Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2024/037

Quebec region

Results of the Mackerel (*Scomber scombrus* L.) Egg Surveys Conducted in the Southern Gulf of St. Lawrence from 1979 to 2022

Caroline Lehoux, Elisabeth Van Beveren, Stéphane Plourde

Fisheries and Oceans Canada
Maurice-Lamontagne Institute
850, Route de la Mer
Mont-Joli, Quebec Canada G5H 3Z4

June 2024

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

© His Majesty the King in Right of Canada, as represented by the Minister of the Department of Fisheries and Oceans, 2024
ISSN 1919-5044
ISBN 978-0-660-71962-7 Cat. No. Fs70-5/2024-037E-PDF

Correct citation for this publication:

Lehoux, C., Van Beveren, E., and Plourde, S. 2024. Results of the Mackerel (*Scomber scombrus* L.) Egg Surveys Conducted in the Southern Gulf of St Lawrence from 1979 to 2022. DFO Can. Sci. Advis. Sec. Res. Doc. 2024/037. v + 47 p.

Aussi disponible en français :

*Lehoux, C., Van Beveren, E., et Plourde, S. 2024. Résultats des relevés d'oeufs de maquereau (*Scomber scombrus *L.) dans le sud du golfe Saint-Laurent de 1979 à 2022. Secr. can. des avis sci. du MPO. Doc. de rech. 2024/037. v + 49 p.*

TABLE OF CONTENTS

ABSTRACT

Atlantic mackerel Total Egg Production (TEP) in the southern Gulf of St. Lawrence is a key indicator of the stock's state, as egg production is related to spawning stock biomass through stock composition and fecundity. We revisited the TEP calculations to increase reproducibility, incorporate methodological improvements, assess the statistical robustness of the results to various assumptions and parameter uncertainties, and to develop quantitative indicators of bias. The revised index was highly similar to the index presented during the 2021 assessment, but performed better in a year with fewer samples. Sensitivity analyses suggested that the TEP time series is robust against a variety of assumptions and uncertainties. Values for sampling years 1991 and 1999 were identified as potentially biased based on the indicators of bias presented herein, and were therefore excluded from the assessment. There was no evidence of bias for years 2006, 2017 and 2019, when eggs were sampled outside of the peak spawning period (when 70% of spawning had occurred in that year). The 2022 fishery closure and associated lack of adult samples from early in the spawning period resulted in relatively higher uncertainty around this year's index value, but again without any indication of bias. The revised index presented here has been used in the 2023 stock assessment as it outperformed the previously used method.

1. INTRODUCTION

A cornerstone of the northern contingent West-Atlantic mackerel (*Scomber scombrus*) assessment is the annual egg survey, covering its main spawning area in the southern Gulf of Saint-Lawrence (sGSL) in June (Van Beveren et al. 2023a). Egg surveys can provide reliable indices of stock trend, as observed egg densities are directly linked to stock abundance through estimates of fecundity (Bernal et al. 2012). Because of the importance of the Total Egg Production (TEP) index to the mackerel stock assessment, it is essential that potential sources of bias are assessed and that uncertainties are fully understood. The goals of this research document are therefore to present a more transparent and reproducible framework to calculate the TEP index and to perform a range of sensitivity analyses to assess the robustness of the index to its various assumptions and uncertainties.

The sGSL mackerel egg survey was first completed in 1979 using the standardized protocol still practised today, after multiple years of formative attempts (Lett and Marshall 1978; Maguire 1979a). Based on observed egg densities and temperature data, an index of relative Spawning Stock Biomass (SSB) has been computed for each stock assessment conducted since 1979(e.g., Maguire, 1979b; Ouellet, 1987). The statistical approach applied, however, evolved. For instance, D'Amours and Grégoire (1992) proposed an analytical correction for oversampling of mackerel eggs at the surface, when the net is towed for longer than the planned duration for a standardized tow. Although this correction was included in subsequent assessments (Grégoire and Lévesque 1994), it was discontinued later on. Likewise, in some but not all years, an instantaneous natural mortality rate was used in the estimation of the number of eggs released initially (Grégoire and Lévesque 1994). Different equations have further been applied to estimate incubation time, and hence the likelihood of observing stage one eggs, in function of temperature. A variety of spatial approaches have been used to extrapolate missing values, or to make predictions on a finer grid spanning the sGSL (e.g., Grégoire and Lévesque 1994 versus Doniol-Valcroze *et al.* 2019). Because new estimates are typically used to extend the existing series, it is unclear how each data point of the time series was obtained or how specific methodological choices have affected the results. Since the 2019 management strategy evaluation (Van Beveren et al. 2020), the assessment model for mackerel has been directly calibrated with a more elemental index of TEP rather than the derived SSB. In this application, TEP was back-calculated from the available SSB index rather than being estimated based on the available data.

Raw egg survey data (1982–2022) were recently quality-controlled and stored into a large database. We build on this effort to estimate a consistent TEP time series, and assess how specific methodological choices might affect our perception of TEP. Methods and results were reviewed and used in the 2023 stock assessment.

2. MATERIAL AND METHODS

2.1. SAMPLING AT SEA

Since 1979, 66 stations have been sampled near-annually in the sGSL (Figure 1 and Table 1). The survey has not been conducted in 1995 and 1997 (budget limitations) and 2020 (COVID pandemic). Survey data for 1980–1982 were never archived and are thus unavailable for analysis. Furthermore, the data from the 1982 survey had issues related to the protocol and was never used. These years were not included in the analyses, a decision that is consistent with previous practices.

The surveys aimed to sample each of the 66 stations at least once in June, when spawning activity has historically been the highest. Mackerel eggs and larvae were collected using a 0.6 m diameter Bongo equipped with 333 µm mesh size nets. A General Oceanics™ flowmeter was attached near the opening of each net to measure the volume of water filtered. Tows, lasting around 10 minutes each, followed a saw-tooth profile between the surface and a maximum depth of 50 m, or up to 5 m from the bottom for shallower stations. Since 2013, the tow profile, volume of water filtered and position of the nets in the water column have been monitored in real time using electronic equipment (BIONET™) attached to the Bongo frame. A CTD (Sea-Bird SBE-19) also attached to the frame provided the temperature and salinity profiles for the sampled portion of the water column. Back on deck, the nets were suspended and rinsed with sea water. Plankton samples from one of the two nets, by default from port side, were preserved in 4–5% formalin.

2.2. LABORATORY ANALYSES

Sorting of mackerel eggs and larvae was done at the Maurice Lamontagne Institute (Fisheries and Oceans Canada, Mont-Joli). Samples were fractionated according to van Guelpen's et al. (1982) beaker method. The identification criteria for Atlantic mackerel eggs and larvae (and other fish species encountered) were taken from Fritzsche (1978), Elliott and Jimenez (1981), and Fahay (1983, 2007). Atlantic mackerel eggs were sorted and counted by stage of development (Girard 2000).

2.3. ESTIMATION OF THE TEP INDEX

Data on egg counts per developmental stage, sampling station and year were extracted from the BioChem database (Devine et al. 2014). Egg density is generally higher in the upper 10 meters of the water column, above the thermocline (Ware and Lambert 1985, Grégoire et al. 1995). The corresponding CTD-based average temperature was determined for the upper 10 meters of water column (0-10 m) and provided by data managers (DFO, Maurice-Lamontagne Institute). Previous assessments used station names and/or positions to assign temperature to a Bongo tow. For the 2023 assessment, we assigned CTD to Bongo tows for all surveys based solely on position to ensure no error are introduced because of potential past and future inconsistencies in station identification. Each Bongo tow was associated with the CTD cast closest in space (within 40 km or the first neighbouring station) and time (within 24 hours) of the Bongo tows. When no CTD cast was available in the area, the average temperature in the corresponding spatial stratum, as defined by Ouellet (1987), was used instead, as done in the previous stock assessments.

Below we provide a step-by-step description of the algorithm used to estimate the TEP index. For each step, we provide the statistical details, note any differences with the previous approaches, and perform key sensitivity analyses. R scripts for each step are available on Github.

2.3.1. CALCULATION OF EGG DENSITY (N*M^{-2}) BY STATION

For each station, egg counts for eggs at stage 1 (N1, newly spawned eggs) and 5 (N5, mostly stage 1 broken, dead or unfertilized eggs) were converted into densities (ED, number * m^{-2}) by taking into account the sorted fraction (F), volume of filtered water (V, m^3) and maximum depth sampled (Z, m) for each station *s* and year *y*:

$$ED_{s,y} = \frac{\frac{N1_{s,y} + N5_{s,y}}{F_{s,y}}}{V_{s,y}} * Z_{s,y} \quad (2.3.1)$$

The above equation has consistently been applied as the first step in the calculation of the egg index (Gregoire et al. 2013).

The inclusion of stages 1 and 5 eggs is based on Maguire (1981) and is consistent with the U.S. approach to estimate egg production (Richardson et al. 2020). Including stages 2 to 4 would decrease the signal-to-noise ratio, because incubation time and especially post-spawning mortality for these stages are more uncertain than for newly spawned eggs.

Observation error associated with the egg counts could theoretically be estimated through repeated measurements. There were, however, only nine stations where both the port and starboard Bongo samples were analyzed, and therefore, observation error could not be estimated. Alternatively, one station was sampled 30 times consecutively in 1994. The measured densities were low and it remains unclear whether the coefficient of variation across these measurements is an adequate metric of egg density estimate accuracy and is applicable to other stations.

The sorted fraction of the sample can be determined precisely. Since 2013, the use of electronic equipment to record net position and water volume in real time has also reduced the potential for errors in filtered water volume and maximum depth sampled. Given the likely higher inaccuracy in both variables prior to 2013, and the direct impact any bias in these parameters can have on estimates of egg production, we developed criteria to identify potentially erroneous values. The sensitivity of the index to these outliers was assessed analytically (see sensitivity runs 2.3.1.1 and 2.3.1.2).

Further, errors in filtered water volume or signs of net clogging can be identified from the relationship with tow duration. Two stations had a high water volume filtered, for a relatively average tow duration (Figure A.1. 1). Comments by data managers suggested problems during sampling, and these two stations (station 3.3 of the first pass in 1988 and station 4.7 in 2022) were therefore excluded from further analyses.

2.3.1.1. Sensitivity run: Water Volume filtered

Based on the overall relationship between tow duration and filtered water volume (Figure A.1. 1), a standard 10-minute tow is expected to result in a filtered volume of approximately 250 m^3 on average. Many samples with water volumes less than 80 m^3 were identified as indicators of net clogging in the database. We therefore subjectively established that at least 80 m^3, or one third of the expected volume, should be filtered to obtain a desirable level of precision. All samples with water volumes below this threshold were considered as outliers and removed. We likewise removed four extra data points with long tow duration for relatively low filtered volume, as this is indicative of clogging or error in the recorded filtered volume. All removed values were subsequently imputed with a spatial model (see section 2.5 for details) to evaluate their impact on the egg index.

2.3.1.2. Sensitivity run: Maximum depth sampled

The maximum depth targeted is 50 m or 5 m above bottom in shallower waters. Maximum depth of the CTD attached to the Bongo was used for most years between 2002 and 2012, and maximum depth recorded by BIONET has been used since 2013. Sampled depths were, however, likely less precise prior to 2002 in rare suboptimal sampling conditions, because of the use of less technologically advanced equipment.

Bongo tows for which the depth exceeded the target depth should not be a source of bias. Mackerel eggs are concentrated within the first 10 m of the water column, and densities decrease exponentially with depth (Ware and Lambert, 1985, Grégoire et al. 1995). No or only extremely low egg densities should be present below 50 m. If we reasonably assume that no

eggs are present below 50 m, then oversampling such depths would result in the same estimated egg densities (number of eggs per square metre). That is, equation 2.3.1 implies that any added depth below 50 m would be cancelled out by a proportional increase in volume. However, sampling below 50 m results in proportionally less time spend sampling the targeted depths, which for a fixed tow duration should result in less accurate determination of the local egg density (i.e., it is the equivalent of having a lower filtered volume).

Bongo tows that did not reach the target depth could introduce a degree of bias if changing egg densities were present between the realized and the target depth. Because all but one station (9.65 m) covered at least the upper 10 m, where egg densities are expected to be highest, bias associated with depth sampled should be small.

To assess the effect of imprecise maximum sampled depths, we removed outliers in a new sensitivity run (Figure A.1. 2 and Figure A.1. 3). For each station, a 99.9% confidence interval for depth (mean± 3.291*sd) was estimated across 2002 to 2022 (years with relative precise maximum sampled depths). Stations (1979–2022) with a maximum sampled depth outside this range and at least 10 m relative to the targeted depth were considered outliers (if eggs were present, as zeros are considered robust). Removed values were again imputed with a spatial model.

2.4. CALCULATION OF DAILY EGG PRODUCTION (N*M^{-2}*DAY^{-1}) BY STATION

Station-specific egg densities (ED, n*m^{-2}) were converted into station-specific daily egg production estimates (DEP, n*m^{-2}*day^{-1}), based on the local incubation time (I, hours) of the eggs:

$$DEP_{s,y} = \frac{ED_{s,y}}{I_{s,y} * 24} \quad (2.4.1)$$

Incubation times of egg stages 1 and 5 are temperature (T, °C, 0-10 m) dependent and were calculated according to the Lockwood et al. (1977) model for mackerel in the Northeast Atlantic. This model is described as follows:

$$I_{s,y} = e^{[-1.614*Ln(T_{s,y})+7.759]} \quad (2.4.2)$$

The above steps, including Lockwood's equation, follow Gregoire et al. (2013).

2.4.1. Sensitivity run: Egg development

The sensitivity of the results to the estimated incubation time of stage 1 eggs was evaluated by using an alternative equation, provided by Mendiola et al. (2006):

$$I_{s,y} = e^{[-1.313*Ln(T_{s,y})+6.902]} \quad (2.4.3)$$

Mendiola et al. (2006) incubated eggs at lower densities than Lockwood et al. (1977) and sampled more frequently (Figure A.1. 4). Both studies were conducted in the Northeast Atlantic. Other studies reporting mackerel egg incubation times in function of temperature (Lanctot 1980,Worley 1933) focused on the period up to hatching rather than up to the end of development stage 1, and are thus inappropriate.

2.5. CALCULATION OF MEAN DAILY EGG PRODUCTION (N *M^{-2}*DAY^{-1}) OVER THE ENTIRE SAMPLED AREA

DEP values must be available for all of the 66 stations to estimate the mean daily egg production over the entire area consistently over time. In 34 out of the 38 survey years, not all

stations could, however, be visited (technical problems, time limitations, etc.) and values needed to be imputed. Although in the large majority of years (25 out of 34), the number of unsampled stations was low (3 or less), there were two years in which that number exceeded 10 (14 stations in 1999 and 13 in 2008). Note that when stations needed to be skipped due to time limitations, mission scientific leads based their decision on the historical egg distribution and elected to skip stations where egg densities have historically been the lowest. Therefore, imputation at unsampled stations should have a relatively small effect on the index.

DEP at unsampled stations was estimated with a newly developed Bayesian Generalized Linear Mixed Model (GLMM).This model avoided the shortcomings associated with the previously applied kriging approach (e.g., inconsistencies and difficulties with fitting of annual variograms, Grégoire and Bourdages 2000). The new model included station as a fixed effect and an annual spatial random effect. The R-INLA library (v.22.05.07, Lindgren and Rue 2015) was used to fit the Bayesian GLMM, which uses a stochastic partial differential equation (SPDE) approach to estimate the spatial random field (Lindgren et al. 2011) and the Integrated Nested Laplace Approximation for Bayesian inference.

The SPDE approach requires the creation of an artificial spatial mesh, so that the spatial autocorrelation between observations can be calculated. Information from variograms fitted by Gregoire and Bourdages (2000) was used to set the resolution of the mesh obtained by triangulation around station locations and for setting penalized complexity (PC) priors. The average range of spatial correlation determined by Gregoire and Bourdages (2000) was 150 km. Given this information, the mesh was built using a maximum triangle edge of 30 km (i.e., range/5) and 150 km inside and outside of the boundary, respectively. The cutoff or minimum allowed distance between points, below which they were replaced by a single vertex, was 15 km (i.e., range/10). The resulting mesh is presented in Figure A.2. 1.We specified PC prior probability distributions (Fuglstad et al. 2016, Simpson et al. 2015) for the spatial correlation range r and the standard deviation σ so that P(r < 150) = 0.05 and P(σ > 1) = 0.05.

DEP followed a zero-altered Gamma (ZAG) distribution with a log link:

$$\mathrm{DEP}_{s,y}=\mathrm{ZAG}\left(\pi_{s,y}\mu_{s,y}\right) \tag{2.5.1}$$

$$E\left(\mathrm{DEP}_{s,y}\right) = \pi_{s,y} \times \mu_{s,y} \tag{2.5.2}$$

$$var\left(DEP_{s,y}\right) = \frac{\pi_{s,y} \times r + \pi_{s,y} - \pi_{s,y}^2 \times r}{r} \times \mu_{s,y}^2 \tag{2,5,3}$$

$$logit\left(\pi_{s,y}\right) = Station_y + w_y + \varepsilon_{s,y} \tag{2,5,4}$$

$$\log\left(\mu_{s,y}\right) = Station_y + w_y + \varepsilon_{s,y} \tag{2,5,5}$$

In the above equations, *w* is the random Gaussian Field with a Matérn covariance structure (Gaussian Markov Random Field; Lindgren et al. 2011), π_{sy} is the probability of egg presence in station *s* and year *y* and has a Bernoulli error distribution, μ_{sy} is the DEP given eggs are present and has a Gamma error distribution. Station is a factor giving a baseline for expected DEP for each station. The model error structure is thus generated by running the model twice (DEP presence/absences, eq. 2.5.4; DEP>0, eq. 2.5.5) and then combining predictions (eq. 2.5.2).

The model was fitted without a temporal correlation structure (AR1) for the spatial random effect. The spatial effect in one year is thus independent from the spatial effect in the next year. The Watanabe-Akaike Information Criterion (WAIC; Watanabe 2010) of our proposed model was 9403 for the Bernoulli run and 16686 for the Gamma run. An alternative model fit that included temporal correlation between consecutive years (with AR1) showed a lower performance (WAIC = 11266 for the Bernoulli run and WAIC – 18167 for the Gamma run).

For each station without an associated value of DEP, predictions were made with the above-described model. The annual mean daily egg production was then calculated as the average DEP across all 66 stations, both observed and predicted. All stations were given equal weight as they represent a quasi-identical surface area, resulting from the fixed distance sample design.

$$\overline{DEP_y} = \frac{\sum_{s=1}^{S=66} DEP_{s,y}}{66} \quad (2.5.6)$$

The DEP (n*day^{-1}) for the whole sGSL was obtained as:

$$DEP_y = \overline{DEP_y} \times 6.945^{10} \quad (2.5.7)$$

where 6.945^{10} m^2 is the study surface area. No changes have been made to this value compared to previous assessments (e.g., Gregoire et al. 2013). Because the egg survey index is relative and the survey surface area is a constant, it has no impact on the estimation of stock state.

2.5.1. Validation

Model quality was inspected visually through the examination of standard plots (residuals, predicted versus observed values, etc.). To validate the predictive performance of the R-INLA approach, a random set of 10 stations (representing the number of stations missed in years of poor coverage) was removed for each year that had at most 1 station missing, prior to model refitting. Estimated values for these stations were compared to the observations using a linear regression.

2.5.2. Sensitivity run: Predicted grid

The baseline approach presented above uses a model-only approach to predict DEP in missing stations. Alternatively, it would be possible to predict DEP in each station, prior to determining the mean DEP over the entire area. Doing so assumes that there is less error in the model predictions than the observations.

