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

2026-03-02

Michaela Kirstine Hjelm Hansen[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 & Baelt, 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 opportunistic sources, such as stranding and bycatch records. While invaluable, these data may be subject to selective demographic filtering due to drift dynamics, detection probability, 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 classified harbour porpoise strandings (of 16,181 total records) in Danish and North Sea waters (1990–2017), bycatch and mixed-mortality records from four independent source datasets, and hunting captures from Greenland (1988–1989, 1995) and Denmark (1941–1944). Deviations from SSD were assessed using goodness-of-fit tests, distributional distance metrics (Keyfitz's Δ and Hellinger distance), and tests for juvenile over-representation.

Strandings showed strong and consistent departures from SSD expectations, with juveniles markedly over-represented across spatial and temporal scales. These patterns were robust to sensitivity analyses that addressed sample-size thresholds, missing age classes, and uncertainty in SSD estimates. Distance metrics indicated moderate to strong divergence from the asymptotic stage structure, with pronounced spatial heterogeneity and significant positive temporal trends 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.

Our results show that harbour porpoise stranding, bycatch, and even hunting data should not be assumed to be demographically representative of living populations. Researchers and managers relying on stranding or bycatch data for demographic inference should treat stage composition as a selective, biased signal rather than a population-representative sample.

Introduction

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

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

Multiple mechanisms could introduce demographic bias into stranding and bycatch datasets. Carcass drift, driven by ocean currents, wind, and the buoyancy of decomposing bodies, may transport animals far from their sites of death, creating a spatial 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). A demographically isolated population inhabits the Baltic Sea Proper, where abundance 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; 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: if mortality risk and recovery probability are non-selective with respect to age, the composition of any mortality sample would be expected to approach SSD. Departures from SSD 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. This includes 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.

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, Danish waters, and the western North Atlantic off Greenland — with the primary analytical focus on the North Sea and Danish waters proper. 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. 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:

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

where F_A is female neonate production per adult female per year, P denotes stage stasis, and G denotes forward transition. Adult natural survival was set to $S_A = 0.94$, based on Danish life-table estimates (Kinze 1990; reviewed in 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), $S_0 = S_A^2 = 0.8836$, 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 $S_J = 0.90$, this gives $G_J = 0.30$ and $P_J = 0.60$.

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 $F_A = 0.90 \times 0.466 = 0.4194$.

Under this annual post-breeding formulation, we set neonate stasis to zero ($P_N = 0$) and the neonatal forward transition to $G_N = S_0$, so that all surviving neonates advance to the juvenile stage by the next census.

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. 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
P_N	0.000	Neonate stasis probability (matrix element)	Assumed under annual post-breeding census ($P_N = 0$; surviving neonates advance to juvenile stage)
F_A	0.419	Female neonates per adult female per year (matrix element)	Derived as $F_A = b * p_f = 0.900 * 0.466$
G_N	0.884	Neonate-to-juvenile transition probability (matrix element)	Set as $G_N = S_0$; S_0 from square rule $S_0 = S_A$ (Woodley and Read 1991; Reilly and Barlow 1986)
P_J	0.600	Juvenile stasis probability (matrix element)	Derived as $P_J = S_J(1 - g)$ with $S_J = 0.900$ and $g = 1/3$
G_J	0.300	Juvenile-to-adult transition probability (matrix element)	Derived as $G_J = S_J\, g$ with $S_J = 0.900$ and $g = 1/3$
P_A	0.940	Adult stasis probability (matrix element)	Set equal to adult survival ($P_A = S_A$)
S_A	0.940	Adult annual natural survival	Kinze (1990), reviewed in Lockyer (2003)
S_0	0.884	Neonate (0–1 year) survival	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	0.900	Juvenile annual survival	Assumed; constrained by $S_0 < S_J \leq S_A$; tested via sensitivity analyses
g	1/3	Annual maturation probability	Derived from juvenile duration (ages 1–3); Lockyer (2003); Lockyer and Kinze (2003)
b	0.900	Calves per adult female per year	Baseline value within reported pregnancy-rate range (0.740–0.986); 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

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 One-sided proportion test for juvenile over-representation, and Keyfitz's Δ.

One-sided proportion test for juvenile over-representation: Motivated by a priori expectations of elevated early-life mortality, we conducted a targeted directional test to assess whether the observed proportion of juveniles exceeds the SSD expectation. This was implemented as a one-sample proportion test with the alternative hypothesis that $p_{\text{juvenile}} > p_{\text{SSD,juvenile}}$.

Keyfitz's ⊿: 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.

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

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.

