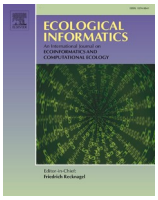
Ecological Informatics 84 (2024) 102852

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Autonomous data sampling for high-resolution spatiotemporal fish biomass estimates

Astrid A. Carlsen [a,*], Michele Casini [a,b], Francesco Masnadi [c], Olof Olsson [d], Aron Hejdström [e], Jonas Hentati-Sundberg [a]

[a] Swedish University of Agricultural Sciences, Department of Aquatic Resources, Institute of Marine Research, Uppsala, Sweden
[b] University of Bologna, Department of Biological, Geological and Environmental Sciences, Bologna, Italy
[c] Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden
[d] Stockholm Resilience Center, Stockholm University, Stockholm, Sweden
[e] Baltic Seabird Project, Sanda Västerby 568, 62379 Klintehamn, Sweden

ARTICLE INFO

Keywords:
Species distribution modelling
Remote sensing
USV
Hydro-acoustic
Spatiotemporal modelling
Small pelagic community

ABSTRACT

Many key ecological dynamics such as biomass distributions are only detectable on a fine spatiotemporal scale. Autonomous data collection with Unmanned Surface Vehicles (USV) creates new possibilities for cost efficient and high-resolution aquatic data sampling. However, the spatial coverage and sampling resolution remain uncertain due to the novelty of the technology. Further, there is no established method for analysing such fine-scale autocorrelated data without aggregation, potentially compromising data resolution. We here used a USV with an echosounder, a conductivity-temperature sensor and a flourometer to collect data from April–July 2019–2023 in a 60x80km area in the central Baltic Sea. The USV covered a total distance of 8000 nmi, over 42–81 days per year, with an average speed of 0.5 m/s. We combined the hydroacoustic data with publicly available oceanographic data from Copernicus Marine Service Information (CMSI) to describe seasonal distribution dynamics of a small pelagic fish community. Key oceanographic variables collected by the USV were correlated with CMSI estimates at daily/monthly resolution, respectively, to test for suitability to scale (Temperature 0.99/0.97; Salinity −0.77/−0.26; Chlorophyll-a 0.12/0.28). We investigated two approaches of Species Distribution Models (SDMs): generalized additive models (GAM) versus spatiotemporal generalized linear mixed effect models (GLMM). The GLMMs explained the observed data better than the GAMs (R^2 0.31 and 0.20, respectively). The addition of environmental variables increased the explanatory capability of GAM and GLMM by 25 % and ~ 3 %, respectively. Due to the high data resolution, we found significant amounts of positive autocorrelation (R: 0.05–0.30) across more than 50 sequential observations (>6 hours). However, we found that diel patterns in fish detection strongly affected the abundance estimates due to vertically migrating species hiding in the 'acoustic dead zone' near the seabed. Such dynamics could only be estimated and corrected for in predictions on the high-resolution data, complicating the trade-off between autocorrelation and high-resolution for SDMs. We compared estimates and effect sizes/directions in identical SDMs on 2x2km/month aggregated (i.e non-autocorrelated) observations and non-aggregated (i.e. autocorrelated) observations, and found relatively little difference in spatiotemporal estimates ($r = 0.80$). For the first time, we predicted the distribution of a small pelagic fish community at a high spatial resolution, in an area essential to breeding top predators, opening up for new applications in ecological studies locally and globally.

1. Introduction

The spatiotemporal distribution of organisms is key in understanding population dynamics. Such distributions are influenced by the organisms' dispersal capabilities (Chaalali et al., 2016; Pulliam, 2000), the presence of interacting species (Wisz et al., 2013) and specific habitat requirements (Guisan and Zimmermann, 2000). As such, organisms distribute in an environment based on factors such as resource

* Corresponding author.
E-mail addresses: astrid.carlsen@slu.se (A.A. Carlsen), michele.casini@slu.se, m.casini@unibo.it (M. Casini), francesco.masnadi@su.se (F. Masnadi), olof.olsson@folkuniversitetet.se (O. Olsson), Jonas.sundberg@slu.se (J. Hentati-Sundberg).

https://doi.org/10.1016/j.ecoinf.2024.102852
Received 19 April 2024; Received in revised form 7 October 2024; Accepted 7 October 2024
Available online 14 October 2024

availability (Charnov, 1976), competition (Fretwell and Calver, 1969; Křivan et al., 2008), predation (Benoit-Bird et al., 2017; Fauchald, 2009; Moody et al., 1996), life stage (La Mesa et al., 2010; Maathuis et al., 2023) and environmental conditions (Elith and Leathwick, 2009). Because animals interact with each other, disperse and migrate, reproduce and die, population numbers and spatial distributions vary through time. This variation leads to the persisting challenge of mapping species at a high temporal resolution accurately and efficiently (Giske et al., 1998; Hughes et al., 2021; J. Miller, 2010; Patterson et al., 2008; Waldock et al., 2022). However, small-scale and short-term spatiotemporal dynamics in abundance of species can be crucial for ecosystem functions (Elith and Leathwick, 2009), e.g. during reproduction for predators (Durant et al., 2005; Hilborn et al., 2017; Keogan et al., 2018).

In surveys of aquatic environments, pelagic species demography is assessed by combining measurements of hydro-acoustic backscatter with biological sampling, e.g. by pelagic trawling (Giske et al., 1998; Pennino et al., 2020; Simmonds and MacLennan, 2005). However, the main aim of these surveys is to provide estimates of the fish population indices (usually abundance or biomass) to be used in stock assessment for fisheries management (WGFAST, ICES, 2023.2). The surveys are typically short, covering extensive areas with a low spatial and temporal resolution, often without resampling within the same year due to the high running costs (e.g. vessel time, crew, instrumentation). This results in loss of details in the inherent dynamics of the population changes that may be of high ecological significance (Durant et al., 2005; Elith and Leathwick, 2009; Hilborn et al., 2017; Phillips et al., 2022; Robinson et al., 2011). In contrast, long-term continuous hydro-acoustic data is frequently sampled by moored equipment at a fixed position, providing important insights in fish ecology (e.g. Egerton et al., 2018; Kaartvedt et al., 2023; Maathuis et al., 2023), but the area covered is inherently restricted. Therefore, both large scale, short-term vessel surveys and long-term studies at fixed positions present limitations for studies where high spatiotemporal resolution is needed.

In the recent decade, various types of unmanned surface vehicle (USV) have become available for scientific data collection (Ghani et al., 2014; Swart et al., 2016), providing an alternative way to monitor large areas with high resolution over long time periods (De Robertis et al., 2021). USVs can collect a range of data types continuously, without demanding on-site crew besides during deployment and retrieval, dramatically decreasing the costs of operation (Liu et al., 2016). Further, they can often operate in areas that larger vessels cannot access, such as shallow water and fragile ecosystems (Liu et al., 2016), decreasing spatial sampling bias (Hughes et al., 2021). Equipped with several different sensors, USVs can simultaneously sample a range of environmental variables such as salinity, temperature, phytoplankton and depth of water column (Swart et al., 2016), important for generating informative spatial predictions of fish over time (Panzeri et al., 2023; Pennino et al., 2020; Rooper and Zimmermann, 2007). While large public databases (e.g. from E.U. Copernicus Marine Service Information; hereafter CMSI) aid with important information on physical and biogeochemical oceanographic variables, they mainly constitute of model estimates with coarse spatiotemporal resolution (Ghani et al., 2014; Michener, 2015). This creates an uncertainty tied to the compatibility to own resolution, and so in-situ sampled data can provide a valuable opportunity to compare.

Species distribution models (hereafter SDMs) comes in a wide range of approaches that integrate abundance and oceanographic data (Robinson et al., 2011), with the ability to predict if species are likely to occur in non-sampled locations or time periods (Panzeri et al., 2023; Pennino et al., 2020). SDMs are routinely used in Ecosystem-Based Fisheries Management, to provide valuable information on Essential Fish Habitats and Vulnerable Marine Ecosystems, as well as to inform protection and restoration strategies (Lauria et al., 2017; Panzeri et al., 2023). They also help defining stock changes (Orio et al., 2019) and habitat suitability under projected climate change scenarios (Palermino et al., 2024; Panzeri et al., 2024).

When implemented in SDMs, high-resolution data comes with a significant challenge of spatiotemporally autocorrelation (Robinson et al., 2011). For weaker correlations, adding correlation structures (i.e. for temporal and/or spatial autocorrelation) to complex SDM model frameworks (Robinson et al., 2011), or including relevant predictive variables (i.e. finer time or space variables) or order/group the correlated data in a meaningful way (Carlsen et al., 2023) can be sufficient. For stronger or longer lasting correlations, the most common approach is to simply aggregate the data (ICES, 2021c) consequentially compromising the high resolution and important ecological dynamics in it. Thus, defining meaningful ways to conserve high data resolution while producing trustworthy predictions is essential.

We here investigate the utility of USV-based collection of spatiotemporally high-resolution hydroacoustic and environmental data as a new tool for studies of ecological dynamics. For the first time, we investigate high-resolution SDMs for a community of small pelagic fish over an extended period of time and in a $\sim$ 480 km^2 area essential for vulnerable seabird species. We perform a stepwise procedure of fish distribution analyses in an attempt to conserve the high resolution of the data, and investigate the effects of autocorrelation on a spatial prediction. This paper specifically:

1) Report the coverage and utility of a USV for scientific monitoring.
2) Correlate environmental data sampled by USV with environmental variable estimates from CMSIs database.
3) Produce and compare SDMs of different complexities (i.e. model type, explanatory variables, error structures, data sources and data resolution) for a community of small pelagic fish, to investigate the gains and limitations of the information retrieved.
4) Estimate the spatiotemporal distribution and variations in biomass, detailing out the effects of each variable, and contrast two different data aggregations (i.e. the full-resolution versus 2x2km spatial aggregations within the month of the year), to evaluate the autocorrelated high-resolution model estimates.
5) Present spatiotemporal fish distribution predictions for future ecological studies.

2. Methods

2.1. Study design and general description of study site

2.1.1. Study design

We collected data using a remotely operated USV, Sailbuoy, (Offshore Sensing AS, Bergen, Norway) equipped with an echosounder, a conductivity- and temperature sensor (CT), and from 2021 on, a fluorometer (see *2.2 Scientific sensors* below). The study was performed in the Central Baltic Sea (International Council for the Exploration of the Sea, ICES, subdivision 27; statistical squares 42G7-44G7), off the island of Stora Karlsö, Sweden (Fig. 1), that hosts the largest seabird colonies in the Baltic Sea. Sampling was performed in transects that were typically sailed in bows between four virtual geolocation cursors, with a range of ~60–80 km from, and back to $\geq$2 km from Stora Karlsö (Fig. 2). The sampling transects were designed to cover the area in which the breeding seabirds at Stora Karlsö perform their foraging trips (Evans et al., 2013; Isaksson et al., 2019) from late April to late July, years 2019–2023.

2.1.2. Upper trophic levels of the study system

Island of Stora Karlsö hosts 26,000 pairs of common guillemot (*Uria aalge*), 12,000 pairs of razorbill (*Alca torda*), and 2000 pairs of great cormorant (*Phalacrocorax carbo sinensis*), amongst other species. In addition a significant population of grey seals (*Halichoerus grypus*) is present in the area (Ahlgren et al., 2022), with daily observations at Stora Karlsö in the summer months. These top predator species forage on small pelagic fish (Hentati-Sundberg et al., 2018; Kadin et al., 2012; Lundström et al., 2010) of which the three most abundant species in the

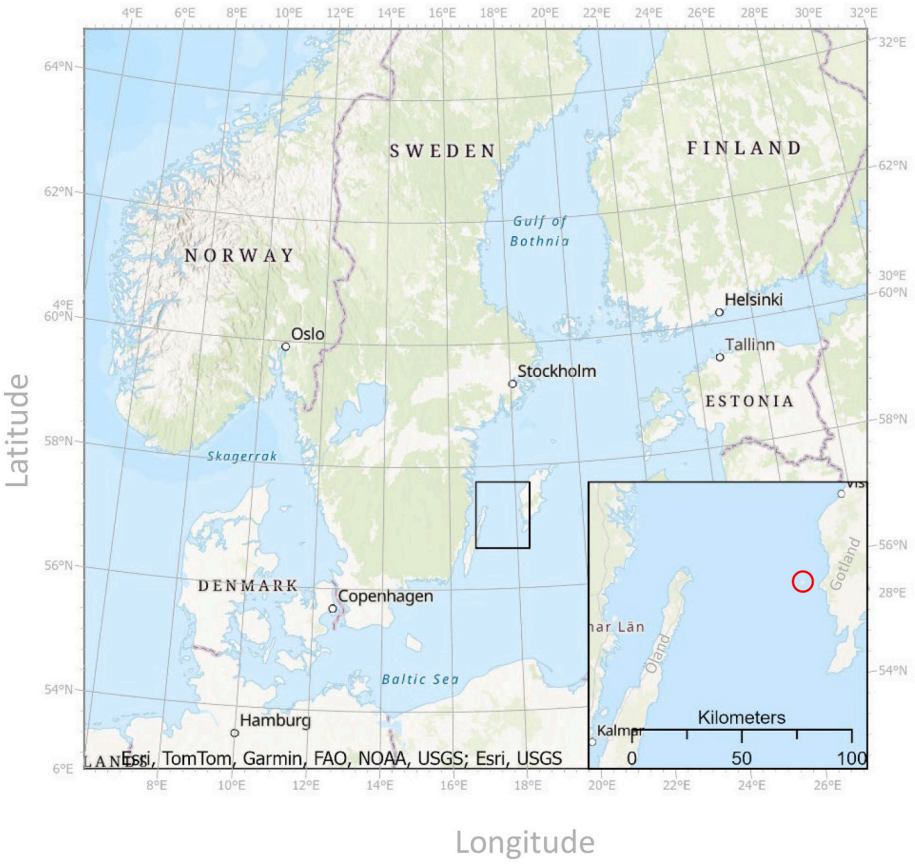


Fig. 1. Map of area and zoom in on survey area and with Stora Karlsö, 57°17′1″N 17°58′19″E (red circle). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

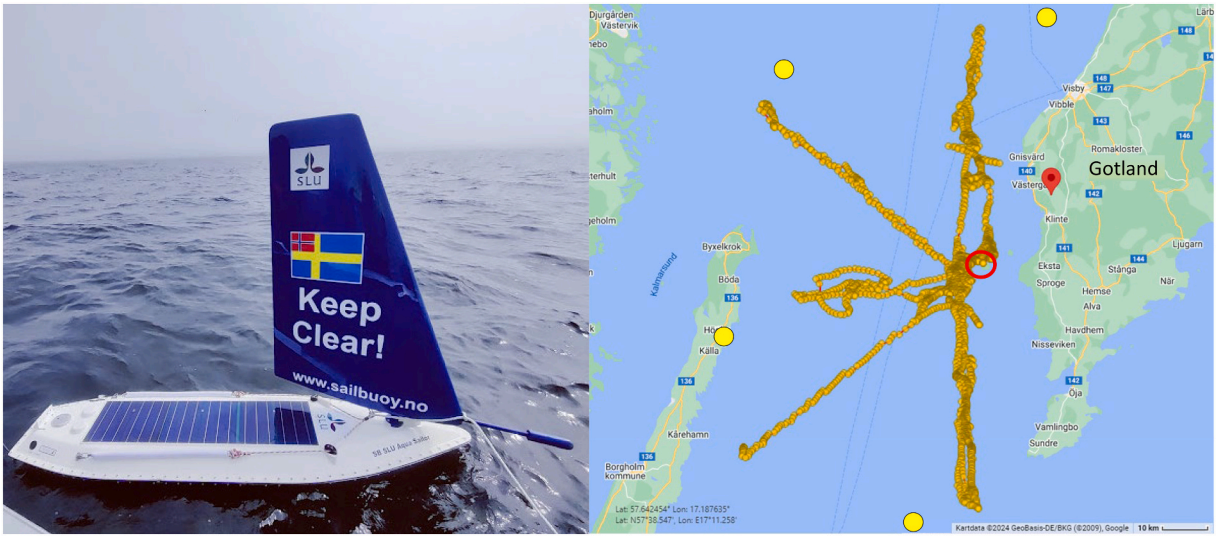


Fig. 2. The USV Sailbuoy (Offshore Sensing AS) (left) and five example sampling transects performed by the USV in May 2021 (right) from and back to Stora Karlsö (red circle). Yellow circles indicate the four virtual geolocation cursors used to define sampling transects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

study area are the two commercially sought clupeid species sprat *Sprattus sprattus* and herring *Clupea harengus,* and three-spined stickleback *Gasterosteus aculeatus*. No other fish species are nearly as numerous in the study region (ICES, 2021b). The entire Baltic Sea spawning stock biomass (SSB) was estimated to be 1,380,565 t for sprat in 2022 (ICES, 2023), and 364,981 t for herring in 2020 (ICES, 2021a). Although there is no stock assessment for sticklebacks, the relative abundance has increased drastically in recent years (Bergström et al., 2015; Olin et al., 2022; Olsson et al., 2019). All three species interchangeably utilize the area for spawning and feeding during the period of this study (Candolin et al., 2008; Jørgensen et al., 2005b; Ojaveer and Kalejs, 2010). Spring spawning herring utilize the shallow seabed areas from March to May

(Jørgensen et al., 2005a, 2005b) partially alongside sticklebacks which spawn continuously from May to July (i.e. coast-wards migration starts in April). Sprat on the other hand perform spawning migrations to and from deeper basin areas from March to July where one individual can spawn several times over the season (Aro, 1989). While the Baltic Sea also host autumn (from July) spawning herring, their population size is low and their use of this area is unknown (ICES, 2013).

2.2. Data collection

2.2.1. USV operation

The USV was 2 m long and weighed ~60 kg. It moved by sail, with an electronic rudder. Every twenty minute, the position of the sail and rudder was reassessed, sailing in one of four possible directions (i.e. 'tacks') in relation to the wind (two headwind and two downwind). The USV was primarily operated in autopilot mode, where a waypoint was given, and an algorithm based on wind direction and sailing speed automatically chose the best tack. The route was limited to a given radius from the transect as to decrease chances of sailing onto land. All remote communication with the USV was accomplished through Iridium Communication Incorporation's global satellite interface, where both steering, vessel and sensor settings could be adjusted, and data from both the vessel log and scientific sensors (see 2.2. *Scientific sensors* below) could be inspected in near real time (ten minutes delay). For more information about the USV structure, mechanical and electronic features see Wullenweber et al. (2022).

Using wind for propulsion, the USV would usually not move with wind strengths <2 m/s, leading to the decision of integrating data by time instead of distance (Ghani et al., 2014). All electronic parts of the USV were solar powered, and therefore the entire USV operation was solely driven on renewable energy.

2.2.2. Scientific sensors

Hydro-acoustic data for fish distribution estimates were collected using a Simrad Wide Band Transceiver (WBT)-mini scientific echosounder with an ES-200CDK transducer, produced by Kongsberg Maritime. The echosounder was mounted on a gimbal in the hull at 0.6 m depth. All hydroacoustic data acquisitions was planned in Simrad EK Mission Planner 3.3.x. The echosounder was run in wideband mode (frequency sweep 185–255 kHz) with a ping rate of 1/1.4 s to suit the low speed of the USV. Data was recorded to 100 m depth, as a permanent stratification (halocline) and anoxia in the study area prevents fish from distributing below 70-80 m (Weidner et al., 2020). The echosounder was calibrated using a 38.1 mm tungsten sphere, with the same settings as used in the data collection, one-two times per year in the study area following internationally standard methodology (ICES, 2021c).

Salinity and water temperature were collected with a Cabled CT Sensor produced by NBOSI, at 0.3 m depth. Fluorescence and turbidity were collected with an ECO-triplet-w fluorometer with a self-cleaning system, produced by Sea-Bird Scientific, at ~20 cm depth facing downwards. The frequencies collected were Chlorophyll-a (Chl) (470/695 nm, excitation/emission), Phycocyanin (Pc) (630/680 nm, excitation/emission), and turbidity/backscatter (700 nm). The data was collected in ten minutes acquisition loops, with one sensor at the time (order: (1) CT-sensor; (2) fluorometer, (3) echosounder continuously for 7 min, (4) full shut-off of all sensors).

2.3. Data treatment and dataset information

All raw echosounder data treatment, including calculation of calibration values, were performed in Echoview Software Pty Ltd. v. 13. Our hydroacoustic data cleaning followed a 3-step approach. In Step 1, we visually inspected echograms to familiarize ourselves with the possible NASC range of fish data. Step 2 involved generating surface (Fraser et al., 2017) and bottom exclusion lines using statistical methods, which we tested across various depths and habitats. In Step 3, we reviewed all exclusion lines visually on exported echograms and adjusted them as necessary, iterating back to Step 2 when needed. Echo integration (Simmonds and MacLennan, 2005) was performed for depth layers of 4 m over 7 min intervals, each constituting of 300 pings. The integrals were summarized as Nautical Area Scattering Coefficient (hereafter NASC), with unit $m^2\ nmi^{-2}$, which is used as a proxy for biomass in this study (see Table 1). We examined a range of the highest values in the dataset, returning to Step 1 to identify likely extreme integration cells for direct inspection, removing any cells or intervals where values were unlikely to represent fish. We set the upper NASC limit at 10,000 m^2nmi^{-2} per integrated cell (4 m depth * 10 min time range) for two reasons: (1) visual inspection of raw data often showed cells reaching 6000–8000 m^2nmi^{-2}, suggesting higher values were possible, and (2) it represented a natural cut-off in the data distribution, as detailed in Appendix A1. All further statistical analyses were performed using R versions 3.6.3 and 4.2.2 (R Core Team). The GPS-position for each data point gives the mean position of all pings within the seven minutes of sampling. All depth layers of the water column within the same time interval were cumulated, giving a dataset of NASC per mean time and position for each interval, in total 41,292 observations. For analyses on aggregated data, the mean of all observations within 2 × 2 km squares per month of the year was extracted, resulting in a dataset of 1899 observations. All dates and times were set to local time (CEST).

Data gathered from the European Union Copernicus Marine Service Information (hereby CMSI; last download 09.10.2023) was limited to a rectangle slightly exceeding size of the area sampled with USV in latitude (56.50824°-57.80822°) and longitude (16.93013°-18.23579°), while the temporal selection was limited to the earliest and latest date sampled with the USV. The datasets were downloaded in two resolutions, monthly average- and daily average estimates (Table 1), and the bathymetry was also retrieved from CMSI. CMSI model estimates for years 2022 and 2023 had not been corrected by the same standard as models from previous years at the time of this study (i.e. due to a 2 year lag in verification, see Le Traon et al., 2019), and may thus still contain occasional errors. However, the data was visually inspected for outliers, and none were identified.

2.4. Statistical analyses

Correlation tests were performed using Pearson correlation analyses (i.e. through the 'R base' function cor()) for observations and model estimates, Variance Inflation Factor ('car' function vif()) for model predictors and temporal autocorrelations assessments ('R base' function acf()), along with visual inspections (e.g. Appendix A7 and A8.3).

