Research

Ecography 40: 549–560, 2017
doi: 10.1111/ecog.01633

Subject Editor: Carsten Dormann. Editor-in-Chief: Miguel Araújo. Accepted 26 February 2016

Variable selection and accurate predictions in habitat modelling: a shrinkage approach

Matthieu Authier, Claire Saraux and Clara Péron

M. Authier (authier@gmail.com), Observatoire PELAGIS UMS 3462, Univ. de La Rochelle, La Rochelle, France. – C. Saraux, Ifremer (Inst. Français de Recherche pour l'Exploitation de la Mer), UMR MARBEC, Sète, France. – C. Péron, Inst. for Marine and Antarctic Studies, Univ. of Tasmania and Australian Antarctic Division, Kingston, TAS, Australia. MA and CP also at: Écologie Spatiale des Populations, Centre d'Écologie Fonctionnelle et Évolutive, Montpellier, France.

Habitat modelling is increasingly relevant in biodiversity and conservation studies. A typical application is to predict potential zones of specific conservation interest. With many environmental covariates, a large number of models can be investigated but multi-model inference may become impractical. Shrinkage regression overcomes this issue by dealing with the identification and accurate estimation of effect size for prediction. In a Bayesian framework we investigated the use of a shrinkage prior, the Horseshoe, for variable selection in spatial generalized linear models (GLM). As study cases, we considered 5 datasets on small pelagic fish abundance in the Gulf of Lion (Mediterranean Sea, France) and 9 environmental inputs. We compared the predictive performances of a simple kriging model, a full spatial GLM model with independent normal priors for regression coefficients, a full spatial GLM model with a Horseshoe prior for regression coefficients and 2 zero-inflated models (spatial and non-spatial) with a Horseshoe prior. Predictive performances were evaluated by cross-validation on a hold-out subset of the data: models with a Horseshoe prior performed best, and the full model with independent normal priors worst. With an increasing number of inputs, extrapolation quickly became pervasive as we tried to predict from novel combinations of covariate values. By shrinking regression coefficients with a Horseshoe prior, only one model needed to be fitted to the data in order to obtain reasonable and accurate predictions, including extrapolations.

Habitat models have become major tools in biodiversity and conservation studies (Record et al. 2013, Swanson et al. 2013). Habitat models or species distribution models (Guisan and Thuiller 2005, Elith and Leathwick 2009) aim at quantifying the statistical links between a set of relevant environmental covariates (inputs) and a species' occurrence, abundance or biomass (Reineking and Schröder 2005). The inferential target is the distribution of a focal species over a spatio-temporal domain. The ecological literature on habitat modelling is large, encompassing case studies (Wilson et al. 2010, Buckland et al. 2014) and technical papers about estimation methods (Reineking and Schröder 2005, Merow et al. 2014). Among practical hurdles, spatially-correlated errors (Dormann et al. 2007), variable selection, collinearity (Dormann et al. 2007, 2013), and the accurate quantification of predictive uncertainties (Rocchini et al. 2011) are recurrent. These issues are related to the generalizability of models outside their comfort zone (that is, the data used for calibration) to obtain accurate predictions for specific purposes (Vaughan and Ormerod 2005). An example of the latter would be a conservation study looking for current abundance maps of an endangered species (Conn et al. 2015). Another important application is predicting future change (Péron et al. 2012, Record et al. 2013, Swanson et al. 2013): models are used to extrapolate since future environmental conditions are expected to differ from past ones.

The generalizability issue betrays a tension between explanation and prediction (Mac Nally 2000). If recent works focused more explicitly on prediction (Elith and Leathwick 2009), explaining species distributions often remains an implicit goal. Ecologists wish to keep models interpretable in light of an underlying theory (niche theory; Austin 2007), and select models with respect to their predictive ability. Following model selection (Buckland et al. 1997, Burnham and Anderson 2002), predictions to unsampled areas are often attempted (Wilson et al. 2010, Mannocci et al. 2014). The rationale for the dual goals of explanation and prediction stems from the assumption that a causal model must have good prediction abilities (Shmuéli 2010). However, these two aims need not necessarily coincide: the best explanatory model can be different from the best predictive model (Betts et al. 2009, Shmuéli 2010).

Shmuéli (2010) lists 4 major sources of discrepancy between explanatory and predictive modelling, of which we will only focus on one: the bias-variance trade-off. Under a statistical model relating a response variable y to a set of predictors x such that the expectation $E[y] = f(x)$ (see notations below), the expected prediction error under a quadratic

loss function of a new observation y_{new} can be decomposed into (Shmuéli 2010): 1) a bias term quantifying the mismatch between the true value y_{new} and its predicted value $\hat{f}(x_{new})$; and 2) a variance term quantifying the uncertainties resulting from estimating $\hat{f}$ from a finite sample rather than knowing f.

The bias term quantifies the intrinsic error that arises even when f is known before hand or accurately estimated (Shmuéli 2010). Estimating $\hat{f}$ from a sample is the model selection problem, which generates additional uncertainty to be accounted for (Buckland et al. 1997, Burnham and Anderson 2002). The decomposition of expected prediction error highlights the difference between explanatory modelling whose aim is to minimize bias with respect to theory, and predictive modelling which is more concerned about minimizing both variance and bias (Betts et al. 2009, Shmuéli 2010). For prediction, introducing bias may be acceptable if it is more than offset by a reduction in variance (Shmuéli 2010): an obviously 'wrong' model can sometimes provide better predictions than a causally more correct model (Bahn and McGill 2007, Betts et al. 2009, Shmuéli 2010).

Willingly introducing bias to improve predictions is known as shrinkage regression (Reineking and Schröder 2005, Carvalho et al. 2010, Dahlgren 2010, Polson and Scott 2010). From a Bayesian perspective, all posterior estimates are biased: they are a weighted average between the data and the prior. With an appropriately chosen prior, regression coefficients can be shrunk toward zero to achieve variable selection. Such shrinkage priors encapsulate two assumptions: sparsity and robustness (Griffin and Brown 2013). Sparsity is the assumption that many coefficients may be set to zero without degrading the predictive ability of a model (Griffin and Brown 2013). Sparsity relates to parsimony and acts against over-fitting by muting irrelevant predictors. Robustness relates to the prior distribution tail fatness: a heavy-tailed prior allows a few regression coefficients to escape the gravitational pull towards 0 if the data suggest otherwise (Polson and Scott 2010). Shrinkage priors favour a parsimonious solution to the variable selection problem: the majority of regression coefficients will be close to zero, but a handful may be away from zero if they relate to influential predictors.

We will compare and discuss the use of shrinkage regression for predicting the habitat of three species of small pelagic fish in the Gulf of Lions, western Mediterranean, France. The Gulf of Lions underwent an ecosystem shift in 2007 with lower biomasses of important commercial fish species (sardine and anchovy) causing an important fishery crisis while sprat abundance rose (Saraux et al. 2014, Van Beveren et al. 2014). Habitat models can be useful tools to predict the spatial distribution of fish biomass and detect potential long-term changes driven by environmental forcing. Our focus was on the predictive accuracy of different habitat models evaluated on hold-out data. We considered spatially explicit models (Wikle 2003): ignoring spatial autocorrelation can result in unreliable inferences, biased environmental correlations and inaccurate predictions (Dormann 2007). To better understand the predictive ability of models, we explicitly distinguished between interpolation and extrapolation (Bahn and McGill 2013, Merow et al. 2014, Conn et al. 2015). We quantified how the relative interpolation to extrapolation ratio varies as more and more inputs were considered in habitats models with a non-parametric method based on Gower's distance (King and Zeng 2007 but see Conn et al. 2015).

Material

Small pelagic fish biomass

Yearly standardized acoustic and trawl surveys of small pelagic fish (PELMED, 'PÉLagiques MEDiterranée') have been carried out continuously every July since 1993 in the Gulf of Lions onboard RV L'Europe. Data were systematically collected along 9 parallel transects perpendicular to the coastline, 12 nautical miles (nm) apart (22 km). Transects are longer in the middle of the Gulf (max. 80 km), than on the edges (min. 20 km). The vessel navigated at depths between 20 and 200 m: beyond the 200 m isobaths, small pelagic fish abundance drops dramatically. Acoustic data were recorded every 1 nm (Elementary Unit Distance Sampling Unit, EDSU) using multi-frequency echosounders (Simrad EK500 and ER60), while traveling at constant speed of 8 nm h^{-1}. All frequencies were visualized during sampling to decide when to trawl for species identification.

