Supporting Information
Artificial Neural Network Analysis of the Catalytic Efficiency of Platinum Nanoparticles

Michael Fernandez*, Hector Barron, and Amanda S. Barnard

Molecular and Materials Modelling
Data61 CSIRO, Door 34 Goods Shed, Village St
Docklands VIC 3008 Australia.
E-mail: michael.fernandezllamosa@data61.csiro.au

1 Artificial Neural Networks (ANNs)

ANNs is a computer-based model in which a number of processing elements, also called neurons, units or nodes are interconnected by links in a netlike structure forming "layers".[1] A variable value is assigned to every neuron. The neurons can be one of three different kinds. The input neurons form the input layer, which receives their values by direct assignation and are associated with independent variables, with the exception of the bias neuron. The hidden neurons collect values from the input neurons, giving a result that is passed to a non-input neuron. Finally, the output neurons collect values from other units and correspond to different dependent variables, forming the output layer. The links between units have associated values, named weights that condition the values assigned to the neurons. There exist additional weights assigned to bias values that act as neuron value offsets. The weights are adjusted through a training process in order to minimize network error. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output.

2 Training the ANN models

To correlate the electron properties to the autocorrelation vectors, we used different machine learning techniques. The training and test sets were randomly generated from 50% and 50% of the data set, respectively. The training set was used to calibrated the machine learning models whilst the test set was used to test the prediction ability of the models.

The quality of the fit is described by an R^2 value where N is the number of nanoparticles, Y_i and P_i are the machine learning predicted and phenomenological model-derived catalytic activity and stability of Pt nanoparticle i, respectively. The average property value over all nanoparticles is given by $\bar{P}$. When computed on the training set, R^2 measures how well the model fits the phenomenological model data. To check for the possibility of overfitting, we applied a technique known as internal three-fold-out (TFO) cross-validation. Here we divide the training set into three subsets and then remove one while the other two are used to fit the regression model. The resulting model is then compared against the phenomenological model data for the left-out subset. This process is repeated until all the subsets have been validated against each other. Equation 1 is used to measure the fit of the model against each left-out subset except that we use the symbol R^2_{TFO} to distinguish from the case where the whole training set is considered. Different number of neurons in the hidden layer were evaluated and the network with the highest R^2_{TFO} was selected as optimum.

$$R^2 = \frac{\sum_{i=1}^{N} (Y_i - P_i)^2}{\sum_{i=1}^{N} (Y_i - \bar{P})^2} \tag{1}$$

3 Single scatter plots of the molar catalytic activity and stability of the Pt nanocatalysts

We first analyse the behaviour of the molar catalytic activity and stability of the Pt nanocatalysts as functions of the global structural features. Figure S1 illustrates that the correlation patterns of both properties strongly dependent on the configurations of the nanoparticle facets.

