Brief Overview of `unmarked`

Outline

A. Features of **`unmarked`**

B. Models included, and not yet included

C. Data structures

D. Exercise

Features of `unmarked`

Unified framework for

- Data manipulation
- Data exploration
- Model fitting (maximum likelihood)
- Model selection
- Model averaging
- Goodness-of-fit tests
- Prediction
- Bootstrapping
- Empirical Bayes estimators
- Publication-quality graphics
- Species distribution maps
- Power analysis
- Simulation studies
- etc...

What models are available in `unmarked`?

Emphasis on hierarchical models of spatial and temporal variation in abundance or occurrence probability when detection is imperfect

- Single-season site occupancy model (MacKenzie et al., 2002)
- Royle-Nichols model (Royle and Nichols, 2003)
- Binomial N-mixture models (Royle, 2004b)
- Multinomial N-mixture models (Royle, 2004a)
- Distance sampling models (Royle et al., 2004)
- "Open population" versions of the above: (MacKenzie et al., 2003; Chandler et al., 2011; Dail and Madsen, 2011)

Models we would like to add

- Occupancy models with spatial correlation (Hines et al., 2010)
- Multi-state occupancy models (Nichols et al., 2007)
- False-positive occupancy models (Royle and Link, 2006; Miller et al., 2011)
- Temporary emigration binomial N-mixture model (Chandler et al., 2011)
- Dail-Madsen model for distance sampling or capture-recapture data
- Occupancy models for inference about species interactions (MacKenzie et al., 2004; Waddle et al., 2010)
- Multi-species occupancy models for inference about community parameters (Dorazio and Royle, 2005; Dorazio et al., 2006)
- Models with additional random effects
- Individual covariate models

Data structure

- Sample units are called "sites"
- We record multiple "observations" at each site.
  - In occupancy studies we make multiple "visits" to each site.
  - In double-observer sampling, we have 2 observations per site

	detection data		
	visit1	visit2	visit3
site1	1	1	1
site2	0	0	0
site3	1	0	0
site4	0	0	1

Data structure

- Sample units are called "sites"
- We record multiple "observations" at each site.
  - In occupancy studies we make multiple "visits" to each site.
  - In double-observer sampling, we have 2 observations per site
- Site-specific or observation-specific covariates

	detection data			site covariate	observation covariate		
	visit1	visit2	visit3	habitat	date1	date2	date3
site1	1	1	1	good	3	6	10
site2	0	0	0	good	1	7	11
site3	1	0	0	bad	2	9	12
site4	0	0	1	bad	5	6	10

Data structure

All data must be put in an object called an `unmarkedFrame`

```
umf <- unmarkedFrame(y=detectionData,
                     siteCovs=siteData,
                     obsCovs=list(wind=windData, date=dateData))
```

Why bother?

- The data structure is not amenable to simple `data.frame`s and `list`s.
- Metadata such as distance-interval cutpoints or measurement units can be bundled with the data
- Makes it easy to detect errors when creating the `unmarkedFrame`
- Allows for custom tools to summarize and visualize data

Data structure

Most `unmarkedFrame`'s have the following compontents in common:

Component	Description
`y`	nSite×nObs matrix of observations (e.g. counts)
`siteCovs`	nSite×nCovs data.frame of site-specific covariates
`obsCovs`	A list with nCovs components. Each component is nSite×nObs data.frame of covariate values.

Data structure

Most `unmarkedFrame`'s have the following compontents in common:

Component	Description
`y`	nSite×nObs matrix of observations (e.g. counts)
`siteCovs`	nSite×nCovs data.frame of site-specific covariates
`obsCovs`	A list with nCovs components. Each component is nSite×nObs data.frame of covariate values.

Others `unmarkedFrame`'s have additional components:

Component	Description
`numPrimary`	The number of primary periods (seasons, years, etc...)
`yearlySiteCovs`	A list with nCovs components. Each component is nSite×nYear matrix of covariate values.
`dist.breaks`	A vector defining the distance intervals for a distance sampling analysis
`unitsIn`	Measurement units for distance data

Data structure

Unique flavors of `unmarkedFrame`'s exist for each class of models.

