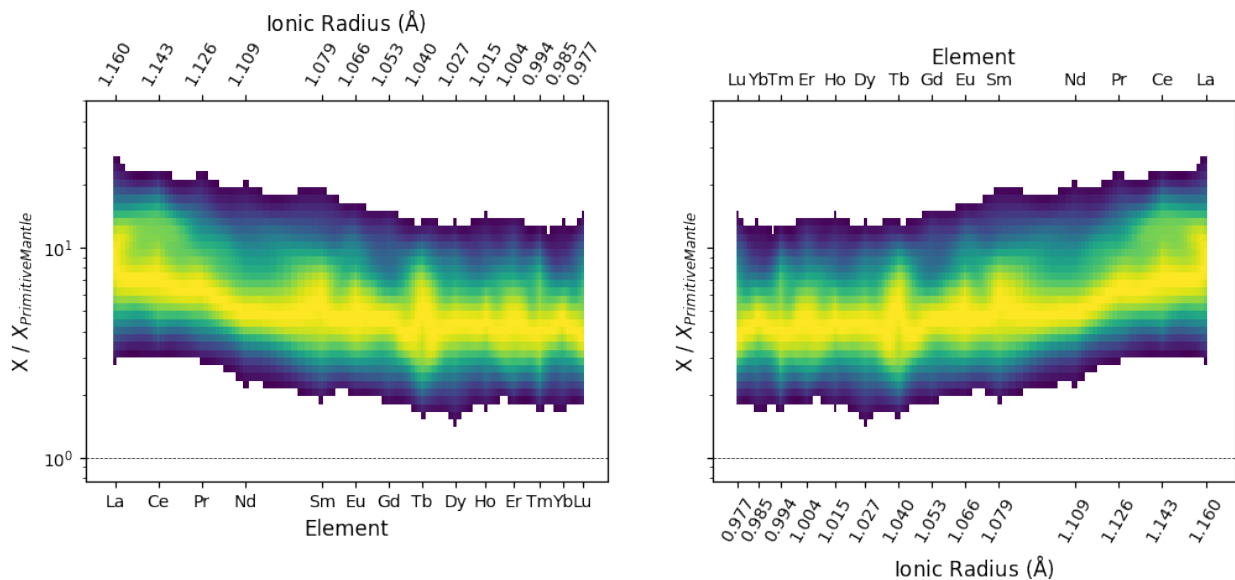
```
ax.scatter(
```

## REE Density Plots

Note that this can also be used for REE-indexed plots, in both configurations. Here we first specify a set of common keyword-argument configurations and use them for both plots:

```
REE_config = dict(unity_line=True, mode="binkde", vmin=0.05, resolution=10)
```

```
fig, ax = plt.subplots(1, 2, sharey=True, figsize=(12, 4))
normdf.pyroplot.REE(ax=ax[0], **REE_config)
normdf.pyroplot.REE(ax=ax[1], index="radii", **REE_config)
[a.set_ylabel("X / X_{PrimitiveMantle}") for a in ax]
plt.show()
```



```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
  ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
  ax.set_ylim(_ymin, _ymax)
```

See also:

[Heatscatter Plots, Density Diagrams](#)

**Total running time of the script:** (0 minutes 11.819 seconds)



### 3.3.2 Geochemistry Examples

#### Mineral Database

pyrolite includes a limited mineral database which is useful for looking up endmember compositions.

```
import pandas as pd

from pyrolite.mineral.mindb import (
    get_mineral,
    get_mineral_group,
    list_formulae,
    list_groups,
    list_minerals,
)

pd.set_option("display.precision", 3) # smaller outputs
```

From the database, you can get the list of its contents using a few utility functions:

```
list_groups()
```

```
['garnet', 'feldspar', 'pyroxene', 'olivine', 'spinel', 'epidote', 'mica', 'amphibole']
```

```
list_minerals()
```

```
['richterite', 'chromoceladonite', 'kaersutite', 'eastonite', 'almandine', 'paragonite',
→ 'anthophyllite', 'forsterite', 'spessartine', 'johannsenite', 'epidote',
→ 'ferroeckermanite', 'majorite', 'chromphyllite', 'manganiceladonite', 'winchite',
→ 'anorthite', 'aegirine', 'muscovite', 'gedrite', 'tephroite', 'polyolithionite',
→ 'glaucophane', 'celadonite', 'andradite', 'allanite', 'spinel', 'hercynite', 'uvarovite',
→ 'diopside', 'magnesioarfvedsonite', 'esseneite', 'spodumene', 'microcline',
→ 'arvedsonite', 'chromite', 'magnesiohastingsite', 'albite', 'katophorite', 'margarite',
→ 'magnesioreibeckite', 'pyrope', 'eckermanite', 'grossular', 'fayalite', 'ferroedenite',
→ 'clintonite', 'magnesioferrite', 'phengite', 'barroisite', 'phlogopite',
→ 'ferrotschermakite', 'hastingsite', 'annite', 'morimotoite', 'aluminoceladonite',
→ 'jadeite', 'ferrokaersutite', 'enstatite', 'tschermakite', 'clinozoisite', 'pargasite',
→ 'ferroceladonite', 'taramite', 'magnesiochromite', 'kosmochlor', 'magnesiohornblende',
→ 'ferrosilite', 'ferroaluminoceladonite', 'riebeckite', 'ferropargasite',
→ 'ferrorichterite', 'piemontite', 'liebenbergite', 'ferrokatophorite', 'namansilite',
→ 'tremolite', 'magnetite', 'trilithionite', 'edenite', 'siderophyllite', 'hedenbergite',
→ 'ferrohornblende']
```

```
list_formulae()
```

```
['Na(Na2)(Fe{2+}4Fe{3+})(Si8)O22(OH)2', 'KAlSi3O8', 'K2(Fe{2+}6)(Si6Al2)O20(OH)4',
→ 'K2(Al4)(Si6Al2)O20(OH)4', '(Na2)(Fe3Fe{3+}2)(Si8)O22(OH)2', 'K2(Al2Li2)(Si8)O20(OH)4',
→ 'Ca2Al2Mn{3+}(Si2O7)(SiO4)O(OH)', '(NaCa)(Mg4Al)(Si8)O22(OH)2',
→ '(Ca2)(Fe4Al)(Si7Al)O22(OH)2', 'NaAlSi3O8', 'Mg3(MgSi)(SiO4)3',
→ 'Na(Ca2)(Fe5)(Si7Al)O22(OH)2', '(Ca2)(Mg5)(Si8)O22(OH)2', 'Fe2SiO4', 'Mg3Al2(SiO4)3',
→ 'Na(Na2)(Mg4Al)(Si8)O22(OH)2', 'NaCrSi2O6', '(Na2)(Mg3Fe{3+}2)(Si8)O22(OH)2', 'K2(Mn
→ {3+}2Mg2)(Si8)O20(OH)4', 'K2(Al3Li3)(Si6Al2)O20(OH)4', 'K2(Fe{2+}2Al2)(Si8)O20(OH)4',
```

(continues on next page)

(continued from previous page)

```

→ 'Na(Ca2)(Mg4Al)(Si6Al2)O22(OH)2', 'Na(NaCa)(Fe4Al)(Si7Al)O22(OH)2', 'K2(Fe{2+}
→ 4)(Si4Al6)O20(OH)4', 'K2(Fe{3+}2Fe{2+}2)(Si8)O20(OH)4', 'Fe{2+}Al2O4', 'NaMn{3+}Si2O6',
→ 'Mn3Al2(SiO4)3', 'Na2(Al4)(Si6Al2)O20(OH)4', 'Na(Na2)(Fe4Al)(Si8)O22(OH)2', 'CaFeSi2O6
→ ', 'Na(NaCa)(Mg3Al2)(Si6Al2)O22(OH)2', 'Na(Na2)(Mg4Fe{3+})(Si8)O22(OH)2', 'K2(Mg2Cr{3+}
→ 2)(Si8)O20(OH)4', 'LiAlSi2O6', '(Ca2)(Mg3Al2)(Si6Al2)O22(OH)2', 'CaMgSi2O6',
→ 'Na(Ca2)(Fe{2+}4Fe{3+})(Si6Al2)O22(OH)2', 'NaAlSi2O6', 'CaAl2Si2O8',
→ 'Na(Ca2)(Mg4Ti)(Si6Al2)O22(OH)2', 'Fe{2+}Cr{3+}2O4', '(Mg2)(Mg3Al2)(Si6Al2)O22(OH)2',
→ 'NaFe{3+}Si2O6', 'Na(NaCa)(Mg5)(Si8)O22(OH)2', 'Na(Ca2)(Fe4Al)(Si6Al2)O22(OH)2',
→ 'K2(Mg2Al2)(Si8)O20(OH)4', 'Mg2SiO4', 'K2(Mg4)(Si4Al6)O20(OH)4', 'MgCr{3+}2O4',
→ 'Ca2(Mg4Al2)(Si2Al6)O20(OH)4', 'K2(Fe{3+}2Mg2)(Si8)O20(OH)4',
→ '(NaCa)(Mg3Al2)(Si7Al)O22(OH)2', 'Ca3Cr2(SiO4)3', '(Mg2)(Mg5)(Si8)O22(OH)2', 'MgFe{3+}
→ 2O4', 'Fe2Si2O6', '(Ca2)(Mg4Al)(Si7Al)O22(OH)2', 'CaCe{3+}Al2Fe{2+}(Si2O7)(SiO4)O(OH)',
→ 'Ca3Fe{3+}2(SiO4)3', 'Ca2Al3(Si2O7)(SiO4)O(OH)', 'Ca3Al2(SiO4)3', 'Ca2Al2Fe{3+}
→ (Si2O7)(SiO4)O(OH)', 'K2(Mg6)(Si6Al2)O20(OH)4', 'Ni1.5Mg0.5SiO4',
→ 'K2(Al3Mg)(Si7Al)O20(OH)4', 'CaMnSi2O6', 'Mn2SiO4', 'Na(Ca2)(Mg5)(Si7Al)O22(OH)2',
→ 'Na(Ca2)(Fe4Ti)(Si6Al2)O22(OH)2', '(Na2)(Mg3Al2)(Si8)O22(OH)2',
→ 'Ca2(Al4)(Si4Al4)O20(OH)4', 'Fe{2+}Fe{3+}2O4', 'Na(NaCa)(Fe5)(Si8)O22(OH)2', 'K2(Cr{3+}
→ 4)(Si6Al2)O20(OH)4', 'Na(Ca2)(Mg4Fe{3+})(Si6Al2)O22(OH)2', 'Fe{2+}3Al2(SiO4)3', 'CaAlFe
→ {3+}SiO6', 'Na(NaCa)(Mg4Al)(Si7Al)O22(OH)2', 'MgAl2O4', 'Ca3(TiFe{2+})(SiO4)3',
→ '(Ca2)(Fe3Al2)(Si6Al2)O22(OH)2', 'Mg2Si2O6']

```

You can also directly get the composition of specific minerals by name:

```
get_mineral("forsterite")
```

```

name      forsterite
group      olivine
formula    Mg2SiO4
Mg         0.346
Si         0.2
O          0.455
Fe         0.0
Mn         0.0
Ni         0.0
Ca         0.0
Al         0.0
Fe{3+}     0.0
Na         0.0
Mn{3+}     0.0
Cr         0.0
Li         0.0
Cr{3+}     0.0
Fe{2+}     0.0
K          0.0
H          0.0
Ti         0.0
Ce{3+}     0.0
dtype: object

```

If you want to get compositions for all minerals within a specific group, you can use `get_mineral_group()`:

```
get_mineral_group("olivine")
```

Total running time of the script: (0 minutes 0.019 seconds)

## Element-Oxide Transformation

One of pyrolite's strengths is converting mixed elemental and oxide data to a new form. The simplest way to perform this is by using the `convert_chemistry()` function. Note that by default pyrolite assumes that data are in the same units.

```
import pandas as pd

import pyrolite.geochem

pd.set_option("display.precision", 3) # smaller outputs
```

Here we create some synthetic data to work with, which has some variables in Wt% and some in ppm. Notably some elements are present in more than one column (Ca, Na):

```
from pyrolite.util.synthetic import normal_frame

df = normal_frame(columns=["MgO", "SiO2", "FeO", "CaO", "Na2O", "Te", "K", "Na"]) * 100
df.pyrochem.elements *= 100 # elements in ppm
```

```
df.head(2)
```

As the units are heterogeneous, we'll need to convert the data frame to a single set of units (here we use Wt%):

```
df.pyrochem.elements = df.pyrochem.elements.pyrochem.scale("ppm", "wt%") # ppm to wt%
```

We can transform this chemical data to a new set of compositional variables. Here we i) convert CaO to Ca, ii) aggregate Na2O and Na to Na and iii) calculate mass ratios for Na/Te and MgO/SiO2. Note that you can also use this function to calculate mass ratios:

```
df.pyrochem.convert_chemistry(
    to=["MgO", "SiO2", "FeO", "Ca", "Te", "Na", "Na/Te", "MgO/SiO2"]
).head(2)
```

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[3.67297884 3.50325632 3.
↳ 66224102 3.25764225 3.52937765 3.9315398
↳ 3.84124843 3.39657993 4.14447134 3.84324796]' has dtype incompatible with int64, please
↳ explicitly cast to a compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
```

You can also specify molar ratios for iron redox, which will result in multiple iron species within the single dataframe:

```
df.pyrochem.convert_chemistry(to=[{"FeO": 0.9, "Fe2O3": 0.1}]).head(2)
```

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[0.66875538 0.6503792 0.
↳ 68473922 0.63454072 0.59756727 0.67044164
↳ 0.69027087 0.70640098 0.70099976 0.66848851]' has dtype incompatible with int64, please
```

(continues on next page)

(continued from previous page)

```
↪ explicitly cast to a compatible dtype first.  
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
```

**Total running time of the script:** (0 minutes 0.152 seconds)

## Normalization

A selection of reference compositions are included in pyrolite, and can be easily accessed with `pyrolite.geochem.norm.get_reference_composition()` (see the list at the bottom of the page for a complete list):

```
import matplotlib.pyplot as plt  
import pandas as pd  
  
import pyrolite.plot  
from pyrolite.geochem.ind import REE  
from pyrolite.geochem.norm import all_reference_compositions, get_reference_composition
```

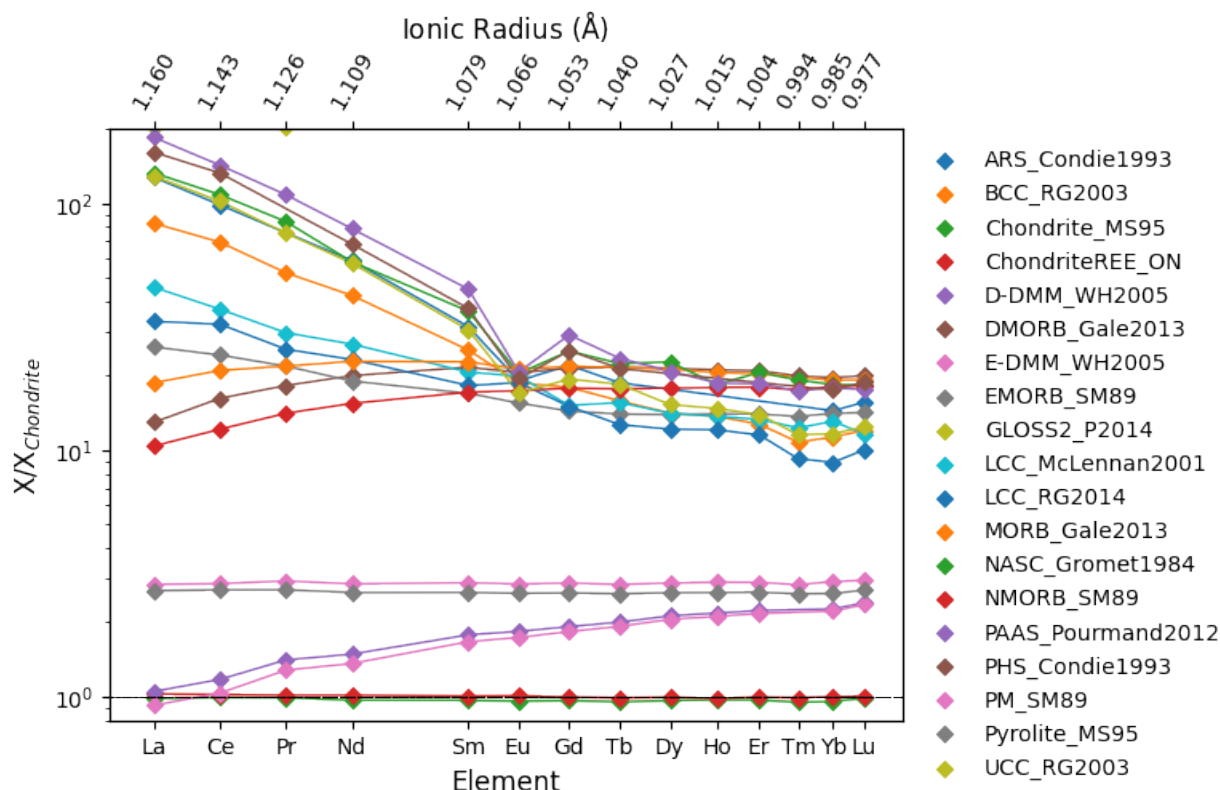
```
chondrite = get_reference_composition("Chondrite_PON")
```

To use the compositions with a specific set of units, you can change them with `set_units()`:

```
CI = chondrite.set_units("ppm")
```

The `normalize_to()` method can be used to normalise DataFrames to a given reference (e.g. for spiderplots):

```
fig, ax = plt.subplots(1)  
  
for name, ref in list(all_reference_compositions().items())[::2]:  
    if name != "Chondrite_PON":  
        ref.set_units("ppm")  
        df = ref.comp.pyrochem.REE.pyrochem.normalize_to(CI, units="ppm")  
        df.pyroplot.REE(unity_line=True, ax=ax, label=name)  
  
ax.set_ylabel("X/X$_{Chondrite}$")  
ax.legend()  
plt.show()
```



```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→will be ignored.
    ax.set_ylim(_ymin, _ymax)

```

(continues on next page)

(continued from previous page)

```
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↪plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↪will be ignored.
    ax.set_ylim(_ymin, _ymax)
```

**See also:**

**Examples:**

[lambdas: Parameterising REE Profiles, REE Radii Plot](#)

Currently available models include:

**Total running time of the script:** (0 minutes 2.377 seconds)

## Geochemical Indexes and Selectors

```
import pandas as pd

import pyrolite.geochem

pd.set_option("display.precision", 3) # smaller outputs
```

```
from pyrolite.util.synthetic import normal_frame

df = normal_frame(
    columns=[
        "CaO",
        "MgO",
        "SiO2",
        "FeO",
        "Mn",
        "Ti",
        "La",
        "Lu",
        "Y",
        "Mg/Fe",
        "87Sr/86Sr",
        "Ar40/Ar36",
    ]
)
```

```
df.head(2).pyrochem.oxides
```

```
df.head(2).pyrochem.elements
```

```
df.head(2).pyrochem.REE
```

```
df.head(2).pyrochem.REY
```

```
df.head(2).pyrochem.compositional
```

```
df.head(2).pyrochem.isotope_ratios
```

```
df.pyrochem.list_oxides
```

```
['CaO', 'MgO', 'SiO2', 'FeO']
```

```
df.pyrochem.list_elements
```

```
['Mn', 'Ti', 'La', 'Lu', 'Y']
```

```
df.pyrochem.list_REE
```

```
['La', 'Lu']
```

```
df.pyrochem.list_compositional
```

```
['CaO', 'MgO', 'SiO2', 'FeO', 'Mn', 'Ti', 'La', 'Lu', 'Y']
```

```
df.pyrochem.list_isotope_ratios
```

```
['87Sr/86Sr', 'Ar40/Ar36']
```

All elements (up to U):

```
from pyrolite.geochem.ind import REE, REY, common_elements, common_oxides
```

```
common_elements() # string return
```

```
['H', 'He', 'Li', 'Be', 'B', 'C', 'N', 'O', 'F', 'Ne', 'Na', 'Mg', 'Al', 'Si', 'P', 'S',
→ 'Cl', 'Ar', 'K', 'Ca', 'Sc', 'Ti', 'V', 'Cr', 'Mn', 'Fe', 'Co', 'Ni', 'Cu', 'Zn', 'Ga',
→ 'Ge', 'As', 'Se', 'Br', 'Kr', 'Rb', 'Sr', 'Y', 'Zr', 'Nb', 'Mo', 'Tc', 'Ru', 'Rh', 'Pd',
→ 'Ag', 'Cd', 'In', 'Sn', 'Sb', 'Te', 'I', 'Xe', 'Cs', 'Ba', 'La', 'Ce', 'Pr', 'Nd',
→ 'Pm', 'Sm', 'Eu', 'Gd', 'Tb', 'Dy', 'Ho', 'Er', 'Tm', 'Yb', 'Lu', 'Hf', 'Ta', 'W', 'Re',
→ 'Os', 'Ir', 'Pt', 'Au', 'Hg', 'Tl', 'Pb', 'Bi', 'Po', 'At', 'Rn', 'Fr', 'Ra', 'Ac',
→ 'Th', 'Pa', 'U']
```

All elements, returned as a list of *~periodictable.core.Formula*:

```
common_elements(output="formula") # periodictable.core.Formula return
```

```
[H, He, Li, Be, B, C, N, O, F, Ne, Na, Mg, Al, Si, P, S, Cl, Ar, K, Ca, Sc, Ti, V, Cr,
→ Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Kr, Rb, Sr, Y, Zr, Nb, Mo, Tc, Ru, Rh, Pd,
→ Ag, Cd, In, Sn, Sb, Te, I, Xe, Cs, Ba, La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er,
→ Tm, Yb, Lu, Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg, Tl, Pb, Bi, Po, At, Rn, Fr, Ra, Ac, Th,
→ Pa, U]
```

Oxides for elements with positive charges (up to U):

```
common_oxides()
```

```
['In2O', 'InO', 'In2O3', 'As2O', 'AsO', 'As2O3', 'AsO2', 'As2O5', 'Ac2O3', 'Pt2O', 'PtO',
→ 'Pt2O3', 'PtO2', 'Pt2O5', 'PtO3', 'Ga2O', 'GaO', 'Ga2O3', 'Y2O', 'YO', 'Y2O3', 'Ta2O',
→ 'TaO', 'Ta2O3', 'TaO2', 'Ta2O5', 'PrO', 'Pr2O3', 'PrO2', 'Pr2O5', 'Hf2O', 'HfO',
→ 'Hf2O3', 'HfO2', 'S2O', 'SO', 'S2O3', 'SO2', 'S2O5', 'SO3', 'SmO', 'Sm2O3', 'At2O',
→ 'At2O3', 'At2O5', 'At2O7', 'RnO', 'RnO3', 'Th2O', 'ThO', 'Th2O3', 'ThO2', 'Sb2O', 'SbO',
→ 'Sb2O3', 'SbO2', 'Sb2O5', 'Os2O', 'OsO', 'Os2O3', 'OsO2', 'Os2O5', 'OsO3', 'Os2O7',
→ 'OsO4', 'YbO', 'Yb2O3', 'Tb2O', 'TbO', 'Tb2O3', 'TbO2', 'Mo2O', 'MoO', 'Mo2O3', 'MoO2',
→ 'Mo2O5', 'MoO3', 'Ge2O', 'GeO', 'Ge2O3', 'GeO2', 'Rb2O', 'CeO', 'Ce2O3', 'CeO2', 'Na2O',
→ 'Gd2O', 'GdO', 'Gd2O3', 'K2O', 'Pa2O3', 'PaO2', 'Pa2O5', 'TmO', 'Tm2O3', 'Te2O',
→ 'TeO', 'Te2O3', 'TeO2', 'Te2O5', 'TeO3', 'Se2O', 'SeO', 'Se2O3', 'SeO2', 'Se2O5', 'SeO3',
→ 'ErO', 'Er2O3', 'DyO', 'Dy2O3', 'DyO2', 'Bi2O', 'BiO', 'Bi2O3', 'BiO2', 'Bi2O5',
→ 'HoO', 'Ho2O3', 'Nb2O', 'NbO', 'Nb2O3', 'NbO2', 'Nb2O5', 'Ru2O', 'RuO', 'Ru2O3', 'RuO2',
→ 'Ru2O5', 'RuO3', 'Ru2O7', 'RuO4', 'RaO', 'P2O', 'PO', 'P2O3', 'PO2', 'P2O5', 'Li2O',
```

(continues on next page)



(continued from previous page)

```

→ 'La2O', 'LaO', 'La2O3', 'EuO', 'Eu2O3', 'Cl2O', 'ClO', 'Cl2O3', 'ClO2', 'Cl2O5', 'ClO3
→ ', 'Cl2O7', 'Fe2O', 'FeO', 'Fe2O3', 'FeO2', 'Fe2O5', 'FeO3', 'Fe2O7', 'Ca2O', 'CaO',
→ 'PoO', 'PoO2', 'Po2O5', 'PoO3', 'Fr2O', 'Cd2O', 'CdO', 'Tc2O', 'TcO', 'Tc2O3', 'TcO2',
→ 'Tc2O5', 'TcO3', 'Tc2O7', 'H2O', 'Hg2O', 'HgO', 'HgO2', 'U2O', 'UO', 'U2O3', 'UO2',
→ 'U2O5', 'UO3', 'Mg2O', 'MgO', 'Pb2O', 'PbO', 'Pb2O3', 'PbO2', 'Cs2O', 'Ag2O', 'AgO',
→ 'Ag2O3', 'AgO2', 'Ti2O', 'TiO', 'Ti2O3', 'TiO2', 'NdO', 'Nd2O3', 'NdO2', 'W2O', 'WO',
→ 'W2O3', 'WO2', 'W2O5', 'WO3', 'Br2O', 'Br2O3', 'BrO2', 'Br2O5', 'Br2O7', 'Zn2O', 'ZnO',
→ 'Cu2O', 'CuO', 'Cu2O3', 'CuO2', 'C2O', 'CO', 'C2O3', 'CO2', 'B2O', 'BO', 'B2O3', 'N2O
→ ', 'NO', 'N2O3', 'NO2', 'N2O5', 'Be2O', 'BeO', 'Ir2O', 'IrO', 'Ir2O3', 'IrO2', 'Ir2O5',
→ 'IrO3', 'Ir2O7', 'IrO4', 'Ir2O9', 'Sc2O', 'ScO', 'Sc2O3', 'Co2O', 'CoO', 'Co2O3',
→ 'CoO2', 'Co2O5', 'Cr2O', 'CrO', 'Cr2O3', 'CrO2', 'Cr2O5', 'CrO3', 'V2O', 'VO', 'V2O3',
→ 'VO2', 'V2O5', 'Mn2O', 'MnO', 'Mn2O3', 'MnO2', 'Mn2O5', 'MnO3', 'Mn2O7', 'Pd2O', 'PdO',
→ 'Pd2O3', 'PdO2', 'Pd2O5', 'PdO3', 'Rh2O', 'RhO', 'Rh2O3', 'RhO2', 'Rh2O5', 'RhO3',
→ 'Re2O', 'ReO', 'Re2O3', 'ReO2', 'Re2O5', 'ReO3', 'Re2O7', 'Au2O', 'AuO', 'Au2O3',
→ 'Au2O5', 'Ni2O', 'NiO', 'Ni2O3', 'NiO2', 'Tl2O', 'TlO', 'Tl2O3', 'Al2O', 'AlO', 'Al2O3
→ ', 'Sn2O', 'SnO', 'Sn2O3', 'SnO2', 'Ba2O', 'BaO', 'Zr2O', 'ZrO', 'Zr2O3', 'ZrO2', 'I2O
→ ', 'I2O3', 'IO2', 'I2O5', 'IO3', 'I2O7', 'Si2O', 'SiO', 'Si2O3', 'SiO2', 'LuO', 'Lu2O3
→ ', 'PmO', 'Pm2O3', 'Sr2O', 'SrO', 'FeOT', 'Fe2O3T', 'LOI']

```

```
REE()
```

```
['La', 'Ce', 'Pr', 'Nd', 'Sm', 'Eu', 'Gd', 'Tb', 'Dy', 'Ho', 'Er', 'Tm', 'Yb', 'Lu']
```

```
REY()
```

```
['La', 'Ce', 'Pr', 'Nd', 'Sm', 'Eu', 'Gd', 'Tb', 'Dy', 'Y', 'Ho', 'Er', 'Tm', 'Yb', 'Lu']
```

**Total running time of the script:** (0 minutes 0.295 seconds)

## Mineral Endmember Decomposition

A common task when working with mineral chemistry data is to take measured compositions and decompose these into relative proportions of mineral endmember compositions. pyrolite includes some utilities to achieve this and a limited mineral database for looking up endmember compositions.

```

import numpy as np
import pandas as pd

from pyrolite.mineral.mindb import get_mineral
from pyrolite.mineral.normative import endmember_decompose

```

First we'll start with a composition of an unknown olivine:

```
comp = pd.Series({"MgO": 42.06, "SiO2": 39.19, "FeO": 18.75})
```

We can break this down into olivine endmembers using the `endmember_decompose()` function:

```

ed = endmember_decompose(
    pd.DataFrame(comp).T, endmembers="olivine", order=1, molecular=True
)
ed

```

```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[0.57294068      nan
↳      nan 0.10480888]' has dtype incompatible with int64, please explicitly cast to a
↳ compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[0.42705932 0.29485884 0.
↳ 29750712 0.31249035]' has dtype incompatible with int64, please explicitly cast to a
↳ compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[      nan 0.70514116
↳      nan      nan]' has dtype incompatible with int64, please explicitly cast to a
↳ compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]

```

Equally, if you knew the likely endmembers beforehand, you could specify a list of endmembers:

```

ed = endmember_decompose(
    pd.DataFrame(comp).T, endmembers=["forsterite", "fayalite"], order=1, molecular=True
)
ed

```

```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[0.57294068      nan]'
↳ has dtype incompatible with int64, please explicitly cast to a compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[0.42705932 0.29485884]'
↳ has dtype incompatible with int64, please explicitly cast to a compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[      nan 0.70514116]'
↳ has dtype incompatible with int64, please explicitly cast to a compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]

```

We can check this by recombining the components with these proportions. We can first lookup the compositions for our endmembers:

```

em = pd.DataFrame([get_mineral("forsterite"), get_mineral("fayalite")])
em.loc[:, ~(em == 0).all(axis=0)] # columns not full of zeros

```

First we have to convert these element-based compositions to oxide-based compositions:

```

emvalues = (
    em.loc[:, ["Mg", "Si", "Fe"]]
    .pyrochem.to_molecular()

```

(continues on next page)

(continued from previous page)

```

    .fillna(0)
    .pyrochem.convert_chemistry(to=["MgO", "SiO2", "FeO"], molecular=True)
    .fillna(0)
    .pyrocomp.renormalise(scale=1)
)
emvalues

```

```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[66.66666667      nan]
↳ ' has dtype incompatible with int64, please explicitly cast to a compatible dtype
↳ first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[33.33333333 33.33333333]
↳ ' has dtype incompatible with int64, please explicitly cast to a compatible dtype
↳ first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[      nan 66.66666667]
↳ ' has dtype incompatible with int64, please explicitly cast to a compatible dtype
↳ first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]

```

These can now be used with our endmember proportions to regenerate a composition:

```

recombined = pd.DataFrame(ed.values.flatten() @ emvalues).T.pyrochem.to_weight()
recombined

```

To make sure these compositions are within 0.01 percent:

```

assert np.allclose(recombined.values, comp.values, rtol=10**-4)

```

**Total running time of the script:** (0 minutes 0.215 seconds)

## Ionic Radii

pyrolite includes a few sets of reference tables for ionic radii in aangstroms (Å) from [?] and [?], each with tables indexed by element, ionic charge and coordination. The easiest way to access these is via the `get_ionic_radii()` function. The function can be used to get radii for individual elements:

```

from pyrolite.geochem.ind import REE, get_ionic_radii

Cu_radii = get_ionic_radii("Cu")
print(Cu_radii)

```

```

index
Cu2+IV      0.57
Cu2+IVSQ    0.57

```

(continues on next page)

(continued from previous page)

```
Cu2+V      0.65
Cu2+VI     0.73
Name: ionicradius, dtype: float64
```

Note that this function returned a series of the possible radii, given specific charges and coordinations of the Cu ion. If we completely specify these, we'll get a single number back:

```
Cu2plus6fold_radii = get_ionic_radii("Cu", coordination=6, charge=2)
print(Cu2plus6fold_radii)
```

```
0.73
```

You can also pass lists to the function. For example, if you wanted to get the Shannon ionic radii of Rare Earth Elements (REE) in eight-fold coordination with a valence of +3, you should use the following:

```
shannon_ionic_radii = get_ionic_radii(REE(), coordination=8, charge=3)
print(shannon_ionic_radii)
```

```
[1.16, 1.143, 1.126, 1.109, 1.079, 1.066, 1.053, 1.04, 1.027, 1.015, 1.004, 0.994, 0.985,
 → 0.977]
```

The function defaults to using the Shannon ionic radii consistent with [?], but you can adjust to use the set you like with the *pauling* boolean argument (`pauling=False` to use Shannon's 'Crystal Radii') or the *source* argument (`source='Whittaker'` to use the [?] dataset):

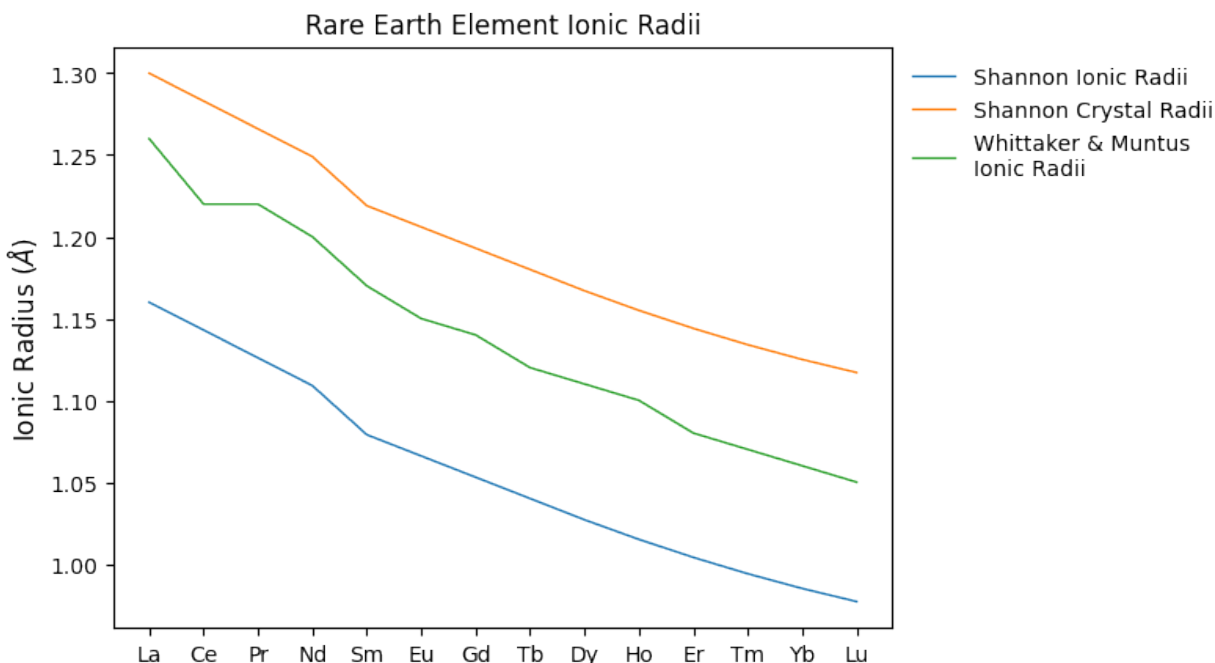
```
shannon_crystal_radii = get_ionic_radii(REE(), coordination=8, charge=3, pauling=False)
whittaker_ionic_radii = get_ionic_radii(
    REE(), coordination=8, charge=3, source="Whittaker"
)
```

We can see what the differences between these look like across the REE:

```
import matplotlib.pyplot as plt

fig, ax = plt.subplots(1)

ax.plot(shannon_ionic_radii, label="Shannon Ionic Radii")
ax.plot(shannon_crystal_radii, label="Shannon Crystal Radii")
ax.plot(whittaker_ionic_radii, label="Whittaker & Muntus\nIonic Radii")
{a: b for (a, b) in zip(REE(), whittaker_ionic_radii)}
ax.set_xticks(range(len(REE())))
ax.set_xticklabels(REE())
ax.set_ylabel("Ionic Radius ($\AA$)")
ax.set_title("Rare Earth Element Ionic Radii")
ax.legend()
plt.show()
```



See also:

#### Examples:

`lambdas`: Parameterising REE Profiles, REE Radii Plot

#### Functions:

`get_ionic_radii()`, `pyrolite.geochem.ind.REE()`, `lambda_lnREE()`,

#### References

**Total running time of the script:** (0 minutes 0.345 seconds)

#### Unit Scaling

```
import numpy as np
import pandas as pd

import pyrolite.geochem

pd.set_option("display.precision", 3) # smaller outputs
```

Here we create an example dataframe to work from, containing some synthetic data in the form of oxides and elements, each with different units (wt% and ppm, respectively).

```
from pyrolite.util.synthetic import normal_frame

df = normal_frame(
    columns=["CaO", "MgO", "SiO2", "FeO", "Ni", "Ti", "La", "Lu"], seed=22
)
df.pyrochem.oxides *= 100 # oxides in wt%
df.pyrochem.elements *= 10000 # elements in ppm
```

In this case, we might want to transform the Ni and Ti into their standard oxide equivalents NiO and TiO<sub>2</sub>:

```
df.pyrochem.convert_chemistry(to=["NiO", "TiO2"]).head(2)
```

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[1235.19505234 1301.
↳ 09283284 1297.67084898 1332.78130758 1231.28287236
1223.09485893 1213.13011719 1220.86333288 1269.99495233 1193.1419757 ]' has dtype
↳ incompatible with int64, please explicitly cast to a compatible dtype first.
  _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[2388.2200631 2164.
↳ 62285792 2286.2934774 2443.58980729 2245.75825742
2269.68007197 2130.01547469 2282.00522338 2257.57443799 2515.3361174 ]' has dtype
↳ incompatible with int64, please explicitly cast to a compatible dtype first.
  _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
```

But here because Ni and Ti have units of ppm, the results are a little non-sensical, especially when it's combined with the other oxides:

```
df.pyrochem.convert_chemistry(to=df.pyrochem.list_oxides + ["NiO", "TiO2"]).head(2)
```

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[1235.19505234 1301.
↳ 09283284 1297.67084898 1332.78130758 1231.28287236
1223.09485893 1213.13011719 1220.86333288 1269.99495233 1193.1419757 ]' has dtype
↳ incompatible with int64, please explicitly cast to a compatible dtype first.
  _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[2388.2200631 2164.
↳ 62285792 2286.2934774 2443.58980729 2245.75825742
2269.68007197 2130.01547469 2282.00522338 2257.57443799 2515.3361174 ]' has dtype
↳ incompatible with int64, please explicitly cast to a compatible dtype first.
  _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
```

There are multiple ways we could convert the units, but here we're going to first convert the elemental ppm data to wt%, then perform our oxide-element conversion. To do this, we'll use the built-in function `scale()`:

```
from pyrolite.util.units import scale

df.pyrochem.elements *= scale("ppm", "wt%")
```

We can see that this then gives us numbers which are a bit more sensible:

```
df.pyrochem.convert_chemistry(to=df.pyrochem.list_oxides + ["NiO", "TiO2"]).head(2)
```

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[0.12351951 0.13010928 0.
```

(continues on next page)

(continued from previous page)

```

↳12976708 0.13327813 0.12312829 0.12230949
0.12131301 0.12208633 0.1269995 0.1193142 ]' has dtype incompatible with int64, please_
↳explicitly cast to a compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is_
↳deprecated and will raise in a future error of pandas. Value '[0.23882201 0.21646229 0.
↳22862935 0.24435898 0.22457583 0.22696801
0.21300155 0.22820052 0.22575744 0.25153361]' has dtype incompatible with int64, please_
↳explicitly cast to a compatible dtype first.
    _df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]

```

## Dealing with Units in Column Names

Often our dataframes will start containing column names which pyrolite doesn't recognize natively by default (work in progress, this is an item on the roadmap). Here we can create an example of that, and go through some key steps for using this data in pyrolite:

```

df = normal_frame(
    columns=["CaO", "MgO", "SiO2", "FeO", "Ni", "Ti", "La", "Lu"], seed=22
)
df.pyrochem.oxides *= 100 # oxides in wt%
df.pyrochem.elements *= 10000 # elements in ppm
df = df.rename(
    columns={
        **{c: c + "_wt%" for c in df.pyrochem.oxides},
        **{c: c + "_ppm" for c in df.pyrochem.elements},
    }
)
df.head(2)

```

If you just wanted to rescale some columns, you can get away without renaming your columns, e.g. converting all of the ppm columns to wt%:

```

df.loc[:, [c for c in df.columns if "_ppm" in c]] *= scale("ppm", "wt%")
df.head(2)

```

However, to access the full native capability of pyrolite, we'd need to rename these columns to use things like `convert_chemistry()`:

```

units = { # keep a copy of the units, we can use these to map back later
    c: c[c.find("_") + 1 :] if "_" in c else None for c in df.columns
}
df = df.rename(
    columns={c: c.replace("_wt%", "").replace("_ppm", "") for c in df.columns}
)
df.head(2)

```

We could then perform our chemistry conversion, rename our columns to include units, and optionally export to e.g. CSV:

```
converted_wt_pct = df.pyrochem.convert_chemistry(  
    to=df.pyrochem.list_oxides + ["NiO", "TiO2"]  
)  
converted_wt_pct.head(2)
```

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/  
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is_  
↳ deprecated and will raise in a future error of pandas. Value '[0.12351951 0.13010928 0.  
↳ 12976708 0.13327813 0.12312829 0.12230949  
0.12131301 0.12208633 0.1269995 0.1193142 ]' has dtype incompatible with int64, please_  
↳ explicitly cast to a compatible dtype first.  
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]  
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/  
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is_  
↳ deprecated and will raise in a future error of pandas. Value '[0.23882201 0.21646229 0.  
↳ 22862935 0.24435898 0.22457583 0.22696801  
0.21300155 0.22820052 0.22575744 0.25153361]' has dtype incompatible with int64, please_  
↳ explicitly cast to a compatible dtype first.  
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
```

Here we rename the columns before we export them, just so we know explicitly what the units are:

```
converted_wt_pct = converted_wt_pct.rename(  
    columns={c: c + "_wt%" for c in converted_wt_pct.pyrochem.list_oxides}  
)  
converted_wt_pct.head(2)
```

```
converted_wt_pct.to_csv("converted_wt_pct.csv")
```

**Total running time of the script:** (0 minutes 0.320 seconds)

## Lattice Strain Calculations

---

**Note:** This example follows that given during a Institute of Advanced Studies Masterclass with Jon Blundy at the University of Western Australia on the 29<sup>th</sup> April 2019, and is used here with permission.

---

Pyrolite includes a function for calculating relative lattice strain<sup>1</sup>, which together with the tables of Shannon ionic radii and Young’s modulus approximations for silicate and oxide cationic sites enable relatively simple calculation of ionic partitioning in common rock forming minerals.

This example below uses previously characterised calcium and sodium partition coefficients between plagioclase ( $CaAl_2Si_2O_8 - NaAlSi_3O_8$ ) and silicate melt to estimate partitioning for other cations based on their ionic radii.

A model parameterised using sodium and calcium partition coefficients<sup>2</sup> is then used to estimate the partitioning for lanthanum into the trivalent site (largely occupied by  $Al^{3+}$ ), and extended to other trivalent cations (here, the Rare Earth Elements). The final section of the example highlights the mechanism which generates plagioclase’s hallmark ‘europium anomaly’, and the effects of variable europium oxidation state on bulk europium partitioning.

---

<sup>1</sup> Blundy, J., Wood, B., 1994. Prediction of crystal–melt partition coefficients from elastic moduli. *Nature* 372, 452. doi: [10.1038/372452A0](https://doi.org/10.1038/372452A0)

<sup>2</sup> Dohmen, R., Blundy, J., 2014. A predictive thermodynamic model for element partitioning between plagioclase and melt as a function of pressure, temperature and composition. *American Journal of Science* 314, 1319–1372. doi: [10.2475/09.2014.04](https://doi.org/10.2475/09.2014.04)



```
import matplotlib.pyplot as plt
import numpy as np

from pyrolite.geochem.ind import REE, get_ionic_radii
from pyrolite.mineral.lattice import strain_coefficient
```

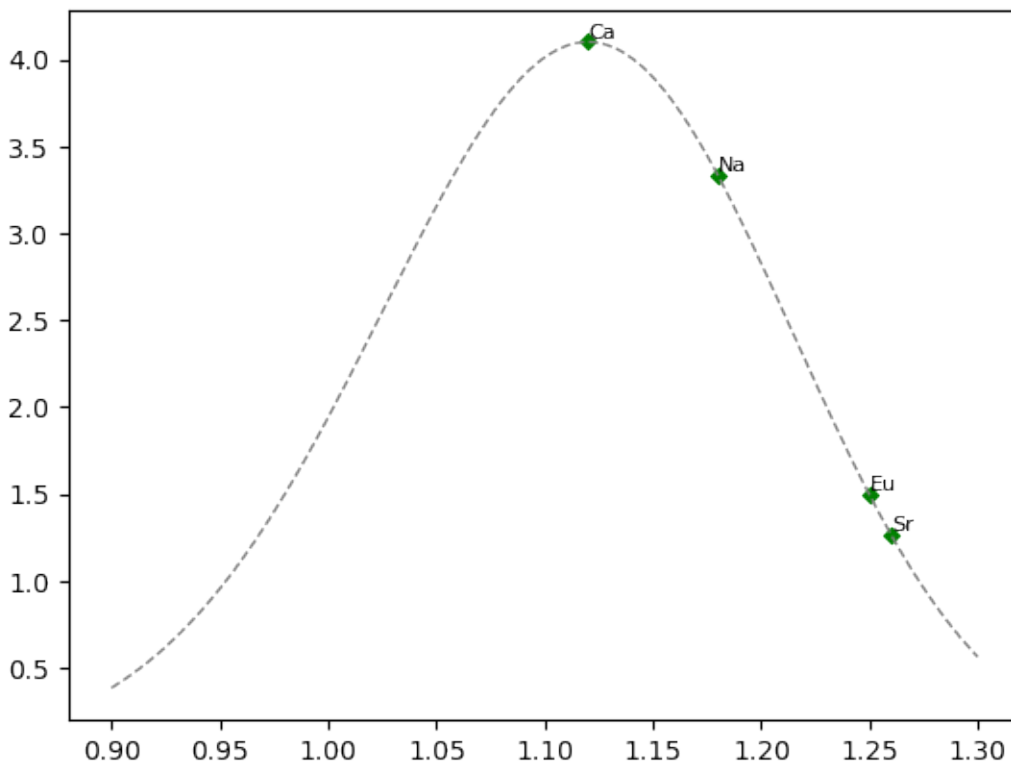
First, we need to define some of the necessary parameters including temperature, the Young's moduli for the  $X^{2+}$  and  $X^{3+}$  sites in plagioclase ( $E_2$ ,  $E_3$ ), and some reference partition coefficients and radii for calcium and sodium:

```
D_Na = 1.35 # Partition coefficient Plag-Melt
D_Ca = 4.1 # Partition coefficient Plag-Melt
Tc = 900 # Temperature, °C
Tk = Tc + 273.15 # Temperature, K
E_2 = 120 * 10**9 # Youngs modulus for 2+ site, Pa
E_3 = 135 * 10**9 # Youngs modulus for 3+ site, Pa
r02, r03 = 1.196, 1.294 # fictive ideal cation radii for these sites
rCa = get_ionic_radii("Ca", charge=2, coordination=8)
rLa = get_ionic_radii("La", charge=3, coordination=8)
```

We can calculate and plot the partitioning of  $X^{2+}$  cations relative to  $Ca^{2+}$  at a given temperature using their radii and the lattice strain function:

```
fontsize = 8
fig, ax = plt.subplots(1)

site2labels = ["Na", "Ca", "Eu", "Sr"]
# get the Shannon ionic radii for the elements in the 2+ site
site2radii = [
    get_ionic_radii("Na", charge=1, coordination=8),
    *[get_ionic_radii(el, charge=2, coordination=8) for el in ["Ca", "Eu", "Sr"]],
]
# plot the relative partitioning curve for cations in the 2+ site
site2Ds = D_Ca * np.array(
    [strain_coefficient(rCa, rx, r0=r02, E=E_2, T=Tk) for rx in site2radii]
)
ax.scatter(site2radii, site2Ds, color="g", label="$X^{2+}$ Cations")
# create an index of radii, and plot the relative partitioning curve for the site
xs = np.linspace(0.9, 1.3, 200)
curve2Ds = D_Ca * strain_coefficient(rCa, xs, r0=r02, E=E_2, T=Tk)
ax.plot(xs, curve2Ds, color="0.5", ls="--")
# add the element labels next to the points
for l, r, d in zip(site2labels, site2radii, site2Ds):
    ax.annotate(
        l, xy=(r, d), xycoords="data", ha="left", va="bottom", fontsize=fontsize
    )
fig
```



<Figure size 640x480 with 1 Axes>

When it comes to estimating the partitioning of  $X^{3+}$  cations, we'll need a reference point - here we'll use  $D_{La}$  to calculate relative partitioning of the other Rare Earth Elements, although you may have noticed it is not defined above. Through a handy relationship, we can estimate  $D_{La}$  based on the easier measured  $D_{Ca}$ ,  $D_{Na}$  and temperature<sup>Page 76, 2</sup>:

```
D_La = (D_Ca**2 / D_Na) * np.exp((529 / Tk) - 3.705)
D_La # 0.48085
```

```
0.48084614946362086
```

Now  $D_{La}$  is defined, we can use it as a reference for the other REE:

```
site3labels = REE(dropPm=True)
# get the Shannon ionic radii for the elements in the 3+ site
site3radii = [get_ionic_radii(x, charge=3, coordination=8) for x in REE(dropPm=True)]
site3Ds = D_La * np.array(
    [strain_coefficient(rLa, rx, r0=r03, E=E_3, T=Tk) for rx in site3radii]
)
# plot the relative partitioning curve for cations in the 3+ site
ax.scatter(site3radii, site3Ds, color="purple", label="$X^{3+}$ Cations")
# plot the relative partitioning curve for the site
curve3Ds = D_La * strain_coefficient(rLa, xs, r0=r03, E=E_3, T=Tk)
ax.plot(xs, curve3Ds, color="0.5", ls="--")
# add the element labels next to the points
for l, r, d in zip(site3labels, site3radii, site3Ds):
    ax.annotate(
        l, xy=(r, d), xycoords="data", ha="right", va="bottom", fontsize=fontsize
```

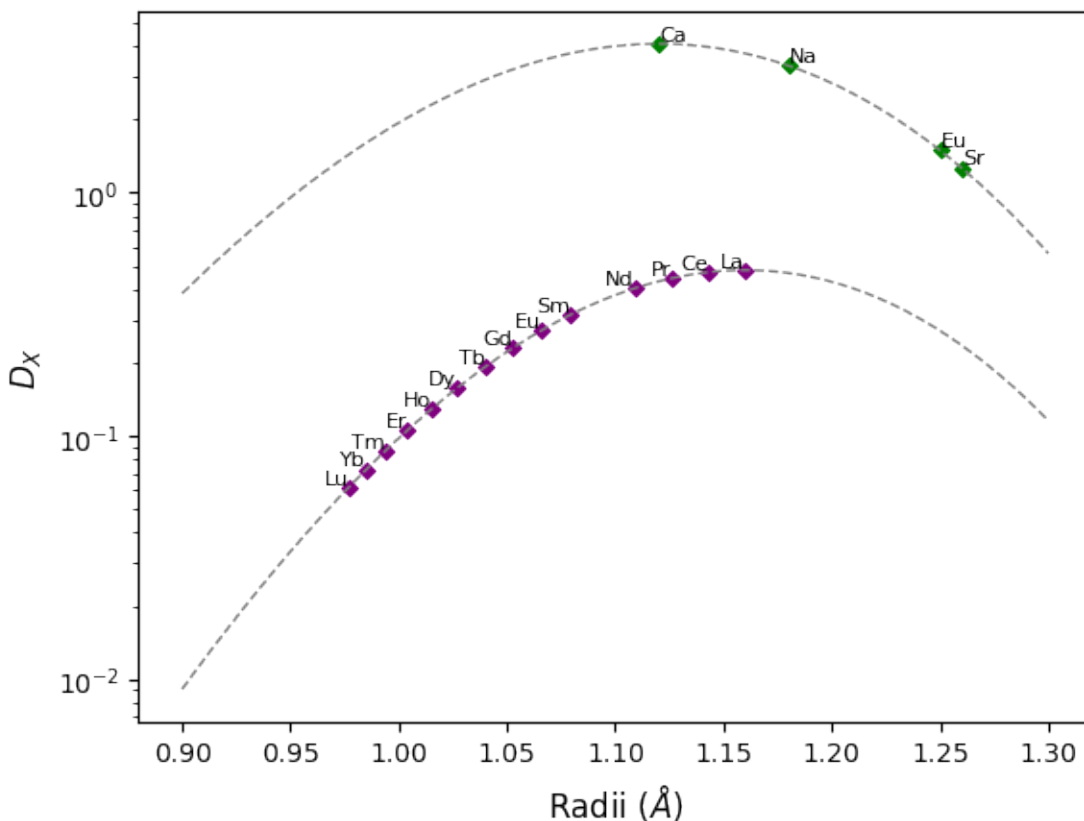
(continues on next page)

(continued from previous page)

```

)
ax.set_yscale("log")
ax.set_ylabel("$D_X$")
ax.set_xlabel("Radii ($\AA$)")
fig

```



<Figure size 640x480 with 1 Axes>

As europium is commonly present as a mixture of both  $Eu^{2+}$  and  $Eu^{3+}$ , the effective partitioning of Eu will be intermediate between that of  $D_{Eu^{2+}}$ . Using a 60:40 mixture of  $Eu^{3+} : Eu^{2+}$  as an example, this effective partition coefficient can be calculated:

```

X_Eu3 = 0.6
# calculate D_Eu3 relative to D_La
D_Eu3 = D_La * strain_coefficient(
    rLa, get_ionic_radii("Eu", charge=3, coordination=8), r0=r03, E=E_3, T=Tk
)
# calculate D_Eu2 relative to D_Ca
D_Eu2 = D_Ca * strain_coefficient(
    rCa, get_ionic_radii("Eu", charge=2, coordination=8), r0=r02, E=E_2, T=Tk
)
# calculate the effective partition coefficient
D_Eu = (1 - X_Eu3) * D_Eu2 + X_Eu3 * D_Eu3
# show the effective partition coefficient relative to the 2+ and 3+ endmembers
radii, ds = (

```

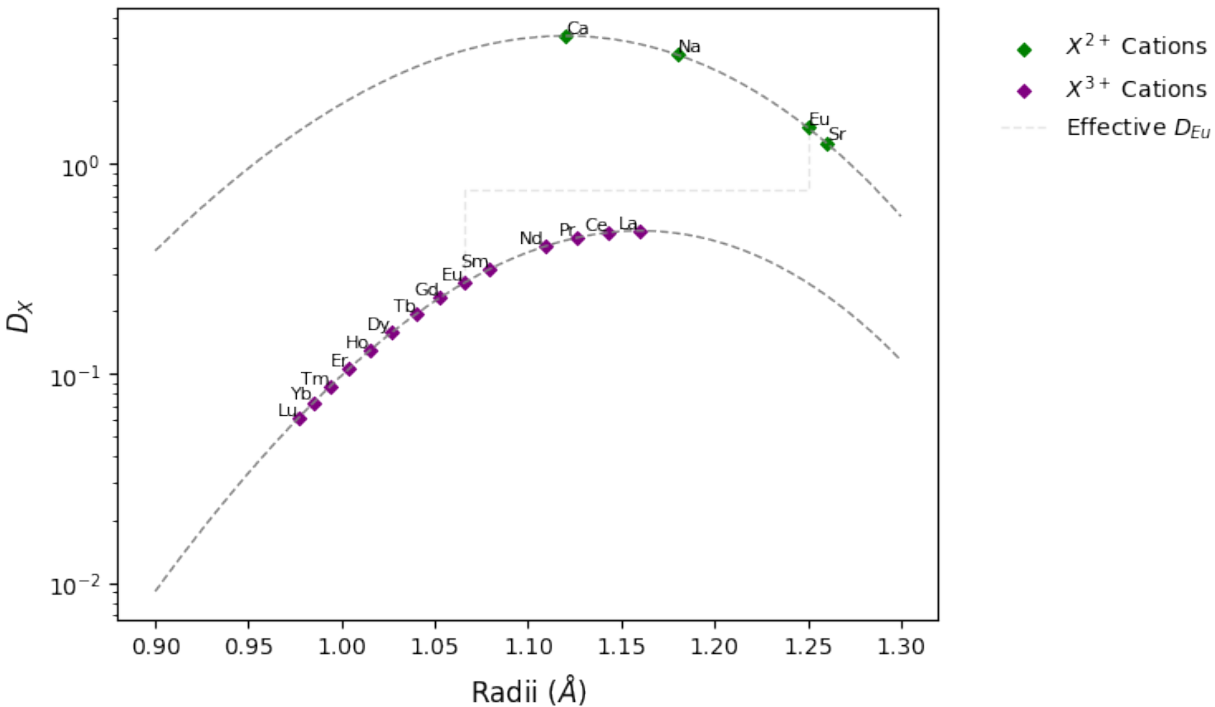
(continues on next page)

(continued from previous page)

```

[get_ionic_radii("Eu", charge=c, coordination=8) for c in [3, 3, 2, 2]],
[D_Eu3, D_Eu, D_Eu, D_Eu2],
)
ax.plot(radii, ds, ls="--", color="0.9", label="Effective $D_{Eu}$", zorder=-1)
ax.legend(bbox_to_anchor=(1.05, 1))
fig

```



&lt;Figure size 640x480 with 1 Axes&gt;

**See also:****Examples:**[Shannon Radii, REE Radii Plot](#)**Functions:**`strain_coefficient()`, `youngs_modulus_approximation()`, `get_ionic_radii()`**Total running time of the script:** (0 minutes 0.938 seconds)

## CIPW Norm

The CIPW (W. Cross, J. P. Iddings, L. V. Pirsson, and H. S. Washington) Norm was introduced as a standard procedure for the estimation of rock-forming mineral assemblages of igneous rocks from their geochemical compositions [?]. This estimation process enables the approximate classification of microcrystalline and partially crystalline rocks using a range of mineralogically-based classification systems (e.g. most IUGS classifications), and the generation of normative-mineral modifiers for geochemical classification systems.

A range of updated, modified and adjusted Norms were published in the century following the original publication of the CIPW Norm, largely culminating in Surendra Verma's 2003 paper "A revised CIPW norm" which enumerates an algorithm for the estimation of an anhydrous Standard Igneous Norm (SIN) [?]. This was subsequently updated with



(continued from previous page)

[illegible]

(continues on next page)

(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)



[illegible]

(continued from previous page)

[illegible]

(continues on next page)

(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)

[illegible]

(continues on next page)

(continued from previous page)

```

nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan 6.56159396e-04
3.28079698e-04 3.28079698e-04 1.09359899e-04 1.96847819e-03
1.09359899e-04 3.28079698e-04 1.09359899e-04 1.42167869e-03
5.46799496e-04 5.46799496e-04 6.56159396e-04 6.56159396e-04
4.37439597e-04 6.56159396e-04 1.09359899e-04 5.46799496e-04
nan nan 1.64039849e-03 2.65744555e-02
1.89192626e-02 9.95175083e-03 8.74879194e-03 6.56159396e-03
8.09263255e-03 8.08169656e-04 3.25236340e-03 3.33547693e-04
3.24798901e-03 nan nan nan
4.06818825e-02 4.25410008e-02 2.60276560e-02 2.37310981e-02
2.55902164e-02 3.45577282e-02 3.70730059e-02 3.93695637e-02
4.25410008e-02 3.95882835e-02 2.05596611e-02 1.05751023e-01
1.73882240e-02 4.36345998e-02 1.14390455e-01 1.78256636e-02
9.29559144e-03 nan 1.32106758e-02 8.00514463e-03
1.24342205e-02 1.03673185e-02 1.25862308e-02 8.69848639e-03
1.77381757e-02 1.54295882e-02 6.72016581e-03 nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan 2.26374992e-02 2.79961342e-02
2.66838154e-02 2.74493347e-02 3.08394916e-02]' has dtype incompatible with int64,
↳ please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[' nan
↳ nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan

```

(continues on next page)

[illegible]

(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)



(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)

(continues on next page)

### 3.3. Examples

(continued from previous page)

[illegible]

(continues on next page)



(continued from previous page)

[illegible]

(continues on next page)

[illegible]

(continues on next page)

(continued from previous page)

```

nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan 2.19233366e-03
7.30777885e-04 3.21542270e-03 1.46155577e-04 1.46155577e-04
1.46155577e-04 4.38466731e-04 4.38466731e-04 2.92311154e-04
2.92311154e-04 2.92311154e-04 5.84622308e-04 7.30777885e-04
4.38466731e-04 5.84622308e-04 4.38466731e-04 2.92311154e-04
nan nan nan 9.50011251e-03
1.12539794e-02 1.16924462e-02 1.11078239e-02 1.06693571e-02
1.18386017e-02 5.33467856e-04 8.97395243e-04 3.25926937e-04
7.32239441e-04 6.95700547e-04 nan nan
4.18004950e-02 2.10464031e-02 8.62317905e-03 6.02160977e-02
2.61618483e-02 5.14467631e-02 9.50011251e-03 4.51620733e-02
4.75005625e-02 4.89621183e-02 2.42618258e-02 3.81466056e-02
4.07774060e-02 1.85617583e-02 2.48464481e-03 6.32853649e-02
6.50392318e-02 nan 8.57787082e-02 2.29464256e-01
8.00201784e-02 1.98771585e-01 1.84156027e-01 2.14848698e-01
4.25020418e-02 1.20066807e-01 1.68078914e-01 nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan 1.03770460e-02 1.09616683e-02
9.20780135e-03 6.72315654e-03 5.11544520e-03]' has dtype incompatible with int64,
↳ please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[ nan nan
↳ nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan

```

(continues on next page)



[illegible]

(continued from previous page)

```

nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan 0.00014711
0.027804 0.003972 0.00044133 0.04280933 0.02648 0.00014711
0.00823822 0.00559022 0.01279867 0.00029422 0.00720844 0.00382489
0.01412267 0.01162178 0.02044844 0.00823822 nan nan
nan 0.01324 0.01368133 0.01206311 0.01015067 0.01044489
0.010592 nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
0.014564 0.01162178 0.00941511 0.01088622 0.00647289]' has dtype incompatible with
↳int64, please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
```

We can quickly check that this includes mineralogical data:

```
NORM.columns
```

```
Index(['quartz', 'zircon', 'potassium metasilicate', 'anorthite',
      'sodium metasilicate', 'acmite', 'thenardite', 'albite', 'orthoclase',
      'perovskite', 'nepheline', 'leucite', 'dicalcium silicate',
      'kaliophilite', 'apatite', 'fluroapatite', 'fluorite', 'pyrite',
      'chromite', 'ilmenite', 'calcite', 'corundum', 'rutile', 'magnetite',
      'hematite', 'forsterite', 'fayalite', 'clinoferrosilite',
      'clinoenstatite', 'ferrosilite', 'enstatite', 'wollastonite',
      'cancrinite', 'halite', 'titanite', 'diopside', 'hypersthene',
      'olivine'],
```

(continues on next page)



(continued from previous page)

[illegible]

(continues on next page)

---

(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)

(continued from previous page)

```

nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan 0.74007099
0.79991055 0.40823347 0.99728646 0.71133382 0.50295972 0.88210124
0.65468728 1.03122708 0.84816062 1.62087137 1.31836233 1.2019945
1.2019945 0.74858571 0.55251066 0.31244288 0.28642569 0.24964683
0.19548376 0.41804905 0.10832615 0.08822195 0.11293829 0.11707739
0.06941861 0.05380829 0.07166555 0.12571037 0.04032666 0.04292838
nan nan 0.12902165 0.2024611 0.12618341 0.19205422
0.1646179 0.1995046 0.09153323 0.13493464 0.13245118 0.15007192
0.1758526 0.12582863 0.11530349 0.18306646 0.60111551 0.11033657
0.13020424 nan 0.17153611 0.18566818 0.13712245 0.09372104
0.08842299 0.14463196 0.12884426 0.1018727 0.11991563 nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
0.03287628 0.02235114 0.0266085 0.02459808 0.017739 ]' has dtype incompatible with
↳ int64, please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[' nan nan
↳
↳ nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan

```

(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)



(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)

(continues on next page)

### 3.3. Examples

(continued from previous page)

[illegible]

(continues on next page)

[illegible]

(continued from previous page)

[illegible]

(continues on next page)



(continued from previous page)

```
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
0.03357944 0.03551672 0.03487096 0.03185742 0.03078116]' has dtype incompatible with_
int64, please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is_
```

(continues on next page)



[illegible]

(continued from previous page)

[illegible]

(continues on next page)

119

(continued from previous page)

[illegible]

(continues on next page)

[illegible]

(continued from previous page)

[illegible]

---

(continues on next page)

[illegible]

(continued from previous page)

```

nan nan nan nan nan 0.00014711
0.027804 0.003972 0.00044133 0.04280933 0.02648 0.00014711
0.00823822 0.00559022 0.01279867 0.00029422 0.00720844 0.00382489
0.01412267 0.01162178 0.02044844 0.00823822 nan nan
nan 0.01324 0.01368133 0.01206311 0.01015067 0.01044489
0.010592 nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
0.014564 0.01162178 0.00941511 0.01088622 0.00647289]' has dtype incompatible with
int64, please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]

```

For the purpose of establishing the congruency of our algorithm with Verma's, we'll use `adjust_all_Fe = True` and LeMaitre's correction. Notably, this won't make too much difference to the format of the output, but it will adjust the estimates of normative mineralogy depending on oxidation state.

```
NORM = CIPW_norm(
    df.pyrochem.compositional,
    adjust_all_Fe=True,
    Fe_correction="LeMaitre",
    Fe_correction_mode="volcanic",
)
```

[illegible]

---

(continues on next page)



[illegible]

(continued from previous page)

[illegible]

(continues on next page)



(continued from previous page)

```

nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan 0.74007099
0.79991055 0.40823347 0.99728646 0.71133382 0.50295972 0.88210124
0.65468728 1.03122708 0.84816062 1.62087137 1.31836233 1.2019945
1.2019945 0.74858571 0.55251066 0.31244288 0.28642569 0.24964683
0.19548376 0.41804905 0.10832615 0.08822195 0.11293829 0.11707739
0.06941861 0.05380829 0.07166555 0.12571037 0.04032666 0.04292838
nan nan 0.12902165 0.2024611 0.12618341 0.19205422
0.1646179 0.1995046 0.09153323 0.13493464 0.13245118 0.15007192
0.1758526 0.12582863 0.11530349 0.18306646 0.60111551 0.11033657
0.13020424 nan 0.17153611 0.18566818 0.13712245 0.09372104
0.08842299 0.14463196 0.12884426 0.1018727 0.11991563 nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
0.03287628 0.02235114 0.0266085 0.02459808 0.017739 ]' has dtype incompatible with
↳int64, please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/

```

(continues on next page)



(continued from previous page)

```

nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan 0.18679143
0.2200633 0.59398112 0.0849661 0.07033987 0.06096122 0.19449532
0.09847582 0.08909717 0.14715547 0.08373794 0.08284473 0.06833016
0.08898552 0.05180587 0.05616024 0.59911704 nan 0.00223301
0.01462623 0.55155389 0.10003892 0.04588839 0.03014566 0.05113597
0.03260197 nan 0.00840952 0.00204209 0.00281025 0.00136325
nan nan 0.25590314 0.4677043 0.1992963 0.3140731
0.28783521 0.39033045 0.11466515 0.31630611 0.28817016 0.30268474
0.21515068 0.19315551 0.24429148 0.27131092 0.37737898 0.39200521
0.36487411 nan 0.20443222 0.23820652 0.17367249 0.11533506
0.09330528 0.10178737 0.16792248 0.11416272 0.076944 nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
0.07167968 0.03673304 0.07648065 0.07357774 0.05191752]' has dtype incompatible with
↳ int64, please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[' nan
↳ nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan
nan nan nan nan nan

```

(continues on next page)

[illegible]

(continued from previous page)

[illegible]

(continues on next page)



(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)

[illegible]

(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)

[illegible]

(continued from previous page)

[illegible]

(continues on next page)

[illegible]

(continues on next page)

(continued from previous page)

```

nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan 1.35077173e-04
7.83447602e-03 6.75385863e-03 1.21569455e-03 1.35077173e-04
5.94339560e-03 1.35077173e-04 1.75600324e-03 2.43138911e-03
1.35077173e-04 1.35077173e-04 1.35077173e-04 4.14686920e-02
1.35077173e-04 6.20004223e-02 3.03923639e-02 4.05231518e-04
nan nan nan 1.08061738e-02
2.16123476e-02 1.91809585e-02 1.74249553e-02 1.79652640e-02
1.86406498e-02 5.69215206e-01 3.51200649e-03 1.48584890e-03
4.32246953e-03 2.97169780e-03 nan nan
6.45668885e-02 5.52465636e-02 2.99871323e-02 7.13207472e-02
4.16037692e-02 6.76736635e-02 3.71462225e-02 4.21440779e-02
4.18739235e-02 4.25493094e-02 3.90373029e-02 3.76865312e-02
3.05274410e-02 4.86277822e-02 4.53859300e-02 1.33456247e-01
1.21164224e-01 nan 8.32332030e-02 1.94011343e-02
6.37010439e-02 3.55279980e-02 5.18210065e-02 9.48836092e-02
1.01915727e-01 1.63848610e-01 9.63843166e-02 nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan 4.52508528e-02 4.28194637e-02
3.93074572e-02 3.84969942e-02 2.93117465e-02]' has dtype incompatible with int64,
↳ please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is

```

(continues on next page)





(continued from previous page)

[illegible]

(continues on next page)

(continued from previous page)

```

nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan 2.19233366e-03
7.30777885e-04 3.21542270e-03 1.46155577e-04 1.46155577e-04
1.46155577e-04 4.38466731e-04 4.38466731e-04 2.92311154e-04
2.92311154e-04 2.92311154e-04 5.84622308e-04 7.30777885e-04
4.38466731e-04 5.84622308e-04 4.38466731e-04 2.92311154e-04
nan nan nan 9.50011251e-03
1.12539794e-02 1.16924462e-02 1.11078239e-02 1.06693571e-02
1.18386017e-02 5.33467856e-04 8.97395243e-04 3.25926937e-04
7.32239441e-04 6.95700547e-04 nan nan
4.18004950e-02 2.10464031e-02 8.62317905e-03 6.02160977e-02
2.61618483e-02 5.14467631e-02 9.50011251e-03 4.51620733e-02
4.75005625e-02 4.89621183e-02 2.42618258e-02 3.81466056e-02
4.07774060e-02 1.85617583e-02 2.48464481e-03 6.32853649e-02
6.50392318e-02 nan 8.57787082e-02 2.29464256e-01
8.00201784e-02 1.98771585e-01 1.84156027e-01 2.14848698e-01
4.25020418e-02 1.20066807e-01 1.68078914e-01 nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan nan nan
nan nan 1.03770460e-02 1.09616683e-02
9.20780135e-03 6.72315654e-03 5.11544520e-03]' has dtype incompatible with int64,
↳ please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ geochem/transform.py:351: FutureWarning: Setting an item of incompatible dtype is
↳ deprecated and will raise in a future error of pandas. Value '[' nan nan
↳
nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan

```

(continues on next page)

(continued from previous page)

[illegible]

(continues on next page)

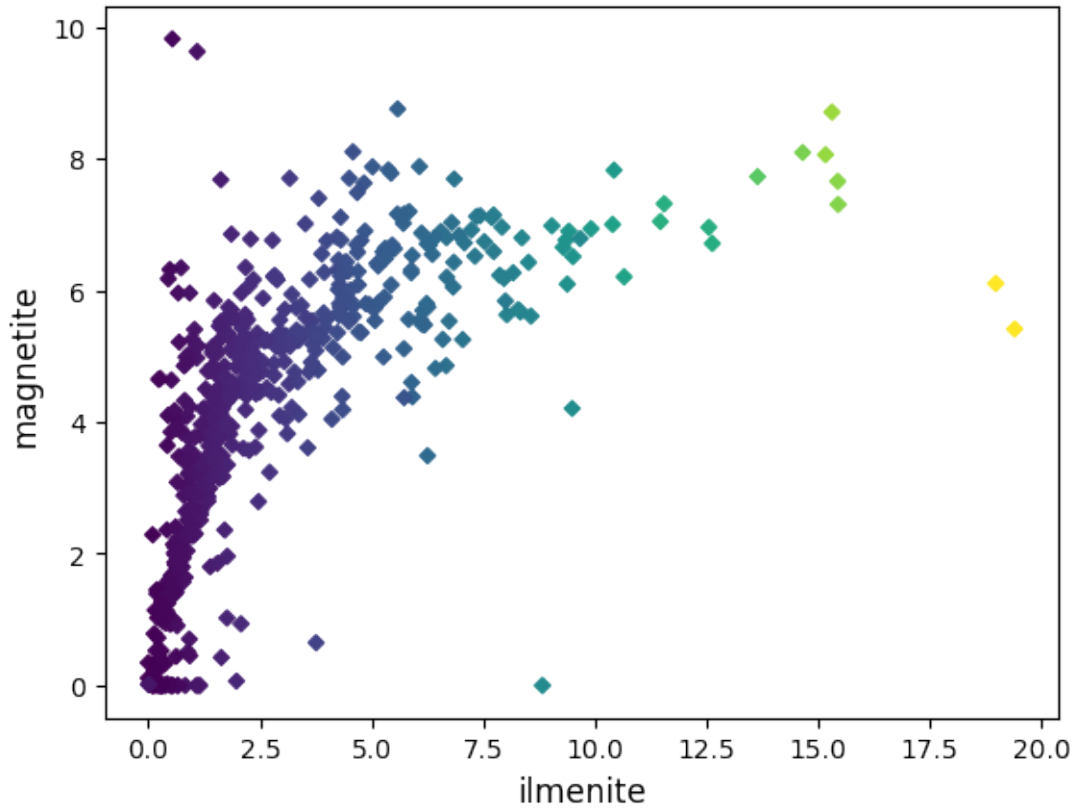
(continued from previous page)

```

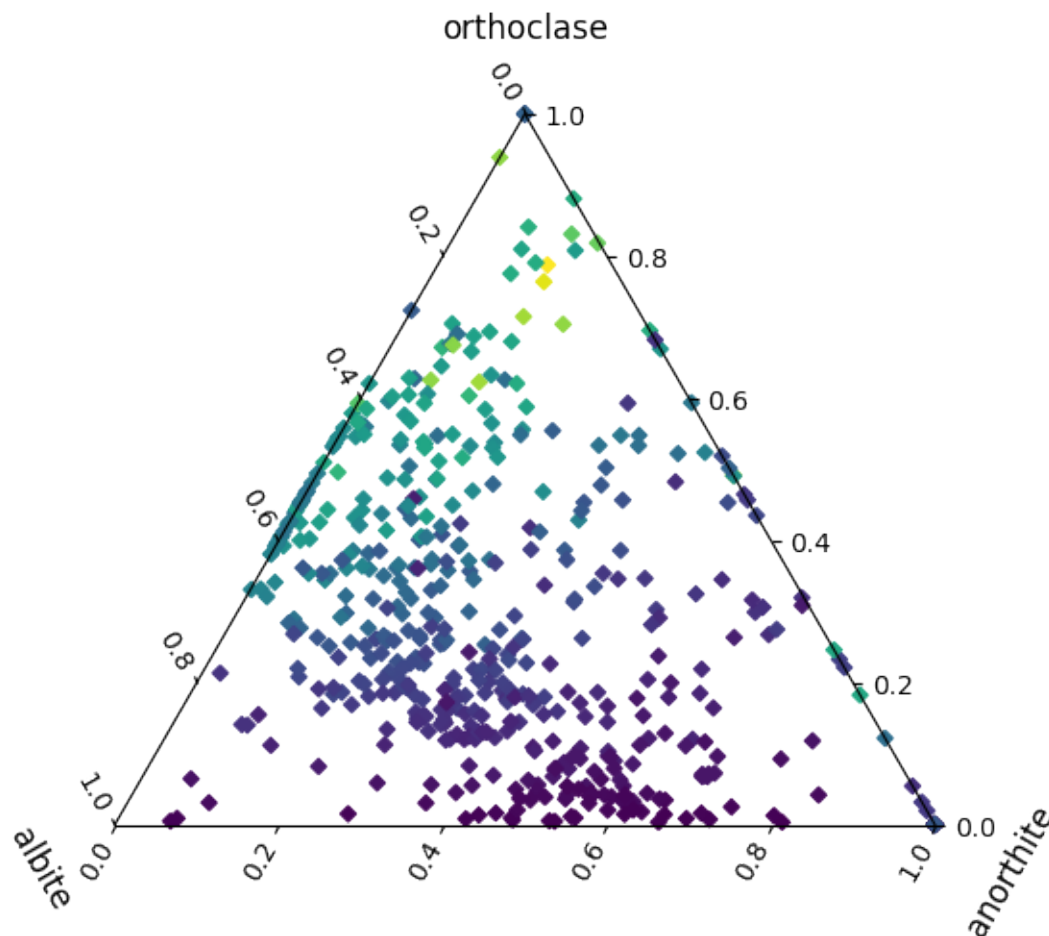
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan 0.00014711
0.027804 0.003972 0.00044133 0.04280933 0.02648 0.00014711
0.00823822 0.00559022 0.01279867 0.00029422 0.00720844 0.00382489
0.01412267 0.01162178 0.02044844 0.00823822 nan nan
nan 0.01324 0.01368133 0.01206311 0.01015067 0.01044489
0.010592 nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
nan nan nan nan nan nan
0.014564 0.01162178 0.00941511 0.01088622 0.00647289]' has dtype incompatible with
↳int64, please explicitly cast to a compatible dtype first.
_df.loc[:, targetnames] = subsum.values[:, np.newaxis] @ coeff[np.newaxis, :]
```

Now we have the normative mineralogical outputs, we can have a look to see how these compare to some relevant geochemical inputs:

```
ax = NORM[["ilmenite", "magnetite"]].pyroplot.scatter(clip_on=False, c=df["TiO2"])
plt.show()
```



```
ax = NORM[["orthoclase", "albite", "anorthite"]].pyroplot.scatter(  
    clip_on=False, c=df["K20"]  
)  
plt.show()
```



### Coherency with SINCLAS / IgRoCS

Given we're reproducing an existing algorithm, it's prudent to check how closely the results match for a specific dataset to check whether there might be any numerical or computational errors. Below we go through this exercise for the test dataset we loaded above (which already includes the output of SINCLAS), comparing the original software to the pyrolite implementation.

**Note:** Currently there are inconsistent results for a small number of samples (deviations colormapped, and inconsistent results shown in red below), likely related to the handling of iron components and their conversion.

The output of SINCLAS has slightly different column naming than that of pyrolite, which provides full mineral names in the output dataframe columns. For this reason, we'll need to translate our NORM output columns to the SINCLAS column names. For this we can use the dictionary of minerals used in the CIPW Norm (NORM\_MINERALS)

```
from pyrolite.mineral.normative import NORM_MINERALS

translation = {
    d["name"]: (d.get("SINCLAS_abbrev", None) or k.upper()) + "_NORM"
    for k, d in NORM_MINERALS.items()
    if (d.get("SINCLAS_abbrev", None) or k.upper()) + "_NORM" in df.columns
}
```

(continues on next page)

(continued from previous page)

```
}
translation
```

```
{'quartz': 'Q_NORM', 'zircon': 'ZIR_NORM', 'potassium metasilicate': 'KS_NORM',
→ 'anorthite': 'AN_NORM', 'sodium metasilicate': 'NS_NORM', 'acmite': 'AC_NORM',
→ 'thenardite': 'TH_NORM', 'albite': 'AB_NORM', 'orthoclase': 'OR_NORM', 'perovskite':
→ 'PER_NORM', 'nepheline': 'NE_NORM', 'leucite': 'LC_NORM', 'dicalcium silicate': 'CS_
→ NORM', 'kaliophilite': 'KP_NORM', 'apatite': 'AP_NORM', 'fluorite': 'FR_NORM', 'pyrite
→ ': 'PYR_NORM', 'chromite': 'CHR_NORM', 'ilmenite': 'IL_NORM', 'calcite': 'CC_NORM',
→ 'corundum': 'C_NORM', 'rutile': 'RU_NORM', 'magnetite': 'MT_NORM', 'forsterite': 'FO_
→ NORM', 'fayalite': 'FA_NORM', 'clinoferrrosilite': 'DIF_NORM', 'clinoenstatite': 'DIM_
→ NORM', 'ferrosilite': 'HYF_NORM', 'enstatite': 'HYM_NORM', 'wollastonite': 'WO_NORM',
→ 'cancrinite': 'CAN_NORM', 'halite': 'HL_NORM', 'titanite': 'SPH_NORM', 'diopside': 'DI_
→ NORM', 'hypersthene': 'HY_NORM', 'olivine': 'OL_NORM'}
```

First we'll collect the minerals which appear in both dataframes, and then iterate through these to check how close the implementations are.

```
minerals = {
    k: v for (k, v) in translation.items() if (df[v] > 0).sum() and (NORM[k] > 0).sum()
}
```

To compare SINCLAS and the pyrolite NORM outputs, we'll construct a grid of plots which compare the respective mineralogical norms relative to a 1:1 line, and highlight discrepancies. As we'll do it twice below (once for samples labelled as volcanic, and once for everything else), we may as well make a function of it.

