Modeling and mapping species distributions

Richard Chandler

March 29, 2012

Abstract

A species' distribution can be characterized by the probability that the species occurs at some location in space. Alternatively, distribution could be described using a spatially-explicit density map. These definitions of species distribution avoid the ambiguity surrounding the indices of occurrence or abundance produced by many presence-only algorithms. The **unmarked** package contains methods of fitting occurrence and abundance models, and can be used to produce distribution maps with the help of the **raster** package (van Etten, 2012) as is demonstrated in this vignette. Unlike many other tools for modeling species distributions, the models in **unmarked** account for bias due to spatial or temporal heterogeneity in detection probability. Furthermore, **unmarked** includes models of population dynamics, allowing one to map quantities such as local colonization or extinction probability.

Mapping Occurrence Probability

In this example, we fit the dynamic occupancy model of (MacKenzie et al., 2003) to data on the European crossbill (*Loxia curvirostra*) collected in 267 1-km^2 sample quadrats in Switzerland, 1999-2007 (Schmid et al., 2004). We then use the model to compute the expected probability of occurrence at each pixel in a raster defining the Swiss landscape.

First we load the crossbill data[1], which is a list with two components. The first component, `crossbill`, is a data.frame containing the detection/non-detection data and some covariates such as the percent cover of forest at each survey location. The following commands format the data and fit the model using the `colext` function. For more information about this model, see the "colext" vignette that comes with **unmarked**.

```
> data(crossbill)
> crossbill <- crossbill.data$crossbill
> years <- as.character(1999:2007)
> years <- matrix(years, nrow(crossbill), 9, byrow=TRUE)
> umf <- unmarkedMultFrame(y=as.matrix(crossbill[,5:31]),
    siteCovs=crossbill[,2:3], yearlySiteCovs=list(year=years),
    numPrimary=9)
> (fm <- colext(~ele + forest, ~ele + forest, ~1, ~1, umf))

colext(psiformula = ~ele + forest, gammaformula = ~ele + forest,
    epsilonformula = ~1, pformula = ~1, data = umf)

Initial:
            Estimate       SE     z  P(>|z|)
(Intercept) -3.28709 0.519357 -6.33 2.47e-10
ele          0.00107 0.000359  2.97 2.93e-03
forest       0.02944 0.006087  4.84 1.32e-06


Colonization:
            Estimate       SE     z  P(>|z|)
(Intercept) -2.378039 0.411148 -5.78 7.30e-09
ele         -0.000323 0.000287 -1.12 2.61e-01
forest       0.023013 0.004290  5.36 8.14e-08


Extinction:
 Estimate    SE     z  P(>|z|)
    -1.46 0.154 -9.45 3.29e-21


Detection:
 Estimate     SE    z P(>|z|)
   0.0955 0.0579 1.65  0.0992


AIC: 5123.448
```

---

[1] This dataset has been temporarily removed from **unmarked** until full permission can be secured.

Now that we have our fitted model, we can use the estimates to compute the expected probability of occurrence at each pixel in the landscape. The `raster` package makes this easy. Here are the commands to create raster objects from the covariate matrices that come with the crossbill data.

```
> library(raster)
> elevation <- raster(crossbill.data$switzerland[[1]])
> layerNames(elevation) <- "ele"
> forest <- raster(crossbill.data$switzerland[[2]])
> layerNames(forest) <- "forest"
```

The landscape extent and the projection were not specified simply because they are not relevant to our purposes. The reason for assigning the layerNames is that the `predict` function, which will be used shortly, requires that the rasters have names identical to the names of the variables used in the model fitting process.

The `predict` function is useful for computing spatially-referenced model predictions, standard errors, and confidence intervals, but it is computationally demanding when there are many pixels in the raster. Thus, if measures of uncertainty are not required, the following code can be used to quickly produce the species distribution map shown in Fig.1.

```
> (beta <- coef(fm, type="psi"))
     psi(Int)      psi(ele)   psi(forest)
-3.287090570  0.001067963  0.029443422
> logit.psi <- beta[1] + beta[2]*elevation + beta[3]*forest
> psi <- exp(logit.psi) / (1 + exp(logit.psi))
> plot(psi, axes=FALSE)
```

The same can be done for any other parameter. For example, if we modeled density, perhaps using the `distsamp` function, we could compute the expected number of individuals in each pixel. Another option with the crossbill data is to map expected colonization probabilities, which can be accomplished using the following code.
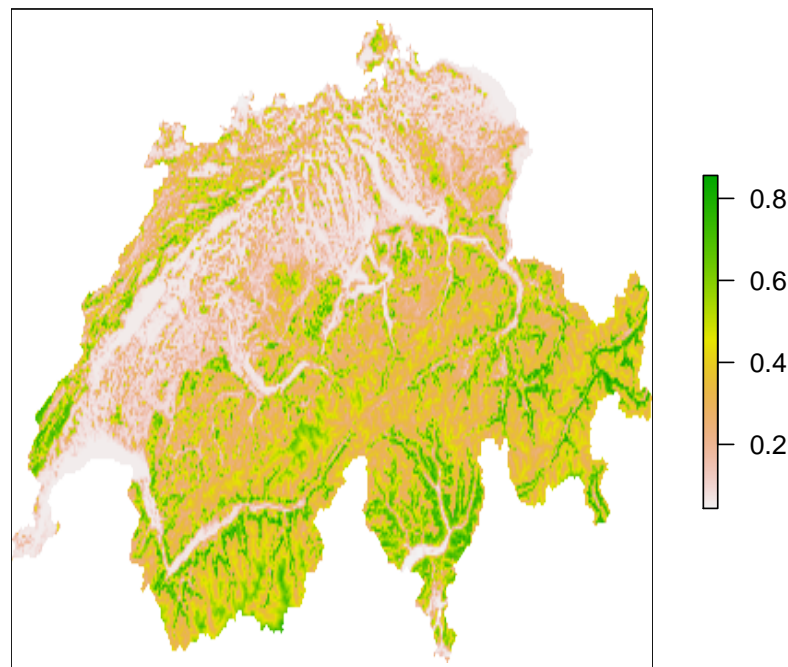