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 ≥ 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), we used the age-determined subset (n = 122). 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 over-representation 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	Age-determined subset from Figure 2 (mixed bycatch + strandings)	217	Age-determined Figure 2 subset (n=122), age-bin lower bounds mapped to stages
Mixed (bycatch + strandings)	Murphy et al. (2020)	Celtic/Irish Sea (2000-2012)	2000-2012	Age-determined subset from Figure 2 (mixed bycatch + strandings)	172	Age-determined Figure 2 subset (n=122), age-bin lower bounds mapped to stages
Mixed (bycatch + strandings)	Murphy et al. (2020)	North Sea (1990-1999)	1990-1999	Age-determined subset from Figure 2 (mixed bycatch + strandings)	155	Age-determined Figure 2 subset (n=122), age-bin lower bounds mapped to stages
Mixed (bycatch + strandings)	Murphy et al. (2020)	North Sea (2000-2012)	2000-2012	Age-determined subset from	101	Age-determined Figure 2 subset (n=122), age-bin

Data Type	Source	Site / Stratum	Years	Unit Type	n (used)	Stage Mapping
				Figure 2 (mixed bycatch + strandings)		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 three 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.

Simulation-based plausibility checks

As a final robustness check, we asked: how often do unconstrained-but-biologically-plausible MPMs produce SSDs similar to 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 $\Delta \leq 0.10$). 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

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 18.6%, 31.3%, and 50.1% — a juvenile elevation of 26.8 percentage points and an adult deficit of 21.4 percentage points (Figure 1A). The one-sided juvenile proportion test was highly significant ($p < 0.001$), and overall divergence was substantial (Keyfitz's $\Delta = 0.268$, bootstrap 95% CI 0.259-0.277), consistent with our prediction of systematic juvenile over-representation.

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 (0.325) exceeded mean strandings Keyfitz's Δ (0.196), and the one-sided test for lower bycatch divergence was not supported ($p = 0.979$; Figures 2 and 3).

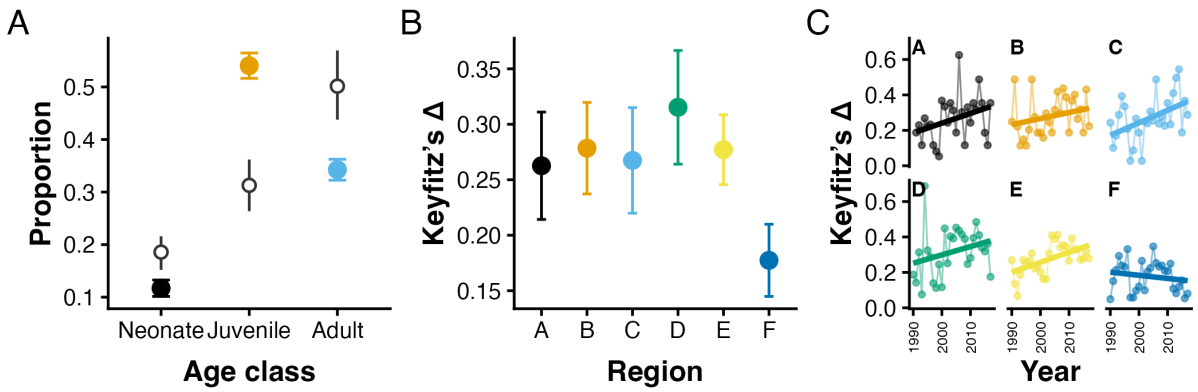

Figure 1: Stage composition in strandings differs systematically from SSD expectations, with divergence varying across regions and increasing over time in some areas. (A) Mean observed stage proportions across region-years

(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 ⊿ by strandings region (A-F), with 95% confidence intervals, summarises persistent regional differences in divergence from SSD. (C) Annual Keyfitz's ⊿ trajectories by region (facets A-F) with fitted linear trends show how SSD divergence changes through time within each region.

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 (0.315) and lowest in Region F (0.177), 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 (0.138 from lowest to highest mean), and should be interpreted alongside substantial within-region year-to-year variation (median region-specific SD = 0.120).

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 0.313 and one-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 0.252 to 0.263, and median Keyfitz's Δ from 0.251 to 0.253 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 0.309 to 0.315 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 0.261 to 0.267, with medians from 0.251 to 0.258 (Supplementary Table S8), confirming substantial mismatch even under conservative SSD assumptions.

Temporal trends in demographic bias

Keyfitz's Δ varied over time; 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 (0.003, $p = 0.191$), and Region F was the exception with a near-zero negative, non-significant slope (-0.002, $p = 0.397$; 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 0.263.

Table 3: Per-region temporal trends in SSD divergence (Keyfitz's ⊿ ~ Year). This shows which regions have increasing, decreasing, or stable divergence over time.