After that, let's take a look at the volcanic samples in isolation, which are the the key samples for which the NORM should be applied:

```
from pyrolite.plot.color import process_color

def compare_NORMs(SINCLAS_outputs, NORM_outputs, name=""):
    """
    Create a grid of axes comparing the outputs of SINCLAS and `pyrolite`'s NORM,
    after translating the column names to the appropriate form.
    """
    ncols = 4
    nrows = len(minerals.keys()) // ncols + (1 if len(minerals.keys()) % ncols else 0)

    fig, ax = plt.subplots(nrows, ncols, figsize=(ncols * 2.5, nrows * 2))
    fig.suptitle(
        " - ".join(
            ["Comparing pyrolite's CIPW Norm to SINCLAS/IgRoCS"] + [name]
            if name
            else []
        ),
        fontsize=16,
        y=1.01,
    )
    ax = ax.flat
    for ix, (b, a) in enumerate(minerals.items()):
        ax[ix].set_title("\n".join(b.split()), y=0.9, va="top")
```

(continues on next page)



(continued from previous page)

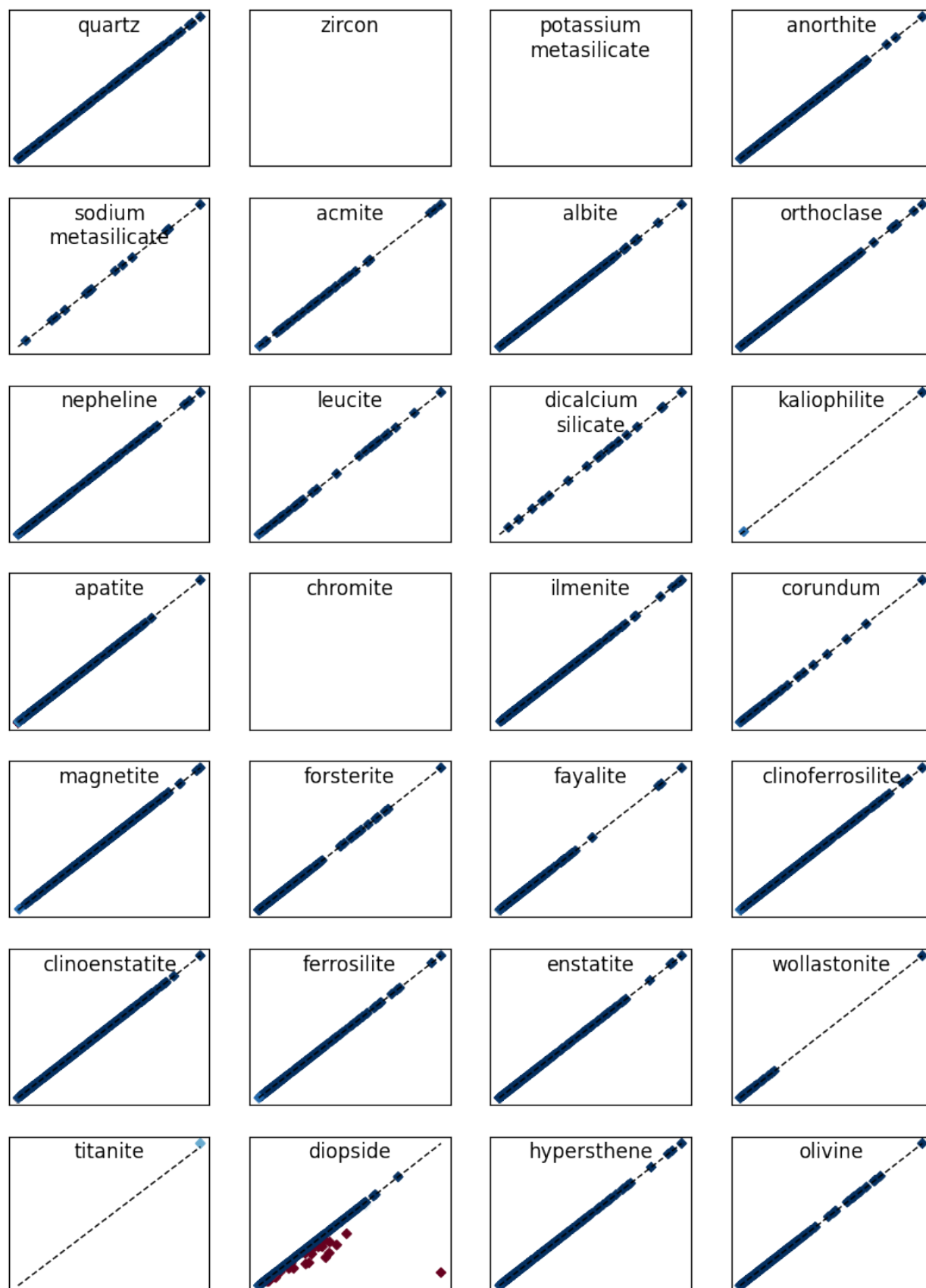
```

    if a in SINCLAS_outputs.columns and b in NORM_outputs.columns:
        # colour by deviation from unity
        c = process_color(
            np.abs((SINCLAS_outputs[a] / NORM_outputs[b]) - 1),
            cmap="RdBu_r",
            norm=plt.Normalize(vmin=0, vmax=0.1),
        )["c"]
        ax[ix].scatter(SINCLAS_outputs[a], NORM_outputs[b], c=c)
    # add a 1:1 line
    ax[ix].plot(
        [0, SINCLAS_outputs[a].max()],
        [0, SINCLAS_outputs[a].max()],
        color="k",
        ls="--",
    )

    for a in ax:
        a.set(xticks=[], yticks=[]) # turn off the ticks
        if not a.collections: # turn off the axis for empty axes
            a.axis("off")
    return fig, ax

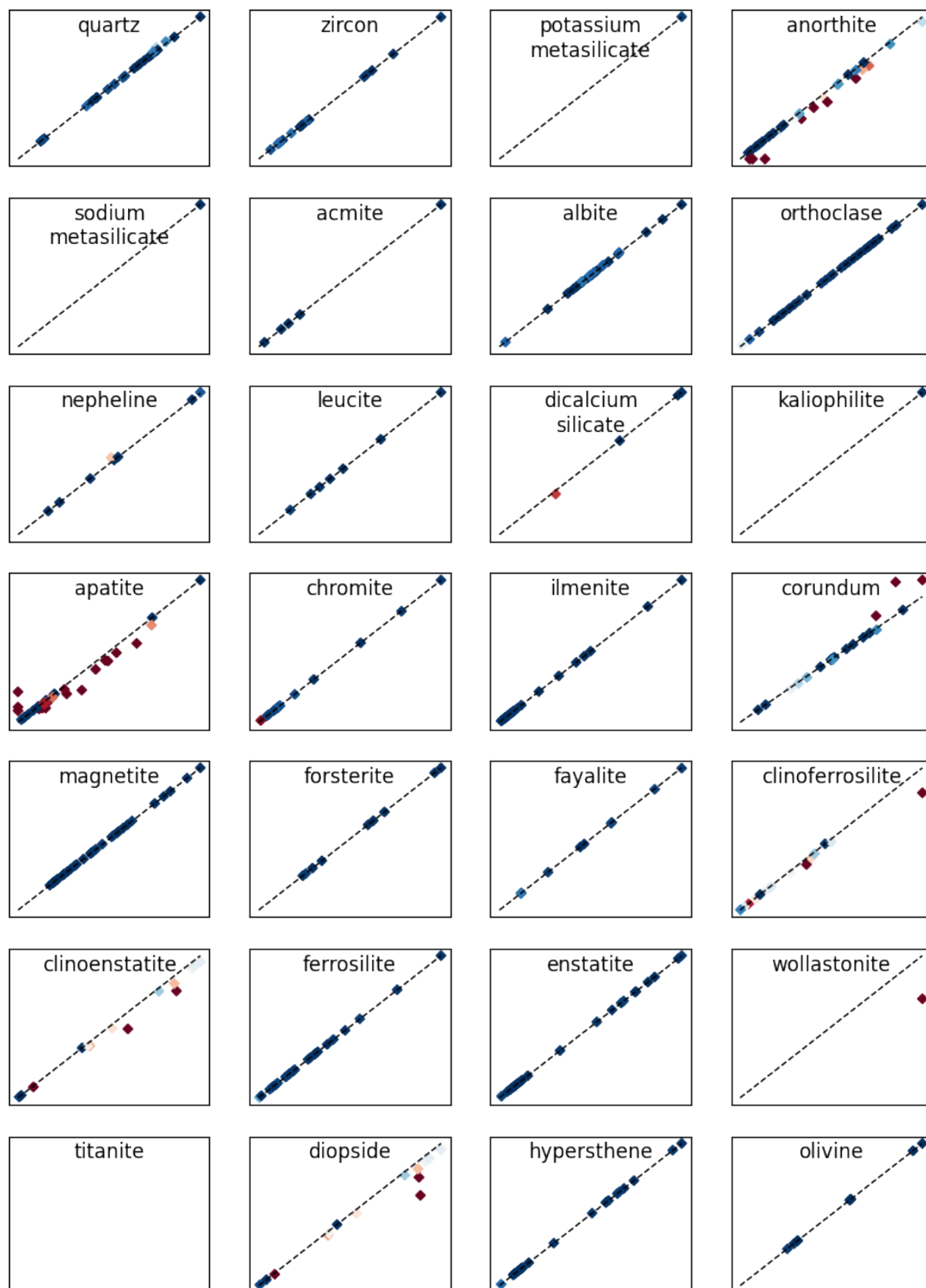
volcanic_filter = df.loc[:, "ROCK_TYPE"].str.lower().str.startswith("volc")
fig, ax = compare_NORMs(df.loc[volcanic_filter, :], NORM.loc[volcanic_filter])

```



And everything else:

```
fig, ax = compare_NORMs(df.loc[~volcanic_filter, :], NORM.loc[~volcanic_filter])  
plt.show()
```



These normative mineralogical components could be input into mineralogical classifiers, as mentioned above. For example, the IUGS QAP classifier:

```
from pyrolite.util.classification import QAP

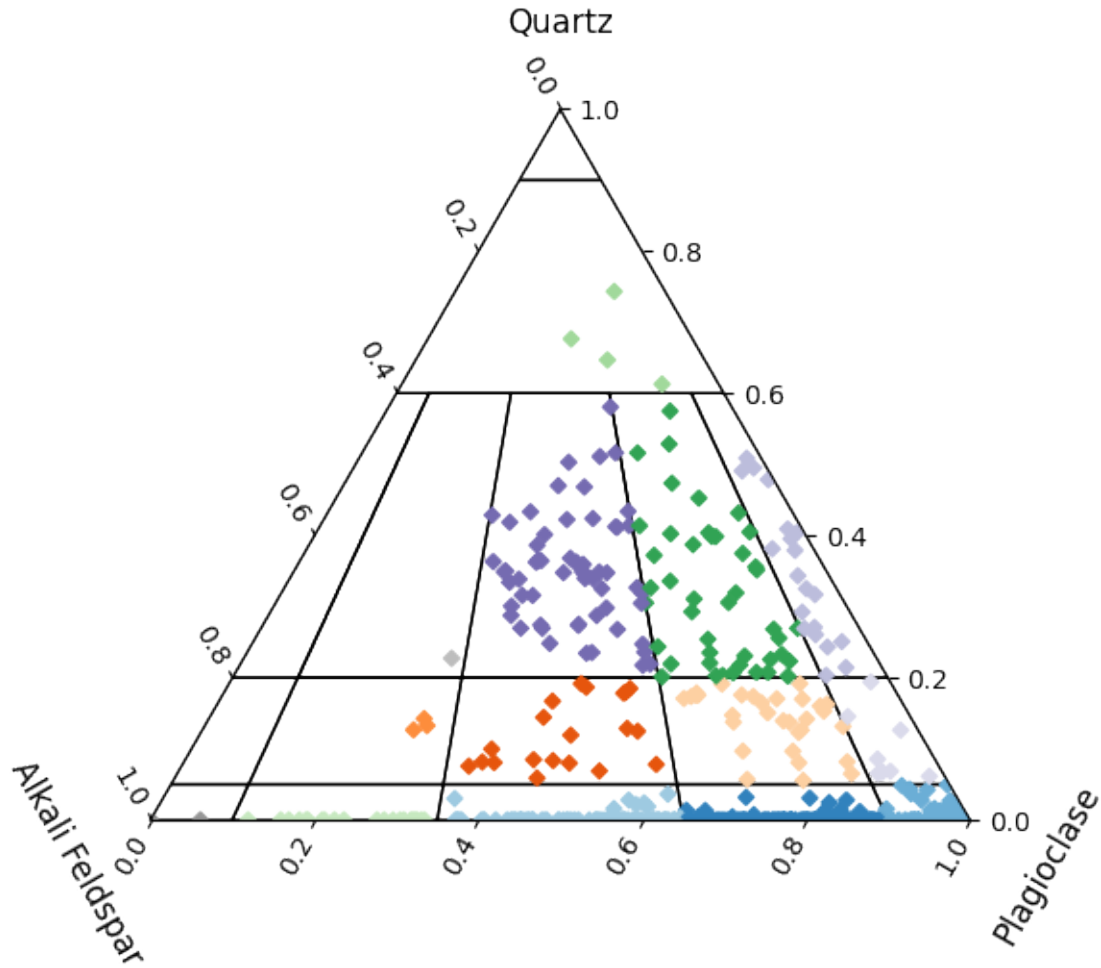
clf = QAP() # build a QAP classifier

qap_data = NORM.loc[:, ["quartz", "orthoclase"]] #
qap_data["plagioclase"] = NORM.loc[:, ["albite", "anorthite"]].sum(axis=1)
# predict which lithological class each mineralogical composiiton belongs in
# we add a small value to zeros here to ensure points fit in polygons
predicted_classes = clf.predict(qap_data.replace(0, 10e-6).values)
predicted_classes.head()
```

```
0      monzodiorite-monzogabbro
1      diorite-gabbro-anorthosite
2      monzodiorite-monzogabbro
3      monzodiorite-monzogabbro
4      monzodiorite-monzogabbro
dtype: object
```

We can use these predicted classes as a color index also, within the QAP diagram or elsewhere:

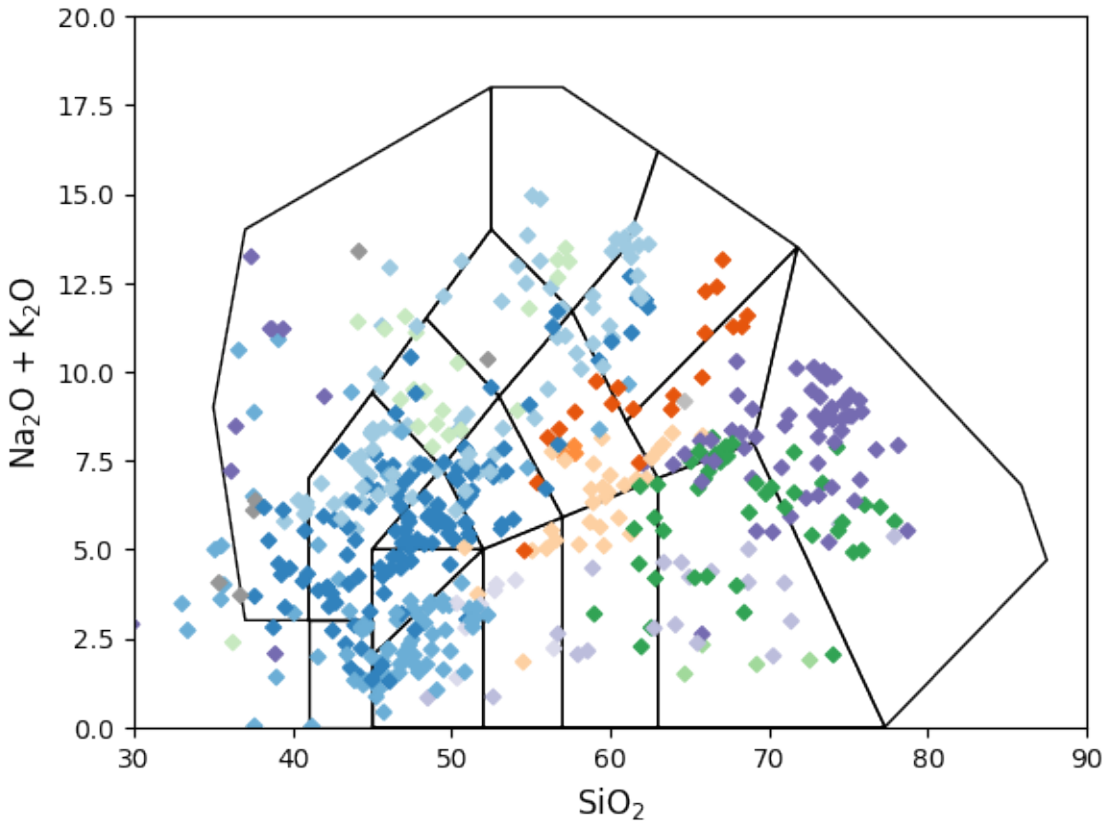
```
ax = clf.add_to_axes()
qap_data.pyroplot.scatter(ax=ax, c=predicted_classes, axlabels=False, cmap="tab20c")
plt.show()
```



We could also compare how these mineralogical distinctions map into chemical ones like the TAS diagram:

```
from pyrolite.plot.templates import TAS

ax = TAS()
components = df.loc[:, ["SiO2"]]
components["alkali"] = df.loc[:, ["Na2O", "K2O"]].sum(axis=1)
# add the predictions from normative mineralogy to the TAS diagram
components.pyroplot.scatter(ax=ax, c=predicted_classes, cmap="tab20c", axlabels=False)
plt.show()
```



## References

**Total running time of the script:** (0 minutes 11.312 seconds)

## lambdas: Parameterising REE Profiles

Orthogonal polynomial decomposition can be used for dimensional reduction of smooth function over an independent variable, producing an array of independent values representing the relative weights for each order of component polynomial. This is an effective method to parameterise and compare the nature of smooth profiles.

In geochemistry, the most applicable use case is for reduction Rare Earth Element (REE) profiles. The REE are a collection of elements with broadly similar physicochemical properties (the lanthanides), which vary with ionic radii. Given their similar behaviour and typically smooth function of normalised abundance vs. ionic radii, the REE profiles and their shapes can be effectively parameterised and dimensionally reduced (14 elements summarised by 3-4 shape parameters).

---

**Note:** A publication discussing this implementation of *lambdas* together with fitting tetrads for REE patterns has been published in [Mathematical Geosciences](#)!

---

Here we generate some example data, reduce these to lambda values, and visualise the results.

```
import matplotlib.pyplot as plt
import numpy as np
```

(continues on next page)

(continued from previous page)

```
import pandas as pd

import pyrolite.plot

np.random.seed(82)
```

First we'll generate some example synthetic data based around Depleted MORB Mantle:

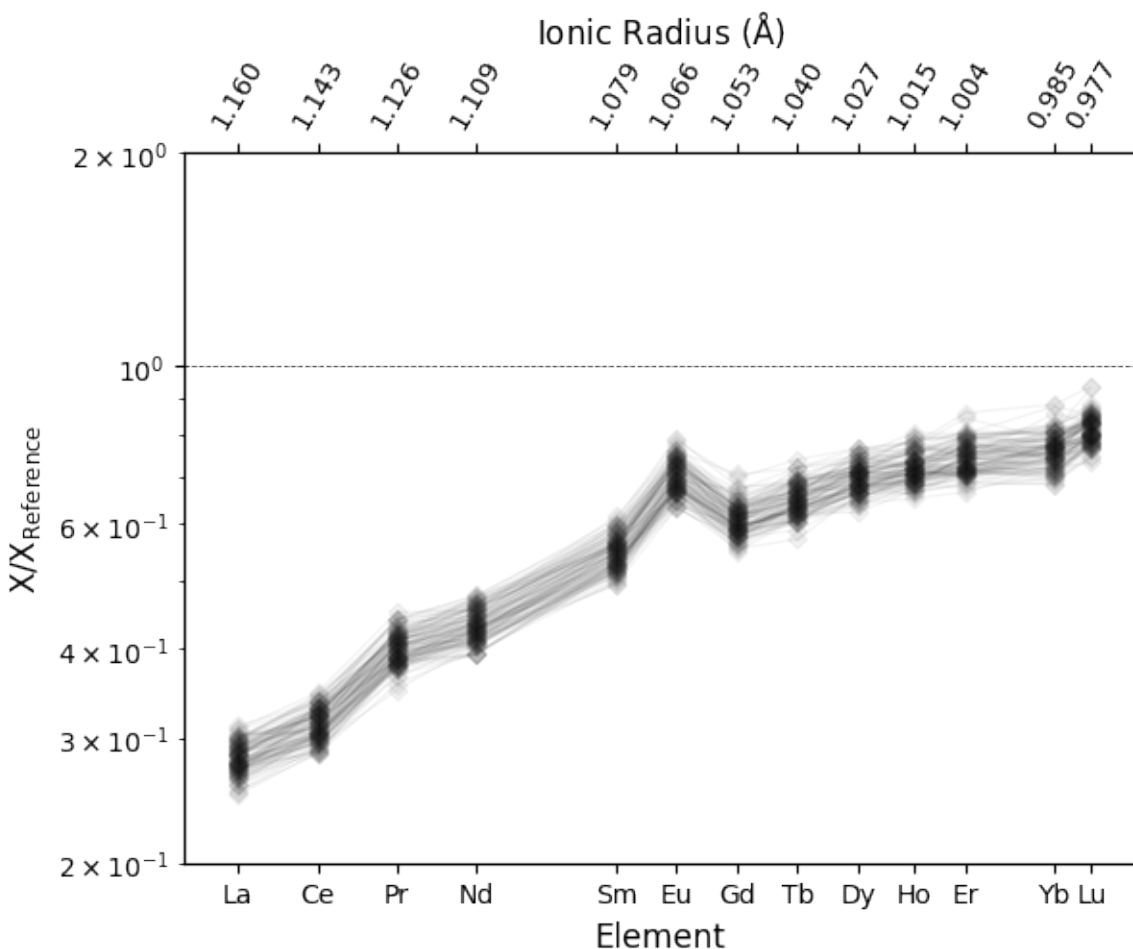
```
from pyrolite.util.synthetic import example_spider_data

df = example_spider_data(
    noise_level=0.05,
    size=100,
    start="DMM_WH2005",
    norm_to="Chondrite_PON",
    offsets={"Eu": 0.2},
).pyrochem.REE.pyrochem.denormalize_from("Chondrite_PON")
df.head(2)
```

Let's have a quick look at what this REE data looks like normalized to Primitive Mantle:

```
df.pyrochem.normalize_to("PM_PON").pyroplot.REE(alpha=0.05, c="k", unity_line=True)
plt.show()
```





From this REE data we can fit a series of orthogonal polynomials, and subsequently used the regression coefficients ('lambdas') as a parameterisation of the REE pattern/profile:

```
ls = df.pyrochem.lambda_lnREE(degree=4)
ls.head(2)
```

So what's actually happening here? To get some idea of what these coefficients correspond to, we can pull this process apart and visualise our REE profiles as the sum of the series of orthogonal polynomial components of increasing order. As lambdas represent the coefficients for the regression of log-transformed normalised data, to compare the polynomial components and our REE profile we'll first need to normalize it to the appropriate composition (here "ChondriteREE\_ON") before taking the logarithm.

With our data, we've then fit a function of ionic radius with the form  $f(r) = \lambda_0 + \lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3 \dots$  where the polynomial components of increasing order are  $f_1 = (r - \beta_0)$ ,  $f_2 = (r - \gamma_0)(r - \gamma_1)$ ,  $f_3 = (r - \delta_0)(r - \delta_1)(r - \delta_2)$  and so on. The parameters  $\beta$ ,  $\gamma$ ,  $\delta$  are pre-computed such that the polynomial components are indeed independent. Here we can visualise how these polynomial components are summed to produce the regressed profile, using the last REE profile we generated above as an example:

```
from pyrolite.util.lambdas.plot import plot_lambdas_components

ax = (
    df.pyrochem.normalize_to("ChondriteREE_ON")
    .iloc[-1, :]
```

(continues on next page)

(continued from previous page)

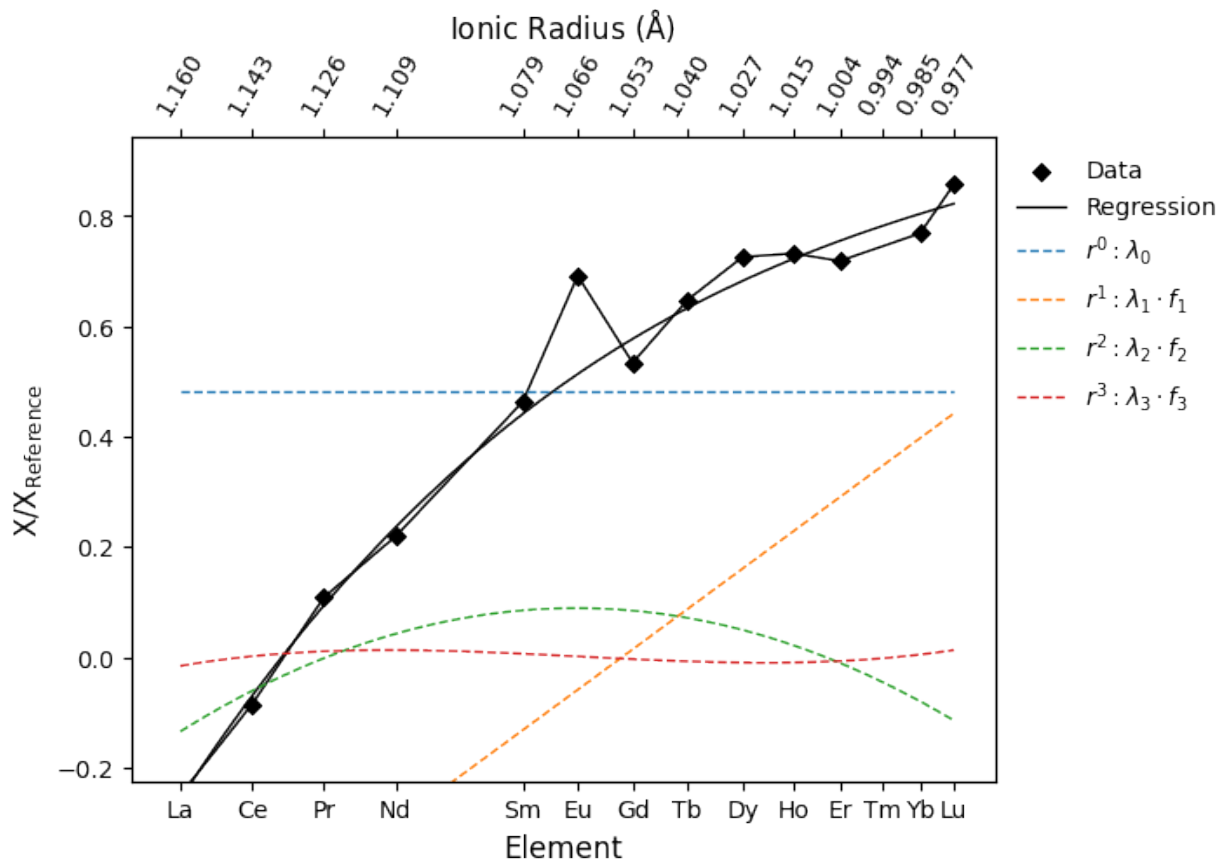
```

        .apply(np.log)
        .pyroplot.REE(color="k", label="Data", logy=False)
    )

    plot_lambdas_components(ls.iloc[-1, :], ax=ax)

    ax.legend()
    plt.show()

```



```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
util/plot/axes.py:219: RuntimeWarning: More than 20 figures have been opened. Figures
created through the pyplot interface (`matplotlib.pyplot.figure`) are retained until
explicitly closed and may consume too much memory. (To control this warning, see the
rcParam `figure.max_open_warning`). Consider using `matplotlib.pyplot.close()`.
fig, ax = plt.subplots(1, **subkwargs(kwargs, plt.subplots, plt.figure))

```

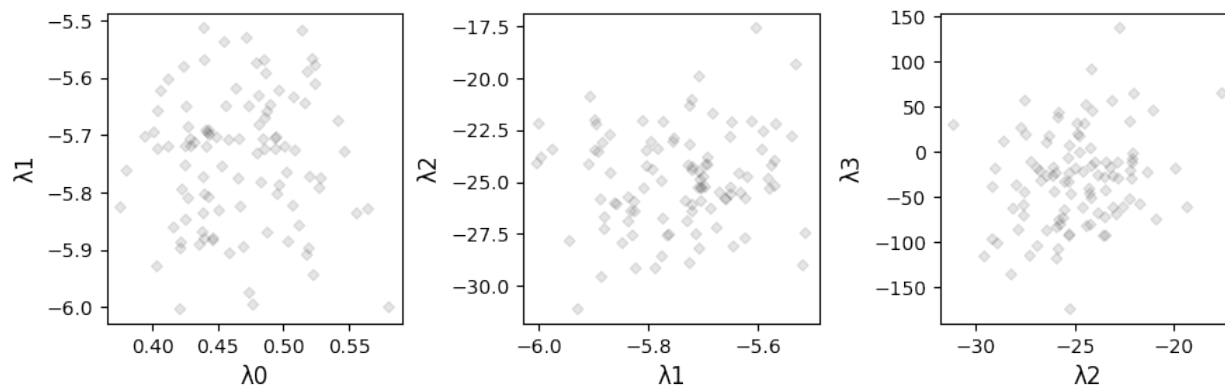
Now that we've gone through a brief introduction to how the lambdas are generated, let's quickly check what the coefficient values themselves look like:

```

fig, ax = plt.subplots(1, 3, figsize=(9, 3))
for ix in range(ls.columns.size - 1):
    ls[ls.columns[ix : ix + 2]].pyroplot.scatter(ax=ax[ix], alpha=0.1, c="k")

plt.tight_layout()

```



But what do these parameters correspond to? From the deconstructed orthogonal polynomial above, we can see that  $\lambda_0$  parameterises relative enrichment (this is the mean value of the logarithm of Chondrite-normalised REE abundances),  $\lambda_1$  parameterises a linear slope (here, LREE enrichment), and higher order terms describe curvature of the REE pattern. Through this parameterisation, the REE profile can be effectively described and directly linked to geochemical processes. While the amount of data we need to describe the patterns is lessened, the values themselves are more meaningful and readily used to describe the profiles and their physical significance.

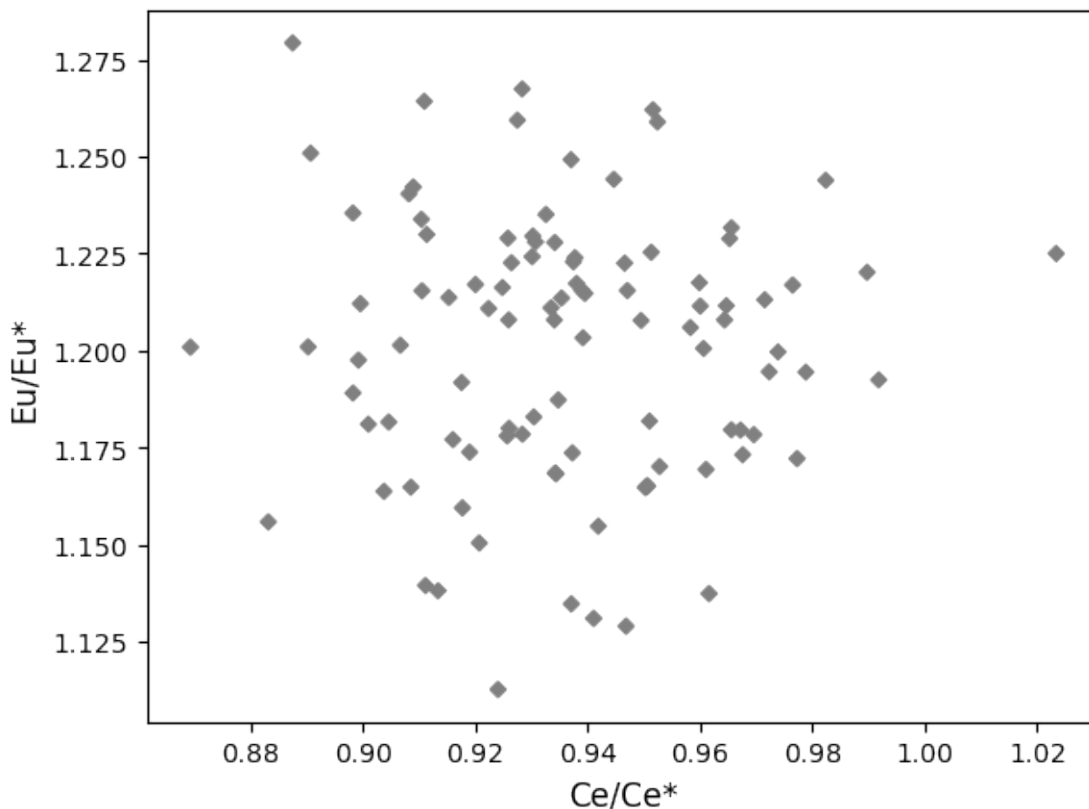
The visualisation of  $\lambda_1$ - $\lambda_2$  can be particularly useful where you're trying to compare REE profiles.

We've used a synthetic dataset here which is by design approximately normally distributed, so the values themselves here are not particularly revealing, but they do illustrate the expected magnitudes of values for each of the parameters.

## Dealing With Anomalies

Note that we've not used Eu in this regression - Eu anomalies are a deviation from the 'smooth profile' we need to use this method. Consider this if your data might also exhibit significant Ce anomalies, you might need to exclude this data. For convenience there is also functionality to calculate anomalies derived from the orthogonal polynomial fit itself (rather than linear interpolation methods). Below we use the `anomalies` keyword argument to also calculate the  $\frac{Ce}{Ce^*}$  and  $\frac{Eu}{Eu^*}$  anomalies (note that these are excluded from the fit):

```
ls_anomalies = df.pyrochem.lambda_lnREE(anomalies=["Ce", "Eu"])
ax = ls_anomalies.iloc[:, -2:].pyroplot.scatter(color="0.5")
plt.show()
```



### Coefficient Uncertainties and Fit Quality

In order to determine the relative significance of the parameterisation and ‘goodness of fit’, the functions are able to estimate uncertainties on the returned coefficients (lambdas and taus) and will also return the chi-square value ( $\chi^2$ ; equivalent to the MSWD) where requested. This will be appended to the end of the dataframe. Note that if you do not supply an estimate of observed value uncertainties a default of 1% of the log-mean will be used.

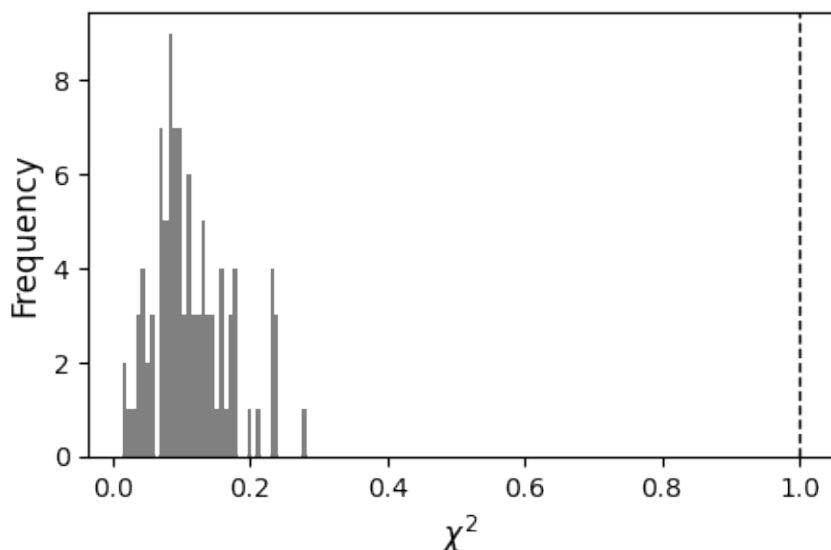
To append the reduced chi-square for each row, the keyword argument `add_X2=True` can be used; here we’ve estimated 10% uncertainty on the REE:

```
ls = df.pyrochem.lambda_lnREE(add_X2=True, sigmas=0.1, anomalies=["Eu", "Ce"])
ls.columns
```

```
Index(['0', '1', '2', '3', 'X2', 'Eu/Eu*', 'Ce/Ce*'], dtype='object')
```

We can have a quick look at the  $\chi^2$  values look like for the synthetic dataset, given the assumed 10% uncertainties. While the fit appears reasonable for a good fraction of the dataset (~2 and below), for some rows it is notably worse:

```
fig, ax = plt.subplots(1, figsize=(5, 3))
ax = ls["X2"].plot.hist(ax=ax, bins=40, color="0.5")
ax.set(xlabel="$\\chi^2$")
ax.axvline(1, color="k", ls="--")
plt.show()
```



We can also examine the estimated uncertainties on the coefficients from the fit by adding the keyword argument `add_uncertainties=True` (note: these do not explicitly propagate observation uncertainties):

```
ls = df.pyrochem.lambda_lnREE(add_uncertainties=True)
ls.columns
```

```
Index(['0', '1', '2', '3', '0_', '1_', '2_', '3_'], dtype='object')
```

Notably, on the scale of natural dataset variation, these uncertainties may end up being smaller than symbol sizes. If your dataset happened to have a great deal more noise, you may happen to see them - for demonstration purposes we can generate a noisier dataset and have a quick look at what these uncertainties *could* look like:

```
ls = (df + 3 * np.exp(np.random.randn(*df.shape))).pyrochem.lambda_lnREE(
    add_uncertainties=True
)
```

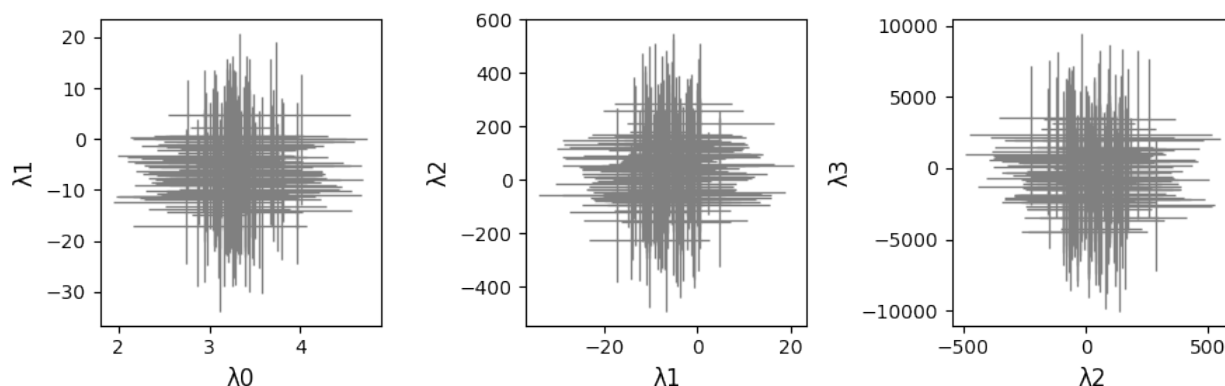
With this ‘noisy’ dataset, we can see some of the errorbars:

```
fig, ax = plt.subplots(1, 3, figsize=(9, 3))
ax = ax.flat
dc = ls.columns.size // 2
for ix, a in enumerate(ls.columns[:3]):
    i0, i1 = ix, ix + 1
    ax[ix].set(xlabel=ls.columns[i0], ylabel=ls.columns[i1])
    ax[ix].errorbar(
        ls.iloc[:, i0],
        ls.iloc[:, i1],
        xerr=ls.iloc[:, i0 + dc] * 2,
        yerr=ls.iloc[:, i1 + dc] * 2,
        ls="none",
        ecolor="0.5",
        markersize=1,
        color="k",
    )
```

(continues on next page)

(continued from previous page)

```
plt.tight_layout()
plt.show()
```



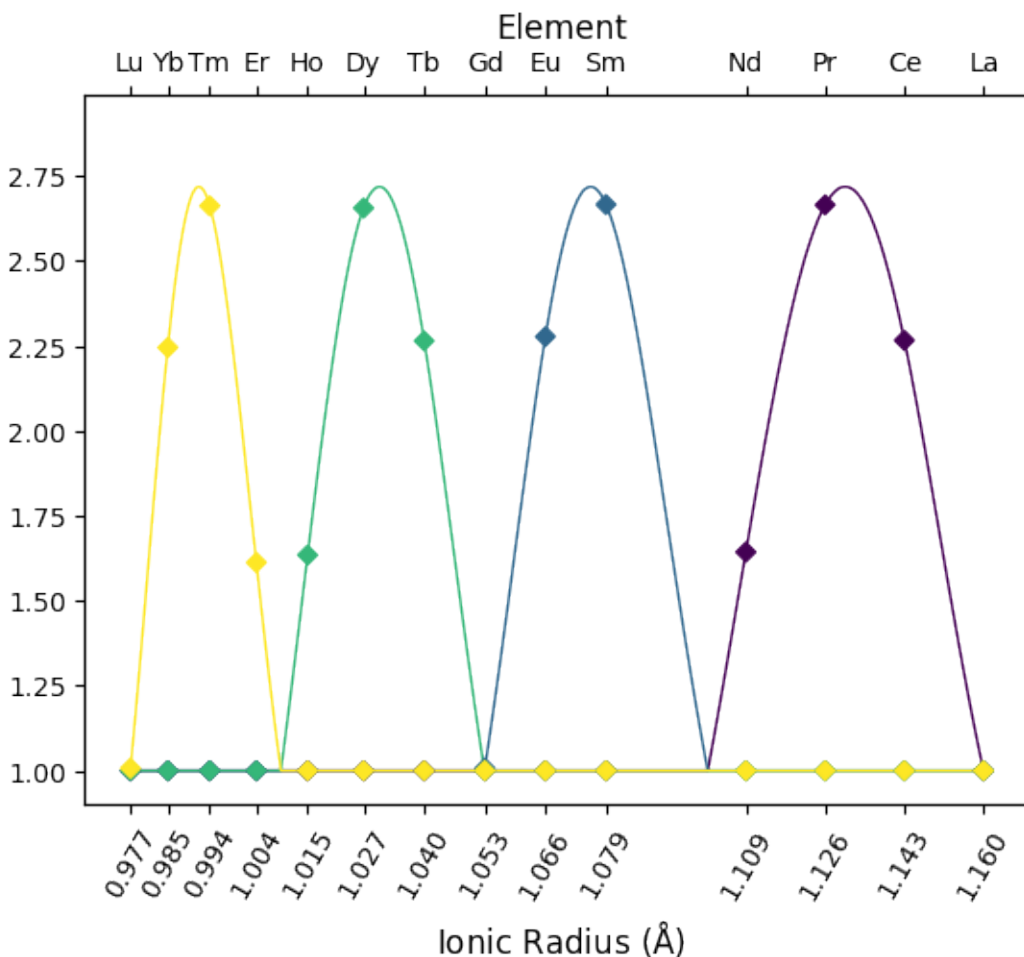
## Fitting Tetrads

In addition to fitting orothogonal polynomial functions, the ability to fit tetrad functions has also recently been added. This supplements the  $\lambda$  coefficients with  $\tau$  coefficients which describe subtle electronic configuration effects affecting sequential subsets of the REE. Below we plot four profiles - each describing only a single tetrad - to illustrate the shape of these function components. Note that these are functions of  $z$ , and are here transformed to plot against radii.

```
from pyrolite.util.lambdas.plot import plot_profiles

# let's first create some synthetic pattern parameters
# we want lambdas to be zero, and each of the tetrads to be shown in only one pattern
lambdas = np.zeros((4, 5))
tetrads = np.eye(4)
# putting it together to generate four sets of combined parameters
fit_parameters = np.hstack([lambdas, tetrads])

ax = plot_profiles(
    fit_parameters,
    tetrads=True,
    color=np.arange(4),
)
plt.show()
```



In order to also fit these function components, you can pass the keyword argument `fit_tetrads=True` to `lambda_lnREE()` and related functions:

```
lts = df.pyrochem.lambda_lnREE(degree=4, fit_tetrads=True)
```

We can see that the four extra  $\tau$  Parameters have been appended to the right of the lambdas within the output:

```
lts.head(2)
from pyrolite.geochem.ind import REE
```

Below we'll look at some of the potential issues of fitting lambdas and tetrads together - by examining the effects of i) fitting tetrads where there are none and ii) not fitting tetrads where they do indeed exist using some synthetic datasets.

```
from pyrolite.util.synthetic import example_patterns_from_parameters

ls = np.array(
    [
        [2, 5, -30, 100, -600, 0, 0, 0, 0], # lambda-only
        [3, 15, 30, 300, 1500, 0, 0, 0, 0], # lambda-only
        [1, 5, -50, 0, -1000, -0.3, -0.7, -1.4, -0.2], # W-pattern tetrad
        [5, 15, 50, 400, 2000, 0.6, 1.1, 1.5, 0.3], # M-pattern tetrad
    ]
)
```

(continues on next page)

(continued from previous page)

```

)
# now we use these parameters to generate some synthetic log-scaled normalised REE
# patterns and add a bit of noise
pattern_df = pd.DataFrame(
    np.vstack([example_patterns_from_parameters(l, includes_tetrads=True) for l in ls]),
    columns=REE(),
)
# We can now fit these patterns and see what the effect of fitting and not-Fitting
# tetrads might look like in these (slightly extreme) cases:
fit_ls_only = pattern_df.pyrochem.lambda_lnREE(
    norm_to=None, degree=5, fit_tetrads=False
)
fit_ts = pattern_df.pyrochem.lambda_lnREE(norm_to=None, degree=5, fit_tetrads=True)

```

We can now examine what the differences between the fits are. Below we plot the four sets of synthetic REE patterns (lambda-only above and lambda+tetrad below) and examine the relative accuracy of fitting some of the higher order lambda parameters where tetrads are also fit:

```

from pyrolite.util.plot.axes import share_axes

x, y = 2, 3
categories = np.repeat(np.arange(ls.shape[0]), 100)
colors = np.array([str(ix) * 2 for ix in categories])
l_only = categories < 2

ax = plt.figure(figsize=(12, 7)).subplot_mosaic(
    """
    AAABBBCC
    DDDEEFF
    """
)
share_axes([ax["A"], ax["D"]])
share_axes([ax["B"], ax["C"], ax["E"], ax["F"]])

ax["B"].set_title("lambdas only Fit")
ax["C"].set_title("lambdas+tetrads Fit")

for a, fltr in zip(["A", "D"], [l_only, ~l_only]):
    pattern_df.iloc[fltr, :].pyroplot.spider(
        ax=ax[a],
        label="True",
        unity_line=True,
        alpha=0.5,
        color=colors[fltr],
    )

for a, fltr in zip(["B", "E"], [l_only, ~l_only]):
    fit_ls_only.iloc[fltr, [x, y]].pyroplot.scatter(
        ax=ax[a],
        alpha=0.2,
        color=colors[fltr],
    )

```

(continues on next page)



(continued from previous page)

```

)

for a, fltr in zip(["C", "F"], [l_only, ~l_only]):
    fit_ts.iloc[fltr, [x, y]].pyroplot.scatter(
        ax=ax[a],
        alpha=0.2,
        color=colors[fltr],
    )

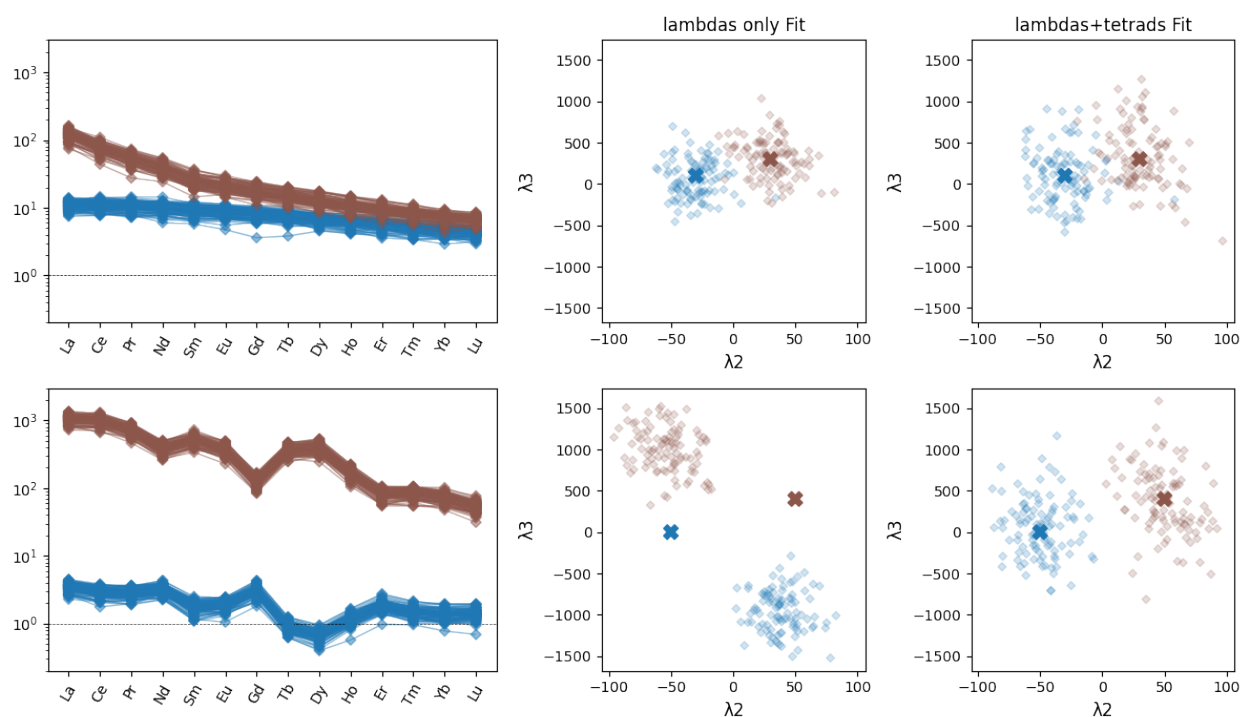
)

true = pd.DataFrame(ls[:, [x, y]], columns=[fit_ls_only.columns[ix] for ix in [x, y]])
for ix, a in enumerate(["B", "C", "E", "F"]):
    true.iloc[np.array([ix < 2, ix < 2, ix >= 2, ix >= 2]), :].pyroplot.scatter(
        ax=ax[a],
        color=np.array([str(ix) * 2 for ix in np.arange(ls.shape[0] // 2)]),
        marker="X",
        s=100,
    )

)

plt.tight_layout()
plt.show()

```



```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)

```

## More Advanced Customisation

Above we've used default parameterisations for calculating *lambdas*, but *pyrolite* allows you to customise the parameterisation of both the orthogonal polynomial components used in the fitting process as well as what data and algorithm is used in the fit itself.

To exclude some elements from the *fit* (e.g. Eu which is excluded by default, and potentially Ce), you can either i) filter the dataframe such that the columns aren't passed, or ii) explicitly exclude them:

```
# filtering the dataframe first
target_columns = [i for i in df.columns if i not in ["Eu", "Ce"]]
ls_noEuCe_filtered = df[target_columns].pyrochem.lambda_lnREE()
# excluding some elements
ls_noEuCe_excl = df.pyrochem.lambda_lnREE(exclude=["Eu", "Ce"])

# quickly checking equivalence
np.allclose(ls_noEuCe_excl, ls_noEuCe_filtered)
```

```
True
```

While the results should be numerically equivalent, *pyrolite* does provide two algorithms for fitting *lambdas*. The first follows almost exactly the original formulation (`algorithm="ONeill"`; this was translated from VBA), while the second simply uses a numerical optimization routine from *scipy* to achieve the same thing (`algorithm="opt"`; this is a fallback for where singular matrices pop up):

```
# use the original version
ls_linear = df.pyrochem.lambda_lnREE(algorithm="ONeill")

# use the optimization algorithm
ls_opt = df.pyrochem.lambda_lnREE(algorithm="opt")
```

To quickly demonstrate the equivalence, we can check numerically (to within 0.001%):

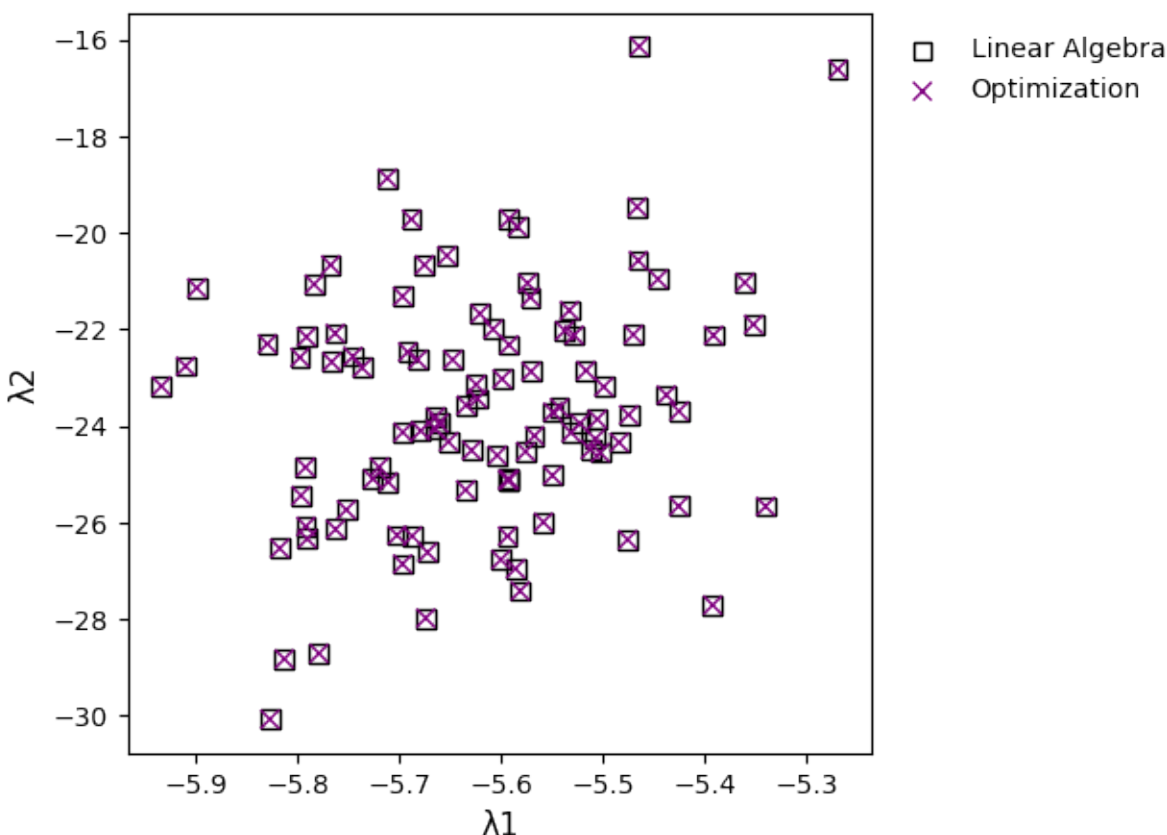
```
np.allclose(ls_linear, ls_opt, rtol=10e-5)
```

```
True
```

Or simply plot the results from both:

```
fig, ax = plt.subplots(1, figsize=(5, 5))
ax.set_title("Comparing  $\lambda$  Estimation Algorithms", y=1.1)
ls_linear.iloc[:, 1:3].pyroplot.scatter(
    ax=ax, marker="s", c="k", facecolors="none", s=50, label="Linear Algebra"
)
ls_opt.iloc[:, 1:3].pyroplot.scatter(
    ax=ax, c="purple", marker="x", s=50, label="Optimization"
)
ax.legend()
plt.show()
```

## Comparing $\lambda$ Estimation Algorithms



You can also use orthogonal polynomials defined over a different set of REE, by specifying the parameters using the keyword argument *params*:

```
# this is the original formulation from the paper, where Eu is excluded
ls_original = df.pyrochem.lambda_lnREE(params="ONeill2016")

# this uses a full set of REE
ls_fullREE_polynomials = df.pyrochem.lambda_lnREE(params="full")
```

Note that as of pyrolite v0.2.8, the original formulation is used by default, but this will cease to be the case as of the following version, where the full set of REE will instead be used to generate the orthogonal polynomials.

While the results are similar, there are small differences. They're typically less than 1%:

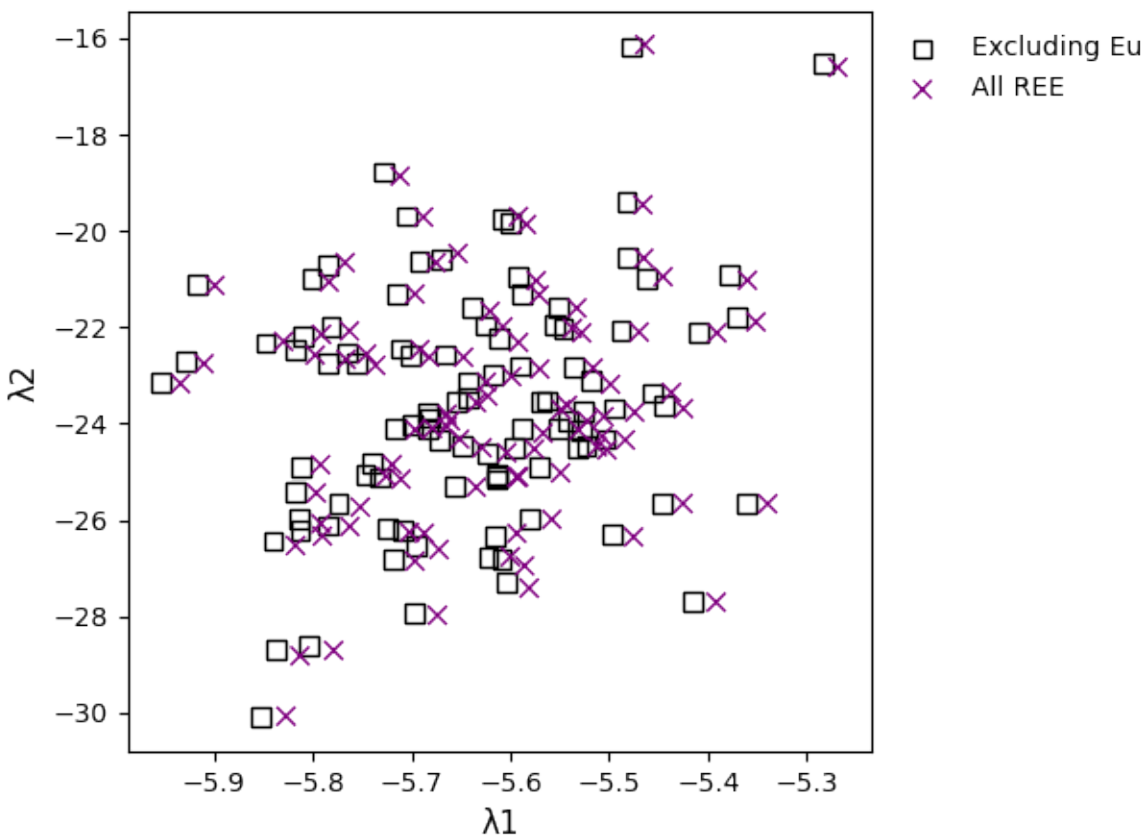
```
np.abs((ls_original / ls_fullREE_polynomials) - 1).max() * 100
```

```
0    8.265e-01
1    4.306e-01
2    7.584e-01
3    1.322e-06
dtype: float64
```

This can also be visualised:

```
fig, ax = plt.subplots(1, figsize=(5, 5))
ax.set_title("Comparing Orthogonal Polynomial Bases", y=1.1)
ls_original.iloc[:, 1:3].pyroplot.scatter(
    ax=ax, marker="s", c="k", facecolors="none", s=50, label="Excluding Eu"
)
ls_fullREE_polynomials.iloc[:, 1:3].pyroplot.scatter(
    ax=ax, c="purple", marker="x", s=50, label="All REE"
)
ax.legend()
plt.show()
```

Comparing Orthogonal Polynomial Bases



## References & Citation

For more on using orthogonal polynomials to describe geochemical pattern data, dig into the paper which introduced the method to geochemists:

O'Neill, H.S.C., 2016. The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. *J Petrology* 57, 1463–1508. doi: [10.1093/petrology/egw047](https://doi.org/10.1093/petrology/egw047).

If you're using pyrolite's implementation of *lambdas* in your research, please consider citing a recent publication directly related to this:

Anenburg, M., & Williams, M. J. (2021). Quantifying the Tetrad Effect, Shape Components, and Ce–Eu–Gd Anomalies in Rare Earth Element Patterns. *Mathematical Geosciences*. doi: [10.1007/s11004-021-09959-5](https://doi.org/10.1007/s11004-021-09959-5)

See also:

Examples:

[Ionic Radii, REE Radii Plot](#)

Functions:

`lambda_lnREE()`, `get_ionic_radii()`, `pyrolite.plot.pyroplot.REE()`

Total running time of the script: (0 minutes 19.004 seconds)

## 3.3.3 Compositional Data Examples

### Log-transforms

pyrolite includes a few functions for dealing with compositional data, at the heart of which are i) closure (i.e. everything sums to 100%) and ii) log-transforms to deal with the compositional space. The commonly used log-transformations include the Additive Log-Ratio (*ALR()*), Centred Log-Ratio (*CLR()*), and Isometric Log-Ratio (*ILR()*)<sup>12</sup>.

This example will show you how to access and use some of these functions in pyrolite.

First let's create some example data:

```
from pyrolite.util.synthetic import normal_frame, random_cov_matrix

df = normal_frame(
    size=100,
    cov=random_cov_matrix(sigmas=[0.1, 0.05, 0.3, 0.6], dim=4, seed=32),
    seed=32,
)
df.describe()
```

Let's have a look at some of the log-transforms, which can be accessed directly from your dataframes (via `pyrolite.comp.pyrocomp`), after you've imported `pyrolite.comp`. Note that the transformations will return *new* dataframes, rather than modify their inputs. For example:

```
import pyrolite.comp

lr_df = df.pyrocomp.CLR() # using a centred log-ratio transformation
```

<sup>1</sup> Aitchison, J., 1984. The statistical analysis of geochemical compositions. *Journal of the International Association for Mathematical Geology* 16, 531–564. doi: [10.1007/BF01029316](https://doi.org/10.1007/BF01029316)

<sup>2</sup> Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology* 35, 279–300. doi: [10.1023/A:1023818214614](https://doi.org/10.1023/A:1023818214614)

The transformations are implemented such that the column names generally make it evident which transformations have been applied (here using default simple labelling; see below for other examples):

```
lr_df.columns
```

```
Index(['CLR(SiO2/G)', 'CLR(CaO/G)', 'CLR(MgO/G)', 'CLR(FeO/G)', 'CLR(TiO2/G)'], dtype=
  ↪ 'object')
```

To invert these transformations, you can call the respective inverse transform:

```
back_transformed = lr_df.pyrocomp.inverse_CLR()
```

Given we haven't done anything to our dataframe in the meantime, we should be back where we started, and our values should all be equal within numerical precision. To verify this, we can use `numpy.allclose()`:

```
import numpy as np
```

```
np.allclose(back_transformed, df)
```

```
True
```

In addition to easy access to the transforms, there's also a convenience function for taking a log-transformed mean (log-transforming, taking a mean, and inverse log transforming; `logratiomean()`):

```
df.pyrocomp.logratiomean()
```

```
SiO2    0.241
CaO     0.395
MgO     0.094
FeO     0.104
TiO2    0.166
dtype: float64
```

While this function defaults to using `clr()`, you can specify other log-transforms to use:

```
df.pyrocomp.logratiomean(transform="CLR")
```

```
SiO2    0.241
CaO     0.395
MgO     0.094
FeO     0.104
TiO2    0.166
dtype: float64
```

Notably, however, the logratio means should all give you the same result:

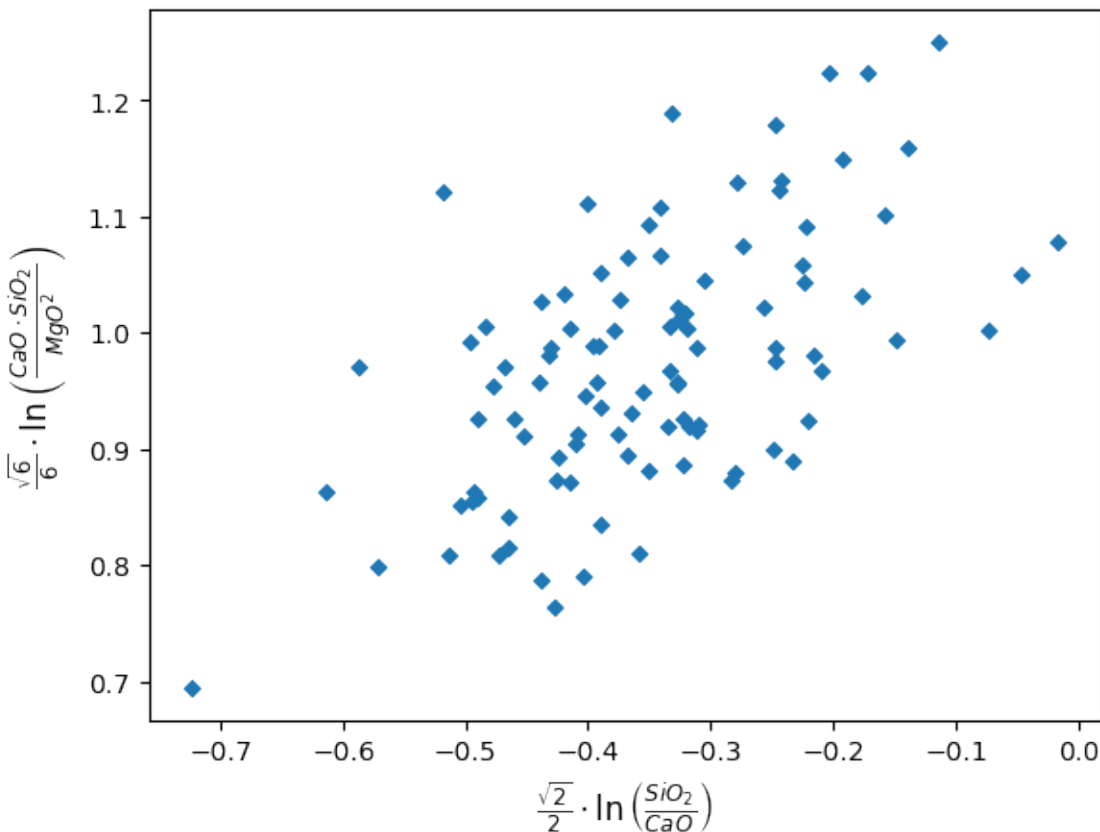
```
np.allclose(
    df.pyrocomp.logratiomean(transform="CLR"),
    df.pyrocomp.logratiomean(transform="ALR"),
) & np.allclose(
    df.pyrocomp.logratiomean(transform="CLR"),
    df.pyrocomp.logratiomean(transform="ILR"),
)
```

```
True
```

To change the default labelling outputs for column names, you can use the `label_mode` parameter, for example to get nice labels for plotting:

```
import matplotlib.pyplot as plt

df.pyrocomp.ILR(label_mode="latex").iloc[:, 0:2].pyroplot.scatter()
plt.show()
```



Alternatively if you simply want numeric indexes which you can use in e.g. a ML pipeline, you can use `label_mode="numeric"`:

```
df.pyrocomp.ILR(label_mode="numeric").columns
```

```
Index(['ILR0', 'ILR1', 'ILR2', 'ILR3'], dtype='object')
```

**See also:**

**Examples:**

[Log Ratio Means, Compositional Data, Ternary Plots](#)

**Tutorials:**

[Ternary Density Plots, Making the Logo](#)

**Modules and Functions:**

`pyrolite.comp.codata`, `boxcox()`, `renormalise()`

**Total running time of the script:** (0 minutes 0.750 seconds)

## Spherical Coordinate Transformations

While logratio methods are typically more commonly used to deal with compositional data, they are not infallible - their principal weakness is being able to deal with true zeros like we have the example dataset below. True zeros are common in mineralogical datasets, and are possible in geochemical datasets to some degree (at least at the level of atom counting..).

An alternative to this is to use a spherical coordinate transform to handle our compositional data. This typically involves treating each covariate as a separate dimension/axis and each composition as a unit vector in this space. The transformation involves the iterative calculation of an angular representation of these vectors.  $N$ -dimensional vectors transformed to  $N - 1$  dimensional angular equivalents (as the first angle is that between the first axis and the second). At least two variants of this type of transformation exist - a spherical cosine method and a spherical sine method; these are complementary and the sine method is used here (where angles close to  $\pi/2$  represent close to zero abundance, and angles closer to zero/aligning with the respective component axis represent higher abundances). See below for an example of this demonstrated graphically.

First let's create some example mineralogical abundance data, where at least one of the minerals might occasionally have zero abundance:

```
import numpy as np

from pyrolite.util.synthetic import normal_frame

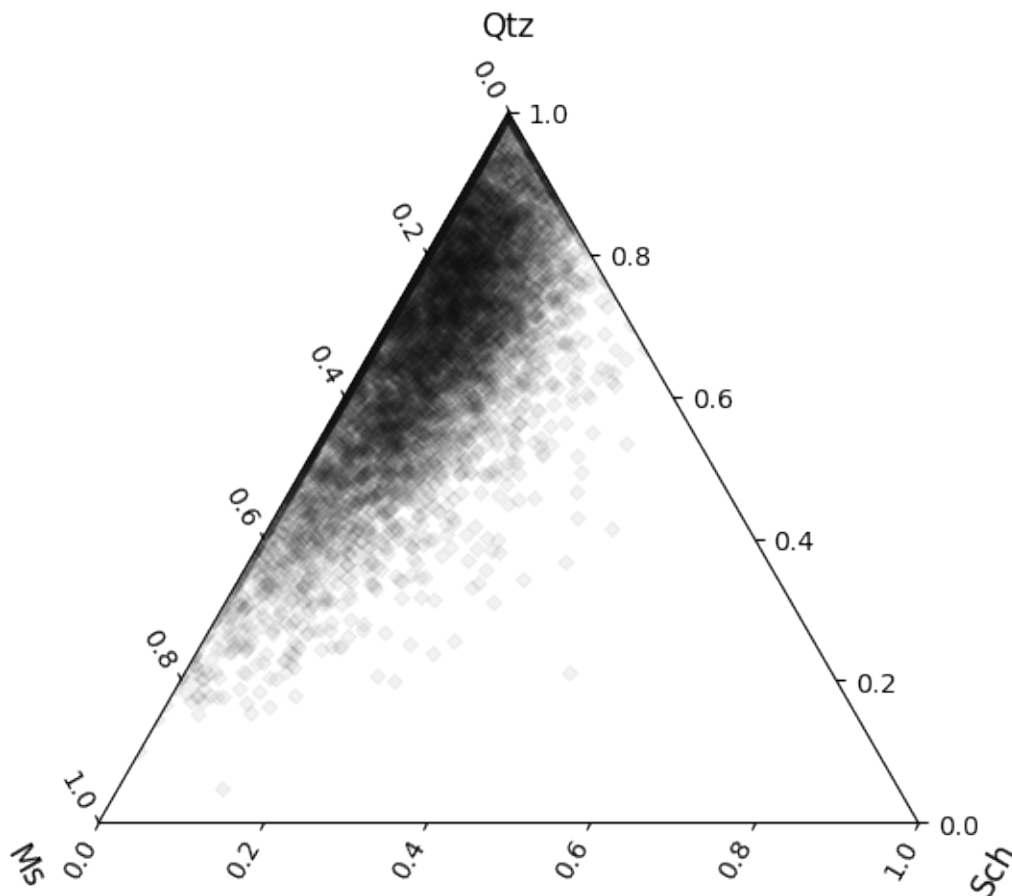
comp = normal_frame(
    columns=["Ab", "Qtz", "Ms", "Sch"],
    mean=[0.5, 1, 0.3, 0.05],
    cov=np.eye(3) * np.array([0.02, 0.5, 0.5]),
    size=10000,
    seed=145,
)
comp += 0.05 * np.random.randn(*comp.shape)
comp[comp <= 0] = 0
comp = comp.pyrocomp.renormalise(scale=1)
```

We can quickly visualise this to see that it does indeed have some true zeros:

```
import matplotlib.pyplot as plt

comp[["Qtz", "Ms", "Sch"]].pyroplot.scatter(alpha=0.05, c="k")
plt.show()
```





The spherical coordinate transform functions can be found within `pyrolite.comp.codata`, but can also be accessed from the `pyrocomp` dataframe accessor:

```
import pyrolite.comp

angles = comp.pyrocomp.sphere()
angles.head()
```

The inverse function can be accessed in the same way:

```
inverted_angles = angles.pyrocomp.inverse_sphere()
inverted_angles.head()
```

We can see that the inverted angles are within precision of the original composition we used:

```
np.isclose(inverted_angles, comp.values).all()
```

```
True
```

To better understand what's going on here, visualising our data is often the best first step. Below we use a helper function from `pyrolite.util.plot.helpers` to create a 3D axis on which to plot our angular data.

```
from pyrolite.plot.color import process_color
from pyrolite.util.plot.helpers import init_spherical_octant
```

(continues on next page)

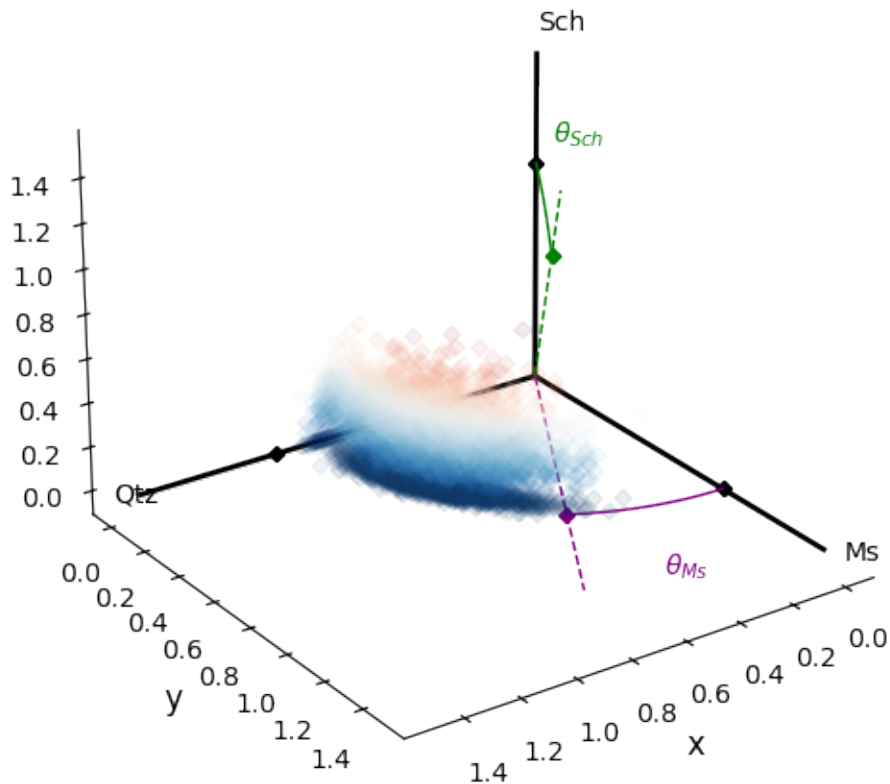
(continued from previous page)

```

ax = init_spherical_octant(labels=[c[2:] for c in angles.columns], figsize=(6, 6))

# here we can color the points by the angle from the Schorl axis
colors = process_color(angles["_Sch"], cmap="RdBu", alpha=0.1)["c"]
ax.scatter(*np.sqrt(comp.values[:, 1:]).T, c=colors)
plt.show()

```

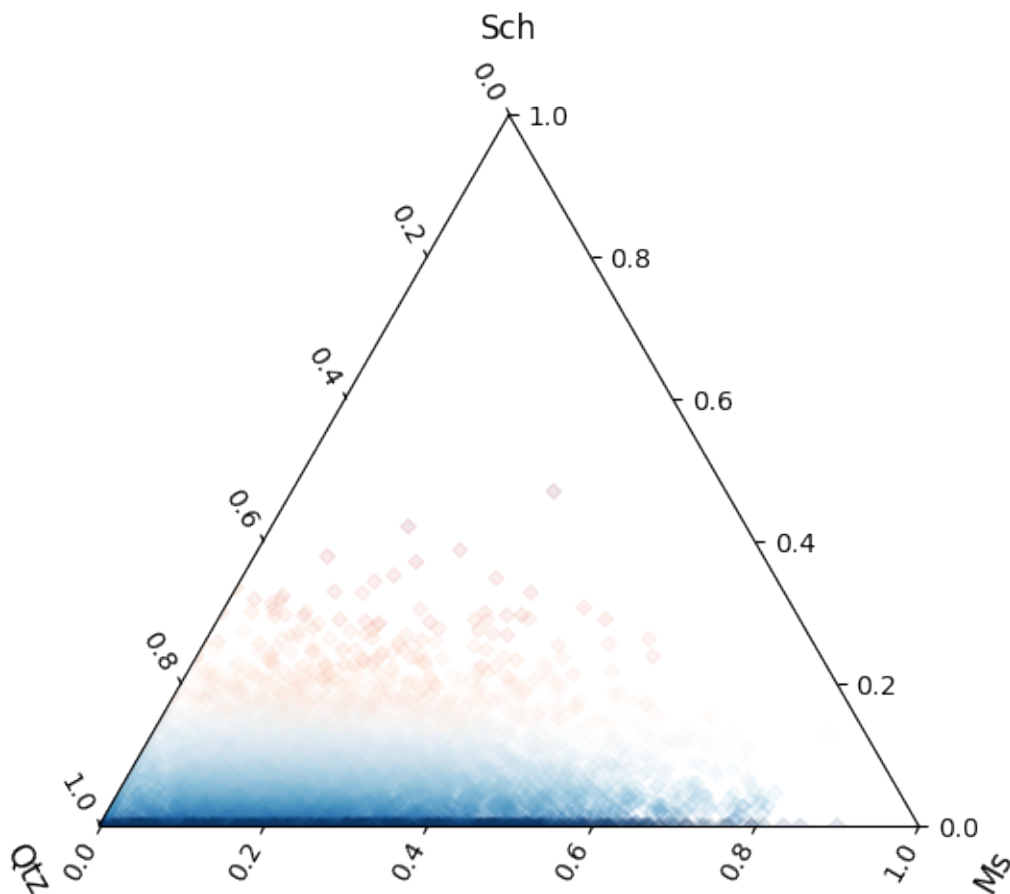


We can compare this to the equivalent ternary projection of our data; note we need to reorder some columns in order to make this align in the same way:

```

tcolumns = np.array([c[2:] for c in angles.columns])[[2, 0, 1]]
comp[tcolumns].pyroplot.scatter(c=colors)
plt.show()

```



Total running time of the script: (0 minutes 7.184 seconds)

### Log Ratio Means

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

import pyrolite.comp
from pyrolite.comp.codata import ILR, close, inverse_ILR
from pyrolite.plot import pyroplot
from pyrolite.util.synthetic import random_cov_matrix

np.random.seed(82)
```

```
def random_compositional_trend(m1, m2, c1, c2, resolution=20, size=1000):
    """
    Generate a compositional trend between two compositions with independent
    variances.
    """
    # generate means intermediate between m1 and m2
```

(continues on next page)

(continued from previous page)

```
mv = np.vstack([ILR(close(m1)).reshape(1, -1), ILR(close(m2)).reshape(1, -1)])
ms = np.apply_along_axis(lambda x: np.linspace(*x, resolution), 0, mv)
# generate covariance matrices intermediate between c1 and c2
cv = np.vstack([c1.reshape(1, -1), c2.reshape(1, -1)])
cs = np.apply_along_axis(lambda x: np.linspace(*x, resolution), 0, cv)
cs = cs.reshape(cs.shape[0], *c1.shape)
# generate samples from each
samples = np.vstack(
    [
        np.random.multivariate_normal(m.flatten(), cs[ix], size=size // resolution)
        for ix, m in enumerate(ms)
    ]
)
# combine together.
return inverse_ILR(samples)
```

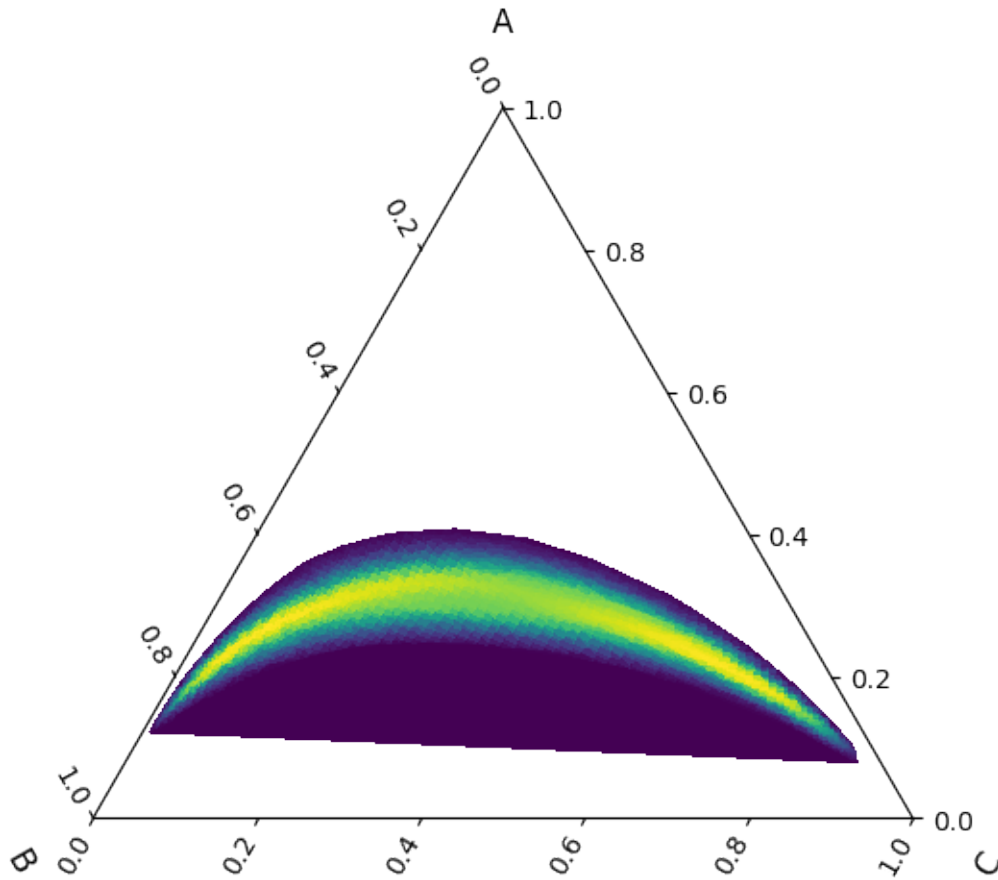
First we create an array of compositions which represent a trend.

