

Do harbour porpoise mortality records reflect living population structure? A matrix population model diagnostic

2026-10-13

Michaela Kirstine S. M. H. Søgaard^{1,2}, Magnus Wahlberg^{1,2,3}, and Owen R. Jones^{1*}

¹ Department of Biology, University of Southern Denmark, Odense, Denmark

² Marine Biological Research Centre, University of Southern Denmark, Kerteminde, Denmark

³ Fjord & Bælt, Kerteminde, Denmark

* Corresponding author: jones@biology.sdu.dk

Abstract

Effective conservation of marine mammals depends on reliable demographic information, yet acquiring such data for highly mobile cetaceans is challenging. Harbour porpoises (*Phocoena phocoena*) are widely used as sentinel species, but much of what is known about their demography comes from stranding and bycatch records. While invaluable, these data may be filtered by drift, detection, reporting effort, and age-specific vulnerability.

We developed a matrix population model (MPM)-based diagnostic framework to quantify how observed stage compositions in mortality datasets deviate from asymptotic demographic expectations. Using a stage-structured MPM parameterised from published vital rates, we derived the stable stage distribution (SSD) and compared it with age-class distributions from 10,863 harbour porpoise strandings (of 16,181 total records) in Danish and North Sea waters, bycatch and mixed-mortality records, and hunting captures from Greenland and Denmark. Deviations from SSD, and from the composition expected among deaths, were assessed using goodness-of-fit tests, distance metrics, and tests of the juvenile proportion.

Strandings showed strong and consistent departures from SSD expectations, with juveniles markedly over-represented across spatial and temporal scales. Juveniles remained over-represented relative to the composition expected among deaths under baseline rates, and the departure from the living population was robust to variation in vital rates. Divergence from the asymptotic stage structure was moderate to strong, varied across space, and increased significantly over time in 3 of 6 regions. Bycatch, mixed-mortality, and hunting captures showed comparable SSD divergence, indicating that juvenile-heavy mortality composition is not an artefact of stranding sampling alone.

34 Our results show that harbour porpoise stranding, bycatch, and even hunting data should not be
35 assumed to be demographically representative of living populations. Separating sampling bias
36 from higher juvenile mortality needs independent estimates of juvenile survival.

37

38 **Introduction**

39 The ability to monitor and manage wild populations hinges on reliable demographic information,
40 including knowledge of survival, reproduction, and population structure across age or stage
41 classes. For marine mammals, and cetaceans in particular, obtaining such data presents
42 formidable challenges. These animals are highly mobile, spend most of their time beneath the
43 surface and out of direct view, and are distributed across expansive and often inaccessible marine
44 habitats. Systematic demographic studies typically require dedicated surveys at considerable
45 logistical and financial cost, and even these efforts provide only snapshot views of population
46 status at specific times and locations.

47 Against this backdrop, opportunistic data sources become critically important. Chief among
48 these are stranding and bycatch records: observations of dead or dying animals that wash ashore
49 or are caught in fishing nets, and are then reported, recovered, and examined. Strandings and
50 bycatch records have long been recognised as valuable windows into cetacean biology, providing
51 material for studies of distribution, diet, reproductive status, contaminant burdens, disease, and
52 cause of death (e.g., Kesselring et al. 2017, Kinze et al. 2021, Peltier et al. 2013, Siebert et
53 al. 2020). However, the extent to which stranding and bycatch data can reliably inform our
54 understanding of demographic processes in living populations remains an open question. The
55 fundamental concern is representativeness: do the animals that strand or are bycaught reflect the
56 age structure, sex ratios, and vital rates of the populations from which they come, or are certain
57 demographic groups systematically over- or under-represented?

58 Multiple mechanisms could introduce demographic bias into stranding and bycatch datasets, if
59 their effect differs among age classes, so that population parameters derived from them do not
60 represent the wild population. Carcass drift, driven by ocean currents, wind, and the buoyancy of
61 decomposing bodies, may transport animals far from their sites of death, creating a spatial
62 mismatch between mortality and recovery (Peltier et al. 2013). Detection probability varies with

coastline accessibility, human population density, and dedicated search effort, introducing sampling heterogeneity that is rarely quantified. Reporting effort fluctuates across regions and over time, influenced by public awareness campaigns, institutional capacity, and changing social priorities. Perhaps most importantly, different age classes may exhibit differential vulnerability to mortality and recovery. Young animals, particularly neonates and juveniles, may be more susceptible to environmental stressors, predation, and starvation, and their smaller size and reduced blubber stores could affect buoyancy and drift behaviour. Young animals may also be naive when encountering fishing gear for the first time. Conversely, older adults might be under-represented if they die in offshore waters, where carcasses sink before reaching shore, and prior exposure to fishing gear may reduce their susceptibility to entanglement.

The harbour porpoise (*Phocoena phocoena*) is ideal for investigating these questions. It is the most abundant cetacean in European shelf waters, with high densities in the North Sea and along Danish coasts (Rogan et al. 2017; Scheidat et al. 2008). Porpoises across the North Atlantic, including the North Sea, are largely one population with some isolation by distance and some sub-structure in the southern North Sea, and West Greenland may hold a separate population (Olsen et al. 2022; Celemin et al. 2025; Autenrieth et al. 2024). The Belt Sea population is considered separate (Owen et al. 2024). The abundance of porpoises in the Baltic Sea proper is substantially lower and conservation concern correspondingly higher (Cervin et al. 2020).

Harbour porpoises face a range of anthropogenic pressures, including habitat degradation, incidental capture in fisheries (bycatch), chemical and acoustic pollution, prey depletion, and climate-driven ecosystem changes. Given these multiple stressors, robust demographic monitoring is essential for assessing population trends and informing management decisions.

Harbour porpoise life history is typical of marine mammals: a relatively long lifespan, delayed maturity, low annual reproductive output, and limited capacity for rapid population recovery after decline. Females typically reach sexual maturity between 3 and 5 years of age and produce a

single calf annually or biennially (Møhl-Hansen 1954; Clausen and Andersen 1988; Lockyer 2003; Lockyer and Kinze 2003). First-year mortality is elevated, as in many mammal species, and survival increases with age through the juvenile period before stabilising in adulthood.

Despite the ecological and conservation importance of this species, systematic demographic surveys remain spatially and temporally limited. Much of the available information on population structure, trends, and mortality patterns comes from stranding records. Notably, IJsseldijk et al. (2020) compiled and analysed more than 16,000 harbour porpoise stranding records from the North Sea region spanning nearly three decades (1990–2017). Their comprehensive study documented spatiotemporal patterns in stranding frequency, identified seasonal peaks in neonatal strandings, described variation in sex and age structure across regions, and highlighted biologically meaningful geographic heterogeneity in mortality signals. However, their analytical focus was on describing and modelling stranding patterns rather than on formally testing whether the demographic composition of strandings reflects that of living populations.

Our study addresses this complementary question. Rather than modelling stranding frequency or spatial distribution, we ask whether the observed age-class composition across multiple harbour porpoise mortality datasets (strandings, bycatch records, and hunting captures) is consistent with theoretical expectations derived from population ecology. To do so, we employ matrix population models (MPMs), a well-established framework for describing and projecting the dynamics of stage- or age-structured populations (Caswell 2001). MPMs organise demographic information — survival, growth, and reproduction — into a projection matrix that governs how population size and structure evolve over discrete time intervals. A central property of any ergodic MPM is the stable stage distribution (SSD): the asymptotic proportional representation of each stage class under the assumed vital rates. We therefore use SSD as a mechanistic reference expectation for mortality samples. Under the assumed demography, age-specific mortality is part of the biological process that gives rise to the expected stage structure; the

critical sampling question is whether post-mortem recovery processes such as drift, detection, and reporting are approximately non-selective with respect to age. If recovery is non-selective, the composition of recovered mortality samples would be expected to approach SSD. Departures from SSD therefore indicate selective demographic filtering and/or a mismatch with asymptotic assumptions. Because natural populations can show transient structure under environmental variability and episodic recruitment, SSD is treated here as a reference benchmark rather than as a claim about the realised population state.

We predict, based on documented early-life vulnerability in marine mammals and terrestrial ungulates (Cook et al. 1971; Linnell et al. 1995), that juveniles will be systematically over-represented in mortality records relative to SSD expectations, and that this pattern will be detectable across data types despite uncertainty in model parameterisation. A secondary prediction is that strandings will show greater demographic bias than bycatch, because the processes linking death and recovery are more indirect and spatially displaced in strandings than in direct bycatch observation. Whether bias differs further between bycatch and hunting captures remains an open question. The spatial and temporal structure of bias within the stranding record is less predictable: whether divergence is consistent across regions, or varies in ways that reflect localised differences in mortality pressure, environmental conditions, or reporting effort. Quantifying this heterogeneity and assessing whether it has changed over the nearly three-decade stranding record are additional empirical targets.

Our analytical approach comprises two complementary components. First, we conduct hypothesis tests and compute distributional distance metrics to assess divergence between observed stage compositions across strandings, bycatch, and hunting records and SSD expectations, with particular attention to juvenile over-representation. Second, we implement sensitivity and robustness analyses to evaluate how strongly our conclusions depend on model assumptions, parameterisation choices, and data completeness. We treat published vital rates as

uncertain assumptions, not fixed values, and this includes varying them over wide ranges, propagating uncertainty in SSD estimates through simulation, testing the impact of missing age classes and low sample sizes, and employing alternative distance metrics to ensure inferential stability.

Throughout, ‘demographic bias’ means a departure of the observed stage composition from that of the living population; it does not by itself show whether the cause is sampling bias or simply higher mortality at some stages.

By explicitly quantifying demographic bias and rigorously propagating uncertainty, we aim to provide both a substantive answer for harbour porpoises and a reusable analytical template for systems where representativeness is uncertain. The framework requires only two inputs: stage-composition data from mortality sources (for example, strandings or bycatch) and demographically informed priors on vital rates. Both are increasingly available for species of conservation concern. The result is an explicit, quantitative diagnostic of departure from asymptotic demographic expectations that can inform how mortality records should be weighted when making inferences about living populations.

Methods

Study area and data sources

Our study draws on harbour porpoise mortality data from across the northeastern Atlantic and adjacent seas — including the North Sea, the Belt Sea, and the western North Atlantic off Greenland — with the primary analytical focus on the North Sea, where the stranding data come from; the Belt Sea region is represented by the Möhl-Hansen hunting data and the bycatch data from the south-western Baltic. This study area is a biogeographically complex mosaic of habitats, varying in depth, salinity, primary productivity, and intensity of anthropogenic use, spanning multiple management jurisdictions and hosting genetically differentiated harbour porpoise

populations. We draw on three classes of mortality data: stranding records from coordinated North Sea monitoring networks, bycatch and mixed-mortality records from studies covering Danish, Norwegian, and UK waters, and historical hunting captures from two sources including Funen, Denmark and West Greenland. These sources differ in geographic scope, sampling period, and collection method.

The primary stranding dataset is compiled and described by IJsseldijk et al. (2020), comprising nearly three decades of coordinated monitoring across multiple national stranding networks. The dataset includes records from Denmark, Germany, the Netherlands, Belgium, the United Kingdom, and other North Sea states. For each stranded individual, observers recorded morphometric measurements (total body length, girth, condition), sex, estimated decomposition state, and where possible, age. Age was determined from tooth growth layer counts or, where these were unavailable, assigned from body length using established life-history criteria. We focus on individuals with available body length measurements, as length serves as a reliable and operational proxy for demographic stage in this species.

Bycatch records were drawn from four studies: Brennecke et al. (2021) from the south-western Baltic, Frie & Lindström (2024) from Norwegian waters, IJsseldijk et al. (2022) from the Netherlands, and Murphy et al. (2020) from the Celtic/Irish Sea and North Sea coast of the UK. These vary in sampling period, geographic coverage, and the form in which age or stage information is reported. Importantly, the Murphy North Sea management-unit component partially overlaps with the IJsseldijk et al. (2020) stranding coverage (regions A-C and part of D), so this source is treated as partially non-independent in interpretation. Historical hunting captures come from two sources: Møhl-Hansen (1954), based on drive hunts at Gamborg Fjord, Funen, Denmark between 1941 and 1944, and Lockyer et al. (2001), based on catches from West Greenland between 1988 and 1995. Full details of sample sizes, sampling units, and stage-harmonisation procedures for all sources are given in Table 2.

Stage-structured matrix population model

We developed a female-only, stage-structured matrix population model for harbour porpoise based on a post-breeding annual census. Because the observational data include both sexes, this introduces a minor mismatch; however, if sex ratios in mortality datasets are approximately 1:1, the female-only SSD provides a close approximation to the all-sex stage structure, and departures from equal sex ratios would not alter the directional conclusion of juvenile over-representation. The population was divided into three life-history stages: neonates (age <1 year), juveniles (age ≥ 1 and <4 years), and adults (age ≥ 4 years). These boundaries align with published growth and maturity schedules — derived from biological sampling including bycatch, directed catches, and strandings — for North Sea porpoises (Lockyer 2003; Lockyer and Kinze 2003), and are consistent with stage groupings commonly used in stranding-based demographic analyses (IJsseldijk et al. 2020).

