

Appendix A. Supplementary material for "Model validation for compositional data in stock assessment models: calculating residuals with correct properties"

Vanessa Trijoulet^{a,*}, Christoffer Moesgaard Albertsen^a, Kasper Kristensen^a,
Christopher M. Legault^b, Timothy J. Miller^b, Anders Nielsen^a

^a*National Institute of Aquatic Resources, Technical University of Denmark, Kemitorvet
201, DK-2800 Kgs. Lyngby, Denmark*

^b*Northeast Fisheries Science Center, National Marine Fisheries Service, National
Oceanic and Atmospheric Administration, 166 Water Street, Woods Hole, MA 02543,
USA*

A1. Additional results for the OSA residual validation examples

The probability of the p-values of the normality and correlation tests for the validation exercise are given in Table A1.1. The distributions of the mean and standard deviation of the residuals are given in Figures A1.1-A1.3. Figures A1.4-A1.6 show the distribution of the correlation between compositional groups and Figures A1.7-A1.9 show the distribution of the lag 1 year correlation per compositional group.

Regarding the OSA residuals under the true model, the OSA residuals have mean 0 and standard deviation 1 (Figures A1.1-A1.3) except for the Dirichlet model for which the standard deviation of the OSA residuals has a

*Corresponding author

Email addresses: vtri@aqu.dtu.dk (Vanessa Trijoulet), cmoe@aqu.dtu.dk (Christoffer Moesgaard Albertsen), kaskr@aqu.dtu.dk (Kasper Kristensen), chris.legault@noaa.gov (Christopher M. Legault), timothy.j.miller@noaa.gov (Timothy J. Miller), an@aqu.dtu.dk (Anders Nielsen)

slight positive bias (Figure A1.2). This is due to the estimation of residual variance. Indeed, this is corrected when the true parameters are used for the OSA residual calculation (Figure A1.10). The Figure A1.10 also illustrates how using the maximum likelihood estimator reduces the variation in the mean and standard deviation of the residuals and this occurs for all distributions (Figures A1.11-A1.12). As a result, the residuals can appear more precise than they actually should under the true model. The results of the correlation tests under the true model are as expected with rejection probabilities close to 0.05 for all correlations and distributions (Table A1.1) and with compositional groups that are clearly independent (Figures A1.4-A1.9). The normality tests results are overall good, but for the Kolmogorov-Smirnov test (as implemented in the R function `ks.test`), the rejection probability is too small (0 instead of 0.05). This is due to the fact that the residuals are calculated using maximum likelihood estimated parameters, while the rejection region of the Kolmogorov-Smirnov tests is based on true parameters. As a result, the probability of rejecting the test can be smaller than expected for the Kolmogorov-Smirnov given that this test checks for standard normality (mean 0 and standard deviation 1) and is therefore more specific than the other two. In contrast, the rejection regions of the Shapiro-Wilks (as implemented in the R function `shapiro.test`) and Anderson-Darling (as implemented in the function `ad.test` in the R-package `nortest` Gross and Ligges, 2015) tests are based on the empirical mean and variance of the residuals and are corrected for estimation. This is clearly illustrated in Table A1.1 with a rejection probability close to 0.05 for the Kolmogorov-Smirnov test when the true parameters are used when estimating the residuals.

Regarding the OSA residuals under the false model, the power of the correlation tests is low for the multinomial model. This is due to the fact that the simulated over-dispersion creates correlation between data groups (most visible between groups 2 and 3 in Figure A1.4) but the residuals are still independent across samples (lag 1 year correlation) since the over-dispersion does not induce a sample correlation (e.g. correlation over time). When a selectivity change was simulated for the first two data groups (Dirichlet and Dirichlet-multinomial), both the lag 1 year correlation and the correlation between the two compositional groups are clearly identified with maximal power (1) (Table A1.1). The high correlation is also clearly visible in Figures A1.5-A1.6 for the groups 1 and 2 and in Figures A1.8-A1.9 where the correlation due to the change in selectivity over time for the two groups is evident. The power of the tests for normality differs depending on the statistical test used under the false model. This is due to the violations that were created in the study. Indeed, in the case of the multinomial distribution, the Kolmogorov-Smirnov does best because the over-dispersion mainly affects the standard deviation of the residuals, which becomes larger than 1 (Figure A1.1). Given that the two other tests only test for normality and not standard normality, the false model is not rejected enough and the power of the two tests is low. This is expected, since the larger standard deviation is not a violation of the hypothesis tested for in the Shapiro-Wilks and Anderson-Darling tests. For the Dirichlet and Dirichlet-multinomial distributions for which a change of mean has been simulated over time, both the mean and the standard deviation of the residuals are affected (Figures A1.2-A1.3). The Shapiro-Wilks and Anderson-Darling tests do best in these cases. This is due

to the fact that the cumulative distribution function of the false model is very close to the one of the true model so the power of the Kolmogorov-Smirnov test is low.

Table A1.1: Proportions of rejected statistical tests (p-values ≤ 0.05) over the 1000 iterations. Ideally, for the OSA residuals under the true model, the rejection probability should be 0.05. If the statistical tests are powerful, the OSA residuals under the false model should have a probability close to 1. The residuals using the true parameters were also calculated to see how this can impact the rejection probabilities. KS stands for Kolmogorov-Smirnov, SW for Shapiro-Wilks, and AD for Anderson-Darling.

| | | Case | Test for normality | | | Correlation groups | | | Correlation lag 1 | | |
|-------|------------|-----------------------|--------------------|------|------|--------------------|-------|-------|-------------------|------|------|
| Model | Parameters | Distribution | KS | SW | AD | 1 & 2 | 1 & 3 | 2 & 3 | 1 | 2 | 3 |
| True | Estimated | Multinomial | 0.00 | 0.05 | 0.05 | 0.05 | 0.04 | 0.05 | 0.05 | 0.05 | 0.06 |
| True | Estimated | Dirichlet | 0.00 | 0.05 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.03 | 0.04 |
| True | Estimated | Dirichlet-multinomial | 0.00 | 0.04 | 0.04 | 0.06 | 0.06 | 0.05 | 0.04 | 0.06 | 0.05 |
| False | Estimated | Multinomial | 0.93 | 0.07 | 0.06 | 0.05 | 0.10 | 0.29 | 0.05 | 0.04 | 0.04 |
| False | Estimated | Dirichlet | 0.92 | 1.00 | 1.00 | 1.00 | 0.05 | 0.05 | 1.00 | 1.00 | 0.04 |
| False | Estimated | Dirichlet-multinomial | 0.24 | 0.98 | 0.96 | 1.00 | 0.06 | 0.06 | 1.00 | 1.00 | 0.06 |
| True | True | Multinomial | 0.05 | 0.04 | 0.04 | 0.05 | 0.04 | 0.05 | 0.04 | 0.05 | 0.06 |
| True | True | Dirichlet | 0.04 | 0.06 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.03 | 0.04 |
| True | True | Dirichlet-multinomial | 0.05 | 0.04 | 0.04 | 0.06 | 0.06 | 0.05 | 0.04 | 0.06 | 0.05 |

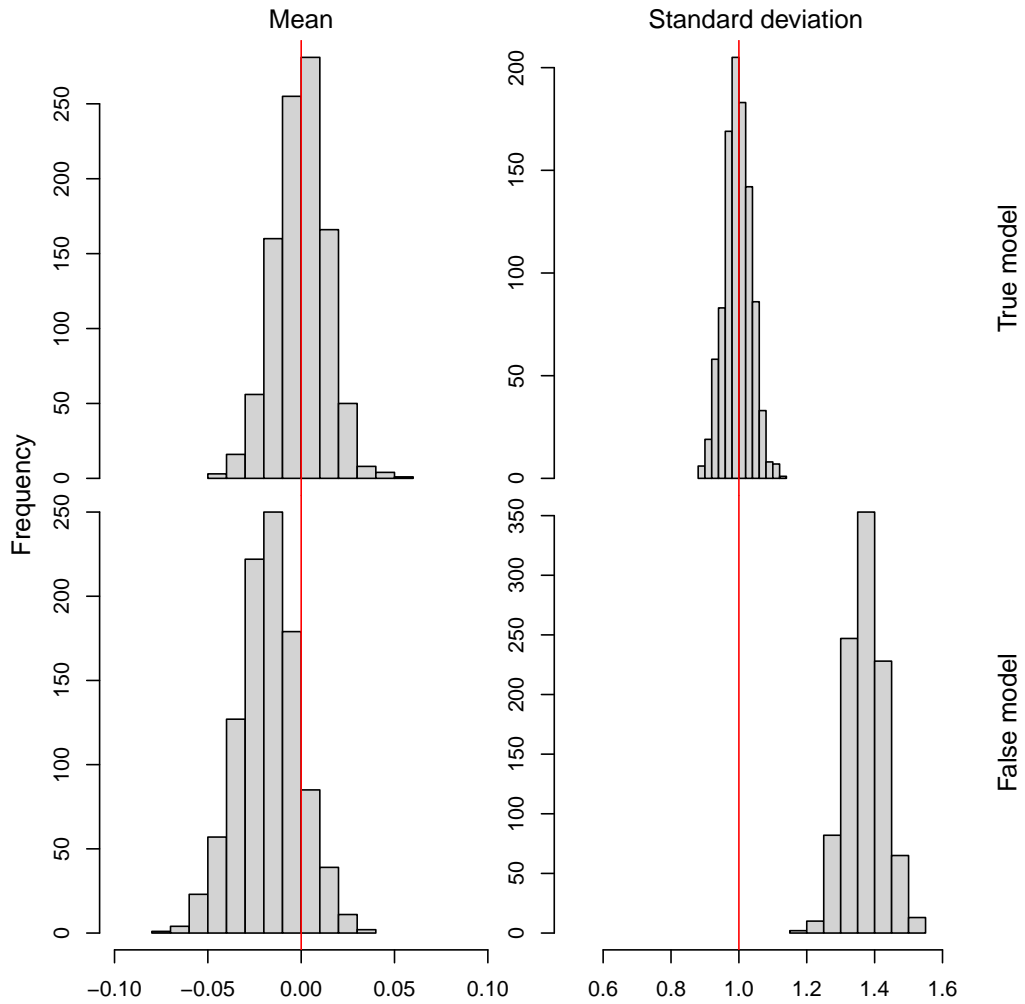


Figure A1.1: Histogram of the mean and standard deviation across the 1000 replicates of OSA residual calculations for the multinomial validation example. Perfect residuals should have a mean of 0 and a standard deviation of 1.

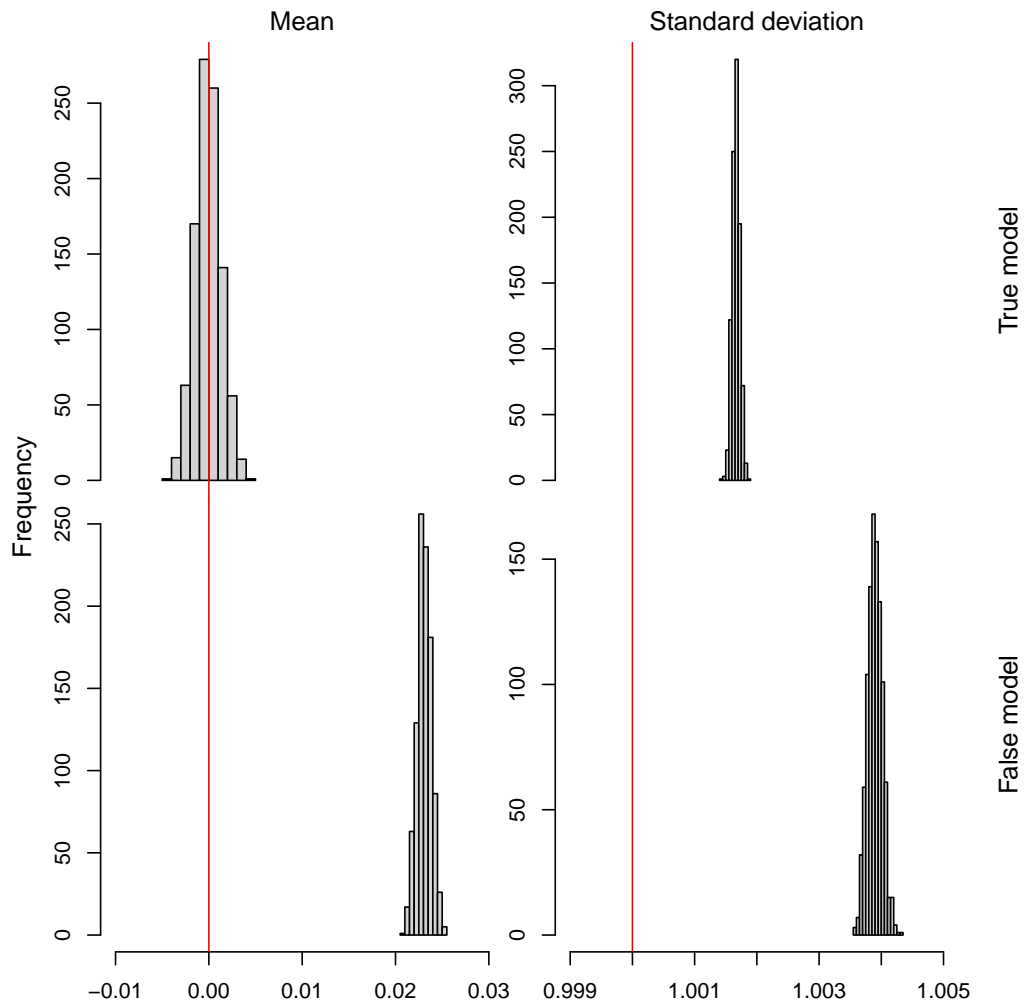


Figure A1.2: Histogram of the mean and standard deviation across the 1000 replicates of OSA residual calculations for the Dirichlet validation example. Perfect residuals should have a mean of 0 and a standard deviation of 1.