2.6. CALCULATION OF THE PROPORTION OF EGGS SPAWNED DAILY

Scaling up daily egg production to annual egg production has for West-Atlantic mackerel previously and consistently been done through the estimation of the proportion of eggs spawned at the time of the survey, relative to the entire spawning period. The proportion of eggs released by the stock at any given day can be estimated through the progression of the gonadosomatic index (GSI, %), which was calculated as follows :

$$GSI = \frac{Gonad\ weight}{Total\ fish\ weight} \times 100 \quad (2.6.1)$$

The relative decline in GSI over time (and thus the number of eggs spawned) is best approximated by a logistic model (F. Grégoire, DFO, unpubl. comparison of 6 models). We replaced the 4-parameter logistic model used in previous assessments (2021 assessment and prior) with an equivalent 3-parameter version, as a more parsimonious model with meaningful parameters was preferred. Furthermore, instead of fitting one model per year (Gregoire et al. 2013), a single mixed model that included all years and an AR1 process across years was fitted. Exploratory analyses suggested that the explicit incorporation of temporal autocorrelation resulted in more biologically plausible model fits for data-poor years (see 2022). Although eggs were not sampled in 1995, 1997 and 2020, fish samples were available and thus used to inform the logistic model. The following logistic model was fitted in the *nlme* package in R (Pinheiro and Bates 2000; Pinheiro et al. 2022):

$$GSI = \frac{Asym}{1 + e^{\frac{xmid - x}{scal}}} \quad (2.6.2)$$

where *Asym* is the plateau (asymptote) of GSI at the start of the spawning season, *xmid* is the sigmoid function's midpoint and corresponds to the peak spawning day, *scal* is the steepness of the curve determining spawning season duration and x is the date when each sample was collected. The starting values were estimated with nls.multstart over 500 iterations (Padfield and Matheson 2020). Starting values for *Asym* were evaluated over the range of observed GSI values (0.04-35). For *xmid*, we considered the earliest to latest sample date across all years (days of year 139–312). Starting *scal* values were evaluated between -100 and 0, corresponding to an extremely short and long spawning season, respectively.

Data used to fit the logistic model were obtained from the commercial port sampling program, as well as from supplementary sources of biological data ("biological database", Van Beveren et al. 2023b).

Biological samples were primarily obtained from from NAFO division 4T (sGSL). However, samples from NAFO divisions 4V and 4W were also included as fish in 4V and 4W were considered in migration to spawn in the sGSL and samples from the onset of spawning in 4T were often not available. Samples from the beginning of the spawning season are critical for model fitting and to adequately assess spawning seasonality. Only fish of maturity stages 5 (ready to spawn), 6 (spawning) and 7–8 (post-spawning, recuperating) were selected (Pelletier et al. 1986). The majority of samples were from NAFO division 4T (Figure A.3. 1). In 4V and 4W, only fish of maturity stages 5 or 6 were considered. All maturity stages were well represented, except in 2022 following the commercial fishing closure. In response to this closure, a new sampling program was set up and sampling started later than during previous years (Figure A.3. 2). See Van Beveren et al. (2024) for more details on fish samples.

Outliers were visually identified and removed from the GSI ~ day of year relationship and from the weight ~ length relationship. We removed GSI > 5% after day of year 250. For years 1994, 1999, 2018, 2019 and 2020, we removed GSI > 10% after day of year 200. Fish of length < 20 cm or fish with a GSI > 40% were likewise removed, as were fish with a weight outside the 0.0001 (0.9999) quantile, by age group.

The logistic model essentially described the expected decrease in the stock's pool of eggs, from the beginning to the end of the spawning period. The rate of change over a given time period is equal to the proportion of eggs spawned. Thus, daily GSI values were predicted from the logistic model, and the standardized slopes over one day periods (d) were computed to obtain the proportion of eggs spawned daily (*S*) for each year (y), following Gregoire et al. (2013):

$$S_{y,d} = \frac{GSI_{y,d+0.5} - GSI_{y,d-0.5}}{\sum_{d=1}^{D}\left(GSI_{y,d+0.5} - GSI_{y,d-0.5}\right)} \quad (2.6.3)$$

The annual proportion of eggs spawned at the median date of the survey ($S_{y,d=m}$) is extracted from this series .

To obtain confidence intervals around predictions of GSI and $S_{y,d}$, a nonparametric bootstrap procedure coupled with global residual bootstrap was used (Thai et al. 2013). Specifically, residuals from equation 2.6.2 were extracted and were resampled for 100 iterations, before being added randomly to the predictions. The model was fitted and the GSI and proportion of eggs spawned were predicted for each iteration. The 2.5 and 97.5% quantiles of the predictions were used as confidence intervals. Considering the high computational cost of running the logistic model at each iteration, we additionally validated that the number of iteration used (N=100) was sufficient for the estimated confidence intervals to stabilize.

2.6.1. Validation

Fish sampled early in the spring were not available for 2022 and no stage 5 fish were identified. To verify the effect of a lack of samples at the beginning of the spawning season, we performed a simulation scenario, where for each year between 1984 and 2019, data collection only started at day 179 (June 28: date of the first 2022 sample). For each simulation scenario, we removed the data collected prior to day 179 for one year and re-fitted the model. We then compared the predicted proportion of spawning at the median date of the egg survey with the full model.

Further external validation was done using the proportion of mackerel larvae (*Lp*) within the samples:

$$Lp_y = \frac{\sum_{s=1}^{S=66} \frac{Larval\ density_{s,y}}{ED_{s,y} + Larval\ density_{s,y}}}{66} \quad (2.6.4)$$

which was calculated for each station *s* and averaged by year *y*. *Lp* was compared against the proportion of spawning before the median date of the survey (area under the curve before the median date). The more advanced the spawning season at the time of the survey, the higher the proportion of larvae that should be observed.

2.7. CALCULATION OF THE TOTAL EGG PRODUCTION

Total annual egg production was calculated as:

$$TEP_y = \frac{DEP_y}{S_{y,d=m}} \quad (2.7.1)$$

2.7.1. Sensitivity run: Timing of the survey

The timing of the survey in relation to the spawning season is a key concern in the estimation of the index. A second sampling of the grid (referred to as a pass) was performed in thirteen years prior to 2001. In this document, we consider the second survey of the year for sensitivity analyses. We calculated the TEP using DEP calculated for the second pass when available. We retained only the seven years with comparable coverage between the two passes (1988, 1989, 1990, 1992, 1994, 1996, 1998,Table 1). The coverage of the second pass was insufficient for 1987 (n=38) and 1993 (n=29) and station-specific temperature data were not available for the second pass of survey years 1983,1984,1985 and 2000.

2.7.2. Sensitivity run: Station $S_{y,d=x}$

The $S_{y,d}$ has traditionally been extracted for the median survey date ($S_{y,d=m}$). This would represent a (small) source of bias if the order in which stations were visited by the vessel is inverse to the order in which mackerel spawns in these locations, and the highest egg densities were observed near the start or end of the survey. An alternative approach would be to apply station-specific estimates of $S_{y,d}$. This would imply that the daily proportion of eggs spawned is identical for each station, or that spawning seasonality is homogeneous across the sGSL. Although this assumption oversimplifies spawning dynamics, we nonetheless tested the approach as it demonstrates the potential impact of estimating the proportion of eggs spawned using one specific date. From the estimated time series of $S_{y,d}$ (eq. 2.6.3), we extracted for each station the $S_{y,d}$ at the date the station was sampled ($S_{y,d=x}$). The total egg production was then obtained as:

$$TEPy = \frac{\sum_{s=1}^{S=66}\left(\frac{DEP_{s,y}}{S_{y,d=x}}\right)}{66} * 6.945^{10} \quad (2.7.2)$$

3. RESULTS

3.1. DAILY EGG PRODUCTION ($N*M^{-2}$)

Incubation time of stage 1 (and 5) eggs at all stations and years varied between 23.5 and 146 hours (mean ± sd = 53.1±16.2 h) based on the upper 0-10 m temperature (Lockwood et al. 1977). The average incubation time across all years and stations at which eggs were observed decreased by 14% when applying the Mendiola et al. (2006) equation (range: 23.55-104.3 h, 45.4 ± 11.1 h). Consequently, $DEP_{s,y}$ increased by 13% based on the Mendiola et al. (2006) equation. Specifically, mean DEP ranged from 0 to 1494 eggs * m^{-2} *day^{-1} with the Lockwood et al. (1977) equation (72.14 ± 171 eggs * m^{-2} *day^{-1}) and from 0 to 1733 eggs * m^{-2} *$day^{-1}$with the Mendiola et al. (2006) equation (81.7±193 eggs * m^{-2} *day^{-1}).

DEP_y has been low since 2008. The maximum observed DEP in 2021 was 85 and 96 eggs * m^{-2} *day^{-1} for the Lockwood et al. (1977) and the Mendiola et al. (2016) equations, respectively. For 2022, these values were respectively 75 and 83 eggs * m^{-2} *day^{-1}.

Maps of $DEP_{s,y}$, including spatial model predictions, are shown in Figure 2. Model output (see spatial random fields in Figures A.2. 2 and A.2. 3) and validation (Figure A.2. 5) suggested that the model complied with the underlying assumptions. Cross-validation demonstrated that predicted DEP did not deviate much from observed DEP, although small values were slightly overestimated (Figure A.2. 4 and Figure A.2. 5). This is typical of ZAG distributions because they typically don't predict absences. Because only a small number of stations generally required predictions during any given year, and the trend in egg production is mostly defined by stations with high egg densities, this bias was considered negligible.

3.2. PROPORTION OF EGGS SPAWNED DAILY

The logistic model fitted the data well, but predictions for 1991 and 1999 were considered poor (Figure 3, Figure A.3. 3, see further). Results are discussed excluding these two years.

The peak day of spawning occurred on average on the 171st (June 20) day of the year (range = 161-186 or June 10 to July 1st). Spawning started (2.5% quantile) on average on the 136th day (May 16) and ended (97.5% quantile) on day 206 (July 25), resulting in an average spawning duration of 70 days. The percentage of eggs spawned at the median survey date (first passes or baseline surveys) ranged from 1% to 4.32% and averaged 2.5% (Table 1).

The level of uncertainty associated with the estimation of daily spawning proportions depends on the timing of the survey relative to the timing of spawning, as well as on biological data availability. Survey timing usually corresponded well to the main period of spawning (Figure 4 and Table 1). The median survey date of the first pass corresponded on average to day 171 (June 20) and ranged between day 162 (June 11) and day 186 (July 5). The second pass was usually less aligned with the peak of spawning, except for years 1990 and 1994 (Figure 4). Surveys were considered to fall outside the main spawning season if the median survey date was outside of the period when 70% of eggs were spawned (Table 1). Using this threshold, the 2017 and 2019 surveys were identified to have occurred before the main spawning season (<15% of eggs spawned) whereas the 2006 survey was relatively late compared to spawning (>85% of eggs spawned). Note that predictions of $S_{y,d=m}$ are more uncertain at the onset of spawning than for the end (Figure 3). Although the timing of the survey was relatively constant, because of the variability in the timing of spawning and the decrease in spawning duration, the likelihood of conducting the survey outside the main spawning period has increased over time (Figure 5) which has the potential to increase the uncertainty around TEP. Specifically,

spawning duration (period over which 95% of egg production occurs) has on average decreased by 0.94 day/year since 1979.

Data availability for the estimation of the spawning season was deemed sufficient for the majority of years (>325 individuals collected over most of the spawning season). In 2022 no early-season biological samples were collected (first sample was collected on June 28), but there were no issues identified with the model fit (e.g., implausible predictions, residual patterns). The 95% confidence intervals around the 2022 estimate of $S_{y,d=m}$ were, however, the widest of the time series, ranging from 0.014 to 0.0289 (Table 1). Furthermore, a validation in which the time series were truncated at day 179 for each year individually, before the model was refitted, showed large changes in the $S_{y,d=m}$ (Figure 6), indicating that the estimated confidence interval for years with truncated data might be underestimated (despite sufficient bootstrap iterations; Figure A.3. 4). There was indeed little correlation (R^2 =0.03) between the predictions based on the truncated time series and the full data set, suggesting that early season data of maturity stage 5 are necessary to accurately evaluate the spawning season's parameters.

Inaccurate estimates are not necessarily biased. Because large bias in $S_{y,d=m}$ during some years would directly and proportionally affect our estimation of TEP, we developed independent bias indicators to validate the estimated annual spawning seasonality. First, the date at which stage 6 fish abundance peaks in the samples (averaged day of year weighted by the proportion of stage 6 in the samples) should be approximately aligned with the estimated date of peak egg production (Figure 7). This was not the case for 1999, suggesting an issue with the model predictions. Years 1991 and 1999 were further characterized by an unusually early and long spawning season. The same bias indicator could not be calculated for 2022 because of the later fish sampling.

Validation of the estimated spawning seasonality was, in addition, done based on the proportion of larvae in the samples, relative to the egg abundances. This ratio increased exponentially with the progression of the spawning season (Figure 8). For 1991 and 1999, the average proportion of larvae in the samples was lower than could be expected based on the estimated spawning seasonality, adding more support to the suspected issues with the logistic fit for these years. The proportion of larvae in 2022 matched well with what could be expected based on the estimated spawning seasonality.

The proportions of eggs spawned at the median date of the survey estimated in this document were significantly correlated (R^2=0.66) to the proportions of eggs spawned estimated by Gregoire et al. (2013), although values estimated herein were generally lower (Figure A.3. 5).

3.3. TOTAL ANNUAL EGG PRODUCTION

Total egg production was maximal in 1987 and has been decreasing since 1993, reaching one of the lowest values of the time series in 2001 (Figure 9 and Table 2). Following a small increase in 2002–2004, TEP has remained low since 2006. TEP averaged 5.14 x 10^{13} eggs from 2006–2022 compared to 3.91 x 10^{14} before 2006.

The TEP in 1991 was three times higher than the value of 1990, which is not consistent with the population dynamics. The poor logistic model predictions to determine spawning seasonality in 1991 (see previous section) could be the main cause of this unrealistic value. For this particular year, Gregoire et al. (2013) made a correction to the logistic fit by forcing a plateau (fixed $Asym_{y=1991}$). This approach incorporates a significant level of subjectivity and we therefore elected to exclude this year from the time series.

3.3.1. Sensitivity analyses

TEP for each sensitivity run is shown in Figure 9, and the relative changes in TEP compared to the baseline run are shown in Figure 10. Figure 9 also includes the TEP reported in the last assessment (Smith et al. 2022).

3.3.1.1. Volume filtered

Whether stations with a low filtered volume were included or excluded had a relatively small impact on TEP. In general, annual differences were less than 5%, especially from 2013 onwards, when electronic equipment was used to monitor filtered volume. The largest relative change in TEP (-16%) due to stations with low filtered volume occurred in 2001, when TEP was low and small change in the absolute value thus resulted in large relative change.

3.3.1.2. Maximum depth sampled

The removal of maximum sampled depth outliers had a larger relative impact on the index values prior to 2001, but did not exceed 50% during that period and did not result in meaningful differences. It is, however, difficult to separate the error caused by incorrectly sampled or recorded depth from the error associated with RINLA predictions. A large number of stations were excluded in some years and the RINLA prediction error thus increased (Figure A.1. 2). The comparison of predicted and observed values did not show the expected directional trend toward higher predictions relative to observations that would be expected if there was a bias in the sampling (Figure A.1. 5). Instead, the slope of the relation between predictions and observations was similar to the relation obtained during cross-validation (Figure A.2. 4) suggesting that the relative difference in the TEP index compared to the baseline scenario might be more strongly related to RINLA uncertainty than to the bias in the sampling.

3.3.1.3. Egg development

The equation used to calculate egg incubation time also had a minor relative effect on TEP. TEP was higher by 14% on average when the Mendiola et al. (2006) equation was used as opposed to the Lockwood et al. (1977) equation. Because the TEP index is relative, and the choice of an incubation time equation mostly rescales this index rather than affects its interannual trend, this factor can effectively be removed from further consideration.

3.3.1.4. Predicted grid

Making model-based predictions of daily egg production for all stations, rather than using direct observations to the extent possible, changed annual TEP by as much as 25% (average: 9%). Changes were not directional over time and essentially resulted in a slight smoothing of the series.

3.3.1.5. Timing of the survey

The first and second pass of the survey were used to assess the impact of the timing of the survey on TEP. First and second survey passes, for all seven years with sufficient sampling on both passes (1988–1990, 1992, 1994,1996,1998), have consistently occurred during the main spawning period (70% of eggs spawned; Table 1). Nevertheless, survey timing had the most notable effect on our estimate of TEP. Although for 3 out of 7 years a quasi-identical output was generated (< 5% relative difference), for the other 4 years the second pass resulted in a relative difference in TEP of 45% to 65% (3 negative differences, i.e., lower TEP on the second pass and 1 positive difference i.e., higher TEP on the second pass). Exploration of possible reasons why some years yielded similar or different values for the two passes suggested that the duration and the timing of the survey in relation to the spawning season can, in some but not all instances, explain part of the difference (result not shown). Years when the second pass was

more aligned with the spawning season (1990 and 1994) indeed resulted in the smallest differences.

3.3.1.6. Station $S_{y,d=x}$

Using $S_{y,d=x}$, instead of $S_{y,d=m}$, changed the TEP index by up to 34%. The average change across all years was, however, only around 5%. The impact of selecting one approach over the other notably increased from 2014 onwards. Years with a shorter spawning season or a survey timing further outside the peak spawning period were indeed more likely to be affected by this change in methodology (Figure A.3. 6). In recent years, the absolute effect of using station-specific spawning proportions was nonetheless small, considering that TEP during these years was low.

4. DISCUSSION

This report presented an improved, more robust and reproducible approach to estimating mackerel TEP in the sGSL. For the first time, the robustness of the index to various sources of uncertainty and subjective choices was assessed through a range of sensitivity analyses. The improved baseline algorithm was used during the 2023 assessment (Van Beveren et al. 2023b).

The first key methodological change presented here was made to address previous concerns with the technique used to perform spatial predictions. The RINLA approach was deemed superior to annual kriging because it explicitly accounted for and had more flexibility with respect to the statistical error distribution around DEP and spatial autocorrelation had interannual consistency. Overall, RINLA therefore performed better during cross-validation (see Figure 13 from Grégoire and Bourdages, 2000, versus Figure A.2. 5) and did not allow for negative predictions. One of the key consequences of this change was that spatial modelling was here only used to fill in gaps, whereas in previous assessments the average DEP was estimated over the entire kriged and thus predicted area. A sensitivity run considering predictions at each station instead of observations, as well as a comparison with the TEP index from 2021 (Smith et al. 2022), demonstrated that the scale at which predictions are performed should not meaningfully impact the assessment.

The second improvement to the TEP estimation algorithm involved quantifying of the proportion of eggs spawned at the time of the survey. This is arguably the most important parameter in the estimation of TEP. Because this parameter is multiplicative, if for example, we assumed a proportion of 0.01 instead of 0.04, TEP would be off by a factor of 4. The new method presented in this research document consisted of a single model with an AR1 process as opposed to yearly independent models, and is more robust to insufficient data for some years. Specifically, without acknowledging that spawning seasonality between years is correlated, no index for 2022 could have been estimated. The new model does, however, not eliminate the possibility for substantial bias. That is, a model simulation scenario with truncated data (day=179) did not always succeed in predicting spawning seasonality as assessed with the full dataset. A comparison with previous $S_{y,d=m}$ estimates, a sensitivity run based on second survey passes as well as the confidence intervals around $S_{y,d=m}$ for certain years all suggested that the determination of $S_{y,d=m}$ might indeed be the most influential step in the TEP index calculation. Because of the importance of properly estimating $S_{y,d=m}$, we defined two independent criteria to flag large biases, one based on the proportion of spawning (stage 6) females in the samples and another based on the proportion of larvae in the samples relative to the number of eggs. These criteria were sufficiently sensitive to justify the removal of 1991 and 1999, the latter which was also excluded from previous stock assessments, albeit based on obscure reasons. Despite that 2022 was characterized by large uncertainty in $S_{y,d=m}$, neither of the two bias indicators

provided grounds for excluding this year from the assessment. The described criteria should be scrutinized in all upcoming TEP calculations to ensure a maximal level of confidence.

The egg survey index drives the trend in SSB estimated during the assessment. As a pillar of the assessment, it is essential that the index be robust to methodological change or process uncertainty. In this research document, we detailed all the steps involved in the calculation of the index and quantified the impact of possible alternative approaches. Although the index is most sensitive to the timing of the survey and dependent on an accurate estimation of spawning seasonality, both factors did in general not generate directional bias nor errors exceeding the observed multi-year variability in the index. Thus, the estimated long-term decline in TEP is highly reliable. Richardson et al. (2020), who analyzed the robustness of the US mackerel egg survey index to a similar suite of uncertainties, came to analogous conclusions. The relative effect of variations in estimated incubation time on TEP was lower in the warmer US waters (~ 5%) compared to the colder sGSL waters (~ 15%), which is expected considering that the Lockwood et al (1977) and Mendiola et al. (2006) equations diverge most notably at temperatures colder than 10°C. Additional sensitivities considered by these authors, including natural mortality between spawning and the moment of sampling, were not included here but are known to mostly influence scale (similar to incubation time). Based on our work, results from Richardson et al. (2020), and the lack of evidence of significant spawning outside the sGSL (Van Beveren et al. 2023a), we conclude that the egg index is a robust and reliable indicator of stock state.

