Solving Capelin Time Series Ecosystem Problem using Hybrid Artificial Neural Networks- Genetic Algorithms Model

Karam M. Eghnam, Sulieman Bani-Ahmad
Information Technology Department, Faculty of Science and Information Technology
Al-Balqa Applied University
Salt, Jordan

Alaa F. Sheta
Computer Science Department,
The World Islamic Sciences and Education University (WISE),
Amman, Jordan

ABSTRACT
The Capelin stock in the Barents Sea is the largest in the world. It maintained a fishery with annual catches of up to 3 million tons. The Capelin stock problem has an impact in fish stock development. In this paper, the stock prediction problem of the Barents Sea capelin is attacked using Artificial Neural Network (ANNs) and Multiple Linear model Regression (MLR) model. The weights of ANNs are adapted using the Genetic Algorithm (GA).The models are compared against each other and empirical work has shown that the ANN-GA model can have better overall accuracy over (MLR). It performs 21% over MLR model..

General Terms
Artificial Intelligence, Algorithms.

Keywords
Forecasting, Capelin stock, Neural Networks, Genetic Algorithm, Ecosystem.

1. INTRODUCTION
Capelin (pelagic fish species) is an Arctic salmon fish that spends most of the year swimming around in the Arctic Ocean (Fig.1). In the Atlantic, the capelin is located in the Barents Sea. The Capelin stock problem has an impact in fish stock development. This problem is commonly attacked using many methods such as the virtual population analysis (VPA) or other techniques that apply fish harvest and estimates of natural mortality and growth to project stock development [1].

Here the general and stock prediction problem of the Barents Sea capelin specifically are attacked using two distinct models the first one is Neural Network(NN) [2] and the second is Multiple Linear Regression Model (MLR).

ANNs apply principles from neurology to find patterns in complex data and have successfully been used to predict yields of the Japanese sardine population [3] and African lake fisheries [4]. A NN is a powerful data modelling tool that is able to capture and represent complex input-output relationships. The motivation for the development of Neural Network technology stemmed from the desire to develop an artificial system that could perform "intelligent" tasks similar to those performed by the human brain. NN resemble the human brain in the following two ways: (1) A NN acquires knowledge through learning. (2) A NN’s knowledge is stored within inter-neuron connection strengths known as synaptic weights. GA has been applied to a number of problems in NN [5]. This is significant for two reasons. Firstly, many of the problems in Neural Networks are important in their own right and do not presently have any wholly satisfactory means of resolution. A good example of this is the network weight optimization as we presented in this research. Secondly, the failure modes of the GA seen in NN applications are common to a broader class of problems, and their study can yield more general insights.

The models used here based on factors that represent an ecological importance to the capelin stock. Ecosystems are spatially heterogeneous and spatial patterns and processes that represent an importance to ecosystem structure and function [6] for this reason the problem in capelins stock is influenced by the stock development, therefore we have motivated to do this research to contribute in solving the fishery estimation problem.

2. THE CAPELIN ECOSYSTEM
The capelin has a northerly circumpolar distribution, and it plays a key role in the arctic food [5] (Fig. 1). Since 1979, the Barents Sea capelin fishery has been regulated by a bilateral fishery management agreement between Russia and Norway. During its autumn 2003 meeting the Mixed Russian Norwegian Fishery Commission decided that no fishing should take place on Barents Sea capelin for the winter season 2004.

Capelin overlaps spatially with cod, herring, and capelin itself at different stages of it is life history. Capelin is the most important food item for the cod (a large fish often lives close to the seafloor). The only food item of similar abundance and energy content is herring. Herring may replace capelin in the diet of cod during capelin collapses [7], for example herring present in the Barents Sea in part of the period when the capelin was practically absent [8]. When abundant in the Barents Sea, Herring often causes recruitment failure and eventually population crashes in the capelin stock [3]. The presence of 0-group herring has little effect on capelin recruitment compared to significant amounts of One-Year-old capelin and young herring, also indicating a negative interaction between these two species.

Since the biomass of capelin in the coming year will depend on the current one, abundance of capelin may be an important input factor in the model. The average weight of the two-year old capelin further indicates the current feeding conditions, which may impact on stock development [9]. The capelin Ecosystem input factors are shown in Table 1. This is the inputs used for our experiments.