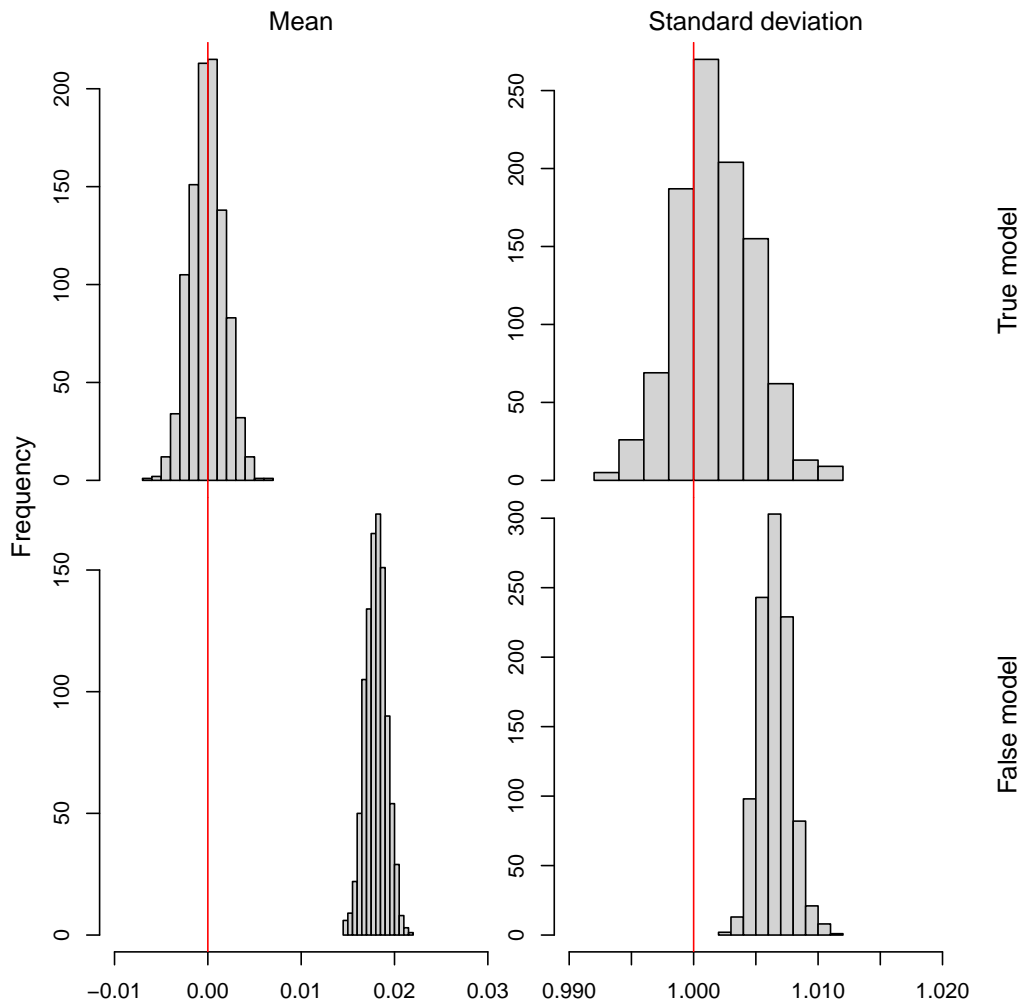


Figure A1.3: Histogram of the mean and standard deviation across the 1000 replicates of OSA residual calculations for the Dirichlet-multinomial validation example. Perfect residuals should have a mean of 0 and a standard deviation of 1.

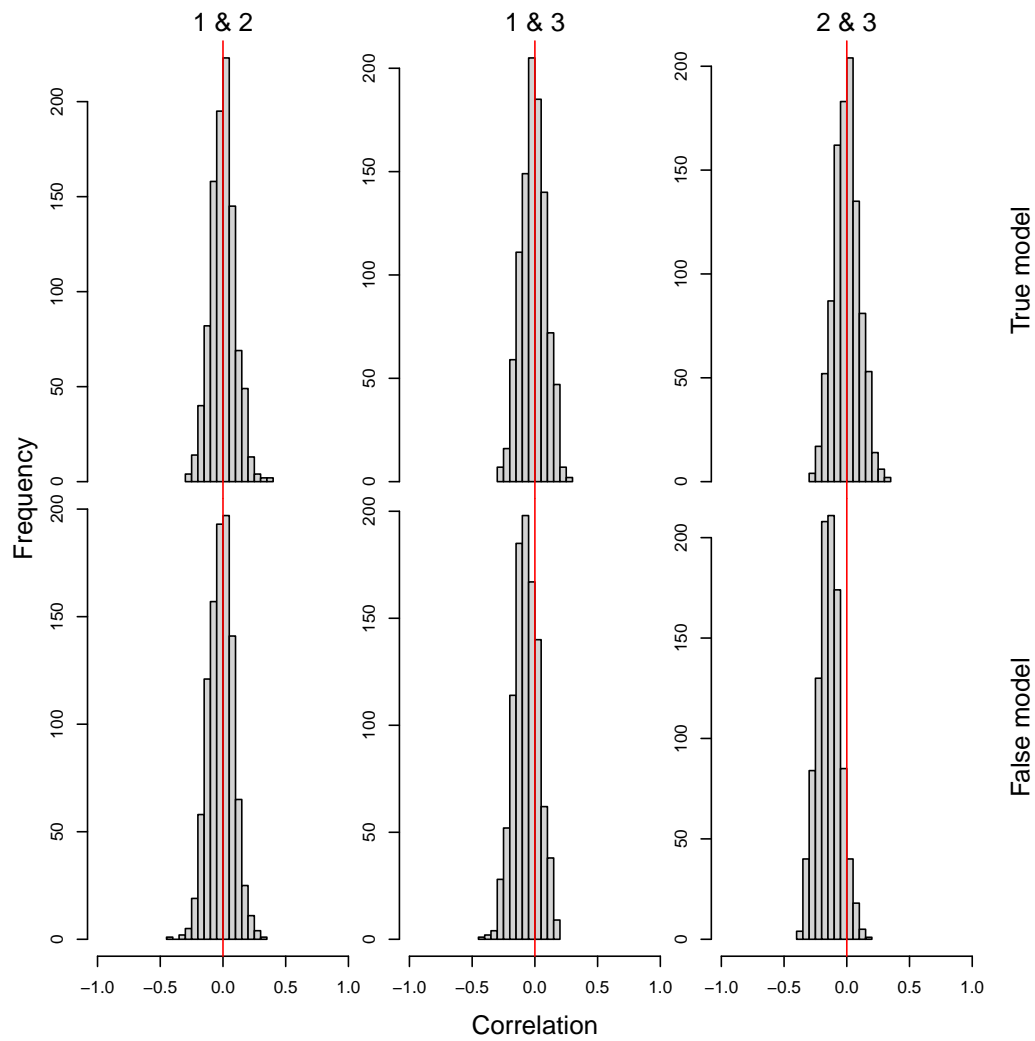


Figure A1.4: Histogram of the correlation between compositional groups of the 1000 replicates of OSA residual calculations for the multinomial validation example. Independent residuals should have a correlation of 0.

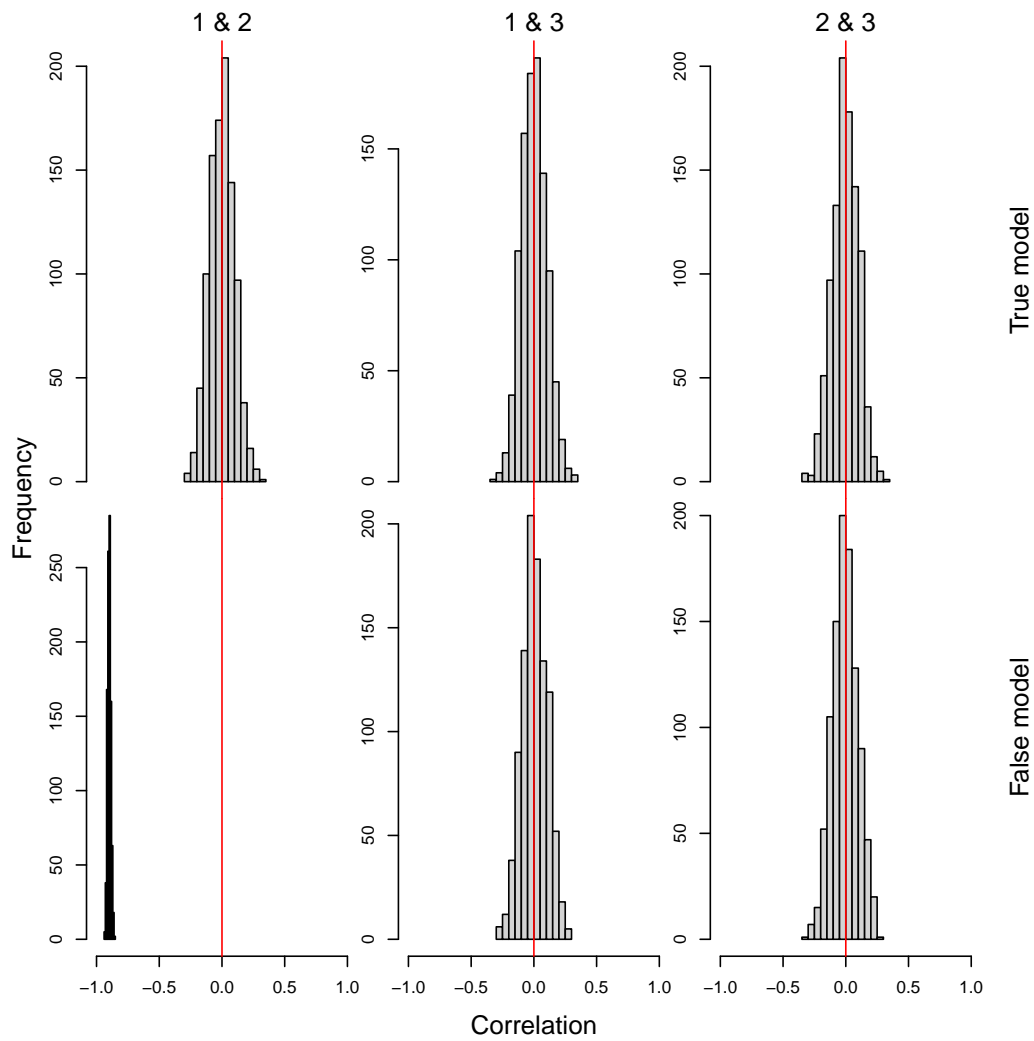


Figure A1.5: Histogram of the correlation between compositional groups of the 1000 replicates of OSA residual calculations for the Dirichlet validation example. Independent residuals should have a correlation of 0.