When the echosounder detected sufficiently long fish traces (≥ 2 nm) or a change in the echotrace characteristics, a pelagic trawl was deployed to assess the species composition. The content was sorted and the total weight of every species caught measured. European anchovy *Engraulis encrasicolus*, European sardine *Sardina pilchardus* and European sprat *Sprattus sprattus* were classified into 0.5 cm size classes, and total weight and abundance were measured afterwards. When total catch was too large, a random subsample (5–10 kg) was taken. Only data collected in 2011 were used in this study, corresponding to a cruise of 1704 nm and 42 trawls performed between 26 June and 31 July.

To estimate fish density, only energies from the 38 kHz (typical frequency used for fish) channel were used. Acoustic data analyses, such as bottom correction, were performed using software Movies+ (Weill et al. 1993). Species discrimination and echo-partitioning were performed by the combination of echotrace classification and trawl outputs (Simmonds and MacLennan 2005). In particular, trawl-EDSU association was done by expert knowledge according to a combination of echotrace classification and minimizing distance. Fish size distributions or repartitions between adult and juvenile were attributed to each acoustic EDSU. Adult and juvenile fish were separated according to length-at-maturity (9 cm, 11 cm and 9 cm for anchovies, sardines and sprats respectively; Saraux unpubl.). Adult European sprats were too few in 2011 to be included in this study. Biomass was estimated from acoustic energy using specific target strength (Doray et al. 2010). Data are biomass per species and per stage (adult or juvenile) in each acoustic sampling point on the 9 transects (Table 1).

Table 1. Descriptive statistics of fish biomass (in tons).

Species	Stage	Dataset	Mean	Median	% zeros	Moran's I observed	Moran's I expected	p value
Anchovy	Juv.	calibration	2.7	0.1	11.0	−0.032	−0.005	2.0×10^{-16}
Anchovy	Ad.	calibration	3.9	1.0	22.5	−0.009	−0.005	2.0×10^{-01}
Sardine	Juv.	calibration	8.7	0.0	39.5	−0.026	−0.005	4.9×10^{-11}
Sardine	Ad.	calibration	2.3	0.1	27.5	−0.007	−0.005	4.8×10^{-01}
Sprat	Juv.	calibration	7.2	1.2	18.5	−0.036	−0.005	5.8×10^{-21}
Anchovy	Juv.	validation	2.7	0.2	5.0	−0.049	−0.014	1.5×10^{-04}
Anchovy	Ad.	validation	3.0	1.0	16.7	−0.009	−0.014	6.4×10^{-01}
Sardine	Juv.	validation	13.3	0.1	38.9	−0.033	−0.014	3.9×10^{-02}
Sardine	Ad.	validation	2.7	0.1	30.5	−0.025	−0.014	2.2×10^{-01}
Sprat	Juv.	validation	10.1	1.8	13.9	−0.055	−0.014	9.0×10^{-06}

Environmental inputs

Environmental inputs were cautiously selected either due to previous findings on sardine and anchovy distribution in the Mediterranean Sea (Giannoulaki et al. 2011, 2013, Tugores et al. 2011) or because they represented proxies of potentially important phenomena, such as mixing or frontal structures known to concentrate phytoplankton and zooplankton. Bathymetry (*Bathy*, in meters), its gradient (*GBathy*) and seabed substrate (*SED*, 5 categories) are topographic features affecting fish distribution (Giannoulaki et al. 2006). Inputs related to chlorophyll a concentrations (*CHLA* concentration in mg m^{-3}, gradient of *CHLA*, *GCHLA* and standard deviation of *CHLA*, *SDCHLA*) were proxies of the spatio-temporal dynamic of primary productivity which controls the distribution of zooplankton, the main prey of small pelagic fish in the Gulf of Lions (Palomera et al. 2007). The remaining inputs were related to sea surface temperature (*SST* in°C; gradient *GSST* and variance *SDSST*). They are dynamic predictors typically used to describe water masses that affect the distribution of primary productivity and low trophic levels.

Dynamic variables were downloaded on a monthly basis to match the response variable. Weekly grids of *SST* and *CHLA* were downloaded to compute monthly *SDSST* and *SDCHLA*, which reflected water masses' stability. Squared terms of *SST* and *CHLA* included to account for optima rather than linear increases or decreases. Spatial gradients were calculated with software Surfer. A low-pass filter (Gaussian 3×3 moving window) was first applied to *SST* and *CHLA* to remove high-frequency noise (measurements error) in satellite images. A gradient calculator was then used across the filtered *SST* and *CHLA* grids as well as *Bathy* grids using a 3×3 moving window. When environmental variables had different spatial resolution (Supplementary material Appendix 1, Table A1), raster grids were upscaled using bilinear interpolation to match a grid of $0.05\times0.05°$ (5.5 km) cell size. All predictors were assumed to be measured without error and values were extracted for each haul from the $0.05\times0.05°$ grid before standardisation. In total, we considered 9 different inputs: *Bathy*, *GBathy*, *SST*, *GSST*, *SDSST*, *CHLA*, *GCHLA*, *SDCHLA*, *SED* and squared terms for *SST* and *CHLA*, resulting in 15 predictors. Following King and Zeng (2007), inputs were distinguished from predictors: for example *SST* or *CHLA* were inputs contributing 2 predictors (*SST* and SST^2, *CHLA* and $CHLA^2$), and *SED* contributed 5 predictors.

Methods

We modelled the spatial distribution of small pelagic fish biomass (in tons) within a Bayesian framework. To assess collinearity, we computed the condition number of the predictor correlation matrix (κ), which is the largest to the smallest eigenvalue ratio. A large κ (>100) indicates collinearity. Biomass data were highly skewed with a large proportion of zeros or small values, and few large ones. Variance increased with the mean, a proportionality effect we removed with a logarithmic transform:

$$y = \log(1 + \text{Biomass}) \quad (1)$$

For each species and each size class, we partition the data (response variable and inputs) into 2 different sets: 1) a calibration dataset ($n = 200$, $\kappa = 22$) to estimate model parameters; and 2) a validation dataset ($n = 72$, $\kappa = 33$) to assess model predictive ability. One datum every 4 collected along transects was kept for validation. The 4:1 ratio between calibration and validation data is arbitrary, but was chosen to balance precise estimation and accurate validation of models. Upon model fitting, we predicted the response variable from inputs, and compared this prediction to actual biomass value. The latter was not seen by the model and provided 'independent' (with respect to parameter estimation) data to assess predictive ability. To further assess the quality of prediction, 10-fold cross validation was also performed by randomly splitting the data into calibration and validation datasets.

Finally, we used a third dataset for pure prediction ($n = 542$, $\kappa = 33$). We restricted this dataset to a grid of 0.05° cell size above the −200 m isobath. The final sample size of the prediction dataset was determined by the resolution of the chosen grid, and the availability of inputs. All these datasets showed moderate collinearity with condition numbers greater than 20 but lower than 100.

Notation

E[*y*] and *V*[*y*] are respectively the mean and variance of the random variable *y*. $U(l,u)$ denotes the uniform distribution with lower bound l and upper bound u. $N(\mu,\sigma)$ denotes the normal distribution of mean μ and scale σ.$C(\mu,\sigma)$ denotes the Cauchy distribution of mean μ and scale σ. $C^+(\mu,\sigma)$ denotes the positive half-Cauchy distribution of mean μ and scale σ. $T(\mu,\sigma,\nu)$ denotes the Student-t distribution of mean μ, scale σ and ν degrees of freedom.

Model building

We considered 5 spatial generalized linear mixed models. Spatial effects $z_{(s)}$ were modelled with a stationary Gaussian Process (*GP*) of mean 0 and isotropic Matèrn covariance function of order 3/2 (Juntunen et al. 2012). Schmidt et al. (2008) concluded of a limited influence of different correlation functions on predictions. The spatial correlation cor(s,s') between two locations s and s' separated by geodetic distance $d_{(s,s')}$ (Banerjee 2005) is:

$$\mathrm{cor}(s,s') = \left(1 + \frac{(d_{(s,s')} \times \sqrt{3})}{\rho}\right) \times e^{-(d_{(s,s')} \times \sqrt{3})/\rho} \quad (2)$$

where $\rho > 0$ is the range. The spatial covariance is $\mathrm{cov}(s,s') = \sigma^2_{\mathrm{sill}} \times \mathrm{cor}(s,s')$. Spatial effects $z_{(s)}$ were modelled (Banerjee 2005):

$$z_{(s)} \sim GP(0, \sigma^2_{\mathrm{sill}} \times \mathrm{cor}(.)) \quad (3)$$

Let $y_{(s)}$ denotes the value of the response variable and $x_{k_{(s)}}$ the value of the k^{th} predictor variable (or input) x at location s.