Closed population models

Model	Fitting function	unmarkedFrame
Site-occupancy	occu	unmarkedFrameOccu
Royle-Nichols	occuRN	unmarkedFrameOccu
N-mixture models	pcount	unmarkedFramePCount
Removal models	multinomPois	unmarkedFrameMPois
Double-observer models	multinomPois	unmarkedFrameMPois
Arbitrary multinomial models	multinomPois	unmarkedFrameMPois
Distance sampling models	distsamp	unmarkedFrameDS

Open population models

Model	Fitting function	unmarkedFrame
Site-occupancy	colext	unmarkedMultFrame
N-mixture models	pcountOpen	unmarkedFramePCO
Removal models	gmultmix	unmarkedFrameGMM
Double-observer models	gmultmix	unmarkedFrameGMM
Arbitrary multinomial models	gmulmix	unmarkedFrameGMM
Distance sampling models	gdistsamp	unmarkedFrameGDS

Data structure

Unique flavors of `unmarkedFrame`'s exist for each class of models.

Closed population models

Model	Fitting function	`unmarkedFrame`
Site-occupancy	`occu`	`unmarkedFrameOccu`
Royle-Nichols	`occuRN`	`unmarkedFrameOccu`
N-mixture models	`pcount`	`unmarkedFramePCount`
Removal models	`multinomPois`	`unmarkedFrameMPois`
Double-observer models	`multinomPois`	`unmarkedFrameMPois`
Arbitrary multinomial models	`multinomPois`	`unmarkedFrameMPois`
Distance sampling models	`distsamp`	`unmarkedFrameDS`

Open population models

Model	Fitting function	`unmarkedFrame`
Site-occupancy	`colext`	`unmarkedMultFrame`
N-mixture models	`pcountOpen`	`unmarkedFramePCO`
Removal models	`gmultmix`	`unmarkedFrameGMM`
Double-observer models	`gmultmix`	`unmarkedFrameGMM`
Arbitrary multinomial models	`gmultmix`	`unmarkedFrameGMM`
Distance sampling models	`gdistsamp`	`unmarkedFrameGDS`

