1   **Methodological Issues in Building, Training, and Testing Artificial Neural Networks**

2

3   Stacy L. Özesmi[a]*, Uygar Özesmi[a], and Can O. Tan[b]

4

5   [a]Environmental Science Chair, Department of Environmental Engineering, Erciyes

6   University, 38039 Kayseri Turkey

7   [b]Department of Biology, Middle East Technical University, 06531 Ankara Turkey

8   * Corresponding author. Environmental Science Chair, Department of Environmental

9   Engineering, Erciyes University, 38039 Kayseri Turkey

10   Tel: 90 352 437 4937 /32800 Fax: 90 352 437 4404 email: stacy@erciyes.edu.tr

11

12   **Abstract**

13   We review the use of artificial neural networks, particularly the feedforward multilayer

14   perceptron with back-propagation for training (MLP), in ecological modelling. In MLP

15   modeling, there are no assumptions about the underlying form of the data that must be met as

16   in standard statistical techniques. Instead the researchers should make clear the process of

17   modelling, because this is what is most critical to how the model performs and how the results

18   can be interpreted. Overtraining on data or giving vague references to how it was avoided is

19   the major problem. Various methods can be used to determine when to stop training in

20   artificial neural networks: 1) early stopping based on cross-validation, 2) stopping after a

21   analyst defined error is reached or after the error levels off, 3) use of a test data set. We do not

22   recommend the third method as the test data set is then not independent of model

23   development. Many studies used the testing data to optimize the model and training.

24   Although this method may give the best model for that set of data it does not give

25   generalizability or improve understanding of the study system. The importance of an

26   independent data set cannot be overemphasized as we found dramatic differences in model

27   accuracy assessed with prediction accuracy on the training data set, as estimated with

28   bootstrapping, and from use of an independent data set. The comparison of the artificial

29   neural network with a general linear model (GLM) as a standard procedure is recommended

30   because a GLM may perform as well or better than the MLP. If the MLP model does not

31   predict better than a GLM, then there are no interactions or nonlinear terms that need to be

32   modelled and it will save time to use the GLM. MLP models should not be treated as black

33   box models but instead techniques such as sensitivity analyses, input variable relevances,

34   neural interpretation diagrams, randomization tests, and partial derivatives should be used to

35   make the model more transparent, and further our ecological understanding which is an

36   important goal of the modelling process. Based on our experience we discuss how to build an

37   MLP model and how to optimize the parameters and architecture. The process should be

38   explained explicitly to make the MLP models more readily accepted by the ecological

39   research community at large, as well as to make it possible to replicate the research.

40

41   *Keywords*: Artificial neural networks; Backpropagation; Modelling; Model; Model

42   generalizability; Ecology, Habitat selection.

43

44

45   **1. Introduction**

46

47   The earliest papers on the use of artificial neural networks in ecology began appearing

48   in the mid 1990's. Reported advantages of artificial neural networks over more traditional

49   methods include: 1) form of relationships need not be specified (no assumptions need to be

50   made about the distribution of data); 2) nonlinear relationships or interactions among variables

51   are easily modelled; 3) performance is usually better than general linear models; 4) complex

52   data patterns can be handled because of their nonlinear nature; 5) ability to generalize to new

53   data. Thus, neural networks have many advantages for ecological studies where data rarely

54   meet parametric statistical assumptions and where nonlinear relationships are prevalent.

55   However, artificial neural networks also have disadvantages: 1) they are computationally

56   intensive; 2) many parameters must be determined with few guidelines and no standard

57   procedure to define the architecture; 3) analyst expertise is required; 4) no global method

58   exists for determining when to stop training and thus overtraining is problematic; 5) sensitive

59   to composition of the training data set; 6) sensitivity of training to initial network parameters;

60   7) black box models.

