When to use pcount vs distsamp vs gdistsamp

941 views

Subscribe



Amy J's profile photo

Amy J

Jun 23, 2016, 1:39:08 PM







to unmarked

Hi All,

I have been struggling with deciding which model in unmarked to use for my dataset as I took several different measures to account for detection probability while conducting grassland bird surveys for my dissertation research.

In summary, I used 10-min point counts to collect information on grassland bird abundance/density in fields (n=~70) over a 4-year time period.  For each field, I have both spatial (n=3 point count stations) as well as within-season temporal (n=4 visits) replication of counts for at least 2 years.  I divided each 10-minute point count into two 5-minute intervals (to incorporate a removal method) and two distance bands (0-49m, 50-100m).  Habitat variables were measured at least once for each site and detection covariates were measured for each station on each survey.   
  
I’d like to look at the effects of habitat covariates on grassland bird abundance/density while accounting for detection probability of each species. I keep going back and forth between distsamp, pcount, and gdistamp and am wondering if anyone has any insight as to which method is best for this type of dataset?

Thank you in advance for your input.

Amy

Tomas Telensky's profile photo

Tomas Telensky

Jun 23, 2016, 2:23:11 PM







to unma...@googlegroups.com

Hi Amy,

this is very interesting dataset really. I did something similar but without the removal smapling. I would definitely go for gdistsamp which combines both distance sampling and multiple visits. However, as far as I know, it doesn't directly support removal sampling. I wonder if there is something like the  NA trick Marc recommended here recently, that you put NA in the second 5 minutes if the species was observed in the first 5 mins, but this obviously doesn't work for counts. And you have hierarchy of three replicate levels: between years, within season and within one visit, they are of course not independent, the closer the more correlated. You definitely need fixed effect for year here,  and AFAIK this is not possible in gdistsamp. I would either go for modifying gdistsamps likelihood function, or writing a model in JAGS. Please note though that there will be big convergence issues in both variants! Even in my case, which was much simpler, many times i didn't achieve convergence in neither gdistsamp nor JAGS.

Anyway, you have data from only 3 points? I am afraid that wont be enough for these models, Buckland says you need at least 60-100 observations for bare distance sampling and these models are much more complex..

Just my experience, perhaps you get more exact reference from Marc or Andy. I would be also curious myself :-)

Regards!

Tomas





--  
You received this message because you are subscribed to the Google Groups "unmarked" group.  
To unsubscribe from this group and stop receiving emails from it, send an email to [unmarked+u...@googlegroups.com](https://groups.google.com/).  
For more options, visit <https://groups.google.com/d/optout>.

Kery Marc's profile photo

Kery Marc

Jun 24, 2016, 3:23:22 AM







to unma...@googlegroups.com

Dear Amy,  
  
I agree with Tomas that is a sampling design of impressive complexity. When thinking about the structure in it that needs modeling, it is useful to conceptually keep apart design features of the "ecological process" and design features of the "observation process".  
  
First, you have a complex sampling design with nesting and repeated measurements in the ecological process: 70 fields, with 3 sampling points in each of them, and this is repeated over 4 years. So this gives you a lot of power to make comparisons in space (between fields and between sampling points) as well as some in time or space-time (between years). Clearly, your main biological hypothesis live here: you want to understand correlates of spatiotemporal variation in abundance.  
  
Then, to quantify measurement error (imperfect detection), your design has two or three features:  
- a time-removal component (2 occasions)  
- a distance sampling component (2 distance bands)  
- and the repeated counts within a season  
  
The former two give information about detection, while the repeated counts could be used either to provide info about detection (hence, observation process) or else about within-season changes in availability (this is more ecological part). This part of the model is your measurement error model.  
  
Here are some possible ways of analysis.  
  
**Analysis in unmarked**If you want to do an analysis in unmarked, then on the observation model side, use of gdistsamp would let you use the DS + point count aspect, while gmultmix would let you use the time-removal + point count aspect. You cannot accommodate all three features of the observation process in a model currently implemented in unmarked, so you would have to ignore either the time-removal or the distance sampling component. I have no intuition which of time-removal or distance sampling data are "stronger" (contain more information in some sense), but clearly you should retain the more informative kind of data.  
  
On the ecological side of the model, you have both nesting (3 points nested per field) and then repeated measures for 4 years. In a traditional ANOVA context, you would model points and to account for non-independence of points in the same field add a random field effect, and likewise add a random year effect as well. It is not possible in unmarked to add more random effects to the hierarchical models, so you would perhaps have to ignore the field factor and treat year also as a ("fixed-effects") factor. This is not entirely satisfactory, but I don't know of any software that has canned models to accommodate all features of your design or even more than you can do with unmarked.  
  