Region	slope	SE (slope)	p-value	R^2	n
A	0.006	0.003	0.081	0.117	27
B	0.003	0.003	0.191	0.065	28
C	0.007	0.003	0.015	0.206	28
D	0.005	0.003	0.150	0.078	28
E	0.006	0.002	0.003	0.299	28
F	-0.002	0.002	0.397	0.028	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 ($r = 0.466$) and D-C ($r = 0.449$). 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 = 0.313, one-sided $p < 0.001$), and overall Keyfitz's Δ was 0.213. In paired year-level comparisons, mean bycatch Keyfitz's Δ (0.325) exceeded mean strandings Keyfitz's Δ (0.196), and the one-sided test for lower bycatch divergence was not supported ($p = 0.979$; Figures 2 and 3). Hunting captures from Møhl-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 = 85.0), with no cells below 5, so sparse-cell corrections were not required; full chi-square results are provided in Table S9.

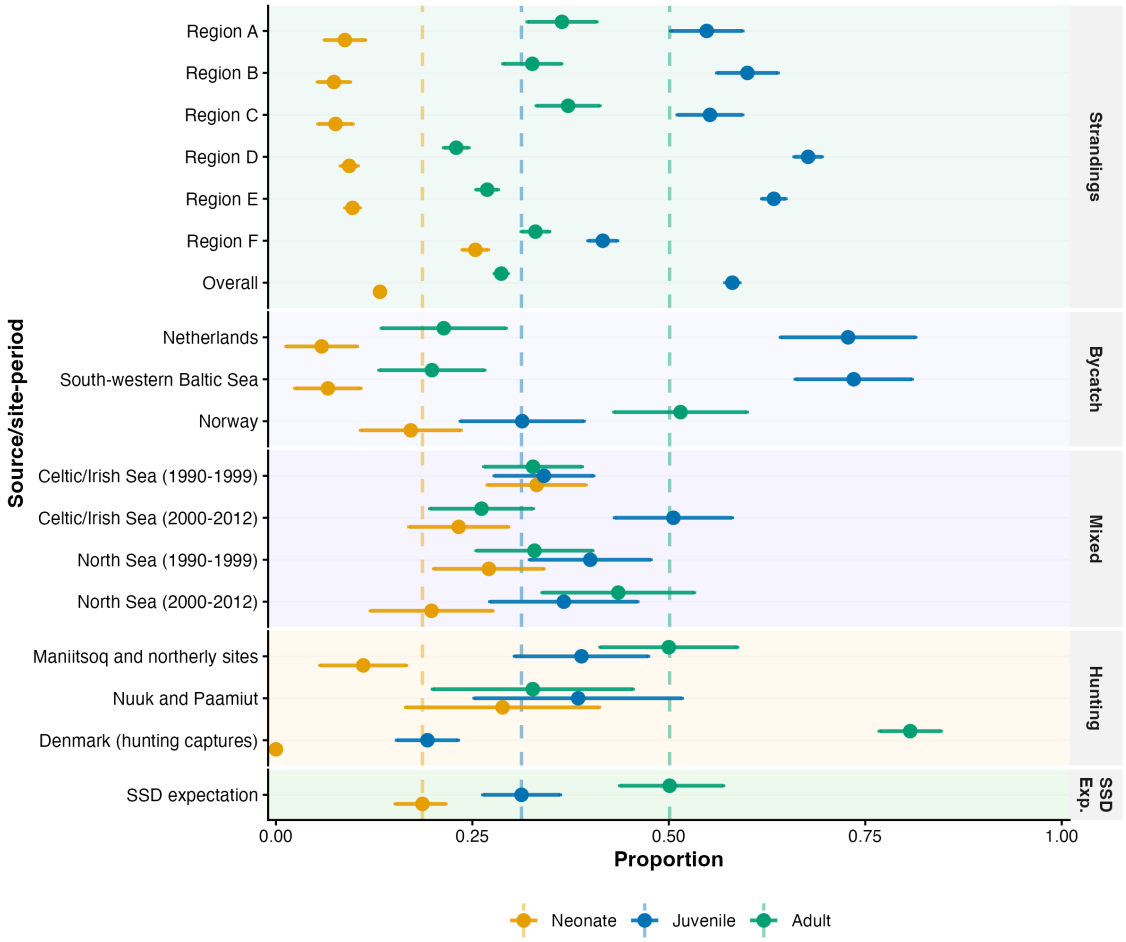

Figure 2: Strandings, bycatch, mixed-mortality, and hunting age-class proportions across sources, with SSD reference panel. Points are observed proportions by age class (Neonate/Juvenile/Adult) and horizontal error bars are 95% intervals for each source/site-period. Rows are grouped into strandings,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

