Machine learning in volcanology

Luca Caricchi, Maurizio Petrelli, Corin Jorgenson, Oliver Higgins
We do not get direct access to magma chambers
Trees and minerals

Trees respond to their environment; in years with lots of precipitation, they grow faster than in years with less precipitation.

Scientists build tree-ring chronologies by starting with living trees and then finding progressively older specimens—including archaeological wood—whose outer rings overlap with the inner rings of more-recent specimens.

Wood Cross Sections

And so on, back to ancient times.

Thinner annual rings reflect years of lower precipitation. Thicker rings reflect years of higher precipitation.
Experiments
How do the data look like?

~1400 experiments

<table>
<thead>
<tr>
<th>Pressure (GPa)</th>
<th>Temperature (°C)</th>
<th>SiO₂-mineral (wt.%)</th>
<th>Al₂O₃-mineral (wt.%)</th>
<th>...</th>
<th>SiO₂-magma (wt.%)</th>
<th>Al₂O₃-magma (wt.%)</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1250</td>
<td>49</td>
<td>11</td>
<td></td>
<td>52</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

![Graph showing Na₂O + K₂O vs SiO₂](image1.png)

![Histogram showing SiO₂ probability density](image2.png)
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
Temperature

![Graph showing temperature distribution](image)

---

**Graph Description:**
- The left graph displays a scatter plot with the predicted temperature on the x-axis and the experimental temperature on the y-axis. The 1:1 line indicates perfect agreement between the two.
- The right graph is a histogram showing the frequency distribution of the standard error of the estimate. The x-axis represents the standard error in degrees Celsius, and the y-axis shows the frequency. The bars indicate the number of occurrences within each interval of the standard error.
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?