Stage dynamics follow $\mathbf{n}_{t+1} = \mathbf{A}\mathbf{n}_t$, with Lefkovitch projection matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & F_A \\ S_0 & P_J & 0 \\ 0 & G_J & P_A \end{bmatrix}$$

where F_A is female neonate production per adult female per year, S_0 is survival from birth to age 1, P denotes stage stasis (remaining in a stage), and G denotes forward transition to the next stage. Adult natural survival was set to $S_A = 0.867$, the annual survival rate from the Danish life table of Kinze (1990) as reported by Lockyer (2003), and adult stasis was therefore set to $P_A = S_A$.

Neonate survival (birth to age 1) was parameterised using the square rule of Woodley and Read (1991), $S_0 = S_A^2 = 0.752$, to reflect elevated calf vulnerability and potential mortality associated with maternal loss. Woodley and Read (1991), citing Reilly and Barlow (1986), present this formulation as an upper-bound heuristic for first-year survival and note a lower bound near 0.5

based on values typical of pinnipeds and long-lived terrestrial mammals. We adopt this upper-bound value as a conservative baseline — one that, if anything, understates early-life mortality and therefore makes detected juvenile over-representation harder to explain by model parameterisation alone.

Juvenile progression was modelled with a constant maturation probability (g) conditional on survival, such that $G_J = S_J g$ and $P_J = S_J(1 - g)$. Setting $g = 1/3$ yields an expected juvenile-stage duration of three years (ages 1–3), consistent with maturation around age 4 under a Lefkovitch stage-structured formulation (Caswell 2001). With baseline juvenile survival $S_J = 0.81$ (the midpoint of S_0 and S_A , because juvenile survival is expected to lie between first-year and adult survival), this gives $G_J = 0.27$ and $P_J = 0.54$. This implies a geometric distribution of juvenile duration (mean three years); we test a fixed-duration alternative in the Supplement (see *Sensitivity and robustness analyses*).

Fertility was parameterised as $F_A = b p_f$, where b is calves per adult female per year and $p_f = \frac{1}{1+r}$ is the female birth fraction derived from the male:female sex ratio r . Assuming single-calf reproduction, we interpret pregnancy rate as expected calves per adult female per year. Using the foetal sex ratio reported by Lockyer (2003) ($r = 1.1$ – 1.2), which we treat as a proxy for sex ratio at birth, we used the midpoint approximation $p_f = 0.466$, giving baseline fecundity $F_A = 0.90 \times 0.466 = 0.419$.

Under this annual post-breeding formulation the neonate stage lasts exactly one projection interval, so neonate stasis is zero (the top-left element of the matrix) and neonate survival (S_0) is the forward transition, so that all surviving neonates advance to the juvenile stage by the next census.

With these baseline values the model has a population growth rate of $\lambda = 1.03$, that is, a population growing by about 3% a year, which is within the 5–10% maximum reported for porpoises (Lockyer 2003). Because the true growth rate of the populations sampled is not known, we also report results with the model adjusted to $\lambda = 1$ (see *Sensitivity and robustness analyses*).

Stage-specific natural survival beyond the first year is uncertain and varies among regions and sampling sources; we therefore treat juvenile survival (S_J) as an explicit model assumption, constrained between neonate and adult survival, and evaluated through sensitivity analyses. More generally, we treat published vital rates as uncertain assumptions, not fixed values: they differ among populations and may themselves be biased if they come from opportunistic samples, and a central aim of this study is to show how much our conclusions depend on them. We parameterise a baseline demographic model intended to approximate natural demographic processes, recognising that published estimates may still reflect some anthropogenic impacts.

Table 1: Baseline matrix population model parameter values, descriptions, and sources used for SSD estimation.

Parameter	Value	Description	Source / justification
F_A	0.419	Female neonates per adult female per year (matrix element)	Derived as $F_A = b * p_f = 0.90 * 0.466$
P_J	0.540	Juvenile stasis probability (matrix element)	Derived as $P_J = S_J(1 - g)$ with $S_J = 0.81$ and $g = 1/3$
G_J	0.270	Juvenile-to-adult transition probability (matrix element)	Derived as $G_J = S_J g$ with $S_J = 0.81$ and $g = 1/3$
P_A	0.867	Adult stasis probability (matrix element)	Set equal to adult survival ($P_A = S_A$)
S_A	0.867	Adult annual natural survival	Annual survival from the Danish life table of Kinze (1990), as reported by Lockyer (2003)
S_0	0.752	Neonate (0–1 year) survival; also the neonate-to-juvenile transition in the matrix (neonate stasis is zero because the neonate stage lasts one year)	Square rule ($S_0 = S_A \times S_A$); Woodley and Read (1991), citing Reilly and Barlow (1986); qualitative support in Lockyer (2003)
S_J	0.81	Juvenile annual survival	Assumed (midpoint of S_0 and S_A); constrained by $S_0 < S_J \leq S_A$; tested via sensitivity analyses

Parameter	Value	Description	Source / justification
g	1/3	Annual maturation probability	Derived from juvenile duration (ages 1–3); Lockyer (2003); Lockyer and Kinze (2003)
b	0.90	Calves per adult female per year	Baseline value within reported pregnancy-rate range (0.74–0.986); Lockyer (2003)
p_f	0.466	Female birth fraction	$p_f = 1/(1+r)$, with Lockyer (2003) foetal sex ratio $r = 1.1$ – 1.2 (male:female), midpoint approximation

Statistical comparison of observed and expected stage distributions

Unless stated otherwise, ‘expected’ refers to the stable stage distribution (SSD), which describes the living population; the expected composition of deaths is introduced in a later subsection.

To assess whether observed stage distributions in strandings differ from SSD expectations, we employed a multi-faceted statistical approach that combined formal hypothesis tests, distributional distance metrics, and targeted directional tests. Observed stage distributions were computed at multiple aggregation levels: overall (pooling all strandings across space and time), by region, by year, and by region-year combinations. For each grouping, we calculated the proportions of neonates, juveniles, and adults.

Our primary statistical tests and metrics were a two-sided proportion test for the juvenile proportion, and Keyfitz’s Δ .

Proportion test for juveniles: Motivated by a priori expectations of elevated early-life mortality, we tested whether the observed proportion of juveniles differs from the expected proportion. This was implemented as a one-sample, two-sided proportion test of the null hypothesis that $p_{\text{juvenile}} = p_{\text{expected,juvenile}}$. We use a two-sided test for three reasons: the aim is to detect bias in either direction, and a juvenile deficit would be as informative as an excess; the direction relative to the SSD is not fixed in advance, because the model already includes higher early-life mortality; and the test is applied to many sources, for which a pre-specified direction is hard to justify.

Keyfitz's Δ : This metric quantifies the overall magnitude of divergence between observed and expected distributions. Values range from 0 (perfect match) to 1 (complete mismatch). Keyfitz's Δ has an intuitive interpretation: it represents the minimum proportion of individuals that would need to be reassigned to different stages to make the observed distribution match the expected distribution (Keyfitz 1968; Jones 2021). Values above 0.10–0.15 indicate departures of practical significance.

$$\Delta = \frac{1}{2} \sum_i |p_{i,\text{obs}} - p_{i,\text{SSD}}|,$$

where $p_{i,\text{obs}}$ and $p_{i,\text{SSD}}$ are observed and expected stage proportions.

As concordance checks, we also computed chi-square goodness-of-fit statistics and Hellinger distance. To keep the main text focused on Keyfitz's Δ , full definitions, equations, and implementation details for these secondary diagnostics are provided in the Supplementary Information (Supplementary Methods; Table S9).

To quantify and display uncertainty we used two distinct error-bar definitions, depending on the plotted quantity. For stage-proportion plots (e.g., observed Neonate/Juvenile/Adult proportions), we used normal-approximation 95% confidence intervals for a binomial proportion:

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}},$$

with bounds truncated to the feasible range [0, 1]. These intervals quantify sampling uncertainty in the estimated stage proportions for each source/group.

For Keyfitz's Δ summaries, we used non-parametric bootstrap 95% intervals. Within each source/group, we resampled stage counts using multinomial draws of size n with probabilities

equal to the observed stage proportions, recalculated Keyfitz's Δ for each of 2,000 bootstrap replicates, and reported the 2.5th and 97.5th percentiles. These intervals therefore represent uncertainty in divergence from SSD, not uncertainty in SSD itself.

Comparison with the expected composition of deaths

The SSD describes the living population, whereas strandings, bycatch and hunting records are animals that have died. If mortality differs among stages, stages will not be equally represented among the dead and the living. We therefore also calculated the stage composition expected among deaths in a population with a stable structure. For each stage we multiplied the proportion of the population in that stage (the SSD) by the probability that an animal in that stage dies during a projection interval (one minus the stage's survival, i.e. one minus the column sum of the survival and transition part of the matrix), and rescaled the three values to sum to one. In the baseline model, annual mortality is 0.248 for neonates, 0.190 for juveniles, and 0.133 for adults. All comparisons (juvenile proportion test, Keyfitz's Δ , and chi-square tests) were made against both the SSD and this death-based expectation.

Temporal and spatial analyses

The IJsseldijk et al. (2020) data permit temporal and spatial analysis. We conduct these analyses at the regional scale, using the same region coding (A–F) defined in that paper. In brief: Region A covers northeastern Scotland (Thurso to St Fergus), including Orkney; Region B spans from St Fergus to Newcastle; Region C spans from Newcastle to Great Yarmouth; Region D includes the remaining English coast, plus Belgium and the Dutch Delta area; Region E includes the mainland Netherlands and the Dutch Wadden area; and Region F includes the North Sea coasts of Schleswig-Holstein (Germany) and Denmark.

Specifically, to assess whether demographic bias varies across space and time, we computed region-year-specific Keyfitz's Δ values and examined their temporal trends. For each region with

sufficient multi-year data, we fitted simple linear regressions of Keyfitz's Δ on year, testing for significant positive or negative trends in divergence over the study period. We also assessed regional synchrony by calculating pairwise Pearson correlation coefficients among region-specific Keyfitz's Δ time series, to determine whether regions show coordinated temporal patterns in demographic bias.

Bycatch and hunting data

To directly compare mortality datasets within a common framework, we repeated the SSD-divergence workflow for each additional mortality source, standardising all records into the same three stage classes (Neonate/Juvenile/Adult). Where sources reported stage labels directly, we harmonised terminology (Calf→Neonate, Immature→Juvenile, Mature→Adult). Where sources reported individual ages or age bins, we mapped to stages using explicit thresholds: age < 1 year → Neonate; age 1 to < 4 years → Juvenile; age $\geq$ 4 years → Adult. We also harmonised sampling strata where needed, retaining source-specific area/period groupings such as the Murphy et al. (2020) area-by-period strata. For IJsseldijk et al. (2022), stage composition was already provided as neonate/juvenile/adult percentages and was used directly for the bycatch-only subset (n = 103). For Murphy et al. (2020), individual-level data were not available, so we used age-frequency counts (ages 0–23 years) that we read from the bars of their Figure 2 for each management unit and period (n = 645 aged animals, sexes combined). The total and the number of animals aged under one year matched those given in the paper, so reading error is likely to be very small. We could not separate stranded from bycaught animals. For Lockyer et al. (2001), we used the age-structured counts reported in Table 1 of that paper. For Møhl-Hansen (1954), 389 measured animals were available for stage-structure analysis.

We then repeated the main analyses against the same SSD reference used for strandings: juvenile proportion tests and Keyfitz's Δ , with secondary chi-square checks reported in the Supplementary Information. For source-year units with low expected counts, Monte Carlo chi-

square p-values were used in those supplementary diagnostics. To assess comparative divergence, we compared year-level Keyfitz's Δ values between bycatch and strandings using a paired one-sided Wilcoxon test (alternative: bycatch < strandings), restricted to years with minimum sample sizes in both datasets.

Table 2: Mortality data sources used for stage-structure comparison against SSD. This shows the provenance, sampling unit, sample size, and stage-harmonisation approach for each source included in the comparative analyses.

Data Type	Source	Site / Stratum	Years	Unit Type	n (used)	Stage Mapping
Strandings	IJsseldijk et al. (2020)	North Sea regions A-F	1990–2017	Individual-level records	10,863	Source-defined length-based Neonate/Juvenile/Adult classes
Bycatch	Brennecke et al. (2021)	South-western Baltic Sea	Oct 1987–Nov 2016 (all months represented)	Individual-level records	136	Stage labels harmonised to Neonate/Juvenile/Adult
Bycatch	Frie & Lindström (2024)	Norway	Autumn 2016; Spring 2017	Individual-level records	134	Age-to-stage mapping (<1, 1 to <4, ≥4 years)
Bycatch	IJsseldijk et al. (2022)	Netherlands	2008–2019 (all months)	Aggregate percentages (bycatch-only subset)	103	Reported Neonate/Juvenile/Adult proportions
Mixed (bycatch + strandings)	Murphy et al. (2020)	Celtic/Irish Sea (1990–1999)	1990–1999	Counts read from published Figure 2 (mixed bycatch + strandings)	217	Counts read from published Figure 2 (n=645), age-bin lower bounds mapped to stages
Mixed (bycatch + strandings)	Murphy et al. (2020)	Celtic/Irish Sea (2000–2012)	2000–2012	Counts read from published Figure 2 (mixed)	172	Counts read from published Figure 2 (n=645), age-bin lower bounds mapped to stages

Data Type	Source	Site / Stratum	Years	Unit Type	n (used)	Stage Mapping
				bycatch + strandings)		
Mixed (bycatch + strandings)	Murphy et al. (2020)	North Sea (1990-1999)	1990-1999	Counts read from published Figure 2 (mixed bycatch + strandings)	155	Counts read from published Figure 2 (n=645), age-bin lower bounds mapped to stages
Mixed (bycatch + strandings)	Murphy et al. (2020)	North Sea (2000-2012)	2000-2012	Counts read from published Figure 2 (mixed bycatch + strandings)	101	Counts read from published Figure 2 (n=645), age-bin lower bounds mapped to stages
Hunting	Lockyer et al. (2001)	Maniitsoq and northerly sites	Jul–Sep 1988–1989; 1995	Aggregate counts	126	Age-year counts mapped to stages
Hunting	Lockyer et al. (2001)	Nuuk and Paamiut	Jul–Sep 1988–1989; 1995	Aggregate counts	52	Age-year counts mapped to stages
Hunting	Møhl-Hansen (1954)	Denmark (hunting captures)	Nov–Feb 1941–1944	Aggregate counts	389	Length thresholds (<91, 91–130, >130 cm)

Sensitivity and robustness analyses

To evaluate the robustness of our conclusions, we conducted the following sensitivity analyses.