61   Perhaps because it is one of the easiest neural networks to understand, the feedforward

62   multilayer perceptron with back-propagation for training, has been the most commonly used

63   neural network in ecology. More details on how this type of neural network works can be

64   found elsewhere (i.e., Lek and Geugan, 1999 or in texts such as Anderson, 1995; Weiss and

65   Kulikowski, 1991; Bishop, 1995; or Ripley, 1996).

66   In this article we review the use of the MLP, or feedforward multilayer perceptron

67   with back-propagation for training, in ecological modelling and how it is practiced. Based on

68   our experience we discuss how to build MLP models and how to optimize the parameters and

69   architecture. We make recommendations for best practices, which include the importance of

70 avoiding overtraining, use of an independent test data set, and use of sensitivity analyses,

71 neural interpretation diagrams, input variable relevances, and other methods to open up the

72 black box model. Although in this article we focus on the MLP, some of our

73 recommendations are also relevant to other types of artificial neural networks.

74

75 **2. Literature review**

76

77   The MLP has been used in ecological studies by Brey et al. (1996) who predicted

78 benthic invertebrate production/biomass ratios, Levine et al. (1996) who classified soil

79 structure from soil sample data, Tan and Smeins (1996) who predicted changes in the

80 dominant species of grassland communities based on climatic input variables, and Poff et al.

81 (1996) who modelled streamflow response based on average daily precipitation and

82 temperature inputs. Paruelo and Tomasel (1997) predicted normalized difference vegetation

83 index (NDVI) used in remote sensing. Phytoplankton production (Scardi, 1996; Scardi and

84 Harding, 1999; Scardi, 2001) and phytoplankton occurrence and succession (Recknagel et al.,

85 1997; Karul et al., 2000) have been modelled with the MLP. Fish abundance based on habitat

86 variables (Baran et al., 1996; Lek et al., 1996), fish yield (Lae et al., 1999), and fish and

87 microhabitat use (Reyjol et al., 2001) have also been modelled. The MLP has been used to

88 predict, based on habitat variables, presence or absence of macroinvertebrates (Hoang et al.,

89 2001), birds (Manel et al. 1999), golden eagle nest sites (Fielding, 1999b), interacting marsh

90 breeding bird nests (Özesmi and Özesmi, 1999), and cyanobacteria (Maier et al., 1998). Bird

91 abundance (Lusk et al., 2001) and macro-invertebrate abundance and species richness (Lek-

92 Ang et al., 1999) has been modelled. The MLP has been used to predict damage to

93   agricultural fields by flamingo (Tourenq et al., 1999) and wild boar (Spitz and Lek, 1999). It

94   has also been used to predict lead concentrations in grasses (Dimopoulos et al., 1999).

95

96   *2.1 Criticisms of modelling with MLP*

97

98   From a literature review, we saw some problems with the reporting on the use of MLP.

99   Sometimes the modelling process was not clearly described. For example, some research did

100   not report why certain variables were chosen for a final model. Others did not tell how the

101   parameters were set or how the architecture, the number of hidden units, was determined. The

102   number of samples used to train, validate and test the model was not always given.

103

104   *2.2 Overtraining*

105

106   However, the major problem was overtraining on data or giving vague references on

107   how it was avoided. An exception is Paruelo and Tomasel (1997), who provide a discussion

108   of their experience with overtraining. Unfortunately it seems that often studies do not make

109   sufficient effort to avoid overtraining. One of the biggest disadvantages of using MLP is that

110   there is no perfect method for determining the number of training iterations. There are

111   basically three methods: 1) choose a user defined error level; 2) use an early stopping method

112   such as autotrain (Goodman, 1996); or 3) use the test data. The problem with the first method

113   is that it is difficult to decide on what this error level should be. Often this level is chosen

114   when the error levels off and does not change. The error usually drops until a certain number

115   of iterations where it levels off and does not get much smaller; however at this point the

116   network may be overtrained. With early stopping methods, part of the training data is held out

117   from the training and used to test the model performance. The error goes down on the training

118   data as the training proceeds. The error also initially goes down on the holdout data but then

119   the error level rises again as the model becomes overtrained. While we prefer this method, it

120   requires more data, which often is not available. Another problem with this method is that it

121   doesn't guarantee that the minimum error found is a global minimum rather than a local

122   minimum. If the test data set is used to determine when to stop training, this means that it is

123   not an independent test of the model.