```
m1, m2 = np.array([[0.3, 0.1, 2.1]]), np.array([[0.5, 2.5, 0.05]])
c1, c2 = (
    random_cov_matrix(2, sigmas=[0.15, 0.05]),
    random_cov_matrix(2, sigmas=[0.05, 0.2]),
)

trend = pd.DataFrame(
    random_compositional_trend(m1, m2, c1, c2, resolution=100, size=5000),
    columns=["A", "B", "C"],
)
```

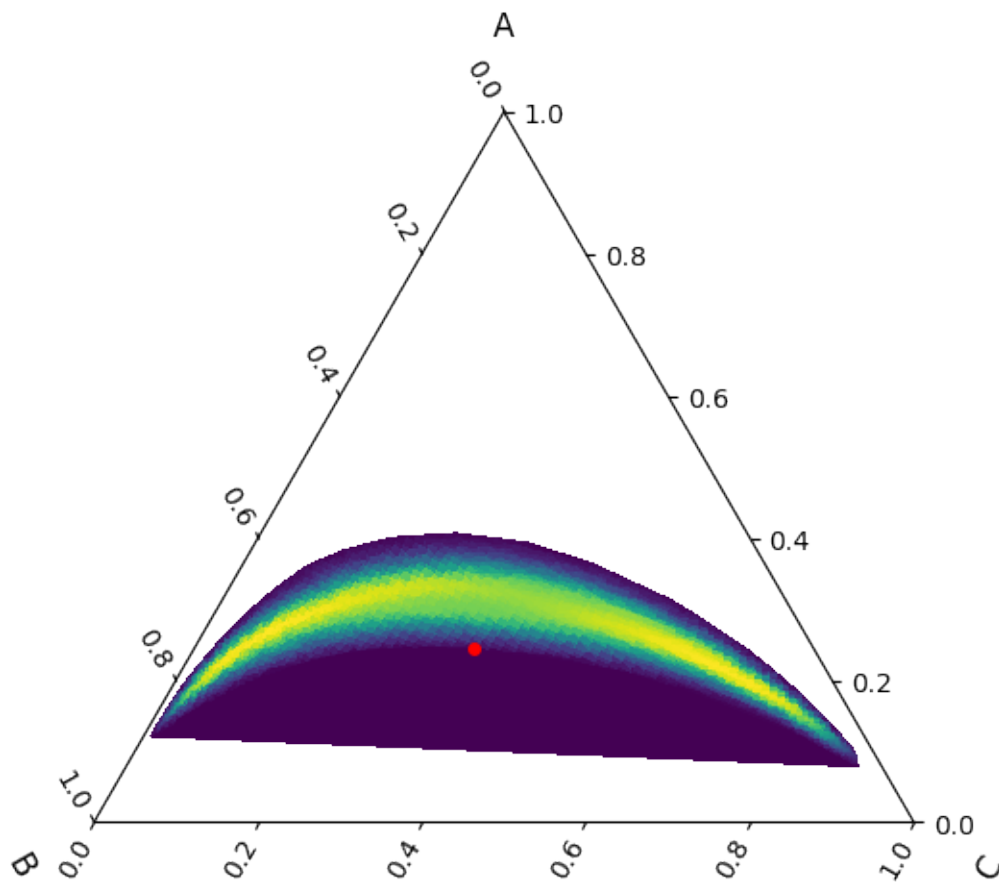
We can visualise this compositional trend with a density plot.

```
ax = trend.pyroplot.density(mode="density", bins=100)
plt.show()
```



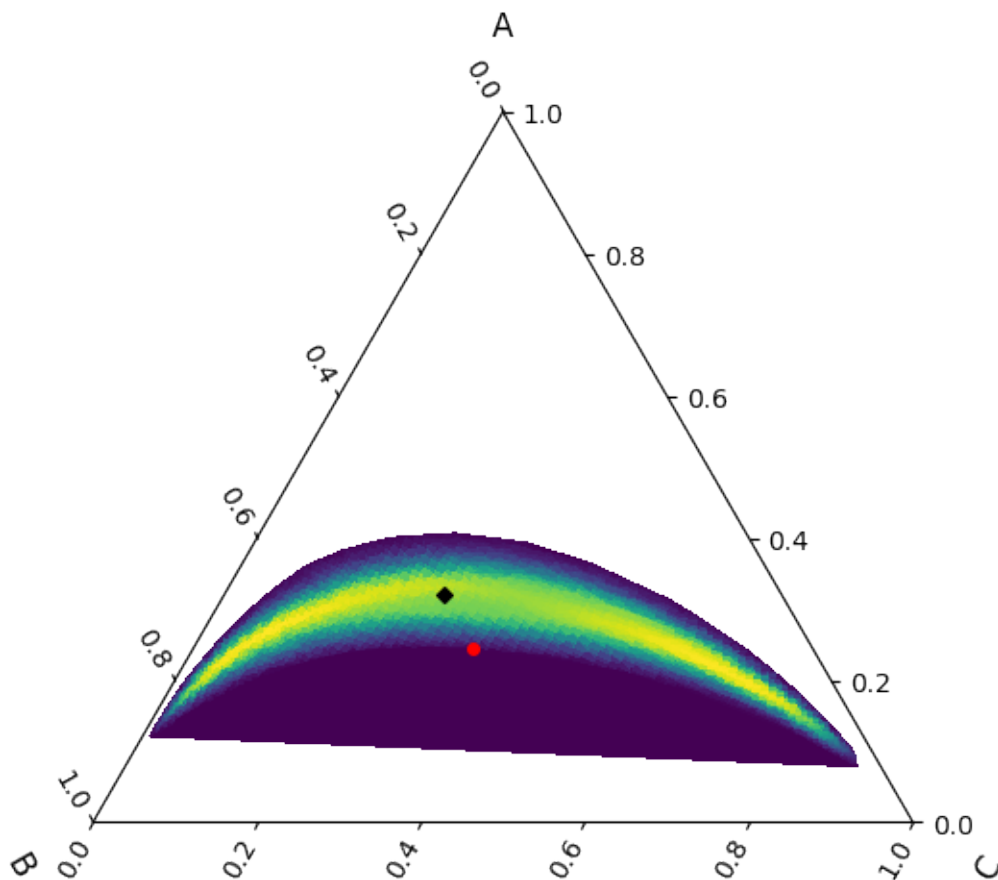
First we can see where the geometric mean would fall:

```
geomean = trend.mean(axis=0).to_frame().T
ax = geomean.pyroplot.scatter(ax=ax, marker="o", color="r", zorder=2, label="GeoMean")
plt.show()
```



Finally, we can also see where the logratio mean would fall:

```
ILRmean = trend.pyrocomp.logratiomean(transform="ILR")
ax = ILRmean.pyroplot.scatter(ax=ax, color="k", label="LogMean")
plt.show()
```



See also:

#### Examples:

[Log Transforms, Compositional Data, Ternary Plots](#)

#### Tutorials:

[Ternary Density Plots, Making the Logo](#)

#### Modules and Functions:

`pyrolite.comp.codata`, `renormalise()`

Total running time of the script: (0 minutes 10.536 seconds)

### Compositional Data?

pyrolite comes with a few datasets from Aitchison (1984) built in which we can use as examples:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from pyrolite.plot import pyroplot
from pyrolite.data.Aitchison import load_kongite

df = load_kongite()
```

For compositional data, everything is relative (thanks to the closure property), so we tend to use ratios to express differences or changes between things. However, if we make incorrect assumptions about the nature of our data, we can get some incorrect answers. Say you want to know the average ratio between A and B:

```
A_on_B = df["A"] / df["B"]
A_on_B.mean() # 2.8265837788402983
```

```
2.8265837788402983
```

Equally, you could have chosen to calculate the average ratio between B and A

```
B_on_A = df["B"] / df["A"]
B_on_A.mean() # 0.4709565704852008
```

```
0.4709565704852008
```

You expect these to be invertible, such that  $A\_on\_B = 1 / B\_on\_A$ ; but not so!

```
A_on_B.mean() / (1 / B_on_A.mean()) # 1.3311982026717262
```

```
1.3311982026717262
```

Similarly, the relative variances are different:

```
np.std(A_on_B) / A_on_B.mean() # 0.6295146309597085
np.std(B_on_A) / B_on_A.mean() # 0.5020948201979953
```

```
0.5020948201979953
```

This improves when using logratios in place of simple ratios, prior to exponentiating means

```
logA_on_B = (df["A"] / df["B"]).apply(np.log)
logB_on_A = (df["B"] / df["A"]).apply(np.log)
```

The logratios are invertible:

```
np.exp(logA_on_B.mean()) # 2.4213410747400514
1 / np.exp(logB_on_A.mean()) # 2.421341074740052
```

```
2.421341074740052
```

The logratios also have the same variance:

```
(np.std(logA_on_B) / logA_on_B.mean()) ** 2 # 0.36598579018127086
(np.std(logB_on_A) / logB_on_A.mean()) ** 2 # 0.36598579018127086
```

```
0.36598579018127086
```

These peculiarities result from incorrect assumptions regarding the distribution of the data: ratios of compositional components are typically *lognormally* distributed, rather than *normally* distributed, and the compositional components themselves commonly have a *Poisson distribution*. These distributions contrast significantly with the normal distribution at the core of most statistical tests. We can compare distributions with similar means and variances but different forms, and note that the normal distribution has one immediate failure, in that it has non-zero probability density below 0, and we know that you can't have negative atoms!



```

from scipy.stats import norm, poisson, lognorm

means = [[10, 10], [10, 20], [20, 100], [1000, 50]]
fig, ax = plt.subplots(len(means), 4, figsize=(11, 8))
ax[0, 0].set_title("A")
ax[0, 1].set_title("B")
ax[0, 2].set_title("Normal Fit to B/A")
ax[0, 3].set_title("Lognormal Fit to B/A")
ax[-1, 0].set_xlabel("A")
ax[-1, 1].set_xlabel("B")
ax[-1, 2].set_xlabel("B/A")
ax[-1, 3].set_xlabel("B/A")
for ix, (m1, m2) in enumerate(means):
    p1, p2 = poisson(mu=m1), poisson(mu=m2)
    y1, y2 = p1.rvs(2000), p2.rvs(2000)
    ratios = y2[y1 > 0] / y1[y1 > 0]

    y1min, y1max = y1.min(), y1.max()
    y2min, y2max = y2.min(), y2.max()
    ax[ix, 0].hist(
        y1,
        color="0.5",
        alpha=0.6,
        label="A",
        bins=np.linspace(y1min - 0.5, y1max + 0.5, (y1max - y1min) + 1),
    )
    ax[ix, 1].hist(
        y2,
        color="0.5",
        alpha=0.6,
        label="B",
        bins=np.linspace(y2min - 0.5, y2max + 0.5, (y2max - y2min) + 1),
    )

    # normal distribution fit
    H, binedges, patches = ax[ix, 2].hist(
        ratios, color="Purple", alpha=0.6, label="Ratios", bins=100
    )
    loc, scale = norm.fit(ratios, loc=0)
    pdf = norm.pdf(binedges, loc, scale)
    twin2 = ax[ix, 2].twinx()
    twin2.set_ylim(0, 1.1 * np.max(pdf))
    twin2.plot(binedges, pdf, color="k", ls="--", label="Normal Fit")

    # log-normal distribution fit
    H, binedges, patches = ax[ix, 3].hist(
        ratios, color="Green", alpha=0.6, label="Ratios", bins=100
    )
    s, loc, scale = lognorm.fit(ratios, loc=0)
    pdf = lognorm.pdf(binedges, s, loc, scale)
    twin3 = ax[ix, 3].twinx()
    twin3.set_ylim(0, 1.1 * np.max(pdf))
    twin3.plot(binedges, pdf, color="k", ls="--", label="Lognormal Fit")

```

(continues on next page)

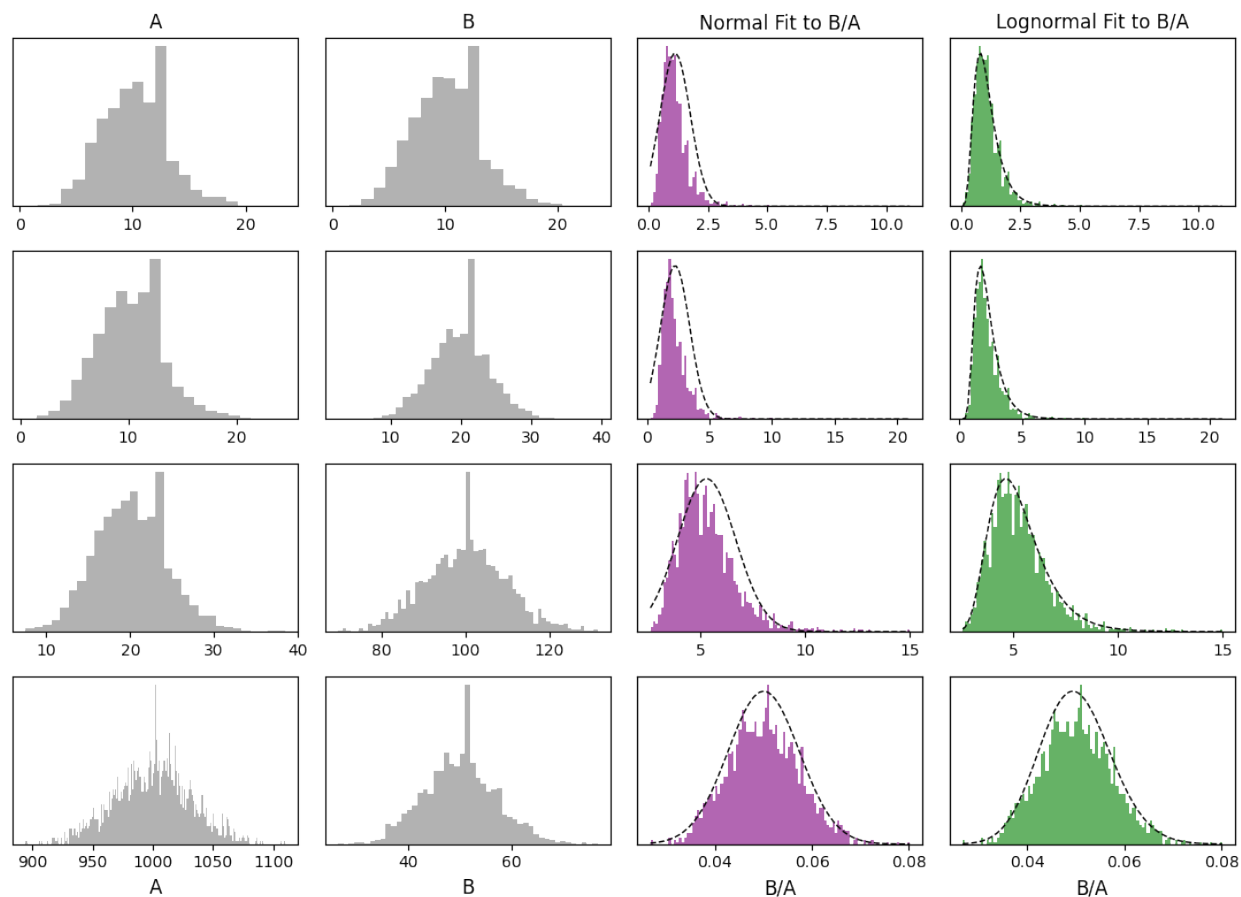
(continued from previous page)

```

for a in [*ax[ix, :], twin2, twin3]:
    a.set_yticks([])

plt.tight_layout()

```



The form of these distributions is a reflection of the fact that geochemical data is at its core a measure of relative quantities of atoms. Quantities of atoms have discrete distributions (i.e. you can have precisely 0, 1 or  $6.02 \times 10^{23}$  atoms, but 1.23 atoms is not a sensible state of affairs); if you were to count them in a shiny machine, the amount of atoms you might measure over a given period will have a Poisson distribution. If you measure two components, the probability density distribution of the ratio is well approximated by a lognormal distribution (note this doesn't consider inherent covariance):

```

from pyrolite.util.plot.axes import share_axes, subaxes
from pyrolite.util.distributions import lognorm_to_norm, norm_to_lognorm

# starting from a normal distribution, then creating similar non-normal distributions
mean, sd = 2.5, 1.5 #
logmu, logs = norm_to_lognorm(mean, sd) # parameters for equival
normrv = norm(loc=mean, scale=sd)
lognormrv = lognorm(s=logs, scale=logmu)
poissonrv = poisson(mu=mean)

```

We can visualise the similarities and differences between these distributions:

```

fig, ax = plt.subplots(2, 3, figsize=(8, 4))
ax = ax.flat
for a in ax:
    a.subax = subaxes(a, side="bottom")

share_axes(ax[:3], which="x")
share_axes(ax[3:], which="x")
ax[0].set_xlim(-2, 10)
ax[3].set_xscale("log")
ax[3].set_xlim(0.1, 10)

for a in ax:
    a.axvline(0, color="k", lw=0.5, ls="--")

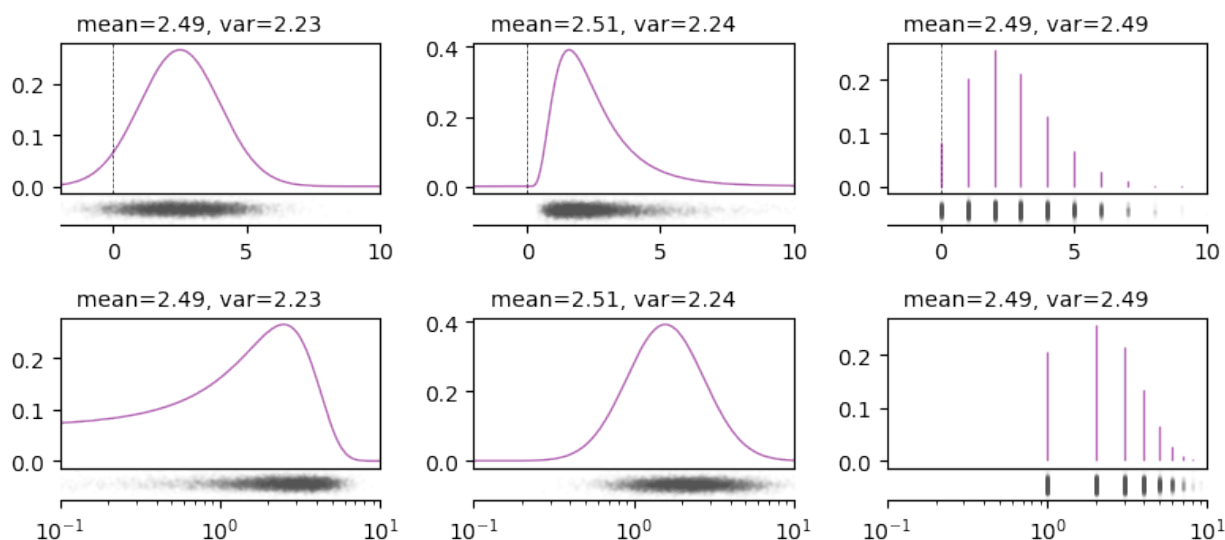
# xs at which to evaluate the pdfs
x = np.linspace(-5, 15.0, 1001)

for ix, dist in enumerate([normrv, lognormrv, poissonrv]):
    _xs = dist.rvs(size=10000) # random sample
    _ys = -0.05 + np.random.randn(10000) / 100 # random offsets for visualisation
    for a in [ax[ix], ax[ix + 3]]:
        a.annotate(
            "mean={:.2f}, var={:.2f}".format(np.mean(_xs), np.var(_xs)),
            xy=(0.05, 1.05),
            ha="left",
            va="bottom",
            xycoords=a.transAxes,
        )
        a.subax.scatter(_xs, _ys, s=2, color="k", alpha=0.01)
        if dist != poissonrv: # cont. distribution
            a.plot(x, dist.pdf(x), color="Purple", alpha=0.6, label="pdf")
        else: # discrete distribution
            a.vlines(
                x[x >= 0],
                0,
                dist.pmf(x[x >= 0]),
                color="Purple",
                alpha=0.6,
                label="pmf",
            )

fig.suptitle("Data Distributions: Normal, Lognormal, Poisson", y=1.1)
plt.tight_layout()

```

## Data Distributions: Normal, Lognormal, Poisson



Accounting for these inherent features of geochemical data will allow you to accurately estimate means and variances, and from this enables the use of standardised statistical measures - as long as you're log-transforming your data. When performing multivariate analysis, use log-ratio transformations (including the additive logratio `alr()`, centred logratio `clr()` and isometric logratio `ilr()`). In this case, the logratio-mean is implemented for you:

```
from pyrolite.comp.codata import logratio_mean
import itertools

fig, ax = plt.subplots(2, 2, figsize=(12, 12), subplot_kw=dict(projection="ternary"))
ax = ax.flat

for columns, a in zip(itertools.combinations(["A", "B", "C", "D"], 3), ax):
    columns = list(columns)

    df.loc[:, columns].pyroplot.scatter(
        ax=a, color="k", marker=".", label=df.attrs["name"], no_ticks=True
    )

    df.mean().loc[columns].pyroplot.scatter(
        ax=a,
        edgecolors="red",
        linewidths=2,
        c="none",
        s=50,
        label="Arithmetic Mean",
        no_ticks=True,
    )

    logratio_mean(df.loc[:, columns]).pyroplot.scatter(
        ax=a,
        edgecolors="k",
```

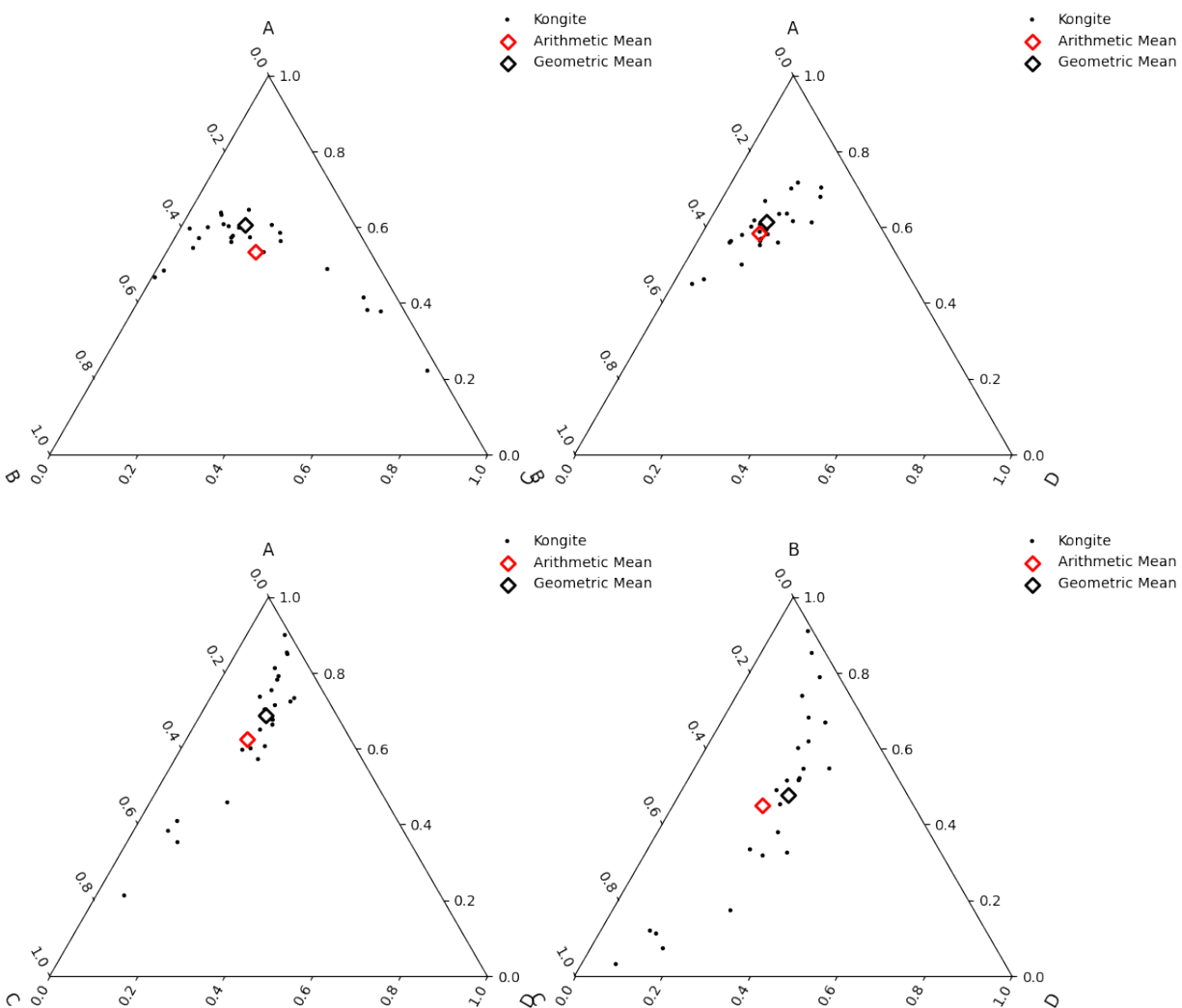
(continues on next page)

(continued from previous page)

```

linewidths=2,
c="none",
s=50,
label="Geometric Mean",
axlabels=True,
no_ticks=True,
)
a.legend(loc=(0.8, 0.5))

```

**See also:****Examples:**[Log Transforms, Logratio Means, Ternary Plots](#)**Tutorials:**[Ternary Density Plots, Making the Logo](#)**Modules and Functions:**`pyrolite.comp.codata, renormalise()`**Total running time of the script:** (0 minutes 18.853 seconds)

### 3.3.4 Utility Examples

pyrolite includes a range of utilities for everything from dealing with the web to plotting, synthetic data and machine learning. While most of these are used as part of the core functions of pyrolite, you may also find other uses for them, and this section provides some simple examples for some of these.

#### Using Manifolds for Visualisation

Visualisation of data which has high dimensionality is challenging, and one solution is to provide visualisations in low-dimension representations of the space actually spanned by the data. Here we provide an example of visualisation of classification predictions and relative prediction certainty (using entropy across predicted probability for each individual class) for a toy `sklearn` dataset.

```
import matplotlib.pyplot as plt
import numpy as np
import sklearn.datasets

from pyrolite.util.plot import DEFAULT_DISC_COLORMAP
from pyrolite.util.sklearn.pipeline import SVC_pipeline
from pyrolite.util.sklearn.vis import plot_mapping

np.random.seed(82)
```

```
wine = sklearn.datasets.load_wine()
data, target = wine["data"], wine["target"]
```

```
# data = data[:, np.random.random(data.shape[1]) > 0.4] # randomly remove fraction of
↳ dimensionality
```

```
svc = SVC_pipeline(probability=True)
gs = svc.fit(data, target)
```

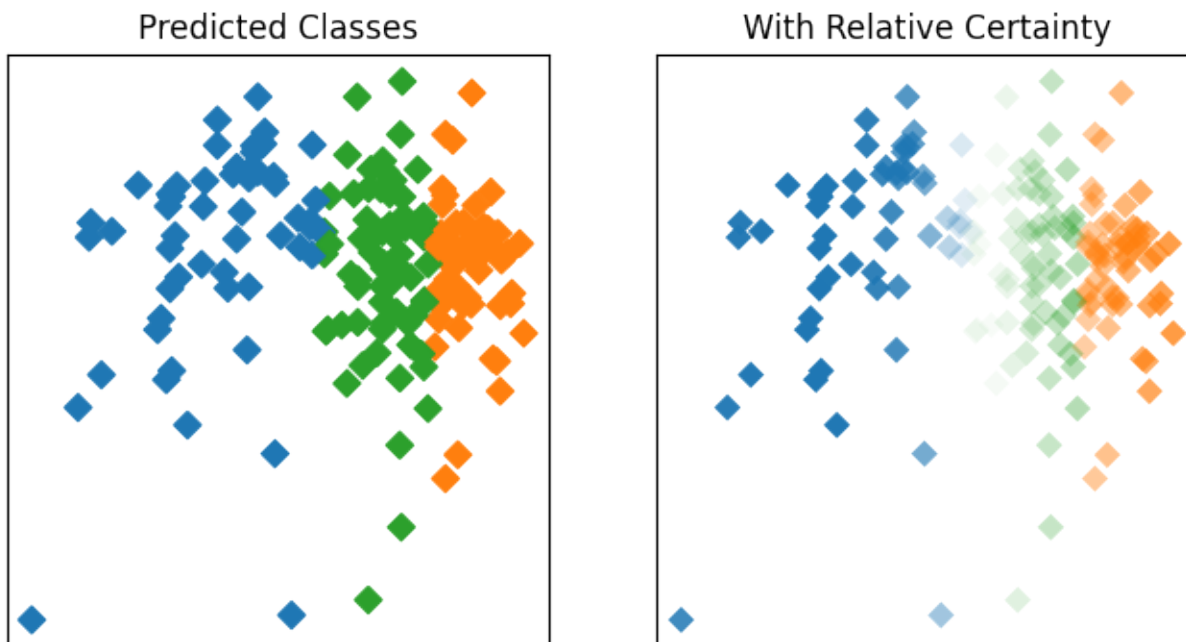
Fitting 10 folds for each of 1 candidates, totalling 10 fits

```
fig, ax = plt.subplots(1, 2, figsize=(8, 4))

a, tfm, mapped = plot_mapping(data, gs.best_estimator_, ax=ax[1], s=50, init="pca")
ax[0].scatter(*mapped.T, c=DEFAULT_DISC_COLORMAP(gs.predict(data)), s=50)

ax[0].set_title("Predicted Classes")
ax[1].set_title("With Relative Certainty")

for a in ax:
    a.set_xticks([])
    a.set_yticks([])
```



Total running time of the script: (0 minutes 3.206 seconds)

### TAS Classifier

Some simple discrimination methods are implemented, including the Total Alkali-Silica (TAS) classification.

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

from pyrolite.util.classification import TAS
from pyrolite.util.synthetic import normal_frame, random_cov_matrix
```

We'll first generate some synthetic data to play with:

```
df = (
    normal_frame(
        columns=["SiO2", "Na2O", "K2O", "Al2O3"],
        mean=[0.5, 0.04, 0.05, 0.4],
        size=100,
        seed=49,
    )
    * 100
)

df.head(3)
```

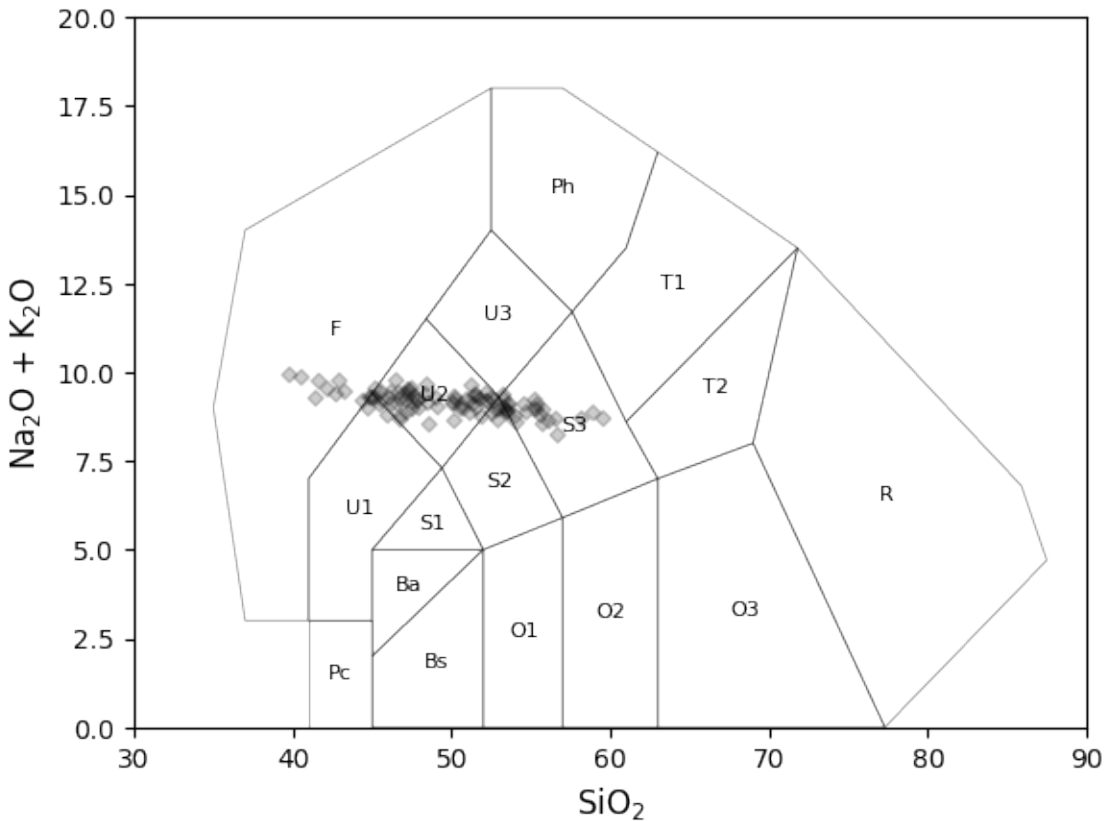
We can visualise how this chemistry corresponds to the TAS diagram:

```
df["Na2O + K2O"] = df["Na2O"] + df["K2O"]
cm = TAS()
```

(continues on next page)

(continued from previous page)

```
fig, ax = plt.subplots(1)
cm.add_to_axes(ax, alpha=0.5, linewidth=0.5, zorder=-1, add_labels=True)
df[["SiO2", "Na2O + K2O"]].pyroplot.scatter(ax=ax, c="k", alpha=0.2, axlabels=False)
plt.show()
```



We can now classify this data according to the fields of the TAS diagram, and add this as a column to the dataframe. Similarly, we can extract which rock names the TAS fields correspond to:

```
df["TAS"] = cm.predict(df)
df["Rocknames"] = df.TAS.apply(lambda x: cm.fields.get(x, {"name": None})["name"])
df["Rocknames"].sample(10) # randomly check 10 sample rocknames
```

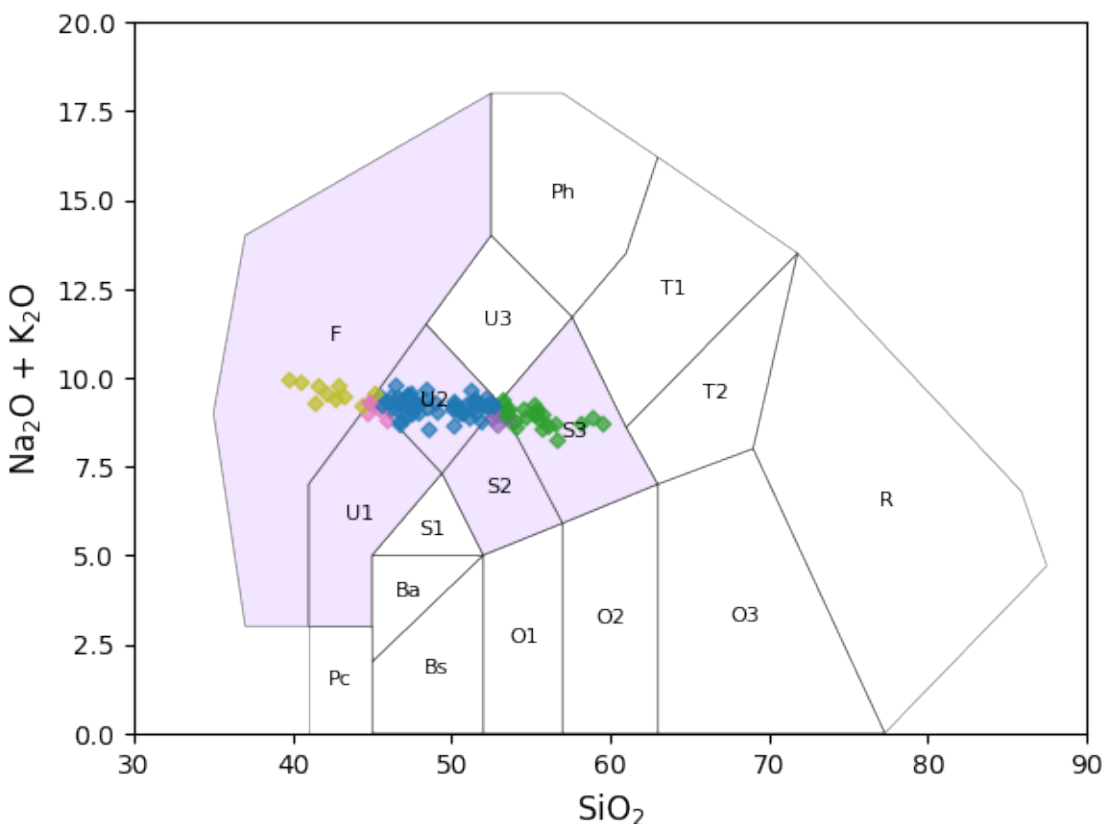
```
78 [Phonotephrite, Foid Monzodiorite]
2 [Phonotephrite, Foid Monzodiorite]
56 [Phonotephrite, Foid Monzodiorite]
41 [Phonotephrite, Foid Monzodiorite]
93 [Phonotephrite, Foid Monzodiorite]
60 [Phonotephrite, Foid Monzodiorite]
61 [Phonotephrite, Foid Monzodiorite]
47 [Phonotephrite, Foid Monzodiorite]
4 [Trachy-\nandesite, Monzonite]
12 [Trachy-\nandesite, Monzonite]
Name: Rocknames, dtype: object
```



We could now take the TAS classes and use them to colorize our points for plotting on the TAS diagram, or more likely, on another plot. Here the relationship to the TAS diagram is illustrated, coloring also the populated fields:

```
fig, ax = plt.subplots(1)

cm.add_to_axes(ax, alpha=0.5, linewidth=0.0, zorder=-2, add_labels=False,
               which_ids=np.unique(df["TAS"]), fill=True, facecolor=[0.9, 0.8, 1.0])
cm.add_to_axes(ax, alpha=0.5, linewidth=0.5, zorder=-1, add_labels=True)
df[["SiO2", "Na2O + K2O"]].pyroplot.scatter(ax=ax, c=df["TAS"], alpha=0.7,
      ↪axlabels=False)
```



```
<Axes: xlabel='SiO$_2$', ylabel='Na$_2$O + K$_2$O'>
```

### Variations of the Diagram

To use a different variation of the TAS diagram, you can pass the relevant keyword `which_model` allowing you to access the other available variants.

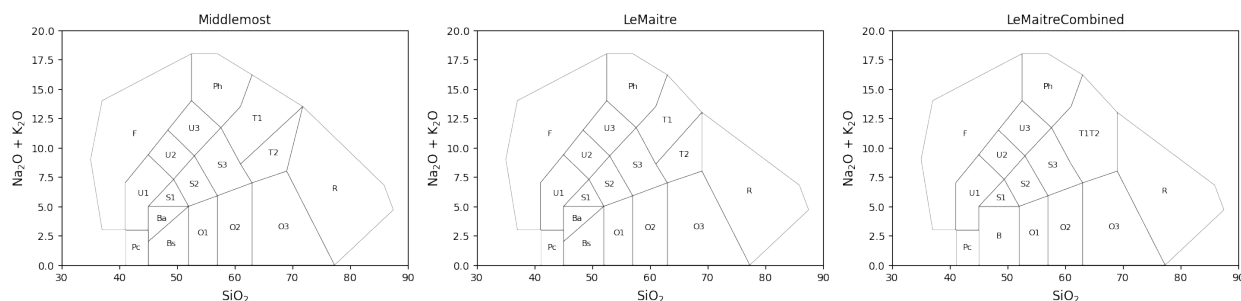
Currently, the Le Bas/Le Maitre alternative and a T1-T2 combined variant of it are available as alternatives to the Middlemost version. Each of these can be used as a classifier model as for the default above.

```
fig, ax = plt.subplots(1, 3, figsize=(20, 4))
for a, model in zip(ax, ["Middlemost", "LeMaitre", "LeMaitreCombined"]):
    a.set_title(model if model is not None else "")
    cm = TAS(which_model=model)
```

(continues on next page)

(continued from previous page)

```
cm.add_to_axes(a, alpha=0.5, linewidth=0.5, add_labels=True)
plt.show()
```



## References & Citation

For a few references on the TAS diagram as used here:

Middlemost, E. A. K. (1994). Naming materials in the magma/igneous rock system. *Earth-Science Reviews*, 37(3), 215–224. doi: [10.1016/0012-8252\(94\)90029-9](https://doi.org/10.1016/0012-8252(94)90029-9).

Le Bas, M.J., Le Maitre, R.W., Woolley, A.R. (1992). The construction of the Total Alkali-Silica chemical classification of volcanic rocks. *Mineralogy and Petrology* 46, 1–22. doi: [10.1007/BF01160698](https://doi.org/10.1007/BF01160698).

Le Maitre, R.W. (2002). *Igneous Rocks: A Classification and Glossary of Terms : Recommendations of International Union of Geological sciences Subcommittee on the Systematics of Igneous Rocks*. Cambridge University Press, 236pp. doi: [10.1017/CBO9780511535581](https://doi.org/10.1017/CBO9780511535581).

**Total running time of the script:** (0 minutes 1.338 seconds)

## Geological Timescale

pyrolite includes a simple geological timescale, based on a recent version of the International Chronostratigraphic Chart<sup>1</sup>. The *Timescale* class can be used to look up names for specific geological ages, to look up times for known geological age names and to access a reference table for all of these.

First we'll create a timescale:

```
from pyrolite.util.time import Timescale

ts = Timescale()
```

From this we can look up the names of ages (in million years, or Ma):

```
ts.named_age(1212.1)
```

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
util/time.py:261: FutureWarning: Series.__getitem__ treating keys as positions is
deprecated. In a future version, integer keys will always be treated as labels
(consistent with DataFrame behavior). To access a value by position, use `ser.
iloc[pos]`
max([ix for (ix, r) in enumerate(counts) if r == counts[0]])
```

(continues on next page)

<sup>1</sup> Cohen, K.M., Finney, S.C., Gibbard, P.L., Fan, J.-X., 2013. The ICS International Chronostratigraphic Chart. *Episodes* 36, 199–204.

(continued from previous page)

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→util/time.py:54: FutureWarning: Series.__getitem__ treating keys as positions is
→deprecated. In a future version, integer keys will always be treated as labels
→(consistent with DataFrame behavior). To access a value by position, use `ser.
→iloc[pos]`
    nameguess = agenamelist[-1]

'Ectasian'
```

As geological age names are hierarchical, the name you give an age depends on what level you're looking at. By default, the timescale will return the most specific non-null level. The levels accessible within the timescale are listed as an attribute:

```
ts.levels
```

```
['Eon', 'Era', 'Period', 'Superepoch', 'Epoch', 'Age']
```

These can be used to refine the output names to your desired level of specificity (noting that for some ages, the levels which are accessible can differ; see the chart):

```
ts.named_age(1212.1, level="Epoch")
```

```
'Ectasian'
```

The timescale can also do the inverse for you, and return the timing information for a given named age:

```
ts.text2age("Holocene")
```

```
(0.0117, 0.0)
```

We can use this to create a simple template to visualise the geological timescale:

```
import pandas as pd
import matplotlib.pyplot as plt

fig, ax = plt.subplots(1, figsize=(5, 10))

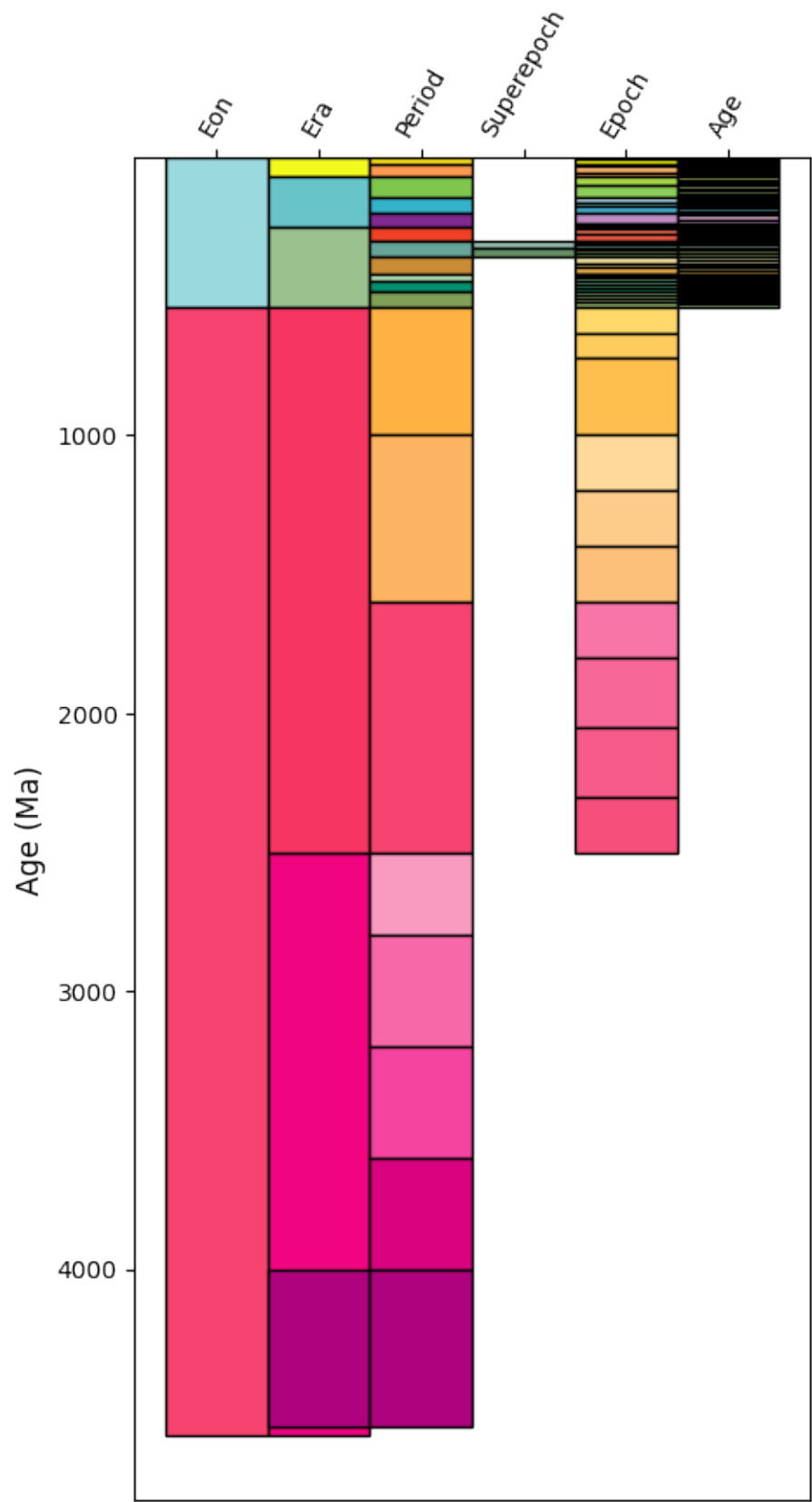
for ix, level in enumerate(ts.levels):
    ldf = ts.data.loc[ts.data.Level == level, :]
    for pix, period in ldf.iterrows():
        ax.bar(
            ix,
            period.Start - period.End,
            facecolor=period.Color,
            bottom=period.End,
            width=1,
            edgecolor="k",
        )

ax.set_xticks(range(len(ts.levels)))
ax.set_xticklabels(ts.levels, rotation=60)
ax.xaxis.set_ticks_position("top")
```

(continues on next page)

(continued from previous page)

```
ax.set_ylabel("Age (Ma)")
ax.invert_yaxis()
```



This doesn't quite look like the geological timescale you may be used to. We can improve on the output somewhat with a bit of customisation for the positioning. Notably, this is less readable, but produces something closer to what we're after. Some of this may soon be integrated as a *Timescale* method, if there's interest.

```
import numpy as np
from matplotlib.patches import Rectangle

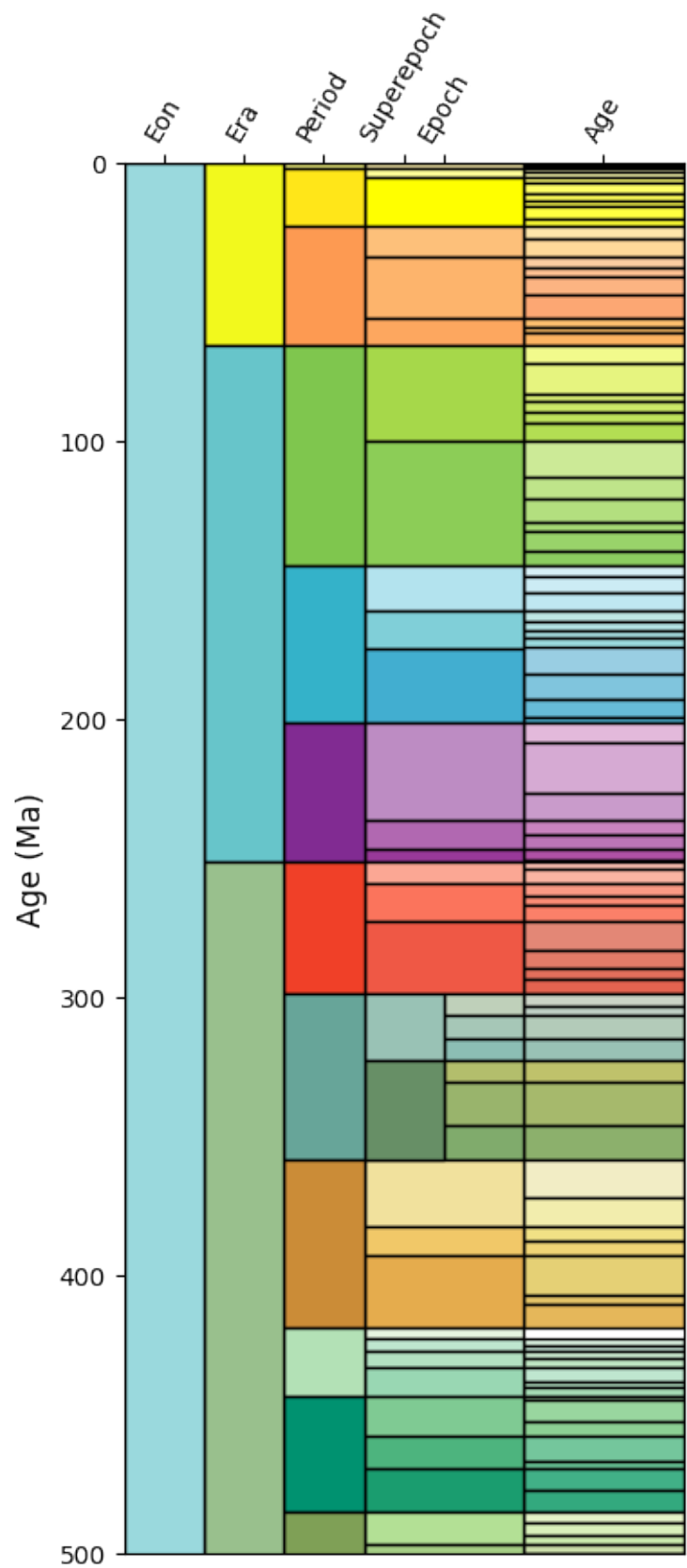
# first let's set up some x-limits for the different timescale levels
xlims = {
    "Eon": (0, 1),
    "Era": (1, 2),
    "Period": (2, 3),
    "Superepoch": (3, 4),
    "Epoch": (3, 5),
    "Age": (5, 7),
}

fig, ax = plt.subplots(1, figsize=(4, 10))

for ix, level in enumerate(ts.levels[::-1]):
    ldf = ts.data.loc[ts.data.Level == level, :]
    for pix, period in ldf.iterrows():
        left, right = xlims[level]
        if ix != len(ts.levels) - 1:
            time = np.mean(ts.text2age(period.Name))
            general = None
            _ix = ix
            while general is None:
                try:
                    general = ts.named_age(time, level=ts.levels[::-1][_ix + 1])
                except:
                    pass
                _ix += 1
            _l, _r = xlims[ts.levels[::-1][_ix]]
            if _r > left:
                left = _r

        rect = Rectangle(
            (left, period.End),
            right - left,
            period.Start - period.End,
            facecolor=period.Color,
            edgecolor="k",
        )
        ax.add_artist(rect)

ax.set_xticks([np.mean(xlims[lvl]) for lvl in ts.levels])
ax.set_xticklabels(ts.levels, rotation=60)
ax.xaxis.set_ticks_position("top")
ax.set_xlim(0, 7)
ax.set_ylabel("Age (Ma)")
ax.set_ylim(500, 0)
```



```
(500.0, 0.0)
```

**Total running time of the script:** (0 minutes 2.023 seconds)

## 3.4 Tutorials

This page is home to longer examples which incorporate multiple components of `pyrolite` or provide examples of how to integrate these into your own workflows.

---

**Note:** This page is a work in progress. Feel free to request tutorials or examples with a feature request.

---

### 3.4.1 Formatting and Cleaning Up Plots

---

**Note:** This tutorial is a work in progress and will be gradually updated.

---

In this tutorial we will illustrate some straightforward formatting for your plots which will allow for greater customisation as needed. As `pyrolite` heavily uses and exposes the API of `matplotlib` for the visualisation components (and also `mpltern` for ternary diagrams), you should also check out their documentation pages for more in-depth guides, examples and API documentation.

First let's pull in a simple dataset to use throughout these examples:

```
from pyrolite.util.synthetic import normal_frame

df = normal_frame(columns=["SiO2", "CaO", "MgO", "Al2O3", "TiO2", "27Al", "d11B"])
```

#### Basic Figure and Axes Settings

`matplotlib` makes it relatively straightforward to customise most settings for your figures and axes. These settings can be defined at creation (e.g. in a call to `subplots()`), or they can be defined after you've created an axis (with the methods `ax.set_<parameter>()`). For example:

```
import matplotlib.pyplot as plt

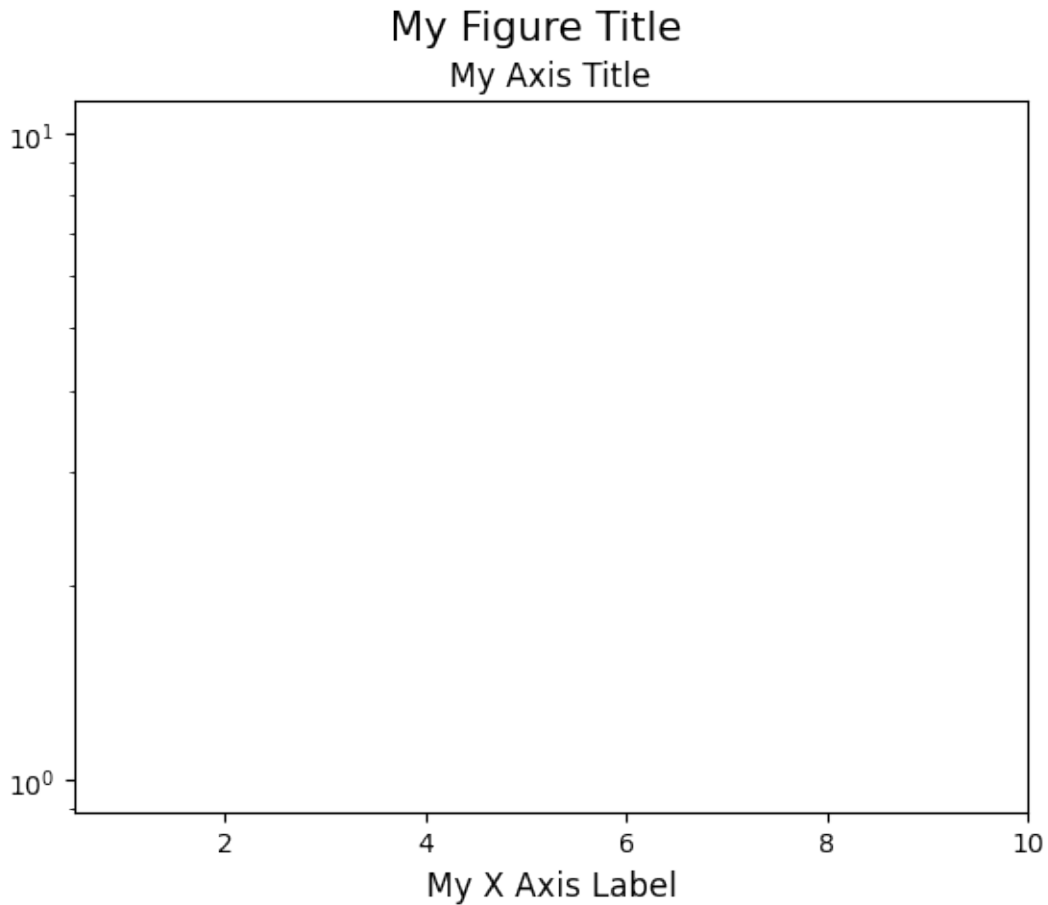
fig, ax = plt.subplots(1)

ax.set_xlabel("My X Axis Label")
ax.set_title("My Axis Title", fontsize=12)
ax.set_yscale("log")
ax.set_xlim((0.5, 10))

fig.suptitle("My Figure Title", fontsize=15)

plt.show()
```

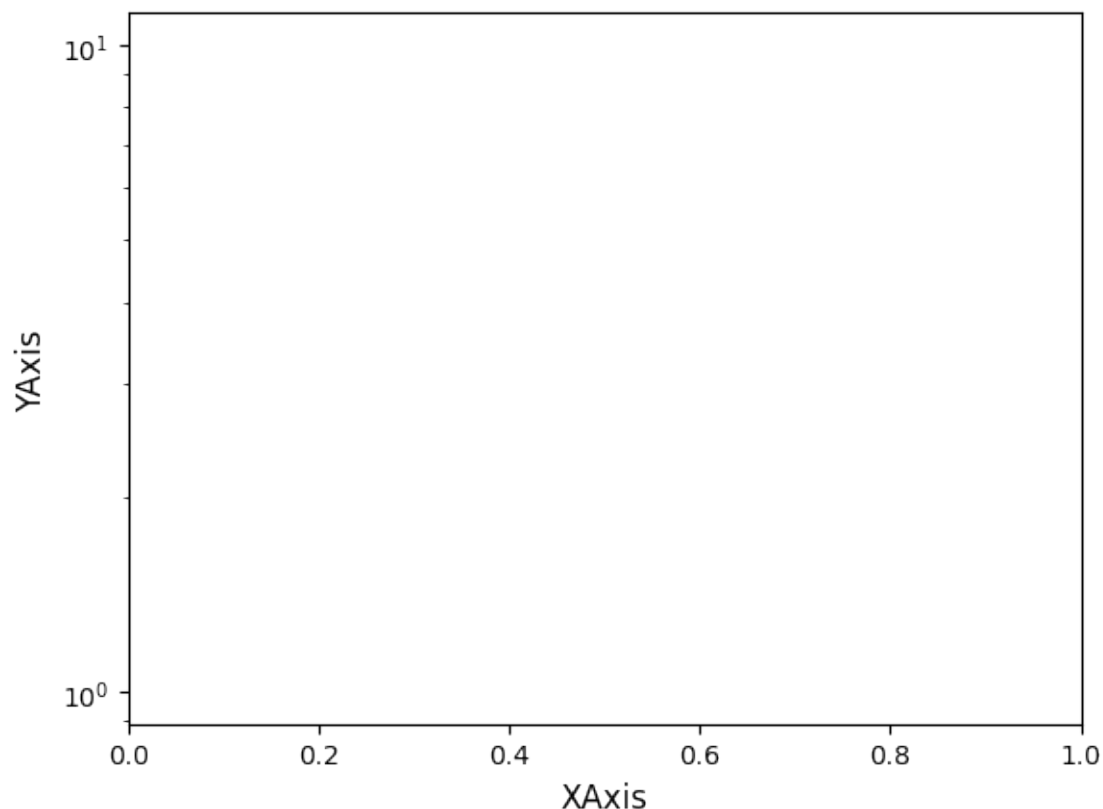




You can use a single method to set most of these things: `set()`. For example:

```
import matplotlib.pyplot as plt

fig, ax = plt.subplots(1)
ax.set(yscale="log", xlim=(0, 1), ylabel="YAxis", xlabel="XAxis")
plt.show()
```



## Labels and Text

`matplotlib` enables you to use  $\text{T}_{\text{E}}\text{X}$  within all text elements, including labels and annotations. This can be leveraged for more complex formatting, incorporating math and symbols into your plots. Check out the [mod:matplotlib tutorial](#), and for more on working with text generally in `matplotlib`, check out the [relevant tutorials gallery](#).

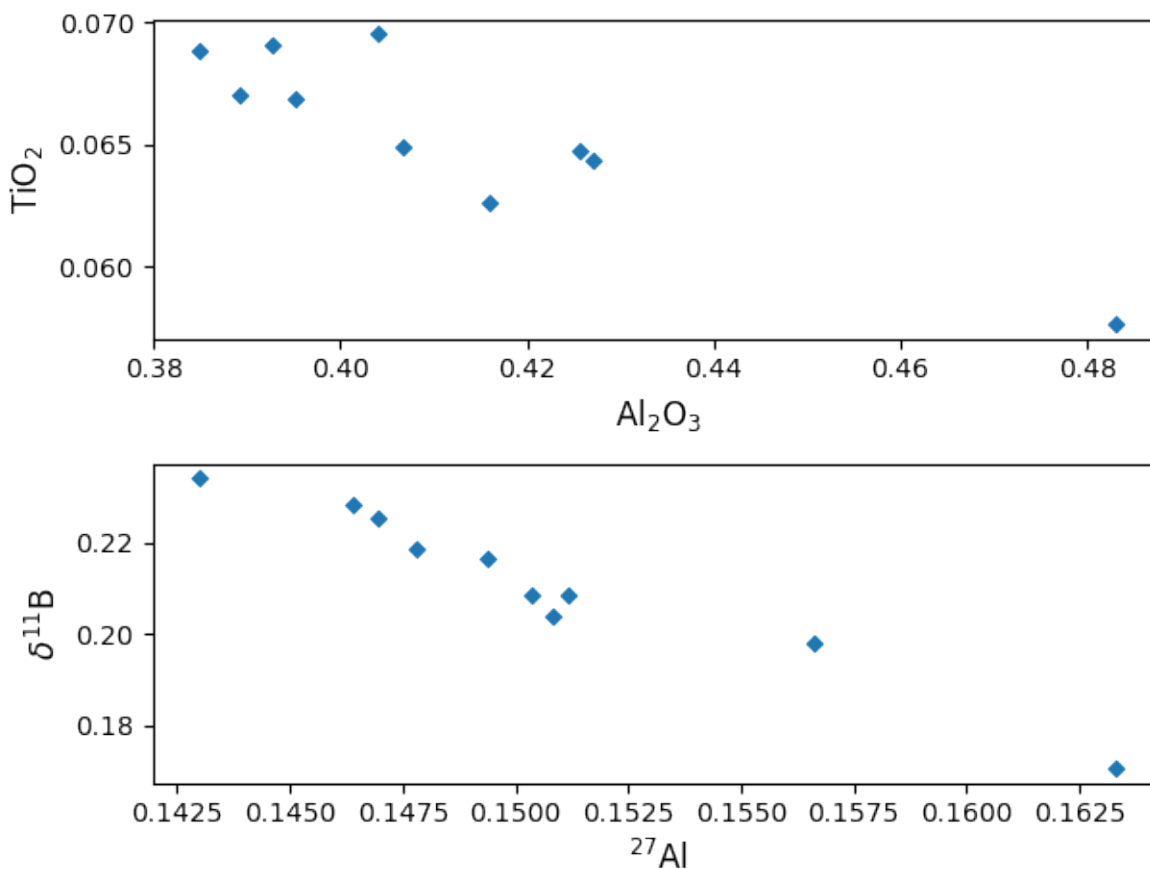
The ability to use TeX syntax in `matplotlib` text objects can also be used for typesetting, like for subscripts and superscripts. This is particularly relevant for geochemical oxides labels (e.g.  $\text{Al}_2\text{O}_3$ , which would ideally be rendered as  $\text{Al}_2\text{O}_3$ ) and isotopes (e.g.  $\text{d}_{11}\text{B}$ , which should be  $\delta^{11}\text{B}$ ). At the moment, `pyrolite` won't do this for you, so you may want to adjust the labelling after you've made them. For example:

```
import pyrolite.plot
import matplotlib.pyplot as plt

fig, ax = plt.subplots(2, 1)
df[["Al2O3", "TiO2"]].pyroplot.scatter(ax=ax[0])
ax[0].set_xlabel("Al$_2$O$_3$")
ax[0].set_ylabel("TiO$_2$")

df[["27Al", "d11B"]].pyroplot.scatter(ax=ax[1])
ax[1].set_xlabel("$^{27}$Al")
ax[1].set_ylabel("$\delta^{11}$B")

plt.tight_layout() # rearrange the plots to fit nicely together
plt.show()
```

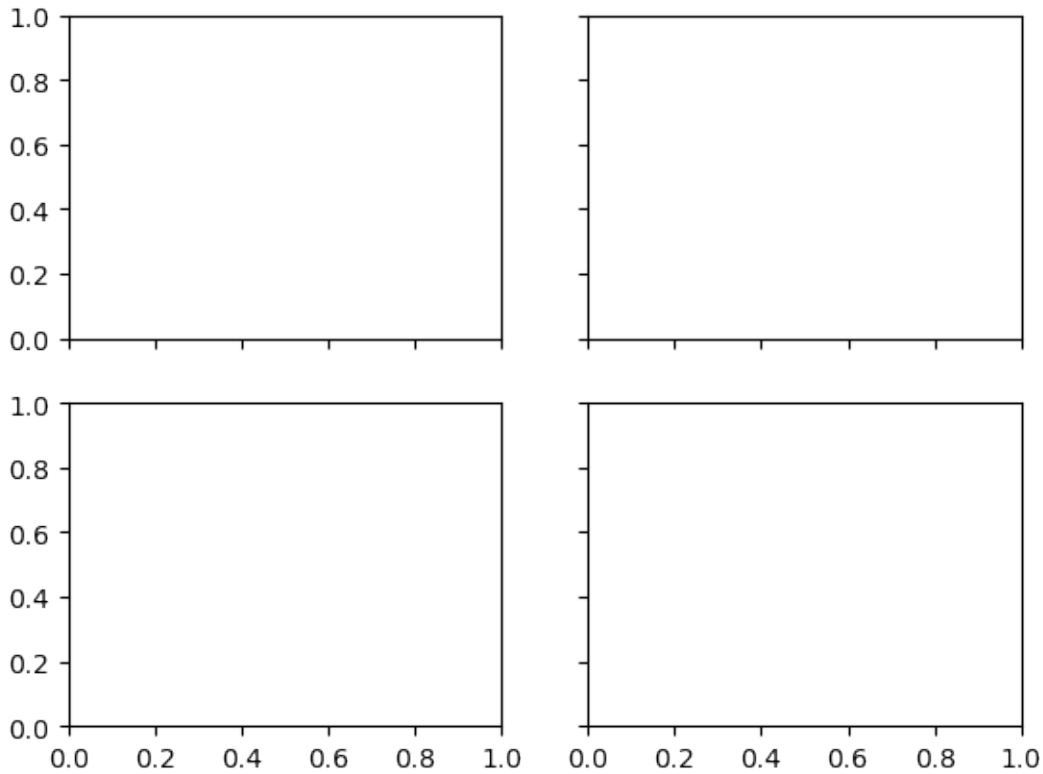


## Sharing Axes

If you're building figures which have variables which are re-used, you'll typically want to 'share' them between your axes. The `matplotlib.pyplot` API makes this easy for when you want to share among *all* the axes as you create them:

```
import matplotlib.pyplot as plt

fig, ax = plt.subplots(2, 2, sharex=True, sharey=True)
plt.show()
```

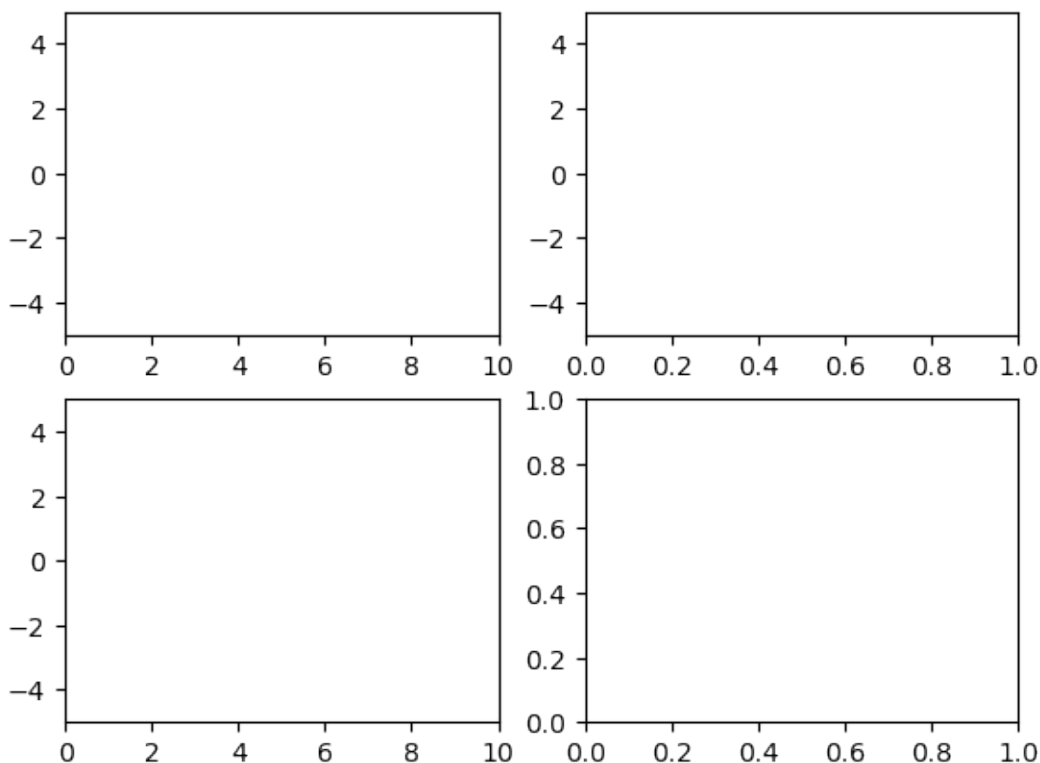


However, if you want to share axes in a way which is less standard, it can be difficult to set up using this function. `pyrolite` has a utility function which can be used to share axes after they're created in slightly more arbitrary ways. For example, imagine we wanted to share the first and third x-axes, and the first three y-axes, you could use:

```
import matplotlib.pyplot as plt
from pyrolite.util.plot.axes import share_axes

fig, ax = plt.subplots(2, 2)
ax = ax.flat # turn the (2,2) array of axes into one flat axes with shape (4,)
share_axes([ax[0], ax[2]], which="x") # share x-axes for 0, 2
share_axes(ax[0:3], which="y") # share y-axes for 0, 1, 2

ax[0].set_xlim((0, 10))
ax[1].set_ylim((-5, 5))
plt.show()
```

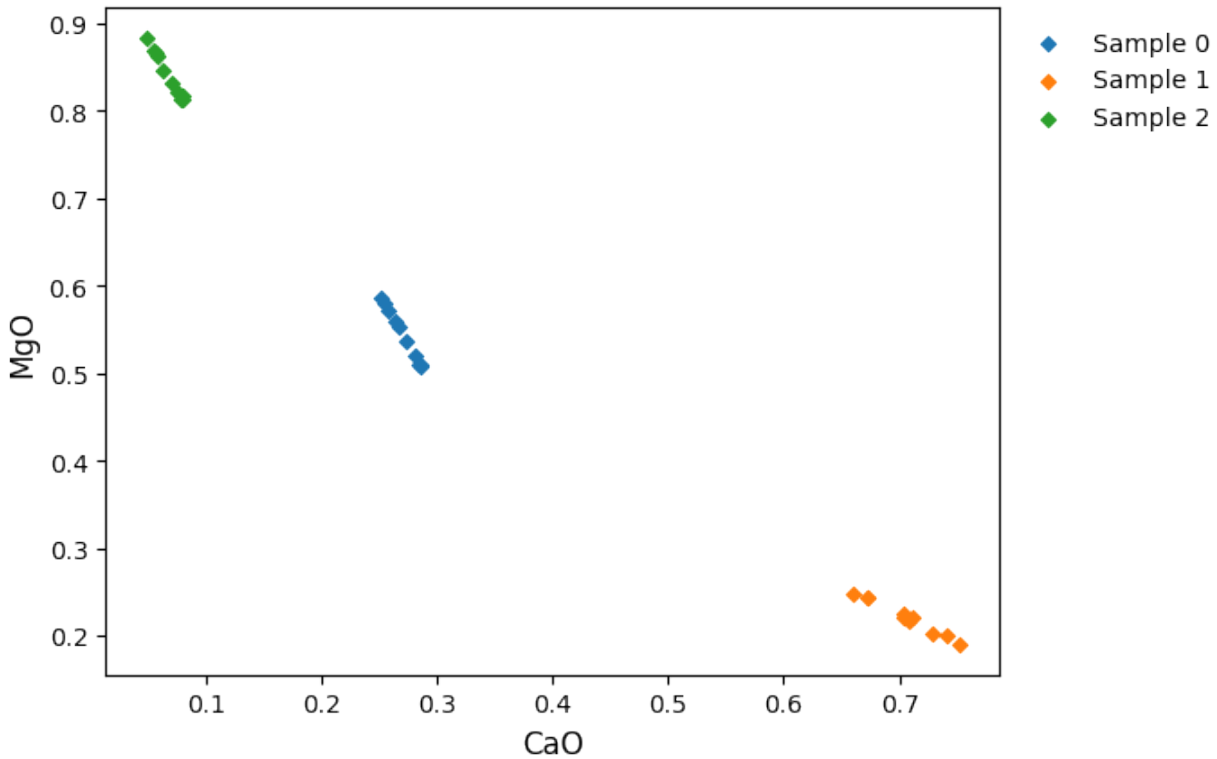


## Legends

While it's simple to set up basic legends in `matplotlib` (see the docs for `matplotlib.axes.Axes.legend()`), often you'll want to customise your legends to fit nicely within your figures. Here we'll create a few synthetic datasets, add them to a figure and create the default legend:

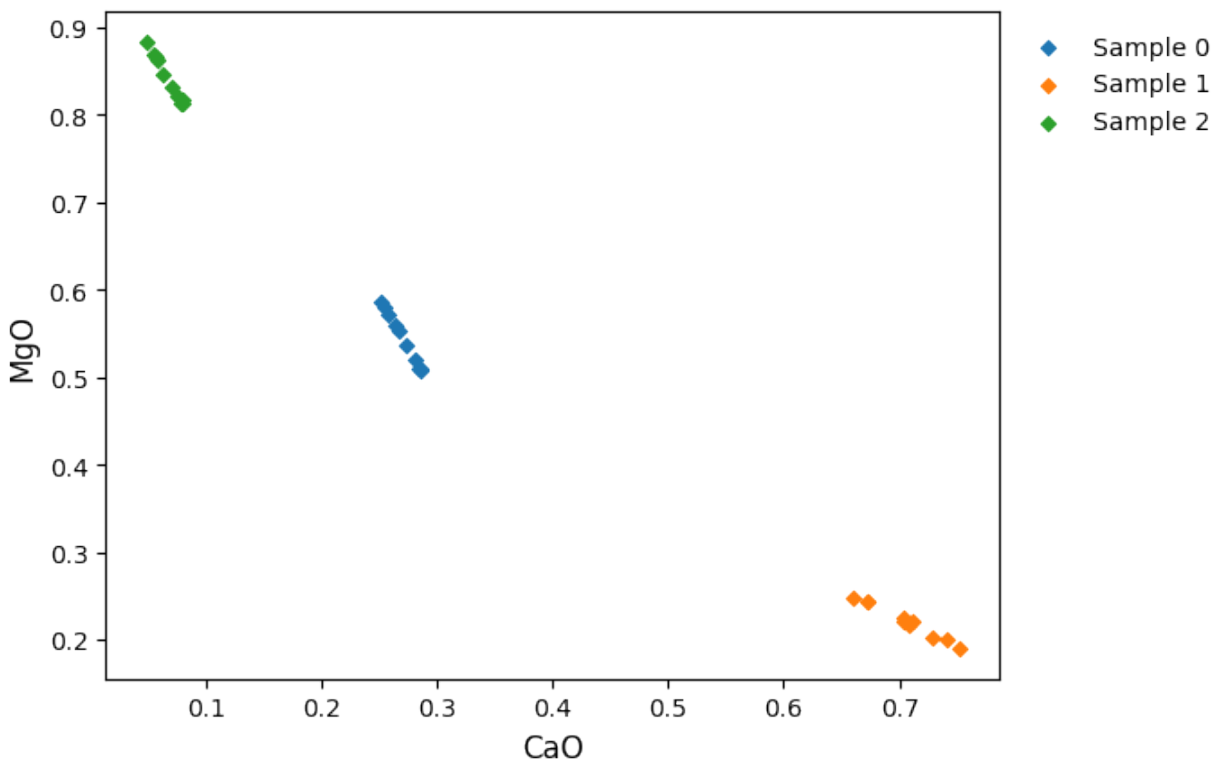
```
import pyrolite.plot
import matplotlib.pyplot as plt

fig, ax = plt.subplots(1)
for i in range(3):
    sample_data = normal_frame(columns=["CaO", "MgO", "FeO"]) # a new random sample
    sample_data[["CaO", "MgO"]].pyroplot.scatter(ax=ax, label="Sample {:d}".format(i))
ax.legend()
plt.show()
```



On many of the `pyrolite` examples, you'll find legends formatted along the lines of the following to clean them up a little (these are the default styles):

```
ax.legend(  
    facecolor=None, # have a transparent legend background  
    frameon=False, # remove the legend frame  
    bbox_to_anchor=(1, 1), # anchor legend's corner to the axes' top-right  
    loc="upper left", # use the upper left corner for the anchor  
)  
plt.show()
```



Check out the [matplotlib legend guide](#) for more.

## Ternary Plots

The ternary plots in `pyrolite` are generated using `mpltern`, and while the syntax is very similar to the `matplotlib` API, as we have three axes to deal with sometimes things are little different. Here we demonstrate how to complete some common tasks, but you should check out the `mpltern` documentation if you want to dig deeper into customising your ternary diagrams (e.g. see the [example gallery](#)), which these examples were developed from.

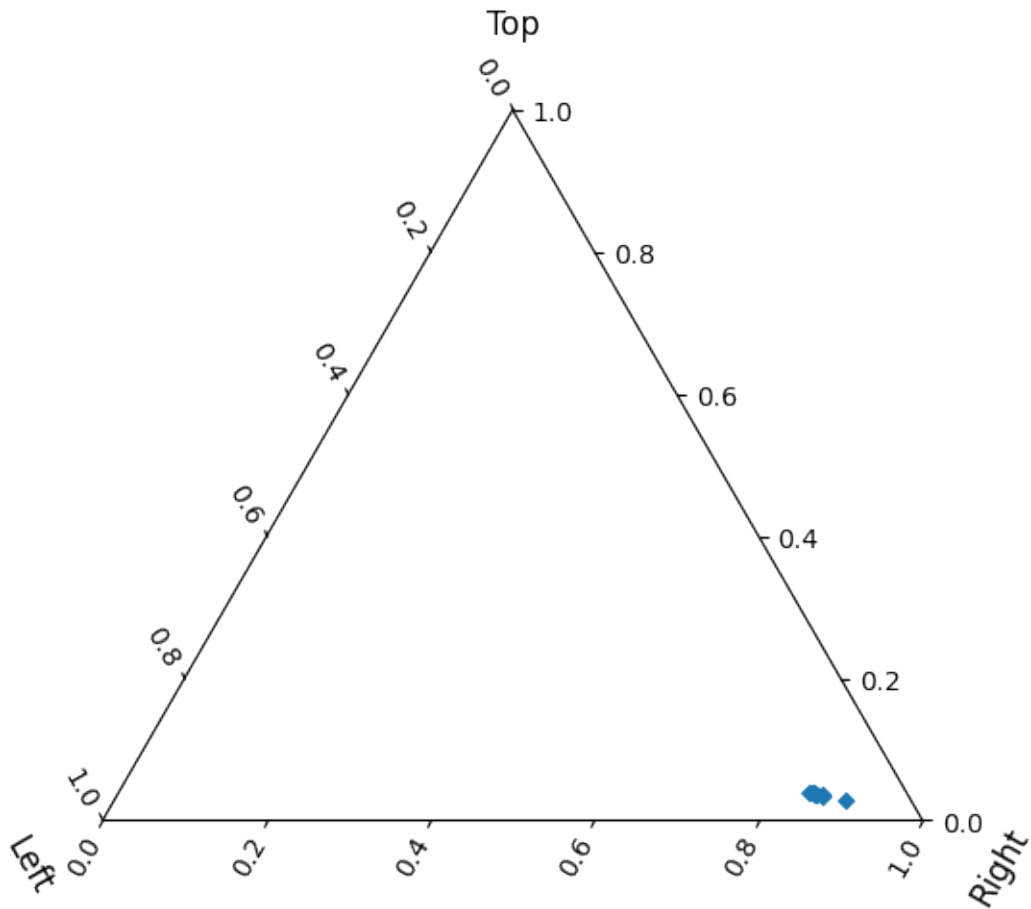
One of the key things to note in `mpltern` is that you have *top*, *left* and *right* axes.

### Ternary Plot Axes Labels

Labelling ternary axes is done similarly to in `matplotlib`, but using the axes prefixes *t*, *l* and *r* for top, left and right axes, respectively:

```
import pyrolite.plot
import matplotlib.pyplot as plt

ax = df[["CaO", "MgO", "Al2O3"]].pyroplot.scatter()
ax.set_tlabel("Top")
ax.set_llabel("Left")
ax.set_rlabel("Right")
plt.show()
```



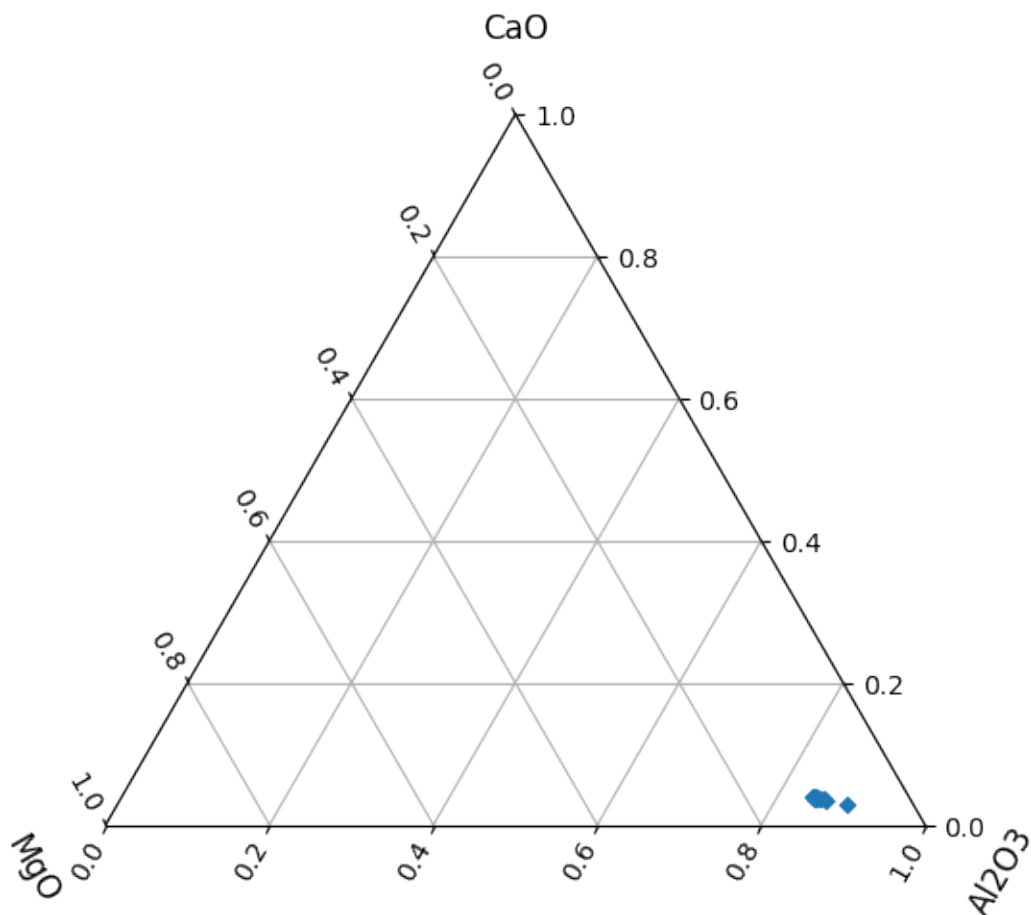
### Ternary Plot Grids

To add a simple grid to your ternary plot, you can use `grid()`:

```
import pyrolite.plot
import matplotlib.pyplot as plt

ax = df[["CaO", "MgO", "Al2O3"]].pyroplot.scatter()
ax.grid()
plt.show()
```

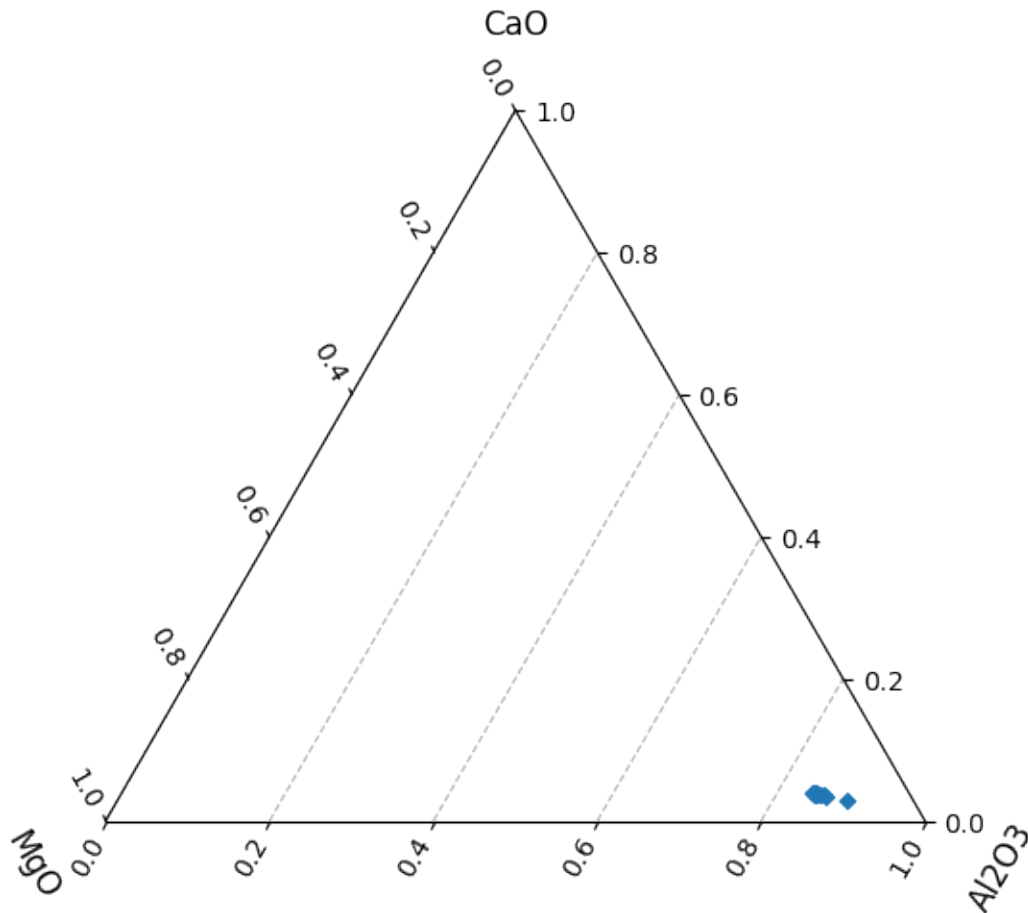




With this method, you can also specify an *axis*, which tickmarks you want to use for the grid ('major', 'minor' or 'both') and a *linestyle*:

```
import pyrolite.plot
import matplotlib.pyplot as plt

ax = df[["CaO", "MgO", "Al2O3"]].pyroplot.scatter()
ax.grid(axis="r", which="both", linestyle="--")
plt.show()
```



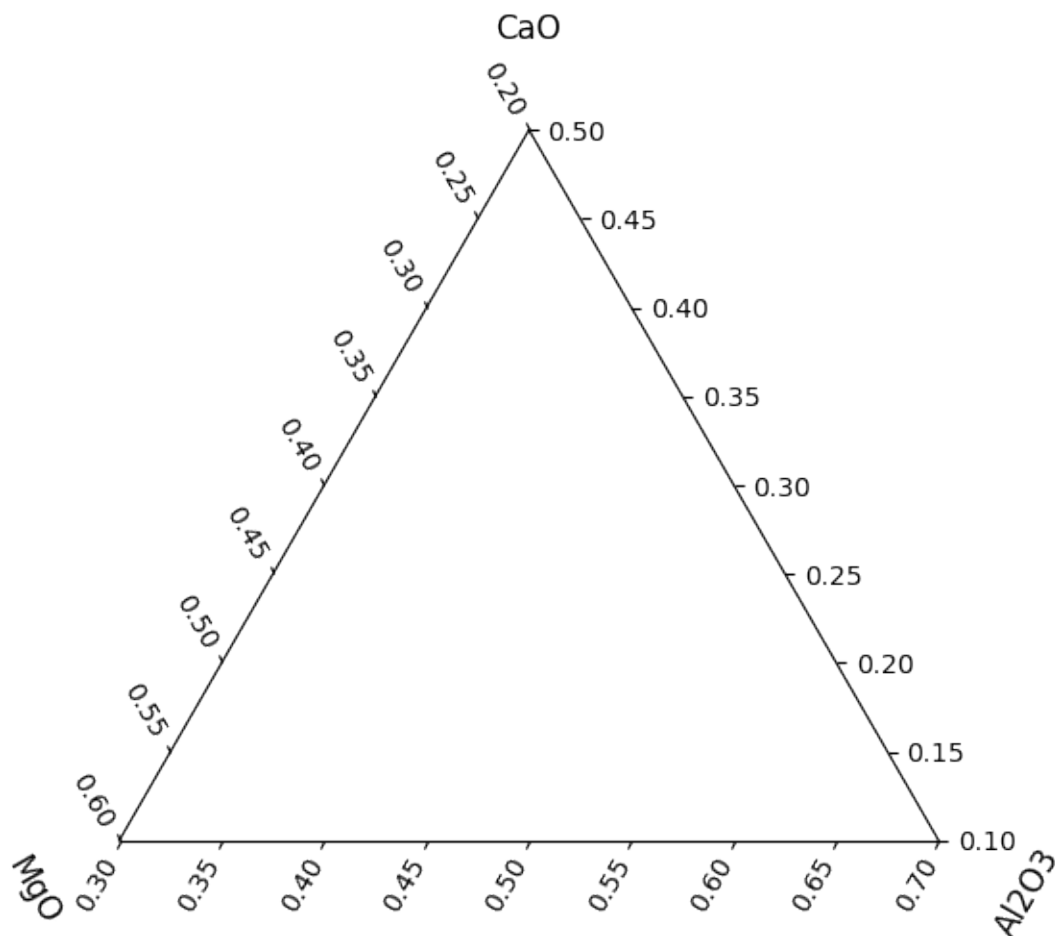
### Ternary Plot Limits

To focus on a specific area, you can reset the limits of your ternary axes with `set_ternary_lim()`.

Also check out the [mpltern inset axes example](#) if you're after ways to focus on specific regions.