1) Model M_1 is a benchmark model without covariates but with a spatial effect ('simple kriging', Wikle 2003):

$$y_{(s)} \sim N(\beta_0 + z_{(s)}, \sigma_{\mathrm{nugget}}) \quad (4)$$

where β_0 is an intercept with a Student-t prior $\beta_0 \sim T(0,10,7)$ (Gelman et al. 2008). $\sigma^2_{\mathrm{nugget}}$ is the residual variance, or nugget effect in geostatistics.

2) Model M_2 is a full model including all p predictors and a spatial effect ('universal kriging', Wikle 2003):

$$y_{(s)} \sim N(\beta_0 + \sum\nolimits_{k=1}^{p} \beta_k \times x_{k_{(s)}} + z_{(s)}, \sigma_{\mathrm{nugget}}) \quad (5)$$

where β_k are regression coefficients (slope) with independent Student-t priors $\beta_k \sim T(0,2.5,7)$ (Gelman et al. 2008).

3) Model M_3 is similar to model M_2 with a shrinkage prior on β_k. A well-known shrinkage prior is the LASSO (Park and Casella 2008). One shortcoming of the latter is its light tails which can lead to excessive shrinkage. Priors with heavy tails, such as the Horseshoe prior (Carvalho et al. 2010), avoid this pitfall. The latter owes its name to the shape of the distribution it induces on the shrinkage coefficient (Supplementary material Appendix 1, Fig. A1). It has an infinite spike at zero (to screen out the noise) and heavy tails to accommodate true signals (which are left unshrunk). The Horseshoe prior favours a parsimonious variable selection by shrinking toward zero small β_k effects yet leaving larger β_k effects unshrunk. It belongs to the class of Global–Local shrinkage priors (Polson and Scott 2010). It involves a global variance $\sigma^2_{\mathrm{global}}$ that handles the noise, and local variances $\sigma^2_{\mathrm{local}_k}$ to identify robust signals; both having a standard half-Cauchy distribution:

$$\begin{cases} \beta_k \mid \sigma_{\mathrm{nugget}}, \sigma_{\mathrm{global}}, \sigma_{\mathrm{local}_k} \sim N(0, \sigma_{\mathrm{nugget}} \times \sigma_{\mathrm{global}} \times \sigma_{\mathrm{local}_k}) \\ \sigma_{\mathrm{local}_k} \sim C^+(0,1) \\ \sigma_{\mathrm{global}} \sim C^+(0,1) \end{cases} \quad (6)$$

4) Model M_4 is an elaboration of M_3, allowing for an excess of zeros (Martin et al. 2005): it is a zero-inflated spatial GLM with a Horseshoe prior.

$$\begin{cases} y_{(s)} \sim \theta_{(s)} \times \delta_0 + (1 - \theta_{(s)}) \times N(\beta_0 + \sum_{k=1}^{p} \beta_k \times x_{k_{(s)}} + Z_{(s)}, \sigma\mathrm{nugget}) \\ \theta_{(s)} \sim \mathrm{Bernoulli}(\Phi(\alpha_0 + \sum_{k=1}^{p} \alpha_k \times x_{k_{(s)}})) \end{cases} \quad (7)$$

where δ_0 is a degenerate distribution with all its mass at 0; $\Phi(.)$ is the standard normal cdf (probit link); α_0 is an intercept, and α_k are slope parameters. No spatial structure was assumed for the process generating excess zeros. Separate Horseshoe priors were assumed for α_k and β_k.

5) Finally, model M_5 is a simplified version of M_4 with no spatial effects:

$$\begin{cases} y_{(s)} \sim \theta_{(s)} \times \delta_0 + (1 - \theta_{(s)}) \times N(\beta_0 + \sum_{k=1}^{p} \beta_k \times x_{k_{(s)}}, \sigma_{\mathrm{nugget}}) \\ \theta_{(s)} \sim \mathrm{Bernoulli}(\Phi(\alpha_0 + \sum_{k=1}^{p} \alpha_k \times x_{k_{(s)}})) \end{cases} \quad (8)$$

Horseshoe priors were again assumed for slope parameters. M_5 is another benchmark model to assess whether the predictors x_k were sufficient for prediction.

Calibration data sample size was 200 with 9 inputs for fitting different model structure (spatial, zero-inflation): the range of the datum-to-parameter ratio for estimation is roughly between 50 (model M_1) and 6 (model M_4). A standard half-Cauchy prior was always assumed for σ_{nugget}. For spatial models, a standard half-cauchy prior was assumed for σ_{sill} and a uniform prior for $\rho \sim u$ (0.1,200). For M_2, M_3 and M_4, we assumed random spatial effects were uncorrelated to any of the inputs. Zero-inflated models M_4 and M_5 were heavily parametrized (large p, small n setting) and we only considered shrinkage priors.

Model fitting

We used cmdstan 2.8.0 (Stan Development Team 2015). Three chains were initialized with random starting values. After burn-in (1000 iterations), convergence was assessed using the Gelman–Rubin potential scale reduction factor $\hat{r}$ (Cowles and Carlin 1996). Inferences are based on a posterior sample of 3000 iterations. STAN codes are available as supplementary materials. Predictions were performed with R 3.2.1 (R Development Core Team) on a HP Compaq LA2306x desktop (Intel (R), Xeon (R) CPU E5-2630, 2.30 GHZ, 32 Go RAM).

Predictive ability

For each combination of species-stage, 75% of data were used for calibration, and the remaining 25% kept for cross-validation. To assess the usefulness of spatial effects, the validation sample was not selected at random, but chosen to reflect the data collection mechanism (Wang and Gelman 2014) in order to test specifically whether taking into account neighbouring locations improved predictions. We also performed 10-fold cross-validation by randomly splitting data, thereby ignoring the data collection design to confirm results. Model M predictive ability was assessed with posterior predictive loss (Gelfand and Ghosh 1998): $D_M = G_M + P_M$, where G_M measures goodness of fit and P_M penalizes models that make imprecise predictions.

For both the calibration and validation datasets, goodness of fit was:

$$G_M = \sum\nolimits_{(s)} (y^{\mathrm{obs}}_{(s)} - y^{\mathrm{pred}}_{(s)})^2 \quad (9)$$

and penalty (Gelfand and Ghosh 1998, Schmidt et al. 2008):

$$P_M = \sum_{(s)} V[y_{(s)}^{\text{pred}}] \quad (10)$$

G_M and P_M quantify respectively the bias and variance of model M predictions: D_M thus reflects a bias-variance trade-off (Shmuéli 2010).

Predictions: inter- or extrapolation

The relative contribution of inter- and extrapolation when predicting fish biomass at unsampled locations was quantified with King and Zeng (2007)' approach implemented in R package WhatIf (Stoll et al. 2009). Using Gower's distance, a combination of input values is determined to lie inside or outside the calibration dataset convex hull. Figure 1 illustrates how, albeit most data can fall within the range of each individual input x_k, the combination of several inputs can rapidly result in predicting outside the convex hull. We assessed how the ratio of interpolation to extrapolation changed as inputs were added. To reduce computation burden, we excluded seabed substrate.

Conn et al. (2015) assessed extrapolation using model-based convex hulls defined from a generalization of Cook's distance. In contrast the nonparametric approach of King and Zeng (2007) does not require any model fitting: only the inputs are required and the convex hull is model-independent.

Results

Convergence and model selection

Twenty-five models were fitted (5 for each species-stage combination), plus 10×25 for the 10-fold cross-validation. In each model, all parameters converged ($\hat{r} < 1.1$). Model selection (excluding 10-fold cross-validation) is summarized in Table 2, 3 and 4 for European anchovies, sardines and sprats respectively. Semi-variograms revealed that anchovy, juvenile sardine and sprat biomasses were spatially structured (Fig. 2, Supplementary material Appendix 1, Fig. A4–A8). In contrast, no spatial pattern was visible for adult sardines (Fig. 2, Supplementary material Appendix 1, Fig. A7). Models M_3 and M_4, with the Horseshoe prior for automatic variable selection, performed well both for the calibration and validation datasets (Fig. 3). The full model M_2 performed reasonably well with the calibration data (small G_M), but its predictive performance deteriorated with the validation data, especially when predicting outside the convex hull.