Typical work flow

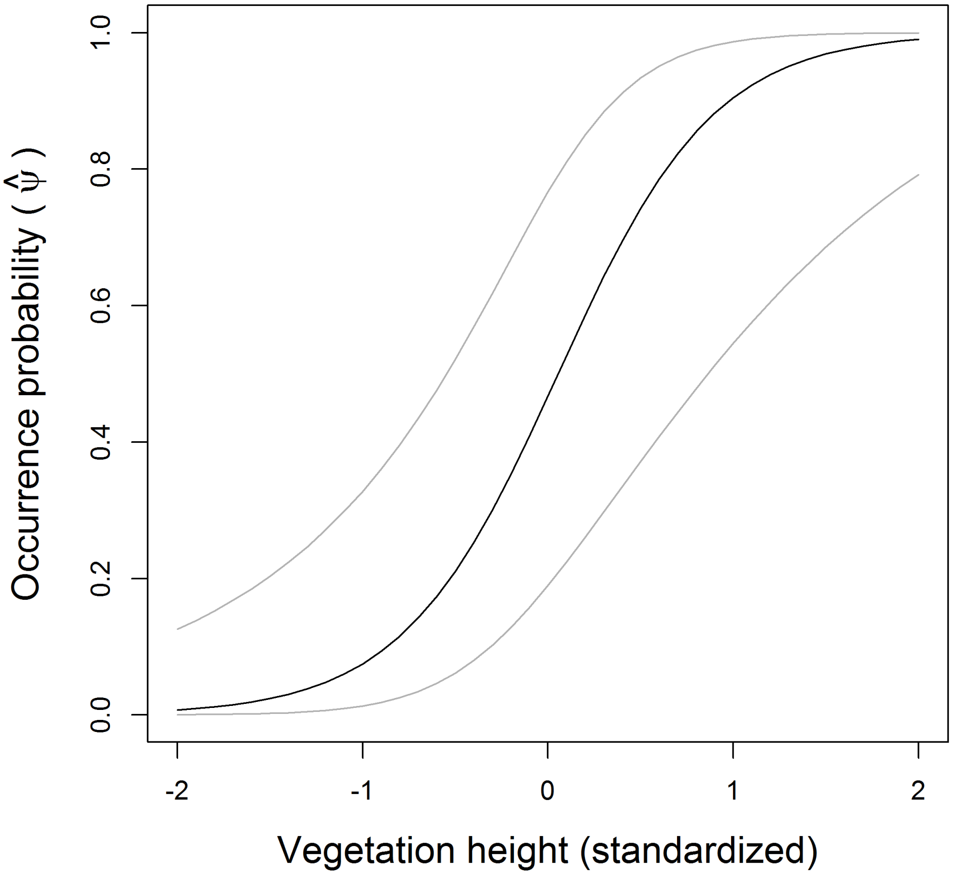
3 Parts

- Import data, format it, create `unmarkedFrame`
- Fit some models
- Analyze the results. Assess model fit, make predictions, map abundance, etc...

Phony example

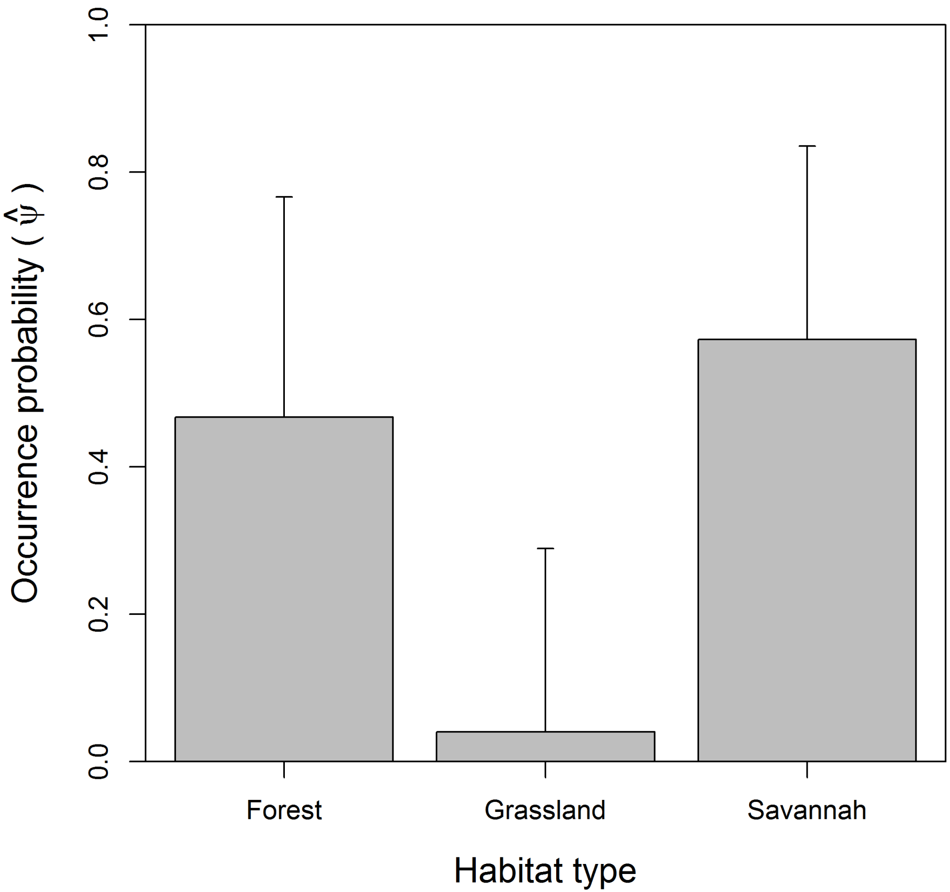
```
# PART I (import data and create unmarkedFrame)
library(unmarked)
pointCountData <- read.csv("myPointCountData.csv")
siteCovariates <- read.csv("mySiteCovariate.csv")
pointCountUMF <- unmarkedFramePCount(y=pointCountData,
    siteCovs = siteCovariates)

# PART II (fit 2 models)
fm1 <- pcount(~1 ~1, data=pointCountUMF)
fm2 <- pcount(~vegHt ~vegHt, data=pointCountUMF)

# PART III (model selection, model fit, prediction, mapping)
fms <- fitList(fm1, fm2)
modSel(fms)  # model selection
parboot(fm2) # goodness of fit
predictions <- predict(fm2, type="state", newdata=mapdata)
levelplot(Predicted ~ x.coord+y.coord, data=predictions)
```

