British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

A dropout-regularised neural network for mapping arsenic enrichment in SW England

using MXNet

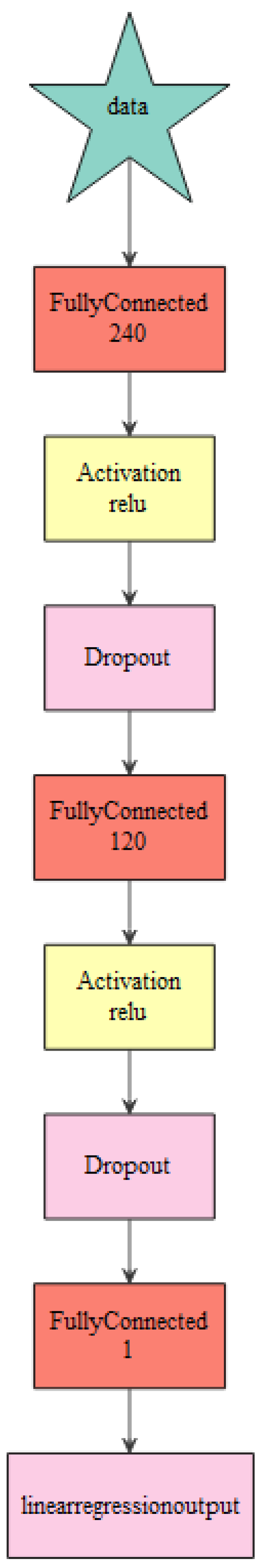
1. Why arsenic?

Mapping arsenic is important: at sufficiently elevated concentrations it presents a hazard to the health of humans and the environment. However, it's impact is not all negative: arsenic is also a useful guide when exploring for precious metal deposits. This is because arsenic behaves similarly to precious metals in hydrothermal mineralisation systems, but is generally more abundant. This abundance means that measurements can be made with a higher signal-to-noise ratio, providing better data from which to guide exploration.

However, accurately modelling and mapping the distribution of arsenic (and other hydrothermally mobile metals) is a challenge; these elements tend to have highly skewed distributions – high concentrations are rare occurrences which tend to be associated with 'nuggety' hydrothermal mineralisation governed by complex processes which are difficult to predict.

This poster presents a neural network solution to the modelling and mapping of arsenic in the stream sediments of south west England. By providing both point-sampled arsenic data and high resolution geophysical survey data to a neural network, a model is constructed through which arsenic concentrations are predicted using a combination of geophysical context and location. The output is a map in which geological and topographic features are captured, thus offering the user a detailed and accurate visualisation from which to perceive the controls on arsenic distribution.

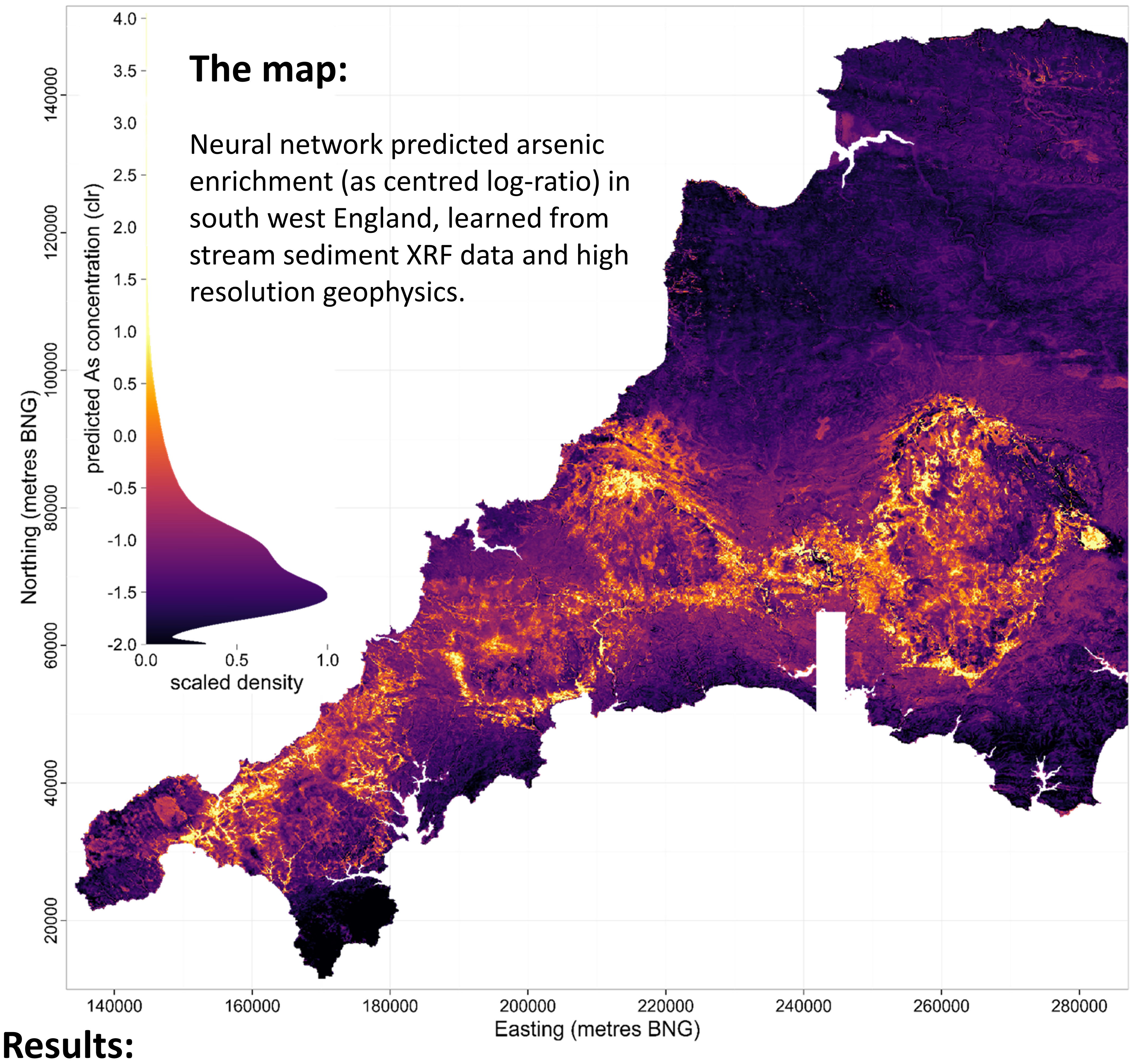
2. The network:



