Automation of ALK substitution

For many years now, Age Length Key supplementation has been done manually, following the procedure described here (link). This procedure states that for each species and area, if total measured individuals for a certain age are less than 25, this area ALK should be supplemented with the data from another area following a predefined scheme of neighbour areas:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RFA | sub1 | sub2 | sub3 | sub4 | sub5 |
| 1 | 2 | 3 | 4 | 7 | 6 |
| 2 | 1 | 3 | 4 | 6 | 7 |
| 3 | 1 | 2 | 4 | 5 | 6 |
| 4 | 2 | 3 | 5 | 6 | 7 |
| 5 | 4 | 6 | 2 | 3 | 7 |
| 6 | 2 | 4 | 5 | 7 | 3 |
| 7 | 2 | 6 | 8 | 3 | 4 |
| 8 | 9 | 7 | 6 | - | - |
| 9 | 8 | 7 | 2 | - | - |
| -10 | 5 | 6 | 4 | 2 | 3 |

Table 1: Borrowing areas scheme for NS-IBTS Round Fish Areas

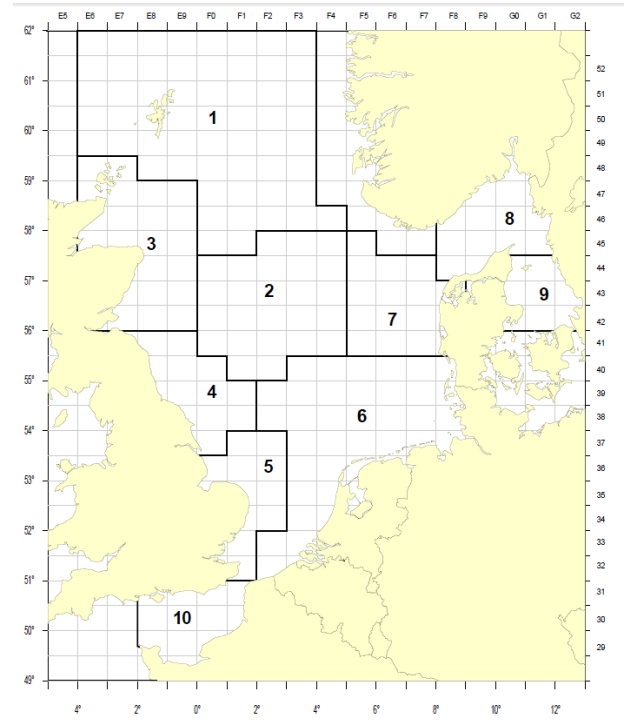


Figure 1: Standard Roundfish Areas used for roundfish since 1980, for all standard species since 1991. Additional RFA 10 added in 2009.

However, this procedure is open to interpretation, if most ages in one area have enough measures, the substitution might not be done. This procedure follows the logic of the person performing the substitution, but it is difficult for the final user to trace back.

The present procedure aims to provide an automated way of substituting ALK and to compare the resulting indices with those calculated with the old manual procedure of substitution.



Figure x: NS-IBTS in 2017 quarter 1 original ALK as submitted to Datras. Each row corresponds to one of the 10 target species and each column corresponds to the 10 RFA (Figure x). In red, the age classes that for that area have less than 25 data points, so should be supplemented with nearby areas.



Figure y: NS-IBTS for 2017 quarter 1 ALK after manual substitution. In red are age classes that have been supplemented. In this case, all age classes in one area appear supplemented, because the old procedure adds up the whole ALK, not taking into account the different Age classes.

In the automated procedure, for each species, area and age class, it is checked whether there are at least 25 observations. If not, that age class (not all of them as before) will be supplemented with the data from the nearest area (see Table 1) and so on until all age classes in all areas have at least 25 data points (or until all possible supplementations have been performed).



Figure z: NS-IBTS 2017 quarter 1 after automated substitution. Please note that for each area, only age classes with less than 25 observations (in red color in figure x, original ALK), have been supplemented.

The effect in the indexes is small:

(figures)

The manual procedure seems to provide quite similar indices results to those resulting from the automated substitution.

The manual procedure is more flexible, in the sense that in some cases the automated procedure might supplement too much when most ALK is well populated.

However, when the substitution is done with the manual procedure, the supplementation affects the whole area, resulting in some cases in exact ALKs across all areas.

The advantages of the automated procedure mostly refer to transparency and traceability. The substitution procedure can be stored in TAF, so users can replicate it if needed. Furthermore it is much faster than the manual procedure, so it would allow some more time for working groups to check it and in case ask for modifications (in plus groups, in borrowing areas scheme etc). This proposed modifications could be coded and replicated every year.

The full procedure to produce these figures is in github: