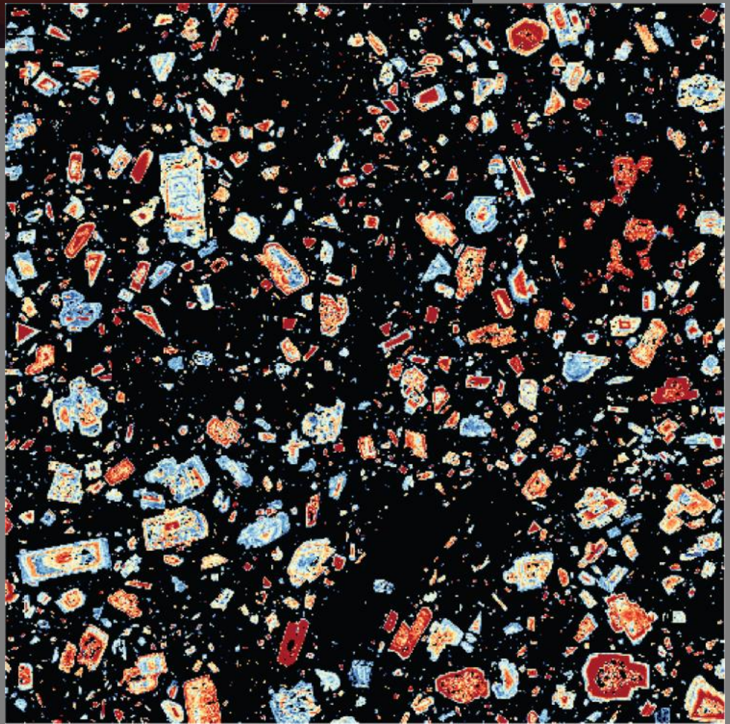


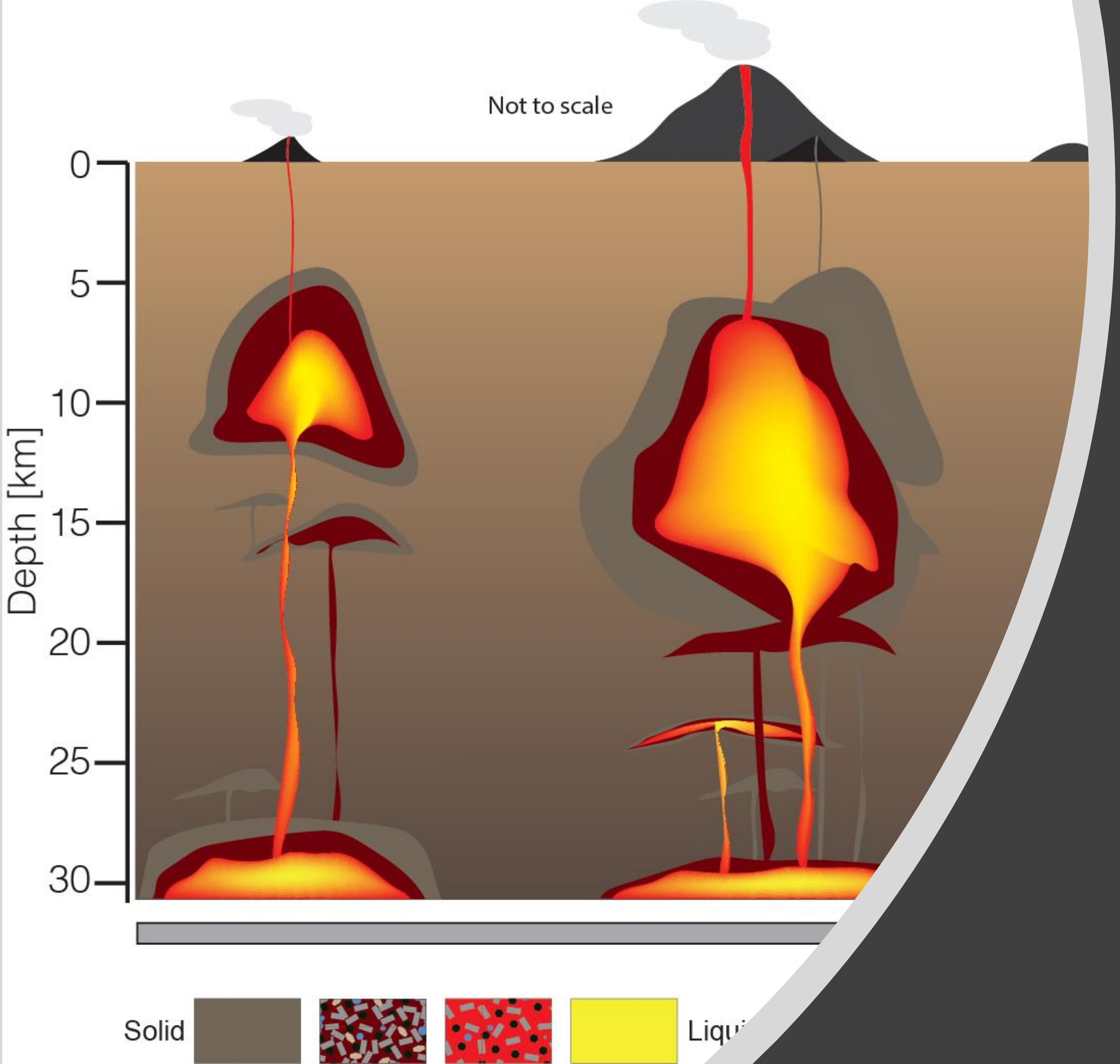
Machine learning in volcanology

Luca Caricchi, Maurizio