APPENDIX E: SUPPLEMENTARY MATERIAL FOR CHAPTER 4

E.1: Summary of Neural Network Applications

E.2: Supplementary Figures

E.3: Discussion of fit metrics for comparing NN and MLR*

*In main body of thesis

APPENDIX E.1:
Summary of Neural Network Applications

Table E.1: Notes on important modelling decisions and rationale (where discussed) for applications of neural networks, with a focus on regression-type problems in marine ecology.

Reference	Inputs	Outputs	Number of Hidden Nodes	Train/Test Procedure	Activation & Error Functions, etc.	Scaling & Transformation	Weights & Variable Importance	Goodness of Fit	Comparison
Lek et al., 1995 Marine and Freshwater Research	6 and 8 Characteristics of the fish: asymptotic weight of the species, morphological ratio, mean annual temperature, 3 discrete diet variables Additional: ratio of standard length/height of the body (D), ratio of total height of the tail/height of the body (P)	1 Annual consumption of food relative to biomass of fish species (Q/B)	8 (Rationale not discussed)	Backpropagation 108 observations Random division of observations: Training- 80% Test- 20% Model fit using training data, and predictive performance assessed on test set. Stopped training: 2000 iterations (because R reached values above 0.9 after 1000) Procedure repeated three times	Activation function: Sigmoid Error function: Not discussed	NN: No variable transformations Inputs: standardized by mean and standard deviation (to make the scales of measurement uniform) Output: Converted to the range [0,1] MLR: Log transformation of the dependent variable & 3 independent variables	Initial weights: random values on interval [-0.3, 0.3] Variable importance: Sensitivity analysis	Prediction error: $Q/B_{observed} - Q/B_{estimated}$ Adjusted-R^2 of observed vs. modelled output values, p-values Predictive Index: [(R_learn + R_test)/2]/abs(R _learn – R_test)	Compared NN to existing MLR equation with same variables (Palomares, 1991) In all cases the predictive performance was better for NN than for MLR
Baran et al. 1996 Hydrobiologia	11 Physical stream characteristics: Width, gradient, mean depth,	1 Brown trout abundance (biomass or density)	Number not stated	Backpropagation 220 observations First trained using all	Activation: Sigmoid Error: SSE	NN: Inputs: standardized (to obtain similar range of variation)	Not discussed	Adjusted-R^2 of observed vs. modelled output values, p-values	Compared to MLR models using stepwise selection of variables

	coefficient of variation of depth, mean bottom velocity, Coefficient of variation of bottom velocity, Froude number, area of cover, area of shelter, deep water area, elevation			observations for comparison to MLR Next, tested predictive ability. Random division of observations 75%: Training 25%: Test Stopped after 1000 iterations		Output: converted to the range [0,1] (to adapt to the demands of the transfer function) MLR: Variables transformed to have best correlation coefficients (square and fourth root transformations to stabilize variance; Logarithmic and inverse transformations to normalize data)		For all observations and test values	R was significantly higher for the NN than MLR for all observations and for test values
Brey et al., 1996 Marine Ecology Progress Series	16 Biotic & abiotic variables: Binary: Vagile-sessile, epifauna-Infauna, carnivorous, omnivorous, herbivorous, lake, river, marine, Mollusca,	1 Annual production per biomass (P/B)	Not discussed Used NEURALWARE software, which "performs semi-automated data analysis, variable selection, and network construction, using elements of fuzzy logic and	Backpropagation 899 observations Random division of observations: Training: ~85% Test: ~15% Repeated procedure 10 times	Not discussed	NN: Transformed continuous variables to achieve more even distributions using Box-Cox algorithm MLR: transformed variables	Not discussed	R^2 of observed vs. modelled output values	Compared to "classical" MLR approach R^2 of NN were significantly higher than those of MLR. NN may be more useful when emphasis is on predicting rather

	Crustacea, Polychaeta, Echinodermata, Insect larvae Continuous: water depth, temperature, mean body mass		genetic algorithms.”	The 10 NN differed in number and type of input variables selected.		according to theoretical considerations and to empirical evidence			than relationships between dependent and independent variables.
Lek et al., 1996 Ecological Modelling	10 Habitat variables: wetted width, area with suitable spawning gravel, surface velocity, water gradient, flow/width, mean depth, standard deviation of depth, bottom velocity, standard deviation of bottom velocity, mean speed/mean depth	1 Density of brown trout spawning sites/linear meter of steam bed	8 Varied number of hidden nodes and repeated train/test procedure 5 times and recording the average R. No improvement after 8 hidden nodes	Backpropagation 250 observations First trained using all observations for comparison to MLR Next, tested predictive ability. Random division of observations: 75%: Training 25%: Test Procedure repeated 5 times Stopped training: 1000 iterations (when SSE & R stabilized)	Activation: Sigmoid Error: SSE	NN & MLR: Inputs: standardized (to standardize the scales of measurement) NN: Output: Standardized and then converted to the interval [0,1] (to adapt to the transfer function) MLR: Output: Standardized and nonlinearly transformed Developed models using untransformed variables and then transformed variables	Sensitivity Analysis	R and R^2 of observed vs. modelled output values Slope of the regression between observed and modelled output values Analysis of residuals	Compared predictive capacity of NN and MLR Developed models using stepwise regression and all variables Unlike MLR models, a clear improvement of the NN results was obtained using the raw data and by including additional variables. NN are clearly more performant than MR