Figure 1: A species distribution map for the European crossbill in Switzerland for the year 1999. The colors represent the probability of occurrence.

```
> (beta <- coef(fm, type="col"))
> logit.col <- beta[1] + beta[2]*elevation + beta[3]*forest
> col <- exp(logit.col) / (1 + exp(logit.col))
> plot(col)
```

As of version 0.9-6, the `predict` method in `unmarked` can make predictions using an object of class `RasterStack` from the `raster` package. As mentioned previously, the rasters must be named, perhaps by using the `layerNames(someraster) <- somename` method. The object returned by `predict` is another raster stack with rasters for the expected values of the parameter of interest, the standard errors, and the upper and lower confidence intervals. The following example is very slow because there are many of pixels in the raster. The resulting map is shown in Fig.2.

```
> rasters <- stack(elevation, forest)
> E.psi <- predict(fm, type="psi", newdata=rasters)
> plot(E.psi, axes=FALSE)
```

Users should be cautious when predicting from models that have categorical predictor variables, *i.e.* `factor`s. The `raster` package does not have advanced methods for handling factors, and thus it is not easy to automatically create dummy variables from them as can typically be done using `model.matrix`. The safest option is to create the dummy variables manually before fitting the models, and to use the same variables as rasters for prediction.
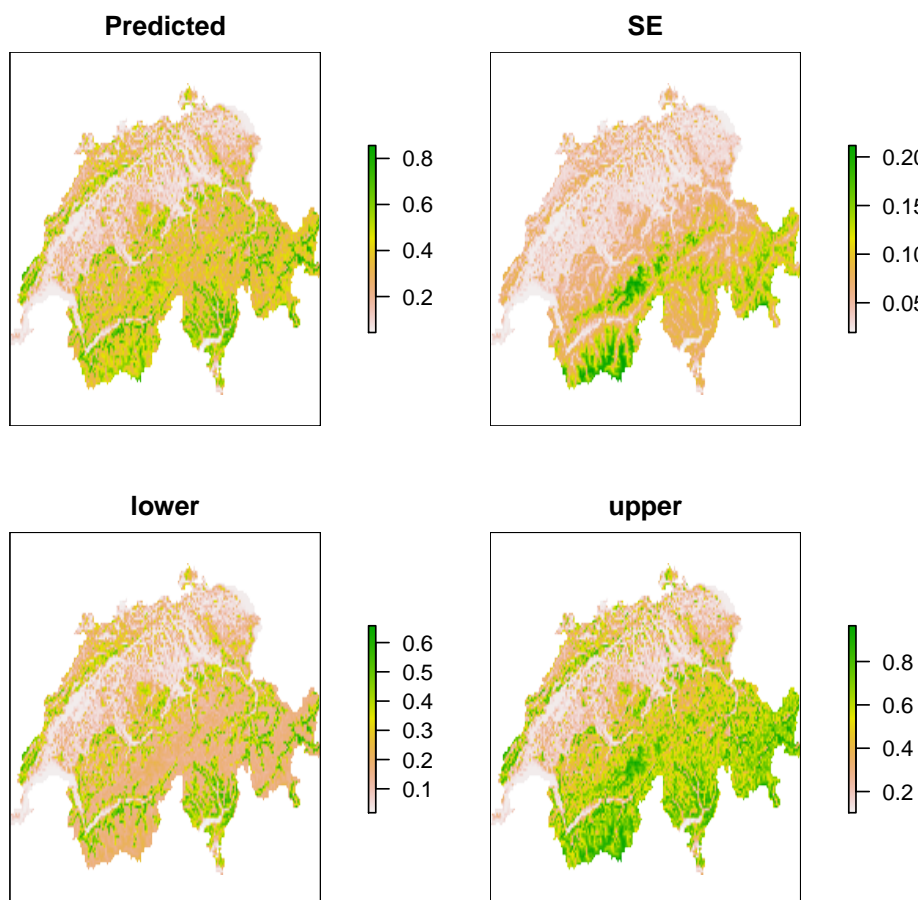

Figure 2: Expected occurrence probability along with standard errors and the limits of the asymptotic 95% confidence interval.

3

References

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin, 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* **84**:2200–2207.

Schmid, H., N. Zbinden, and V. Keller, 2004. Überwachung der bestandsentwicklung häufiger brutvögel in der schweiz. *Swiss Ornithological Institute Sempach Switzerland* .

van Etten, R. J. H. . J., 2012. raster: Geographic analysis and modeling with raster data. R package version 1.9-67.