Supporting Information for "Deep ocean learning of turbulence"

Ali Mashayek[1], Fangming Zhai[1], Nick Reynard[1], Adam Jelley[2], Colm Cauldfield[3], Alberto Naveira Garabato[4]

[1]Imperial College London, UK

[2]University of Edinburgh, UK

[3]University of Cambridge, UK

[4]University of Southampton, UK

Contents of this file

1. Text S1 to S2
2. Supporting references
3. Figures S1 to S7

S1. Global Observational Surveys Figure S1a-d show the coverage of the global observational surveys that provide T, S, Z data (in addition to other fields) that can be used to infer estimates of turbulent mixing either through finescale parameterizations(Polzin et al., 2014) or through data-driven methods as in this work. These field programs do not contain direct turbulent measurements. Of relevance to this work is the hydrographic surveying

component of these experiments, which provide high-quality conductivity-temperature-pressure profiles to construct a climatological temperature-salinity-depth database. Figure S1e,f show the high-resolution topography data that is key to turbulence prediction, due to the importance of the bottom boundary in generating propagating waves as well as non-propagating boundary turbulence. Gravity data provides coarser topographic information than the direct echo-sounding surveys, which cover only 30% of the seafloor, but are extending their coverage at an accelerating rate. High-resolution seafloor mapping is also commonly provided by deep-ocean surveying research cruises, and is integrated in global topographic data (e.g. https://www.gebco.net/https://www.gebco.net/).

S2. Neural Network Architectures. Figure S2 shows a schematic of the set ups of the neural networks used in this study. A series of fully connected layers are applied with batch normalisation and Leaky ReLu activation applied to the output of each layer. The number of neurons of the fully connected layers increase from 9 in the input to a maximum hidden layer size of 192 before reducing from 192 to 96 to 48 to 1 for the output. A mean squared error (MSE) loss function is applied to the output with linear activation. An Adam optimiser with linearly decreasing learning rate initialised at 0.005 was applied for 50 epochs (to convergence). Around $\sim$ 161000 data samples (standard rescaled, in batches of 128) were used to train the network, which lies at the low end of data size typically needed for deep learning.

Adding convolutional layers between the input and fully connected layers in an attempt to incorporate local information was also investigated. In order to input the oceanographic data into a CNN, the 1D profiles were first reshaped into an empirical 9×9 'images'. Given

that $\sim$ 161000 data samples were used here, $\sim 161000/9 \approx 18000$ images were constructed and input to the CNN, which is similarly at the low end of the data requirements for deep learning. Following Figure S2, a series of 3×3 convolutional filters (with a stride of 1 and padding set to 0), batch normalisation, Leaky ReLu functions and 2×2 max pooling filters are applied so that the eventual activation reaches a dimension of 2×2×96. As mentioned in the main text, the performance of CNN was not superior to that of FNN and so FNN is discussed in the main text.

A number of other network architectures were investigated prior to focusing on the ones just described. A number of activation functions (including both standard ReLu and clipped ReLu) were investigated, along with a range of other key hyperparameters such as the learning rate and number of epochs. Figure S7 shows the sum of MSE for 8 model designs, against the number of iterations during the training of the neural network. At the end of the neural network training, there is no significant difference in the sum of the MSE between the model designs, although the MSE decreases at a faster rate and is more stable for the FNN compared to the CNNs.

References

Davis, R. E., Talley, L. D., Roemmich, D., Owens, W. B., Rudnick, D. L., Toole, J., ... Barth, J. A. (2019). 100 years of progress in ocean observing systems. *Meteorological Monographs*, *59*, 3–1.

GEOTRACERS. (2019). Geotracers. *https://www.geotraces.org/*. Retrieved from `https://www.geotraces.org/`

GO-SHIP. (2018). Go-ship. *http://www.go-ship.org/*. Retrieved from `http://www.go`

-ship.org/

Gouretski, V., & Koltermann, K. P. (2004). {WOCE} global hydrographic climatology. *Berichte des BSH*, *35*, 1–52.

Polzin, K. L., Garabato, A. C. N., Huussen, T. N., Sloyan, B. M., & Waterman, S. (2014). Finescale parameterizations of turbulent dissipation. *Journal of Geophysical Research: Oceans*, *119*(2), 1383–1419.

Sandwell, D. T., Müller, R. D., Smith, W. H., Garcia, E., & Francis, R. (2014). New global marine gravity model from cryosat-2 and jason-1 reveals buried tectonic structure. *Science*, *346*(6205), 65–67.