Exclusion of incomplete age-class records: We repeated all analyses after excluding region-years in which one or more stage classes were entirely absent. This tests whether conclusions are sensitive to the way missing data are handled.

Minimum count thresholds: We imposed minimum sample size requirements (e.g., ≥ 10 or ≥ 20 total strandings per group) and re-evaluated key findings to ensure they are not artefacts of small-sample instability.

SSD uncertainty propagation: Because our baseline SSD is derived from a single MPM parameterisation, we propagated uncertainty in matrix elements via simulation. Using `compute_ci` function from the `mpmsim` R package (Jones et al. 2025), we generated perturbed U and F matrices around the baseline model, recalculated SSD for each simulated matrix, and derived confidence bounds for SSD proportions. We then re-evaluated juvenile over-representation and Keyfitz’s Δ summaries against these bounds. Supplementary Figure S1 and Supplementary Table S1 show the full uncertainty spread around SSD expectations, clarifying why the main conclusions remain robust to plausible vital-rate variation.

Sensitivity to vital rates: Published vital rates are uncertain, so we tested how much the expected composition depends on them. We drew 20,000 random sets of vital rates (Latin hypercube sampling), each value independently from a plausible range: adult survival 0.80–0.99, juvenile survival 0.60–0.99, neonate survival 0.50–0.99, pregnancy rate 0.60–1.00, and the annual probability that a juvenile matures 0.25–0.50. For each set we calculated the SSD and the death-based expectation and compared them with the observed stranding composition, for all sets and for sets with λ between 0.95 and 1.05 or between 0.98 and 1.02. We also changed each rate on its own with the others at baseline (Supplementary Tables S12 and S13).

Population growth rate: We repeated the comparisons with the model adjusted to $\lambda = 1$, by lowering either fertility (to 0.65 calves per adult female per year) or adult survival (to 0.83), with other rates at baseline (Supplementary Table S11).

Working backwards from the data: To see what the data would imply if they were an unbiased sample of deaths, we fitted adult, neonate and juvenile survival to each observed composition (maximum likelihood, with λ between 0.97 and 1.03). Three rates are fitted to two independent proportions, so we use this only to show how far survival would have to differ from published values, not as an estimate of mortality (Supplementary Table S14).

Fixed juvenile duration: A constant yearly maturation probability implies a wide spread of juvenile durations. To check whether this matters, we built an age-by-age (Lefkovitch) model in which juveniles (ages 1–3) move up exactly one age per year and adults start at age 4, with survival and fertility as in the baseline model, and repeated it with adulthood starting at age 3 or 5 (Supplementary Tables S15–S17).

Simulation-based plausibility checks

To address uncertainty in the baseline parameterisation without assuming a single fixed set of vital rates, we asked whether any biologically plausible MPM, regardless of specific parameter values, produces an SSD resembling the observed stranding composition. To address this, we used the **mpmsim** R package (Jones et al. 2025) to generate 500 random 3-stage Lefkovitch matrices (archetype 4: monotonic forward stage progression without retrogression), constrained only to plausible population growth rates (λ between 0.7 and 1.3). For each simulated matrix, we computed the SSD and Keyfitz’s Δ relative to the observed overall stranding composition, then summarised the proportion of matrices meeting resemblance thresholds (Keyfitz’s $\Delta \leq 0.15$, and more conservatively Keyfitz’s $\Delta \leq 0.10$). In these matrices all survival, transition and fecundity values varied freely; the only constraint was on λ . Full simulation settings are provided in the Supplementary Methods; Supplementary Figure S3 provides an additional PCA projection of this simulation space.

All analyses were performed in R version 4.3.0, using the **popbio** (Stubben and Milligan 2007), **mpmsim** (Jones 2025), **tidyverse** (Wickham et al. 2019), and **ggplot2** (Wickham 2016) packages. Code and derived data are available in the project repository (see *Data and Code Availability*).

Results

We present results in two stages. The stranding dataset, being the largest and most spatially and temporally resolved, forms the primary focus; bycatch and hunting records are then analysed within the same framework to assess whether patterns of demographic bias are specific to strandings or general across mortality data types.

Overall stage distribution and divergence from SSD

The SSD describes the living population, so this first comparison asks how the animals that died differ from the living population; the comparison with the expected composition of deaths follows below.

Among strandings with assigned age class ($n = 10,863$ of 16,181 total records), the observed stage distribution differed markedly from the SSD predicted by our baseline MPM. The overall observed composition was 13.2% neonates, 58.1% juveniles, and 28.7% adults, compared with SSD expectations of 20.1%, 30.5%, and 49.4% — a juvenile elevation of 27.6 percentage points and an adult deficit of 20.7 percentage points (Figure 1A). The two-sided juvenile proportion test was highly significant ($p < 0.001$), and overall divergence was substantial (Keyfitz's Δ , the share of animals that would have to change stage for the observed and expected compositions to match, $= 0.276$, bootstrap 95% CI 0.266-0.284), consistent with our prediction of systematic juvenile over-representation.

Simulation-based plausibility of alternative MPMs

To assess whether the observed mortality composition could plausibly arise under any reasonable demographic parameterisation, we generated 500 random biologically plausible 3-stage Lefkovitch matrices. Only 5 (1.0%) produced SSDs within Keyfitz's $\Delta \leq 0.15$ of the observed stranding composition, and only 2 (0.4%) met the stricter Keyfitz's $\Delta \leq 0.10$ threshold. Most plausible matrices were therefore substantially farther from the observed composition

(median Keyfitz's $\Delta = 0.557$, range 0.075-0.832), indicating that the observed juvenile-heavy mortality composition is inconsistent with demographic equilibrium across the broader space of plausible vital-rate combinations, not merely under our specific baseline parameterisation (Supplementary Figure S3).

Alternative metrics and tests (Hellinger distance and chi-square diagnostics) were in broad agreement with Keyfitz's Δ and are reported in Supplementary Tables S2-S3 and S9. Descriptive context for region-year sampling structure and stage-composition trajectories through time is provided in Supplementary Table S4 and Supplementary Figure S5. The full region-year distribution of Keyfitz's Δ is shown in Supplementary Figure S2 and summarised in Supplementary Table S2.

Using the same stage framework and SSD reference, bycatch showed the same directional pattern (juvenile elevation and adult deficit relative to SSD). In paired year-level comparisons, mean bycatch Keyfitz's Δ (0.331) exceeded mean strandings Keyfitz's Δ (0.199), and the one-sided test for lower bycatch divergence was not supported ($p = 0.979$; Figures 2 and 3).

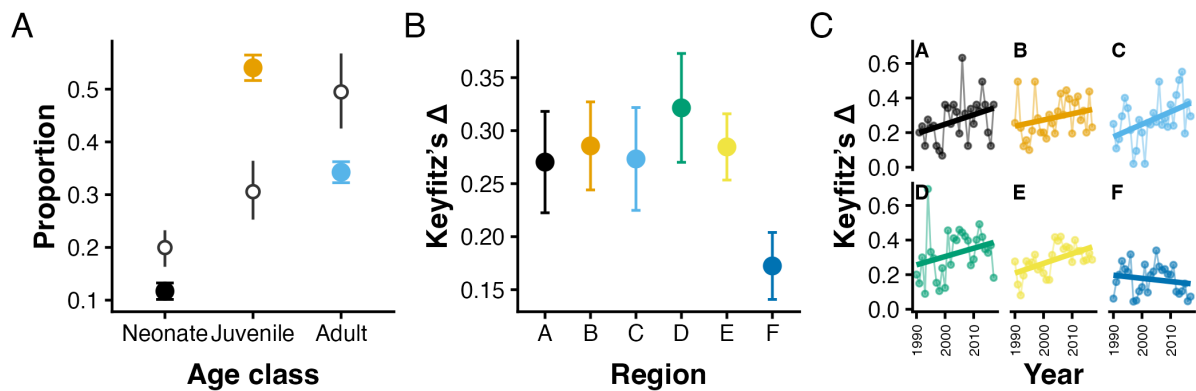


Figure 1: Stage composition in strandings differs systematically from SSD expectations, with divergence varying across regions and increasing over time in some areas. (A) Mean observed stage proportions across region-years (Neonate, Juvenile, Adult) are shown as coloured points with 95% confidence intervals; SSD expected proportions and uncertainty intervals are overlaid for direct comparison. (B) Mean Keyfitz's Δ by strandings

region (A-F), with 95% confidence intervals, summarises persistent regional differences in divergence from SSD.

(C) Annual Keyfitz's Δ trajectories by region (facets A-F) with fitted linear trends show how SSD divergence

changes through time within each region. References to sources of data are found in the text.

Comparison with the expected composition of deaths

Among living animals, the baseline model expects 20.1% neonates, 30.5% juveniles and 49.4% adults; among deaths it expects 28.7%, 33.4% and 37.9%, because neonates and juveniles die at higher rates than adults. Strandings still contained far more juveniles than either expectation (Supplementary Table S10): 58.1%, against 30.5% for the living population and 33.4% for deaths. Keyfitz's Δ fell from 0.28 (SSD) to 0.25 (deaths), and the juvenile proportion differed significantly from both (two-sided $p < 0.001$), as it did in every region.

The two large bycatch samples (Brennecke et al. 2021; IJsseldijk et al. 2022) were 73–74% juveniles and also differed from both references, and the Danish hunting sample (Møhl-Hansen 1954) had a significant juvenile deficit. Most of the remaining bycatch, hunting and mixed-mortality samples were consistent with the death expectation; the exception was the Murphy et al. (2020) Celtic and Irish Seas 2000–2012 sample, which still had more juveniles than expected (Supplementary Table S10). Two-sided tests removed two significant results from the previous one-sided analysis, both in smaller samples (Greenland, Maniitsoq, against the SSD; Murphy et al. North Sea 1990–1999 against the death expectation).

A stable population does not remove the pattern

In the baseline model, the population grows at about 3% a year. Lowering fertility or adult survival so that the population is stable ($\lambda = 1$) did not reduce the mismatch with strandings: Keyfitz's Δ was 0.31 and 0.27 against the SSD and 0.27 and 0.30 against the death expectation, respectively (baseline: 0.28 and 0.25; Supplementary Table S11). A stable population needs fertility below the published pregnancy-rate range (0.65, against 0.74–0.99) or adult survival of

0.83, within the range of published estimates (0.80–0.87), so our baseline probably overstates growth slightly.

Sensitivity to vital rates

Across the 20,000 random sets of vital rates, the expected juvenile share of the living population ranged from 14% to 44% and never reached the 58.1% seen in strandings (Supplementary Table S12), including in sets with λ between 0.95 and 1.05 or between 0.98 and 1.02. The death-based expectation was more variable (1–92% juveniles), and 3–8% of sets reached the observed juvenile share, generally those with low juvenile survival (median about 0.68). Changing one rate at a time (Supplementary Table S13), the juvenile share of the living population stayed within 25–35%, whereas that of deaths ranged from 6% to 76%.

If strandings were an unbiased sample of deaths, the best-fitting model would have juvenile survival of 0.62 and adult survival of 0.89 (Supplementary Table S14), the juvenile value well below the one we assumed (0.81). We read this as a sign that strandings are not an unbiased sample of deaths, not as an estimate of true survival, because the fit is not unique.

Fixing the length of the juvenile period

In the age-by-age model with juveniles aged 1–3 years, the expected juvenile share was 34.2% for the living population and 37.2% for deaths (three-stage model: 30.5% and 33.4%; Supplementary Tables S15 and S16), and strandings still differed strongly from both (Keyfitz's $\Delta = 0.24$ and 0.21; Supplementary Table S17). The expected juvenile share depended more on the length of the juvenile period than on whether it was fixed: it was 27% (living) and 29% (deaths) if adults start at age 3, and 40% and 43% if they start at age 5. Strandings differed significantly from expectations in all cases.

Geographic variation in demographic bias

Regional analyses revealed pronounced heterogeneity in divergence from SSD, with stable between-region differences persisting over time (Figure 1B & C). Using the region coding of IJsseldijk et al. (2020), mean region-year divergence was highest in Region D (0.321) and lowest in Region F (0.172), with the remaining regions falling within an intermediate range. Although between-region differences are statistically clear, the absolute spread in regional mean Keyfitz's Δ is moderate (0.149 from lowest to highest mean), and should be interpreted alongside substantial within-region year-to-year variation (median region-specific SD = 0.119).