2.4.1. Models setup and relevant variables

Generalized additive model (GAMs; mgcv), and spatiotemporal generalized mixed effect model (GLMM; sdmTMB, see Anderson et al., 2022) with spatiotemporal fields and smooths/priors (i.e. mimicking GAMMs), were used as SDMs to model the distribution of the small pelagic fish community in the sampled area. While GAMs could be constructed with spatiotemporal random fields (i.e. through Stochastic Partial Differential Equations, or SPDE, see D. L. Miller et al., 2020), GLMM models within the framework of sdmTMB hold great advantages in explaining data with complex structures. This is especially due to the variety of options for spatiotemporal field realizations such as the possibility to define suitable spatial meshes to the specific data, and the use of separable versus non-separable time-space effects (Anderson et al., 2022). However, GAMs have advantages in being highly flexible in defining non-linear relationships without assumptions, and in explaining complex data with less information (e.g. non-parametric). Thus, the simplest models (i.e. without spatiotemporal structure) here were performed as GAMs, while more complex spatiotemporal models were fundamentally GLMMs. All models used in this study are presented in Table 2.

All models tested used 'NASC' as response variable, which was log-

Table 1
Data variable information. See *Literature* for dataset sources Anonymous a–d.

Variable	Unit	Sensitivity	Resolution	Total sensor range	Variable range	Source	Collection method	Variable status
Nautical Area Scattering Coefficient (NASC)	m^2 nmi^{-2}		Hourly mean		0.0001–8000	USV	Simrad WBT mini, 70 cm depth	Observation
Salinity	Mmhos /cm	+/− 0.005	Hourly mean	0–60	0–11.9	USV	NBOSI CT 40 cm depth	Observation, Sea surface
Temperature (Temp)	°C	+/− 0.002	Hourly mean	0–30	3.1–25.9	USV	NBOSI CT 40 cm depth	Observation, Sea surface
Chlorophyll (Chl)	µg/l	+ − 0.025	Hourly mean	0–50	0–7.3	USV	Sea Bird ECO triplet 20 cm depth	Observation, Sea surface
Salinity	PSU (~Mmhos/ cm)	−0.01	Monthly and daily on 2x2km grid		6.3–7.4	CMSI (Anonymous c, Anonymous d)	Moored monitoring stations, CTD and ferry boxes	Estimate, Sea surface
Temperature (Temp)	°C	0-5 m: −0.4 5-30 m: 0.1 30–80:m 0.3 80–200: 0.3	Monthly and daily on 2x2km grid		3.3–20.6	CMSI (Anonymous c, Anonymous d)	Moored monitoring stations, CTD, L3 satellite imagery and ferry boxes	Estimate, Sea surface
Chlorophyll (Chl)		−0.72	Monthly and daily on 2x2km grid		0.5–4.1	CMSI (Anonymous a, Anonymous b)	Moored monitoring stations, CTD, L3 satellite imagery and ferry boxes	Estimate, Sea surface
North-South currents (NS_current)	m/s	−0.08 - 0.23 m/s −17° - 38°	Monthly on 2x2km grid		−0.1-0.1	CMSI (Anonymous c, Anonymous d)	Moored monitoring stations	Estimate, weighted average of water column
East-West currents (EW_current)	m/s	−0.08 - 0.23 m/s −17° - 38°	Monthly on 2x2km grid		−0.1-0.1	CMSI (Anonymous c, Anonymous d)	Moored monitoring stations	Estimate, weighted average of water column
Current speed	m/s	−0.08 - 0.23 m/s	Monthly on 2x2km grid		0.003–0.128	CMSI (Anonymous c, Anonymous d)	Moored monitoring stations	Estimate
Depth of seabed (Depth)	meter	−749 - 0	2x2km grid		−166.4 - 0	CMSI (Anonymous a)		Observation
Week	Julian week				16–31	USV		Observation
Hour	Cyclic 24 h				00–23	USV		Observation
Year	Julian year				2019–2023	USV		Observation
Month	Julian month				4–7	USV		Observation
X,Y	UTM coordinates, CRS 33 N				X: 619.4–697.1 Y: 6290–6311	Calculated from USV latitude, longitude		Observation

transformed due to its wide range in values and their near Poisson distribution (see fig in Appendix A1). Explanatory variables were chosen based on literature for fish distribution (e.g. Aro, 1989; Cardinale et al., 2003; Giske et al., 1998; Maravelias & Reid, 1995; Pennino et al., 2020; Sabatini et al., 2004; Schaeffer et al., 2008; Watson et al., 2015) and availability of environmental variables from the area and time of the study (Table 1). While temperature (Nøttestad et al., 2007), salinity and chlorophyll are established variables for determining abundance in SDMs of many aquatic species, hour of the day was expected to affect the detection rather than the true abundance (Cardinale et al., 2003; Mello and Rose, 2009). Both 'Week' and 'Month' was included for different scales of seasonal dynamics, as they were not correlated (i.e. as tested by Variance Inflation Factor, VIF = 0.03) while improved the models. Water currents were included as they can affect availability of food for planktivores (Sabatini et al., 2004; Schaeffer et al., 2008; Watson et al., 2015), organisms metabolic cost and movement (Maathuis et al., 2023; Simmonds and MacLennan, 2005). Currents can also affect organisms distance to the seabed /topographic structures (Maravelias, 1999), along with the stratification of the water- and thus the detectability by echosonar (Mello and Rose, 2009). The effect of, 'depth' (i.e. of the seabed) is likely to interact with the time of day, as benthic hiding/acoustic dead-zone is primarily a problem when depth of the seabed is shallower than the anoxic zone. However, to reduce model complexity and avoid over-parametrization we chose 'hour' to detail out the cyclic pattern in fish detection.

The GLMMs were fitted on a spatial mesh based on the coordinates of observations across all years. The temporal field of the model was by months as the entire survey area was aimed to be resampled once per month. In addition, all environmental variables from CMSI were monthly averages in fish biomass models. The variable Julian 'week' was added to capture finer scale changes in fish biomass due for example to short-term weather patterns and behaviour in general. Models were initially tested against simplified models in terms of fixed and random effects included, and smooth structures, where the models with best log likelihood (LL) were selected for further analyses. The final model evaluation was based on the statistical estimates of log likelihood, maximum absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2) of the highest ranked models. Best models were selected based on explanatory ability through cross validations of 70/30 % training/test data (see *Results*). Autocorrelation structures of models were checked with the function *acf()* (Appendix A7). Explanatory variables were checked for correlation using R base function *cor()* (see Appendix A3) and variables representing the same environmental variable (e.g. 'Temperature' from the USV versus CMSI) were never present in the same model. While water column averages of CMSI estimates for temperature and chlorophyll explained the fish distribution better than surface estimates (Appendix A3), we proceeded with the surface estimates due to the wish to correlate the variables against the observations made by the USV (i.e. which was only able to collect data at the surface). Residuals and goodness of fit were visually inspected, and several distribution families for log NASC were tested before student distribution with 4 degrees of freedom was selected as the

Table 2
Numbered overview of all the models. The smoothers were b-spline 's(…)', cyclic s(…,bs='cc'), random intercept (1|…). For spatiotemporal and spatial models, the random fields are given in [brackets], and the correlation structure (i.e. Autoregression = AR1; Identical and independent = IID) is indicated.

Model name	Model formula/structure	Rationale	Model type
Model 1	*log_NASC ~ s(Week) + s (Hour, bs = 'cc') + s(Depth) + Year*	Baseline GAM model	GAM
Model 2	*log_NASC ~ s(Week) + s (Hour, bs = 'cc') + s(Depth) + s(NS_current) + EW_current + Current_speed + s(Temp) + s(Chl) + s(Salinity) + Year*	Model 1 with CMSI environmental variables	GAM
Model 3	*log_NASC ~ s(Week) + s (Hour, bs = 'cc') + s(Depth) + (1|Year); [Month, X/Y]; Correlation structure AR1;*	Baseline GLMM model	GLMM
Model 4	*log_NASC ~ s(Week) + s (Hour, bs = 'cc') + s(Depth) + s(NS_current) + EW_current + Current_speed + s(Temp) + s(Chl) + s(Salinity) + (1| Year); [Month, X/Y]; Correlation structure AR1;*	Model 3 with CMSI environmental variables	GLMM
Model 5	*log_NASC ~ s(Week) + s (Hour, bs = 'cc') + s(Depth) + s(NS_current) + EW_current + Current_speed + s(Temp) + s(Chl) + s(Salinity) + (1| Year); [Month, X/Y]; Correlation structure AR1;*	*Model 4 with 'Temperature' and 'chlorophyll' from USV, years 2021–2023*	GLMM
Model 6	*log_NASC ~ s(Week) + s (Hour, bs = 'cc') + s(Depth) + s(NS_current) + EW_current + Current_speed + s(Temp) + s(Chl) + s(Salinity) + (1| Year); [Month, X/Y]; Correlation structure AR1;*	*Model 4 on years 2021–2023*	GLMM
Model 7	*log_NASC ~ s(Depth) + NS_current + EW_current + Current_speed + s(Temp) + s (Chl) + s(Salinity) + (1|Year) + (1|Month); [X/Y]; Correlation structure IID;*	*Spatial Model 4 without 'Week' and 'Hour' for 2X2km data aggregation*	GLMM
Model 8	*log_NASC ~ s(Depth) + s (NS_current) + EW_current + Current_speed + s(Temp) + s (Chl) + s(Salinity) + (1| Year); [Month, X/Y]; Correlation structure AR1;*	*Model 4 without 'Hour' and 'Week', for comparison with Model 7*	GLMM
Model 9	*log_NASC ~ s(Hour, bs = 'cc') + s(Depth) + s(NS_current) + EW_current + Current_speed + s(Temp) + s(Chl) + s (Salinity) + (1|Year); [Month, X/Y]; Correlation structure AR1;*	Model 4 without 'Week' for simplified time structure in prediction	GLMM

best available option (see Appendix A1). Space-time was separable for all full-data GLMMs through 'autoregressive' correlation structure (AR1), but inseparable for the aggregated data GLMM through 'identical and individual distributed' autocorrelation (IID) (Lindgren et al., 2023). The decisions were based on residual fit and the time and space span between observations (i.e. ≥ 10 min in full-data and 1 month in the aggregated data).

2.4.2. Analyses workflow

First, we tested how well the fish distribution in 2019–2023 could be estimated by two different model types: a GAM and a spatiotemporal GLMM (Table 2). Two different complexities were investigated, with and without additional environmental variables (GAM Models 1 and 3 and GLMM Models 2 & 4 respectively), to see if information on monthly average of a set of environmental parameters (i.e. CMSI provided salinity, temperature, chlorophyll-a and current direction/speed) could improve the two different model types.

The discrepancy between USV observations and CMSI model estimates of temperature, salinity and chlorophyll-a was evaluated in two ways. First, a correlation test, with CMSI estimates versus USV observations were performed between the high-resolution daily averages of CMSI estimates to the USV observations, contrasted with low-resolution monthly averages of CMSI estimates. As the year 2021 was the only year with both Chlorophyll data from the USV and verified CMSI daily estimates (Le Traon et al., 2019), the fine resolution correlation was based solely on this year, but with cross references to the correlations of monthly estimates to all available years (2019–2023). Second, the significance of the discrepancy was tested through model performances in two contrasting GLMMs using USV (Model 5) and CMSI (Model 6) sea-surface variables respectively (monthly CMSI estimates), with the baseline structure as provided below. The variables compared in the GLMMs were temperature and chlorophyll-a values from the years 2021 to 2023. As the local salinity measures from the USV deviated significantly from the CMSI estimates (and also from the Swedish Metrological and Hydrological Institute data, see Results), we chose to proceed with CMSI values for salinity also in Model 5.

To address the potential effects of autocorrelation, we compared the model estimates from a simplified spatiotemporal GLMM (Model 8) with a model on spatiotemporally aggregated data, using mean values for 2x2km by month/year (Model 7), to verify direction and sizes of variable effects on a non-correlated dataset. The only difference between the models were the correlation structure (see Table 2).

The final prediction of biomass distribution was produced using Model 9, where 'week' was selected out for a simplified time structure in the prediction. The predictive grid was constituted of CMSI values for a slightly increased area (as compared to observations). The chosen grid value for 'hour' was 01 (CEST) to utilize the maximum biomass estimates during diel vertical migrations (Mello and Rose, 2009).

3. Results

3.1. Data collections

The number of operating days per year for the USV were 42 days in 2019; 81 days in 2020; 67 days in 2021; 81 days in 2022; 52 days in 2023 (Fig. 3a). The average sailing speed of the USV in the Baltic Sea was 0.55 ms^{-1}; 1.1 knots (Range: 0-3 ms^{-1}, see Appendix A4) and the total distance sailed was ~8000 nmi (14,752 km) across all five years. The shallowest areas the USV was operated in was <2 m deep and the deepest were > 150 m. Each round-trip, from Stora Karlsö to the outer edge of the study area and back, typically took 2–10 days, and thus, approximately five such trips were sampled each month. The USV had a running cost of 100 SEK per day, and a purchase cost of 1,750,000 SEK (in 2019), leading to an overall operation cost of ~5500 SEK per day. The calculation does not include piloting, which typically took 5–15 min per day, and 4 h per deployment/retrieval for one person (on average 6 per season), including calibrations.

The log NASC values observed ranged from: lower extreme −2.32, 1st quartile 3.49, median 4.26, mean 4.31, 3rd quartile 5.04 and upper extreme 9.63 (Fig. 4). These observations include all pelagic species, and are exemplified as echograms in Fig. 4. The lower extreme essentially represent no fish detected, and 1st quartile and mean shows small and medium dense fish aggregations <50 m depth. The 3rd quartile shows two layers of fish aggregations around 10-20 m and 60 m depth respectively, and the upper extreme shows one extremely dense layer of fish around 50-60 m depth.

3.2. Surface environment variables in pelagic fish biomass/distribution

The correlation tests between USV observations and CMSI estimates

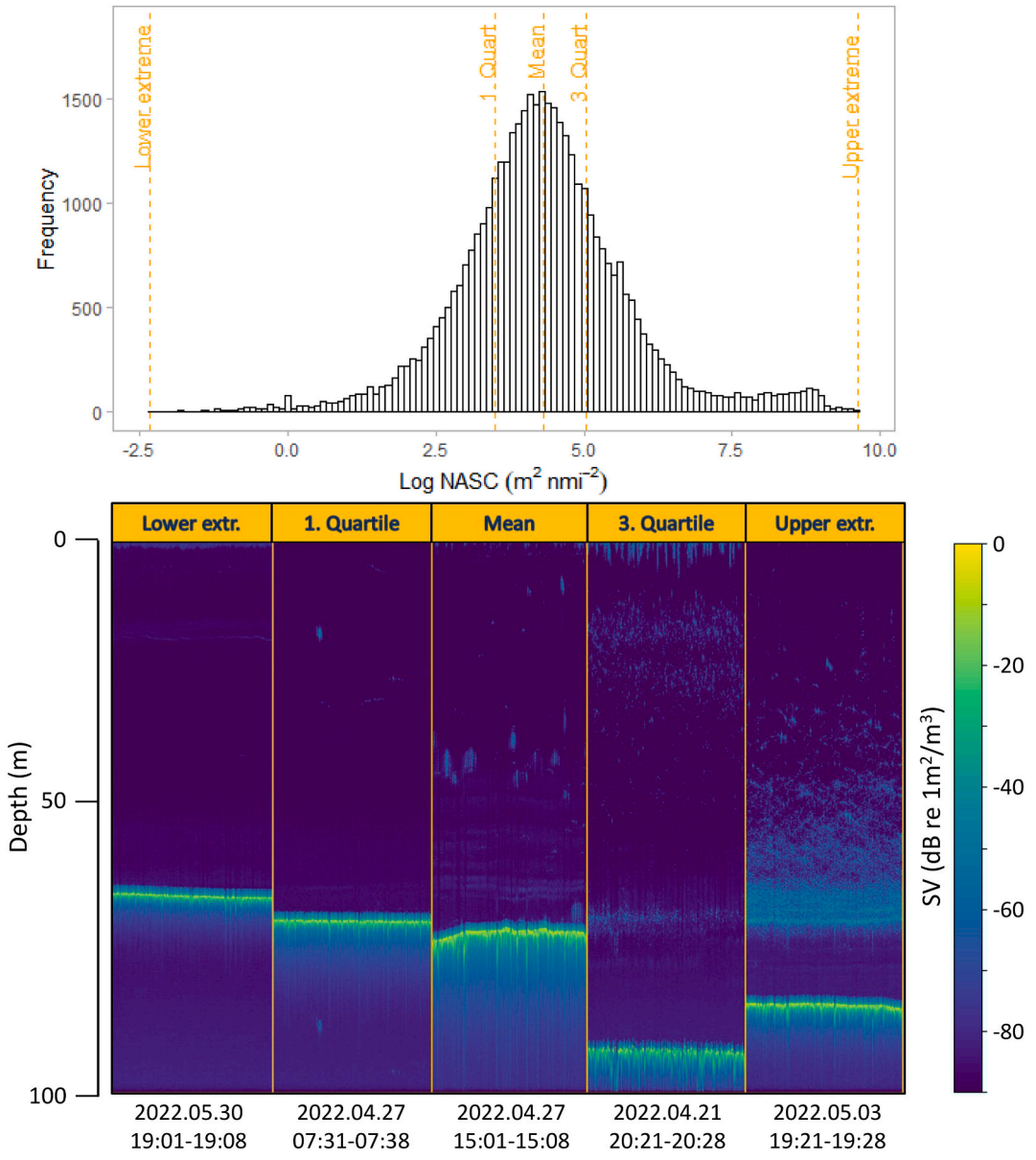


Fig. 3. Number of acoustic observations by the USV per Julian day in each year of operation.

shows that daily averages of temperature (Table 3 and Fig. 5a) were highly correlated between the USV and CMSI ($R = 0.99$), while the monthly averages still represent the observations well ($R = 0.92$) (Table 3 and Fig. 5d). There was a rather low correlation in chlorophyll-a values, due to a mismatch in the estimated versus observed peak bloom timing and level in June/July (Fig. 5c). The monthly average estimates from CMSI (Fig. 5f), which included two more years (2021–2023 versus 2021), was more correlated to the USV observations than the daily averages ($R = 0.28$ versus 0.12). Especially the correlation in salinity data stood out (Table 4 and Fig. 5b and e), with a strongly negative correlation both in the monthly average and especially in the daily average of CMSI data against USV observations ($R = -0.26$ and -0.77 respectively). The difference increased over the season. With reference to observations from an SMHI-station outside of Stora Karlsö (BY38) we found that the error was tied to the USVs measurements rather than the CMSI estimate (see Appendix A5), where the USVs CT-sensor had a consistent time drift over season, reoccurring in all years (see Appendix A5.2). The discrepancy between the two high-resolution data sources (daily average of CMSI estimates versus USV observations, 2021) is shown in Fig. 5.

When testing for the effects of monthly CMSI estimates in SDMs, we found that the GAM was improved by about 25 % (R^2: 0.16 for Model 1 versus 0.20 for Model 2), while the corresponding spatiotemporal GLMM was improved only by <3 % (R^2: 0.305 for Model 3 versus 0.313 for Model 4) upon inclusion of these three variables (see Table 4). However, the GLMM (Model 4) gave an overall better fit, as determined through LL, MAE and RMSE (Table 4), and were deemed important also in GLMM (Model 4) regardless of the variable effect sizes and significance. When contrasting the model effects of CMSI monthly averages (Model 5) with identical models based on USV collected versions of the environmental surface variables we found that the models' abilities to estimate the observations differed very little (R^2: 0.380 for CMSI, Model 5, versus 0.383 for USV, Model 6; <1 % difference). While R^2 was higher for the model with USV observations, all other metrics were slightly better for the CMSI based model.

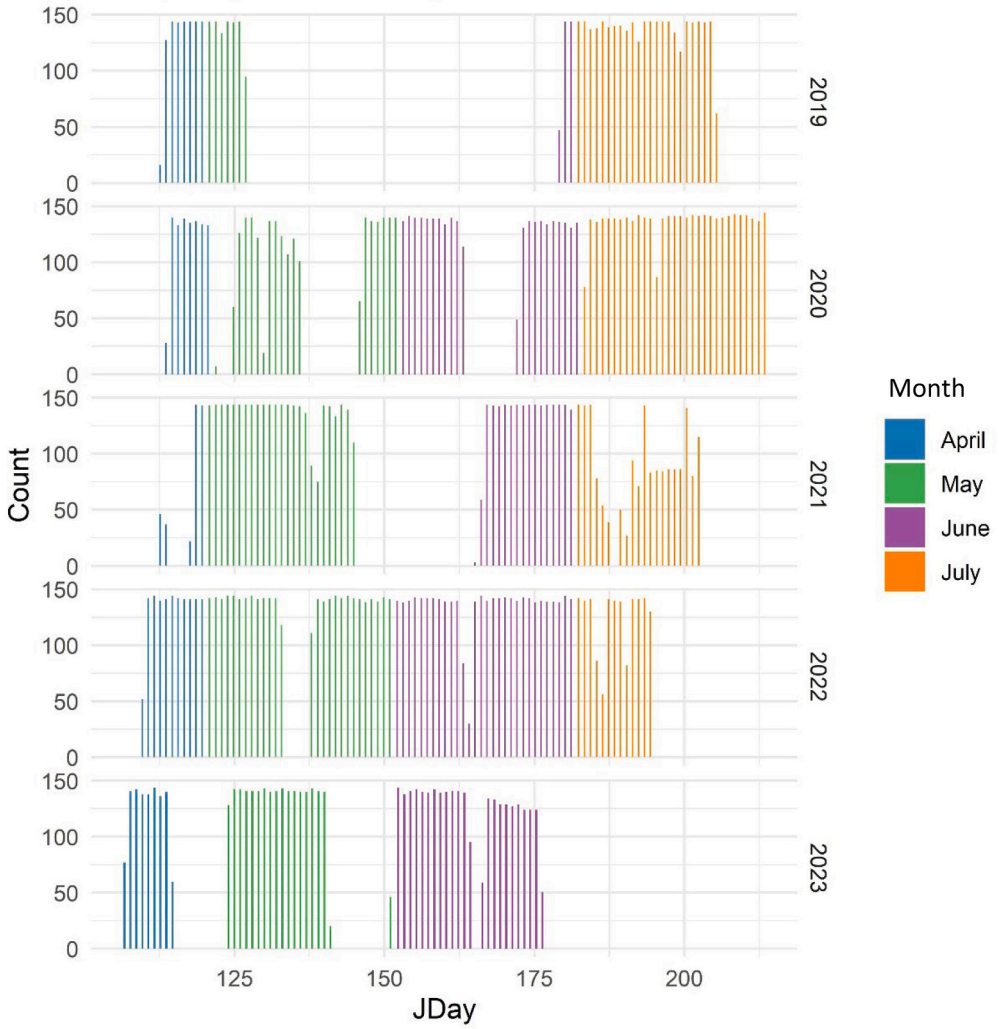


Fig. 4. Absolute distribution of Log NASC (m^2 nmi^{-2}) values (above) and example echograms of fish aggregations (below) at lower extreme (−2.32), 1st quartile (3.49), mean (4.31), 3rd quartile (5.04) and upper extreme (9.63). Strong lower echo line reflects the bottom topography, and in 3rd quartile there was surface turbulence; these echo's are not included in the calculations.

Table 3
Correlation between sea surface values from CMSI versus USV, with two CMSI data resolutions: daily averages from 2021 and monthly averages from all years of available USV data. See Fig. 5a-c of daily estimates versus observations, and Fig. 5d-f of the discrepancy between USV and CMSI on a monthly resolution.