*michael.fernandezllamosa@data61.csiro.au

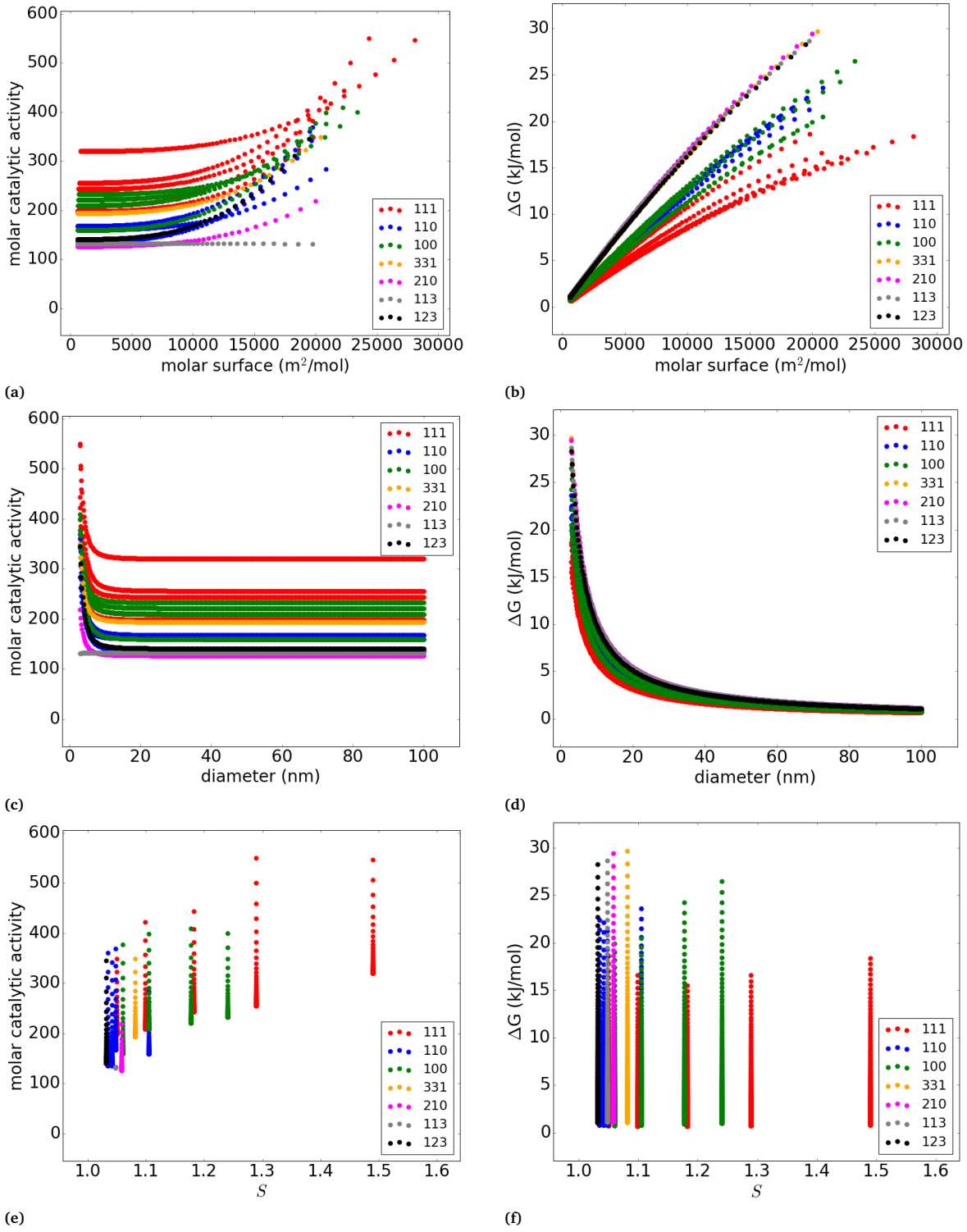


Figure S1 Scatter plots of the (a, c and e) catalytic activity and (b, d and f) stability at 25 °C of Pt nanoparticles as a function of (a and b) surface area, (c and d) diameter and and (c and d) sphericity.

4 Single scatter plots of the molar catalytic activity and stability of the Pt nanocatalysts

We first analyse the behaviour of the molar catalytic activity and stability of the Pt nanocatalysts as functions of the global structural features. Figure S1 illustrates that the correlation patterns of both properties strongly dependent on the configurations of the nanoparticle facets.

5 Learning curves of the machine learning models of the catalytic efficiency

Learning curve plots of root mean square errors of crossvalidation ($RMSE_{CV}$) for increasing number of training examples appear in Figure S2 for mean predictor (meanp), multilinear regression (MLR), ridge regression (Ridge), decision tree regression (Tree) and artificial neural networks (ANN) as described in[2]. These plots illustrate that the optimum predictions are yielded by the ANN models.

