**USER MANUAL : AGE-READING ERROR MATRIX ESTIMATOR (AGEMAT)**

André E. Punt

School of Aquatic and Fishery Sciences, Box 355020

University of Washington

Seattle, WA 98195-5020

USA

(August 2014)

**1. INTRODUCTION**

The program “Agemat” is used to compute age-reading error matrices given data on the otoliths (or any form of ageing structure) which have been read multiple times, either by the same age-reader or different age-readers. It can deal with data for which the true age is known, in addition to data for which the true age is unknown. The program can utilize multiple data sets (defined as sets of otoliths read by the same set of readers). The method, as originally published, is described in Punt *et al*. (2008). An alternative implementation has been developed by Toshi Kitakato and was used to analyse data from an age-reading experiment for Southern Hemisphere minke whales (Kitakado *et al.*, 2010).

The primary input file is “AGEMAT.DAT”. The input in AGEMAT.DAT is echoed to the file ECHO.OUT – this file should be checked if there appear to be errors specifying the inputs.

**2. INPUTS FOR AGEMAT.DAT**

The file AGEMAT.DAT provides the specifications for the data to be analysed as well as the models which are to be fitted to those data. The program is written in ADMB so comments can be included in the file by adding a line of the form “# comment”.

The first inputs specify the number of readers and the number of data sets. In the example below, there are three age-readers (“1”, “2”, and “3” and three data sets “A”, “B”, and “C”). Data set A involves otoliths read by age-readers #1 and 2, data set B involves otoliths read by age-readers #1, 2 and 3, and data set C involves otoliths read by age-readers #1 and 3.

# Maximum number of readers

3

# Number of data sets

3

The second set of inputs specifies the number of lines of input for each data set (what a “line of input” is is discussed below) and the number of readers for each data set. For example, if there are 120, 100 and 97 lines of input for each of data sets A, B and C, and 2 readers for data set A, 3 for data set B, and 2 for data set C:

# Number of points per data set

120 100 97

# Readers per data set

2 3 2

The third input specifies which readers read which data set (and hence the order in which the data will be specified). For our example, the input is specified as:

# Age reader indices per data set

1 2

1 2 3

1 3

The fourth input specifies the minimum and maximum ages and well as the reference ages. The range of ages considered must include the range of ages in the data sets. It is good practice to consider the sensitivity of the results to different choices for these ages. There must be is a reference age for each data set. The reference age should be an age which is well-represented within the data set. The reference ages are needed because the (true) proportions in each age-class are only defined up to a constant.

# Minimum and maximum ages

0 13

# Reference ages

7 9 7

The next two inputs specify the lowest and highest ages for which proportions are to be estimated (the proportions for ages between the minimum and lowest age and those for ages between the maximum and highest age are interpolated [see Equation D of the technical description for details]).

# Minus groups

5 5 5

# Plus groups

10 11 5

The next two inputs are used to specify the functional forms for ageing bias and random ageing imprecision. The available options for bias are:

* -x. Assume that the relationship between expected age and true age is the same as that for reader x [note that the “x” must be lower than the number of the reader for which bias is being defined].
* 0. Age-estimates are unbiased.
* 1. The expected age of an animal of age *a*, , is a linear function of its true age, i.e.:

. (1)

* 2. The expected age of an animal of age *a*, , is a given by:

 (2)

The available options for random ageing precision are:

* -x. Assume that the relationship between the variance of age-reading error and true age is the same as that for reader x [note that “x” must be lower than the number of the reader for which bias is being defined].
* 1. The standard deviation of random age-reading error, , is a linear function of true age, i.e.:

  (3)

* 2. The standard deviation of random age-reading error, , is given by:

  (4)

* 3. The coefficient of variation of random age-reading error, , is given by:

  (5)

* 4. The estimates of age are exact (use this option for “known age” individuals).
* 5. The standard deviation of random age-reading error, , is a spline function of age.
* 6. The standard deviation of random age-reading error, , is a piecewise linear function of age.