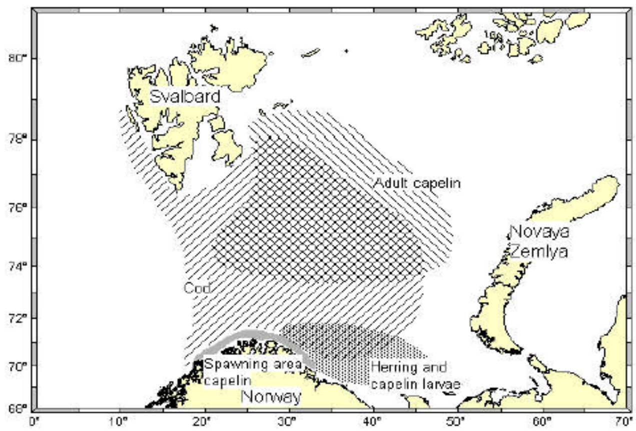

Fig. 1. Map of the Barents Sea with the main features of the distribution of various age groups of capelin as well as cod and herring. Predation by juvenile herring (1-3 years old) on larval capelin seems to have great impact on capelin recruitment at times when herring is abundant in the Barents Sea. Cod predation infers high mortality on the adult capelin, especially during capelin spawning in March [9].

Table 1. Input data in the models. T refers to the year of making the prognosis.

1	0-group T-1
2	Capelin 2 T
3	Weight 2 T
4	Herring T-1
5	Herring T
6	Cod T

3. MODEL (1): ARTIFICIAL NEURAL NETWORK- GENETIC ALGORITHM MODEL

Selecting weights for a NN itself is an optimization problem, and a GA can be applied to it, using an inverse error as the measure of utility (fitness). Whitley and his co-workers [10, 11, and 12] have done much work in this area, and the study by Montana and Davis [13] is especially ingenious and noteworthy. Hybrid approaches have also been discussed [14, 16], and there have been studies in which GA have been used to tune the parameters of other training schemes, including initial weight configurations In this research, we applied the concept of hybrid model to combine NN with GA to solve the problem of Capelin Ecosystem and we tried to predict the next year biomass of capelin. The hybrid model (Fig. 2) which is used to solve the problem consists of the following main steps. They are:

3.1 Train the NN

In this step we present data to the network. The neural network consists of a set of financial inputs. (Fig. 3)

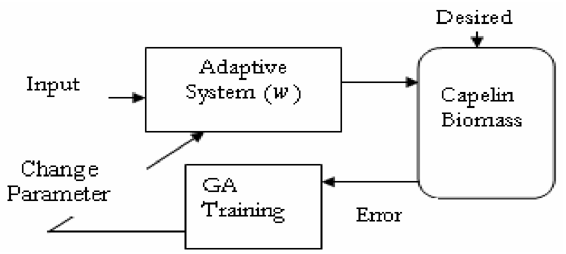

Fig. 2. Proposed Neural Network based GA model

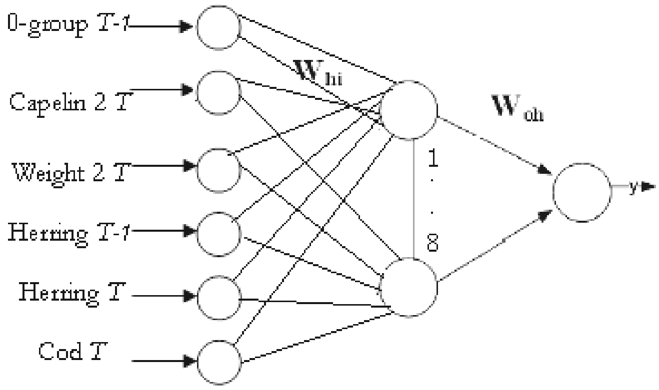

Fig. 3 NN and GA hybrid Model

Next is the NN weight optimization step. In this step the representation of weights is decided (Table 2). W_{hi} , is 6×8 matrix of synaptic weights connecting the inputs and hidden layers and W_{oh} , is 8×1 matrix of synaptic weights connecting the hidden and output layers. A floating point representation of connection weights will be used in our case for the ANN weights. Thus, the chromosome representation can be given in Table 2.