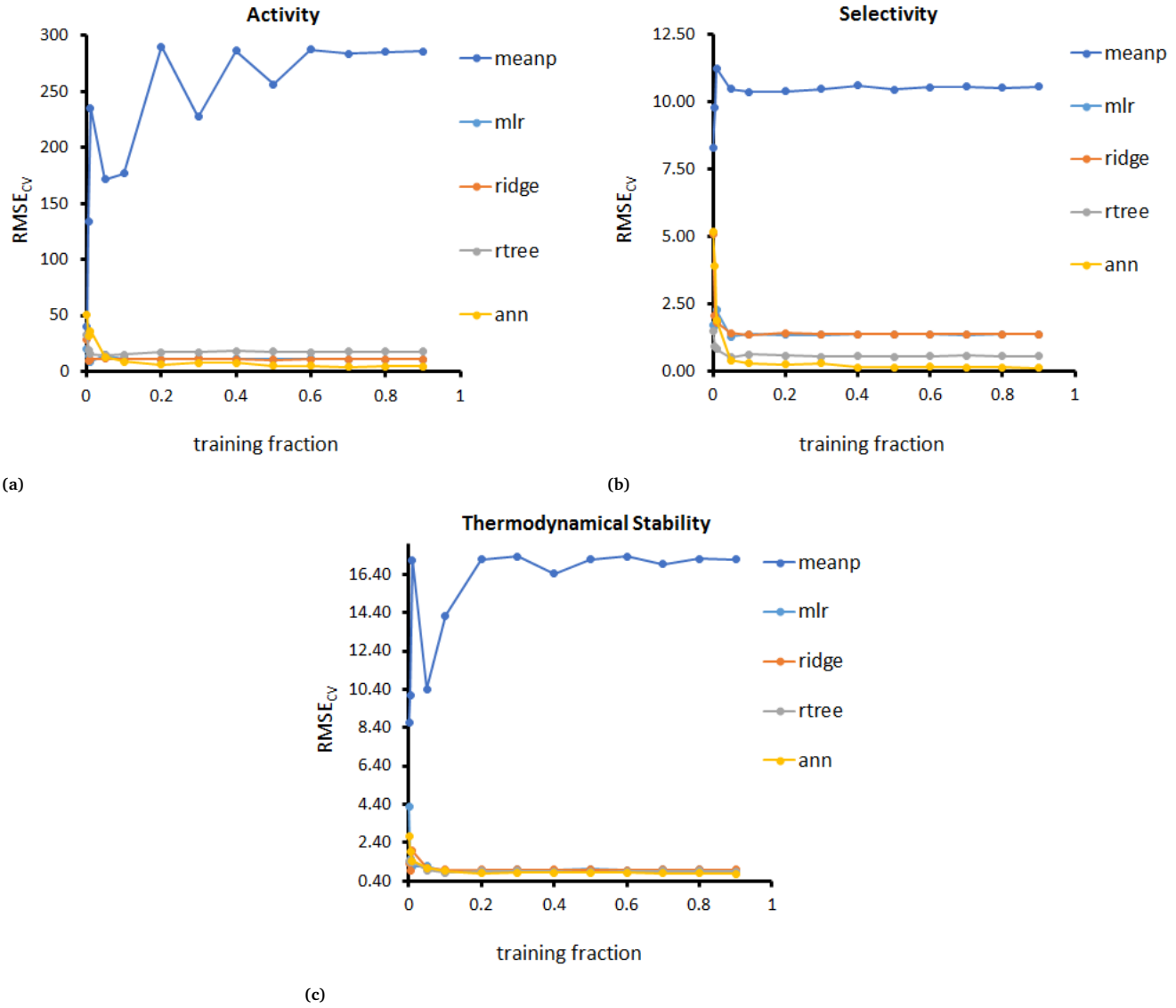
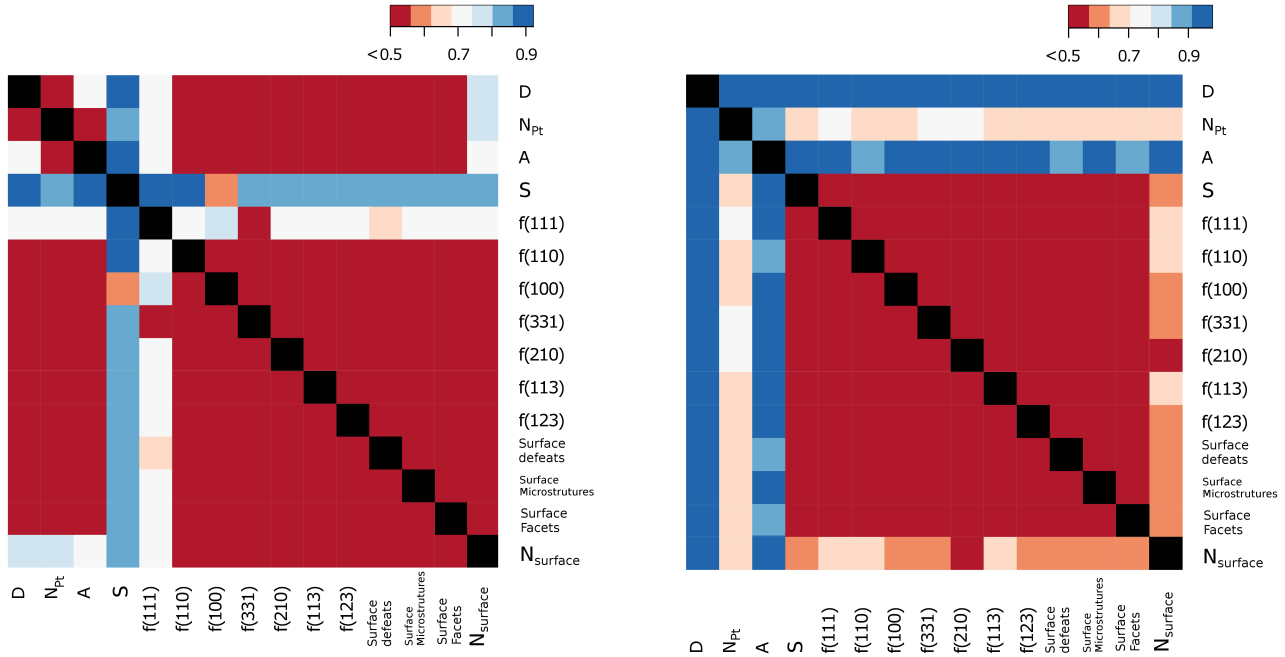