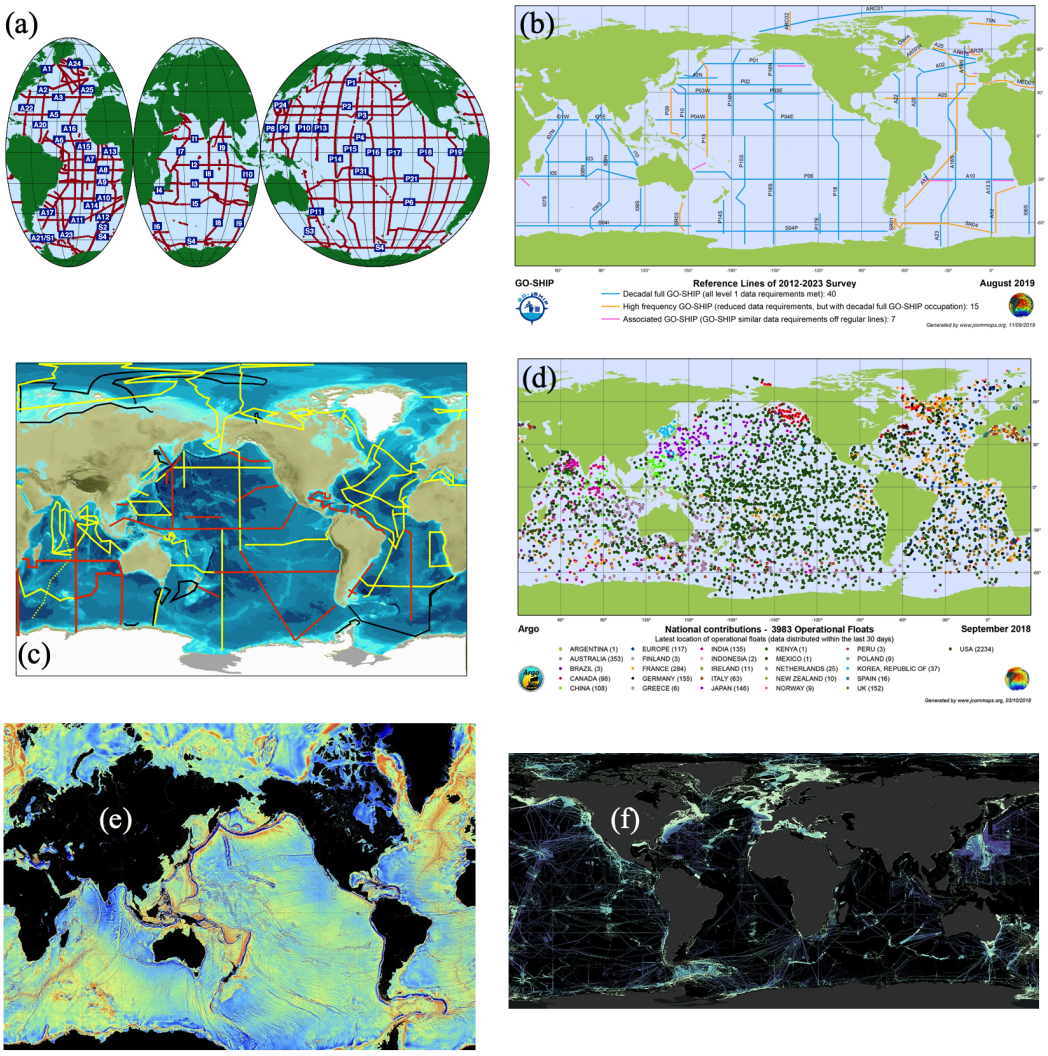

Figure S1. International surveys have provided invaluable hydrographic and bathymetric information required to quantify oceanic turbulent processes. (a) WOCE Hydrographic Programme survey stations [1985–97](Gouretski & Koltermann, 2004; Davis et al., 2019). (b) GO-SHIP hydrographic sections [GO-SHIP 2018](GO-SHIP, 2018; Davis et al., 2019). (c) GEOTRACES sections [from 2018](GEOTRACERS, 2019; Davis et al., 2019). (d) Global Argo array coverage [as of 2018](Davis et al., 2019).(e) Satellite-measured marine gravity, revealing the ocean bathymetric features.(Sandwell et al., 2014) (f) Black regions represent the areas yet to be measured with echo-sounders, whereas lines represent already sampled regions (One-fifth as of 2020) [from NIPPON FOUNDATION-GEBCO SEABED 2030 PROJECT].

