ECOGRAPHY

Research Article

Species distribution modeling with expert elicitation and Bayesian calibration

Karel Kaurila[1], **Sanna Kuningas**[2], **Antti Lappalainen**[2] **and Jarno Vanhatalo**[1,3]

[1]Department of Mathematics and Statistics, Faculty of Science, University of Helsinki, Helsinki, Finland
[2]Natural Resources Institute Finland, Helsinki, Finland
[3]Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Correspondence: Jarno Vanhatalo (jarno.vanhatalo@helsinki.fi) and Karel Kaurila (karel.kaurila@helsinki.fi)

Ecography
2026: e08173
doi: 10.1002/ecog.08173
Subject Editor:
F. Guillaume Blanchet
Editor-in-Chief: Miguel Araújo
Accepted 22 November 2025

Species distribution models (SDM) are key tools in ecology, conservation, and natural resources management. They are traditionally trained with data on direct species observations. However, if collecting species data is difficult or expensive, complementary information sources on species distributions are needed. Expert knowledge has been demonstrated to improve SDM predictions in a number of such applications but there is still no consensus on methods to integrate information from several experts into a single coherent species distribution prediction. Moreover, since expert assessments are inherently subjective and prone to biases, expert-driven SDMs should calibrate their assessments. We propose a method to tackle these challenges by extending the hierarchical Bayesian integrated species distribution modeling framework to expert informed species distribution modeling. We treated map-like expert assessments as data and integrated them with calibration data on species recordings. Our integrated SDM has model components to estimate experts' reliability and to adjust for potential biases in their assessments. After integrated inference, we used the model to make predictions over a study area. We tested our approach with an extensive simulation study and a real world case study comprising ten expert assessments and survey data on pikeperch larvae from a coastal area of the Gulf of Finland. Expert assessments significantly improved species distribution predictions compared to predictions conditioned on survey data only. They also improved parameter inference, thus strengthening the ecological interpretation of the results. The skill of the experts, and biases in their assessments, varied considerably in the case study though, emphasizing the importance of formal expert calibration provided by our model. Our results show that expert elicitation can be an efficient tool for improving species distribution model predictions. Our approach is especially useful for applications where any type of species data are expensive to collect but local species experts can easily be reached.

Keywords: Bayesian methods, expert calibration, expert elicitation, hierarchical model, integrated species distribution model, surveys

NORDIC SOCIETY OIKOS

www.ecography.org

Introduction

Species distribution models (SDMs) are key tools in ecology, conservation, and management of natural resources (Elith and Leathwich 2009). Optimally these models are informed by well designed survey data but, since organizing scientific surveys is expensive, there is growing interest in complementary data to train SDMs (Isaac et al. 2020). This interest has catalyzed development of integrated SDMs (ISDMs; Fletcher et al. 2019, Isaac et al. 2020, Foster et al. 2024), which combine different types of data to improve the accuracy and coverage of species distribution inference and predictions. A typical ISDM integrates survey data with presence-only data arising from a citizen science program (Fletcher et al. 2019, Foster et al. 2024). Other examples include, among others, integrating data from different taxonomic resolutions (Adjei et al. 2024) and integrating survey and experimental data (Kotta et al. 2019). Common to current ISDMs is, however, that they rely on recorded species observations. When considering species that are not of general interest to citizen science programs, or when interest is in areas that are especially hard to reach, any type of data might be scarce. In such cases, expert information has been proposed as a solution to improve species distribution assessments (Di Febbraro et al. 2018, Crawford et al. 2020, Pearman-Gillman et al. 2020). Even though expert information can be seen as a type of data (O'Hagan et al. 2006), expert-informed SDMs have not been formalized as ISDMs to date. Here, we propose a method to do this and show how it facilitates elicitation practices and adjusting for inaccuracies and biases in expert assessments.

In the context of species distribution assessments, expert elicitation has mainly been used in conservation and management applications with two main lines of approach. First, experts can provide constraints (Gobbi et al. 2012, Di Febbraro et al. 2018) or, in the context of Bayesian modeling, informative priors (Murray et al. 2009) for the parameters of a statistical model. The assumption underlying this approach is that experts have relevant information on limiting habitat and climatic conditions on species. Second, experts can inform on locations or areas where species are present or absent. These elicitation sites are then used as data to train an SDM. This approach is motivated by the observation that experts are most accurate within regions in which they have personal experience (Pearce et al. 2001, Di Febbraro et al. 2018, Crawford et al. 2020, Hurtado et al. 2023). Experts are also better in assessing their beliefs on real world, observable, variables than on parameters of statistical models (O'Hagan et al. 2006). Since SDM parameters are mathematical abstractions, whose interpretation requires expertise in modeling, experts typically struggle to provide information on them. However, drawing a map of a distribution range of a species, or pointing out locations of presence or absence of a species, is intuitively feasible for a species expert.

To make expert elicitation attainable to SDMs beyond specific case studies, an important open question is: how should we pool information from more than one expert into a single coherent species distribution prediction? Since humans are prone to psychological idiosyncrasies and subjective biases (Tversky and Kahneman 1974, Dias et al. 2018), expert assessments should also be calibrated and 'weighted' based on their trustworthiness (Murray et al. 2009, Di Febbraro et al. 2018, Perälä et al. 2020). The conceptual framework underlying ISDMs solves this challenge. By treating expert assessments as data and combining them with survey based calibration data within an ISDM is conceptually equal to the so-called supra-Bayesian framework for expert elicitation that has been shown to be probabilistically coherent (Lindley and Singpurwalla 1986) and efficient for expert calibration (Perälä et al. 2020). Hence, we propose a method for expert-informed species distribution modeling, that is a special case of a general hierarchical Bayesian ISDM (Miller et al. 2019, Isaac et al. 2020).

In our approach, the true underlying species distribution is modeled with a spatial (Poisson) point process to which the expert assessments and calibration data, obtained by survey sampling, are linked with different likelihood functions (Fig. 1). We collected expert assessments as expert-drawn maps to facilitate easy and intuitive reporting of their beliefs – both properties being important for successful expert elicitation process (O'Hagan et al. 2006, Dias et al. 2018, O'Hagan 2019). In this, our approach is similar to the approach used by Merow et al. (2017). However, unlike them, we did not require experts to assess the whole study region, but they could restrict their reporting into their self-assessed regions of expertise. In this respect, we followed the recommendations by Crawford et al. (2020) and Pearman-Gillman et al. (2020). Moreover, Merow et al. (2017) treated expert maps as offsets to Poisson point process, whereas our approach treats them through a likelihood function. Doing this we can calibrate expert assessments and, if needed, weight them in the inference and prediction based on their information content. We also allowed experts to express uncertainty in their assessments since this provides a more complete picture of their beliefs than hard cut division to presence and absence regions by range maps (O'Hagan et al. 2006, on benefits of eliciting distributions).

To motivate the proposed method and to test it with real world data, we applied it to modeling the distribution of newly hatched pikeperch *Sander lucioperca* larvae in the Gulf of Finland. We collected survey data on pikeperch larvae and elicited information on their distribution areas from ten experts. Our results show that combining these two datasets improves pikeperch larvae distribution predictions while simultaneously providing information on biases and skills of the experts. The performance of our method was also supported by a simulation study. The methods developed here will be useful in situations where survey data are expensive to collect but local expert information is easily available.