5. REFERENCES CITED

Bernal, M., Somarakis, S., Witthames, P.R., van Damme, C.J.G., Uriarte, A., Lo, N.C.H., and Dickey-Collas, M. 2012. Egg production methods in marine fisheries: An introduction. Fish. Res. 117–118: 1–5.

D'Amours, D., and Grégoire, F. 1992. Analytical correction for oversampled Atlantic mackerel *Scomber scombrus* eggs collected with oblique plankton tows. Fish. Bull. 90: 190–196.

Devine, L., Kennedy, M.K., St-Pierre, I., Lafleur, C., Ouellet, M., and Bond, S. 2014. BioChem: the Fisheries and Oceans Canada database for biological and chemical data. Can. Tech. Rep. Fish. Aquat. Sci. 3073: iv + 40 pp.

Doniol-Valcroze, T., Van Beveren, E., Légaré, B., Girard, L., and Castonguay, M. 2019. Atlantic mackerel (*Scomber scombrus* L.) in NAFO Subareas 3 and 4 in 2016. DFO Can. Sci. Advis. Sec. Res. Doc. 2018/062. v + 51 p.

Elliott, E. M. and Jimenez, D. 1981. Laboratory manual for the identification of ichthyoplankton from the Beverly - Salem Harbor area. Dept. of Fisheries, Wildlife and Recreational Vehicles, Massachusetts. 230 pp.

Fahay, M. P. 1983. Guide to the early stages of marine fishes occurring in the western North Atlantic Ocean, Cape Hatteras to the southern Scotian Shelf. J. Northw. Atl. Fish. Sci. 4: 1–423.

Fahay, M. P. 2007. Early stages of fishes in the western north Atlantic Ocean : (Davis Strait, Southern Greenland and Flemish Cap to Cape Hatteras). NAFO. 1696 p.

Fritzsche, R. A. 1978. Development of fishes of the mid-Atlantic Bight, an atlas of egg, larval and juvenile stages. Vol. V. Chaetodontidae through Ophidiidae. U.S. Fish. Wildl. Serv. Biol. Serv. Program. 340 pp.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. 2016. Constructing Priors that Penalize the Complexity of Gaussian Random Fields. J. Am. Stat. Ass. 114 (525) 445-452.

Girard, L. 2000. Identification of mackerel (Scomber scombrus L.) eggs sampled during abundance surveys in the southern Gulf of St. Lawrence. In: The Atlantic mackerel (Scomber scombrus L.) of NAFO Subareas 2 to 6. Chapter 4. Edited by F. Grégoire. DFO Can. Sci. Advis. Sec. Res. Doc. 2000/021. pp. 119–137.

Grégoire, F and Bourdages, H. 2000. Geostatistical evaluation of the Atlantic mackerel (*Scomber scombrus* L.) mean egg density based on abundance surveys conducted from 1982 to 1998. Chapter 7. Edited by F. Grégoire. DFO Can. Sci. Advis. Sec. Res. Doc. 2000/021. pp. 195–262.

Grégoire, F., and Lévesque, C. 1994. Estimate of Gulf of St. Lawrence spawning stock of mackerel (*Scomber scombrus* L.) by total egg production and batch fecundity in 1993. DFO Atl. Fish. Res. Doc. 61: 22.

Grégoire, F., D'Amours D., Lévesque, C. and Thibeault, D. 1995. Estimation of the Gulf of St.Lawrence spawning stock biomass of mackerel (*Scomber scombrus* L.) for 1994. DFO Can. Sci. Advis. Sec. Res. Doc. 1995/118.

Grégoire, F., Gendron, M.-F., Beaulieu, J.-L. and Lévesque, I. 2013. Results of the Atlantic mackerel (*Scomber scombrus* L.) egg surveys conducted in the southern Gulf of St. Lawrence from 2008 to 2011. DFO Can. Sci. Advis. Sec. Res. Doc. 2013/035. v + 57 p.

Lanctot, M. 1980. The Development and Early Growth of Embryos and Larvae of the Atlantic Mackerel (*Scomber scombrus* L.), at Different Temperatures. Can. Tech. Rep. Fish. Aquat. Sci. 927, 101 p.

Lett, C., and Marshall, W.H. 1978. An interpretation of biological factors important in the management of the northwestern Atlantic mackerel stock. CAFSAC Res. Doc. 77: 38.

Lindgren, F., and Rue, H. 2015. Bayesian Spatial Modelling with R-INLA. J. Stat. Softw. 63 (19): 1–25.

Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73(4): 423–498.

Lockwood, S.J., Nichols, J.H., and Coombs, S.H. 1977. The development rates of mackerel (*Scomber scombrus* L.) eggs over a range of temperatures. ICES C.M. J(13): 13 p.

Maguire, J.-J. 1979a. An outline of a method to back calculate the mackerel spawning stock from egg abundance estimates. CAFSAC Res. Doc. 31: 24.

Maguire, J.-J. 1979b. An analytical assessment of SA 3-6 mackerel with information from egg and larval survey. DFO Res. Doc. 46: 42.

Maguire, J.J. 1981. Maturité, fécondité, ponte et évaluation de la taille du stock reproducteur du maquereau atlantique (*Scomber scombrus*) dans le golfe du St-Laurent. Master's thesis, Université Laval, Québec. 137 p.

Mendiola, D., Alvarez, P., Cotano, U., Etxebeste, E., and de Murguia, A.M. 2006. Effects of temperature on development and mortality of Atlantic mackerel fish eggs. Fish. Res. 80 (2–3): 158–168.

Ouellet, P. 1987. Mackerel (*Scomber scombrus* L.) egg abundance in the southern Gulf of St. Lawrence from 1979 to 1986, and the use of the estimate for stock assessment. CAFSAC Res. Doc. 87/62.

Padfield, D., and Matheson, G. 2020. nls.multstart: Robust Non-Linear Regression using AIC Scores.

Pelletier, L. 1986. Fécondité du maquereau bleu, *Scomber scombrus* L., du golfe du Saint-Laurent. Can. Tech. Rep. Fish. Aquat. Sci. 1467: v + 37 pp.

Pinheiro, J.C., and Bates, D.M. 2000. Mixed-Effects Models in S and S-PLUS. Springer, New York.

Pinheiro, J., Bates, D., and R Core Team. 2022. nlme: Linear and Nonlinear Mixed Effects Models.

Richardson, D.E., Carter, L., Curti, K.L., Marancik, K.E., and Castonguay, M. 2020. Changes in the spawning distribution and biomass of atlantic mackerel (*Scomber scombrus*) in the Western atlantic ocean over 4 decades. Fish. Bull. 118(2): 120–134.

Simpson, D., Rue, H., Martins, T., Riebler, A., and Sørbye, S. 2015. Penalising model component complexity: A principled, practical approach to constructing priors. Statist. Sci. 32 (1) 1-28.

Smith, A.D., Girard, L., Boudreau, M., Van Beveren, E., and Plourde, S. 2022. Assessment of the northern contingent of Atlantic Mackerel (*Scomber scombrus*) in 2020. DFO Can. Sci. Advis. Sec. Res. Doc. 2022/045. iv + 44 p.

Thai, H.-T., Mentré, F., Holford, N.H.G., Veyrat-Follet, C., and Comets, E. 2013. A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. Pharm. Stat. 12(3): 129–140.

Van Beveren, E., Marentette, J.R., Duplisea, D., Castonguay, M., Kronlund, A.R., and Smith, A. 2020. Evaluation of Rebuilding Strategies for northwestern Atlantic Mackerel (NAFO Subareas 3 and 4). Can. Sci. Advis. Secr. Res. Doc. 2020/021: v + 56 p.

Van Beveren, E., Plourde, S., Pepin, P., Cogliati, K., and Castonguay, M. 2023a. A review of the importance of various areas for northern contingent West-Atlantic mackerel spawning. ICES J. Mar. Sci. 80(1): 1–15.

Van Beveren, E., Boudreau, M., Lévesque, L., Lehoux, C., Boudreau, M., and Plourde, S. 2023b. Assessment of the Northern Contingent of Atlantic Mackerel (Scomber scombrus) in 2022. DFO Can. Sci. Advis. Sec. Res. Doc. 2023/080. v + 48 p.

Van Beveren, E. 2024. Revision of Catch- and Maturity- at Age Used to Assess the Northern Contingent of Atlantic Mackerel (*Scomber scombrus*). DFO Can. Sci. Advis. Sec. Res. Doc. 2024/012. iv + 28 p.

van Guelpen, L., Markle, D.F., and Duggan, D.J. 1982. An evaluation of accuracy, precision, and speed of several zooplankton subsampling techniques. ICES J. Mar. Sci. 40 (3): 226–236.

Ware D. M. and Lambert, T. C. 1985. Early Life History of Atlantic Mackerel (*Scomber scombrus*) in the Southern Gulf of St. Lawrence. Can. J. Fish. Aquat. Sci. 42 (3): 577-592.

Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. 11: 3571–3594.

Worley, L.G. 1933. Development of the egg of the mackerel at different constant temperatures. J. Gen. Physiol. 16(5): 841-857.

6. ACKNOWLEDGMENTS

We would like to thank all harvesters and fishery organizations that helped collect mackerel samples. We are also grateful to Linda Girard, Mélanie Boudreau, Laurence Lévesque, Roxanne Noël, and Quentin Emblanc for the great amount of time spent analyzing mackerel and egg samples. We would like to thank data managers; Helena Talbot verified and archived the egg survey data and Caroline Lafleur provided the CTD data. We would also like to thank the crew of the CCGS Teleost and the technical support staff at the Maurice Lamontagne Institute (DAISS). We would like to thank Andrew Smith for sharing the R codes from the 2021 assessment. We further thank Jean-Martin Chamberland and Manuelle Beaudry-Sylvestre for the review of this manuscript.

7. TABLES

Table 1. Summary of all mackerel egg surveys and spawning season information. The spawning duration is estimated as the number of days over which 95% of all spawning occurred (difference between days of the year spanning the 95% confidence interval). Years flagged as likely biased (stage 6 fish and larvae bias indicators, inconsistency in logistic fit) and removed from the assessment are indicated in red. Years with larger uncertainty, but no indication of bias, because of the timing of the survey relative to the timing of spawning (outside the period when 70% of eggs are spawned) are indicated in blue. Years with larger uncertainty, but no indication of bias, because of poor sample coverage of the spawning season are indicated in green (2022). Averages were calculated excluding the second passes and years indicated in red. (DOY = day of year)

Year	Survey					Logistic model parameters			Spawning DOY					Proportion of egg spawned at median date ($S_{y,d=m}$)		
	Pass	N stations	Start Date	End Date	Median DOY	Asym	xmid	scal	2.5% CI	15% CI	85 % CI	97.5% CI	Duration	Mean	2.5% CI	97.5% CI
1979	1	59	-	-	166	19.8	168.7	-12.4	123	147	189	213	90	0.0199	0.0188	0.0206
1980						17.2	177.0	-12.9	130	155	198	223	93			
1981			Survey data unavailable			13.2	184.2	-6.1	162	174	194	205	43		Survey data unavailable	
1982						17.0	173.7	-16.4	114	145	201	233	119			
1983	1	63	06-23	07-02	178	13.9	174.0	-5.1	155	165	182	192	37	0.0419	0.0365	0.0425
1984	1	65	06-21	06-29	176	18.3	169.1	-14.3	117	144	193	221	104	0.0165	0.0146	0.0164
1985	1	63	06-22	07-01	178	20.3	166.1	-14.8	112	140	191	219	107	0.0144	0.0129	0.0147
1986	1	58	06-20	06-27	175	16.8	171.0	-13.7	121	147	194	220	99	0.0179	0.0161	0.0185
1987	1	64	06-17	06-26	172	13.8	172.2	-11.6	130	152	191	214	84	0.0215	0.0198	0.0220
1988	1	64	06-18	06-25	173	14.4	174.3	-6.9	149	162	185	199	50	0.0360	0.0336	0.0378
1988	2	65	06-27	07-04	183	14.4	174.3	-6.9	149	162	185	199	50	0.0249	0.0229	0.0251
1989	1	65	06-20	06-26	174	21.3	162.4	-14.9	108	136	187	216	108	0.0144	0.0129	0.0146
1989	2	65	06-27	07-04	181	21.3	162.4	-14.9	108	136	187	216	108	0.0116	0.0103	0.0118
1990	1	65	06-16	06-22	170	14.3	175.7	-9.8	140	159	192	211	71	0.0234	0.0224	0.0242
1990	2	65	06-22	06-28	176	14.3	175.7	-9.8	140	159	192	211	71	0.0254	0.0236	0.0262
1991	1	65	06-17	06-26	171	21.5	163.2	-17.7	98	133	193	227	129	0.0135	0.0122	0.0151
1992	1	65	06-16	06-24	173	14.9	175.3	-11.8	132	155	195	217	85	0.0210	0.0190	0.0218
1992	2	63	06-25	07-02	180	14.9	175.3	-11.8	132	155	195	217	85	0.0204	0.0178	0.0213
1993	1	65	06-15	06-23	170	17.4	173.3	-13.8	123	149	196	223	100	0.0178	0.0164	0.0183
1994	1	65	06-14	06-21	168	13.7	176.4	-11.2	135	157	195	216	81	0.0195	0.0185	0.0203
1994	2	64	06-20	06-27	175	13.7	176.4	-11.2	135	157	195	216	81	0.0223	0.0201	0.0233
1995			No survey			14.0	177.8	-10.2	141	160	194	214	73		No survey	
1996	1	65	06-18	06-24	173	17.7	168.7	-11.0	128	150	187	208	80	0.0218	0.0192	0.0232
1996	2	62	06-24	06-30	179	17.7	168.7	-11.0	128	150	187	208	80	0.0183	0.0160	0.0196
1997			No survey			15.4	174.9	-8.5	144	160	189	205	61		No survey	

Year	Survey					Logistic model parameters			Spawning DOY					Proportion of egg spawned at median date ($S_{y,d=m}$)		
	Pass	N stations	Start Date	End Date	Median DOY	Asym	xmid	scal	2.5% CI	15% CI	85 % CI	97.5% CI	Duration	Mean	2.5% CI	97.5% CI
1998	1	64	06-13	06-19	167	16.9	166.1	-10.5	128	148	183	203	75	0.0238	0.0217	0.0245
1998	2	65	06-17	06-24	172	16.9	166.1	-10.5	128	148	183	203	75	0.0221	0.0194	0.0229
1999	1	52	06-19	06-25	173	24.3	145.2	-15.4	89	118	171	201	112	0.0079	0.0070	0.0090
2000	1	62	06-18	06-26	173	15.3	170.8	-11.2	130	151	189	211	81	0.0221	0.0198	0.0224
2001	1	61	06-11	06-21	166	12.8	172.0	-6.9	147	160	183	196	49	0.0302	0.0292	0.0316
2002	1	63	06-16	06-23	170	15.3	172.0	-12.6	126	150	193	217	91	0.0198	0.0180	0.0202
2003	1	65	06-17	06-24	171	16.6	170.3	-9.0	137	155	185	202	65	0.0278	0.0249	0.0289
2004	1	64	06-15	06-22	171	17.7	170.6	-12.6	125	149	191	216	91	0.0199	0.0177	0.0209
2005	1	61	06-16	06-22	170	18.9	172.9	-10.6	134	154	190	211	77	0.0231	0.0217	0.0238
2006	1	63	06-29	07-08	186	18.1	165.4	-10.4	127	147	182	202	75	0.0102	0.0094	0.0103
2007	1	66	06-21	06-29	177	13.9	177.5	-7.0	152	165	189	202	50	0.0358	0.0329	0.0373
2008	1	53	06-21	06-29	177	13.0	175.6	-6.2	153	165	185	197	44	0.0401	0.0364	0.0413
2009	1	64	06-13	06-21	168	12.6	174.0	-8.0	145	160	187	202	57	0.0272	0.0251	0.0295
2010	1	65	06-10	06-19	166	15.1	170.2	-9.2	136	154	185	203	67	0.0257	0.0242	0.0270
2011	1	66	06-11	06-19	166	17.5	167.9	-7.3	141	155	180	194	53	0.0335	0.0308	0.0355
2012	1	66	06-11	06-19	168	17.0	161.4	-9.0	128	146	176	193	65	0.0243	0.0198	0.0266
2013	1	66	06-11	06-20	167	15.3	168.6	-8.3	138	154	182	198	60	0.0300	0.0272	0.0310
2014	1	64	06-10	06-20	166	16.7	166.7	-7.9	138	153	179	195	57	0.0315	0.0278	0.0337
2015	1	60	06-12	06-20	167	15.9	172.2	-7.2	146	160	184	198	52	0.0305	0.0291	0.0319
2016	1	65	06-11	06-23	169	20.2	163.6	-11.6	121	143	183	205	84	0.0204	0.0182	0.0215
2017	1	65	06-10	06-17	164	14.2	174.1	-4.8	157	166	181	191	34	0.0201	0.0201	0.0232
2018	1	65	06-16	06-24	172	13.2	175.0	-5.3	155	166	183	194	39	0.0432	0.0392	0.0463
2019	1	65	06-07	06-15	162	15.3	176.4	-7.1	150	164	188	201	51	0.0145	0.0140	0.0165
2020			No survey			16.0	172.7	-2.4	164	169	176	181	17		No survey	
2021	1	61	06-12	06-22	169	13.9	167.1	-6.2	144	156	177	189	45	0.0391	0.0350	0.0401
2022	1	65	06-17	06-26	171	18.2	161.2	-9.3	127	145	176	194	67	0.0206	0.0140	0.0289
Average	-	63	-	-	171	16.1	171.2	-9.7	136	154	187	206	70	0.0250	0.0227	0.0263

Table 2. Total egg production estimates and parameters (daily egg production and proportion of eggs spawned at the median date of the survey) for the baseline scenario. Years flagged as likely biased (stage 6 fish and larvae bias indicators, inconsistency in logistic fit) and removed from the assessment are indicated in red. Years with larger uncertainty, but not indication of bias, because of the timing of the survey relative to the timing of spawning (outside the period when 70% of eggs are spawned) are indicated in blue. Years with larger uncertainty, but no indication of bias, because of poor sample coverage of the spawning season are indicated in green (2022).