Sensitivity to missing age classes and low counts

Applying stricter data filters did not materially alter the juvenile-over-representation result. Across scenarios (all data, complete age classes only, minimum $n = 10$, and both filters), the observed juvenile proportion remained at 0.581 (Supplementary Table S5), while the expected SSD juvenile proportion remained 0.305 and two-sided tests remained highly significant (all $p < 0.001$). Likewise, the magnitude of divergence was stable under filtering. Mean region-year Keyfitz's Δ varied only from 0.256 to 0.268, and median Keyfitz's Δ from 0.256 to 0.258 across scenarios (Supplementary Table S6), indicating that the central inference is not driven by missing classes or small- n groups.

Sensitivity to SSD uncertainty

When SSD uncertainty bounds were propagated into hypothesis testing, the expected juvenile proportion ranged from 0.300 to 0.313 across SSD scenarios (Supplementary Table S7). The observed juvenile proportion remained at 0.581 and continued to exceed expectation under all bounds (all $p < 0.001$). Distance-based results were similarly robust. Mean Keyfitz's Δ across region-years remained high, ranging from 0.262 to 0.272, with medians from 0.252 to 0.262

(Supplementary Table S8), confirming substantial mismatch even under conservative SSD assumptions.

Temporal trends in demographic bias

Keyfitz's Δ varied over time (Table 3); only Regions C and E showed significant positive slopes in divergence from the SSD (C: 0.007; E: 0.006 Keyfitz's Δ units per year; both $p < 0.05$). Region B was positive but non-significant (0.004, $p = 0.179$), and Region F was the exception with a near-zero negative, non-significant slope (-0.002, $p = 0.370$; Figure 1C). In practical terms, an increase of ~ 0.007 Keyfitz's Δ units per year implies an increase of about 0.189 over the 27-year study period, relative to the baseline mean region-year Keyfitz's Δ of 0.268.

Table 3: Per-region temporal trends in SSD divergence (Keyfitz's $\Delta \sim \text{Year}$). This shows which regions have increasing, decreasing, or stable divergence over time.

Region	slope	SE (slope)	p-value	R-squared	n
A	0.006	0.003	0.078	0.119	27
B	0.004	0.003	0.179	0.068	28
C	0.007	0.003	0.015	0.207	28
D	0.005	0.003	0.141	0.081	28
E	0.006	0.002	0.003	0.299	28
F	-0.002	0.002	0.370	0.031	28

Regional synchrony in demographic bias

Pairwise correlations among region-specific Keyfitz's Δ time series were generally modest. The strongest positive correlations were C-A ($r = 0.464$) and E-D ($r = 0.452$). The full regional correlation matrix is provided in Supplementary Figure S4.

Comparative strandings, bycatch and hunting results

Using the standardised bycatch dataset (four harmonised sources mapped to the shared Neonate/Juvenile/Adult framework), juvenile over-representation remained pronounced

(observed juvenile proportion = 0.526, SSD expectation = 0.305, two-sided $p < 0.001$), and overall Keyfitz's Δ was 0.221. In paired year-level comparisons, mean bycatch Keyfitz's Δ (0.331) exceeded mean strandings Keyfitz's Δ (0.199), and the one-sided test for lower bycatch divergence was not supported ($p = 0.979$; Figures 2 and 3). Hunting captures from Möhl-Hansen (1954) and Lockyer et al. (2001), converted to the same stage structure, occupied the same divergence range as the other mortality sources. Across all chi-square goodness-of-fit tests, expected counts were high (minimum expected = 91.4), with no cells below 5, so sparse-cell corrections were not required; full chi-square results are provided in Table S9.

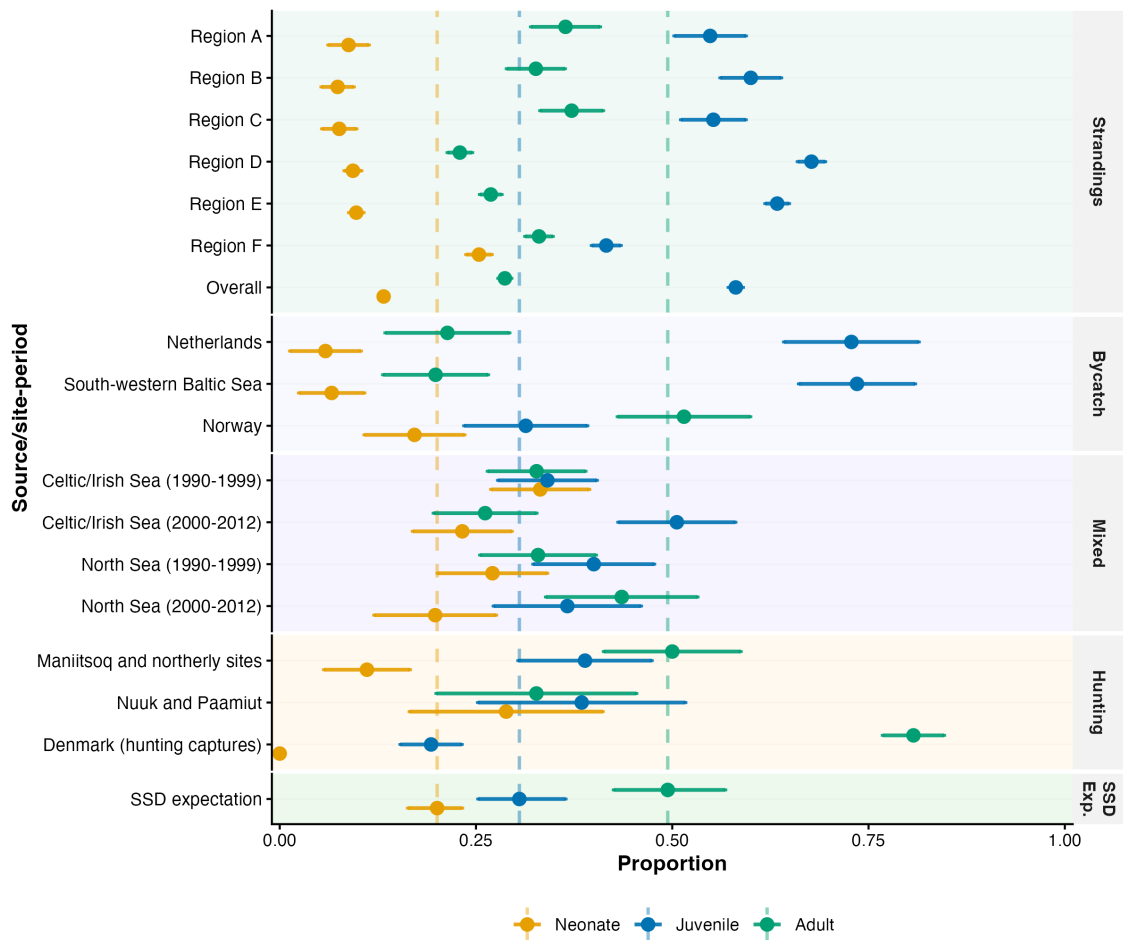


Figure 2: Strandings, bycatch, mixed-mortality, and hunting age-class proportions across sources, with SSD reference panel. Points are observed proportions by age class (Neonate/Juvenile/Adult) and horizontal error bars are 95% intervals for each source/site-period. Rows are grouped into strandings, bycatch, mixed (bycatch +

strandings), hunting, and SSD panels with light background shading to separate data types; the SSD panel shows expected proportions with uncertainty intervals from the SSD simulation. References for sources of data are found in the text.

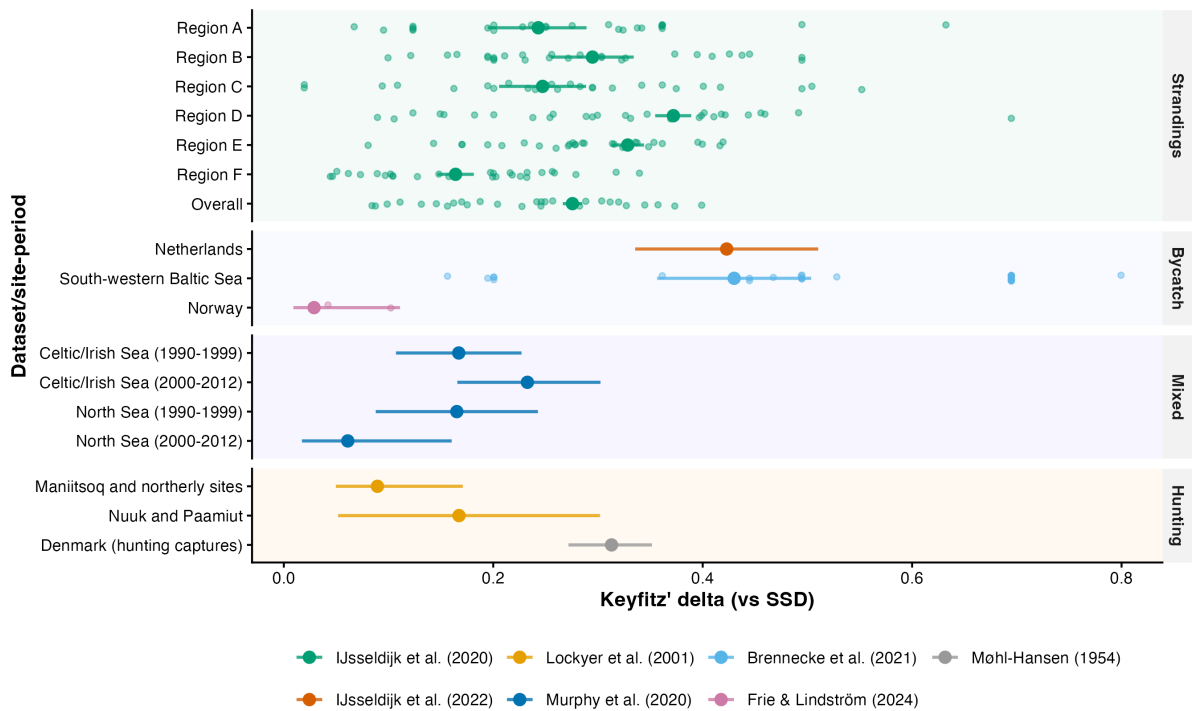


Figure 3: Combined SSD-divergence comparison across strandings, bycatch and hunting sources. Points show Keyfitz's Δ estimates, and horizontal error bars show bootstrap 95% intervals. Faint jittered points indicate repeated year-level values where available. Rows are grouped by data type (strandings, bycatch, mixed bycatch + strandings, hunting) with light background shading, and points are colour-coded by source article (IJsseldijk et al. (2020, 2022), Lockyer et al. (2001), Murphy et al. (2020), Brennecke et al. (2021), Frie & Lindström (2024), Möhl-Hansen (1954)). Site and period labels are shown where relevant.

In summary, strandings and bycatch both differed significantly from SSD expectations, with juveniles over-represented in both datasets. Bycatch showed divergence at least as great as strandings. The magnitude of bias varied considerably across stranding regions and over time.

Discussion

Our results provide clear and robust evidence that harbour porpoise mortality datasets show a systematic departure from asymptotic demographic expectations, as reflected by the SSD predicted by a stage-structured matrix population model. Strandings showed strong SSD divergence, and bycatch data were not less biased away from SSD: in paired years, mean bycatch Keyfitz's Δ (0.331) exceeded mean strandings Keyfitz's Δ (0.199), with no support for lower bycatch divergence (one-sided $p = 0.979$). Both data types should therefore not be treated as demographic cross-sections of living populations.

The most striking and consistent signal is pronounced juvenile over-representation relative to SSD expectations. In the baseline comparison, the observed juvenile proportion exceeded the SSD expectation by 27.6 percentage points, while adult proportion was 20.7 percentage points below SSD expectation. This pattern is evident in strandings across all spatial scales and in sensitivity analyses, and it is also evident in bycatch (0.526 observed vs 0.305 expected). Together, these results suggest that a juvenile-heavy composition is not confined to a single sampling method and should be accounted for in any demographic inference drawn from mortality records.

Why might juveniles be over-represented? Several mechanisms, possibly acting in concert, could explain this pattern. First, juvenile harbour porpoises may experience elevated mortality beyond the baseline MPM parameterisation. Early-life vulnerability is well documented in marine mammals and other taxa (Cook et al. 1971; Linnell et al. 1995), driven by factors such as immunological immaturity, energetic stress during and shortly after weaning, inexperience in foraging or predator avoidance, and higher susceptibility to disease, contaminants, or fisheries interactions. If juvenile mortality is indeed higher than adult mortality, particularly if it is substantially higher, then juveniles will be over-represented in any mortality sample.

Second, differential detection and recovery probabilities could contribute to juvenile over-representation. Smaller-bodied juveniles might drift differently from larger adults, influenced by buoyancy dynamics, wind, and ocean currents, making them more likely to wash ashore in accessible locations. Alternatively, juveniles might be more visible or more readily reported by beachgoers, though we consider this less likely given that body size differences between juveniles and adults are not extreme. A third possibility is that our baseline model, an asymptotic reference derived from assumed vital rates, is misspecified. A higher juvenile share in the living population would need lower juvenile mortality or higher adult mortality than we assumed (or higher fertility), but across the ranges we tested the juvenile share of the living population stayed between 25% and 35%, well below the 58% seen in strandings (Supplementary Table S13). For the composition of deaths the effect goes the other way: juveniles would form a larger share of deaths if juvenile mortality were higher, or adult mortality lower, than we assume. Because the composition of deaths is more sensitive to these rates, we compare the data with both the living population and the expected composition of deaths. Our sensitivity analyses with varied vital rates (Supplementary Tables S11–S14) and with simulated MPMs suggest that the qualitative conclusion that juveniles are over-represented is robust across a wide range of plausible demographic parameter space, although with the death-based expectation a minority of parameter sets (3–8%) can reproduce the observed juvenile share, mainly with low juvenile survival. Matching the strandings would need juvenile survival of about 0.62 (Supplementary Table S14), well below the value we assumed, so independent estimates of juvenile survival are the key missing information for deciding whether the juvenile excess reflects sampling bias or higher juvenile mortality. If published survival rates are themselves derived from opportunistic samples, they may carry the same bias, and independent estimates of age-specific mortality are needed.