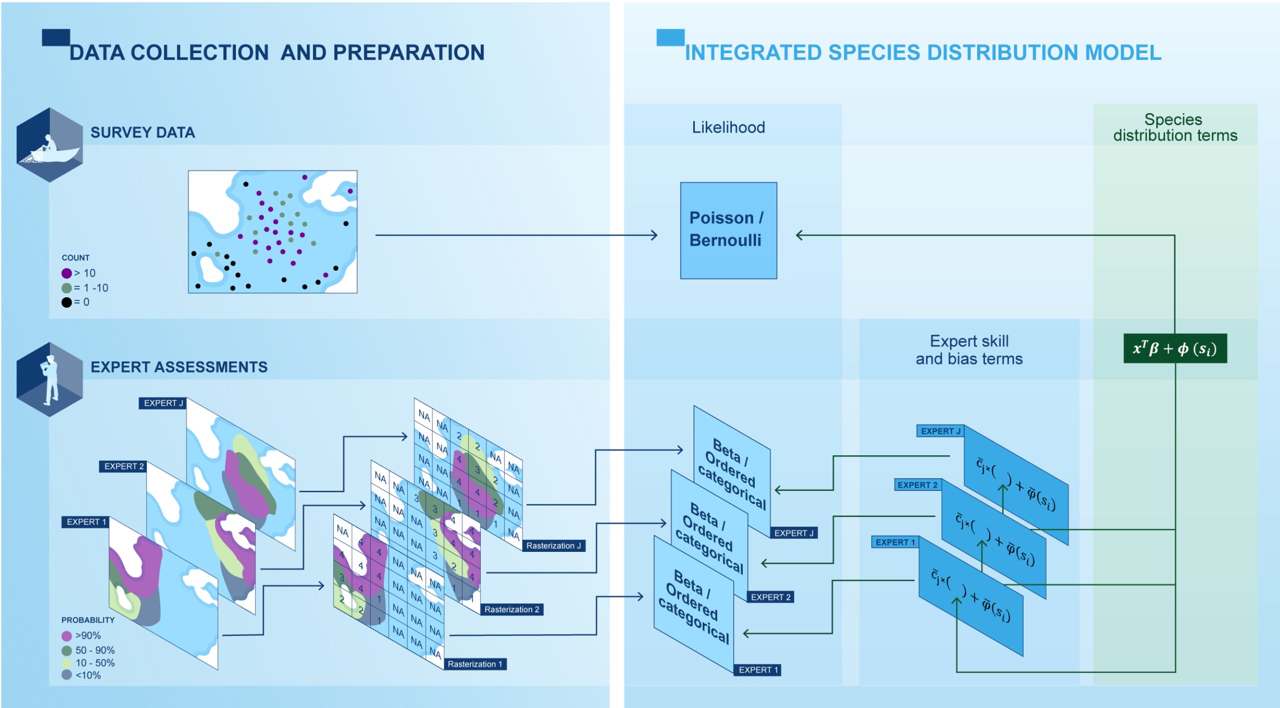

Figure 1. Visual representation of the expert elicitation process, survey data, and the integrated species distribution model used for analyzing these data. The survey data are point-wise abundance or presence–absence observations whereas expert assessments are drawings of areas within which the probability of a species presence is under 10, 10–50, 50–90%, or over 90%. The likelihood, expert skill and bias terms, and species distribution terms represent model components specified in the 'Integrated species distribution model' (we have omitted the offset and intercept parameters for clarity) section.

Motivating case study and data

Pikeperch survey data and environmental covariates

Pikeperch is a central species in the coastal ecosystem of the Baltic Sea and forms an important fish stock for commercial and recreational fisheries in the region. Knowledge about its spawning areas is therefore important for both economical and conservation reasons. We collected abundance data on pikeperch for this study from the Porvoo–Sipoo fisheries region (Fig. 2) by conducting a field survey of the surface water layer in June 2017 (Supporting information). We collected 156 samples dispersed over the entire study area, covering all main habitat types (Fig. 2). The survey design was optimized with respect to sampling weeks and spatial locations so that it would inform well the predictive surface for pikeperch larvae (Liu and Vanhatalo 2020). As a result, the sampling was scheduled around the peak larval season and dispersed roughly along the depth gradient with emphasis in shallow and intermediate depths. The exact sampling dates were adjusted to have minimal variability in catchability due to weather conditions.

The case study area consists of coastal environment types ranging from open water to sheltered bays. Pikeperch typically selects shallow (< 10 m deep), vegetated, and sheltered bays for its reproduction (Veneranta et al. 2011, Kallasvuo et al. 2017). Hence, we used depth (m), distance to deep water (> 10 m), and shoreline density (km m^{-2}) to characterize the environment in our case study. The latter two covariates are proxies for how close to shoreline and how sheltered a location is. Each covariate was available throughout the study area as a raster map with 50 m resolution (Kallasvuo et al. 2017).

Expert elicitation

In spring 2018, we elicited information concerning the pikeperch spawning areas from ten commercial fishermen (experts, hereafter) who had several decades of experience in active fishing in the study area and, hence, could be considered as regional experts (Supporting information). Even though fishermen cannot typically observe pikeperch spawning, they should be able to assess spawning areas based on their experience on where pikeperch concentrate during a spawning season. We first asked each expert to draw the borders of their assessment region to a map. This was the region within which an expert was confident to state their beliefs. After this, we asked each expert to mark areas within their assessment region where they believe the species does or does not exist, and to state the strength of their belief with probabilities (Fig. 1). However, since filling in a questionnaire map can be tedious, we sought for a compromise between assessment feasibility and the level of detail, ending up with four probability categories:

- Category 1: the probability of presence is less than 10%
- Category 2: the probability of presence is 10–50%

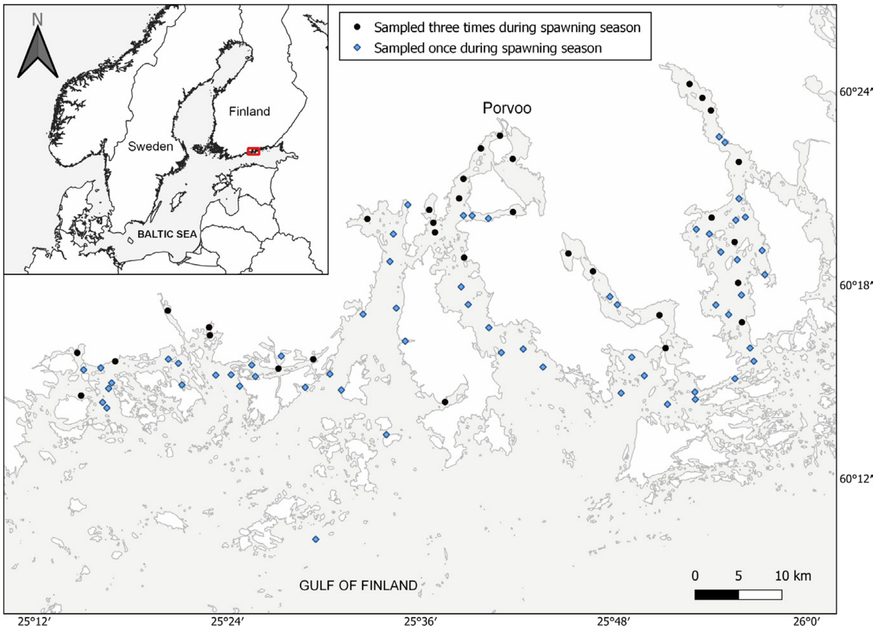

Figure 2. The pikeperch larvae case study area in Porvoo–Sipoo archipelago, in the Baltic Sea. The dots show the locations where we collected the survey data.