```
import pyrolite.plot
import matplotlib.pyplot as plt

ax = df[["CaO", "MgO", "Al2O3"]].pyroplot.scatter()
ax.set_ternary_lim(
    0.1, 0.5, 0.2, 0.6, 0.3, 0.7 # tmin # tmax # lmin # lmax # rmin # rmax
)
plt.show()
```



Total running time of the script: (0 minutes 16.190 seconds)

### 3.4.2 Making the Logo

Having some funky ellipses in a simplex inspired some interest when I put the logo together for pyrolite, so I put together a cleaned-up example of how you can create these kinds of plots for your own data. These examples illustrate different methods to show distribution of (homogeneous, or near so) compositional data for exploratory analysis.

```
import matplotlib
import matplotlib.cm
import matplotlib.colors
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

import pyrolite.plot
from pyrolite.comp.codata import *
from pyrolite.util.plot.helpers import plot_pca_vectors, plot_stdev_ellipses
from pyrolite.util.skl.transform import ALRTransform, ILRTransform
from pyrolite.util.synthetic import random_composition
```

(continues on next page)

(continued from previous page)

```

np.random.seed(82)

# ignore sphinx_gallery warnings
import warnings

warnings.filterwarnings("ignore", "Matplotlib is currently using agg")

```

First we choose some colors, create some log-distributed synthetic data. Here I've generated a synthetic dataset with four samples having means equidistant from the log-space centre and with varying covariance. This should illustrate the spatial warping of the simplex nicely. Additionally, I chose a log-transform here to go from and to compositional space (ILRTransform, which uses the isometric log-ratio function `ilr()`). Choosing another transform will change the distortion observed in the simplex slightly. This synthetic dataset is added into a `DataFrame` for convenient access to plotting functions via the pandas API defined in `pyrolite.plot.pyroplot`.

```

t10b3 = [ # tableau 10 colorblind safe colors, a selection of 4
    (0, 107, 164),
    (171, 171, 171),
    (89, 89, 89),
    (95, 158, 209),
]
t10b3 = [(r / 255.0, g / 255.0, b / 255.0) for r, g, b in t10b3]

d = 1.0 # distance from centre
sig = 0.1 # scale for variance
# means for logspace (D=2)
means = np.array(np.meshgrid([-1, 1], [-1, 1])).T.reshape(-1, 2) * d
# means = np.array([(-d, -d), (d, -d), (-d, d), (d, d)])
covs = ( # covariance for logspace (D=2)
    np.array(
        [
            [[1, 0], [0, 1]],
            [[0.5, 0.15], [0.15, 0.5]],
            [[1.5, -1], [-1, 1.5]],
            [[1.2, -0.6], [-0.6, 1.2]],
        ]
    )
    * sig
)

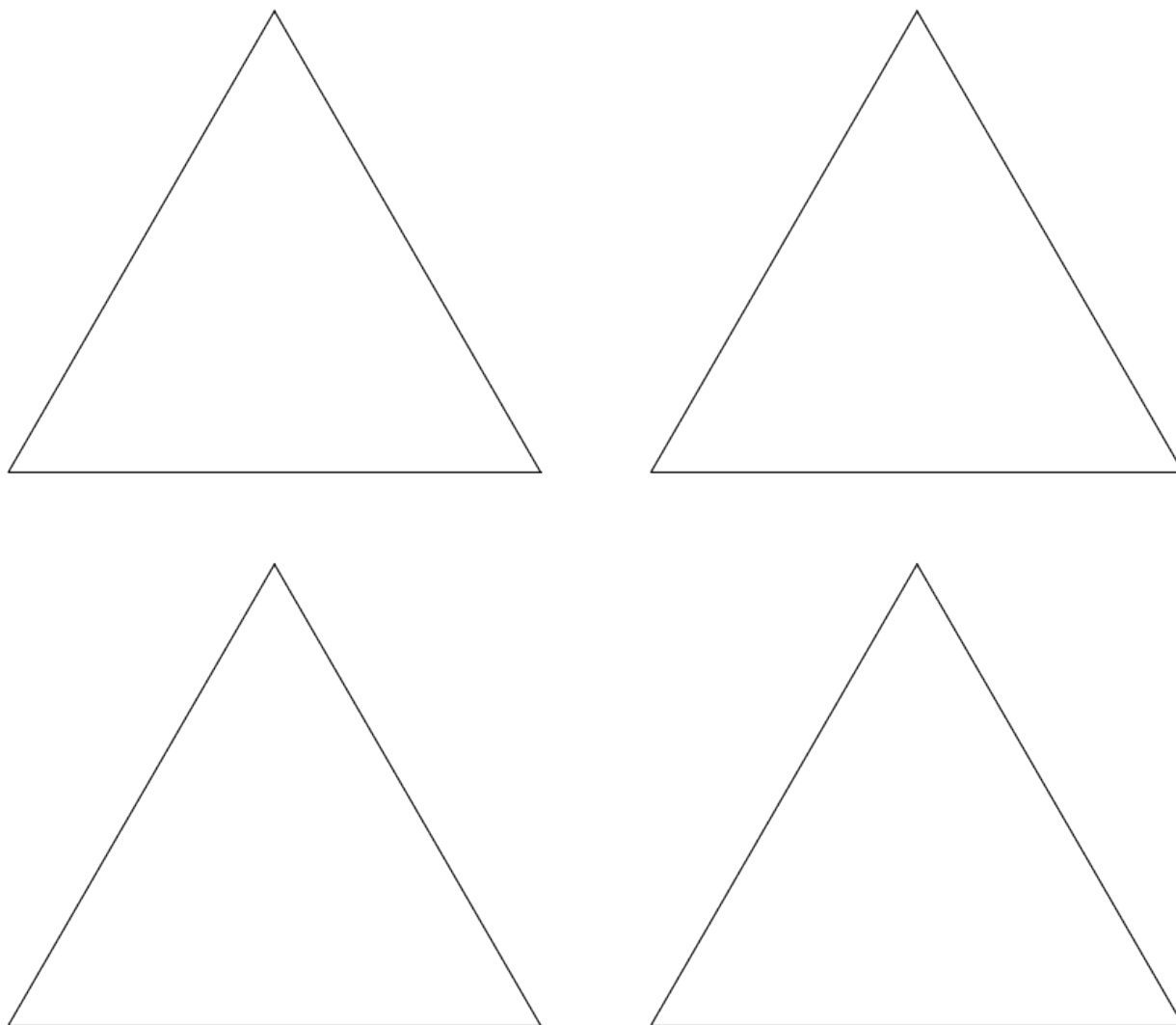
means = ILRTransform().inverse_transform(means) # compositional means (D=3)
size = 2000 # logo @ 10000
pts = [random_composition(mean=M, cov=C, size=size) for M, C in zip(means, covs)]

T = ILRTransform()
to_log = T.transform
from_log = T.inverse_transform

df = pd.DataFrame(np.vstack(pts))
df.columns = ["SiO2", "MgO", "FeO"]
df["Sample"] = np.repeat(np.arange(df.columns.size + 1), size).flatten()
chem = ["MgO", "SiO2", "FeO"]

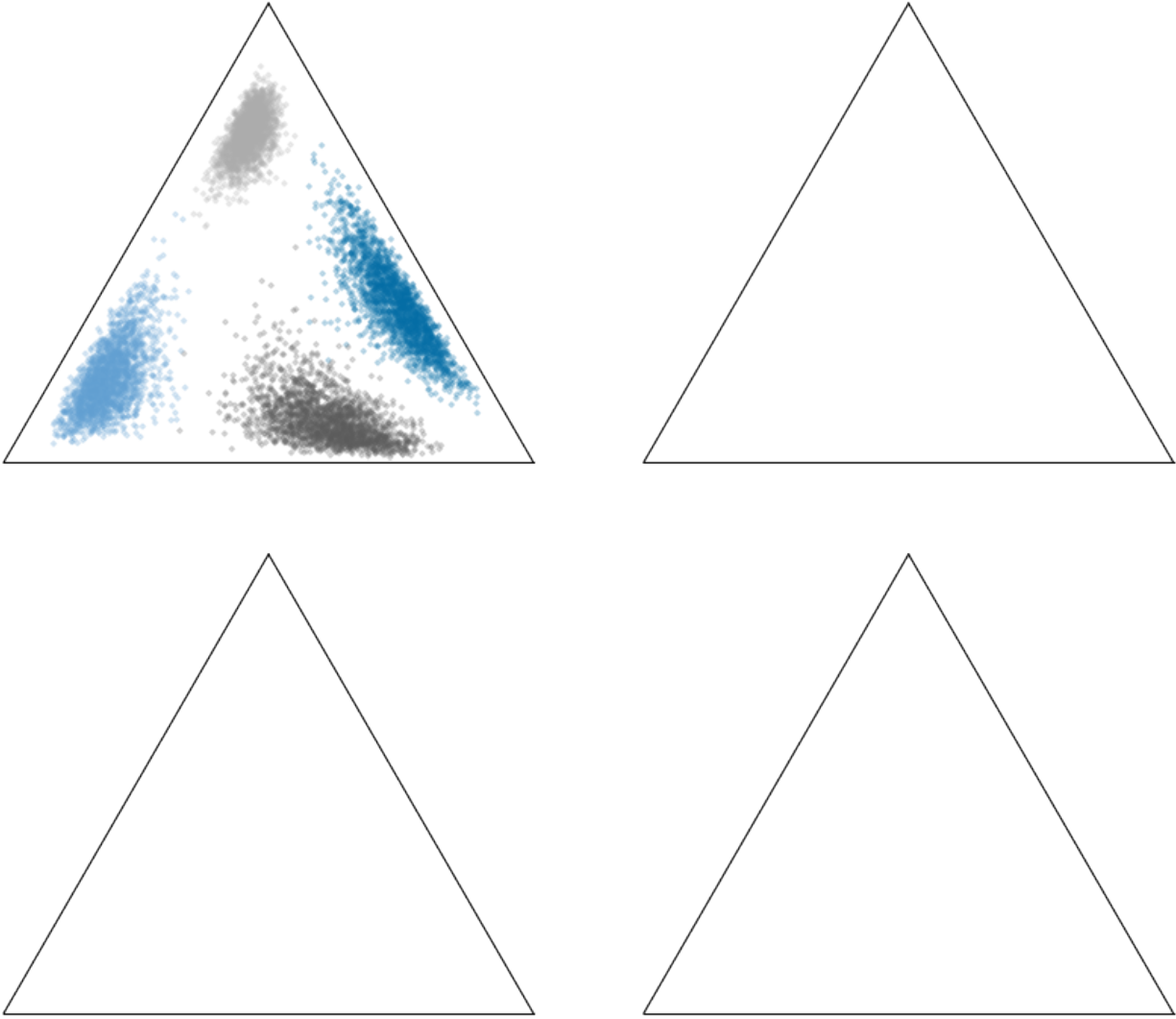
```

```
fig, ax = plt.subplots(
    2, 2, figsize=(10, 10 * np.sqrt(3) / 2), subplot_kw=dict(projection="ternary")
)
ax = ax.flat
_ = [[x.set_ticks([]) for x in [a.taxis, a.laxis, a.raxis]] for a in ax]
```



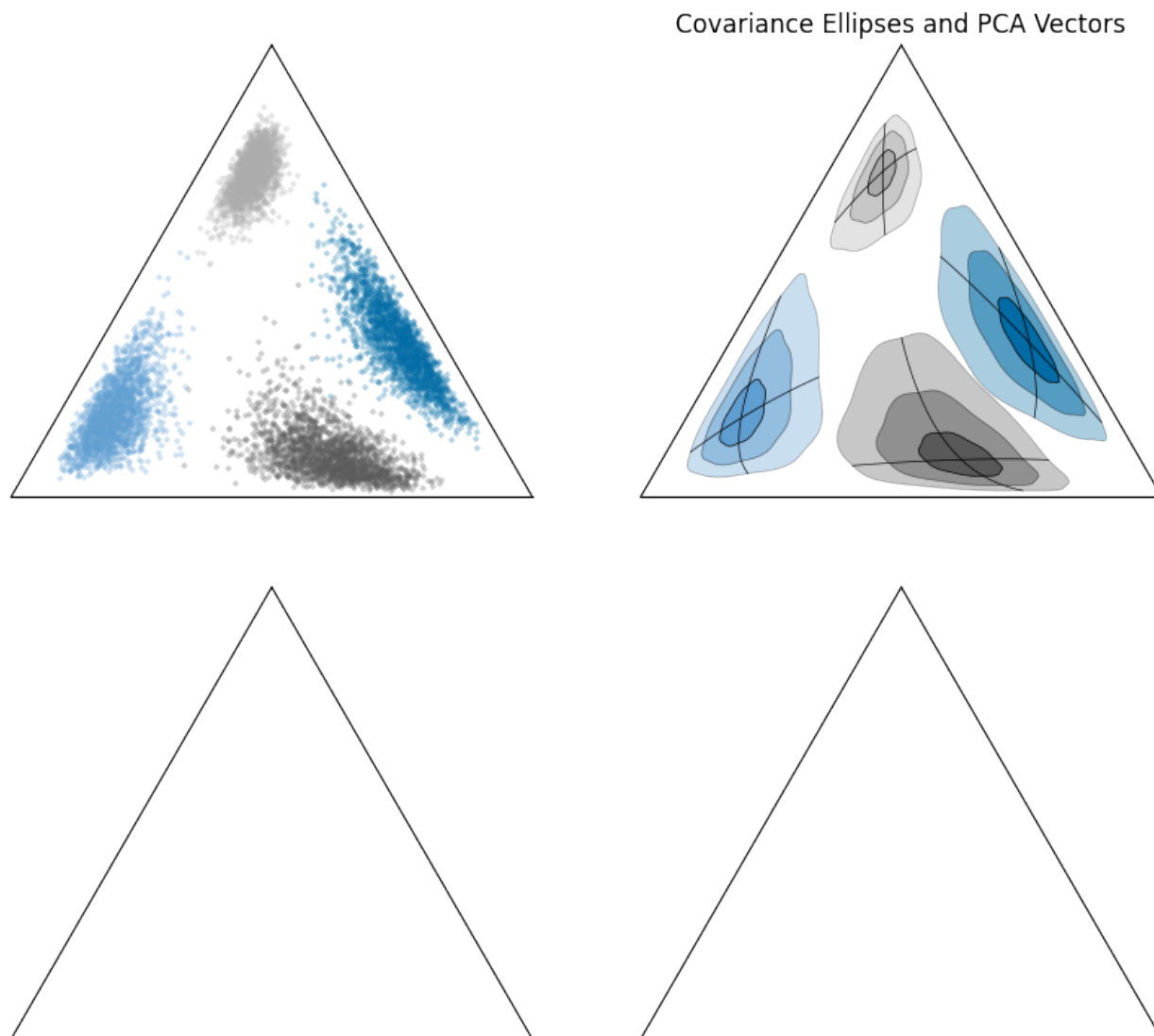
First, let's look at the synthetic data itself in the ternary space:

```
kwargs = dict(alpha=0.2, s=3, no_ticks=True, axlabels=False)
for ix, sample in enumerate(df.Sample.unique()):
    comp = df.query("Sample == {}".format(sample))
    comp.loc[:, chem].pyroplot.scatter(ax=ax[ix], c=t10b3[ix], **kwargs)
plt.show()
```



We can take the mean and covariance in log-space to create covariance ellipses and vectors using principal component analysis:

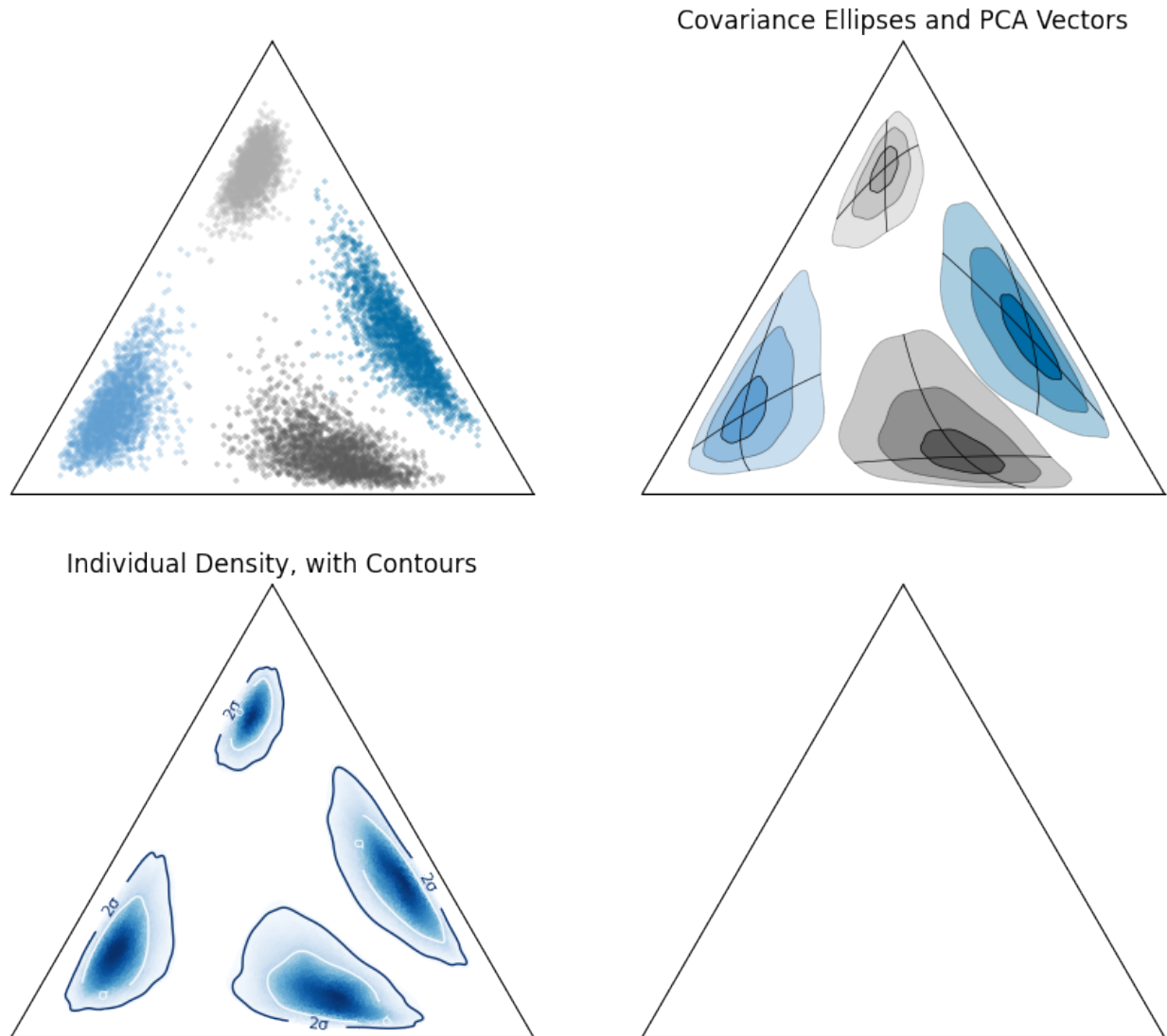
```
kwargs = dict(ax=ax[1], transform=from_log, nstds=3)
ax[1].set_title("Covariance Ellipses and PCA Vectors")
for ix, sample in enumerate(df.Sample.unique()):
    comp = df.query("Sample == {}".format(sample))
    tcomp = to_log(comp.loc[:, chem])
    plot_stdev_ellipses(tcomp.values, color=t10b3[ix], resolution=1000, **kwargs)
    plot_pca_vectors(tcomp.values, ls="--", lw=0.5, color="k", **kwargs)
plt.show()
```



We can also look at data density (here using kernel density estimation) in logratio-space:

```
kwargs = dict(ax=ax[-2], bins=100, axlabels=False)
ax[-2].set_title("Individual Density, with Contours")

for ix, sample in enumerate(df.Sample.unique()):
    comp = df.query("Sample == {}".format(sample))
    comp.loc[:, chem].pyroplot.density(cmap="Blues", vmin=0.05, **kwargs)
    comp.loc[:, chem].pyroplot.density(
        contours=[0.68, 0.95],
        cmap="Blues_r",
        contour_labels={0.68: "", 0.95: "2"},
        **kwargs,
    )
plt.show()
```

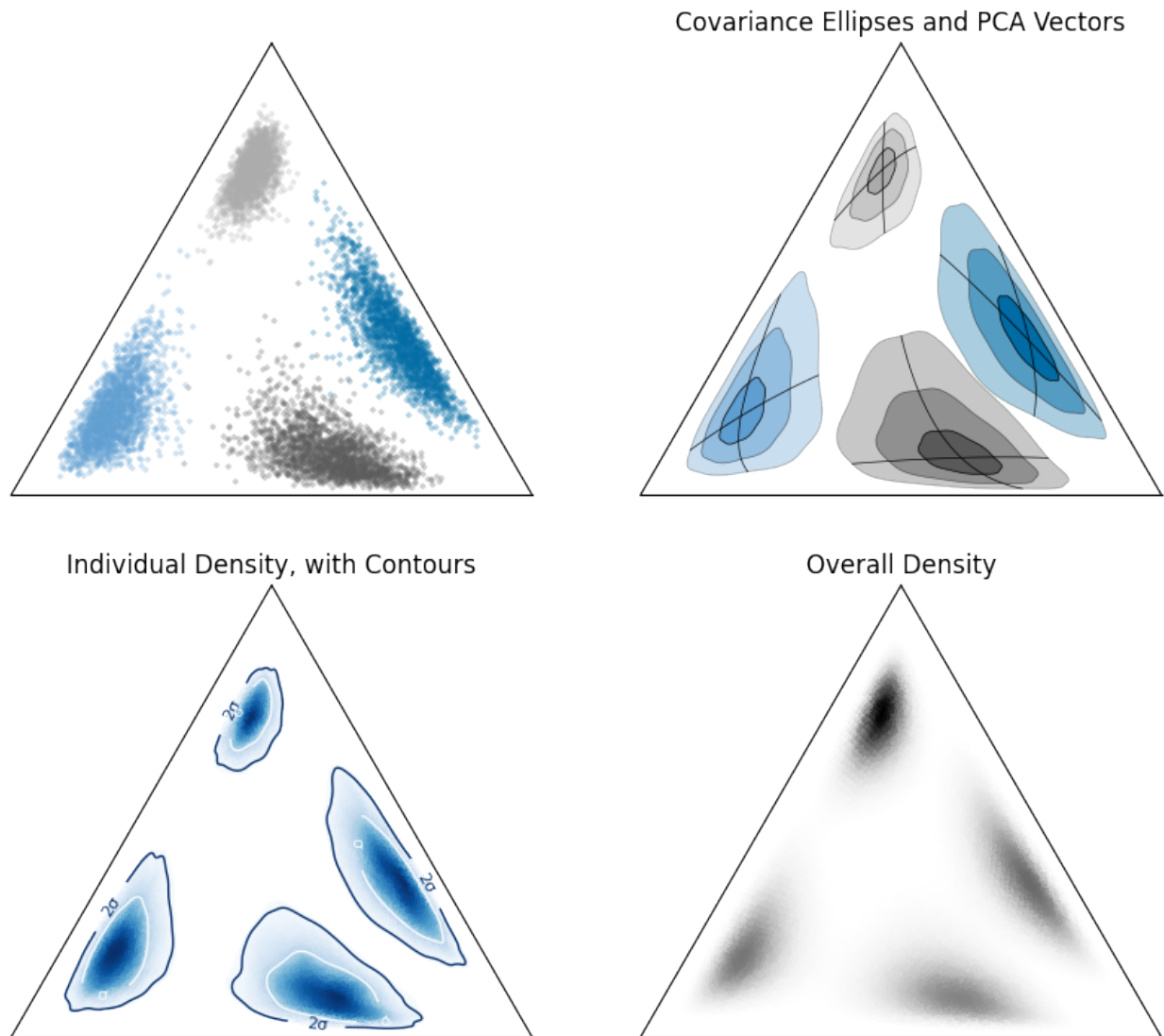


```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/  
→comp/codata.py:177: RuntimeWarning: invalid value encountered in log  
    Y = np.log(X) # Log operation  
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/  
→comp/codata.py:177: RuntimeWarning: invalid value encountered in log  
    Y = np.log(X) # Log operation
```

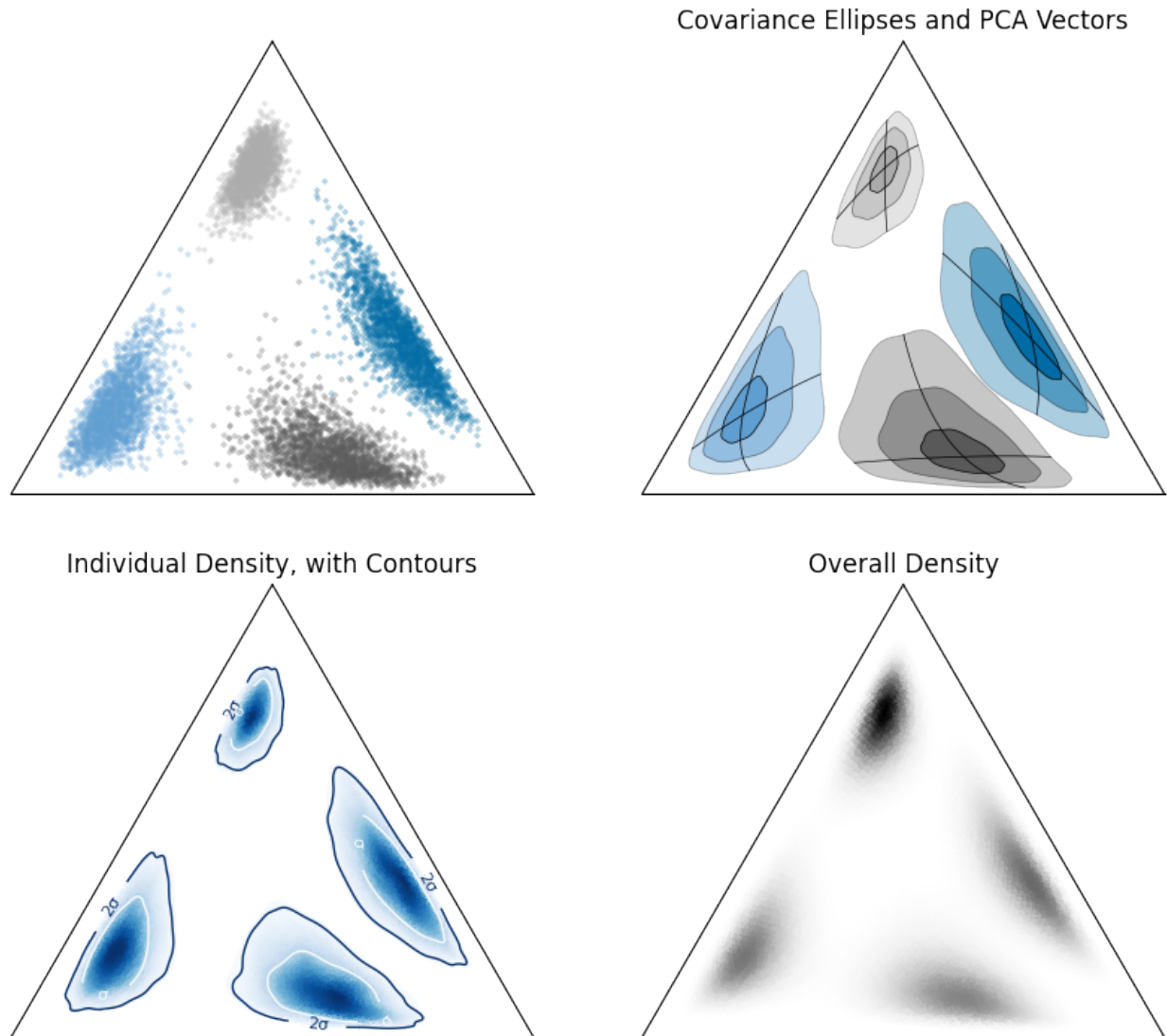
We can also do this for individual samples, and estimate percentile contours:

```
kwargs = dict(ax=ax[-1], axlabels=False)  
ax[-1].set_title("Overall Density")  
df.loc[:, chem].pyroplot.density(bins=100, cmap="Greys", **kwargs)  
plt.show()
```





```
for a in ax:  
    a.set_aspect("equal")  
    a.patch.set_visible(False)  
plt.show()
```



**Total running time of the script:** (0 minutes 11.255 seconds)

### 3.4.3 One Way to Do Ternary Heatmaps

There are multiple ways you can achieve ternary heatmaps, but those based on the cartesian axes (e.g. a regularly spaced rectangular grid, or even a regularly spaced triangular grid) can result in difficulties and data misrepresentation.

Here we illustrate how the ternary heatmaps for pyrolite are constructed using an irregular triangulated grid and log transforms, and how this avoids some of the potential issues of the methods mentioned above.

Let's first get some data to deal with. `mpltern` has a convenient dataset which we can use here:

```
import numpy as np
import pandas as pd
from mpltern.ternary.datasets import get_scatter_points

np.random.seed(43)
```

(continues on next page)

(continued from previous page)

```
df = pd.DataFrame(np.array([*get_scatter_points(n=80)]).T, columns=["A", "B", "C"])
df = df.loc[(df > 0.1).all(axis=1), :]
```

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/envs/develop/lib/python3.10/
↳ site-packages/mpltern/ternary/datasets.py:9: UserWarning: `mpltern.ternary.datasets.
↳ py` has been moved to `mpltern.datasets.py` and will be removed from the present
↳ directory in mpltern 0.6.0.
warnings.warn(msg)
```

From this dataset we'll generate a `ternary_heatmap()`, which is the basis for ternary density diagrams via `density()`:

```
from pyrolite.comp.codata import ILR, inverse_ILR
from pyrolite.plot.density.ternary import ternary_heatmap

coords, H, data = ternary_heatmap(
    df.values,
    bins=10,
    mode="density",
    remove_background=True,
    transform=ILR,
    inverse_transform=inverse_ILR,
    grid_border_frac=0.2,
)
```

This function returns more than just the coordinates and histogram/density estimate, which will come in handy for exploring how it came together. The data variable here is a dictionary with contains the grids and coordinates used to construct the histogram/density diagram. We can use these to show how the ternary log-grid is constructed, and then transformed back to ternary space before being triangulated and interpolated for the ternary heatmap:

```
import matplotlib.pyplot as plt

import pyrolite.plot
from pyrolite.util.math import flattengrid
from pyrolite.util.plot.axes import axes_to_ternary, share_axes

fig, ax = plt.subplots(3, 2, figsize=(6, 9))
ax = ax.flat

share_axes([ax[1], ax[2], ax[3]])
ax = axes_to_ternary([ax[0], ax[4], ax[5]])

ax[0].set_title("data", y=1.2)
df.pyroplot.scatter(ax=ax[0], c="k", alpha=0.1)
ax[0].scatter(*data["tern_bound_points"].T, c="k")

ax[1].set_title("transformed data", y=1.2)
ax[1].scatter(*data["tfm_tern_bound_points"].T, c="k")
ax[1].scatter(*data["grid_transform"](df.values).T, c="k", alpha=0.1)

ax[2].set_title("log grid", y=1.2)
ax[2].scatter(*flattengrid(data["tfm_centres"]).T, c="k", marker=".", s=5)
```

(continues on next page)

(continued from previous page)

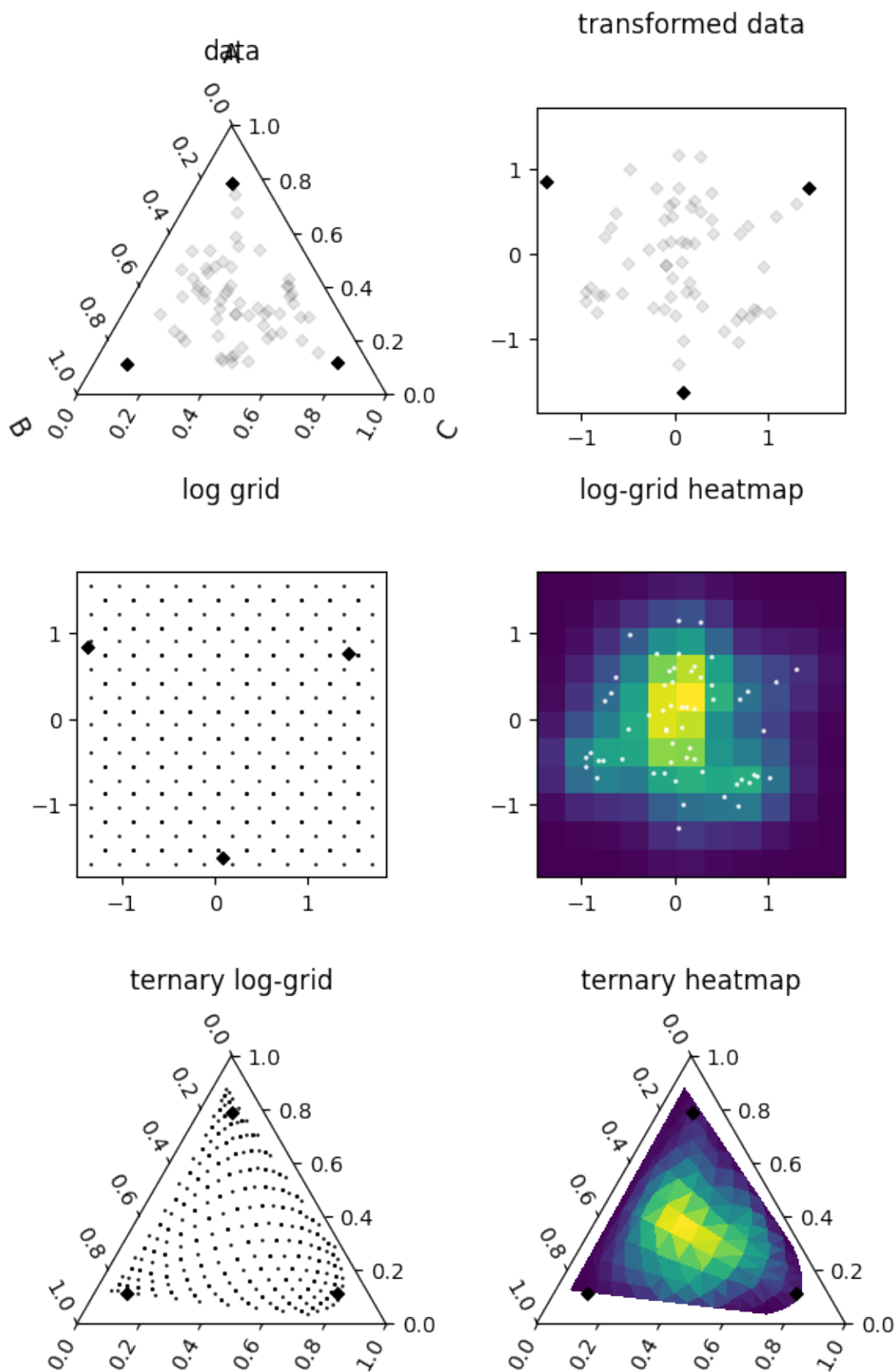
```
ax[2].scatter(*flattengrid(data["tfm_edges"]).T, c="k", marker=".", s=2)
ax[2].scatter(*data["tfm_tern_bound_points"].T, c="k")

ax[3].set_title("log-grid heatmap", y=1.2)
ax[3].pcolormesh(*data["tfm_edges"], H)
ax[3].scatter(*data["grid_transform"](df.values).T, c="white", alpha=0.8, s=1)

ax[4].set_title("ternary log-grid", y=1.2)
ax[4].scatter(*data["tern_centres"].T, c="k", marker=".", s=5)
ax[4].scatter(*data["tern_edges"].T, c="k", marker=".", s=2)
ax[4].scatter(*data["tern_bound_points"].T, c="k")

ax[5].set_title("ternary heatmap", y=1.2)
ax[5].tripcolor(*coords.T, H.flatten())
ax[5].scatter(*data["tern_bound_points"].T, c="k")

plt.tight_layout()
```



```
plt.close("all") # let's save some memory..
```

We can see how this works almost exactly the same for the histograms:

```
coords, H, data = ternary_heatmap(
    df.values,
    bins=10,
    mode="histogram",
    remove_background=True,
    transform=ILR,
    inverse_transform=inverse_ILR,
    grid_border_frac=0.2,
)
```

```
fig, ax = plt.subplots(3, 2, figsize=(6, 9))
ax = ax.flat

share_axes([ax[1], ax[2], ax[3]])
ax = axes_to_ternary([ax[0], ax[4], ax[5]])

ax[0].set_title("data", y=1.2)
df.pyroplot.scatter(ax=ax[0], c="k", alpha=0.1)
ax[0].scatter(*data["tern_bound_points"].T, c="k")

ax[1].set_title("transformed data", y=1.2)
ax[1].scatter(*data["tfm_tern_bound_points"].T, c="k")
ax[1].scatter(*data["grid_transform"](df.values).T, c="k", alpha=0.1)

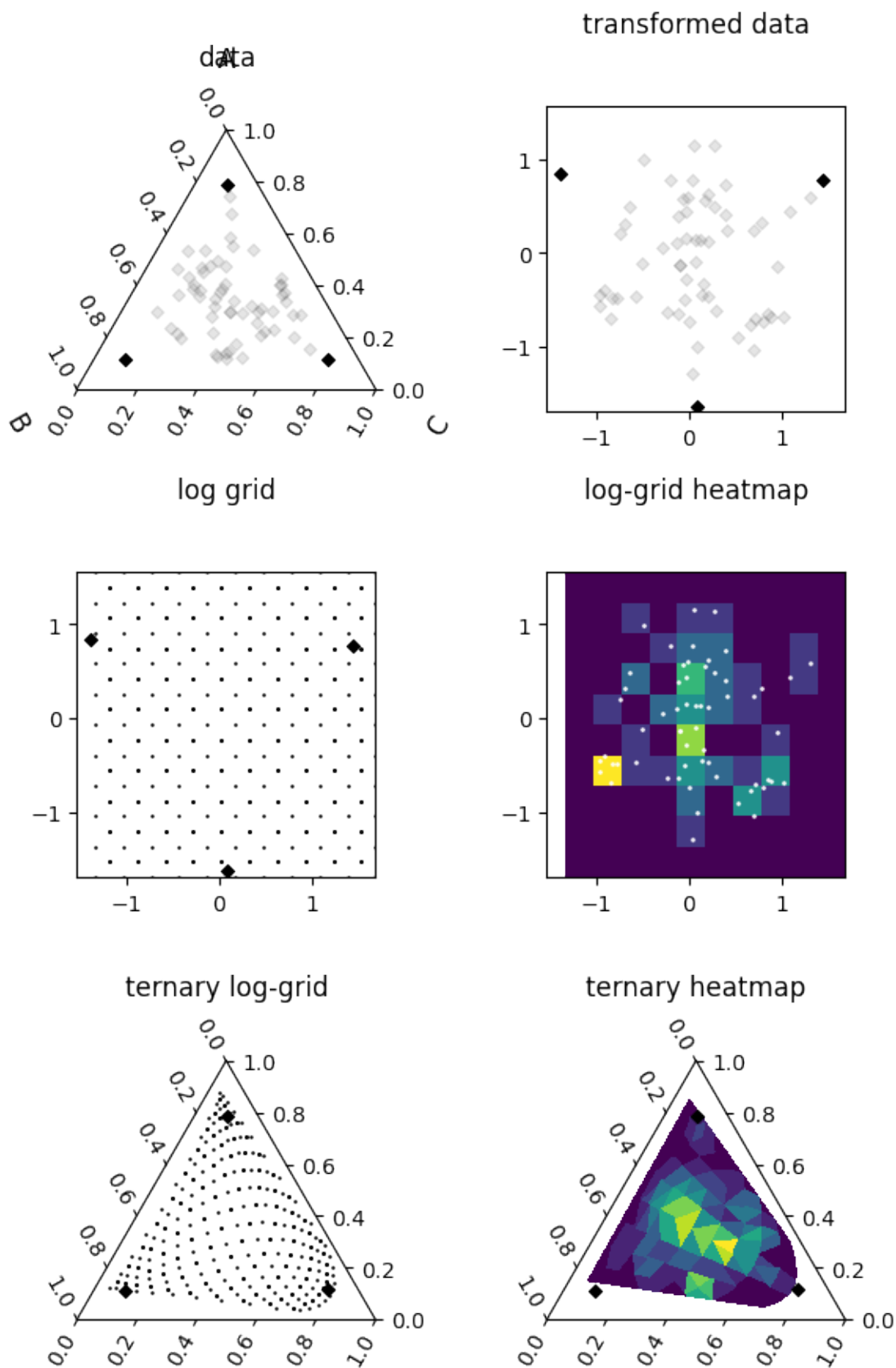
ax[2].set_title("log grid", y=1.2)
ax[2].scatter(*flattengrid(data["tfm_centres"]).T, c="k", marker=".", s=5)
ax[2].scatter(*flattengrid(data["tfm_edges"]).T, c="k", marker=".", s=2)
ax[2].scatter(*data["tfm_tern_bound_points"].T, c="k")

ax[3].set_title("log-grid heatmap", y=1.2)
ax[3].pcolormesh(*data["tfm_edges"], H)
ax[3].scatter(*data["grid_transform"](df.values).T, c="white", alpha=0.8, s=1)

ax[4].set_title("ternary log-grid", y=1.2)
ax[4].scatter(*data["tern_centres"].T, c="k", marker=".", s=5)
ax[4].scatter(*data["tern_edges"].T, c="k", marker=".", s=2)
ax[4].scatter(*data["tern_bound_points"].T, c="k")

ax[5].set_title("ternary heatmap", y=1.2)
ax[5].tripcolor(*coords.T, H.flatten())
ax[5].scatter(*data["tern_bound_points"].T, c="k")

plt.tight_layout()
```



Total running time of the script: (0 minutes 6.700 seconds)

## 3.5 Citation

---

**Note:** pyrolite began as a personal project, but as the project develops there are likely to be other contributors. Check the [Contributors list](#) and add major contributors as authors.

---

If you use pyrolite extensively for your research, citation of the software would be particularly appreciated. It helps quantify the impact of the project (assisting those contributing through paid and volunteer work), and is one way to get the message out and help build the pyrolite community.

pyrolite has been published in the Journal of Open Source Software (JOSS), and generally this paper is what you should cite. This paper was published in reference to pyrolite v0.2.7; if you're using a different version please note that in the citation.

While the exact format of your citation will vary with wherever you're publishing, it should take the general form:

Williams et al., (2020). pyrolite: Python for geochemistry. Journal of Open Source Software, 5(50), 2314, doi: [10.21105/joss.02314](https://doi.org/10.21105/joss.02314)

Or, if you wish to cite a specific version of the archive:

Williams et al. (). pyrolite v , Zenodo, doi:[10.5281/zenodo.2545106](https://doi.org/10.5281/zenodo.2545106)

If you're after a BibTeX citation for pyrolite, I've added one below.

```
@article{Williams2020,
  title = {pyrolite: Python for geochemistry},
  author = {Morgan J. Williams and
    Louise Schoneveld and
    Yajing Mao and
    Jens Klump and
    Justin Gosses and
    Hayden Dalton and
    Adam Bath and
    Steve Barnes},
  year = {2020},
  journal = {Journal of Open Source Software}
  doi = {10.21105/joss.02314},
  url = {https://doi.org/10.21105/joss.02314},
  publisher = {The Open Journal},
  volume = {5},
  number = {50},
  pages = {2314},
}
```

If you're using pyrolite's implementation of *lambdas*, please consider citing a more recent publication directly related to this:

Anenburg, M., & Williams, M. J. (2021). Quantifying the Tetrad Effect, Shape Components, and Ce–Eu–Gd Anomalies in Rare Earth Element Patterns. Mathematical Geosciences. doi: [10.1007/s11004-021-09959-5](https://doi.org/10.1007/s11004-021-09959-5)

Or, if you're using BibTeX:



```
@article{Anenburg_Williams_2021,  
  title={Quantifying the Tetrad Effect, Shape Components,  
        and Ce-Eu-Gd Anomalies in Rare Earth Element Patterns},  
  author={Anenburg, Michael and Williams, Morgan J.},  
  year={2021},  
  journal={Mathematical Geosciences},  
  doi={10.1007/s11004-021-09959-5},  
  url={https://doi.org/10.1007/s11004-021-09959-5},  
  ISSN={1874-8953},  
  month={Jul}  
}
```



## DEVELOPMENT

### 4.1 Development

#### 4.1.1 Development History and Planning

- [Changelog](#)
- [Future](#)

#### 4.1.2 Contributing

- [Contributing](#)
- [Contributors](#)
- [Code of Conduct](#)

#### 4.1.3 Development Installation

To access and use the development version, you can either [clone the repository](#) or install via pip directly from GitHub:

```
pip install --user git+https://github.com/morganjwilliams/pyrolite.git@develop  
↪#egg=pyrolite
```

#### 4.1.4 Tests

If you clone the source repository, unit tests can be run using `pytest` from the root directory after installation with development dependencies (`pip install -e .[dev]`):

```
python setup.py test
```

If instead you only want to test a subset, you can call `pytest` directly from within the pyrolite repository:

```
pytest ./test/<path to test or test folder>
```

## 4.2 Changelog

All notable changes to this project will be documented here.

### 4.2.1 Development

---

**Note:** Changes noted in this subsection are to be released in the next version. If you're keen to check something out before its released, you can use a [development install](#).

---

#### 4.2.2 0.3.5

- **New Contributor:** [Malte Mues](#)
- **New Contributor:** [Bob Myhill](#)
- **Bugfix:** Fixed a bug with the index structure for the documentation example galleries.
- Expanded test suite for a text coverage bump.
- Started testing against and supporting Python 3.12.

#### `pyrolite.geochem`

- **PR Merged:** [Malte Mues](#) contributed a PR to allow series and dataframes with *Pm* data to be used with the `pyrolite.pyrochem` REE accessors (i.e., the column will not be dropped if it exists; [#100](#)). This PR also included an update for the `pyrolite.util.lambdas` functions to be more flexible in terms of access, including as individual series (allowing performant usage with [pandarallel](#)).
- Updates to `pyrolite.pyrochem` accessors to be more flexible, generally allowing usage with both DataFrame and Series objects (e.g. `df.pyrochem.REE`, `ser.pyrochem.REE`).

#### `pyrolite.plot`

- **PR Merged:** [Bob Myhill](#) contributed a series of pull requests to fix formatting of TAS diagram labels ([#91](#)), update the TAS field names in the JSON template ([#92](#)), improve the formatting and scaling of TAS diagrams ([#93](#)), add an option to add field labels in the visual centre of polygons (rather than the 'centroid'; [#94](#)), update some usage of `matplotlib` ([#96](#)), and allow selective plotting of individual fields when adding a classification diagram to axes ([#98](#)).

#### `pyrolite.mineral`

- Suppressed `pandas` performance warnings related to sequential construction of `pandas.DataFrame`'s in `:func:`pyrolite.mineral.normative.CIPW_norm``.

### 4.2.3 0.3.4

- **Bugfix:** Tom Buckle contributed a PR with some minor bugfixes for the CIPW Norm (#87).
- Various maintenance updates, including migrating the package to use *pyproject.toml*.

### 4.2.4 0.3.3

- **New Contributor:** Sarah Shi
- **New Contributor:** Ondrej Lexa
- **Bugfix:** Updated docs builds to be compatible with recent versions of *sphinx-gallery*.
- **Bugfix:** Updated some *pandas* assignment, aggregation and similar operations to be compatible with more recent versions of Pandas.
- Added explicit Python 3.10 and 3.11 support.
- Removed Python 3.7 support (now end of life).

#### pyrolite.mineral

- **PR Merged:** Added an option to get expanded outputs for the CIPW Norm (from Tom Buckle; #80).
- **Bugfix:** Fixes and updated tests for CIPW Norm outputs.

#### pyrolite.plot

- **PR Merged:** Sarah Shi contributed a PR to add variations on the TAS diagram from Le Maitre ( #79). These can be accessed by providing a *which\_model* keyword argument to the *TAS* constructor (or plot template).
- **PR Merged:** Ondrej Lexa contributed a PR to add sandstone bulk geochemistry discrimination diagrams (*Pettijohn, Herron*; #82).
- **Bugfix:** Fixed issue with handling *vmin* and *vmax* for colormapping in *pyrolite.plot.color*.
- Suppressed warnings for 'division by zero'/'invalid value encountered in divide' in ternary diagram scatter plots.
- Added explicit support for colormapping categorical data in *pyrolite.plot.color*, such that ordering is preserved/consistent in e.g. legends.

#### pyrolite.util

- **Feature:** Added new version of ICS International Chronostratigraphic Chart (2022-10; *pyrolite.util.time.Timescale*).
- **Bugfix:** Corrected TAS diagram references, and fixed an issue where only the ID names were able to we added to diagrams.
- Updated axes-sharing utility function *share\_axes()* to reflect more recent versions of *matplotlib*.
- Fixed issue in path interpolation for contours (*pyrolite.util.plot.interpolation.get\_contour\_paths*) for recent *matplotlib* versions.
- Updated figure export utility function to use *pathlib* syntax for suffixes, which should avoid potential for double suffixes (e.g. *figure\_name.png.png*).

### 4.2.5 0.3.2

- **New Contributor:** [Angela Rodrigues](#)
- **Bugfix:** Edited docstrings and added ignore-warning for `numpydoc` warnings.
- **Bugfix:** Updated installation instructions and Binder configuration to use secure protocols when installing via *git* (i.e. <https://>)
- **Bugfix:** Update CI builds so that tests can be run on MacOS.

#### `pyrolite.mineral`

- **Feature:** Added a TAS-based iron correction following Middlemost (1989).
- **Bugfix:** Fixed some errors in mineral formulae and handling leading to inaccurate outputs from CIPW Norm.
- Split out volcanic from intrusive samples in the CIPW Norm volcanic rock comparison.
- Added SINCLAS abbreviations to the mineral dictionary associated with the CIPW Norm, so alternate mineral naming systems can be compared.

#### `pyrolite.util`

- **PR Merged:** [Louise Schoneveld](#) submitted a pull request to add bivariate and ternary classifier models for spinel compositions (*SpinelFeBivariate*, *SpinelTrivalentTernary*).
- **PR Merged:** [Angela Rodrigues](#) submitted a pull request to add the Jensen ternary cation classifier model for subalkalic volcanic rocks (*JensenPlot*).
- Updated `pyrolite.util.sklearn.plot_confusion_matrix()` to be able to plot on existing axes, use an explicit class order and use rotation for e.g. long x-axis class label names.
- Updated references to `scipy.stats.gaussian_kde()` after namespace deprecation.

### 4.2.6 0.3.1

- **New Contributor:** [Martin Bentley](#)
- **New Contributor:** [Chetan Nathwani](#)
- **New Contributor:** [Tom Buckle](#)
- **New Contributor:** [Nicolas Piette-Lauziere](#)
- Removed a redundant `pathlib` dependency (which is standard library as of Python 3.4). This will fix an issue blocking setting up a *conda-forge* recipe (#51).
- Updated instances of redundant `numpy` types throughout to silence deprecation warnings (using base types `float`, `int` except where specific `numpy` types are required).
- Added a minimum `sympy` version requirement (v1.7) to avoid potential import errors.
- Updated minimum versions for `matplotlib` and `mpltern` to address potential version conflicts.
- A user installation is now recommended by default. This solves some potential issues on \*-nix and MacOS systems.
- Fixed broken links to documentation in the README (thanks to [Alessandro Gentilini](#)).
- Fixed a bad documentation link the PyPI package information.

- Updated supported Python versions (Python 3.7-3.9).
- **Bugfix:** Updated use of `tinydb` databases to default to read-only access except where write access is explicitly needed. This should solve issues with permissions during installation and use of `pyrolite` on some systems ( #61). Thanks to [Antoine Ouellet](#) for bringing this to attention, and both Sam Bradley and Alex Hunt for following up with the idea for the current solution.

### `pyrolite.geochem`

- **Feature:** [Nicolas Piette-Lauziere](#) contributed two new functions for `pyrolite.geochem.alteration`: The chlorite-carbonate-pyrite index of Large et al. (2001; `CCPI()`) and the Alteration index of Ishikawa (1976; `IshikawaAltIndex()`).
- **Bugfix:** Fixed a bug where incomplete results were being returned for calls to `lambda_lnREE()` using the O'Neill algorithm. In this instance only the rows with the least missing data (typically those with no missing data) would return lambda values, other rows would contain null values. Thanks to Mark Pearce for identifying this one! In the meantime, using `df.pyrochem.lambda_lnREE(algorithm='opt')` will allow you to avoid the issue.
- **Bugfix:** Modified a few of the `pyrolite.geochem.pyrochem` methods to avoid a bug due to assignment of the dataframe ( `to_molecular()`, `to_weight()`, `recalculate_Fe()` ). This bug seems to be contained to the dataframe accessor, the individual functions from `pyrolite.geochem.transform` appear to work as expected outside of this context. Thanks to [Chetan Nathwani](#) for highlighting this one!
- Renamed (private) package variables `__common_oxides__` and `__common_elements__` to `_common_oxides` and `_common_elements`

### `pyrolite.mineral`

- **Feature:** CIPW function added to `pyrolite.mineral.normative`, largely from contributions by both [Chetan Nathwani](#) and [Tom Buckle](#) ( #53). Note that the implementation still has a bug or two to be ironed out; it will currently raise a warning when used to make sure you're aware of this. An [example](#) has been added demonstrating the intended functionality and demonstrating how coherent this is with existing implementations of CIPW (e.g. SINCLAS).

### `pyrolite.comp`

- Updated `pyrolite.comp.codata.close()` to better deal with zeros (avoiding unnecessary warnings).
- Added spherical coordinate transformation to `pyrolite.comp.pyrocomp` and `pyrolite.comp.codata` (see `pyrolite.comp.pyrocomp.sphere()`).

### `pyrolite.plot`

- **Feature:** Added ternary classification plot templates `USDAsoilTexture`, `FeldsparTernary` and `QAP` ( #49; idea and implementation of the latter thanks to [Martin Bentley](#) !). The idea for implementing the ternary diagram came from a discussion with Jordan Lubbers and Penny Wieser (of the [Thermobar](#) team, who are working in similar spaces); they've now implemented a version using `python-ternary` (rather than `mpltern`, which `pyrolite` is currently using).
- Updated examples and documentation for density and contour plots.
- Added autoscaling for standard `spider()` and related plots to address ( #55)

- `process_color()` has been updated to better deal with data explicitly declared to be of a 'category' data type (as a `pandas.Series`), and also to better handle variation in mapping notations. Ordering of categorical variables will now be preserved during color-mapping.
- Added the option to have a 'bad' color to be used in categorical color-mapping where a category cannot be found.
- Inconsistent color specifications (e.g. a list or array of multiple types) will now result in an error when passed to `process_color()`.
- `parallel()` has been updated to align with other plotting functions (taking an optional `components` keyword argument).

## pyrolite.util

- **Feature:** Added ternary classification models for *USDA Soil Texture*, *Feldspar Ternary* and *QAP* ([#49](#); idea and implementation of the latter thanks to [Martin Bentley](#)).
- Added some functionality to `pyrolite.util.classification` to allow classifier fields to be precisely specified by ratios (useful in ternary systems), for multiple 'modes' of diagrams to be contained a single configuration file, and fixed some issues with labelling (arguments `add_labels` and `which_labels` can now be used to selectively add either field IDs/abbreviations or field names to classification diagrams).
- Limits are no longer explicitly required for bivariate templates (`xlim`, `ylim`) in `pyrolite.util.classification`.
- Update default parameterisation to "full" for lambdas, using all REE to generate orthogonal polynomial functions.
- Expanded `pyrolite.util.text.int_to_alpha()` to handle integers which are greater than 25 by adding multiple alphabetical characters (e.g. `26 > aa`), and to use the built-in `string.ascii_lowercase`.
- `save_figure()` will now create the directory it's given if it doesn't exist.
- Citation information for *lambdas* updated to include recent publications.
- Updated `plot_pca_vectors()` to accept line `colors` and `linestyles` arguments.
- Updated `init_spherical_octant()` to accept a `fontsize` argument.
- Added `example` for coloring ternary diagrams and ternary scatter points based on a ternary color system.
- Added helper for generating PCA component labels from a *scikit-learn* PCA object (`get_PCA_component_labels()`)
- Updated confusion matrix visualisation helper `plot_confusion_matrix()` to remove grid and provide more useful default colormap normalization options.
- Moved the `manifold visualisation` example to utility examples from plotting examples.
- Added a `fnt_string` argument to *LogTransform* for use in automated naming of transformed columns; this may be expanded to other transformers soon.
- Fixed some string issues for `pyrolite.util.text`.



### 4.2.7 0.3.0

- **New Contributor:** Lucy Mathieson
- Continuous Integration has been migrated from Travis to GitHub Actions.
- Added an `environment.yml` file for development environment consistency.
- Removed some tests dependent on `xlrd` due to external issues with reading `.xls` and `.xlsx` files with some OS-Python version combinations.
- Fixed some broken documentation links.
- Added `psutil` to requirements.

#### `pyrolite.plot`

- **Bugfix:** Fixed a bug where there scatter and line arguments would conflict for `spider()` (#46). To address this, `spider()` and related functions will now accept the keyword arguments `line_kw` and `scatter_kw` to explicitly configure the scatter and line aspects of the spider plot - enabling finer customization. An [extra example](#) has been added to the docs to illustrate the use of these parameters. Thanks go to Lucy Mathieson for raising this one!
- Added the `set_ticks` keyword argument to `spider()` and associated functions, allowing ticks to be optionally set (`set_ticks=False` if you don't want to set the x-ticks).
- Updated `pyrolite.plot.color.process_color()` to better handle colour mapping and added [examples illustrating this](#). You can also now use RGBA colours when using the `color_mappings` keyword argument.
- Updated automated pyrolite `matplotlib` style export to be more reliable.
- Changed the default shading for `density()` to suppress error about upcoming `matplotlib` deprecation.
- Ordering for contours, contour names and contour styles is now preserved for `density()` and related functions.
- Updated `pyrolite.plot.templates.pearce` to use ratios from Sun & McDonough (1989), as in the Pearce (2008) paper.

#### `pyrolite.geochem`

- **Bugfix:** Fixed a bug where Eu was unnecessarily excluded from the `lambda_InREE()` fit in all cases.
- **Bugfix:** Fixed a bug where ratio-based normalisation was not implemented for `get_ratio()` and related functions (#34)
- Added a local variable to `pyrolite.geochem.ind` to allow referencing of indexing functions (e.g. `by_incompatibility()`) by name, allowing easier integration with `spider()`.
- Added `by_number()` for indexing a set of elements by atomic number.

### pyrolite.comp

- Updated the docstring for `pyrolite.comp.impute.EMCOMP()`.
- Minor updates for `pyrolite.comp.codata` labelling, and reflected changes in `pyrolite.util.sk1.transform`. Issues were identified where the column name 'S' appears, and a workaround has been put in place for now.

### pyrolite.util

- Expanded `pyrolite.util.lambdas` to allow fitting of tetrad functions, anomalies and estimation of parameter uncertainties for all three algorithms.
- Added `pyrolite.util.resampling` for weighted spatiotemporal bootstrap resampling and estimation, together with added a number of updates to `pyrolite.util.spatial` to provide required spatial-similarity functionality.
- Updated the geological timescale in `pyrolite.util.time` to use the 2020/03 version of the International Chronostratigraphic Chart (#45).
- Added `alphalabel_subplots()` for automatic alphabetic labelling of subplots (e.g. for a manuscript figure).
- Fixed an low-precision integer rollover issue in a combinatorial calculation for `pyrolite.util.missing` by increasing precision to 64-bit integers.
- Added `example_patterns_from_parameters()` to work with `pyrolite.util.lambdas` and generate synthetic REE patterns based on lambda and/or tetrad-parameterised curves.
- Moved `get_centroid()` from `pyrolite.util.classification` to `pyrolite.util.plot.helpers`
- Made a small change to `density` to allow passing contour labels as a list.
- `mappable_from_values()` will not accept a `norm` keyword argument, allowing use of colormap normalisers like `matplotlib.colors.Normalize`. This function was also updated to better handle `Series` objects.
- Fixed a small bug for `TAS` instantiation which didn't allow passing the variables to be used from a `pandas.DataFrame`. If you have different variable names, you can now pass them as a list with the `axes` keyword argument (e.g. `TAS(axes=['sio2', 'alkali'])`).
- Homogenised logging throughout the package - now all managed through `pyrolite.util.log`. The debugging and logging streaming function `stream_log()` can now also be accessed here (`pyrolite.util.log.stream_log()`).

## 4.2.8 0.2.8

- Updated citation information.
- Added specific testing for OSX for Travis, and updated the install method to better pick up issues with pip installations.
- **Feature:** Added a [gallery](#) of pages for each of the datasets included with `pyrolite`. This will soon be expanded, especially for the reference compositions (to address #38).

## pyrolite.geochem

- **PR Merged:** Kaarel Mand submitted a [pull request](#) to add a number of shale and crustal compositions to the reference database.
- **Bugfix:** Fixed a bug where lambdas would only be calculated for rows without missing data. Where missing data was present, this would result in an assertion error and hence no returned values.
- **Bugfix:** Fixed a bug where missing data wasn't handled correctly for calculating lambdas. The functions now correctly ignore the potential contribution of elements which are missing when parameterising REE patterns. Thanks to Steve Barnes for the tip off which led to identifying this issue!
- **Feature:** Added `pyrolite.geochem.ind.REY()`, `list_REY()`, and `REY()` to address ([#35](#)). This issue was also opened by Kaarel Mand!
- As a lead-in to a potential change in default parameterisation, you can now provide additional specifications for the calculation of *lambdas* to `lambda_lnREE()` and `calc_lambdas()` to determine the basis over which the individual orthogonal polynomials are defined (i.e. which REE are included to define the orthonormality of these functions). For the keyword argument `params`, (as before) you can pass a list of tuples defining the constants representing the polynomials, but you can now alternatively pass the string "ONeill2016" to explicitly specify the original parameterisation, or "full" to use all REE (including Eu) to define the orthonormality of the component functions (i.e. using `params="full"`). To determine which elements are used to perform the *fit*, you can either filter the columns passed to these functions or specifically exclude columns using the `exclude` keyword argument (e.g. the default remains `exclude=["Eu"]` which excludes Eu from the fitting process). Note that the default for fitting will remain, but going forward the default for the definition of the polynomial functions will change to use all the REE by default (i.e. change to `params="full"`).
- Significant performance upgrades for `lambda_lnREE()` and associated functions (up to 3000x for larger datasets).
- Added a minimum number of elements, configurable for `lambda_lnREE()`. This is currently set to seven elements (about half of the REE), and probably lower than it should be ideally. If for some reason you want to test what lambdas (maybe just one or two) look like with less elements, you can use the `min_elements` keyword argument.
- Added `list_isotope_ratios()` and corresponding selector `isotope_ratios()` to subset isotope ratios.
- Added `parse_chem()` to translate geochemical columns to a standardised (and pyrolite-recognised) column name format.

## pyrolite.plot

- **Bugfix:** Fixed a bug where arguments processing by `pyrolite.plot.color` would consume the 'alpha' parameter if no colour was specified (and as such it would have no effect on the default colors used by `matplotlib`)
- **Bugfix:** `pyrolite.plot.color` now better handles colour and value arrays.
- **Bugfix:** Keyword arguments passed to `pyrolite.plot.density` will now correctly be forwarded to respective functions for histogram and hexbin methods.
- Customised `matplotlib` styling has been added for pyrolite plotting functions, including legends. This is currently relatively minimal, but could be expanded slightly in the future.
- The `bw_method` argument for `scipy.stats.kde.gaussian_kde()` can now be parsed by pyrolite density-plot functions (e.g. `density()`, `heatscatter()`). This means you can modify the default bandwidth of the gaussian kernel density plots. Future updates may allow non-Gaussian kernels to also be used for these purposes - keep an eye out!

- You can now specify the y-extent for conditional spider plots to restrict the range over which the plot is generated (and focus the plot to where your data actually is). For this, feed in a (`min`, `max`) tuple for the `yextent` keyword argument.
- The `ybins` argument for `spider()` and related functions has been updated to `bins` to be in line with other functions.
- Conditional density `REE()` plots now work as expected, after some fixes for generating reverse-ordered indexes and bins
- Added a filter for ternary density plots to ignore true zeroes.
- Some updates for `pyrolite.plot.color` for alpha handling and colour arrays .

#### pyrolite.comp

- Updated transform naming to be consistent between functions and class methods. From this version use capitalised versions for the transform name acronyms (e.g. `ILR` instead of `ilr`).
- Added for transform metadata storage within DataFrames for `pyrocomp`, and functions to access transforms by name.
- Added labelling functions for use with `pyrolite.comp.pyrocomp` and `codata` to illustrate the precise relationships depicted by the logratio metrics (specified using the `label_mode` parameter supplied to each of the respective `pyrocomp` logratio transforms).

#### pyrolite.util

- Revamped `pyrolite.util.classification` to remove cross-compatibility bugs with OSX/other systems. This is now much simpler and uses JSON for serialization.
- Small fix for `mappable_from_values()` to deal with NaN values.
- Added `pyrolite.util.log` for more streamlined logging (from `pyrolite-meltsutil`)
- Added `pyrolite.util.spatial.levenshtein_distance()` for comparing sequence differences/distances between 1D iterables (e.g. strings, lists).

### 4.2.9 0.2.7

- Bugfix to include radii data in `MANIFEST.in`

### 4.2.10 0.2.6

- **New Contributors:** [Kaarel Mand](#) and [Laura Miller](#)
- **PR Merged:** [Louise Schoneveld](#) submitted a pull request to fill out the newly-added [Formatting and Cleaning Up Plots](#) tutorial. This tutorial aims to provide some basic guidance for common figure and axis formatting tasks as relevant to `pyrolite`.
- Added `codacy` for code quality checking, and implemented numerous clean-ups and a few new tests across the package.
- Performance upgrades, largely for the documentation page. The docs page should build and load faster, and have less memory hang-ups - due to smaller default image sizes/DPI.
- Removed dependency on `fancyimpute`, instead using functions from `scikit-learn`

### pyrolite.geochem

- **Bugfix:** pyrolite lambdas differ slightly from [?] (#39). Differences between the lambda coefficients of the original and pyrolite implementations of the lambdas calculation were identified (thanks to [Laura Miller](#) for this one). With further investigation, it's likely the cost function passed to `scipy.optimize.least_squares()` contained an error. This has been remedied, and the relevant pyrolite functions now by default should give values comparable to [?]. As part of this, the reference composition *ChondriteREE\_ON* was added to the reference database with the REE abundances presented in [?].
- **Bugfix:** Upgrades for `convert_chemistry()` to improve performance (#29). This bug appears to have resulted from caching the function calls to `pyrolite.geochem.ind.simple_oxides()`, which is addressed with [18fed0](#).
- **Feature:** Added the [?] ionic radii for use in silicate geochemistry ( #41), which can optionally be used with `pyrolite.geochem.ind.get_ionic_radii()` using the `source` keyword argument (`source='Whittaker'`). Thanks to [Charles Le Losq](#) for the suggestion!
- **Bugfix:** Removed an erroneous zero from the GLOSS reference composition (*GLOSS\_P2014* value for Pr).
- Updated `REE()` to default to `dropPm=True`
- Moved `pyrolite.mineral.ions` to `pyrolite.geochem.ions`

### pyrolite.mineral

- **Bugfix:** Added the mineral database to *MANIFEST.in* to allow this to be installed with pyrolite (fixing a bug where this isn't present, identified by [Kaarel Mand](#)).

### pyrolite.plot

- **Bugfix:** Updated `pyrolite.plot` to use `pandas.DataFrame.reindex()` over `pandas.DataFrame.loc()` where indexes could include missing values to deal with #31.
- Updated `spider()` to accept `logy` keyword argument, defaulting to `True`

### pyrolite.util

- Broke down `pyrolite.util.plot` into submodules, and updated relevant imports. This will result in minimal changes to API usage where functions are imported explicitly.
- Split out `pyrolite.util.lambdas` from `pyrolite.util.math`
- Added a minimum figure dimension to `init_axes()` to avoid having null-dimensions during automatic figure generation from empty datasets.
- Added `example_spider_data()` to generate an example dataset for demonstrating spider diagrams and associated functions. This allowed detailed synthetic data generation for `spider()` and `pyrolite.plot.pyroplot.REE()` plotting examples to be cut down significantly.
- Removed unused submodule `pyrolite.util.wfs`

### 4.2.11 0.2.5

- **PR Merged:** @lavender22 updated the spider diagram example to add a link to the normalisation example (which lists different reservoirs you can normalise to).
- Added an ‘Importing Data’ section to the docs [Getting Started](#) page.
- Disabled automatic extension loading (e.g. for `pyrolite_meltsutil`) to avoid bugs during version mismatches.

#### `pyrolite.comp`

- Updated the `pyrolite.comp.pyrocomp` dataframe accessor API to include reference to compositional data log transform functions within `pyrolite.comp.codata`

#### `pyrolite.plot`

- Added support for spider plot index ordering added with the keyword `index_order` (#30)
- Added support for color indexing in `color` using `pandas.Series`, and also for list-like arrays of categories
- Added a workaround for referring to axes positions where the projection is changed to a ternary projection (displacing the original axis), but the reference to the original axes object (now booted from `fig.axes/fig.orderedaxes`) is subsequently used.
- Updated `process_color()` processing of auxillary color keyword arguments (fixing a bug for color arguments in `stem()`)
- Added support for a `color_mappings` keyword argument for mapping categorical variables to specific colors.
- Updated the effect of `relin` keyword argument of `density()` to remove the scaling (it will no longer log-scale the axes, just the grid/histogram bins).
- Updated `Grid` to accept an x-y tuple to specify numbers of bins in each direction within a grid (e.g. `bins=(20, 40)`)
- Updated the grids used in some of the `density()` methods to be edges, lining up the arrays such that shading parameters will work as expected (e.g. `shading='gouraud'`)

#### `pyrolite.geochem`

- Added sorting function `~pyrolite.geochem.ind.by_incompatibility` for incompatible element sorting (based on BCC/PM relative abundances).

#### `pyrolite.mineral`

- Minor bugfix for `update_database()`

## pyrolite.util

- Moved `check_perl()` out of `pyrolite` into `pyrolite_meltsutil`

## 4.2.12 0.2.4

- Removed Python 3.5 support, added Python 3.8 support.
- Updated ternary plots to use `mpltern` (#28)
- Added a [ternary heatmap tutorial](#)

## pyrolite.plot

- Added `pyrolite.plot.pyroplot.plot()` method
- Removed `pyrolite.plot.pyroplot.ternary()` method (ternary plots now served through the same interface as bivariate plots using `pyrolite.plot.pyroplot.scatter()`, `pyrolite.plot.pyroplot.plot()`, and `pyrolite.plot.pyroplot.plot()`)
- Added `pyrolite.plot.color` for processing color arguments.
- Moved `pyrolite.plot.density` to its own sub-submodule, including `pyrolite.plot.density.ternary` and `pyrolite.plot.density.grid`

## pyrolite.util

- Updated `time` to include official colors.
- Added `pyrolite.util.time` [example](#)
- Updated `stream_log()` to deal with logger duplication issues.
- Various updates to `pyrolite.util.plot`, noted below:
- Added universal axes initiation for bivariate/ternary diagrams using `init_axes()` and axes labelling with `label_axes()`,
- Added keyword argument processing functions `scatterkwargs()`, `linekwargs()`, and `patchkwargs()`
- Added functions for replacing non-projected axes with ternary axes, including `replace_with_ternary_axis()`, `axes_to_ternary()` (and `get_axes_index()` to maintain ordering of new axes)
- Added `get_axis_density_methods()` to access the relevant histogram/density methods for bivariate and ternary axes
- Renamed private attributes for default colormaps to `DEFAULT_DISC_COLORMAP` and `DEFAULT_CONT_COLORMAP`
- Updated `add_colorbar()` to better handle colorbars for ternary diagrams

### 4.2.13 0.2.3

- Added [Getting Started](#) page

#### `pyrolite.mineral`

- Updated database for `pyrolite.mineral.mindb` to include epidotes, garnets, micas

#### `pyrolite.plot`

- Minor updates for `pyrolite.plot.templates`, added functionality to `pyrolite.plot.templates.TAS()` stub.
- Fixed a bug for `vmin` in `pyrolite.plot.spider` density modes

#### `pyrolite.geochem`

- `pyrolite.geochem.parse` now also includes functions which were previously included in `pyrolite.geochem.validate`
- Fixed some typos in reference compositions from Gale et al. (2013)

#### `pyrolite.util`

- Added `pyrolite.util.plot.set_ternary_labels()` for setting and positioning ternary plot labels

### 4.2.14 0.2.2

#### `pyrolite.geochem`

- Added `SCSS()` for modelling sulfur content at sulfate/sulfide saturation.

#### `pyrolite.mineral`

- Added [mineral](#) database and [mineral endmember decomposition](#) examples

### 4.2.15 0.2.1

- Updated and refactored documentation
  - Added [Development](#), [Debugging](#) section, [Extensions](#)
  - Added [sphinx\\_gallery](#) with binder links for examples
  - Removed duplicated examples
  - Amended [citation guidelines](#)
- Removed extensions from pyrolite (`pyrolite.ext.datarepo`, `pyrolite.ext.alphamelts`). These will soon be available as separate extension packages. This enabled faster build and test times, and removed extraneous dependencies for the core pyrolite package.
- Added `stats_require` as optional requirements in `setup.py`



### pyrolite.geochem

- Added `get_ratio()` and `pyrolite.geochem.pyrochem.get_ratio()`
- Added `pyrolite.geochem.pyrochem.compositional()` selector

### pyrolite.plot

- `parallel()` now better handles `pyplot` figure and subplot arguments
- `ternary()` and related functions now handle label offsets and label font sizes
- Minor bugfixes for `density`
- Added `unity_line` argument to `spider()` to be consistent with `REE_v_radii()`

### pyrolite.mineral

- Added a simple `pyrolite.mineral.mindb` database
- Added `pyrolite.mineral.transform` to house mineral transformation functions
- Expanded `pyrolite.mineral.normative` to include `unmix()` and `pyrolite.mineral.normative.endmember_decompose()` for composition-based mineral endmember decomposition

### pyrolite.util

- Added `pyrolite.util.plot.mappable_from_values()` to enable generating `ScalarMappable` objects from an array of values, for use in generating colorbars

## 4.2.16 0.2.0

- Added alt-text to documentation example images
- Updated contributing guidelines
- Added Python 3.8-dev to Travis config (not yet available)
- Removed `pandas-flavor` decorators from `pyrolite.geochem` and `pyrolite.comp`, eliminating the dependency on `pandas-flavor`

### pyrolite.geochem

- Expanded `pyrolite.geochem.pyrochem` DataFrame accessor and constituent methods
- Updates and bugfixes for `pyrolite.geochem.transform` and `pyrolite.geochem.norm`
- Updated the `normalization example`

#### pyrolite.comp

- Added *pyrolite.comp.pyrocomp* DataFrame accessor with the *pyrolite.comp.codata.renormalise()* method.
- Removed unused imputation and aggregation functions.