The magnitude of departure from asymptotic stage structure, as quantified primarily by Keyfitz's Δ (with concordant alternative metrics; Supplementary Tables S3 and S9), varied considerably

across regions and over time. Sustained increases in Regions C and E over the 27-year study period are consistent with changing fisheries pressure or habitat change altering age-specific mortality exposure, rather than reflecting static sampling bias alone. This heterogeneity suggests that demographic bias is driven by regional differences in mortality processes, population structure, and environmental conditions rather than a single uniform mechanism. The modest regional synchrony in divergence through time (Supplementary Figure S4) is also more consistent with region-specific drivers than with a single basin-wide forcing acting uniformly across all regions. Regions with particularly high divergence may be experiencing elevated juvenile mortality due to localised stressors (e.g., intensive fisheries activity, habitat degradation, prey depletion), or they may differ in detection and reporting dynamics that selectively filter age classes. Notably, the strongest divergence is concentrated in Regions D–E (English/Belgian/Dutch sectors) rather than Region F, which is nearest the Baltic transition zone, suggesting that spatial variation in mortality composition reflects regional differences in mortality processes, carcass transport, and detection effort rather than a simple Baltic-proximity gradient.

The bycatch result is notable because it challenges our initial expectation that bycatch would be less demographically biased than strandings. Instead, bycatch was at least as divergent from SSD and, in paired years, often more so. Several non-exclusive mechanisms could explain this. First, bycatch risk is likely gear-selective: certain net types, mesh sizes, soak times, and deployment depths may disproportionately intercept younger/smaller porpoises (Kindt-Larsen et al. 2023). Second, habitat overlap can raise bycatch risk independently of population stage structure: porpoise foraging areas and coastal gillnet effort can co-occur in prey-rich shelf habitats, creating local exposure hotspots (Maeda et al. 2021; Brennecke et al. 2021). Third, the bycatch dataset used here pools records across multiple regions, periods, and gear contexts, so compositional heterogeneity can accumulate into a strong aggregate juvenile bias. Taken together, these points

suggest that the demographic bias documented here is not a property of stranding data specifically, but a feature of opportunistic mortality sampling more broadly.

Strandings remain a valuable source of information on cetacean biology, health, and mortality, providing material for necropsy, tissue sampling, contaminant analysis, diet studies, reproductive assessments, and genetic work, all of which would be difficult or impossible to obtain from living animals. The present findings concern representativeness, not utility: the issue is not whether stranding data are informative, but whether they can be treated as unbiased demographic samples of living populations. Strandings also offer early warning signals of emerging threats, such as disease outbreaks, harmful algal blooms, fisheries interactions, or environmental contaminant spikes. Our results indicate that strandings should not be uncritically treated as unfiltered or demographically representative samples of living populations. When making inferences about population age structure, stage-specific survival rates, or overall population trends, the systematic biases we have documented must be acknowledged and, where possible, accounted for.

How can this be done? One approach is to use stranding data primarily for surveillance and hypothesis generation rather than for direct demographic parameter estimation. Strandings can alert us to potential problems, such as unusual mortality events, spatial clustering of deaths, and temporal pulses in specific age classes, without necessarily providing precise estimates of vital rates or population growth. A second approach is to integrate stranding data with other data sources in formal statistical models that explicitly account for bias. For example, integrated population models (IPMs) can combine stranding records with survey-based abundance estimates, telemetry data, and bycatch records, using hierarchical modelling frameworks to partition observation error, process error, and sampling bias. If stranding detection probabilities or stage-specific vulnerabilities can be estimated or bounded, they can be incorporated as covariates or priors in the modelling framework.

A third and complementary approach, which we have pursued here, is to develop explicit diagnostics that quantify the magnitude and direction of bias. By comparing observed compositions with asymptotic, model-based expectations, we provide a calibration tool to inform how much weight to place on stranding-derived demographic inferences. If divergence from asymptotic stage structure (SSD) is modest, strandings may be cautiously used for demographic inference, with appropriate caveats. If divergence is large and persistent, as we found for juveniles, alternative data sources or strong corrective models are needed.

The framework requires only two inputs: age-composition data (which can be derived from body size if direct age estimates are unavailable) and a demographically informed MPM (which can be parameterised using the literature, data from related species, or expert opinion). Both inputs are increasingly available for species of conservation concern.

We emphasise that deviation from SSD should not be interpreted as evidence that strandings are necessarily misrepresenting the true population structure; a living population in a transient demographic state will also depart from SSD, independently of any sampling bias. This seems to be the case of the Belt Sea population of porpoises, which has undergone a seemingly drastic decline during the past decade (Owen et al. 2024), although part of this change may also reflect a geographic shift in where porpoises occur. What our framework quantifies is inconsistency between observed mortality composition and asymptotic expectations under the assumed demography, a signal informative about bias, but one that cannot fully separate demographic filtering from genuine transient population dynamics.

This interpretive constraint is one of several limitations worth acknowledging. A potential concern is that the detected deviation from SSD reflects misspecification of the baseline model rather than genuine demographic bias. Our simulation analysis directly addresses this: only 5 of 500 randomly sampled biologically plausible matrices (1.0%) produced SSDs resembling the observed mortality composition under the Keyfitz's $\Delta \leq 0.15$ criterion. This indicates that the

juvenile over-representation signal cannot readily be attributed to a particular choice of vital rates, and that the conclusion is robust to parametric uncertainty in a way that sensitivity analyses around a single baseline cannot fully establish. Even so, vital rates do vary among populations, fluctuate over time in response to environmental conditions, and are difficult to estimate precisely in long-lived marine mammals. Future work should ideally integrate direct estimates of vital rates from telemetry studies, mark-recapture data, or longitudinal monitoring programmes where available. Relatedly, our MPM is parameterised primarily from North Sea-focused literature, whereas Baltic harbour porpoises are genetically distinct and may have different vital rates; this could influence the absolute magnitude of the inferred mismatch in Region F (the Baltic-adjacent sector) in particular, although it does not change the broader finding of substantial divergence from the modelled SSD reference.

We also used body length as a proxy for demographic stage. While this is a practical and widely used approach, it is imperfect. Growth rates vary among individuals, and some animals may reach reproductive maturity at sizes outside our defined thresholds. More refined staging, perhaps incorporating additional morphometric or physiological indicators (e.g., tooth wear, reproductive tract development, hormone levels), could improve precision. However, such data are rarely available for large samples of stranded animals, and the body-length approach we employed is a practical compromise between biological realism and operational feasibility.

We did not explicitly model detection probability, reporting effort, or drift dynamics. These factors likely vary across regions and over time, introducing heterogeneity beyond what our framework captures. In addition, the Murphy mixed-mortality source partially overlaps the IJsseldijk North Sea stranding geography (regions A-C and part of D), so some cross-source comparisons are not fully independent. Where available, cause-of-death information could also provide important context: strandings due to bycatch might show different demographic patterns from those due to disease or starvation.

In our baseline model the population grows at about 3% a year, within the 5–10% maximum reported for porpoises, but it is not matched to the local population trend in each region and year, and expected compositions differ between growing and declining populations. We also treat the animals as one randomly mating population, although porpoises show some population structure; if strandings come from groups with different vital rates, the expected composition would differ between areas. Finally, the juvenile period is uncertain: a longer period gives a larger expected juvenile share and a smaller, though still significant, mismatch (Supplementary Table S17).

Conclusions

Harbour porpoise mortality records from both strandings and bycatch show substantial and consistent departures from asymptotic demographic expectations, as represented by the stable stage distribution predicted by a stage-structured matrix population model. Juveniles are significantly over-represented across all spatial and temporal scales examined relative to asymptotic expectations (the excess is smaller when compared with the expected composition of deaths, and some bycatch, mixed and hunting samples are consistent with it), a pattern that is robust to multiple sensitivity analyses and alternative analytical choices. This departure from the living population has important implications for the interpretation and use of mortality data in demographic inference, risk assessment, and conservation planning. Whether the juvenile excess reflects sampling bias or simply higher juvenile mortality cannot be resolved without independent estimates of juvenile survival.

Stranding and bycatch data remain valuable for surveillance, early warning, and biological investigation, but they should not be treated as unfiltered or demographically representative samples of living populations. The MPM-based diagnostic framework we present provides a practical, quantitative tool for assessing representativeness in opportunistic wildlife datasets,

including other marine mammal species and other non-random data streams where demographic bias is suspected but not formally quantified.

Future work should focus on integrating explicit models of detection probability and reporting effort, incorporating covariates such as coastline length, human population density, stranding network coverage, and oceanographic conditions (currents and wind), incorporating cause-of-death information, and combining stranding data with other demographic data sources in formal integrated population models. In practice, the framework enables more cautious and calibrated use of mortality data in bycatch-risk assessment and conservation planning.

Acknowledgements

We thank the many field workers, veterinarians, observers, and stranding network coordinators who contributed to the collection and curation of the harbour porpoise mortality datasets used in this study. We also acknowledge the researchers and institutions who compiled and made available the source datasets, including stranding, bycatch, mixed-mortality, and historical hunting records. This work originated as the master's thesis of MKHH at the Department of Biology, University of Southern Denmark.

Funding

This research received no specific grant funding.

Data Availability

Stranding data used in this study are derived from the compiled dataset from IJsseldijk et al. (2020). Bycatch data were sourced from IJsseldijk et al. (2022), Brennecke et al. (2021), and Frie & Lindström (2024); mixed bycatch + stranding mortality data were sourced from Murphy et al. (2020). Historical hunting data were sourced from Lockyer et al. (2001) and Møhl-Hansen

(1954). All analytical code, input data, and derived outputs required to reproduce the analyses are archived on Zenodo: <https://doi.org/10.5281/zenodo.18664347>.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

Supplementary materials consist of the following:

- Supplementary Methods: definitions and implementation details for secondary diagnostics (chi-square and Hellinger) and full settings for the simulation-based plausibility analysis (including archetype, constraints, seed, and thresholds).
- Figure S1: SSD uncertainty distribution from Monte Carlo MPM sampling.
- Figure S2: Distribution of Keyfitz's Δ across region-years, showing that SSD divergence is heterogeneous across region-years.
- Figure S3: PCA (PC1–PC3) view of life-history transition-rate space for unconstrained biologically plausible simulated MPMs, showing the structure of sampled transition combinations in an alternative PCA projection.
- Figure S4: Pairwise correlation matrix of regional Keyfitz's Δ time series.
- Figure S5: Strandings population structure through time by region.
- Table S1: SSD uncertainty summary by stage class, showing that SSD uncertainty intervals are bounded and interpretable across stage classes.
- Table S2: Summary statistics for Keyfitz's Δ across region-years, showing that average and upper-range divergence remain substantial.

- 768 • Table S3: Summary statistics for Hellinger distance across region-years showing that
769 conclusions are concordant across distance metrics.
- 770 • Table S4: Region-year descriptive context for stranded harbour porpoises, showing that
771 sampling intensity and temporal coverage vary across regions (A-F).
- 772 • Table S5: Juvenile bias sensitivity across filtering scenarios, showing that the juvenile
773 over-representation signal is robust to alternative data filters.
- 774 • Table S6: Keyfitz's Δ sensitivity across filtering scenarios, showing that divergence from
775 SSD remains substantial under sensitivity filters.
- 776 • Table S7: Juvenile bias sensitivity to SSD uncertainty bounds, showing that juvenile over-
777 representation persists under SSD uncertainty.
- 778 • Table S8: Keyfitz's Δ sensitivity to SSD uncertainty bounds, showing that divergence
779 magnitudes remain elevated across plausible SSD variants.
- 780 • Table S9: Chi-square goodness-of-fit results across strandings and bycatch groupings,
781 showing strong and directionally concordant deviation from SSD expectations.
- 782 • Table S10: Observed juvenile percentage and Keyfitz's Δ for each source against the
783 living-population SSD and the expected composition of deaths.
- 784 • Table S11: Keyfitz's Δ with the model calibrated to a stable population ($\lambda = 1$).
- 785 • Table S12: Sensitivity of expected stage composition to vital rates (20,000 random
786 parameter sets).
- 787 • Table S13: Effect of changing one vital rate at a time.
- 788 • Table S14: Survival rates that would make the model's death composition match each
789 observed sample.
- 790 • Table S15: Three-stage model compared with age-by-age models with a fixed juvenile
791 period.