124

125   *2.3 Independent test data*

126

127   Another problem we saw with the use of neural networks in ecological research was

128   the lack of independent test data sets. Two main approaches exist for evaluating (testing)

129   model performance (Guisan and Zimmerman, 2000). The first approach is to use a single data

130   set to train and test the model using cross-validation, leave-one-out jack-knife, or

131   bootstrapping. These techniques are used most often when the available data set is small and

132   all the data is needed to train the model. The second approach is to use an independent data

133   set for testing. If the data set is divided into two parts with one part for a test, it is called a

134   split sample approach. However, the independent test is optimal if the two data sets originate

135   from two different sampling strategies.

136   Some studies used the test data to optimize the model and the amount of training (i.e.

137   Levine et al., 1996 ; Karul et al., 2000; Tourenq et al., 1999; Reyjolet al. 2001). Although this

138   method may give the best model for that set of data it does not give generalizability or

139   improve understanding of the study system.

140   Studies with independent test data set(s) are rare (i.e. Recknagel et al., 1996; Poff et

141   al., 1996; Paruelo and Tomasel, 1997; Özesmi and Özesmi, 1999). Often the jackknifing

142   (leave-one-out) of the data set is used (i.e. Lae et al., 1999; Manel et al., 1999; Spitz and Lek,

143   1999) or cross validation (Levine et al., 1996; Lek et al., 1996; Baran et al., 1996; Paruelo and

144   Tomasel, 1997; Manel et al., 1999).

145   Hirzel et al. (2001) even argue that one independent test is not sufficient but that more

146   tests are needed. However, independent test data is usually in short supply or non-existent in

147   many ecological studies. Therefore they generated a virtual species and used simulated data

148   sets to compare their models. Similar techniques may be possible in other ecological studies

149   as well.

150

151   *2.4 Comparison with general linear models*

152

153   Many studies have compared the MLP with general linear models (i.e. Baran et al.,

154   1996; Brey et al., 1996; Scardi, 1996; Paruelo and Tomasel, 1997; Fielding, 1999b; Karul et

155   al., 1999; Lae et al., 1999; Manel et al., 1999; Özesmi and Özesmi, 1999). Generally these

156   studies have found that the MLP performs better than general linear models such as multiple

157   linear or logistic regression (i.e. Brey et al., 1996; Baran et al., 1996; Paruelo and Tomasel,

158   1997; Özesmi and Özesmi, 1999). However the MLP does not always outperform linear

159   techniques (Fielding, 1999b; Manel et al., 1999).

160   We recommend the comparison of the artificial neural network with a general linear

161   model as a standard procedure because general linear models may perform as well or better

162   than MLP. If the MLP model does not predict better, then there are no interactions or

163   nonlinear terms that need to be modelled and it will save time to use the general linear model.

164   A quick method is to make a neural network without a hidden layer to create the general linear

165   model. If this model performs as well as the MLP with a hidden layer, then there is no need to

166   use the MLP. In addition, by connecting each input to only one hidden unit and then to the

167   output unit a transformation-only neural network model can be made. If the model

168   performance of the MLP with a hidden layer is the same as a transformation-only model, then

169   the variables can simply be transformed to remove nonlinearities in a general linear model

170   (Goodman, 1995).

171

172   *2.5 Opening the black box*

173

174   Finally, artificial neural network models should not be treated as black box models but

175   by using various techniques the box can be opened (Scardi, 2001). Available techniques such

176   sensitivity analyses (Lek et al., 1996; Scardi, 1996; Recknagel et al., 1997), input variable

177   relevances and neural interpretation diagrams (Özesmi and Özesmi, 1999), randomization

178   tests of significance (Olden 2000; Olden and Jackson, 2000), and partial derivatives

179   (Dimopoulos et al., 1999; Reyjol et al., 2001) should be used to make the model more

180   transparent. Use of these techniques, which are described below, will further our ecological

181   understanding, which is an important goal of the modelling process.