#### pyrolite.plot

- Added *heatscatter()* and *example*.
- Updates and bugfixes for *pyrolite.plot.spider.REE\_v\_radii()*, including updating spacing to reflect relative ionic radii

#### pyrolite.util

- Added *pyrolite.util.plot.get\_twins()*

### 4.2.17 0.1.21

#### pyrolite.plot

- Added parallel coordinate plots: *pyrolite.plot.pyroplot.parallel()*
- Updated *scatter()* and *ternary()* to better deal with colormaps

#### pyrolite.ext.alphamelts

- Updated *pyrolite.ext.alphamelts* interface:
  - Docs
  - Updated to default to tables with percentages (Wt%, Vol%)
  - Updated *plottemplates* y-labels
  - Fixed automation grid bug

### 4.2.18 0.1.20

#### pyrolite.geochem

- Stub for *pyrolite.geochem.pyrochem* accessor (yet to be fully developed)
- Convert reference compositions within of *pyrolite.geochem.norm* to use a JSON database

### `pyrolite.util.skf`

- Added `pyrolite.util.skf.vis.plot_mapping()` for manifold dimensional reduction
- Added `pyrolite.util.skf.vis.alphas_from_multiclass_prob()` for visualising multi-class classification probabilities in scatter plots

### `pyrolite.plot`

- Added `pyrolite.plot.biplot` to API docs
- Updated default y-aspect for ternary plots and axes patches

### `pyrolite.ext.alphamelts`

- Updated `pyrolite.ext.alphamelts.automation`, `pyrolite.ext.alphamelts.meltsfile`, `pyrolite.ext.alphamelts.tables`
- Updated docs to use `pyrolite.ext.alphamelts.automation.MeltsBatch` with a parameter grid

## 4.2.19 0.1.19

- Added this changelog
- Require `pandas`  $\geq$  v0.23 for DataFrame accessors

### `pyrolite.geochem`

- Moved normalization into `pyrolite.geochem`
- Improved support for molecular-based calculations in `pyrolite.geochem`
- Added `pyrolite.geochem` section to API docs
- Added the `convert_chemistry()` docs example

### `pyrolite.ext.alphamelts`

- Improvements for `pyrolite.ext.alphamelts.download`
- Completed `pyrolite.ext.alphamelts.automation.MeltsBatch`
- Added the `pyrolite.ext.alphamelts.web` docs example
- Added `pyrolite.ext.alphamelts.plottemplates` to API docs
- Added `pyrolite.ext.alphamelts.tables.write_summary_phaselist()`
- Added `pyrolite.ext.alphamelts.automation.exp_name()` for automated alphaMELTS experiment within batches

## pyrolite.util

- Added `pyrolite.util.meta.ToLogger` output stream for logging
- Added `pyrolite.util.multip.combine_choices()` for generating parameter combination grids

## 4.2.20 0.1.18

- Require `scipy`  $\geq 1.2$

## pyrolite.plot

- Automatic import of dataframe accessor `df.pyroplot` removed; import `pyrolite.plot` to use `pyrolite.plot.pyroplot` dataframe accessor
- Updated label locations for `pyrolite.plot.biplot`
- Default location of the y-axis updated for `pyrolite.plot.stem.stem()`

## pyrolite.geochem

- Added stub for `pyrolite.geochem.quality`

## pyrolite.util

- Moved `pyrolite.classification` to `pyrolite.util.classification`
- Added `pyrolite.util.plot.marker_cycle()`

## 4.2.21 0.1.17

- Update status to Beta

## pyrolite.geochem

- Added database for geochemical components (`geochemdb.json`) for faster import via `common_elements()` and `common_oxides()`
- Added stub for `pyrolite.geochem.isotope`
- Update to using `pyrolite.util.transform.aggregate_element()` rather than `aggregate_cation`

## pyrolite.plot

- Expanded use of `pyrolite.plot.pyroplot` dataframe accessor
- Added `pyrolite.plot.pyrochem.cooccurrence()`
- Added `pyrolite.plot.biplot`
- Added support for conditional density spiderplots within `spider()` and `REE_v_radii()`
- Updated keyword argument parsing for `spider()`

### `pyrolite.mineral`

- Removed automatic import of mineral structures to reduce delay
- Updated `pyrolite.mineral.lattice.strain_coefficient()`
- Added stub for `pyrolite.mineral.normative()`
- Updated `pyrolite.mineral.sites.Site`

### `pyrolite.util`

- Added functions for interpolating paths and patches (e.g. contours) and exporting these: `interpolate_path()`, `interpolated_patch_path()`, `get_contour_paths()`, `path_to_csv()`
- Added `util.plot._mpl_sp_kw_split()`
- Added `util.text.remove_suffix()`
- Added `util.text.int_to_alpha()`

### `pyrolite.ext`

- Updated alphaMELTS interface location to external package interface rather than utility (from `pyrolite.util` to `pyrolite.ext`)
- Added `pyrolite.ext.datarepo` stub

## 4.2.22 0.1.16

### `pyrolite.mineral`

- Added `pyrolite.mineral.lattice` example
- Added `pyrolite.mineral.lattice.youngs_modulus_approximation()`

### `pyrolite.ext.alphamelts`

- Added `pyrolite.ext.alphamelts` Monte Carlo uncertainty estimation example
- Added `pyrolite.ext.alphamelts.automation.MeltsExperiment.callstring()` to facilitate manual reproducibility of pyrolite calls to alphaMELTS.
- Improved alphaMELTS interface termination
- Added `pyrolite.ext.alphamelts.plottemplates.phase_linestyle()` to for auto-differentiated linestyles in plots generated from alphaMELTS output tables
- Added `pyrolite.ext.alphamelts.plottemplates.table_by_phase()` to generate axes per phase from a specific output table

### `pyrolite.geochem`

- Added MORB compositions from Gale et al. (2013) to Reference Compositions
- Updated `pyrolite.geochem.ind.get_radii` to `pyrolite.geochem.ind.get_ionic_radii()`
- `dropPm` parameter added to `pyrolite.geochem.ind.REE()`

### `pyrolite.plot`

- Updated `pyrolite.plot.spider.REE_radii_plot` to `pyrolite.plot.spider.REE_v_radii()`
- Updated `pyrolite.util.meta.steam_log()` to take into account active logging handlers

### `pyrolite.util`

- Added `pyrolite.util.pd.drop_where_all_empty()`
- Added `pyrolite.util.pd.read_table()` for simple `.csv` and `.xlsx/.xls` imports
- Added `pyrolite.util.plot.rect_from_centre()`
- Added `pyrolite.util.text.slugify()` for removing spaces and non-alphanumeric characters

## 4.2.23 0.1.15

### `pyrolite.ext.alphamelts`

- Bugfixes for automation and download
- Add a `permissions` keyword argument to `pyrolite.util.general.copy_file()`

## 4.2.24 0.1.14

- Added Contributor Covenant Code of Conduct

### `pyrolite.plot`

- Added `pyrolite.plot.stem.stem()` example
- Added `pyrolite.plot.stem`
- Added `pyrolite.plot.stem` to API docs
- Added `pyrolite.plot.stem` example

### `pyrolite.mineral`

- Added `pyrolite.mineral.lattice` for lattice strain calculations
- Added `pyrolite.mineral` to API docs

### `pyrolite.ext.alphamelts`

- Improved `pyrolite.ext.alphamelts.automation` workflows, process tracking and termination
- Incorporated `MeltsProcess` into `MeltsExperiment`
- Added `MeltsBatch` stub
- Added `read_meltsfile()` and `read_envfile()`
- Added `pyrolite.ext.alphamelts.plottemplates`
- Added `pyrolite.ext.alphamelts.tables.get_experiments_summary()` for aggregating alphaMELTS experiment results across folders

### `pyrolite.util`

- Added manifold uncertainty example with `pyrolite.util.sklearn.vis.plot_mapping()`
- Updated `pyrolite.util.distributions.norm_to_lognorm`
- Added `pyrolite.util.general.get_process_tree()` to extract related processes
- Added `pyrolite.util.pd.zero_to_nan()`

## 4.2.25 0.1.13

### `pyrolite.ext.alphamelts`

- Updated `pyrolite.ext.alphamelts.automation.MeltsProcess` workflow
- Updated `pyrolite.ext.alphamelts.download` local installation
- Added `pyrolite.ext.alphamelts.install` example
- Added `pyrolite.ext.alphamelts.tables` example
- Added `pyrolite.ext.alphamelts.automation` example
- Added `pyrolite.ext.alphamelts.env` example

## 4.2.26 0.1.12

### `pyrolite.util.pd`

- Bugfix for `pyrolite.util.pd.to_frame()`

#### 4.2.27 0.1.11

- Added [citation](#) page to docs
- Added [contributors](#) page to docs
- Updated docs [future](#) page
- Updated docs config and logo

#### `pyrolite.geochem`

- Added stub for `pyrolite.geochem.isotope`, `pyrolite.geochem.isotope.count`

#### `pyrolite.comp`

- Added compositional data example
- Added `pyrolite.comp.codata.logratiomean()`
- Added `pyrolite.data.Aitchison` and associated data files

#### `pyrolite.ext.alphamelts`

- Added `pyrolite.ext.alphamelts` to API docs
- Added `pyrolite.ext.alphamelts.automation`

#### `pyrolite.util`

- Expanded `pyrolite.util` API docs
- Added `pyrolite.util.distributions`
- Moved `pyrolite_datafolder` from `pyrolite.util.general` to `pyrolite.util.meta.pyrolite_datafolder()`
- Added `share_axes()`, `ternary_patch()`, `subaxes()`
- Added `pyrolite.util.units`, moved `pyrolite.geochem.norm.scale_multiplier` to `pyrolite.util.units.scale()`
- Updated `pyrolite.util.synthetic.random_cov_matrix()` to optionally take a `sigmas` keyword argument

#### 4.2.28 0.1.10

- Updated [installation](#) docs



### pyrolite.util

- Added `pyrolite.util.types`
- Added `pyrolite.util.web`
- Added manifold uncertainty example with `pyrolite.util.sklearn.vis.plot_mapping()`
- Moved `stream_log` to `pyrolite.util.meta.stream_log()`
- Added `pyrolite.util.meta.take_me_to_the_docs()`
- Updated `pyrolite.util.sklearn.vis`

### pyrolite.ext.datarepo

- Updated `pyrolite.ext.datarepo.georoc` (then `pyrolite.util.repositories.georoc`)

## 4.2.29 0.1.9

### pyrolite.plot

- Added `pyrolite.plot.templates`, and related API docs
- Added Pearce templates under `pyrolite.plot.templates.pearce`
- Update default color schemes in scatter plots within `pyrolite.plot` to fall-back to `matplotlib.pyplot` cycling

### pyrolite.util

- Added conditional import for `PCA` and `statsmodels` within `pyrolite.util.plot`
- Refactored `sklearn` utilities to submodule `pyrolite.util.sklearn`
- Added `pyrolite.util.meta.sphinx_doi_link()`
- Updated `pyrolite.util.meta.inargs()`
- Updated `pyrolite.util.meta.stream_log()` (then `pyrolite.util.general.stream_log`)
- Added conditional import for `imblearn` under `pyrolite.util.sklearn.pipeline`

### pyrolite.ext.alphamelts

- Added `pyrolite.ext.alphamelts` (then `pyrolite.util.alphamelts`)
- Bugfix for Python 3.5 style strings in `pyrolite.ext.alphamelts.parse`

### 4.2.30 0.1.8

- Bugfixes for `pyrolite.plot.spider` and `pyrolite.util.plot.conditional_prob_density`

### 4.2.31 0.1.7

#### `pyrolite.plot`

- Added `cooccurrence()` method to `pyrolite.plot.pyroplot` DataFrame accessor

#### `pyrolite.util`

- Added `pyrolite.util.missing.cooccurrence_pattern()`
- Moved `pyrolite.util.sklearn.plot_cooccurrence` to `pyrolite.util.plot.plot_cooccurrence()`
- Updated `pyrolite.util.plot.conditional_prob_density()`, `pyrolite.util.plot.bin_edges_to_centres()` and `pyrolite.util.plot.bin_centres_to_edges()`

### 4.2.32 0.1.6

#### `pyrolite.plot`

- Update `spider()` to use `contours` keyword argument, and pass these to `pyrolite.util.plot.plot_Z_percentiles()`

#### `pyrolite.util`

- Bugfixes for invalid steps in `pyrolite.util.math.linspc_()`, `pyrolite.util.math.logspc_()`

### 4.2.33 0.1.5

- Updated docs [future](#) page

#### `pyrolite.geochem`

- Bugfix for iron redox recalculation in `pyrolite.geochem.transform.convert_chemistry()`

#### `pyrolite.plot`

- Added `mode` keyword argument to `pyrolite.plot.spider.spider()` to enable density-based visualisation of spider plots.
- Update `pyrolite.plot.pyroplot.spider()` to accept `mode` keyword argument
- Update `pyrolite.plot.pyroplot.REE()` to use a `index` keyword argument in the place of the previous `mode`; `mode` is now used to switch between line and density base methods of visualising spider plots consistent with `spider()`
- Added `spider()` [examples for conditional density plots](#) using `conditional_prob_density()`

- Bugfix for `set_under` in `density()`
- Updated [logo example](#)

#### `pyrolite.util`

- Updated `pyrolite.util.meta`
- Added `pyrolite.util.plot.conditional_prob_density()`; added conditional `statsmodels` import within `pyrolite.util.plot` to access `KDEMultivariateConditional`
- Added keyword argument `logy` to `pyrolite.util.math.interpolate_line()`
- Added `pyrolite.util.math.grid_from_ranges()` and `pyrolite.util.math.flattengrid()`
- Added support for differential x-y padding in `pyrolite.util.plot.get_full_extent()` and `pyrolite.util.plot.save_axes()`
- Added `pyrolite.util.sklearn.pipeline.fit_save_classifier()` (then `pyrolite.util.sklearn.fit_save_classifier()`)

### 4.2.34 0.1.4

#### `pyrolite.plot`

- Updated relevant docs and references for `pyrolite.plot` and the `pyrolite.plot.pyroplot` DataFrame accessor

#### `pyrolite.comp`

- Expanded `pyrolite.comp.impute` and improved `pyrolite.comp.impute.EMCOMP()`
- Added EMCOMP example (later removed in 0.2.5, pending validation and improvements for EMCOMP).

#### `pyrolite.util`

- Updated `pyrolite.util.meta` with docstring utilities `numpydoc_str_param_list()` and `get_additional_params()`

### 4.2.35 0.1.2

- Fixed logo naming issue in docs

#### `pyrolite.plot`

- Bugfixes for `pyrolite.plot.density.density()` (then `pyrolite.plot.density`) and `pyrolite.plot.util.ternary_heatmap()`

## 4.2.36 0.1.1

### pyrolite.plot

- Added `logo` example
- Refactored `pyrolite.plot` to use the `pyrolite.plot.pyroplot` DataFrame accessor:
  - Renamed `pyrolite.plot.spiderplot` to `pyrolite.plot.spider.spider()`
  - Renamed `pyrolite.plot.spider.REE_radii_plot` to `pyrolite.plot.spider.REE_v_radii()`
  - Renamed `pyrolite.plot.ternaryplot` to `pyrolite.plot.tern.ternary()`
  - Renamed `pyrolite.plot.densityplot` to `pyrolite.plot.density.density()`
- Updated `pyrolite.plot.density.density()` and `pyrolite.plot.tern.ternary()`

### pyrolite.comp

- Bugfixes and improvements for `pyrolite.comp.impute`

### pyrolite.geochem

- Updated `oxide_conversion()` and `convert_chemistry()`

### pyrolite.util

- Added `plot_stdev_ellipses()` and `plot_pca_vectors()`
- Updated `pyrolite.util.plot.plot_Z_percentiles()`
- Updated `pyrolite.util.plot.ternary_heatmap()`
- Updated `pyrolite.util.plot.vector_to_line()`

## 4.2.37 0.1.0

### pyrolite.plot

- Updates to `pyrolite.plot.density.density()` to better deal with linear/log spaced and a ternary heatmap

### pyrolite.comp

- Added `EMCOMP()` to `pyrolite.comp.impute`
- Renamed `inv_alr`, `inv_clr`, `inv_ilr` and `inv_boxcox` to `inverse_alr()`, `inverse_clr()`, `inverse_ilr()` and `inverse_boxcox()`

## pyrolite.util

- Added `pyrolite.util.synthetic`
- Moved `pyrolite.util.pd.normal_frame` and `pyrolite.util.pd.normal_series` to `pyrolite.util.synthetic.normal_frame()` and `pyrolite.util.synthetic.normal_series()`
- Added `pyrolite.util.missing` and `pyrolite.util.missing.md_pattern()`
- Added `pyrolite.util.math.eigsorted()`, `pyrolite.util.math.augmented_covariance_matrix()`, `pyrolite.util.math.interpolate_line()`

---

**Note:** Releases before 0.1.0 are available via [GitHub](#) for reference, but were alpha versions which were never considered stable.

---

## 4.3 Future

This page details some of the under-development and planned features for `pyrolite`. Note that while no schedules are attached, features under development are likely to be completed with weeks to months, while those ‘On The Horizon’ may be significantly further away (or in some cases may not make it to release).

### 4.3.1 Under Development

These features are either under development or planned to be implemented and should be released in the near future.

- **Interactive Plotting Backend Options:** `pyrolite` visualisation is currently based entirely on static plot generation via `matplotlib`. While this works well for publication-style figures, it may be possible to leverage `pandas`-based frameworks to provide options for alternative backends, some of which are more interactive and amendable to data exploration (e.g. `hvplot`, `plotly`). See the `pandas` extension docs for one option for implementing this ([plotting-backends](#)).
- **Units and Uncertainty Aware Geochemistry DataFrames:** Work has started on a prototype implementation to bring in native data types related to units, and incorporate uncertainties which can be propagated through calculations. This is principally based around `pint`, `pint_pandas` and `uncertainties`.

### 4.3.2 On the Horizon, Potential Future Updates

These are a number of features which are in various stages of development, which are planned be integrated over the longer term.

#### `pyrolite.mineral`

There are a few components which will make better use of mineral chemistry data, and facilitate better comparison of whole-rock and mineral data ([Issue #5](#)):

- Normative mineral calculations
- Mineral formulae recalculation, site partitioning, endmember calculations

#### `pyrolite.geochem.isotope`

- Stable isotope calculations
- Simple radiogenic isotope system calculations and plots

`pyrolite.comp.impute`

Expansion of compositional data imputation algorithms beyond EMCOMP ([Issue #6](#)).

`pyrolite.geochem.magma`

Utilities for simple melting and fractionation models.

`pyrolite.geochem.quality`

**Utilities for:**

- assessing data quality
- identifying potential analytical artefacts
- assessing uncertainties

### 4.3.3 Governance and Documentation

- Depending on how the community grows, and whether `pyrolite` brings with it a series of related tools, the project and related tools may be migrated to an umbrella organization on GitHub (e.g. `pyrolite/pyrolite` ) so they can be collectively managed by a community.
- **Internationalization:** While the `pyrolite` source is documented in English, it would be good to be able to provide translated versions of the documentation to minimise hurdles to getting started.
- **Teaching Resources:** `pyrolite` is well placed to provide solutions and resources for use in under/post-graduate education. While we have documentation sections dedicated to examples and tutorials, perhaps we could develop explicit sections for educational resources and exercises.

## 4.4 Code of Conduct

### 4.4.1 Our Pledge

In the interest of fostering an open and welcoming environment, we as contributors and maintainers pledge to making participation in our project and our community a harassment-free experience for everyone, regardless of age, body size, disability, ethnicity, sex characteristics, gender identity and expression, level of experience, education, socio-economic status, nationality, personal appearance, race, religion, or sexual identity and orientation.

### 4.4.2 Our Standards

Examples of behaviour that contributes to creating a positive environment include:

- Using welcoming and inclusive language
- Being respectful of differing viewpoints and experiences
- Gracefully accepting constructive criticism
- Focusing on what is best for the community
- Showing empathy towards other community members

Examples of unacceptable behaviour by participants include:

- The use of sexualized language or imagery and unwelcome sexual attention or advances
- Trolling, insulting/derogatory comments, and personal or political attacks

- Public or private harassment
- Publishing others' private information, such as a physical or electronic address, without explicit permission
- Other conduct which could reasonably be considered inappropriate in a professional setting

### 4.4.3 Our Responsibilities

Project maintainers are responsible for clarifying the standards of acceptable behaviour and are expected to take appropriate and fair corrective action in response to any instances of unacceptable behaviour.

Project maintainers have the right and responsibility to remove, edit, or reject comments, commits, code, wiki edits, issues, and other contributions that are not aligned to this Code of Conduct, or to ban temporarily or permanently any contributor for other behaviours that they deem inappropriate, threatening, offensive, or harmful.

### 4.4.4 Scope

This Code of Conduct applies both within project spaces and in public spaces when an individual is representing the project or its community. Examples of representing a project or community include using an official project e-mail address, posting via an official social media account, or acting as an appointed representative at an online or offline event. Representation of a project may be further defined and clarified by project maintainers.

### 4.4.5 Enforcement

Instances of abusive, harassing, or otherwise unacceptable behaviour may be reported by contacting the project admins. All complaints will be reviewed and investigated and will result in a response that is deemed necessary and appropriate to the circumstances. The project team is obligated to maintain confidentiality with regard to the reporter of an incident. Further details of specific enforcement policies may be posted separately.

Project maintainers who do not follow or enforce the Code of Conduct in good faith may face temporary or permanent repercussions as determined by other members of the project's leadership.

### 4.4.6 Attribution

This Code of Conduct is adapted from the [Contributor Covenant](https://www.contributor-covenant.org/version/1/4/code-of-conduct.html) Version 1.4, available at <https://www.contributor-covenant.org/version/1/4/code-of-conduct.html>. The Contributor Covenant is released under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

For answers to common questions about this code of conduct, see <https://www.contributor-covenant.org/faq>.

## 4.5 Contributing

The long-term aim of this project is to be designed, built and supported by (and for) the geochemistry community. In the present, the majority of the work involves incorporating geological knowledge and frameworks into a practically useful core set of tools which can be later be expanded. As such, requests for features and bug reports are particularly valuable contributions, in addition to code and expanding the documentation. All individuals contributing to the project are expected to follow the [Code of Conduct](#), which outlines community expectations and responsibilities.

Also, be sure to add your name or GitHub username to the [contributors list](#).

---

**Note:** This project is currently in *beta*, and as such there's much work to be done.

---

### 4.5.1 Feature Requests

If you're new to Python, and want to implement a specific process, plot or framework as part of `pyrolite`, you can submit a [Feature Request](#). Perhaps also check the [Issues Board](#) first to see if someone else has suggested something similar (or if something is in development), and comment there.

### 4.5.2 Bug Reports

If you've tried to do something with `pyrolite`, but it didn't work, and googling error messages didn't help (or, if the error messages are full of `pyrolite.XX.xx`), you can submit a [Bug Report](#). Perhaps also check the [Issues Board](#) first to see if someone else is having the same issue, and comment there.

### 4.5.3 Contributing to Documentation

The [documentation](#) and [examples](#) for `pyrolite` are gradually being developed, and any contributions or corrections would be greatly appreciated. Currently the examples are patchy, and any 'getting started' guides would be a helpful addition.

**These pages serve multiple purposes:**

- A human-readable reference of the source code (compiled from docstrings).
- A set of simple examples to demonstrate use and utility.
- A place for developing extended examples<sup>1</sup>

### 4.5.4 Contributing Code

Code contributions are always welcome, whether it be small modifications or entire features. As the project gains momentum, check the [Issues Board](#) for outstanding issues, features under development. If you'd like to contribute, but you're not so experienced with Python, look for good `first issue` tags or email the maintainer for suggestions.

To contribute code, the place to start will be forking the source for `pyrolite` from [GitHub](#). Once forked, clone a local copy and from the repository directory you can install a development (editable) copy via `python setup.py develop`. To incorporate suggested changes back to into the project, push your changes to your remote fork, and then submit a pull request onto [pyrolite/develop](#) or a relevant feature branch.

---

**Note:**

- See [Development](#) for directions for installing extra dependencies for development and for information on development environments and tests.
- `pyrolite` development roughly follows a [gitflow workflow](#). `pyrolite/main` is only used for releases, and large separable features should be build on `feature` branches off `develop`.
- Contributions introducing new functions, classes or entire features should also include appropriate tests where possible (see *Writing Tests*, below).
- `pyrolite` uses [Black](#) for code formatting, and submissions which have passed through `Black` are appreciated, although not critical.

---

<sup>1</sup> Such examples could easily be distributed as educational resources showcasing the utility of programmatic approaches to geochemistry



### 4.5.5 Writing Tests

There is currently a broad unit test suite for `pyrolite`, which guards against breaking changes and assures baseline functionality. `pyrolite` uses continuous integration via [GitHub Actions](#), where the full suite of tests are run for each commit and pull request, and test coverage output to [Coveralls](#).

Adding or expanding tests is a helpful way to ensure `pyrolite` does what is meant to, and does it reproducibly. The unit test suite one critical component of the package, and necessary to enable sufficient trust to use `pyrolite` for scientific purposes.

## 4.6 Contributors

This list includes people who have contributed to the project in the form of code, comments, testing, bug reports, or feature requests.

- [Morgan Williams](#)
- [Hayden Dalton](#)
- [Louise Schoneveld](#)
- [Adam Bath](#)
- [Yajing Mao](#)
- [Justin Gosses](#)
- [Kaarel Mand](#)
- [Laura Miller](#)
- [Steve Barnes](#)
- [Lucy Mathieson](#)
- [Nicolas Piette-Lauziere](#)
- [Chetan Nathwani](#)
- [Martin Bentley](#)
- [Tom Buckle](#)
- [Angela Rodrigues](#)
- [Sarah Shi](#)
- [Ondrej Lexa](#)
- [Bob Myhill](#)
- [Malte Mues](#)



## 5.1 API

### 5.1.1 pyrolite.plot

Submodule with various plotting and visualisation functions.

#### pyrolite.plot.pyroplot (Pandas Interface)

**class** `pyrolite.plot.pyroplot(obj)`

**cooccurrence**(*ax=None, normalize=True, log=False, colorbar=False, \*\*kwargs*)

Plot the co-occurrence frequency matrix for a given input.

**Parameters**

- **ax** (`matplotlib.axes.Axes`, `None`) – The subplot to draw on.
- **normalize** (`bool`) – Whether to normalize the cooccurrence to compare disparate variables.
- **log** (`bool`) – Whether to take the log of the cooccurrence.
- **colorbar** (`bool`) – Whether to append a colorbar.

**Returns**

Axes on which the cooccurrence plot is added.

**Return type**

`matplotlib.axes.Axes`

**density**(*components: Optional[list] = None, ax=None, axlabels=True, \*\*kwargs*)

Method for plotting histograms (mode='hist2d'/'hexbin') or kernel density estimates from point data. Convenience access function to `density()` (see *Other Parameters*, below), where further parameters for relevant `matplotlib` functions are also listed.

**Parameters**

- **components** (`list`, `None`) – Elements or compositional components to plot.
- **ax** (`matplotlib.axes.Axes`, `None`) – The subplot to draw on.
- **axlabels** (`bool`, `True`) – Whether to add x-y axis labels.

---

**Note:** The following additional parameters are from `pyrolite.plot.density.density()`.

---

**Other Parameters**

- **arr** (`numpy.ndarray`) – Dataframe from which to draw data.

- **logx** (`bool`, `False`) – Whether to use a logspaced *grid* on the x axis. Values strictly >0 required.
- **logy** (`bool`, `False`) – Whether to use a logspaced *grid* on the y axis. Values strictly >0 required.
- **bins** (`int`, 20) – Number of bins used in the gridded functions (histograms, KDE evaluation grid).
- **mode** (`str`, 'density') – Different modes used here: ['density', 'hexbin', 'hist2d']
- **extent** (`list`) – Predetermined extent of the grid for which to from the histogram/KDE. In the general form (xmin, xmax, ymin, ymax).
- **contours** (`list`) – Contours to add to the plot, where mode='density' is used.
- **percentiles** (`bool`, `True`) – Whether contours specified are to be converted to percentiles.
- **relim** (`bool`, `True`) – Whether to relimit the plot based on xmin, xmax values.
- **cmap** (`matplotlib.colors.Colormap`) – Colormap for mapping surfaces.
- **vmin** (`float`, 0.) – Minimum value for colormap.
- **shading** (`str`, 'auto') – Shading to apply to pcolormesh.
- **colorbar** (`bool`, `False`) – Whether to append a linked colorbar to the generated mappable image.

**Returns**

Axes on which the density diagram is plotted.

**Return type**

`matplotlib.axes.Axes`

**heatscatter**(*components*: *Optional*[*list*] = *None*, *ax*=*None*, *axlabels*=*True*, *logx*=*False*, *logy*=*False*, *\*\*kwargs*)

Heatmapped scatter plots using the pyroplot API. See further parameters for *matplotlib.pyplot.scatter* function below.

**Parameters**

- **components** (`list`, *None*) – Elements or compositional components to plot.
- **ax** (`matplotlib.axes.Axes`, *None*) – The subplot to draw on.
- **axlabels** (`bool`, `True`) – Whether to add x-y axis labels.
- **logx** (`bool`, `False`) – Whether to log-transform x values before the KDE for bivariate plots.
- **logy** (`bool`, `False`) – Whether to log-transform y values before the KDE for bivariate plots.

**Returns**

Axes on which the heatmapped scatterplot is added.

**Return type**

`matplotlib.axes.Axes`

**parallel**(*components*=*None*, *rescale*=*False*, *legend*=*False*, *ax*=*None*, *\*\*kwargs*)

Create a *pyrolite.plot.parallel.parallel()*. coordinate plot from the columns of the *DataFrame*.

**Parameters**

- **components** (`list`, *None*) – Components to use as axes for the plot.
- **rescale** (`bool`) – Whether to rescale values to [-1, 1].
- **legend** (`bool`, `False`) – Whether to include or suppress the legend.
- **ax** (`matplotlib.axes.Axes`, *None*) – The subplot to draw on.

---

**Note:** The following additional parameters are from `pyrolite.plot.parallel.parallel()`.

---

#### Other Parameters

**df** (`pandas.DataFrame`) – Dataframe to create a plot from.

#### Returns

Axes on which the parallel coordinates plot is added.

#### Return type

`matplotlib.axes.Axes`

---

#### Todo:

- Adapt figure size based on number of columns.
- 

**plot**(*components: Optional[list] = None, ax=None, axlabels=True, \*\*kwargs*)

Convenience method for line plots using the pyroplot API. See further parameters for `matplotlib.pyplot.scatter` function below.

#### Parameters

- **components** (`list`, `None`) – Elements or compositional components to plot.
  - **ax** (`matplotlib.axes.Axes`, `None`) – The subplot to draw on.
  - **axlabels** (`bool`, `True`) – Whether to add x-y axis labels.
- 

**Note:** The following additional parameters are from `matplotlib.pyplot.plot()`.

---

#### Other Parameters

- **x, y** (*array-like or scalar*) – The horizontal / vertical coordinates of the data points. *x* values are optional and default to `range(len(y))`.

Commonly, these parameters are 1D arrays.

They can also be scalars, or two-dimensional (in that case, the columns represent separate data sets).

These arguments cannot be passed as keywords.

- **fmt** (*str, optional*) – A format string, e.g. 'ro' for red circles. See the *Notes* section for a full description of the format strings.

Format strings are just an abbreviation for quickly setting basic line properties. All of these and more can also be controlled by keyword arguments.

This argument cannot be passed as keyword.

- **data** (*indexable object, optional*) – An object with labelled data. If given, provide the label names to plot in *x* and *y*.

---

**Note:** Technically there's a slight ambiguity in calls where the second label is a valid *fmt*. `plot('n', 'o', data=obj)` could be `plt(x, y)` or `plt(y, fmt)`. In such cases, the former interpretation is chosen, but a warning is issued. You may suppress the warning by adding an empty format string `plot('n', 'o', '', data=obj)`.

---

#### Returns

Axes on which the plot is added.

**Return type**`matplotlib.axes.Axes`

**REE**(*index*='elements', *ax*=None, *mode*='plot', *dropPm*=True, *scatter\_kw*={}, *line\_kw*={}, *\*\*kwargs*)

Pass the pandas object to `pyrolite.plot.spider.REE_v_radrii()`.

**Parameters**

- **ax** (`matplotlib.axes.Axes`, None) – The subplot to draw on.
- **index** (`str`) – Whether to plot radii ('radii') on the principal x-axis, or elements ('elements').
- **mode** (`str`, :code:`["plot", "fill", "binkde", "ckde", "kde", "hist"]`) – Mode for plot. Plot will produce a line-scatter diagram. Fill will return a filled range. Density will return a conditional density diagram.
- **dropPm** (`bool`) – Whether to exclude the (almost) non-existent element Promethium from the REE list.
- **scatter\_kw** (`dict`) – Keyword parameters to be passed to the scatter plotting function.
- **line\_kw** (`dict`) – Keyword parameters to be passed to the line plotting function.

---

**Note:** The following additional parameters are from `pyrolite.plot.spider.REE_v_radrii()`.

---

**Other Parameters**

- **arr** (`numpy.ndarray`) – Data array.
- **ree** (`list`) – List of REE to use as an index.
- **logy** (`bool`) – Whether to use a log y-axis.
- **tl\_rotation** (`float`) – Rotation of the numerical index labels in degrees.
- **unity\_line** (`bool`) – Add a line at y=1 for reference.
- **set\_labels** (`bool`) – Whether to set the x-axis ticklabels for the REE.
- **set\_ticks** (`bool`) – Whether to set the x-axis ticks according to the specified index.

**Returns**

Axes on which the REE plot is added.

**Return type**`matplotlib.axes.Axes`

**scatter**(*components*: *Optional*[*list*] = None, *ax*=None, *axlabels*=True, *\*\*kwargs*)

Convenience method for scatter plots using the pyroplot API. See further parameters for `matplotlib.pyplot.scatter` function below.

**Parameters**

- **components** (`list`, None) – Elements or compositional components to plot.
- **ax** (`matplotlib.axes.Axes`, None) – The subplot to draw on.
- **axlabels** (`bool`, True) – Whether to add x-y axis labels.

---

**Note:** The following additional parameters are from `matplotlib.pyplot.scatter()`.

---

**Other Parameters**

- **x, y** (*float or array-like, shape (n, )*) – The data positions.
- **s** (*float or array-like, shape (n, ), optional*) – The marker size in points\*\*2 (typographic points are 1/72 in.). Default is

```
rcParams['lines.markersize'] ** 2.
```

The linewidth and edgecolor can visually interact with the marker size, and can lead to artifacts if the marker size is smaller than the linewidth.

If the linewidth is greater than 0 and the edgecolor is anything but *'none'*, then the effective size of the marker will be increased by half the linewidth because the stroke will be centered on the edge of the shape.

To eliminate the marker edge either set *linewidth=0* or *edgecolor='none'*.

- **c** (*array-like or list of colors or color, optional*) – The marker colors. Possible values:
  - A scalar or sequence of n numbers to be mapped to colors using *cmap* and *norm*.
  - A 2D array in which the rows are RGB or RGBA.
  - A sequence of colors of length n.
  - A single color format string.

Note that *c* should not be a single numeric RGB or RGBA sequence because that is indistinguishable from an array of values to be colormapped. If you want to specify the same RGB or RGBA value for all points, use a 2D array with a single row. Otherwise, value-matching will have precedence in case of a size matching with *x* and *y*.

If you wish to specify a single color for all points prefer the *color* keyword argument.

Defaults to *None*. In that case the marker color is determined by the value of *color*, *facecolor* or *facecolors*. In case those are not specified or *None*, the marker color is determined by the next color of the Axes' current "shape and fill" color cycle. This cycle defaults to `rcParams["axes.prop_cycle"]`.

- **marker** (*~matplotlib.markers.MarkerStyle, default: rcParams["scatter.marker"]*) – The marker style. *marker* can be either an instance of the class or the text shorthand for a particular marker. See `matplotlib.markers` for more information about marker styles.
- **cmap** (*str or ~matplotlib.colors.Colormap, default: rcParams["image.cmap"]*) – The Colormap instance or registered colormap name used to map scalar data to colors.

This parameter is ignored if *c* is RGB(A).

- **norm** (*str or ~matplotlib.colors.Normalize, optional*) – The normalization method used to scale scalar data to the [0, 1] range before mapping to colors using *cmap*. By default, a linear scaling is used, mapping the lowest value to 0 and the highest to 1.

If given, this can be one of the following:

- An instance of *.Normalize* or one of its subclasses (see `Colormap normalization`).
- A scale name, i.e. one of "linear", "log", "symlog", "logit", etc. For a list of available scales, call `matplotlib.scale.get_scale_names()`. In that case, a suitable *.Normalize* subclass is dynamically generated and instantiated.

This parameter is ignored if *c* is RGB(A).

- **vmin, vmax** (*float, optional*) – When using scalar data and no explicit *norm*, *vmin* and *vmax* define the data range that the colormap covers. By default, the colormap covers the complete value range of the supplied

data. It is an error to use *vmin/vmax* when a *norm* instance is given (but using a *str norm* name together with *vmin/vmax* is acceptable).

This parameter is ignored if *c* is RGB(A).

- **alpha** (*float, default: None*) – The alpha blending value, between 0 (transparent) and 1 (opaque).
  - **linewidths** (*float or array-like, default: rcParams["lines.linewidth"]*) – The linewidth of the marker edges. Note: The default *edgecolors* is 'face'. You may want to change this as well.
  - **edgecolors** (*{ 'face', 'none', None } or color or sequence of color, default: rcParams["scatter.edgecolors"]*) – The edge color of the marker. Possible values:
    - 'face': The edge color will always be the same as the face color.
    - 'none': No patch boundary will be drawn.
    - A color or sequence of colors.
- For non-filled markers, *edgecolors* is ignored. Instead, the color is determined like with 'face', i.e. from *c*, *colors*, or *facecolors*.
- **plotnonfinite** (*bool, default: False*) – Whether to plot points with non-finite *c* (i.e. *inf*, *-inf* or *nan*). If *True* the points are drawn with the *bad* colormap color (see *.Colormap.set\_bad*).

#### Returns

Axes on which the scatterplot is added.

#### Return type

`matplotlib.axes.Axes`

**spider**(*components: Optional[list] = None, indexes: Optional[list] = None, ax=None, mode='plot', index\_order=None, autoscale=True, scatter\_kw={}, line\_kw={}, \*\*kwargs*)

Method for spider plots. Convenience access function to *spider()* (see *Other Parameters*, below), where further parameters for relevant *matplotlib* functions are also listed.

#### Parameters

- **components** (*list, None*) – Elements or compositional components to plot.
- **indexes** (*list, None*) – Elements or compositional components to plot.
- **ax** (*matplotlib.axes.Axes, None*) – The subplot to draw on.
- **index\_order** – Function to order spider plot indexes (e.g. by incompatibility).
- **autoscale** (*bool*) – Whether to autoscale the y-axis limits for standard spider plots.
- **mode** (*str, :code:`["plot", "fill", "binkde", "ckde", "kde", "hist"]`*) – Mode for plot. Plot will produce a line-scatter diagram. Fill will return a filled range. Density will return a conditional density diagram.
- **scatter\_kw** (*dict*) – Keyword parameters to be passed to the scatter plotting function.
- **line\_kw** (*dict*) – Keyword parameters to be passed to the line plotting function.



---

**Note:** The following additional parameters are from `pyrolite.plot.spider.spider()`.

---

#### Other Parameters

- **arr** (`numpy.ndarray`) – Data array.
- **label** (`str`, `None`) – Label for the individual series.
- **logy** (`bool`) – Whether to use a log y-axis.
- **yextent** (`tuple`) – Extent in the y direction for conditional probability plots, to limit the gridspace over which the kernel density estimates are evaluated.
- **unity\_line** (`bool`) – Add a line at y=1 for reference.
- **set\_ticks** (`bool`) – Whether to set the x-axis ticks according to the specified index.

#### Returns

Axes on which the spider diagram is plotted.

#### Return type

`matplotlib.axes.Axes`

---

#### Todo:

- Add ‘compositional data’ filter for default components if `None` is given
- 

**stem**(*components*: *Optional*[*list*] = *None*, *ax*=*None*, *orientation*='horizontal', *axlabels*=*True*, *\*\*kwargs*)

Method for creating stem plots. Convenience access function to `stem()` (see *Other Parameters*, below), where further parameters for relevant `matplotlib` functions are also listed.

#### Parameters

- **components** (`list`, `None`) – Elements or compositional components to plot.
- **ax** (`matplotlib.axes.Axes`, `None`) – The subplot to draw on.
- **orientation** (`str`) – Orientation of the plot (horizontal or vertical).
- **axlabels** (`bool`, `True`) – Whether to add x-y axis labels.

---

**Note:** The following additional parameters are from `pyrolite.plot.stem.stem()`.

---

#### Other Parameters

**x**, **y** (`numpy.ndarray`) – 1D arrays for independent and dependent axes.

#### Returns

Axes on which the stem diagram is plotted.

**Return type**`matplotlib.axes.Axes`**pyrolite.plot.spider**

```
pyrolite.plot.spider.spider(arr, indexes=None, ax=None, label=None, logy=True,
                             yextent=None, mode='plot', unity_line=False,
                             scatter_kw={}, line_kw={}, set_ticks=True,
                             autoscale=True, **kwargs)
```

Plots spidergrams for trace elements data. Additional arguments are typically forwarded to respective `matplotlib` functions `plot()` and `scatter()` (see Other Parameters, below).

**Parameters**

- **arr** (`numpy.ndarray`) – Data array.
- **indexes** (: `numpy.ndarray`) – Numerical indexes of x-axis positions.
- **ax** (`matplotlib.axes.Axes`, `None`) – The subplot to draw on.
- **label** (`str`, `None`) – Label for the individual series.
- **logy** (`bool`) – Whether to use a log y-axis.
- **yextent** (`tuple`) – Extent in the y direction for conditional probability plots, to limit the gridspace over which the kernel density estimates are evaluated.
- **mode** (`str`, [`"plot"`, `"fill"`, `"binkde"`, `"ckde"`, `"kde"`, `"hist"`]) – Mode for plot. Plot will produce a line-scatter diagram. Fill will return a filled range. Density will return a conditional density diagram.
- **unity\_line** (`bool`) – Add a line at y=1 for reference.
- **scatter\_kw** (`dict`) – Keyword parameters to be passed to the scatter plotting function.
- **line\_kw** (`dict`) – Keyword parameters to be passed to the line plotting function.
- **set\_ticks** (`bool`) – Whether to set the x-axis ticks according to the specified index.
- **autoscale** (`bool`) – Whether to autoscale the y-axis limits for standard spider plots.

---

**Note:** The following additional parameters are from `matplotlib.pyplot.scatter()`.

---

**Other Parameters**

- **x, y** (*float or array-like, shape (n, )*) – The data positions.
- **s** (*float or array-like, shape (n, ), optional*) – The marker size in points\*\*2 (typographic points are 1/72 in.). Default is `rcParams['lines.markersize'] ** 2`.

The linewidth and edgcolor can visually interact with the marker size, and can lead to artifacts if the marker size is smaller than the linewidth.

If the linewidth is greater than 0 and the edgcolor is anything but *'none'*, then the effective size of the marker will be increased by half the linewidth because the stroke will be centered on the edge of the shape.

To eliminate the marker edge either set *linewidth=0* or *edgcolor='none'*.

- **c** (*array-like or list of colors or color, optional*) – The marker colors. Possible values:
  - A scalar or sequence of n numbers to be mapped to colors using *cmap* and *norm*.
  - A 2D array in which the rows are RGB or RGBA.
  - A sequence of colors of length n.
  - A single color format string.

Note that *c* should not be a single numeric RGB or RGBA sequence because that is indistinguishable from an array of values to be colormapped. If you want to specify the same RGB or RGBA value for all points, use a 2D array with a single row. Otherwise, value-matching will have precedence in case of a size matching with *x* and *y*.

If you wish to specify a single color for all points prefer the *color* keyword argument.

Defaults to *None*. In that case the marker color is determined by the value of *color*, *facecolor* or *facecolors*. In case those are not specified or *None*, the marker color is determined by the next color of the Axes' current "shape and fill" color cycle. This cycle defaults to `rcParams["axes.prop_cycle"]`.

- **marker** (*~matplotlib.markers.MarkerStyle*, default: `rcParams["scatter.marker"]`) – The marker style. *marker* can be either an instance of the class or the text shorthand for a particular marker. See `matplotlib.markers` for more information about marker styles.
- **cmap** (*str or ~matplotlib.colors.Colormap*, default: `rcParams["image.cmap"]`) – The Colormap instance or registered colormap name used to map scalar data to colors.

This parameter is ignored if *c* is RGB(A).

- **norm** (*str or ~matplotlib.colors.Normalize*, optional) – The normalization method used to scale scalar data to the [0, 1] range before mapping to colors using *cmap*. By default, a linear scaling is used, mapping the lowest value to 0 and the highest to 1.

If given, this can be one of the following:

- An instance of *.Normalize* or one of its subclasses (see `Colormap normalization`).

- A scale name, i.e. one of “linear”, “log”, “symlog”, “logit”, etc. For a list of available scales, call `matplotlib.scale.get_scale_names()`. In that case, a suitable `.Normalize` subclass is dynamically generated and instantiated.

This parameter is ignored if *c* is RGB(A).

- **vmin, vmax** (*float, optional*) – When using scalar data and no explicit *norm*, *vmin* and *vmax* define the data range that the colormap covers. By default, the colormap covers the complete value range of the supplied data. It is an error to use *vmin/vmax* when a *norm* instance is given (but using a *str norm* name together with *vmin/vmax* is acceptable).

This parameter is ignored if *c* is RGB(A).

- **alpha** (*float, default: None*) – The alpha blending value, between 0 (transparent) and 1 (opaque).
- **linewidths** (*float or array-like, default: rcParams["lines.linewidth"]*) – The linewidth of the marker edges. Note: The default *edgecolors* is ‘face’. You may want to change this as well.
- **edgecolors** (*{‘face’, ‘none’, None} or color or sequence of color, default: rcParams["scatter.edgecolors"]*) – The edge color of the marker. Possible values:
  - ‘face’: The edge color will always be the same as the face color.
  - ‘none’: No patch boundary will be drawn.
  - A color or sequence of colors.

For non-filled markers, *edgecolors* is ignored. Instead, the color is determined like with ‘face’, i.e. from *c*, *colors*, or *facecolors*.

- **plotnonfinite** (*bool, default: False*) – Whether to plot points with nonfinite *c* (i.e. `inf`, `-inf` or `nan`). If `True` the points are drawn with the *bad* colormap color (see `.Colormap.set_bad`).

---

**Note:** The following additional parameters are from `matplotlib.pyplot.plot()`.

---

### Other Parameters

- **x, y** (*array-like or scalar*) – The horizontal / vertical coordinates of the data points. *x* values are optional and default to `range(len(y))`.

Commonly, these parameters are 1D arrays.

They can also be scalars, or two-dimensional (in that case, the columns represent separate data sets).

These arguments cannot be passed as keywords.

- **fmt** (*str, optional*) – A format string, e.g. ‘ro’ for red circles. See the *Notes* section for a full description of the format strings.

Format strings are just an abbreviation for quickly setting basic line properties. All of these and more can also be controlled by keyword arguments.

This argument cannot be passed as keyword.

- **data** (*indexable object, optional*) – An object with labelled data. If given, provide the label names to plot in *x* and *y*.

---

**Note:** Technically there's a slight ambiguity in calls where the second label is a valid *fmt*. `plot('n', 'o', data=obj)` could be `plt(x, y)` or `plt(y, fmt)`. In such cases, the former interpretation is chosen, but a warning is issued. You may suppress the warning by adding an empty format string `plot('n', 'o', '', data=obj)`.

---

### Returns

Axes on which the spiderplot is plotted.

### Return type

`matplotlib.axes.Axes`

### Notes

By using separate lines and scatterplots, values between two missing items are still presented.

---

### Todo:

- Might be able to speed up lines with `~matplotlib.collections.LineCollection`.
  - Legend entries
- 

### See also:

Functions:

```
matplotlib.pyplot.plot()      matplotlib.pyplot.scatter()
REE_v_radii()
```

```
pyrolite.plot.spider.REE_v_radii(arr=None, ax=None, ree=['La', 'Ce', 'Pr', 'Nd',
    'Sm', 'Eu', 'Gd', 'Tb', 'Dy', 'Ho', 'Er', 'Tm', 'Yb',
    'Lu'], index='elements', mode='plot', logy=True,
    tl_rotation=60, unity_line=False,
    scatter_kw={}, line_kw={}, set_labels=True,
    set_ticks=True, **kwargs)
```

Creates an axis for a REE diagram with ionic radii along the x axis.

### Parameters

- **arr** (`numpy.ndarray`) – Data array.
- **ax** (`matplotlib.axes.Axes`, `None`) – Optional designation of axes to reconfigure.
- **ree** (`list`) – List of REE to use as an index.

- **index** (`str`) – Whether to plot using radii on the x-axis ('radii'), or elements ('elements').
- **mode** (`str`, ["plot", "fill", "binkde", "ckde", "kde", "hist"]) – Mode for plot. Plot will produce a line-scatter diagram. Fill will return a filled range. Density will return a conditional density diagram.
- **logy** (`bool`) – Whether to use a log y-axis.
- **tl\_rotation** (`float`) – Rotation of the numerical index labels in degrees.
- **unity\_line** (`bool`) – Add a line at y=1 for reference.
- **scatter\_kw** (`dict`) – Keyword parameters to be passed to the scatter plotting function.
- **line\_kw** (`dict`) – Keyword parameters to be passed to the line plotting function.
- **set\_labels** (`bool`) – Whether to set the x-axis ticklabels for the REE.
- **set\_ticks** (`bool`) – Whether to set the x-axis ticks according to the specified index.

---

**Note:** The following additional parameters are from `pyrolite.plot.spider.spider()`.

---

#### Other Parameters

- **indexes** (: `numpy.ndarray`) – Numerical indexes of x-axis positions.
- **label** (`str`, `None`) – Label for the individual series.
- **yextent** (`tuple`) – Extent in the y direction for conditional probability plots, to limit the gridspace over which the kernel density estimates are evaluated.
- **autoscale** (`bool`) – Whether to autoscale the y-axis limits for standard spider plots.

#### Returns

Axes on which the REE\_v\_radii plot is added.

#### Return type

`matplotlib.axes.Axes`

---

#### Todo:

- Turn this into a plot template within `pyrolite.plot.templates` submodule
- 

#### See also:

Functions:

```
matplotlib.pyplot.plot()      matplotlib.pyplot.scatter()
spider() pyrolite.geochem.transform.lambda_lnREE()
```

## pyrolite.plot.density

Kernel density estimation plots for geochemical data.

```
pyrolite.plot.density.density(arr, ax=None, logx=False, logy=False, bins=25,
                             mode='density', extent=None, contours=[],
                             percentiles=True, relim=True,
                             cmap=<matplotlib.colors.ListedColormap object>,
                             shading='auto', vmin=0.0, colorbar=False,
                             **kwargs)
```

Creates diagrammatic representation of data density and/or frequency for either binary diagrams (X-Y) or ternary plots. Additional arguments are typically forwarded to respective `matplotlib` functions `pcolormesh()`, `hist2d()`, `hexbin()`, `contour()`, and `contourf()` (see Other Parameters, below).

### Parameters

- **arr** (`numpy.ndarray`) – Dataframe from which to draw data.
- **ax** (`matplotlib.axes.Axes`, `None`) – The subplot to draw on.
- **logx** (`bool`, `False`) – Whether to use a logspaced *grid* on the x axis. Values strictly >0 required.
- **logy** (`bool`, `False`) – Whether to use a logspaced *grid* on the y axis. Values strictly >0 required.
- **bins** (`int`, 20) – Number of bins used in the gridded functions (histograms, KDE evaluation grid).
- **mode** (`str`, 'density') – Different modes used here: ['density', 'hexbin', 'hist2d']
- **extent** (`list`) – Predetermined extent of the grid for which to from the histogram/KDE. In the general form (xmin, xmax, ymin, ymax).
- **contours** (`list`) – Contours to add to the plot, where mode='density' is used.
- **percentiles** (`bool`, `True`) – Whether contours specified are to be converted to percentiles.
- **relim** (`bool`, `True`) – Whether to relimit the plot based on xmin, xmax values.
- **cmap** (`matplotlib.colors.Colormap`) – Colormap for mapping surfaces.
- **vmin** (`float`, 0.) – Minimum value for colormap.
- **shading** (`str`, 'auto') – Shading to apply to `pcolormesh`.
- **colorbar** (`bool`, `False`) – Whether to append a linked colorbar to the generated mappable image.

---

**Note:** The following additional parameters are from `matplotlib.pyplot.pcolormesh()`.

---

### Other Parameters

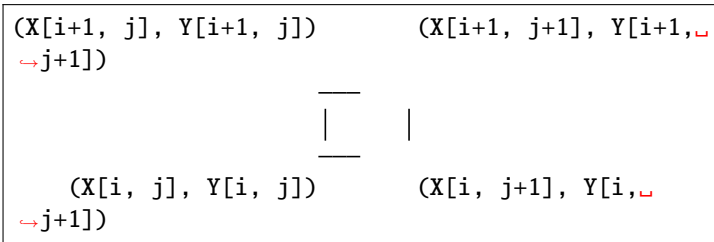
- **C** (*array-like*) – The mesh data. Supported array shapes are:

- (M, N) or M\*N: a mesh with scalar data. The values are mapped to colors using normalization and a colormap. See parameters *norm*, *cmap*, *vmin*, *vmax*.
- (M, N, 3): an image with RGB values (0-1 float or 0-255 int).
- (M, N, 4): an image with RGBA values (0-1 float or 0-255 int), i.e. including transparency.

The first two dimensions (M, N) define the rows and columns of the mesh data.

- **X, Y** (*array-like, optional*) –

The coordinates of the corners of quadrilaterals of a `pcolormesh`:



Note that the column index corresponds to the x-coordinate, and the row index corresponds to y. For details, see the [Notes](#) section below.

If `shading='flat'` the dimensions of *X* and *Y* should be one greater than those of *C*, and the quadrilateral is colored due to the value at *C*[*i*, *j*]. If *X*, *Y* and *C* have equal dimensions, a warning will be raised and the last row and column of *C* will be ignored.

If `shading='nearest'` or `'gouraud'`, the dimensions of *X* and *Y* should be the same as those of *C* (if not, a `ValueError` will be raised). For `'nearest'` the color *C*[*i*, *j*] is centered on  $(X[i, j], Y[i, j])$ . For `'gouraud'`, a smooth interpolation is carried out between the quadrilateral corners.

If *X* and/or *Y* are 1-D arrays or column vectors they will be expanded as needed into the appropriate 2D arrays, making a rectangular grid.

- **norm** (str or `~matplotlib.colors.Normalize`, optional) – The normalization method used to scale scalar data to the [0, 1] range before mapping to colors using *cmap*. By default, a linear scaling is used, mapping the lowest value to 0 and the highest to 1.

If given, this can be one of the following:

- An instance of `.Normalize` or one of its subclasses (see [Colormap normalization](#)).
- A scale name, i.e. one of “linear”, “log”, “symlog”, “logit”, etc. For a list of available scales, call `matplotlib.scale.get_scale_names()`. In that case, a suitable `.Normalize` subclass is dynamically generated and instantiated.
- **vmin, vmax** (*float, optional*) – When using scalar data and no explicit *norm*, *vmin* and *vmax* define the data range that the colormap covers. By default, the colormap covers the complete value range of



the supplied data. It is an error to use *vmin/vmax* when a *norm* instance is given (but using a *str norm* name together with *vmin/vmax* is acceptable).

- **edgecolors** (*{'none', None, 'face', color, color sequence}, optional*)
  - The color of the edges. Defaults to 'none'. Possible values:
    - 'none' or '': No edge.
    - *None*: `rcParams["patch.edgecolor"]` will be used. Note that currently `rcParams["patch.force_edgecolor"]` has to be `True` for this to work.
    - 'face': Use the adjacent face color.
    - A color or sequence of colors will set the edge color.

The singular form *edgecolor* works as an alias.

- **alpha** (*float, default: None*) – The alpha blending value, between 0 (transparent) and 1 (opaque).
- **snap** (*bool, default: False*) – Whether to snap the mesh to pixel boundaries.
- **rasterized** (*bool, optional*) – Rasterize the `pcolormesh` when drawing vector graphics. This can speed up rendering and produce smaller files for large data sets. See also `/gallery/misc/rasterization_demo`.

---

**Note:** The following additional parameters are from `matplotlib.pyplot.hist2d()`.

---

#### Other Parameters

- **x, y** (*array-like, shape (n, )*) – Input values
- **range** (*array-like shape(2, 2), optional*) – The leftmost and rightmost edges of the bins along each dimension (if not specified explicitly in the bins parameters): `[[xmin, xmax], [ymin, ymax]]`. All values outside of this range will be considered outliers and not tallied in the histogram.
- **density** (*bool, default: False*) – Normalize histogram. See the documentation for the *density* parameter of `~Axes.hist` for more details.
- **weights** (*array-like, shape (n, ), optional*) – An array of values `w_i` weighing each sample `(x_i, y_i)`.
- **cmin, cmax** (*float, default: None*) – All bins that has count less than *cmin* or more than *cmax* will not be displayed (set to `NaN` before passing to `~Axes.pcolormesh`) and these count values in the return value count histogram will also be set to `nan` upon return.

---

**Note:** The following additional parameters are from `matplotlib.pyplot.hexbin()`.

---

### Other Parameters

- **x, y** (*array-like*) – The data positions. *x* and *y* must be of the same length.
- **C** (*array-like, optional*) – If given, these values are accumulated in the bins. Otherwise, every point has a value of 1. Must be of the same length as *x* and *y*.
- **gridsize** (*int or (int, int), default: 100*) – If a single int, the number of hexagons in the *x*-direction. The number of hexagons in the *y*-direction is chosen such that the hexagons are approximately regular.

Alternatively, if a tuple (*nx, ny*), the number of hexagons in the *x*-direction and the *y*-direction. In the *y*-direction, counting is done along vertically aligned hexagons, not along the zig-zag chains of hexagons; see the following illustration.

To get approximately regular hexagons, choose  $n_x = \sqrt{3} n_y$ .

- **xscale** (*{'linear', 'log'}, default: 'linear'*) – Use a linear or log10 scale on the horizontal axis.
- **yscale** (*{'linear', 'log'}, default: 'linear'*) – Use a linear or log10 scale on the vertical axis.
- **mincnt** (*int >= 0, default: \*None\**) – If not *None*, only display cells with at least *mincnt* number of points in the cell.
- **marginals** (*bool, default: \*False\**) – If *marginals* is *True*, plot the marginal density as colormapped rectangles along the bottom of the *x*-axis and left of the *y*-axis.

---

**Note:** The following additional parameters are from `matplotlib.pyplot.contour()`.

---

### Other Parameters

- **X, Y** (*array-like, optional*) – The coordinates of the values in *Z*.  
*X* and *Y* must both be 2D with the same shape as *Z* (e.g. created via `numpy.meshgrid`), or they must both be 1-D such that `len(X) == N` is the number of columns in *Z* and `len(Y) == M` is the number of rows in *Z*.  
*X* and *Y* must both be ordered monotonically.  
If not given, they are assumed to be integer indices, i.e. `X = range(N)`, `Y = range(M)`.
- **Z** (*(M, N) array-like*) – The height values over which the contour is drawn. Color-mapping is controlled by *cmap*, *norm*, *vmin*, and *vmax*.
- **levels** (*int or array-like, optional*) – Determines the number and positions of the contour lines / regions.

If an int  $n$ , use `~matplotlib.ticker.MaxNLocator`, which tries to automatically choose no more than  $n+1$  “nice” contour levels between minimum and maximum numeric values of  $Z$ .

If array-like, draw contour lines at the specified levels. The values must be in increasing order.

---

**Note:** The following additional parameters are from `matplotlib.pyplot.contourf()`.

---

### Other Parameters

- **X, Y** (*array-like, optional*) – The coordinates of the values in  $Z$ .

$X$  and  $Y$  must both be 2D with the same shape as  $Z$  (e.g. created via `numpy.meshgrid`), or they must both be 1-D such that `len(X) == N` is the number of columns in  $Z$  and `len(Y) == M` is the number of rows in  $Z$ .

$X$  and  $Y$  must both be ordered monotonically.

If not given, they are assumed to be integer indices, i.e. `X = range(N)`, `Y = range(M)`.

- **Z** ( $(M, N)$  *array-like*) – The height values over which the contour is drawn. Color-mapping is controlled by `cmap`, `norm`, `vmin`, and `vmax`.
- **levels** (*int or array-like, optional*) – Determines the number and positions of the contour lines / regions.

If an int  $n$ , use `~matplotlib.ticker.MaxNLocator`, which tries to automatically choose no more than  $n+1$  “nice” contour levels between minimum and maximum numeric values of  $Z$ .

If array-like, draw contour lines at the specified levels. The values must be in increasing order.

### Returns

- `matplotlib.axes.Axes` – Axes on which the densityplot is plotted.
- `.. seealso::` –

Functions:

```
matplotlib.pyplot.pcolormesh()      matplotlib.  
pyplot.hist2d() matplotlib.pyplot.contourf()
```

## Notes

The default density estimates and derived contours are generated based on kernel density estimates. Assumptions around e.g. 95% of points lying within a 95% contour won't necessarily be valid for non-normally distributed data (instead, this represents the approximate 95% percentile on the kernel density estimate). Note that contours are currently only generated; for `mode="density"`; future updates may allow the use of a histogram basis, which would give results closer to 95% data percentiles.

---

### Todo:

- **Allow generation of contours from histogram data, rather than just the kernel density estimate.**
  - **Implement an option and filter to 'scatter' points below the minimum threshold or maximum percentile contours.**
- 

## pyrolite.plot.density.grid

```
class pyrolite.plot.density.grid.DensityGrid(x, y, extent=None, bins=50,
                                              logx=False, logy=False,
                                              coverage_scale=1.2)

    calculate_grid()
    get_ystep()
    get_xstep()
    extent_from_xy(x, y, coverage_scale=None)
    get_xrange()
    get_yrange()
    get_extent()
    get_range()
    update_grid_centre_ticks()
    update_grid_edge_ticks()
    get_centre_grid()
    get_edge_grid()
    get_hex_extent()

    kdefrom(xy, xtransform=<function DensityGrid.<lambda>>, ytransform=<function
        DensityGrid.<lambda>>, mode='centres', bw_method=None)
        Take an x-y array and sample a KDE on the grid.
```

**pyrolite.plot.density.ternary**

```
pyrolite.plot.density.ternary.ternary_heatmap(data, bins=20, mode='density',
                                              transform=<function ILR>,
                                              inverse_transform=<function
                                              inverse_ILR>,
                                              ternary_min_value=0.0001,
                                              grid_border_frac=0.1,
                                              grid=None, **kwargs)
```

Heatmap for ternary diagrams. This invokes a 3D to 2D transform such as a log transform prior to creating a grid.

**Parameters**

- **data** (`numpy.ndarray`) – Ternary data to obtain heatmap coords from.
- **bins** (`int`) – Number of bins for the grid.
- **mode** (`str`, {'histogram', 'density'}) – Which mode to render the histogram/KDE in.
- **transform** (callable | `sklearn.base.TransformerMixin`) – Callable function or Transformer class.
- **inverse\_transform** (callable) – Inverse function for *transform*, necessary if transformer class not specified.
- **ternary\_min\_value** (`float`) – Optional specification of minimum values within a ternary diagram to draw the transformed grid.
- **grid\_border\_frac** (`float`) – Size of border around the grid, expressed as a fraction of the total grid range.
- **grid** (`numpy.ndarray`) – Grid coordinates to sample at, if already calculated. For the density mode, this is a (nsamples, 2) array. For histograms, this is a two-member list of bin edges.

**Returns**

- **t, l, r** (tuple of `numpy.ndarray`) – Ternary coordinates for the heatmap.
- **H** (`numpy.ndarray`) – Histogram/density estimates for the coordinates.
- **data** (`dict`) – Data dictionary with grid arrays and relevant information.

**Notes**

Zeros will not render in this heatmap, consider replacing zeros with small values or imputing them if they must be incorporated.

## pyrolite.plot.parallel

```
pyrolite.plot.parallel.parallel(df, components=None, classes=None,  
                               rescale=True, legend=False, ax=None,  
                               label_rotate=60, **kwargs)
```

Create a parallel coordinate plot across dataframe columns, with individual lines for each row.

### Parameters

- **df** (`pandas.DataFrame`) – Dataframe to create a plot from.
- **components** (`list`) – Subset of dataframe columns to use as indexes along the x-axis.
- **rescale** (`bool`) – Whether to rescale values to [-1, 1].
- **legend** (`bool`, `False`) – Whether to include or suppress the legend.
- **ax** (`matplotlib.axes.Axes`) – Axis to plot on (optional).

---

### Todo:

- A multi-axis version would be more compatible with independent rescaling and zoom
- Enable passing a list of colors

Rather than just a list of numbers to be converted to colors.

---

## pyrolite.plot.stem

```
pyrolite.plot.stem.stem(x, y, ax=None, orientation='horizontal', **kwargs)
```

Create a stem (or ‘lollipop’) plot, with optional orientation.

### Parameters

- **x, y** (`numpy.ndarray`) – 1D arrays for independent and dependent axes.
- **ax** (`matplotlib.axes.Axes`, `None`) – The subplot to draw on.
- **orientation** (`str`) – Orientation of the plot (horizontal or vertical).

### Returns

Axes on which the stem diagram is plotted.

### Return type

`matplotlib.axes.Axes`

**pyrolite.plot.biplot**

`pyrolite.plot.biplot.compositional_SVD(X: ndarray)`

Breakdown a set of compositions to vertexes and cases for adding to a compositional biplot.

**Parameters**

**X** (`numpy.ndarray`) – Compositional array.

**Returns**

**vertexes, cases**

**Return type**

`numpy.ndarray, numpy.ndarray`

`pyrolite.plot.biplot.plot_origin_to_points(xs, ys, labels=None, ax=None, origin=(0, 0), color='k', marker='o', pad=0.05, **kwargs)`

Plot lines radiating from a specific origin. Formulated for creation of biplots (`covariance_biplot()`, `compositional_biplot()`).

**Parameters**

- **xs, ys** (`numpy.ndarray`) – Coordinates for points to add.
- **labels** (`list`) – Labels for verticies.
- **ax** (`matplotlib.axes.Axes`) – Axes to plot on.
- **origin** (`tuple`) – Origin to plot from.
- **color** (`str`) – Line color to use.
- **marker** (`str`) – Marker to use for ends of vectors and origin.
- **pad** (`float`) – Fraction of vector to pad text label.

**Returns**

Axes on which radial plot is added.

**Return type**

`matplotlib.axes.Axes`

`pyrolite.plot.biplot.compositional_biplot(data, labels=None, ax=None, **kwargs)`

Create a compositional biplot.

**Parameters**

- **data** (`numpy.ndarray`) – Coordinates for points to add.
- **labels** (`list`) – Labels for verticies.
- **ax** (`matplotlib.axes.Axes`) – Axes to plot on.

**Returns**

Axes on which biplot is added.

**Return type**

`matplotlib.axes.Axes`

## pyrolite.plot.templates

A utility submodule for standardised plot templates to be added to matplotlib axes.

---

### Todo:

- Make use of new `ax.axline` features ([https://matplotlib.org/3.3.1/users/whats\\_new.html#new-axes-axline-method](https://matplotlib.org/3.3.1/users/whats_new.html#new-axes-axline-method))
- 

`pyrolite.plot.templates.pearceThNbYb(ax=None, relim=True, color='k',  
 **kwargs)`

Adds the Th-Nb-Yb delimiter lines from Pearce (2008)<sup>1</sup> to an axes.

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axes to add the template onto.
- **relim** (`bool`) – Whether to relimit axes to fit the built in ranges for this diagram.
- **color** (`str`) – Line color for the diagram.

### References

#### Returns

**ax**

#### Return type

`matplotlib.axes.Axes`

`pyrolite.plot.templates.pearceTiNbYb(ax=None, relim=True, color='k',  
 annotate=True, **kwargs)`

Adds the Ti-Nb-Yb delimiter lines from Pearce (2008)<sup>2</sup> to an axes.

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axes to add the template onto.
- **relim** (`bool`) – Whether to relimit axes to fit the built in ranges for this diagram.
- **color** (`str`) – Line color for the diagram.

---

<sup>1</sup> Pearce J. A. (2008) Geochemical fingerprinting of oceanic basalts with applications to ophiolite classification and the search for Archean oceanic crust. *Lithos* 100, 14–48. doi: 10.1016/j.lithos.2007.06.016

<sup>2</sup> Pearce J. A. (2008) Geochemical fingerprinting of oceanic basalts with applications to ophiolite classification and the search for Archean oceanic crust. *Lithos* 100, 14–48. doi: {pearce2008}



## References

### Returns

**ax**

### Return type

`matplotlib.axes.Axes`

`pyrolite.plot.templates.JensenPlot(ax=None, add_labels=False, color='k',  
**kwargs)`

Jensen Plot for classification of sub-alkaline volcanic rocks<sup>3</sup>.

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axes to add the template on to.
- **add\_labels** (`bool`) – Whether to include the labels for the diagram.
- **color** (`str`) – Line color for the diagram.

### Returns

**ax**

### Return type

`matplotlib.axes.Axes`

## References

### Notes

Diagram used for the classification of subalkalic volcanic rocks. The diagram is constructed for molar cation percentages of Al, Fe+Ti and Mg, on account of these elements' stability upon metamorphism. This particular version uses updated labels relative to Jensen (1976), in which the fields have been extended to the full range of the ternary plot.

`pyrolite.plot.templates.TAS(ax=None, add_labels=False, which_labels='ID',  
relim=True, color='k', which_model=None, **kwargs)`

Adds the TAS diagram to an axes. Diagram from Middlemost (1994)<sup>4</sup>, a closed-polygon variant after Le Bas et al (1992)<sup>5</sup>.

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axes to add the template on to.
- **add\_labels** (`bool`) – Whether to add labels at polygon centroids.
- **which\_labels** (`str`) – Which labels to add to the polygons (e.g. for TAS, 'volcanic', 'intrusive' or the field 'ID').
- **relim** (`bool`) – Whether to relimit axes to fit the built in ranges for this diagram.
- **color** (`str`) – Line color for the diagram.

<sup>3</sup> Jensen, L. S. (1976). A new cation plot for classifying sub-alkaline volcanic rocks. Ontario Division of Mines. Miscellaneous Paper No. 66.

<sup>4</sup> Middlemost, E. A. K. (1994). Naming materials in the magma/igneous rock system. Earth-Science Reviews, 37(3), 215–224. doi: 10.1016/0012-8252(94)90029-9

<sup>5</sup> Le Bas, M.J., Le Maitre, R.W., Woolley, A.R. (1992). The construction of the Total Alkali-Silica chemical classification of volcanic rocks. Mineralogy and Petrology 46, 1–22. doi: 10.1007/BF01160698

- **which\_model** (*str*) – The name of the model variant to use, if not Middlemost.

**Returns**

**ax**

**Return type**

`matplotlib.axes.Axes`

**References**

`pyrolite.plot.templates.USDAsoilTexture(ax=None, add_labels=False, color='k', **kwargs)`

United States Department of Agriculture Soil Texture classification model<sup>67</sup>.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axes to add the template on to.
- **add\_labels** (*bool*) – Whether to include the labels for the diagram.
- **color** (*str*) – Line color for the diagram.

**Returns**

**ax**

**Return type**

`matplotlib.axes.Axes`

**References**

`pyrolite.plot.templates.QAP(ax=None, add_labels=False, which_labels='ID', color='k', **kwargs)`

IUGS QAP ternary classification diagram<sup>89</sup>.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Ternary axes to add the diagram to.
- **add\_labels** (*bool*) – Whether to add labels at polygon centroids.
- **which\_labels** (*str*) – Which data to use for field labels - field 'name' or 'ID'.
- **color** (*str*) – Color for the polygon edges in the diagram.

---

<sup>6</sup> Soil Science Division Staff (2017). Soil Survey Manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

<sup>7</sup> Thien, Steve J. (1979). A Flow Diagram for Teaching Texture-by-Feel Analysis. Journal of Agronomic Education 8:54–55. doi: 10.2134/jae.1979.0054

<sup>8</sup> Streckeisen, A. Classification and nomenclature of plutonic rocks recommendations of the IUGS subcommission on the systematics of Igneous Rocks. Geol Rundsch 63, 773–786 (1974). doi: 10.1007/BF01820841

<sup>9</sup> Le Maitre, R.W. 2002. Igneous Rocks: A Classification and Glossary of Terms : Recommendations of International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks. Cambridge University Press, 236pp

## References

```
pyrolite.plot.templates.FeldsparTernary(ax=None, add_labels=False,
                                         which_labels='ID',
                                         mode='miscibility-gap', color='k',
                                         **kwargs)
```

Simplified feldspar classification diagram, based on a version printed in the second edition of ‘An Introduction to the Rock Forming Minerals’ (Deer, Howie and Zussman).  
[#pyrolite.plot.templates.feldspar.FeldsparTernary\_1]

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Ternary axes to add the diagram to.
- **add\_labels** (`bool`) – Whether to add labels at polygon centroids.
- **which\_labels** (`str`) – Which data to use for field labels - field ‘name’ or ‘ID’.
- **mode** (`str`) – Which mode of the diagram to use; the two implemented for the feldspar ternary diagram are ‘default’ and ‘miscibility-gap’, the second of which provides a simplified approximation of the miscibility gap between k-feldspar and plagioclase.
- **color** (`str`) – Color for the polygon edges in the diagram.

## References

```
pyrolite.plot.templates.SpinelFeBivariate(ax=None, add_labels=False,
                                           which_labels='ID', color='k',
                                           **kwargs)
```

Fe-Spinel classification, designed for data in atoms per formula unit.

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axes to add the diagram to.
- **add\_labels** (`bool`) – Whether to add labels at polygon centroids.
- **which\_labels** (`str`) – Which data to use for field labels - field ‘name’ or ‘ID’.
- **color** (`str`) – Color for the polygon edges in the diagram.

```
pyrolite.plot.templates.SpinelTrivalentTernary(ax=None, add_labels=False,
                                                which_labels='ID', color='k',
                                                **kwargs)
```

Spinel Trivalent Ternary classification - designed for data in atoms per formula unit.

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Ternary axes to add the diagram to.
- **add\_labels** (`bool`) – Whether to add labels at polygon centroids.
- **which\_labels** (`str`) – Which data to use for field labels - field ‘name’ or ‘ID’.
- **color** (`str`) – Color for the polygon edges in the diagram.

```
pyrolite.plot.templates.Pettijohn(ax=None, add_labels=False,  
                                   which_labels='ID', relim=True, color='k',  
                                   **kwargs)
```

Adds the Pettijohn (1973) [#pyrolite.plot.templates.sandstones.Pettijohn\_1] sandstones classification diagram.

#### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axes to add the template on to.
- **add\_labels** (`bool`) – Whether to add labels at polygon centroids.
- **which\_labels** (`str`) – Which data to use for field labels - field 'name' or 'ID'.
- **relim** (`bool`) – Whether to relimit axes to fit the built in ranges for this diagram.
- **color** (`str`) – Line color for the diagram.

#### Returns

**ax**

#### Return type

`matplotlib.axes.Axes`

#### References

```
pyrolite.plot.templates.Herron(ax=None, add_labels=False, which_labels='ID',  
                                relim=True, color='k', **kwargs)
```

Adds the Herron (1988) [#pyrolite.plot.templates.sandstones.Herron\_1] sandstones classification diagram.

#### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axes to add the template on to.
- **add\_labels** (`bool`) – Whether to add labels at polygon centroids.
- **which\_labels** (`str`) – Which data to use for field labels - field 'name' or 'ID'.
- **relim** (`bool`) – Whether to relimit axes to fit the built in ranges for this diagram.
- **color** (`str`) – Line color for the diagram.

#### Returns

**ax**

#### Return type

`matplotlib.axes.Axes`

## References

### pyrolite.plot.color

`pyrolite.plot.color.get_cmode(c=None)`

Find which mode a color is supplied as, such that it can be processed.

#### Parameters

**c** (`str` | `list` | `tuple` | `numpy.ndarray`) – Color arguments as typically passed to `matplotlib.pyplot.scatter()` or `matplotlib.pyplot.plot()`.

`pyrolite.plot.color.process_color(c=None, color=None, cmap=None, alpha=None, norm=None, bad='0.5', cmap_under=(1, 1, 1, 0.0), color_converter=<function to_rgba>, color_mappings={}, size=None, **otherkwargs)`

Color argument processing for pyrolite plots, returning a standardised output.

#### Parameters

- **c** (`str` | `list` | `tuple` | `numpy.ndarray`) – Color arguments as typically passed to `matplotlib.pyplot.scatter()`.
- **color** (`str` | `list` | `tuple` | `numpy.ndarray`) – Color arguments as typically passed to `matplotlib.pyplot.plot()`
- **cmap** (`str` | `ScalarMappable`) – Colormap for mapping unknown color values.
- **alpha** (`float`) – Alpha to modulate color opacity.
- **norm** (`Normalize`) – Normalization for the colormap.
- **cmap\_under** (`str` | `tuple`) – Color for values below the lower threshold for the cmap.
- **color\_converter** – Function to use to convert colors (from strings, hex, tuples etc).
- **color\_mappings** (`dict`) – Dictionary containing category-color mappings for individual color variables, with the default color mapping having the key 'color'. For use where categorical values are specified for a color variable.
- **size** (`int`) – Size of the data array along the first axis.

#### Returns

C – Color returned in standardised RGBA format.

#### Return type

`tuple` | `numpy.ndarray`

### Notes

As formulated here, the addition of unused styling parameters may cause some properties (associated with 'c') to be set to None - and hence revert to defaults. This might be mitigated if the context could be checked - e.g. via checking keyword argument membership of `scatterkwargs()` etc.

## 5.1.2 pyrolite.geochem

Submodule for working with geochemical data.

### pyrolite.geochem.pyrochem (Pandas Interface)

```
class pyrolite.geochem.pyrochem(obj)
```

#### property list\_elements

Get the subset of columns which are element names.

##### Return type

`list`

### Notes

The list will have the same ordering as the source DataFrame.

#### property list\_isotope\_ratios

Get the subset of columns which are isotope ratios.

##### Return type

`list`

### Notes

The list will have the same ordering as the source DataFrame.

#### property list\_REE

Get the subset of columns which are Rare Earth Element names.

##### Return type

`list`

### Notes

The returned list will reorder REE based on atomic number.

#### property list\_REY

Get the subset of columns which are Rare Earth Element names.

##### Return type

`list`

### Notes

The returned list will reorder REE based on atomic number.

#### **property list\_oxides**

Get the subset of columns which are oxide names.

##### **Return type**

`list`

### Notes

The list will have the same ordering as the source DataFrame.

#### **property list\_compositional**

#### **property elements**

Get an elemental subset of a DataFrame.

##### **Return type**

`pandas.DataFrame`

#### **property REE**

Get a Rare Earth Element subset of a DataFrame.

##### **Return type**

`pandas.DataFrame`

#### **property REY**

Get a Rare Earth Element + Yttrium subset of a DataFrame.

##### **Return type**

`pandas.DataFrame`

#### **property oxides**

Get an oxide subset of a DataFrame.

##### **Return type**

`pandas.DataFrame`

#### **property isotope\_ratios**

Get an isotope ratio subset of a DataFrame.

##### **Return type**

`pandas.DataFrame`

#### **property compositional**

Get an oxide & elemental subset of a DataFrame.

##### **Return type**

`pandas.DataFrame`

### Notes

This will not include isotope ratios.

**parse\_chem**(*abbrv*=['ID', 'IGSN'], *split\_on*='[\s\_]+')

Convert column names to pyrolite-recognised elemental, oxide and isotope ratio column names where valid names are found.

**check\_multiple\_cation\_inclusion**(*exclude*=['LOI', 'FeOT', 'Fe2O3T'])

Returns cations which are present in both oxide and elemental form.

#### Parameters

**exclude** (*list*, ["LOI", "FeOT", "Fe2O3T"]) – List of components to exclude from the duplication check.

#### Returns

Set of elements for which multiple components exist in the dataframe.

#### Return type

*set*

**to\_molecular**(*renorm*=True)

Converts mass quantities to molar quantities.

#### Parameters

**renorm** (*bool*, True) – Whether to renormalise the dataframe after converting to relative moles.

### Notes

Does not convert units (i.e. mass% → mol%; mass-ppm → mol-ppm).

#### Returns

Transformed dataframe.

#### Return type

*pandas.DataFrame*

**to\_weight**(*renorm*=True)

Converts molar quantities to mass quantities.

#### Parameters

**renorm** (*bool*, True) – Whether to renormalise the dataframe after converting to relative moles.

### Notes

Does not convert units (i.e. mol% → mass%; mol-ppm → mass-ppm).

#### Returns

Transformed dataframe.

#### Return type

*pandas.DataFrame*



**devolatilise**(*exclude*=['H2O', 'H2O\_PLUS', 'H2O\_MINUS', 'CO2', 'LOI'],  
*renorm*=True)

Recalculates components after exclusion of volatile phases (e.g. H2O, CO2).

#### Parameters

- **exclude** ([list](#)) – Components to exclude from the dataset.
- **renorm** ([bool](#), True) – Whether to renormalise the dataframe after devolatilisation.

#### Returns

Transformed dataframe.

#### Return type

[pandas.DataFrame](#)

**elemental\_sum**(*component*=None, *to*=None, *total\_suffix*='T', *logdata*=False,  
*molecular*=False)

Sums abundance for a cation to a single series, starting from a dataframe containing multiple components with a single set of units.

#### Parameters

- **component** ([str](#)) – Component indicating which element to aggregate.
- **to** ([str](#)) – Component to cast the output as.
- **logdata** ([bool](#), False) – Whether data has been log transformed.
- **molecular** ([bool](#), False) – Whether to perform a sum of molecular data.

#### Returns

Series with cation aggregated.

#### Return type

[pandas.Series](#)

**aggregate\_element**(*to*, *total\_suffix*='T', *logdata*=False, *renorm*=False,  
*molecular*=False)

Aggregates cation information from oxide and elemental components to either a single species or a designated mixture of species.

#### Parameters

- **to** ([str](#) | [Element](#) | [Formula](#) | [dict](#)) – Component(s) to convert to. If one component is specified, the element will be converted to the target species.  
  