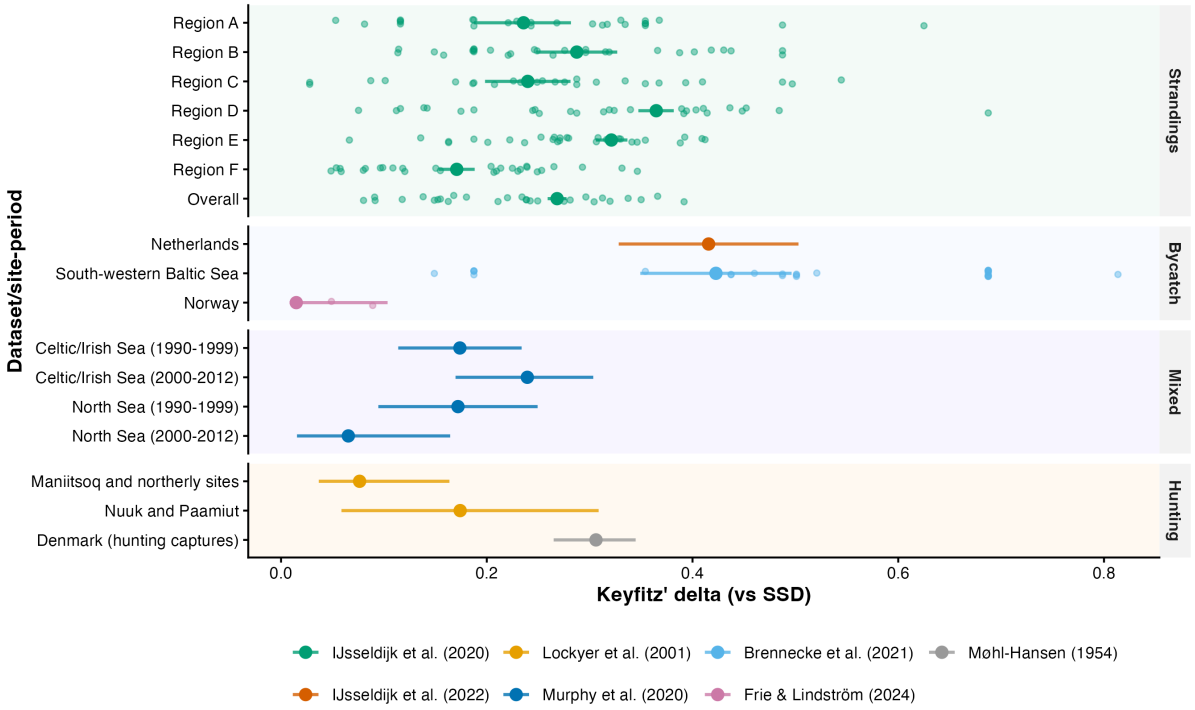

Figure 3: Combined SSD-divergence comparison across strandings, bycatch and hunting sources. Points show Keyfitz's ⊿ estimates, and horizontal error bars show bootstrap 95% intervals. Faint jittered points indicate repeated year-level values where available. Rows are grouped by data type (strandings, 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), Møhl-Hansen (1954)). Site and period labels are shown where relevant.

Simulation-based plausibility of alternative MPMs

Under unconstrained draws from biologically plausible MPM space (n = 500), only 5 matrices (1.0%) produced SSDs with Keyfitz's $\Delta \leq 0.15$, and 2 (0.4%) met the stricter Keyfitz's $\Delta \leq 0.10$ criterion. Most plausible matrices were therefore substantially farther from the observed composition (median Keyfitz's $\Delta = 0.557$, range 0.075-0.832), supporting the interpretation that observed mortality composition is not readily recovered from generic plausible demographic parameterisations (Supplementary Figure S3).

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 (0.325) exceeded mean strandings Keyfitz's Δ (0.196), with no support for lower bycatch divergence (one-sided $p = 0.979$). Both data types should therefore be treated as selectively filtered mortality samples rather than 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 26.8 percentage points, while adult proportion was 21.4 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 0.313 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 SSD baseline, an asymptotic reference derived from assumed vital rates, is misspecified. If adult survival is higher than we assume, or juvenile survival lower, the true SSD would include a higher juvenile proportion, reducing the apparent bias. Similarly, if fertility is misspecified, the asymptotic stage structure would shift accordingly. However, our sensitivity analyses using simulated MPMs with varying vital rates suggest that the qualitative conclusion that juveniles are over-represented is robust across a wide range of plausible demographic parameter space.

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 inner Danish water population of porpoises, which has undergone a seemingly drastic decline during the past decade (Owen et al. 2024). 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. First, our MPM parameterisation, while grounded in published estimates and biological realism, remains uncertain. Vital rates vary among populations, fluctuate over time in response to environmental conditions, and are inherently difficult to measure precisely in long-lived marine mammals. We addressed this uncertainty through simulation and sensitivity analyses, but we cannot rule out the possibility that our baseline SSD is biased in ways that affect our conclusions. 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.

Second, we 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.

Third, 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.

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, a pattern that is robust to multiple sensitivity analyses and alternative analytical choices. This bias has important implications for the interpretation and use of mortality data in demographic inference, risk assessment, and conservation planning.

