Does your model know its limits?

Leveraging uncertainty estimates in compound prioritisation

Matthew Segall, Samar Mahmoud, Charlotte Wharrick, Bailey Montefiore
Optibrium Limited, Cambridge, UK

Introduction

Accurate uncertainty quantification is critical for the effective, reliable application of *in silico* models in drug discovery. If a model's estimates of prediction uncertainty correlate with the observed accuracy of predictions, this means that low uncertainty predictions can be used with confidence in decision-making. The applications of uncertainty estimates can go far beyond this: identifying unlikely experimental results that can reveal false negatives and missed opportunities, revealing novel high-potential chemical space where the model is extrapolating, and prioritising new experimental measurements that will add the greatest additional information to improve a model and extend its domain of applicability, i.e. active learning.

Estimating prediction uncertainties

Models should estimate the uncertainties in every prediction. Figure 1 shows an example of a Gaussian processes model [1].

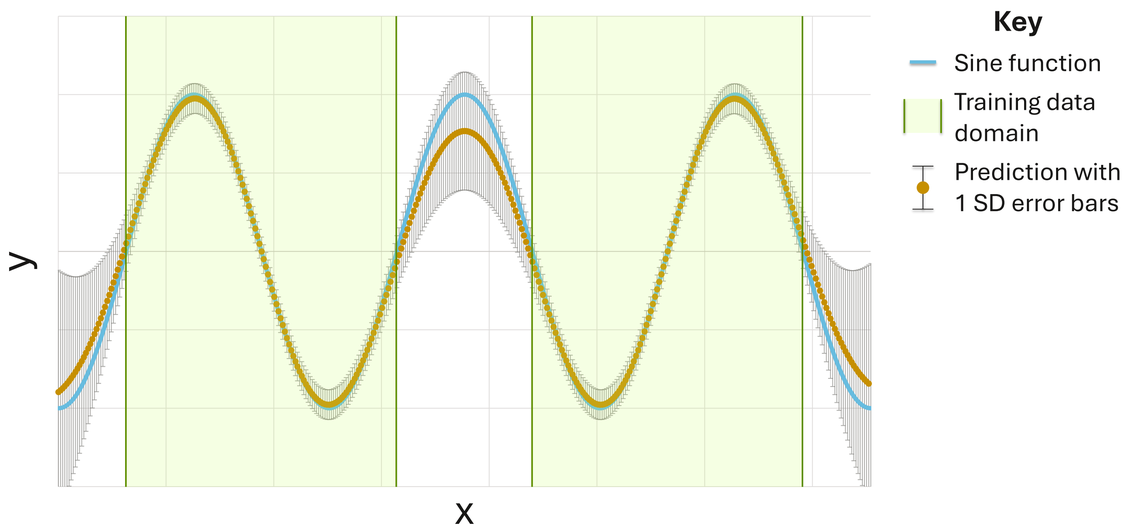

Figure 1. An example one-dimensional Gaussian processes model trained with data from $y=\sin(x) + \varepsilon$, where ε is a Gaussian noise term. Here, we can see that the model correctly identifies that the uncertainties increase when extrapolating outside the training data domain.

Assessing the quality of uncertainty estimates

Confident predictions should be more accurate on average. We can assess the calibration of uncertainty estimates using statistics such as the expected calibration error (ECE) [2], or visually, as shown in Figure 2.