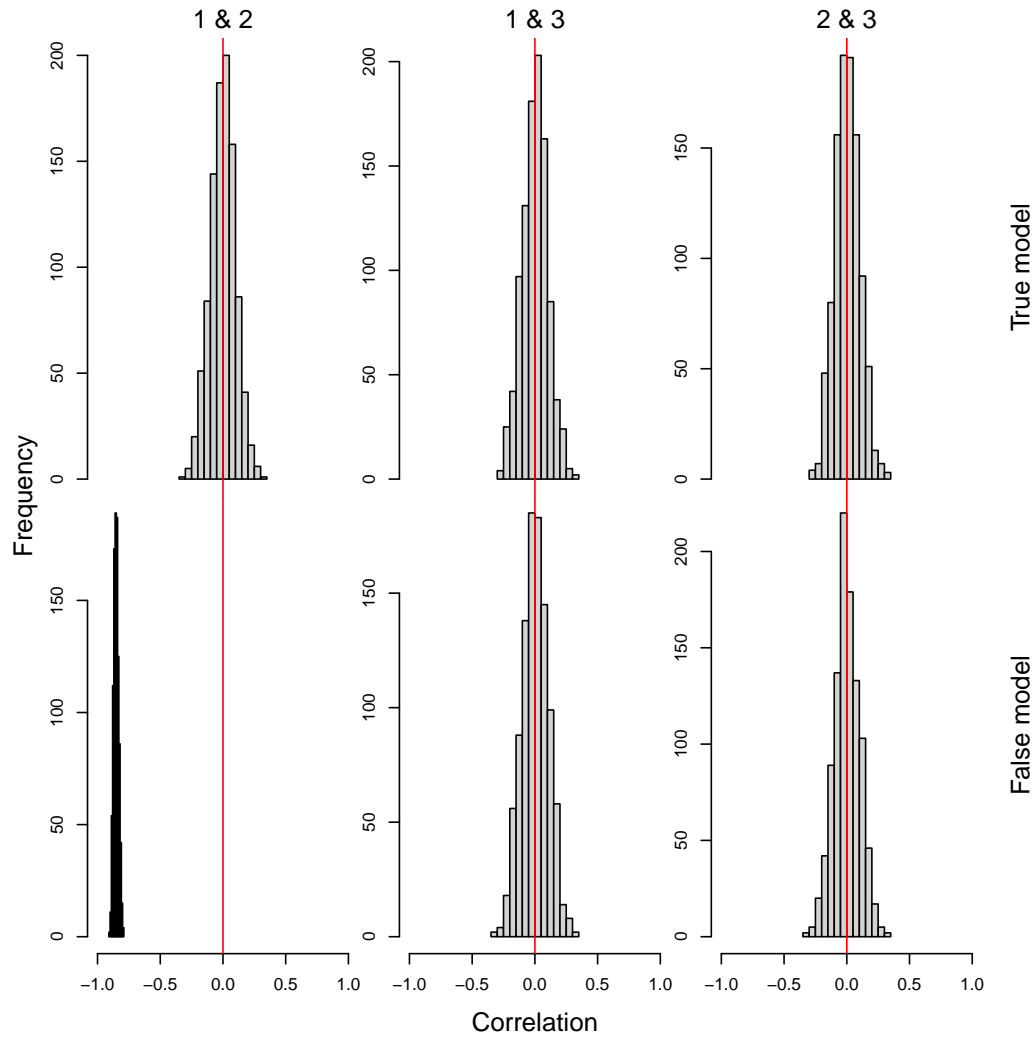


Figure A1.6: Histogram of the correlation between compositional groups of the 1000 replicates of OSA residual calculations for the Dirichlet-multinomial validation example. Independent residuals should have a correlation of 0.

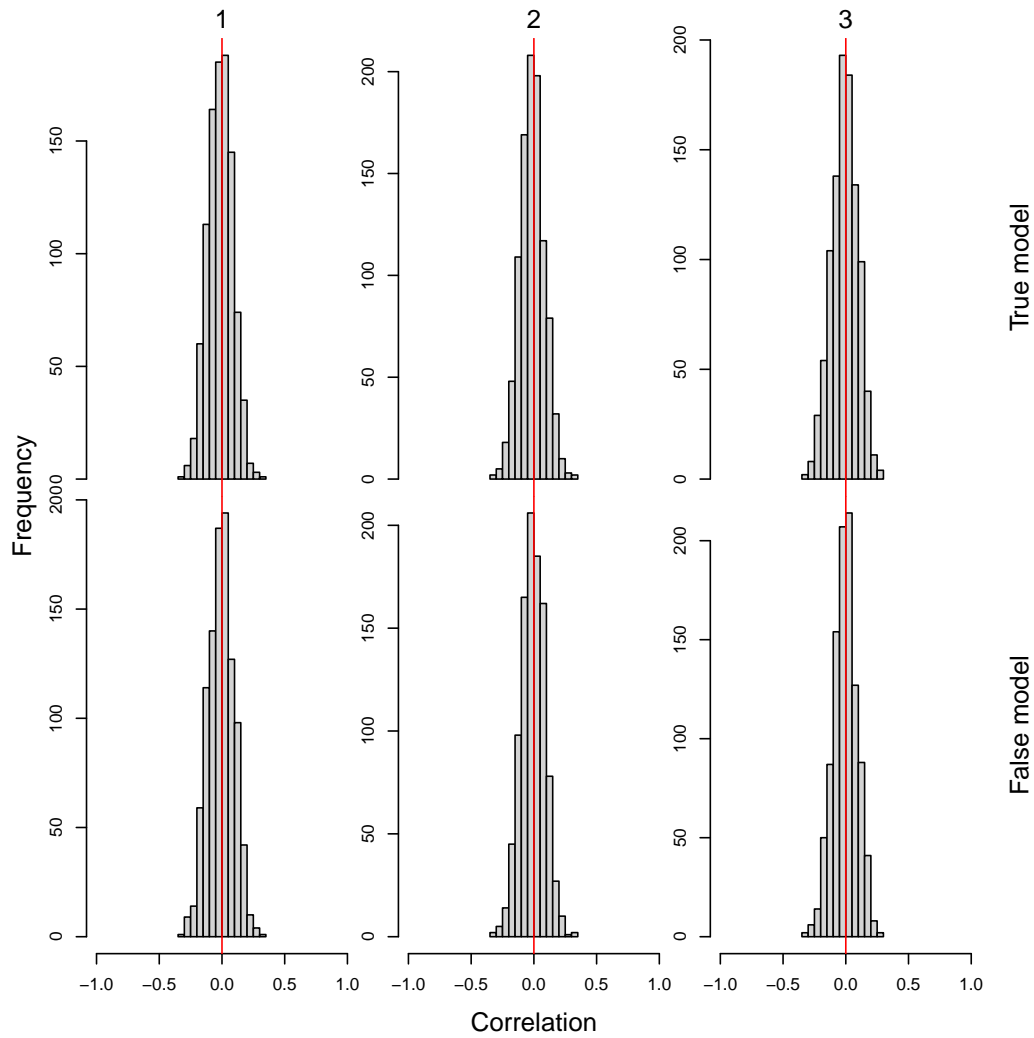


Figure A1.7: Histogram of the lag 1 year correlation for each compositional group of the 1000 replicates of OSA residual calculations for the multinomial validation example. Independent residuals should have a correlation of 0.

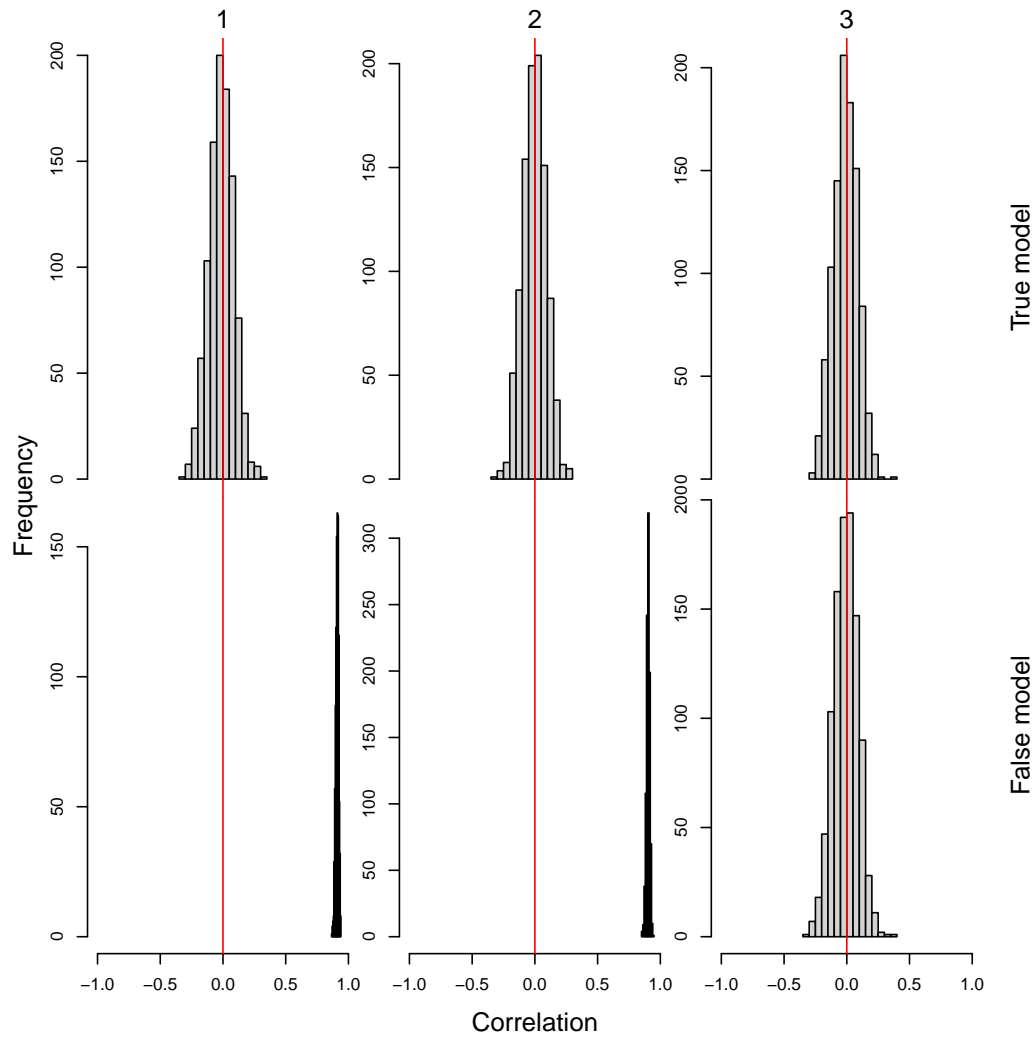


Figure A1.8: Histogram of the lag 1 year correlation for each compositional group of the 1000 replicates of OSA residual calculations for the Dirichlet validation example. Independent residuals should have a correlation of 0.

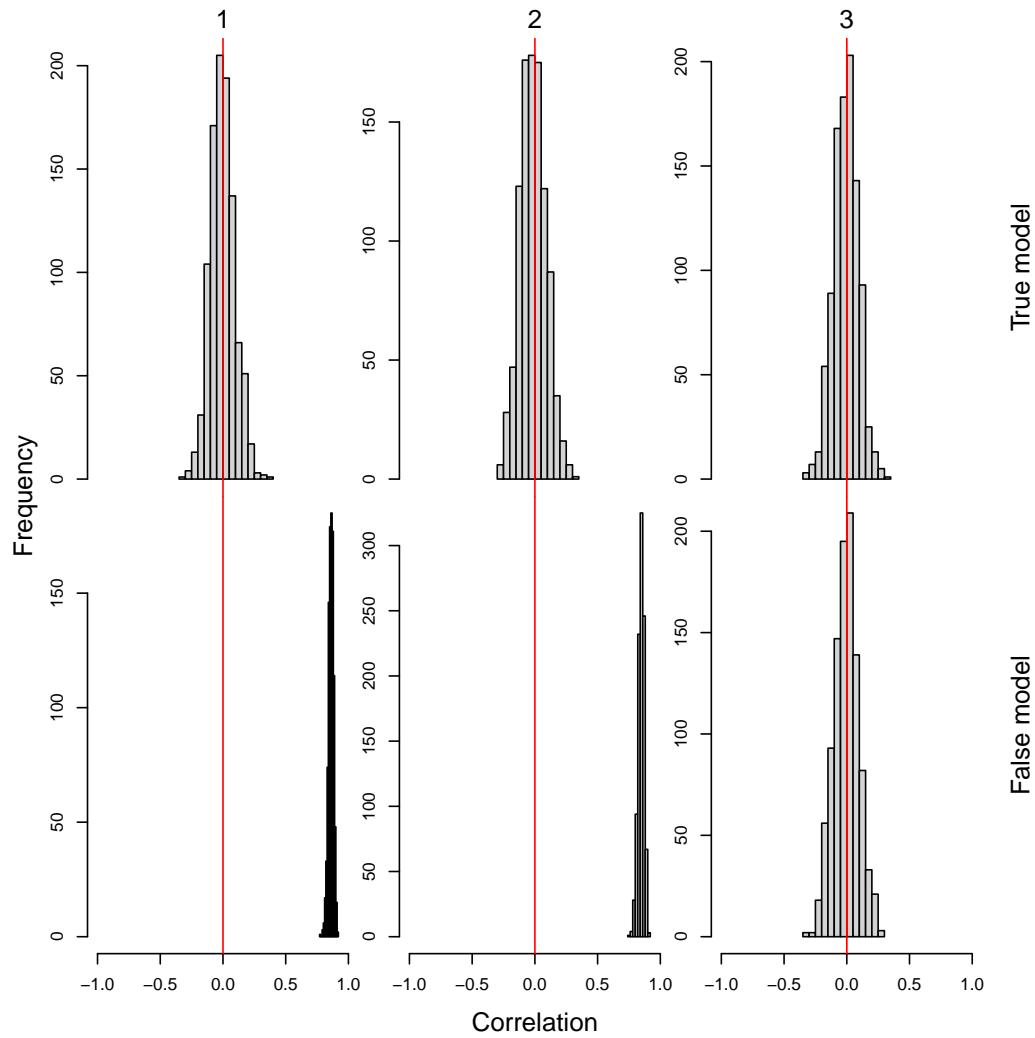


Figure A1.9: Histogram of the lag 1 year correlation for each compositional group of the 1000 replicates of OSA residual calculations for the Dirichlet-multinomial validation example. Independent residuals should have a correlation of 0.