Table 2. Chromosome Representation

W_1	W_2	…	W_{n*p}	$W_{(n*p)+1}$	$W_{(n*p)+2}$	…	$W_{(n*p)+p}$

After that we start the evolutionary process. Selection, crossover, and mutation operation are applied by GA. The best individuals survive to the next generation.

Next we compute the fitness function. The fitness of these connection weights (chromosome) is computed by constructing the corresponding feed-forward neural network through decoding each chromosome and computing its fitness functions and Mean Square Error (MSE) function. In our case we adopted the Mean Square Error (MSE) as the fitness criterion as given in Equation 1.

$$MSE = \frac{1}{n}\sum_{I=1}^{n}\left(y - \hat{y}\right)^2 \quad (1)$$

The fitness of an individual is determined by the total MSE. The higher the error is the lower the fitness. y is the actual neurons output for the inputs training samples at the output layer, while $\hat{y}$ as its desired response.

After that, the built network computes an output. A one-dimensional actual output vector $y = \{y_1, y_2, ... , y_8\}$.

For simplicity, we assume that $\hat{y}$ is the desired response of the NN.

This networks output compared to desired Output based on Equation 1. Based on the computed error, network weights are modified to reduce errors.

3.2 Use the Neural Network

After training and computing the weights of the NN, we present new data to the network. The capelin Ecosystem input data (Table 1) was divided in two main subsets. Half of them were used for training the NN and the other were used for testing the NN. Network computes an output based on its training.

In our model, a population of weight set standardized using Min-Max standardization, then the weight set initiated randomly (uniform distributed in the interval [-1 1]). The GA mutation operator randomly selecting a fraction of the alleles (i.e., locations in the string) and replacing their values with values randomly selected from the set of possible values. Floating-Point-based crossover was provided , in order to take the advantage of the low-order schemata tend to be formed from neighbouring alleles, as is the case when real numbers are encoded using a binary representation. And also we run GA for 500 generations to obtain the optimal set of weights.

4. MODEL (2): MULTIPLE LINEAR REGRESSION (MLR) MODEL

MLR is used to predict one variable from one or more other variables [15]. It provides the scientist with powerful tool, allowing prediction about past, present, or future events to be made with information about past or present events. Regression analysis is one of the most widely used techniques for analyzing multifactor data. This is because of its ability to assess which factors to include and which to exclude, in order to develop alternate models with different factors. In order to construct a regression model, the following steps can be undertaken: (1) Both the information (Table 1) which is going to be used to make the prediction and the information which is to be predicted must be obtained from a sample of objects or individuals. (2) The relationship between two pieces of information is then modeled with a linear transformation. (3) In future, only the first information is necessary, and the regression model is used to transform this information in to the predicted (it is necessary to have information on both variables before the model can be constructed).

In this method, the MLR models can be represented using Equation 2.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_6 x_6 + \varepsilon \quad (2)$$

where,

- y is the price of the general index for,
- x_1 0-group *T-1*,
- x_2 Capelin 2 *T*,
- x_3 Weight 2 *T*,
- x_4 Herring *T-1*,
- x_5 Herring *T*,
- x_6 Cod *T*
- β_0 Intercept of the regression equation,
- β_i Regression coefficients, $i =$ 1, 2,3,4,5,6 and
- ε Independent $N(0, \sigma^2)$.

We may write the sample regression model corresponding to Equation 3.

$$y = \beta_0 + \sum_{j=1}^{5} \beta_j x_{ij} + \varepsilon_i \quad (3)$$

5. EXPERIMENTAL RESULTS AND DISCUSSION

The next year capelin biomass in 1979-1999 was collected from VPA for technical analysis of Capelins biomass. The first 13 years entries were used as training data. The rest 13 were testing data. The raw data is pre-processed using Min-Max Standardization. Six technical factors were selected as inputs of the model: 0-group *T-1*, Capelin 2 *T*, Weight 2 *T*, Herring *T-1*, Herring *T*, Cod *T*, and also the next year capelin biomass was predicted with different performance capability.

In order to evaluate the performance of our models, the Variance-Account-For (VAF) was chosen as given in Equation 4

$$VAF = 1 - \frac{\operatorname{var}(y - \hat{y})}{\operatorname{var}(y)} \quad (4)$$

Model (1) provides a better fit with observations than model (2), the simple one (Fig.4a, B and Fig.5a, b). Model (1) has a better overall accuracy than model (2) (Fig. 4.a., b and Fig. 5a, b).on the other hand, Model (2) provided a good productivity in training case, but it predicted poorly in testing case (See Table 3).