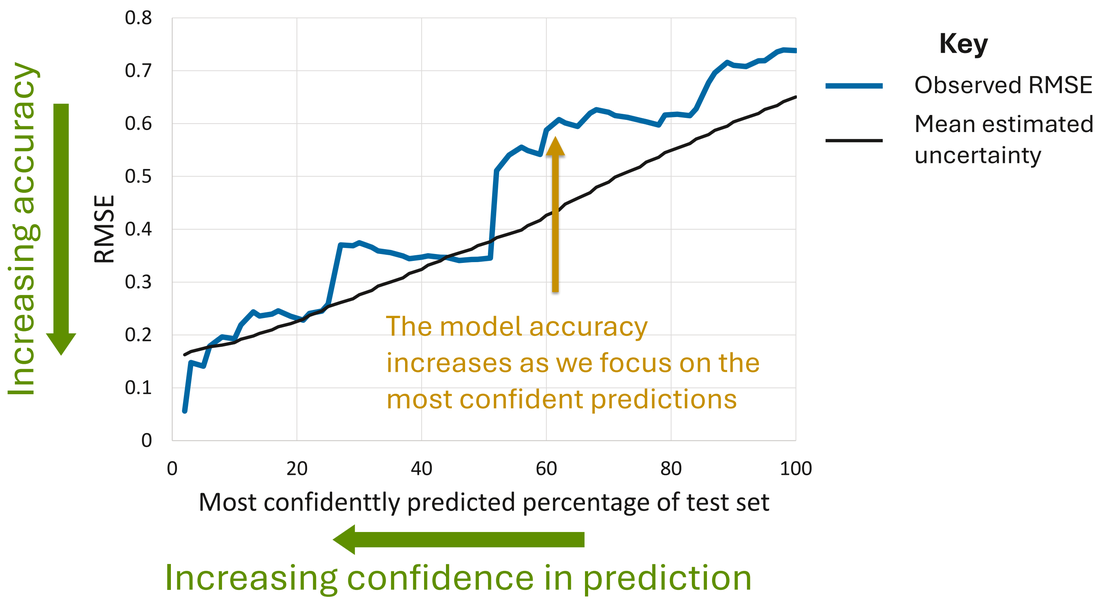

Figure 2. An example of a 'Rollback plot' generated with Cerella™ that demonstrates a strong correlation between a model's uncertainty estimates and the observed accuracy on a test set.

Choose your best compounds more effectively

Prediction uncertainties provide valuable information for compound selection. Figure 3, shows an example using Cerella™ [3]; the model has a poor overall correlation but can still **effectively prioritise compounds with low *in vivo* clearance** when prediction uncertainties are taken into account.

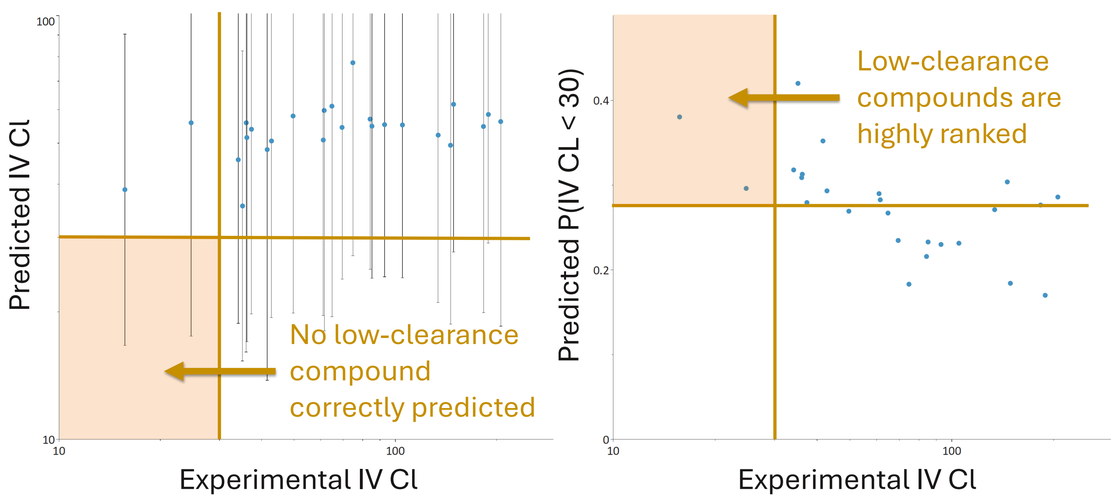

Figure 3. The predictive model of *in vivo* clearance has a low correlation on the test set (Left). However, the model has higher confidence in predictions of high clearance, meaning that it is very effective at prioritising low-clearance compounds (Right).

Rescuing missed opportunities

A confident prediction that differs significantly from an experimentally measured value can indicate an experimental error, as illustrated in Figure 4.

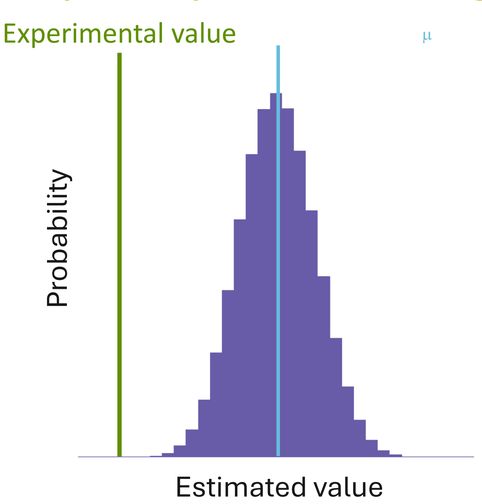

Figure 4. An illustration of a potential false negative. Here, the predicted value (light blue) differs significantly ($p<0.05$) from the measured value (green) when accounting for the probability distribution of the predicted value (purple).

Retesting potential false negatives can rescue opportunities hidden in our existing data. Figure 5 shows an example of such an application.

