Machine learning predictions of superalloy microstructure

Patrick L. Taylor[1,*], Gareth Conduit[1]

[1] Cavendish Laboratory, University of Cambridge, * pt409@cam.ac.uk

Introduction

Nickel-base superalloys are a class of alloys that have found widespread use in the aerospace and energy sectors due to their excellent high temperature properties that arise due to the their two-phase matrix-precipitate microstructure. For this reason predicting the amount and composition of each phase is a critical first-step towards making further property predictions.

Typically this is done by using the **Cal**culation of **Pha**se **D**iagrams (CALPHAD) method. This thermodynamic approach minimises the free energy of the system with respect to the variables of interest [1]. This means the problem of fitting the model to experimental data is the problem of correctly modelling the free energy of each phase—an inverse problem.

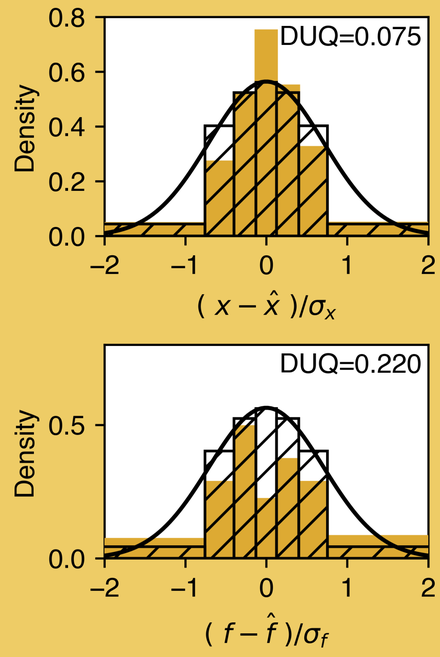
Distribution of Uncertainty Quality

Machine learning delivers uncertainty estimates for each of its predictions, so evaluating how close the uncertainties are to the real-life validation data is crucial. The typical error in a prediction should be normally distributed with standard deviation equal to its associated uncertainty. Binning all of the N errors/uncertainty for a dataset and comparing the resulting histogram to an ideal gaussian allows us to quantify the quality of our predictions. We call this metric the distribution of uncertainty quality (DUQ) [2]. It is defined so that its value is between 0 and 1.

$$\mathrm{DUQ} = \frac{1}{2}\frac{N_{\mathrm{bins}}}{N_{\mathrm{bins}}-1}\sum_{m=1}^{N_{\mathrm{bins}}}\left|\frac{n_m}{N}-\frac{1}{N_{\mathrm{bins}}}\right|$$