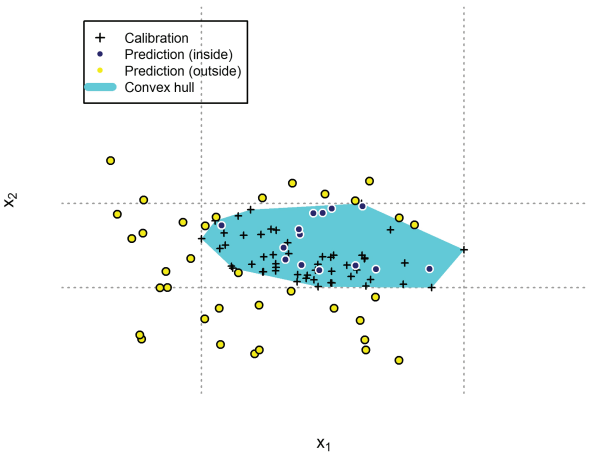

Figure 1. Interpolation versus extrapolation (from simulated data): the convex hull is the smallest convex polygon containing the calibration data (crosses) used for parameter estimation. Points symbolize a simulated prediction dataset: predictions falling inside the convex hull are interpolations (in blue); whereas predictions outside the convex hull are extrapolations (in yellow), that is predictions based on a combination of values not seen in the calibration data.

Predictive ability with the validation data

For validation data, 56 and 16 points fell inside and outside the convex hull respectively. For spatially structured species, spatial random effects unsurprisingly improved predictions. M_5 (without spatial effects) often performed worse with validation data, both outside and inside the hull. Exceptions were juvenile sardines (validation outside) and adult sardines (validation inside), probably due to sardines showing less spatial structuring than anchovies or sprats. M_2 was the worst model (Fig. 3). The simple M_1 was often best when predicting outside the convex hull (Fig. 3). Spatial models with a Horseshoe prior (M_3 and M_1) had a more stable rank across

Table 2. Model selection for predicting European anchovy biomass in the Gulf of Lions. G_M measures bias and P_M variance in predictions. Their sum is the posterior predictive loss D_M. The model with the smallest D_M has the best predictive ability (in bold).

Species	Age	Model	Data	Hull	G_M	P_M	D_M
Anchovy	**Juvenile**	**M_1**	**calibration**	**inside**	**58.1**	**16.5**	**74.6**
Anchovy	Juvenile	M_2	calibration	inside	69.0	12.4	81.4
Anchovy	Juvenile	M_3	calibration	inside	63.6	13.6	77.2
Anchovy	Juvenile	M_4	calibration	inside	63.9	13.6	77.5
Anchovy	Juvenile	M_5	calibration	inside	110.7	6.2	116.9
Anchovy	Juvenile	M_1	validation	inside	23.7	5.4	29.1
Anchovy	Juvenile	M_2	validation	inside	23.5	4.2	27.7
Anchovy	Juvenile	M_3	validation	inside	22.5	4.7	27.2
Anchovy	**Juvenile**	**M_4**	**validation**	**inside**	**22.2**	**4.7**	**26.9**
Anchovy	Juvenile	M_5	validation	inside	38.1	2.1	40.2
Anchovy	**Juvenile**	**M_1**	**validation**	**outside**	**4.4**	**1.4**	**5.7**
Anchovy	Juvenile	M_2	validation	outside	5.2	1.5	6.6
Anchovy	Juvenile	M_3	validation	outside	4.6	1.4	6.0
Anchovy	Juvenile	M_4	validation	outside	4.6	1.3	5.9
Anchovy	Juvenile	M_5	validation	outside	6.3	0.6	6.9
Anchovy	Adult	M_1	calibration	inside	62.2	24.7	86.9
Anchovy	Adult	M_2	calibration	inside	71.7	24.8	96.5
Anchovy	Adult	M_3	calibration	inside	63.4	24.4	87.8
Anchovy	**Adult**	**M_4**	**calibration**	**inside**	**62.2**	**23.6**	**85.8**
Anchovy	Adult	M_5	calibration	inside	160.6	11.0	171.7
Anchovy	Adult	M_1	validation	inside	19.9	7.7	27.5
Anchovy	Adult	M_2	validation	inside	19.7	8.5	28.2
Anchovy	**Adult**	**M_3**	**validation**	**inside**	**19.1**	**7.8**	**26.8**
Anchovy	**Adult**	**M_4**	**validation**	**inside**	**19.1**	**7.7**	**26.8**
Anchovy	Adult	M_5	validation	inside	37.8	3.5	41.3
Anchovy	**Adult**	**M_1**	**validation**	**outside**	**3.9**	**2.1**	**6.0**
Anchovy	Adult	M_2	validation	outside	7.7	3.3	11.0
Anchovy	Adult	M_3	validation	outside	4.5	2.4	6.9
Anchovy	Adult	M_4	validation	outside	5.0	2.2	7.2
Anchovy	Adult	M_5	validation	outside	11.1	0.9	12.0

Table 3. Model selection for predicting European sardine biomass in the Gulf of Lions. G_M measures bias and P_M variance in predictions. Their sum is the posterior predictive loss D_M. The model with the smallest D_M has the best predictive ability (in bold).

Species	Age	Model	Data	Hull	G_M	P_M	D_M
Sardine	Juvenile	M_1	calibration	inside	225.8	20.0	245.7
Sardine	**Juvenile**	$\boldsymbol{M_2}$	**calibration**	**inside**	**156.5**	**51.6**	**208.1**
Sardine	Juvenile	M_3	calibration	inside	164.1	47.3	211.4
Sardine	Juvenile	M_4	calibration	inside	164.3	47.9	212.3
Sardine	Juvenile	M_5	calibration	inside	252.7	16.9	269.6
Sardine	**Juvenile**	$\boldsymbol{M_1}$	**validation**	**inside**	**76.1**	**5.7**	**81.8**
Sardine	Juvenile	M_2	validation	inside	75.9	14.6	90.5
Sardine	Juvenile	M_3	validation	inside	72.0	10.5	82.5
Sardine	Juvenile	M_4	validation	inside	75.2	11.1	86.3
Sardine	Juvenile	M_5	validation	inside	88.1	5.5	93.5
Sardine	**Juvenile**	$\boldsymbol{M_1}$	**validation**	**outside**	**11.0**	**1.7**	**13.1**
Sardine	Juvenile	M_2	validation	outside	12.7	5.6	18.3
Sardine	Juvenile	M_3	validation	outside	11.8	3.6	15.4
Sardine	Juvenile	M_4	validation	outside	11.8	3.3	15.1
Sardine	Juvenile	M_5	validation	outside	12.7	1.3	13.9
Sardine	Adult	M_1	calibration	inside	35.7	27.4	63.1
Sardine	Adult	M_2	calibration	inside	33.0	28.6	61.6
Sardine	**Adult**	$\boldsymbol{M_3}$	**calibration**	**inside**	**27.5**	**27.7**	**55.2**
Sardine	Adult	M_4	calibration	inside	30.1	27.3	57.4
Sardine	Adult	M_5	calibration	inside	131.2	4.0	139.2
Sardine	Adult	M_1	validation	inside	48.0	6.3	54.3
Sardine	Adult	M_2	validation	inside	50.5	7.8	58.2
Sardine	Adult	M_3	validation	inside	47.9	5.8	53.7
Sardine	Adult	M_4	validation	inside	47.9	6.4	54.3
Sardine	**Adult**	$\boldsymbol{M_5}$	**validation**	**inside**	**42.0**	**1.1**	**43.1**
Sardine	**Adult**	$\boldsymbol{M_1}$	**validation**	**outside**	**2.5**	**1.8**	**4.3**
Sardine	Adult	M_2	validation	outside	3.7	2.9	6.6
Sardine	Adult	M_3	validation	outside	2.7	1.8	4.5
Sardine	Adult	M_4	validation	outside	2.6	1.8	4.4
Sardine	Adult	M_5	validation	outside	4.1	0.4	4.5

the different datasets. Models with a Horseshoe prior, spatial random effects and zero-inflation gave the best tradeoff between calibration and validation (Fig. 3).

Predictive ability with the prediction data

For adult anchovies, predictions from M_1, M_3 and M_4 were similar. In contrast, M_2 made unrealistic predictions close to the Camargue (estuary of the Rhône river, Supplementary material Appendix 1, Fig. A5). M_5 suffered from the same defect, though the Horseshoe prior attenuated the extreme predictions by an order of magnitude. For adult sardines, predictions from M_1, M_3 and M_4 were similar with large values predicted close to the shelf edge. M_2 made again unrealistic predictions close to the Camargue (Supplementary material Appendix 1, Fig. A7). Predictions from M_5 were on a very different scale. All predictions from M_3 and M_4 had narrower 95% credibility interval width than M_2 or M_5 (not shown), but were comparable to those obtained with M_1. M_2 always predicted a few implausibly extreme values (Supplementary material Appendix 1). For prediction, 101 and 441 data points were respectively inside and outside the convex hull. Models with covariates quickly extrapolated as the number of inputs increased (Fig. 4). For any two inputs, the average amount of interpolation was 90%; however for 5 inputs it was down to $<50\%$.