182   In sensitivity analyses, the response of the model to each of the input variables is

183   determined by applying a typical range of values to one variable at a time while holding the

184   other variables constant (Lek et al., 1996). The variables that are held constant are set an

185   arbitrary level. The level they are held at influences the results so they can be set at their

186   minimum, first quartile, median or mean, third quartile, and maximum values successively.

187   The resulting plots allow one to examine how the variables influence the model response.

188   By examining the input variable relevances, we can see how much each input variable

189   contributes to the model (Özesmi and Özesmi, 1999). The relevance of an input variable is

190   the sum square of weights for that input variable divided by the total sum square of weights

191   for all input variables. Variables with high relevances are more important in the model.

192   Neural interpretation diagrams (NIDs) can be drawn to understand how the model is

193   weighting different input variables and how the input variables interact to give the model

194   response (Özesmi and Özesmi, 1999). NIDs are drawn by scaling the thickness of the lines

195   connecting the units according to the relative values of their weights. Black lines represent

196   positive signals and gray lines represent negative signals. Thus in one diagram, we can look

197   at the thickness of the connections coming out of the input units to see which variables are

198   most important. We can see how the input variables interact and their contribution to model

199   output by looking at the hidden layer. However neural interpretation diagrams are most

200   helpful when the number of units and connections is limited. Diagrams with 20 or more

201   variables are too complicated to gain any insights. For example, twelve variables are typical

202   in analysis of cognitive maps (Buede and Ferrell, 1993).

203   A randomization test has been developed to assess the statistical significance of

204   connection weights and input variable relevances (Olden, 2000; Olden and Jackson, 2000). In

205   this approach the response variable is randomized, a neural network is constructed using the

206   randomized data, and all the input-hidden-output connection weights (product of the input-

207   hidden and hidden-output weights) are recorded. This procedure is repeated a large number of

208   times to generate a null distribution for each input-hidden-output connection weight. This

209   value is then compared to the actual model value to calculate the significance level. With this

210   randomization test, the neural network can be pruned by eliminating connection weights that

211   have little influence on the network output. With the insignificant connection weights

212   removed, it is easier to interpret how the model makes predictions with a NID. In addition,

213   the randomization test identifies the independent variables that significantly contribute to

214   model prediction.

215   The partial derivatives of the network output with respect to input variables can be

216   used to show the influence of the variables in the model (Dimopoulos et al., 1999; Reyjol et

217   al., 2001). By plotting the partial derivatives of the network output with respect to an input

218   variable, how the network output changes with increasing values of the input variable can be

219   seen. Somewhat similar to relevances, the importance of the variables in the model can be

220   determined by calculating the sensitivity of the MLP output for the data set with respect to

221   each input variable. The sensitivity is the summation of all the squared partial derivatives for

222   each input variable. By using techniques such as these, MLP models can be easier to interpret

223   and help to improve our understanding of the study system.

224

225   **3. ANN modelling process**

226

227   Based on our experience we discuss how to build a MLP model and how to optimize the

228   network parameters and architecture. See Tan et al. (2002) for an example of our modeling

229   process where we follow the guidelines and recommendations presented in this paper. Our

230   MLP modelling process generally proceeds as follows. First it is necessary to determine the

231   form of inputs and outputs for the data, the pre and post-processing of the data. Usually the

232   input variables are standardized so that they are all on the same order of magnitude. We have

233   found that standardizing the input variables, to means of zero and units of standard deviations,

234   has consistently led to better results. Then we determine the network parameters such as learn

235   rate, weight range, etc. Next we optimize the architecture, the number of hidden layers and

236   number of hidden units in the hidden layers. Then we optimize the parameters together with

237   the chosen architecture. We use techniques such as neural interpretation diagrams, input

238   variable relevances, and sensitivity analyses to understand how our model is making

239   predictions. Finally we conduct an independent test of the model. When our output is binary

240   and depends on threshold we prefer ROC curves and c-index (which is an estimate of the area

241   under the ROC curve) for assessing model accuracy. See Fielding (1999a) for a discussion of

242   different ways to assess model accuracy.