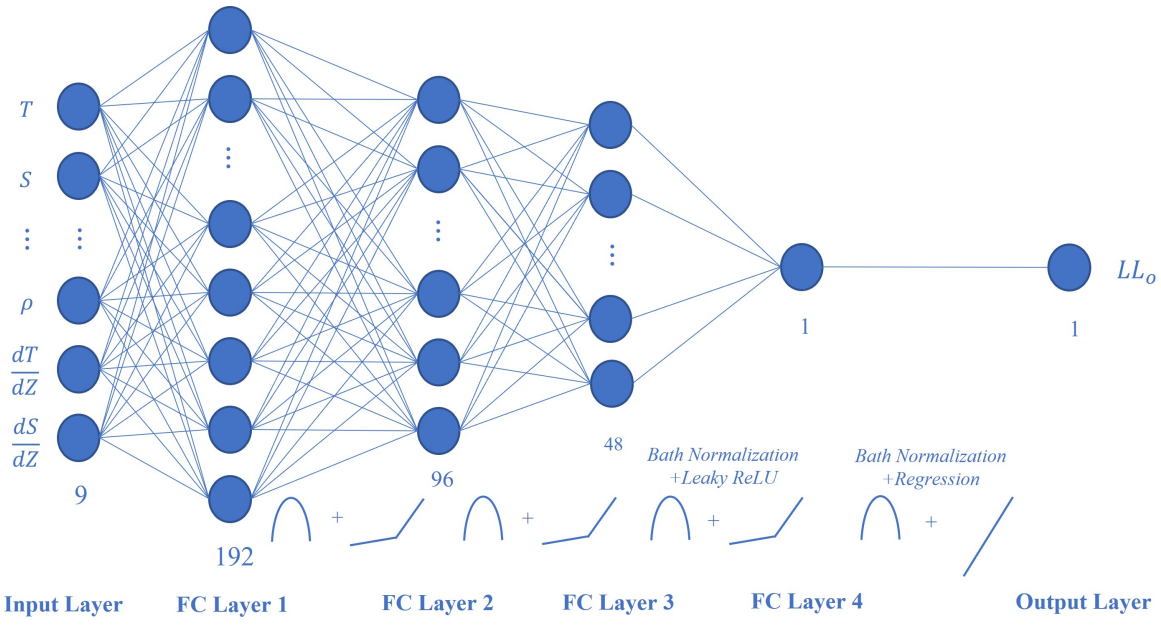

Figure S2. A schematic representation of the Feedforward Neural Network (FNN) designed in this study for the deep-learning approach.