Scardi, 1996 Marine Ecology Progress Series	3 - 5 Surface irradiance, mean chlorophyll concentration, depth of photic zone, light extinction coefficient, station depth, bay (binary variable of location)	1 Phytoplankton production (PP)	5 – 8 Compared performance of NNs with 1 to 12 hidden nodes. Chose number with the lowest MSE (varied depending on number and type of input variables Performances of the NN paper should be considered minimal estimates because training was intentionally limited & further improvement is certainly possible (e.g more epochs, different network initialization, etc.)	Backpropagation 27 observations Random division of observations (with replacement): 50% - Train 100% - Test Training stopped after 50,000 epochs. Weights corresponding to the lowest MSE were saved.	Activation: Sigmoid Error: Not discussed Simple network for comparison: Learning rate: Constant Momentum term: None	Inputs & Outputs: Converted to range [0,1]. Scaled using values larger than the maximum observed rather the range; training becomes easier if values are not too close to limits of the sigmoid function. Small positive offset added to raw PP data to avoid scaled values too close to 0.	Not discussed	R^2 of observed vs. modelled output values The meaning of R^2 as a measure of goodness of fit is not the same for linear models vs. NNs. For MLR: it has a unique, exact value; for NN it is affected by the random and arbitrary factors involved in network training.	Compared NN to two existing empirical models NN-based empirical models of PP were far more effective than linear empirical models (as evidenced by higher R^2 values)
Guegan et al., 1998 Letters to Nature	3 River characteristics + productivity: Surface of the drainage area, flow regime, and	1 Species richness (SR; global scale)	5 Best compromise between bias and variance (i.e. compromise between over and under-fitting)	Backpropagation 183 observations n-fold cross validation	Not discussed	Not discussed	Variable importance: Garson's algorithm	R^2 of observed vs. modelled output values, p-values	Some discrepancies with previous work (linear models) regarding variable importance

	net primary productivity								NN explained more of the variation in SR than previous linear models
Aoki and Komatsu, 1999 Oceanologica	15 Hydrographic, biological, and climatic variables	1 Catch of sardine recruits (proxy for recruitment index)	5 Chosen empirically as one-third of the number of input units	Backpropagation 19 observations Test data: 4 observations (~20%) from: (i) beginning, (ii) end, and (iii) steep increase	Activation: Sigmoid Error: Not discussed	Inputs: Converted to the interval [0,1] Output: Not discussed	Initial weights: random values on interval [-0.3, 0.3] Variable importance: 2-step weight analysis Used this analysis to reduce number of inputs and re-run model	Mean absolute error	N/A
Brosse et al., 1999 Ecological Modelling	8 Environmental variables: Depth, distance from bank, slope of the bottom, flooded vegetation cover, and percentage of boulders, pebbles, gravel, and mud	1 Fish density (for 6 different species) Chose to use six different models instead of one model with six outputs to facilitate analysis of variable contributions to each species	10 Empirically selected (lowest error in training and test sets with minimal computing time)	Backpropagation 306 observations Trained using all data Validated with n-fold cross-validation 10 models evaluated for each species	Activation: Not discussed Error: Not discussed	Inputs: Not discussed Outputs: Log10(x + 1) transformation (to reduce influence of outliers)	Garson's algorithm	R and R^2 of observed vs. modelled output values For cross-validation: Performance index (PI) Sum of squared errors (SSE) (did not want to use R or R2 because of the lack of high	Compared to MLR and generalized additive models (GAMs) Found that NN are more suitable for predicting fish abundance at the population scale than MLR Models' predictions improved with

								values of fish density)	GAM, which justifies the use of NN
Chen and Ware, 1999 Canadian Journal of Fisheries and Aquatic Sciences	5 Ecological and environmental variables lagged 3 years: Spawner biomass, predator biomass, mean annual SST, spring salinity, and summer salinity	1 Biomass of herring recruits	2 Evaluated 1 to 5 using fuzzy logic	Backpropagation 41 observations Test data: 5 observations (~12%) from (i) random, (ii) low, (iii) high, and (iv) medium periods of biomass Forecast data: last 4 years	Activation: Logistic (because most common) Error: SSE (because most common)	NN: Not addressed MLR: Logged all variables (to comply with regression assumptions of homogeneity and normality)	Trained networks with different starting values and noted different convergences. Used Ripley's regularization (weight decay) of error function to improve optimization Variable importance: 2 step weight analysis	R^2 of observed vs. modelled output values Mean prediction error (MPE) Variance of prediction error (VAR) Mean absolute percent error (MAPE) Evaluated training & test data Used fuzzy logic to evaluate over all criteria	Compared to MLR and existing recruitment model NN performance was "far superior" MLR and Ricker climate stock–recruitment model