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 harbour porpoise mortality data we use. We acknowledge the researchers and institutions who compiled and published the source datasets used here, including stranding, bycatch, mixed-mortality, and historical hunting records. This research received no specific grant funding and originated as the master's thesis of MKHH.

Data and code 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; age-determined Figure 2 subset). 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 (DOI: 10.5281/zenodo.18664348).

Supplementary information

Supplementary analyses are provided in Supplementary Materials. These 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.
- Table S3: Summary statistics for Hellinger distance across region-years showing that conclusions are concordant across distance metrics.
- Table S4: Region-year descriptive context for stranded harbour porpoises, showing that sampling intensity and temporal coverage vary across regions (A-F).
- Table S5: Juvenile bias sensitivity across filtering scenarios, showing that the juvenile over-representation signal is robust to alternative data filters.

- Table S6: Keyfitz's Δ sensitivity across filtering scenarios, showing that divergence from SSD remains substantial under sensitivity filters.
- Table S7: Juvenile bias sensitivity to SSD uncertainty bounds, showing that juvenile over-representation persists under SSD uncertainty.
- Table S8: Keyfitz's Δ sensitivity to SSD uncertainty bounds, showing that divergence magnitudes remain elevated across plausible SSD variants.
- Table S9: Chi-square goodness-of-fit results across strandings and bycatch groupings, showing strong and directionally concordant deviation from SSD expectations.

References

Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*. 2nd edn. Sinauer Associates, Sunderland, MA, USA.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2026-03-02

Michaela Kirstine Hjelm Hansen[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 & Baelt, 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:

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

where $p_{i,\text{obs}}$ and $p_{i,\text{SSD}}$ 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.

SSD uncertainty

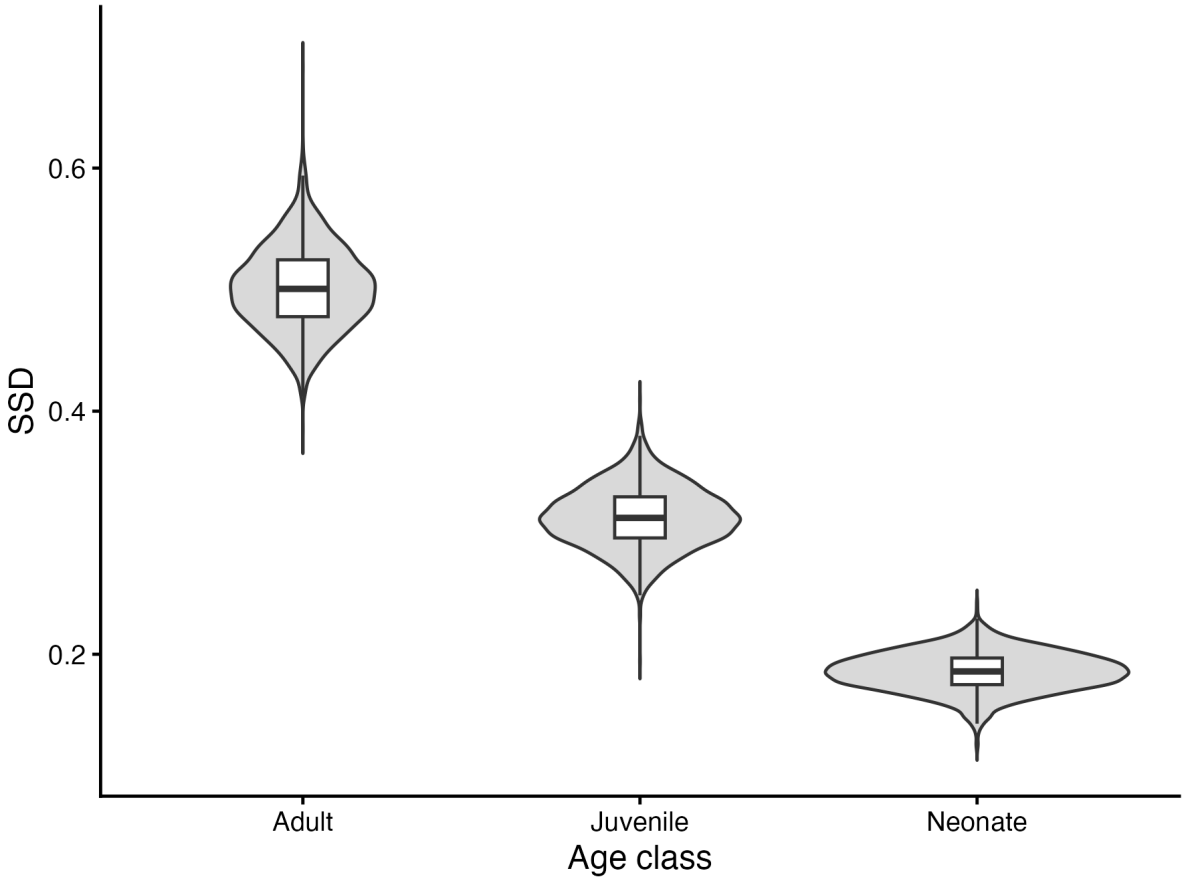

Figure S1: SSD uncertainty distribution from Monte Carlo MPM sampling. This shows that plausible SSD stage proportions vary within bounded ranges under vital-rate uncertainty, supporting robustness interpretation in the main manuscript.