Year	Daily egg production	Spawning proportion	Total egg production	Year	Daily egg production	Spawning proportion	Total egg production
1979	119	0.0199	4.16E+14	2001	10	0.0302	2.40E+13
1980				2002	76	0.0198	2.66E+14
1981		Data unavailable		2003	98	0.0278	2.46E+14
1982				2004	77	0.0199	2.67E+14
1983	65	0.0419	1.08E+14	2005	40	0.0231	1.19E+14
1984	85	0.0165	3.59E+14	2006	7	0.0102	4.60E+13
1985	141	0.0144	6.78E+14	2007	45	0.0358	8.71E+13
1986	263	0.0179	1.02E+15	2008	57	0.0401	9.84E+13
1987	168	0.0215	5.43E+14	2009	27	0.0272	6.98E+13
1988	236	0.0360	4.55E+14	2010	10	0.0257	2.57E+13
1989	112	0.0144	5.38E+14	2011	14	0.0335	2.95E+13
1990	119	0.0234	3.54E+14	2012	4	0.0243	1.09E+13
1991	210	0.0135	1.08E+15	2013	17	0.0300	3.84E+13
1992	176	0.0210	5.83E+14	2014	22	0.0315	4.77E+13
1993	175	0.0178	6.83E+14	2015	20	0.0305	4.56E+13
1994	89	0.0195	3.16E+14	2016	15	0.0204	4.95E+13
1995		No survey		2017	22	0.0201	7.56E+13
1996	30	0.0218	9.43E+13	2018	28	0.0432	4.54E+13
1997		No survey		2019	21	0.0145	9.96E+13
1998	26	0.0238	7.47E+13	2020		No survey	
1999	32	0.0079	2.85E+14	2021	9	0.0391	1.64E+13
2000	36	0.0221	1.13E+14	2022	11	0.0206	3.72E+13

8. FIGURES

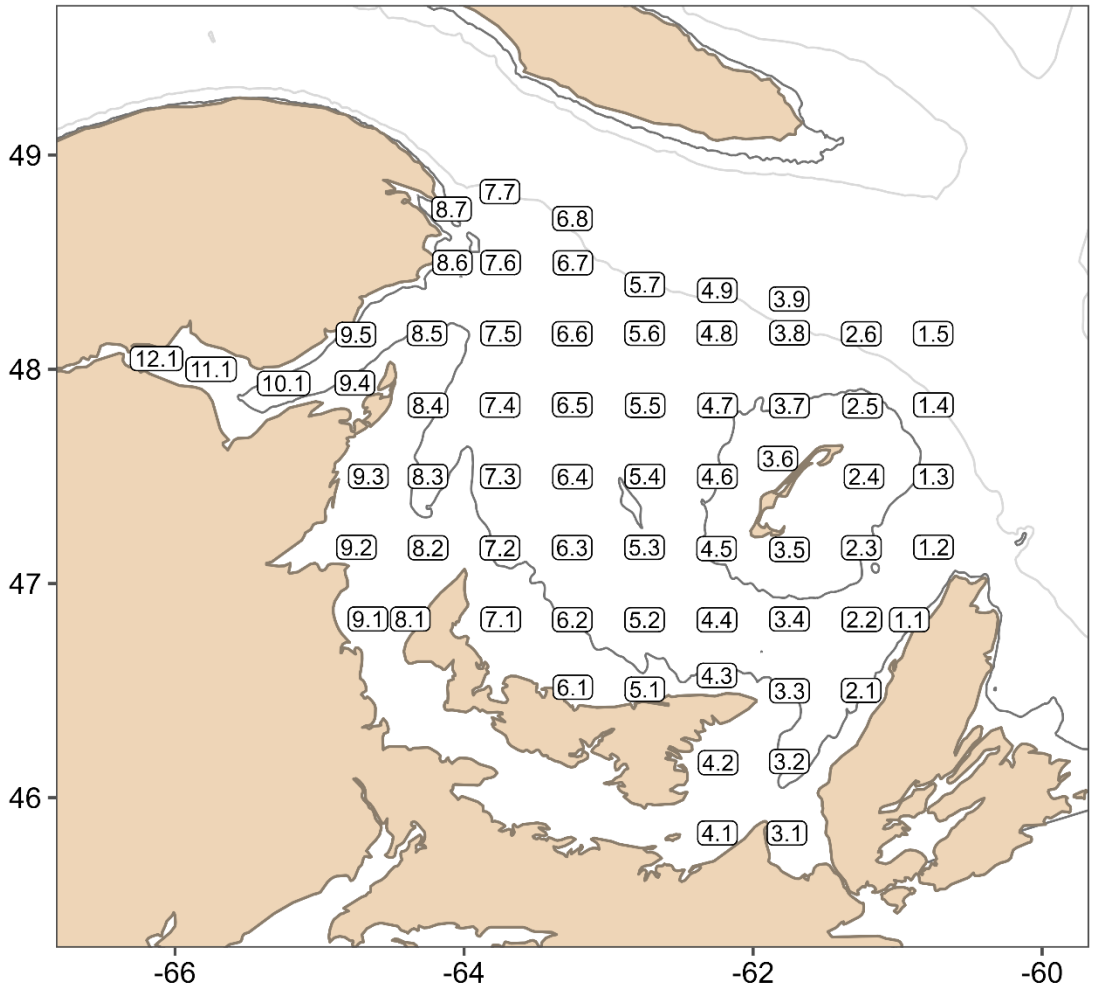

Figure 1. Stations sampled during the June egg survey. The dark grey and light grey lines represent the 50 m and 200 m isobaths, respectively.

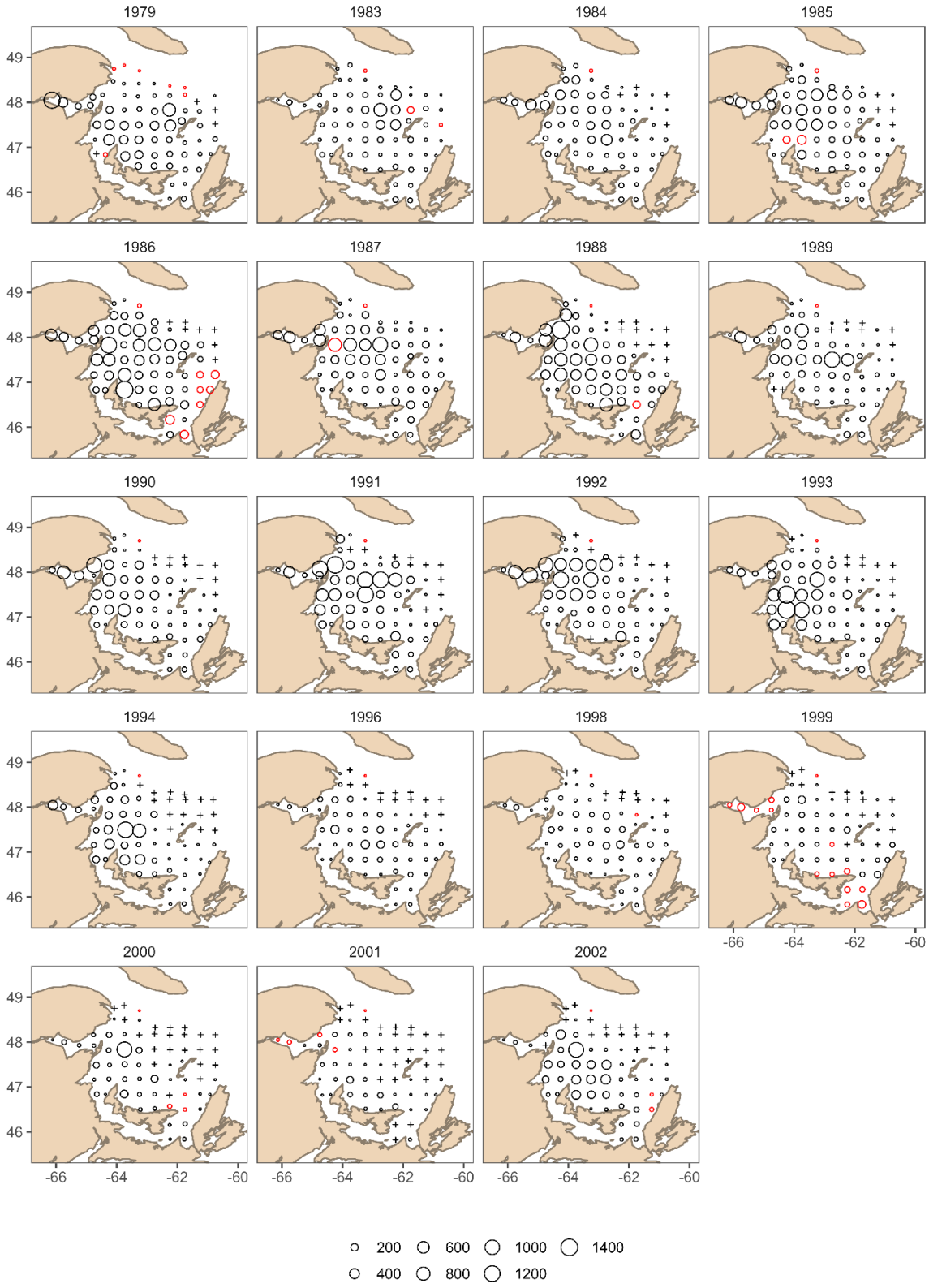

*Figure 2. Daily egg production (n*m^{-2}*day^{-1}). Absences of eggs are indicated with crosses. Predicted values for unsampled stations are indicated in red.*

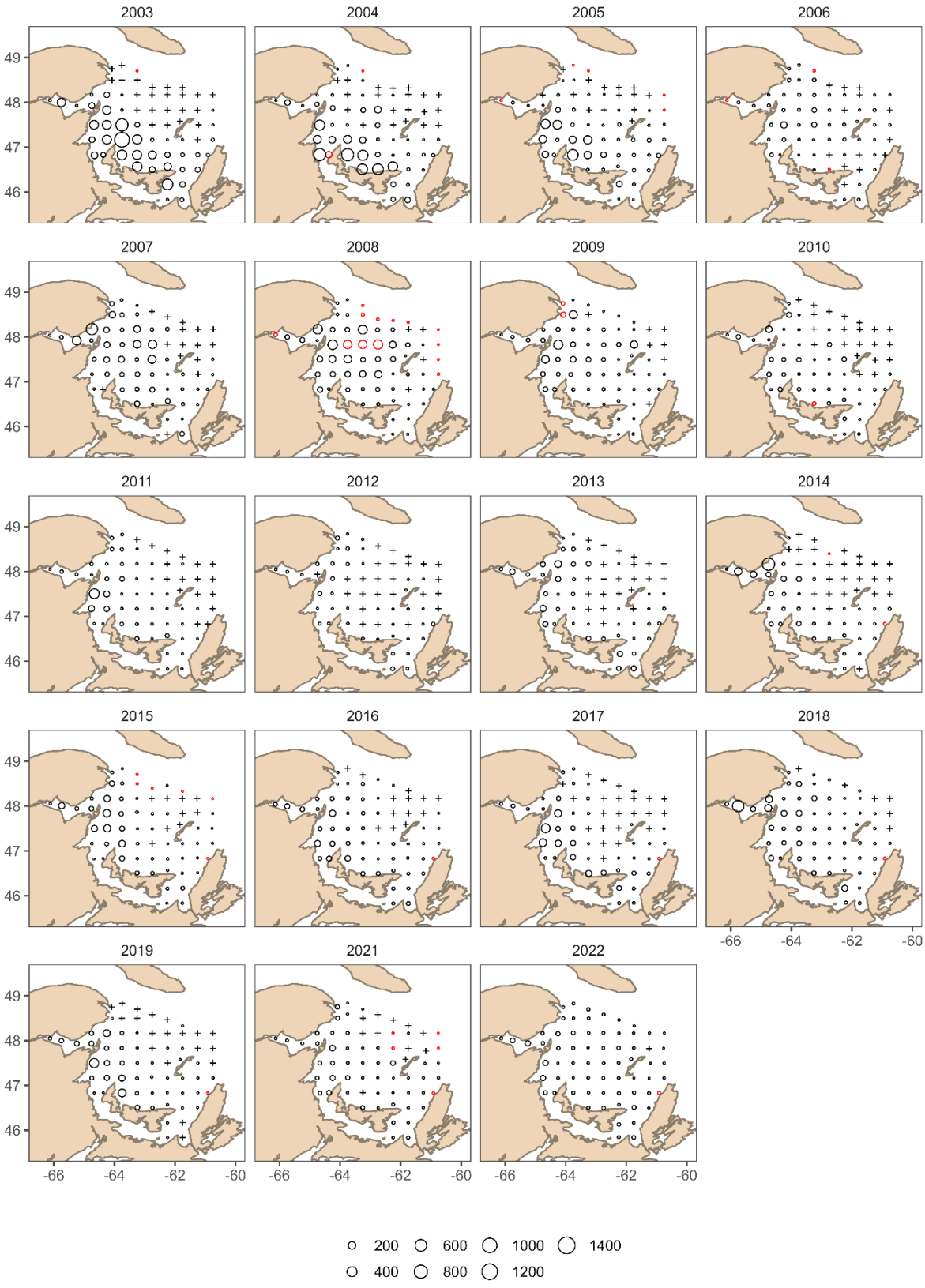

Figure 2 (continued).

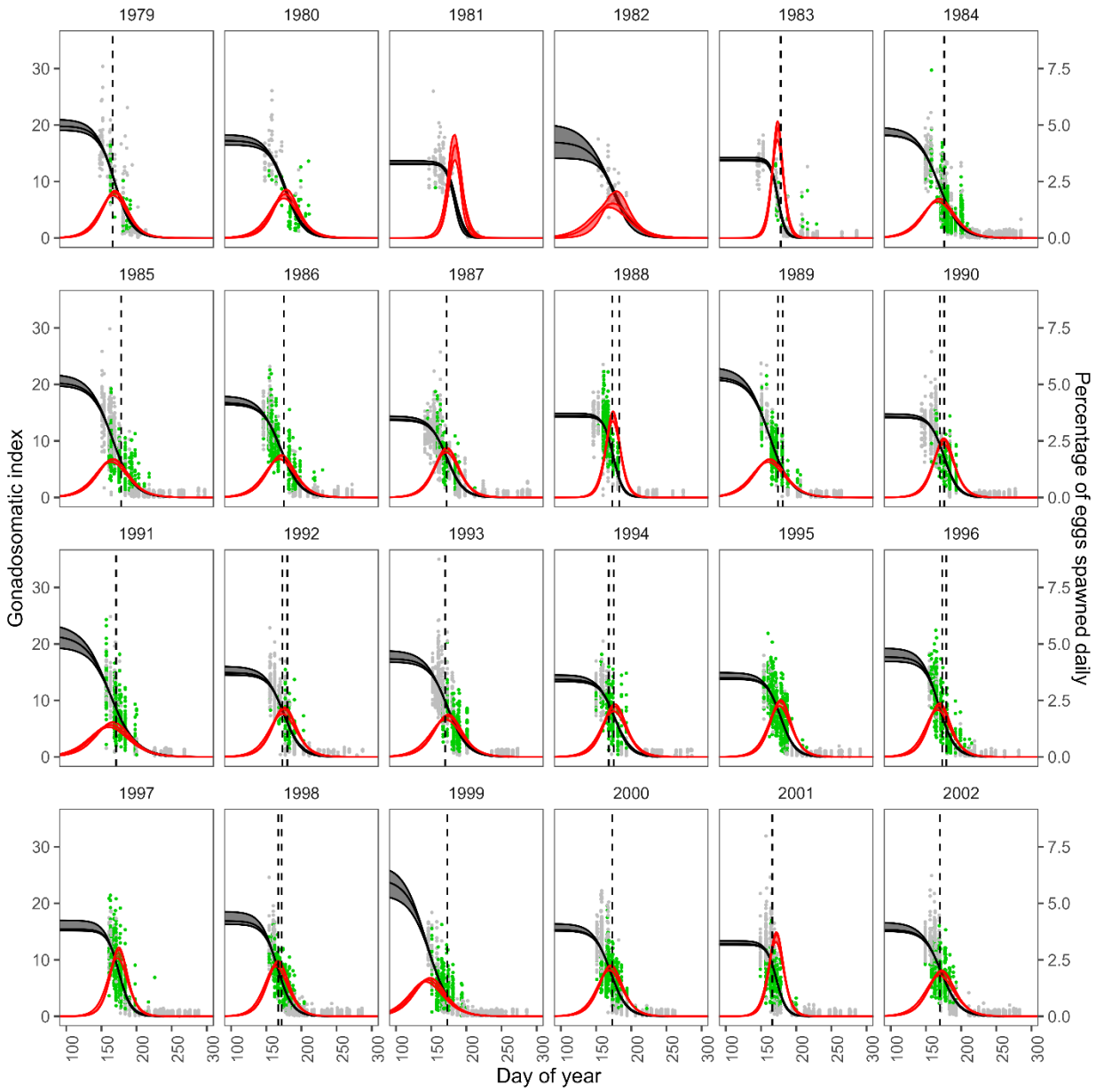

Figure 3. Logistic model fit (black line and shaded 95% confidence intervals) relating the gonadosomatic index (GSI %) to the day of year. The observations are indicated by the green (maturity stage 6) and grey (maturity stages 5, 7 and 8) dots. The black dashed vertical lines represent the median date of the survey(s). The red curve represents the daily percentage of eggs spawned, with a 95% confidence interval.

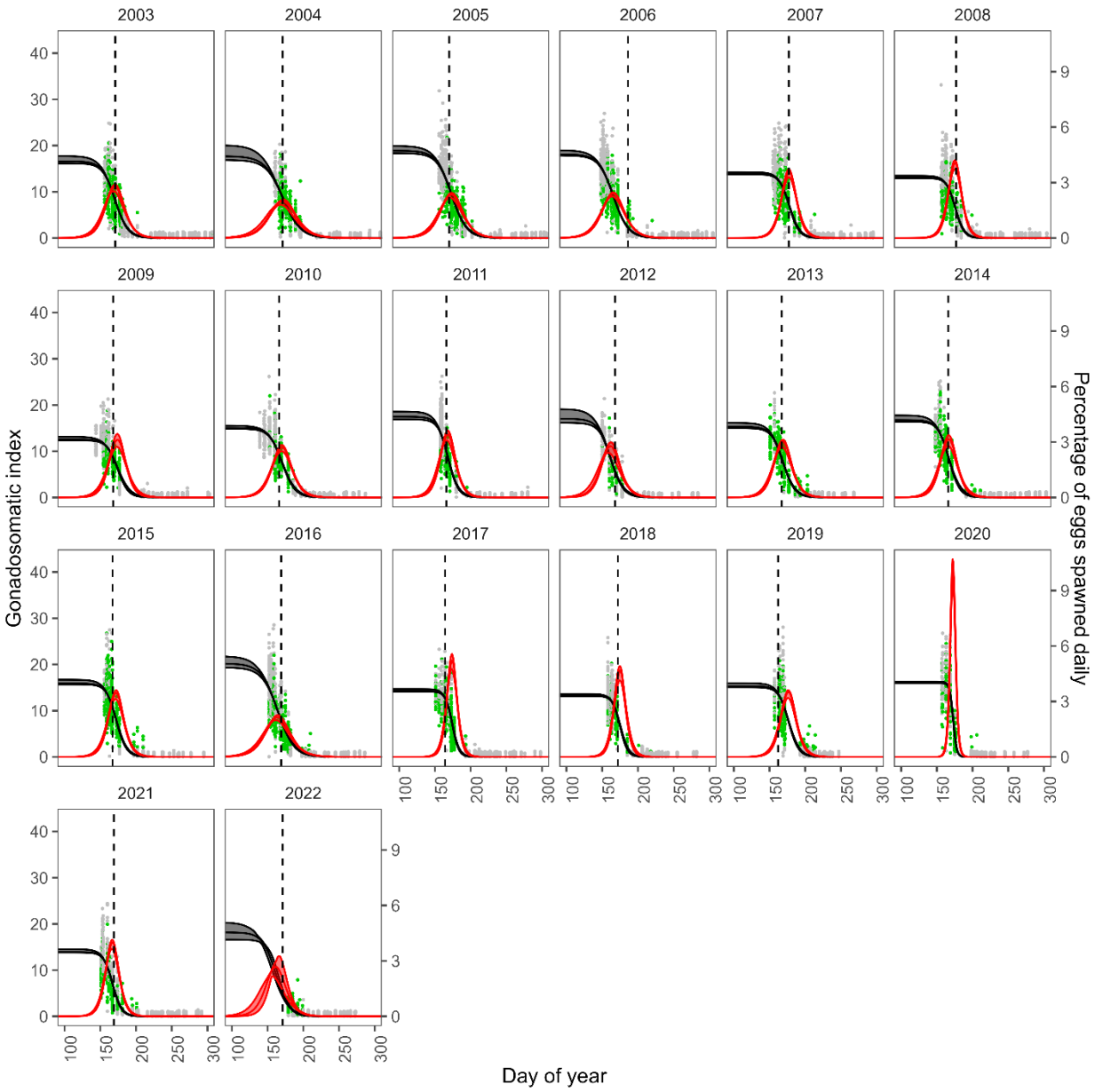

Figure 3. (continued)

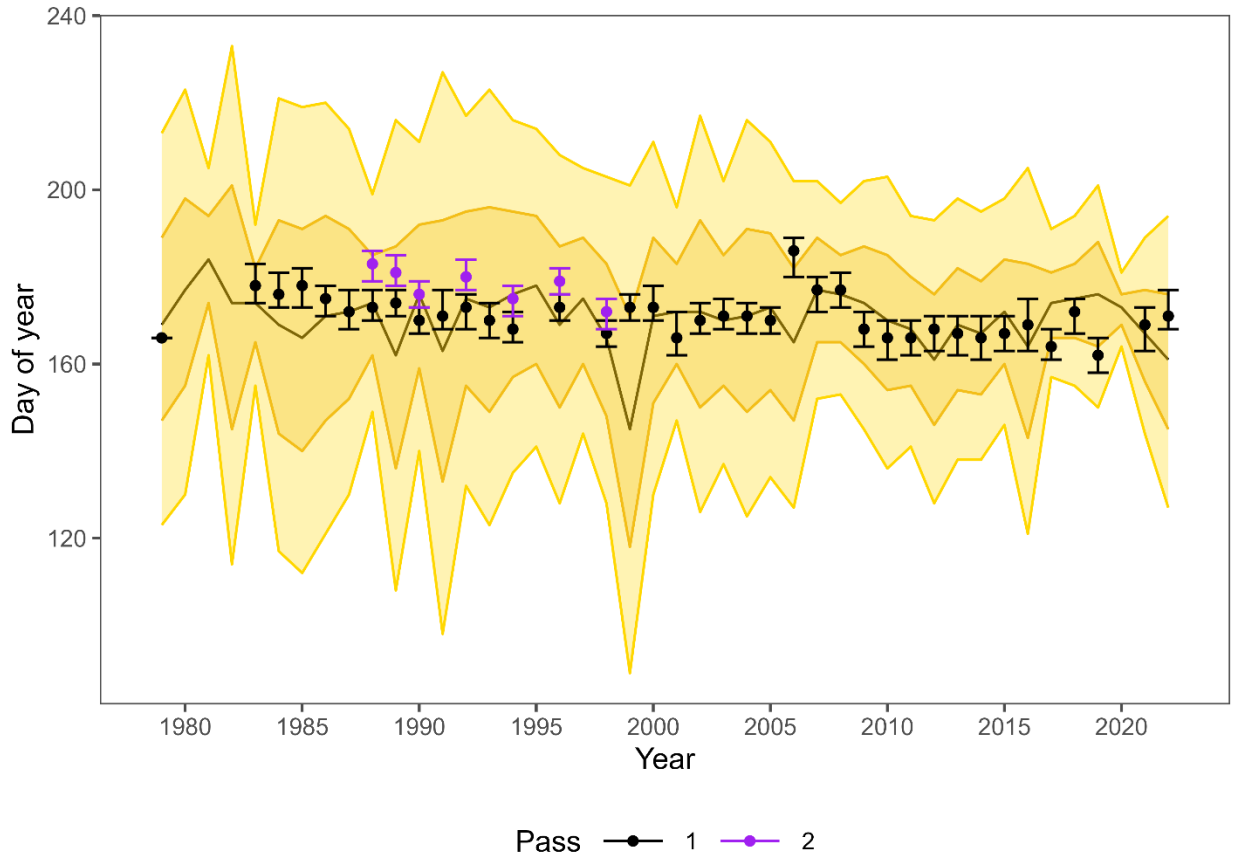

Figure 4. Timing of the survey (circles indicate the median date and bars the start and end date) relative to the spawning season (black line indicates the peak and the light and dark orange ribbons indicate the period over which 95% and 70% of eggs are spawned, respectively). The first and second survey passes are indicated in black and purple, respectively. Only the median date was available for 1979.