Dimopoulos et al., 1999 Ecological Modelling	8 Urban descriptors: Vegetation density, vegetation height, wind velocity, building height, distance of adjacent street, traffic volume.	1 Lead concentration in grass	3 Trial and error Chose number with "optimal generalization capability"	Backpropagation 140 observations (1) k-fold cross validation: 60% - Train 20% - Test 20% - Validation (2) Non-linear k-fold cross validation 10 folds	Activation: Sigmoid Error: not discussed	Inputs: Standardized (to standardize the scales of measurement) Output: Centred, reduced, and converted to the interval [0,1] (because activation function in output nodes adjusts response values between 0 and 1)	Variable Importance: Partial derivatives	MSE R^2 of observed vs. modelled output values	Compared NN to full MLR model and stepwise regression NN models had better explanatory power than MLR models (regardless of type of cross validation used for training)
Lae et al., 1999 Ecological Modelling	6 Environmental variables: Catchment area/maximum area, fishing effort, conductivity, depth, altitude & latitude	1 Fish yield	5 Empirical approach. Chose smallest number of hidden nodes with "satisfying" results	Backpropagation 59 observations Fit with all 59 obs. To avoid overfitting: chose configuration with minimal dimension & satisfying results; Limited iterations to 500 n-fold cross validation	Activation: Sigmoid Error: Not discussed	For MLR: All variables transformed by log10	Initial weights: random values on interval [-0.3, 0.3] Variable Importance: Sensitivity analysis	R, Adjusted-R^2 of observed vs. modelled output values, p-values	Compared to stepwise MLR Showed that ANN models are viable when compared to traditional statistical methodologies.

Ozesmi & Ozesmi, 1999 Ecological Modelling	6 Habitat descriptors: Vegetation durability, stem density, stem height, distance to open water, distance to edge, water depth	1 - 3 Red-winged blackbird nesting probability (RWN) Marsh wren nesting probability (MWN) No RWN or MWN	6 Started with more complex, and then reduced the number of layers and hidden units. Most complex: two hidden layers with 200 hidden units in each. Did not get better results with more hidden layers. Started with 300 hidden nodes in one layer and reduced number until error increased	Backpropagation Trained on data from one wetland; validated on data from another wetland Training stopped when error on the training data reached a steady state. Minimum error of validation error recorded as target error. The model was then rerun using all the data to this target error.	Activation Hidden units: Logistic function on interval [−0.5, 0.5] Output units: Asymmetric logistic with a range on interval [0,1] (so the output was a probability between 0 and 1) Error: Cross entropy	Inputs: Standardized Outputs: already on range of [0,1] (because probabilities)	Initial weights: Random values on interval [-0.1, 0.1] (range where all the models were able to run) Variable Importance: Relevances, Sensitivity analysis, Neural interpretation diagrams (NIDs)	Average cross entropy, Concordance index (c-index), Percent better than random	Compared to stepwise logistic regression model Found that NN predicts habitat selection better and that using relevances, sensitivity analyses, and NIDs can lead to a better understanding of the mechanisms of habitat selection. Logistic model performed better in the cases of interspecies interactions
Olden and Jackson, 2001 Transactions of the American Fisheries Society	8 Habitat variables: Surface area, total shoreline perimeter, maximum depth, total dissolved solids (TDS), pH, lake elevation, occurrence of	1 Probability of occurrence of fish species and species abundance (separate model for each type of output and species)	Determined empirically by comparing networks with 1–20 hidden neurons and choosing the one with the "best predictive performance."	128 observations from Madawaska River drainage; 32 observations from Oxtongue River drainage Two methods to evaluate predictive performance:	Activation: Sigmoid Error: SSE	Inputs: Converted to the interval [0,1] Outputs: standardized by mean and standard deviation (to standardize the measurement	Variable importance: Randomization test Removed input and hidden nodes that were not significant and re-trained. Predictability	For presence/absence models: Confusion matrices (1) Percentage of observations correctly classified (2) Ability to predict species	Compared to logistic regression NNs has greater predictive power for almost all species. However: experiment with simulated data

