Paige Copenhaver-Parry
May 26, 2017

Seedling Metamodel Description

Naïve Metamodel

The naïve metamodel (θ_m) is a logistic regression SDM parameterized on presence/absence data (X_m) for each of 6 tree species (**Table 1**) across the U.S. states of Montana, Idaho, Wyoming, Utah, Colorado, New Mexico and Arizona. All data are from the FIA database. Covariates (D_m) are climate data from the Moscow Forest Sciences Laboratory data set downscaled according to the ANUSPLIN thin-plate spline model. The best-fitting covariates for each species were selected from all possible combinations of non-collinear covariates, their interactions, and quadratic terms based on DIC and BIC. The covariates to be used for each species are specified in a separate file (**Table 2**). The naïve metamodel predicts the probability of presence from presence/absence data using a logit link function.

$$p(\theta_m|X_m, D_m) \propto p(X_m|\theta_m, D_m)p(\theta_m)$$

Likelihood:

$$P(X_m = 1) = \frac{\exp{(\alpha + \beta_m D_m)}}{1 + \exp{(\alpha + \beta_m D_m)}}$$

The model is specified in the "naiveSDM_stan.stan" file and executed in the "naiveSDM_r.r" file. This R script is written to iterate through all 6 species and to save posterior output and workspaces for each model. The script also expands the covariate matrix to include interaction and quadratic terms and scales all covariates.

Table 1. Species included in all data sets (FIA pres/abs, FIA seedling population growth, seedling growth experiment)

Name	Acronym	FIA code
Abies lasiocarpa	ABLA	19
Picea engelmannii	PIEN	93
Pinus contorta	PICO	108
Pinus ponderosa	PIPO	122
Pseudotsuga menziesii	PSME	202
Populus tremuloides	POTR	746

Table 2. Files needed to execute the naïve metamodel

File Name	Description
naiveSDM_stan.stan	model specification
naiveSDM_r.r	data prep and model execution
mods_to_test.csv	covariate specification
SDM_mod_dat_adult_fullcovs.csv	data (pres/abs and covariates)

Seedling Population Sub-model

The seedling population model (θ_s) estimates the log of the proportional change in seedling counts at an FIA seedling plot location between consecutive census. In the process, the log of the intrinsic population growth rate (r_s) is estimated for each set of climate covariates (D_s; uses the same covariates as the naïve SDM), which is then transformed into an estimation of presence/absence (X_s) by assuming that an estimated $\log(r)>0$ represents species presence, while an estimated $\log(r)\leq0$ represents an absence. Due to the high number of zeros in the seedling count data, a zero-inflated model approach is utilized, which assumes that absences may arise either because the species is absent from the sample or because the seedling count did not change between censuses.

$$p(\theta_s|X_s, D_s) \propto p(X_s|\theta_s, D_s)p(\theta_s)$$

$$y_s = \begin{cases} 0 & with\ probability\ \rho \\ \max(0, \omega_s) & 1 - \rho \times P(\omega_s) \end{cases}$$

$$\omega_s \sim N(a_s, \sigma_s)$$

$$a_s = \alpha + \beta D_s$$

$$\log(r_s) = \frac{a_s}{\Delta t}$$

$$X_s = \begin{cases} 0 & \log(r_s) \leq 0 \\ 1 & \log(r_s) > 0 \end{cases}$$

The fitted seedling population model can then be used to simulate a new presence/absence data set (Ψ_s) across the full range of environmental conditions represented by D_m by generating predictions from θ_s. The probability of this prediction can then be integrated into the metamodel by evaluating the probability of the data according to the metamodel:

$$p(\Psi_s|\theta_m)$$

The model is specified in the "naivepop_ZIM4_stan.stan" file and executed in the "naivepop_r.R" file (**Table 3**). Similarly to the naïve SDM, this R script is written to iterate through models through all 6 species and to save posterior output and workspaces for each run. The script also expands the covariate matrix to include interaction and quadratic terms and scales all covariates.

Table 3. Files needed to execute the seedling population sub-model

File Name	Description
naivepop_ZIM4_stan.stan	model specification
naivepop_r.R	data prep and model execution
mods_to_test.csv	covariate specification
FIA_mod_dat_fullcovs.csv	data (seedling counts and covariates)

Seedling Survival Sub-model

A second sub-model (θ_i) is used to estimate probability of survival ($p(X_i=1)$) of seedlings grown in experimental plots across a temperature gradient using a logistic regression. Covariates (D_i) include the temperature variables included in the naïve metamodel and the seedling population sub-model as well as seedling size.

$$p(\theta_s | X_s, D_s)$$

Likelihood:

$$p(X_i = 1) = \frac{\exp(\alpha_i + \beta_i D_i)}{1 + \exp(\alpha_i + \beta_i D_i)}$$

The fitted survival can then be used generated a predicted presence/absence data set, Ψ_i, by assuming that survival probability exceeding a particular threshold represents seedling occurrence under a given set of conditions defined by the covariates. In this case, the probability threshold (TROC) is set to maximize sensitivity and specificity (determined from receiver-operating curve). These binary predictions can be evaluated according to the metamodel for integration with the seedling population model and the naïve SDM:

$$\hat{p}_i = f(\theta_i, D_i)$$

$$\varphi_i = \begin{cases} 1 & \hat{p}_i > T_{ROC} \\ 0 & otherwise \end{cases}$$

$$p(\Psi_i | \theta_m)$$

The model is specified in a stan file and executed in an R script that iterates through all 6 speices and saves posterior output and workspaces. (**Table 4**).

Table 4. Files needed to execute the seedling IPM sub-model

File Name	Description
survival_r.R	model execution
survival_stan.stan	model specification
IPM_seedling_data.csv	data (seedling size and covariates)