Look ahead



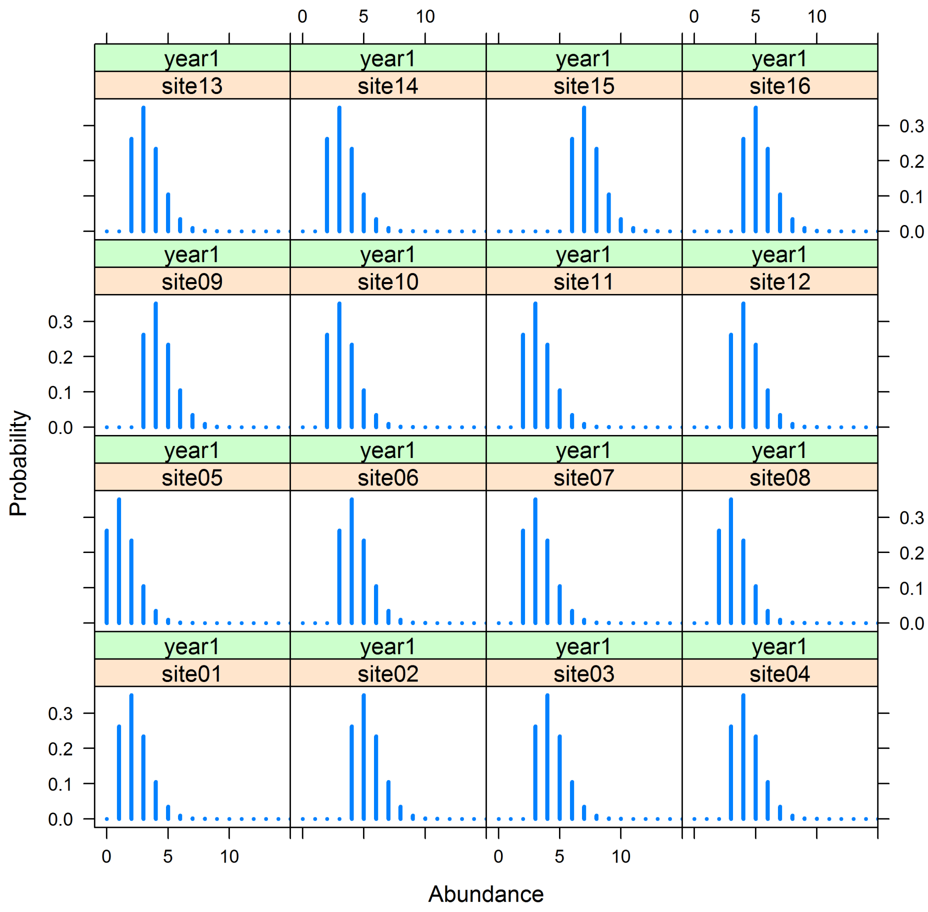
Occurrence probability ($\hat{\psi}$)
1.0
0.8
0.6
0.4
0.2
0.0
-2
-1
0
1
2
Vegetation height (standardized)