- Category 3: the probability of presence is 50–90%
- Category 4: the probability of presence is over 90%

We considered areas outside an expert's assessment region as missing information from that expert.

For logistical reasons, experts drew their answers on a printed map with crayons (Supporting information). However, a computer program could have been used as well (Murray et al. 2009, Pearman-Gillman et al. 2020). We digitized the expert-drawn paper maps by scanning them and using QGIS software to align them with the maps of the environmental covariates. We then formatted the expert maps as raster maps on the same lattice grid that we used for the environmental covariates (section 'Pikeperch survey data and environmental covariates'). Finally, we labeled grid cells on the assessment maps as one of the above four categories or NA for areas outside an expert's assessment region (Fig. 1).

Integrated species distribution model

To perform model-based integration of experts' map-like species distribution information with point-wise survey data, we followed the general approach in ISDMs whereby a point process model is used to describe the true distribution of a species, and the different types of data – survey data and expert assessments here – are linked to it by likelihood functions (Fig. 1; Fletcher et al. 2019, Miller et al. 2019, Isaac et al. 2020, Ahmad Suhaimi et al. 2021, Foster et al. 2024). We introduce this approach in the following subsections.

Survey data models

We modeled species abundance with a log Gaussian Cox process (LGCP) with an intensity function $\lambda(\mathbf{s},\mathbf{x}(\mathbf{s}))$, where $\mathbf{s}$ denotes a location inside the study area, and $\mathbf{x}(\mathbf{s})$ is a vector of environmental covariates at that location (Illian et al. 2013, Isaac et al. 2020). The specific form of the (log) intensity function varies among applications but as a baseline we considered a common linear mixed effects form:

$$\ln\lambda(\mathbf{s},\mathbf{x}) = \alpha + \mathbf{x}^T\beta + \phi(\mathbf{s}), \tag{1}$$

where α is the intercept, $\boldsymbol{\beta}$ is a vector of linear weights, and $\phi(\mathbf{s})$ is a spatial random effect capturing spatially correlated patterns not explained by the covariates. The model extends also to other types of covariate effects (e.g. polynomials or splines) and random effects (section 'Case study experiment on pikeperch larvae mapping'; Illian et al. 2013, Kallasvuo et al. 2017).

The observation model for the count data $\mathbf{y} = [y_1, \dots y_n]^T$ is then:

$$p(\mathbf{y} \mid \lambda, \mathrm{V}) = \prod_{i=1}^{n} \mathrm{Poisson}(y_i \mid V_i\lambda_i) \tag{2}$$

where V_i is an offset that accounts for among observations heterogeneity in sampling effort (Foster et al. 2024) – such as the volume of sampled water in our case study (Supporting information). Ecological data often exhibit excess zeros or over-dispersion compared to the variability captured by the Poisson distribution. These can be accounted for by replacing the Poisson distribution with a negative-binomial distribution, so that $p(\mathbf{y} \mid \lambda, \mathrm{V}, r) = \prod_{i=1}^{n} \mathrm{Negative-Binomial}(y_i \mid V_i\lambda_i, r)$, where r is an over-dispersion parameter (Lindén and

Mäntyniemi 2011). With small values of r the negative-binomial allows larger variance compared to Poisson distribution (i.e. captures overdispersion) whereas when r increases the distribution approaches Poisson distribution (Liu and Vanhatalo 2020; Supporting information).

A common alternative to count survey data are presence–absence data whereby $\mathbf{y}_i \in \{0,1\}$, which are typically modeled with a Bernoulli-likelihood:

$$y_i \sim \text{Bernoulli}\left(\pi\left(\mathbf{s}_i, \mathbf{x}_i\right)\right), \tag{3}$$

where $\pi(\mathbf{s}_i,\mathbf{x}_i)$ denotes the probability of presence for a species. This probability is related to the mean of a Poisson distribution via the well-known relation that motivates the complementary log-log link function $\ln\left(-\ln\left(1-\pi\left(\mathbf{s}_i,\mathbf{x}_i\right)\right)\right) = \ln\left(V_i\right) + \alpha + \mathbf{x}_i^T\beta + \phi\left(\mathbf{s}_i\right)$ (Phillips et al. 2017, Fletcher et al. 2019). However, if species counts follow a negative-binomial distribution, a more accurate relation between the expected count and the probability of presence is the logistic link, $\pi\left(\mathbf{s}_i,\mathbf{x}_i\right) = \text{logit}^{-1}\left(\ln\left(V_i\right) + \alpha + \mathbf{x}_i^T\beta + \phi\left(\mathbf{s}_i\right)\right)$, which gives the exact relationship between $V_i\lambda_i$ and π_i when the overdispersion parameter $r=1$ (Supporting information). In Bernoulli models, the log sampling effort, $\ln(v_i)$, is typically treated as a covariate (Foster et al. 2024).

Throughout this work, we assume covariates are standardized. Thus, we gave a zero mean Gaussian prior with a variance of four for the intercept and linear weights, encoding vague prior that restricts the scale of the covariate effects to be ecologically sensible. We gave a Gaussian process prior for the spatial random effects (sections 'Simulation experiments' and 'Case study experiment on pikeperch larvae mapping').

Expert assessment models

Cumulative Beta distribution model

We coded expert assessments as raster maps, in which the assessment from the j'th expert for a grid cell i at location $\mathbf{s}_i$ is a categorical variable $z_{ji} \in \{1,2,3,4\}$ if the grid cell belongs to the expert's assessment region (section 'Expert elicitation', Fig. 1). Otherwise $z_{ji}=\text{NA}$ and the grid cell is excluded from the likelihood. We ordered the categories so that $z_{ji}=1$ corresponds to the lowest (< 10%) and $z_{ji}=4$ corresponds to the highest (> 90%) subjective probability of presence of an expert.

To construct a model for expert assessments, we first denoted by $\bar{\pi}_{ji}$ the jth expert's subjective probability that a species is present at a location s_i. We then modeled this probability with a Beta distribution:

$$\bar{\pi}_{ji} \sim \text{Beta}\left(\bar{\mu}_{ji}, \bar{s}_j\right), \tag{4}$$

where $\bar{\mu}_{ji} = E\left[\bar{\pi}_{ji}\right]$, and $\bar{s}_j$ is the precision parameter governing the spread of the Beta distribution (Supporting information). We fixed $\bar{s}_j = 2$ to encode vague prior predictive distribution for expert opinions. After this, we calculated the probability that an expert reports a category. For example, for the first category $\Pr\left(z_{ji}=1\right) = \Pr\left(\bar{\pi}_{ji} \le 0.1\right) = F_{\text{Beta}}\left(0.1 \mid \bar{\mu}_{ji}, \bar{s}_j\right)$, where $F_{\text{Beta}}(\bullet|\mu_{ji},s_j)$ is the cumulative Beta distribution function. We calculated the observation probabilities for other categories analogously, giving us the full likelihood for expert assessments:

$$\Pr\left(z_{ji} \mid \bar{\mu}_{ji}, \bar{s}_j\right) = \begin{cases} F_{\text{Beta}}\left(0.1 \mid \bar{\mu}_{ji}, \bar{s}_j\right), & \text{if } z_{ji}=1 \\ F_{\text{Beta}}\left(0.5 \mid \bar{\mu}_{ji}, \bar{s}_j\right) - F_{\text{Beta}}\left(0.1 \mid \bar{\mu}_{ji}, \bar{s}_j\right), & \text{if } z_{ji}=2 \\ F_{\text{Beta}}\left(0.9 \mid \bar{\mu}_{ji}, \bar{s}_j\right) - F_{\text{Beta}}\left(0.5 \mid \bar{\mu}_{ji}, \bar{s}_j\right), & \text{if } z_{ji}=3 \\ 1 - F_{\text{Beta}}\left(0.9 \mid \bar{\mu}_{ji}, \bar{s}_j\right), & \text{if } z_{ji}=4. \end{cases} \tag{5}$$