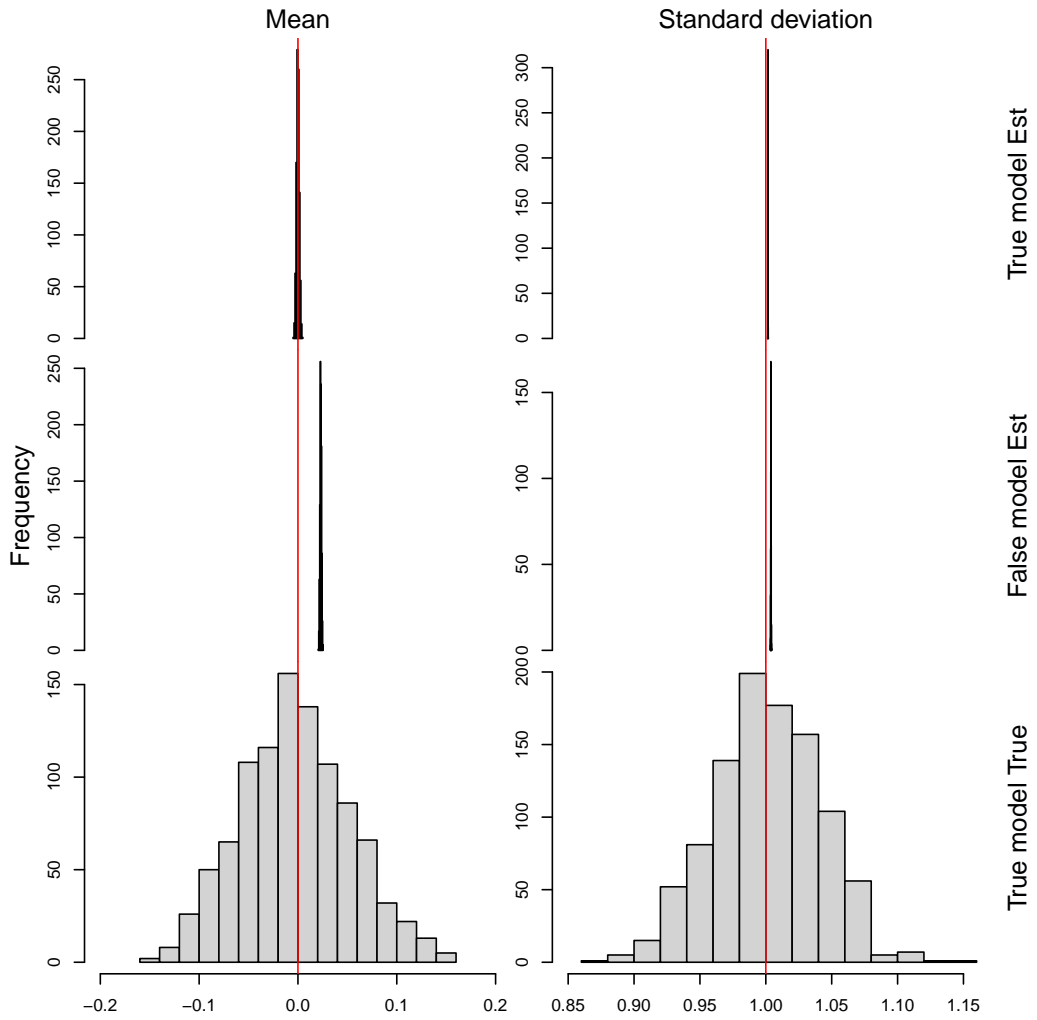


Figure A1.10: Histogram of the mean and standard deviation across the 1000 replicates of OSA residual calculations for the Dirichlet validation example including the results when the true parameters are used in the residual calculation. Perfect residuals should have a mean of 0 and a standard deviation of 1.

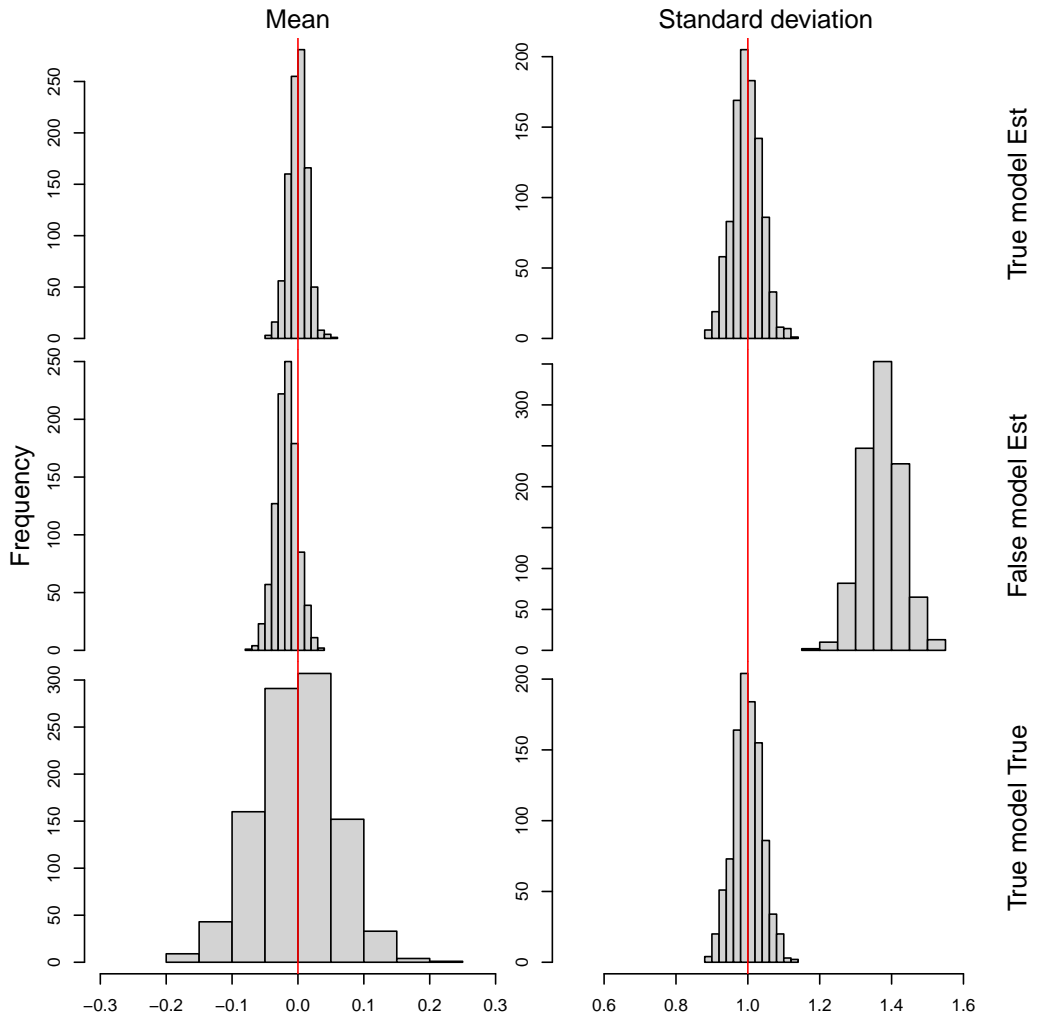


Figure A1.11: Histogram of the mean and standard deviation across the 1000 replicates of OSA residual calculations for the multinomial validation example including the results when the true parameters are used in the residual calculation. Perfect residuals should have a mean of 0 and a standard deviation of 1.

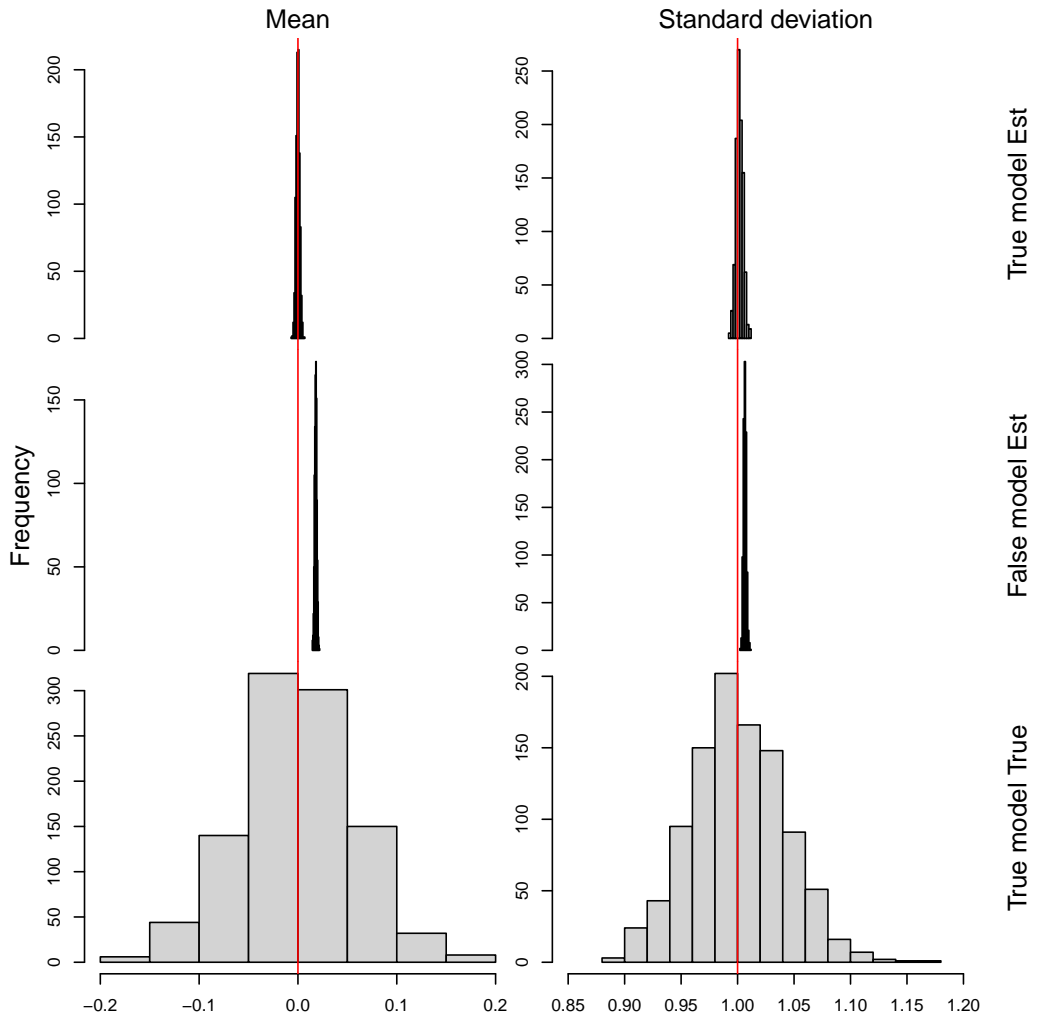


Figure A1.12: Histogram of the mean and standard deviation across the 1000 replicates of OSA residual calculations for the Dirichlet-multinomial validation example including the results when the true parameters are used in the residual calculation. Perfect residuals should have a mean of 0 and a standard deviation of 1.

A2. Additional figures for the Gulf of Maine haddock example with simulated observations

The residuals estimated from the simulated example informed by the Gulf of Maine haddock assessment for the first iteration of fleet 1 are given in Figure 4. The results for all the other fleet replicates are given in Figures A2.1-A2.14.