Petrelli, Corin Jorgenson, Oliver Higgins

UNIVERSITÉ DE GENÈVE

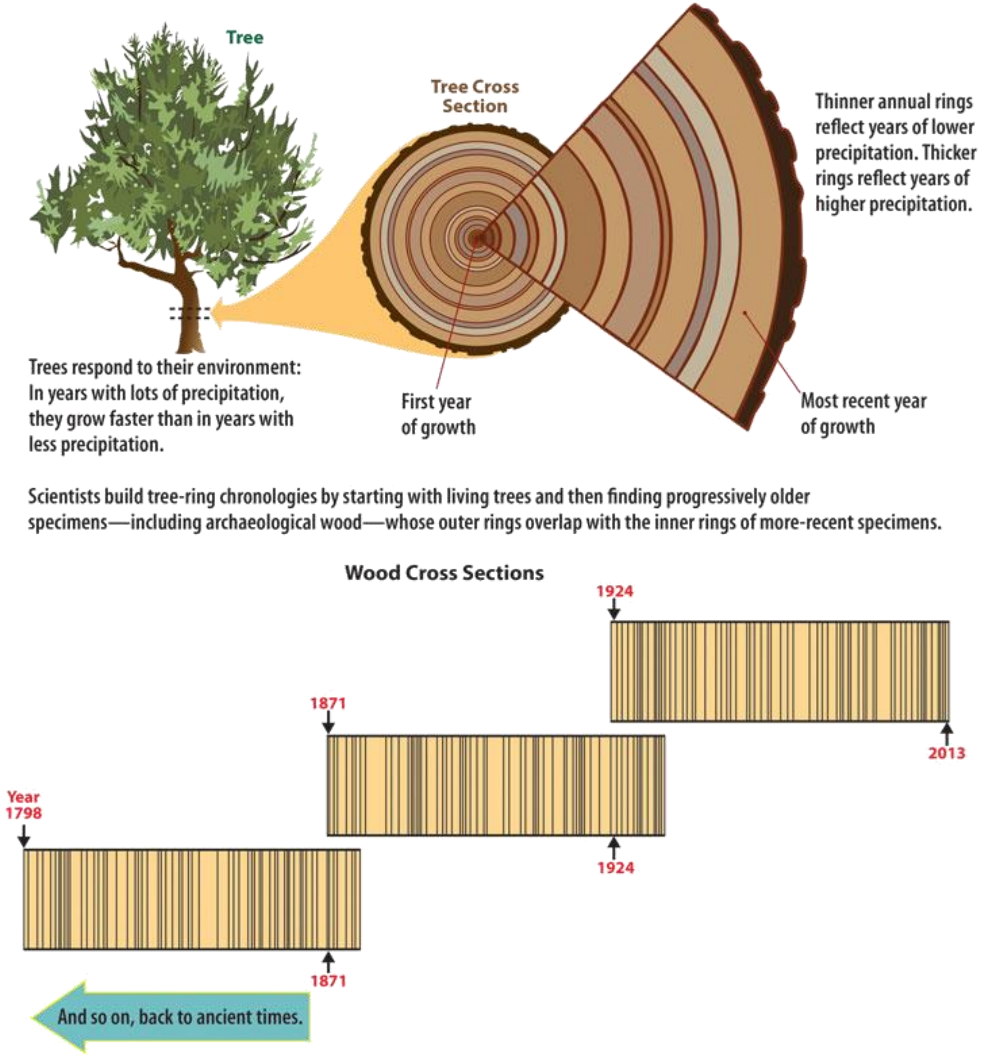
FACULTY OF SCIENCE
Department of Earth Sciences