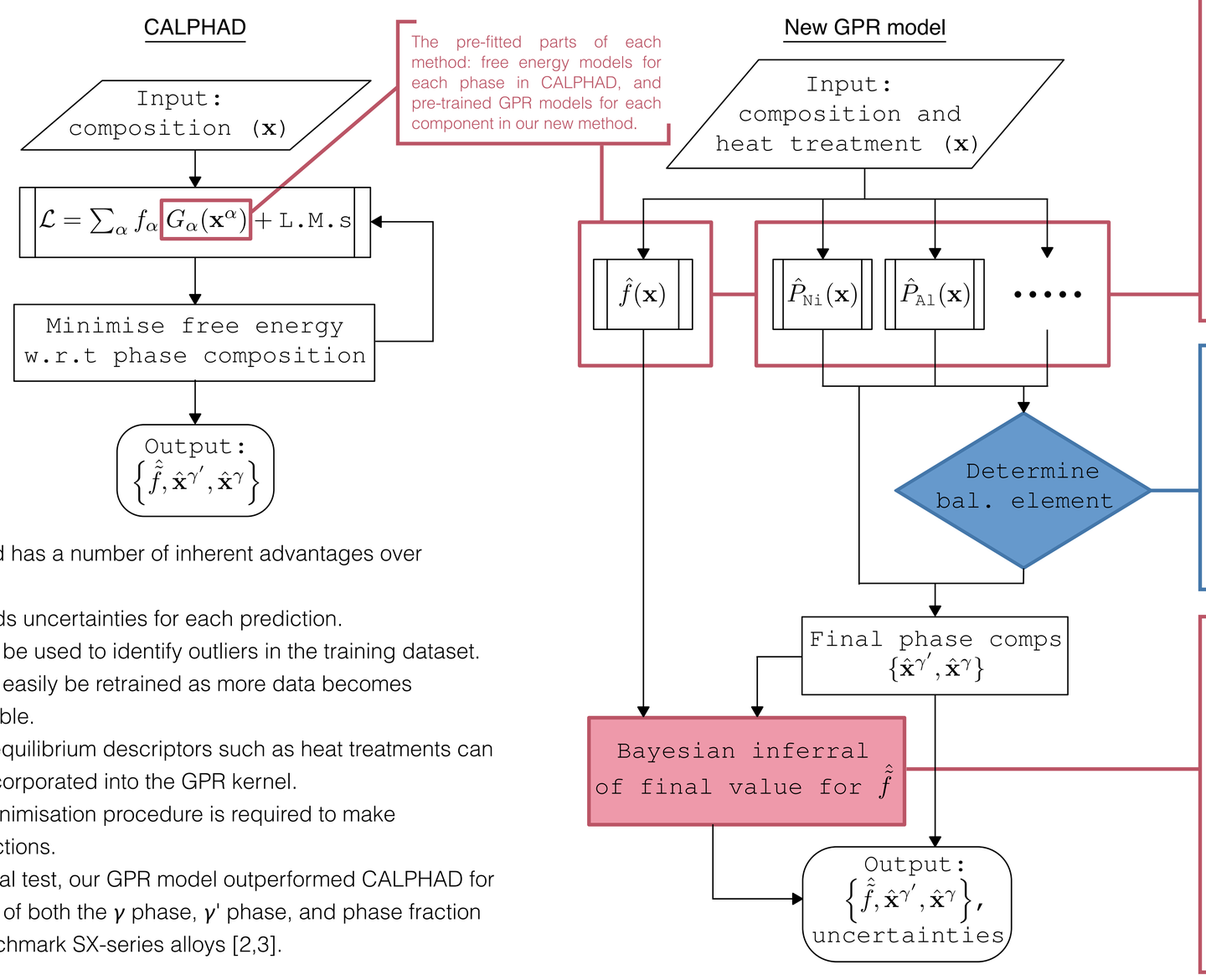
$$\frac{1}{N_{\mathrm{bins}}} = \text{Bin area}$$

$$n_m = \#\text{ data points in bin}$$



Method

Our approach to modelling alloy composition is to use gaussian process regression (GPR) models of the phase fractions and the log partitioning coefficients for each element in each phase [2].



Our method has a number of inherent advantages over CALPHAD:

- It yields uncertainties for each prediction.
- It can be used to identify outliers in the training dataset.
- It can easily be retrained as more data becomes available.
- Non-equilibrium descriptors such as heat treatments can be incorporated into the GPR kernel.
- No minimisation procedure is required to make predictions.

In a practical test, our GPR model outperformed CALPHAD for predictions of both the γ phase, γ' phase, and phase fraction of four benchmark SX-series alloys [2,3].

Each component of the alloy microstructure has its own GPR model. Interpreting the GPR kernel as a measure of alloy similarity, we are able to incorporate some simple physics into the model. To begin with we split each alloy's descriptor into a composition and heat treatment vector, $\mathbf{x}_c$ and $\mathbf{x}_h$. We then use the following kernel:

$$k(\mathbf{x},\mathbf{x}') = b + a_0\,\mathbf{x}_c\cdot\mathbf{x}_c' + a_1 e^{-\|B(\mathbf{x}_c-\mathbf{x}_c')\|^2/l_1^2}\cdot e^{-\|\mathbf{x}_h-\mathbf{x}_h'\|^2/l_2^2}$$

Most of the variance is due to composition. We capture this with a simple linear kernel term.

The heat treatment is coupled with a composition term to capture its influence on alloy similarity. In this context we can interpret the product as an AND operation [4].