CMSI resolution	Sea surface variables	Correlation (R)	Years
Daily	Temperature	0.99	2021
Monthly	Temperature	0.92	2019–2023
Daily	Salinity	−0.77	2021
Monthly	Salinity	−0.26	2019–2023
Daily	Chlorophyll-a	0.12	2021
Monthly	Chlorophyll-a	0.28	2021–2023

3.3. Spatiotemporal biomass estimates and prediction

3.3.1. Final model estimates

The final spatiotemporal fish distribution estimate was based on Model 4 (Fig. 6). The predictor 'week' (Fig. 6a) detailed out the general pattern observed over the season where fish biomasses were highest at the beginning and end of the season, but with significantly lower levels in weeks 20–25, from mid-May to mid-June. The effects of 'hour' (Fig. 6b) and 'depth of seabed' (Fig. 6c) on fish biomass was the most stable across models in terms of effect sizes and regression (Appendix A8). 'Hour' showed a distinct cyclic pattern of higher levels detected during dusk/dawn/night hours (19–05 CEST) than during the day (05–19 CEST). Both temperature (Fig. 6d) and salinity (Fig. 6e) had overall positive effects on fish biomass, but note that temperature in particular was very even across the area (Appendix A13). Chlorophyll (Fig. 6f) did not have a clear effect on biomass, returning an undulating regression. We estimated more fish during stronger currents from the north (Fig. 6g), south and west (Fig. 6h), but lower total current speed (Fig. 6i) seemed to overall increase fish detections. Note that lower current speeds were the dominant type in this area with mean; 0.04 m/s; range: 0.003–0.13 m/s (See Appendix A14). The models log NASC estimates (mean: 4.23, range: −0.69–7.12) fitted well with the observations (mean: 4.31, range: −2.32 - 9.63), though with a slight underestimation and a narrower distribution (Fig. 6j).

3.3.2. Autocorrelation and data aggregation

As expected, model 4 returned a consistent autocorrelation across >50 lags (Appendix A7), starting at $R = 0.30$, reduced to <0.20 after 4 lags (i.e. 4 × 10 minutes). To test the robustness of the variable effect directions and sizes, a spatial model on aggregated data (Model 7; identical to model 4 without the fine-scale time variables 'Week' and 'Hour') was compared to a full data version (Model 8), which returned R^2's of 0.310 and 0.364, respectively. The aggregated data resulted in much improved residual fit (Appendix A1), no autocorrelation across

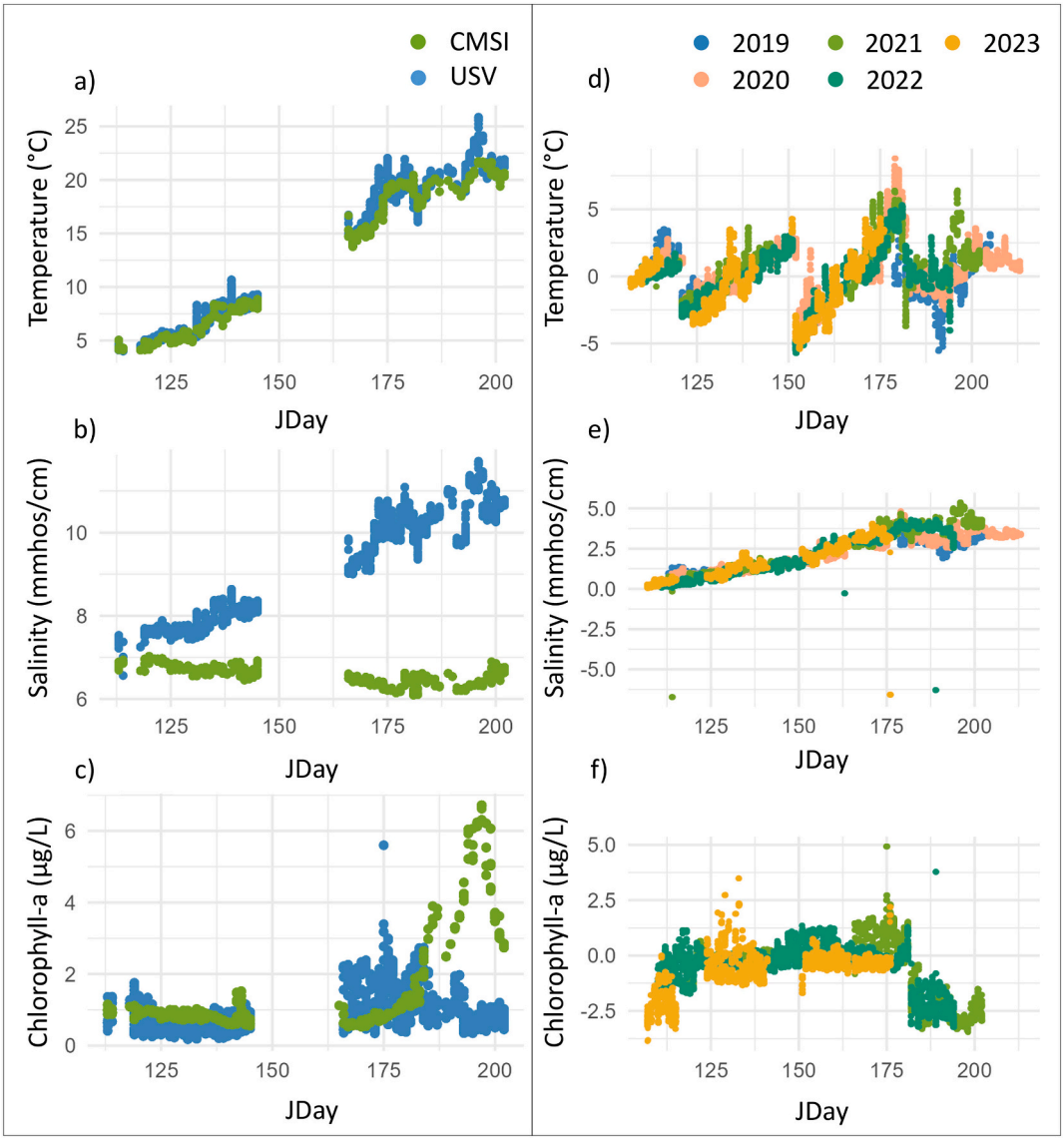


Fig. 5. Seasonal trends in daily surface environmental estimates from CMSI (green) versus in situ observations by the USV (blue) in 2021, for a) Temperature, b) salinity and c) chlorophyll-a. Further, the discrepancy between CMSI monthly surface environmental estimates subtracted from in-situ observations by the USV for all years (2019–2023), for d) Temperature, e) salinity and f) chlorophyll-a. See Appendix A5 for more information. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Model statistics for goodness of fit by Log Likelihood, coefficient of correlation (R^2), maximum absolute error (MAE) and root mean square error (RMSE). Cross validation of 70/30 % training/test data, values mean of 5 reiterations. See all model structures in Table 2. Datasets refer to foundational data for comparability of log likelihood: A = full data; B = subset of 2021–2023; C = 2x2km aggregated NASC values.

Model name	Log Likelihood	R^2	MAE	RMSE	Datasets	SDM type
Model 1	−52,625.3	0.164	0.928	1.325	A	GAM
Model 2	−51,802.9	0.204	0.897	1.292	A	GAM
Model 3	−45,988	0.305	0.783	1.210	A	GLMM
Model 4	−45,673.6	0.313	0.775	1.204	A	GLMM
Model 5	−28,260.5	0.380	0.784	1.204	B	GLMM
Model 6	−28,269.3	0.383	0.783	1.201	B	GLMM
Model 7	−2015.2	0.364	0.466	0.645	C	GLMM
Model 8	−45,845.6	0.310	0.780	1.208	A	GLMM
Model 9	−46,587.0	0.292	0.798	1.222	A	GLMM

observations (Appendix A7), and yet it showed similar variable interpretations as the non-aggregated data model in terms of direction and range of effects (Appendix A7). All effects sizes were higher in the aggregated data model (Model 8). The predicted biomass estimates had a mean of 4.38 log NASC, and the range 2.76 to 6.05 log NASC, which matched the observations (mean: 4.38, range: 1.51 to 7.27) well, though the estimated range was much narrower than observed (Appendix A8). In comparison, the corresponding values from the non-aggregated data model (Model 8) were: mean 4.22 log NASC, range − 0.60 to 7.61 log NASC, when the non-aggregated dataset had a mean of 5.04 log NASC, and a range of −1.32 to 9.63 log NASC.

When predicted on a spatiotemporal grid, the mean estimate of Model 7 was 4.46, ranging from 2.43 to 6.88 (Appendix A8). Model 8 had a mean of 4.81 log NASC, and range of −0.54 to 8.77, returning a much wider range corresponding better to the fine scale observations (Appendix A8). The two predictions from Model 7 and 8 had a correlation of $r = 0.80$. When inspecting the direction and sizes of effects, Model 7 came out with slightly stronger effects and larger ranges than Model 8, but overall with very similar directions of effects for all

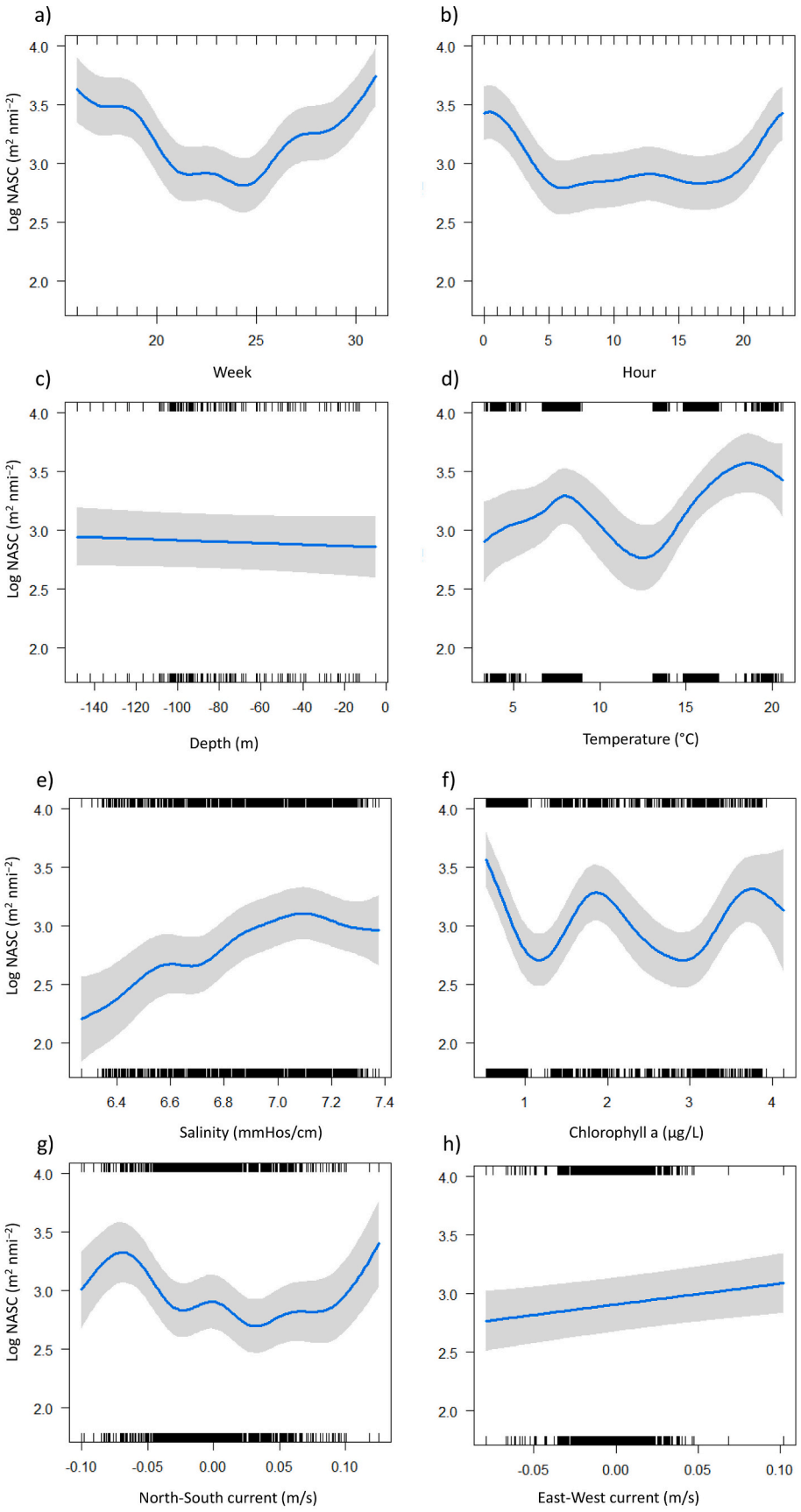


Fig. 6. Effect sizes of each explanatory variable on log NASC, from spatiotemporal GLMM with environmental variables from CMSI monthly averages (Model 4). R^2 = 0.313. See Appendix A6 for model summary.

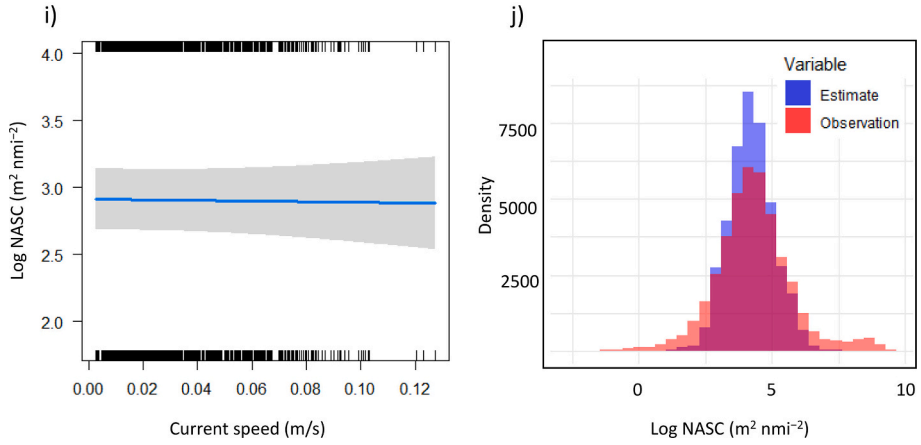


Fig. 6. (*continued*).

explanatory variables (Appendix A8). However, as shown in Fig. 6, the hour of the day strongly affects the detection of pelagic fish, and > 30 % of the observations were made during daytime (between hour 04–21) and at depths shallow enough (<80 m) to expect acoustic dead zones (50 % of the data was sampled in areas shallower than 100 m during daytime). This means high risk of under-estimation of true biomass values that cannot be corrected for under a grid prediction. Thus, we here continue with the non-aggregated data for a spatiotemporal grid prediction (see Appendix A9 for aggregated data grid prediction and estimate discrepancies between Model 7 and Model 8).

3.3.3. *Spatial prediction*

When the final fish distributions were predicted on a grid (Fig. 7), 'Week' was excluded to simplify the time structure (Model 9). The highest biomasses were detected in shallow areas (e.g. close to Gotland and Öland, ref. Fig. 1), but not strictly. Some areas, like north-east of Öland consistently had higher levels of fish than others. There were significant seasonal patterns of change across months, as reflected on a finer scale in the predictor 'week' from Model 4, where overall higher levels of fish were found in the beginning and end of the season, with an estimated dip in June. The model predicts very similar trends in months across years (see Appendix 13 for environmental variable trends), where the same general areas were predicted to produce upper quantile levels of fish (see Appendix A14 for distribution of observed upper quantiles). Strongest of all, however, were the 'year' effects, where years with high biomass such as 2023 contrasts with poorer years, like 2022 (See Appendix A1 for year values). As no data was gathered in or after July 2023 we have chosen not to include this month in the prediction (see Appendix A14).

4. Discussion

We here show how we collected and analysed high-resolution spatiotemporal hydroacoustic data for estimating fish biomass distributions at a community level over time. The area selected and the speed of the USV facilitated the possibility to cover a large and diverse area with high revisiting rate across the season each year, which gave a solid foundation for describing trends in pelagic fish distribution and amounts. Further, it enhanced the possibilities to predict also in non-sampled areas and times, and even to estimate changes in detected fish abundance on a scale as fine as by hours. The details picked up in the hydroacoustic data were on a remarkably fine scale (as shown in Fig. 3), from 1 to 100 m depths, with next to no noise due to the low speed and lack of propulsion from the USV (DuFour et al., 2021). In all, the cost effectiveness of the observations, and the extension in space and shallow water depths possible to monitor provide great opportunities for ecological studies on yet under-sampled habitat types (e.g. upper water column, shallow-water areas and fragile ecosystems). Further, the sampling design of always starting and returning to the same area (i.e. close to Stora Karlsö, see Fig. 1) provided an important advantage: high but discrete re-visitation rate to a concentrated area, providing a stronger foundation for estimating fish biomass in similar areas that were less frequently observed. In addition, the travel from east to west sampled the range of water depths several times weekly (See Appendix A12), under different environmental conditions such as currents and hour of the day. In 2019 there was a break in data collection of ~1.5 months (May–June), which impacted the ability to explain the distribution trends (i.e. R^2 Model 4: 0.31; Model 5: 0.38), underlining the value of high resolution and frequent resampling to inform spatiotemporal models.

We found high consistency between the CMSI estimates of surface temperature and our in-situ observations. However, the CMSI estimates for chlorophyll-a was less consistent with our observations. The low correlation was likely due to the annual late summer chlorophyll-a bloom which did not match the observations well, neither in timing nor level (see Fig. 5 and Appendix A5). The match in the initial season may explain why the models containing chlorophyll-a as a variable still did better in cross validations than the ones without (the salinity observations are dealt with below). Notably, the variation in USV-observations of temperature and chlorophyll-a was partly due to the fine resolution of hour and level of solar irradiance. High temperature/irradiance leads to sensors exaggerating chlorophyll measurements (Rousso et al., 2021), but on a USV there is no possibility for in-situ sensor corrections nor determination of calibration values. While the environmental variables, the semi-random movement pattern and speed of the USV, and the time of day decreased problematic autocorrelation (Appendix A7), it was not completely removed. For that, we needed to aggregate the data before modelling (Model 7), a procedure resulting in more accurate estimates but on fewer and more conservative observations (i.e. smaller CIs). Nevertheless, when comparing the aggregated model prediction (from Model 7) with the non-aggregated one (Model 8), the general trends were similarly described. The aggregated data variant gave stronger effect sizes and better residual fits, along with non-significant autocorrelation. However, the aggregated models also returned larger variations in effect sizes though a smaller overall range in estimates, diffusing the significance of variables which made it

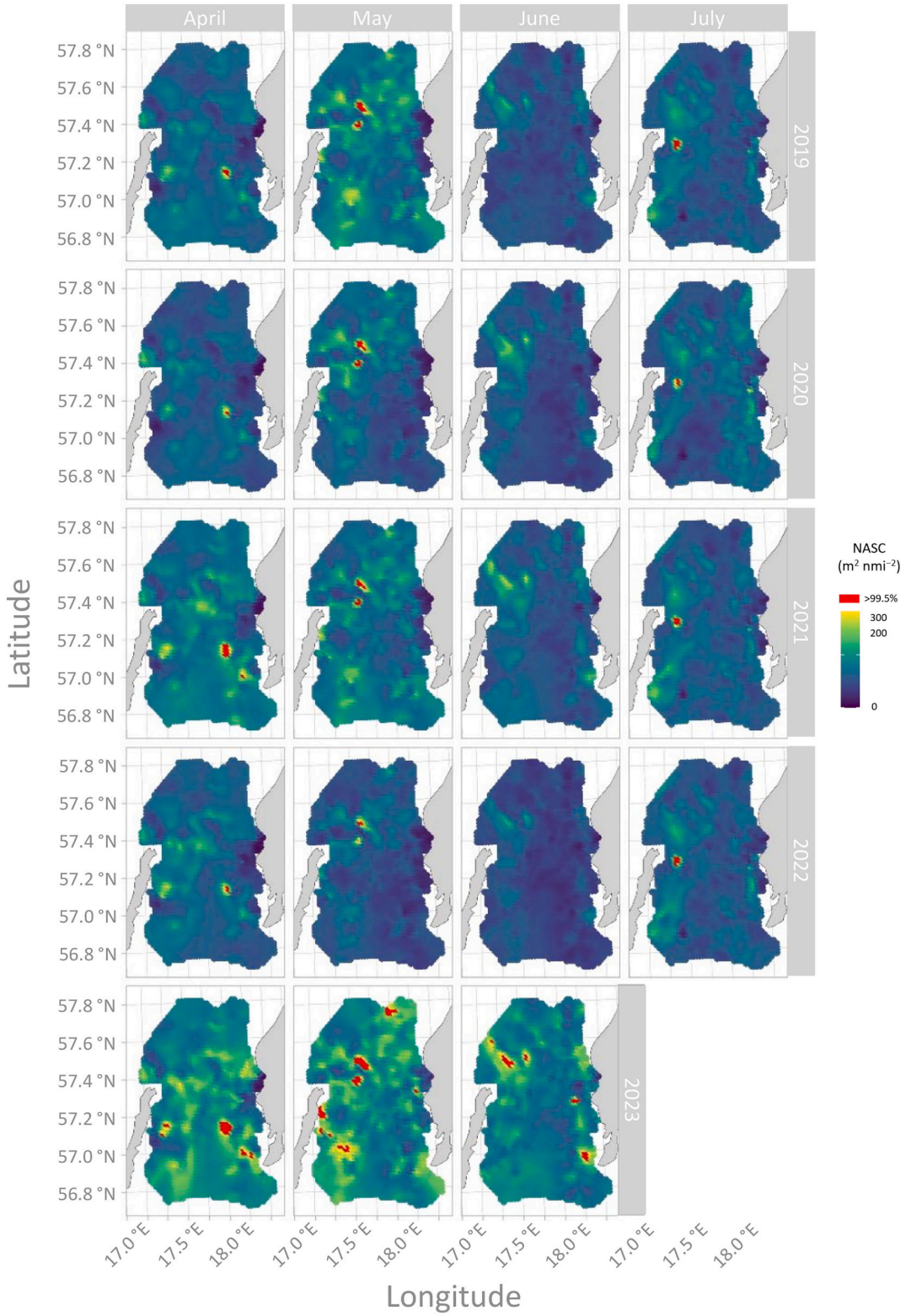


Fig. 7. Predicted fish biomass distribution in NASC ($m^2 nmi^{-2}$) with upper 0.005 quantile marked in red, by months (April–July) per year (2019–2023). Prediction was based on 'hour' 01, using Model 9 (See Table 2) with full-data and autoregressive correlation structure. See Appendix A6 for model summary. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

unclear how successful this aggregation was in terms of unit size (i.e. 2x2km). Without a reference system for biologically significant entities (Schneider and Piatt, 1986; Weimerskirch, 2007), we aggregated to the same resolution as the by-month CMSI data. Preliminary tests found that other time aggregations (e.g. week or even day of the year) resulted in similar number of observations as by month (by JDay: $n = 3220$; Week: $n = 2458$; Month: $n = 1920$; no aggregation: 41292), while any finer scale (e.g. 1, 3 and 6 h intervals, respectively) still had significant autocorrelation.