Table3. Computed VAF

Model (1): NN-GA (VAF)	
Training Data	Testing Data
81%	77%
Model (2): MLR (VAF)	
Training Data	Testing Data
86%	56%

In model (1), Note that the curve sub-area (Fig. 4a) numbered by 2 to 7(1979-1987) has shown a high degree of predict, while the other half of the curve area has shown some miss prediction years such as in 1989 and 1991, but in general the total biomass of capelin is well predicted by model (1) with a high degree of VAF (77%). It performed 21% better than model (2) in the testing case.

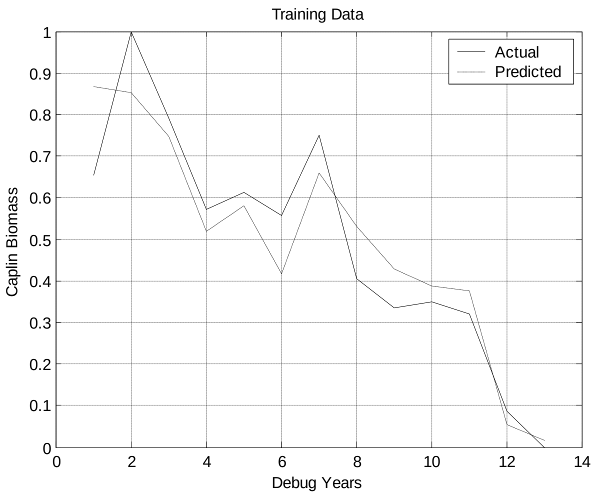

Fig. 4a. Actual and predicted Capelin biomass training case based NN-GA: model (1)

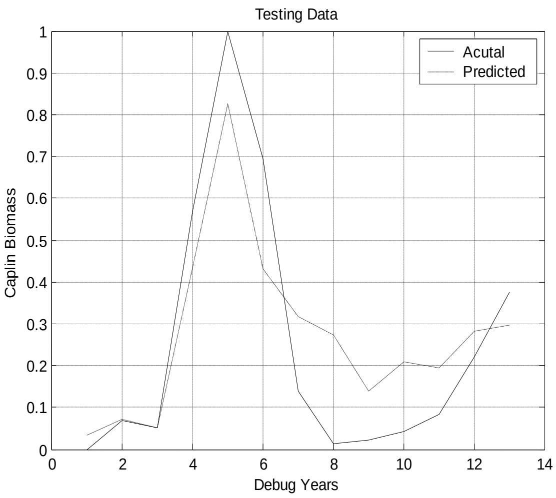

Fig. 4b. Actual and predicted Capelin biomass testing case based NN-GA: model (1)

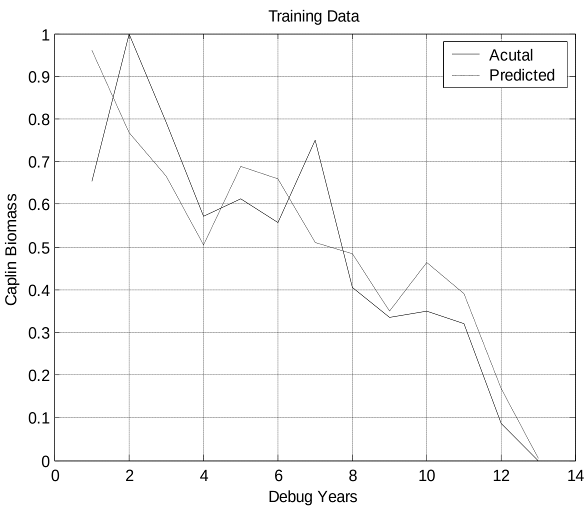

Fig. 5a. Actual and predicted Capelin biomass training case based MLR: model (2)

Resource assessment for the capelin stock is presently based on two major sources of data: From the fisheries and from scientific surveys. The capelin Stock data between 1997 and 1999 are used as input to the model. The input data was very few for the models to work efficiently, but this is the entire available data from both the fisheries and scientific surveys. In order to face this problem we decided to predict the biomass of capelin using two different models.