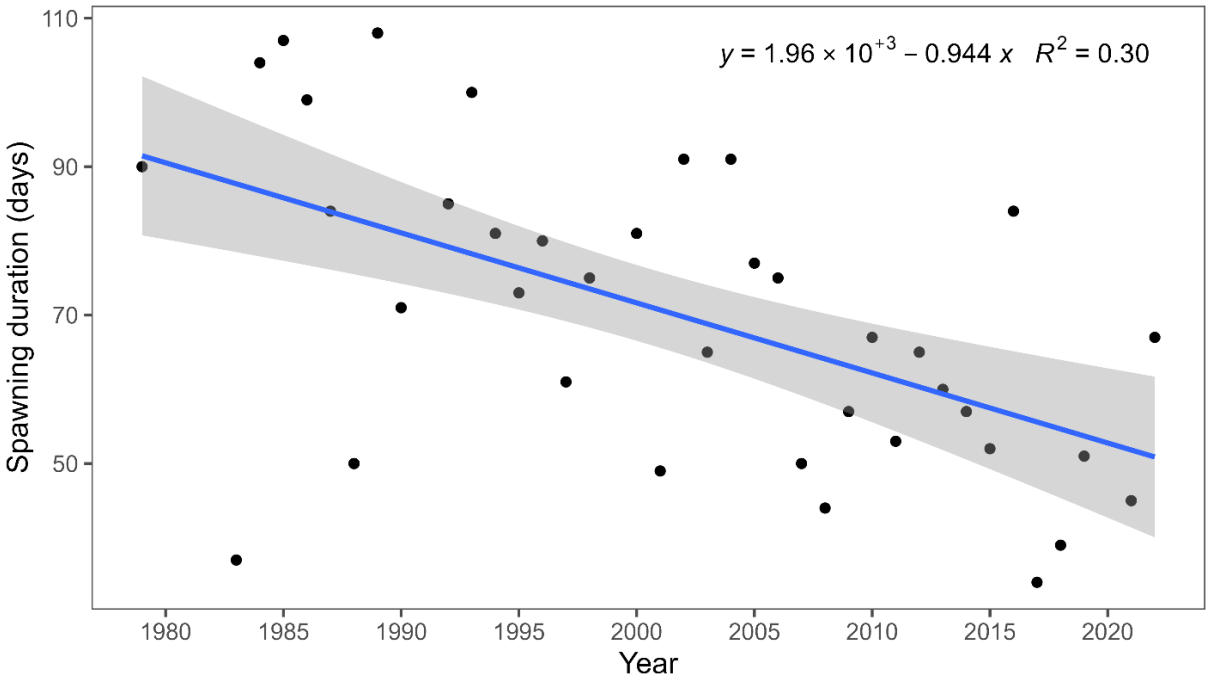

Figure 5. Linear regression between spawning duration (number of days over which 95% of eggs were spawned) and time. Estimates for 1991 and 1999 were excluded because of the poor fit of the logistic model.

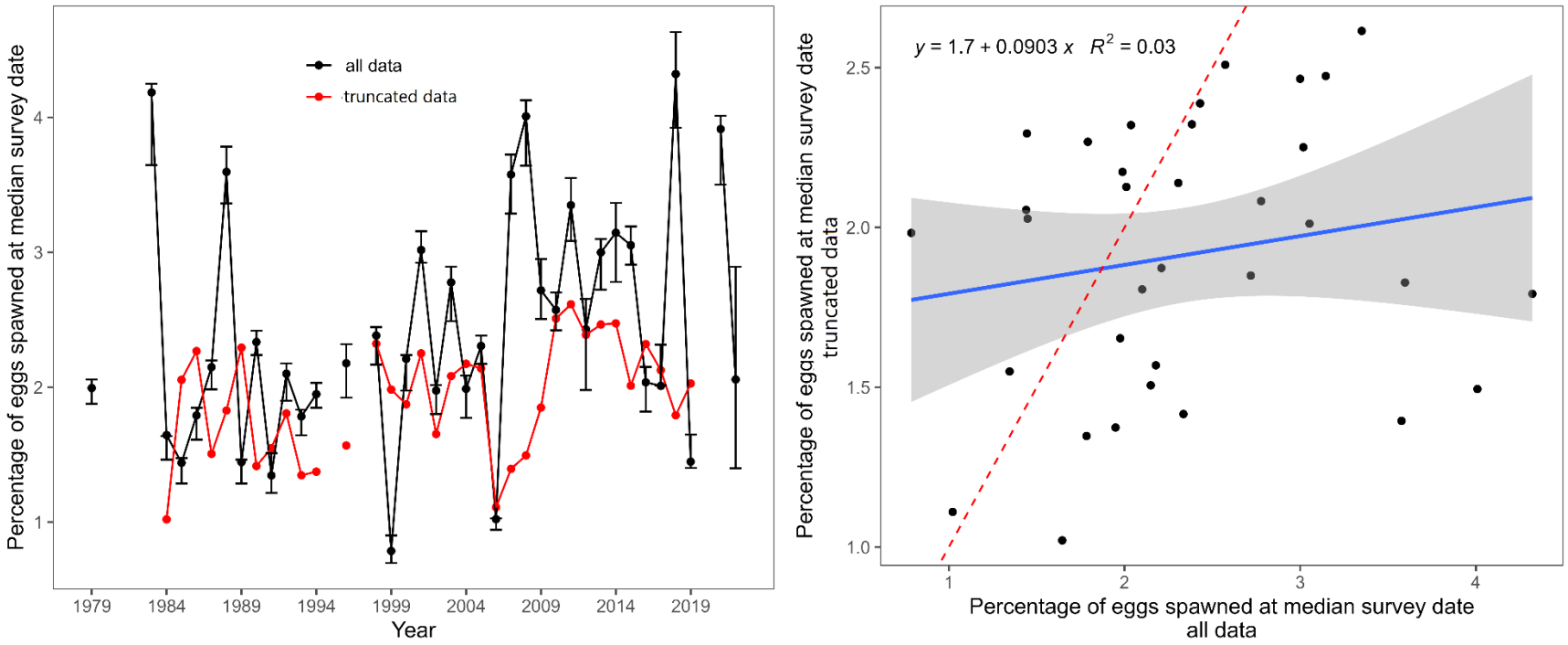

Figure 6. Comparison of logistic model predictions (mean with 95% confidence interval) between the baseline scenario (all data available) and a scenario in which for the given year gonad data was only available from day 179 onwards (mimicking 2022).

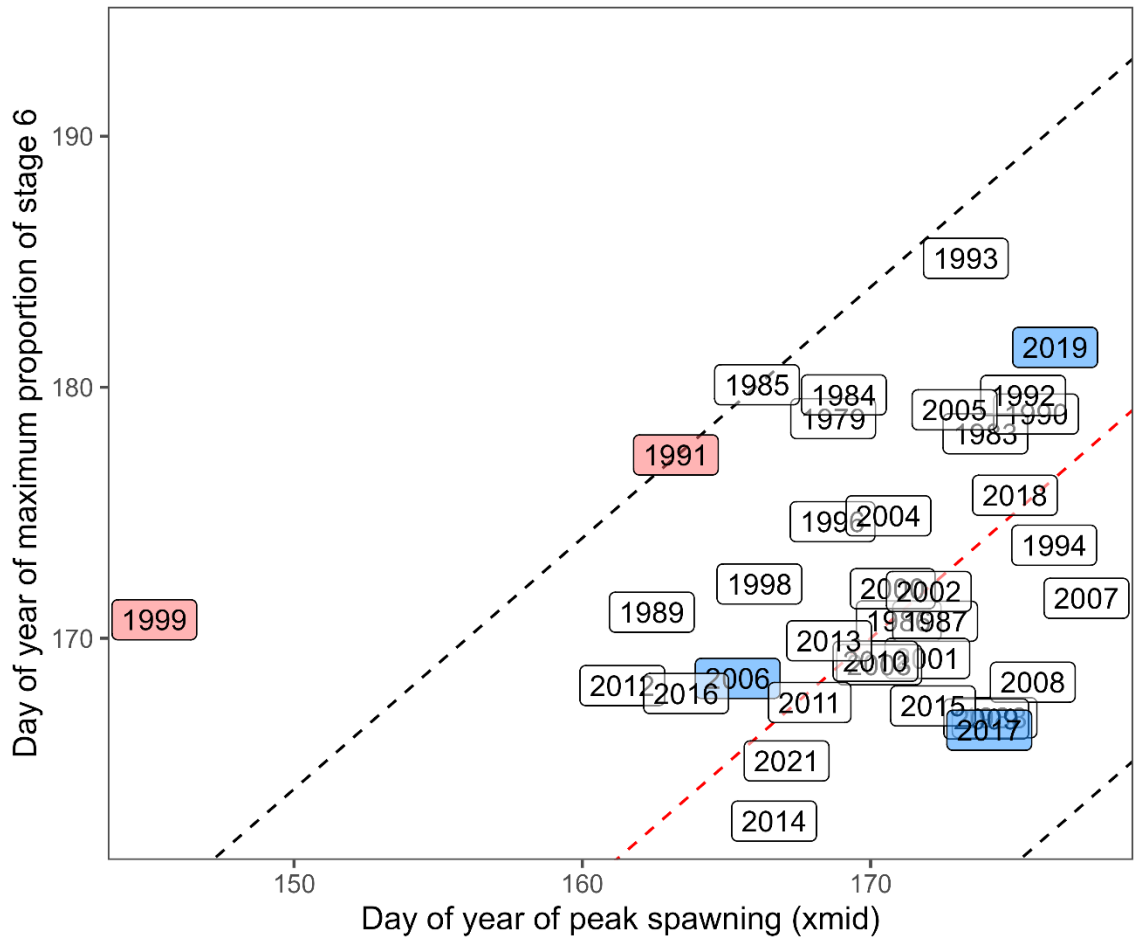

*Figure 7. Independent validation of the estimated spawning season, based on the day of highest proportion of stage 6 mackerel in the samples. The red dashed line indicates a perfect fit between the day of the highest proportion of stage 6 and the day of peak spawning (slope=1, intercept=0). The black dashed line represents relationships with intercepts of ±1.96 * the standard deviation of the difference in days between the peak of stage 6 and the peak of spawning (slope=1). Years flagged as likely biased (larvae bias indicator, inconsistency in logistic fit) and removed from the assessment are indicated in red. Years with larger uncertainty, but not indication of bias, because of the timing of the survey relative to the timing of spawning (outside the period when 70% of eggs are spawned) are indicated in blue. Years with larger uncertainty, but no indication of bias, because of poor sample coverage of the spawning season are not shown (2022).*

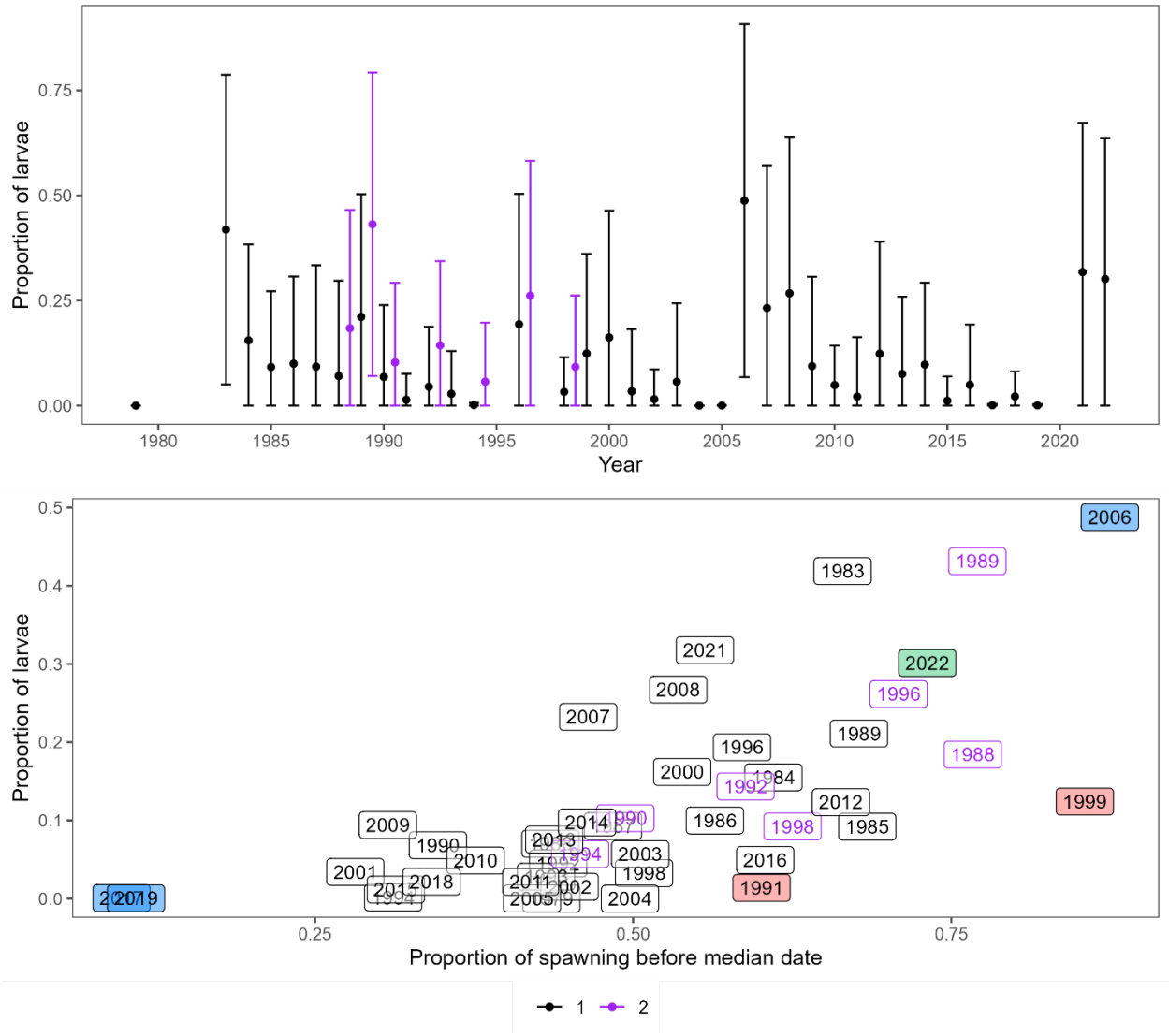

Figure 8. Independent validation of the estimated spawning season, based on the proportion of larvae observed in the samples (eq 2.6.4). The top panel shows the annual proportion of larvae (mean ± sd across stations) for the first (black) and, when available, the second survey pass (purple). The bottom panel shows the mean proportion of larvae observed annually in function of the estimated cumulative proportion of eggs spawned prior to the survey median date. Years flagged as likely biased (larvae bias indicator, inconsistency in logistic fit) and removed from the assessment are indicated in red. Years with larger uncertainty, but not indication of bias, because of the timing of the survey relative to the timing of spawning (outside the period when 70% of eggs are spawned) are indicated in blue. Years with larger uncertainty, but no indication of bias, because of poor sample coverage of the spawning season are indicated in green (2022).

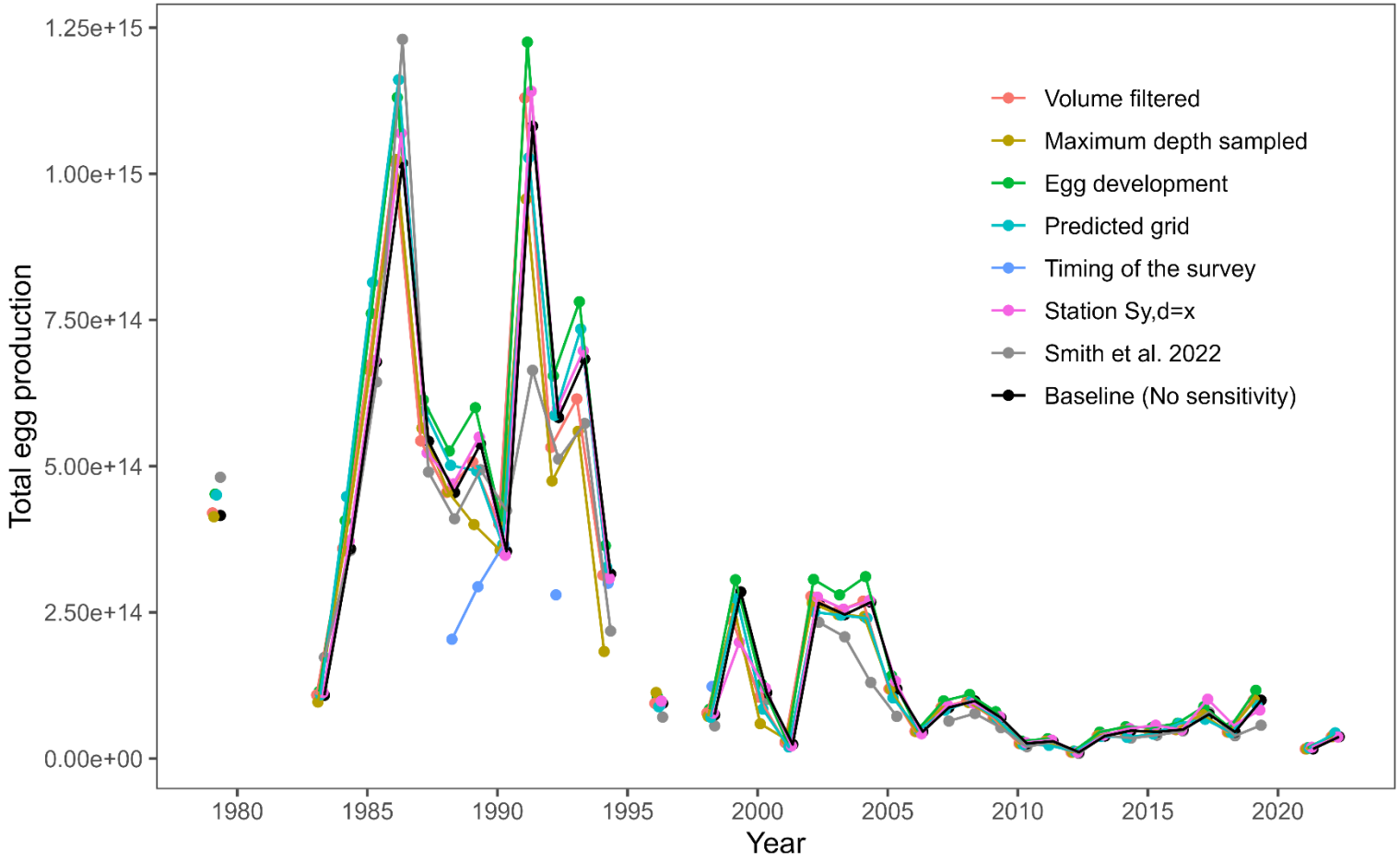

Figure 9. Annual total egg production, as estimated under the baseline scenario (black line), the different sensitivity runs (coloured lines), and during the 2021 mackerel stock assessment (grey line; Smith et al. 2022).