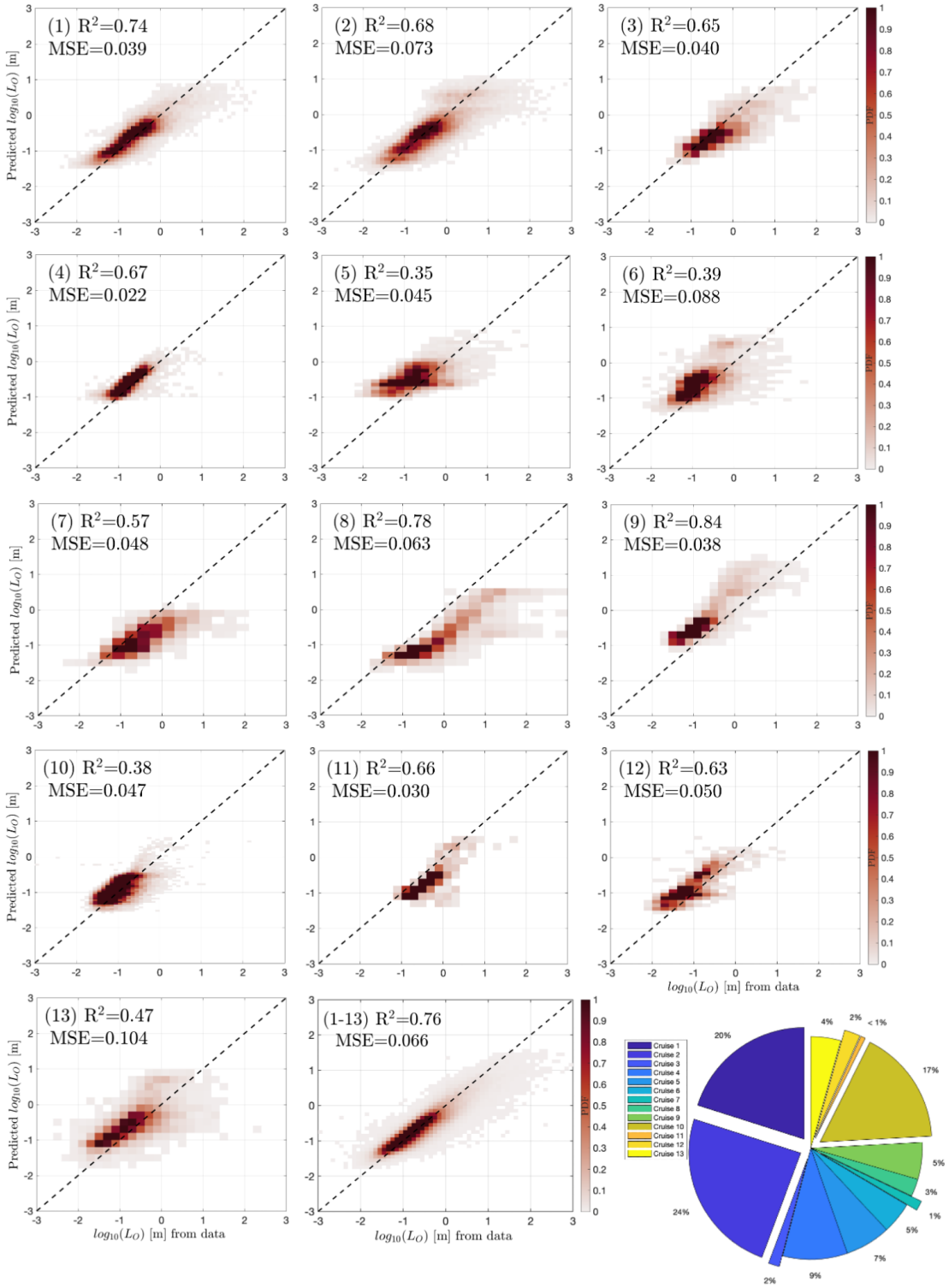

Figure S3. Similar to Fig. 3a in the main text, but for individual experiments. As in Fig. 4 in the main text, for each experiment, a bagging tree model was trained based on the data from all other 12 experiments, **excluding** the data from the given experiment itself, to avoid overfitting.