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

Ageclass	n	mean	median	sd	CI low	CI high	min	max
Adult	2,000	0.502	0.501	0.034	0.438	0.569	0.385	0.683
Juvenile	2,000	0.313	0.312	0.025	0.264	0.362	0.195	0.409
Neonate	2,000	0.186	0.186	0.016	0.152	0.216	0.122	0.243

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.

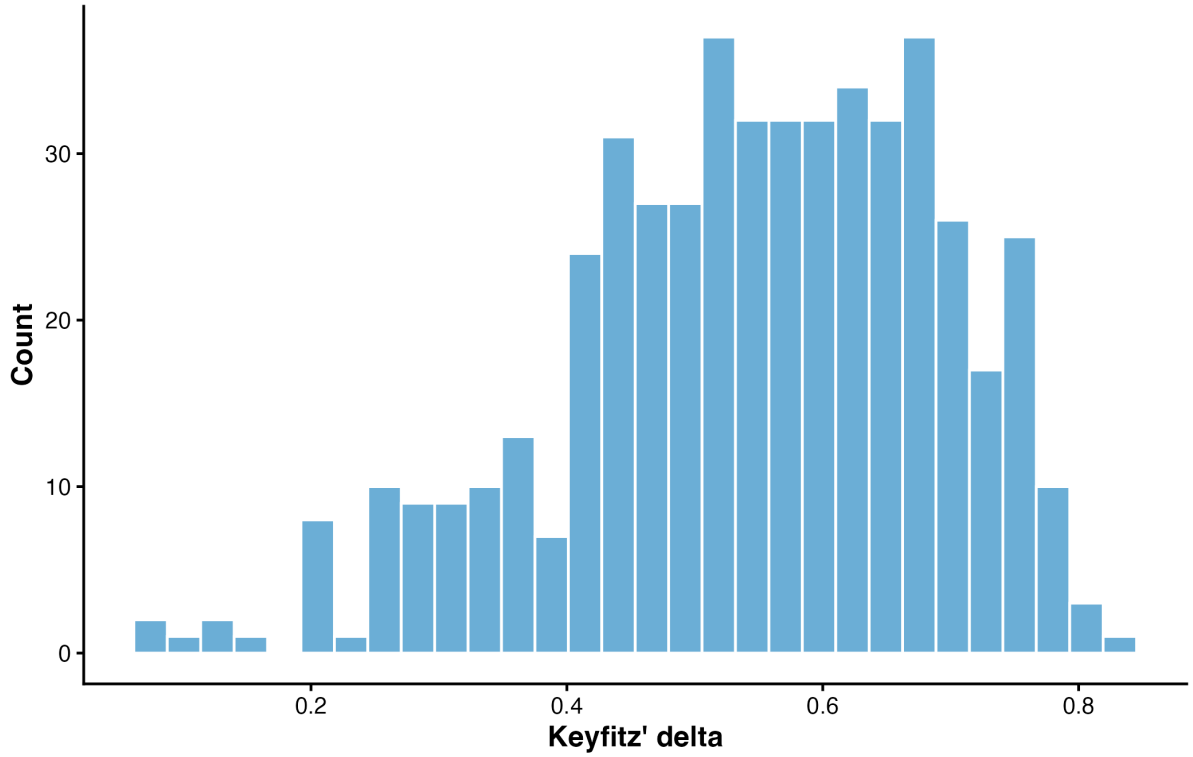

Figure S2: Distribution of Keyfitz's ⊿ across region-years, showing that SSD divergence is heterogeneous across region-years.

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

n	mean	median	sd	min	max
167	0.263	0.253	0.121	0.028	0.687

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	0.222	0.216	0.104	0.028	0.664