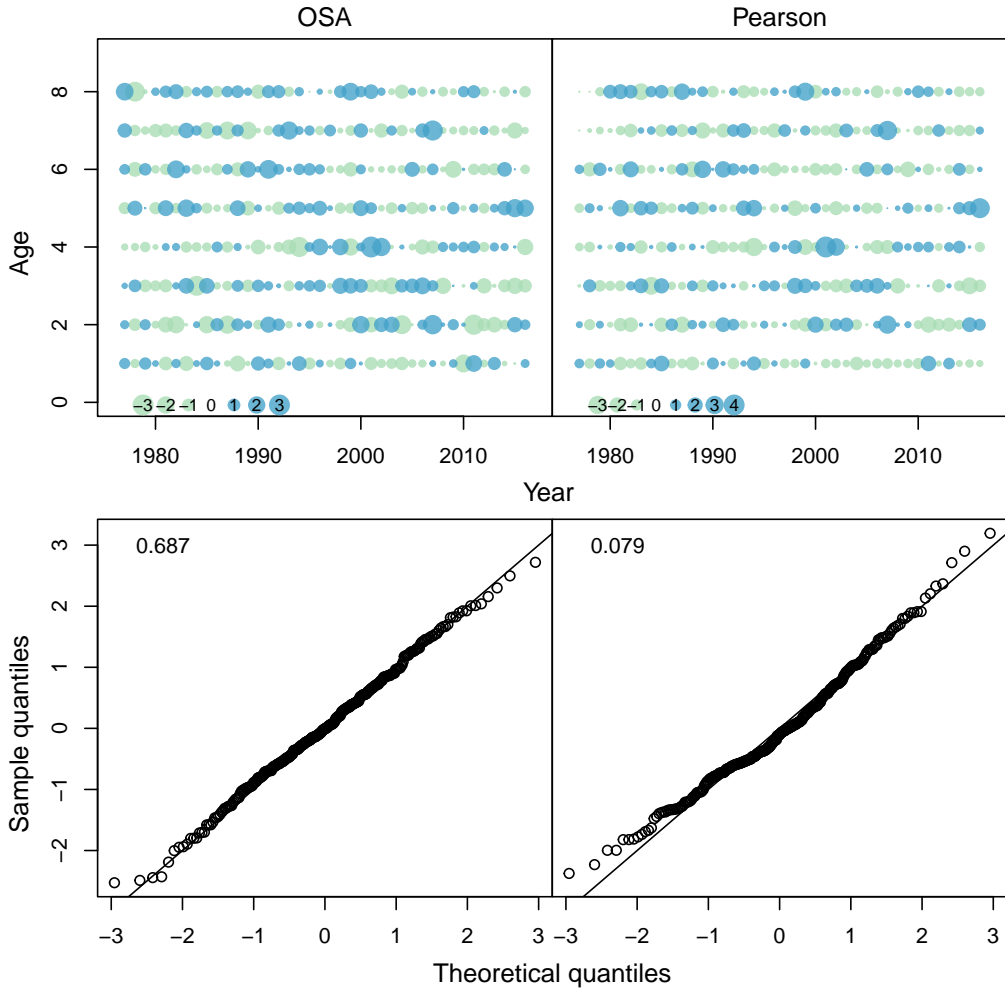


Figure A2.1: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 1st replicate of fleet 1 for the Gulf of Maine haddock simulated example. Each bubble is scaled to the residual's size and colored given its sign (positive in blue or negative in green). The p-value of the Kolmogorov-Smirnov test for normality is given in the top left of the Q-Q plots and the line is the identity line.

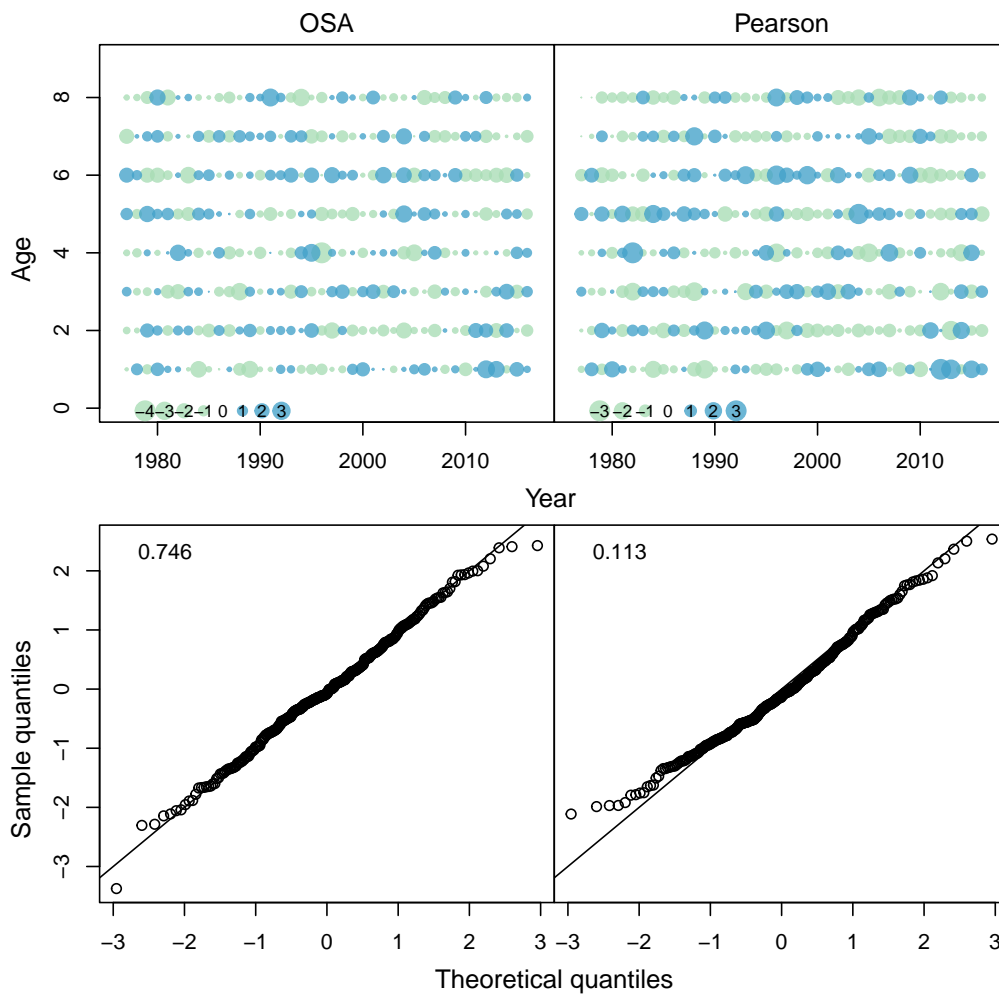


Figure A2.2: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 1st replicate of fleet 3 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

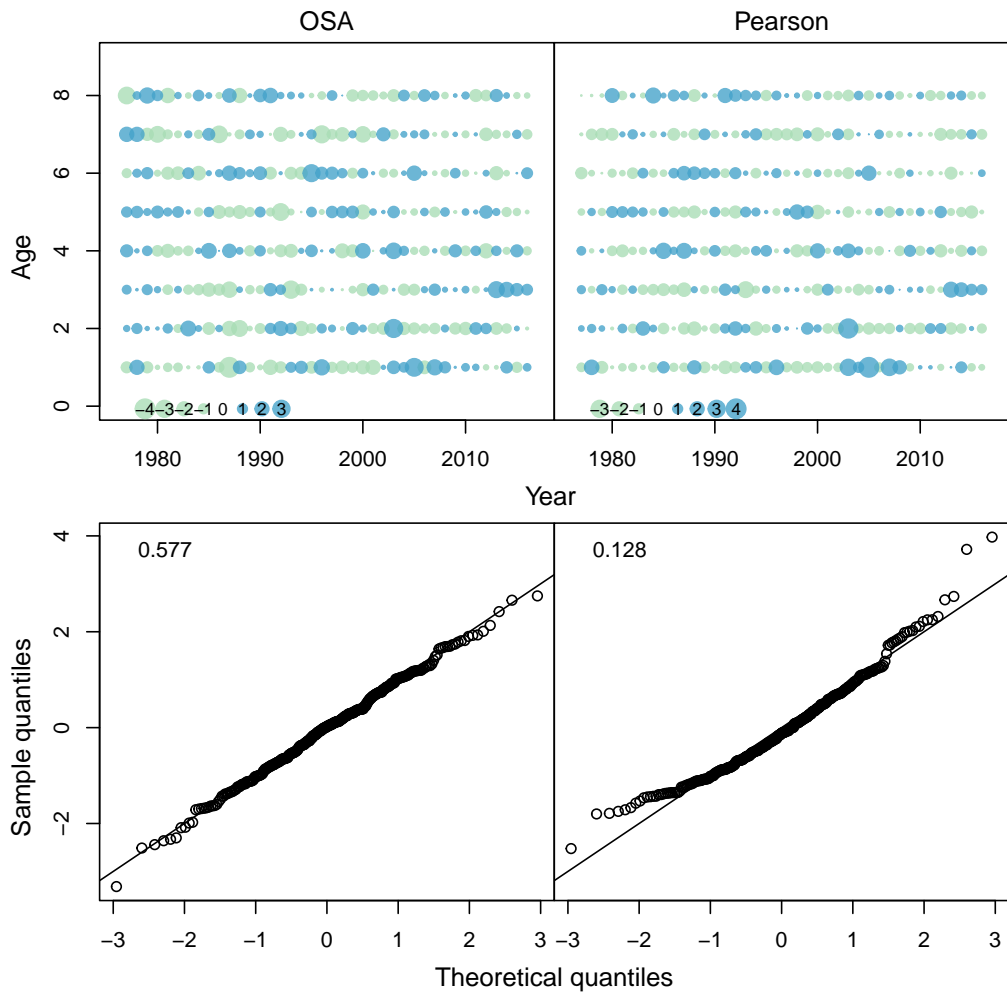


Figure A2.3: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 2nd replicate of fleet 1 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

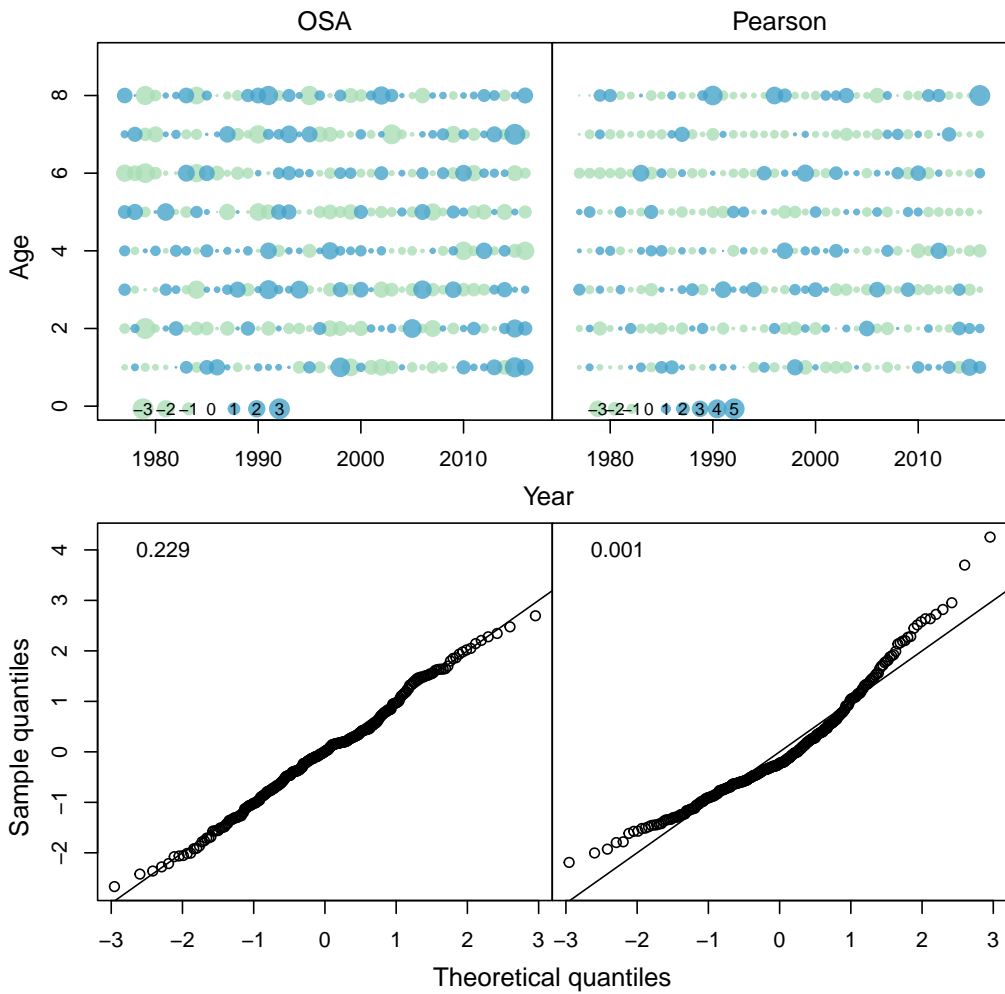


Figure A2.4: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 2nd replicate of fleet 2 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