243

244   *3.1 Network parameters*

245

246   Learn rate and weight range are network parameters that influence the performance of

247   the model by affecting the weights. The learn rate and weight range can be set at default

248   values and if the model is unstable made smaller until it stabilizes. We have found that

249   changing the weight range or the learn rate does not result in large changes in model accuracy.

250   The changes in accuracy are in the same range as those that result from changing the random

251   start, which initializes the weights. Because of the variation in model performance caused by

252   different initial weights, we run all network configurations at least five times using the same

253   predetermined random seeds, produced by a random number generator. Thus we optimize the

254   network parameters and network architecture based on the average of the five random starts.

255

256   *3.2 Architecture optimization*

257

258   Next we optimize the architecture, the number of hidden layers and number of

259   hidden units in the hidden layers. Heuristics exist for determining the number of units in the

260   hidden layer(s). However, for each application it is basically a process of trial and error.

261   We systematically run different models to optimize the network architecture. First we create

262   a general linear model (GLM), or a network with no hidden layer. Second we create a

263   transformation only model where each input is connected to only one hidden unit because if

264   the transformation model performs as well as an ANN with a hidden layer then the input

265   variables need only be transformed to remove nonlinearities. Then we create models with

266   one hidden layer having different numbers of hidden units. For a well generalized ANN

267   model, there should be about 10 times as many training data points as there are weights in

268   the network (Bishop, 1995). By using this heuristic, we can set an upper limit on the

269   number of hidden units in the model. We have found with our data that one hidden layer has

270   been as accurate or more accurate than networks with 2 hidden layers (Figure 2). The

271   accuracy of the networks with hidden layers generally increases and then levels off after a

272   certain number of hidden units. When choosing the final architecture, the model with fewer

273   hidden nodes should be chosen, because for two networks with similar errors on the training

274   sets, the simpler one is likely to predict better on new cases (Bishop, 1995).

275

276   *3.3 Avoiding overtraining*

277

278   We have had good results with training a MLP and deciding when to stop training by

279   using a holdout set (Özesmi and Özesmi, 1999). However this technique requires more data,

280   which is often in short supply in ecological studies. More recently we have trained MLPs and

281   decided when to stop training by determining when the accuracy leveled off using

282   bootstrapping. We used c-index, which is approximately the area under the Receiver

283   Operating Characteristic (ROC) curve, to assess the accuracy. A c-index of 1 indicates a

284   perfect model and a c-index of 0.5 indicates a model that predicts no better than a random

285   model. We trained the same network to various numbers of epochs and calculated the c-index

286   and corrected c-index with 150 bootstraps. With each bootstrap the model is trained with

287   approximately 66% of the data that is randomly selected from the total set. The model

288   accuracy (c-index) is calculated using that subset of the data and the whole set. The optimistic

289   bias of each bootstrap is determined as the difference between these two accuracies. The

290   corrected c-index is then determined by subtracting the average optimistic bias for all of the

291   bootstraps from the full model c-index. When the corrected c-index levelled off we thought

292   that this would be a well-generalized model. For example, the number of epochs versus c-

293   index is shown in Figure 1. While the c-index continues to increase as the number of epochs

294   increases, the corrected c-index levels off at 70 epochs. This is an indication of overtraining

295   beyond 70 epochs.

296

3.4 Independent test data set

298

299   Although it has been stated before many times (i.e. Fielding, 1999a), we cannot

300   overemphasize the importance of an independent testing set. For example, in our work, the c-

301   index of a model based on the training data was 0.746 (Tan et al., 2002). We used 150

302   bootstraps to estimate how generalizable the model was and the corrected c-index as

303   determined from these bootstraps was 0.663. Finally we tested our model on an independent

304   data set. The c-index was 0.511 or about random. Thus our training data indicated the model

305   was explaining some of the variation in the data. The bootstrapping indicated it was not quite

306   as generalizable but the test data indicated our model was performing about the same as a

307   random model. Thus while very good results might be obtained with training data, and still

308   good results might be indicated by bootstrapping (or some other data splitting technique such

309   as jack-knifing), the real test is the independent data set.

310

4. Conclusions

312

313   We recommend that the following information be included in every published research

314   using MLP. These should be included to make the MLP modelling process more transparent

315   and thus more readily accepted by the ecological research community at large, as well as to

316   make it possible to replicate the research. When using standard statistical techniques, the

317   researchers must justify that their data meet the assumptions of those techniques. However in

318   the MLP modeling, there are no assumptions about the underlying form of the data that must

319   be met. Instead the researchers should make clear the process of modelling, because this is

320   what is most critical to how the MLP model performs and how the results can be interpreted.

321   A clear explanation of the modeling processing is necessary including which variables were

322   chosen for the final model and why they were chosen. A description of the form of the input

323   and output variables is needed. For example, input variables are usually standardized so that

324   they are all in the same order of magnitude. An explanation of how the network parameters

325   (learn rate, weight range, momentum) were chosen, and the values that were used in the final

326   model(s) should be stated. How the network architecture was optimized should be included

327   and the number of hidden layers and hidden units that were chosen for the final model(s). The

328   number of samples used to train, validate and test the model should be clearly indicated. This

329   information could be included in an appendix or in a table, but they should be part of every

330   published research.

331   We have found that one hidden layer is sufficient in our MLP models to achieve high

332   accuracy on the training data and that two hidden layers does not increase this accuracy. The

333   accuracy level levels off after a certain number of hidden units. However to avoid

334   overtraining, the number of training epochs should be limited as well as the number of hidden

335   units in the hidden layer.

336   Based on the literature review of the use of MLP in ecological research and our own

337   experience, we suggest that more effort be made to interpret the results of the neural network

338   models using techniques such as input variable relevances, sensitivity analyses, neural

339   interpretation diagrams, randomization tests, and partial derivatives. We also recommend that

340   independent test data sets be used for assessing model accuracy, as we found dramatic

341 differences between model performance based on training data, bootstrapping, and use of an

342 independent test data set.

343

344 **References**

345 Anderson, J.A., 1995. An Introduction to Neural Networks. MIT, Cambridge, Massachusetts,

346   p 650.

347 Baran, P., Lek, S., Delacoste, M., Belaud, A., 1996. Stochastic models that predict trout

348   population density or biomass on a mesohabitat scale. Hydrobiol. 337, 1–9.

349 Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press,

350   Oxford, p 482.

351 Brey, T., Jarre-Teichmann, A., Borlich, O., 1996. Artificial neural network versus multiple

352   linear regression: predicting P:B ratios from empirical data. Mar. Ecol. Prog. Ser. 140,

353   251–256.

354 Buede, D.M., Ferrell, D.O., 1993. Convergence in problem solving: a prelude to quantitative

355   analysis. IEEE Transactions on Systems, Man, and Cybernetics 23, 746-765.

356 Dimopoulos, I., Chronopoulos, J., Chronopoulou-Sereli, A., Lek, S., 1999. Neural network

357   models to study relationships between lead concentration in grasses and permanent

358   urban descriptors in Athens city (Greece). Ecol. Model. 120, 157–165.

359 Fielding, A.H., 1999a. How should accuracy be measured? In: Fielding AH, editor. Machine

360   Learning Methods for Ecological Applications. Kluwer, Dordrecht, p 209-223.

361 Fielding, A.H., 1999b. An introduction to machine learning methods. In: Fielding AH, editor.

362   Machine Learning Methods for Ecological Applications. Kluwer, Dordrecht, p 1–35.

363   Goodman, P. H., 1996. NevProp Software, Version 3. University of Nevada, Reno, NV.

364   http://brain.unr.edu/index.php

365   Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol.

366   Model. 135, 147–186.

367   Hirzel, A.H., Helfer, V., Metral, F., 2001. Assessing habitat-suitability models with a virtual

368   species. Ecol. Model. 145, 111-121.