Finally, to link the above likelihood function to the true intensity of a species, $\lambda(\mathbf{s},\mathbf{x}(\mathbf{s}))$ (or the true probability of presence), we wrote the expected value of an expert's probability as a function of the true intensity, so that:

$$\text{logit}\left(\bar{\mu}_{ji}\right) = \bar{\alpha}_j + \bar{c}_j\left(\mathbf{x}_i^T\beta + \phi\left(\mathbf{s}_i\right)\right) + \bar{\varphi}_j\left(\mathbf{s}_i\right). \tag{6}$$

Here, the parameter $\bar{\alpha}_j$ is an intercept, $\bar{c}_j$ is a parameter for the expert's skill, and $\bar{\varphi}_j\left(\mathbf{s}_i\right)$ is a residual error. Note that $\ln\lambda_i - \alpha = \mathbf{x}_i^T\beta + \phi\left(\mathbf{s}_i\right)$, so $\text{logit}(\mu_{ji})$ is proportional to $\ln\lambda_i$ (or $\text{logit}(\pi_i)$) but we have subtracted the intercept, α, to improve parameter identifiability. The parameter $\bar{c}_j$, thus, describes how strongly an expert's assessment follows the true underlying species intensity or presence probability, so that a positive $\bar{c}_j$ indicates that an expert has information about species distribution. The error terms $\bar{\varphi}_j\left(\mathbf{s}_i\right)$ adjust for spatially correlated local biases in expert assessments that can result from actual bias in an expert's opinion but also from inaccuracies in expressing them. For example, experts might draw the maps with coarser resolution than the true resolution in the species distribution pattern, leading to spatially correlated error which is captured by $\bar{\varphi}_j\left(\mathbf{s}_i\right)$. We have summarized priors for the expert skill and bias terms in section 'Priors for the expert skill and bias parameters'. Note that we could also use the complementary log–log link function in Eq. (6). However, as ecological data more often exhibit overdispersion than not, we chose to use the logit-link function, because of its connection to the negative-binomial count model (Supporting information).

As a special case of our model, we also considered binary expert statements whereby experts provide hard-cut presence or absence assessments. In this case, we treated an assessment for grid cell i as a presence if the expert gave it over 50% probability for species presence (i.e. $z_{ji} \in \{3,4\}$) and as an absence otherwise. The observation model for a binary presence assessment was then:

$$\Pr\left(z_{ji} \mid \bar{\mu}_{ji}, \bar{s}_j\right) = \begin{cases} F_{\text{Beta}}\left(0.5 \mid \bar{\mu}_{ji}, \bar{s}_j\right), & \text{if } z_{ji} \in \{1,2\} \\ 1 - F_{\text{Beta}}\left(0.5 \mid \bar{\mu}_{ji}, \bar{s}_j\right), & \text{if } z_{ji} \in \{3,4\}, \end{cases} \tag{7}$$

where we modeled parameters $\bar{\mu}_{ji}$ and $\bar{s}_j$ as in the previous four category case.

Ordered categorical likelihood for expert assessments

The cumulative Beta distribution likelihood (Eq. 5) is close to an ordered categorical model (Agresti 2010), also known as the ordered logit model. Hence, we considered it as an alternative approach for modeling the expert assessments. In this approach, we modelled assessments z_{ji} as:

$$z_{ji} \sim \text{Ordered} - \text{Categorical}\left(\bar{\mu}_{ij}, \boldsymbol{d}, \tilde{s} = 1\right), \tag{8}$$

where $\bar{\mu}_{ji}$ is the linear predictor as in the previous models, $\boldsymbol{d}$ is a vector of increasing cut-off values, such that $-\infty = d_1 < d_2 < d_3 < d_4 < d_5 = \infty$, and $\tilde{s}$ is a scale parameter. The resulting observation model for assessments is thus:

$$\Pr\left(z_{ji} = k \mid \bar{\mu}_{ji}, \boldsymbol{d}, \tilde{s}\right) = \text{logit}^{-1}\left(d_{k+1} - \bar{\mu}_{ji}\right) - \text{logit}^{-1}\left(d_k - \bar{\mu}_{ji}\right) = \frac{1}{1 + e^{\left(-d_{k+1} + \bar{\mu}_{ji}\right)/\tilde{s}}} - \frac{1}{1 + e^{\left(-d_k + \bar{\mu}_{ji}\right)/\tilde{s}}}. \tag{9}$$

We used the same cut-offs and scales for all experts to avoid overparametrizing the model, though the model could be extended to have heterogeneous experts with their own cut-offs d_{kj}. Following the implementation of the model in 'R-INLA' (section 'Posterior inference and model comparison'), the prior for the class probabilities was a weakly informative symmetric Dirichlet prior with concentration parameter equal to 2.

The Beta (Eq. 5) and the ordered categorical (Eq. 9) likelihoods differ to some extent in their functional form (Supporting information). However, a conceptual difference between the two lies in the cut-off hyperparameters: the probability limits are assumed known in the Beta likelihood whereas the cut-off parameters in the ordered categorical are estimated. Thus, the former encodes more prior information than the latter.

Priors for the expert skill and bias parameters

An important question in ISDMs is how to address biases arising from variation in the quality and amount of different data sources (Fletcher et al. 2019). In case of a conflict between an expert assessment and survey data, or among expert assessments, we wanted our model to down-weight the influence of the conflicting assessment in the joint likelihood (section 'Integrated likelihood'). To attain this behavior we gave relatively stricter prior for the expert skill parameters than for expert bias variables. The prior for skill parameters $\bar{c}_j$ was $N(0,0.5^2)$, implying 95% prior probability that $\bar{c}_j$ is less than one, and the prior for the expert bias function $\bar{\varphi}_j(\mathbf{s}_i)$ followed the so called Besag–York–Mollié (BYM) -spatial random field model (Besag et al. 1991) with wide prior variance (Supporting information). We used Gaussian priors for the skill parameters since they were supported by the integrated nested Laplace approximation (INLA) R package ('R-INLA'; Rue et al. 2009), which we used for implementing the model (section 'Posterior inference and model comparison'). This choice means that a skill parameter can become also negative which could happen if an expert would assess the relative differences between the areas opposite to the underlying truth.

Integrated likelihood

Conditionally on the model parameters, the survey data and expert assessments are mutually independent. Hence, for example, in case of count survey observations the joint model for all data over the whole study region is:

$$p\left(\mathbf{y}, \mathbf{z}_1, \ldots, \mathbf{z}_J \mid \bar{\alpha}, \bar{c}, \bar{\varphi}, \alpha, \beta, \varphi, r\right) = \left[\prod_{i=1}^{n} p\left(y_i \mid V_i \lambda_i, r\right)\right] \prod_{j=1}^{J} \prod_{k=1}^{n_j} p\left(z_{jk} \mid \bar{\alpha}_j, \bar{c}_j, \bar{\varphi}_j, \beta, \varphi\right). \tag{10}$$

where J is the number of experts, n_j is the number of expert mesh nodes inside the assessment area of the jth expert, $\bar{\alpha} = \{\bar{\alpha}_1, \ldots, \bar{\alpha}_J\}$, $\bar{c} = \{\bar{c}_1, \ldots, \bar{c}_J\}$ and $\bar{\varphi} = \{\bar{\varphi}_1, \ldots, \bar{\varphi}_J\}$. The full model with the presence–absence survey observations is analogous.