4.1. Environmental variables and SDMs on community level

Any prediction extending in space and/or time from observations depend on models containing explanatory variable values comparable to what is available on the predictive grid. The correlation procedures were intended to verify the suitability of scale (i.e. monthly resolution and 2x2km means in CMSI variables versus the hourly means of USV variables), in case the effects deviated from the expectations (Andersson et al., 2023; Candolin et al., 2008; Fey, 2001; Jørgensen et al., 2005a, 2005b; Lefébure et al., 2011, 2014; Novotny et al., 2022). We spotted a critical time drift in our own salinity measurements. The USVs CT sensor, without filters or flushing system, was likely cumulating a biological film (i.e. bacterial/algae) over time (Ando et al., 2005). This is a direct consequence of miniature equipment along with the lack of maintenance by an on-site crew. In contrast, the fluorometer had a mechanical cleaning wiper to prevent growth, but still returned a large difference in chlorophyll-a measurement as compared to CMSI estimates. As the USV moved over large areas the problem is unlikely to be due to a spatial mismatch or short-term local levels (see Appendix A12). Notably, the model containing USV-collected temperature and chlorophyll-a variables (Model 6) performed slightly better than its CMSI counterpart (Model 5) in the evaluation process. In models, the environmental variables in general correlated positively (i.e. depth inverted) with NASC levels, except the non-significant chlorophyll-a and current speed. Whilst the general direction of the environmental effects remained similar across models and data aggregation, the effect sizes and thus their significance needs to be evaluated with caution due to the uncertainty of the underlying data (i.e. autocorrelation in fish observations, and mismatch between CMSI estimates and USV observations). Though the models presented here explained the distribution of the pelagic fish community reasonably well ($R^2 > 0.30$), the environmental effects could not be viewed as essential. However, there may be alternative reasons to why the variable effects were less clear than expected. Firstly, surface variable estimates for chlorophyll-a, salinity and temperature are unlikely to be the best predictors for pelagic fish distributions (see Appendix A3), especially in highly stratified waters like the Baltic Sea (Liblik and Lips, 2019; Muchowski et al., 2023). Secondly, the models were based on aggregated NASC values from a community of at least three fish species, when even different species of cohabitant clupeids can respond to hydro-climatic forces in significantly different ways (Pennino et al., 2020). In addition, the species in this community have different spawning times and conditional demands during the studied time-period and, even within species, the dynamics depended on size and reproductive state (Andersson et al., 2023; Cardinale et al., 2003). In order to disentangle data by species, size compositions and proportions for realistic biomass estimates (i.e. kg/km^2), biological samples such as trawling are still essential and cannot be performed by the USV. Alternatively, species and size composition can be inferred by Target Strength (TS) equations to the raw acoustic data (Didrikas and Hansson, 2004). For now, methods are primarily under development for species of commercial importance (Fässler et al., 2008; Ona, 2003). Further development on this topic would be desirable to improve the results of trawling-independent acoustic monitoring such as USV surveys.

4.2. Autocorrelation and data aggregation

Spatiotemporal autocorrelation was unavoidable in the USV observations, with no standardized method for handling such data without heavily aggregating it. However, beyond compromising the resolution, aggregating the data is problematic for several reasons. Firstly, the speed of the USV varied, and the areas sampled were opportunistic based on wind directions, meaning that the number of observations within each spatial and temporal aggregation would vary drastically amongst areas. While weighting estimates for the number of observations could aid, we would still face the problem of detected fish biomass varying drastically with hour of the day (Fig. 6). This variation reflects one of the main challenges in hydroacoustic data treatment: distinguishing fish close to the seabed (Mello and Rose, 2009). Two of the species highest represented in the area, herring and sprat, perform diel vertical migrations (Cardinale et al., 2003). During daylight hours they can be inseparable from the seabed, even by visual inspection. Due to the anoxic zone, fish were rarely detected in deeper depths, and so whenever the seabed was deeper, a loss of fish detection was unlikely. However, more than 1/3 of our observations were made during low-detection hours (see Fig. 6) and at depths shallower than the typical depths of the anoxic zone. While the models were trained on all hours of the day, we decided to predict on the estimated night distribution (01 am CEST), to compensate for the change in detection rate by time of day. For future studies, the key areas for change in biomass by time of day could be estimated prior to spatiotemporal modelling, to correct for observations in areas and times where an acoustic 'dead zone' (Mello and Rose, 2009) is likely to exclude fish. Though a large proportion of the clupeids often remains close to the seabed also during the night (Cardinale et al., 2003), finding a way to, at least partially, compensate for non-detected fish before aggregating data would be ideal.

4.3. Final remarks and potential contributions to future research

Despite the complexity of environmental effects, fish behaviour and autocorrelation, the final distribution of small pelagic community predicted in this study is in line with what is expected for the Baltic during late spring-summer (Candolin et al., 2008; Jørgensen et al., 2005; Ojaveer and Kalejs, 2010), as for similar species around the globe (e.g. sardines and anchovy in Mediterranean Sea: Pennino et al., 2020; sprat in the Adriatic Sea: Palermino et al., 2024). The varying spawning times and bathymetric requirements, means that movement between suitable spawning sites is expected with strong month specific patterns and similarities across years.

There are possibilities for technical development in smaller USVs and miniature sensors, especially with focus on autonomous calibrations and maintenance for longer lasting surveys. The continuity of the data makes it potentially highly correlated, whilst robust tools for handling such data is still under development. In the meantime, stepwise analysis can clarify the strengths and weaknesses in the data, and can reveal meaningful dynamics and insights on the way. While data aggregation still proved the best way to reduce problematic data structures like autocorrelation, crucial details were lost in the process. However, both with aggregated and non-aggregated data, we estimated similar effects and distributions, which revealed important ecological dynamics and highlighted that some areas consistently offered favourable habitats for small pelagic species. Overall, hydroacoustic surveys can become much more cost effective with the inclusion of USVs for pre-surveys, simultaneous co-operative surveys (e.g. pre-sailing transects) though with limitations in comparability between obtained values between larger vessels and USVs. The majority of biomass in our study was observed in water too shallow (<15 m) for larger ships to monitor effectively, and significant abundances were found in shallow-seabed areas (<20 m) that are often inaccessible to larger vessels. Even in comparable depths, aligning the ping rate to survey speed between vessels would be difficult. Finally, the USVs variance in speed with fixed ping rate results in greater data variability as compared to large vessel surveys. In any case, the USV would be highly useful for extending surveyed areas and habitat types, and for coping with issues of larger vessels, such as disturbance to the observed environment. The possibility to model fish distribution on a

fine scale could be a valuable tool for ecosystem-based management of small pelagic fish, and for detailed ecological studies of their role in marine ecosystems (Cury et al., 2011; Hilborn et al., 2017; Pikitch et al., 2014). Here, we estimated the detailed spatial distribution of the small pelagic fish community in high resolution, in an area important to avian predators throughout their reproductive period (Galatius et al., 2020; Österblom et al., 2006). The next steps should be to find a solution to the effects of time of day, and match the spatiotemporal distribution of prey to the foraging sites of the local top predators dependent on them.

CRediT authorship contribution statement

Astrid A. Carlsen: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Michele Casini:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Investigation, Funding acquisition. **Francesco Masnadi:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Conceptualization. **Olof Olsson:** Project administration, Methodology, Funding acquisition. **Aron Hejdström:** Data curation. **Jonas Hentati-Sundberg:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Data curation.

Declaration of competing interest

None.

Data availability

All data are publically available through the Swedish National Data Service (SND) https://doi.org/10.5878/9m27-vg06. R-codes are published at Github Inc., https://github.com/Astidac/Autonomous-data-sampling-for-high-resolution-spatiotemporal-fish-biomass-estimates.

Acknowledgements

We owe many thanks to everyone involved in the Baltic Seabird Project, especially those tied to the USV work. We'd also like to thank the crew at RV Svea for aid and assistance during the voyages of the Sailbuoy. Thank you to Max Lindmark for feedback in the modelling procedure, and Agnes Olin for feedback on the manuscript and on the modelling procedure. AAC was supported by Marcus and Marianne Wallenbergs stiftelse, JHS was supported by Vetenskapsrådet and Formas. Field work was supported by WWF Sweden and Milkywire. We also wish to thank three anonymous reviewers for valuable revision of the manuscript.

Jonas Hentati-Sundberg, Olof Olsson and Astrid Carlsen were supported by Marcus and Marianne Wallenbergs stiftelse (grant number 2018-0093). Jonas Hentati-Sundberg was also supported by Vetenskapsrådet (grant number 2021- 03892) and Formas (grant number 2021-02639).

Appendix A. Appendix

A.1. Distribution of response variable, NASC

The biomasses observed ranged from 0.1 to 15,174.9 m2 nmi − 2, with 1st quartile of 32.9 m2 nmi − 2, median of 70.74 m2 nmi − 2, mean of 293.8 m2 nmi − 2 and 3rd quartile of 154.6 m2 nmi − 2, meaning strongly Poisson distributed. For all analysis, NASC values were log transformed to approach the assumption of normal distribution (Fig. A1.1). There was still a slight upper skew likely reflects larger aggregations of fish.

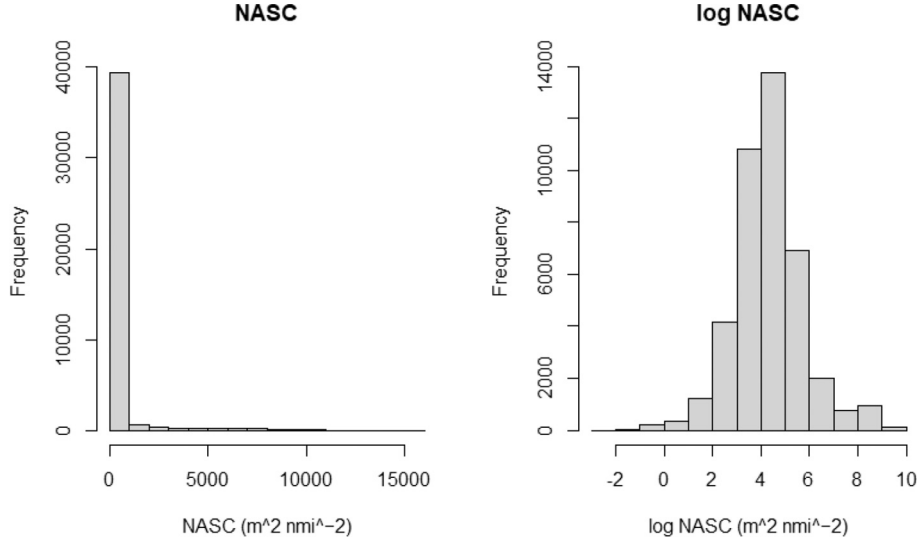


Fig. A1.1. Distribution of response variable NASC before and after log transforming.

Distribution family for the models were chosen after stepwise tests and inspecting residual distributions. The final best distribution for the response variable log-NASC was student distribution, with df = 4 (fig. A1.2).

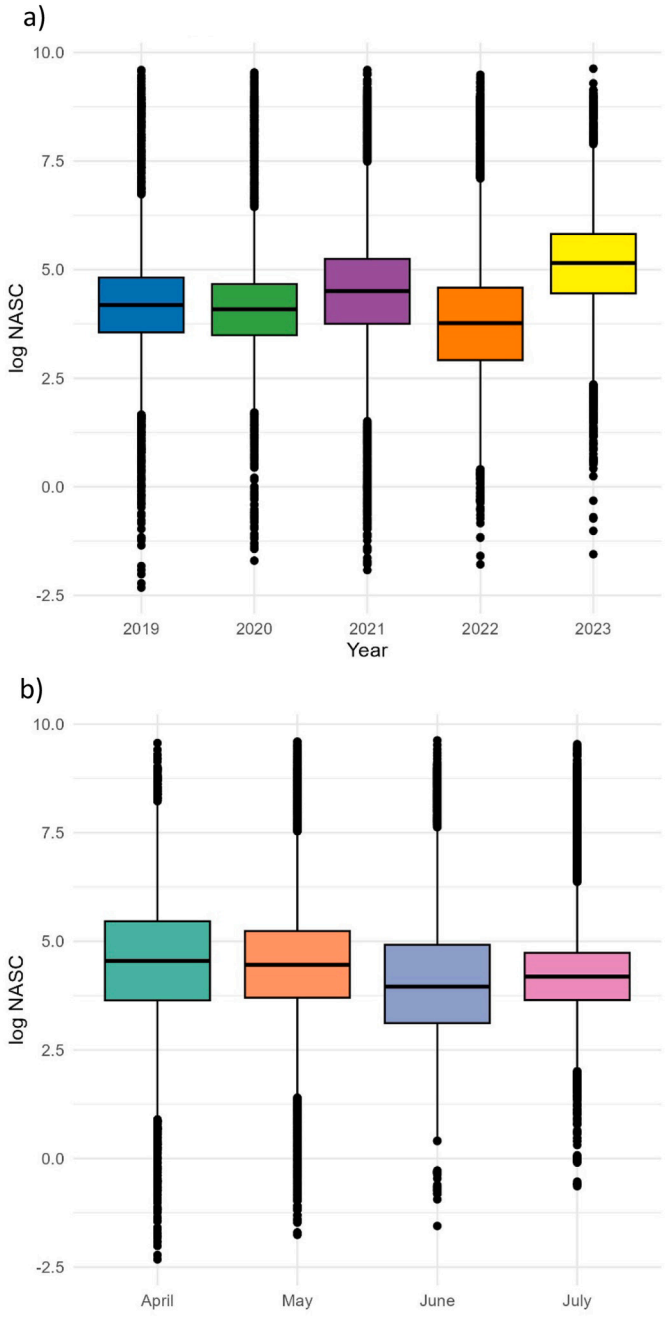


Fig. A1.2. Distribution of log-transformed NASC values across (a) years and (b) months.

A.2. Model residuals

Model residuals were inspected visually to select the best distribution family for each model.

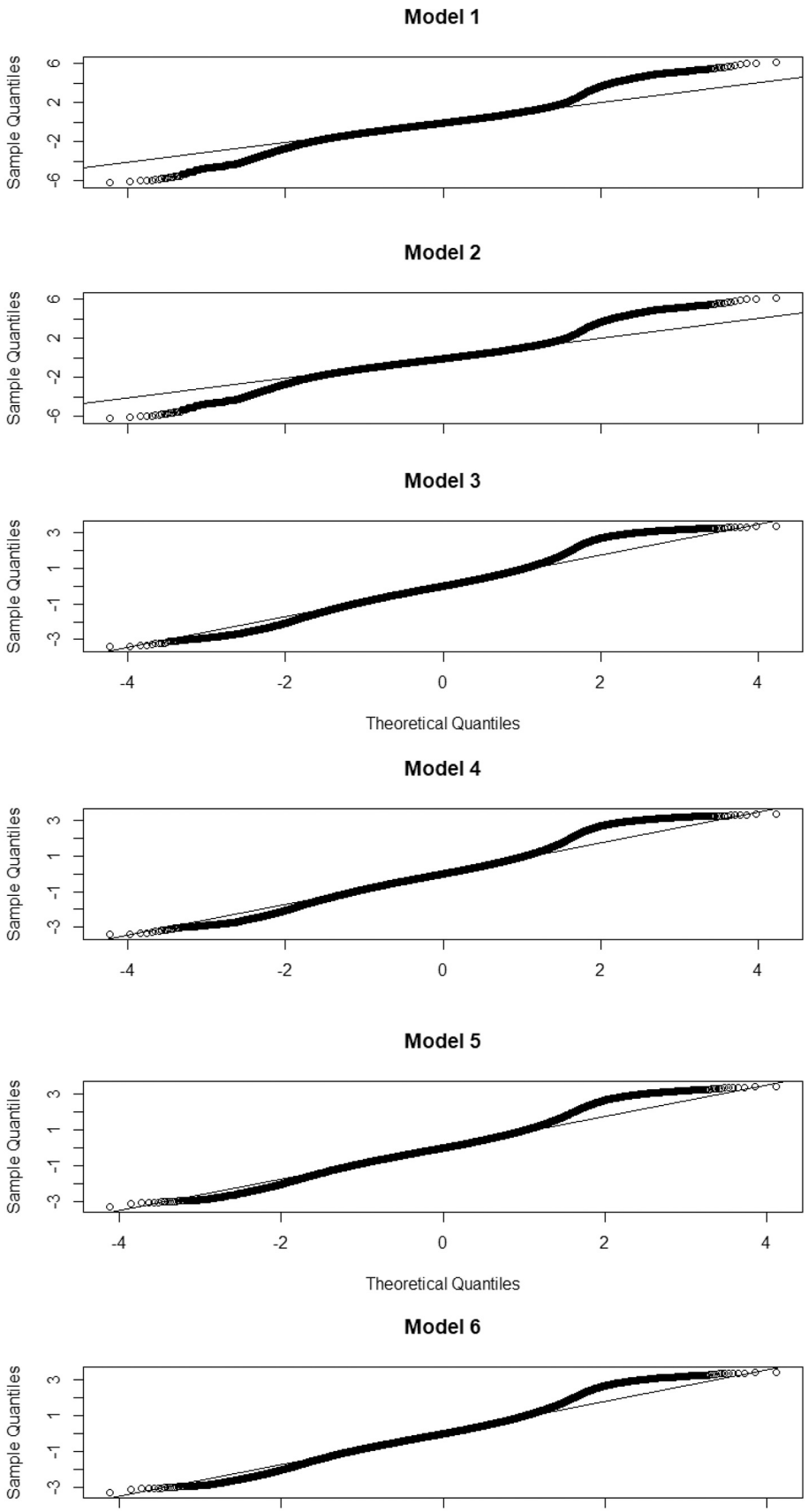


Fig. A2. Residual distributions for Model 1–9, all fitted with student distribution (link 'identity', df = 4), models 1–6 and 8–9 had correlation structure 'AR1', Model 7 had correlation structure 'iid'.

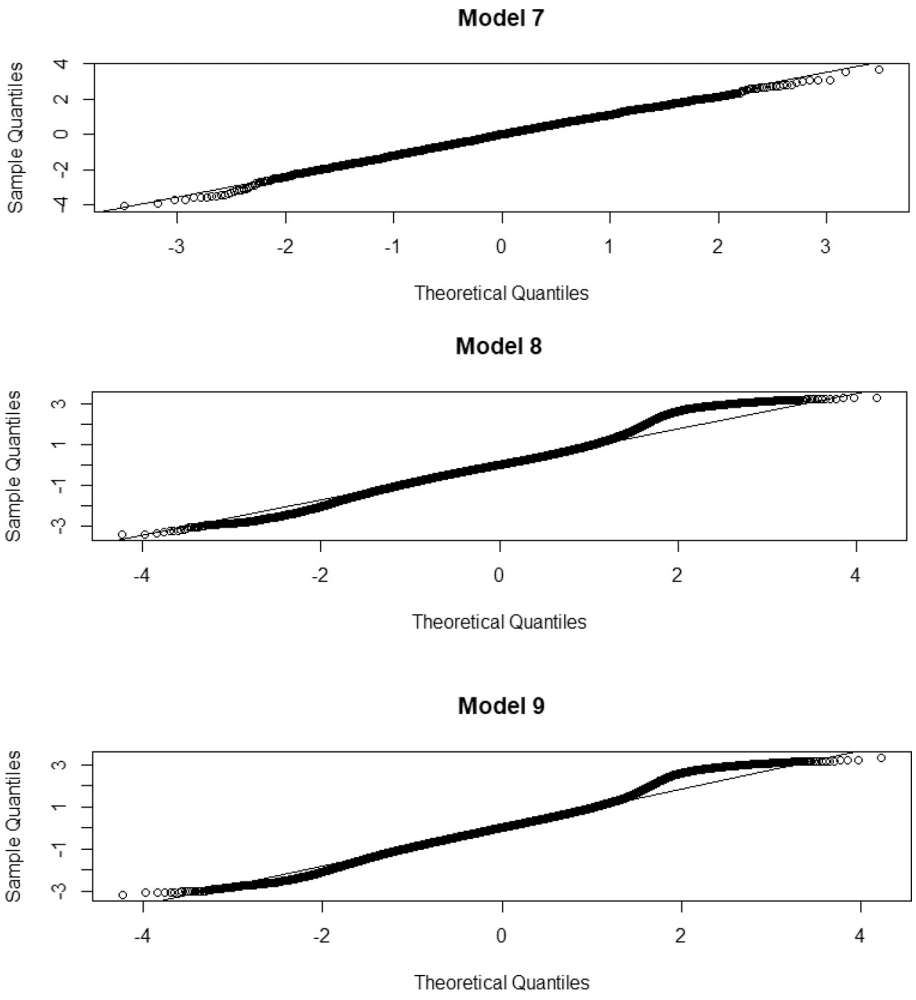


Fig. A2. (*continued*).

A.3. Correlation tests predictive variables

General correlation between mean water column estimates and sea –surface estimates of chlorophyll-a, salinity and temperature were produced, to test how similar they were (Table A3.1). While water column estimates for temperature and chlorophyll explained the fish distribution better than surface estimates (Table A3.2), water column estimates should be used with cation due to their high level of uncertainty as e.g. remote sensing is less accurate. Further, the USV can per now only perform surface observations, and so for any discrepancy test on values for salinity, temperature and chlorophyll-a we'd have to rely on surface values from CMSI anyways. The correlation between water column means (weighted for depth, where surface values are weighted heavier) shows that there is a rather large discrepancy between them, inferring the strong stratification of the water column in the Baltic sea. This reflects why it may be problematic to base models for fish distribution of pelagic species on surface variables, and why the models didn't improve more when these core variables are included.

Table A3.1
Correlations between surface and water column averages by CMSI.

Variables	Correlation
Chlorophyll: water column vs surface, CMSI	0.52
Salinity: water column vs surface, CMSI	−0.05
Temperature: water column vs surface, CMSI	0.72

Table A3.2
Estimated effects and prediction power of distribution model when using mean of the water column versus sea-surface versions of the variables chlorophyll-a, salinity and temperature from CMSI.

Variable	Effect (SE)	Pr(>|t|)	R2-Adj.
Mean water column Chlorophyll	0.14 (0.013)	<2e-16	0.003
Sea surface Chlorophyll	0.06 (0.007)	8.59e-15	0.001

(*continued on next page*)

Table A3.2 (*continued*)

Variable	Effect (SE)	Pr(>|t|)	R2-Adj.
Mean water column salinity	−0.11 (0.013)	<2e-16	0.002
Sea surface salinity	−0.63 (0.029)	<2e-16	0.011
Mean water column temperature	−0.07 (0.003)	<2e-16	0.011
Sea surface temperature	−0.03 (0.002)	<2e-16	0.007

A.4. Sailing velocity of USV

While the drone were in movement for most of the time, more than 700 observations were made laying completely still, and more than 2500 were made with a speed of more than 1 m per second. With a speed of 0.5 m per second, 7 min of sampling aggregated contains fish detections over 210 m, while a speed of 1 m/s an observation would cover a distance of 420 m. Distance covered must be seen as a conservative measurement as the drone sometimes would sail in a straight line, and sometimes in zig-zag to maintain the direction set by waypoint, meaning that including distance and/or speed directly in any model is far from straight forward.

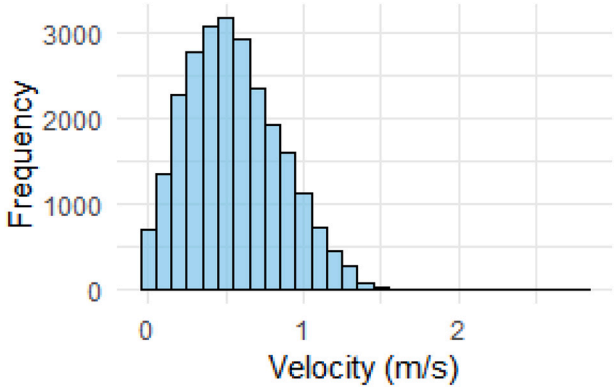


Fig. A4. Velocity of USV per NASC observation across all years.