Table 4. Model selection for predicting European sprat biomass in the Gulf of Lions. G_M measures bias and P_M variance in predictions. Their sum is the posterior predictive loss D_M. The model with the smallest D_M has the best predictive ability.

Species	Age	Model	Data	Hull	G_M	P_M	D_M
Sprat	Juvenile	M_1	calibration	inside	57.2	39.9	97.1
Sprat	Juvenile	M_2	calibration	inside	53.8	42.3	96.1
Sprat	**Juvenile**	$\boldsymbol{M_3}$	**calibration**	**inside**	**54.6**	**39.5**	**94.0**
Sprat	Juvenile	M_4	calibration	inside	56.5	39.6	96.1
Sprat	Juvenile	M_5	calibration	inside	223.9	15.9	239.8
Sprat	**Juvenile**	$\boldsymbol{M_1}$	**validation**	**inside**	**35.8**	**12.1**	**47.8**
Sprat	Juvenile	M_2	validation	inside	36.6	13.3	49.9
Sprat	Juvenile	M_3	validation	inside	36.2	11.8	48.0
Sprat	Juvenile	M_4	validation	inside	36.2	12.1	48.3
Sprat	Juvenile	M_5	validation	inside	68.7	5.3	74.0
Sprat	**Juvenile**	$\boldsymbol{M_1}$	**validation**	**outside**	**19.7**	**3.4**	**23.2**
Sprat	Juvenile	M_2	validation	outside	20.2	5.3	25.5
Sprat	Juvenile	M_3	validation	outside	20.7	3.7	24.4
Sprat	Juvenile	M_4	validation	outside	21.3	3.8	25.0
Sprat	Juvenile	M_5	validation	outside	38.4	1.4	39.8

Estimation of regression goefficients

Regression coefficients differed between species and stages. Important predictors were *GCHLA* for adult anchovies; *SDSST*, *GCHLA* and *SED* for juvenile anchovies; *CHLA*, *GCHLA* and *SED* for juvenile sardine; and *GBathy* and *SDSST* for juvenile sprat (Fig. 5). Figure 5 compares the estimated regression coefficients between M_2 and $M_{3–5}$ (Supplementary material Appendix 1, Fig. A9–A10). Regression coefficients for adult sardines were very noisy (M_2), or heavily shrunk ($M_{3–5}$). Shrunk regression coefficients consistently enabled more accurate predictions (Fig. 3). Specifically, the Horseshoe prior shrank regression coefficients β_k by an average factor of 1.8, 5.8, 3.2, 11.1 and 11.1 for juvenile anchovies, adult anchovies, juvenile sardines, adult sardines and juvenile sprats respectively. The bad predictive performance of M_2 was thus partly linked to overestimated effect size of coefficients. Taking into account spatial random effects was also important and affected estimated regression coefficients, even in models with Horseshoe priors.

Discussion

We investigated shrinkage regression for habitat modelling of small pelagic fish in the western Mediterranean Sea. Accounting for spatial structure resulted in better predictions. Record et al. (2013) and Swanson et al. (2013) argued of the importance of spatial random effects: their inclusion in habitat models leads to conservative predictions. A spatial model will predict a similar pattern to that seen in the calibration data, unless the simultaneous inclusion of inputs really improves predictions. M_3 and M_4, with both a spatial structure and a Horseshoe prior, were competitive with M_1. They were usually better when predicting inside the convex hull, but M_1 was often just as good, or better, for predictions outside the convex hull (Fig. 3). In other words, a simple explicitly spatial model with no environmental covariates often generated accurate predictions (Bahn and McGill 2007). Reasons why a simple kriging model performed better may include collinearity issues with the predictors

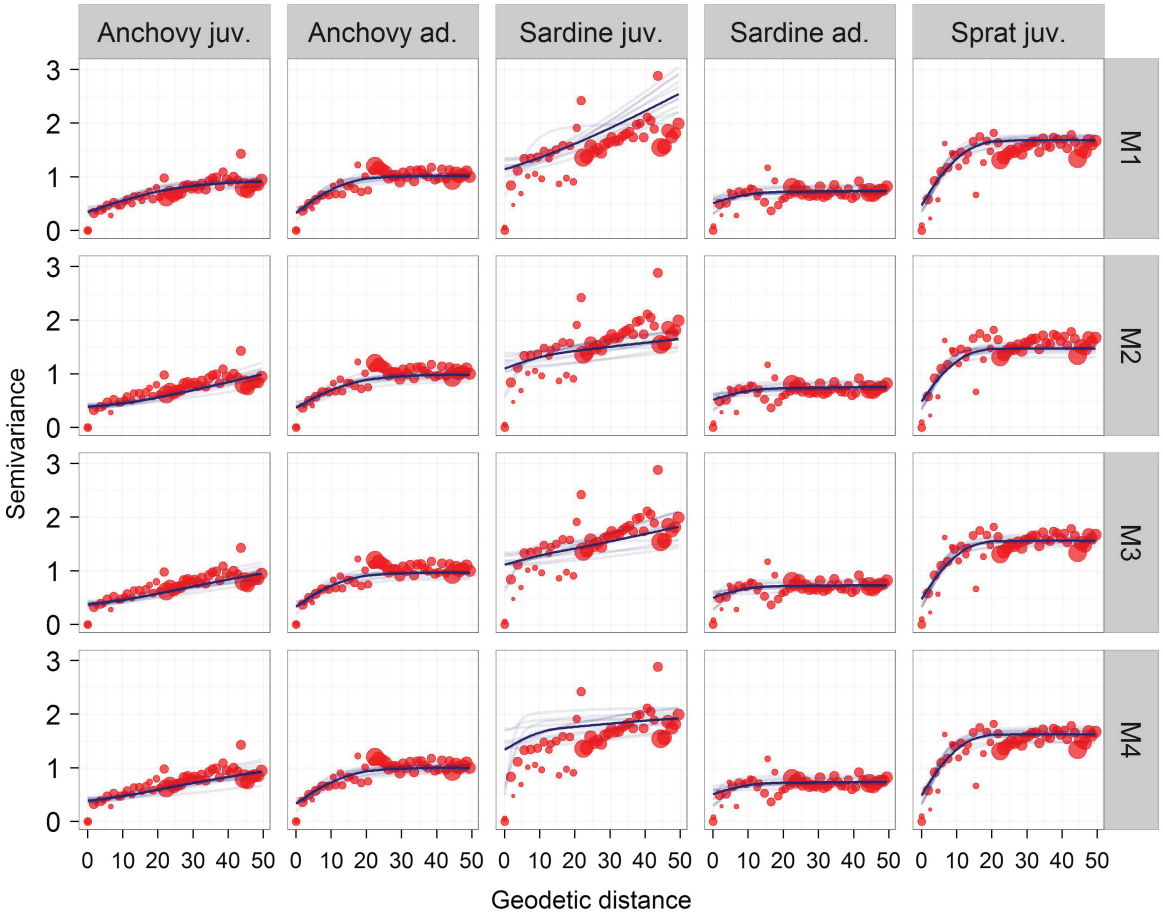

Figure 2. Estimated semi-variance curves are represented by transparent solid blue line for each cross-validation. The dark blue line is a loess curve. Red dots are observed semi-variances, with size proportional to the number of pairs of points. Rows correspond to the different models, and columns to species-stage combinations.

(Dormann et al. 2007, 2013) and overfitting, that is capitalizing on noise in the calibration data that did not reccur in the validation data (Babyak 2004, Merow et al. 2014).

Interpolation, extrapolation and overfitting

Predictions were made on the same geographical region, using covariates that were from the same source as those used in the calibration data, and restricting the range of an individual covariate to be on average $>$ 90% within the range of the calibration data (Fig. 4). Despite this favorable setting, M_2 performance was disappointing (Fig. 3) and particularly poor with validation data. This may be 1) due to overestimation (type-M errors, Gelman and Tuerlinckx 2000) of regression coefficients (Fig. 5); or 2) because the scope for overfitting became larger with more inputs as predictions became extrapolations rather than interpolations (Fig. 4). Albeit all the values of inputs occurred in the calibration data, entirely novel combinations appeared in both the validation and prediction datasets (Mesgaran et al. 2014). This problem was most acute with M_2, which clearly overfitted and predicted unrealistically large biomasses (Supplementary material Appendix 1, Fig. A4–A8). M_2 was a full model without variable selection: its predictions were both unrealistically large and imprecise.