	summer stratification Also included index of predation for some models			(1) n-fold cross validation on data from Madawaska (2) Train using Madawaska data (80%); Test on Oxtongue data (20%)		scales of the inputs)	generally not affected. Olden method	presence (model sensitivity); (3) Ability to predict species absence (model specificity); For abundance models: (1) R of observed vs. modelled output (2) RMSE of the predicted values	showed that when assumptions are met for a traditional statistical approach, it may perform as well as NN
Gevrey et al., 2003 Ecological Modelling	10 Habitat variables: wetted width, area with suitable spawning gravel, surface velocity, water gradient, flow/width, mean depth, standard deviation of depth, bottom velocity, standard deviation of bottom velocity, mean speed/mean depth	1 Density of brown trout spawning sites	5 Evaluated different model configurations. Fit model with training data, and tested on test data. Chose number of nodes with the best performance on the test set.	Backpropagation 205 observations Random division of observations: 75%: Training 25%: Test First used training and test data to inform choice of number of hidden nodes. Next, fit with all data for the comparison of different methods for input variable contributions	Activation: Logistic Error: Not discussed	Not discussed	Compared 7 methods: (i) Partial Derivatives; (ii) Weights method is a computation using the connection weights (Garson's algorithm); (iii) Perturb method (iv) Profile method (Sensitivity analysis?) (v) classical stepwise (vi) Improved stepwise a; (vii) Improved stepwise b. Partial	Not discussed	MLR used to judge the prediction quality of the NNs; stepwise MLR used to define significant variables NN had better prediction than MLR models, "confirming the non-linearity of the relationship between the variables"

							Derivatives method was found to be the most useful; stepwise methods gave the poorest results.		
Olden, 2003 Conservation Biology	9 Lake habitat variables Surface area, maximum depth, volume, total shoreline perimeter, elevation, total dissolved solids, pH, growing-degree days, occurrence of summer stratification.	27 Probability of occurrence of fish species (1 node for each species)	7 Compared performances of cross-validated networks with 1 to 25 hidden nodes. Chose the number that produced "the greatest network performance." Validated with n-fold cross-validation.	Backpropagation 286 observations	Activation: Logistic Error: Cross-entropy	Inputs: Converted to z scores (to standardize the measurement scales of the inputs, so that the same percentage change in the weighted sum of the inputs causes a similar percentage change in the unit output.)	Variable importance: Connection weights method	Not discussed	NA
Zhou, 2003 North American Journal of Fisheries Management	4 Historical (lagged) escapement data Selection of predictors was based on the characteristics of chinook salmon life history and data availability.	1 Salmon escapement (amount of salmon that return to their spawning habitat) Two different stocks (different	1 – 3 Started with two hidden nodes and then tested networks with one more and one less. One hidden neuron slightly outperformed	Backpropagation 15 observations All observations included to examine learning capability. Output for each year was compared with observed value.	Activation: Logistic Error: Difference between observed and modelled values	Not discussed	Used an ensemble method: Different training runs might result in networks with different weights, which would result in different predictions with the same inputs.	Mean absolute error (MAE) for the trained data Mean absolute percent error for the test data	Compared to the "traditional forecast method" (Moving Average) and ARIMA The NNs generally outperformed the MA method for both stocks analyzed.

		model for each stock)	those with two or three hidden neurons.	nfold cross validation to test the forecast capability. Found that forecast precision was lower than that of the trained fit.			Each NN was trained multiple times with the same training data set. Prediction outputs were obtained from the networks, and the mean and variance of the predictions were estimated.		ARIMA had better forecast for some years for one of the stocks.
Joy and Death, 2004 Freshwater Biology	31 Landscape scale data: geospatial landuse, geomorphologic, climatic, and geographic information system E.g., latitudinal & elevational position of the site reach, catchment area, average air temperature, vegetation type, land use proportions of the catchment, and	14 Probability of occurrence of fish species (1 node for each species)	70 Compared networks with 20–120 (in intervals of 20) hidden nodes and varied the number of iterations from 50 to 250 (in intervals of 50) and selected the combination with the “greatest predictive accuracy.”	Backpropagation 379 observations? Used n-fold cross validation to to ensure model was not overtrained/to evaluate the predictive accuracy of the model.	Not discussed	Inputs: Converted to z scores.	Input variable importance: Connection weights method	Classification metrics derived from confusion matrices: (i) Overall classification: percentage of sites where model correctly predicted the presence/absenc e of each species; (ii) Model sensitivity: percentage of site presences correctly predicted; (iii) Model specificity: ability to correctly predict species absences; (iv) Cohen’s Kappa	NA

	catchment geology							coefficient' of agreement; (v) Receiver-operating characteristic (ROC) plots	
Olden et al., 2006 Ecological Applications	24 Reach and catchment scale habitat variables E.g., latitude, distance from sea, catchment area, surface rock	16 Probability of occurrence of fish species (1 node for each species)	Not indicated Number of hidden nodes was chosen by comparing the performances of networks with 5-100 hidden neurons (in increments of five). Chose the number with the "greatest network performance"	Backpropagation 379 observations Max 500 iterations to determine optimum weights n-fold cross validation to assess model classification performance	Activation: Logistic Error: Cross entropy	Inputs: Converted to z-scores (to standardize the measurement scales of inputs to the network)	Input variable importance: Connection weights method	Simple matching coefficient Jaccard's similarity coefficient	Compared to two "traditional" approaches: (1) species-by-species approach (logistic regression); (2) a "classification-then-modelling" approach NN outperformed both traditional methods, exhibiting greater precision and accuracy for predictions. On average, correctly predicted community composition in nearly twice as many sites compared to the other methods.