A.5. Salinity by month of year at Karlsö by SMHI 2021–2023

To verify the deviating salinity levels observed by the USV versus CMSI estimates, we obtained observations from the local hydrology station of Swedens Metrological and Hydrological Institute (SMHI). The hydrological measuring station is situated just outide of Stora Karlsö, within the study area. The observations by SMHI, which is also included in the data for estimates of surface temperature by CMSI, closely resembles the CMSI estimates, with the same clear deviation from the USV observations, supporting the impression that we had technical issues with our equipment.

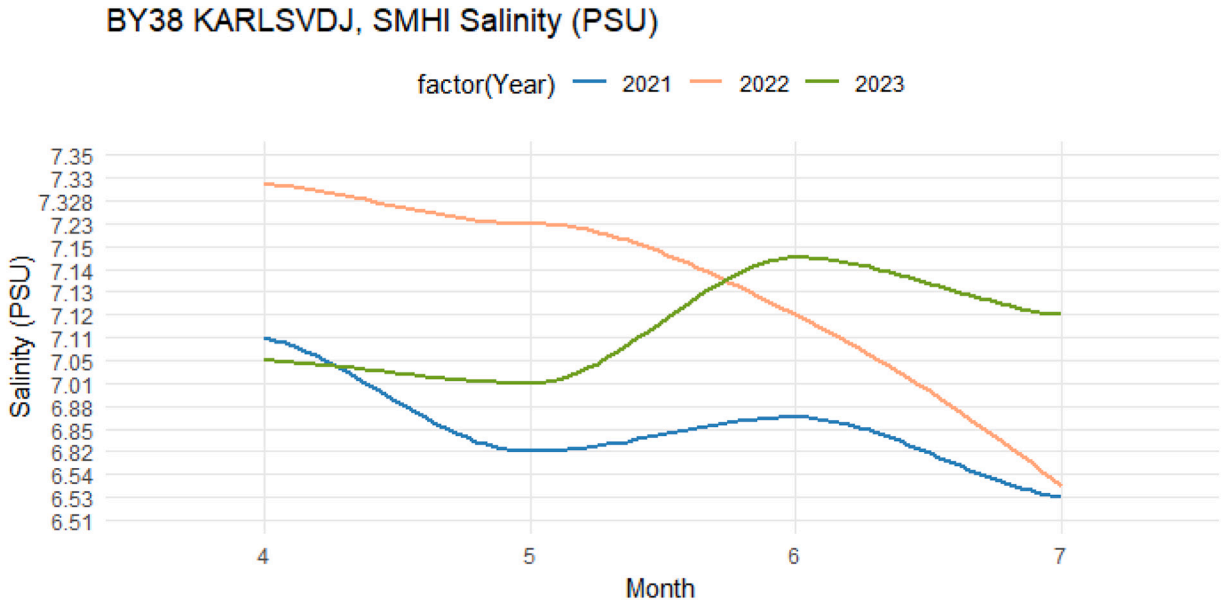


Fig. A5.1. Local sea-surface salinity measurements by SMHI across three years for reference in range and direction of effects as compared to the daily resolution 2021 estimates by CMSI and the USV.

A.6. Model output /results

Model output for the models described in more detail in the result section: Model 4, Model 7, Model 8 and Model 9. Model outputs include information about the general structure of the model, effect sizes and information on model validation summarized. Visualisations of variable effects for model 4 is given in *Results* Fig. 6, and for Model 7 and 8 Appendix A8 Fig. A8.1 and A8.2, respectively.

Table A6
Model output for Key SDMs: Model 4, Model 7, Model 8 and model 9.

Model 4:		
Spatiotemporal model fit by ML [‘sdmTMB’]		
Time column: Month		
Family: student(link = ‘identity’)		
	coef.est.	coef.se
(Intercept)	4.35	0.24
Mean_EW_uo_Wclm	1.79	0.63
Mean_current_speed	−0.24	1.22
sWeek	0.18	0.16
sDepth	−0.02	0.02
sMean_thetao_SST	0.02	0.31
sSO_surf	0.12	0.16
sChl_surf	−1.14	0.37
sMean_NS_vo_Wclm	0.7	0.11
Smooth terms:		
	Std. Dev.	
sds(Week)	2.23	
sds(Hour)	0.08	
sds(Depth)	0	
sds(Mean_thetao_SST)	1.12	
sds(SO_surf)	1.4	
sds(Chl_surf)	5.32	
sds(Mean_NS_vo_Wclm)	2.53	
Random intercepts:		
	Std. Dev.	
Year	0.5	
Dispersion parameter:	0.72	
Spatiotemporal AR1 correlation (rho):	0.12	
Matérn range:	7.3	
Spatial SD:	0.31	
Spatiotemporal marginal AR1 SD:	0.77	
ML criterion at convergence:	60,590.05	

Model 7:		
Spatiotemporal model fit by ML [‘sdmTMB’]		
Time column: Month		
Family: student(link = ‘identity’)		
	coef.est.	coef.se
(Intercept)	4.57	0.26
Mean_EW_uo_Wclm	4.28	1.45
Mean_current_speed	−1.09	2.66
sDepth	−0.27	0.23
sMean_thetao_SST	0.13	0.43
sSO_surf	0.19	0.06
sChl_surf	−0.52	0.53
sMean_NS_vo_Wclm	0.49	0.27
Smooth terms:		
	Std. Dev.	
sds(Depth)	0.75	
sds(Mean_thetao_SST)	0.97	
sds(SO_surf)	0	
sds(Chl_surf)	5.14	
sds(Mean_NS_vo_Wclm)	1.53	
Random intercepts:		
	Std. Dev.	
Year	0.5	
Month	0.08	
Dispersion parameter:	0.49	
Matérn range:	20.96	
Spatial SD:	0.2	
Spatiotemporal IID SD:	0.45	
ML criterion at convergence:	2012.305	

Model 8:	**Model 9:**

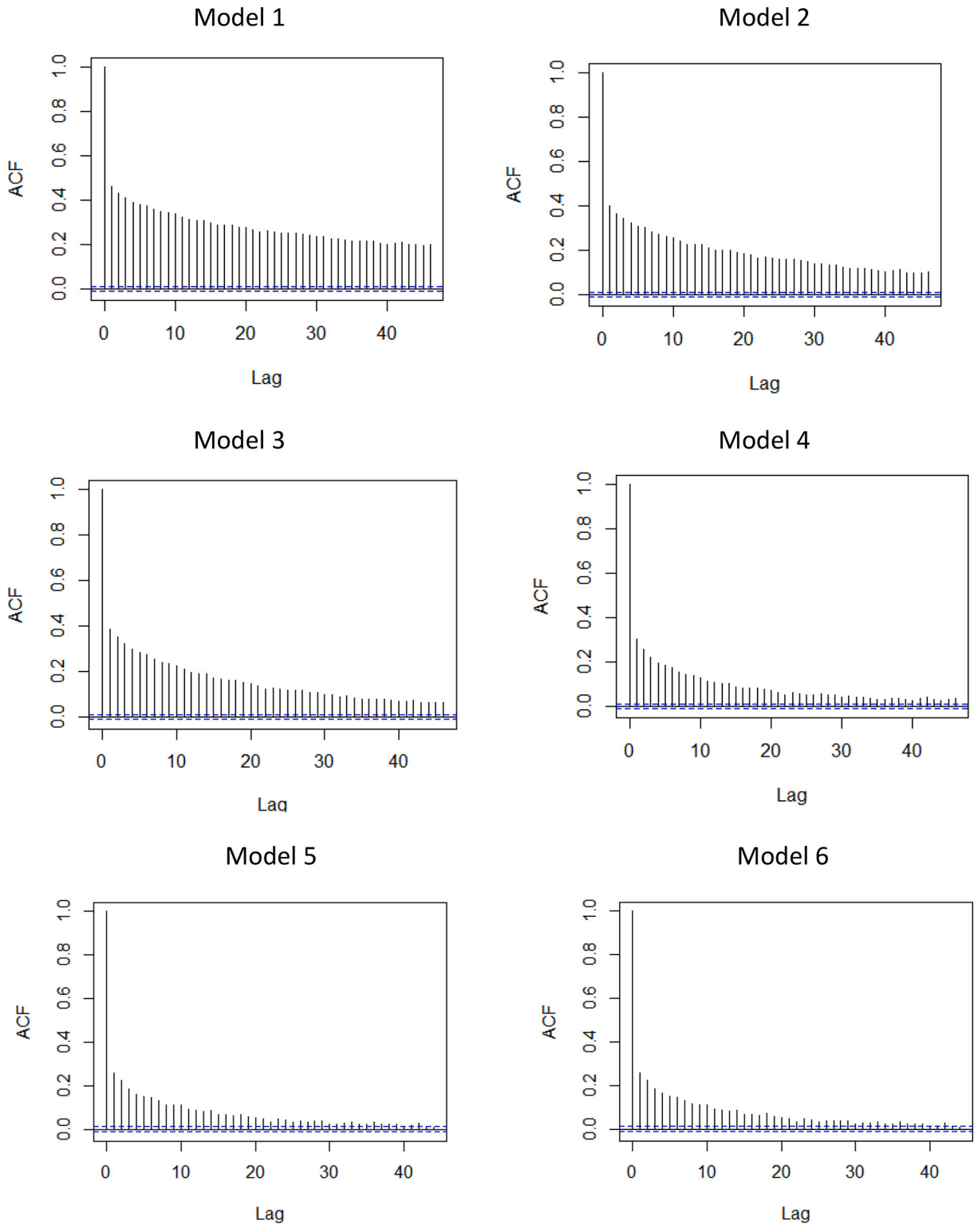
(*continued on next page*)

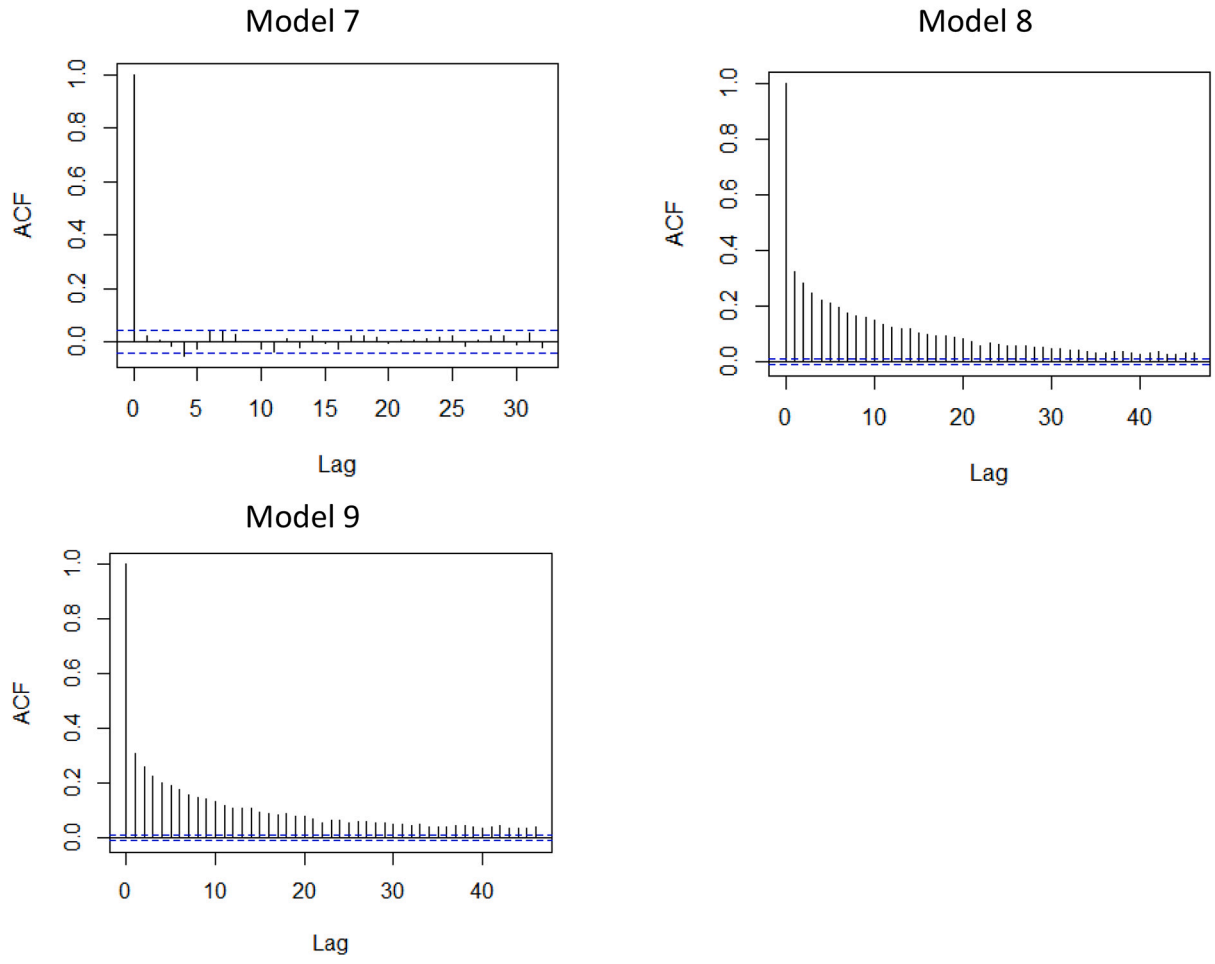
Table A6 (*continued*)

Model 8:			**Model 9:**		
Spatiotemporal model fit by ML ['sdmTMB'] Time column: Month Family: student(link = 'identity')			Spatiotemporal model fit by ML ['sdmTMB'] Time column: Month Family: student(link = 'identity')		
	coef.est.	coef.se		coef.est.	coef.se
(Intercept)	2.46	0.69	(Intercept)	4.37	0.24
Mean_thetao_SST	−0.01	0.01	Mean_EW_uo_Wclm	1.83	0.63
SO_surf	0.29	0.09	Mean_current_speed	0.1	1.25
Chl_surf	−0.04	0.02	sDepth	0	0.02
Mean_EW_uo_Wclm	0.41	0.62	sMean_thetao_SST	−0.16	0.37
Mean_current_speed	0.65	1.28	sSO_surf	0.03	0.18
sDepth	−0.09	0.18	sChl_surf	−1.79	0.38
sMean_NS_vo_Wclm	0.75	0.11	sMean_NS_vo_Wclm	0.74	0.11
Smooth terms:			Smooth terms:		
	Std. Dev.			Std. Dev.	
sds(Depth)	1.02		sds(Hour)	0.08	
sds(Mean_NS_vo_Wclm)			sds(Depth)	0	
			sds(Mean_thetao_SST)	1.87	
Random intercepts:			sds(SO_surf)	1.61	
	Std. Dev.		sds(Chl_surf)	6.25	
Year	0.45		sds(Mean_NS_vo_Wclm)	2.72	
Dispersion parameter:	0.75		Random intercepts:		
Spatiotemporal AR1 correlation (rho):	0.05			Std. Dev.	
Matérn range:	7.17		Year	0.51	
Spatial SD:	0.37				
Spatiotemporal marginal AR1 SD:	0.83		Dispersion parameter:	0.72	
ML criterion at convergence:	62,028.04		Spatiotemporal AR1 correlation (rho):	0.12	
			Matérn range:	7.21	
			Spatial SD:	0.3	
			Spatiotemporal marginal AR1 SD:	0.77	
			ML criterion at convergence:	60,867.51	

A.7. Autocorrelation structures in models

The autocorrelation across 50 observations of NASC values for models 1–9. The autocorrelation is highest in models without environmental variables (Model 1 and 3), and further decreased for models including variables for high resolution time structures, i.e. 'hour' and 'week' (Model 4–6 and 9). All models had highly significant autocorrelation structures except the aggregated data model (Model 7).


Model 1
ACF
Lag
Model 2
ACF
Lag
Model 3
ACF
Lag
Model 4
ACF
Lag
Model 5
ACF
Lag
Model 6
ACF
Lag



. (*continued*).

A.8. Variable regressions from comparative models with different data aggregation levels

To validate the results from a spatial prediction model based on autocorrelated data, we compared the effect sizes and directions in two models with the same model structure, based on non-autocorrelated aggregated data (Model 7) versus non-aggregated autocorrelated data (Model 8). See *Results* for more details. The model output for Models 7 and 8 is given in Table A6.

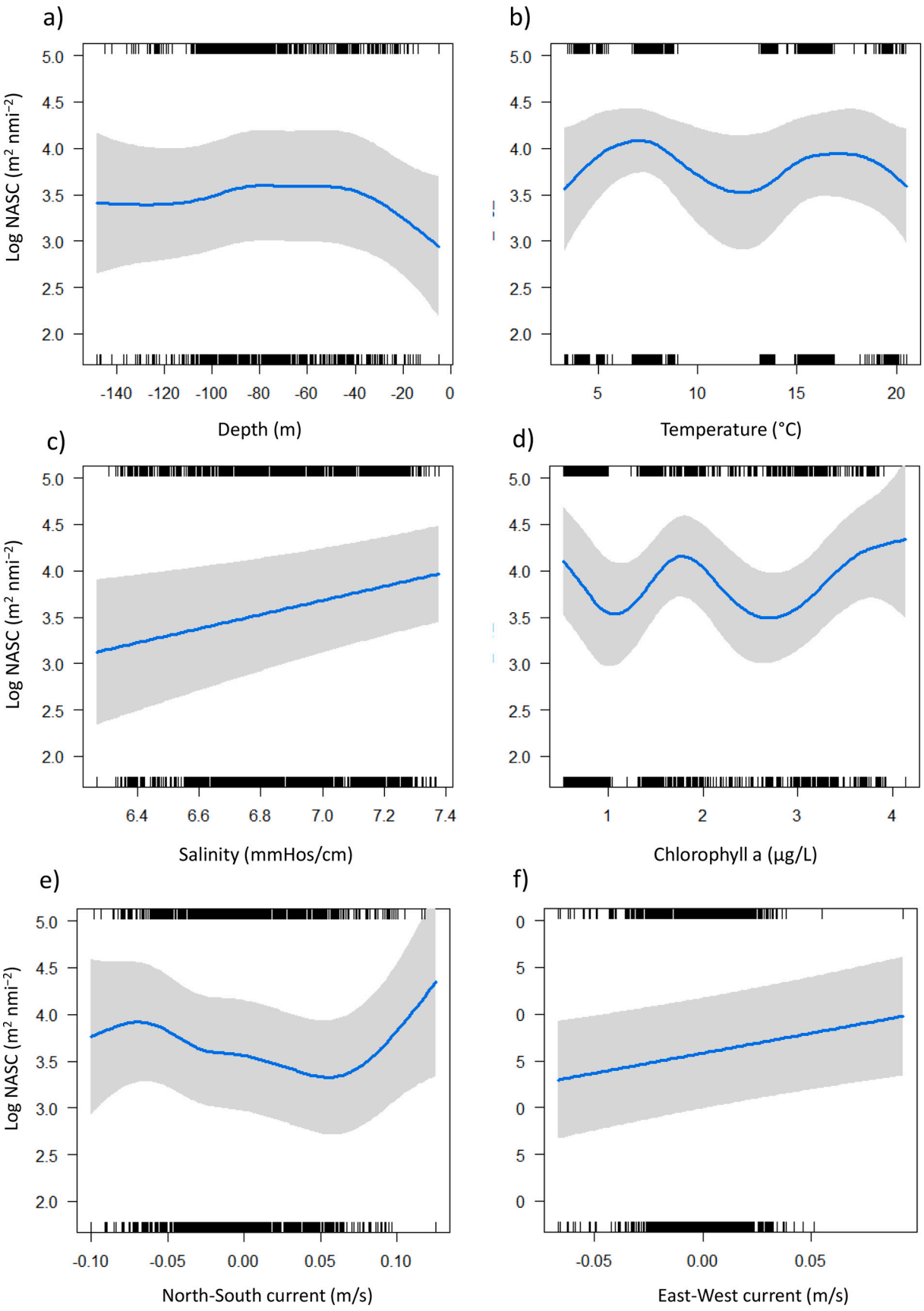


Fig. A8.1. Model 7 (2×2 km spatially aggregated data) effect sizes of each explanatory variable on log NASC, from the GLMM with environmental estimates from CMSI. See Appendix A6 for model summary and Table 4 for model validation and performance.

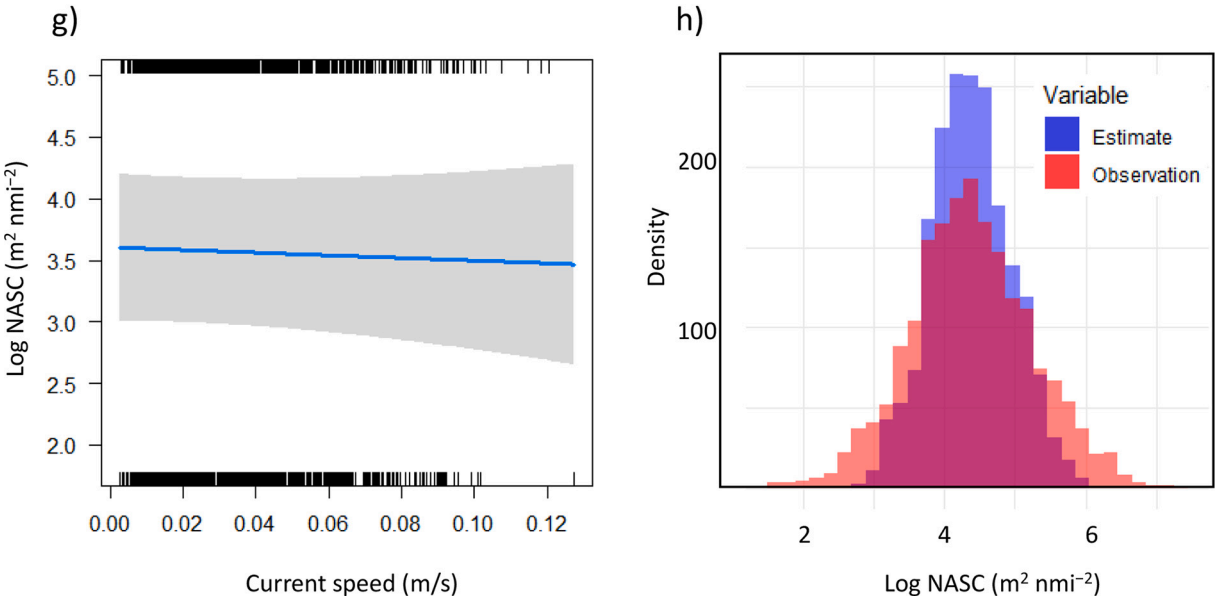


Fig. A8.1. (*continued*).