If more than one component is specified with proportions in a dictionary (e.g. {'FeO': 0.9, 'Fe2O3': 0.1}), the components will be split as a fraction of the elemental sum.
- **renorm** ([bool](#), True) – Whether to renormalise the dataframe after recalculation.
- **total\_suffix** ([str](#), 'T') – Suffix of 'total' variables. E.g. 'T' for FeOT, Fe2O3T.
- **logdata** ([bool](#), False) – Whether the data has been log transformed.

- **molecular** (`bool`, `False`) – Whether to perform a sum of molecular data.

## Notes

This won't convert units, so need to start from single set of units.

### Returns

Series with cation aggregated.

### Return type

`pandas.Series`

**recalculate\_Fe**(*to*='FeOT', *renorm*=`False`, *total\_suffix*='T', *logdata*=`False`, *molecular*=`False`)

Recalculates abundances of iron, and normalises a dataframe to contain either a single species, or multiple species in certain proportions.

### Parameters

- **to** (`str` | `Element` | `Formula` | `dict`) – Component(s) to convert to.  
  
If one component is specified, all iron will be converted to the target species.  
  
If more than one component is specified with proportions in a dictionary (e.g. `{ 'FeO': 0.9, 'Fe2O3': 0.1 }`), the components will be split as a fraction of Fe.
- **renorm** (`bool`, `False`) – Whether to renormalise the dataframe after recalculation.
- **total\_suffix** (`str`, 'T') – Suffix of 'total' variables. E.g. 'T' for FeOT, Fe2O3T.
- **logdata** (`bool`, `False`) – Whether the data has been log transformed.
- **molecular** (`bool`, `False`) – Flag that data is in molecular units, rather than weight units.

### Returns

Transformed dataframe.

### Return type

`pandas.DataFrame`

**get\_ratio**(*ratio*: `str`, *alias*: *Optional*[`str`] = `None`, *norm\_to*=`None`, *molecular*=`False`)

Add a ratio of components A and B, given in the form of string 'A/B'. Returned series be assigned an alias name.

### Parameters

- **ratio** (`str`) – String decription of ratio in the form A/B[\_n].
- **alias** (`str`) – Alternate name for ratio to be used as column name.
- **norm\_to** (`str` | `pyrolite.geochem.norm.Composition`, `None`) – Reference composition to normalise to.

- **molecular** (`bool`, `False`) – Flag that data is in molecular units, rather than weight units.

**Returns**

Dataframe with ratio appended.

**Return type**

`pandas.DataFrame`

**See also:**

`add_MgNo()`

**add\_ratio**(*ratio*: *str*, *alias*: *Optional[str]* = *None*, *norm\_to*=*None*, *molecular*=*False*)

Add a ratio of components A and B, given in the form of string 'A/B'. Returned series be assigned an alias name.

**Parameters**

- **ratio** (*str*) – String decription of ratio in the form A/B[\_n].
- **alias** (*str*) – Alternate name for ratio to be used as column name.
- **norm\_to** (*str* | *pyrolite.geochem.norm.Composition*, *None*) – Reference composition to normalise to.
- **molecular** (`bool`, `False`) – Flag that data is in molecular units, rather than weight units.

**Returns**

Dataframe with ratio appended.

**Return type**

`pandas.DataFrame`

**See also:**

`add_MgNo()`

**add\_MgNo**(*molecular*=*False*, *use\_total\_approx*=*False*, *approx\_Fe203\_frac*=*0.1*, *name*='Mg#')

Append the magnesium number to a dataframe.

**Parameters**

- **molecular** (`bool`, `False`) – Whether the input data is molecular.
- **use\_total\_approx** (`bool`, `False`) – Whether to use an approximate calculation using total iron rather than just FeO.
- **approx\_Fe203\_frac** (*float*) – Fraction of iron which is oxidised, used in approximation mentioned above.
- **name** (*str*) – Name to use for the Mg Number column.

**Returns**

Dataframe with ratio appended.

**Return type**

`pandas.DataFrame`

**See also:**

`add_ratio()`

**lambda\_lnREE**(*norm\_to*='ChondriteREE\_ON', *exclude*=['Pm', 'Eu'], *params*=None, *degree*=4, *scale*='ppm', *sigmas*=None, *\*\*kwargs*)

Calculates orthogonal polynomial coefficients (lambdas) for a given set of REE data, normalised to a specific composition<sup>1</sup>. Lambda factors are given for the radii vs.  $\ln(\text{REE}/\text{NORM})$  polynomial combination.

#### Parameters

- **norm\_to** (*str* | *Composition* | *numpy.ndarray*) – Which reservoir to normalise REE data to (defaults to "ChondriteREE\_ON").
- **exclude** (*list*, ["Pm", "Eu"]) – Which REE elements to exclude from the *fit*. May wish to include Ce for minerals in which Ce anomalies are common.
- **params** (*list* | *str*, None) – Pre-computed parameters for the orthogonal polynomials (a list of tuples). Optionally specified, otherwise defaults the parameterisation as in O'Neill (2016). If a string is supplied, "O'Neill (2016)" or similar will give the original defaults, while "full" will use all of the REE (including Eu) as a basis for the orthogonal polynomials.
- **degree** (*int*, 4) – Maximum degree polynomial fit component to include.
- **scale** (*str*) – Current units for the REE data, used to scale the reference dataset.
- **sigmas** (*float* | *numpy.ndarray* | *pandas.Series*) – Value or 1D array of fractional REE uncertainties (i.e.  $\sigma_{\text{REE}}/\text{REE}$ ).

#### References

See also:

`get_ionic_radii()`, `calc_lambdas()`, `orthogonal_polynomial_constants()`, `REE_radii_plot()`

**convert\_chemistry**(*to*=[], *logdata*=False, *renorm*=False, *molecular*=False)

Attempts to convert a dataframe with one set of components to another.

#### Parameters

- **to** (*list*) – Set of columns to try to extract from the dataframe.  
Can also include a dictionary for iron speciation. See `pyrolite.geochem.recalculate_Fe()`.
- **logdata** (*bool*, False) – Whether chemical data has been log transformed. Necessary for aggregation functions.
- **renorm** (*bool*, False) – Whether to renormalise the data after transformation.
- **molecular** (*bool*, False) – Flag that data is in molecular units, rather than weight units.

---

<sup>1</sup> O'Neill HSC (2016) The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. *J Petrology* 57:1463–1508. doi: 10.1093/petrology/egw047

**Returns**

Dataframe with converted chemistry.

**Return type**

`pandas.DataFrame`

---

**Todo:**

- Check for conflicts between oxides and elements
  - Aggregator for ratios
  - Implement generalised redox transformation.
  - Add check for dictionary components (e.g. Fe) in tests
- 

**normalize\_to**(*reference=None, units=None, convert\_first=False*)

Normalise a dataframe to a given reference composition.

**Parameters**

- **reference** (`str` | *Composition* | `numpy.ndarray`) – Reference composition to normalise to.
- **units** (`str`) – Units of the input dataframe, to convert the reference composition.
- **convert\_first** (`bool`) – Whether to first convert the reference composition before normalisation. This is useful where elements are presented as different components (e.g. Ti, TiO<sub>2</sub>).

**Returns**

Dataframe with normalised chemistry.

**Return type**

`pandas.DataFrame`

**Notes**

This assumes that dataframes have a single set of units.

**denormalize\_from**(*reference=None, units=None*)

De-normalise a dataframe from a given reference composition.

**Parameters**

- **reference** (`str` | *Composition* | `numpy.ndarray`) – Reference composition which the composition is normalised to.
- **units** (`str`) – Units of the input dataframe, to convert the reference composition.

**Returns**

Dataframe with normalised chemistry.

**Return type**

`pandas.DataFrame`

## Notes

This assumes that dataframes have a single set of units.

**scale**(*in\_unit*, *target\_unit*='ppm')

Scale a dataframe from one set of units to another.

### Parameters

- **in\_unit** (*str*) – Units to be converted from
- **target\_unit** (*str*, "ppm") – Units to scale to.

### Returns

Dataframe with new scale.

### Return type

`pandas.DataFrame`

## pyrolite.geochem.ind

Collections and indexes of elements and oxides, and functions for obtaining these and relevant properties (e.g. radii).

---

### Todo:

- Incompatibility indexes for spider plot ordering.
- 

**pyrolite.geochem.ind.common\_elements**(*cutoff*=92, *output*='string', *order*=None, *as\_set*=False)

Provides a list of elements up to a particular cutoff (by default including U).

### Parameters

- **cutoff** (*int*) – Upper cutoff on atomic number for the output list. Defaults to stopping at uranium (92).
- **output** (*str*) – Whether to return output list as formulae ('formula') or strings (anything else).
- **order** (callable) – Sorting function for elements.
- **as\_set** (*bool*, False) – Whether to return a `set` (True) or `list` (False).

### Returns

List of elements.

### Return type

`list` | `set`

## Notes

Formulae cannot be used as members of a set, and hence sets returned will instead consist only of strings.

---

### Todo:

- Implement ordering for e.g. incompatibility.
- 

`pyrolite.geochem.ind.REE(output='string', dropPm=True)`

Provides a list of Rare Earth Elements.

#### Parameters

- **output** (`str`) – Whether to return output list as formulae ('formula') or strings (anything else).
- **dropPm** (`bool`) – Whether to exclude the (almost) non-existent element Promethium from the REE list.

#### Returns

List of REE.

#### Return type

`list` | `set`

`pyrolite.geochem.ind.REY(output='string', dropPm=True)`

Provides a list of Rare Earth Elements, with the addition of Yttrium.

#### Parameters

- **output** (`str`) – Whether to return output list as formulae ('formula') or strings (anything else).

#### Returns

List of REE+Y.

#### Return type

`list` | `set`

## Notes

This currently modifies the hardcoded list of `REE()`, but could be adapted for different element ordering.

`pyrolite.geochem.ind.common_oxides(elements: list = [], output='string', addition: list = ['FeOT', 'Fe2O3T', 'LOI'], exclude=['O', 'He', 'Ne', 'Ar', 'Kr', 'Xe'], as_set=False)`

Creates a list of oxides based on a list of elements.

#### Parameters

- **elements** (`list`, `[]`) – List of elements to obtain oxide forms for.
- **output** (`str`) – Whether to return output list as formulae ('formula') or strings (anything else).
- **addition** (`list`, `[]`) – Additional components to append to the list.

- **exclude** (`list`) – Elements to not produce oxide forms for (e.g. oxygen, noble gases).
- **as\_set** (`bool`) – Whether to return a `set` (True) or `list` (False).

**Returns**

List of oxides.

**Return type**

`list` | `set`

**Notes**

Formulae cannot be used as members of a set, and hence sets returned will instead consist only of strings.

---

**Todo:**

- Element verification
  - Conditional additional components on the presence of others (e.g. Fe - FeOT)
- 

`pyrolite.geochem.ind.simple_oxides(cation, output='string')`

Creates a list of oxides for a cationic element (oxide of ions with c=1+ and above).

**Parameters**

- **cation** (`str` | `periodictable.core.Element`) – Cation to obtain oxide forms for.
- **output** (`str`) – Whether to return output list as formulae ('formula') or strings (anything else).

**Returns**

List of oxides.

**Return type**

`list` | `set`

`pyrolite.geochem.ind.get_cations(component: str, exclude=[], total_suffix='T')`

Returns the principal cations in an oxide component.

**Parameters**

- **component** (`str` | `periodictable.formulas.Formula`) – Component to obtain cations for.
- **exclude** (`list`) – Components to exclude, i.e. anions (e.g. O, Cl, F).

**Returns**

List of cations.

**Return type**

`list`

---

**Todo:**

- Consider implementing `periodictable.core.Element` return.
-



`pyrolite.geochem.ind.get_isotopes(ratio_text)`

Regex for isotope ratios.

#### Parameters

**ratio\_text** (`str`) – Text to extract isotope ratio components from.

#### Returns

Isotope ration numerator and denominator.

#### Return type

`list`

`pyrolite.geochem.ind.by_incompatibility(els, reverse=False)`

Order a list of elements by their relative ‘incompatibility’ given by a proxy of the relative abundances in Bulk Continental Crust over a Primitive Mantle Composition.

#### Parameters

- **els** (`list`) – List of element names to be reordered.
- **reverse** (`bool`) – Whether to reverse the ordering.

#### Returns

Reordered list of elements.

#### Return type

`list`

### Notes

Some elements are missing from this list, as as such will be omitted.

`pyrolite.geochem.ind.by_number(els, reverse=False)`

Order a list of elements by their atomic number.

#### Parameters

- **els** (`list`) – List of element names to be reordered.
- **reverse** (`bool`) – Whether to reverse the ordering.

#### Returns

Reordered list of elements.

#### Return type

`list`

`pyrolite.geochem.ind.get_ionic_radii(element, charge=None, coordination=None, variant=[], source='shannon', pauling=True, **kwargs)`

Function to obtain ionic radii for a given ion and coordination<sup>12</sup>.

#### Parameters

- **element** (`str` | `list`) – Element to obtain a radii for. If a list is passed, the function will be applied over each of the items.

<sup>1</sup> Shannon RD (1976). Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. Acta Crystallographica Section A 32:751–767. doi: [10.1107/S0567739476001551](https://doi.org/10.1107/S0567739476001551)

<sup>2</sup> Whittaker, E.J.W., Muntus, R., 1970. Ionic radii for use in geochemistry. Geochimica et Cosmochimica Acta 34, 945–956. doi: [10.1016/0016-7037\(70\)90077-3](https://doi.org/10.1016/0016-7037(70)90077-3)

- **charge** (`int`) – Charge of the ion to obtain a radii for. If unspecified will use the default charge from `pyrolite.mineral.ions`.
- **coordination** (`int`) – Coordination of the ion to obtain a radii for.
- **variant** (`list`) – List of strings specifying particular variants (here ‘squareplanar’ or ‘pyramidal’, ‘highspin’ or ‘lowspin’).
- **source** (`str`) – Name of the data source for ionic radii (‘shannon’<sup>?</sup> or ‘whittaker’<sup>?</sup>).
- **pauling** (`bool`) – Whether to use the radii consistent with Pauling (1960)<sup>3</sup> from the Shannon (1976) radii dataset<sup>?</sup>.

**Returns**

Series with viable ion charge and coordination, with associated radii in angstroms. If the ion charge and coordination are completely specified and found in the table, a single value will be returned instead.

**Return type**

`pandas.Series` | `float`

**Notes**

Shannon published two sets of radii. The first (‘Crystal Radii’) were using Shannon’s value for  $r(O_{VI}^{2-})$  of 1.26 Å, while the second (‘Ionic Radii’) is consistent with the Pauling (1960) value of  $r(O_{VI}^{2-})$  of 1.40 Å<sup>?</sup>.

**References**

---

**Todo:**

- Implement interpolation for coordination +/- charge.
- 

**pyrolite.geochem.transform**

Functions for converting, transforming and parameterizing geochemical data.

`pyrolite.geochem.transform.to_molecular(df: DataFrame, renorm=True)`

Converts mass quantities to molar quantities of the same order.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe to transform.
- **renorm** (`bool`, `True`) – Whether to renormalise the dataframe after converting to relative moles.

**Returns**

Transformed dataframe.

**Return type**

`pandas.DataFrame`

---

<sup>3</sup> Pauling, L., 1960. The Nature of the Chemical Bond. Cornell University Press, Ithaca, NY.

## Notes

Does not convert units (i.e. mass% → mol%; mass-ppm → mol-ppm).

`pyrolite.geochem.transform.to_weight(df: DataFrame, renorm=True)`

Converts molar quantities to mass quantities of the same order.

### Parameters

- **df** (`pandas.DataFrame`) – Dataframe to transform.
- **renorm** (`bool`, `True`) – Whether to renormalise the dataframe after converting to relative moles.

### Returns

Transformed dataframe.

### Return type

`pandas.DataFrame`

## Notes

Does not convert units (i.e. mol% → mass%; mol-ppm → mass-ppm).

`pyrolite.geochem.transform.devolatilise(df: DataFrame, exclude=['H2O', 'H2O_PLUS', 'H2O_MINUS', 'CO2', 'LOI'], renorm=True)`

Recalculates components after exclusion of volatile phases (e.g. H2O, CO2).

### Parameters

- **df** (`pandas.DataFrame`) – Dataframe to devolatilise.
- **exclude** (`list`) – Components to exclude from the dataset.
- **renorm** (`bool`, `True`) – Whether to renormalise the dataframe after devolatilisation.

### Returns

Transformed dataframe.

### Return type

`pandas.DataFrame`

`pyrolite.geochem.transform.oxide_conversion(oxin, oxout, molecular=False)`

Factory function to generate a function to convert oxide components between two elemental oxides, for use in redox recalculations.

### Parameters

- **oxin** (`str` | `Formula`) – Input component.
- **oxout** (`str` | `Formula`) – Output component.
- **molecular** (`bool`, `False`) – Whether to apply the conversion for molecular data.

### Returns

- Function to convert a `pandas.Series` from one element-oxide
- *component to another.*

```
pyrolite.geochem.transform.elemental_sum(df: DataFrame, component=None,  
                                         to=None, total_suffix='T',  
                                         logdata=False, molecular=False)
```

Sums abundance for a cation to a single series, starting from a dataframe containing multiple components with a single set of units.

#### Parameters

- **df** (`pandas.DataFrame`) – DataFrame for which to aggregate cation data.
- **component** (`str`) – Component indicating which element to aggregate.
- **to** (`str`) – Component to cast the output as.
- **logdata** (`bool`, `False`) – Whether data has been log transformed.
- **molecular** (`bool`, `False`) – Whether to perform a sum of molecular data.

#### Returns

Series with cation aggregated.

#### Return type

`pandas.Series`

```
pyrolite.geochem.transform.aggregate_element(df: DataFrame, to,  
                                              total_suffix='T', logdata=False,  
                                              renorm=False,  
                                              molecular=False)
```

Aggregates cation information from oxide and elemental components to either a single species or a designated mixture of species.

#### Parameters

- **df** (`pandas.DataFrame`) – DataFrame for which to aggregate cation data.
- **to** (`str` | `Element` | `Formula` | `dict`) – Component(s) to convert to. If one component is specified, the element will be converted to the target species.  
  
If more than one component is specified with proportions in a dictionary (e.g. `{ 'FeO': 0.9, 'Fe2O3': 0.1 }`), the components will be split as a fraction of the elemental sum.
- **renorm** (`bool`, `True`) – Whether to renormalise the dataframe after recalculation.
- **total\_suffix** (`str`, `'T'`) – Suffix of ‘total’ variables. E.g. ‘T’ for FeOT, Fe2O3T.
- **logdata** (`bool`, `False`) – Whether the data has been log transformed.
- **molecular** (`bool`, `False`) – Whether to perform a sum of molecular data.

## Notes

This won't convert units, so need to start from single set of units.

### Returns

Dataframe with cation aggregated to the desired species.

### Return type

`pandas.DataFrame`

```
pyrolite.geochem.transform.get_ratio(df: DataFrame, ratio: str, alias:
Optional[str] = None, norm_to=None,
molecular=False)
```

Get a ratio of components A and B, given in the form of string 'A/B'. Returned series be assigned an alias name.

### Parameters

- **df** (`pandas.DataFrame`) – Dataframe to append ratio to.
- **ratio** (`str`) – String decription of ratio in the form A/B[\_n].
- **alias** (`str`) – Alternate name for ratio to be used as column name.
- **norm\_to** (`str` | `pyrolite.geochem.norm.Composition`, `None`) – Reference composition to normalise to.
- **molecular** (`bool`, `False`) – Flag that data is in molecular units, rather than weight units.

### Returns

Dataframe with ratio appended.

### Return type

`pandas.DataFrame`

---

## Todo:

- Use elemental sum from reference compositions
- Use sympy-like functionality to accept arbitrary input for calculation

e.g. "MgNo = Mg / (Mg + Fe)"

---

## See also:

`add_MgNo()`

```
pyrolite.geochem.transform.add_MgNo(df: DataFrame, molecular=False,
use_total_approx=False,
approx_Fe2O3_frac=0.1, name='Mg#')
```

Append the magnesium number to a dataframe.

### Parameters

- **df** (`pandas.DataFrame`) – Input dataframe.
- **molecular** (`bool`, `False`) – Whether the input data is molecular.
- **use\_total\_approx** (`bool`, `False`) – Whether to use an approximate calculation using total iron rather than just FeO.

- **approx\_Fe203\_frac** (`float`) – Fraction of iron which is oxidised, used in approximation mentioned above.
- **name** (`str`) – Name to use for the Mg Number column.

**Returns**

Dataframe with ratio appended.

**Return type**

`pandas.DataFrame`

**See also:**

`get_ratio()`

```
pyrolite.geochem.transform.lambda_lnREE(df, norm_to='ChondriteREE_ON',  
                                         exclude=['Pm', 'Eu'], params=None,  
                                         degree=4, scale='ppm',  
                                         allow_missing=True, min_elements=7,  
                                         algorithm='O'Neill', sigmas=None,  
                                         **kwargs)
```

Calculates orthogonal polynomial coefficients (lambdas) for a given set of REE data, normalised to a specific composition<sup>1</sup>. Lambda coefficients are given for the polynomial regression of  $\ln(\text{REE}/\text{NORM})$  vs radii.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe to calculate lambda coefficients for.
- **norm\_to** (`str` | *Composition* | `numpy.ndarray`) – Which reservoir to normalise REE data to (defaults to "ChondriteREE\_ON").
- **exclude** (`list`, `["Pm", "Eu"]`) – Which REE elements to exclude from the *fit*. May wish to include Ce for minerals in which Ce anomalies are common.
- **params** (`list` | `str`, `None`) – Pre-computed parameters for the orthogonal polynomials (a list of tuples). Optionally specified, otherwise defaults the parameterisation as in O'Neill (2016). If a string is supplied, "O'Neill (2016)" or similar will give the original defaults, while "full" will use all of the REE (including Eu) as a basis for the orthogonal polynomials.
- **degree** (`int`, 4) – Maximum degree polynomial fit component to include.
- **scale** (`str`) – Current units for the REE data, used to scale the reference dataset.
- **allow\_missing** (`True`) – Whether to calculate lambdas for rows which might be missing values.
- **min\_elements** (`int`) – Minimum columns present to return lambda values.
- **algorithm** (`str`) – Algorithm to use for fitting the orthogonal polynomials.

---

<sup>1</sup> O'Neill HSC (2016) The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. *J Petrology* 57:1463–1508. doi: 10.1093/petrology/egw047

- **sigmas** (`float` | `numpy.ndarray` | `pandas.Series`) – Value or 1D array of fractional REE uncertainties (i.e.  $\sigma_{REE}/REE$ ).

---

**Todo:**

- Operate only on valid rows.
  - Add residuals, Eu, Ce anomalies as options.
- 

**References****See also:**

`get_ionic_radii()`, `calc_lambdas()`, `orthogonal_polynomial_constants()`, `REE_radii_plot()`

```
pyrolite.geochem.transform.convert_chemistry(input_df, to=[], total_suffix='T',
                                             renorm=False,
                                             molecular=False,
                                             logdata=False, **kwargs)
```

Attempts to convert a dataframe with one set of components to another.

**Parameters**

- **input\_df** (`pandas.DataFrame`) – Dataframe to convert.
- **to** (`list`) – Set of columns to try to extract from the dataframe.  
Can also include a dictionary for iron speciation. See `aggregate_element()`.
- **total\_suffix** (`str`, 'T') – Suffix of 'total' variables. E.g. 'T' for FeOT, Fe2O3T.
- **renorm** (`bool`, False) – Whether to renormalise the data after transformation.
- **molecular** (`bool`, False) – Flag that data is in molecular units, rather than weight units.
- **logdata** (`bool`, False) – Whether chemical data has been log transformed. Necessary for aggregation functions.

**Returns**

Dataframe with converted chemistry.

**Return type**

`pandas.DataFrame`

---

**Todo:**

- Check for conflicts between oxides and elements
  - Aggregator for ratios
  - Implement generalised redox transformation.
  - Add check for dictionary components (e.g. Fe) in tests
-

## pyrolite.geochem.norm

Reference compositions and compositional normalisation.

`pyrolite.geochem.norm.all_reference_compositions(path=None)`

Get a dictionary of all reference compositions indexed by name.

**Parameters**

**path** (`str` | `pathlib.Path`)

**Return type**

`dict`

`pyrolite.geochem.norm.get_reference_composition(name)`

Retrieve a particular composition from the reference database.

**Parameters**

**name** (`str`) – Name of the reference composition model.

**Return type**

`pyrolite.geochem.norm.Composition`

`pyrolite.geochem.norm.get_reference_files(directory=None, formats=['csv'])`

Get a list of the reference composition files.

**Parameters**

- **directory** (`str`, `None`) – Location of reference data files.
- **formats** (`list`, `["csv"]`) – List of potential data formats to draw from. Currently only csv will work.

**Return type**

`list`

`pyrolite.geochem.norm.update_database(path=None, encoding='cp1252',  
**kwargs)`

Update the reference composition database.

## Notes

This will take all csv files from the geochem/refcomp pyrolite data folder and construct a document-based JSON database.

`class pyrolite.geochem.norm.Composition(src, name=None, reference=None,  
reservoir=None, source=None,  
**kwargs)`

`set_units(to='wt%')`

Set the units of the dataframe.

**Parameters**

**to** (`str`, `"wt%"`)

`describe(verbose=True, **kwargs)`



## pyrolite.geochem.parse

Functions for parsing, formatting and validating chemical names and formulae.

`pyrolite.geochem.parse.is_isotoperatio(s)`

Check if text is plausibly an isotope ratio.

**Parameters**

**s** (`str`) – String to validate.

**Return type**

`bool`

---

**Todo:**

- Validate the isotope masses vs natural isotopes
- 

`pyrolite.geochem.parse.repr_isotope_ratio(isotope_ratio)`

Format an isotope ratio pair as a string.

**Parameters**

**isotope\_ratio** (`tuple`) – Numerator, denominator pair.

**Return type**

`str`

---

**Todo:** Consider returning additional text outside of the match (e.g.  $^{87}\text{Sr}/^{86}\text{Sr}$  should include the 'i').

---

`pyrolite.geochem.parse.ischem(s)`

Checks if a string corresponds to chemical component (compositional). Here simply checking whether it is a common element or oxide.

**Parameters**

**s** (`str`) – String to validate.

**Return type**

`bool`

---

**Todo:**

- Implement checking for other compounds, e.g. carbonates.
- 

`pyrolite.geochem.parse.tochem(strings: list, abbrv=['ID', 'IGSN'],  
split_on='[\s_]+')`

Converts a list of strings containing some chemical compounds to appropriate case.

**Parameters**

- **strings** (`list`) – Strings to convert to 'chemical case'.
- **abbr** (`list`, `["ID", "IGSN"]`) – Abbreviated phrases to ignore in capitalisation.
- **split\_on** (`str`, `"[s_]+"`) – Regex for character or phrases to split the strings on.

**Return type**`list | str`

`pyrolite.geochem.parse.check_multiple_cation_inclusion(df, exclude=['LOI', 'FeOT', 'Fe2O3T'])`

Returns cations which are present in both oxide and elemental form.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe to check duplication within.
- **exclude** (`list`, `["LOI", "FeOT", "Fe2O3T"]`) – List of components to exclude from the duplication check.

**Returns**

Set of elements for which multiple components exist in the dataframe.

**Return type**`set`

---

**Todo:**

- Options for output (string/formula).
- 

## pyrolite.geochem.magma

Submodule for calculating and modelling melt chemistry. Includes common functions for predicting and accounting for melt evolution.

`pyrolite.geochem.magma.FeAt8MgO(FeOT: float, MgO: float) → float`

To account for differences in the slopes and curvature of liquid lines of descent as a function of parental magma composition<sup>12</sup> (after<sup>3</sup>).

**Parameters**

- **FeOT** (`float`) – Iron oxide content.
- **MgO** (`float`) – Magnesium oxide content.

## References

`pyrolite.geochem.magma.NaAt8MgO(Na2O: float, MgO: float) → float`

To account for differences in the slopes and curvature of liquid lines of descent as a function of parental magma composition<sup>45</sup> (after<sup>6</sup>).

**Parameters**

---

<sup>1</sup> Castillo PR, Klein E, Bender J, et al (2000). Petrology and Sr, Nd, and Pb isotope geochemistry of mid-ocean ridge basalt glasses from the 11°45'N to 15°00'N segment of the East Pacific Rise. *Geochemistry, Geophysics, Geosystems* 1:1. doi: [10.1029/1999GC000024](https://doi.org/10.1029/1999GC000024)

<sup>2</sup> Klein EM, Langmuir CH (1987). Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *Journal of Geophysical Research: Solid Earth* 92:8089–8115. doi: [10.1029/JB092iB08p08089](https://doi.org/10.1029/JB092iB08p08089)

<sup>3</sup> Langmuir CH, Bender JF (1984). The geochemistry of oceanic basalts in the vicinity of transform faults: Observations and implications. *Earth and Planetary Science Letters* 69:107–127. doi: [10.1016/0012-821X\(84\)90077-3](https://doi.org/10.1016/0012-821X(84)90077-3)

<sup>4</sup> Castillo PR, Klein E, Bender J, et al (2000). Petrology and Sr, Nd, and Pb isotope geochemistry of mid-ocean ridge basalt glasses from the 11°45'N to 15°00'N segment of the East Pacific Rise. *Geochemistry, Geophysics, Geosystems* 1:1. doi: [10.1029/1999GC000024](https://doi.org/10.1029/1999GC000024)

<sup>5</sup> Klein EM, Langmuir CH (1987). Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *Journal of Geophysical Research: Solid Earth* 92:8089–8115. doi: [10.1029/JB092iB08p08089](https://doi.org/10.1029/JB092iB08p08089)

<sup>6</sup> Langmuir CH, Bender JF (1984). The geochemistry of oceanic basalts in the vicinity of transform faults: Observations and implications. *Earth and Planetary Science Letters* 69:107–127. doi: [10.1016/0012-821X\(84\)90077-3](https://doi.org/10.1016/0012-821X(84)90077-3)

- **Na2O** (`float`) – Iron oxide content.
- **MgO** (`float`) – Magnesium oxide content.

## References

`pyrolite.geochem.magma.SCSS(df, T, P, kelvin=False, grid=None, outunit='wt%')`  
 Obtain the sulfur content at sulfate and sulfide saturation<sup>7,8</sup>.

### Parameters

- **df** (`pandas.DataFrame`) – Dataframe of compositions.
- **T** (`float` | `numpy.ndarray`) – Temperature
- **P** (`float` | `numpy.ndarray`) – Pressure (kbar)
- **kelvin** (`bool`) – Whether temperature values are in kelvin (True) or celsius (False)
- **grid** (`None`, 'geotherm', 'grid') – Whether to consider temperature and pressure as a geotherm (geotherm), or independently (as a grid, grid).

### Returns

**sulfate, sulfide** – Arrays of mass fraction sulfate and sulfide abundances at saturation.

### Return type

`numpy.ndarray, numpy.ndarray`

## Notes

For anhydrite-saturated systems, the sulfur content at sulfate saturation is given by the following:

$$\ln(X_S) = 10.07 - 1.151 \cdot (10^4/T_K) + 0.104 \cdot P_{kbar} - 7.1 \cdot X_{SiO_2} - 14.02 \cdot X_{MgO} - 14.164 \cdot X_{Al_2O_3} \quad (5.1)$$

(5.3)

For sulfide-liquid saturated systems, the sulfur content at sulfide saturation is given by the following:

$$\ln(X_S) = -1.76 - 0.474 \cdot (10^4/T_K) + 0.021 \cdot P_{kbar} + 5.559 \cdot X_{FeO} + 2.565 \cdot X_{TiO_2} + 2.709 \cdot X_{Al_2O_3} - 3.192 \cdot X_{SiO_2} - 3.049 \cdot X_{H_2O} \quad (5.4)$$

(5.7)

<sup>7</sup> Li, C., and Ripley, E.M. (2009). Sulfur Contents at Sulfide-Liquid or Anhydrite Saturation in Silicate Melts: Empirical Equations and Example Applications. *Economic Geology* 104, 405–412. doi: [gsecongeo.104.3.405](https://doi.org/10.1016/j.econgeo.2009.03.005)

<sup>8</sup> Smythe, D.J., Wood, B.J., and Kiseeva, E.S. (2017). The S content of silicate melts at sulfide saturation: New experiments and a model incorporating the effects of sulfide composition. *American Mineralogist* 102, 795–803. doi: [10.2138/am-2017-5800CCBY](https://doi.org/10.2138/am-2017-5800CCBY)

## References

---

### Todo:

- Produce an updated version based on log-regressions?
  - Add updates from Smythe et al. (2017)?
- 

## pyrolite.geochem.alteration

Functions for calculating indexes of chemical alteration.

`pyrolite.geochem.alteration.CIA(df: DataFrame)`

Chemical Index of Alteration (molecular)<sup>1</sup>

### Parameters

`df` (`pandas.DataFrame`) – DataFrame to calculate index from.

### Returns

Alteration index series.

### Return type

`pandas.Series`

## References

`pyrolite.geochem.alteration.CIW(df: DataFrame)`

Chemical Index of Weathering (molecular)<sup>2</sup>

### Parameters

`df` (`pandas.DataFrame`) – DataFrame to calculate index from.

### Returns

Alteration index series.

### Return type

`pandas.Series`

## References

`pyrolite.geochem.alteration.PIA(df: DataFrame)`

Plagioclase Index of Alteration (molecular)<sup>3</sup>

### Parameters

`df` (`pandas.DataFrame`) – DataFrame to calculate index from.

### Returns

Alteration index series.

---

<sup>1</sup> Nesbitt HW, Young GM (1984). Prediction of some weathering trends of plutonic and volcanic rocks based on thermodynamic and kinetic considerations. *Geochimica et Cosmochimica Acta* 48:1523–1534. doi: [10.1016/0016-7037\(84\)90408-3](https://doi.org/10.1016/0016-7037(84)90408-3)

<sup>2</sup> Harnois L (1988). The CIW index: A new chemical index of weathering. *Sedimentary Geology* 55:319–322. doi: [10.1016/0037-0738\(88\)90137-6](https://doi.org/10.1016/0037-0738(88)90137-6)

<sup>3</sup> Fedo CM, Nesbitt HW, Young GM (1995). Unraveling the effects of potassium metasomatism in sedimentary rocks and paleosols, with implications for paleoweathering conditions and provenance. *Geology* 23:921–924. doi: [10.1130/0091-7613\(1995\)023<0921:UTEOPM>2.3.CO;2](https://doi.org/10.1130/0091-7613(1995)023<0921:UTEOPM>2.3.CO;2) [https://dx.doi.org/10.1130/0091-7613\(1995\)023<0921:UTEOPM>2.3.CO;2](https://dx.doi.org/10.1130/0091-7613(1995)023<0921:UTEOPM>2.3.CO;2)

**Return type**`pandas.Series`**References**`pyrolite.geochem.alteration.SAR(df: DataFrame)`

Silica-Alumina Ratio (molecular)

**Parameters**`df` (`pandas.DataFrame`) – DataFrame to calculate index from.**Returns**

Alteration index series.

**Return type**`pandas.Series``pyrolite.geochem.alteration.SiTiIndex(df: DataFrame)`Silica-Titania Index (molecular)<sup>4</sup>**Parameters**`df` (`pandas.DataFrame`) – DataFrame to calculate index from.**Returns**

Alteration index series.

**Return type**`pandas.Series`**References**`pyrolite.geochem.alteration.WIP(df: DataFrame)`Weathering Index of Parker (molecular)<sup>5</sup>**Parameters**`df` (`pandas.DataFrame`) – DataFrame to calculate index from.**Returns**

Alteration index series.

**Return type**`pandas.Series`**References**`pyrolite.geochem.alteration.IshikawaAltIndex(df: DataFrame)`Alteration Index of Ishikawa (wt%)<sup>6</sup>**Parameters**`df` (`pandas.DataFrame`) – DataFrame to calculate index from.

---

<sup>4</sup> Jayawardena U de S, Izawa E (1994). A new chemical index of weathering for metamorphic silicate rocks in tropical regions: A study from Sri Lanka. *Engineering Geology* 36:303–310. doi: [10.1016/0013-7952\(94\)90011-6](https://doi.org/10.1016/0013-7952(94)90011-6)

<sup>5</sup> Parker A (1970). An Index of Weathering for Silicate Rocks. *Geological Magazine* 107:501–504. doi: [10.1017/S0016756800058581](https://doi.org/10.1017/S0016756800058581)

<sup>6</sup> Ishikawa Y, Sawaguchi T, Iwaya S, Horiuchi M (1976). Delineation of prospecting targets for Kuroko deposits based on modes of volcanism of underlying dacite and alteration halos. *Mining Geology* 27:106-117.

**Returns**

Alteration index series.

**Return type**

`pandas.Series`

**References**

`pyrolite.geochem.alteration.CCPI(df: DataFrame)`

Chlorite-carbonate-pyrite index of Large et al. (wt%)<sup>7</sup>.

**Parameters**

**df** (`pandas.DataFrame`) – DataFrame to calculate index from.

**Returns**

Alteration index series.

**Return type**

`pandas.Series`

**Notes**

Here FeO is taken as total iron (FeO+Fe<sub>2</sub>O<sub>3</sub>).

**References****pyrolite.geochem.ions**

`pyrolite.geochem.ions.set_default_ionic_charges(charges=None)`

Set the default ionic charges for each element.

**Parameters**

**charges** (`dict`) – Dictionary of elements : charges.

**pyrolite.geochem.isotope**

Submodule for calculation and transformation of mass spectrometry data (particularly for ion-counting and isotope ratio data). Currently in the early stages of development.

---

**Todo:**

- Dodson ratios<sup>1</sup>
- Add noise-corrected quasi-unbiased ratios of Coath et al. (2012)<sup>2</sup>

---

<sup>7</sup> Large RR, Gemmell, JB, Paulick, H (2001). The alteration box plot: A simple approach to understanding the relationship between alteration mineralogy and lithogeochemistry associated with volcanic-hosted massive sulfide deposits. *Economic Geology* 96:957-971. doi:10.2113/gsecongeo.96.5.957

<sup>1</sup> Dodson M. H. (1978) A linear method for second-degree interpolation in cyclical data collection. *Journal of Physics E: Scientific Instruments* 11, 296. doi: 10.1088/0022-3735/11/4/004

<sup>2</sup> Coath C. D., Steele R. C. J. and Lunn W. F. (2012). Statistical bias in isotope ratios. *J. Anal. At. Spectrom.* 28, 52–58. doi: 10.1039/C2JA10205F

- Deadtime methods<sup>345</sup>
- Synthetic data, file format parsing
- Capability for dealing with `pyrolite.geochem.isotopepeerence` materials

This would be handy for `pyrolite.geochem` in general, and could be implemented in a similar way to how `pyrolite.geochem.norm` handles `pyrolite.geochem.isotopepeerence` compositions.

- Stable isotope calculations
  - Simple radiogenic isotope system calculations and plots
  - U-Pb, U-series data reduction and uncertainty propagation
- 

## References

### 5.1.3 pyrolite.comp

Submodule for working with compositional data.

#### pyrolite.comp.pyrocomp (Pandas Interface)

```
class pyrolite.comp.pyrocomp(obj)
```

```
renormalise(components: list = [], scale=100.0)
```

Renormalises compositional data to ensure closure.

##### Parameters

- **components** (`list`) – Option subcompositon to renormalise to 100. Useful for the use case where compositional data and non-compositional data are stored in the same dataframe.
- **scale** (`float`, 100.) – Closure parameter. Typically either 100 or 1.

##### Returns

Renormalized dataframe.

##### Return type

`pandas.DataFrame`

---

<sup>3</sup> Tyler B. J. (2014). The accuracy and precision of the advanced Poisson dead-time correction and its importance for multivariate analysis of high mass resolution ToF-SIMS data. *Surface and Interface Analysis* 46, 581–590. doi: [10.1002/sia.5543](https://doi.org/10.1002/sia.5543)

<sup>4</sup> Takano A., Takenaka H., Ichimaru S. and Nonaka H. (2012). Comparison of a new dead-time correction method and conventional models for SIMS analysis. *Surface and Interface Analysis* 44, 1287–1293. doi: [10.1002/sia.5004](https://doi.org/10.1002/sia.5004)

<sup>5</sup> Müller J. W. (1991). Generalized dead times. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 301, 543–551. doi: [10.1016/0168-9002\(91\)90021-H](https://doi.org/10.1016/0168-9002(91)90021-H)

## Notes

This won't modify the dataframe in place, you'll need to assign it to something. If you specify components, those components will be summed to 100%, and others remain unchanged.

**ALR**(*components=[]*, *ind=-1*, *null\_col=False*, *label\_mode='simple'*)

Additive Log Ratio transformation.

### Parameters

- **ind** (*int*, *str*) – Index or name of column used as denominator.
- **null\_col** (*bool*) – Whether to keep the redundant column.

### Returns

ALR-transformed array, of shape (N, D-1).

### Return type

`pandas.DataFrame`

**inverse\_ALR**(*ind=None*, *null\_col=False*)

Inverse Additive Log Ratio transformation.

### Parameters

- **ind** (*int*, *str*) – Index or name of column used as denominator.
- **null\_col** (*bool*, *False*) – Whether the array contains an extra redundant column (i.e. shape is (N, D)).

### Returns

Inverse-ALR transformed array, of shape (N, D).

### Return type

`pandas.DataFrame`

**CLR**(*label\_mode='simple'*)

Centred Log Ratio transformation.

### Parameters

**label\_mode** (*str*) – Labelling mode for the output dataframe (*numeric*, *simple*, *LaTeX*). If you plan to use the outputs for automated visualisation and want to know which components contribute, use *simple* or *LaTeX*.

### Returns

CLR-transformed array, of shape (N, D).

### Return type

`pandas.DataFrame`

**inverse\_CLR**()

Inverse Centred Log Ratio transformation.

### Returns

Inverse-CLR transformed array, of shape (N, D).

### Return type

`pandas.DataFrame`



**ILR**(*label\_mode='simple'*)

Isometric Log Ratio transformation.

**Parameters**

**label\_mode** (*str*) – Labelling mode for the output dataframe (*numeric*, *simple*, *LaTeX*). If you plan to use the outputs for automated visualisation and want to know which components contribute, use *simple* or *LaTeX*.

**Returns**

ILR-transformed array, of shape (N, D-1).

**Return type**

`pandas.DataFrame`

**inverse\_ILR**(*X=None*)

Inverse Isometric Log Ratio transformation.

**Parameters**

**X** (*numpy.ndarray*, *None*) – Optional specification for an array from which to derive the orthonormal basis, with shape (N, D).

**Returns**

Inverse-ILR transformed array, of shape (N, D).

**Return type**

`pandas.DataFrame`

**boxcox**(*lmbda=None*, *lmbda\_search\_space=(-1, 5)*, *search\_steps=100*, *return\_lmbda=False*)

Box-Cox transformation.

**Parameters**

- **lmbda** (*numpy.number*, *None*) – Lambda value used to forward-transform values. If none, it will be calculated using the mean
- **lmbda\_search\_space** (*tuple*) – Range tuple (min, max).
- **search\_steps** (*int*) – Steps for lambda search range.

**Returns**

Box-Cox transformed array.

**Return type**

`pandas.DataFrame`

**inverse\_boxcox**(*lmbda=None*)

Inverse Box-Cox transformation.

**Parameters**

**lmbda** (*float*) – Lambda value used to forward-transform values.

**Returns**

Inverse Box-Cox transformed array.

**Return type**

`pandas.DataFrame`

**sphere**()

Spherical coordinate transformation for compositional data.

**Returns**

– Array of angles in radians ( $(0, \pi/2]$ )

**Return type**

`pandas.DataFrame`

**inverse\_sphere**(*variables=None*)

Inverse spherical coordinate transformation to revert back to compositional data in the simplex.

**Parameters**

**variables** (`list`) – List of names for the compositional data variables, optionally specified (for when they may not be stored in the dataframe attributes through the *pyrocomp* functions).

**Returns**

**df** – Dataframe of original compositional (simplex) coordinates, normalised to 1.

**Return type**

`pandas.DataFrame`

**logratiomean**(*transform=<function CLR>*, *inverse\_transform=<function inverse\_CLR>*)

Take a mean of log-ratios along the index of a dataframe.

**Parameters**

**transform** (callable : `str`) – Log transform to use.

**Returns**

Mean values as a pandas series.

**Return type**

`pandas.Series`

**invert\_transform**(*\*\*kwargs*)

Try to inverse-transform a transformed dataframe.

## pyrolite.comp.codata

`pyrolite.comp.codata.close`(*X: ~numpy.ndarray*, *sumf=<function sum>*)

Closure operator for compositional data.

**Parameters**

- **X** (`numpy.ndarray`) – Array to close.
- **sumf** (callable, `numpy.sum()`) – Sum function to use for closure.

**Returns**

Closed array.

**Return type**

`numpy.ndarray`

## Notes

- Does not check for non-positive entries.

```
pyrolite.comp.codata.renormalise(df: DataFrame, components: list = [],
                                scale=100.0)
```

Renormalises compositional data to ensure closure.

### Parameters

- **df** (*pandas.DataFrame*) – Dataframe to renormalise.
- **components** (*list*) – Option subcompositon to renormalise to 100. Useful for the use case where compositional data and non-compositional data are stored in the same dataframe.
- **scale** (*float*, 100.) – Closure parameter. Typically either 100 or 1.

### Returns

Renormalized dataframe.

### Return type

*pandas.DataFrame*

```
pyrolite.comp.codata.ALX(X: ndarray, ind: int = -1, null_col=False)
```

Additive Log Ratio transformation.

### Parameters

- **X** (*numpy.ndarray*) – Array on which to perform the transformation, of shape (N, D).
- **ind** (*int*) – Index of column used as denominator.
- **null\_col** (*bool*) – Whether to keep the redundant column.

### Returns

ALR-transformed array, of shape (N, D-1).

### Return type

*numpy.ndarray*

```
pyrolite.comp.codata.inverse_ALR(Y: ndarray, ind=-1, null_col=False)
```

Inverse Centred Log Ratio transformation.

### Parameters

- **Y** (*numpy.ndarray*) – Array on which to perform the inverse transformation, of shape (N, D-1).
- **ind** (*int*) – Index of column used as denominator.
- **null\_col** (*bool*, False) – Whether the array contains an extra redundant column (i.e. shape is (N, D)).

### Returns

Inverse-ALR transformed array, of shape (N, D).

### Return type

*numpy.ndarray*

`pyrolite.comp.codata.CLR(X: ndarray)`

Centred Log Ratio transformation.

**Parameters**

**X** (`numpy.ndarray`) – 2D array on which to perform the transformation, of shape (N, D).

**Returns**

CLR-transformed array, of shape (N, D).

**Return type**

`numpy.ndarray`

`pyrolite.comp.codata.inverse_CLR(Y: ndarray)`

Inverse Centred Log Ratio transformation.

**Parameters**

**Y** (`numpy.ndarray`) – Array on which to perform the inverse transformation, of shape (N, D).

**Returns**

Inverse-CLR transformed array, of shape (N, D).

**Return type**

`numpy.ndarray`

`pyrolite.comp.codata.ILR(X: ndarray, psi=None, **kwargs)`

Isometric Log Ratio transformation.

**Parameters**

- **X** (`numpy.ndarray`) – Array on which to perform the transformation, of shape (N, D).
- **psi** (`numpy.ndarray`) – Array or matrix representing the ILR basis; optionally specified.

**Returns**

ILR-transformed array, of shape (N, D-1).

**Return type**

`numpy.ndarray`

`pyrolite.comp.codata.inverse_ILR(Y: ndarray, X: Optional[ndarray] = None, psi=None, **kwargs)`

Inverse Isometric Log Ratio transformation.

**Parameters**

- **Y** (`numpy.ndarray`) – Array on which to perform the inverse transformation, of shape (N, D-1).
- **X** (`numpy.ndarray`, None) – Optional specification for an array from which to derive the orthonormal basis, with shape (N, D).
- **psi** (`numpy.ndarray`) – Array or matrix representing the ILR basis; optionally specified.

**Returns**

Inverse-ILR transformed array, of shape (N, D).

**Return type**

`numpy.ndarray`

`pyrolite.comp.codata.logratiomean(df, transform=<function CLR>)`

Take a mean of log-ratios along the index of a dataframe.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe from which to compute a mean along the index.
- **transform** (callable) – Log transform to use.
- **inverse\_transform** (callable) – Inverse of log transform.

**Returns**

Mean values as a pandas series.

**Return type**

`pandas.Series`

`pyrolite.comp.codata.get_ALR_labels(df, mode='simple', ind=-1, **kwargs)`

Get symbolic labels for ALR coordinates based on dataframe columns.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe to generate ALR labels for.
- **mode** (`str`) – Mode of label to return (LaTeX, simple).

**Returns**

List of ALR coordinates corresponding to dataframe columns.

**Return type**

`list`

**Notes**

Some variable names are protected in `sympy` and if used can result in errors. If one of these column names is found, it will be replaced with a title-cased duplicated version of itself (e.g. 'S' will be replaced by 'Ss').

`pyrolite.comp.codata.get_CLR_labels(df, mode='simple', **kwargs)`

Get symbolic labels for CLR coordinates based on dataframe columns.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe to generate CLR labels for.
- **mode** (`str`) – Mode of label to return (LaTeX, simple).

**Returns**

List of CLR coordinates corresponding to dataframe columns.

**Return type**

`list`

## Notes

Some variable names are protected in `sympy` and if used can result in errors. If one of these column names is found, it will be replaced with a title-cased duplicated version of itself (e.g. 'S' will be replaced by 'Ss').

`pyrolite.comp.codata.get_ILR_labels(df, mode='latex', **kwargs)`

Get symbolic labels for ILR coordinates based on dataframe columns.

### Parameters

- **df** (`pandas.DataFrame`) – Dataframe to generate ILR labels for.
- **mode** (`str`) – Mode of label to return (LaTeX, simple).

### Returns

List of ILR coordinates corresponding to dataframe columns.

### Return type

`list`

## Notes

Some variable names are protected in `sympy` and if used can result in errors. If one of these column names is found, it will be replaced with a title-cased duplicated version of itself (e.g. 'S' will be replaced by 'Ss').

`pyrolite.comp.codata.boxcox(X: ndarray, lmbda=None, lmbda_search_space=(-1, 5), search_steps=100, return_lmbda=False)`

Box-Cox transformation.

### Parameters

- **X** (`numpy.ndarray`) – Array on which to perform the transformation.
- **lmbda** (`numpy.number`, `None`) – Lambda value used to forward-transform values. If none, it will be calculated using the mean
- **lmbda\_search\_space** (`tuple`) – Range tuple (min, max).
- **search\_steps** (`int`) – Steps for lambda search range.
- **return\_lmbda** (`bool`) – Whether to also return the lambda value.

### Returns

Box-Cox transformed array. If `return_lmbda` is true, tuple contains data and lambda value.

### Return type

`numpy.ndarray | numpy.ndarray` (:class:`float)`

`pyrolite.comp.codata.inverse_boxcox(Y: ndarray, lmbda)`

Inverse Box-Cox transformation.

### Parameters

- **Y** (`numpy.ndarray`) – Array on which to perform the transformation.
- **lmbda** (`float`) – Lambda value used to forward-transform values.

**Returns**

Inverse Box-Cox transformed array.

**Return type**

`numpy.ndarray`

`pyrolite.comp.codata.sphere(ys)`

Spherical coordinate transformation for compositional data.

**Parameters**

**ys** (`numpy.ndarray`) – Compositional data to transform (shape (n, D)).

**Returns**

– Array of angles in radians ( $(0, \pi/2]$ )

**Return type**

`numpy.ndarray`

**Notes**

`numpy.arccos()` will return angles in the range  $(0, \pi)$ . This shouldn't be an issue for this function given that the input values are all positive.

`pyrolite.comp.codata.inverse_sphere()`

Inverse spherical coordinate transformation to revert back to compositional data in the simplex.

**Parameters**

(`numpy.ndarray`) – Angular coordinates to revert.

**Returns**

**ys** – Compositional (simplex) coordinates, normalised to 1.

**Return type**

`numpy.ndarray`

`pyrolite.comp.codata.compositional_cosine_distances(arr)`

Calculate a distance matrix corresponding to the angles between a number of compositional vectors.

**Parameters**

**arr** (`numpy.ndarray`) – Array of n-dimensional compositions of shape (n\_samples, n).

**Returns**

Array of angular distances of shape (n\_samples, n\_samples).

**Return type**

`numpy.ndarray`

`pyrolite.comp.codata.get_transforms(name)`

Lookup a transform-inverse transform pair by name.

**Parameters**

**name** (`str`) – Name of of the transform pairs (e.g. :code:'CLR').

**Returns**

**tfm, inv\_tfm** – Transform and inverse transform functions.

**Return type**

callable

**pyrolite.comp.aggregate**

`pyrolite.comp.aggregate.get_full_column(X: ndarray)`

Returns the index of the first array column which contains only finite numbers (i.e. no missing data, nan, inf).

**Parameters**

**X** (`numpy.ndarray`) – Array for which to find the first full column within.

**Returns**

Index of the first full column.

**Return type**

`int`

`pyrolite.comp.aggregate.weights_from_array(X: ndarray)`

Returns a set of equal weights with size equal to that of the first axis of an array.

**Parameters**

**X** (`numpy.ndarray`) – Array of compositions to produce weights for.

**Returns**

Array of weights.

**Return type**

`numpy.ndarray`

`pyrolite.comp.aggregate.nan_weighted_mean(X: ndarray, weights=None)`

Returns a weighted mean of compositions, where weights are renormalised to account for missing data.

**Parameters**

- **X** (`numpy.ndarray`) – Array of compositions to take a weighted mean of.
- **weights** (`numpy.ndarray`) – Array of weights.

**Returns**

Array mean.

**Return type**

`numpy.ndarray`

`pyrolite.comp.aggregate.compositional_mean(df, weights=[], **kwargs)`

Implements an aggregation using a compositional weighted mean.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe of compositions to aggregate.
- **weights** (`numpy.ndarray`) – Array of weights.

**Returns**

Mean values along index of dataframe.

**Return type**

`pandas.Series`



```
pyrolite.comp.aggregate.nan_weighted_compositional_mean(X: ndarray,  
                                                         weights=None,  
                                                         ind=None,  
                                                         renorm=True,  
                                                         **kwargs)
```

Implements an aggregation using a weighted mean, but accounts for nans. Requires at least one non-nan column for ALR mean.

When used for internal standardisation, there should be only a single common element - this would be used by default as the divisor here. When used for multiple-standardisation, the [specified] or first common element will be used.

#### Parameters

- **X** (`numpy.ndarray`) – Array of compositions to aggregate.
- **weights** (`numpy.ndarray`) – Array of weights.
- **ind** (`int`) – Index of the column to use as the ALR divisor.
- **renorm** (`bool`, `True`) – Whether to renormalise the output compositional mean to unity.

#### Returns

An array with the mean composition.

#### Return type

`numpy.ndarray`

```
pyrolite.comp.aggregate.cross_ratios(df: DataFrame)
```

Takes ratios of values across a dataframe, such that columns are denominators and the row indexes the numerators, to create a square array. Returns one array per record.

#### Parameters

**df** (`pandas.DataFrame`) – Dataframe of compositions to create ratios of.

#### Returns

A 3D array of ratios.

#### Return type

`numpy.ndarray`

```
pyrolite.comp.aggregate.np_cross_ratios(X: ndarray, debug=False)
```

Takes ratios of values across an array, such that columns are denominators and the row indexes the numerators, to create a square array. Returns one array per record.

#### Parameters

**X** (`numpy.ndarray`) – Array of compositions to create ratios of.

#### Returns

A 3D array of ratios.

#### Return type

`numpy.ndarray`

```
pyrolite.comp.aggregate.standardise_aggregate(df: DataFrame, int_std=None,  
                                              fixed_record_idx=0,  
                                              renorm=True, **kwargs)
```

Performs internal standardisation and aggregates dissimilar geochemical records. Note:

this changes the closure parameter, and is generally intended to integrate major and trace element records.

#### Parameters

- **df** (`pandas.DataFrame`) – Dataframe of compositions to aggregate of.
- **int\_std** (`str`) – Name of the internal standard column.
- **fixed\_record\_idx** (`int`) – Numeric index of a specific record's for which to retain the internal standard value (e.g for standardising trace element data).
- **renorm** (`bool`, `True`) – Whether to renormalise to unity.

#### Returns

A series representing the internally standardised record.

#### Return type

`pandas.Series`

### pyrolite.comp.impute

```
pyrolite.comp.impute.EMCOMP(X, threshold=None, tol=0.0001,  
                             convergence_metric=<function <lambda>>,  
                             max_iter=30)
```

EMCOMP replaces rounded zeros in a compositional data set based on a set of thresholds. After Palarea-Albaladejo and Martín-Fernández (2008)<sup>1</sup>.

#### Parameters

- **X** (`numpy.ndarray`) – Dataset with rounded zeros
- **threshold** (`numpy.ndarray`) – Array of threshold values for each component as a proportion.
- **tol** (`float`) – Tolerance to check for convergence.
- **convergence\_metric** (callable) – Callable function to check for convergence. Here we use a compositional distance rather than a maximum absolute difference, with very similar performance. Function needs to accept two `numpy.ndarray` arguments and third tolerance argument.
- **max\_iter** (`int`) – Maximum number of iterations before an error is thrown.

#### Returns

- **X\_est** (`numpy.ndarray`) – Dataset with rounded zeros replaced.
- **prop\_zeros** (`float`) – Proportion of zeros in the original data set.
- **n\_iters** (`int`) – Number of iterations needed for convergence.

---

<sup>1</sup> Palarea-Albaladejo J. and Martín-Fernández J. A. (2008) A modified EM ALR-algorithm for replacing rounded zeros in compositional data sets. Computers & Geosciences 34, 902–917. doi: 10.1016/j.cageo.2007.09.015

## Notes

- At least one component without missing values is needed for the divisor. Rounded zeros/missing values are replaced by values below their respective detection limits.
- This routine is not completely numerically stable as written.

---

## Todo:

- Implement methods to deal with variable depection limits (i.e thresholds are array shape (N, D))
  - Conisder non-normal models for data distributions.
  - Improve numerical stability to reduce the chance of `np.inf` appearing.
- 

## References

### 5.1.4 pyrolite.mineral

A submodule for working with mineral data.

---

**Note:** This module is a work in progress and will be gradually updated.

---

#### pyrolite.mineral.template

```
class pyrolite.mineral.template.MineralTemplate(name, *components)
```

```
    set_structure(*components)
```

Set the structure of the mineral template.

#### Parameters

**components** – Argument list consisting of each of the structural components. Can consist of any mixture of Sites or argument tuples which can be passed to Site `__init__`.

```
    copy()
```

```
class pyrolite.mineral.template.Mineral(name=None, template=None,
                                          composition=None,
                                          endmembers=None)
```

```
    set_endmembers(endmembers=None)
```

Set the endmbmer components for a mineral.

```
    add_endmember(em, name=None)
```

Add a single endmember to the database.

```
    set_template(template, name=None)
```

Assign a mineral template to the mineral.

**set\_composition**(*composition=None*)

Parse and assign a composition to the mineral.

**Parameters**

**composition** – Composition to assign to the mineral. Can be provided in any form which is digestable by `parse_composition`.

**recalculate\_cations**(*composition=None, ideal\_cations=None, ideal\_oxygens=None, Fe\_species=['FeO', 'Fe', 'Fe2O3'], oxygen\_constrained=False*)

Recalculate a composition to give an elemental ionic breakdown.

**Parameters**

- **composition** – Composition to recalculate. If not provided, will try to use the mineral composition as set.
- **ideal\_cations** (*int*) – Ideal number of cations to use for formulae calculations. Will only be used if oxygen is constrained (i.e. multiple Fe species present or `oxygen_constrained=True`).
- **ideal\_oxygens** (*int*) – Ideal number of oxygens to use for formulae calculations. Will only be used if oxygen is not constrained (i.e. single Fe species present and `oxygen_constrained=False`).
- **Fe\_species** (*list*) – List of iron species for identifying redox-defined compositions.
- **oxygen\_constrained** (*bool, False*) – Whether the oxygen is a closed or open system for the specific composition.

**apfu**()

Get the atoms per formula unit.

**endmember\_decompose**(*det\_lim=0.01*)

Decompose a mineral composition into endmember components.

**Parameters**

**det\_lim** (*float*) – Detection limit for individual

## Notes

Currently implemented using optimization based on mass fractions.

---

**Todo:** Implement site-based endmember decomposition, which will enable more checks and balances.

---

**calculate\_occupancy**(*composition=None, error=1e-05, balances=[['Fe{2+}', 'Mg{2+}']]*)

Calculate the estimated site occupancy for a given composition. Ions will be assigned to sites according to affinities. Sites with equal affinities should receive equal assignment.

**Parameters**

- **composition** – Composition to calculate site occupancy for.

- **error** (*float*) – Absolute error for floating point occupancy calculations.
- **balances** (*list*) – List of iterables containing ions to balance across multiple sites. Note that the partitioning will occur after non-balanced cations are assigned, and that ions are only balanced between sites which have defined affinities for all of the particular ions defined in the ‘balance’.

**get\_site\_occupancy()**

Get the site occupancy for the mineral.

## pyrolite.mineral.normative

`pyrolite.mineral.normative.unmix(comp, parts, order=1, det_lim=0.0001)`

From a composition and endmember components, find a set of weights which best approximate the composition as a weighted sum of components.

### Parameters

- **comp** (`numpy.ndarray`) – Array of compositions (shape  $n_S, n_C$ ).
- **parts** (`numpy.ndarray`) – Array of endmembers (shape  $n_E, n_C$ ).
- **order** (`int`) – Order of regularization, defaults to L1 for sparsity.
- **det\_lim** (`float`) – Detection limit, below which minor components will be omitted for sparsity.

### Returns

Array of endmember modal abundances (shape  $n_S, n_E$ )

### Return type

`numpy.ndarray`

`pyrolite.mineral.normative.endmember_decompose(composition, endmembers=[], drop_zeros=True, molecular=True, order=1, det_lim=0.0001)`

Decompose a given mineral composition to given endmembers.

### Parameters

- **composition** (`DataFrame` | `Series` | `Formula` | `str`) – Composition to decompose into endmember components.
- **endmembers** (`str` | `list` | `dict`) – List of endmembers to use for the decomposition.
- **drop\_zeros** (`bool`, `True`) – Whether to omit components with zero estimated abundance.
- **molecular** (`bool`, `True`) – Whether to *convert* the chemistry to molecular before calculating the decomposition.
- **order** (`int`) – Order of regularization passed to `unmix()`, defaults to L1 for sparsity.
- **det\_lim** (`float`) – Detection limit, below which minor components will be omitted for sparsity.

**Return type**`pandas.DataFrame``pyrolite.mineral.normative.MiddlemostOxRatio(df)`

Apply a TAS classification to a dataframe, and get estimated Fe<sub>2</sub>O<sub>3</sub>/FeO ratios from Middlemost (1989).

**Parameters**

**df** (`pandas.DataFrame`) – Dataframe to get Fe<sub>2</sub>O<sub>3</sub>/FeO ratios for, containing the required oxides to calculate TAS diagrams from (i.e. SiO<sub>2</sub>, Na<sub>2</sub>O, K<sub>2</sub>O).

**Returns**

**ratios** – Series of estimated Fe<sub>2</sub>O<sub>3</sub>/FeO ratios, based on TAS classification.

**Return type**`pandas.Series`**References**

Middlemost, Eric A. K. (1989). Iron Oxidation Ratios, Norms and the Classification of Volcanic Rocks. *Chemical Geology* 77, 1: 19–26. [https://doi.org/10.1016/0009-2541\(89\)90011-9](https://doi.org/10.1016/0009-2541(89)90011-9).

`pyrolite.mineral.normative.LeMaitreOxRatio(df, mode=None)`**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe containing compositions to calibrate against.
- **mode** (`str`) – Mode for the correction - ‘volcanic’ or ‘plutonic’.

**Returns**

Series with oxidation ratios.

**Return type**`pandas.Series`**Notes**

This is a  $\text{FeO}/(\text{FeO} + \text{Fe}_2\text{O}_3)$  mass ratio, not a standard molar ratio  $\text{Fe}^{2+}/(\text{Fe}^{2+} + \text{Fe}^{3+})$  which is more straightforwardly used; data presented should be in mass units. For the calculation, SiO<sub>2</sub>, Na<sub>2</sub>O and K<sub>2</sub>O are expected to be present.

**References**

Le Maitre, R. W (1976). Some Problems of the Projection of Chemical Data into Mineralogical Classifications. *Contributions to Mineralogy and Petrology* 56, no. 2 (1 January 1976): 181–89. <https://doi.org/10.1007/BF00399603>.

`pyrolite.mineral.normative.Middlemost_Fe_correction(df)`**Parameters**

**df** (`pandas.DataFrame`) – Dataframe containing compositions to calibrate against.

**Returns**

Series with two corrected iron components (FeO, Fe<sub>2</sub>O<sub>3</sub>).

**Return type**

`pandas.DataFrame`

**References**

Middlemost, Eric A. K. (1989). Iron Oxidation Ratios, Norms and the Classification of Volcanic Rocks. *Chemical Geology* 77, 1: 19–26. [https://doi.org/10.1016/0009-2541\(89\)90011-9](https://doi.org/10.1016/0009-2541(89)90011-9).

```
pyrolite.mineral.normative.LeMaitre_Fe_correction(df, mode='volcanic')
```

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe containing compositions to correct iron for.
- **mode** (`str`) – Mode for the correction - 'volcanic' or 'plutonic'.

**Returns**

Series with two corrected iron components (FeO, Fe<sub>2</sub>O<sub>3</sub>).

**Return type**

`pandas.DataFrame`

**References**

Le Maitre, R. W (1976). Some Problems of the Projection of Chemical Data into Mineralogical Classifications. *Contributions to Mineralogy and Petrology* 56, no. 2 (1 January 1976): 181–89. <https://doi.org/10.1007/BF00399603>.

Middlemost, Eric A. K. (1989). Iron Oxidation Ratios, Norms and the Classification of Volcanic Rocks. *Chemical Geology* 77, 1: 19–26. [https://doi.org/10.1016/0009-2541\(89\)90011-9](https://doi.org/10.1016/0009-2541(89)90011-9).

```
pyrolite.mineral.normative.CIPW_norm(df, Fe_correction=None,
                                     Fe_correction_mode=None,
                                     adjust_all_Fe=False,
                                     return_adjusted_input=False,
                                     return_free_components=False,
                                     rounding=3)
```

Standardised calculation of estimated mineralogy from bulk rock chemistry. Takes a dataframe of chemistry & creates a dataframe of estimated mineralogy. This is the CIPW norm of Verma et al. (2003). This version only uses major elements.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe containing compositions to transform.
- **Fe\_correction** (`str`) – Iron correction to apply, if any. Will default to 'LeMaitre'.
- **Fe\_correction\_mode** (`str`) – Mode for the iron correction, where applicable.

- **adjust\_all\_Fe** (*bool*) – Where correcting iron compositions, whether to adjust all iron compositions, or only those where singular components are specified.
- **return\_adjusted\_input** (*bool*) – Whether to return the adjusted input chemistry with the output.
- **return\_free\_components** (*bool*) – Whether to return the free components in the output.
- **rounding** (*int*) – Rounding to be applied to input and output data.

**Return type**

`pandas.DataFrame`

## References

Verma, Surendra P., Ignacio S. Torres-Alvarado, and Fernando Velasco-Tapia (2003). A Revised CIPW Norm. Swiss Bulletin of Mineralogy and Petrology 83, 2: 197–216.  
Verma, S. P., & Rivera-Gomez, M. A. (2013). Computer Programs for the Classification and Nomenclature of Igneous Rocks. Episodes, 36(2), 115–124.

---

**Todo:**

- Note whether data needs to be normalised to 1 or 100?
- 

## Notes

The function expect oxide components to be in wt% and elemental data to be in ppm.

## pyrolite.mineral.transform

`pyrolite.mineral.transform.formula_to_elemental(formula, weight=True)`

Convert a periodictable.formulas.Formula to elemental composition.

`pyrolite.mineral.transform.merge_formulae(formulas)`

Combine multiple formulae into one. Particularly useful for defining oxide mineral formulae.

**Parameters**

**formulas** (*iterable*) – Iterable of multiple formulae to merge into a single larger molecular formulae.

`pyrolite.mineral.transform.recalc_cations(df, ideal_cations=4,  
ideal_oxygens=6,  
Fe_species=['FeO', 'Fe', 'Fe2O3'],  
oxygen_constrained=False)`

Recalculate a composition to a.p.f.u.



## pyrolite.mineral.sites

```
class pyrolite.mineral.sites.Site(name=None, coordination=0, affinities={},
                                  mode='cation')

class pyrolite.mineral.sites.MX(name='M', coordination=8, *args, **kwargs)
    Octahedrally coordinated M site.

class pyrolite.mineral.sites.TX(name='T', coordination=4, affinities={'Al{3+}':
                              1, 'Fe{3+}': 2, 'Si{4+}': 0}, *args, mode='cation',
                              **kwargs)
    Tetrahedrally coordinated T site.

class pyrolite.mineral.sites.IX(name='I', coordination=12, *args, **kwargs)
    Dodecahedrally coordinated I site.

class pyrolite.mineral.sites.VX(name='V', coordination=0, *args, **kwargs)
    Vacancy site.

class pyrolite.mineral.sites.OX(name='O', coordination=0, affinities={'O{2-}':
                              0}, *args, **kwargs)
    Oxygen site.

class pyrolite.mineral.sites.AX(name='A', coordination=0, *args, **kwargs)
    Anion site.
```

## pyrolite.mineral.lattice

Submodule for calculating relative ion partitioning based on the lattice strain model<sup>123</sup>.

### Todo:

- Bulk modulus and Youngs modulus approximations<sup>245</sup>.

## References

```
pyrolite.mineral.lattice.strain_coefficient(ri, rx, r0=None, E=None,
                                             T=298.15, z=None, **kwargs)
```

Calculate the lattice strain associated with an ionic substitution<sup>67</sup>.

### Parameters

<sup>1</sup> Brice, J.C., 1975. Some thermodynamic aspects of the growth of strained crystals. *Journal of Crystal Growth* 28, 249–253. doi: [10.1016/0022-0248\(75\)90241-9](https://doi.org/10.1016/0022-0248(75)90241-9)

<sup>2</sup> Blundy, J., Wood, B., 1994. Prediction of crystal–melt partition coefficients from elastic moduli. *Nature* 372, 452. doi: [10.1038/372452a0](https://doi.org/10.1038/372452a0)

<sup>3</sup> Wood, B.J., Blundy, J.D., 2014. Trace Element Partitioning: The Influences of Ionic Radius, Cation Charge, Pressure, and Temperature. *Treatise on Geochemistry* (Second Edition) 3, 421–448. doi: [10.1016/B978-0-08-095975-7.00209-6](https://doi.org/10.1016/B978-0-08-095975-7.00209-6)

<sup>4</sup> Anderson, D.L., Anderson, O.L., 1970. Brief report: The bulk modulus-volume relationship for oxides. *Journal of Geophysical Research* (1896-1977) 75, 3494–3500. doi: [10.1029/JB075i017p03494](https://doi.org/10.1029/JB075i017p03494)

<sup>5</sup> Hazen, R.M., Finger, L.W., 1979. Bulk modulus—volume relationship for cation-anion polyhedra. *Journal of Geophysical Research: Solid Earth* 84, 6723–6728. doi: [10.1029/JB084iB12p06723](https://doi.org/10.1029/JB084iB12p06723)

<sup>6</sup> Brice, J.C., 1975. Some thermodynamic aspects of the growth of strained crystals. *Journal of Crystal Growth* 28, 249–253. doi: [10.1016/0022-0248\(75\)90241-9](https://doi.org/10.1016/0022-0248(75)90241-9)

<sup>7</sup> Blundy, J., Wood, B., 1994. Prediction of crystal–melt partition coefficients from elastic moduli. *Nature* 372, 452. doi: [10.1038/372452a0](https://doi.org/10.1038/372452a0)

- **ri** (`float`) – Ionic radius to calculate strain relative to, in angstroms (Å).
- **rj** (`float`) – Ionic radius to calculate strain for, in angstroms (Å).
- **r0** (`float`, `None`) – Fictive ideal ionic radii for the site. The value for **ri** will be used in its place if none is given, and a warning issued.
- **E** (`float`, `None`) – Young’s modulus (stiffness) for the site, in pascals (Pa). Will be estimated using `youngs_modulus_approximation()` if none is given.
- **T** (`float`) – Temperature, in Kelvin (K).
- **z** (`int`) – Optional specification of cationic valence, for calculation of approximate Young’s modulus using `youngs_modulus_approximation()`, where the modulus is not specified.

### Returns

The strain coefficient  $e^{\frac{-\Delta G_{strain}}{RT}}$ .

### Return type

`float`

### Notes

The lattice strain model relates changes in partitioning to differences in ionic radii for ions of a given cationic charge, and for a for a specific site (with Young’s modulus  $E$ ). This is calculated using the work done to expand a spherical shell centred on the lattice site, which alters the  $\Delta G$  for the formation of the mineral. This can be related to changes in partition coefficients using the following<sup>?</sup>:

$$D_{j^{n+}} = D_{A^{n+}} \cdot e^{\frac{-4\pi EN \left( \frac{r_0}{2} (r_j - r_0)^2 + \frac{1}{3} (r_j - r_0)^3 \right)}{RT}}$$

Where  $D_{A^{n+}}$  is the partition coefficient for the ideal ion A, and N is Avagadro’s number (6.023e23 atoms/mol). This can also be calculated relative to an ‘ideal’ fictive ion which has a maximum  $D$  where this data are available. This relationship arises via i) the integration to calculate the strain energy mentioned above ( $4\pi E(\frac{r_0}{2}(r_j - r_0)^2 + \frac{1}{3}(r_j - r_0)^3)$ ), and ii) the assumption that the changes in  $\Delta G$  occur only to size differences, and the difference is additive. The ‘segregation coefficient’  $K_j$  can be expressed relative to the non-doped equilibrium constant  $K_0$ <sup>?</sup>:

$$K_j = e^{\frac{-\Delta G_0 - \Delta G_{strain}}{RT}} \quad (5.8)$$

$$= e^{\frac{-\Delta G_0}{RT}} \cdot e^{\frac{-\Delta G_{strain}}{RT}} \quad (5.9)$$

$$= K_0 \cdot e^{\frac{-\Delta G_{strain}}{RT}} \quad (5.10)$$

$$(5.11)$$

The model assumes that the crystal is elastically isotropic.

## References

`pyrolite.mineral.lattice.youngs_modulus_approximation(z, r)`

Young's modulus approximation for cationic sites in silicates and oxides<sup>8910</sup>.

### Parameters

- **z** (`integer`) – Cationic valence.
- **r** (`float`) – Ionic radius of the cation (Å).

### Returns

**E** – Young's modulus for the cationic site, in Pascals (Pa)

### Return type

`float`

## Notes

The bulk modulus  $K$  for an an ionic crystal is esimated using<sup>?</sup>:

$$K = \frac{AZ_aZ_ce^2(n-1)}{9d_0V_0}$$

Where  $A$  is the Madelung constant,  $Z_c$  and  $Z_a$  are the anion and cation valences,  $e$  is the charge on the electron,  $n$  is the Born power law coefficient, and  $d_0$  is the cation-anion distance<sup>?</sup>. Using the Shannon ionic radius for oxygen (1.38 Å), this is approximated for cations coordinated by oxygen in silicates and oxides using the following relationship<sup>?</sup>:

$$K = 750Z_cd^{-3}$$

Where  $d$  is the cation-anion distance (Å),  $Z_c$  is the cationic valence (and  $K$  is in GPa). The Young's modulus  $E$  is then calculated through the relationship<sup>?</sup>:

$$E = 3K(1 - 2\sigma)$$

Where  $\sigma$  is Poisson's ratio, which in the case of minerals can be approimxated by 0.25<sup>?</sup>, and hence:

$$E \approx 1.5K \tag{5.12}$$

$$E \approx 1025Z_c^2d^3$$

---

### Todo:

- Add links to docstring
- 

---

<sup>8</sup> Anderson, D.L., Anderson, O.L., 1970. Brief report: The bulk modulus-volume relationship for oxides. *Journal of Geophysical Research* (1896-1977) 75, 3494–3500. doi: [10.1029/JB075i017p03494](https://doi.org/10.1029/JB075i017p03494)

<sup>9</sup> Hazen, R.M., Finger, L.W., 1979. Bulk modulus—volume relationship for cation-anion polyhedra. *Journal of Geophysical Research: Solid Earth* 84, 6723–6728. doi: [10.1029/JB084iB12p06723](https://doi.org/10.1029/JB084iB12p06723)

<sup>10</sup> Wood, B.J., Blundy, J.D., 2014. Trace Element Partitioning: The Influences of Ionic Radius, Cation Charge, Pressure, and Temperature. *Treatise on Geochemistry* (Second Edition) 3, 421–448. doi: [10.1016/B978-0-08-095975-7.00209-6](https://doi.org/10.1016/B978-0-08-095975-7.00209-6)

## References

### pyrolite.mineral.mindb

Submodule for accessing the rock forming mineral database.

#### Notes

Accessing and modifying the database across multiple with multiple threads/processes *could* result in database corruption (e.g. through repeated truncation etc).

`pyrolite.mineral.mindb.list_groups()`

List the mineral groups present in the mineral database.

#### Return type

`list`

`pyrolite.mineral.mindb.list_minerals()`

List the minerals present in the mineral database.

#### Return type

`list`

`pyrolite.mineral.mindb.list_formulae()`

List the mineral formulae present in the mineral database.

#### Return type

`list`

`pyrolite.mineral.mindb.get_mineral(name="", dbpath=None)`

Get a specific mineral from the database.

#### Parameters

- **name** (`str`) – Name of the desired mineral.
- **dbpath** (`pathlib.Path`, `str`) – Optional overriding of the default database path.

#### Return type

`pd.Series`

`pyrolite.mineral.mindb.parse_composition(composition, drop_zeros=True)`

Parse a composition reference to provide an ionic elemental version in the form of a `Series`. Currently accepts `pandas.Series`, `periodictable.formulas.Formula` and structures which will directly convert to `pandas.Series` (list of tuples, dict).

#### Parameters

- **composition** (`str` | `periodictable.formulas.Formula` | `pandas.Series`) – Name of a mineral, a formula or composition as a series
- **drop\_zeros** (`bool`) – Whether to drop compositional zeros.

#### Returns

**mineral** – Composition formatted as a series.

#### Return type

`pandas.Series`

```
pyrolite.mineral.mindb.get_mineral_group(group="")
```

Extract a mineral group from the database.

**Parameters**

**group** (`str`) – Group to extract from the mineral database.

**Returns**

Dataframe of group members and compositions.

**Return type**

`pandas.DataFrame`

```
pyrolite.mineral.mindb.update_database(path=None, **kwargs)
```

Update the mineral composition database.

**Parameters**

**path** (`str` | `pathlib.Path`) – The desired filepath for the JSON database.

**Notes**

This will take the ‘mins.csv’ file from the mineral pyrolite data folder and construct a document-based JSON database.

## 5.1.5 pyrolite.util

Various utilities used by other submodules.

### pyrolite.util.general

```
class pyrolite.util.general.Timewith(name="")
```

**property elapsed**

**checkpoint** (`name=""`)

```
pyrolite.util.general.temp_path(suffix="")
```

Return the path of a temporary directory.

```
pyrolite.util.general.flatten_dict(d, climb=False, safemode=False)
```

Flattens a nested dictionary containing only string keys.

This will work for dictionaries which don’t have two equivalent keys at the same level. If you’re worried about this, use `safemode=True`.

Partially taken from <https://stackoverflow.com/a/6043835>.

**Parameters**

- **climb** (`bool`, `False`) – Whether to keep trunk or leaf-values, for items with the same key.
- **safemode** (`bool`, `True`) – Whether to keep all keys as a tuple index, to avoid issues with conflicts.

**Returns**

Flattened dictionary.

**Return type**`dict``pyrolite.util.general.swap_item(startlist: list, pull: object, push: object)`

Swap a specified item in a list for another.

**Parameters**

- **startlist** (`list`) – List to replace item within.
- **pull** – Item to replace in the list.
- **push** – Item to add into the list.

**Return type**`list``pyrolite.util.general.copy_file(src, dst, ext=None, permissions=None)`

Copy a file from one place to another. Uses the full filepath including name.

**Parameters**

- **src** (`str` | `pathlib.Path`) – Source filepath.
- **dst** (`str` | `pathlib.Path`) – Destination filepath or directory.
- **ext** (`str`, `None`) – Optional file extension specification.

`pyrolite.util.general.remove_tempdir(directory)`

Remove a specific directory, contained files and sub-directories.

**Parameters**

**directory** (`str`, `Path`) – Path to directory.

## pyrolite.util.pd

`pyrolite.util.pd.drop_where_all_empty(df)`

Drop rows and columns which are completely empty.

**Parameters**

**df** (`pandas.DataFrame` | `pandas.Series`) – Pandas object to ensure is in the form of a series.

`pyrolite.util.pd.read_table(filepath, index_col=0, **kwargs)`

Read tabluar data from an excel or csv text-based file.

**Parameters**

**filepath** (`str` | `pathlib.Path`) – Path to file.

**Return type**`pandas.DataFrame``pyrolite.util.pd.column_ordered_append(df1, df2, **kwargs)`

Appends one dataframe to another, preserving the column order of the first and adding new columns on the right. Also accepts and passes on standard keyword arguments for `pd.DataFrame.append`.

**Parameters**

- **df1** (`pandas.DataFrame`) – The dataframe for which columns order is preserved in the output.

- **df2** (`pandas.DataFrame`) – The dataframe for which new columns are appended to the output.

**Return type**`pandas.DataFrame`

```
pyrolite.util.pd.accumulate(dfs, ignore_index=False, trace_source=False,
                             names=[])
```

Accumulate an iterable containing multiple `pandas.DataFrame` to a single frame.

**Parameters**

- **dfs** (`list`) – Sequence of dataframes.
- **ignore\_index** (`bool`) – Whether to ignore the indexes upon joining.
- **trace\_source** (`bool`) – Whether to retain a reference to the source of the data rows.
- **names** (`list`) – Names to use in place of indexes for source names.

**Returns**

Accumulated dataframe.

**Return type**`pandas.DataFrame`

```
pyrolite.util.pd.to_frame(ser)
```

Simple utility for converting to `pandas.DataFrame`.

**Parameters**

**ser** (`pandas.Series` | `pandas.DataFrame`) – Pandas object to ensure is in the form of a dataframe.

**Return type**`pandas.DataFrame`

```
pyrolite.util.pd.to_ser(df)
```

Simple utility for converting single column `pandas.DataFrame` to `pandas.Series`.

**Parameters**

**df** (`pandas.DataFrame` | `pandas.Series`) – Pandas object to ensure is in the form of a series.

**Return type**`pandas.Series`

```
pyrolite.util.pd.to_numeric(df, errors: str = 'coerce', exclude=['float', 'int'])
```

Converts non-numeric columns to numeric type where possible.