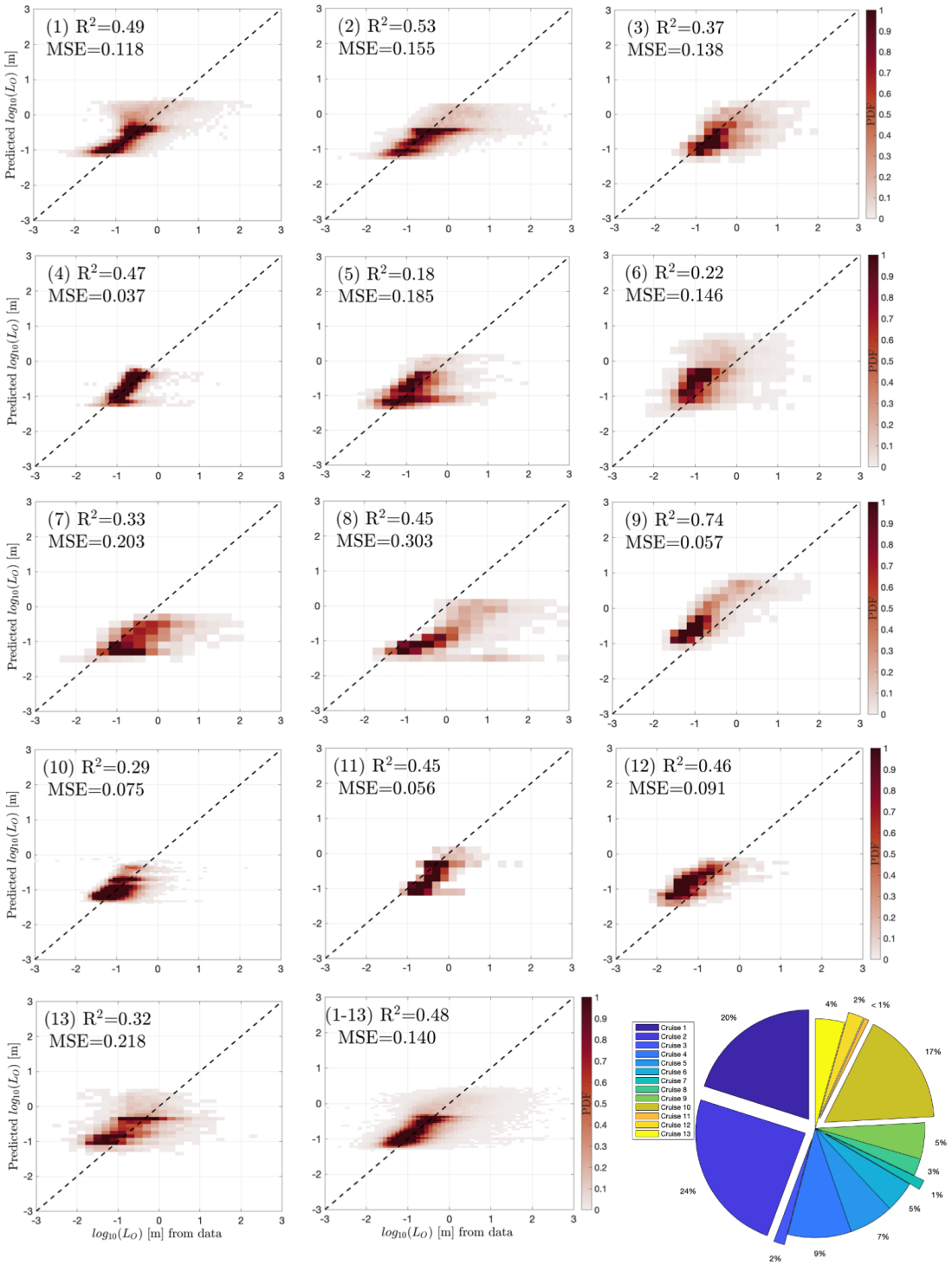

Figure S4. Similar to Fig. 3b in the main text, but for individual experiments. As in Fig. 4 in the main text, for each experiment, a CNN model was trained based on the data from all other 12 experiments, **excluding** the data from the given experiment itself, to avoid overfitting.

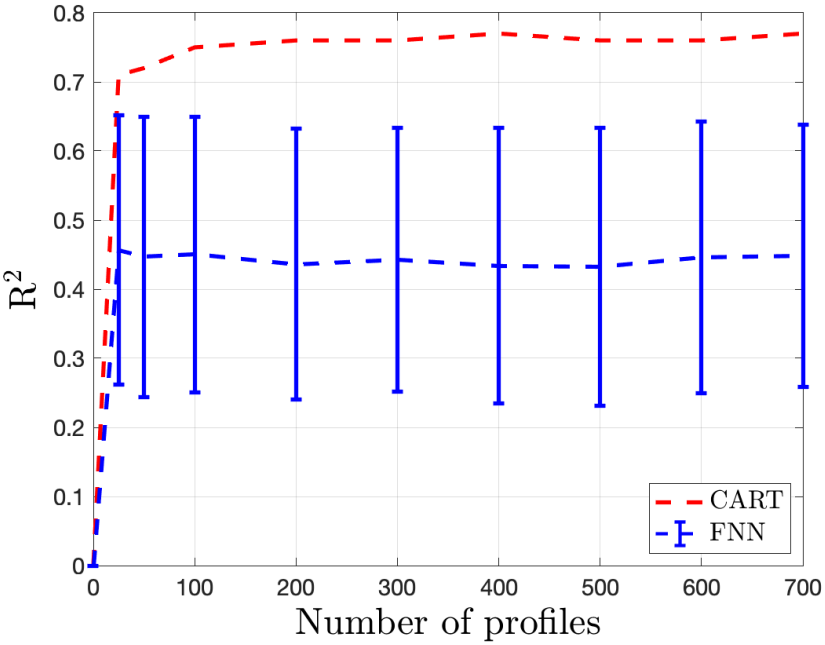

Figure S5. Calculated R^2 values for trained models as a function of the number of samples used to train them for both CART and FNN. For the FNN, for each iteration number, the model was trained 500 times and the mean and the standard deviation are shown with the dashed blue line and its associated errorbars.