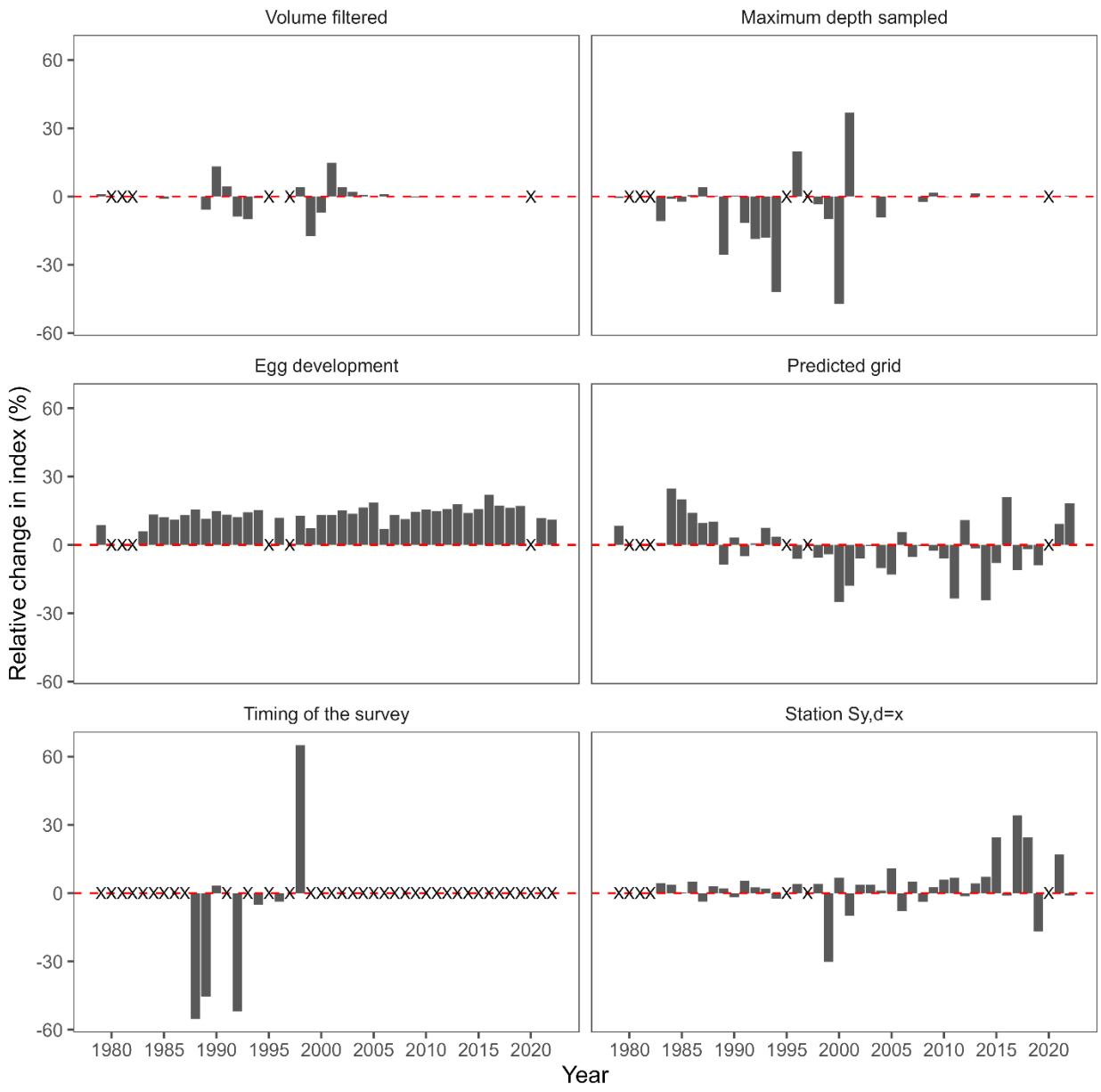

Figure 10. Percentage change in annual total egg production for each sensitivity run, relative to the baseline scenario (Red dashed line= no change). Years for which no comparison was made are marked with a X.

APPENDIX 1 - SENSITIVITY

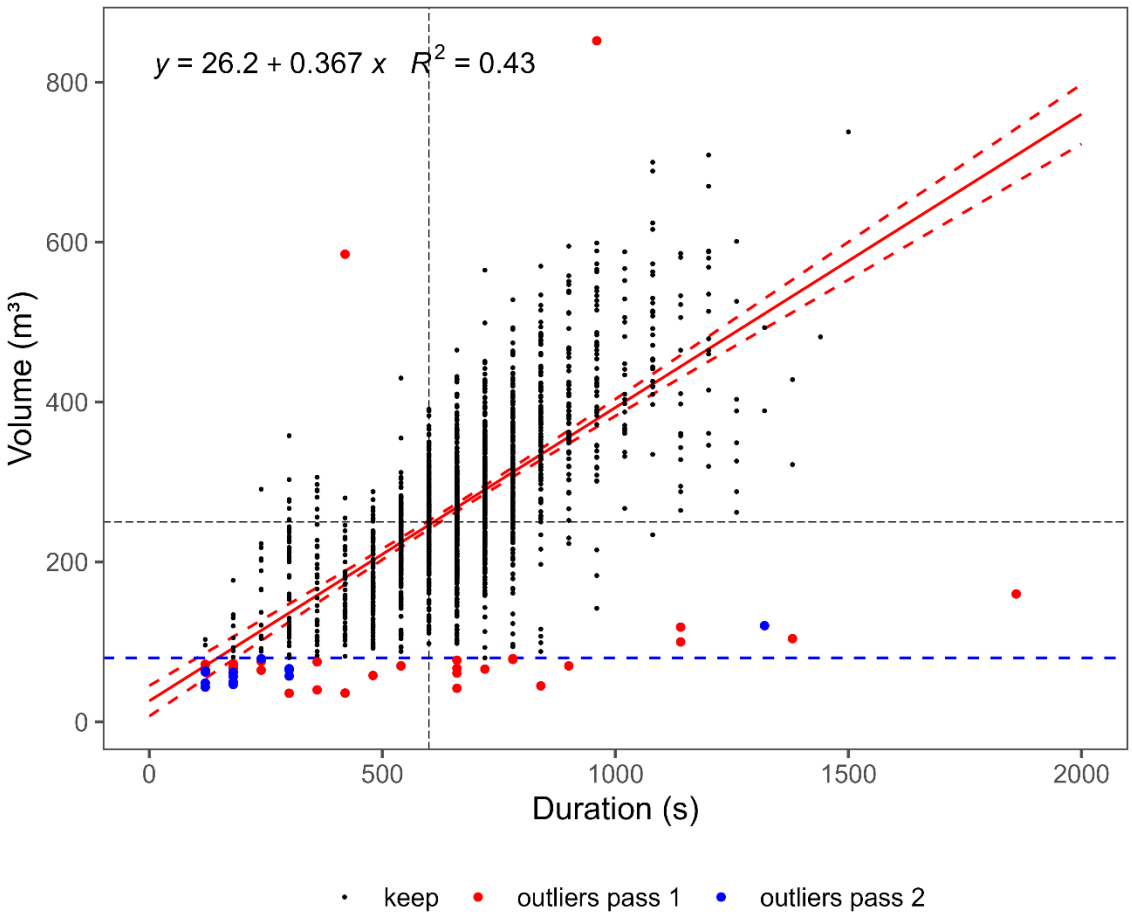

Figure A.1.1. Relationship between sampled volume and tow duration for ichthyoplankton surveys conducted over 1979-2022. The fitted linear relationship with 95% confidence intervals are represented by the solid red and dashed lines respectively. The grey dashed lines indicate the expected filtered volume for a standard 10-minute tow. Potential outliers for the first and second passes are identified in red and blue circles respectively. Only outliers from the first pass were removed from the volume filtered sensitivity run.

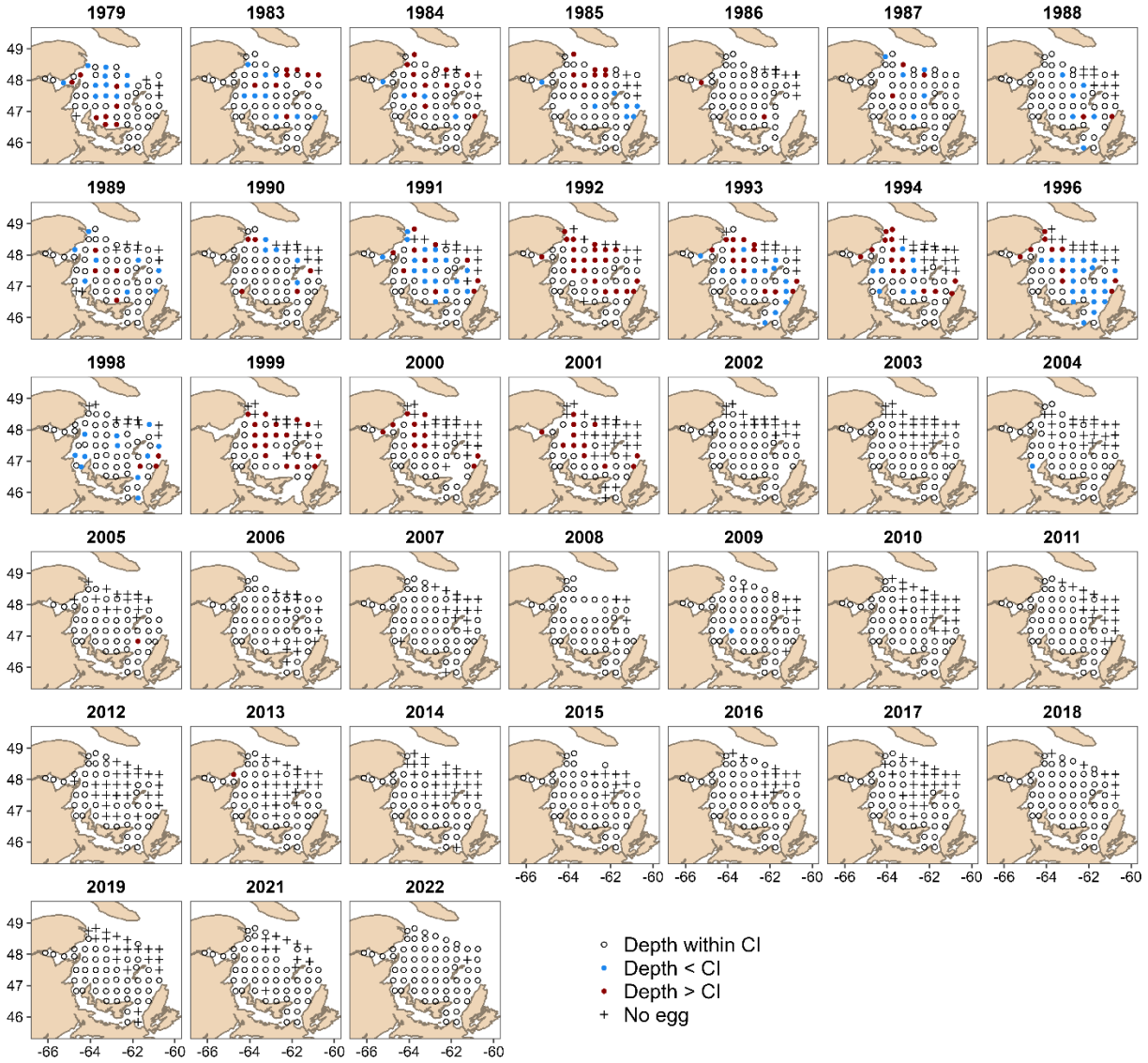

*Figure A.1.2. Sampled depth outliers (coloured dots). Stations for which the deepest point sampled (sampled depth) is outside the 99.9% confidence interval (CI) determined for that station across 2002-2022 (mean ± 3.291*sd) are indicated in blue (exceptionally shallow) or red (exceptionally deep). Stations where no eggs were observed were systematically not defined as outliers, as they should not be impacted by an incongruous sampled depth (black crosses).*

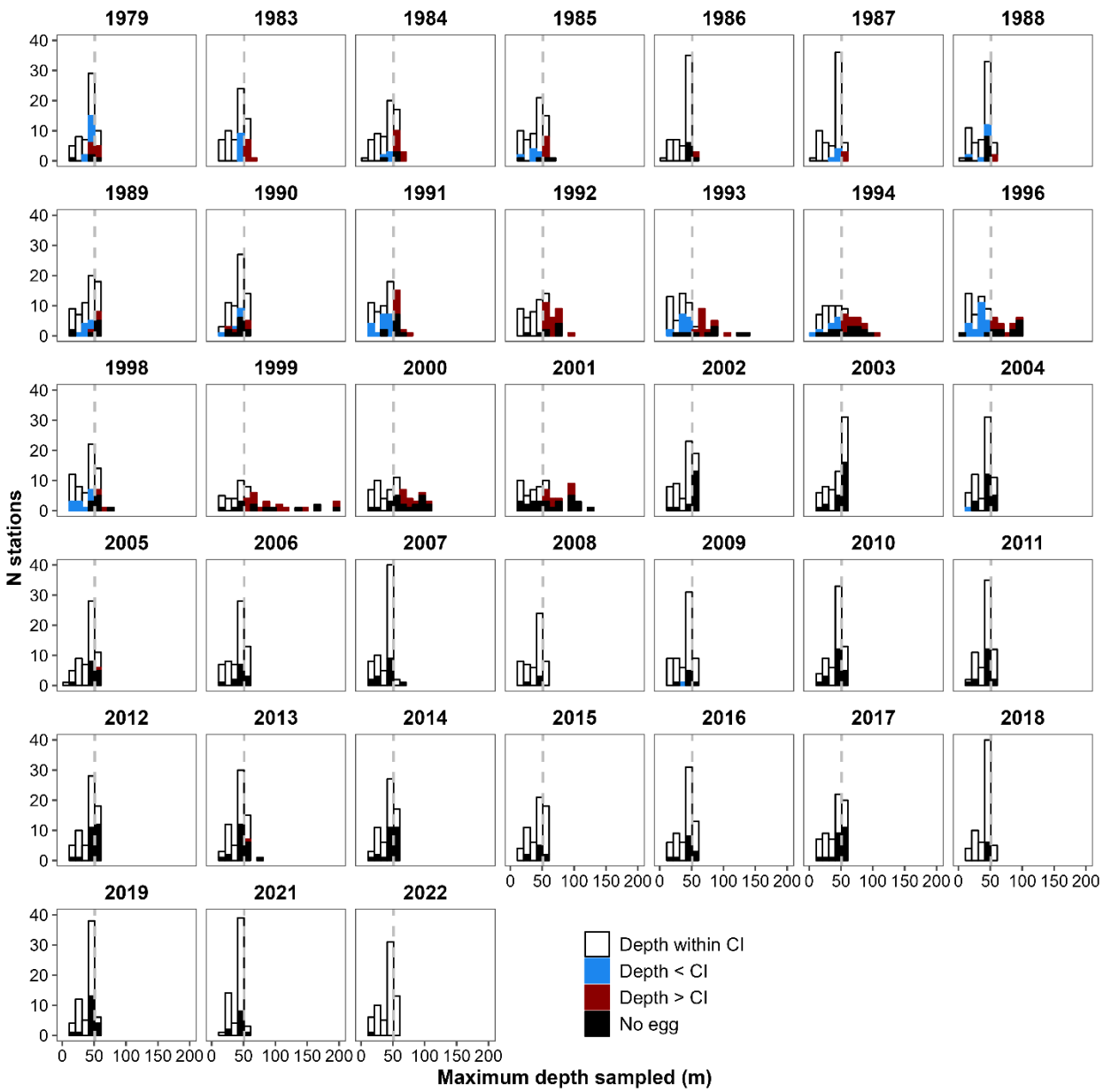

*Figure A.1.3. Verification of recorded sampled depths. Barplots show the distribution of sampled depths within a given year. Stations for which the deepest point sampled (sampled depth) is outside the 99.9% confidence interval determined for that station across 2002-2022(mean ± 3.291*sd) are indicated in blue (exceptionally shallow) or red (exceptionally deep).The grey dashed vertical line represents the targeted depth for deep-water stations (50 m). Stations where no eggs were observed (black) were systematically not defined as outliers, as they should not be impacted by an incongruous sampled depth.*

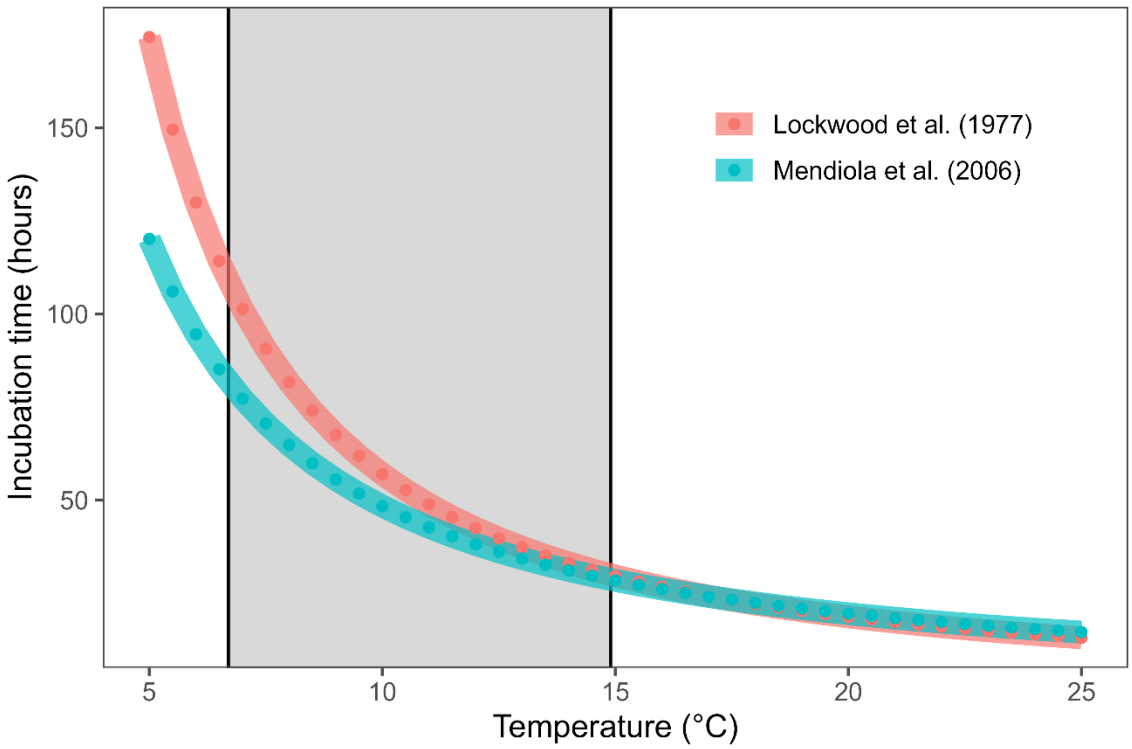

Figure A.1.4. Predicted incubation time (hours) of stage 1 eggs in function of temperature, following the equations of Lockwood et al. (1977) and Mendiola et al. (2006). The grey rectangle delineates the temperature range measured in 95% of the stations throughout all years (1979-2022).