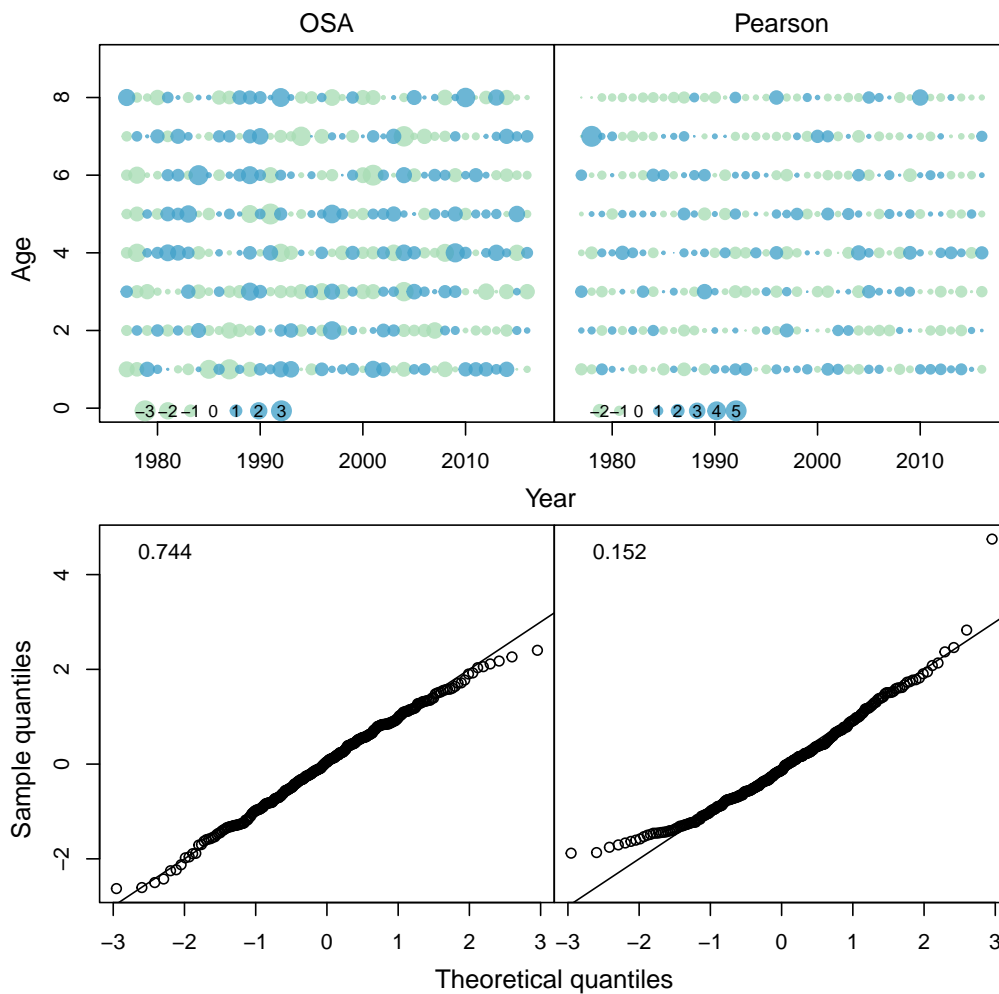


Figure A2.5: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 2nd replicate of fleet 3 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

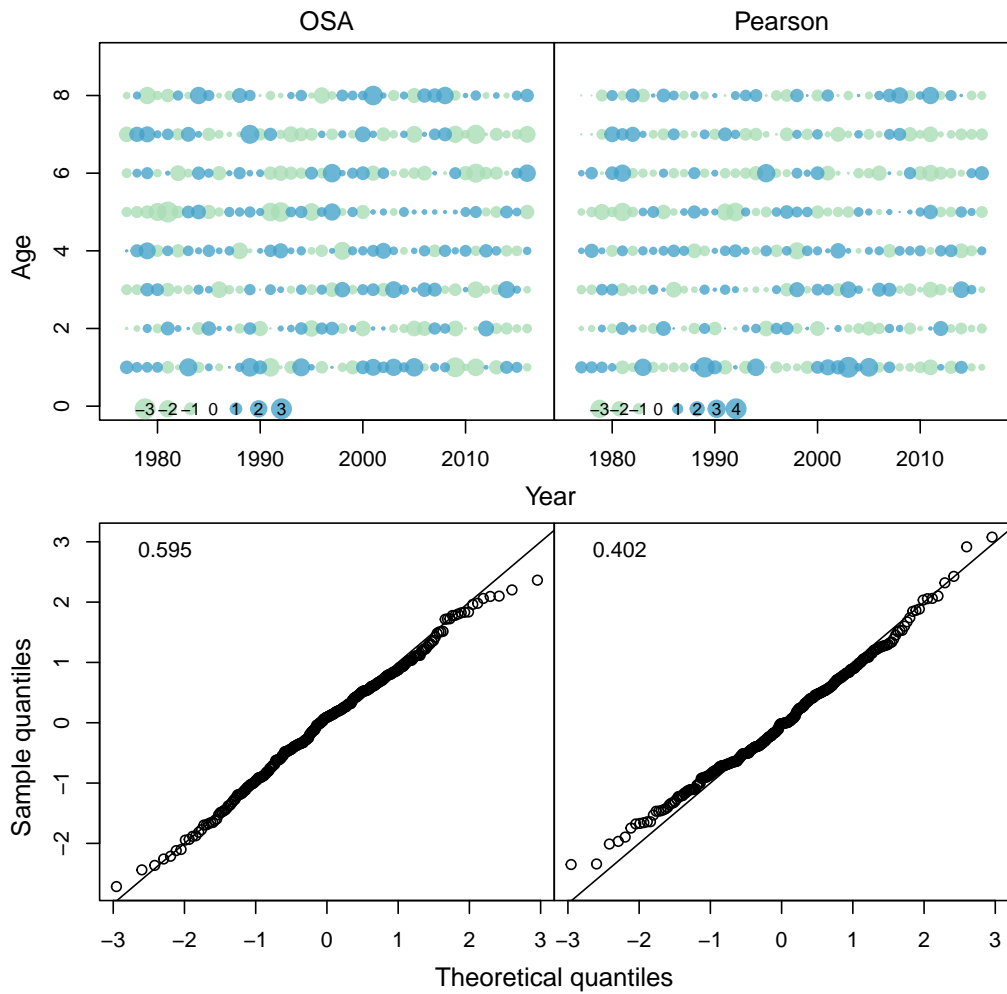


Figure A2.6: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 3rd replicate of fleet 1 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

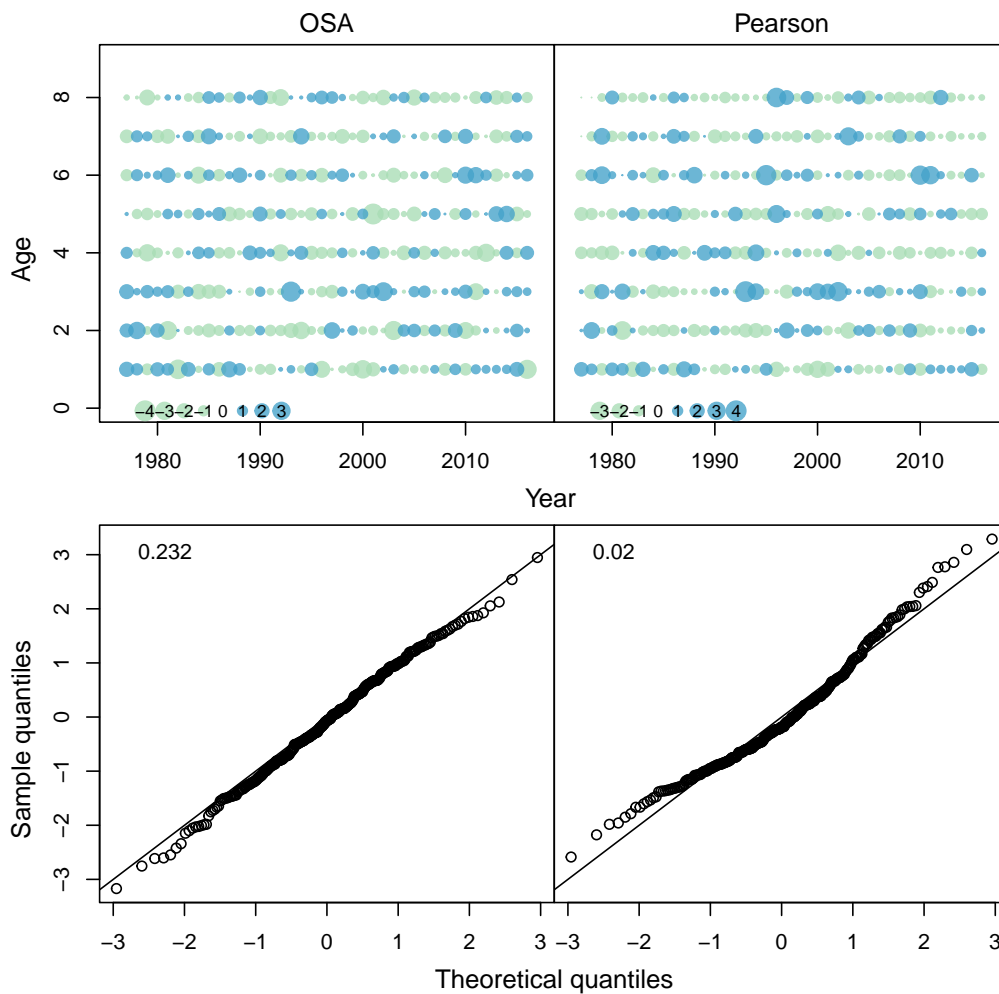


Figure A2.7: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 3rd replicate of fleet 2 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

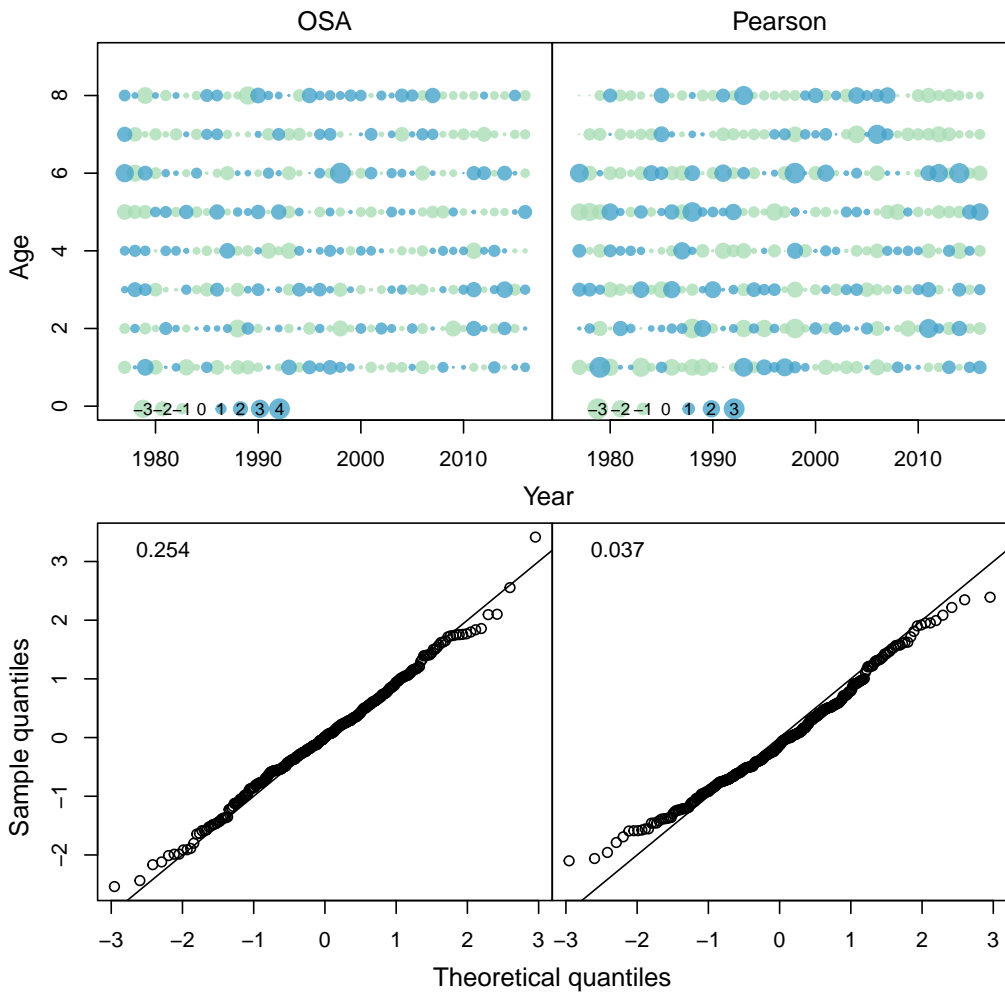


Figure A2.8: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 3rd replicate of fleet 3 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