**Notes**

Avoid using `.loc` or `.iloc` on the LHS to make sure that data dtypes are propagated.

```
pyrolite.util.pd.zero_to_nan(df, rtol=1e-05, atol=1e-08)
```

Replace floats close, less or equal to zero with `np.nan` in a dataframe.

**Parameters**

- **df** (`pandas.DataFrame`) – DataFrame to censor.
- **rtol** (`float`) – The relative tolerance parameter.

- **atol** (`float`) – The absolute tolerance parameter.

**Returns**

Censored DataFrame.

**Return type**

`pandas.DataFrame`

```
pyrolite.util.pd.outliers(df, cols=[], detect=<function <lambda>>,
                           quantile_select=(0.02, 0.98), logquantile=False,
                           exclude=False)
```

```
pyrolite.util.pd.concat_columns(df, columns=None, astype=<class 'str'>,
                                **kwargs)
```

Concatenate strings across columns.

**Parameters**

- **df** (`pandas.DataFrame`) – Dataframe to concatenate.
- **columns** (`list`) – List of columns to concatenate.
- **astype** (`type`) – Type to convert final concatenation to.

**Return type**

`pandas.Series`

```
pyrolite.util.pd.uniques_from_concat(df, columns=None, hashit=True)
```

Creates ideally unique keys from multiple columns. Optionally hashes string to standardise length of identifier.

**Parameters**

- **df** (`pandas.DataFrame`) – DataFrame to create indexes for.
- **columns** (`list`) – Columns to use in the string concatenation.
- **hashit** (`bool`, `True`) – Whether to use a hashing algorithm to create the key from a typically longer string.

**Return type**

`pandas.Series`

```
pyrolite.util.pd.df_from_csvs(csvs, dropna=True, ignore_index=False, **kwargs)
```

Takes a list of .csv filenames and converts to a single DataFrame. Combines columns across dataframes, preserving order of the first entered.

E.g. SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, MgO, MnO, CaO SiO<sub>2</sub>, MgO, FeO, CaO SiO<sub>2</sub>, Na<sub>2</sub>O, Al<sub>2</sub>O<sub>3</sub>, FeO, CaO => SiO<sub>2</sub>, Na<sub>2</sub>O, Al<sub>2</sub>O<sub>3</sub>, MgO, FeO, MnO, CaO - Existing neighbours take priority (i.e. FeO won't be inserted bf Al<sub>2</sub>O<sub>3</sub>) - Earlier inputs take priority (where ordering is ambiguous, place the earlier first)

---

**Todo:** Attempt to preserve column ordering across column sets, assuming they are generally in the same order but preserving only some of the information.

---



## pyrolite.util.plot

Utility functions for working with matplotlib.

### Parameters

- **DEFAULT\_CONT\_COLORMAP** (matplotlib.colors.ScalarMappable) – Default continuous colormap.
- **DEFAULT\_DICS\_COLORMAP** (matplotlib.colors.ScalarMappable) – Default discrete colormap.
- **USE\_PCOLOR** (bool) – Option to use the `matplotlib.pyplot.pcolor()` function in place of `matplotlib.pyplot.pcolormesh()`.

## pyrolite.util.plot.axes

Functions for creating, ordering and modifying Axes.

`pyrolite.util.plot.axes.get_ordered_axes(fig)`

Get the axes from a figure, which may or may not have been modified by pyrolite functions. This ensures that ordering is preserved.

`pyrolite.util.plot.axes.get_axes_index(ax)`

Get the three-digit integer index of a subplot in a regular grid.

### Parameters

**ax** (matplotlib.axes.Axes) – Axis to get the gridspec index for.

### Returns

Rows, columns and axis index for the gridspec.

### Return type

tuple

`pyrolite.util.plot.axes.replace_with_ternary_axis(ax)`

Replace a specified axis with a ternary equivalent.

### Parameters

**ax** (Axes)

### Returns

tax

### Return type

TernaryAxes

`pyrolite.util.plot.axes.label_axes(ax, labels=[], **kwargs)`

Convenience function for labelling rectilinear and ternary axes.

### Parameters

- **ax** (Axes) – Axes to label.
- **labels** (list) – List of labels: [x, y] | or [t, l, r]

`pyrolite.util.plot.axes.axes_to_ternary(ax)`

Set axes to ternary projection after axis creation. As currently implemented, note that this will replace and reorder axes as accessed from the figure (the ternary axis will then be at the end), and as such this returns a list of axes in the correct order.

**Parameters**

**ax** (*Axes* | *list* (*Axes*)) – Axis (or axes) to convert projection for.

**Returns**

*axes*

**Return type**

*list*' (:class:~matplotlib.axes.Axes,  
class:~mpltern.ternary.TernaryAxes)

`pyrolite.util.plot.axes.check_default_axes(ax)`

Simple test to check whether an axis is empty of artists and hasn't been rescaled from the default extent.

**Parameters**

**ax** (*matplotlib.axes.Axes*) – Axes to check for artists and scaling.

**Return type**

*bool*

`pyrolite.util.plot.axes.check_empty(ax)`

Simple test to check whether an axis is empty of artists.

**Parameters**

**ax** (*matplotlib.axes.Axes*) – Axes to check for artists.

**Return type**

*bool*

`pyrolite.util.plot.axes.init_axes(ax=None, projection=None, minsize=1.0,  
**kwargs)`

Get or create an Axes from an optionally-specified starting Axes.

**Parameters**

- **ax** (*Axes*) – Specified starting axes, optional.
- **projection** (*str*) – Whether to create a projected (e.g. ternary) axes.
- **minsize** (*float*) – Minimum figure dimension (inches).

**Returns**

*ax*

**Return type**

*Axes*

`pyrolite.util.plot.axes.share_axes(axes, which='xy')`

Link the x, y or both axes across a group of *Axes*.

**Parameters**

- **axes** (*list*) – List of axes to link.
- **which** (*str*) – Which axes to link. If *x*, link the x-axes; if *y* link the y-axes, otherwise link both.

`pyrolite.util.plot.axes.get_twins(ax, which='y')`

Get twin axes of a specified axis.

**Parameters**

- **ax** (*matplotlib.axes.Axes*) – Axes to get twins for.

- **which** (`str`) – Which twins to get (shared 'x', shared 'y' or the concatenation of both, 'xy').

**Return type**`list`**Notes**

This function was designed to assist in avoiding creating a series of duplicate axes when replotting on an existing axis using a function which would typically create a twin axis.

`pyrolite.util.plot.axes.subaxes(ax, side='bottom', width=0.2, moveticks=True)`

Append a sub-axes to one side of an axes.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axes to append a sub-axes to.
- **side** (`str`) – Which side to append the axes on.
- **width** (`float`) – Fraction of width to give to the subaxes.
- **moveticks** (`bool`) – Whether to move ticks to the outer axes.

**Returns**

Subaxes instance.

**Return type**`matplotlib.axes.Axes`

`pyrolite.util.plot.axes.add_colorbar(mappable, **kwargs)`

Adds a colorbar to a given mappable object.

Source: <http://joseph-long.com/writing/colorbars/>

**Parameters**

**mappable** – The Image, ContourSet, etc. to which the colorbar applies.

**Return type**`matplotlib.colorbar.Colorbar`

---

**Todo:**

- Where no mappable specified, get most recent axes, and check for collections etc
- 

**pyrolite.util.plot.density**

Functions for dealing with kernel density estimation.

**ivar USE\_PCOLOR**

Option to use the `matplotlib.pyplot.pcolor()` function in place of `matplotlib.pyplot.pcolormesh()`.

**vartype USE\_PCOLOR**`bool`

`pyrolite.util.plot.density.get_axis_density_methods(ax)`

Get the relevant density and contouring methods for a given axis.

**Parameters**

**ax** (`matplotlib.axes.Axes` | `mpltern.ternary.TernaryAxes`) – Axis to check.

**Returns**

Relevant functions for this axis.

**Return type**

`pcolor`, `contour`, `contourf`

`pyrolite.util.plot.density.percentile_contour_values_from_meshz(z, percentiles=[0.95, 0.66, 0.33], resolution=1000)`

Integrate a probability density distribution  $Z(X,Y)$  to obtain contours in  $Z$  which correspond to specified percentile contours. Contour values will be returned with the same order as the inputs.

**Parameters**

- **z** (`numpy.ndarray`) – Probability density function over  $x, y$ .
- **percentiles** (`numpy.ndarray`) – Percentile values for which to create contours.
- **resolution** (`int`) – Number of bins for thresholds between 0. and  $\max(Z)$

**Returns**

- **labels** (`list`) – Labels for contours (percentiles, if above minimum  $z$  value).
- **contours** (`list`) – Contour height values.

---

**Todo:** This may error for a list of percentiles where one or more requested values are below the minimum threshold. The exception handling should be updated to cater for arrays - where some of the values may be above the minimum.

---

`pyrolite.util.plot.density.plot_Z_percentiles(*coords, zi=None, percentiles=[0.95, 0.66, 0.33], ax=None, extent=None, fontsize=8, cmap=None, colors=None, linewidths=None, linestyle=None, contour_labels=None, label_contours=True, **kwargs)`

Plot percentile contours onto a 2D (scaled or unscaled) probability density distribution  $Z$  over  $X,Y$ .

**Parameters**

- **coords** (`numpy.ndarray`) – Arrays of  $(x, y)$  or  $(a, b, c)$  coordinates.
- **z** (`numpy.ndarray`) – Probability density function over  $x, y$ .
- **percentiles** (`list`) – Percentile values for which to create contours.
- **ax** (`matplotlib.axes.Axes`, `None`) – Axes on which to plot. If none given, will create a new Axes instance.
- **extent** (`list`, `None`) – List or `np.ndarray` in the form  $[-x, +x, -y, +y]$  over which the image extends.
- **fontsize** (`float`) – Fontsize for the contour labels.

- **cmap** (`matplotlib.colors.ListedColormap`) – Color map for the contours and contour labels.
- **colors** (`str` | `list`) – Colors for the contours, can optionally be specified *in place of cmap*.
- **linewidths** (`str` | `list`) – Widths of contour lines.
- **linestyles** (`str` | `list`) – Styles for contour lines.
- **contour\_labels** (`dict` | `list`) – Labels to assign to contours, organised by level.
- **label\_contours** :class:`bool` – Whether to add text labels to individual contours.

**Returns**

Plotted and formatted contour set.

**Return type**

`matplotlib.contour.QuadContourSet`

**Notes**

When the contours are percentile based, high percentile contours tend to get washed out with colormapping - consider adding different controls on coloring, especially where there are only one or two contours specified. One way to do this would be via the string based keyword argument *colors* for `plt.contour`, with an adaption for non-string colours which post-hoc modifies the contour lines based on the specified colours?

```
pyrolite.util.plot.density.conditional_prob_density(y, x=None, logy=False,
                                                    resolution=5, bins=50,
                                                    yextent=None, rescale=True,
                                                    mode='binkde', ret_centres=False,
                                                    **kwargs)
```

Estimate the conditional probability density of one dependent variable.

**Parameters**

- **y** (`numpy.ndarray`) – Dependent variable for which to calculate conditional probability  $P(y | X=x)$
- **x** (`numpy.ndarray`, `None`) – Optionally-specified independent index.
- **logy** (`bool`) – Whether to use a logarithmic bin spacing on the y axis.
- **resolution** (`int`) – Points added per segment via interpolation along the x axis.
- **bins** (`int`) – Bins for histograms and grids along the independent axis.
- **yextent** (`tuple`) – Extent in the y direction.
- **rescale** (`bool`) – Whether to rescale bins to give the same max Z across x.
- **mode** (`str`) –  
Mode of computation.

If mode is "ckde", use `statsmodels.nonparametric.KDEMultivariateConditional()` to compute a conditional kernel density estimate. If mode is "kde", use a normal gaussian

kernel density estimate. If mode is "binkde", use a gaussian kernel density estimate over y for each bin. If mode is "hist", compute a histogram.

- **ret\_centres** (`bool`) – Whether to return bin centres in addition to histogram edges, e.g. for later contouring.

#### Returns

x bin edges `xe`, y bin edges `ye`, histogram/density estimates `Z`. If `ret_centres` is `True`, the last two return values will contain the bin centres `xi`, `yi`.

#### Return type

tuple of `numpy.ndarray`

### pyrolite.util.plot.export

Functions for export of figures and figure elements from matplotlib.

`pyrolite.util.plot.export.save_figure`(`figure`, `name='fig'`, `save_at=""`, `save_fmts=['png']`, `**kwargs`)

Save a figure at a specified location in a number of formats.

`pyrolite.util.plot.export.save_axes`(`ax`, `name='fig'`, `save_at=""`, `save_fmts=['png']`, `pad=0.0`, `**kwargs`)

Save either a single or multiple axes (from a single figure) based on their extent. Uses the `save_figure` procedure to save at a specific location using a number of formats.

---

#### Todo:

- Add legend to items
- 

`pyrolite.util.plot.export.get_full_extent`(`ax`, `pad=0.0`)

Get the full extent of an axes, including axes labels, tick labels, and titles. Text objects are first drawn to define the extents.

#### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axes of which to check items to get full extent.
- **pad** (`float` | `tuple`) – Amount of padding to add to the full extent prior to returning. If a tuple is passed, the padding will be as above, but for x and y directions, respectively.

#### Returns

Bbox of the axes with optional additional padding.

#### Return type

`matplotlib.transforms.Bbox`

`pyrolite.util.plot.export.path_to_csv`(`path`, `xname='x'`, `yname='y'`, `delim=','`, `linesep='\n'`)

Extract the vertices from a path and write them to csv.

#### Parameters

- **path** (`matplotlib.path.Path` | `tuple`) – Path or x-y tuple to use for coordinates.

- **xname** (`str`) – Name of the x variable.
- **yname** (`str`) – Name of the y variable.
- **delim** (`str`) – Delimiter for the csv file.
- **linesep** (`str`) – Line separator character.

**Returns**

String-representation of csv file, ready to be written to disk.

**Return type**

`str`

## pyrolite.util.plot.grid

Gridding and binning functions.

`pyrolite.util.plot.grid.bin_centres_to_edges(centres, sort=True)`

Translates point estimates at the centres of bins to equivalent edges, for the case of evenly spaced bins.

---

**Todo:**

- This can be updated to unevenly spaced bins, just need to calculate outer bins.
- 

`pyrolite.util.plot.grid.bin_edges_to_centres(edges)`

Translates edges of histogram bins to bin centres.

`pyrolite.util.plot.grid.ternary_grid(data=None, nbins=10, margin=0.001, force_margin=False, yscale=1.0, tfm=<function <lambda>>)`

Construct a graphical linearly-spaced grid within a ternary space.

**Parameters**

- **data** (`numpy.ndarray`) – Data to construct the grid around ((samples, 3)).
- **nbins** (`int`) – Number of bins for grid.
- **margin** (`float`) – Proportional value for the position of the outer boundary of the grid.
- **force\_margin** (`bool`) – Whether to enforce the grid margin.
- **yscale** (`float`) – Y scale for the specific ternary diagram.
- **tfm** – Log transform to use for the grid creation.

**Returns**

- **bins** (`numpy.ndarray`) – Bin centres along each of the ternary axes ((samples, 3))
- **binedges** (`numpy.ndarray`) – Position of bin edges.
- **centregrid** (`list` of `numpy.ndarray`) – Meshgrid of bin centres.
- **edgegrid** (`list` of `numpy.ndarray`) – Meshgrid of bin edges.

## pyrolite.util.plot.helpers

matplotlib helper functions for common drawing tasks.

`pyrolite.util.plot.helpers.alphalabel_subplots(ax, fmt='{ }', xy=(0.03, 0.95), ha='left', va='top', **kwargs)`

Add alphabetical labels to a successive series of subplots with a specified format.

### Parameters

- **ax** (`list` | `numpy.ndarray` | `numpy.flatiter`) – Axes to label, in desired order.
- **fmt** (`str`) – Format string to use. To add e.g. parentheses, you could specify "{ }".
- **xy** (`tuple`) – Position of the labels in axes coordinates.
- **ha** (`str`) – Horizontal alignment of the labels ({ "left", "right" }).
- **va** (`str`) – Vertical alignment of the labels ({ "top", "bottom" }).

`pyrolite.util.plot.helpers.get_centroid(poly)`

Centroid of a closed polygon using the Shoelace formula.

### Parameters

**poly** (`matplotlib.patches.Polygon`) – Polygon to obtain the centroid of.

### Returns

**cx, cy** – Centroid coordinates.

### Return type

`tuple`

`pyrolite.util.plot.helpers.get_visual_center(poly, vertical_exaggeration=1)`

Visual center of a closed polygon.

### Parameters

- **poly** (`matplotlib.patches.Polygon`) – Polygon for which to obtain the visual center.
- **vertical\_exaggeration** (`float`) – Apparent vertical exaggeration of the plot (pixels per unit in y direction divided by pixels per unit in the x direction).

### Returns

**cx, cy** – Centroid coordinates.

### Return type

`tuple`

`pyrolite.util.plot.helpers.rect_from_centre(x, y, dx=0, dy=0, **kwargs)`

Takes an xy point, and creates a rectangular patch centred about it.

`pyrolite.util.plot.helpers.draw_vector(v0, v1, ax=None, **kwargs)`

Plots an arrow representing the direction and magnitude of a principal component on a biaxial plot.

Modified after Jake VanderPlas' Python Data Science Handbook <https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html>

---

**Todo:** Update for ternary plots.

---



```
pyrolite.util.plot.helpers.vector_to_line(mu: array, vector: array, variance: float, spans:
                                             int = 4, expand: int = 10)
```

Creates an array of points representing a line along a vector - typically for principal component analysis. Modified after Jake VanderPlas' Python Data Science Handbook <https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html>

```
pyrolite.util.plot.helpers.plot_stdev_ellipses(comp, nstds=4, scale=100,
                                                resolution=1000, transform=None,
                                                ax=None, **kwargs)
```

Plot covariance ellipses at a number of standard deviations from the mean.

#### Parameters

- **comp** (`numpy.ndarray`) – Composition to use.
- **nstds** (`int`) – Number of standard deviations from the mean for which to plot the ellipses.
- **scale** (`float`) – Scale applying to all x-y data points. For intergration with python-ternary.
- **transform** (`callable`) – Function for transformation of data prior to plotting (to either 2D or 3D).
- **ax** (`matplotlib.axes.Axes`) – Axes to plot on.

#### Returns

`ax`

#### Return type

`matplotlib.axes.Axes`

```
pyrolite.util.plot.helpers.plot_pca_vectors(comp, nstds=2, scale=100.0, transform=None,
                                              ax=None, colors=None, linestyle=None,
                                              **kwargs)
```

Plot vectors corresponding to principal components and their magnitudes.

#### Parameters

- **comp** (`numpy.ndarray`) – Composition to use.
- **nstds** (`int`) – Multiplier for magnitude of individual principal component vectors.
- **scale** (`float`) – Scale applying to all x-y data points. For intergration with python-ternary.
- **transform** (`callable`) – Function for transformation of data prior to plotting (to either 2D or 3D).
- **ax** (`matplotlib.axes.Axes`) – Axes to plot on.

#### Returns

`ax`

#### Return type

`matplotlib.axes.Axes`

---

#### Todo:

- Minor reimplementaion of the sklearn PCA to avoid dependency.

[https://en.wikipedia.org/wiki/Principal\\_component\\_analysis](https://en.wikipedia.org/wiki/Principal_component_analysis)

---

`pyrolite.util.plot.helpers.plot_2dhull(data, ax=None, splines=False, s=0, **plotkwargs)`

Plots a 2D convex hull around an array of xy data points.

`pyrolite.util.plot.helpers.plot_cooccurrence(arr, ax=None, normalize=True, log=False, colorbar=False, **kwargs)`

Plot the co-occurrence frequency matrix for a given input.

#### Parameters

- **ax** (`matplotlib.axes.Axes`, `None`) – The subplot to draw on.
- **normalize** (`bool`) – Whether to normalize the cooccurrence to compare disparate variables.
- **log** (`bool`) – Whether to take the log of the cooccurrence.
- **colorbar** (`bool`) – Whether to append a colorbar.

#### Returns

Axes on which the cooccurrence plot is added.

#### Return type

`matplotlib.axes.Axes`

`pyrolite.util.plot.helpers.nan_scatter(xdata, ydata, ax=None, axes_width=0.2, **kwargs)`

Scatter plot with additional marginal axes to plot data for which data is partially missing. Additional keyword arguments are passed to matplotlib.

#### Parameters

- **xdata** (`numpy.ndarray`) – X data
- **ydata** (`class:numpy.ndarray | pd.Series`) – Y data
- **ax** (`matplotlib.axes.Axes`) – Axes on which to plot.
- **axes\_width** (`float`) – Width of the marginal axes.

#### Returns

Axes on which the nan\_scatter is plotted.

#### Return type

`matplotlib.axes.Axes`

`pyrolite.util.plot.helpers.init_spherical_octant(angle_indicated=30, labels=None, view_init=(25, 55), fontsize=10, **kwargs)`

Initialize a figure with a 3D octant of a unit sphere, appropriately labeled with regard to angles corresponding to the handling of the respective compositional data transformation function (`sphere()`).

#### Parameters

- **angle\_indicated** (`float`) – Angle relative to axes for indicating the relative positioning, optional.
- **labels** (`list`) – Optional specification of data/axes labels. This will be used for labelling of both the axes and optionally-added arcs specifying which angles are represented.

#### Returns

**ax** – Initialized 3D axis.

**Return type**`matplotlib.axes.Axes3D`**pyrolite.util.plot.interpolation**

Line interpolation for matplotlib lines and paths.

```
pyrolite.util.plot.interpolation.interpolate_path(path, resolution=100, periodic=False,
                                                  aspath=True, closefirst=False,
                                                  **kwargs)
```

Obtain the interpolation of an existing path at a given resolution. Keyword arguments are forwarded to `scipy.interpolate.splprep()`.

**Parameters**

- **path** (`matplotlib.path.Path`) – Path to interpolate.
- **resolution** :class:`int` – Resolution at which to obtain the new path. The vertices of the new path will have shape *(resolution, 2)*.
- **periodic** (`bool`) – Whether to use a periodic spline.
- **periodic** (`bool`) – Whether to return a `matplotlib.path.Path`, or simply a tuple of x-y arrays.
- **closefirst** (`bool`) – Whether to first close the path by appending the first point again.

**Returns**

Interpolated `Path` object, if *aspath* is True, else a tuple of x-y arrays.

**Return type**`matplotlib.path.Path | tuple`

```
pyrolite.util.plot.interpolation.interpolated_patch_path(patch, resolution=100,
                                                         **kwargs)
```

Obtain the periodic interpolation of the existing path of a patch at a given resolution.

**Parameters**

- **patch** (`matplotlib.patches.Patch`) – Patch to obtain the original path from.
- **resolution** :class:`int` – Resolution at which to obtain the new path. The vertices of the new path will have shape *(resolution, 2)*.

**Returns**

Interpolated `Path` object.

**Return type**`matplotlib.path.Path`

```
pyrolite.util.plot.interpolation.get_contour_paths(src, resolution=100, minsize=3)
```

Extract the paths of contours from a contour plot.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axes to extract contours from.
- **resolution** (`int`) – Resolution of interpolated splines to return.

**Returns**

- **contourspaths** (`list` (`list`)) – List of lists, each representing one line collection (a single contour). In the case where this contour is multimodal, there will be multiple paths for each contour.
- **contournames** (`list`) – List of names for contours, where they have been labelled, and there are no other text artists on the figure.
- **contourstyles** (`list`) – List of styles for contours.

## Notes

This method assumes that contours are the only `matplotlib.collections.LineCollection` objects within an axes; and when this is not the case, additional non-contour objects will be returned.

## pyrolite.util.plot.legend

Functions for creating and modifying legend entries for matplotlib.

---

### Todo:

- Functions for working with and modifying legend entries.  
ax.lines + ax.patches + ax.collections + ax.containers, handle ax.parasites
- 

`pyrolite.util.plot.legend.proxy_rect(**kwargs)`

Generates a legend proxy for a filled region.

#### Return type

`matplotlib.patches.Rectangle`

`pyrolite.util.plot.legend.proxy_line(**kwargs)`

Generates a legend proxy for a line region.

#### Return type

`matplotlib.lines.Line2D`

`pyrolite.util.plot.legend.modify_legend_handles(ax, **kwargs)`

Modify the handles of a legend based for a single axis.

#### Parameters

**ax** (`matplotlib.axes.Axes`) – Axis for which to obtain modified legend handles.

#### Returns

- **handles** (`list`) – Handles to be passed to a legend call.
- **labels** (`list`) – Labels to be passed to a legend call.

## pyrolite.util.plot.style

Functions for automated plot styling and argument handling.

**ivar** `DEFAULT_CONT_COLORMAP`

Default continuous colormap.

**vartype** `DEFAULT_CONT_COLORMAP`

`matplotlib.colors.ScalarMappable`

**ivar** `DEFAULT_DISC_COLORMAP`

Default discrete colormap.

**vartype** `DEFAULT_DISC_COLORMAP`

`matplotlib.colors.ScalarMappable`

`pyrolite.util.plot.style.linekwargs(kwargs)`

Get a subset of keyword arguments to pass to a matplotlib line-plot call.

**Parameters**

**kwargs** (`dict`) – Dictionary of keyword arguments to subset.

**Return type**

`dict`

`pyrolite.util.plot.style.scatterkwargs(kwargs)`

Get a subset of keyword arguments to pass to a matplotlib scatter call.

**Parameters**

**kwargs** (`dict`) – Dictionary of keyword arguments to subset.

**Return type**

`dict`

`pyrolite.util.plot.style.patchkwargs(kwargs)`

`pyrolite.util.plot.style.marker_cycle(markers=['D', 's', 'o', '+', '*'])`

Cycle through a set of markers.

**Parameters**

**markers** (`list`) – List of markers to provide to matplotlib.

`pyrolite.util.plot.style.mappable_from_values(values,`

`cmap=<matplotlib.colors.ListedColormap object>, norm=None, **kwargs)`

Create a scalar mappable object from an array of values.

**Return type**

`matplotlib.cm.ScalarMappable`

`pyrolite.util.plot.style.ternary_color(tlr, alpha=1.0, colors=([1, 0, 0], [0, 1, 0], [0, 0, 1]),  
coefficients=(1, 1, 1))`

Color a set of points by their ternary combinations of three specified colors.

**Parameters**

- **tlr** (`numpy.ndarray`)
- **alpha** (`float`) – Alpha coefficient for the colors; to be applied *multiplicatively* with any existing alpha value (for RGBA colors specified).

- **colors** (`tuple`) – Set of colours corresponding to the top, left and right vertices, respectively.
- **coefficients** (`tuple`) – Coefficients for the ternary data to adjust the centre.

**Returns**

Color array for the ternary points.

**Return type**

`numpy.ndarray`

```
pyrolite.util.plot.style.color_ternary_polygons_by_centroid(ax=None, patches=None,
                                                            alpha=1.0, colors=([1, 0,
0], [0, 1, 0], [0, 0, 1]),
                                                            coefficients=(1, 1, 1))
```

Color a set of polygons within a ternary diagram by their centroid colors.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Ternary axes to check for patches, if patches is not supplied.
- **patches** (`list`) – List of ternary-hosted patches to apply color to.
- **alpha** (`float`) – Alpha coefficient for the colors; to be applied *multiplicatively* with any existing alpha value (for RGBA colors specified).
- **colors** (`tuple`) – Set of colours corresponding to the top, left and right vertices, respectively.
- **coefficients** (`tuple`) – Coefficients for the ternary data to adjust the centre.

**Returns**

**patches** – List of patches, with updated facecolors.

**Return type**

`list`

## pyrolite.util.plot.transform

Transformation utilites for matplotlib.

```
pyrolite.util.plot.transform.affine_transform(mtx=array([[1, 0, 0], [0, 1, 0], [0, 0, 1]]))
```

Construct a function which will perform a 2D affine transform based on a 3x3 affine matrix.

**Parameters**

**mtx** (`numpy.ndarray`)

```
pyrolite.util.plot.transform.tlr_to_xy(tlr)
```

Transform a ternary coordinate system (top-left-right) to an xy-cartesian coordinate system.

**Parameters**

**tlr** (`numpy.ndarray`) – Array of shape (n, 3) in the t-l-r coordinate system.

**Returns**

**xy** – Array of shape (n, 2) in the x-y coordinate system.

**Return type**

`numpy.ndarray`

`pyrolite.util.plot.transform.xy_to_tlr(xy)`

**Parameters**

**xy** (`numpy.ndarray`) – Array of shape (n, 2) in the x-y coordinate system.

**Returns**

**tlr** – Array of shape (n, 3) in the t-l-r coordinate system.

**Return type**

`numpy.ndarray`

`pyrolite.util.plot.transform.ABC_to_xy(ABC, xscale=1.0, yscale=1.0)`

Convert ternary compositional coordinates to x-y coordinates for visualisation within a triangle.

**Parameters**

- **ABC** (`numpy.ndarray`) – Ternary array (samples, 3).
- **xscale** (`float`) – Scale for x-axis.
- **yscale** (`float`) – Scale for y-axis.

**Returns**

Array of x-y coordinates (samples, 2)

**Return type**

`numpy.ndarray`

`pyrolite.util.plot.transform.xy_to_ABC(xy, xscale=1.0, yscale=1.0)`

Convert x-y coordinates within a triangle to compositional ternary coordinates.

**Parameters**

- **xy** (`numpy.ndarray`) – XY array (samples, 2).
- **xscale** (`float`) – Scale for x-axis.
- **yscale** (`float`) – Scale for y-axis.

**Returns**

Array of ternary coordinates (samples, 3)

**Return type**

`numpy.ndarray`

## **pyrolite.util.text**

`pyrolite.util.text.to_width(multiline_string, width=79, **kwargs)`

Uses builtin textwrap for text wrapping to a specific width.

`pyrolite.util.text.normalise_whitespace(strg)`

Substitutes extra tabs, newlines etc. for a single space.

`pyrolite.util.text.remove_prefix(z, prefix)`

Remove a specific prefix from the start of a string.

`pyrolite.util.text.remove_suffix(x, suffix='')`

Remove a specific suffix from the end of a string.

`pyrolite.util.text.quoted_string(s)`

```
pyrolite.util.text.titlecase(s, exceptions=['and', 'in', 'a'], abbrev=['ID', 'IGSN',  
                                'CIA', 'CIW', 'PIA', 'SAR', 'SiTiIndex', 'WIP'],  
                             capitalize_first=True, split_on=['\\s_']+ , delim="")
```

Formats strings in CamelCase, with exceptions for simple articles and omitted abbreviations which retain their capitalization.

---

**Todo:**

- Option for retaining original CamelCase.
- 

```
pyrolite.util.text.string_variations(names, preprocess=['lower', 'strip'],  
                                     swaps=[(' ', '_'), (' ', '_'), ('-', ''), ('_', ' '),  
                                             ('-', ''), ('_', ' ')])
```

Returns equivalent string variations based on an input set of strings.

**Parameters**

- **names** (*list, str*) – String or list of strings to generate name variations of.
- **preprocess** (*list*) – List of preprocessing string functions to apply before generating variations.
- **swaps** (*list*) – List of tuples for str.replace(out, in).

**Returns**

Set (or SortedSet, if sortedcontainers installed) of unique string variations.

**Return type**

set

```
pyrolite.util.text.parse_entry(entry, regex='(\\s)*?(?P<value>[\\w\\W]+)(\\s)*?',  
                               delimiter=',', values_only=True, first_only=True,  
                               errors=None, replace_nan='None')
```

Parses an arbitrary string data entry to return values based on a regular expression containing named fields including 'value' (and any others). If the entry is of non-string type, this will return the value (e.g. int, float, NaN, None).

**Parameters**

- **entry** (*str*) – String entry which to search for the regex pattern.
- **regex** (*str*) – Regular expression to compile and use to search the entry for a value.
- **delimiter** (*str, : , ' , '*) – Optional delimiter to split the string in case of multiple inclusion.
- **values\_only** (*bool, True*) – Option to return only values (single or list), or to instead return the dictionary corresponding to the matches.
- **first\_only** (*bool, True*) – Option to return only the first match, or else all matches
- **errors** – Error value to denote 'no match'. Not yet implemented.



`pyrolite.util.text.split_records(data, delimiter='\r\n')`

Splits records in a csv where quotation marks are used. Splits on a delimiter followed by an even number of quotation marks.

`pyrolite.util.text.slugify(value, delim='-')`

Normalizes a string, removes non-alpha characters, converts spaces to delimiters.

**Parameters**

- **value** (`str`) – String to slugify.
- **delim** (`str`) – Delimiter to replace whitespace with.

**Return type**

`str`

`pyrolite.util.text.int_to_alpha(num)`

Encode an integer into alpha characters, useful for sequences of axes/figures.

**Parameters**

**int** (`int`) – Integer to encode.

**Returns**

Alpha-encoding of a small integer.

**Return type**

`str`

## **pyrolite.util.web**

`pyrolite.util.web.urlify(url)`

Strip a string to return a valid URL.

`pyrolite.util.web.internet_connection(target='pypi.org', secure=True)`

Tests for an active internet connection, based on an optionally specified target.

**Parameters**

**target** (`str`) – URL to check connectivity, defaults to `www.google.com`

**Returns**

Boolean indication of whether a HTTP connection can be established at the given url.

**Return type**

`bool`

`pyrolite.util.web.download_file(url: str, encoding='UTF-8', postprocess=None)`

Downloads a specific file from a url.

**Parameters**

- **url** (`str`) – URL of specific file to download.
- **encoding** (`str`) – String encoding.
- **postprocess** (callable) – Callable function to post-process the requested content.

## pyrolite.util.time

`pyrolite.util.time.listify(df, axis=1)`

Condense text information across columns into a single list.

### Parameters

- **df** (`pandas.DataFrame`) – Dataframe (or slice of dataframe) to condense along axis.
- **axis** (`int`) – Axis to condense along.

`pyrolite.util.time.age_name(agenamelist, prefixes=['Lower', 'Middle', 'Upper'], suffixes=['Stage', 'Series'])`

Condenses an agename list to a specific agename, given a subset of ambiguous\_names.

### Parameters

- **agenamelist** (`list`) – List of name components (i.e. [Eon, Era, Period, Epoch])
- **prefixes** (`list`) – Name components which occur prior to the higher order classification (e.g. "Upper Triassic").
- **suffixes** (`list`) – Name components which occur after the higher order classification (e.g. "Cambrian Series 2").

`pyrolite.util.time.import_colors(filename=PosixPath('/home/docs/checkouts/readthedocs.org/user_builds/pyrolite'), delim='/')`

Import a list of timescale names with associated colors.

`pyrolite.util.time.timescale_reference_frame(filename=PosixPath('/home/docs/checkouts/readthedocs.org/user_builds/pyrolite'), info_cols=['Start', 'End', 'Aliases'], color_info=None)`

Rearrange the text-based timescale dataframe. Utility function for timescale class.

### Parameters

- **filename** (`str` | `pathlib.Path`) – File from which to generate the timescale information.
- **info\_cols** (`list`) – List of columns beyond hierarchial group labels (e.g. Eon, Era..).

### Returns

Dataframe containing timescale information.

### Return type

`pandas.DataFrame`

`class pyrolite.util.time.Timescale(filename=None)`

### build()

Build the timescale from data within file.

`text2age(entry, nulls=[None, 'None', 'none', nan, 'NaN'])`

Converts a text-based age to the corresponding age range (in Ma).

String-based entries return (max\_age, min\_age). Collection-based entries return a list of tuples.

**Parameters**

**entry** (`str`) – String name, or series of string names, for geological age range.

**Returns**

Tuple or list of tuples.

**Return type**

`tuple | list (tuple)`

**named\_age**(*age*, *level*='Specific', *\*\*kwargs*)

Converts a numeric age (in Ma) to named age at a specific level.

**Parameters**

- **age** (`float`) – Numeric age in Ma.
- **level** (`str`, {'Eon', 'Era', 'Period', 'Superepoch', 'Epoch', 'Age', 'Specific'}) – Level of specificity.

**Returns**

String representation for the entry.

**Return type**

`str`

## pyrolite.util.math

`pyrolite.util.math.eigsorted`(*cov*)

Returns arrays of eigenvalues and eigenvectors sorted by magnitude.

**Parameters**

**cov** (`numpy.ndarray`) – Covariance matrix to extract eigenvalues and eigenvectors from.

**Returns**

- **vals** (`numpy.ndarray`) – Sorted eigenvalues.
- **vecs** (`numpy.ndarray`) – Sorted eigenvectors.

`pyrolite.util.math.augmented_covariance_matrix`(*M*, *C*)

Constructs an augmented covariance matrix from means *M* and covariance matrix *C*.

**Parameters**

- **M** (`numpy.ndarray`) – Array of means.
- **C** (`numpy.ndarray`) – Covariance matrix.

**Returns**

Augmented covariance matrix *A*.

**Return type**

`numpy.ndarray`

## Notes

Augmented covariance matrix constructed from mean of shape (D, ) and covariance matrix of shape (D, D) as follows:

$$\begin{array}{c|c} -1 & M.T \\ \hline M & C \end{array}$$

`pyrolite.util.math.interpolate_line(x, y, n=0, logy=False)`

Add intermediate evenly spaced points interpolated between given x-y coordinates, assuming the x points are the same.

### Parameters

- **x** (`numpy.ndarray`) – 1D array of x values.
- **y** (`numpy.ndarray`) – ND array of y values.

`pyrolite.util.math.grid_from_ranges(X, bins=100, **kwargs)`

Create a meshgrid based on the ranges along columns of array X.

### Parameters

- **X** (`numpy.ndarray`) – Array of shape (samples, dimensions) to create a meshgrid from.
- **bins** (`int` | `tuple`) – Shape of the meshgrid. If an integer, provides a square mesh. If a tuple, values for each column are required.

### Return type

`numpy.ndarray`

## Notes

Can pass keyword arg `indexing = {'xy', 'ij'}`

`pyrolite.util.math.flattengrid(grid)`

Convert a collection of arrays to a concatenated array of flattened components. Useful for passing meshgrid values to a function which accepts arguments of shape (samples, dimensions).

### Parameters

**grid** (`list`) – Collection of arrays (e.g. a meshgrid) to flatten and concatenate.

### Return type

`numpy.ndarray`

`pyrolite.util.math.linspc_(_min, _max, step=0.0, bins=20)`

Linear spaced array, with optional step for grid margins.

### Parameters

- **\_min** (`float`) – Minimum value for spaced range.
- **\_max** (`float`) – Maximum value for spaced range.
- **step** (`float`, 0.0) – Step for expanding at grid edges. Default of 0.0 results in no expansion.

- **bins** (*int*) – Number of bins to divide the range (adds one by default).

**Returns**

Linearly-spaced array.

**Return type**

`numpy.ndarray`

`pyrolite.util.math.logspc_(_min, _max, step=1.0, bins=20)`

Log spaced array, with optional step for grid margins.

**Parameters**

- **\_min** (`float`) – Minimum value for spaced range.
- **\_max** (`float`) – Maximum value for spaced range.
- **step** (`float`, 1.0) – Step for expanding at grid edges. Default of 1.0 results in no expansion.
- **bins** (*int*) – Number of bins to divide the range (adds one by default).

**Returns**

Log-spaced array.

**Return type**

`numpy.ndarray`

`pyrolite.util.math.logrng_(v, exp=0.0)`

Range of a sample, where values <0 are excluded.

**Parameters**

- **v** (`list`; list-like) – Array of values to obtain a range from.
- **exp** (`float`, (0, 1)) – Fractional expansion of the range.

**Returns**

Min, max tuple.

**Return type**

`tuple`

`pyrolite.util.math.linrng_(v, exp=0.0)`

Range of a sample, where values <0 are included.

**Parameters**

- **v** (`list`; list-like) – Array of values to obtain a range from.
- **exp** (`float`, (0, 1)) – Fractional expansion of the range.

**Returns**

Min, max tuple.

**Return type**

`tuple`

`pyrolite.util.math.isclose(a, b)`

Implementation of np.isclose with equal nan.

**Parameters**

**a,b** (`float` | `numpy.ndarray`) – Numbers or arrays to compare.

**Return type**`bool``pyrolite.util.math.is_numeric(obj)`

Check for numerical behaviour.

**Parameters**

**obj** – Object to check.

**Return type**`bool``pyrolite.util.math.significant_figures(n, unc=None, max_sf=20, rtol=1e-20)`

Iterative method to determine the number of significant digits for a given float, optionally providing an uncertainty.

**Parameters**

- **n** (`float`) – Number from which to ascertain the significance level.
- **unc** (`float`, `None`) – Uncertainty, which if provided is used to derive the number of significant digits.
- **max\_sf** (`int`) – An upper limit to the number of significant digits suggested.
- **rtol** (`float`) – Relative tolerance to determine similarity of numbers, used in calculations.

**Returns**

Number of significant digits.

**Return type**`int``pyrolite.util.math.signify_digit(n, unc=None, leeway=0, low_filter=True)`

Reformats numbers to contain only significant\_digits. Uncertainty can be provided to digits with relevant precision.

**Parameters**

- **n** (`float`) – Number to reformat
- **unc** (`float`, `None`) – Absolute uncertainty on the number, optional.
- **leeway** (`int`, 0) – Manual override for significant figures. Positive values will force extra significant figures; negative values will remove significant figures.
- **low\_filter** (`bool`, `True`) – Whether to return `np.nan` in place of values which are within precision equal to zero.

**Returns**

Reformatted number.

**Return type**`float`

## Notes

- Will not pad 0s at the end or before floats.

`pyrolite.util.math.most_precise(arr)`

Get the most precise element from an array.

### Parameters

**arr** (`numpy.ndarray`) – Array to obtain the most precise element/subarray from.

### Returns

Returns the most precise array element (for `ndim=1`), or most precise subarray (for `ndim > 1`).

### Return type

`float` | `numpy.ndarray`

`pyrolite.util.math.equal_within_significance(arr, equal_nan=False, rtol=1e-15)`

Test whether elements of an array are equal within the precision of the least precise.

### Parameters

- **arr** (`numpy.ndarray`) – Array to test.
- **equal\_nan** (`bool`, `False`) – Whether to consider `np.nan` elements equal to one another.
- **rtol** (`float`) – Relative tolerance for comparison.

### Return type

`bool` | `numpy.ndarray` (:class:`bool`)

`pyrolite.util.math.helmert_basis(D: int, full=False, **kwargs)`

Generate a set of orthogonal basis vectors in the form of a helmert matrix.

### Parameters

**D** (`int`) – Dimension of compositional vectors.

### Returns

(D-1, D) helmert matrix corresponding to default orthogonal basis.

### Return type

`numpy.ndarray`

`pyrolite.util.math.symbolic_helmert_basis(D, full=False)`

Get a symbolic representation of a Helmert Matrix.

### Parameters

- **D** (`int`) – Order of the matrix. Equivalent to dimensionality for compositional data analysis.
- **full** (`bool`) – Whether to return the full matrix, or alternatively exclude the first row. Analogous to the option for `scipy.linalg.helmert()`.

### Return type

`sympy.matrices.dense.DenseMatrix`

`pyrolite.util.math.on_finite(X, f)`

Calls a function on an array ignoring `np.nan` and `+/- np.inf`. Note that the shape of the output may be different to that of the input.

**Parameters**

- **X** (`numpy.ndarray`) – Array on which to perform the function.
- **f** (Callable) – Function to call on the array.

**Return type**

`numpy.ndarray`

`pyrolite.util.math.nancov(X)`

Generates a covariance matrix excluding nan-components.

**Parameters**

**X** (`numpy.ndarray`) – Input array for which to derive a covariance matrix.

**Return type**

`numpy.ndarray`

`pyrolite.util.math.solve_ratios(*eqs, evaluate=True)`

Solve a ternary system (top-left-right) given two constraints on two ratios, which together describe intersecting lines/a point.

## pyrolite.util.lambdas

`pyrolite.util.lambdas.calc_lambdas(df, params=None, degree=4, exclude=[],  
algorithm='ONeill', anomalies=[],  
fit_tetrads=False, sigmas=None,  
add_uncertainties=False, add_X2=False,  
**kwargs)`

Parameterises values based on linear combination of orthogonal polynomials over a given set of values for independent variable  $x^1$ . This function expects to receive data already normalised and log-transformed.

**Parameters**

- **df** (`pd.DataFrame`) – Dataframe containing REE Data.
- **params** (`list` | `str`) – Pre-computed parameters for the orthogonal polynomials (a list of tuples). Optionally specified, otherwise defaults the parameterisation as in O'Neill (2016).<sup>?</sup> If a string is supplied, "O'Neill (2016)" or similar will give the original defaults, while "full" will use all of the REE (including Eu) as a basis for the orthogonal polynomials.
- **degree** (`int`) – Degree of orthogonal polynomial fit.
- **exclude** (`list`) – REE to exclude from the *fit*.
- **algorithm** (`str`) – Algorithm to use for fitting the orthogonal polynomials.
- **anomalies** (`list`) – List of relative anomalies to append to the dataframe.

---

<sup>1</sup> O'Neill HSC (2016) The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. J Petrology 57:1463–1508. doi: 10.1093/petrology/egw047



- **fit\_tetrads** (`bool`) – Whether to fit tetrad functions in addition to orthogonal polynomial functions. This will force the use of the optimization algorithm.
- **sigmas** (`float` | `numpy.ndarray`) – Single value or 1D array of observed value uncertainties.
- **add\_uncertainties** (`bool`) – Whether to append estimated parameter uncertainties to the dataframe.
- **add\_X2** (`bool`) – Whether to append the chi-squared values (2) to the dataframe.

**Return type**

pd.DataFrame

**See also:***orthogonal\_polynomial\_constants()*, *lambda\_lnREE()*, *normalize\_to()***References****pyrolite.util.lambdas.params**

Functions to generate parameters for the construction of orthogonal polynomials which are used to fit REE patterns.

```
pyrolite.util.lambdas.params.orthogonal_polynomial_constants(xs,
                                                             degree=3,
                                                             rounding=None,
                                                             tol=1e-14)
```

Finds the parameters  $(\beta_0), (\gamma_0, \gamma_1), (\delta_0, \delta_1, \delta_2)$  etc. for constructing orthogonal polynomial functions  $f(x)$  over a fixed set of values of independent variable  $x$ . Used for obtaining lambda values for dimensional reduction of REE data<sup>2</sup>.

**Parameters**

- **xs** (`numpy.ndarray`) – Indexes over which to generate the orthogonal polynomials.
- **degree** (`int`) – Maximum polynomial degree. E.g. 2 will generate constant, linear, and quadratic polynomial components.
- **tol** (`float`) – Convergence tolerance for solver.
- **rounding** (`int`) – Precision for the orthogonal polynomial coefficients.

**Returns**

List of tuples corresponding to coefficients for each of the polynomial components. I.e the first tuple will be empty, the second will contain a single coefficient etc.

**Return type**

list

<sup>2</sup> O'Neill HSC (2016) The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. J Petrology 57:1463–1508. doi: 10.1093/petrology/egw047

## Notes

Parameters are used to construct orthogonal polynomials of the general form:

$$\begin{aligned} f(x) = & a_0 \\ & + a_1 * (x - \beta) \\ & + a_2 * (x - \gamma_0) * (x - \gamma_1) \\ & + a_3 * (x - \delta_0) * (x - \delta_1) * (x - \delta_2) \end{aligned}$$

See also:

`calc_lambdas()`, `lambda_lnREE()`

## References

`pyrolite.util.lambdas.params.parse_sigmas(size, sigmas=None)`

Disambiguate a value or set of sigmas for a dataset for use in lambda-fitting algorithms.

### Parameters

**sigmas** (`float` | `numpy.ndarray`) – 2D array of REE uncertainties.  
Values as fractional uncertainties (i.e.  $\sigma_{REE}/REE$ ).

### Returns

**sigmas** – 1D array of sigmas ( $\sigma_{REE}/REE$ ).

### Return type

`float` | `numpy.ndarray`

## Notes

Note that the y-array is passed here only to infer the shape which should be assumed by the uncertainty array. Through propagation of uncertainties, the uncertainty on the natural logarithm of the normalised REE values are equivalent to  $\sigma_{REE}/REE$  where the uncertainty in the reference composition is assumed to be zero. Thus, standard deviations of 1% in REE will result in  $\sigma = 0.01$  for the log-transformed REE. If no sigmas are provided, 1% uncertainty will be assumed and an array of 0.01 will be returned.

## `pyrolite.util.lambdas.eval`

Generation and evaluation of orthogonal polynomial and tetrad functions from sets of parameters (the sequence of polynomial roots and tetrad centres and widths).

`pyrolite.util.lambdas.eval.lambda_poly(x, ps)`

Evaluate polynomial  $\lambda_{n(x)}$  given a tuple of parameters  $ps$  with length equal to the polynomial degree.

### Parameters

- **x** (`numpy.ndarray`) – X values to calculate the function at.
- **ps** (`tuple`) – Parameter set tuple. E.g. parameters  $(a, b)$  from  $f(x) = (x - a)(x - b)$ .

### Return type

`numpy.ndarray`

`pyrolite.util.lambdas.eval.tetrad(x, centre, width)`

Evaluate  $f(z)$  describing a tetrad with specified centre and width.

#### Parameters

- **x**
- **centre** (`float`)
- **width** (`float`)

`pyrolite.util.lambdas.eval.get_tetrads_function(params=None)`

`pyrolite.util.lambdas.eval.get_lambda_poly_function(lambdas: ndarray,  
params=None,  
radii=None, degree=5)`

Expansion of lambda parameters back to the original space. Returns a function which evaluates the sum of the orthogonal polynomials at given  $x$  values.

#### Parameters

- **lambdas** (`numpy.ndarray`) – Lambda values to weight combination of polynomials.
- **params** (`list ( tuple )`) – Parameters for the orthogonal polynomial decomposition.
- **radii** (`numpy.ndarray`) – Radii values used to construct the lambda values.<sup>3</sup>
- **degree** (`int`) – Degree of the orthogonal polynomial decomposition.<sup>?</sup>

See also:

`calc_lambdas()`, `orthogonal_polynomial_constants()`, `lambda_lnREE()`

#### Notes

`pyrolite.util.lambdas.eval.get_function_components(radii, params=None,  
fit_tetrads=False,  
tetrad_params=None,  
degree=5, **kwargs)`

### pyrolite.util.lambdas.oneill

Linear algebra methods for fitting a series of orthogonal polynomial functions to REE patterns.

`pyrolite.util.lambdas.oneill.get_polynomial_matrix(radii, params=None)`

Create the matrix  $A$  with polynomial components across the columns, and increasing order down the rows.

#### Parameters

- **radii** (`list`, `numpy.ndarray`) – Radii at which to evaluate the orthogonal polynomial.

<sup>3</sup> Only needed if parameters are not supplied

- **params** (`tuple`) – Tuple of constants for the orthogonal polynomial.

**Return type**`numpy.ndarray`**See also:**`orthogonal_polynomial_constants()`

```
pyrolite.util.lambdas.oneill.lambdas_ONEi112016(df, radii, params=None,
                                                sigmas=None,
                                                add_X2=False,
                                                add_uncertainties=False,
                                                **kwargs)
```

Implementation of the original algorithm.<sup>4</sup>

**Parameters**

- **df** (`pandas.DataFrame` | `pandas.Series`) – Dataframe of REE data, with sample analyses organised by row.
- **radii** (`list`, `numpy.ndarray`) – Radii at which to evaluate the orthogonal polynomial.
- **params** (`tuple`) – Tuple of constants for the orthogonal polynomial.
- **sigmas** (`float` | `numpy.ndarray`) – Single value or 1D array of normalised observed value uncertainties ( $\sigma_{REE}/REE$ ).
- **add\_X2** (`bool`) – Whether to append the chi-squared values (2) to the dataframe/series.
- **add\_uncertainties** (`bool`) – Append parameter standard errors to the dataframe/series.

**Return type**`pandas.DataFrame`**See also:**`orthogonal_polynomial_constants()`, `lambda_lnREE()`**References****pyrolite.util.lambdas.opt**

Functions for optimization-based REE profile fitting and parameter uncertainty estimation.

```
pyrolite.util.lambdas.opt.pcov_from_jac(jac)
```

Extract a covariance matrix from a Jacobian matrix returned from `scipy.optimize` functions.

**Parameters**

**jac** (`numpy.ndarray`) – Jacobian array.

**Returns**

**pcov** – Square covariance array; this hasn't yet been scaled by residuals.

---

<sup>4</sup> O'Neill HSC (2016) The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. *J Petrology* 57:1463–1508. doi: 10.1093/petrology/egw047

**Return type**`numpy.ndarray`

```
pyrolite.util.lambdas.opt.linear_fit_components(y, x0, func_components,
                                                sigmas=None)
```

Fit a weighted sum of function components using linear algebra.

**Parameters**

- **y** (`numpy.ndarray`) – Array of target values to fit.
- **x0** (`numpy.ndarray`) – Starting guess for the function weights.
- **func\_components** (`list ( numpy.ndarray )`) – List of arrays representing static/evaluated function components.
- **sigmas** (`float | numpy.ndarray`) – Single value or 1D array of normalised observed value uncertainties ( $\sigma_{REE}/REE$ ).

**Returns**

**B, s, 2** – Arrays for the optimized parameter values (B; (n, d)), parameter uncertainties (s, 1; (n, d)) and chi-chi\_squared (2; (n, 1)).

**Return type**`numpy.ndarray`

```
pyrolite.util.lambdas.opt.optimize_fit_components(y, x0, func_components,
                                                  residu-
                                                  als_function=<function
                                                  _residuals_func>,
                                                  sigmas=None)
```

Fit a weighted sum of function components using `scipy.optimize.least_squares()`.

**Parameters**

- **y** (`numpy.ndarray`) – Array of target values to fit.
- **x0** (`numpy.ndarray`) – Starting guess for the function weights.
- **func\_components** (`list ( numpy.ndarray )`) – List of arrays representing static/evaluated function components.
- **residuals\_function** (*callable*) – Callable function to compute residuals which accepts ordered arguments for weights, target values and function components.
- **sigmas** (`float | numpy.ndarray`) – Single value or 1D array of normalised observed value uncertainties ( $\sigma_{REE}/REE$ ).

**Returns**

**B, s, 2** – Arrays for the optimized parameter values (B; (n, d)), parameter uncertainties (s, 1; (n, d)) and chi-chi\_squared (2; (n, 1)).

**Return type**`numpy.ndarray`

```
pyrolite.util.lambdas.opt.lambdas_optimize(df, radii, params=None,
                                             fit_tetrads=False,
                                             tetrad_params=None,
                                             fit_method='opt', sigmas=None,
                                             add_uncertainties=False,
                                             add_X2=False, **kwargs)
```

Parameterises values based on linear combination of orthogonal polynomials over a given set of values for independent variable  $x$ .<sup>5</sup>

#### Parameters

- **df** (`pandas.DataFrame` | `:class: pandas.Series`) – Target data to fit. For geochemical data, this is typically normalised so we can fit a smooth function.
- **radii** (`list`, `numpy.ndarray`) – Radii at which to evaluate the orthogonal polynomial.
- **params** (`list`, `None`) – Orthogonal polynomial coefficients (see `orthogonal_polynomial_constants()`).
- **fit\_tetrads** (`bool`) – Whether to also fit the patterns for tetrads.
- **tetrad\_params** (`list`) – List of parameter sets for tetrad functions.
- **fit\_method** (`str`) – Which fit method to use: "optimization" or "linear".
- **sigmas** (`float` | `numpy.ndarray`) – Single value or 1D array of observed value uncertainties.
- **add\_uncertainties** (`bool`) – Whether to append estimated parameter uncertainties to the dataframe/series.
- **add\_X2** (`bool`) – Whether to append the chi-squared values (2) to the dataframe/series.

#### Returns

Optimal results for weights of orthogonal polynomial regression (*lambdas*).

#### Return type

`numpy.ndarray` | (`numpy.ndarray`, `numpy.ndarray`)

#### See also:

`orthogonal_polynomial_constants()`, `lambda_lnREE()`

#### References

##### pyrolite.util.lambdas.plot

Functions for the visualisation of reconstructed and deconstructed parameterised REE profiles based on parameterisations using 'lambdas' (and tetrad-equivalent weights 'taus').

`pyrolite.util.lambdas.plot.plot_lambdas_components(lambdas, params=None, ax=None, **kwargs)`

Plot a decomposed orthogonal polynomial from a single set of lambda coefficients.

#### Parameters

- **lambdas** – 1D array of lambdas.

---

<sup>5</sup> O'Neill HSC (2016) The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. J Petrology 57:1463–1508. doi: 10.1093/petrology/egw047

- **params** (`list`) – List of orthongonal polynomial parameters, if defaults are not used.
- **ax** (`matplotlib.axes.Axes`) – Optionally specified axes to plot on.
- **index** (`str`) – Index to use for the plot (one of "index", "radii", "z").

**Return type**`matplotlib.axes.Axes`

```
pyrolite.util.lambdas.plot.plot_tetrads_components(taus,
                                                    tetrad_params=None,
                                                    ax=None, index='radii',
                                                    logy=True, drop0=True,
                                                    **kwargs)
```

Individually plot the four tetrad components for one set of \$ au\$.

**Parameters**

- **taus** (`numpy.ndarray`) – 1D array of \$ au\$ tetrad function coefficients.
- **tetrad\_params** (`list`) – List of tetrad parameters, if defaults are not used.
- **ax** (`matplotlib.axes.Axes`) – Optionally specified axes to plot on.
- **index** (`str`) – Index to use for the plot (one of "index", "radii", "z").
- **logy** (`bool`) – Whether to log-scale the y-axis.
- **drop0** (`bool`) – Whether to remove zeroes from the outputs such that individual tetrad functions are shown only within their respective bounds (and not across the entire REE, where their effective values are zero).

```
pyrolite.util.lambdas.plot.plot_profiles(coefficients, tetrads=False,
                                         params=None, tetrad_params=None,
                                         ax=None, index='radii', logy=False,
                                         **kwargs)
```

Plot the reconstructed REE profiles of a 2D dataset of coefficients (\$lambda\$, and optionally \$tau\$).

**Parameters**

- **coefficients** (`numpy.ndarray`) – 2D array of \$lambda\$ orthogonal polynomial coefficients, and optionally including \$tau\$ tetrad function coefficients in the last four columns (where `tetrads=True`).
- **tetrads** (`bool`) – Whether the coefficient array contains tetrad coefficients (\$tau\$).
- **params** (`list`) – List of orthongonal polynomial parameters, if defaults are not used.
- **tetrad\_params** (`list`) – List of tetrad parameters, if defaults are not used.

- **ax** (`matplotlib.axes.Axes`) – Optionally specified axes to plot on.
- **index** (`str`) – Index to use for the plot (one of "index", "radii", "z").
- **logy** (`bool`) – Whether to log-scale the y-axis.

**Return type**`matplotlib.axes.Axes`**pyrolite.util.lambdas.transform**

Functions for transforming ionic radii to and from atomic number for the visualisation of REE patterns.

```
pyrolite.util.lambdas.transform.REE_z_to_radii(z, fit=None, degree=7,  
                                                **kwargs)
```

Estimate the ionic radii which would be approximated by a given atomic number based on a provided (or calculated) fit for the Rare Earth Elements.

**Parameters**

- **z** (`float` | `list` | `numpy.ndarray`) – Atomic nubmers to be converted.
- **fit** (`callable`) – Callable function optionally specified; if not specified it will be calculated from Shannon Radii.
- **degree** (`int`) – Degree of the polynomial fit between atomic number and radii.

**Returns**

**r** – Approximate atomic nubmers for given radii.

**Return type**`float` | `numpy.ndarray`

```
pyrolite.util.lambdas.transform.REE_radii_to_z(r, fit=None, degree=7,  
                                                **kwargs)
```

Estimate the atomic number which would be approximated by a given ionic radii based on a provided (or calculated) fit for the Rare Earth Elements.

**Parameters**

- **r** (`float` | `list` | `numpy.ndarray`) – Radii to be converted.
- **fit** (`callable`) – Callable function optionally specified; if not specified it will be calculated from Shannon Radii.
- **degree** (`int`) – Degree of the polynomial fit between radii and atomic number.

**Returns**

**z** – Approximate atomic numbers for given radii.

**Return type**`float` | `numpy.ndarray`



## pyrolite.util.distributions

`pyrolite.util.distributions.get_scaler(*fs)`

Generate a function which will transform columns of an array based on input functions (e.g. `np.log` will log-transform the x values, `None`, `np.log` will log-transform the y values but not the x).

### Parameters

**fs** – A series of functions to apply to subsequent axes of an array.

`pyrolite.util.distributions.sample_kde(data, samples, renorm=False, transform=<function <lambda>>, bw_method=None)`

Sample a Kernel Density Estimate at points or a grid defined.

### Parameters

- **data** (`numpy.ndarray`) – Source data to estimate the kernel density estimate; observations should be in rows (`npoints`, `ndim`).
- **samples** (`numpy.ndarray`) – Coordinates to sample the KDE estimate at (`npoints`, `ndim`).
- **transform** – Transformation used prior to kernel density estimate.
- **bw\_method** (`str`, `float`, callable) – Method used to calculate the estimator bandwidth. See `scipy.stats.gaussian_kde()`.

### Return type

`numpy.ndarray`

`pyrolite.util.distributions.sample_ternary_kde(data, samples, transform=<function ILR>)`

Sample a Kernel Density Estimate in ternary space points or a grid defined by samples.

### Parameters

- **data** (`numpy.ndarray`) – Source data to estimate the kernel density estimate (`npoints`, `ndim`).
- **samples** (`numpy.ndarray`) – Coordinates to sample the KDE estimate at (`npoints`, `ndim`).
- **transform** – Log-transformation used prior to kernel density estimate.

### Return type

`numpy.ndarray`

`pyrolite.util.distributions.lognorm_to_norm(mu, s)`

Calculate mean and variance for a normal random variable from the lognormal parameters `mu` and `s`.

### Parameters

- **mu** (`float`) – Parameter `mu` for the lognormal distribution.
- **s** (`float`) – `sigma` for the lognormal distribution.

### Returns

- **mean** (`float`) – Mean of the normal distribution.

- **sigma** (`float`) – Variance of the normal distribution.

`pyrolite.util.distributions.norm_to_lognorm(mean, sigma, exp=True)`

Calculate mu and sigma parameters for a lognormal random variable with a given mean and variance. Lognormal with parameters mean and sigma.

#### Parameters

- **mean** (`float`) – Mean of the normal distribution.
- **sigma** (`float`) – sigma of the normal distribution.
- **exp** (`bool`) – If using the `scipy.stats` parameterisation; this uses `scale = np.exp(mu)`.

#### Returns

- **mu** (`float`) – Parameter mu for the lognormal distribution.
- **s** (`float`) – sigma of the lognormal distribution.

### pyrolite.util.synthetic

Utility functions for creating synthetic (geochemical) data.

`pyrolite.util.synthetic.random_cov_matrix(dim, sigmas=None, validate=False, seed=None)`

Generate a random covariance matrix which is symmetric positive-semidefinite.

#### Parameters

- **dim** (`int`) – Dimensionality of the covariance matrix.
- **sigmas** (`numpy.ndarray`) – Optionally specified sigmas for the variables.
- **validate** (`bool`) – Whether to validate output.

#### Returns

Covariance matrix of shape (dim, dim).

#### Return type

`numpy.ndarray`

---

#### Todo:

- Implement a characteristic scale for the covariance matrix.
- 

`pyrolite.util.synthetic.random_composition(size=1000, D=4, mean=None, cov=None, propnan=0.1, missing_columns=None, missing=None, seed=None)`

Generate a simulated random unimodal compositional dataset, optionally with missing data.

#### Parameters

- **size** (`int`) – Size of the dataset.
- **D** (`int`) – Dimensionality of the dataset.

- **mean** (`numpy.ndarray`, `None`) – Optional specification of mean composition.
- **cov** (`numpy.ndarray`, `None`) – Optional specification of covariance matrix (in log space).
- **propnan** (`float`, `[0, 1)`) – Proportion of missing values in the output dataset.
- **missing\_columns** (`int` | `tuple`) – Specification of columns to be missing. If an integer is specified, interpreted to be the number of columns containin missing data (at a proportion defined by *propnan*). If a tuple or list, the specific columns to contain missing data.
- **missing** (`str`, `None`) – Missingness pattern. If not `None`, one of "MCAR", "MAR", "MNAR".
  - If `missing` = "MCAR", data will be missing at random.
  - If `missing` = "MAR", data will be missing with some relationship to other parameters.
  - If `missing` = "MNAR", data will be thresholded at some lower bound.
- **seed** (`int`, `None`) – Random seed to use, optionally specified.

**Returns**

Simulated dataset with missing values.

**Return type**

`numpy.ndarray`

---

**Todo:**

- Add feature to translate rough covariance in D to logcovariance in D-1
  - Update the `:code: `missing = "MAR"`` example to be more realistic/variable.
- 

```
pyrolite.util.synthetic.normal_frame(columns=['SiO2', 'CaO', 'MgO', 'FeO',  
                                           'TiO2'], size=10, mean=None, **kwargs)
```

Creates a `pandas.DataFrame` with samples from a single multivariate-normal distributed composition.

**Parameters**

- **columns** (`list`) – List of columns to use for the dataframe. These won't have any direct impact on the data returned, and are only for labelling.
- **size** (`int`) – Index length for the dataframe.
- **mean** (`numpy.ndarray`, `None`) – Optional specification of mean composition.

**Return type**

`pandas.DataFrame`

```
pyrolite.util.synthetic.normal_series(index=['SiO2', 'CaO', 'MgO', 'FeO',  
                                           'TiO2'], mean=None, **kwargs)
```

Creates a `pandas.Series` with a single sample from a single multivariate-normal distributed composition.

#### Parameters

- **index** (`list`) – List of indexes for the series. These won't have any direct impact on the data returned, and are only for labelling.
- **mean** (`numpy.ndarray`, `None`) – Optional specification of mean composition.

#### Return type

`pandas.Series`

```
pyrolite.util.synthetic.example_spider_data(start='EMORB_SM89',
                                             norm_to='PM_PON', size=120,
                                             noise_level=0.5, offsets=None,
                                             units='ppm')
```

Generate some random data for demonstrating spider plots.

By default, this generates a composition based around EMORB, normalised to Primitive Mantle.

#### Parameters

- **start** (`str`) – Composition to start with.
- **norm\_to** (`str`) – Composition to normalise to. Can optionally specify `None`.
- **size** (`int`) – Number of observations to include (index length).
- **noise\_level** (`float`) – Log-units of noise (1sigma).
- **offsets** (`dict`) – Dictionary of offsets in log-units (in log units).
- **units** (`str`) – Units to use before conversion. Should have no effect other than reducing calculation times if `norm_to` is `None`.

#### Returns

**df** – Dataframe of example synthetic data.

#### Return type

`pandas.DataFrame`

```
pyrolite.util.synthetic.example_patterns_from_parameters(fit_parameters,
                                                         radii=None,
                                                         n=100, proportional_noise=0.15,
                                                         includes_tetrads=False,
                                                         columns=None)
```

## pyrolite.util.spatial

Basic spatial utility functions.

```
pyrolite.util.spatial.great_circle_distance(a, b=None, absolute=False,
                                             degrees=True, r=6371.0088,
                                             method=None, dtype='float32',
                                             max_memory_fraction=0.25)
```

Calculate the great circle distance between two lat, long points.

### Parameters

- **a, b** (`float` | `numpy.ndarray`) – Lat-Long points or arrays to calculate distance between. If only one array is specified, a full distance matrix (i.e. calculate a point-to-point distance for every combination of points) will be returned.
- **absolute** (`bool`, `False`) – Whether to return estimates of on-sphere distances [`True`], or simply return the central angle between the points.
- **degrees** (`bool`, `True`) – Whether lat-long coordinates are in degrees [`True`] or radians [`False`].
- **r** (`float`) – Earth radii for estimating absolute distances.
- **method** (`str`, `{'vicenty', 'cosines', 'haversine'}`) – Which method to use for great circle distance calculation. Defaults to the Vicenty formula.
- **dtype** (`numpy.dtype`) – Data type for distance arrays, to constrain memory management.
- **max\_memory\_fraction** (`float`) – Constraint to switch to calculating mean distances where `matrix=True` and the distance matrix requires greater than a specified fraction of total available physical memory.

```
pyrolite.util.spatial.piecewise(segment_ranges: list, segments=2,
                                output_fmt=<class 'numpy.float64'>)
```

Generator to provide values of quantizable parameters which define a grid, here used to split up queries from databases to reduce load.

### Parameters

- **segment\_ranges** (`list`) – List of segment ranges to create a grid from.
- **segments** (`int`) – Number of segments.
- **output\_fmt** – Function to call on the output.

```
pyrolite.util.spatial.spatiotemporal_split(segments=4, nan_lims=[nan, nan],
                                             **kwargs)
```

Creates spatiotemporal grid using piecewise function and arbitrary ranges for individual kw-parameters (e.g. `age=(0., 450.)`), and sequentially returns individual grid cell attributes.

### Parameters

- **segments** (`int`) – Number of segments.

- **nan\_lims** (`list` | `tuple`) – Specificaiton of NaN indexes for missing boundaries.

**Yields**

`dict` – Iteration through parameter sets for each cell of the grid.

`pyrolite.util.spatial.NSEW_2_bounds(cardinal, order=['minx', 'miny', 'maxx', 'maxy'])`

Translates cardinal points to xy points in the form of bounds. Useful for converting to the format required for WFS from REST style queries.

**Parameters**

- **cardinal** (`dict`) – Cardinaly-indexed point bounds.
- **order** (`list`) – List indicating order of returned x-y bound coordinates.

**Returns**

x-y indexed extent values in the specified order.

**Return type**

`list`

`pyrolite.util.spatial.levenshtein_distance(seq_one, seq_two)`

Compute the Levenshtein Distance between two sequences with comparable items. Adapted from Wiki pseudocode.

**Parameters**

**seq\_one, seq\_two** (`str` | `list`) – Sequences to compare.

**Return type**

`int`

## pyrolite.util.resampling

Utilities for (weighted) bootstrap resampling applied to geoscientific point-data.

`pyrolite.util.resampling.univariate_distance_matrix(a, b=None, distance_metric=None)`

Get a distance matrix for a single column or array of values (here used for ages).

**Parameters**

- **a, b** (`numpy.ndarray`) – Points or arrays to calculate distance between. If only one array is specified, a full distance matrix (i.e. calculate a point-to-point distance for every combination of points) will be returned.
- **distance\_metric** – Callable function  $f(a, b)$  from which to derive a distance metric.

**Returns**

2D distance matrix.

**Return type**

`numpy.ndarray`

```
pyrolite.util.resampling.get_spatiotemporal_resampling_weights(df, spa-
    tial_norm=1.8,
    tempo-
    ral_norm=38,
    lat-
    long_names=['Latitude',
    'Longi-
    tude'],
    age_name='Age',
    max_memory_fraction=0.25,
    normal-
    ized_weights=True,
    **kwargs)
```

Takes a dataframe with lat, long and age and returns a sampling weight for each sample which is essentially the inverse of the mean distance to other samples.

### Parameters

- **df** (`pandas.DataFrame`) – Dataframe to calculate weights for.
- **spatial\_norm** (`float`) – Normalising constant for spatial measures (1.8 arc degrees).
- **temporal\_norm** (`float`) – Normalising constant for temporal measures (38 Mya).
- **latlong\_names** (`list`) – List of column names referring to latitude and longitude.
- **age\_name** (`str`) – Column name corresponding to geological age or time.
- **max\_memory\_fraction** (`float`) – Constraint to switch to calculating mean distances where `matrix=True` and the distance matrix requires greater than a specified fraction of total available physical memory. This is passed on to `great_circle_distance()`.
- **normalized\_weights** (`bool`) – Whether to renormalise weights to unity.

### Returns

**weights** – Sampling weights.

### Return type

`numpy.ndarray`

### Notes

This function is equivalent to Eq(1) from Keller and Schone:

$$W_i \propto 1 / \sum_{j=1}^n \left( \frac{1}{((z_i - z_j)/a)^2 + 1} + \frac{1}{((t_i - t_j)/b)^2 + 1} \right)$$

```
pyrolite.util.resampling.add_age_noise(df, min_sigma=50, noise_level=1.0,
    age_name='Age',
    age_uncertainty_name='AgeUncertainty',
    min_age_name='MinAge',
    max_age_name='MaxAge')
```

Add gaussian noise to a series of geological ages based on specified uncertainties or age ranges.

#### Parameters

- **df** (`pandas.DataFrame`) – Dataframe with age data within which to look up the age name and add noise.
- **min\_sigma** (`float`) – Minimum uncertainty to be considered for adding age noise.
- **noise\_level** (`float`) – Scaling of the noise added to the ages. By default the uncertainties are unscaled, but where age uncertainties are specified and are the one standard deviation level this can be used to expand the range of noise added (e.g. to 2SD).
- **age\_name** (`str`) – Column name for absolute ages.
- **age\_uncertainty\_name** (`str`) – Name of the column specifying absolute age uncertainties.
- **min\_age\_name** (`str`) – Name of the column specifying minimum absolute ages (used where uncertainties are otherwise unspecified).
- **max\_age\_name** (`str`) – Name of the column specifying maximum absolute ages (used where uncertainties are otherwise unspecified).

#### Returns

**df** – Dataframe with noise-modified ages.

#### Return type

`pandas.DataFrame`

#### Notes

This modifies the dataframe which is input - be aware of this if using outside of the bootstrap resampling for which this was designed.



```
pyrolite.util.resampling.spatiotemporal_bootstrap_resample(df,
                                                            columns=None,
                                                            uncert=None,
                                                            weights=None,
                                                            niter=100,
                                                            cate-
                                                            gories=None,
                                                            trans-
                                                            form=None,
                                                            boot-
                                                            strap_method='smooth',
                                                            add_gaussian_age_noise=True,
                                                            met-
                                                            rics=['mean',
                                                            'var'], de-
                                                            fault_uncertainty=0.02,
                                                            rela-
                                                            tive_uncertainties=True,
                                                            noise_level=1,
                                                            age_name='Age',
                                                            lat-
                                                            long_names=['Latitude',
                                                            'Longitude'],
                                                            **kwargs)
```

Resample and aggregate metrics from a dataframe, optionally aggregating by a given set of categories. Formulated specifically for dealing with resampling to address uneven sampling density in space and particularly geological time.

#### Parameters

- **df** (`pandas.DataFrame`) – Dataframe to resample.
- **columns** (`list`) – Columns to provide bootstrap resampled estimates for.
- **uncert** (`float` | `numpy.ndarray` | `pandas.Series` | `pandas.DataFrame`) – Fractional uncertainties for the dataset.
- **weights** (`numpy.ndarray` | `pandas.Series`) – Array of weights for resampling, if precomputed.
- **niter** (`int`) – Number of resampling iterations. This will be the minimum index size of the output metric dataframes.
- **categories** (`list` | `numpy.ndarray` | `pandas.Series`) – List of sample categories to group the outputs by, which has the same size as the dataframe index.
- **transform** – Callable function to transform input data prior to aggregation functions. Note that the outputs will need to be inverse-transformed.
- **bootstrap\_method** (`str`) – Which method to use to add gaussian noise to the input dataset parameters.
- **add\_gaussian\_age\_noise** (`bool`) – Whether to add gaussian noise to the input dataset ages, where present.
- **metrics** (`list`) – List of metrics to use for dataframe aggregation.

- **default\_uncertainty** (`float`) – Default (fractional) uncertainty where uncertainties are not given.
- **relative\_uncertainties** (`bool`) – Whether uncertainties are relative (True, i.e. fractional proportions of parameter values), or absolute (False)
- **noise\_level** (`float`) – Multiplier for the random gaussian noise added to the dataset and ages.
- **age\_name** (`str`) – Column name for geological age.
- **latlong\_names** (`list`) – Column names for latitude and longitude, or equivalent orthogonal spherical spatial measures.

**Returns**

Dictionary of aggregated Dataframe(s) indexed by statistical metrics. If categories are specified, the dataframe(s) will have a hierarchical index of categories, iteration.

**Return type**

`dict`

**pyrolite.util.missing**

`pyrolite.util.missing.md_pattern(Y)`

Get the missing data patterns from an array.

**Parameters**

**Y** (`numpy.ndarray` | `pandas.DataFrame`) – Input dataset.

**Returns**

- **pattern\_ids** (`numpy.ndarray`) – Pattern ID array.
- **pattern\_dict** (`dict`) – Dictionary of patterns indexed by pattern IDs. Contains a pattern and count for each pattern ID.

`pyrolite.util.missing.cooccurrence_pattern(Y, normalize=False, log=False)`

Get the co-occurrence patterns from an array.

**Parameters**

- **Y** (`numpy.ndarray` | `pandas.DataFrame`) – Input dataset.
- **normalize** (`bool`) – Whether to normalize the cooccurrence to compare disparate variables.
- **log** (`bool`) – Whether to take the log of the cooccurrence.

**Returns**

**co\_occur** – Cooccurrence frequency array.

**Return type**

`numpy.ndarray`

## pyrolite.util.units

`pyrolite.util.units.scale(in_unit, target_unit='ppm')`

Provides the scale difference between two mass units.

### Parameters

- **in\_unit** (`str`) – Units to be converted from
- **target\_unit** (`str`, "ppm") – Units to scale to.

---

### Todo:

- Implement different inputs: `str`, `list`, `pandas.Series`
- 

### Return type

`float`

## pyrolite.util.types

`pyrolite.util.types.iscollection(obj)`

Checks whether an object is an iterable collection.

### Parameters

**obj** (`object`) – Object to check.

### Returns

Boolean indication of whether the object is a collection.

### Return type

`bool`

## pyrolite.util.meta

`pyrolite.util.meta.get_module_datafolder(module='pyrolite', subfolder=None)`

Returns the path of a module data folder.

### Parameters

**subfolder** (`str`) – Subfolder within the module data folder.

### Return type

`pathlib.Path`

`pyrolite.util.meta.pyrolite_datafolder(subfolder=None)`

Returns the path of the pyrolite data folder.

### Parameters

**subfolder** (`str`) – Subfolder within the pyrolite data folder.

### Return type

`pathlib.Path`

`pyrolite.util.meta.take_me_to_the_docs()`

Opens the pyrolite documentation in a webbrowser.

`pyrolite.util.meta.sphinx_doi_link(doi)`

Generate a string with a restructured text link to a given DOI.

**Parameters**

**doi** (`str`)

**Returns**

String with doi link.

**Return type**

`str`

`pyrolite.util.meta.subkwargs(kwargs, *f)`

Get a subset of keyword arguments which are accepted by a function.

**Parameters**

- **kwargs** (`dict`) – Dictionary of keyword arguments.
- **f** (callable) – Function(s) to check.

**Returns**

Dictionary containing only relevant keyword arguments.

**Return type**

`dict`

`pyrolite.util.meta.inargs(name, *funcs)`

Check if an argument is a possible input for a specific function.

**Parameters**

- **name** (`str`) – Argument name.
- **f** (callable) – Function(s) to check.

**Return type**

`bool`

`pyrolite.util.meta.numpydoc_str_param_list(iterable, indent=4)`

Format a list of numpydoc parameters.

**Parameters**

- **iterable** (`list`) – List of numpydoc parameters.
- **indent** (`int`) – Indent as number of spaces.

**Return type**

`str`

`pyrolite.util.meta.get_additional_params(*fs, t='Parameters', header='',  
indent=4, subsections=False,  
subsection_delim='Note')`

Checks the base Parameters section of docstrings to get 'Other Parameters' for a specific function. Designed to incorporate information on inherited or forwarded parameters.

**Parameters**

- **fs** (`list`) – List of functions.
- **t** (`str`) – Target block of docstrings.
- **header** (`str`) – Optional section header.

- **indent** (`int` | `str`) – Indent as number of spaces, or a string of a given length.
- **subsections** (`bool`, `False`) – Whether to include headers specific for each function, creating subsections.
- **subsection\_delim** (`str`) – Subsection delimiter.

**Return type**`str`

---

**Todo:**

- Add delimiters between functions to show where arguments should be passed.
- 

`pyrolite.util.meta.update_docstring_references(obj, ref='ref')`

Updates docstring reference names to strings including the function name. Decorator will return the same function with a modified docstring. Sphinx likes unique names - specifically for citations, not so much for footnotes.

**Parameters**

- **obj** (`func` | `class`) – Class or function for which to update documentation references.
- **ref** (`str`) – String to replace with the object name.

**Returns**

Object with modified docstring.

**Return type**`func` | `class`

## pyrolite.util.skf

Utilities for use with `sklearn`.

## pyrolite.util.skf.vis

```
pyrolite.util.skf.vis.plot_confusion_matrix(*args, ax=None, classes=[],
                                             class_order=None,
                                             normalize=False,
                                             title='Confusion Matrix',
                                             cmap=<matplotlib.colors.LinearSegmentedColormap
                                             object>, norm=None,
                                             xlabelrotation=None)
```

This function prints and plots the confusion matrix. Normalization can be applied by setting `normalize=True`.

**Parameters**

- **args** (`tuple`) –
  - Data to evaluate and visualise a confusion matrix:
  - A single confusion matrix (n x n)
  - A tuple of (y\_test, y\_predict)

- A tuple of (classifier\_model, X\_test, y\_test)
- **ax** (`matplotlib.axes.Axes`) – Axis to plot on, if one exists.
- **classes** (`list`) – List of class names to use as labels, and for ordering (see below). This should match the order contained within the model. Where a classifier model is passed, the classes will be directly extracted.
- **class\_order** (`list`) – List of classes in the desired order along the axes. Should match the supplied classes where classes are given, or integer indices for where no named classes are given.
- **normalize** (`bool`) – Whether to normalize the counts for the confusion matrix to the sum of all cases (i.e. be between 0 and 1).
- **title** (`str`) – Title for the axes.
- **cmap** (`str` | `matplotlib.colors.Colormap`) – Colormap for the visualisation of the confusion matrix.
- **norm** (`bool`) – Normalization for the colormap visualisation across the confusion matrix.
- **xlabelrotation** (`float`) – Rotation in degrees for the xaxis labels.

**Returns**

**ax**

**Return type**

`matplotlib.axes.Axes`

`pyrolite.util.sklearn.vis.plot_gs_results(gs, xvar=None, yvar=None)`

Plots the results from a GridSearch showing location of optimum in 2D.

`pyrolite.util.sklearn.vis.alphas_from_multiclass_prob(probs, method='entropy',  
alpha=1.0)`

Take an array of multiclass probabilities and map to an alpha variable.

**Parameters**

- **probs** (`numpy.ndarray`) – Multiclass probabilities with shape (nsamples, nclasses).
- **method** (`str`, `entropy` | `kl_div`) – Method for mapping probabilities to alphas.
- **alpha** (`float`) – Optional specification of overall maximum alpha value.