The results from the models have shown relatively stable constant growth for the capelin biomass which has a good impact on the capelin stock. The simple prediction model (Model (2)) has good predictive capabilities. But when we used more complicated one (Model (1)), we obtained a better accuracy. Preliminary research results proved that the hybrid model has a better overall accuracy than MLR model. The following reasons are enhance the hybrid model results:

(1) The use of ANNs which are the most effective technique in financial problem. Mainly, its ability to discover nonlinear relationship and irregularities in input data makes them ideal for predicting the stock problems market.

(2) The use of GA was a major reason to overcome the problem of shortage data availability. To avoid over-training, the number of hidden nodes was limited to 8 and the number of generations was kept at 500.

(3) Using the GA procedures over many generations, increasingly better solutions to the problem will emerge.

6. CONCLUSIONS

The experimental results show that it is possible to model Capelin stock based on capelin ecosystem factors by both models. In general, both models reach the same level of accuracy, but Model (1) provides a better fit with observations than model (2). In this work, we demonstrated how GAs can be used to optimize the NN weights, and we concluded that the use of GA was a major reason to overcome the problem of selecting NN weights and the shortage data availability.

7. REFERENCES

[1] Seafood: http://www.seafood.no/Eff/eng/effandfacts. Sf/ May 12, 2005.

[2] H.Yndestad, and A. Stene (2002). System dynamics of the Barents Sea capelin. *ICES Journal of Marine Science*, pp. 1155–1166

[3] K. Eghnam and Alaa F. Sheta (2007), Training artificial neural networks using genetic algorithms to predict the price of the general index for Amman Stock Exchange, *In the Proceedings of the Midwest Artificial Intelligence and Cognitive Science Conference, DePaul University, Chicago, IL, USA, pp. 1-7.*

[4] Vilhjálmsson, H. (1994). The Icelandic Capelin Stock. Journal of Marine Research Institute, Reykjavik, Vol. XIII, No. 1.

[5] K. Richard, Belew, J. McInerny, and N. Schraudolph (1990). Evolving Networks: Using the Genetic Algorithm with Connectionist Learning. Technical Report CS90–174, UCSD (La Jolla).

[6] Anon, (1999). Preliminary report of the international 0-group survey in the Barents Sea and adjacent waters in August-September. ICES Council Meeting.

[7] B. Anon, Ressursoversikten. (1999). Bergen, Norway: Institute of Marine Research.

[8] H. Holland (1975). Adaptation in natural and artificial systems. University of Michigan, pp.183.

[9] H. Gjosaeter, H. Loeng (1987). Growth of the Barents Sea capelin, Mallotus villosus, in relation to climate. Environmental Biology of Fishes, Volume 20, Number 4, 293-300, DOI: 10.1007/BF00005300.

[10] Darrell Whitley. Applying Genetic Algorithms to Neural Network Problems: A Preliminary Report.

[11] D. Whitley and T. Hanson (1989). The Genitor Algorithm: Using Genetic Algorithms to Optimize Neural Networks. Technical Report, pp. S-89-107, Colorado State University.

[12] D. Whitley, T. Starkweather, and C. Bogart (1989). Genetic Algorithms and Neural Networks: Optimizing

Connections and Connectivity. Technical Report, pp. 89-117, Colorado State University.

[14] D.Whitley and T. Hanson (1989). The genetic algorithm: Using genetic algorithms to optimize neural networks. Technical Report, pp. 89-107, University of Colorado state.

[13] N.J. Radcliffe (1991). Genetic Set Recombination and its Application to Neural Network Topology Optimization. *Proceedings of the 4rth International Conference on Genetic Algorithms.* pp. 222–229.

[14] N. J. Radcliffe (1990). *Genetic Neural Networks on MIMD Computers*, PhD thesis, University of Edinburgh, UK.

[15] A.S Chen, M.T. Leung, and H. Daouk (2003). Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading the Taiwan Stock Index. Computers and Operations Research pp. 901-923.

[16] K. Eghnam, A. Sheta, S. Bani-Ahmad. (2011). Optimal Weight Selection of ANN to Predict the Price of the General Index for Amman Stock Exchange. Journal of Computing, Volume 3, Issue 3, March 2011, ISSN 2151-9617.