369   Hoang, H., Recknagel, F., Marshall, J., Choy, S., 2001. Predictive modelling of

370   macroinvertebrate assemblages for stream habitat assessments in Queensland

371   (Australia). Ecol. Model. 195, 195–206.

372   Karul, C., Soyupak, S., Cilesiz, A.F., Akbay, N., Germen, E., 2000. Case studies on the use of

373   neural networks in eutrophication modeling. Ecol. Model. 134, 145–152.

374   Lae, R., Lek, S., Moreau, J., 1999. Predicting fish yield of African lakes using neural

375   networks. Ecol. Model. 120, 325–335.

376   Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M., 1996. Role of some

377   environmental variables in trout abundance models using neural networks. Aquat.

378   Living Resour. 9, 23–29.

379   Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an

380   introduction. Ecol. Model. 120(2-3), 65–73.

381   Lek-Ang, S., Deharveng, L., Lek, S., 1999. Predictive models of collembolan diversity and

382   abundance in a riparian habitat. Ecol. Model. 120(2-3), 247-260.

383   Levine, E.R., Kimes, D.S., Sigillito, V.G., 1996. Classifying soil structure using neural

384   networks. Ecol. Model. 92, 101–108.

385 Lusk, J.J., Guthery, F.S., DeMaso, S.J., 2001. Northern bobwhite (*Colinus virginianus*)

386   abundance in relation to yearly weather and long-term climate patterns. Ecol. Model.

387   146, 3–15.

388 Maier, H.R., Dandy, G.C., Burch, M.D., 1998. Use of artificial neural networks for modelling

389   cyanobacteria Anabaena spp. in the River Murray, South Australia. Ecol. Model.

390   105(2-3), 257–272.

391 Manel, S., Dias, J.-M., Ormerod, S.J., 1999. Comparing discriminant analysis, neural

392   networks and logistic regression for predicting species distributions: a case study with

393   a Himalayan river bird. Ecol. Model. 120, 337–347.

394 Olden, J.D., 2000. An artificial neural network approach for studying phytoplankton

395   succession. Hydrobiol. 436, 131-143.

396 Olden, J.D., Jackson, D.A., 2000. Illuminating the 'black box': a randomization approach for

397   understanding variable contributions in artificial neural networks. Ecol. Model., .

398 Özesmi, S.L., Özesmi, U., 1999. An artificial neural network approach to spatial habitat

399   modelling with interspecific interaction. Ecol. Model. 116, 15–31.

400 Paruelo, J.M., Tomasel, F., 1997. Prediction of functional characteristics of ecosystems: a

401   comparison of artificial neural networks and regression models. Ecol. Model. 98, 173-

402   186.

403 Poff, N.L., Tokar, S., Johnson, P., 1996. Stream hydrological and ecological responses to

404   climate change assessed with an artificial neural network. Limnol. Oceanogr. 41, 857–

405   863.

406 Recknagel, F., French, M., Harkonen, P., Yabunaka, K., 1997. Artificial neural network

407   approach for modelling and prediction of algal blooms. Ecol. Model. 96, 11–28.

408   Reyjol, Y., Lim, P., Belaud, A., Lek, S., 2001. Modelling of microhabitat used by fish in

409   natural and regulated flows in the river Garonne (France). Ecol. Model. 146, 131–142.

410   Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press,

411   Cambridge, p 403.

412   Scardi, M., 1996. Artificial neural networks as empirical models for estimating phytoplankton

413   production. Mar. Ecol. Prog. Ser. 139, 289–299.

414   Scardi, M., 2001. Advances in neural network modeling of phytoplankton primary production.

415   Ecol. Model. 146, 33–45.

416   Scardi, M., Harding, L.W.J., 1999. Developing an empirical model of phytoplankton primary

417   production: a neural network case study. Ecol. Model. 120, 213–223.

418   Spitz, F., Lek, S., 1999. Environmental impact prediction using neural network modelling. An

419   example in wildlife damage. Journal of Applied Ecology 36, 317-326.

420   Tan, S.S., Smeins, F.E., 1996. Predicting grassland community changes with an artificial

421   neural network model. Ecol. Model. 84, 91–97.