Palacz et al., 2013 Biogeosciences	5 Ecological indicators (sea surface temperature, wind speed, photosynthetically available radiation, surface chlorophyll a concentration & mixed layer depth)	4 Phytoplankton functional type (Phyto-PFT) biomass	8 Tested 5 – 15 hidden nodes. Concluded that an 8 hidden-node NN was "well fitted yet general enough to simulate phyto-PFTs," and trained in a relatively short time	Levenberg–Marquardt Random or systematic division of observations: 70%: Training 15%: Test 15%: Evaluation Used an early stopping procedure to avoid over-fitting	Activation Hidden nodes: tangential sigmoidal Output nodes: linear Error: MSE	Transformed variables onto a log-10 scale if distribution was non-normal. (to avoid results biased towards the populated end of the range) Converted all inputs and outputs to common minimum-maximum range (e.g. [-1,1]) to avoid bias towards high values	Initial weights: random values Variable importance: Hinton weight diagrams (from 1 of 10 models in ensemble)	R of observed vs. modelled output values for training, test, and evaluation sets	NA
de Oña and Garrido, 2014 Neural Computing and Applications	12 Variables related to user satisfaction level Information, Punctuality, Safety, Courtesy, Cleanliness, Space, Temperature, Accessibility, Fare, Speed, Frequency, Proximity	1 Quality of service	6 Evaluated 1 – 30 hidden nodes; chose 6 because this architecture minimized the mean MAPE of the test data	Backpropagation Random division of observations: 70%: Training 15%: Validation 15%: Test	Activation: Logistic	A range of values in the interval [0, 1] has been used as input values for every variable, instead of using the original interval [0, 10]. This translation allows to adapt them for subsequent treatment in the NN, since the limits of the	Initial weights: small random values Compared four methods for assessing variable contributions: (i) Perturb Method; (ii) Profile Methods; (iii) Connection Weights Method; (iv) Partial derivatives methods	Mean Absolute Percent Error (MAPE) of test data	NA

						value range of every variable directly coincide with the upper and lower limits of the activation functions.	Methods showed similar rankings when ensemble approach applied Ensemble modeling: test 50 sets of different weights for 1 – 30 hidden nodes		
Krekoukiotis et al., 2016 Frontiers in Marine Science	9 Reproducti e, mortality and habitat variables, lagged 2 years: Reproductive Volume (May and August), Spawning Stock Biomass, Natural Mortality, Fishing Mortality, Egg Mortality, Egg Predation, Egg Abundance, Larval Abundance, Age 2 recruitment	1 Cod recruitment (number of cod recruits to the fishery at age 2)	3 Evaluated 20 models each with 1 to 30 neurons and recorded their average performance on training & test sets. Chose number that minimized test data error (“model configuration with the simplest architecture and highest generalization capability”)	Backpropagation 24 observations Plus 2000 – 2009 as test? 3-fold cross validation used to assess model prediction accuracy Random division of training observations from 3-fold split: 70%: Training 30%: Validation	Activation: Not discussed Error: MSE	Not discussed	Initial weights: Random values Ensemble model approach to account for the variability in model results (from initial weights and random data splitting during training). Trained 35 models (same architecture but different weights Variable contribution: (i) product-of-standardized-weights (ii) connection weights method	MSE of the test data R2 Mean and median values from ensemble reported	Performed better than existing stock-recruitment models