Additional MPM space visualisation

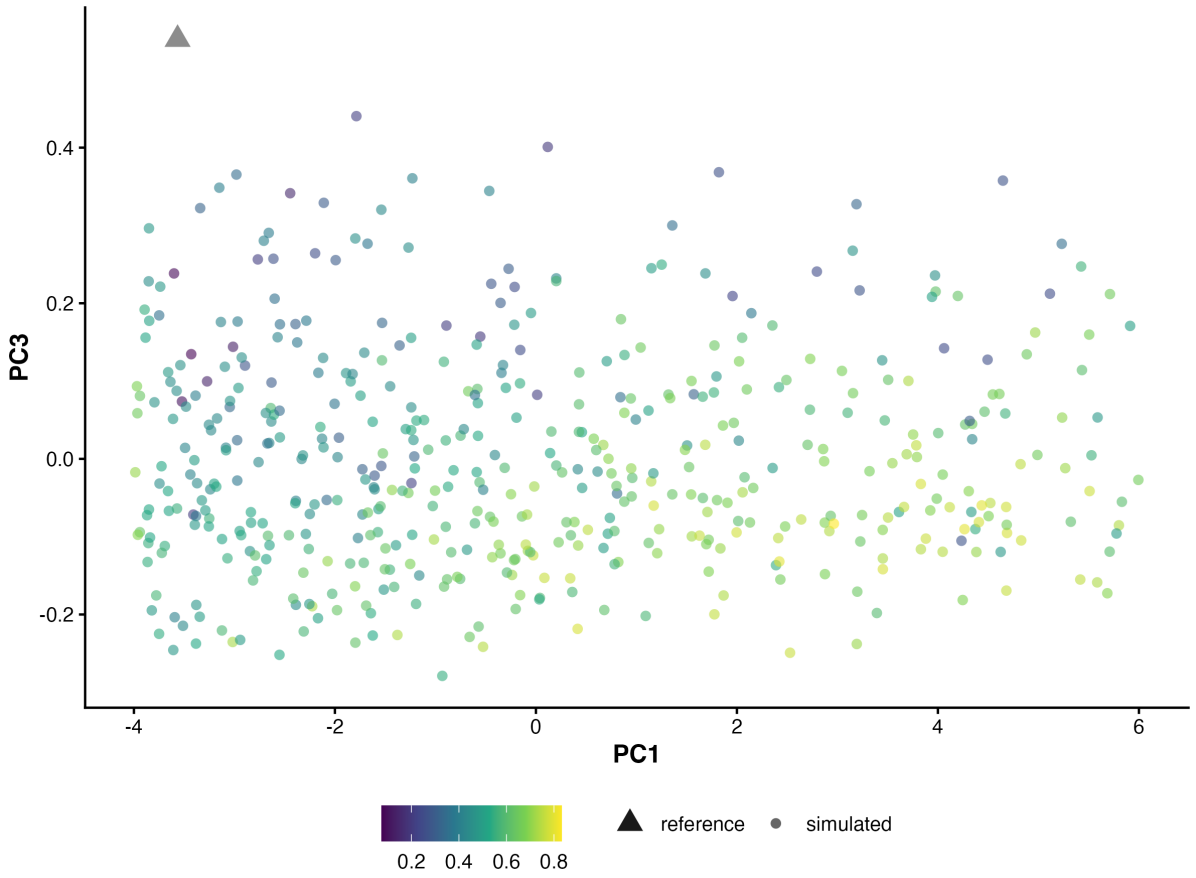

Figure S3: PCA (PC1–PC3) view of life-history transition-rate space for unconstrained biologically plausible simulated MPMs, showing the structure of sampled transition combinations in an alternative PCA projection. Point colour encodes Keyfitz's ⊿ (divergence from observed overall stage composition),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

Regional synchrony matrix

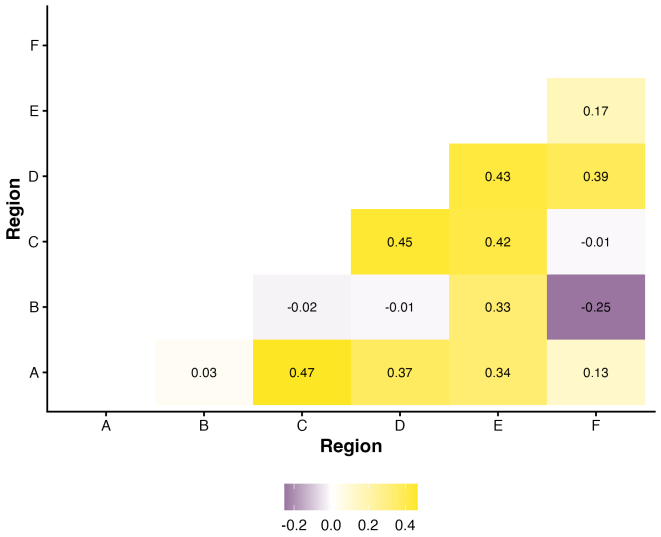

Figure S4: Pairwise correlation matrix of regional Keyfitz's ⊿ time series. Each cell value is a Pearson correlation coefficient between two regional divergence trajectories over time; higher positive values indicate stronger temporal synchrony in SSD divergence between those regions.