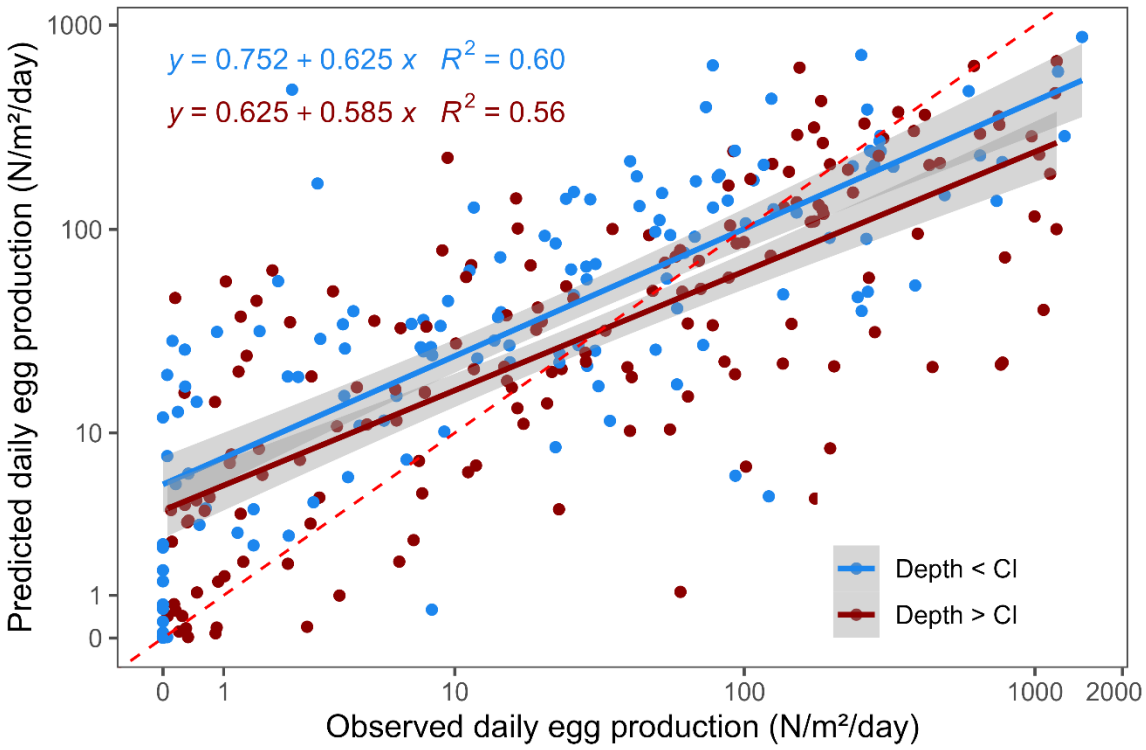

*Figure A.1.5. Prediction of the "depth outliers" by R-INLA compared to the observed values. Stations for which the deepest point sampled (sampled depth) is outside the 99.9% confidence interval determined for that station across all survey years from 1979 to 2022 (mean ± 3.291*sd) are indicated in blue (exceptionally shallow) or red (exceptionally deep).*

APPENDIX 2 - RINLA

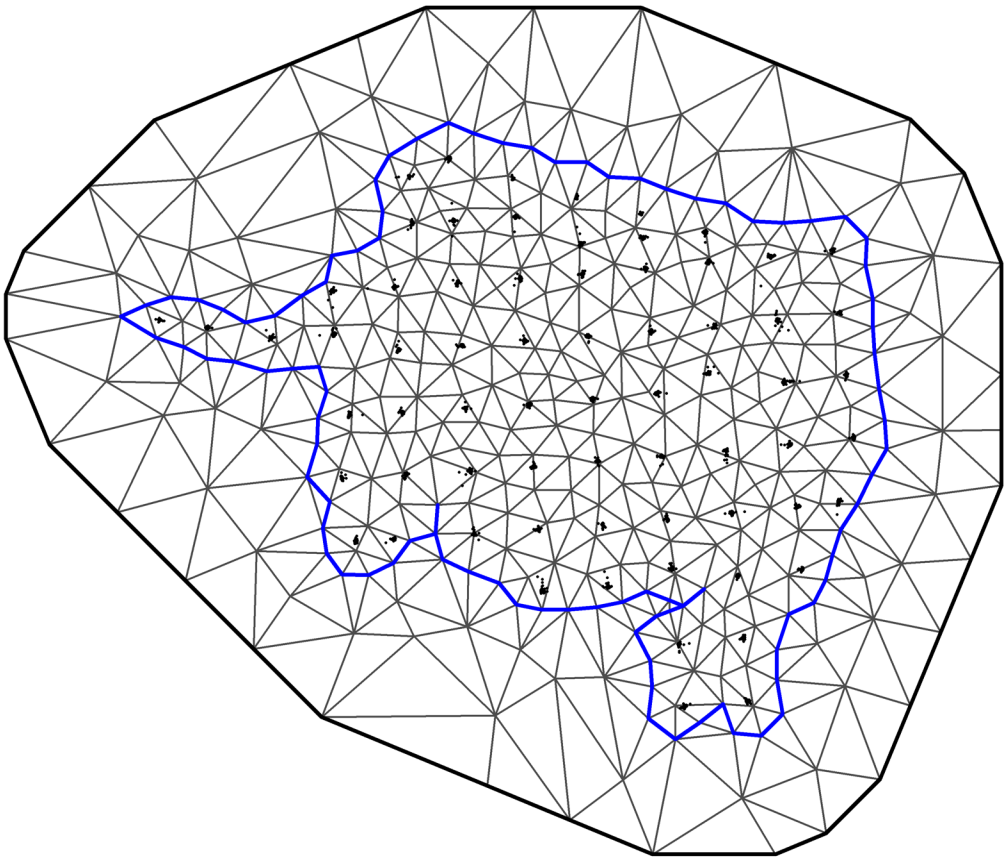

Figure A.2.1. Mesh used for the spatial random field to estimate unsampled stations. The blue line delineates the boundary. Triangles outside of this boundary have longer vertices (i.e., triangle sides). Station positions are indicated with black dots.

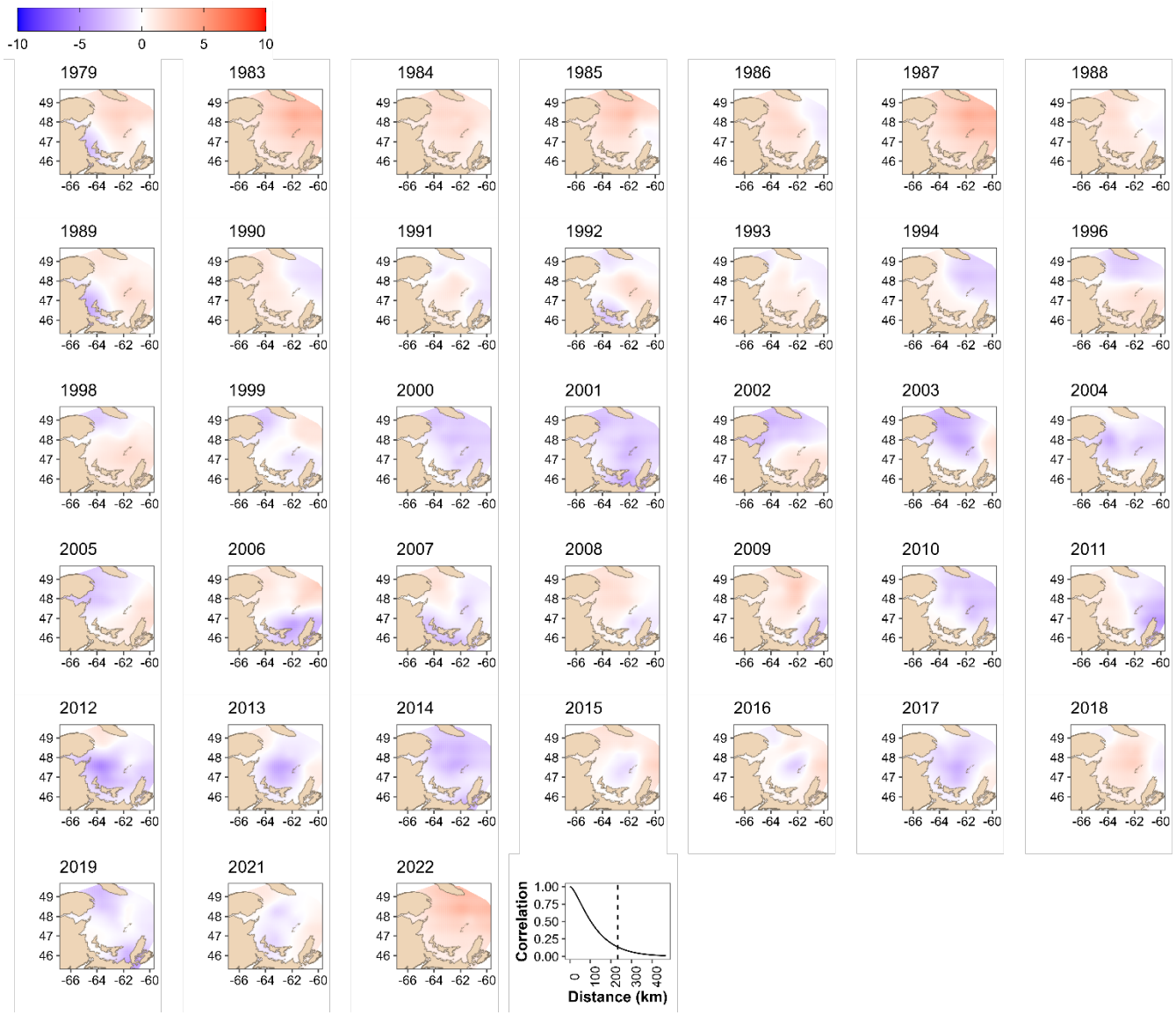

Figure A.2.2. Spatial random fields (w_y) predicted with a Bernoulli distribution. The estimated Matérn spatial correlation function is provided in the last panel, in which the dashed vertical line represents the range of spatial correlation.

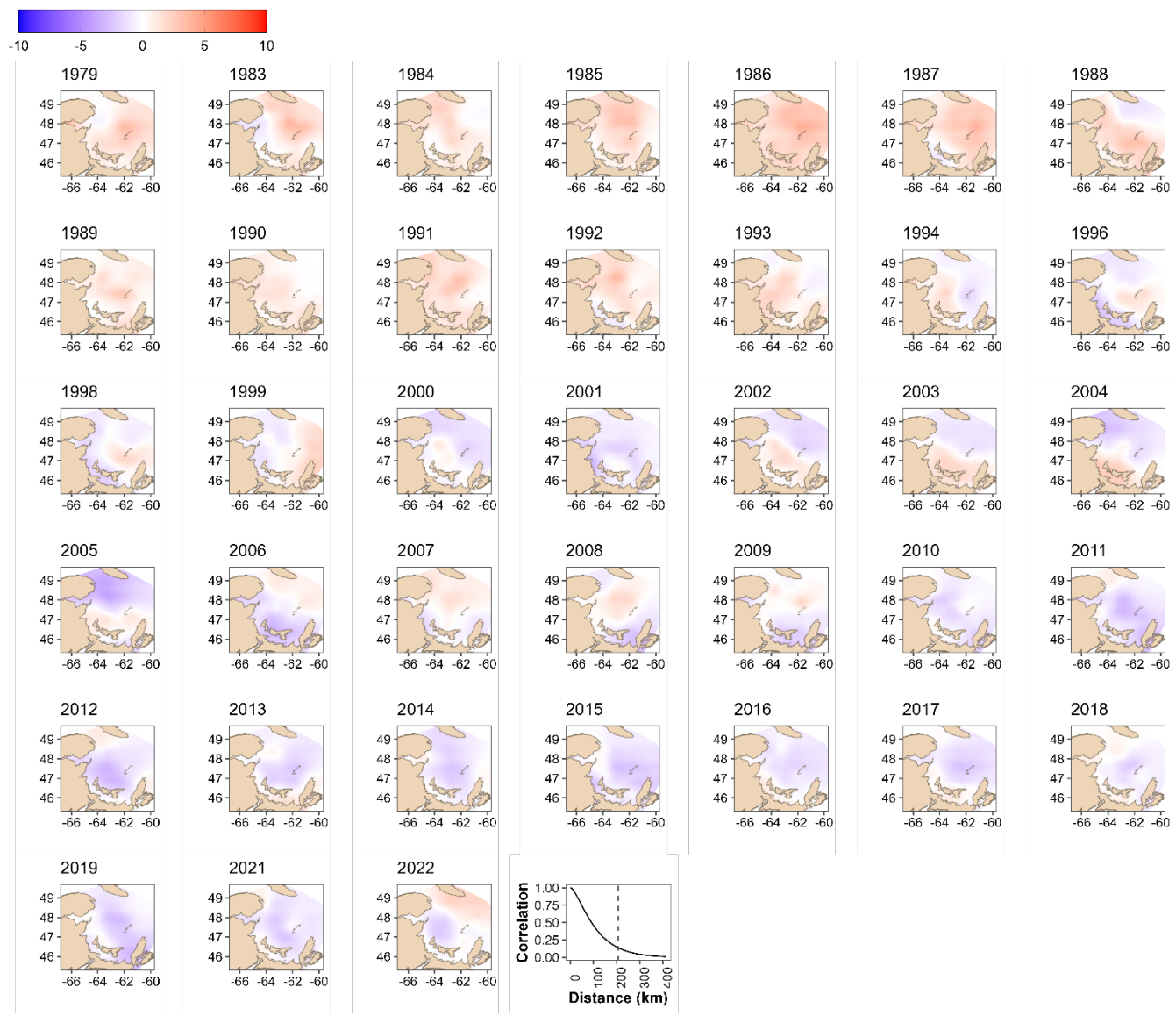

Figure A.2.3. Spatial random fields (w_y) predicted with a Gamma distribution. The estimated Matérn spatial correlation function is provided in the last panel, in which the dashed vertical line represents the range of spatial correlation.

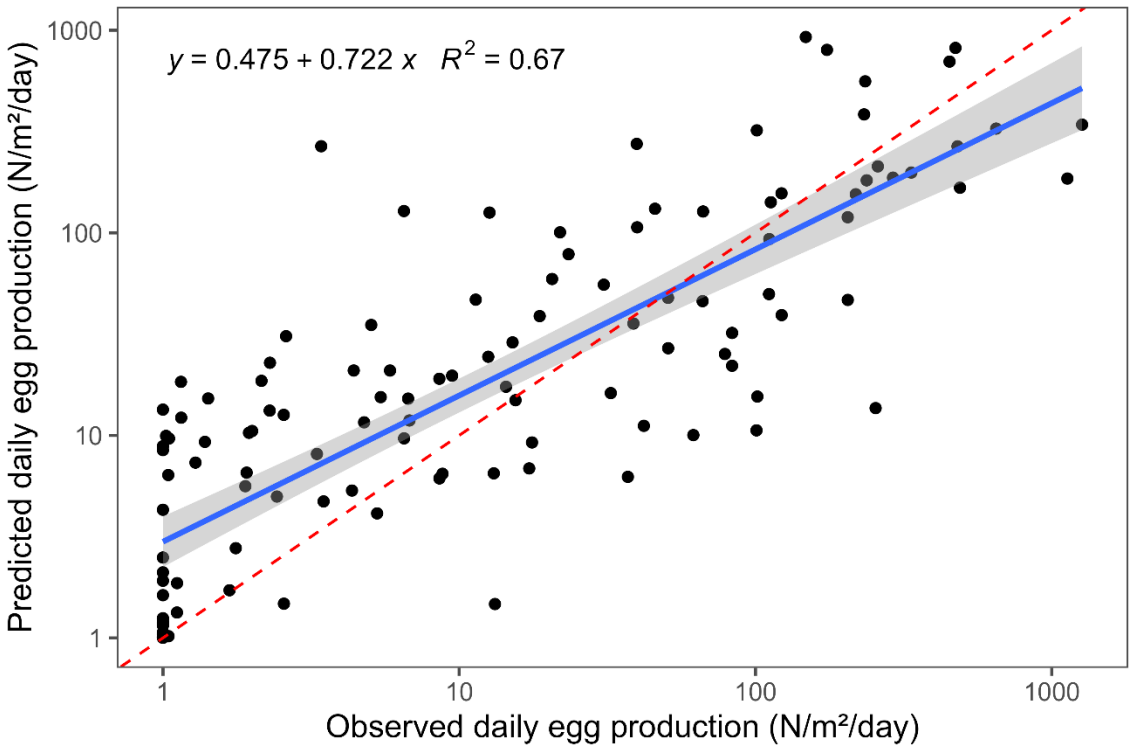

Figure A.2.4. Spatial model validation. For each survey year with at most one missing station, 10 stations were randomly removed before the model was refitted to predict the removed values. The red dashed line represents a perfect fit. The solid blue line represents the actual fit of predictions against observations, with a 95% confidence interval compared to the red dashed equality line (y=x).

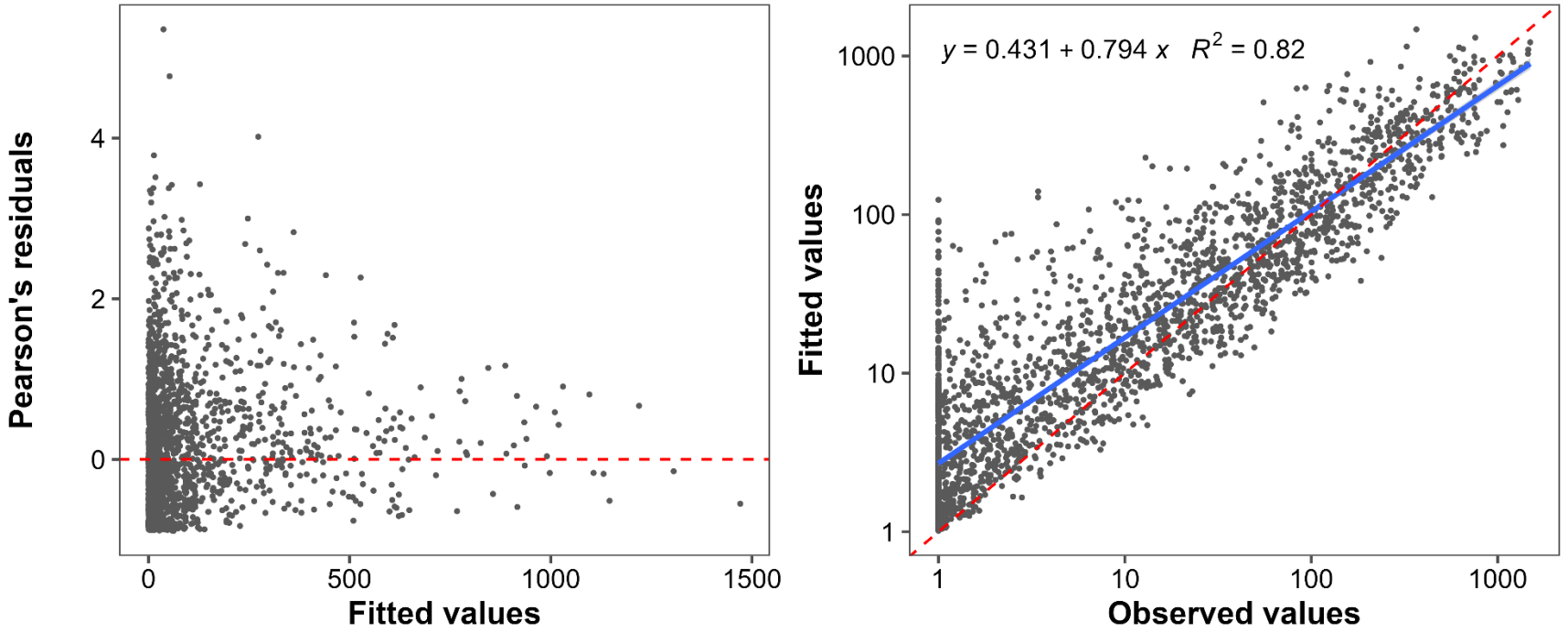

Figure A.2.5. Spatial model validation. Homogeneity of variances is verified using Pearson's residuals in function of predicted daily egg production (left panel). Predictions are validated in function of the observations (right panel). Values are expected to be around the dashed red line. The blue line shows the linear relationship between predictions and observations, with a 95% confidence interval compared to the red dashed equality line (y=x).