Posterior inference and model comparison

We implemented all the models and conducted the posterior inference using the 'R-INLA' package, which provides an efficient computational tool for point process models (Illian et al. 2013) and good flexibility for data integration applications (Isaac et al. 2020, Foster et al. 2024). Since 'R-INLA' does not directly support the exact likelihood functions derived from the Beta model (section 'Cumulative Beta distribution model'), we searched for the closest match to them from among the binomial likelihood functions which are supported by INLA. We explain the INLA implementation in detail in the Supporting information.

After setting up the models and conducting posterior inference, we used the model to predict the species intensity over the whole study area. The INLA approximation gives access to approximate leave-one-out (LOO) predictive densities for all observations. We used them to calculate the LOO cross-validation log predictive density (LPD) for model comparison. Additionally, in simulation experiments (section 'Simulation experiments'), we calculated the root mean squared error (RMSE) between the predicted and the true simulated log-intensity, as well as the LPD for the true log intensity, over the study area. Moreover, to assess the recoverability of the model parameters we also calculated the proportion of true parameter values inside their 95% central posterior credible intervals in the simulation experiments.

Simulation experiments

Simulation study setup

To assess our models in a controlled setting, we conducted two simulation experiments. In the first experiment, to assess models' predictive performance under different expert skills, we generated four alternative expert assessment scenarios: 1) three strongly informative experts (III), 2) three weakly informative experts (WWW), 3) three non-informative experts (NNN), and 4) one expert of each type (IWN). We first generated 100 true species intensity layers over an area of interest using the model (1) with randomly drawn parameter values. With each simulated species intensity, we sampled survey data at 50 randomly chosen locations by using a negative-binomial distribution and expert assessments by using the ordered categorical distribution. We used the case study area of size 48 × 33 km (Fig. 2) and its environmental covariate raster maps for generating the simulation data. The expert assessment maps were simulated with 1 km resolution whereas the survey data coordinates were continuous. We explain the simulation data in detail in the Supporting information. We then fitted four models to each simulated dataset: model for survey count data only (Survey (abu) model), model including survey count data and four class Beta-model for expert assessments (Survey (abu) + Beta model), model including survey count data and presence–absence model for expert assessments (Survey (abu) + Bernoulli model), and model including survey count data and ordered categorical model for expert assessments (Survey (abu) + Categorical model).

In the second simulation experiment, to assess how the amount of survey data impacts model performance, we generated datasets with 25, 50, 75, 100, and 125 survey data locations similarly to the first experiment. Moreover, to assess how the resolution of expert assessment maps impacts the model performance, we accompanied each of these five survey data scenarios with three IWN expert assessments of different resolutions. We generated the high resolution expert assessment maps at the original 1 km resolution; while using spatial moving average of the log species intensity, ln λ for the intermediate and low resolution assessment maps. That is, we generated each grid cell in these maps conditional on the average of 3 km (intermediate resolution) and 5 km (low resolution) neighborhood instead of the focal ln λ only. We demonstrate these neighborhoods and their effect on the simulated expert assessment maps in the Supporting information. As a result, we arrived at a full-factorial design of 15 different simulated data scenarios. We then compared Survey (abu) and Survey (abu) + Beta models with each simulated dataset. Note that, even though we are comparing models with different types of data (presence absence versus abundance survey data and 4-category versus 2-category expert data), the log predictive densities and RMSE errors for the true simulated log intensity are comparable among all models since the model component for the log intensity is the same in all of them.

Simulation study results

In the first simulation experiment, assessing models' predictive performance under different expert skills, the Survey (abu) + Beta and Survey + Bernoulli models had the best and the second best predictive performance in scenarios that included at least weakly informative experts (III, IWN, and WWW) whereas the Survey (abu) model had the best predictive performance in the scenario with only non-informative experts (NNN; Table 1). The Survey (abu) + Categorical had the worst predictive performance in all expert skill scenarios due to being numerically unstable. The instability of Survey (abu) + Categorical model likely arose from its additional threshold parameters ($\boldsymbol{d}$ in Eq. (9)) which were

Table 1. Posterior predictive model comparison for alternative models in the simulation study summarizing the median over 100 simulation runs for root mean squared error (RMSE), log predictive density (LPD), and coverage of the central 95% posterior credible interval for the model parameters.

				Coverage (%)					
		Predictive performance		Skill coefficient			Cov. effects		
Expert data scenario	Model	RMSE	LPD	c_1	c_2	c_3	β_1	β_2	β_3
–	Survey (abu)	1.15	−1.51	–	–	–	94	96	95
III	+ Bernoulli	1.15	−1.41	99	100	100	82	73	74
	+ Beta	0.78	−1.13	100	100	100	92	93	91
	+ Categorical	1.25	−1.72	99	100	100	90	90	69
WWW	+ Bernoulli	1.09	−1.45	100	100	100	91	90	91
	+ Beta	1.01	−1.45	99	99	99	88	91	90
	+ Categorical	1.24	−1.71	99	100	99	85	93	89
NNN	+ Bernoulli	1.69	−2.14	44	54	38	67	68	74
	+ Beta	1.51	−2.01	50	57	48	72	72	79
	+ Categorical	1.84	−2.48	49	59	47	67	62	78
IWN	+ Bernoulli	1.16	−1.45	100	99	42	95	86	90
	+ Beta	1.00	−1.45	100	100	55	83	86	90
	+ Categorical	1.24	−1.66	100	100	59	86	86	88

weakly recovered. Note though that the 'R-INLA' implementation of the ordered categorical model, and specifically the Dirichlet prior for its class probabilities, is still in experimental stage according to the 'R-INLA' documentation. Hence, we expect that the numerical challenges related to the Survey (abu) + Categorical model are related to the implementation details and not necessarily a property of the model itself. All models recovered other model parameters well in all scenarios except in the scenario with only non-informative experts (NNN) where models using expert assessments had worse coverage than the Survey (abu) model (Table 1).

In the second simulation experiment, assessing models' predictive performance under different amounts of survey data and different expert assessment resolutions, the information from expert assessments became less relevant when the amount of survey data increased (Supporting information). Moreover, the added value from expert assessments decreased as the resolution of the expert maps decreased. With high resolution expert maps, the Survey (abu) + Beta model was consistently the best model. With intermediate resolution expert maps, the Survey (abu) + Beta model outperformed Survey (abu) model with small amounts of survey data (25–75 survey locations) after which the Survey (abu) model was equally good or slightly better. With low resolution expert maps, the Survey (abu) + Beta model was consistently worse than the Survey (abu) model.

Case study experiment on pikeperch larvae mapping

Case study setup

To account for temporal variation in survey sampling times in the pikeperch larvae data, we added an independent Gaussian distributed week-wise random effects, $\omega(t_i)$, into the log intensity function so that:

$$\ln\lambda\left(\mathbf{s}_i, \mathbf{x}_i\right) = \alpha + \mathbf{x}_i^T\beta + \phi\left(\mathbf{s}_i\right) + \omega\left(t_i\right), \tag{11}$$

where t_i denotes the week when the i'th survey observation was sampled. However, since the experts were asked only about the spatial distribution of larvae, and not the temporal variation in their abundance, we linked expert means with the spatial components of the log intensity function only; that is, $\mathbf{x}^T\boldsymbol{\beta} + \phi(\mathbf{s})\mathbf{x}^T\beta + \phi\left(\mathbf{s}\right)$ as in Eq. (6). Hence, the temporal random effect can be interpreted as a temporal 'bias' in the survey data. Moreover, since the study area was scattered by islands and peninsulas, forming physical obstacles for pikeperch, we formulated the spatial random effect, $\phi(\mathbf{s})$, through the barrier Gaussian process model (Bakka et al. 2019, Supporting information).