The optimal network architecture is dependent on the nature of the relationships within the data and true optimization is likely to require the use of evolutionary algorithms. For simplicity this study just used a network with two hidden layers of 240 and 120 neurons respectively (16 and 8 times the 15 predictor variables. See computation chart, left).

The network was trained on a dataset consisting of arsenic concentrations at 3395 sample sites, appended with location information (easting, northing and elevation) and a suite of geophysical variables from gravity, magnetic, radiometric and topographic surveys.

To prevent overfitting each hidden layer neuron was given a 10% chance of being temporarily removed from the network during each training iteration. This procedure, called dropout, forces the nodes of the network to learn independently of one another, improving generalisation. In addition, through per-epoch validation against a set of held-out test data, early stopping was used to prevent the network from overtraining.

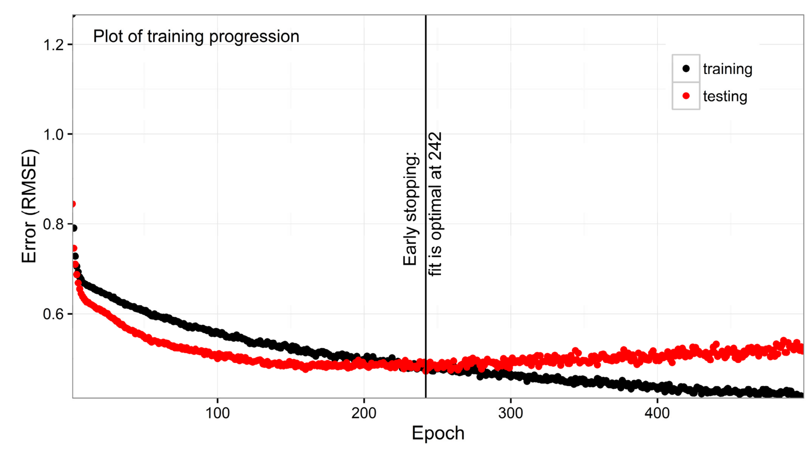


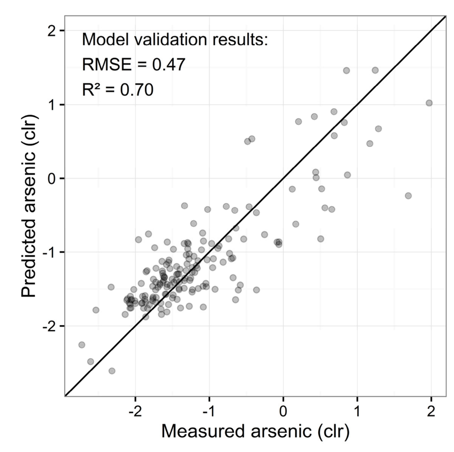
The map:

Neural network predicted arsenic enrichment (as centred log-ratio) in south west England, learned from stream sediment XRF data and high resolution geophysics.