**Returns**

**a** – Alpha values for each sample with shape (nsamples, 1).

**Return type**

`numpy.ndarray`

`pyrolite.util.sklearn.vis.plot_mapping(X, Y, mapping=None, ax=None, cmap=None,  
alpha=1.0, s=10, alpha_method='entropy',  
**kwargs)`

**Parameters**

- **X** (`numpy.ndarray`) – Coordinates in multidimensional space.

- **Y** (`numpy.ndarray` | `sklearn.base.BaseEstimator`) – An array of targets, or a method to obtain such an array of targets via `Y.predict()`. Transformers with probabilistic output (via `Y.predict_proba()`) will have these probability estimates accounted for via the alpha channel.
- **mapping** (`numpy.ndarray` | `TransformerMixin`) – Mapped points or transformer to create mapped points.
- **ax** (`matplotlib.axes.Axes`) – Axes to plot on.
- **cmap** (`matplotlib.cm.ListedColormap`) – Colormap to use for the classification visualisation (ideally this should be a discrete colormap unless the classes are organised).
- **alpha** (`float`) – Coefficient for alpha.
- **alpha\_method** ('entropy' or 'kl\_div') – Method to map class probabilities to alpha. 'entropy' uses a measure of entropy relative to null-scenario of equal distribution across classes, while 'kl\_div' calculates the information gain relative to the same null-scenario.

#### Returns

- **ax** (`Axes`) – Axes on which the mapping is plotted.
- **tfm** (`BaseEstimator`) – Fitted mapping transform.

---

#### Todo:

- Option to generate colors for individual classes

This could be based on the distances between their centres in multidimensional space (or low dimensional mapping of this space), enabling a continuous (n-dimensional) colormap to be used to show similar classes, in addition to classification confidence.

---

## pyrolite.util.sklearn.pipeline

```
pyrolite.util.sklearn.pipeline.fit_save_classifier(clf, X_train, y_train,  
                                                  directory='.', name='clf',  
                                                  extension='.joblib')
```

Fit and save a classifier model. Also save relevant metadata where possible.

#### Parameters

- **clf** (`sklearn.base.BaseEstimator`) – Classifier or gridsearch.
- **X\_train** (`numpy.ndarray` | `pandas.DataFrame`) – Training data.
- **y\_train** (`numpy.ndarray` | `pandas.Series`) – Training true classes.
- **directory** (`str` | `pathlib.Path`) – Path to the save directory.
- **name** (`str`) – Name of the classifier.
- **extension** (`str`) – Extension to give the saved classifier pickled with joblib.

**Returns**

**clf** – Fitted classifier.

**Return type**

`sklearn.base.BaseEstimator`

```
pyrolite.util.sklearn.pipeline.classifier_performance_report(clf, X_test, y_test,  
                                                            classes=[],  
                                                            directory='.',  
                                                            name='clf')
```

Output a performance report for a classifier. Currently outputs the overall classification score, a confusion matrix and where relevant an indication of variation seen across the gridsearch (currently only possible for 2D searches).

**Parameters**

- **clf** (`sklearn.base.BaseEstimator` | `sklearn.model_selection.GridSearchCV`) – Classifier or grid-search.
- **X\_test** (`numpy.ndarray` | `pandas.DataFrame`) – Input data for testing.
- **y\_test** (`numpy.ndarray` | `pandas.Series`) – Labelled/target data for testing.
- **classes** (*list*) – Names of classes. **directory** : `str` | `pathlib.Path`  
Path to the save directory.
- **name** (`str`) – Name of the classifier.

**Returns**

**clf** – Fitted classifier.

**Return type**

`sklearn.base.BaseEstimator`

```
pyrolite.util.sklearn.pipeline.SVC_pipeline(sampler=None, balance=True,  
                                             transform=None, scaler=None,  
                                             kernel='rbf',  
                                             decision_function_shape='ovo',  
                                             probability=False,  
                                             cv=StratifiedKFold(n_splits=10,  
                                             random_state=None, shuffle=True),  
                                             param_grid={}, n_jobs=4,  
                                             verbose=10, cache_size=500,  
                                             **kwargs)
```

A convenience function for constructing a Support Vector Classifier pipeline.

**Parameters**

- **sampler** (`sklearn.base.TransformerMixin`) – Resampling transformer.
- **balance** (`bool`) – Whether to balance the class weights for the classifier.
- **transform** (`sklearn.base.TransformerMixin`) – Preprocessing transformer.



- **scaler** (`sklearn.base.TransformerMixin`) – Scale transformer.
- **kernel** (`str` | callable) – Name of kernel to use for the support vector classifier ('linear'|'rbf'|'poly'|'sigmoid'). Optionally, a custom kernel function can be supplied (see `sklearn` docs for more info).
- **decision\_function\_shape** (`str`, 'ovo' or 'ovr') – Shape of the decision function surface. 'ovo' one-vs-one classifier of `libsvm` (returning classification of shape (samples, classes\*(classes-1)/2)), or the default 'ovr' one-vs-rest classifier which will return classification estimation shape of `(samples, classes)`.
- **probability** (`bool`) – Whether to implement Platt-scaling to enable probability estimates. This must be enabled prior to calling fit, and will slow down that method.
- **cv** (`int` | `sklearn.model_selection.BaseSearchCV`) – Cross validation search. If an integer `k` is provided, results in default `k`-fold cross validation. Optionally, if a `sklearn.model_selection.BaseSearchCV` instance is provided, it will be used directly (enabling finer control, e.g. over sorting/shuffling etc).
- **param\_grid** (`dict`) – Dictionary representing a parameter grid for the support vector classifier. Typically contains 1D arrays of grid indices for `SVC()` parameters each prefixed with `svc__` (e.g. `dict(svc__gamma=np.logspace(-1, 3, 5), svc__C=np.logspace(-0.5, 2, 5))`).
- **n\_jobs** (`int`) – Number of processors to use for the `SVC` construction. Note that providing `n_jobs = -1` will use all available processors.
- **verbose** (`int`) – Level of verbosity for the pipeline logging output.
- **cache\_size** (`float`) – Specify the size of the kernel cache (in MB).

---

**Note:** The following additional parameters are from `sklearn.svm._classes.SVC()`.

---

#### Other Parameters

- **C** (`float`, *default=1.0*) – Regularization parameter. The strength of the regularization is inversely proportional to `C`. Must be strictly positive. The penalty is a squared  $l_2$  penalty.
- **degree** (`int`, *default=3*) – Degree of the polynomial kernel function ('poly'). Must be non-negative. Ignored by all other kernels.
- **gamma** (`{'scale', 'auto'}` or `float`, *default='scale'*) – Kernel coefficient for 'rbf', 'poly' and 'sigmoid'.
  - if `gamma='scale'` (default) is passed then it uses `1 / (n_features * X.var())` as value of `gamma`,

- if ‘auto’, uses  $1 / n\_features$
- if float, must be non-negative.

Changed in version 0.22: The default value of `gamma` changed from ‘auto’ to ‘scale’.

- **coef0** (*float, default=0.0*) – Independent term in kernel function. It is only significant in ‘poly’ and ‘sigmoid’.
- **shrinking** (*bool, default=True*) – Whether to use the shrinking heuristic. See the [User Guide](#).
- **tol** (*float, default=1e-3*) – Tolerance for stopping criterion.
- **cache\_size** (*float, default=200*) – Specify the size of the kernel cache (in MB).
- **class\_weight** (*dict or ‘balanced’, default=None*) – Set the parameter  $C$  of class  $i$  to `class_weight[i]*C` for SVC. If not given, all classes are supposed to have weight one. The “balanced” mode uses the values of  $y$  to automatically adjust weights inversely proportional to class frequencies in the input data as `n_samples / (n_classes * np.bincount(y))`.
- **max\_iter** (*int, default=-1*) – Hard limit on iterations within solver, or -1 for no limit.
- **break\_ties** (*bool, default=False*) – If true, `decision_function_shape='ovr'`, and number of classes  $> 2$ , `predict` will break ties according to the confidence values of `decision_function`; otherwise the first class among the tied classes is returned. Please note that breaking ties comes at a relatively high computational cost compared to a simple `predict`.

New in version 0.22.

- **random\_state** (*int, RandomState instance or None, default=None*) – Controls the pseudo random number generation for shuffling the data for probability estimates. Ignored when `probability` is False. Pass an int for reproducible output across multiple function calls. See [Glossary](#).

#### Returns

**gs** – Gridsearch object containing the results of the SVC training across the parameter grid. Access the best estimator with `gs.best_estimator_` and its parameters with `gs.best_params_`.

#### Return type

`sklearn.model_selection.GridSearchCV`

```
class pyrolite.util.sklearn.pipeline.PdUnion(estimators: list = [])
```

```
fit(X, y=None)
```

```
transform(X)
```

**pyrolite.util.sklearn.select**

```
class pyrolite.util.sklearn.select.TypeSelector(dtype)
    fit(X, y=None)
    transform(X)

class pyrolite.util.sklearn.select.ColumnSelector(columns)
    fit(X, y=None)
    transform(X)

class pyrolite.util.sklearn.select.CompositionalSelector(columns=None,
                                                         inverse=False)
    fit(X, y=None)
    transform(X)

class pyrolite.util.sklearn.select.MajorsSelector(components=None)
    fit(X, y=None)
    transform(X)

class pyrolite.util.sklearn.select.ElementSelector(components=None)
    fit(X, y=None)
    transform(X)

class pyrolite.util.sklearn.select.REESelector(components=None)
    fit(X, y=None)
    transform(X)
```

**pyrolite.util.sklearn.transform**

```
class pyrolite.util.sklearn.transform.DropBelowZero(**kwargs)
    transform(X, *args, **kwargs)
    fit(X, *args)

class pyrolite.util.sklearn.transform.LinearTransform(**kwargs)
    transform(X, *args, **kwargs)
    inverse_transform(Y, *args, **kwargs)
    fit(X, *args)

class pyrolite.util.sklearn.transform.ExpTransform(**kwargs)
    transform(X, *args, **kwargs)
```

```
    inverse_transform(Y, *args, **kwargs)

    fit(X, *args)

class pyrolite.util.sklearn.transform.LogTransform(fmt_string='Ln({}),
                                                    **kwargs)

    transform(X, *args, **kwargs)

    inverse_transform(Y, *args, **kwargs)

    fit(X, *args)

class pyrolite.util.sklearn.transform.ALRTTransform(label_mode='numeric',
                                                    **kwargs)

    transform(X, *args, **kwargs)

    inverse_transform(Y, *args, **kwargs)

    fit(X, *args, **kwargs)

class pyrolite.util.sklearn.transform.CLRTransform(label_mode='numeric',
                                                    **kwargs)

    transform(X, *args, **kwargs)

    inverse_transform(Y, *args, **kwargs)

    fit(X, *args, **kwargs)

class pyrolite.util.sklearn.transform.ILRTTransform(label_mode='numeric',
                                                    **kwargs)

    transform(X, *args, **kwargs)

    inverse_transform(Y, *args, **kwargs)

    fit(X, *args, **kwargs)

class pyrolite.util.sklearn.transform.SphericalCoordTransform(**kwargs)

    transform(X, *args, **kwargs)

    inverse_transform(Y, *args, **kwargs)

    fit(X, *args, **kwargs)

class pyrolite.util.sklearn.transform.BoxCoxTransform(**kwargs)

    transform(X, *args, **kwargs)

    inverse_transform(Y, *args, **kwargs)

    fit(X, *args, **kwargs)

class pyrolite.util.sklearn.transform.Devolatilizer(exclude=['H2O',
                                                            'H2O_PLUS',
                                                            'H2O_MINUS', 'CO2',
                                                            'LOI'], renorm=True)
```

```
fit(X, y=None)

transform(X)

class pyrolite.util.sklearn.transform.ElementAggregator(renorm=True,
                                                         form='oxide')

fit(X, y=None)

transform(X)

class pyrolite.util.sklearn.transform.LambdaTransformer(norm_to='Chondrite_PON',
                                                         exclude=['Pm', 'Eu',
                                                         'Ce'], params=None,
                                                         degree=5)

fit(X, y=None)

transform(X)
```

### pyrolite.util.sklearn.impute

```
class pyrolite.util.sklearn.impute.MultipleImputer(multiple=5, max_iter=10,
                                                    groupby=None, *args,
                                                    **kwargs)
```

```
transform(X, *args, **kwargs)
```

```
fit(X, y=None)
```

Fit the imputers.

#### Parameters

- **X** (`pandas.DataFrame`) – Data to use to fit the imputations.
- **y** (`pandas.Series`) – Target class; optionally specified, and used similarly to `groupby`.

### pyrolite.util.classification

Utilities for rock chemistry and mineral abundance classification.

---

#### Todo:

- Petrological classifiers: QAPF (aphanitic/phaneritic), gabbroic Pyroxene-Olivine-Plagioclase, ultramafic Olivine-Orthopyroxene-Clinopyroxene
- 

```
class pyrolite.util.classification.PolygonClassifier(name=None,
                                                         axes=None,
                                                         fields=None,
                                                         scale=1.0,
                                                         transform=None,
                                                         mode=None,
                                                         **kwargs)
```

A classifier model built from a series of polygons defining specific classes.

**Parameters**

- **name** (`str`) – A name for the classifier model.
- **axes** (`dict`) – Mapping from plot axes to variables to be used for labels.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing 'name' and 'fields' items.
- **scale** (`float`) – Default maximum scale for the axes. Typically 100 (wt%) or 1 (fractional).
- **xlim** (`tuple`) – Default x-limits for this classifier for plotting.
- **ylim** (`tuple`) – Default y-limits for this classifier for plotting.

**predict**(*X*, *data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

**Parameters**

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

**Returns**

Series containing classifier predictions. If a dataframe was input, it inherit the index.

**Return type**

`pandas.Series`

**property axis\_components**

Get the axis components used by the classifier.

**Returns**

Ordered names for axes used by the classifier.

**Return type**

`tuple`

**add\_to\_axes**(*ax=None*, *fill=False*, *axes\_scale=1.0*, *add\_labels=False*,  
*which\_labels='ID'*, *which\_ids=[]*, *\*\*kwargs*)

Add the fields from the classifier to an axis.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field 'name' or 'ID'.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, ['F', 'T1'] to plot

the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

#### Returns

**ax**

#### Return type

`matplotlib.axes.Axes`

**class** `pyrolite.util.classification.TAS`(*which\_model=None*, *\*\*kwargs*)

Total-alkali Silica Diagram classifier from Middlemost (1994)<sup>1</sup>, a closed-polygon variant after Le Bas et al. (1992)<sup>2</sup>.

#### Parameters

- **name** (`str`) – A name for the classifier model.
- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing ‘name’ and ‘fields’ items.
- **scale** (`float`) – Default maximum scale for the axes. Typically 100 (wt%) or 1 (fractional).
- **xlim** (`tuple`) – Default x-limits for this classifier for plotting.
- **ylim** (`tuple`) – Default y-limits for this classifier for plotting.
- **which\_model** (`str`) – The name of the model variant to use, if not Middlemost.

## References

**add\_to\_axes**(*ax=None*, *fill=False*, *axes\_scale=100.0*, *add\_labels=False*, *which\_labels='ID'*, *which\_ids=[]*, *label\_at\_centroid=True*, *\*\*kwargs*)

Add the TAS fields from the classifier to an axis.

#### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which labels to add to the polygons (e.g. for TAS, ‘volcanic’, ‘intrusive’ or the field ‘ID’).
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, [‘F’, ‘T1’] to plot

<sup>1</sup> Middlemost, E. A. K. (1994). Naming materials in the magma/igneous rock system. *Earth-Science Reviews*, 37(3), 215–224. doi: 10.1016/0012-8252(94)90029-9

<sup>2</sup> Le Bas, M.J., Le Maitre, R.W., Woolley, A.R. (1992). The construction of the Total Alkali-Silica chemical classification of volcanic rocks. *Mineralogy and Petrology* 46, 1–22. doi: 10.1007/BF01160698

the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

- **label\_at\_centroid** (`bool`) – Whether to label the fields at the centroid (True) or at the visual center of the field (False).

**Returns**

`ax`

**Return type**

`matplotlib.axes.Axes`

**property axis\_components**

Get the axis components used by the classifier.

**Returns**

Ordered names for axes used by the classifier.

**Return type**

`tuple`

**predict**(*X*, *data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

**Parameters**

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

**Returns**

Series containing classifier predictions. If a dataframe was input, it inherit the index.

**Return type**

`pandas.Series`

**class** `pyrolite.util.classification.USDAsoilTexture`(\*\**kwargs*)

United States Department of Agriculture Soil Texture classification model<sup>45</sup>.

**Parameters**

- **name** (`str`) – A name for the classifier model.
- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing ‘name’ and ‘fields’ items.

---

<sup>4</sup> Soil Science Division Staff (2017). Soil Survey Manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

<sup>5</sup> Thien, Steve J. (1979). A Flow Diagram for Teaching Texture-by-Feel Analysis. Journal of Agronomic Education 8:54–55. doi: 10.2134/jae.1979.0054



## References

**add\_to\_axes**(*ax=None, fill=False, axes\_scale=1.0, add\_labels=False, which\_labels='ID', which\_ids=[], \*\*kwargs*)

Add the fields from the classifier to an axis.

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field 'name' or 'ID'.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, ['F', 'T1'] to plot the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

### Returns

**ax**

### Return type

`matplotlib.axes.Axes`

**property axis\_components**

Get the axis components used by the classifier.

### Returns

Ordered names for axes used by the classifier.

### Return type

`tuple`

**predict**(*X, data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

### Parameters

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

### Returns

Series containing classifier predictions. If a dataframe was input, it inherit the index.

### Return type

`pandas.Series`

**class** `pyrolite.util.classification.QAP`(*\*\*kwargs*)

IUGS QAP ternary classification<sup>67</sup>.

<sup>6</sup> Streckeisen, A. (1974). Classification and nomenclature of plutonic rocks: recommendations of the IUGS subcommission on the systematics of Igneous Rocks. *Geol Rundsch* 63, 773–786. doi: 10.1007/BF01820841

<sup>7</sup> Le Maitre, R.W. (2002). *Igneous Rocks: A Classification and Glossary of Terms : Recommendations of International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks*. Cambridge University Press, 236pp doi: 10.1017/CBO9780511535581

**Parameters**

- **name** (`str`) – A name for the classifier model.
- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing ‘name’ and ‘fields’ items.

**References**

**add\_to\_axes**(*ax=None, fill=False, axes\_scale=1.0, add\_labels=False, which\_labels='ID', which\_ids=[], \*\*kwargs*)

Add the fields from the classifier to an axis.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field ‘name’ or ‘ID’.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, [‘F’, ‘T1’] to plot the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

**Returns**

`ax`

**Return type**

`matplotlib.axes.Axes`

**property axis\_components**

Get the axis components used by the classifier.

**Returns**

Ordered names for axes used by the classifier.

**Return type**

`tuple`

**predict**(*X, data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

**Parameters**

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

**Returns**

Series containing classifier predictions. If a dataframe was input, it inherit the index.

**Return type**

`pandas.Series`

**class** `pyrolite.util.classification.FeldsparTernary`(\*\**kwargs*)

Simplified feldspar diagram classifier, based on a version printed in the second edition of ‘An Introduction to the Rock Forming Minerals’ (Deer, Howie and Zussman).

**Parameters**

- **name** (`str`) – A name for the classifier model.
- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing ‘name’ and ‘fields’ items.
- **mode** (`str`) – Mode of the diagram to use; two are currently available - ‘default’, which fills the entire ternary space, and ‘miscibility-gap’ which gives a simplified approximation of the miscibility gap.

**References**

**add\_to\_axes**(*ax=None*, *fill=False*, *axes\_scale=1.0*, *add\_labels=False*, *which\_labels='ID'*, *which\_ids=[]*, \*\**kwargs*)

Add the fields from the classifier to an axis.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field ‘name’ or ‘ID’.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, [‘F’, ‘T1’] to plot the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

**Returns**

`ax`

**Return type**

`matplotlib.axes.Axes`

**property** `axis_components`

Get the axis components used by the classifier.

**Returns**

Ordered names for axes used by the classifier.

**Return type**`tuple`**predict**(*X*, *data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

**Parameters**

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

**Returns**

Series containing classifier predictions. If a dataframe was input, it inherit the index.

**Return type**`pandas.Series`**class** `pyrolite.util.classification.PeralkalinityClassifier`**predict**(*df: DataFrame*)**class** `pyrolite.util.classification.JensenPlot`(*\*\*kwargs*)

Jensen Plot for classification of subalkaline volcanic rocks<sup>9</sup>.

**Parameters**

- **name** (`str`) – A name for the classifier model.
- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing ‘name’ and ‘fields’ items.

**References****Notes**

Diagram used for the classification of subalkalic volcanic rocks. The diagram is constructed for molar cation percentages of Al, Fe+Ti and Mg, on account of these elements’ stability upon metamorphism. This particular version uses updated labels relative to Jensen (1976), in which the fields have been extended to the full range of the ternary plot.

**add\_to\_axes**(*ax=None*, *fill=False*, *axes\_scale=1.0*, *add\_labels=False*, *which\_labels='ID'*, *which\_ids=[]*, *\*\*kwargs*)

Add the fields from the classifier to an axis.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).

---

<sup>9</sup> Jensen, L. S. (1976). A new cation plot for classifying sub-alkaline volcanic rocks. Ontario Division of Mines. Miscellaneous Paper No. 66.

- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field ‘name’ or ‘ID’.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, [‘F’, ‘T1’] to plot the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

**Returns**`ax`**Return type**`matplotlib.axes.Axes`**property axis\_components**

Get the axis components used by the classifier.

**Returns**

Ordered names for axes used by the classifier.

**Return type**`tuple`**predict**(*X*, *data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

**Parameters**

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

**Returns**

Series containing classifier predictions. If a dataframe was input, it inherit the index.

**Return type**`pandas.Series`**class** `pyrolite.util.classification.SpinelTrivalentTernary`(*\*\*kwargs*)

Spinel Trivalent Ternary classification - designed for data in atoms per formula unit

**Parameters**

- **name** (`str`) – A name for the classifier model.
- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing ‘name’ and ‘fields’ items.

**add\_to\_axes**(*ax=None*, *fill=False*, *axes\_scale=1.0*, *add\_labels=False*,  
*which\_labels='ID'*, *which\_ids=[]*, *\*\*kwargs*)

Add the fields from the classifier to an axis.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.

- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field ‘name’ or ‘ID’.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, [‘F’, ‘T1’] to plot the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

**Returns**`ax`**Return type**`matplotlib.axes.Axes`**property axis\_components**

Get the axis components used by the classifier.

**Returns**

Ordered names for axes used by the classifier.

**Return type**`tuple`**predict**(*X*, *data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

**Parameters**

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

**Returns**

Series containing classifier predictions. If a dataframe was input, it inherit the index.

**Return type**`pandas.Series`**class** `pyrolite.util.classification.SpinelFeBivariate`(\*\**kwargs*)

Fe-Spinel classification, designed for data in atoms per formula unit.

**Parameters**

- **name** (`str`) – A name for the classifier model.
- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing ‘name’ and ‘fields’ items.

**add\_to\_axes**(*ax=None, fill=False, axes\_scale=1.0, add\_labels=False, which\_labels='ID', which\_ids=[], \*\*kwargs*)

Add the fields from the classifier to an axis.

#### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field 'name' or 'ID'.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, ['F', 'T1'] to plot the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

#### Returns

**ax**

#### Return type

`matplotlib.axes.Axes`

#### **property axis\_components**

Get the axis components used by the classifier.

#### Returns

Ordered names for axes used by the classifier.

#### Return type

`tuple`

#### **predict**(*X, data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

#### Parameters

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

#### Returns

Series containing classifier predictions. If a dataframe was input, it inherit the index.

#### Return type

`pandas.Series`

**class** `pyrolite.util.classification.Pettijohn`(*\*\*kwargs*)

Pettijohn (1973) sandstones classification<sup>10</sup>.

#### Parameters

- **name** (`str`) – A name for the classifier model.

<sup>10</sup> Pettijohn, F. J., Potter, P. E. and Siever, R. (1973). Sand and Sandstone. New York, Springer-Verlag. 618p. doi: [10.1007/978-1-4615-9974-6](https://doi.org/10.1007/978-1-4615-9974-6)

- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing 'name' and 'fields' items.

## References

**add\_to\_axes**(*ax=None, fill=False, axes\_scale=1.0, add\_labels=False, which\_labels='ID', which\_ids=[], \*\*kwargs*)

Add the fields from the classifier to an axis.

### Parameters

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field 'name' or 'ID'.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, ['F', 'T1'] to plot the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

### Returns

**ax**

### Return type

`matplotlib.axes.Axes`

## **property axis\_components**

Get the axis components used by the classifier.

### Returns

Ordered names for axes used by the classifier.

### Return type

`tuple`

## **predict**(*X, data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

### Parameters

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

### Returns

Series containing classifier predictions. If a dataframe was input, it inherit the index.



**Return type**`pandas.Series`**class** `pyrolite.util.classification.Herron`(\*\*kwargs)Herron (1988) sandstones classification<sup>11</sup>.**Parameters**

- **name** (`str`) – A name for the classifier model.
- **axes** (`list` | `tuple`) – Names of the axes corresponding to the polygon coordinates.
- **fields** (`dict`) – Dictionary describing individual polygons, with identifiers as keys and dictionaries containing ‘name’ and ‘fields’ items.

**References****add\_to\_axes**(*ax=None, fill=False, axes\_scale=1.0, add\_labels=False, which\_labels='ID', which\_ids=[], \*\*kwargs*)

Add the fields from the classifier to an axis.

**Parameters**

- **ax** (`matplotlib.axes.Axes`) – Axis to add the polygons to.
- **fill** (`bool`) – Whether to fill the polygons.
- **axes\_scale** (`float`) – Maximum scale for the axes. Typically 100 (for wt%) or 1 (fractional).
- **add\_labels** (`bool`) – Whether to add labels for the polygons.
- **which\_labels** (`str`) – Which data to use for field labels - field ‘name’ or ‘ID’.
- **which\_ids** (`list`) – List of field IDs corresponding to the polygons to add to the axes object. (e.g. for TAS, [‘F’, ‘T1’] to plot the Foidite and Trachyte fields). An empty list corresponds to plotting all the polygons.

**Returns****ax****Return type**`matplotlib.axes.Axes`**property** `axis_components`

Get the axis components used by the classifier.

**Returns**

Ordered names for axes used by the classifier.

**Return type**`tuple`

---

<sup>11</sup> Herron, M.M. (1988). Geochemical classification of terrigenous sands and shales from core or log data. *Journal of Sedimentary Research*, 58(5), pp.820-829. doi: 10.1306/212F8E77-2B24-11D7-8648000102C1865D

**predict**(*X*, *data\_scale=None*)

Predict the classification of samples using the polygon-based classifier.

**Parameters**

- **X** (`numpy.ndarray` | `pandas.DataFrame`) – Data to classify.
- **data\_scale** (`float`) – Maximum scale for the data. Typically 100 (wt%) or 1 (fractional).

**Returns**

Series containing classifier predictions. If a dataframe was input, it inherit the index.

**Return type**

`pandas.Series`

## pyrolite.util.log

```
pyrolite.util.log.Handle(logger, handler_class=<class 'logging.StreamHandler'>,
                        formatter='%(%asctime)s %(name)s - %(levelname)s:
                        %(message)s', level=None)
```

Handle a logger with a standardised formatting.

**Parameters**

- **logger** (`logging.Logger` | `str`) – Logger or module name to source a logger from.
- **handler\_class** (`logging.Handler`) – Handler class for the logging messages.
- **formatter** (`str` | `logging.Formatter`) – Formatter for the logging handler. Strings will be passed to the `logging.Formatter` constructor.
- **level** (`str`) – Logging level for the handler.

**Returns**

Configured logger.

**Return type**

`logging.Logger`

```
class pyrolite.util.log.ToLogger(logger, level=None)
```

Output stream which will output to logger module instead of stdout.

```
buf = ''
```

```
logger = None
```

```
level = None
```

```
write(buf)
```

Write string to file.

Returns the number of characters written, which is always equal to the length of the string.

**flush()**

Flush write buffers, if applicable.

This is not implemented for read-only and non-blocking streams.

`pyrolite.util.log.stream_log(logger=None, level='INFO')`

Stream the log from a specific package or subpackage.

**Parameters**

- **logger** (`str` | `logging.Logger`) – Name of the logger or module to monitor logging from.
- **level** (`str`, 'INFO') – Logging level at which to set the handler output.

**Returns**

Logger for the specified package with stream handler added.

**Return type**

`logging.Logger`

See also:

[Extensions](#)

## 5.2 Data

This gallery describes the datasets used by pyrolite, how to access them, and their sources.

### 5.2.1 Mineral Database

pyrolite includes a limited mineral database which is useful for looking up endmember compositions. This part of the package is being actively developed, so expect expansions and improvements soon.

See also:

Examples:

[Mineral Database](#)

**Total running time of the script:** (0 minutes 0.000 seconds)

### 5.2.2 Geological Timescale

pyrolite includes a simple geological timescale, based on a recent version of the International Chronostratigraphic Chart<sup>1</sup>. The *Timescale* class can be used to look up names for specific geological ages, to look up times for known geological age names and to access a reference table for all of these.

```
from pyrolite.util.time import Timescale, age_name

ts = Timescale()

eg = ts.data.iloc[:, :5] # the first five columns of this data table
eg
```

<sup>1</sup> Cohen, K.M., Finney, S.C., Gibbard, P.L., Fan, J.-X., 2013. The ICS International Chronostratigraphic Chart. Episodes 36, 199–204.

## References

See also:

Examples:

[Timescale](#)

Modules, Classes and Functions:

`pyrolite.util.time`, `Timescale`

Total running time of the script: (0 minutes 0.059 seconds)

### 5.2.3 Ionic Radii

pyrolite includes a few sets of reference tables for ionic radii in angstroms (Å) from [?] and [?], each with tables indexed by element, ionic charge and coordination. The easiest way to access these is via the `get_ionic_radii()` function. The function can be used to get radii for individual elements, using a `source` keyword argument to swap between the datasets:

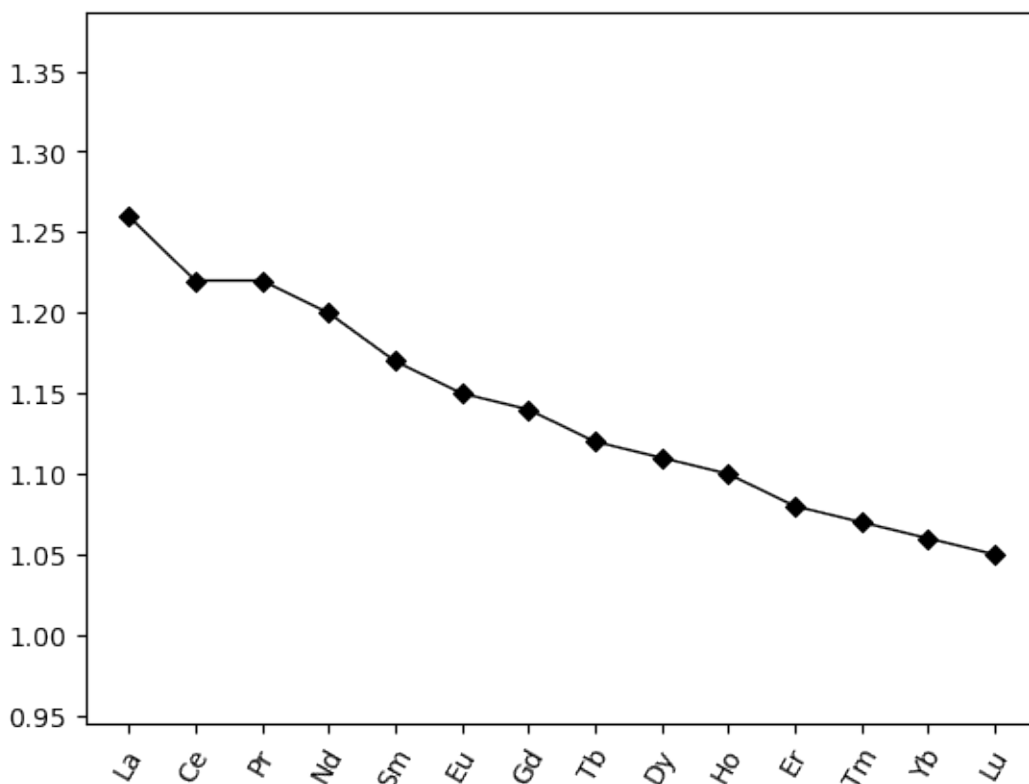
```
import matplotlib.pyplot as plt
import pandas as pd

from pyrolite.geochem.ind import REE, get_ionic_radii

REE_radii = pd.Series(
    get_ionic_radii(REE(), coordination=8, charge=3, source="Whittaker"), index=REE()
)
REE_radii
```

```
La      1.26
Ce      1.22
Pr      1.22
Nd      1.20
Sm      1.17
Eu      1.15
Gd      1.14
Tb      1.12
Dy      1.11
Ho      1.10
Er      1.08
Tm      1.07
Yb      1.06
Lu      1.05
dtype: float64
```

```
REE_radii.pyroplot.spider(color="k", logy=False)
plt.show()
```



## References

See also:

Examples:

[Ionic Radii](#), [lambdas: Parameterising REE Profiles](#), [REE Radii Plot](#)

Functions:

`get_ionic_radii()`, `pyrolite.geochem.ind.REE()`, `lambda_lnREE()`,

Total running time of the script: (0 minutes 0.233 seconds)

## 5.2.4 Aitchison Examples

pyrolite includes four synthetic datasets which are used in [?] which can be accessed using each of the respective functions `load_boxite()`, `load_coxite()`, `load_hongite()` and `load_kongite()` (all returning a `DataFrame`).

```
from pyrolite.data.Aitchison import load_boxite, load_coxite, load_hongite, load_kongite

df = load_boxite()
df.head()
```

```
import matplotlib.pyplot as plt

import pyrolite.plot
```

(continues on next page)

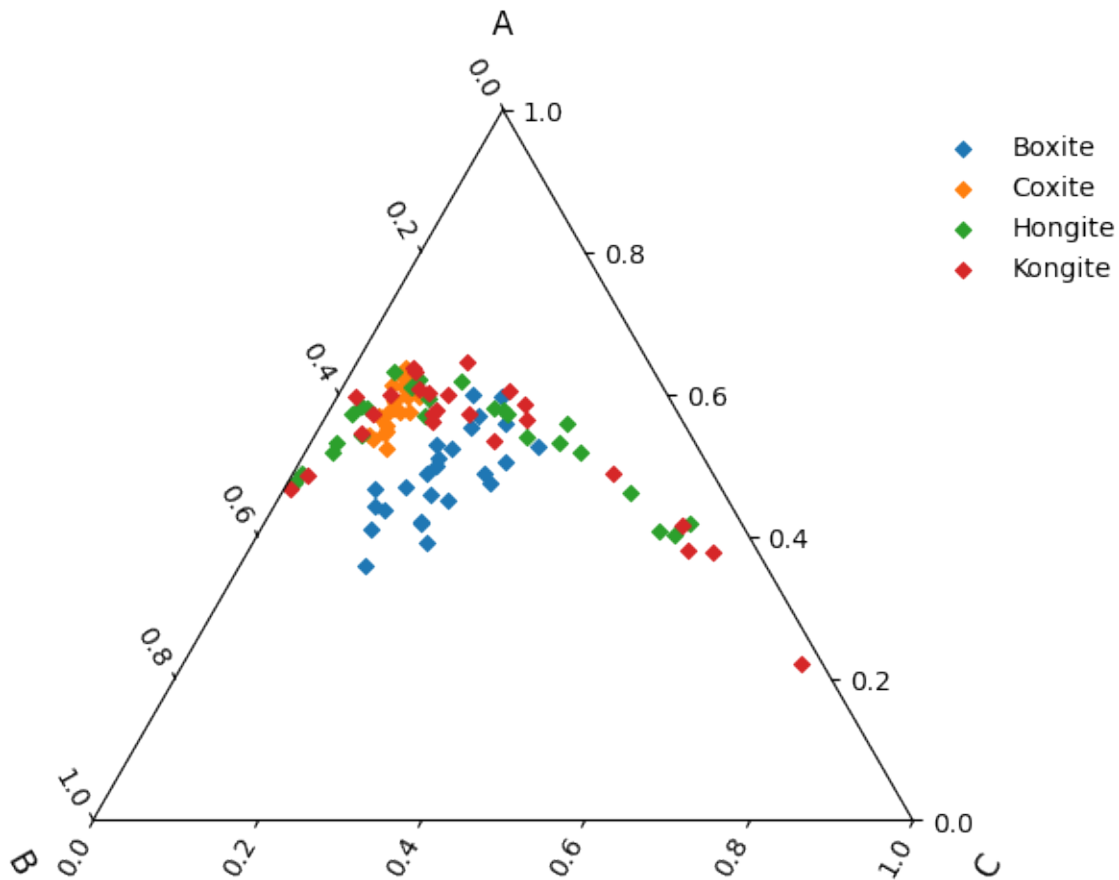
(continued from previous page)

```

fig, ax = plt.subplots(1)
for loader in [load_boxite, load_coxite, load_hongite, load_kongite]:
    df = loader()
    ax = df[["A", "B", "C"]].pyroplot.scatter(ax=ax, label=df.attrs["name"])

ax.legend()
plt.show()

```



## References

See also:

### Examples:

[Log Ratio Means](#), [Log Transforms](#), [Compositional Data](#), [Ternary Plots](#)

### Tutorials:

[Ternary Density Plots](#), [Making the Logo](#)

### Modules and Functions:

`pyrolite.comp.codata`, `renormalise()`

**Total running time of the script:** (0 minutes 3.281 seconds)

## 5.2.5 Reference Compositions

This page presents the range of compositions within the reference composition database accessible within `pyrolite`. It's currently a work in progress, but will soon contain extended descriptions and notes for some of the compositions and associated references.

```
import matplotlib.pyplot as plt

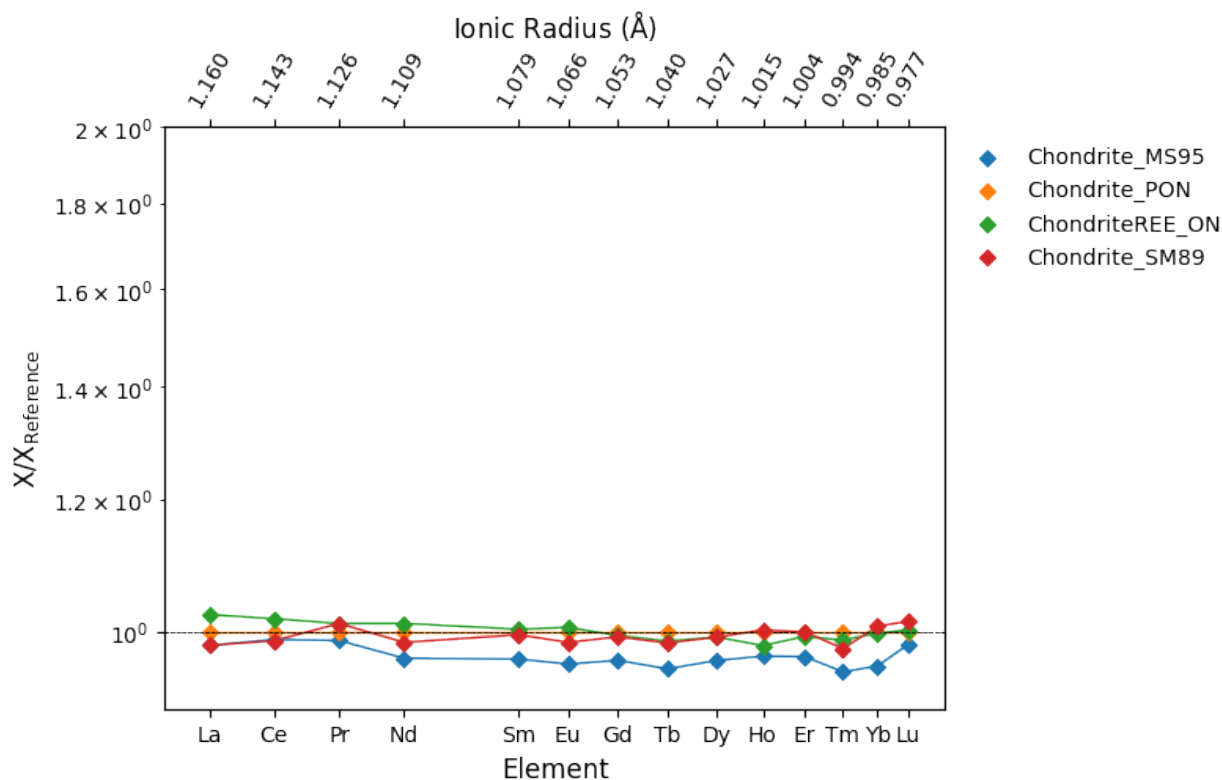
from pyrolite.geochem.norm import all_reference_compositions, get_reference_composition

refcomps = all_reference_compositions()
norm = "Chondrite_PON" # a constant composition to normalise to
```

### Chondrites

```
fltr = lambda c: c.reservoir == "Chondrite"
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()
```

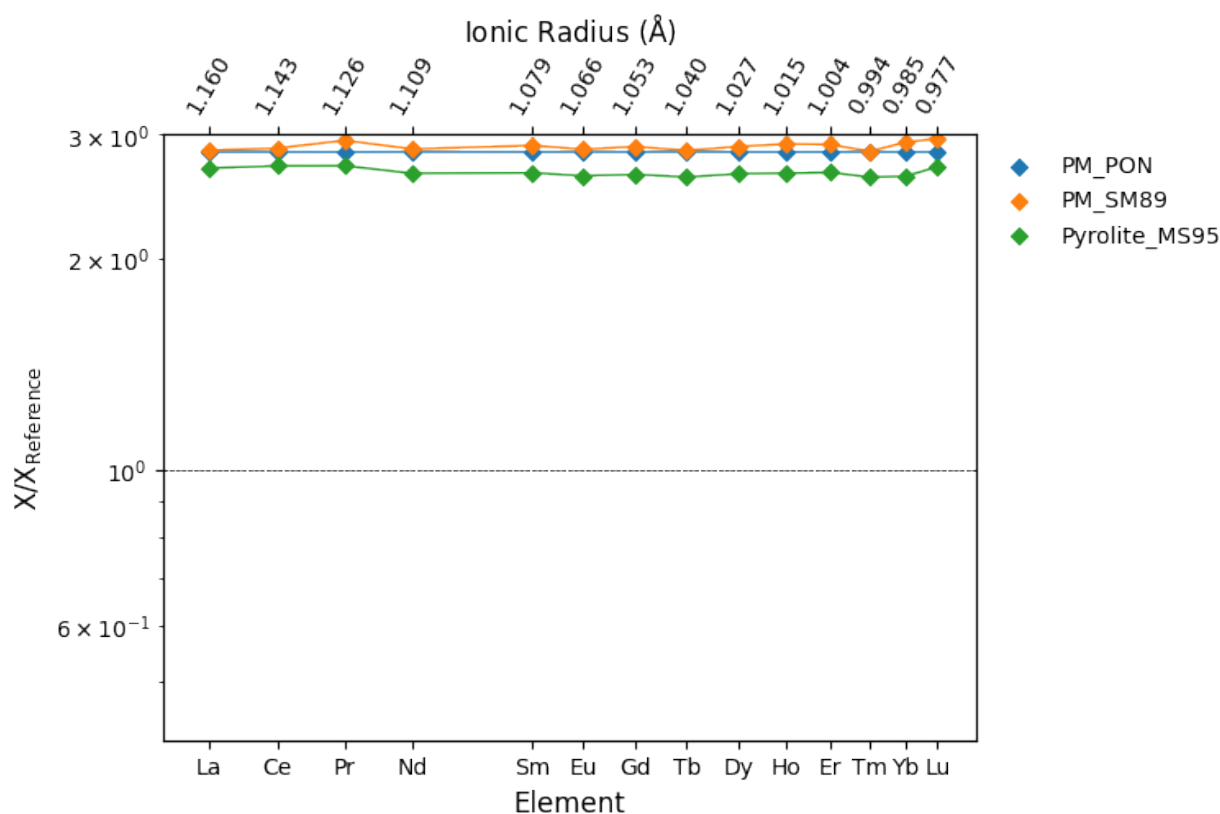


## Mantle

### Primitive Mantle & Pyrolite

```
fltr = lambda c: c.reservoir in ["PrimitiveMantle", "BSE"]
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()
```



```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
```

(continues on next page)



(continued from previous page)

```

↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)

```

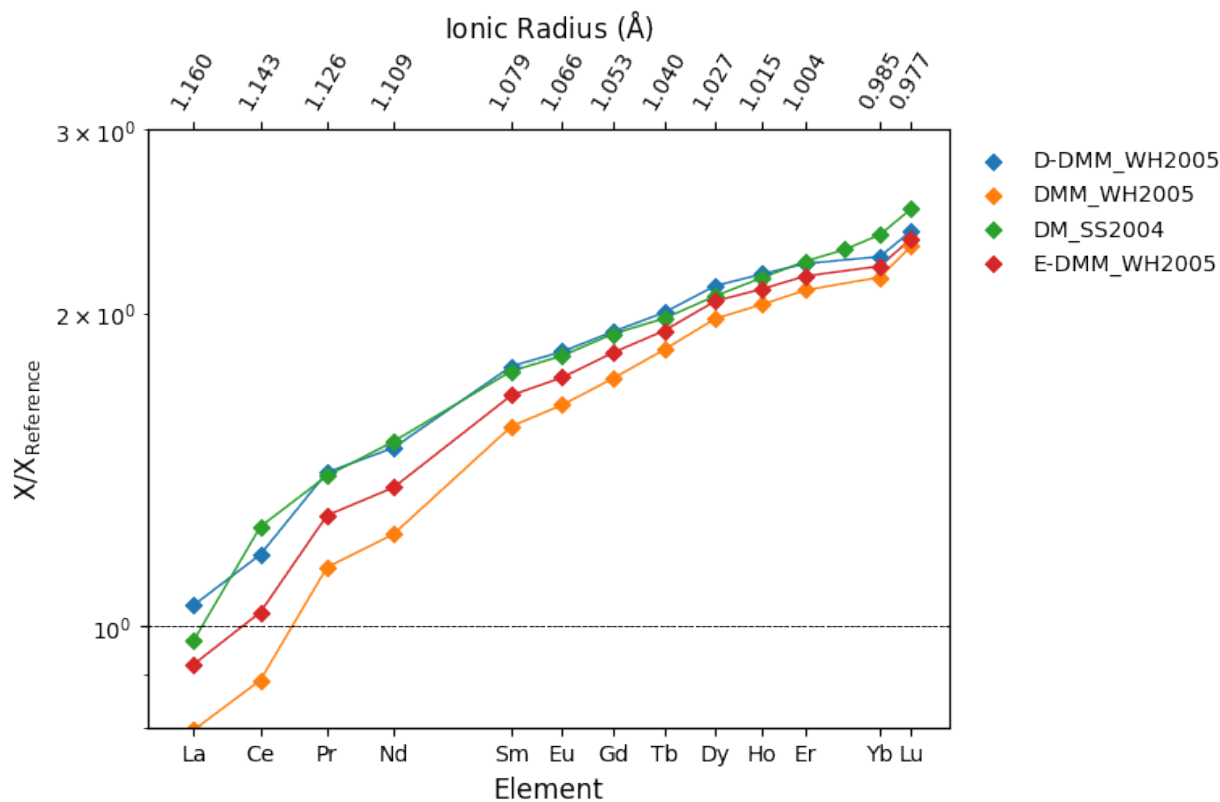
## Depleted Mantle

```

fltr = lambda c: ("Depleted" in c.reservoir) & ("Mantle" in c.reservoir)
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()

```



```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)

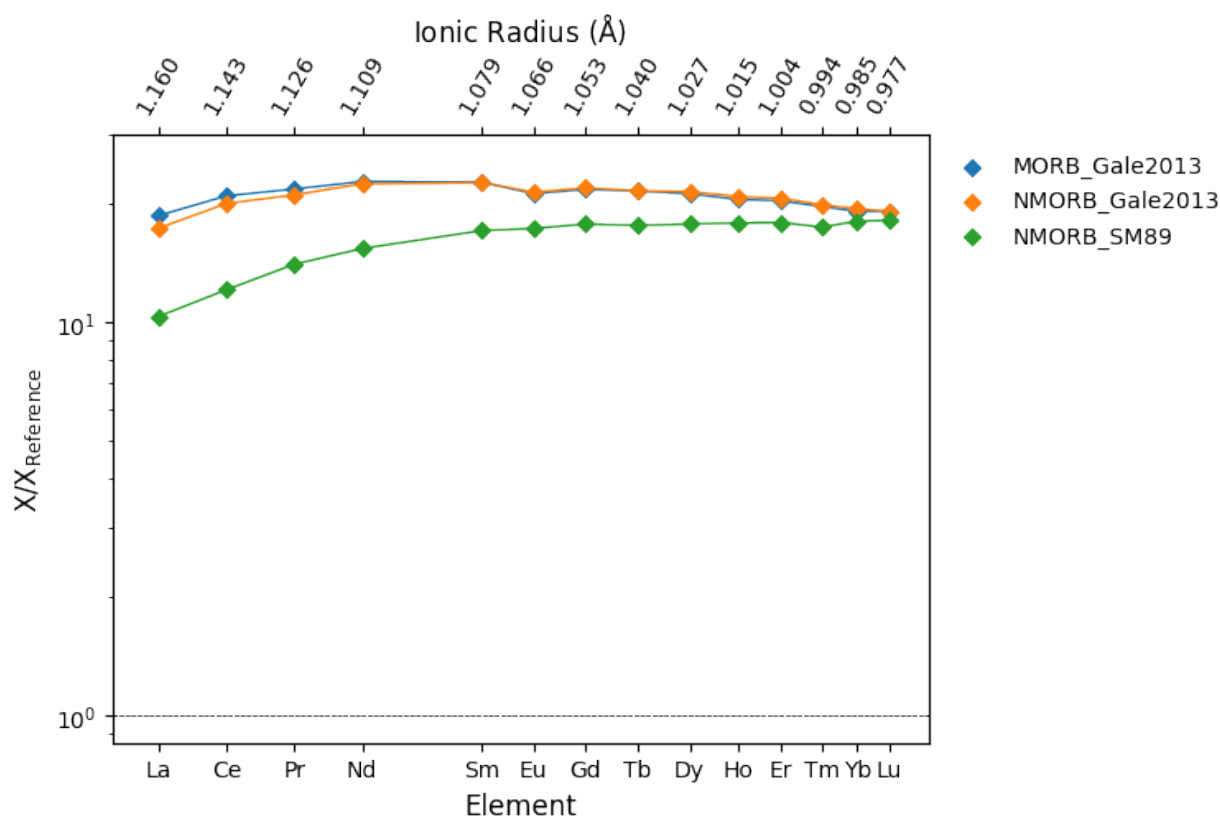
```

## Mid-Ocean Ridge Basalts (MORB)

### Average MORB, NMORB

```
fltr = lambda c: c.reservoir in ["MORB", "NMORB"]
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()
```



```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
```

(continues on next page)

(continued from previous page)

```

→ will be ignored.
    ax.set_ylim(_ymin, _ymax)

```

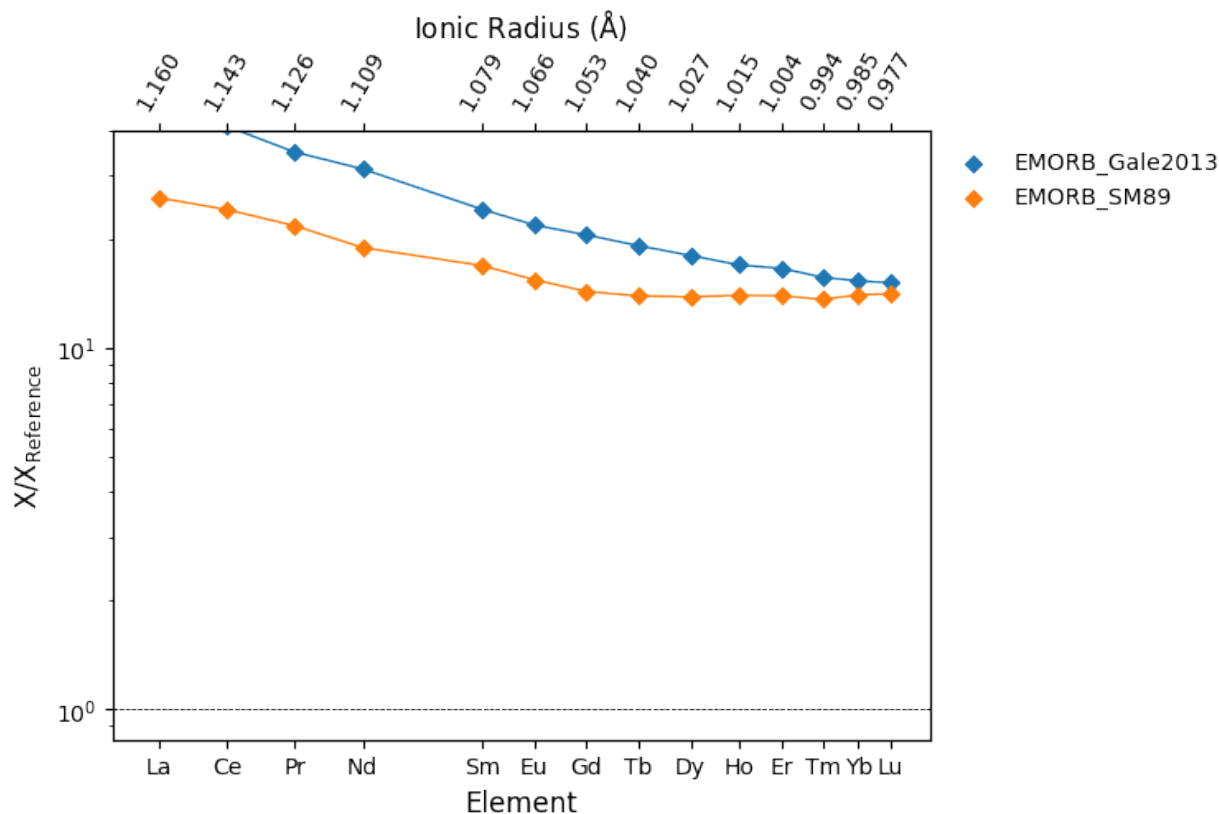
## Enriched MORB

```

fltr = lambda c: "EMORB" in c.reservoir
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()

```



```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
→ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
→ will be ignored.

```

(continues on next page)

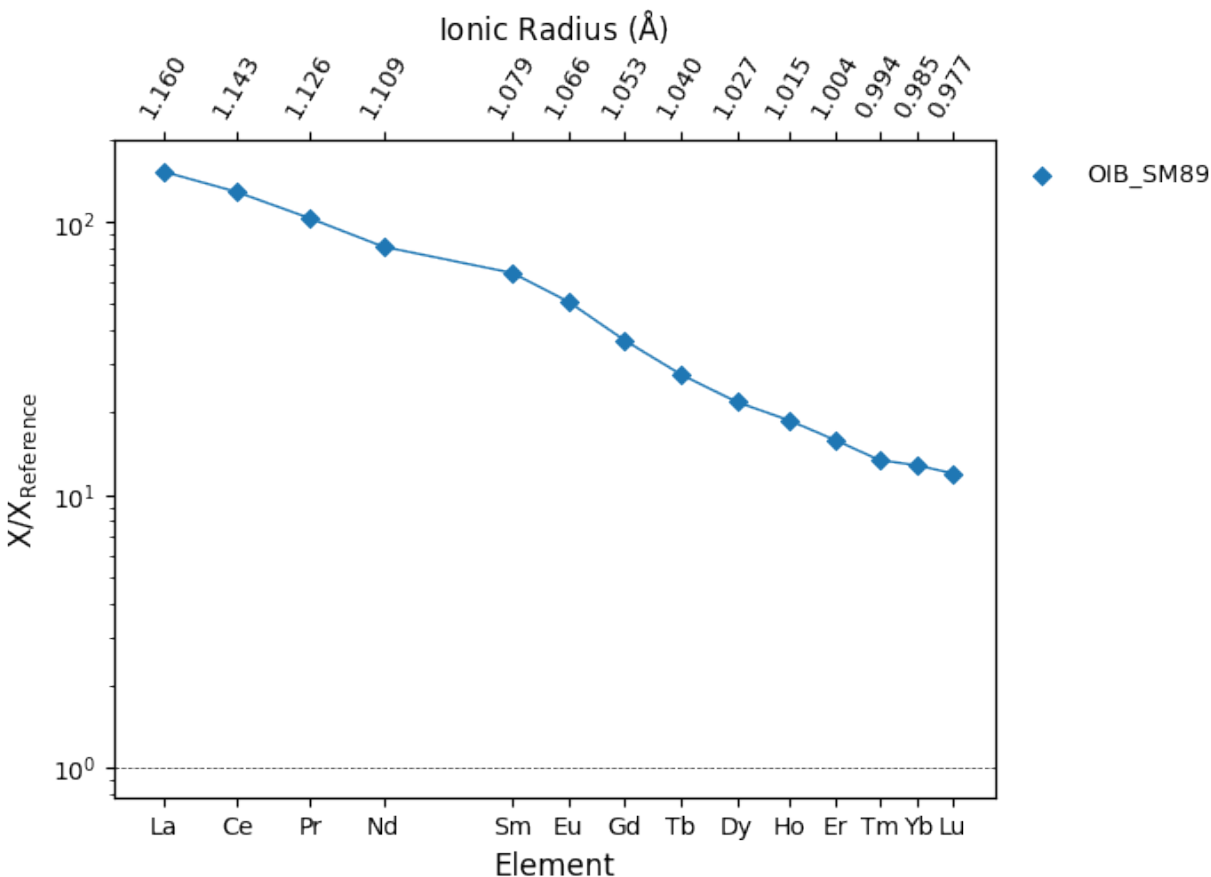
(continued from previous page)

```
ax.set_ylim(_ymin, _ymax)
```

## Ocean Island Basalts

```
fltr = lambda c: "OIB" in c.reservoir
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()
```



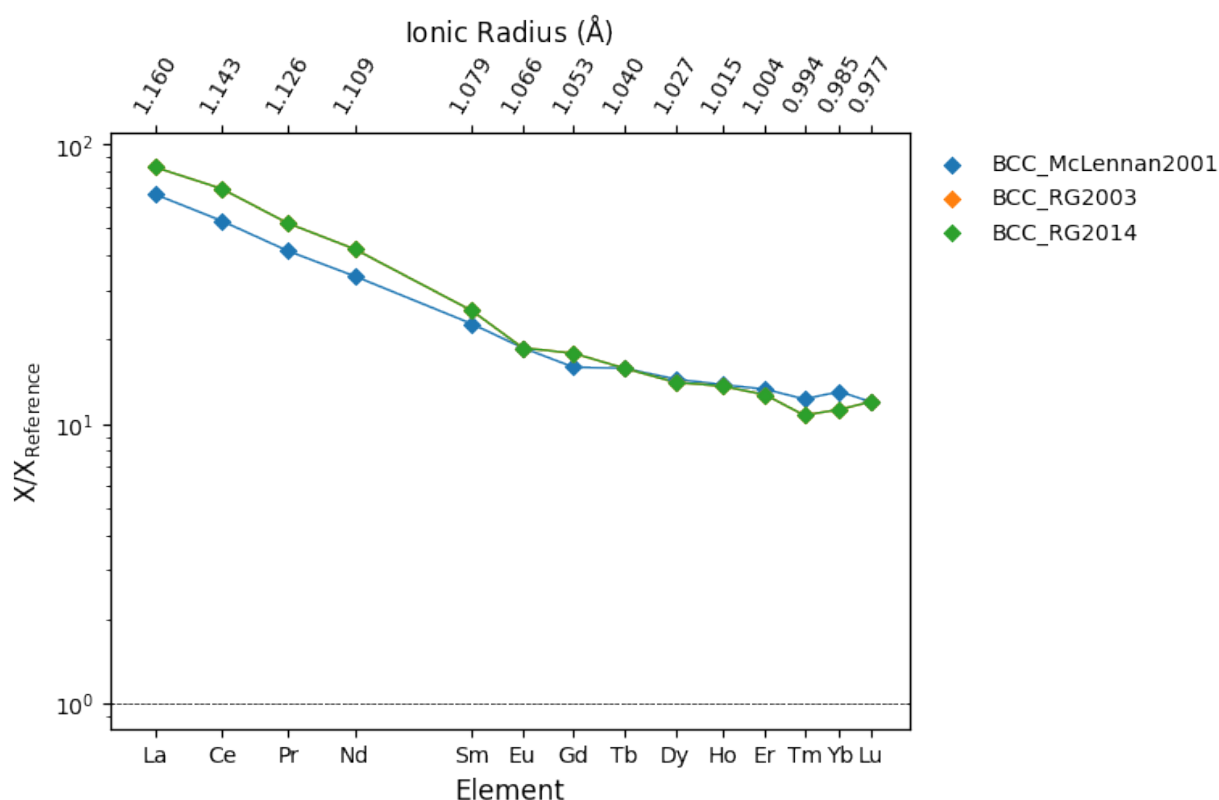
```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
ax.set_ylim(_ymin, _ymax)
```

## Continental Crust

### Bulk Continental Crust

```
fltr = lambda c: c.reservoir == "BulkContinentalCrust"
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()
```



```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
```

(continues on next page)

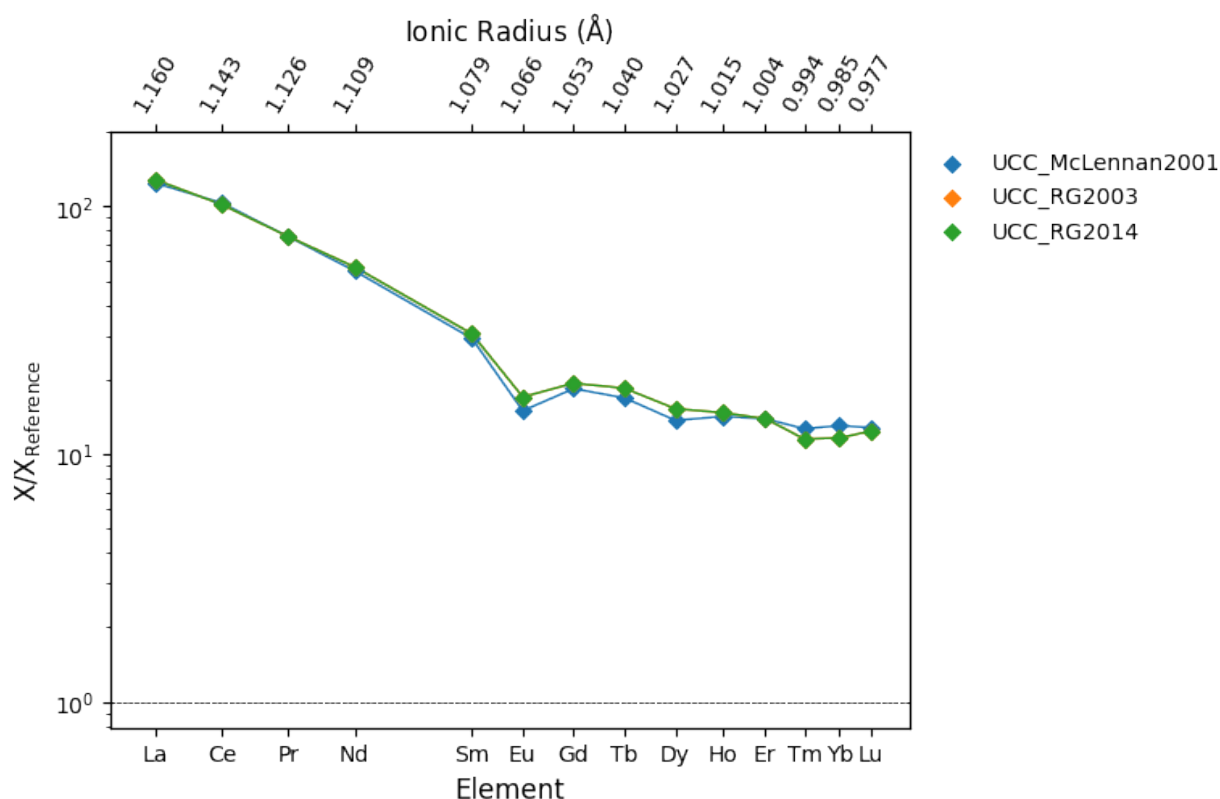
(continued from previous page)

```
ax.set_ylim(_ymin, _ymax)
```

## Upper Continental Crust

```
fltr = lambda c: c.reservoir == "UpperContinentalCrust"
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()
```



```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
ax.set_ylim(_ymin, _ymax)
```

(continues on next page)

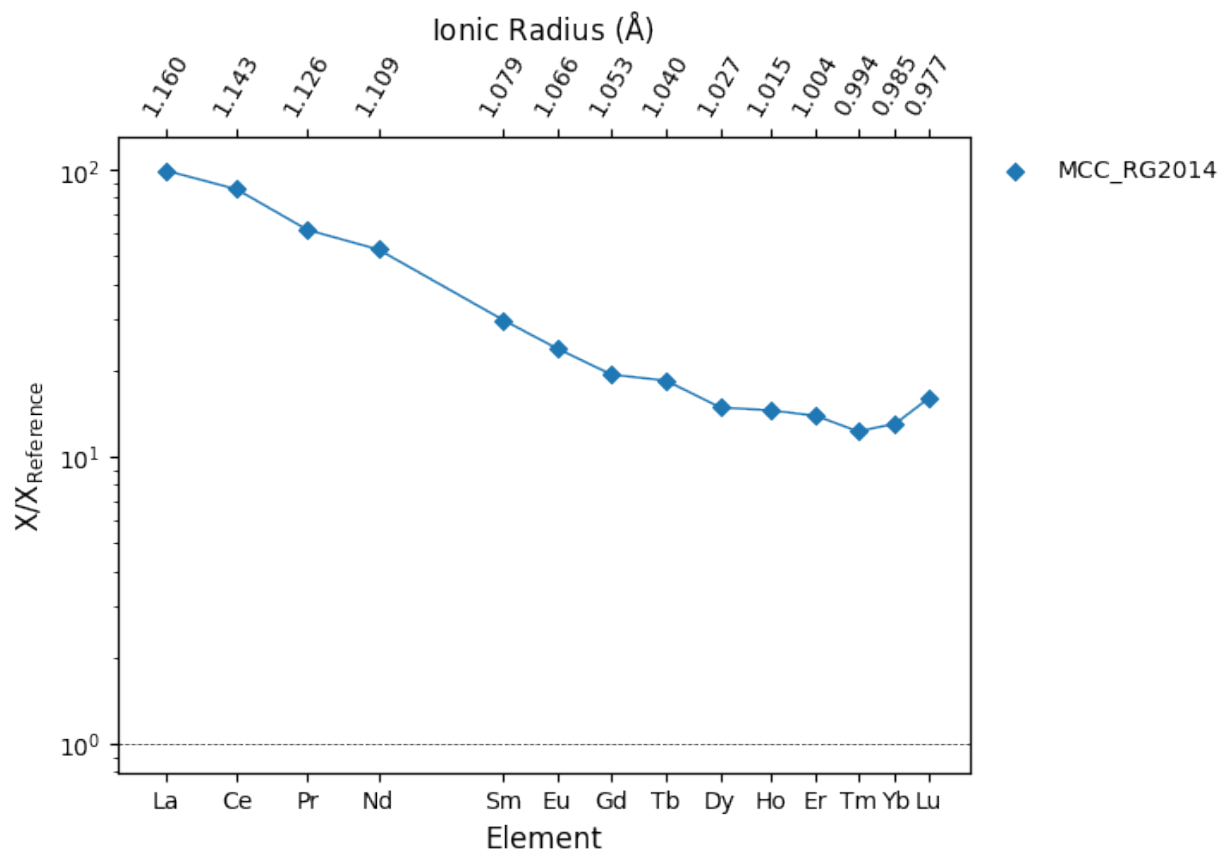
(continued from previous page)

```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)
```

## Mid-Continental Crust

```
fltr = lambda c: c.reservoir == "MidContinentalCrust"
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()
```



```
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)
```

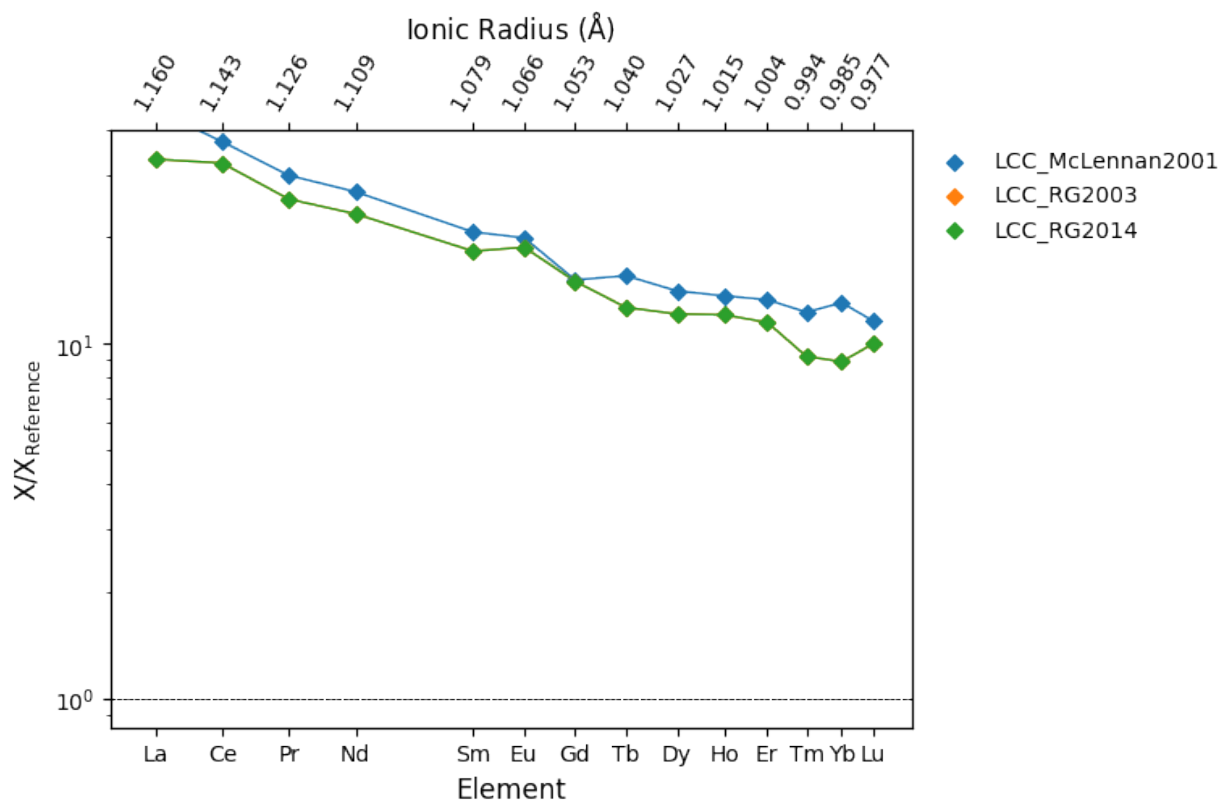
## Lower Continental Crust

```

fltr = lambda c: c.reservoir == "LowerContinentalCrust"
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()

```



```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)

```



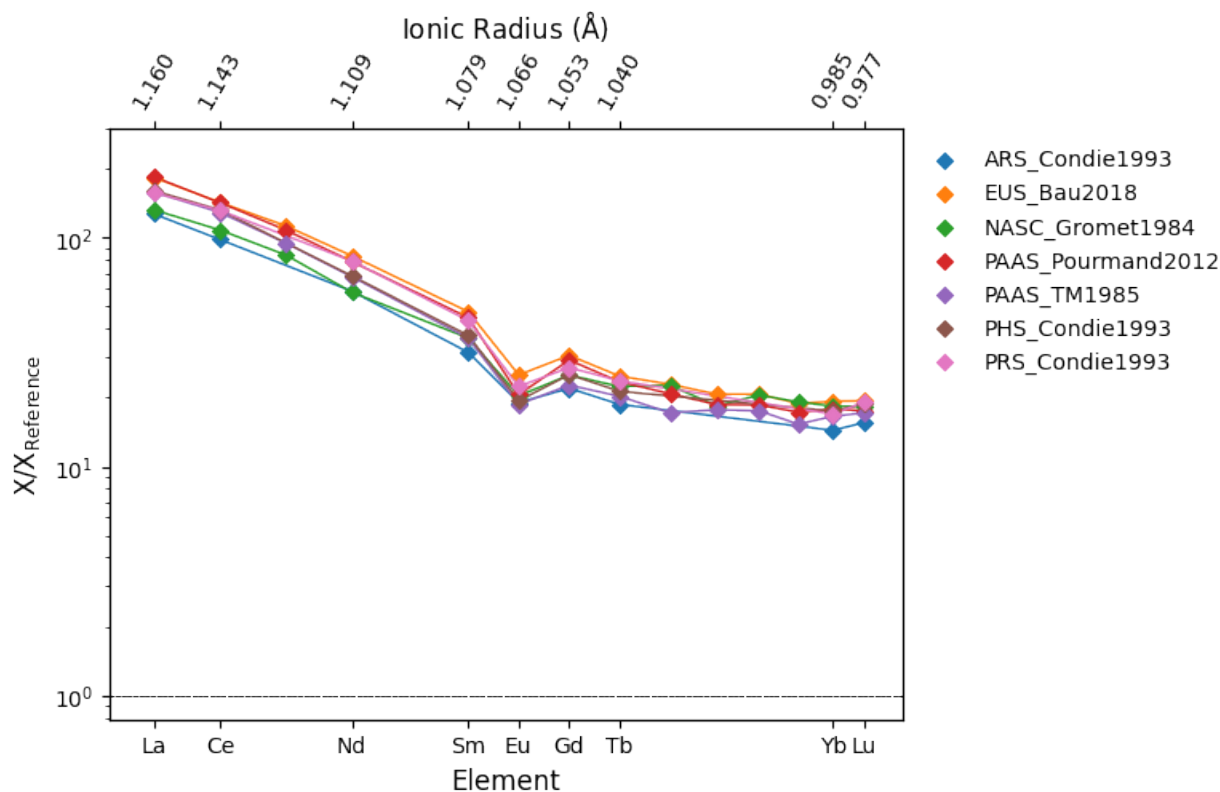
## Shales

```

fltr = lambda c: "Shale" in c.reservoir
compositions = [x for (name, x) in refcomps.items() if fltr(x)]

fig, ax = plt.subplots(1)
for composition in compositions:
    composition.set_units("ppm")
    df = composition.comp.pyrochem.normalize_to(norm, units="ppm")
    df.pyroplot.REE(unity_line=True, ax=ax, label=composition.name)
ax.legend()
plt.show()

```



```

/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳ plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳ will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/

```

(continues on next page)

(continued from previous page)

```
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)
/home/docs/checkouts/readthedocs.org/user_builds/pyrolite/checkouts/develop/pyrolite/
↳plot/spider.py:276: UserWarning: Attempt to set non-positive ylim on a log-scaled axis.
↳will be ignored.
    ax.set_ylim(_ymin, _ymax)
```

## Composition List

See also:

Examples:

[Normalisation](#)

**Total running time of the script:** (0 minutes 12.401 seconds)

## 5.3 Extensions

This page will list packages which have been developed as extensions to pyrolite.

---

**Note:** This page is a work in progress and will updated soon as submodules from `pyrolite.ext` are migrated out of pyrolite to reduce installation overheads and unnecessary dependencies.

---

### 5.3.1 pyrolite-meltsutil

- `pyrolite` extension which wraps the `alphaMELTS` executable
- Functions for working with `alphaMELTS` tables
- Eventually, this will likely instead link to *under-development* `python-MELTS`.

---

**Note:** There are some great things happening on the `MELTS-for-scripting` front; these utilities will be intended to link your data to these tools.

---

## BIBLIOGRAPHY

- [Ringwood1962] Ringwood, A.E. (1962). A model for the upper mantle. *Journal of Geophysical Research* (1896-1977) 67, 857–867. doi: [10.1029/JZ067i002p00857](https://doi.org/10.1029/JZ067i002p00857)
- [Box1976] Box, G.E.P. (1976). Science and Statistics. *Journal of the American Statistical Association* 71, 791–799. doi: [10.1080/01621459.1976.10480949](https://doi.org/10.1080/01621459.1976.10480949)
- [Wickham2014] Wickham, H., 2014. Tidy Data. *Journal of Statistical Software* 59, 1–23. doi: [doi.org/10.18637/jss.v059.i10](https://doi.org/10.18637/jss.v059.i10)
- [Shannon1976] Shannon RD (1976). Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A* 32:751–767. doi: [10.1107/S0567739476001551](https://doi.org/10.1107/S0567739476001551).
- [WhittakerMuntus1970] Whittaker, E.J.W., Muntus, R., 1970. Ionic radii for use in geochemistry. *Geochimica et Cosmochimica Acta* 34, 945–956. doi: [10.1016/0016-7037\(70\)90077-3](https://doi.org/10.1016/0016-7037(70)90077-3).
- [Pauling1960] Pauling, L., 1960. *The Nature of the Chemical Bond*. Cornell University Press, Ithaca, NY.
- [Cross1902] Cross, W., Iddings, J. P., Pirsson, L. V., & Washington, H. S. (1902). A Quantitative Chemico-Mineralogical Classification and Nomenclature of Igneous Rocks. *The Journal of Geology*, 10(6), 555–690. doi: [10.1086/621030](https://doi.org/10.1086/621030)
- [Verma2003] Verma, S. P., Torres-Alvarado, I. S., & Velasco-Tapia, F. (2003). A revised CIPW norm. *Swiss Bulletin of Mineralogy and Petrology*, 83(2), 197–216.
- [Verma2013] Verma, S. P., & Rivera-Gomez, M. A. (2013). Computer Programs for the Classification and Nomenclature of Igneous Rocks. *Episodes*, 36(2), 115–124.
- [LeMaitre1976] Le Maitre, R. W (1976). Some Problems of the Projection of Chemical Data into Mineralogical Classifications. *Contributions to Mineralogy and Petrology* 56, no. 2 (1 January 1976): 181–89. doi: [doi.org/10.1007/BF00399603](https://doi.org/10.1007/BF00399603)
- [Middlemost1989] Middlemost, Eric A. K. (1989). Iron Oxidation Ratios, Norms and the Classification of Volcanic Rocks. *Chemical Geology* 77, 1: 19–26. doi: [doi.org/10.1016/0009-2541\(89\)90011-9](https://doi.org/10.1016/0009-2541(89)90011-9).
- [ONeill2016] O'Neill, H.S.C., 2016. The Smoothness and Shapes of Chondrite-normalized Rare Earth Element Patterns in Basalts. *J Petrology* 57, 1463–1508. doi: [10.1093/petrology/egw047](https://doi.org/10.1093/petrology/egw047).
- [WhittakerMuntus1970] Whittaker, E.J.W., Muntus, R., 1970. Ionic radii for use in geochemistry. *Geochimica et Cosmochimica Acta* 34, 945–956. doi: [10.1016/0016-7037\(70\)90077-3](https://doi.org/10.1016/0016-7037(70)90077-3).
- [Shannon1976] Shannon RD (1976). Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A* 32:751–767. doi: [10.1107/S0567739476001551](https://doi.org/10.1107/S0567739476001551).
- [WhittakerMuntus1970] Whittaker, E.J.W., Muntus, R., 1970. Ionic radii for use in geochemistry. *Geochimica et Cosmochimica Acta* 34, 945–956. doi: [10.1016/0016-7037\(70\)90077-3](https://doi.org/10.1016/0016-7037(70)90077-3).

[Pauling1960] Pauling, L., 1960. The Nature of the Chemical Bond. Cornell University Press, Ithaca, NY.

[Aitchison1984] Aitchison, J., 1984. The statistical analysis of geochemical compositions. Journal of the International Association for Mathematical Geology 16, 531–564. doi: [10.1007/BF01029316](https://doi.org/10.1007/BF01029316)

## PYTHON MODULE INDEX

### p

pyrolite.comp, ??  
pyrolite.comp.aggregate, ??  
pyrolite.comp.codata, ??  
pyrolite.comp.impute, ??  
pyrolite.geochem, ??  
pyrolite.geochem.alteration, ??  
pyrolite.geochem.ind, ??  
pyrolite.geochem.ions, ??  
pyrolite.geochem.isotope, ??  
pyrolite.geochem.magma, ??  
pyrolite.geochem.norm, ??  
pyrolite.geochem.parse, ??  
pyrolite.geochem.transform, ??  
pyrolite.mineral, ??  
pyrolite.mineral.lattice, ??  
pyrolite.mineral.mindb, ??  
pyrolite.mineral.normative, ??  
pyrolite.mineral.sites, ??  
pyrolite.mineral.template, ??  
pyrolite.mineral.transform, ??  
pyrolite.plot, ??  
pyrolite.plot.biplot, ??  
pyrolite.plot.color, ??  
pyrolite.plot.density, ??  
pyrolite.plot.density.grid, ??  
pyrolite.plot.density.ternary, ??  
pyrolite.plot.parallel, ??  
pyrolite.plot.spider, ??  
pyrolite.plot.stem, ??  
pyrolite.plot.templates, ??  
pyrolite.util, ??  
pyrolite.util.classification, ??  
pyrolite.util.distributions, ??  
pyrolite.util.general, ??  
pyrolite.util.lambdas, ??  
pyrolite.util.lambdas.eval, ??  
pyrolite.util.lambdas.oneill, ??  
pyrolite.util.lambdas.opt, ??  
pyrolite.util.lambdas.params, ??  
pyrolite.util.lambdas.plot, ??  
pyrolite.util.lambdas.transform, ??  
pyrolite.util.log, ??  
pyrolite.util.math, ??  
pyrolite.util.meta, ??  
pyrolite.util.missing, ??  
pyrolite.util.pd, ??  
pyrolite.util.plot, ??  
pyrolite.util.plot.axes, ??  
pyrolite.util.plot.density, ??  
pyrolite.util.plot.export, ??  
pyrolite.util.plot.grid, ??  
pyrolite.util.plot.helpers, ??  
pyrolite.util.plot.interpolation, ??  
pyrolite.util.plot.legend, ??  
pyrolite.util.plot.style, ??  
pyrolite.util.plot.transform, ??  
pyrolite.util.resampling, ??  
pyrolite.util.sklearn, ??  
pyrolite.util.sklearn.impute, ??  
pyrolite.util.sklearn.pipeline, ??  
pyrolite.util.sklearn.select, ??  
pyrolite.util.sklearn.transform, ??  
pyrolite.util.sklearn.vis, ??  
pyrolite.util.spatial, ??  
pyrolite.util.synthetic, ??  
pyrolite.util.text, ??  
pyrolite.util.time, ??  
pyrolite.util.types, ??  
pyrolite.util.units, ??  
pyrolite.util.web, ??