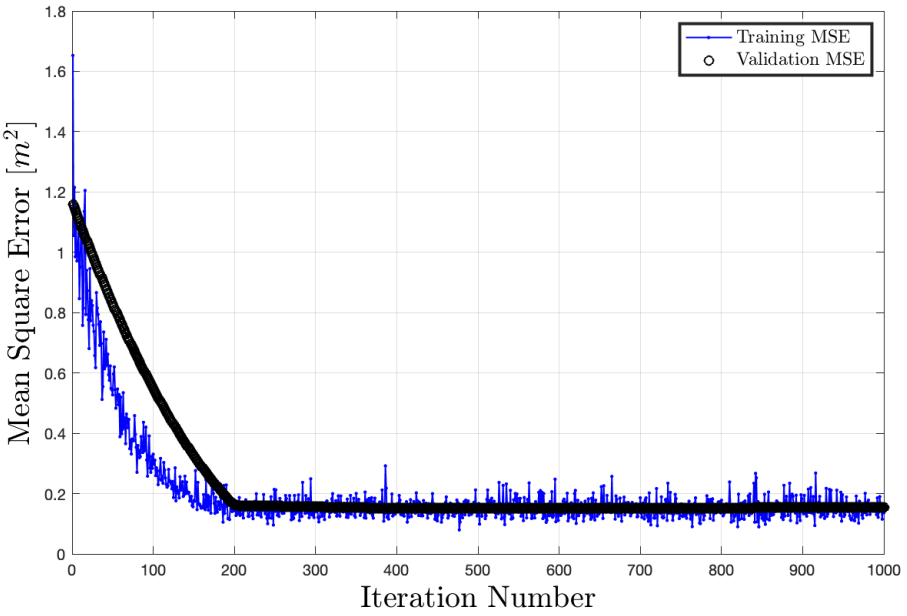

Figure S6. The Mean Square Error (MSE) of the *FNN* model, as a measure of regression Loss, for both the training and validation data. This plot is based on the BBTRE dataset.

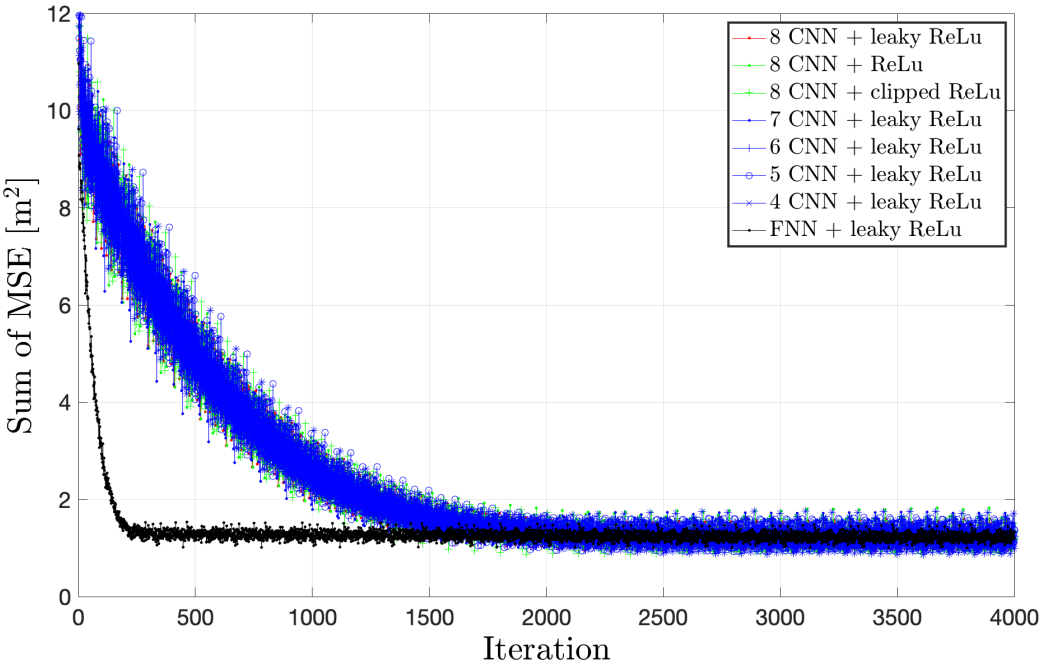

Figure S7. The MSE for 8 model designs, against the number of iterations during the training of the neural network.