To assess the value of expert information in pikeperch larvae predictions, we analyzed the data and calculated the (approximate) LOO cross-validation predictive performance on survey data with six models: Survey (abu) model, a model for survey presence–absence data only (Survey (p/a) model), Survey (abu) + Bernoulli model, Survey (abu) + Beta model, Survey (p/a) + Bernoulli model, and Survey (p/a) + Beta model. We left out the ordered categorical model from the comparison due to its poor numerical performance (section 'Simulation study results'). Note that the LOO-CV log predictive densities on survey data are comparable among all Survey (p/a) and among all Survey (abu) models but not between them, since the survey data in these two groups of models are of different types; in the former we used presence–absence information only whereas in the latter we used count observations.

To further assess the impact of the size of the survey sample on model performance, we trained Survey (abu) and Survey (abu) + Beta models with survey data subsets of sizes 25, 50, 75, 100, and 125 (training data) and compared their predictive performance on the remaining 31 survey samples (test data). To reduce the effect of randomness in the comparison, we repeated the comparison for 100 random splits of the survey data so that, at each iteration, the smaller training datasets were subsets of the larger training datasets. We also conducted prior sensitivity analyses for the bias and skill parameters by altering their relative informativeness (Supporting information).

Case study results

The expert assessments provided additional information on pikeperch larval areas compared to using survey data only. All the models predicted similarly highest larvae intensities in the sheltered bays in the north and smallest intensities in open sea areas in the south (Fig. 3; Supporting information). However, the models combining survey data and expert assessments predicted larger differences between the north and south and reduced the predictive uncertainty compared to Survey only models, thus providing more precise predictions. These differences were also reflected in the parameter estimates, which were sharper in models combining survey data and expert assessments than in the models using survey data only (Fig. 4a), thus strengthening the ecological interpretation of the results. The spatial (barrier) Gaussian process, $\phi(\mathbf{s})$, was also estimated more precisely and had relatively larger effect in explaining the larval distribution in the models combining survey and expert data than in the models using survey data only. This is clearly visible as local differences in larval intensity predictions in a few bays in the north (Fig. 3), demonstrating that experts provided useful local information that was not attainable from spatially sparse survey sampling data only.

Not all experts were informative though, since only five out of ten experts had positive skill parameters (Fig. 4b), whereas the informative experts were mostly unbiased (Supporting information). The areas assessed and uncertainties expressed by the experts also varied considerably. Six of the experts expressed their views for small regions only, whereas four experts covered more than a third of the study area (Supporting information). Likewise, only one expert used all four categories while others used only categories one (< 10% probability), two (10–50% probability), and three (50–90%

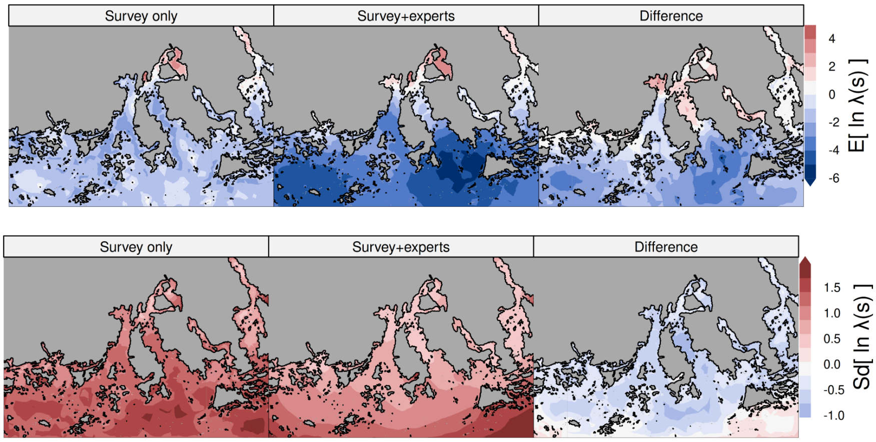

Figure 3. Posterior predictive mean (top row) and standard deviation (bottom row) for pikeperch log larval density in the study area with Survey only model and Survey + experts model and their difference ([Survey + experts] − [Survey only]). The survey data were species counts modeled with negative binomial likelihood and the expert assessments were probabilities of species presence divided into four categories and modeled with a Beta-likelihood.

probability), indicating that experts were more confident to assess pikeperch larvae absence than presence.

In the posterior predictive model comparisons, models with both expert information and survey data outperformed the models with survey data only when survey sample size was small (Table 2). However, all the models had practically equally good LOO cross-validation log predictive density scores with the full data (Supporting information). The prior sensitivity analysis showed that the model's performance was more sensitive to the prior for the expert skill parameters than to the prior for the expert bias terms. As the prior for the expert skill parameters was narrowed from the baseline (the one used in the case study results), model's predictive performance decreased and the results obtained from it started to resemble the Survey only model. However, for any given prior for the expert skill parameters, the model's performance did not change considerably as the prior for the bias term was altered from its baseline. For detailed results see the Supporting information.

Discussion

We have presented a formal Bayesian approach to include expert information into species distribution mapping using the integrated species distribution modeling framework. Our approach integrates map-like expert information with point-wise survey data, while simultaneously calibrating and assessing the reliability of the expert information. In our experiments, expert assessments significantly improved

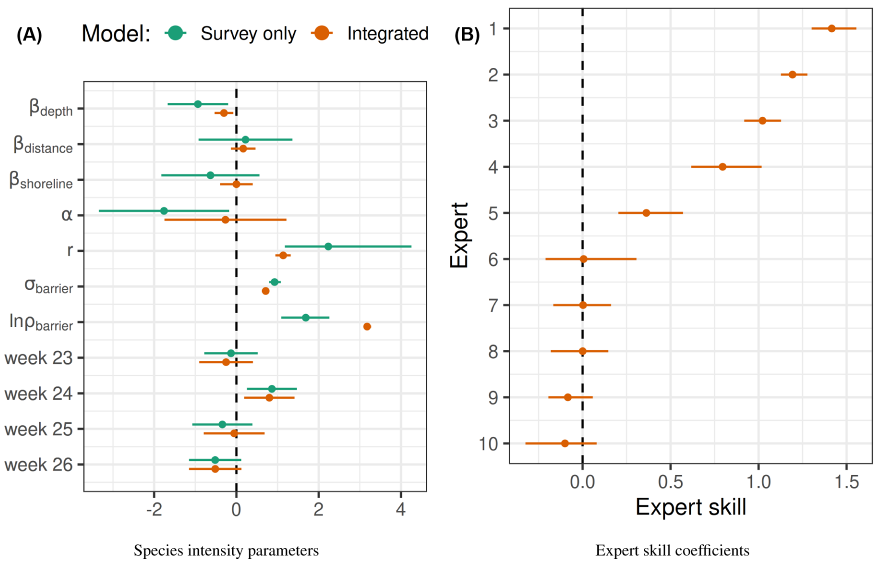

Figure 4. Posterior distributions (median with dot and central 95% credible interval with whiskers) for (A) the parameters of the Survey (abu) + Beta model integrating survey count data and (B) expert assessments in four probability categories.