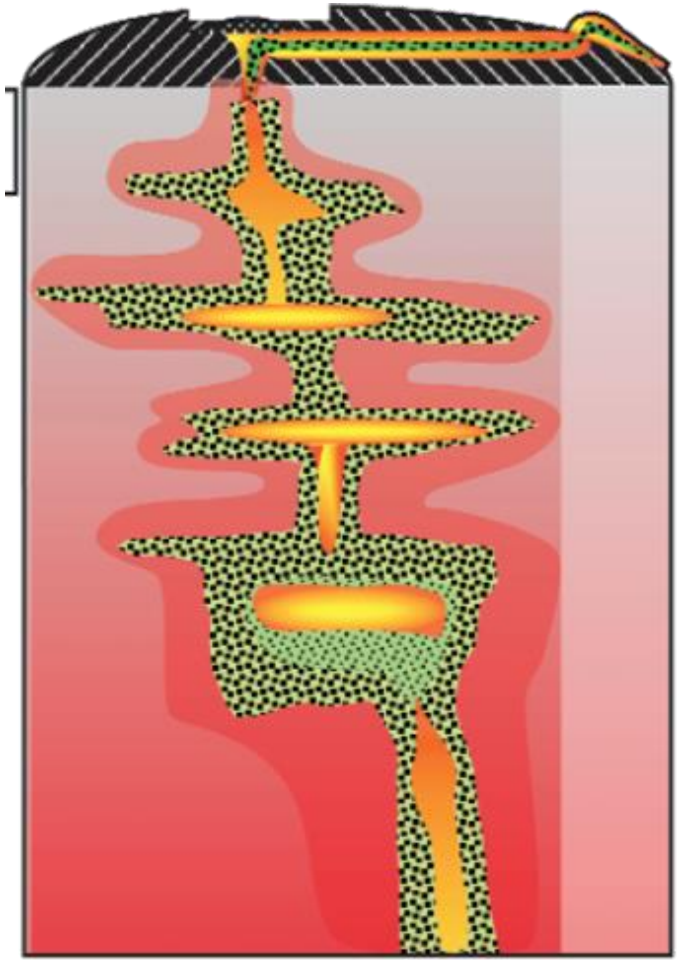
Swiss National Science Foundation



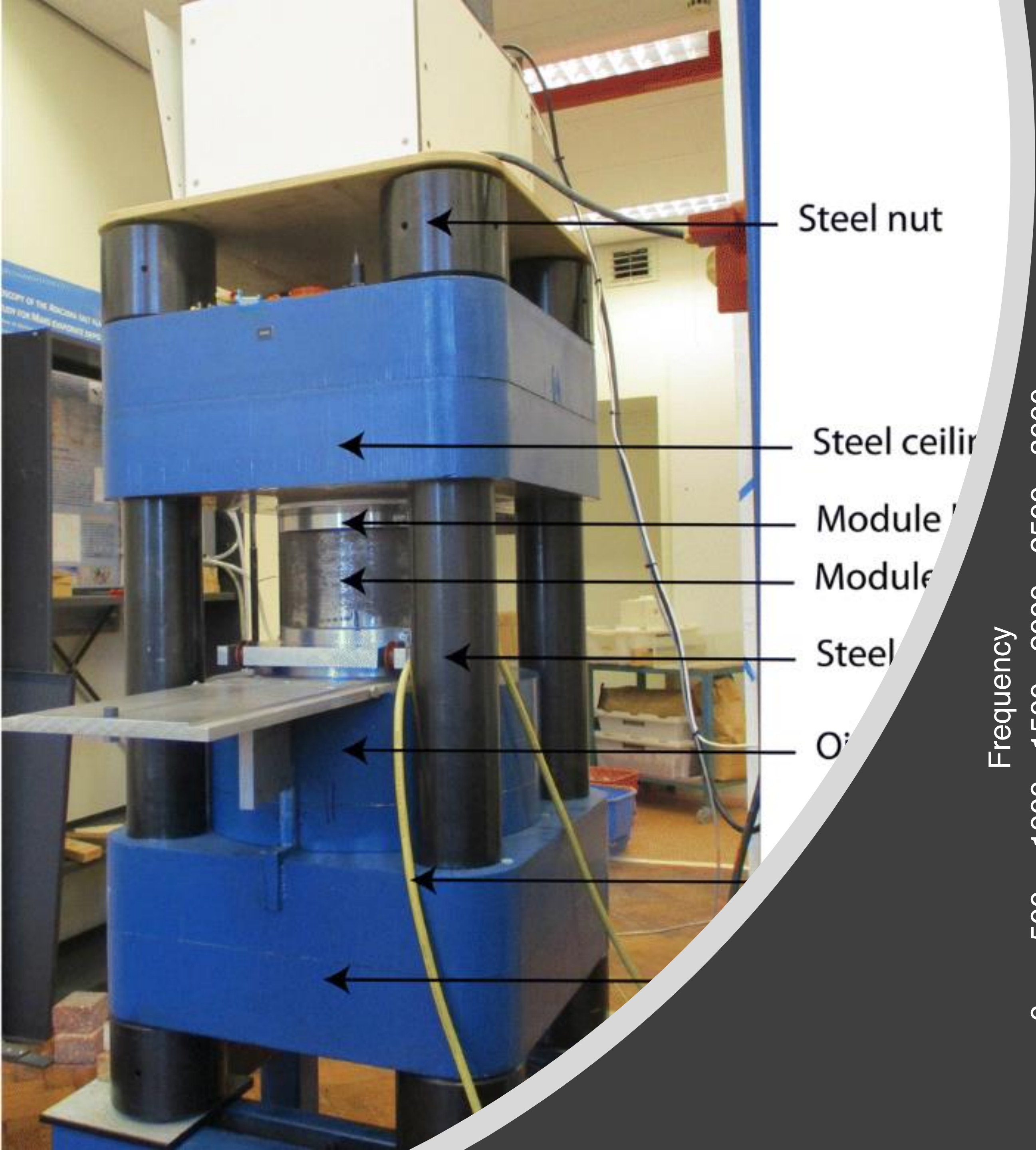
We do not get direct access to magma chambers