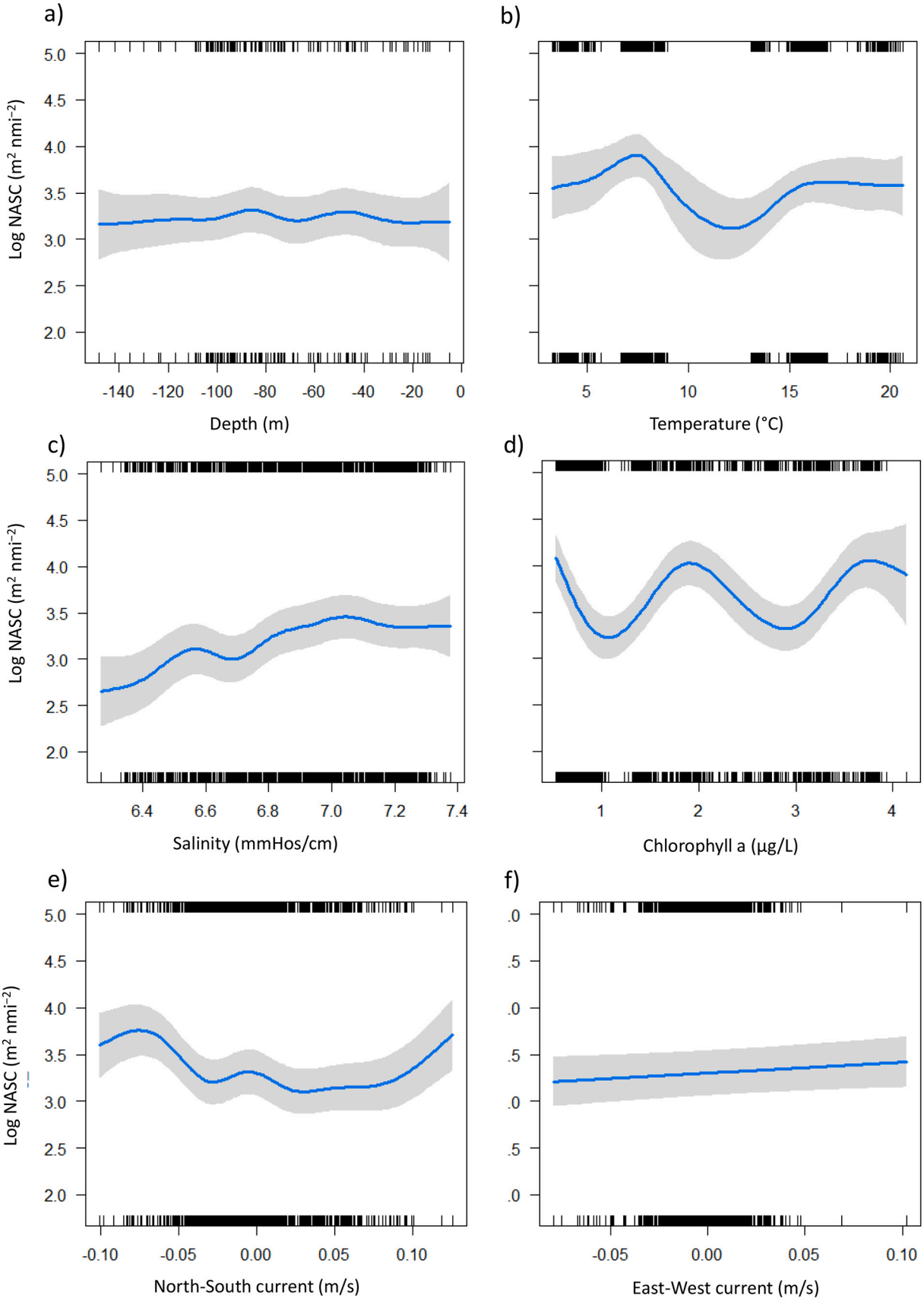


Fig. A8.2. Non-aggregated data model (Model 8) effect sizes of each explanatory variable on log NASC, from the GLMM with environmental estimates from CMSI. See Appendix A6 for model summary and Table 4 for model validation and performance.

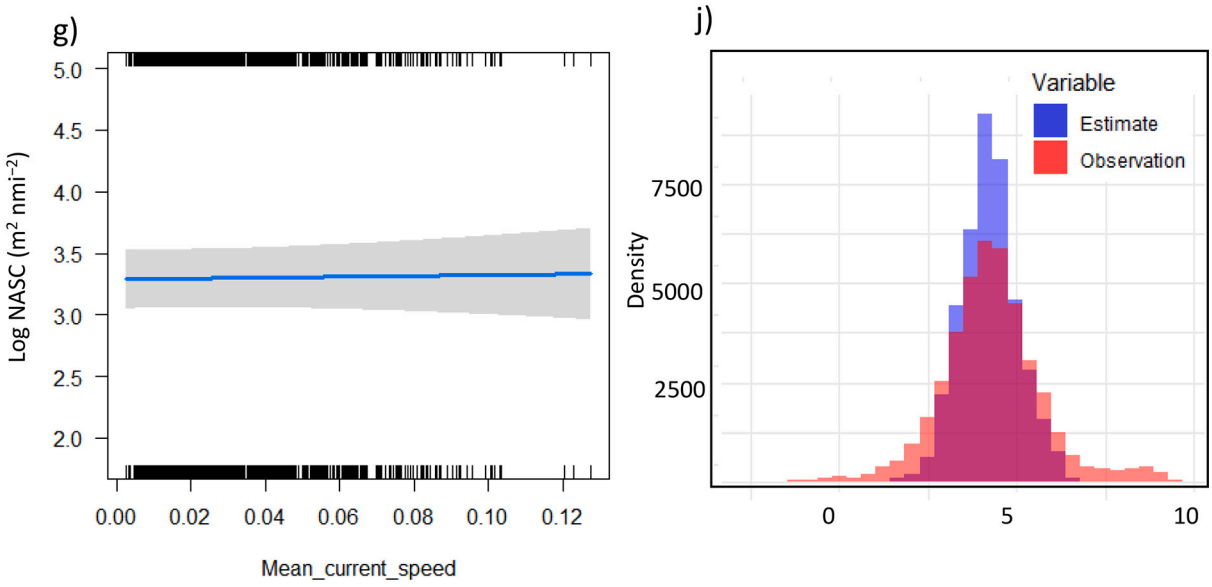


Fig. A8.2. (*continued*).

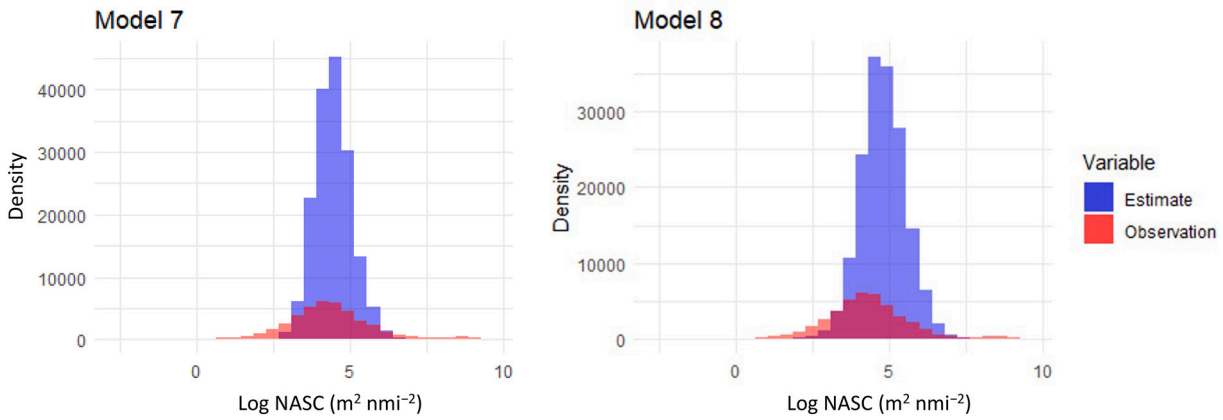


Fig. A8.3. Grid predicted values versus observations for Model 7: prediction on 2 × 2km aggregated data, versus Model 8: prediction on non-aggregated data. See Appendix A6 for model summary and Table 4 for model validation.

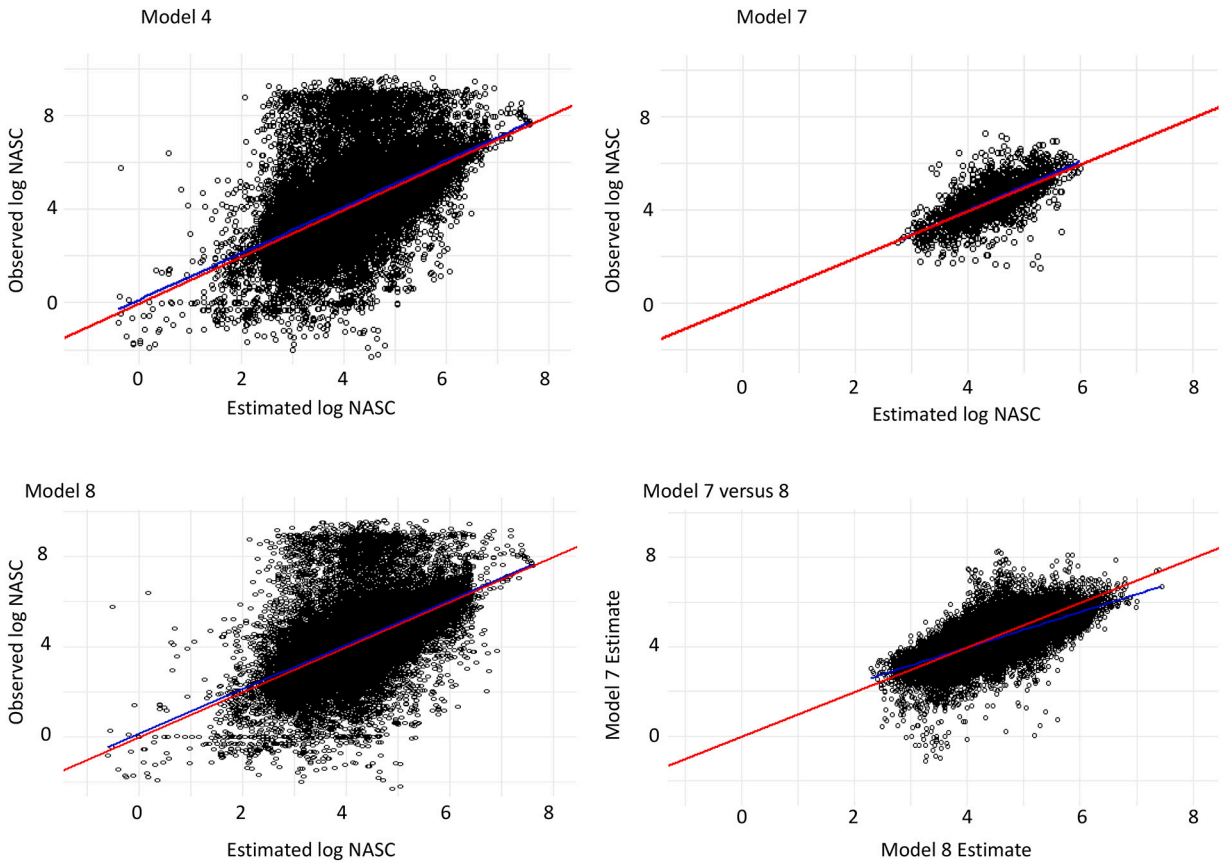


Fig. A8.4. Correlation of prediction versus observations for non-aggregated data model (Model 4), 2 × 2km aggregated data (Model 7), comparative non-aggregated model (Model 8) and Model 7 on Model 8 estimates. Red line indictes a 1-1 linear regression while blue line gives the observed regression.

A.9. Spatiotemporal prediction of Model 7

Spatial prediction based on aggregated data and Model 7. While much less upper quantile predictions are made, the general pattern of high abundance areas is similar to the patterns generated by the non-aggregated data Model 9.

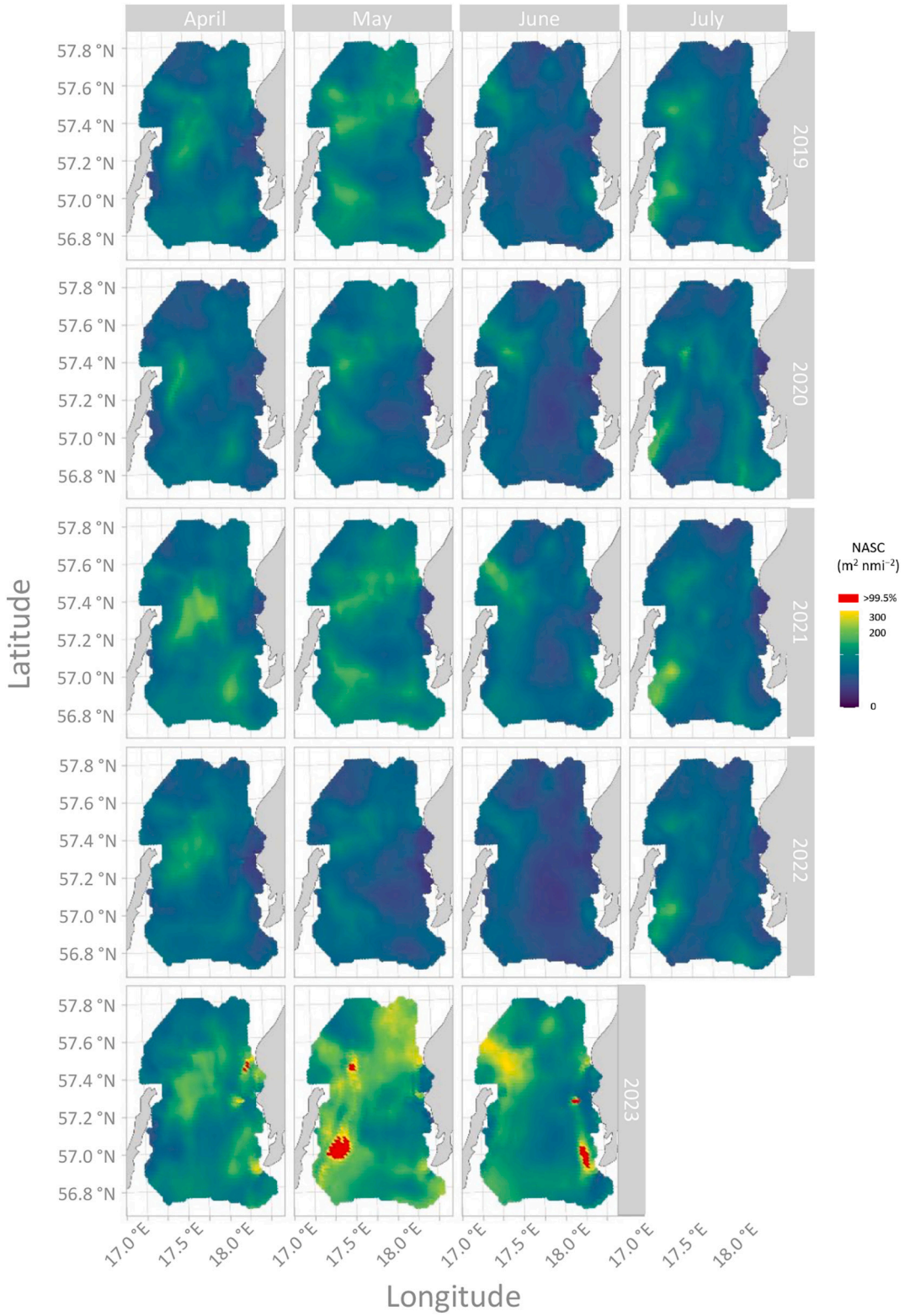


Fig. A9.1. Spatiotemporal prediction of aggregated data, Model 7.

A.10. Spatial distribution of model residuals

Residuals of model estimates

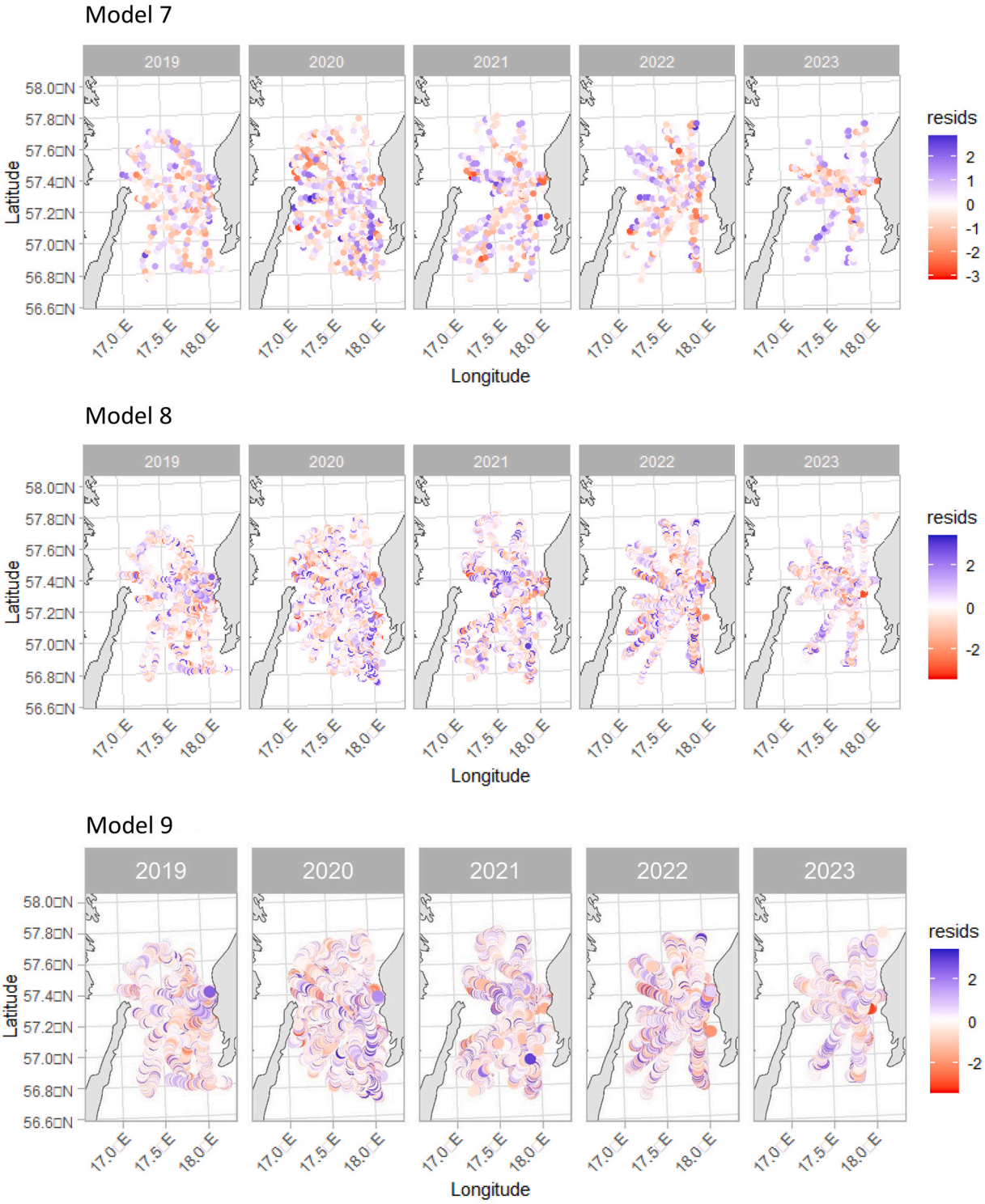


Fig. A10.1. Spatial distribution of residuals of Model 7, 8 and 9.

A.11. Model validations

The stepwise model selection and validation was based on the metrics R2, MAE, RMSE and LL, where we present the mean values in Table 3 (See *Results*). The mean value was based on 5 iterations of each model (Fig. A11.1), after performing 10 iterations and determining that the range of variation across 10 iterations was sufficiently small (Fig. A11.2).

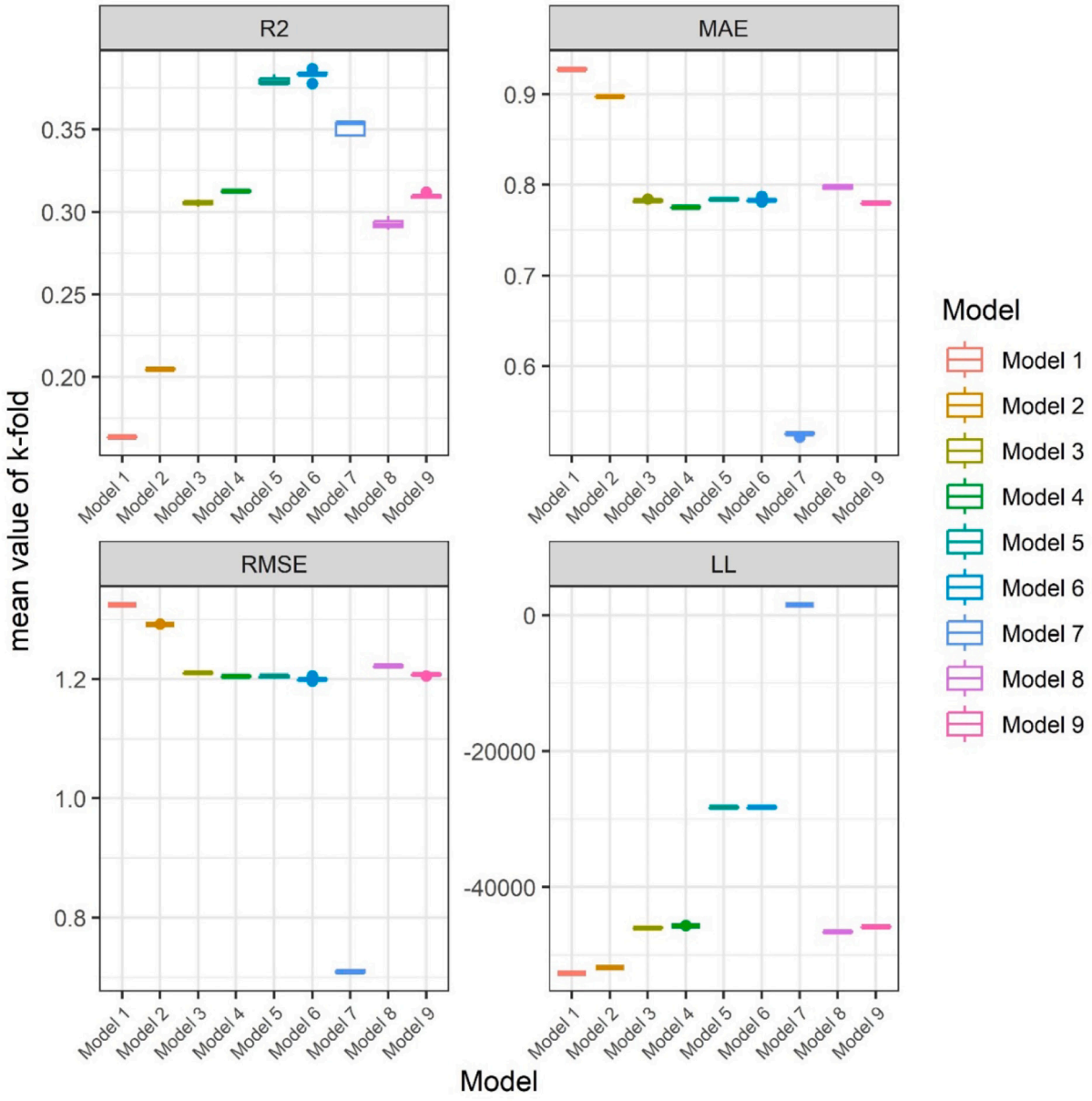


Fig. A11.1. Mean value of k-fold from 5 iterations of each model, with 70/30 % training/test data (See Fig. A11.2 for 10 iterations of first 4 models).