The age-estimates for at least one of the age-readers must be assumed to be unbiased (sensitivity to the choice of this age-reader should be examined if there is uncertainty about which age-reader is most likely to be unbiased). The following inputs cover the case where age-reader #1 is assumed to be unbiased, the bias for age-reader #2 follows Equation 2, and the bias for reader-reader #3 is the same as for age-reader #2 [this would be appropriate if “age-reader #3” was, in fact, the results of the 2nd reads by age-reader #2] while the standard deviation of age-reading error follows Equation 4 for readers #1 and #2 and the standard deviations of age-reading error for age-readers #2 and #3 are assumed to be the same.

# Option for bias

0 2 -2

# Option for standard deviation

2 2 -2

The eighth input indicates the effective sample size to assume for each data set. Entering a “0” here sets the effective samples to the actual sample sizes and setting it to a negative number will allow the program to iteratively update the effective sample sizes.

# Option for effective sample size

0 0 0

The ninth input specifies whether the initial values for the parameters should be read in from the file AGEMAT.PIN (0) or set to the initial values in the file AGEMAT.

# Use Pin File (1=Yes)

0

The tenth input provides the minimum, maximum, initial value and phase for each bias parameter (for Equation 2 the order is , *β*, )

# Bias

# Min, Max, Init, Phase

1. 40.0 1.0 2 # Bias for lowest age
2. 0.5 0.1 2 # Rate parameter

0.0 40.0 10.0 2 # Bias for oldest age

The next input provides additional specifications when the standard deviation is a spline function of age. The input for the case where age-reading error is governed by a spline function for two readers and one spline function is based on three knots and another on five knots is:

# Splines

# Number of knots; phase

3

1 20 35

5

1 5 10 15 35

No values should be provided for this input if none of the readers have standard deviation functions which are spline functions of age. The first and last points should match the minimum and maximum ages.

The twelfth input provides additional specifications when the standard deviation is a piecewise linear function of age. The input for the case where age-reading error is governed by a piecewise linear function with two breakpoints is given by:

# Linear interpolation

3

1 20 35

No values should be provided for this input if none of the readers have standard deviation functions which are piecewise linear functions of age. The first and last points should match the minimum and maximum ages.

The next input provides the minimum, maximum, initial value and phase for each standard deviation parameter (for Equation 3 the order is , *δ*, ; for Equation 4 the order is , *δ*, ; for a spline relationship the order is the value of the function each knot and the derivative at the end points; for a piecewise linear function the order is the value of the function at each point). For the case where the relationship between standard deviation and age for reader #1 is a spline function of age (three knots), that for reader #2 (two breakpoints) is piecewise linear function of age, and that for reader# 3 is Equation 4, the input is:

# SD

# Min, Max, Init, Phase for sigma and bias

# Reader #1

-10.0 10.0 1.0 1 # Spline parameter #1

-10.0 10.0 1.0 1 # Spline parameter #2

-10.0 10.0 1.0 1 # Spline parameter #3

-10.0 10.0 0.0 -1 # Derivative parameter #1

-10.0 10.0 0.0 -1 # Derivative parameter #2

# Reader #2

-10.0 10.0 1.0 1 # Piecewise linear parameter #1

-10.0 10.0 2.0 1 # Piecewise linear parameter #2

-10.0 10.0 4.0 1 # Piecewise linear parameter #3

# Reader #3

0.0 0.4 0.1 2 # Standard deviation for lowest age

0.0 0.5 0.1 2 # Rate parameter for standard deviation

0.0 0.4 0.2 2 # Standard deviation for oldest age

Note that the parameter values for a spline function and a piecewise linear function are in log-space.

The next input indicate the range and phase for the parameters which determine the relative frequency of each age-class in each data set.

# Min, Max, Phase for Probs

-20 20 2

The next input provides the minimum, maximum, initial value and phase for each slope parameter (two parameters for each data set):

# Min, Max, Init, Phase for the slopes

# Data set #1

-10 1.0 0.0 2

-10 1.0 0.0 2

# Data set #2

-10 1.0 0.0 2

-10 1.0 0.0 2

The final inputs are the data themselves. The format of the data (one block of data for each data set) are as follows:

1 2 4

2 4 3

4 4 4

1 5 7

9 6 7

The input is entered as follows. Each row represents a unique combination of ages from each age-reader. The first number on each line is the number of cases in which this combination of ages was observed. The following numbers are ages by each age-reader. The last line above implies that there were nine otoliths for which the first age-reader (who may not be age-reader #1 depending on what was specified at the third input) obtained an age-estimate of 6 and second age-reader obtained an age-estimate of 7. If there were three age-readers then the following line would be interpreted to mean that there were 16 otoliths for which the first two readers obtained an age-estimate of 7 and the third reader obtained an age-reading of 8.