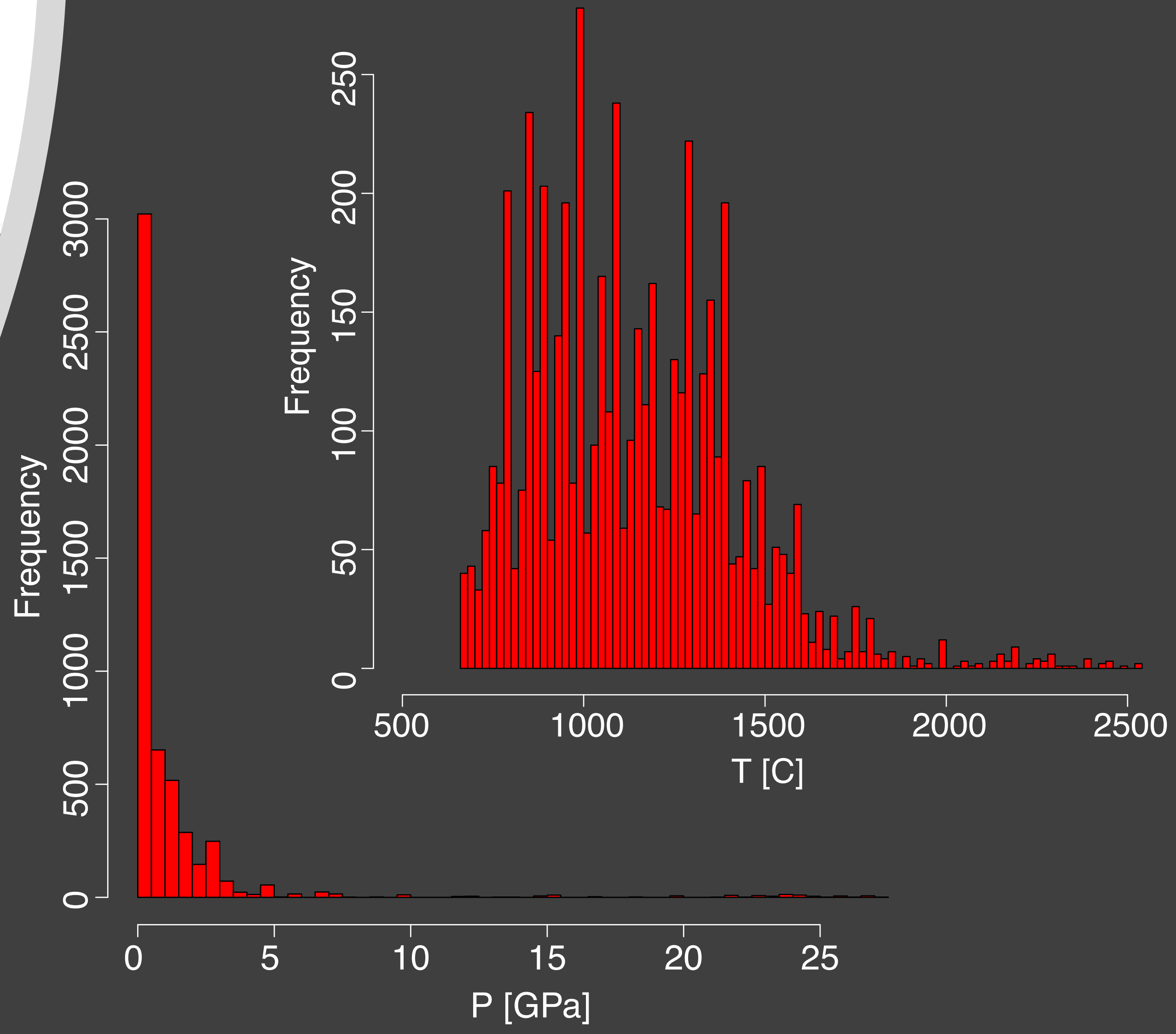
Trees and minerals





Experiments

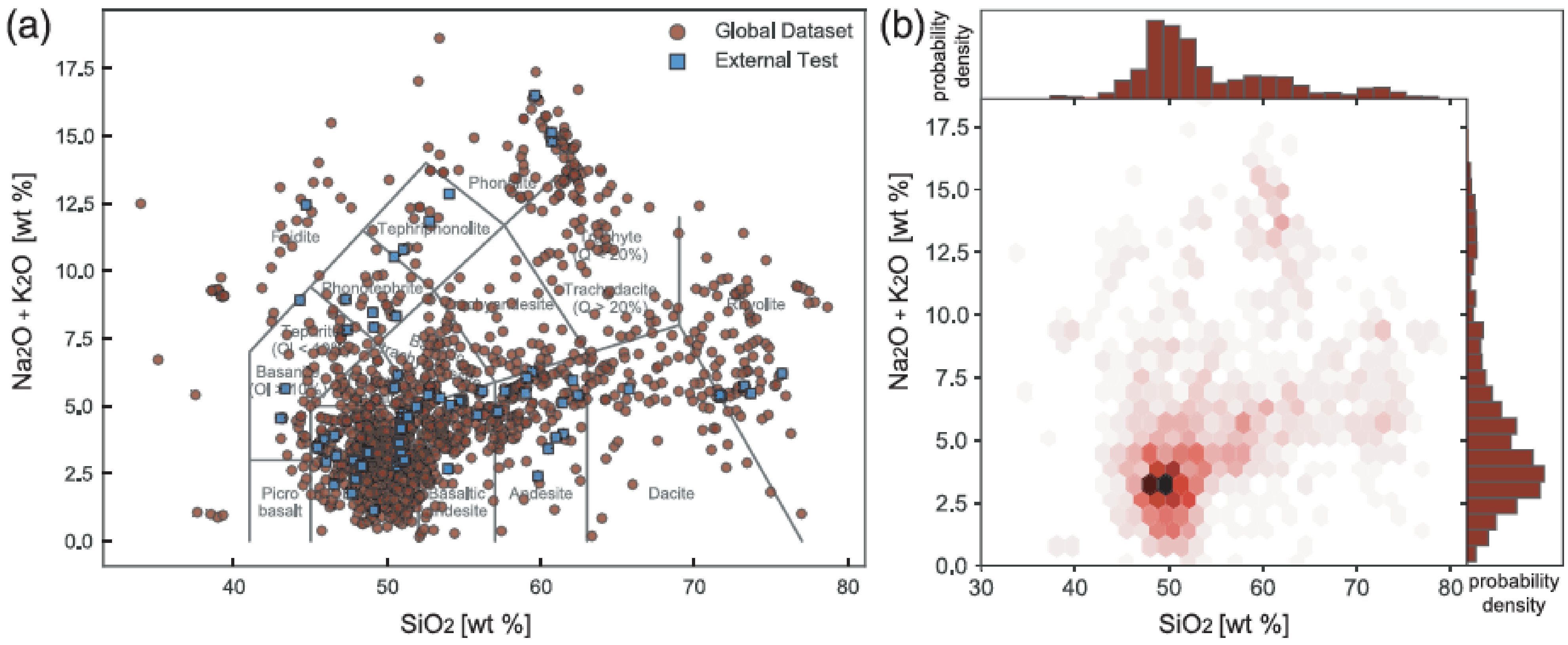




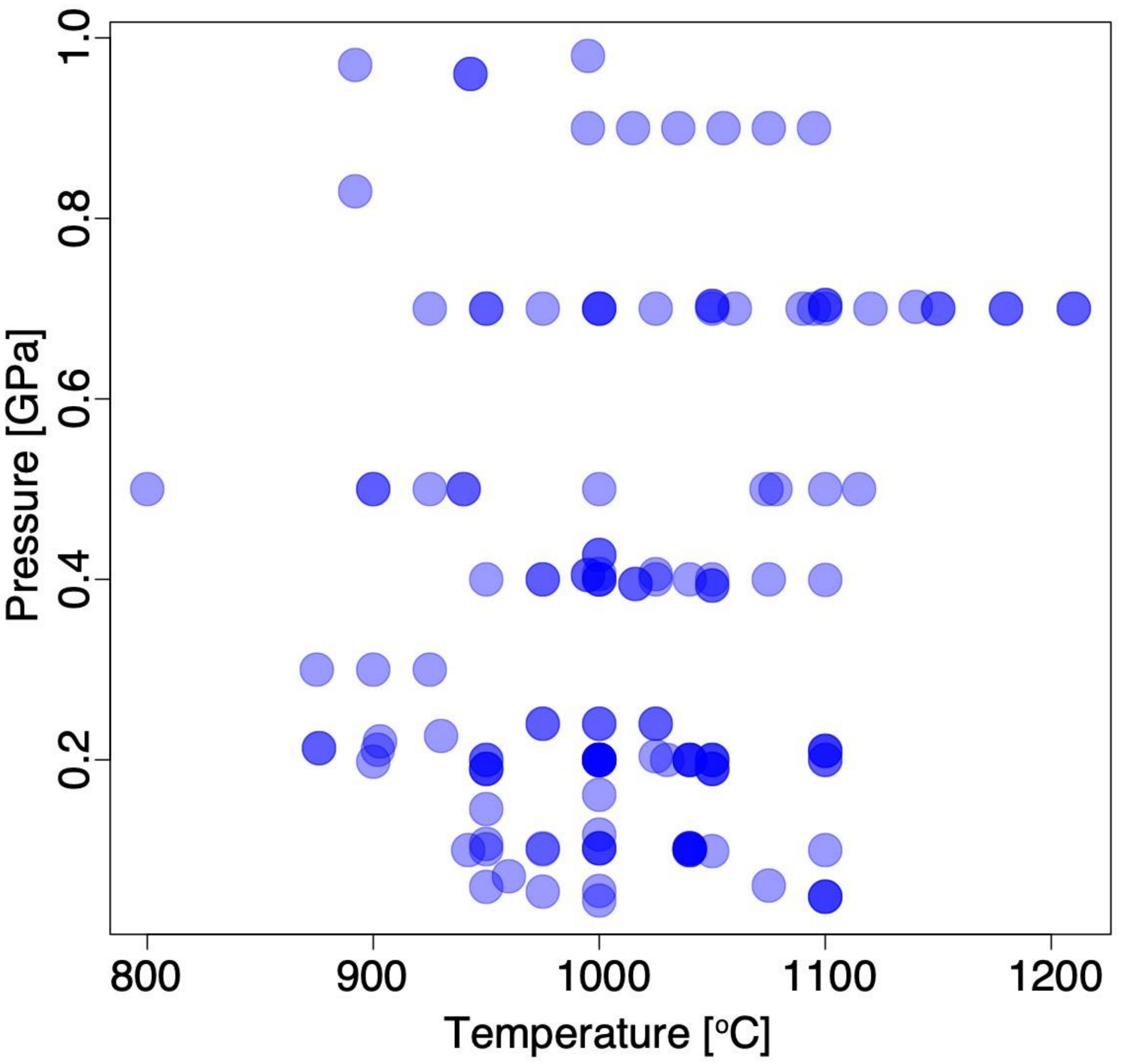
How do the data look like?

~1400 experiments

Pressure (GPa)	Temperature (T^oC)	SiO_2-mineral (wt.%)	Al_2O_3-mineral (wt.%)	…	SiO_2-magma (wt.%)	Al_2O_3-magma (wt.%)	…
2	1250	49	11		52	15	



An example: cpx and plg (129 experiments)