Look ahead



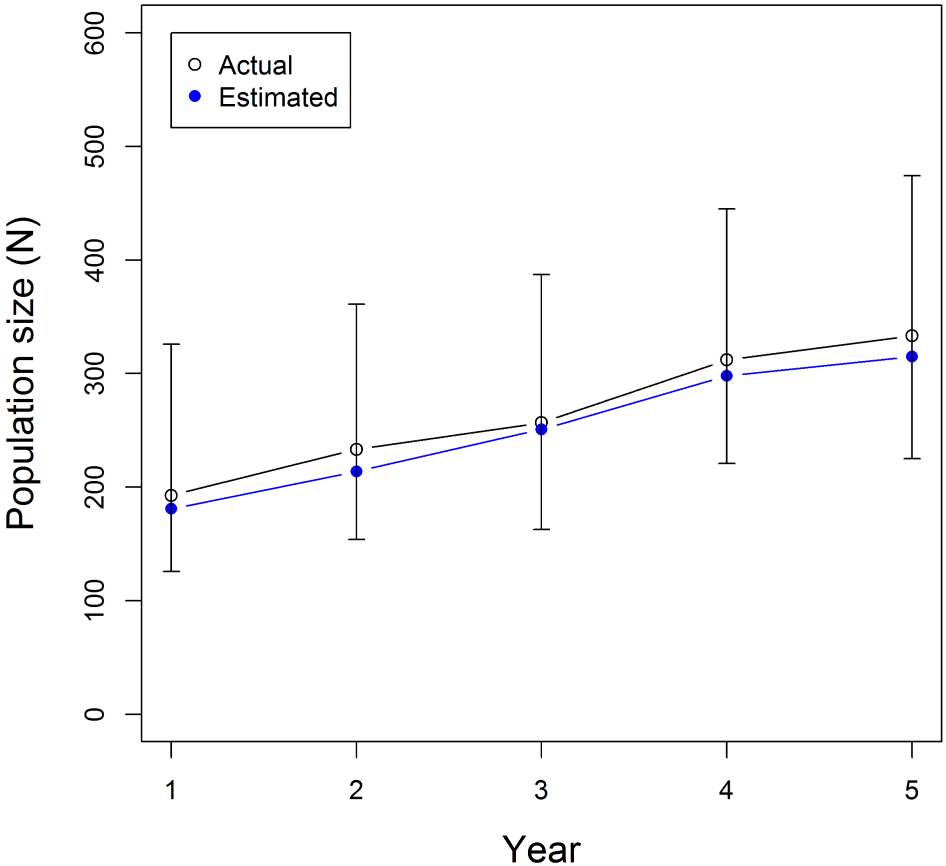
Occurrence probability ($\hat{\psi}$)
0.0
0.2
0.4
0.6
0.8
1.0
Forest
Grassland
Savannah
Habitat type

Look ahead



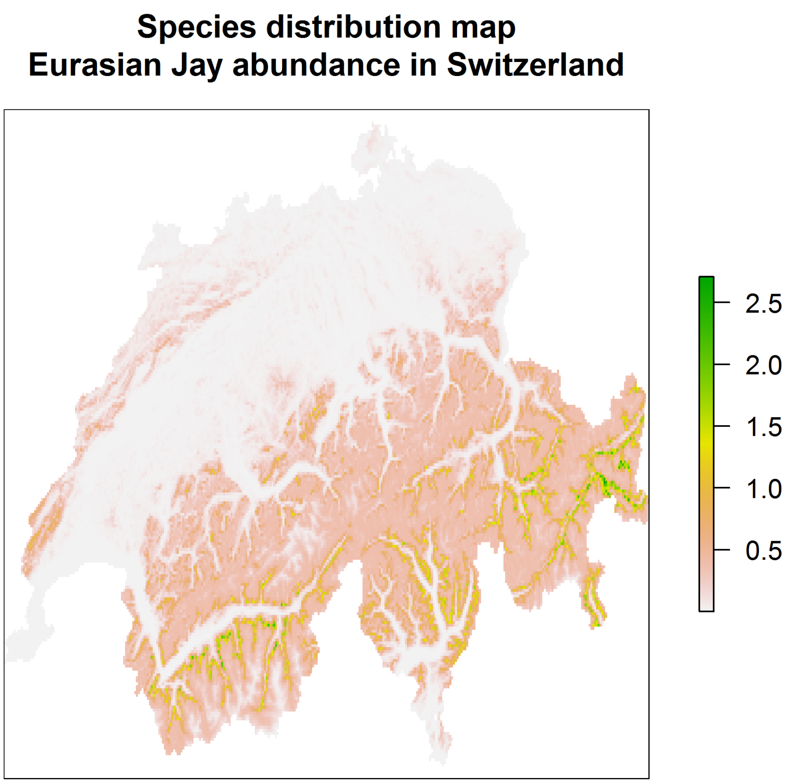
year1
site13
year1
site14
year1
site15
year1
site16
year1
site09
year1
site10
year1
site11
year1
site12
year1
site05
year1
site06
year1
site07
year1
site08
year1
site01
year1
site02
year1
site03
year1
site04
Probability
Abundance
0.0
0.1
0.2
0.3
0
5
10