- 792 • Table S16: Age-by-age share of the living population in the baseline age-structured
793 model.
- 794 • Table S17: Keyfitz's Δ for each source under the three-stage and age-by-age models.
795

References

- Autenrieth, M., Havenstein, K., De Cahsan, B., Canitz, J., Benke, H., Roos, A., Pampoulie, C., Sigurðsson, G. M., Siebert, U., Olsen, M. T., Biard, V., Heide-Jørgensen, M. P., Öztürk, A. A., Öztürk, B., Lawson, J. W., and Tiedemann, R. 2024. Genome-wide analysis of the harbour porpoise (*Phocoena phocoena*) indicates isolation-by-distance across the North Atlantic and potential local adaptation in adjacent waters. *Conservation Genetics* 25(2):563-584. <https://doi.org/10.1007/s10592-023-01589-0>
- Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*. 2nd edn. Sinauer Associates, Sunderland, MA, USA.
- Celemín, E., Autenrieth, M., Roos, A., Pawliczka, I., Quintela, M., Lindstrøm, U., Benke, H., Siebert, U., Lockyer, C., Berggren, P., Öztürk, A. A., Öztürk, B., Lesage, V., and Tiedemann, R. 2025. Evolutionary history and seascape genomics of harbour porpoises (*Phocoena phocoena*) across environmental gradients in the North Atlantic and adjacent waters. *Molecular Ecology Resources* 25:e13860. <https://doi.org/10.1111/1755-0998.13860>
- Cervin, L., Harkonen, T., and Harding, K. C. 2020. Multiple stressors and data deficient populations: a Bayesian framework for incorporating uncertainty in cumulative risk assessments for Baltic harbour porpoises. *Environment International* 144:106076.
- Clausen, B., and Andersen, S. 1988. Evaluation of bycatch and health status of the harbour porpoise (*Phocoena phocoena*) in Danish waters. *Danish Review of Game Biology* 13(5):1-20.
- Cook, R. S., White, M., Trainer, D. O., and Glazener, W. C. 1971. Mortality of young white-tailed deer fawns in south Texas. *The Journal of Wildlife Management* 35:47-56.
- IJsseldijk, L. L., ten Doeschate, M. T. I., Brownlow, A., Davison, N. J., Deaville, R., Galatius, A., Gilles, A., Haelters, J., Jepson, P. D., Keijl, G. O., Kinze, C. C., Olsen, M. T., Siebert, U.,

819 Thostesen, C. B., van den Broek, J., Grone, A., and Heesterbeek, H. 2020. Spatiotemporal
820 mortality and demographic trends in a small cetacean: strandings to inform conservation
821 management. *Biological Conservation* 249:108733.

822 IJsseldijk, L. L., Leopold, M. F., Begeman, L., Kik, M. J. L., Wiersma, L., Morell, M., Bravo
823 Rebolledo, E. L., Jauniaux, T., Heesterbeek, H., and Grone, A. 2022. Pathological findings in
824 stranded harbor porpoises (*Phocoena phocoena*) with special focus on anthropogenic causes.
825 *Frontiers in Marine Science* 9:997388.

826 Jones, O. R. 2021. Life tables: construction and interpretation. In Salguero-Gomez, R., and
827 Gamelon, M. (eds.), *Demographic Methods Across the Tree of Life*. Oxford University Press, Oxford,
828 UK.

829 Jones, O. R. (2025). mpmsim: An R package for simulating matrix population models. *Methods in*
830 *Ecology and Evolution*, 16(5), 904-911.

831 Keyfitz, N. 1968. *Introduction to the Mathematics of Population*. Addison-Wesley, Reading, MA, USA.

832 Kindt-Larsen, L., Glemarec, G., Berg, C. W., Königson, S., Kroner, A. M., Søgaard, M., and
833 Lusseau, D. 2023. Knowing the fishery to know the bycatch: bias-corrected estimates of harbour
834 porpoise bycatch in gillnet fisheries. *Proceedings of the Royal Society B: Biological Sciences*
835 290:20222570.

836 Kesselring, T., Viquerat, S., Brehm, R., and Siebert, U. 2017. Coming of age: do female harbour
837 porpoises (*Phocoena phocoena*) from the North Sea and Baltic Sea have sufficient time to reproduce
838 in a human influenced environment? *PLOS ONE* 12:e0186951.

839 Kinze, C. C. 1990. *The harbour porpoise (Phocoena phocoena) stock identification and migration patterns*
840 *in Danish and adjacent waters*. PhD thesis, University of Copenhagen, Copenhagen, Denmark.

841 Kinze, C. C., Czeck, R., Herr, H., and Siebert, U. 2021. Cetacean strandings along the German
842 North Sea coastline 1604-2017. *Journal of the Marine Biological Association of the United Kingdom*
843 101:483-502.

844 Linnell, J. D. C., Aanes, R., and Andersen, R. 1995. Who killed Bambi? The role of predation in
845 the neonatal mortality of temperate ungulates. *Wildlife Biology* 1:209-223.

846 Lockyer, C. 2003. Harbour porpoises (*Phocoena phocoena*) in the North Atlantic: biological
847 parameters. *NAMMCO Scientific Publications* 5:71-89.

848 Lockyer, C., Heide-Jorgensen, M. P., Jensen, J., Kinze, C. C., and Sorensen, T. B. 2001. Age,
849 length and reproductive parameters of harbour porpoises (*Phocoena phocoena*) (L.) from West
850 Greenland. *ICES Journal of Marine Science* 58:154-162.

851 Lockyer, C., and Kinze, C. C. 2003. Status, ecology and life history of harbour porpoise (*Phocoena*
852 *phocoena*) in Danish waters. *NAMMCO Scientific Publications* 5:143-175.

853 Murphy, S., Petitguyot, M. A., Jepson, P. D., Deaville, R., Lockyer, C., Barnett, J., and Minto, C.
854 2020. Spatio-temporal variability of harbor porpoise life history parameters in the North-East
855 Atlantic. *Frontiers in Marine Science* 7:502352.

856 Møhl-Hansen, U. 1954. Investigations on reproduction and growth of the porpoise (*Phocaena*
857 *phocaena* (L.)) from the Baltic. *Videnskabelige Meddelelser fra Dansk Naturhistorisk Forening* 116:369-
858 396.

859 Olsen, M. T., Nielsen, N. H., Biard, V., Teilmann, J., Ngô, M. C., Vikingsson, C., Gunnlaugsson,
860 T., Stenson, G., Lawson, J., Lah, L., Tiedemann, R., and Heide-Jørgensen, M. P. 2022. Genetic
861 and behavioural data confirm the existence of a distinct harbour porpoise ecotype in West
862 Greenland. *Ecological Genetics and Genomics* 22:100108. <https://doi.org/10.1016/j.egg.2021.100108>

863 Owen, K., Gilles, A., Authier, M., Carlstrom, J., Genu, M., Kyhn, L. A., Nachtsheim, D. A.,
864 Ramirez-Martinez, N. C., Siebert, U., Skold, M., Teilmann, J., Unger, B., and Sveegaard, S. 2024.
865 A negative trend in abundance and an exceeded mortality limit call for conservation action for
866 the vulnerable Belt Sea harbour porpoise population. *Frontiers in Marine Science* 11:1289808.

867 Maeda, S., Sakurai, K., Akamatsu, T., Matsuda, A., Yamamura, O., Kobayashi, M., and Matsuishi,
868 T. F. 2021. Foraging activity of harbour porpoises around a bottom-gillnet in a coastal fishing
869 ground, under the risk of bycatch. *PLOS ONE* 16:e0246838.

870 Peltier, H., Baagoe, H. J., Camphuysen, K. C. J., Czeck, R., Dabin, W., et al. 2013. The stranding
871 anomaly as population indicator: the case of harbour porpoise (*Phocoena phocoena*) in North-
872 Western Europe. *PLOS ONE* 8:e62180.

873 Rogan, E., Canadas, A., Macleod, K., Santos, M. B., Mikkelsen, B., Uriarte, A., Van Canneyt, O.,
874 Vazquez, J. A., and Hammond, P. S. 2017. Distribution, abundance and habitat use of deep-
875 diving cetaceans in the North-East Atlantic. *Deep-Sea Research Part II: Topical Studies in Oceanography*
876 141:8-19.

877 Reilly, S., and Barlow, J. 1986. Rates of increase in dolphin population size. *Fishery Bulletin*
878 84:527-533.

879 Scheidat, M., Gilles, A., Kock, K. H., and Siebert, U. 2008. Harbour porpoise (*Phocoena phocoena*)
880 abundance in the southwestern Baltic Sea. *Endangered Species Research* 5:215-223.

881 Brennecke, D., Wahlberg, M., Gilles, A., and Siebert, U. 2021. Age and lunar cycle predict harbor
882 porpoise bycatch in the south-western Baltic Sea. *PeerJ* 9:e12284.

883 Frie, A. K., and Lindstrom, U. 2024. Exploring the effects of methodological choices on the
884 estimation and biological interpretation of life history parameters for harbour porpoises in
885 Norway and beyond. *PLOS ONE* 19:e0301427.

886 Siebert, U., Blanchet, M. A., Teilmann, J., Andersen Hansen, K., Kristensen, J., Bunschoek, P.,
 887 Dietz, R., Desforges, J. P., Sonne, C., and Desportes, G. 2020. Haematology and clinical blood
 888 chemistry in harbour porpoises (*Phocoena phocoena*) from the inner Danish waters. *Environment*
 889 *International* 143:105937.

890 Woodley, T. H., and Read, A. J. 1991. Potential rates of increase of a harbour porpoise (*Phocoena*
 891 *phocoena*) population subjected to incidental mortality in commercial fisheries. *Canadian Journal of*
 892 *Fisheries and Aquatic Sciences* 48:2429-2435.

893 Stubben, C. J., and Milligan, B. G. 2007. Estimating and analysing demographic models using the
 894 popbio package in R. *Journal of Statistical Software* 22:1-23.

895 Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York, NY, USA.

896 Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., Francois, R.,
 897 Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S.
 898 M., Muller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D.,
 899 Wilke, C., Woo, K., and Yutani, H. 2019. Welcome to the tidyverse. *Journal of Open Source Software*
 900 4:1686.

Supplementary Information: Do harbour porpoise mortality records reflect living population structure? A matrix population model diagnostic

2026-10-13

Michaela Kirstine S. M. H. Søgaard^{1,2}, Magnus Wahlberg^{1,2,3}, and Owen R. Jones^{1*}

¹ Department of Biology, University of Southern Denmark, Odense, Denmark

² Marine Biological Research Centre, University of Southern Denmark, Kerteminde, Denmark

³ Fjord & Bælt, Kerteminde, Denmark

* Corresponding author: jones@biology.sdu.dk

Supplementary Results

This supplement includes additional material referenced in the main text: SSD uncertainty outputs (Figure S1, Table S1), distribution summaries for divergence metrics (Figure S2, Tables S2-S3), an additional MPM-space PCA view (Figure S3), the regional synchrony correlation matrix (Figure S4), descriptive stranding context (Table S4, Figure S5), sensitivity analyses (Tables S5-S8), and additional goodness-of-fit diagnostics supporting concordance with Keyfitz's Δ (Table S9).

Supplementary Methods for secondary diagnostics

In the main manuscript, Keyfitz's Δ is treated as the primary divergence metric. Here we document the secondary diagnostics used to verify robustness of inference.

Chi-square goodness-of-fit tests were computed as:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i},$$

where O_i and E_i are observed and SSD-expected counts for stage i .

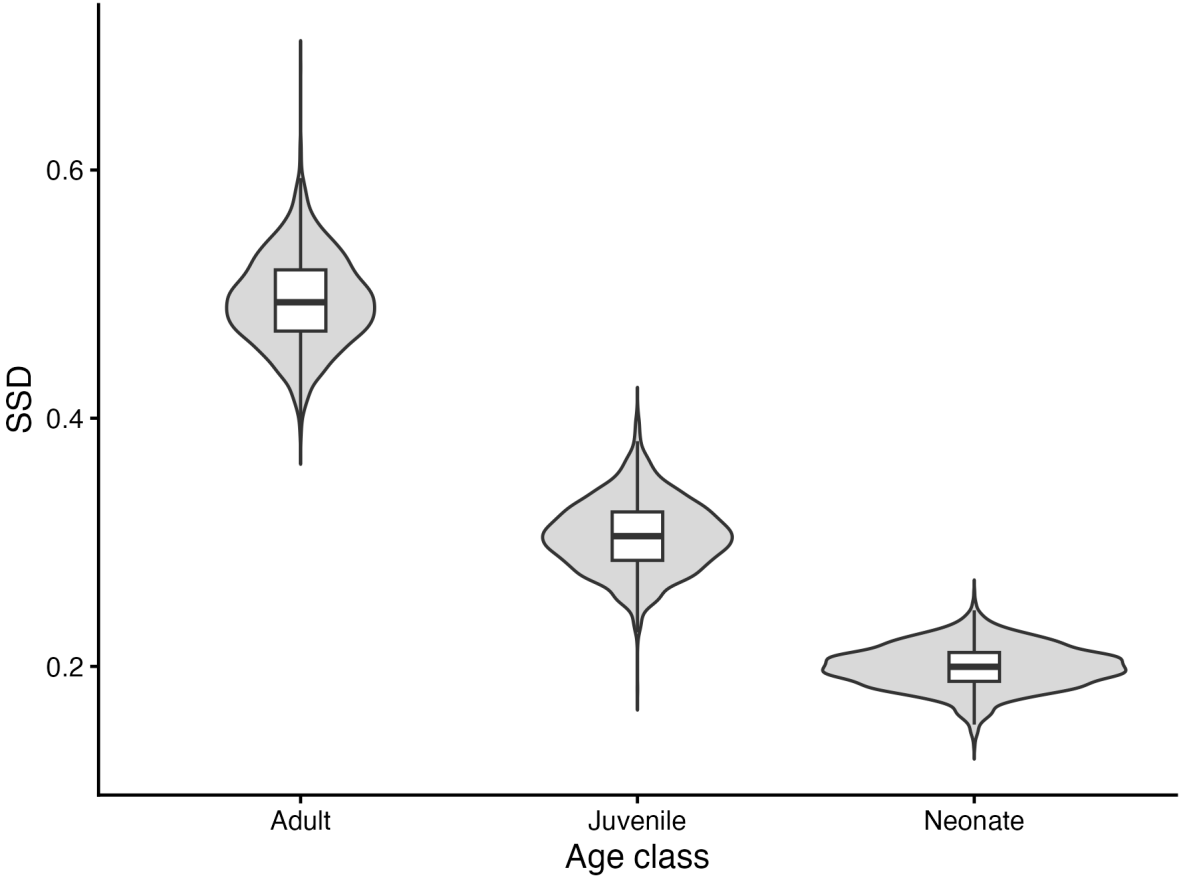
Hellinger distance was computed as:

$$H = \sqrt{1 - \sum_i \sqrt{p_{i,obs} p_{i,SSD}}},$$

where $p_{i,obs}$ and $p_{i,SSD}$ are observed and SSD-expected stage proportions. This metric was used as a complementary distance check and interpreted for concordance with Keyfitz's Δ , not as a replacement for the primary metric.