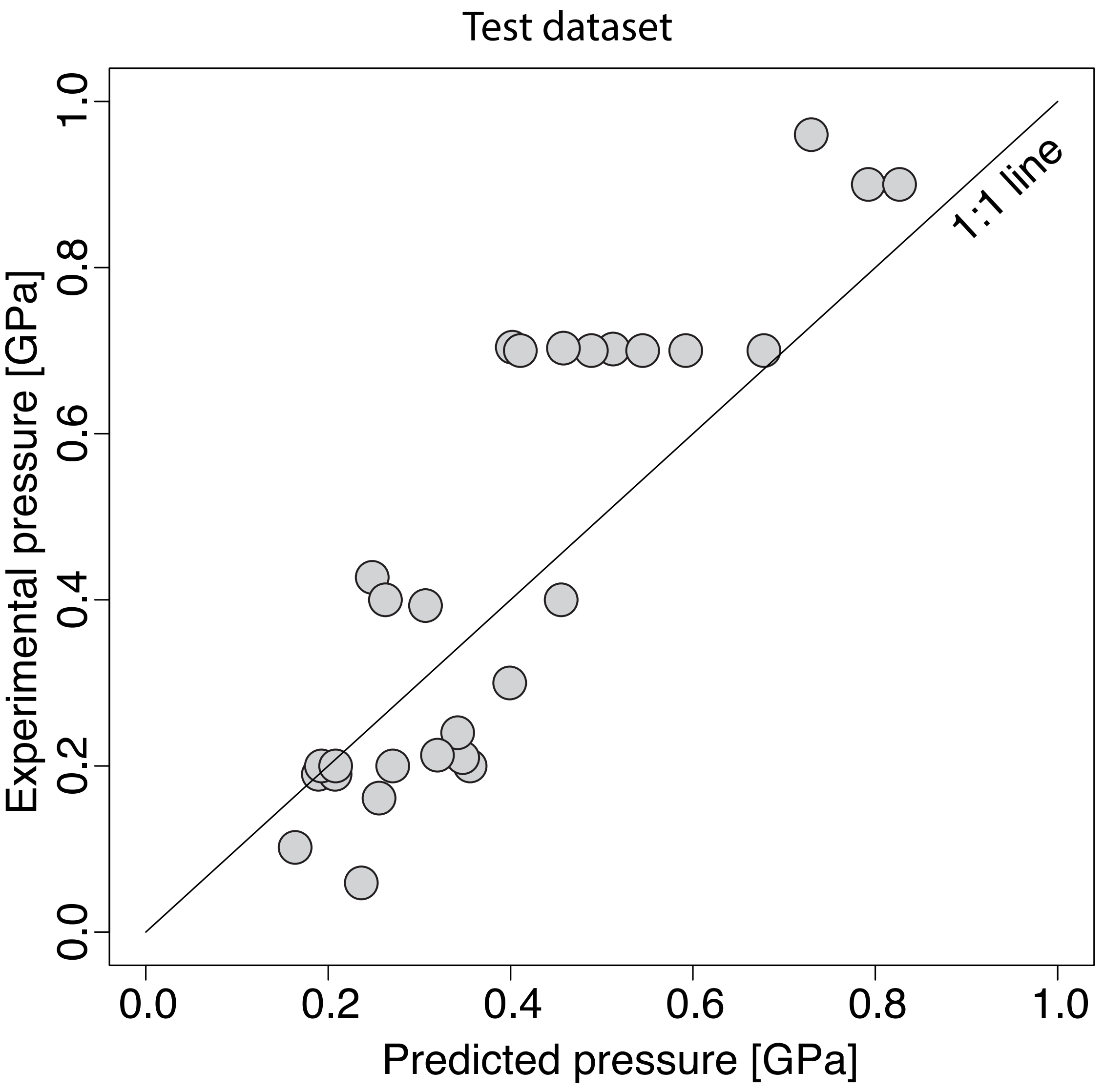


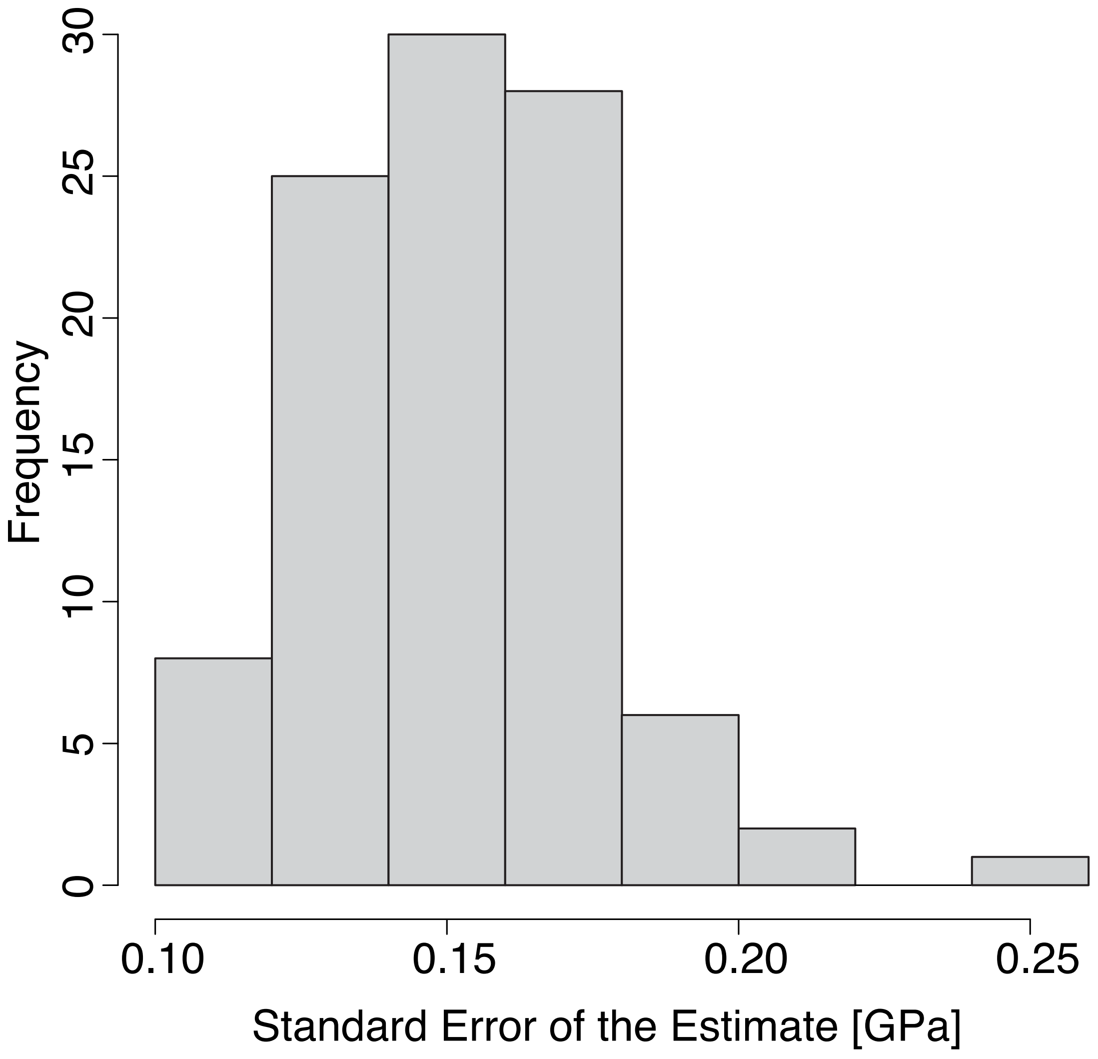
Random forest

1. Split dataset in train (70%), validation (20%) and test (10%) dataset

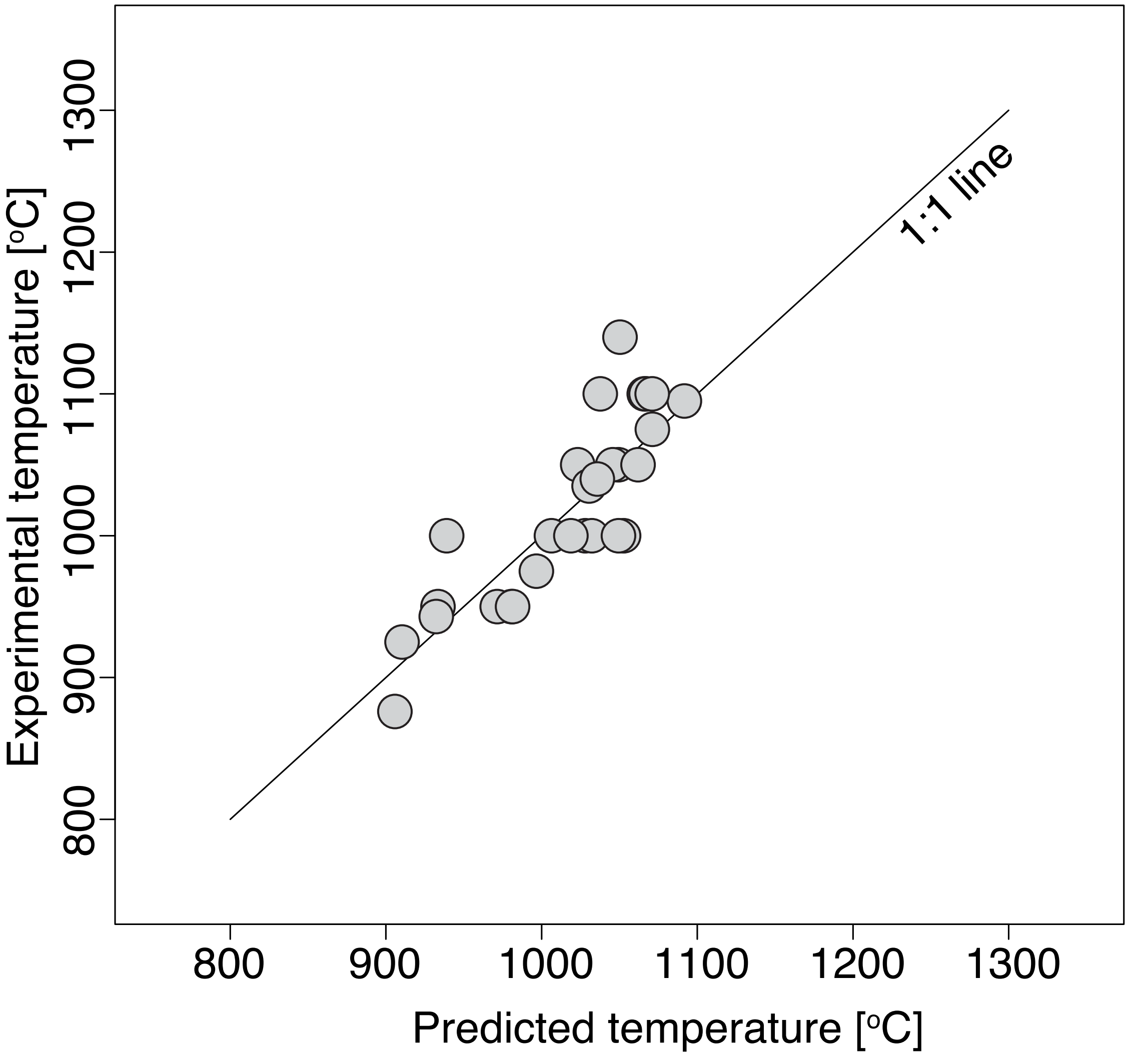
2. Split in a train (80%) and test (20%) dataset