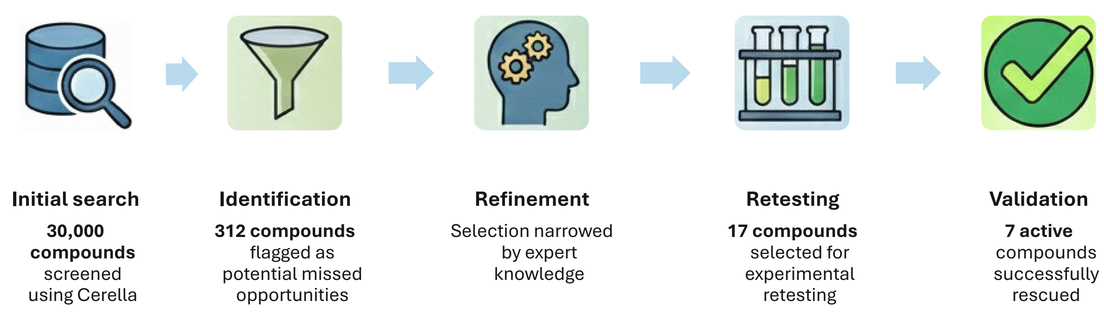

Figure 5. An example of identifying false negatives in a database of 30,000 compounds.

Recovering valuable compounds that were incorrectly discarded delivers higher returns from existing compounds and data than synthesising and testing new molecules. For example, an experimental screen to find seven new active compounds would typically cost **over $70,000***

Reduce synthesis and testing costs with active learning

Accurate uncertainty estimates enable a model to identify the most valuable experimental measurements to improve its predictions. Figure 6 shows the impact of a single active learning step with Cerella™ on identifying potent compounds in a novel chemical space. A small up-front investment yields a highly predictive model and a **45% reduction in synthetic and experimental effort** compared with the project's chronological progression.

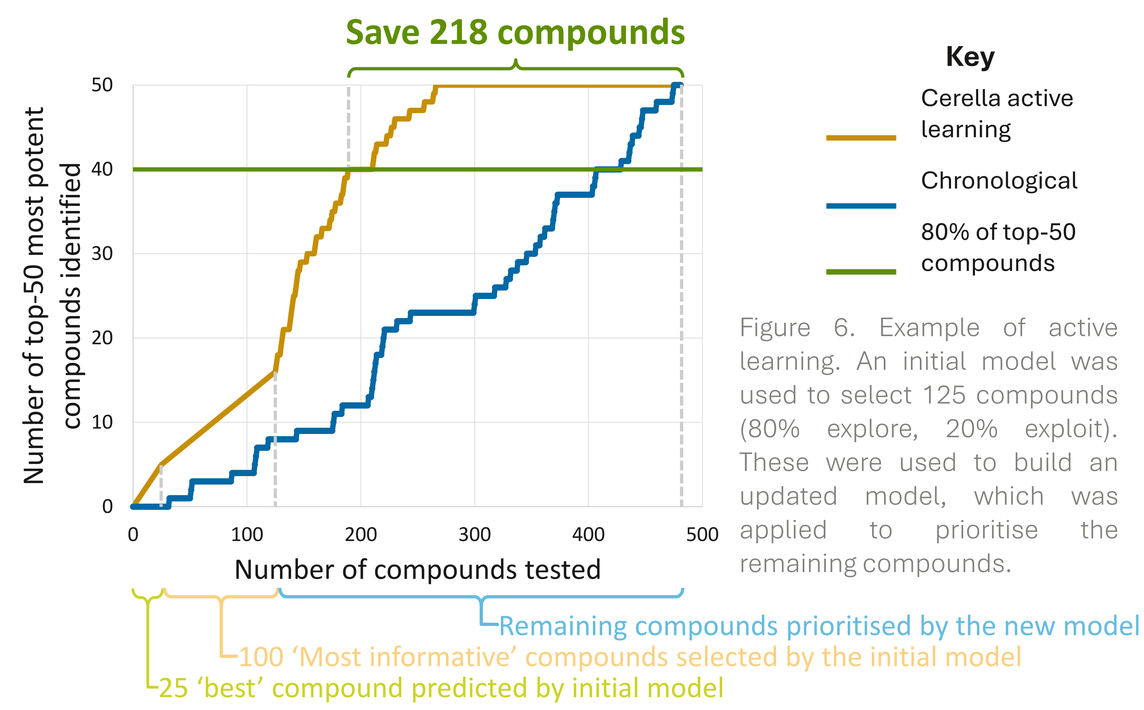

Figure 6. Example of active learning. An initial model was used to select 125 compounds (80% explore, 20% exploit). These were used to build an updated model, which was applied to prioritise the remaining compounds.

References

[1] J. Chem. Inf. Model. (2007) 47 (5): 1847–1857
[2] https://arxiv.org/html/2501.19047v1
[3] https://optibrium.com/products/cerella/

* Assuming a 1% hit rate and a screening cost of $100 per compound