Table 2. Model comparison results in the pikeperch larval area case study. The first 10 rows summarize the mean out-of-sample log predictive densities (LPD) for survey only data models (Survey (abu)) and models integrating survey data and expert information (Survey (abu) + Beta) with increasing amount of training data, n_{train}. The last two rows summarize the leave-one-out (loo) LPD for the same models with the full survey data.

Model	LPD	n_{train}
Survey (abu)	−9.48	25
Survey (abu) + Beta	−7.79	
Survey (abu)	−6.13	50
Survey (abu) + Beta	−4.47	
Survey (abu)	−4.23	75
Survey (abu) + Beta	−3.28	
Survey (abu)	−3.23	100
Survey (abu) + Beta	−2.56	
Survey (abu)	−2.38	125
Survey (abu) + Beta	−1.99	
	loo-LPD	
Survey (abu)	−1.58	Full data
Survey (abu) + Beta	−1.56	Full data

species distribution predictions compared to predictions conditioned on survey data only. The added value from expert information, however, diminished as the amount of survey data increased, suggesting that expert elicitation is most beneficial in applications where direct species observations are expensive to collect.

Our case study results show considerable variation among experts' skill, and thus emphasize the importance of formal calibration and weighting of expert information. This finding aligns with earlier studies that have demonstrated that the reliability of an expert in assessing species distributions depends, for example, on the study area, the species, and the background of an expert (Pearce et al. 2001, Di Febbraro et al. 2018, Crawford et al. 2020, Mainali et al. 2020, Hurtado et al. 2023). While the skill parameter in our model plays a similar role to expert weights in other expert integration approaches, such as weighted expert averaging, due to being a formal model parameter it provides a more coherent and transparent approach to expert weighting. Benefits of model-based expert calibration and weighting have been demonstrated also in other applications (O'Hagan et al. 2006, Burgman et al. 2011, Dias et al. 2018, Perälä et al. 2020). As the spatial extent of the expert assessments varied considerably, our results also accumulate the empirical evidence that experts should be allowed to define the spatial area of their expertise (Crawford et al. 2020, Pearman-Gillman et al. 2020). This would not be possible with methods that use expert maps as offsets in an SDM (Merow et al. 2017). Moreover, there was an asymmetry in experts' uncertainties about species presence versus absence since experts were more confident in assessing the latter than the former – a finding that supports the need for elicitation to capture experts' personal uncertainties (O'Hagan et al. 2006, O'Hagan 2019). Accounting for this asymmetry would not be possible with approaches that treat expert maps as offsets or take simple averages of expert assessments.

Based on our model comparisons, the model integrating survey abundance data and four category expert assessments performed the best both in the simulation and in the case study experiments. Models that treated survey data as presence–absence observations, or expert information as presence–absence statements, produced less accurate results. This is reasonable since compressing abundances to presence–absence observations and modifying a four-category expert statement to two-category statement decrease the information content of data. In the simulation study we were able to alter the informativeness and resolution of the expert assessments in a controlled manner and, thus, provide evidence for the technical performance of the model in alternative expert scenarios. All our model variants were able to identify the informative and uninformative experts in these tests. However, when all experts were uninformative, the predictive performance of the model combining survey data and expert assessments decreased compared to the model using survey data only. This is reasonable considering that the number of model parameters grows with the number of experts while uninformative experts do not improve the prediction. Therefore, we recommend that models combining expert and survey data should be compared to survey-only models to reveal situations where experts contribute little. On the other hand, the simulation tests indicate that already two at least weakly informative experts who assess relatively large area (in the simulations all expert assessments covered approximately 1/3 of the study area) can improve species distribution predictions while three informative experts can provide significant improvement. However, informativeness of an expert depends on the size of the area assessed as well as biases and resolution in the assessment. In the case study, skill parameters were estimated to be close to zero both for experts who provided biased assessments over a large area (e.g. experts 6 and 7) and for experts who provided non-biased assessment over a very small area (e.g. experts 8 and 10).

We implemented our models using the 'R-INLA' software (Rue et al. 2009) due to its flexibility for ISDMs (Isaac et al. 2020, Foster et al. 2024) and support for the barrier spatial random effect model. Because of this choice, we had to approximate the expert likelihood functions with the built-in likelihood functions in INLA (section 'Cumulative Beta distribution model'; Supporting information). We believe this choice had only a minor effect on the results, though. More fundamental considerations are related to the prior distributions and the resolution of the expert-mesh. Too narrow priors for the expert skill parameters and the bias terms would either prevent the model from identifying informative experts or calibrating their assessments. Moreover, the information content of expert assessments decreased as the expert-mesh resolution decreased in the simulation experiments. Hence, the expert-mesh resolution should correspond to the resolution at which the experts are confident in providing their assessment. The elicitation process could also be improved with digital map drawing tools and structured elicitation software (Murray et al. 2009, Pearman-Gillman et al. 2020) – even though, sometimes a low-tech approach can be beneficial to engage experts. Another development item could be to add correlation between the expert skill and bias terms to account for possible correlation among experts' arising from, for example, a shared institution or background knowledge, and to test how these correlations could affect the performance of the model.

In summary, our results are encouraging for further development of expert elicitation methods in species distribution modeling. They also suggest that the methods proposed in this work could be expanded to larger scale applications as well as to other ecosystems and species. Since the backbone of the method is a (point process based) SDM, the theoretical limits of the applicability of the method are similar to any SDM. We believe the method could also be used to integrate point-wise survey data with other types of map-like data with partial coverage over the study area, such as remote sensing maps or maps from earlier SDMs, since the former are conceptually very similar to expert maps in this work. Our approach seems especially useful for applications where any type of species data is expensive to collect but local species experts can be reached easily – such as natural resources management, pest control, and conservation area planning. One additional benefit from expert elicitation is stakeholder engagement. Stakeholders and other interest groups more readily accept management and policy decisions if they have been heard and their knowledge has been incorporated into the decision making process (La Mere et al. 2020). In our work, the experts were also stakeholders (fishermen) and their input clearly improved the larval area assessment. Hence, our method also provides a tool to increase the acceptance of management actions, such as restrictions in fishing and marine area use, in the identified pikeperch reproduction hot spots in our case study area.

Acknowledgements – We thank the executive director of the Porvoo–Sipoo fisheries region and all the fishermen who participated in the expert elicitation for their valuable contribution for this study. We also thank the anonymous reviewers for their insightful comments, which helped us improve the manuscript. Open access publishing facilitated by Helsingin yliopisto, as part of the Wiley - FinELib agreement.
Funding – This work has received funding from the Doctoral Programme in Mathematics and Statistics of the University of Helsinki (KK), Research Council of Finland (grant no. 317255; KK, JV), Jane and Aatos Erkko Foundation (KK, JV), the European Union's Horizon Europe Research and Innovation Programme (grant no. 101081642; KK, JV), the Finnish Operational Program for the European Maritime and Fisheries Fund (EMFF; AL and SK), and European Research Council (ERC Consolidator Grant BEFPREDICT, no. 101087409; JV).
Conflict of interest – The authors declare no conflict of interest.

Author contributions

Karel Kaurila: Data curation (equal); Formal analysis (lead); Investigation (equal); Methodology (equal); Software (lead); Validation (equal); Visualization (lead); Writing – original draft (equal); Writing – review and editing (supporting). **Sanna Kuningas**: Data curation (equal); Investigation (equal); Methodology (supporting); Writing – review and editing (supporting). **Antti Lappalainen**: Data curation (equal); Investigation (equal); Methodology (supporting); Writing – review and editing (supporting). **Jarno Vanhatalo**: Conceptualization (lead); Funding acquisition (lead); Methodology (equal); Project administration (lead); Supervision (lead); Validation (equal); Visualization (supporting); Writing – original draft (equal); Writing – review and editing (lead).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ecog.08173.

