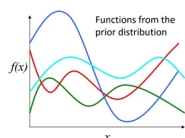


Ed Champness gave this presentation at the ACS Spring National Meeting & Exposition held in San Diego, USA on 16<sup>th</sup> April 2016.

## Abstract

A quantitative structure-activity relationship (QSAR) model is a mathematical function of molecular descriptors. The parameters of this function are found by maximizing the fit of this function to the observed activities of a training set of compounds, using a statistical or machine learning method. Following validation of the resulting model, most methods for estimation of the uncertainty in a prediction focus measures of the 'domain of applicability' or 'distance to model' to identify new compounds that differ significantly from the training set and hence for which the confidence in a prediction will be low.



In contrast to this, the Gaussian Processes method estimates a probability distribution over possible models that fit the observed data. The predicted value for a new compound is the mean of this distribution, while the standard deviation provides a well-defined estimate of the uncertainty for each individual prediction. This naturally takes into account the distance to the training set compounds and also identifies cases where variability in the training set data limits the ability to make a confident prediction, even if the new compound lies within the domain of applicability.

In this talk we will describe the Gaussian Processes method, discuss its strengths and weaknesses and compare its results with other QSAR modelling methods. This will be illustrated by several examples applications to different QSAR modelling problems.

\*With apologies to Douglas Adams for the deliberate misquote!

You can download this presentation as a [PDF](#).