Simulation-based plausibility settings

For the simulation-based plausibility analysis, we generated random 3-stage Lefkovitch matrices using `mpmsim::rand_lefko_set` in `code/analysis/porpoise-modeling.R`. We used archetype 4 (monotonic forward stage progression without retrogression), `output = "Type1"`, and a fixed random seed (`set.seed(43)`). We generated 500 matrices (`n = 500`), constraining only the dominant eigenvalue to a biologically plausible range ($\lambda \in [0.7, 1.3]$). Fecundity was given broad variation via the specified fecundity bounds (F_A from 0 to 10; other fertility entries fixed to 0). For each simulated matrix, we calculated the SSD and then Keyfitz's Δ against the observed overall stranding composition (not the baseline SSD). We report the proportion of simulations meeting the Keyfitz's Δ resemblance thresholds of ($\Delta \leq 0.15$, $\Delta \leq 0.10$) and summary statistics of the simulated Δ distribution.



41

42 *Figure S1: SSD uncertainty distribution from Monte Carlo MPM sampling. This shows that plausible SSD*
43 *stage proportions vary within bounded ranges under vital-rate uncertainty, supporting robustness interpretation in*
44 *the main manuscript.*

45 *Table S1: SSD uncertainty summary by stage class, showing that SSD uncertainty intervals are bounded and*
46 *interpretable across stage classes.*

Ageclass	n	mean	median	sd	CI low	CI high	min	max
Adult	2,000	0.495	0.493	0.036	0.425	0.568	0.384	0.683
Juvenile	2,000	0.306	0.305	0.029	0.253	0.364	0.182	0.408
Neonate	2,000	0.200	0.200	0.017	0.163	0.233	0.136	0.259

47

Distribution of divergence metrics

This section provides the additional metric checks used to verify that conclusions do not depend on Keyfitz’s Δ alone. Hellinger distance follows the same regional and temporal structure as Keyfitz’s Δ , supporting robustness of the main interpretation.

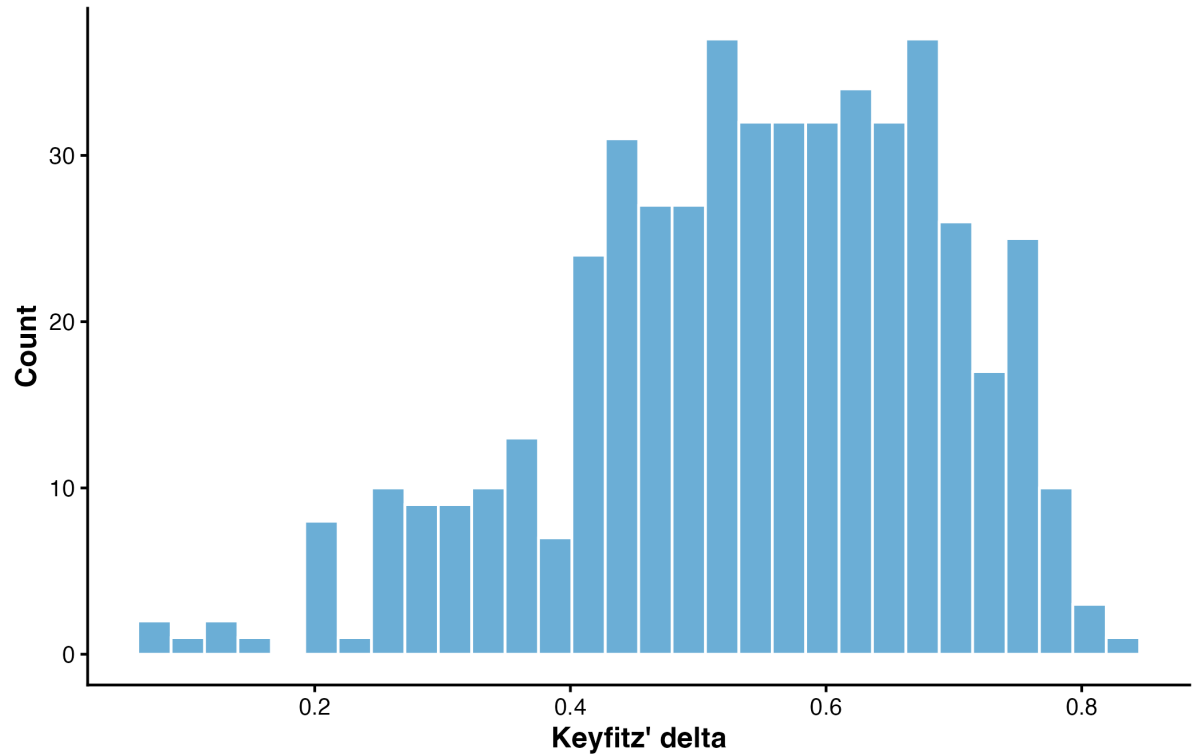


Figure S2: Distribution of Keyfitz’s Δ across region-years, showing that SSD divergence is heterogeneous across region-years.

Table S2: Summary statistics for Keyfitz’s Δ across region-years, showing that average and upper-range divergence remain substantial.

n	mean	median	sd	min	max
167	0.268	0.257	0.122	0.02	0.695

Table S3: Summary statistics for Hellinger distance across region-years showing that conclusions are concordant across distance metrics.

n	mean	median	sd	min	max
167	0.228	0.224	0.107	0.017	0.669

Additional MPM space visualisation

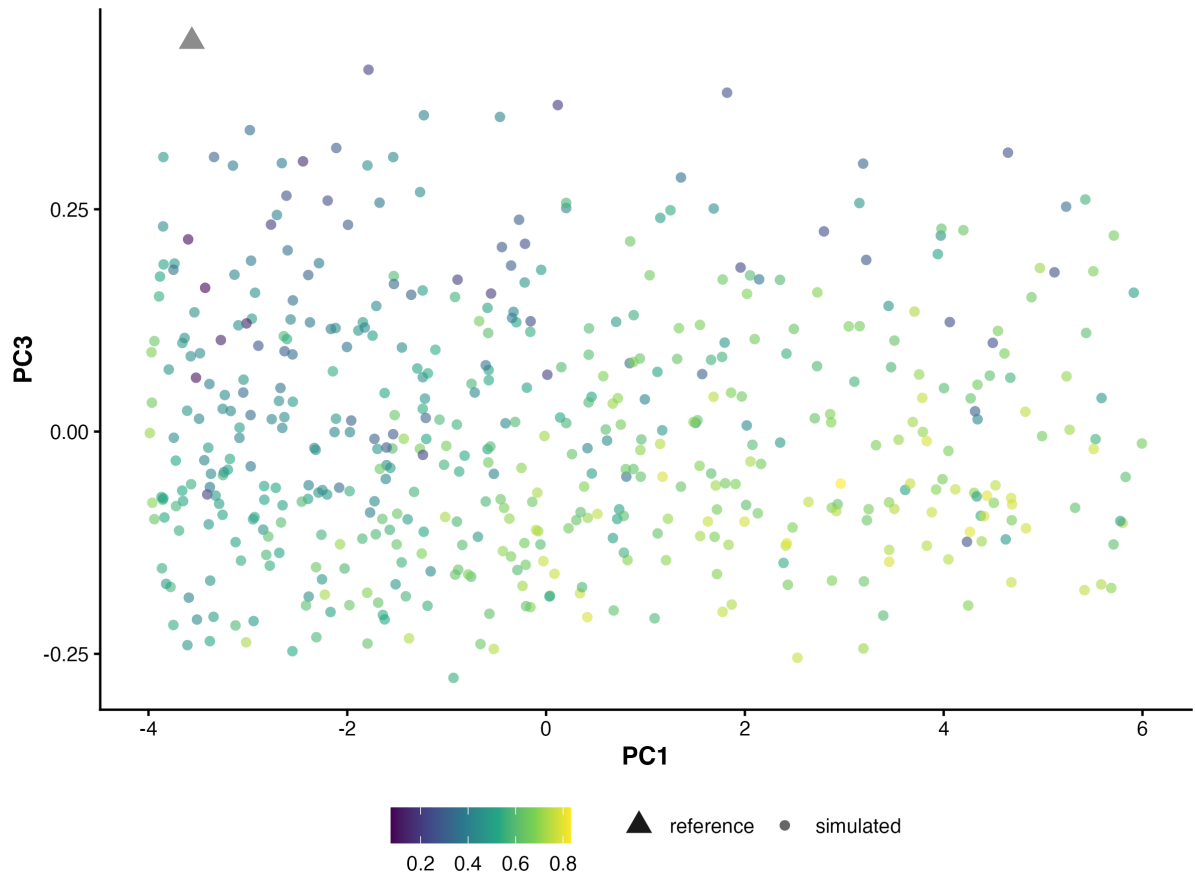
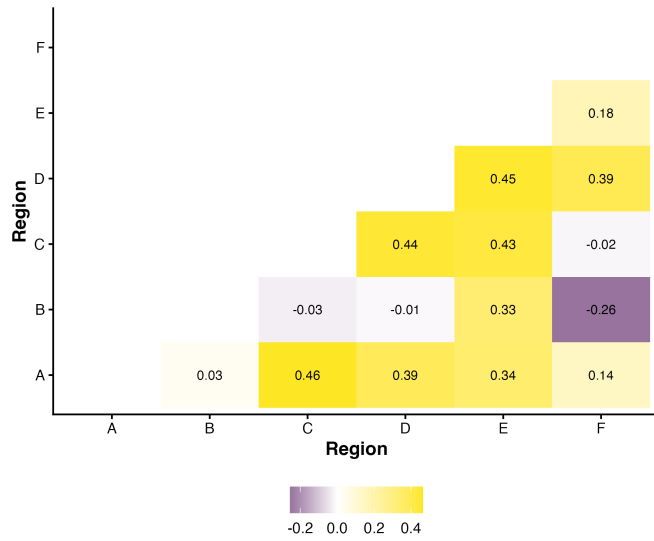


Figure S3: PCA (PC1–PC3) view of life-history transition-rate space for unconstrained biologically plausible simulated MPMs, showing the structure of sampled transition combinations in an alternative PCA projection. Point colour encodes Keyfitz’s Δ (divergence from observed overall stage composition), and point shape indicates matrix type (simulated vs reference).