APPENDIX E.2: Supplementary Figures

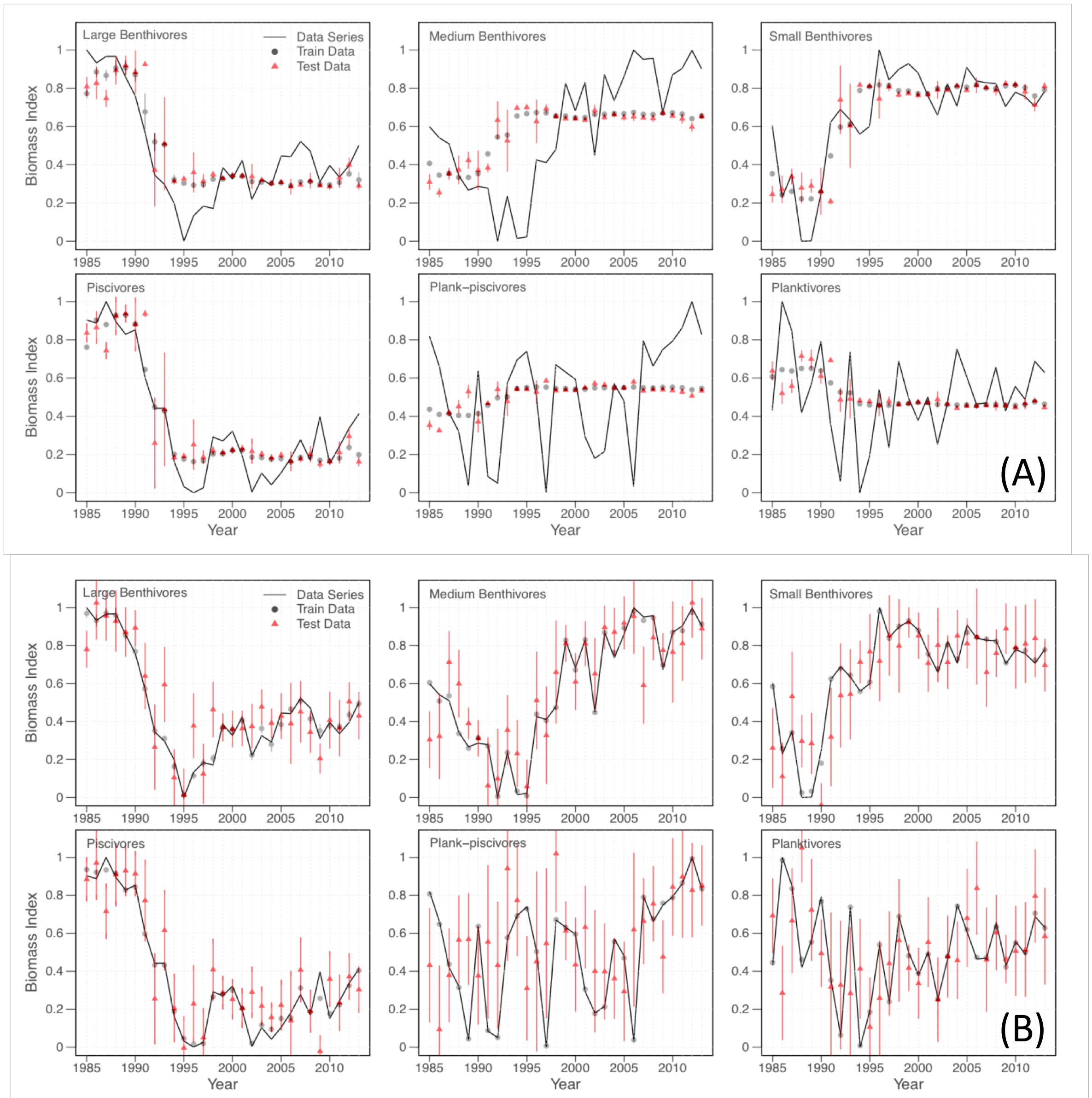

Figure S.1: Average and standard deviation of the modelled training data (black points) and test data (red triangles) for the Full period using (A) 1 hidden node and (B) 10 hidden nodes.