Look ahead



Actual
Estimated
Population size (N)
0
100
200
300
400
500
600
1
2
3
4
5
Year

Look ahead


Species distribution map
Eurasian Jay abundance in Switzerland
2.5
2.0
1.5
1.0
0.5

Online Resources

Materials for this workshop

https://sites.google.com/site/unmarkedinfo/home/workshops/patux2012

Home page

https://sites.google.com/site/unmarkedinfo

Users' group

groups.google.com/group/unmarked

References

Chandler, R., Royle, J., and King, D. (2011), "Inference about density and temporary emigration in unmarked populations," *Ecology*, 92, 1429–1435.

Dail, D. and Madsen, L. (2011), "Models for estimating abundance from repeated counts of an open metapopulation." *Biometrics*, 67, 577–87.

Dorazio, R. and Royle, J. (2005), "Estimating size and composition of biological communities by modeling the occurrence of species," *Journal of the American Statistical Association*, 100, 389–398.

Dorazio, R., Royle, J., Söderström, B., and Glimskär, A. (2006), "Estimating species richness and accumulation by modeling species occurrence and detectability," *Ecology*, 87, 842–854.

Hines, J., Nichols, J., Royle, J., MacKenzie, D., Gopalaswamy, A., Kumar, N., and Karanth, K. (2010), "Tigers on trails: occupancy modeling for cluster sampling," *Ecological Applications*, 20, 1456–1466.

MacKenzie, D., Bailey, L., and Nichols, J. (2004), "Investigating species co-occurrence patterns when species are detected imperfectly," *Journal of Animal Ecology*, 73, 546–555.

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

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002), "Estimating site occupancy rates when detection probabilities are less than one," *Ecology*, 83, 2248–2255.

Miller, D., Nichols, J., McClintock, B., Grant, E., Bailey, L., and Weir, L. (2011), "Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification," *Ecology*, 92, 1422–1428.

Nichols, J. D., Hines, J. E., MacKenzie, D. I., Seamans, M. E., and Gutierrez, R. J. (2007), "Occupancy estimation and modeling with multiple states and state uncertainty," *Ecology*, 88, 1395–1400.

Royle, J. A. (2004a), "Generalized estimators of avian abundance from count survey data," *Animal Biodiversity and Conservation*, 27, 375–386.

— (2004b), "N-mixture models for estimating population size from spatially replicated counts," *Biometrics*, 60, 108–115.

Royle, J. A., Dawson, D. K., and Bates, S. (2004), "Modeling abundance effects in distance sampling," *Ecology*, 85, 1591–1597.

Royle, J. A. and Link, W. A. (2006), "Generalized site occupancy models allowing for false positive and false negative errors," *Ecology*, 87, 835–841.

Royle, J. A. and Nichols, J. D. (2003), "Estimating abundance from repeated presence-absence data or point counts," *Ecology*, 84, 777–790.

Waddle, J., Dorazio, R., Walls, S., Rice, K., Beauchamp, J., Schuman, M., and Mazzotti, F. (2010), "A new parameterization for estimating co-occurrence of interacting species," *Ecological Applications*, 20, 1467–1475.