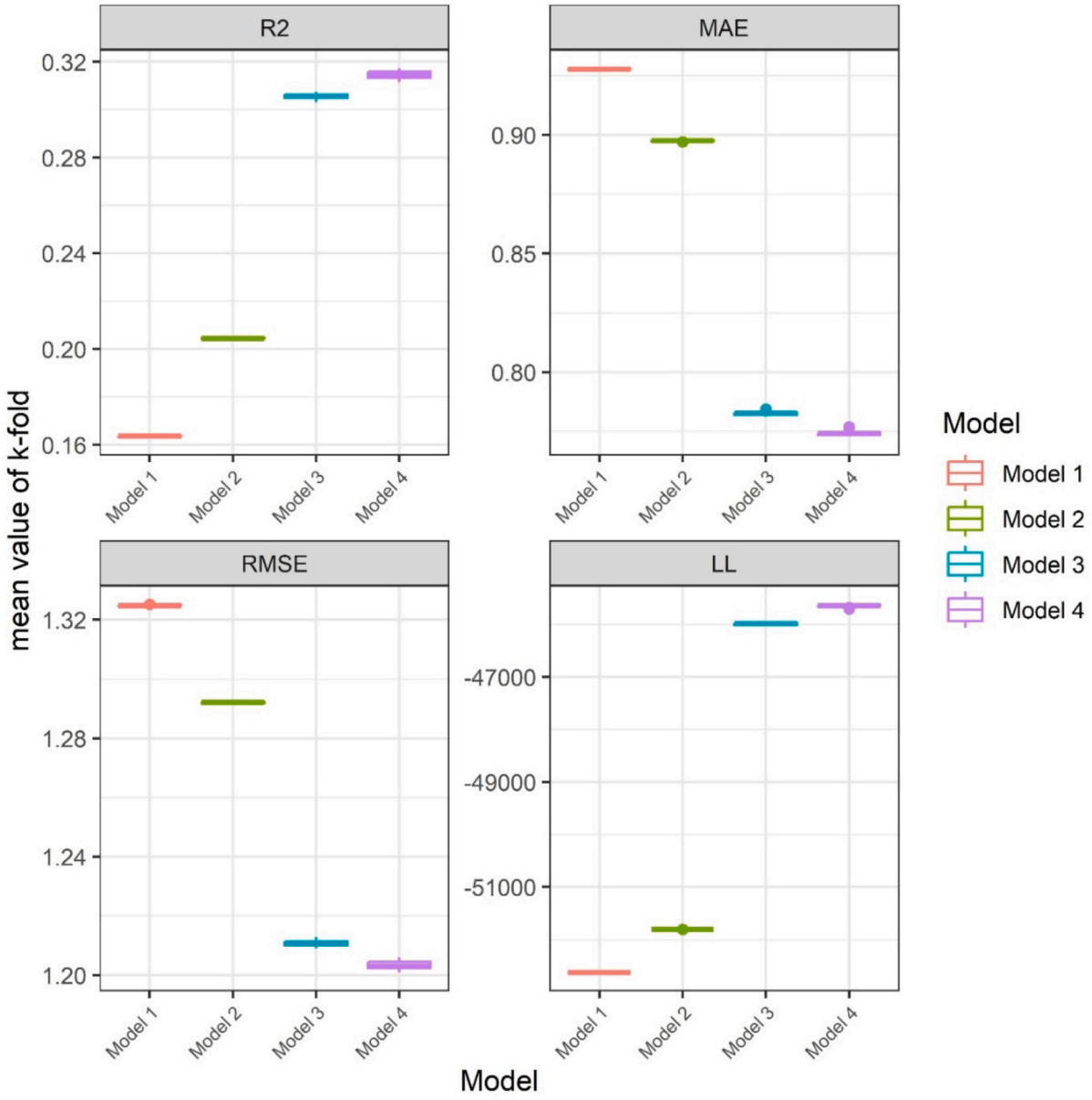


Fig. A11.2. 10 iterations, 70/30 % training/test data. The small variation amongst iterations were used to justify a smaller number of iterations for following models (Fig. A11.1) to save computation time.

A.12. Model validations

The study was performed within ICES statistical rectangles 42G7-44G7, in SD27 of the Central Baltic region.

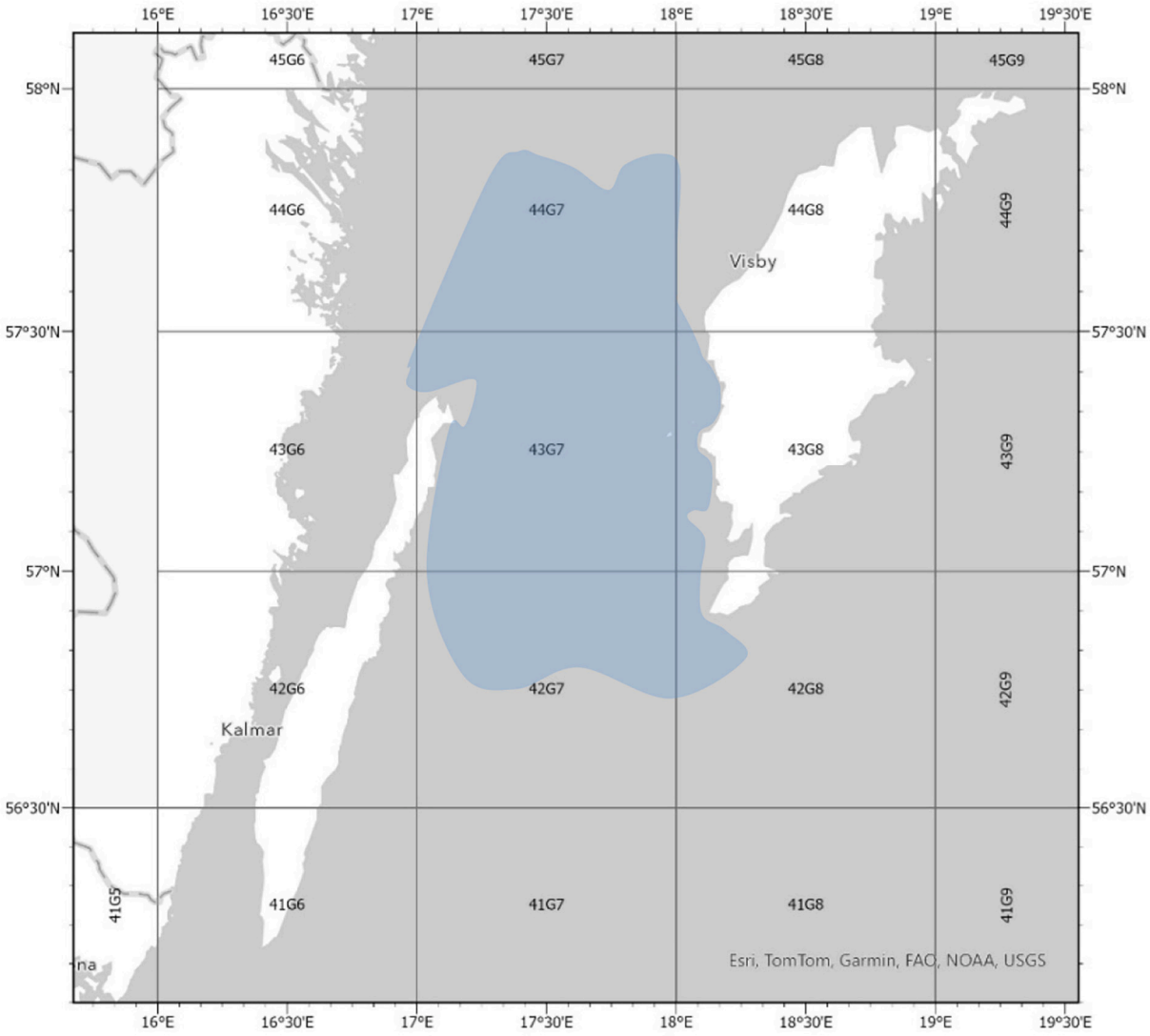


Fig. A12. ICES statistical units in the area surveyed with the USV (blue field) for reference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A.13. CMSI estimates of environmental variables

The CMSI environmental variables visualized on the area for reference.

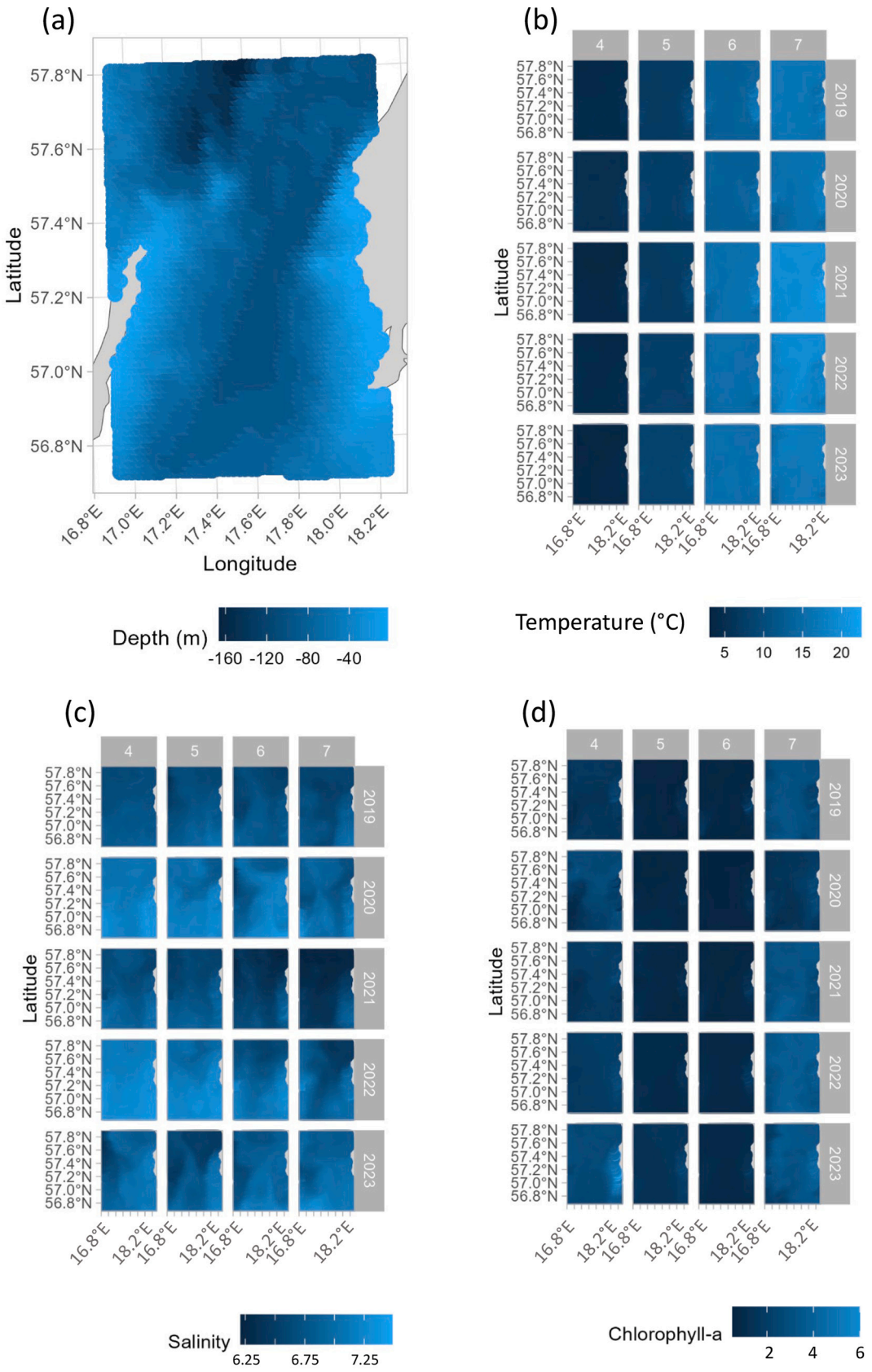


Fig. A13. Environmental variables from CMSI used for spatial prediction of fish distribution with monthly resolution: (a) depth, (b) temperature, (c) salinity, (d) chlorophyll, (e) north-south current, (f) east-west current and (g) current speed.

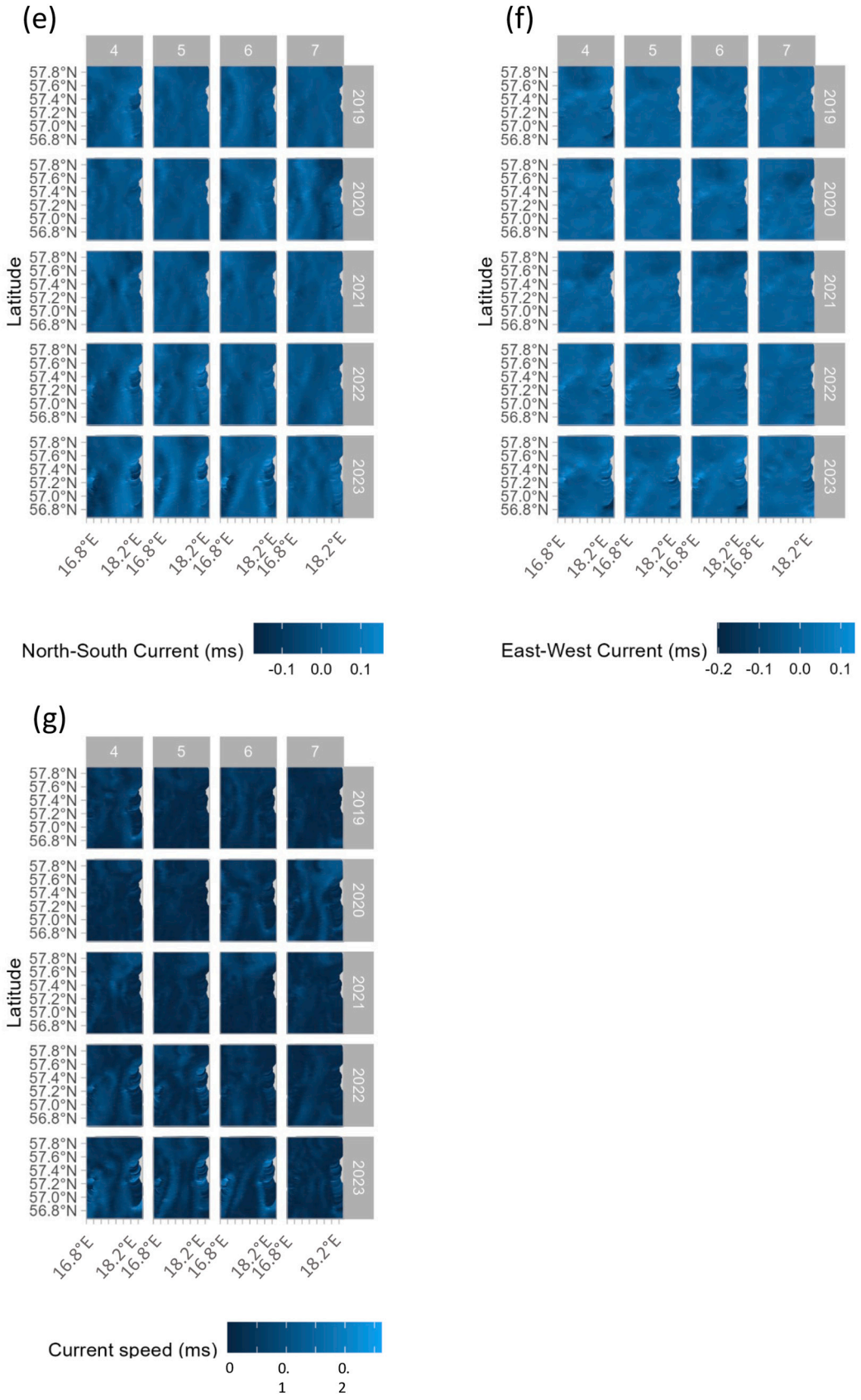


Fig. A13. (*continued*).

A.14. Data collection by month and upper quantiles (0.995) in raw data

The size of the areas sampled varied between months and years, where the wind direction and strength determined how large areas could be covered in the days of operation. Inspection of the distribution of upper quantiles in raw NASC values was needed when their spatial prediction was difficult to make (See Appendix A8 Fig. A8.3 and A8.4).

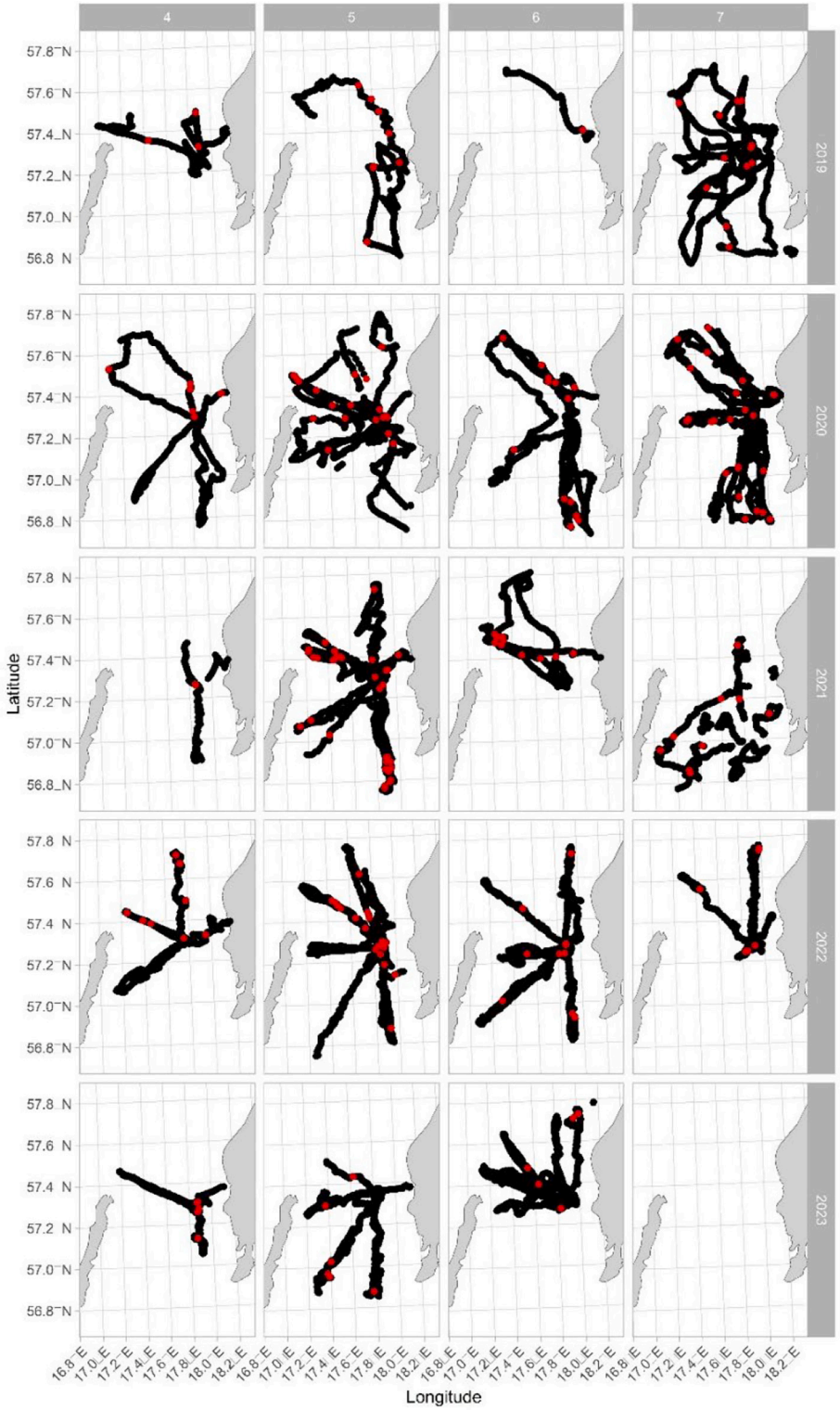


Fig. A14. All observations (black points) and upper 0.995 quantile observations (red points). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

Ahlgren, H., Bro-Jørgensen, M.H., Glykou, A., Schmölcke, U., Angerbjörn, A., Olsen, M. T., Lidén, K., 2022. The Baltic grey seal: a 9000-year history of presence and absence. Holocene 32 (6), 569–577. https://doi.org/10.1177/09596836221080764.

Anderson, S.C., Ward, E.J., English, P.A., Barnett, L.A.K., 2022. sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. BioRxiv 1–17. https://doi.org/10.1101/2022.03.24.485545, 2022.03.24.

Andersson, L., André, C., Johannesson, K., Pettersson, M., 2023. Ecological adaptation in cod and herring and possible consequences of future climate change in the Baltic Sea. Front. Mar. Sci. 10 (March), 1–8. https://doi.org/10.3389/fmars.2023.1101855.

Ando, K., Matsumoto, T., Nagahama, T., Ueki, I., Takatsuki, Y., Kuroda, Y., 2005. Drift characteristics of a moored conductivity – temperature – depth sensor and. J. Atmos. Ocean. Technol. 22, 282–292. https://doi.org/10.1175/JTECH1704.1.

Aro, E., 1989. A review of fish migration patterns in the Baltic. Rapports et Procès-Verbaux Des Réunions / Conseil Permanent International Pour l'Exploration de La Mer 190, 72–96. http://en.scientificcommons.org/26047672.

Benoit-Bird, K.J., Moline, M.A., Southall, B.L., 2017. Prey in oceanic sound scattering layers organize to get a little help from their friends. Limnol. Oceanogr. 62 (6), 2788–2798. https://doi.org/10.1002/lno.10606.

Bergström, U., Olsson, J., Casini, M., Eriksson, B.K., Fredriksson, R., Wennhage, H., Appelberg, M., 2015. Stickleback increase in the Baltic Sea - a thorny issue for coastal predatory fish. Estuar. Coast. Shelf Sci. 163, 134–142. https://doi.org/10.1016/j.ecss.2015.06.017.

Candolin, U., Engström-Öst, J., Salesto, T., 2008. Human-induced eutrophication enhances reproductive success through effects on parenting ability in sticklebacks. Oikos 117 (3), 459–465. https://doi.org/10.1111/j.2007.0030-1299.16302.x.

Cardinale, M., Casini, M., Arrhenius, F., Håkansson, N., 2003. Diel spatial distribution and feeding activity of herring (Clupea harengus) and sprat (Sprattus sprattus) in the Baltic Sea. Aquat. Living Resour. 16 (3), 283–292. https://doi.org/10.1016/S0990-7440(03)00007-X.

Carlsen, A.A., Lorentsen, S.H., Mattisson, J., Wright, J., 2023. Temporal non-independence of foraging dive and surface duration sequences in the European shag Gulosus aristotelis. Ethology 1–15. https://doi.org/10.1111/eth.13362. April 2022.

Chaalali, A., Beaugrand, G., Raybaud, V., Lassalle, G., Saint-Béat, B., Le Loc'h, F., Bopp, L., Tecchio, S., Safi, G., Chifflet, M., Lobry, J., Niquil, N., 2016. From species distributions to ecosystem structure and function: a methodological perspective. Ecol. Model. 334, 78–90. https://doi.org/10.1016/j.ecolmodel.2016.04.022.