3. Results:

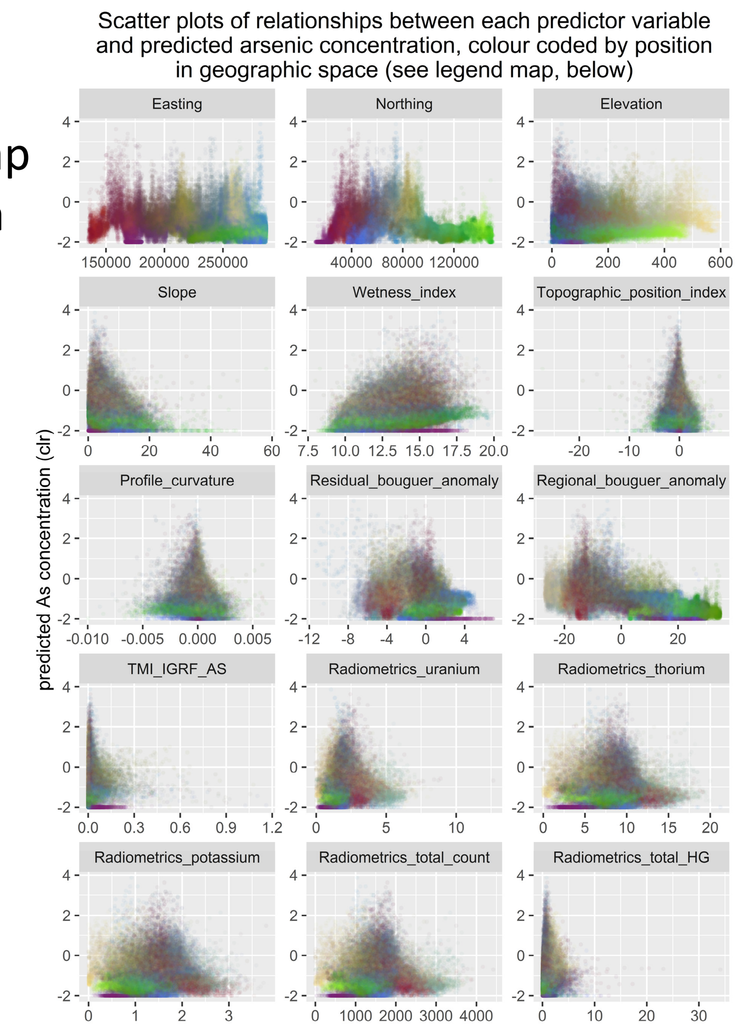
The best model, realised after 242 epochs, achieves an RMSE of 0.47 on the held-out test data, providing a 'real world' R^2 of 0.7 (see training and validation plots, below). This is a good outcome and indicates that the map has good accuracy. For interest, a random forest trained on the same data achieved a slightly inferior RMSE of 0.50.

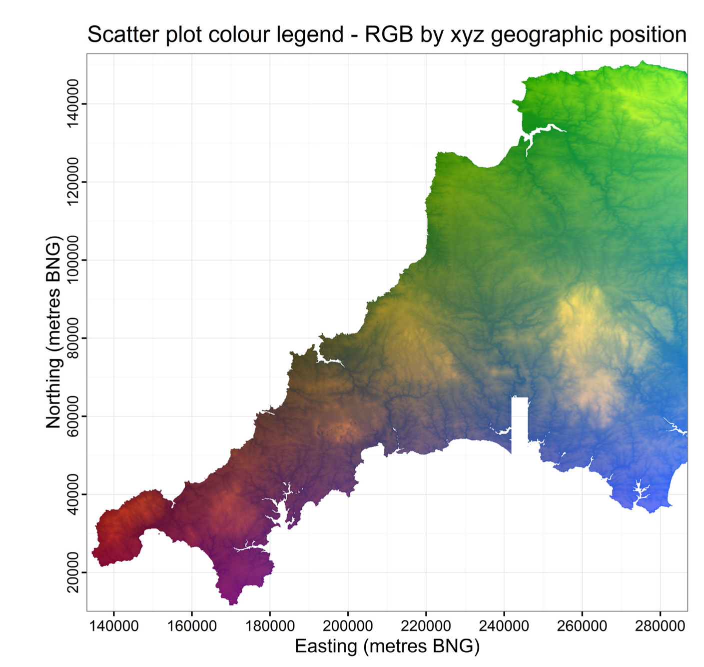




The predicted arsenic map (top) reveals the relationship between arsenic mineralisation and the granite intrusions of the Cornubian Batholith: arsenic has been concentrated by hydrothermal activity around the margins of the granites. Interestingly the network has recognised a particularly pronounced tendency for arsenic enrichment in stream channels around the Carnmenellis Granite in the far south west, the area which has historically been most heavily mined. The map also captures structural features and there is some evidence of variations in arsenic concentrations associated with different granite phases.

The scatter plots to the right give an idea of the complexity of the system that the network has learned. Despite complex interactions between predictor variables and arsenic concentrations the neural network delivers an impressively high degree of accuracy. Could even the most diligent human expert be expected to understand such a system at the level required to offer comparably accurate predictions?





British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, NG12 5GG, UK

Charlie Kirkwood - cwk@bgs.ac.uk