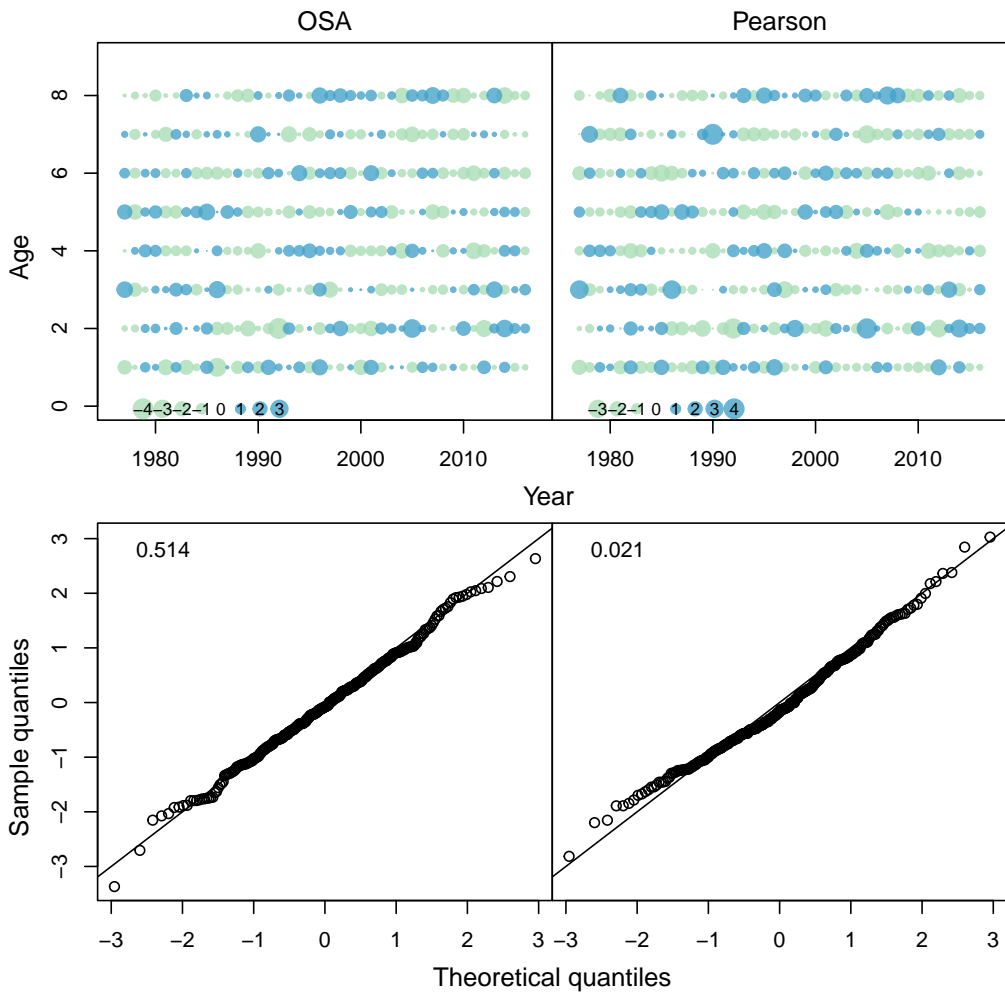


Figure A2.9: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 4th replicate of fleet 1 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

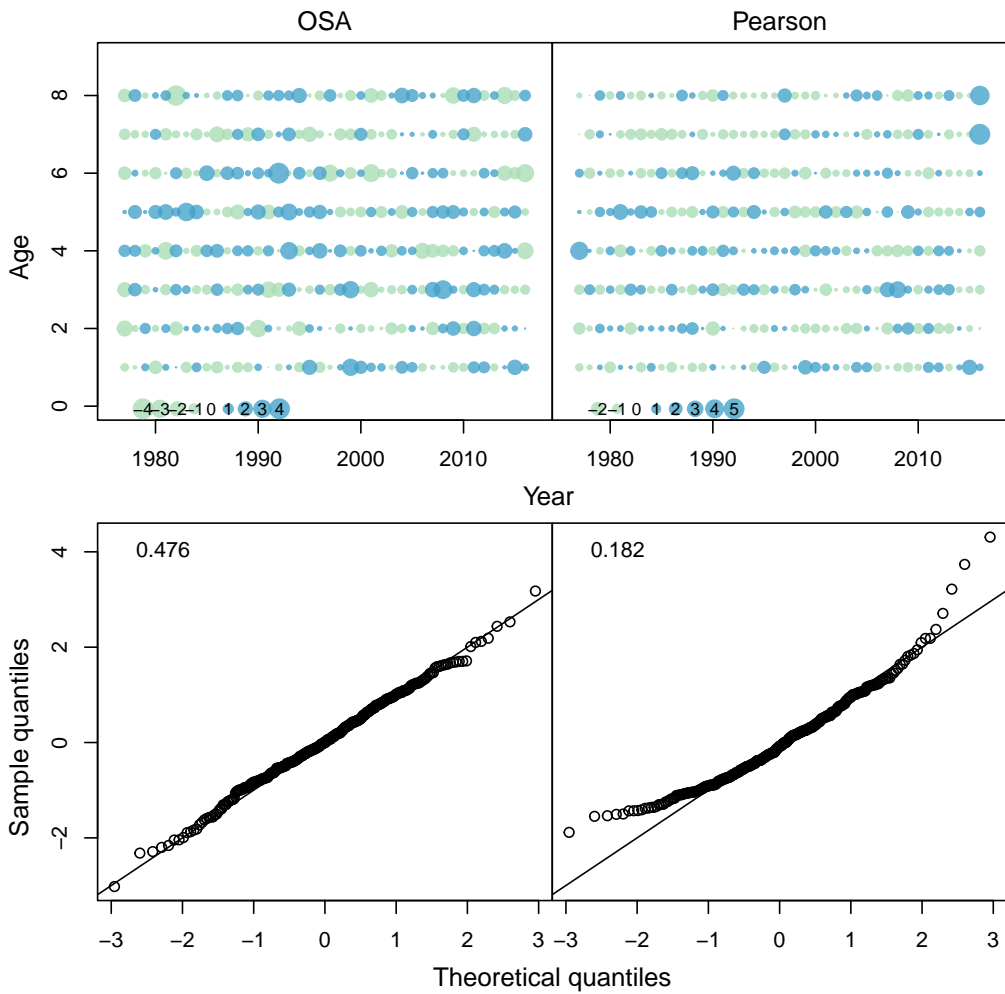


Figure A2.10: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 4th replicate of fleet 2 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

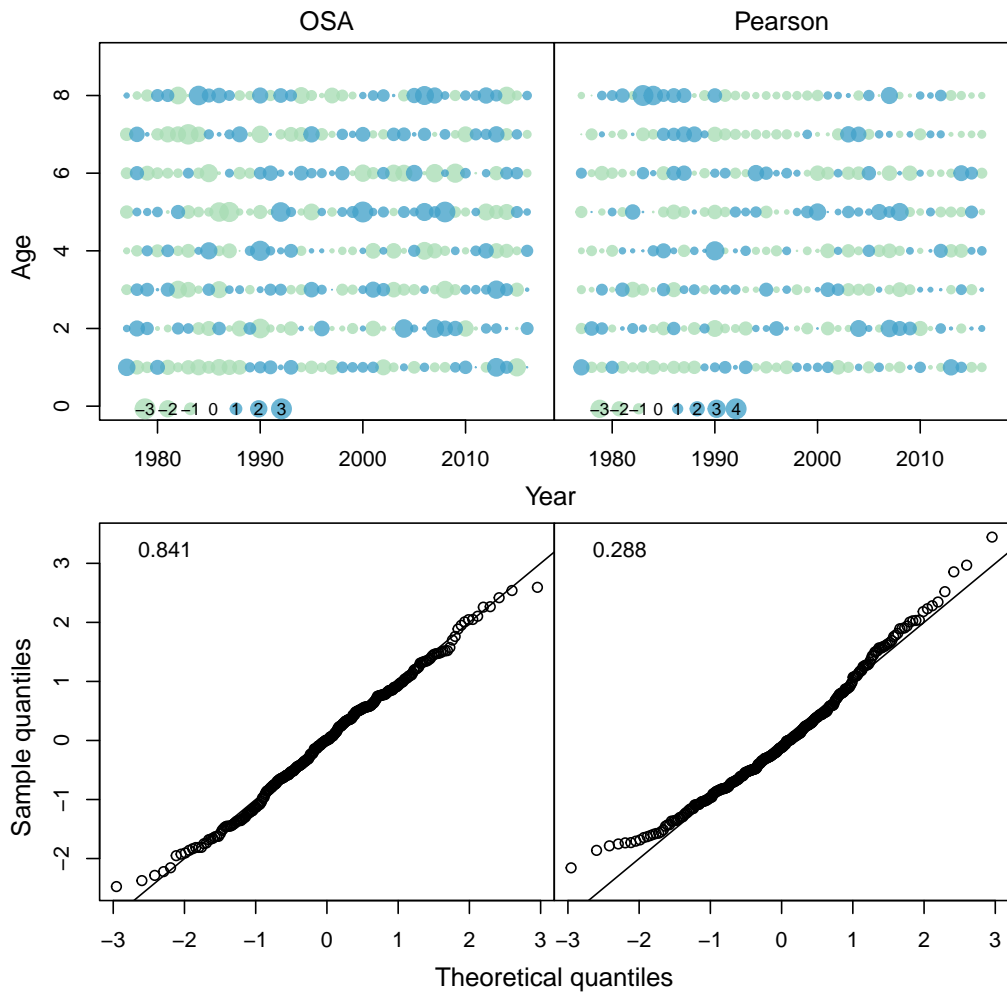


Figure A2.11: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 4th replicate of fleet 3 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

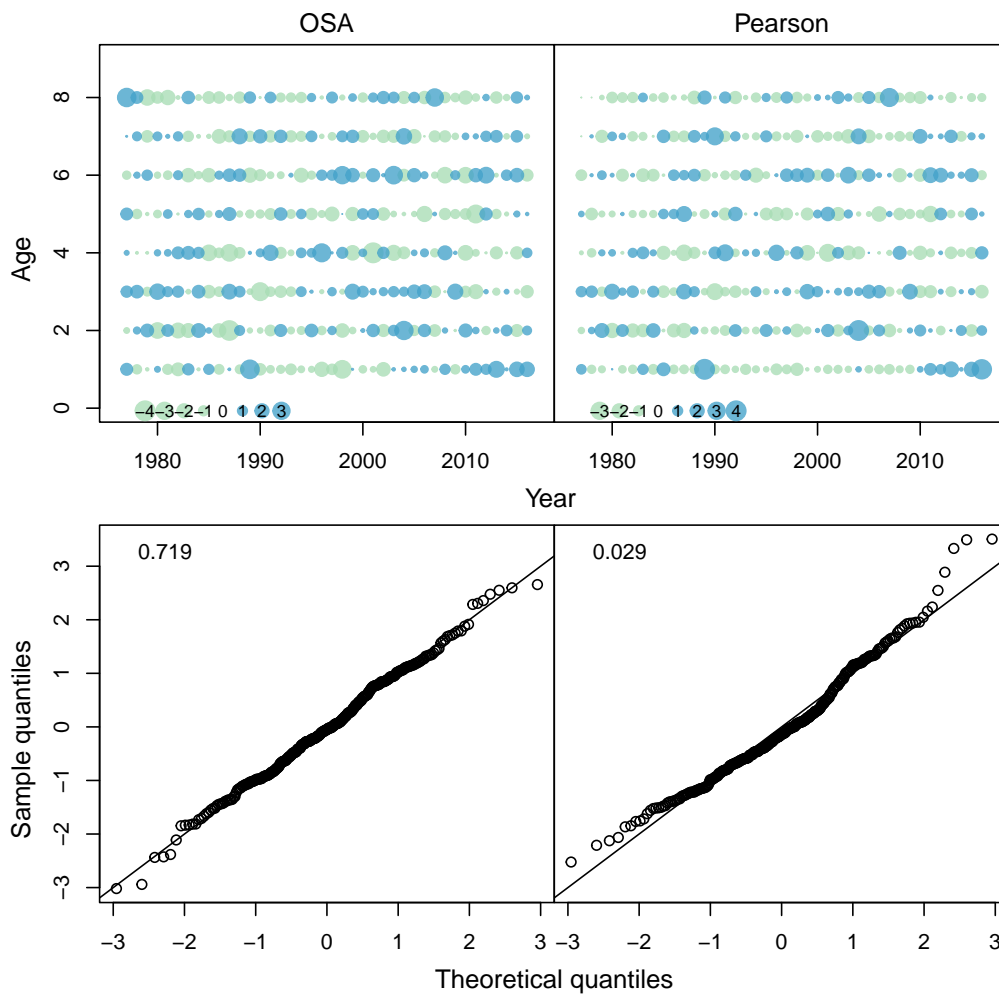


Figure A2.12: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 5th replicate of fleet 1 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