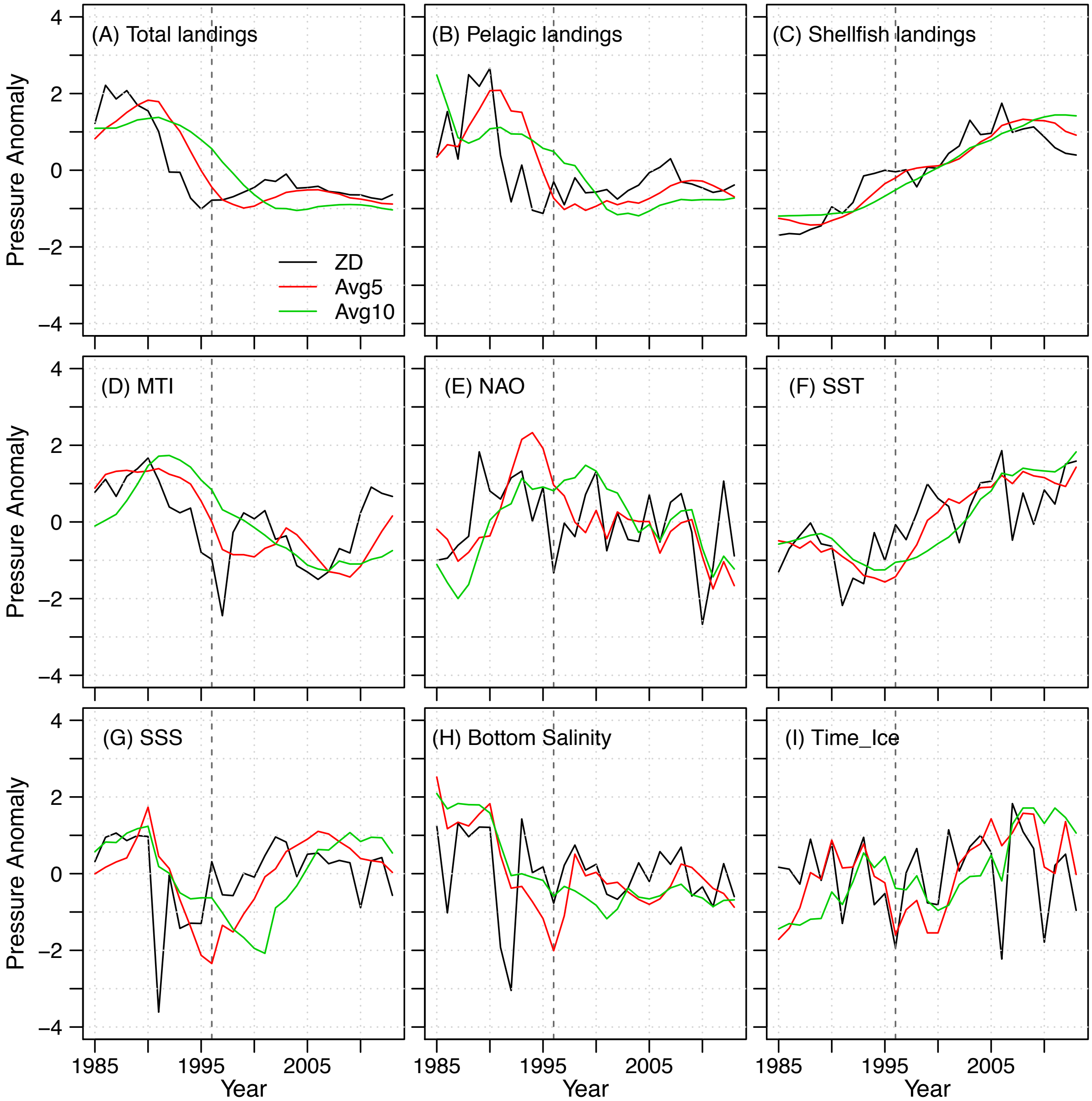

Figure S.2: Moving average predictors for delay lengths $k = 0$ (ZD), $k = 5$ (Avg5), and $k = 10$ (Avg10).

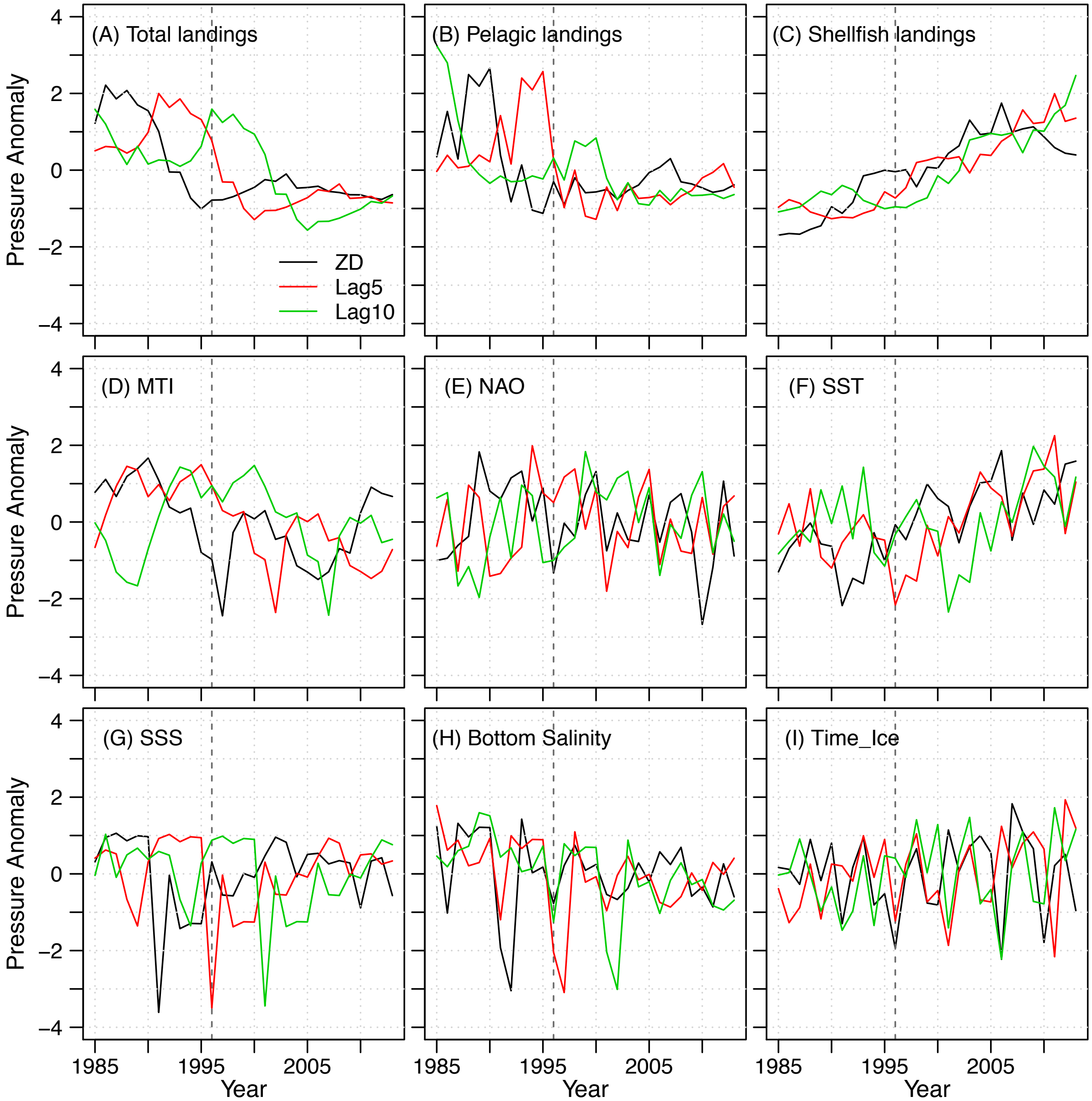

Figure S.3: Lagged predictors for delay lengths $k = 0$ (ZD), $k = 5$ (Lag5), and $k = 10$ (Lag10).