Performing model averaging following all-subsets model fitting (Johnson and Hoeting 2011) or ensemble forecasts (Araújo and New 2007, Buckland et al. 2014) would alleviate overfitting and address type M-errors. However, these procedures are costly computation-wise. Using the Horseshoe prior as a regularization device for variable selection resulted in accurate predictions in one run. Shrunk parameters estimates were also stable when compared by 10-fold cross-validation (Supplementary material Appendix 1, Fig. A10). In a frequentist framework, Reineking and Schröder (2005) conducted a simulation study and concluded of the superior predictive ability of shrinkage approaches. For Bayesian linear models, Carvalho et al. (2010) established that posterior means from a Horseshoe prior were similar to model-averaged estimates. Our results showed that the Horseshoe prior gave accurate and precise spatial predictions (Table 2–4), while avoiding overfitting and costly model selection procedure.

Although shrinkage regression may appear overly technical to ecologists (Dahlgren 2010), we think it is worth the trouble beyond pragmatic variable selection. For example, M_2 predicted extremely large biomass of anchovies, adult sardines but very few juvenile sardines and sprats off the Rhône Estuary (Supplementary material Appendix 1, Fig. A4–A8). If we were to trust these predictions, the Rhône

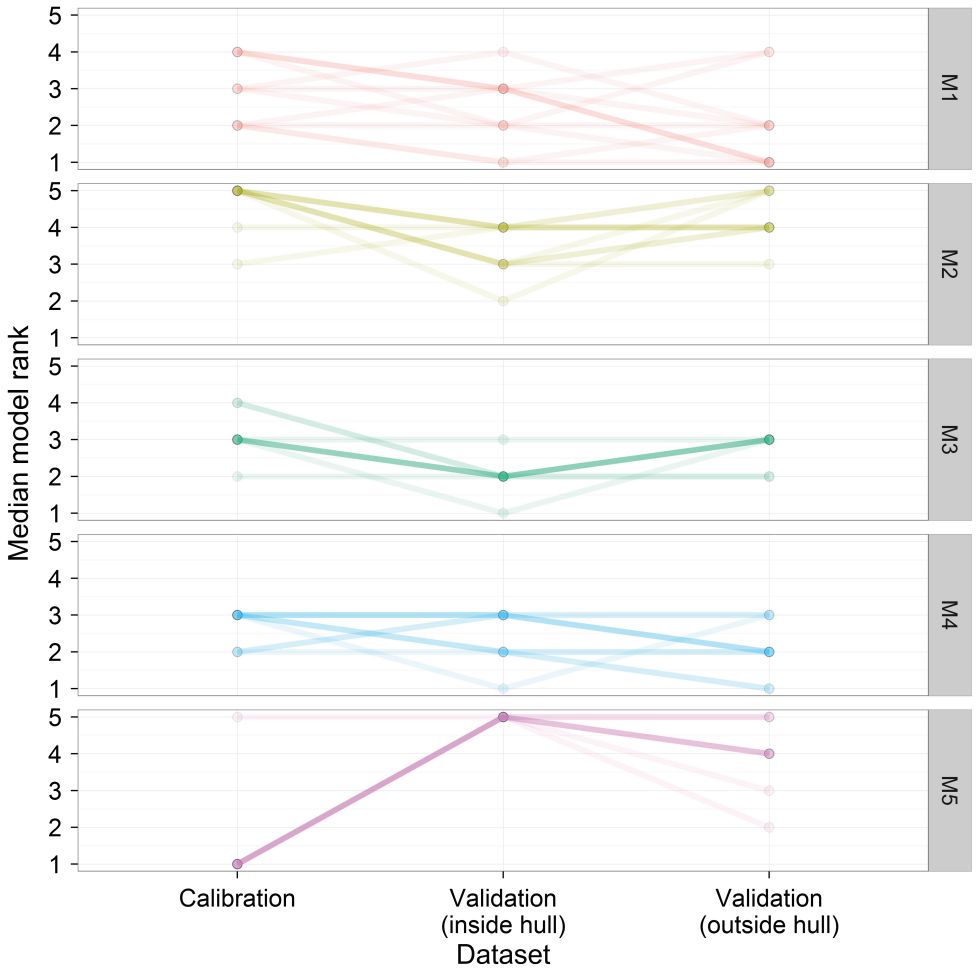

Figure 3. Left: median rank calculated over all species-stage combinations for each model with respect to the smallest posterior smallest posterior predictive loss D_M. The best model across all species-age combinations is ranked first (smallest D_M). Model rank changed across the different datasets: the simplest model was worst for the calibration data, but its predictive ability improved when interpolating or extrapolating. Models with stable rank across the different task (calibration, interpolation or extrapolation) were M_3 and M_4. Each line corresponds to a (cross-)validation dataset. Rows correspond to the different models.

estuary would be a very profitable fishing area for anchovies. These were but extrapolations (Fig. 4) which either a simple kriging model or spatial models with shrinkage greatly attenuated. This attenuation was most dramatic for adult anchovies and adult sardines whose predicted biomasses differed by more than a full order of magnitude between M_2 and shrinkage models. Shrinkage regression appears valuable in fishery or conservation studies to prevent inaccurate predictions that could lead to unfortunate management decisions (overfishing) or resources attribution (research survey in an unsampled areas predicted to concentrate high abundance or biomass).

Prediction or explanation

Predictive modelling is the process of applying a statistical model to data for the purpose of predicting new observations. It is a pragmatic approach that puts a premium on empirical adequacy regardless of theoretical considerations. Explanatory modelling is more concerned with adequacy with a theory about how the data arise. The two goals of prediction and explaination can then lead to different model choices (Betts et al. 2009, Shmuéli 2010).

The tension between explanation and prediction mirrors that between realism and instrumentalism (Sober 1999, Stanford 2006). Realism is the philosophical idea that theories provide a true description of the world (Chang 2012, p. 222). Scientific theories can use unobservable quantities only if these literally exist in nature (Stanford 2006, Chang 2012). For example, spatial random effects are often interpreted realistically as capturing the effect of missing important predictors (Dormann 2007, Dormann et al. 2007), or as accounting for ecological processes, e.g. dispersal or aggregative behaviour (Merow et al. 2014).

'[I]nstrumentalism is the idea that theories are [mere] instruments for making predictions' (Sober 1999). Including predictors that may have no causal link with the response variable, or to use a data dimension reduction techniques (such as principal component analysis), even if the results have no clear interpretation, is acceptable for an instrumentalist if better predictions are obtained. Instrumentalism does not investigate why a pattern arises, but only seeks to predict it most accurately: its aim is predictive success, not truth

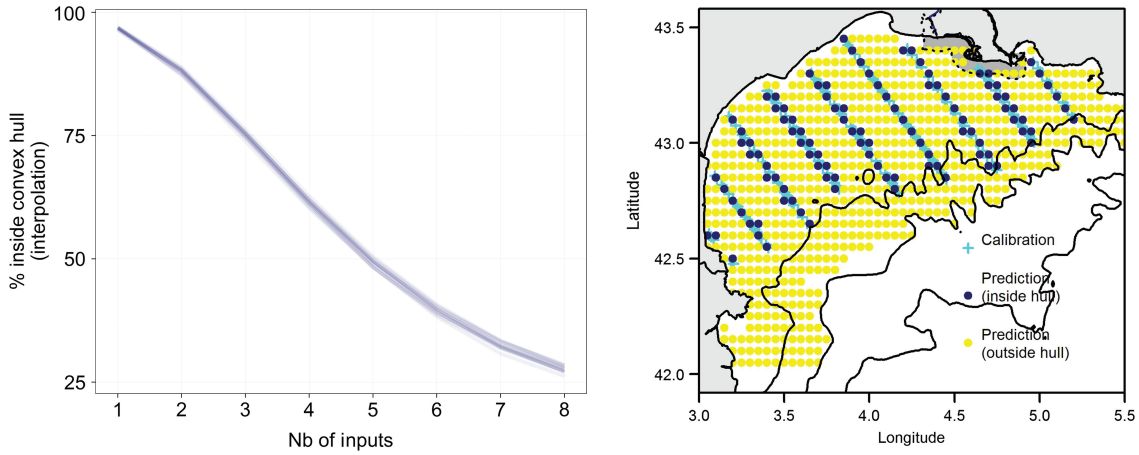

Figure 4. Left panel: interpolation to extrapolation ratio in the prediction dataset. The y-axis shows the % of predictions inside the calibration convex hull averaged over all combinations of k inputs (x-axis). Each line corresponds to a cross-validation. Right panel: predicting pelagic fish biomass in the Gulf of Lions with 9 different inputs. Calibration and prediction data are depicted in the study area. Predictions falling inside the convex hull are interpolations (in blue), whereas predictions outside the convex hull are extrapolations (in yellow). The black dotted line materializes the Camargue Natura 2000 protected area.