16 7 7 8

Data sets often contain missing values. These can be specified by setting the age estimate to “-1”. The software has not been tested when there are many missing values so this option should be used carefully. Moreover, the estimated effective sample sizes (see Section 3) will be somewhat incorrect if there missing values.

The program will check that there are no duplicate lines of input (i.e. the same set of ages). If duplicate lines are detected a corrected data set will be written to ECHO.OUT and program will halt.

The final input in the file AGEMAT.DAT is a check so the software can determine is some inputs are missing

# Test number

123456

**3. OUTPUTS FROM AGEMAT**

The two files with outputs from AGEMAT are AGEMAT.REP and AGEMAT.STD. AGEMAT.STD contains the estimates and standard deviations for the parameters and outputs. The variable “Allout” contains the estimates of the expected age and standard deviation / CV by age and reader.

The file AGEMAT.REP contains more information. The first few lines contains summary information for the analyses. The first major block of outputs provides the expected ages, coefficients of variation and standard deviations for each combination of age and age-reader. The second major block summarizes the estimated underlying distribution of ages in each data set, and the next block lists the estimated ageing reading error matrices. The final block indicates the observed and estimated proportions for each combination of ages and the estimated effective sample sizes. If the model is fitting well (and hence the asymptotic standard errors in the AGEMAT.STD are appropriate) the ratio of the estimated to the actual effective sample size should not be substantially smaller than 1.

**4. REFERENCES**

Kitakado, T. and Punt, A.E. 2010. Examination of the period/reader effect on the age-determination for the Antartic minke whales and its implications to the statistical catch-at-age analyses. Document SC/62/IA2 presented to the IWC Scientific Committee (12pp).

Punt, A.E., Smith, D.C., KrusicGolub, K. and Robertson, S. 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in Australia’s Southern and Eastern Scalefish and Shark Fishery. *Can. J. Fish. Aquat. Sci.* 65:1991-2005.

**TECHNICAL REFERENCE**

**(based primarily on Punt *et al*. 2008)**

Assuming that (i) ageing bias depends on reader and the true age of an animal, (ii) the age-reading error standard deviation depends on true age and reader, and (iii) age-reading error is normally distributed about the expected age (i.e. the expected age given any bias in age-reading) leads to the following model for :

 (A)

where  is expected age when reader *i* determines the age of an animal of true age *a*,  is the standard deviation for reader *i* of the age-reading error for animals whose true age is *a*, and  is the vector of parameters that determines the age-reading error matrix.

The probabilities obtained from Equation 2 are set to zero for values of  < 0 and larger than a pre-specified maximum age. Ageing bias and the variance of ageing error are given by Equations 3, 4 and 5 above. The values for the parameters that determine the age-reading error matrix for each age-reader are estimated by maximizing the following likelihood function if there is some set of *J* ageing structures that are read by all readers:

 (B)

where  is the age assigned by reader ** to the *j*th ageing structure; L and H are respectively the minimum and maximum ages, and **A** is the entire data set of age-readings. The βs are nuisance parameters that can be interpreted as the relative frequency of animals of (true) age *a* in the sample (rather than in the population from which the sample was taken). The likelihood function is more generally the product of eqn B over sets of ageing structures that were all read by the same group of readers and a separate set of βs is estimated for each such set of ageing structures.

Value for the βs need not be estimated for all of the ages (especially when there are a few “outlying” ages). Instead the relative age frequencies between ages *L* and *L*1 and *H*1 and *H* can be set as exponential functions of age, i.e:

 (C)

  (D)

where ,  and  are estimated parameters. One of the s is set to 0 to ensure that the parameters are identifiable.

Equation B is modified if one of sets of age-readings are the true ages of the animals:

 (E)

where  is the true age of the *j*th ageing structure.