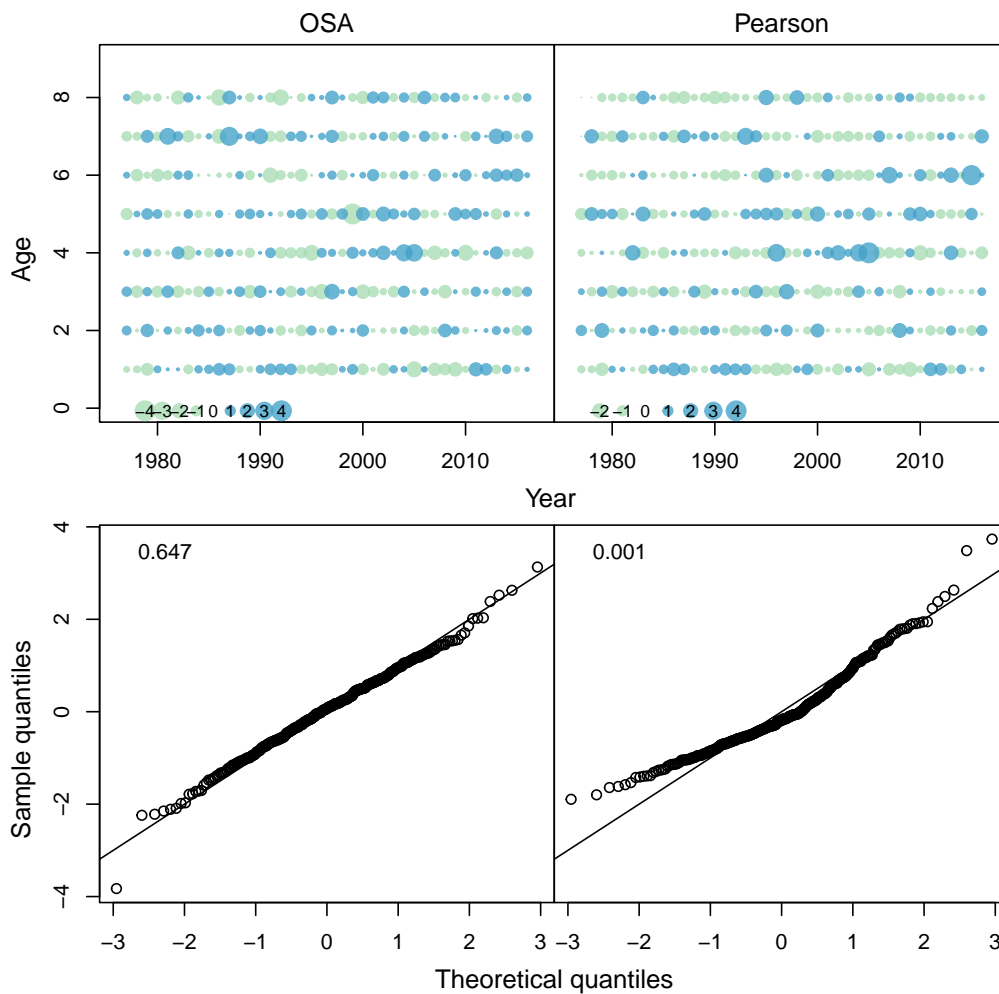


Figure A2.13: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 5th replicate of fleet 2 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

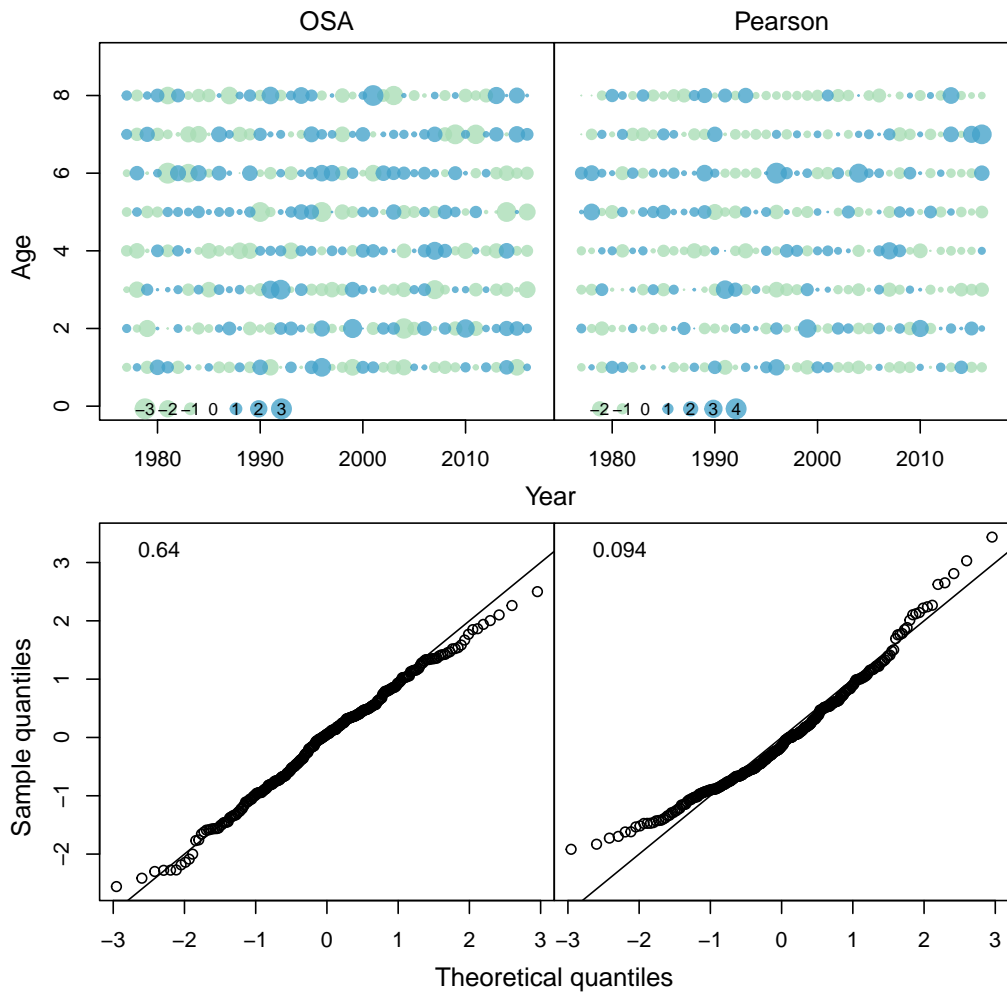


Figure A2.14: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 5th replicate of fleet 3 for the Gulf of Maine haddock simulated example. See caption of Figure A2.1 for a detailed description of the figure.

A3. Additional figures for the Gulf of Maine haddock example with true observations

The residuals estimated from the true observations example for Gulf of Maine haddock for the second fleet is given in Figure 5 and in Figures A3.1 and A3.2 for the 2 other fleets.

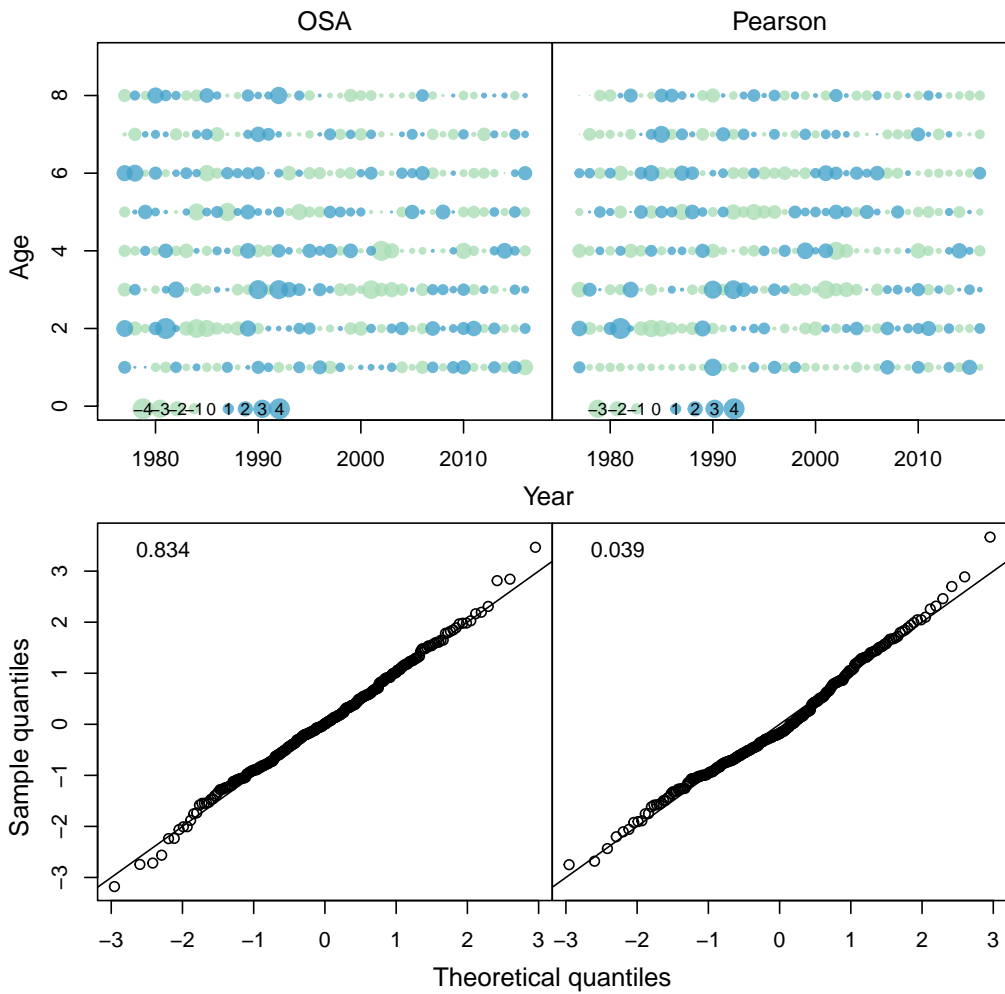


Figure A3.1: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 1st fleet of the Gulf of Maine haddock true observations example. See caption of Figure A2.1 for a detailed description of the figure.

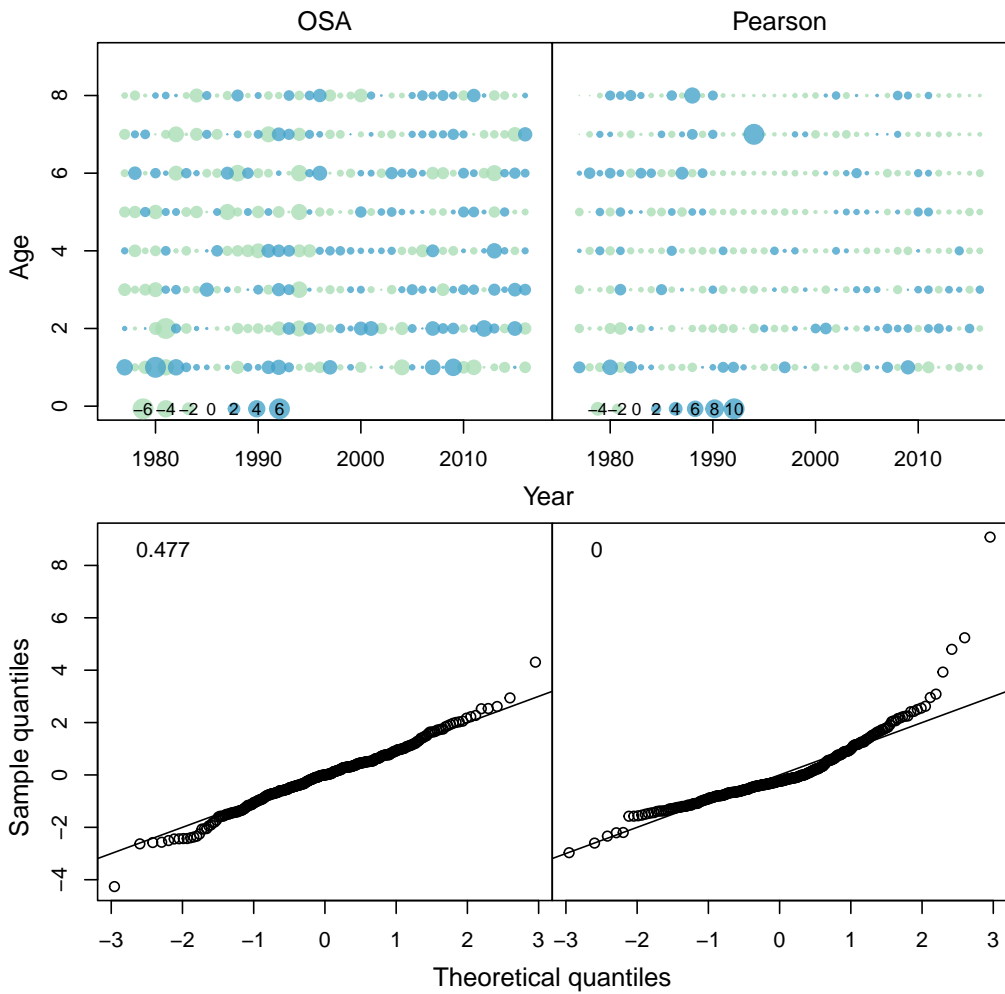


Figure A3.2: Residual bubble plots (top) and normal Q-Q plots (bottom) for the 3rd fleet of the Gulf of Maine haddock true observations example. See caption of Figure A2.1 for a detailed description of the figure.

References

Gross, J., Ligges, U., 2015. nortest: Tests for Normality. URL: <https://CRAN.R-project.org/package=nortest>. r package version 1.0-4.