**Analysis in detect**Peter Solymos' R package detect has the cmulti function, which does the combination of time-removal and distance sampling, but I think it could not cope with all three aspects of your observation model either. Otherwise, in terms of the ecological model, I believe that the same limitations would apply as for unmarked.  
  
**Analysis in BUGS**Possibly the best course of action would be to fit the full model in BUGS, at least if you already know how to program in the BUGS language and have more than say 6 months left in your project (not all of which would have to be dedicated to the BUGS programming though). In BUGS it would be easy to fully accommodate all aspects of non-independence in the ecological part of your model, plus you would enjoy some other benefits of Bayesian inference using MCMC methods (such as ease of inference about functions of parameters). The AHM1 book (section 9.3) has a code example for the combination of distance sampling and time removal and I think it would not be very difficult to accommodate the within-season replication as well.  
  
Best regards  --  Marc

**From:** [unma...@googlegroups.com](https://groups.google.com/) [[unma...@googlegroups.com](https://groups.google.com/)] on behalf of Tomas Telensky [[tomas.t...@gmail.com](https://groups.google.com/)]  
**Sent:** 23 June 2016 20:23  
**To:** [unma...@googlegroups.com](https://groups.google.com/)  
**Subject:** Re: [unmarked] When to use pcount vs distsamp vs gdistsamp



Amy J's profile photo

Amy J

Jun 24, 2016, 10:44:50 AM







to unmarked

Hi Marc and Tomas,

Thank you for your thorough responses.  I was worried you might recommend BUGS as I'm very much a beginner when it comes to that!  I do have the AHM1 book and will give it a shot using the example in chapter 9.3.  I will likely post more questions as they come up!

Best,

Amy



Kery Marc's profile photo

Kery Marc

Jun 24, 2016, 12:36:01 PM







to unma...@googlegroups.com

Dear Amy,  
  
I suggest you don't be afraid of BUGS but ... I think it is a good idea to start with an analysis in unmarked or detect by ignoring one of the features in the observation model. By "stacking" your data (one year on top of the other) and treating year as a factor (as for the field) you can obtain inferences for a fairly reasonable model.  
  
Best regards -- Marc

**From:** [unma...@googlegroups.com](https://groups.google.com/) [[unma...@googlegroups.com](https://groups.google.com/)] on behalf of Amy J [[amy3mj...@gmail.com](https://groups.google.com/)]  
**Sent:** 24 June 2016 16:44  
**To:** unmarked





Amy J's profile photo

Amy J

Jun 24, 2016, 5:22:22 PM







to unmarked

Thank you so much Marc, I will start with that and let you know how it goes.

Enjoy your weekend,

Amy





Gretchen Nareff's profile photo

Gretchen Nareff

Jun 4, 2018, 4:42:01 PM







to unmarked

I just wanted to check in and see if anything has come about in the past two years to model repeated counts with distance and time removal? I have a similar dataset as Amy and I'm curious what she ended up doing. I've been using pcountopen (because I am looking at the effects of harvesting between years), but I started playing around with gdistsamp yesterday.

5 study areas with 10-40 point count stations each

Each point count was sampled 3 times in a breeding season (10-min counts split into five 2-minute intervals and five distance bins)

Each point count was sampled at least one year pre-harvest and at least one year post-harvest

Harvests occurred in different calendar years so I use a years-since-harvest effect instead of year. I have some sites with one or two years of pre-harvest data and two or three years of post-harvest data. To handle that, I am only using data from the season immediately preceding the harvest and then one-year since harvest. I have a separate analysis for post-harvest only data so that I can take advantage of more of the data that I collected. For that I used a stacked pcount approach.

With the gdistsamp I ran post-harvest data and got results that make sense. I am going to try stacking the pre- and post-harvest data and adding a harvest variable.

I want to look at change in songbird abundance pre- vs post-harvest and if that change in abundance is correlated to a particular vegetation or topographical variable (aspect, slope position, basal area, etc.)

Thanks!

Gretchen

Jeffrey Royle's profile photo

Jeffrey Royle

Jun 4, 2018, 6:14:53 PM







to unma...@googlegroups.com

hi Gretchen,

 I'm afraid we haven't added any of these capabilities to unmarked in the mean time.   More likely unmarked is in a static state for the foreseeable future, unless we find an ambitious and motivated grad student who wants to expand some of the capabilities.  (possibly we will add some Dail-Madsen distance sampling models in the next year but that's the only thing "in development").

 The other thing I might do is add some non-Euclidean distance models to the distance sampling functionality (but this is item #27 on my "to do list").

 All of that said, I think you can develop the models you want pretty easily in JAGS in some generality.

 It's possible that you could fake unmarked into fitting the repeated removal/distance sampling model by thinking hard about how to build a custom piFun  (I'm fairly certain this could be done but it's an "advanced unmarked trick"....).

 regards,

andy