(Chang 2012, chapter 4). From a realist perspective, our results showed that having too many inputs is detrimental: extrapolation quickly became pervasive with just 5 different inputs. Habitat models routinely include more than 5 different inputs (Warton et al. 2015). Our ability to reliably test theoretical predictions may degrade as we add more inputs.

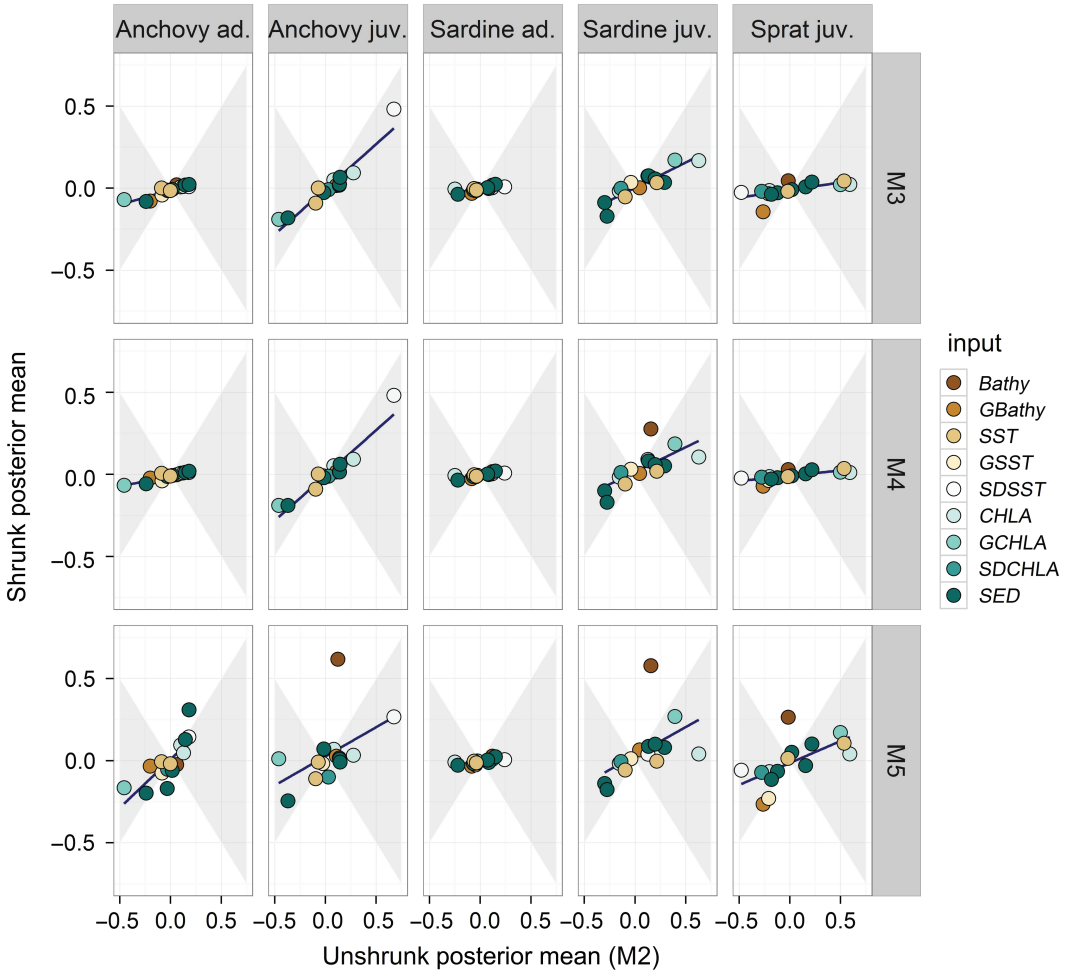

Figure 5. Comparison of estimated regression coefficients β_k for M_2 and other models. The blue line is the regression slope (average shrinkage factor) between two sets of coefficients. Points falling in the grey area illustrate shrinkage compared to estimates from a model with independent priors.

Successful studies referring to niche theory usually include a couple of inputs (Fort et al. 2012) or restrict a priori their maximum number (Mannocci et al. 2014). For explanatory purposes, keeping a small number of inputs improves both model interpretation and understanding.

The philosophy behind habitat modelling oscillates between the end points of realism and instrumentalism: predictions and the production of useful maps over large unsampled areas are pragmatic goals. Yet explanation often remains desirable. When more emphasis is put on realism, concerns about additivity and linear assumptions become important (Austin 2007). When the goal is pure prediction, these concerns are less pressing than accuracy and precision. For nowcasting, our results showed that a simple kriging model is enough (on small pelagic fish, Saraux et al. 2014; on the value of simplicity, Hogarth 2012, Ward et al. 2014). However, including predictors in conjunction with the use of shrinkage priors such as the Horseshoe can give equivalent, or better, predictions than a simple kriging model.

Conclusion

Habitat modelling is a very active research program: its goals are manifold and, sometimes, antagonistic (Merow et al. 2014). Researchers often want to include an ever expanding list of environmental inputs to improve predictions for conservation purposes. A large number of predictors, in conjunction with model complexity, raised additional difficulties (Merow et al. 2014, but see Warton et al. 2015). Accounting for model uncertainty is important to avoid overconfident inferences (Buckland et al. 1997), and to draw accurate ignorance maps (Rocchini et al. 2011). When proper model selection (sensu Burnham and Anderson 2002) becomes impractical, relying on rules such as confidence or credibility intervals (CI) excluding 0 may encourage overfitting: large estimates are away from zero, but their CI may exclude zero simply because these estimates are imprecisely estimated (Link and Sauer 1996). Their large standard errors may actually indicate that the bias-variance tradeoff is too much in favour of reducing bias in the calibration data, at the expense of a large prediction variance: predictions are unreliable extrapolations.

The usual caveat is not helpful: extrapolation to unsampled areas is often of primary interest. Methods that predict well outside the comfort zone of models are needed for conservation purposes (such as fishery time closures or regulation in areas of high juvenile recruitment). Shrinkage regression enables models to predict accurately beyond the calibration data, and could be useful for hind- and fore-casting (Record et al. 2013, Swanson et al. 2013; but see Davis et al. 2014). Models M_3 or M_4 could be used for predicting habitats under climate change for example, which is not possible with a simple kriging model. The use of shrinkage priors is an asset in this case because they 'stabilize estimates and predictions by making fitted models less sensitive to certain details of the data' (Gelman and Shalizi 2013).

However, shrinkage is not a panacea: we are not advocating to include all available inputs and let a shrinkage prior decides of their importance. Careful thinking about processes and model structure remains paramount (Leamer 2012, Nichols et al. 2012). The pragmatic appeal of shrinkage is how it deals with collinearity and overfitting even with many inputs, a common feature of current habitat models. It also provides accurate predictions and tighter CI around predictions (not shown) with only one model fit. We are next planning to perform a multi-year analysis of the distribution of small pelagic fish biomass in the Gulf of Lions (PELMED survey data) using model M_3 as an instrument for estimating regression coefficients.

Acknowledgements – We are grateful to David Grémillet for financial support and scientific advices. We are indebted to the crew of RV L'Europe and everyone who participated in the 2011 PELMED survey, in particular Jean-Louis Bigot chief scientist onboard. PELMED surveys are cofunded by Ifremer and the European Union through the Data Collection Framework. We thank Carsten Dormann, the associate editor, for his patient and constructive comments.

References

Araújo, M. and New, M. 2007. Ensemble forecasting of species distributions. – Trends Ecol. Evol. 22: 42–47.

Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. – Ecol. Model. 200: 1–19.

Babyak, M. 2004. What you see may not be what you get: a brief, non-technical introduction to overfitting in regression-type models. – Psychosom. Med. 66: 411–421.

Bahn, V. and McGill, B. 2007. Can niche-based distribution models outperform spatial interpolation? – Global Ecol. Biogeogr. 16: 733–742.

Bahn, V. and McGill, B. 2013. Testing the predictive performance of distribution models. – Oikos 122: 321–331.

Banerjee, S. 2005. On geodetic distance computations in spatial modelling. – Biometrics 61: 617–625.

Betts, M. et al. 2009. Comment on 'Methods to account for spatial autocorrelation in the analysis of species distributional data: a review'. – Ecography 32: 374–378.

Buckland, S. et al. 1997. Model-selection: an integral part of inference. – Biometrics 53: 603–618.