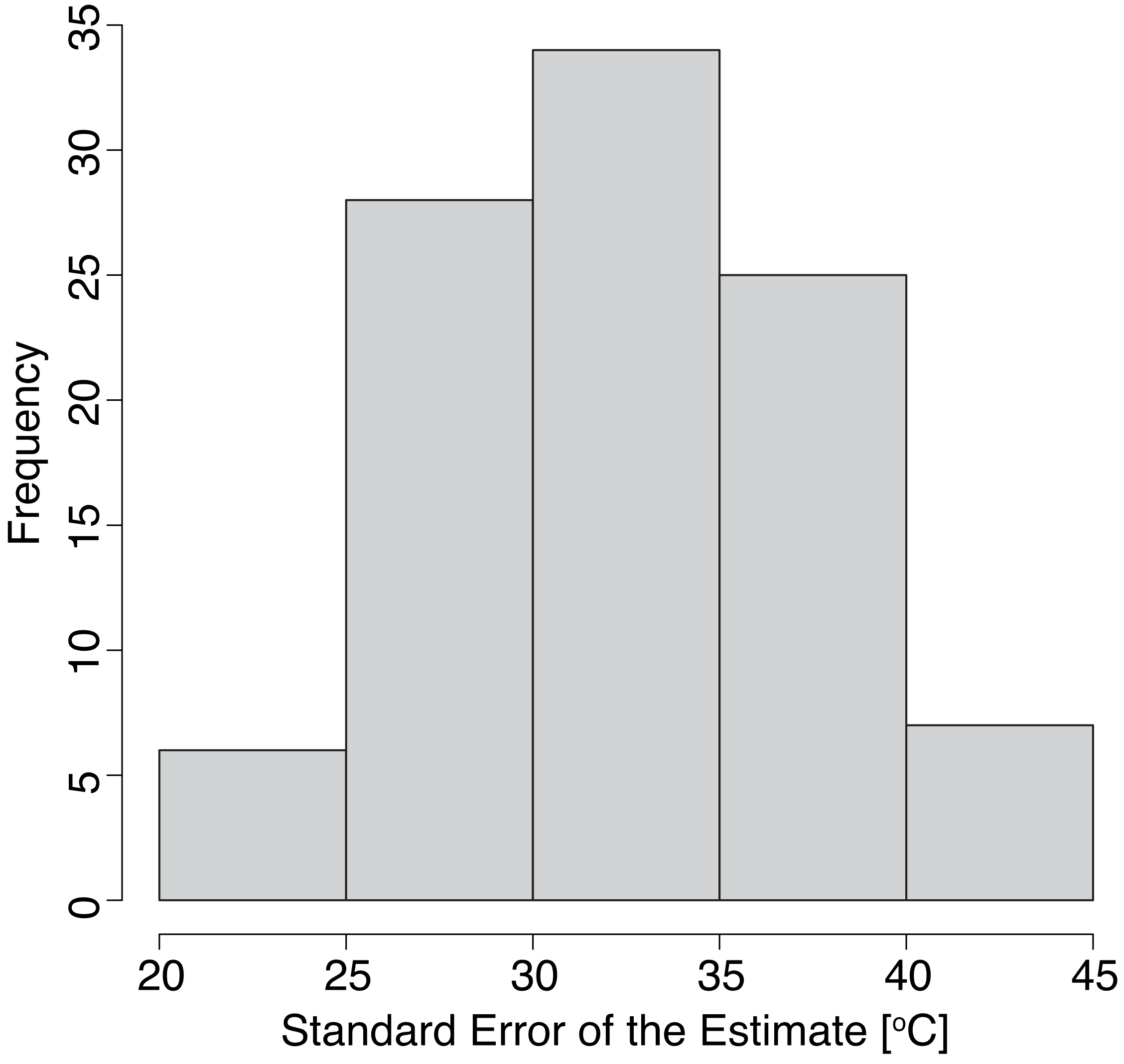
Pressure


Test dataset
1:1 line
Experimental pressure [GPa]
Predicted pressure [GPa]
0.0
0.2
0.4
0.6
0.8
1.0

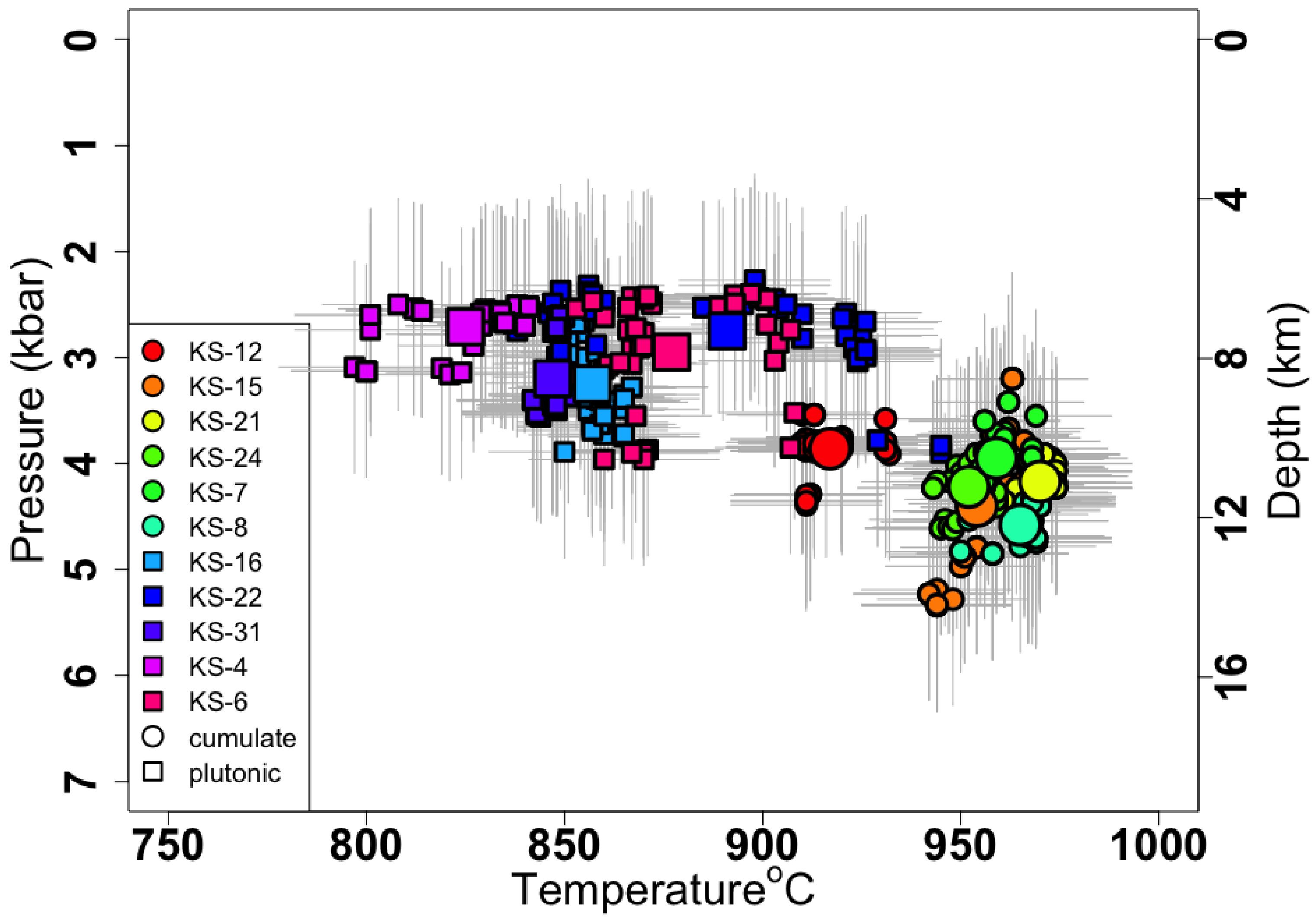

Frequency
Standard Error of the Estimate [GPa]
0
5
10
15
20
25
30
0.10
0.15
0.20
0.25

Temperature





Depth – temperature profile of a magmatic system



1. Since we have few experiments that we use to calibrate our method, "data augmentation" is a salient issue.
2. The range of pressure and temperature is large, but when we study one or two minerals they might be stable only within specific ranges. What is the best approach to calibrate algorithms over the entire pressure and temperature range?
3. Which uncertainty should we use for our models?

More general issues of interest for other projects within our research group

1- How to deal with mixed datasets with images and chemical data?
2- We often have very high quality data (in small numbers) and lower/low-quality data in abundance. What is the best approach to obtain the most information from combining these sort of mixed datasets?