67 *Regional synchrony matrix*



68

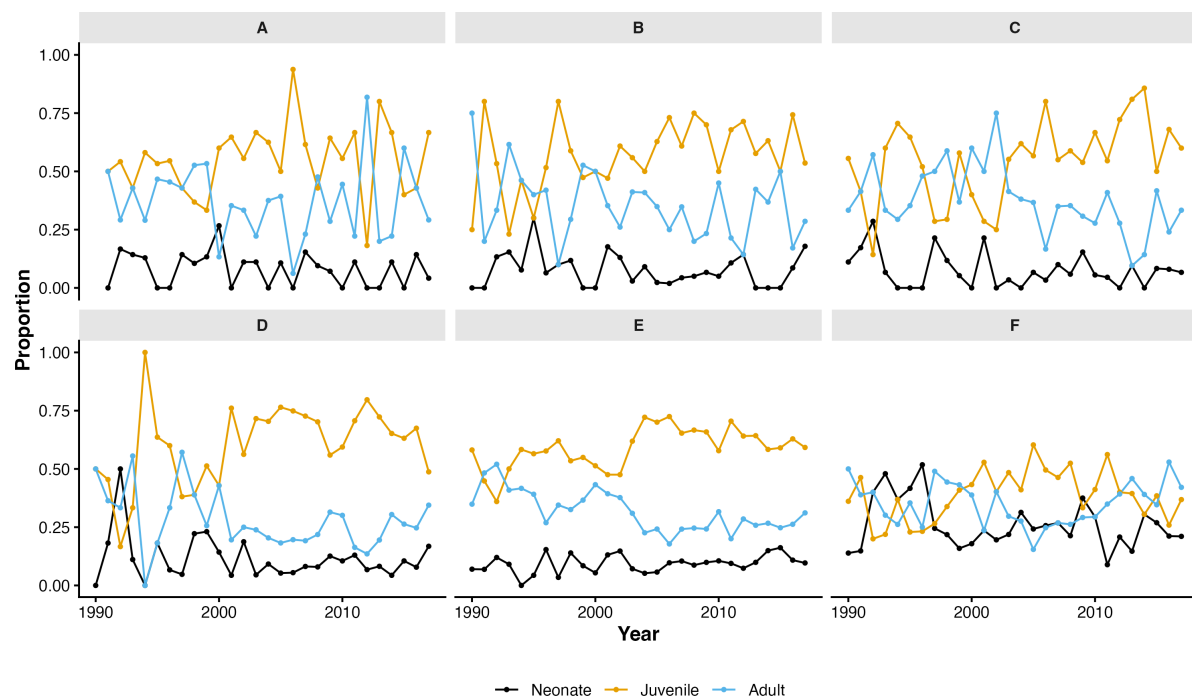
69 *Figure S4: Pairwise correlation matrix of regional Keyfitz's Δ time series. Each cell value is a Pearson correlation*
70 *coefficient between two regional divergence trajectories over time; higher positive values indicate stronger temporal*
71 *synchrony in SSD divergence between those regions.*

72 *Descriptive stranding context by region-year*

73 *Table S4: Region-year descriptive context for stranded harbour porpoises, showing that sampling intensity and*
74 *temporal coverage vary across regions (A-F).*

Region	years	total strandings	median annual n	iqr annual n	min annual n	max annual n
A	1991-2017	456	16.0	7.50	2	35
B	1990-2017	610	20.0	15.50	4	52
C	1990-2017	554	17.5	8.75	7	60
D	1990-2017	2,702	82.0	137.50	5	300
E	1990-2017	3,846	94.5	187.75	12	444
F	1990-2017	2,695	85.0	69.00	36	225

75



76

77 *Figure S5: Strandings population structure through time by region. Each panel represents a region; points are*
78 *observed annual age-class proportions and coloured lines trace temporal trajectories for neonates, juveniles, and*
79 *adults, showing how stranding composition changes through time within regions.*

80 *Sensitivity analyses*

81 *Table S5: Juvenile bias sensitivity across filtering scenarios, showing that the juvenile over-representation signal is*
82 *robust to alternative data filters.*

scenario	estimate	expected	p-value	n
all region year	0.581	0.305	<0.001	10,863
complete ageclasses	0.581	0.305	<0.001	10,481
min count 10	0.581	0.305	<0.001	10,778
complete and min 10	0.581	0.305	<0.001	10,430

83

Table S6: Keyfitz's Δ sensitivity across filtering scenarios, showing that divergence from SSD remains substantial under sensitivity filters.

scenario	n	mean	median	sd	min	max
all region year	167	0.268	0.257	0.122	0.02	0.695
complete ageclasses	139	0.257	0.256	0.116	0.02	0.504
min count 10	154	0.263	0.258	0.117	0.02	0.632
complete and min 10	133	0.256	0.256	0.117	0.02	0.504

Table S7: Juvenile bias sensitivity to SSD uncertainty bounds, showing that juvenile over-representation persists under SSD uncertainty.

scenario	expected	p-value	n
ssd mean	0.306	<0.001	10,863
ssd ci low	0.300	<0.001	10,863
ssd ci high	0.313	<0.001	10,863

Table S8: Keyfitz's Δ sensitivity to SSD uncertainty bounds, showing that divergence magnitudes remain elevated across plausible SSD variants.

scenario	n	mean	median	sd	min	max
ssd mean	167	0.268	0.257	0.122	0.020	0.694
ssd ci low	167	0.272	0.262	0.123	0.020	0.700
ssd ci high	167	0.262	0.252	0.121	0.027	0.687

Chi-square diagnostics

These diagnostics report formal goodness-of-fit test results for the strandings and bycatch datasets. They are presented here to document statistical agreement with the Keyfitz's Δ results while keeping the main manuscript focused on the primary distance metric, Keyfitz's Δ .

Table S9: Chi-square goodness-of-fit results across strandings and bycatch groupings, showing strong and directionally concordant deviation from SSD expectations.

dataset	group	statistic	df	p-value	n
Strandings overall	Overall	3,901.009	2	0.000	10,863
Strandings by region	A	132.776	2	0.000	456
Strandings by region	B	257.310	2	0.000	610
Strandings by region	C	170.561	2	0.000	554
Strandings by region	D	1,763.331	2	0.000	2,702
Strandings by region	E	1,957.532	2	0.000	3,846
Strandings by region	F	293.011	2	0.000	2,695
Strandings by country	Belgium	810.909	2	0.000	877
Strandings by country	Denmark	228.721	2	0.000	998
Strandings by country	England	329.289	2	0.000	1,143
Strandings by country	Germany	312.937	2	0.000	1,697
Strandings by country	Netherlands	2,827.688	2	0.000	5,187
Strandings by country	Scotland	365.316	2	0.000	961
Bycatch overall	Overall	62.631	2	0.000	270
Bycatch by source	Brennecke 2021	118.695	2	0.000	136
Bycatch by source	Frie Lindstrom 2024	0.702	2	0.704	134

Death-based expectation, vital-rate sensitivity and age-structured checks

The tables in this section support the Methods and Results in the main text. Table S10 compares the data with both the living-population SSD and the expected composition of deaths. Table S11 repeats this with the model set to a stable population ($\lambda = 1$). Tables S12 and S13 show how the

expectations change with the vital rates, and Table S14 shows the survival rates that would be needed to match the data. Tables S15–S17 compare the three-stage model with age-by-age models with a fixed juvenile period. All were produced by `code/analysis/death-expectation-sensitivity.R`, `age-structured-check.R` and `revision-tables.R`.

Table S10: Baseline model: observed juvenile percentage and Keyfitz's Δ for each source against the living-population SSD and against the expected composition of deaths ($SSD \times \text{stage-specific mortality}$, rescaled). p -values are from two-sided tests of the juvenile proportion.

juvenile %	Living SSD: expected juvenile %	Living SSD: Keyfitz Δ	Living SSD: p (two-sided)	Deaths:
58.1	30.5	0.28	<0.001	
54.8	30.5	0.24	<0.001	
60.0	30.5	0.29	<0.001	
55.2	30.5	0.25	<0.001	
67.7	30.5	0.37	<0.001	
63.4	30.5	0.33	<0.001	
41.6	30.5	0.16	<0.001	
73.5	30.5	0.43	<0.001	
31.3	30.5	0.03	0.911	
72.8	30.5	0.42	<0.001	
38.9	30.5	0.09	0.052	
38.5	30.5	0.17	0.275	
19.3	30.5	0.31	<0.001	
34.1	30.5	0.17	0.285	
50.6	30.5	0.23	<0.001	
40.0	30.5	0.17	0.013	
36.6	30.5	0.06	0.221	

110

111 *Table S11: Keyfitz's Δ with the model calibrated to a stable population ($\lambda = 1$), either by lowering fertility or by*
112 *lowering adult survival (other rates at baseline). The baseline model has $\lambda = 1.03$.*

Baseline ($\lambda=1.03$): SSD	Baseline: deaths	$\lambda=1$ via fertility: SSD	$\lambda=1$ via fertility: deaths	$\lambda=1$ via adult s
0.28	0.25	0.31	0.27	
0.24	0.21	0.27	0.24	
0.29	0.27	0.33	0.29	
0.25	0.22	0.28	0.24	
0.37	0.34	0.40	0.37	
0.33	0.30	0.36	0.32	
0.16	0.08	0.23	0.11	
0.43	0.40	0.46	0.42	
0.03	0.14	0.04	0.08	
0.42	0.39	0.45	0.42	
0.09	0.18	0.11	0.14	
0.17	0.05	0.23	0.11	
0.31	0.43	0.25	0.37	
0.17	0.05	0.23	0.11	
0.23	0.17	0.30	0.20	
0.17	0.07	0.23	0.11	
0.06	0.09	0.12	0.06	

113

114 *Table S12: Sensitivity of the expected stage composition to vital rates for strandings (all regions), from 20,000*
115 *random parameter sets (Latin hypercube) and subsets with stable-ish population growth. Observed juvenile*
116 *percentage: 58.1%.*

ge)	Expected juvenile %, deaths (range)	% of draws reaching observed 58.1% (SSD)	% of draws reachi
	32.6 (1.2–92.0)	0	
	33.0 (1.2–79.5)	0	

ge)	Expected juvenile %, deaths (range)	% of draws reaching observed 58.1% (SSD)	% of draws reachi
	33.3 (1.2–70.9)	0	

117

118 *Table S13: Changing one vital rate at a time (all others at baseline): ranges of λ , expected juvenile percentage and*

119 *Keyfitz’s Δ for strandings (all regions).*

	Range	λ	Juvenile % living SSD	Juvenile % deaths	Keyfitz Δ SSD
	0.6–1	0.99–1.05	26.7–31.4	30.4–34.1	0.27–0.31
probability	0.25–0.5	1.02–1.05	24.8–34.7	27.4–37.8	0.23–0.33
ate survival = adult ²)	0.8–0.98	0.97–1.14	30.1–30.8	25.5–76.4	0.27–0.28
	0.6–0.98	0.98–1.09	28.1–32.6	5.5–48.2	0.25–0.30
	0.6–0.98	1.01–1.07	27.2–34.5	25.2–49.8	0.24–0.31

120

121 *Table S14: Survival rates that would make the model’s death composition match each observed sample*

122 *(maximum-likelihood fit with λ restricted to 0.97–1.03; fertility and maturation at baseline). Three rates are*

123 *fitted to two independent proportions, so these values are not unique estimates of mortality.*

Source	n	Adult survival	Neonate survival	Juvenile survival	
strandings: all regions	10,863	0.89	0.87	0.62	1
strandings: region A	456	0.86	0.92	0.68	1
strandings: region B	610	0.86	0.93	0.61	1
strandings: region C	554	0.84	0.92	0.67	1
strandings: region D	2,702	0.89	0.90	0.44	0
strandings: region E	3,846	0.89	0.90	0.55	1
strandings: region F	2,695	0.87	0.76	0.72	1
atch: Brennecke 2021	136	0.91	0.93	0.37	0
atch: Frie Lindstrom 2024	134	0.78	0.83	0.83	1
atch: IJsseldijk et al. 2022	103	0.90	0.94	0.39	0

Source	n	Adult survival	Neonate survival	Juvenile survival
hunting: Lockyer 2001, Maniitsoq North	126	0.78	0.88	0.78
hunting: Lockyer 2001, Nuuk Paamiut	52	0.86	0.71	0.71
hunting: Møhl-Hansen 1954	389	0.65	1.00	0.92
ed: Murphy 2020, Celtic Irish 1990-1999	217	0.86	0.66	0.73
ed: Murphy 2020, Celtic Irish 2000-2012	172	0.90	0.78	0.64
ed: Murphy 2020, North Sea 1990-1999	155	0.85	0.71	0.68
ed: Murphy 2020, North Sea 2000-2012	101	0.79	0.78	0.76

124

125 *Table S15: Three-stage model compared with age-by-age (Lefkovich) models in which juveniles take a fixed*

126 *number of years (one-year age classes with an adult plus-group).*

	λ	Living SSD: neonate / juvenile / adult %	Deaths: neonate / juvenile / adult %
geometric juvenile stage, mean 3 y)	1.034	20.1 / 30.5 / 49.4	28.7 / 33.4 / 37.9
d, juveniles ages 1-3 (adult from 4)	1.021	19.2 / 34.2 / 46.7	27.3 / 37.2 / 35.5
d, juveniles ages 1-2 (adult from 3)	1.047	21.0 / 26.7 / 52.3	30.2 / 29.4 / 40.4
d, juveniles ages 1-4 (adult from 5)	1.002	17.8 / 39.9 / 42.4	25.0 / 43.0 / 32.0

127

128 *Table S16: Age-by-age share of the living population in the baseline age-structured model (juveniles aged 1–3*

129 *years, adults from age 4).*

Age	Share of living population %
0	19.2
1	14.1
2	11.2
3	8.9
4+	46.7

130

131 Table S17: Keyfitz's Δ for each data source under the three-stage and age-by-age models, against the living-
 132 population SSD and the expected composition of deaths. 3-stage = baseline; Juv. 1-3 y = juveniles aged 1–3
 133 (adults from age 4); Juv. 1-2 y and Juv. 1-4 y = shorter and longer juvenile periods.

3-stage SSD	3-stage deaths	Juv. 1-3 y SSD	Juv. 1-3 y deaths	Juv. 1-2 y SSD	Juv. 1-2 y deaths
0.28	0.25	0.24	0.21	0.31	0.2
0.24	0.21	0.21	0.18	0.28	0.2
0.29	0.27	0.26	0.23	0.33	0.3
0.25	0.22	0.21	0.20	0.29	0.2
0.37	0.34	0.34	0.31	0.41	0.3
0.33	0.30	0.29	0.26	0.37	0.3
0.16	0.08	0.14	0.04	0.19	0.1
0.43	0.40	0.39	0.36	0.47	0.4
0.03	0.14	0.05	0.16	0.05	0.1
0.42	0.39	0.39	0.36	0.46	0.4
0.09	0.18	0.08	0.16	0.12	0.1
0.17	0.05	0.14	0.03	0.20	0.0
0.31	0.43	0.34	0.45	0.28	0.4
0.17	0.05	0.14	0.06	0.20	0.0
0.23	0.17	0.20	0.13	0.26	0.2
0.17	0.07	0.14	0.03	0.19	0.1
0.06	0.09	0.03	0.08	0.10	0.1