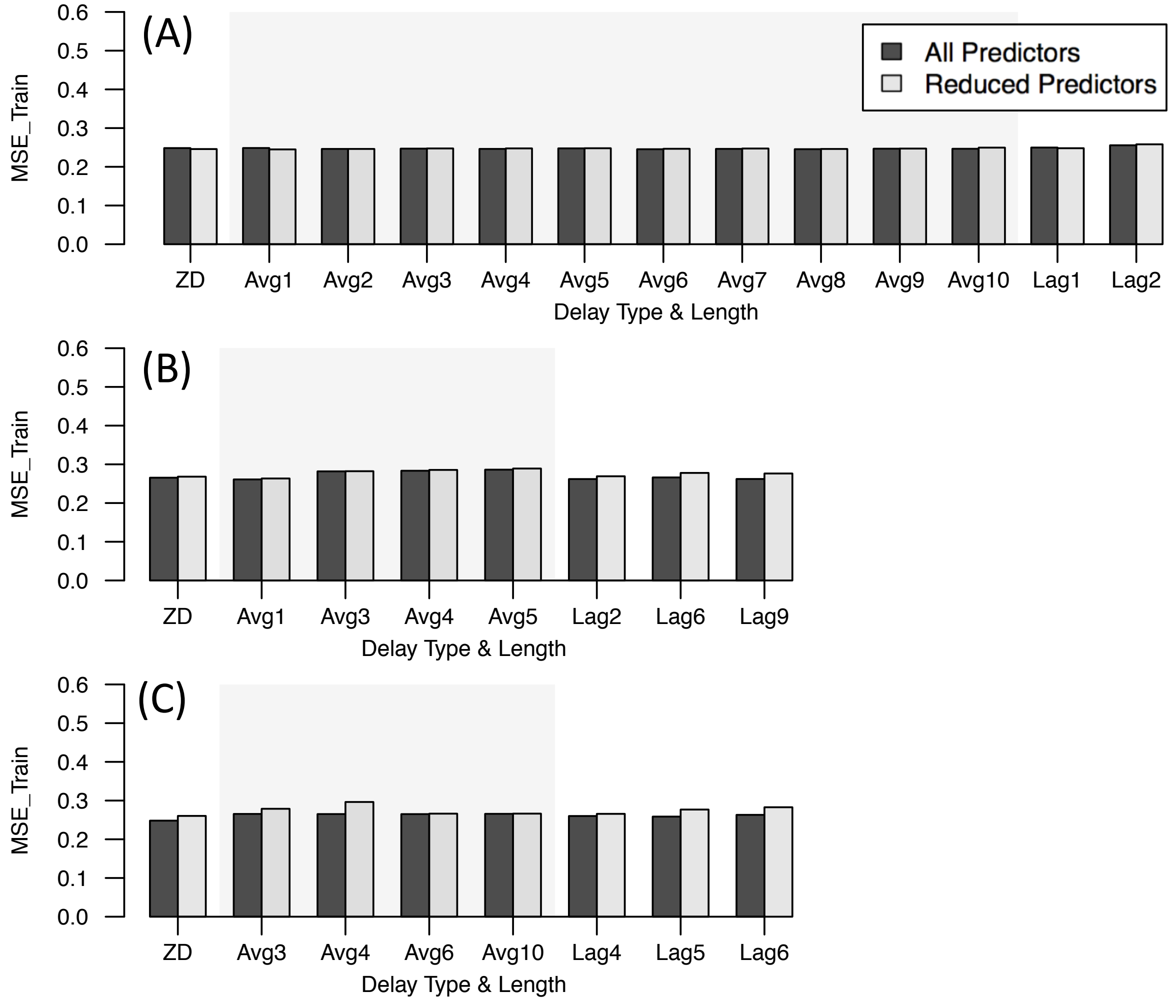

Figure S.4: $\overline{\text{MSE}}_{\text{Train}}$ of the best delays for each period. Dark grey represents the models trained with all predictors; light grey represents the models trained with the reduced predictor set (i.e., only the most influential pressures for the given delay). There are no notable differences in the fit between the models trained with all predictors and the reduced models for any period. (A) Full period; (B) Before period; (C) After period. Faint shaded box indicates the moving average models (to differentiate from ZD and lag).