Data availability statement

Data are available from the Zenodo Digital Repository: https://doi.org/10.5281/zenodo.14781364 (Kaurila et al. 2025). Code is available from: https://doi.org/10.5281/zenodo.17964586.

Supporting information

The Supporting information associated with this article is available with the online version.

References

Adjei, K. P., Carvell, C., Isaac, N. J. B., Mancini, F. and O'Hara, R. B. 2024. Integrating data from different taxonomic resolutions to better estimate community alpha diversity. – Ecography 2024: e07182.

Agresti, A. 2010. Analysis of ordinal categorical data. – John Wiley and Sons.

Ahmad Suhaimi, S. S., Blair, G. S. and Jarvis, S. G. 2021. Integrated species distribution models: a comparison of approaches under different data quality scenarios. – Divers. Distrib. 27: 1066–1075.

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D. and Rue, H. 2019. Non-stationary Gaussian models with physical barriers. – Spat. Stat. 29: 268–288.

Besag, J., York, J. and Mollié, A. 1991. Bayesian image restoration, with two applications in spatial statistics. – Ann. Inst. Stat. Math. 43: 1–20.

Burgman, M., Carr, A., Godden, L., Gregory, R., McBride, M., Flander, L. and Maguire, L. 2011. Redefining expertise and improving ecological judgment. – Conserv. Lett. 4: 81–87.

Crawford, B. A., Maerz, J. C. and Moore, C. T. 2020. Expert-informed habitat suitability analysis for at-risk species assessment and conservation planning. – J. Fish Wildl. Manage. 11: 130–150.

Di Febbraro, M., Sallustio, L., Vizzarri, M., De Rosa, D., De Lisio, L., Loy, A., Eichelberger, B. A. and Marchetti, M. 2018. Expert-based and correlative models to map habitat quality: which gives better support to conservation planning? – Global Ecol. Conserv. 16: e00513.

Dias, L. C., Morton, A. and Quigley, J. 2018. Elicitation. – Springer.

Elith, J. and Leathwich, J. R. 2009. Species distributions models: ecological explanation and predictions across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

Fletcher, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A. and Dorazio, R. M. 2019. A practical guide for combining data to model species distributions. – Ecology 100: e02710.

Foster, S. D., Peel, D., Hosack, G. R., Hoskins, A., Mitchell, D. J., Proft, K., Yang, W. H., Uribe-Rivera, D. E. and Froese, J. G. 2024. 'RISDM': species distribution modelling from multiple data sources in R. – Ecography 2024: e06964.

Gobbi, M., Riservato, E., Bragalanti, N. and Lencioni, V. 2012. An expert-based approach to invertebrate conservation: identification of priority areas in central-eastern Alps. – J. Nat. Conserv. 20: 274–279.

Hurtado, C., Hemming, V. and Burton, C. 2023. Comparing wildlife habitat suitability models based on expert opinion with camera trap detections. – Conserv. Biol. 37: e14113.

Illian, J. B., Martino, S., Sørbye, S. H., Gallego-Fernández, J. B., Zunzunegui, M., Esquivias, M. P. and Travis, J. M. J. 2013. Fitting complex ecological point process models with integrated nested Laplace approximation. – Methods Ecol. Evol. 4: 305–315.

Isaac, N. J., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O. L., Schmucki, R., Simmonds, E. G. and O'Hara, R. B. 2020. Data integration for large-scale models of species distributions. – Trends Ecol. Evol. 35: 56–67.

Kallasvuo, M., Vanhatalo, J. and Veneranta, L. 2017. Modeling the spatial distribution of larval fish abundance provides essential information for management. – Can. J. Fish. Aquat. Sci. 74: 636–649.

Kaurila, K., Kuningas, S., Lappalainen, A. and Vanhatalo, J. 2025. Data from: Species distribution modeling with expert elicitation and Bayesian calibration. – Zenodo Digital Repository, https://doi.org/10.5281/zenodo.14781364.

Kotta, J., Vanhatalo, J., Jänes, H., Orav-Kotta, H., Rugiu, L., Jormalainen, V., Bobsien, I., Viitasalo, M., Virtanen, E., Sandman, A. N., Isaeus, M., Leidenberger, S., Jonsson, P. R. and Johannesson, K. 2019. Integrating experimental and distribution data to predict future species patterns. – Sci. Rep. 9: 1821.

La Mere, K., Mäntyniemi, S. and Haapasaari, P. 2020. The effects of climate change on Baltic salmon: framing the problem in collaboration with expert stakeholders. – Sci. Total Environ. 738: 140068.

Lindén, A. and Mäntyniemi, S. 2011. Using the negative binomial distribution to model overdispersion in ecological count data. – Ecology 92: 1414–1421.

Lindley, D. V. and Singpurwalla, N. D. 1986. Reliability (and fault tree) analysis using expert opinions. – J. Am. Stat. Assoc. 81: 87–90.

Liu, J. and Vanhatalo, J. 2020. Bayesian model based spatiotemporal survey designs and partially observed log Gaussian Cox process. – Spat. Stat. 35: 100392.

Mainali, K., Hefley, T., Ries, L. and Fagan, W. F. 2020. Matching expert range maps with species distribution model predictions. – Conserv. Biol. 34: 1292–1304.

Merow, C., Wilson, A. M. and Jetz, W. 2017. Integrating occurrence data and expert maps for improved species range predictions. – Global Ecol. Biogeogr. 26: 243–258.

Miller, D. A. W., Pacifici, K., Sanderlin, J. S. and Reich, B. J. 2019. The recent past and promising future for data integration methods to estimate species' distributions. – Methods Ecol. Evol. 10: 22–37.

Murray, J. V., Goldizen, A. W., O'Leary, R. A., McAlpine, C. A., Possingham, H. P. and Choy, S. L. 2009. How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. – J. Appl. Ecol. 46: 842–851.

O'Hagan, A. 2019. Expert knowledge elicitation: subjective but scientific. – Am. Stat. 73: 69–81.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. 2006. Uncertain judgements: eliciting experts' probabilities. – John Wiley and Sons.

Pearce, J. L., Cherry, K., Drielsma, M., Ferrier, S. and Whish, G. 2001. Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. – J. Appl. Ecol. 38: 412–424.

Pearman-Gillman, S. B., Katz, J. E., Mickey, R. M., Murdoch, J. D. and Donovan, T. M. 2020. Predicting wildlife distribution patterns in New England USA with expert elicitation techniques. – Global Ecol. Conserv. 21: e00853.

Perälä, T., Vanhatalo, J. and Chrysafi, A. 2020. Calibrating expert assessments using hierarchical Gaussian process models. – Bayesian Anal. 15: 1251–1280.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. and Blair, M. E. 2017. Opening the black box: an open-source release of maxent. – Ecography 40: 887–893.

Rue, H., Martino, S. and Chopin, N. 2009. Approximate bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). – J. R. Stat. Soc. B 71: 319–392.

Tversky, A. and Kahneman, D. 1974. Judgment under uncertainty: heuristics and biases. – Science 185: 1124–1131.

Veneranta, L., Urho, L., Lappalainen, A. and Kallasvuo, M. 2011. Turbidity characterizes the reproduction areas of pikeperch (*Sander lucioperca* L.) in the northern Baltic Sea. – Estuarine Coast. Shelf Sci. 95: 199–206.