Charnov, E.L., 1976. Optimal foraging, the marginal value theorem. Theor. Popul. Biol. 9 (2), 129–136. https://doi.org/10.1016/0040-5809(76)90040-X.

Cury, P.M., Boyd, I.L., Bonhommeau, S., Anker-Nilssen, T., Crawford, R.J.M., Furness, R. W., Mills, J.A., Murphy, E.J., Österblom, H., Paleczny, M., Piatt, J.F., Roux, J.-P., Shannon, L., Sydeman, W.J., 2011. Global seabird response to forage fish depletion—one-third for the birds. Science 334 (December), 1–23. https://doi.org/10.1126/science.1212928.

De Robertis, A., Levine, M., Lauffenburger, N., Honkalehto, T., Ianelli, J., Monnahan, C. C., Towler, R., Jones, D., Stienessen, S., McKelvey, D., 2021. Uncrewed surface vehicle (USV) survey of walleye Pollock, Gadus chalcogrammus, in response to the cancellation of ship-based surveys. ICES J. Mar. Sci. 78 (8), 2797–2808. https://doi.org/10.1093/icesjms/fsab155.

Didrikas, T., Hansson, S., 2004. In situ target strength of the Baltic Sea herring and sprat. ICES J. Mar. Sci. 61 (3), 378–382. https://doi.org/10.1016/j.icesjms.2003.08.003.

DuFour, M.R., Kocovsky, P.M., Deller, J., Simonin, P.W., Rudstam, L.G., 2021. Hydroacoustic survey standardization: inter-vessel differences in fish densities and potential effects of vessel avoidance. Fish. Res. 239 (October 2020), 105948. https://doi.org/10.1016/j.fishres.2021.105948.

Durant, J.M., Hjermann, D., Anker-Nilssen, T., Beaugrand, G., Mysterud, A., Pettorelli, N., Stenseth, N.C., 2005. Timing and abundance as key mechanisms affecting trophic interactions in variable environments. Ecol. Lett. 8 (9), 952–958. https://doi.org/10.1111/j.1461-0248.2005.00798.x.

Egerton, J.P., Johnson, A.F., Turner, J., LeVay, L., Mascareñas-Osorio, I., Aburto-Oropeza, O., 2018. Hydroacoustics as a tool to examine the effects of marine protected areas and habitat type on marine fish communities. Sci. Rep. 8 (1), 1–12. https://doi.org/10.1038/s41598-017-18353-3.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159.

Evans, T.J., Kadin, M., Olsson, O., Åkesson, S., 2013. Foraging behaviour of common murres in the Baltic Sea, recorded by simultaneous attachment of GPS and time-depth recorder devices. Mar. Ecol. Prog. Ser. 475, 277–289. https://doi.org/10.3354/meps10125.

Fässler, S.M.M., Gorska, N., Ona, E., Fernandes, P.G., 2008. Differences in swimbladder volume between Baltic and Norwegian spring-spawning herring: consequences for mean target strength. Fish. Res. 92 (2–3), 314–321. https://doi.org/10.1016/j.fishres.2008.01.013.

Fauchald, P., 2009. Spatial interaction between seabirds and prey: review and synthesis. Mar. Ecol. Prog. Ser. 391, 139–151. https://doi.org/10.3354/meps07818.

Fey, D.P., 2001. Differences in temperature conditions and somatic growth rate of larval and early juvenile spring-spawned herring from the Vistula Lagoon, Baltic Sea manifested in the otolith to fish size relationship. J. Fish Biol. 58 (5), 1257–1273. https://doi.org/10.1006/jfbi.2000.1529.

Fraser, S., Nikora, V., Williamson, B.J., Scott, B.E., 2017. Automatic active acoustic target detection in turbulent aquatic environments. Limnol. Oceanogr. Methods 15 (2), 184–199. https://doi.org/10.1002/lom3.10155.

Fretwell, S.D., Calver, J.S., 1969. On territorial behavior and other factors influencing habitat distribution in birds - II. Sex ratio variation in the Dickcissel (Spiza americana Gmel). Acta Biotheor. 19 (1), 37–44. https://doi.org/10.1007/BF01601954.

Galatius, A., Teilmann, J., Dähne, M., Ahola, M., Westphal, L., Kyhn, L.A., Pawliczka, I., Olsen, M.T., Dietz, R., 2020. Grey seal Halichoerus grypus recolonisation of the southern Baltic Sea, Danish Straits and Kattegat. Wildl. Biol. 2020 (4). https://doi.org/10.2981/wlb.00711.

Ghani, M.H., Hole, L.R., Fer, I., Kourafalou, V.H., Wienders, N., Kang, H.S., Drushka, K., Peddie, D., 2014. The sailBuoy remotely-controlled unmanned vessel: measurements of near surface temperature, salinity and oxygen concentration in the northern gulf of mexico. Methods Oceanogr. 10, 104–121. https://doi.org/10.1016/j.mio.2014.08.001.

Giske, J., Huse, G., Fiksen, Ø., 1998. Modelling spatial dynamics of fish. Rev. Fish Biol. Fish. 8, 57–91. https://doi.org/10.1023/A:1008864517488.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135 (2–3), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9.

Hentati-Sundberg, J., Österblom, H., Evans, T., Hjelm, J., 2018. Fish and seabird spatial distribution and abundance around the largest seabird colony in the Baltic. Mar. Ornithol. 46, 61–68. http://www.marineornithology.org/PDF/46_1/46_1_61-68.pdf.

Hilborn, R., Amoroso, R.O., Bogazzi, E., Jensen, O.P., Parma, A.M., Szuwalski, C., Walters, C.J., 2017. When does fishing forage species affect their predators? Fish. Res. 191, 211–221. https://doi.org/10.1016/j.fishres.2017.01.008.

Hughes, A.C., Orr, M.C., Ma, K., Costello, M.J., Waller, J., Provoost, P., Yang, Q., Zhu, C., Qiao, H., 2021. Sampling biases shape our view of the natural world. Ecography 44 (9), 1259–1269. https://doi.org/10.1111/ecog.05926.

ICES, 2013. ICES WKBALT REPORT 2013 Report of the Benchmark Workshop on Baltic Multispecies Assessments (WKSPRAT). Ices C.M, pp. 11–15. February.

ICES, 2021a. Herring (Clupea harengus) in subdivisions 25–29 and 32, excluding the Gulf of Riga (Central Baltic Sea). In: ICES Advisory Committee. https://doi.org/10.17895/ices.advice.4748.

ICES, 2021b. ICES Working Group on Baltic International Fish Survey (WGBIFS; outputs from 2020 meeting). In: ICES Scientific Reports, 3:02. 05(10). https://doi.org/10.17895/ices.pub.7679.

ICES, 2021c. Manual for acoustic sureys coordinated under ICES working group on acoustic and egg surveys for small pelagic fish (WGACEGG). In: ICES Techniques in Marine and Environmental Sciences. ICES Survey Protocols, Vol. 64. https://doi.org/10.17895/ices.pub.7462. Issue June.

ICES, 2023. Sprat (Sprattus sprattus) in subdivisions 22–32 (Baltic Sea). In: ICES Advisory Committee (Issue May 2023). https://doi.org/10.17895/ices.advice.4754.

Isaksson, N., Evans, T.J., Olsson, O., Åkesson, S., 2019. Foraging behaviour of razorbills Alca torda during chick-rearing at the largest colony in the Baltic Sea. Bird Study 66 (1), 11–21. https://doi.org/10.1080/00063657.2018.1563044.

Jørgensen, H.B.H., Hansen, M.M., Bekkevold, D., Ruzzante, D.E., Loeschcke, V., 2005a. Marine landscapes and population genetic structure of herring (Clupea harengus L.) in the Baltic Sea. Mol. Ecol. 14 (10), 3219–3234. https://doi.org/10.1111/j.1365-294X.2005.02658.x.

Jørgensen, H.B.H., Hansen, M.M., Loeschcke, V., 2005b. Spring-spawning herring (Clupea harengus L.) in the southwestern Baltic Sea: do they form genetically distinct spawning waves? ICES J. Mar. Sci. 62 (6), 1065–1075. https://doi.org/10.1016/j.icesjms.2005.04.007.

Kaartvedt, S., Christiansen, S., Titelman, J., 2023. Mid-summer fish behavior in a high-latitude twilight zone. Limnol. Oceanogr. 68 (7), 1654–1669. https://doi.org/10.1002/lno.12374.

Kadin, M., Österblom, H., Hentati-Sundberg, J., Olsson, O., 2012. Contrasting effects of food quality and quantity on a marine top predator. Mar. Ecol. Prog. Ser. 444, 239–249. https://doi.org/10.3354/meps09417.

Keogan, K., Daunt, F., Wanless, S., Phillips, R.A., Walling, C.A., Agnew, P., Ainley, D.G., Anker-Nilssen, T., Ballard, G., Barrett, R.T., Barton, K.J., Bech, C., Becker, P., Berglund, P.A., Bollache, L., Bond, A.L., Bouwhuis, S., Bradley, R.W., Burr, Z.M., et al., 2018. Global phenological insensitivity to shifting ocean temperatures among seabirds. Nat. Clim. Chang. 8 (4), 313–317. https://doi.org/10.1038/s41558-018-0115-z.

Křivan, V., Cressman, R., Schneider, C., 2008. The ideal free distribution: a review and synthesis of the game-theoretic perspective. Theor. Popul. Biol. 73 (3), 403–425. https://doi.org/10.1016/j.tpb.2007.12.009.

La Mesa, M., Catalano, B., Russo, A., Greco, S., Vacchi, M., Azzali, M., 2010. Influence of environmental conditions on spatial distribution and abundance of early life stages of antarctic silverfish, pleuragramma antarcticum (nototheniidae), in the Ross Sea. Antarct. Sci. 22 (3), 243–254. https://doi.org/10.1017/S0954102009990721.

Lauria, V., Garofalo, G., Fiorentino, F., Massi, D., Milisenda, G., Piraino, S., Russo, T., Gristina, M., 2017. Species distribution models of two critically endangered deep-sea octocorals reveal fishing impacts on vulnerable marine ecosystems in Central Mediterranean Sea. Sci. Rep. 7 (1), 1–14. https://doi.org/10.1038/s41598-017-08386-z.

Le Traon, P.Y., Reppucci, A., Fanjul, E.A., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando, V.E., Kreiner, M.B., Benkiran, M., Carval, T., Ciliberti, S.A., Claustre, H., Clementi, E., Coppini, G., Cossarini, G., De Alfonso Alonso-Muñoyerro, M., Delamarche, A., et al., 2019. From observation to information and users: the Copernicus marine service perspective. Front. Mar. Sci. 6 (May). https://doi.org/10.3389/fmars.2019.00234.

Lefébure, R., Larsson, S., Byström, P., 2011. A temperature-dependent growth model for the three-spined stickleback Gasterosteus aculeatus. J. Fish Biol. 79 (7), 1815–1827. https://doi.org/10.1111/j.1095-8649.2011.03121.x.

Lefébure, R., Larsson, S., Byström, P., 2014. Temperature and size-dependent attack rates of the three-spined stickleback (Gasterosteus aculeatus); are sticklebacks in the Baltic Sea resource-limited? J. Exp. Mar. Biol. Ecol. 451, 82–90. https://doi.org/10.1016/j.jembe.2013.11.008.

Liblik, T., Lips, U., 2019. Stratification has strengthened in the Baltic Sea – an analysis of 35 years of observational data. Front. Earth Sci. 7 (July), 1–15. https://doi.org/10.3389/feart.2019.00174.

Lindgren, F., Bakka, H., Bolin, D., Krainski, E., Rue, H., 2023. A diffusion-based spatio-temporal extension of Gaussian Matérn fields. Sort 48 (1). https://doi.org/10.57645/20.8080.02.13.

Liu, Z., Zhang, Y., Yu, X., Yuan, C., 2016. Unmanned surface vehicles: an overview of developments and challenges. Annu. Rev. Control. 41, 71–93. https://doi.org/10.1016/j.arcontrol.2016.04.018.

Lundström, K., Hjerne, O., Lunneryd, S.G., Karlsson, O., 2010. Understanding the diet composition of marine mammals: Grey seals (Halichoerus grypus) in the Baltic Sea. ICES J. Mar. Sci. 67 (6), 1230–1239. https://doi.org/10.1093/icesjms/fsq022.

Maathuis, M., Couperus, B., van der Molen, J., Poos, J., Tulp, I., Sakinan, S., 2023. Resolving the variability in habitat use by juvenile small pelagic fish in a major tidal system by continuous echosounder measurements. Mar. Ecol. Prog. Ser. SPF2. https://doi.org/10.3354/meps14368.

Maravelias, C.D., 1999. Habitat selection and clustering of a pelagic fish: effects of topography and bathymetry on species dynamics. Can. J. Fish. Aquat. Sci. 56 (3), 437–450. https://doi.org/10.1139/cjfas-56-3-437.

Maravelias, C.D., Reid, D.G., 1995. Relationship between herring (Clupea harengus, L.) distribution and sea surface salinity and temperature in the northern North Sea. Scientia Marina (Espana) 59 (3).

Mello, L.G.S., Rose, G.A., 2009. The acoustic dead zone: theoretical vs. empirical estimates, and its effect on density measurements of semi-demersal fish. ICES J. Mar. Sci. 66 (6), 1364–1369. https://doi.org/10.1093/icesjms/fsp099.

Michener, W.K., 2015. Ecological data sharing. Eco. Inform. 29 (P1), 33–44. https://doi.org/10.1016/j.ecoinf.2015.06.010.

Miller, J., 2010. Species distribution modeling. *Geography*. Compass 4 (6), 490–509. https://doi.org/10.1111/j.1749-8198.2010.00351.x.

Miller, D.L., Glennie, R., Seaton, A.E., 2020. Understanding the stochastic partial differential equation approach to smoothing. J. Agric. Biol. Environ. Stat. 25 (1), 1–16. https://doi.org/10.1007/s13253-019-00377-z.

Moody, A.L., Houston, A.I., Mcnamara, J.M., Ecology, S.B., 1996. Ideal free distributions under predation. Risk 38 (2), 131–143.

Muchowski, J., Jakobsson, M., Umlauf, L., Arneborg, L., Gustafsson, B., Holtermann, P., Humborg, C., Stranne, C., 2023. Observations of strong turbulence and mixing impacting water exchange between two basins in the Baltic Sea. Ocean Sci. 19 (6), 1809–1825. https://doi.org/10.5194/os-19-1809-2023.

Nøttestad, L., Misund, O.A., Melle, W., Hoddevik Ulvestad, B.K., Orvik, K.A., 2007. Herring at the Arctic front: influence of temperature and prey on their spatio-temporal distribution and migration. Mar. Ecol. 28 (Suppl. 1), 123–133. https://doi.org/10.1111/j.1439-0485.2007.00182.x.

Novotny, A., Jan, K.M.G., Dierking, J., Winder, M., 2022. Niche partitioning between planktivorous fish in the pelagic Baltic Sea assessed by DNA metabarcoding, qPCR and microscopy. Sci. Rep. 12 (1), 1–11. https://doi.org/10.1038/s41598-022-15116-7.

Ojaveer, E., Kalejs, M., 2010. Ecology and long-term forecasting of sprat (Sprattus sprattus balticus) stock in the Baltic Sea: a review. Rev. Fish Biol. Fish. 20, 203–217. https://doi.org/10.1007/s11160-009-9130-5.

Olin, A.B., Olsson, J., Eklöf, J.S., Eriksson, B.K., Kaljuste, O., Briekmane, L., Bergström, U., 2022. Increases of opportunistic species in response to ecosystem change: the case of the Baltic Sea three-spined stickleback. ICES J. Mar. Sci. 79 (5), 1419–1434. https://doi.org/10.1093/icesjms/fsac073.

Olsson, J., Jakubavičiūte, E., Kaljuste, O., Larsson, N., Bergström, U., Casini, M., Cardinale, M., Hjelm, J., Byström, P., Anderson, E., 2019. The first large-scale assessment of three-spined stickleback (Gasterosteus aculeatus) biomass and spatial distribution in the Baltic Sea. ICES J. Mar. Sci. 76 (6), 1653–1665. https://doi.org/10.1093/icesjms/fsz078.

Ona, E., 2003. An expanded target-strength relationship for herring. ICES J. Mar. Sci. 60 (3), 493–499. https://doi.org/10.1016/S1054-3139(03)00031-6.

Orio, A., Bergström, U., Florin, A.B., Lehmann, A., Šics, I., Casini, M., 2019. Spatial contraction of demersal fish populations in a large marine ecosystem. J. Biogeogr. 46 (3), 633–645. https://doi.org/10.1111/jbi.13510.

Österblom, H., Casini, M., Olsson, O., Bignert, A., 2006. Fish, seabirds and trophic cascades in the Baltic Sea. Mar. Ecol. Prog. Ser. 323, 233–238. https://doi.org/10.3354/meps323233.

Palermino, A., De Felice, A., Canduci, G., Biagiotti, I., Costantini, I., Centurelli, M., Menicucci, S., Gašparević, D., Tičina, V., Leonori, I., 2024. Modeling of the habitat suitability of European sprat (Sprattus sprattus, L.) in the Adriatic Sea under several climate change scenarios. Front. Mar. Sci. 11 (June), 1–16. https://doi.org/10.3389/fmars.2024.1383063.

Panzeri, D., Russo, T., Arneri, E., Carlucci, R., Cossarini, G., Isajlović, I., Krstulović Šifner, S., Manfredi, C., Masnadi, F., Reale, M., Scarcella, G., Solidoro, C., Spedicato, M.T., Vrgoč, N., Zupa, W., Libralato, S., 2023. Identifying priority areas for spatial management of mixed fisheries using ensemble of multi-species distribution models. Fish Fish. 00 (2023), 1–18. https://doi.org/10.1111/faf.12802.

Panzeri, D., Reale, M., Cossarini, G., Salon, S., Carlucci, R., Spedicato, M.T., Zupa, W., Vrgoč, N., Libralato, S., 2024. Future distribution of demersal species in a warming Mediterranean sub-basin. Front. Mar. Sci. 11 (May), 1–15. https://doi.org/10.3389/fmars.2024.1308325.

Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J., 2008. State-space models of individual animal movement. Trends Ecol. Evol. 23 (2), 87–94. https://doi.org/10.1016/j.tree.2007.10.009.

Pennino, M.G., Coll, M., Albo-Puigserver, M., Fernández-Corredor, E., Steenbeek, J., Giráldez, A., González, M., Esteban, A., Bellido, J.M., 2020. Current and future influence of environmental factors on small pelagic fish distributions in the northwestern Mediterranean Sea. Front. Mar. Sci. 7 (July), 1–20. https://doi.org/10.3389/fmars.2020.00622.

Phillips, L.R., Carroll, G., Jonsen, I., Harcourt, R., Brierley, A.S., Wilkins, A., Cox, M., 2022. Variability in prey field structure drives inter-annual differences in prey encounter by a marine predator, the little penguin. R. Soc. Open Sci. 9 (9). https://doi.org/10.1098/rsos.220028.

Pikitch, E.K., Rountos, K.J., Essington, T.E., Santora, C., Pauly, D., Watson, R., Sumaila, U.R., Boersma, P.D., Boyd, I.L., Conover, D.O., Cury, P., Heppell, S.S., Houde, E.D., Mangel, M., Plagányi, É., Sainsbury, K., Steneck, R.S., Geers, T.M., Gownaris, N., Munch, S.B., 2014. The global contribution of forage fish to marine fisheries and ecosystems. Fish Fish. 15 (1), 43–64. https://doi.org/10.1111/faf.12004.

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecol. Lett. 3 (4), 349–361. https://doi.org/10.1046/j.1461-0248.2000.00143.x.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P., Richardson, A.J., 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Glob. Ecol. Biogeogr. 20 (6), 789–802. https://doi.org/10.1111/j.1466-8238.2010.00636.x.

Rooper, C.N., Zimmermann, M., 2007. A bottom-up methodology for integrating underwater video and acoustic mapping for seafloor substrate classification. Cont. Shelf Res. 27 (7), 947–957. https://doi.org/10.1016/j.csr.2006.12.006.

Rousso, B.Z., Bertone, E., Stewart, R.A., Rinke, K., Hamilton, D.P., 2021. Light-induced fluorescence quenching leads to errors in sensor measurements of phytoplankton chlorophyll and phycocyanin. Water Res. 198, 117133. https://doi.org/10.1016/j.watres.2021.117133.

Sabatini, M., Reta, R., Matano, R., 2004. Circulation and zooplankton biomass distribution over the southern Patagonian shelf during late summer. Cont. Shelf Res. 24 (12), 1359–1373. https://doi.org/10.1016/j.csr.2004.03.014.

Schaeffer, B.A., Morrison, J.M., Kamykowski, D., Feldman, G.C., Xie, L., Liu, Y., Sweet, W., McCulloch, A., Banks, S., 2008. Phytoplankton biomass distribution and identification of productive habitats within the Galapagos marine reserve by MODIS, a surface acquisition system, and in-situ measurements. Remote Sens. Environ. 112 (6), 3044–3054. https://doi.org/10.1016/j.rse.2008.03.005.

Schneider, D., Piatt, J., 1986. Scale-dependent correlation of seabirds with schooling fish in a coastal ecosystem. Mar. Ecol. Prog. Ser. 32, 237–246. https://doi.org/10.3354/meps032237.

Simmonds, J., MacLennan, D., 2005. Fisheries acoustics. In: Fisheries Oceanography, 2nd ed. Blackwell Science.

Swart, S., Zietsman, J.J., Coetzee, J.C., Goslett, D.G., Hoek, A., Needham, D., Monteiro, P.M.S., 2016. Ocean robotics in support of fisheries research and management. Afr. J. Mar. Sci. 38 (4), 525–538. https://doi.org/10.2989/1814232X.2016.1251971.

Waldock, C., Stuart-Smith, R.D., Albouy, C., Cheung, W.W.L., Edgar, G.J., Mouillot, D., Tjiputra, J., Pellissier, L., 2022. A quantitative review of abundance-based species distribution models. Ecography 2022 (1), 1–18. https://doi.org/10.1111/ecog.05694.

Watson, J.R., Stock, C.A., Sarmiento, J.L., 2015. Exploring the role of movement in determining the global distribution of marine biomass using a coupled hydrodynamic - size-based ecosystem model. Prog. Oceanogr. 138, 521–532. https://doi.org/10.1016/j.pocean.2014.09.001.

Weidner, E., Stranne, C., Sundberg, J.H., Weber, T.C., Mayer, L., Jakobsson, M., 2020. Tracking the spatiotemporal variability of the oxic-anoxic interface in the Baltic Sea with broadband acoustics. ICES J. Mar. Sci. 77 (7–8), 2814–2824. https://doi.org/10.1093/icesjms/fsaa153.

Weimerskirch, H., 2007. Are seabirds foraging for unpredictable resources? Deep-Sea Res. II 54, 211–223. https://doi.org/10.1016/j.dsr2.2006.11.013.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., et al., 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol. Rev. 88 (1), 15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x.

Wullenweber, N., Hole, L.R., Ghaffar, P., Graves, I., Tholo, H., Camus, L., 2022. Current measurements. MDPI Sensors 22, 1–18. https://doi.org/10.3390/s22155553.