Descriptive stranding context by region-year

Table S4: Region-year descriptive context for stranded harbour porpoises, showing that sampling intensity and 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

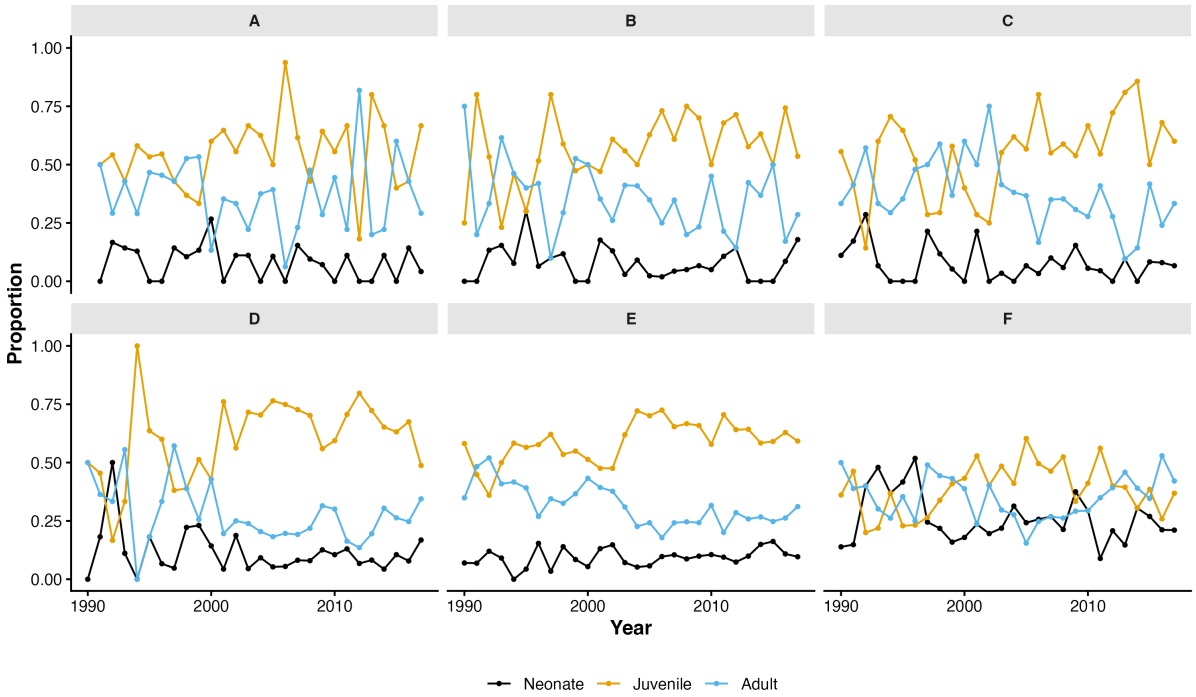

Figure S5: Strandings population structure through time by region. Each panel represents a region; points are observed annual age-class proportions and coloured lines trace temporal trajectories for neonates, juveniles, and adults, showing how stranding composition changes through time within regions.

Sensitivity analyses

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

scenario	estimate	expected	p-value	n
all region year	0.581	0.313	<0.001	10,863
complete ageclasses	0.581	0.313	<0.001	10,481
min count 10	0.581	0.313	<0.001	10,778
complete and min 10	0.581	0.313	<0.001	10,430

Table S6: Keyfitz's ⊿ 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	0.263	0.253	0.121	0.028	0.687
complete ageclasses	139	0.253	0.251	0.115	0.028	0.497
min count 10	154	0.259	0.253	0.115	0.028	0.625
complete and min 10	133	0.252	0.251	0.115	0.028	0.497

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	0.313	<0.001	10,863
ssd ci low	0.309	<0.001	10,863
ssd ci high	0.315	<0.001	10,863

Table S8: Keyfitz's ⊿ 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	0.263	0.252	0.121	0.029	0.687
ssd ci low	167	0.267	0.258	0.121	0.036	0.691
ssd ci high	167	0.261	0.251	0.120	0.030	0.685

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	3,668.977	2	0.000	10,863
Strandings by region	A	121.997	2	0.000	456
Strandings by region	B	240.050	2	0.000	610
Strandings by region	C	156.782	2	0.000	554
Strandings by region	D	1,673.877	2	0.000	2,702
Strandings by region	E	1,846.052	2	0.000	3,846
Strandings by region	F	314.779	2	0.000	2,695
Strandings by country	Belgium	774.150	2	0.000	877
Strandings by country	Denmark	208.019	2	0.000	998
Strandings by country	England	302.160	2	0.000	1,143
Strandings by country	Germany	367.872	2	0.000	1,697
Strandings by country	Netherlands	2,673.870	2	0.000	5,187
Strandings by country	Scotland	339.563	2	0.000	961
Bycatch overall	Overall	57.433	2	0.000	270
Bycatch by source	Brennecke 2021	113.172	2	0.000	136
Bycatch by source	Frie Lindstrom 2024	0.209	2	0.901	134