Figure S2 Learning curves of the (a) catalytic activity, (b) selectivity and (c) stability at 25 °C of Pt nanoparticles

6 Accuracy of two-variable models of molar catalytic activity and stability of Pt nanoparticles

Heatmaps of cross-validation accuracies of the two-variable models of molar catalytic activity and stability of Pt nanoparticles in Figure S3.



(a) (b)

Figure S3 Heatmaps of the accuracy of the two-inputs ANN models of the catalytic activity (a) and stability (b) of the Pt nanoparticles at 25 °C.

7 Decision tree model of the thermodynamical stability

The DT analysis yields a simpler model of molar stability in Figure S4 with only one fundamental rule accounting for the low thermodynamic stability of nanoparticles with less than 1000 total atoms in the surface.

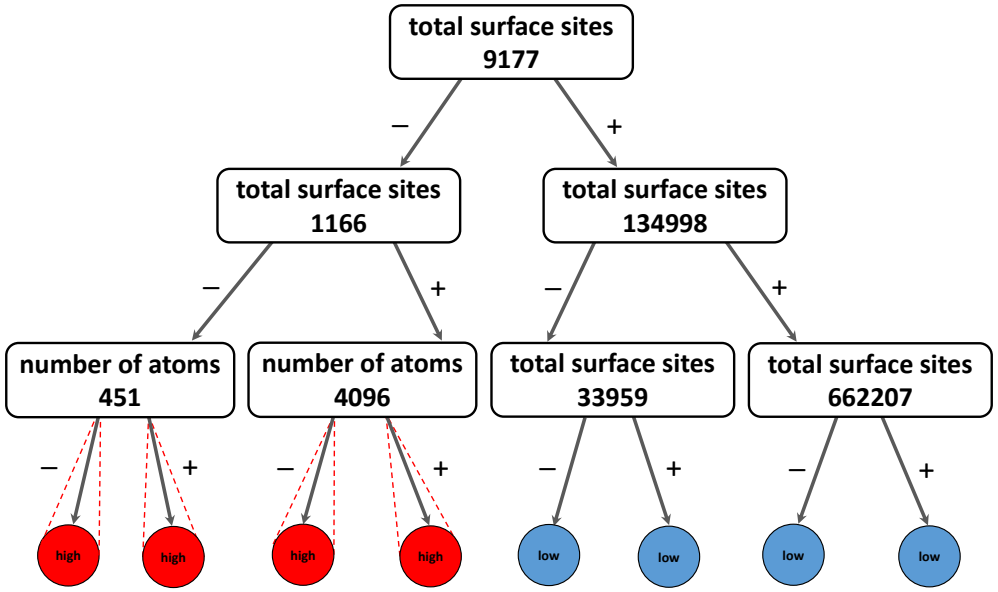


Figure S4 Binary DT model of the molar thermodynamical stability at 25 °C of the Pt nanoparticles.

References

[1] Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press, USA, 1995.

[2] Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Muller, K.-R. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.