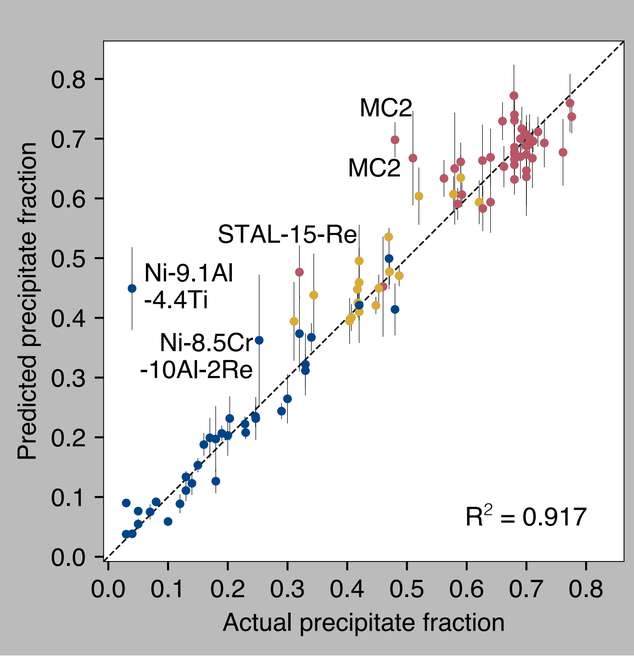
An obvious physical constraint is that the amounts of each element in each phase must sum to 1. We enforce this by making a dynamic choice of balance element for each prediction. This makes use of the uncertainty in the composition predictions. Compared to using Ni as a fixed balance element, the dynamic method improved the predictions not only for Ni but in fact for all elements.

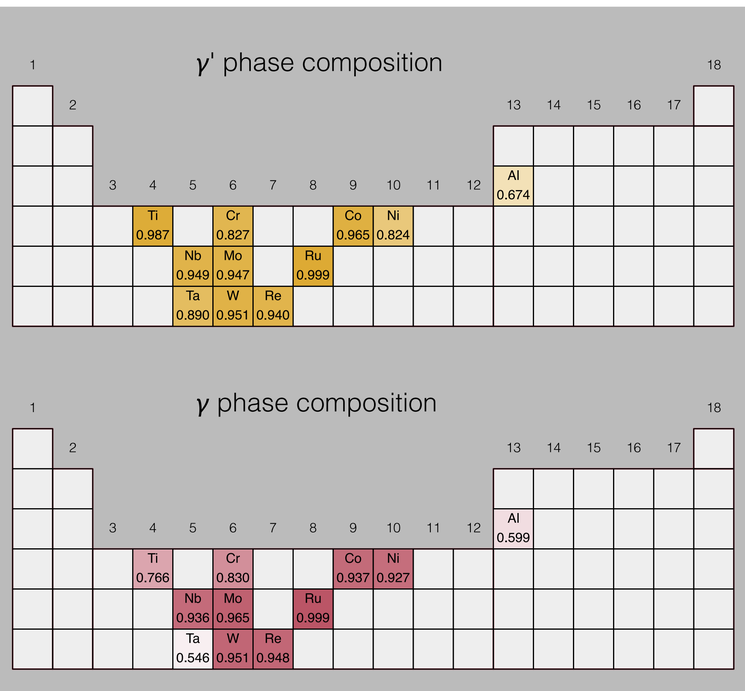
A further physical constraint is that the total amount of an element in the alloy must equal the input composition—for two-phase superalloys this means that:

$$f x_i^{\gamma'} + (1-f)x_i^{\gamma} = x_i$$

Due to how the experimental data that we train on has been collected, this equation typically doesn't hold exactly. For this reason we treat it as a probabilistic constraint rather than a hard one. To do so the GPR prediction for the phase fraction f is taken as a prior and the equation above is rearranged into a likelihood, allowing for a Bayesian inferral of a final value for f.

Results





Left: cross-validation predictions for the γ' phase fraction. Colours are for single crystal, polycrystalline, and lab alloys. **Right:** R^2 values for individual phase components, colour coded such that darker is better.

References

[1] U.R. Kattner, Tecnol. Metal. Mater. Min., **13** (1), 3-15, (2016)

[2] P.L. Taylor and G. Conduit, Comp. Mater. Sci., **201**, 110916, (2022)

[3] S. Sulzer, M. Hasselqvist, H. Murakami, P. Bagot, M. Moody, R. Reed, Metall. Mater. Trans. A Phys. Metall. Mater. Sci., **51**, 4902-4921 (2020)

[4] D. Duvenaud, Automatic model construction with Gaussian processes, [Thesis: Cambridge] (2014)

EPSRC Engineering and Physical Sciences Research Council

DASSAULT SYSTEMES

THE ROYAL SOCIETY