Buckland, S. et al. 2014. Ecological effects of the invasive giant Madagascar day gecko on endemic Mauritian geckos: applications of binomial-mixture and species distribution models. – PLoS One 9: e88798.

Burnham, K. and Anderson, D. 2002. Model selection and multimodel inference. A practical information-theoretic approach, 2nd ed. – Springer.

Carvalho, C. et al. 2010. The horseshoe estimator for sparse signals. – Biometrika 97: 465–480.

Chang, H. 2012. Is water H2O? 1st ed. – Springer.

Conn, P. et al. 2015. On extrapolating past the range of observed data when making statistical predictions in ecology. – PLoS One 10: e0141416.

Cowles, M. and Carlin, B. 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. – J. Am. Stat. Assoc. 91: 883–904.

Dahlgren, J. 2010. Alternative regression methods are not considered in Murtaugh (2009) or by ecologists in general. – Ecol. Lett. 13: E7–E9.

Davis, E. et al. 2014. Ecological niche models of mammal glacial refugia show consistent bias. – Ecography 37: 1133–1138.

Doray, M. et al. 2010. Pelagic fish stock assessment by acoustic methods at Ifremer. – Tech. Rep. R.INT. DOP/DCN/EMH 10-02, Ifremer.

Dormann, C. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. – Global Ecol. Biogeogr. 16: 129–138.
Dormann, C. et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional aata: a review. – Ecography 30: 609–628.
Dormann, C. et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. – Ecography 36: 27–46.
Elith, J. and Leathwick, J. 2009. Species distribution models: ecological explanation and prediction across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.
Fort, J. et al. 2012. Biologging, remotely-sensed oceanography and the continuous plankton recorder reveal the environmental determinants of a seabird wintering hotspot. – PLoS One 7: e41194.
Gelfand, A. and Ghosh, S. 1998. Model choice: a minimum posterior predictive loss approach. – Biometrika 85: 1–11.
Gelman, A. and Tuerlinckx, F. 2000. Type-S error rates for classical and Bayesian single and multiple comparison procedures. – Comput. Stat. 15: 373–390.
Gelman, A. and Shalizi, C. 2013. Philosophy and the practice of Bayesian statistics. – Br. J. Math. Stat. Psychol. 66: 8–38.
Gelman, A. et al. 2008. A weakly informative default prior distribution for logistic and other regression models. – Ann. Appl. Stat. 2: 1360–1383.
Giannoulaki, M. et al. 2006. The effect of coastal topography on the spatial structure of anchovy and sardine. – ICES J. Mar. Sci. 63: 650–662.
Giannoulaki, M. et al. 2011. Habitat suitability modelling for sardine juveniles (*Sardina pilchardus*) in the Mediterranean sea. – Fish. Oceanogr. 20: 367–382.
Giannoulaki, M. et al. 2013. Characterizing the potential habitat of European anchovy *Engraulis encrasicolus* in the Mediterranean sea, at different life stages. – Fish. Oceanogr. 22: 69–89.
Griffin, J. and Brown, P. 2013. Some priors for sparse regression modelling. – Bayesian Analysis 8: 691–702.
Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.
Hogarth, R. 2012. When simple is hard to accept. – In: Todd, P. and Gigerenzer, G. (eds), Ecological rationality: intelligence in the world. Oxford Univ. Press, pp. 61–79.
Johnson, D. and Hoeting, J. 2011. Bayesian multimodel inference for geostatistical regression models. – PLoS One 6: e35677.
Juntunen, T. et al. 2012. Bayesian spatial multispecies modelling to assess pelagic fish stocks from acoustic- and trawl-survey data. – ICES J. Mar. Sci. 69: 95–104.
King, G. and Zeng, L. 2007. When can history be our guide? The pitfalls of counterfactual inference. – Int. Studies Quat. 51: 183–210.
Leamer, E. 2012. The context matters: comment on Jerome H. Friedman, 'Fast sparse regression and classification'. – Int. J. Forecasting 28: 741–748.
Link, W. and Sauer, J. 1996. Extremes in ecology: avoiding the misleading effects of sampling variation in summary analyses. – Ecology 77: 1633–1640.
Mac Nally, R. 2000. Regression and model building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – 'predictive' and 'explanatory' models. – Biodivers. Conserv. 9: 655–671.
Mannocci, L. et al. 2014. Predicting top predator habitat in the South Indian ocean. – Ecography 37: 261–278.
Martin, T. et al. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. – Ecol. Lett. 8: 1235–1246.
Merow, C. et al. 2014. What do we gain from simplicity versus complexity in species distribution models? – Ecography 37: 1267–1281.
Mesgaran, M. et al. 2014. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. – Divers. Distrib. 20: 1147–1159.
Nichols, J. et al. 2012. Studying biodiversity: is a new paradigm really needed. – BioScience 62: 497–502.
Palomera, I. et al. 2007. Small pelagic fish in the north western Mediterranean sea: an ecological review. – Prog. Oceanogr. 74: 377–396.
Park, T. and Casella, G. 2008. The Bayesian lasso. – J. Am. Stat. Assoc. 103: 681–686.
Péron, C. et al. 2012. Projected poleward shift of king penguins' (*Aptenodyptes patagonicus*) foraging range at the Crozet islands, southern Indian Ocean. – Proc. R. Soc. B 279: 2515–2523.
Polson, N. and Scott, J. 2010. Shrink globally, act locally: sparse Bayesian regularization and prediction. – In: Bernardo, J. M. et al. (eds) Bayesian statistics. Oxford Univ. Press, pp. 501–525.
Record, S. et al. 2013. Should species distribution models account for spatial autocorrelation? A test of model projections across eight millenia of climate change. – Global Ecol. Biogeogr. 22: 760–771.
Reineking, B. and Schröder, B. 2005. Constrain to perform: regularization of habitat models. – Ecol. Model. 193: 675–690.
Rocchini, D. et al. 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. – Prog. Phys. Geogr. 35: 211–226.
Saraux, C. et al. 2014. Spatial structure and distribution of small pelagic fish in the north western Mediterranean Sea. – PLoS One 9: e111211.
Schmidt, A. et al. 2008. Investigating the sensitivity of Gaussian processes to the choice of their correlation function and prior specification. – J. Stat. Comput. Simul. 78: 681–699.
Shmuéli, G. 2010. To explain or to predict. – Stat. Sci. 25: 289–310.
Simmonds, E. and MacLennan, D. 2005. Fisheries acoustics – theory and practice. – Blackwell.
Sober, E. 1999. Instrumentalism revisited. – CRÍTICA, Revista Hispanoamericana de Filosofia 31: 3–39.
Stan Development Team 2015. Stan: a c++ library for probability and sampling. – Version 2.2.0.
Stanford, P. 2006. Exceeding our grasp: science, history and the problem of unconceived alternatives, 1st ed. – Oxford Univ. Press.
Stoll, H. et al. 2009. WhatIf: software for evaluating counterfactuals. – J. Stat. Softw. 15: 1–18.
Swanson, A. et al. 2013. Spatial regression methods capture prediction uncertainty in species distribution model projections through time. – Global Ecol. Biogeogr. 22: 242–251.
Tugores, M. et al. 2011. Habitat suitability modelling for sardine *Sardina pilchardus* in a highly diverse ecosystem: the Mediterranean Sea. – Mar. Ecol. Prog. Ser. 443: 181–205.
Van Beveren, E. et al. 2014. Rapid changes in growth, condition, size and age of small pelagic fish in the Mediterranean. – Mar. Biol. 161: 1809–1822.
Vaughan, I. and Ormerod, S. 2005. The continuing challenges of testing species distribution models. – J. Appl. Ecol. 42: 720–730.
Wang, W. and Gelman, A. 2014. Difficulty of selecting among multilevel models using predictive accuracy. – Stat. Interface 7: 1–8.
Ward, E. et al. 2014. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. – Oikos 123: 652–661.

Warton, D. et al. 2015. So many variables: joint modeling in community ecology. – Trends Ecol. Evol. 30: 766–779.
Weill, A. et al. 1993. MOVIES-B: an acoustic detection description software. Application to shoal species' classification. – Aquat. Living Resour. 6: 255–267.
Wikle, C. 2003. Hierarchical models in environmental sciences. – Int. Stat. Rev. 71: 181–199.
Wilson, T. et al. 2010. Hierarchical spatial models for predicting pygmy rabbit distribution and relative abundance. – J. Appl. Ecol. 47: 401–409.

Supplementary material (Appendix ECOG-01633 at <www.ecography.org/appendix/ecog-01633>). Appendix 1.