APPENDIX 3 - SPAWNING

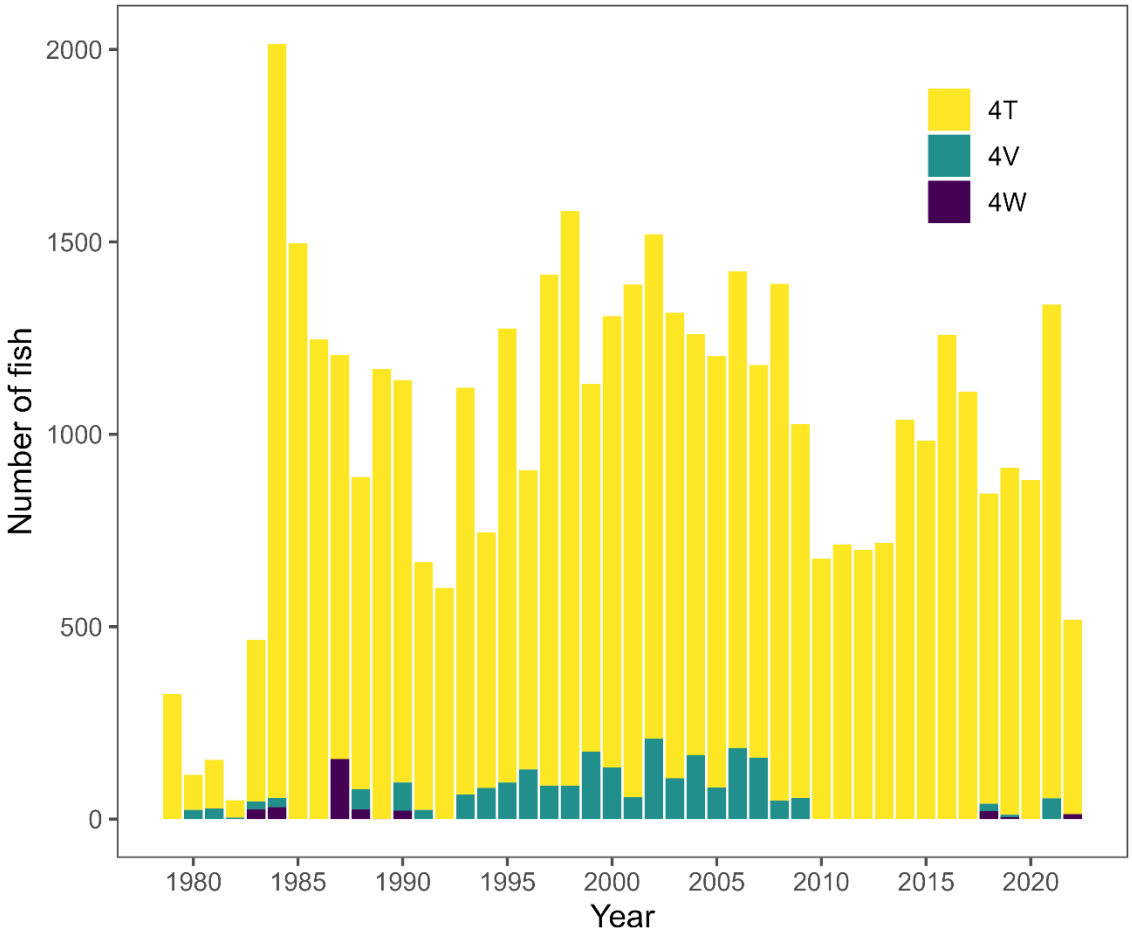

Figure A.3.1. Number of fish available by year and NAFO division used to fit the logistic model describing the progression of the spawning season.

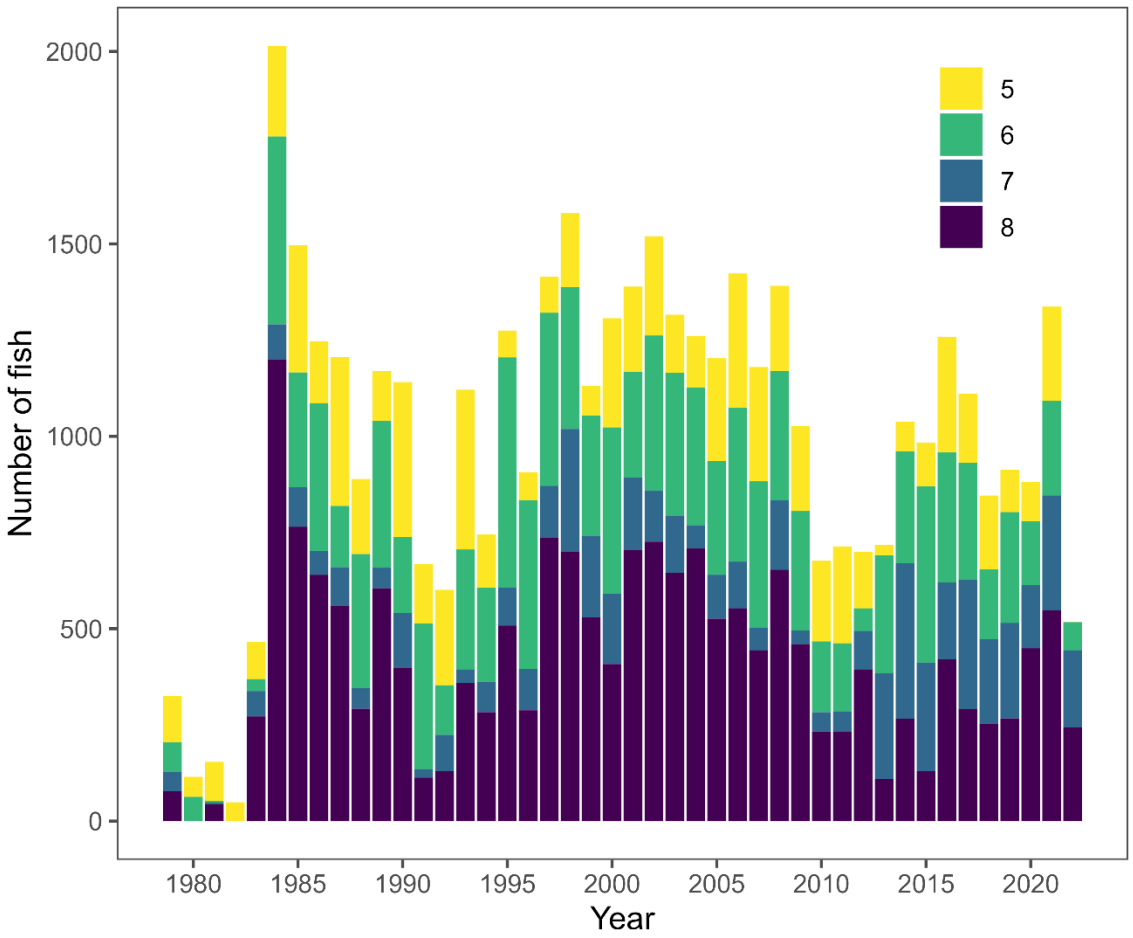

Figure A.3.2. Number of fish available by maturity stage used to fit the logistic model describing the progression of the spawning season (stage 5 = about to spawn, stage 6 = spawning, stage 7 = resting and stage 8 = post-spawning).

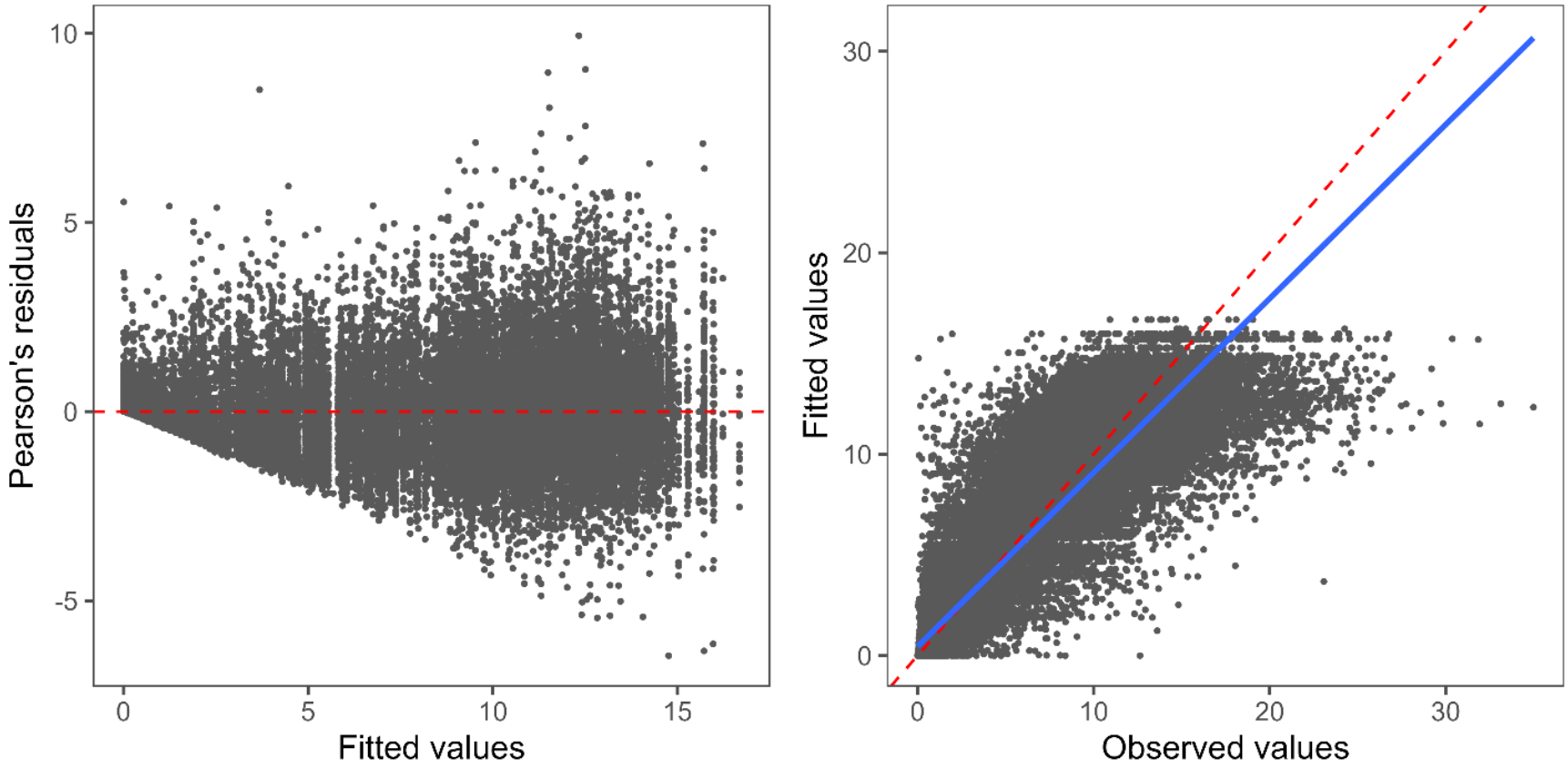

Figure A.3.3. Logistic model validation. Homogeneity of variance is verified using Pearson's residuals in function of predicted daily egg production (left panel). Predictions are validated in function of the observations (right panel). Values are expected to be around the dashed red line. The blue line shows the linear relationship between predictions and observations, with 95% confidence intervals compared to the red dashed equality line (y=x).

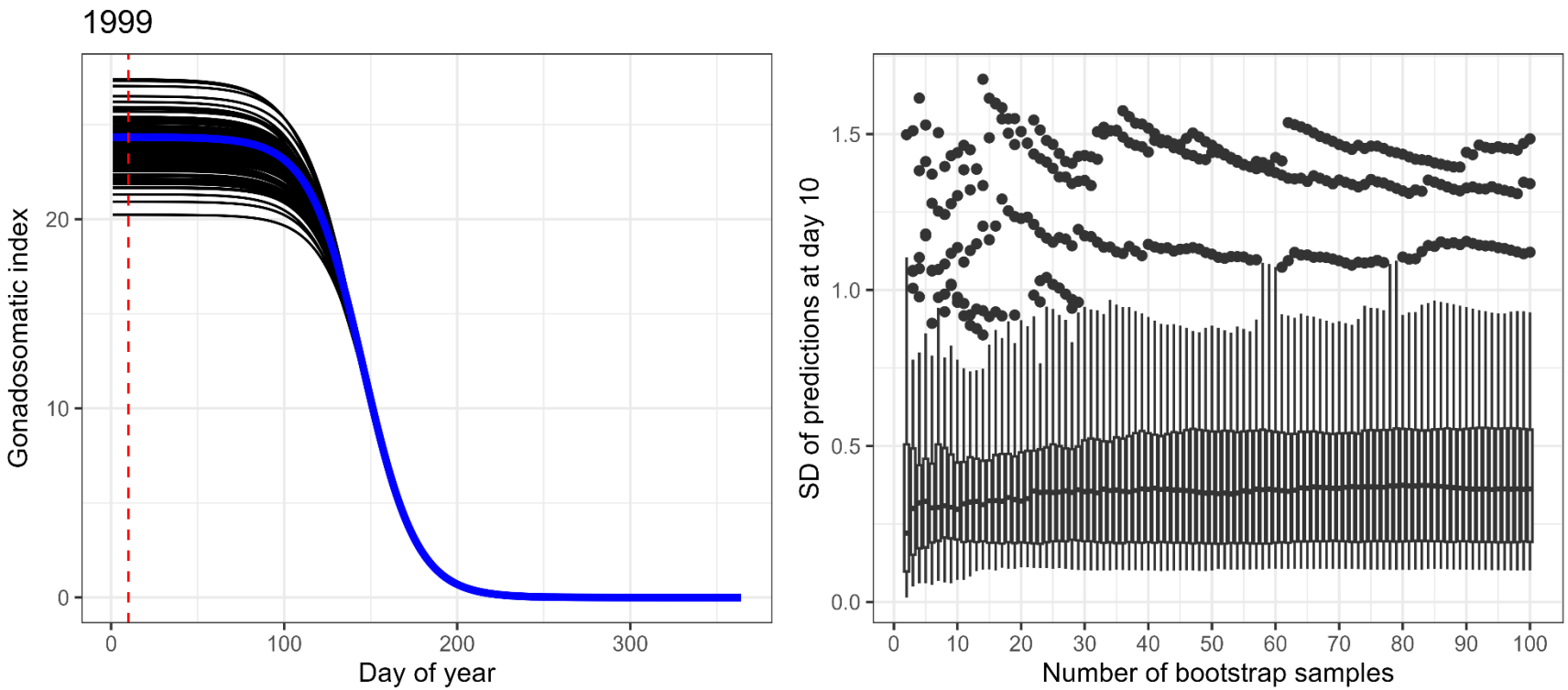

Figure A.3.4. Results of the bootstrapped logistic fit (100 iterations). The left panel shows the logistic model fit using the data (blue) and the 100 fits using bootstrapped samples (black), for 1999 as an example. The red line on the left panel represents the day at which the standard deviation of the predictions is calculated to obtain the right panel. The boxplots in the right panel represent the distribution of standard deviations (all years) in function of the number of bootstraps, and helps identify the number of bootstraps necessary to obtain stable uncertainty estimations.

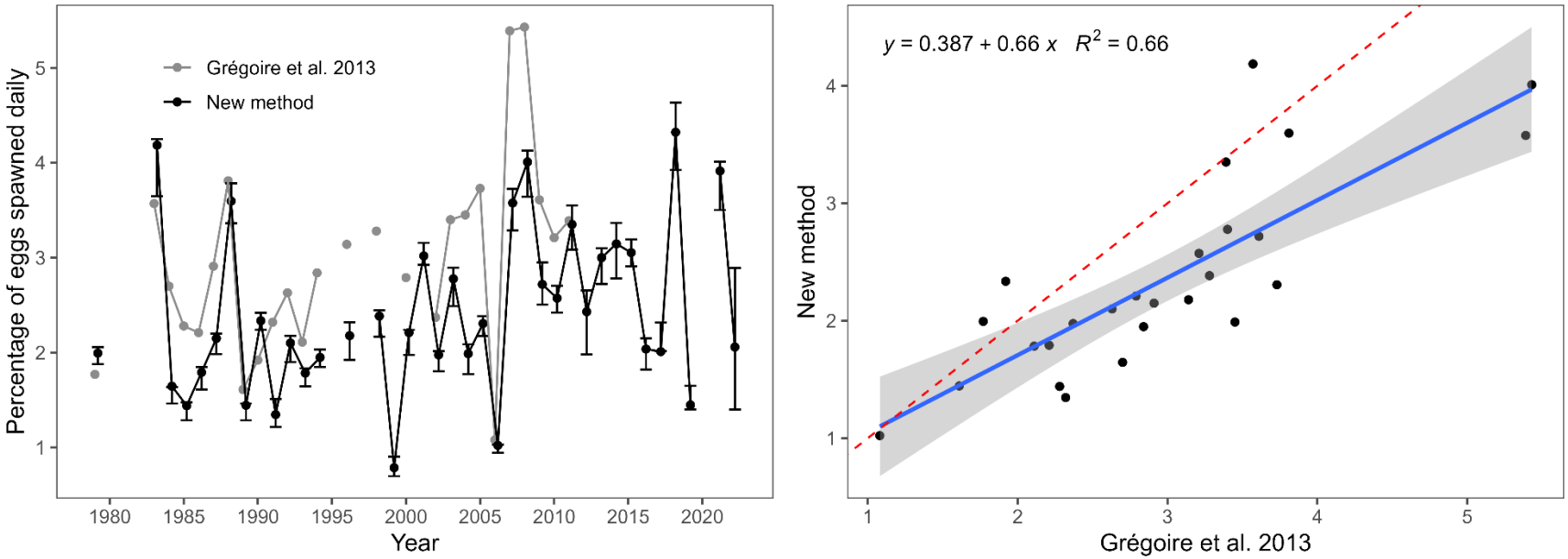

Figure A.3.5. Percentage of eggs spawned at the median date of the survey, estimated with the method presented in this document (including 95% confidence intervals) and the method presented in Grégoire et al. 2013 (left panel). The linear relationship between both time series is presented in the right panel (red = equality line y=x, blue = linear relationship with a 95% confidence interval).

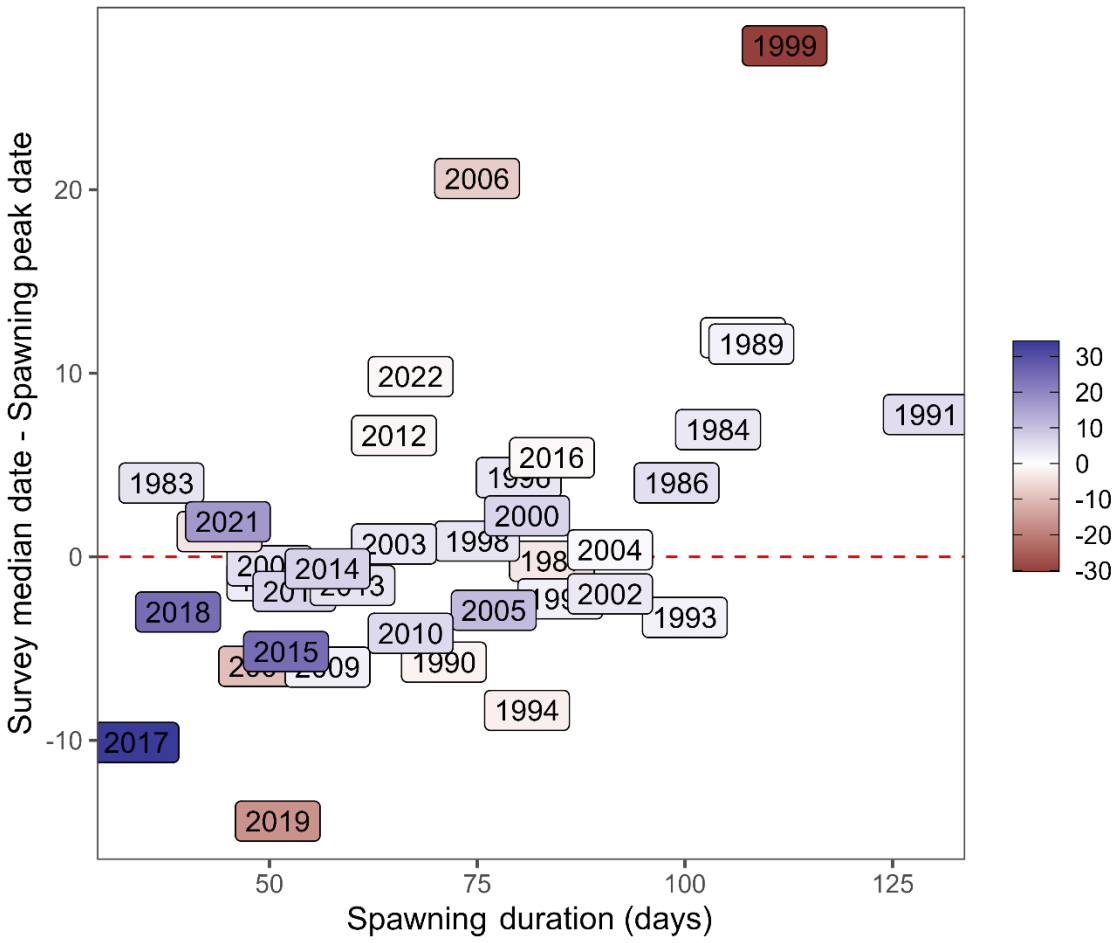

Figure A.3.6. Relationship between the mismatch in survey timing (distance from the peak spawning day) and the duration of spawning. Colours indicate the relative difference in TEP between the baseline and "Station $S_{y,d=x}$" sensitivity run.