422   Tan, C.O., Özesmi, S.L., Özesmi, U., Robertson, R.J. Generalizability of artificial neural

423   network models: I. Independent test for predicting breeding success. Ecol. Model.

424   (submitted)

425   Tourenq, C., Aulagnier, S., Mesleard, F., Durieux, L., Johnson, A., Gonzalez, G., Lek, S.,

426   1999. Use of artificial neural networks for predicting rice crop damage by greater

427   flamingos in the Camargue, France. Ecol. Model. 120, 349–358.

428   Weiss, S.M., Kulikowski, C.A., 1991. Computer Systems that Learn. Morgan Kauffman, San

429   Mateo, CA, p 223.
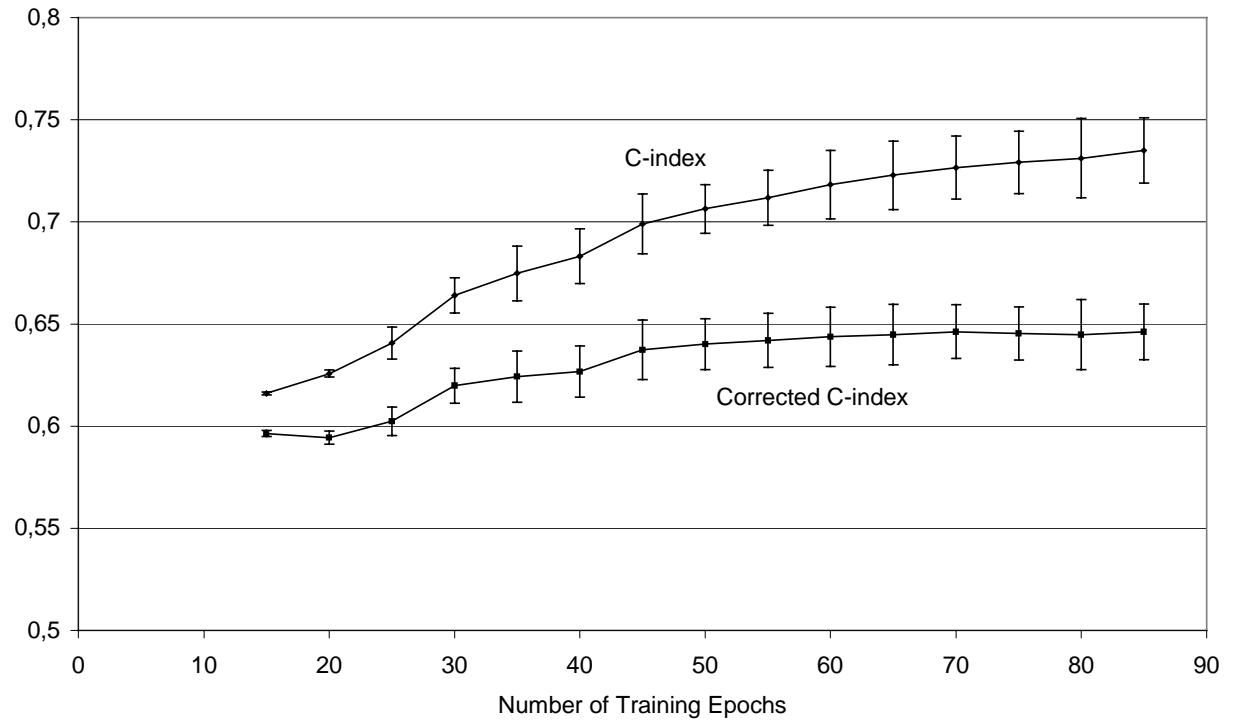
430

430   Figure Captions

431

432   Figure 1.   Average and standard deviation of c-index and corrected c-index versus the number

433   of epochs the MLP model was trained.

434

435   Figure 2. MLP model accuracy on the training data versus the number of hidden units in one

436   hidden layer (top curve) and two hidden layers (bottom curve).

437

438   Figure 1.

439

439

440   Figure 2.

441