**Comments on model selection and assessment chapter**

Content: Good, really interesting and insightful. Form: should be polished a bit. In general, I would try to eliminate a little of the following two things:

- too much relativism: e.g., “we show a couple of methods which might perhaps be useful under some circumstances” (this example is invented)

- too indirect descriptions where a more direct one is entirely sufficient and shorter and hence better. Eg., instead of “There are two distinct classes of problems ...” (line 56) I would simply say “two problems” or “two challenges“ or “two tasks”.

**Specific things**

- lines 5-7: I would perhaps number the tasks like this: (1), (2), ... Or else first simply mention the four of them and only afterwards explain what each means

- 10: comma after latent variables

- 19: indicator variable model selection. Also, to check model adequacy could mention bootstrap around there

- 30: add a few cites here

- 1.1. in title would write model selection, since the same term is also written lower-case in 1.1.1. (or does this have a system ?)

- around 45: could perhaps mention that DIC seems to be particularly unusable for mixture models (that’s what I believe). Also, this makes me feel ashamed that I have still not summarizued that simulation with covariate selection by DIC and AIC for a site-occ model, which I ran almost 6 months ago ...

- 50 and throughout: O’Hara

- At some places I thought that there were lots of good ideas and interesting stuff in there but that they are a little unfinished. For instance, most of 1.1.1. looked interesting to me and I think I could understand some of the things you say, but I still found it fairly vague.

- 64: you mention model Mb for the biological side ? Does this make sense ?

- I made a couple of notes next to the first para of page 3: (1) You say that encounter models are not biological constructs but rather mere descriptions. To me, the terms mechanistic vs. phenomenological are mostly relative, in that no model can ever be totally mechanistic in a sense and will always be merely descriptive in a sense. Then, you say that the standard models are all unrealistic. I wouldn’t say such a thing about a model I am goind to use, it sounds to absolute and might be read almost as if you suggest not to use the model. It sounds too harsh to me. Also, there are lots of unrealistic models that are useful and in many cases adequate. For instance, why should we ever assume anything is linear ? Probably, most things are really non-linear (and not only in a graphical way, but probably also in a mathematical way --- but I don’t really know). Then, I wonder whether you should critique distance sampling so harshly here. They will just start to hate you in S-Andrews and poison your beer when you go there sometime.

- after line 87: partly different font in equation.

- 106 something missig

- 116: 3 S

- code (e.g., top of page 5): would add comments to degree possible. Also, add space before and after assigment arrod, makes it more readable

- bottom of that page: I hate the engineering display of numbers (e.g., 10e-05). I can never understand this without hard thinking. If possible, I would show as normal frational numbers. In R, these can be enfored by typing options(scipen = some large number, e.g. 10 or 40)

- head of table 1.1. Delete period after mask

- 205: delete ‘terminology’ after SECR

- top p. 8: 3 times idea. Prune.

- same place: perhaps call the indicator variable w as in your own papers where you used it ? I sounds like the indicator function to me.

- 264: cite DIC reference here

- 273: delete a period

- start of 1.2.1.: I found this very vague, but unfortunately, I don’t haven’t anything to add. Except perhaps to remind you that I believe that the DIC is particularly inadequate for mixture models (think I read this somewhere).

- 282: seem

- 292: perhaps write theta\_hat instead of theta ?

- last line of that page: p\_D, i.e., D is capital in most places

- 305: add ‘did’

- 306: do you need to say this ? Sounds standard

- 327: I found this interesting, because I just experimented with a similar way to switch on or off components of a big general Nmix model

- Panel 1.1. Perhaps say that this is NOT the actual model in the function, because the whole on and off switch is missing ?

- middle of p. 11: made the note that DIC needs sometimes very long runs to stabilise

- 392: add comma before ‘say’. Also, would prefer w to I

- 395: let’s, then individual-level covariate

- 396: individual-, trap and occasion-specific (i.e., add comma)

- 398: individual-level (not sure about this one, but I believe it should be so)

- 403: ‘as we demonstrate shorly.’

- 422: mixture from the prior (when a value of the indicator variable of 0 is sampled) and draws informed by the data (i.e., from the posterior, when a 1 is drawn for I).

- 435: why toad btw ?

- 449-450: I would give same number of digits

- 452: qualitatively

- 481: might add ‘(e.g., half-normal versus hazard rate)’

- 482: twiddle is missing

- around 499: here I would refer to some place in the book or else to a paper or other book where algebraic expressions for these functions can be found

- 504: why quotes around intercept ? I would use quotes sparsely (not that I always do...)

- 577: these values of pd and DIC look suspicious to me

- 584: adequacy

- 591: evaluating fit

- 591-601: In know that you could write several books about this, but I thought that more could be said about this. Also, at the end of this I was thinking of that other famous stats quote from Tukey, which goes about like this: “A heap of data and a burning desire to analyse them does not guarantee that something sensible comes out of it” (or something like this)

- 602: evaluation of model fit

- 604: here could perhaps cite some statistica VIPs ? Like Rubin or Gelman et al. (StatSin 1996) ?

- 607: too many ‘then’

- 617: this is not closed pop models, but you could perhaps cite the classical Chisquare decompositions in program RELEASE for CJS assumptions (and I believe that this is described in the Lebreton et al. 1991 EcolMono paper).

- 618: ‘similar idea’ seems to refer back to “test with known theoretical properties”, which is does not exclusively

- lines 665 and 670 sound like a contradiction to me

- in the equation after line 687, couldn’t you wite G instead of #cells ? Same also on 691

- and then around the same place, I would explicitly mention “variance to mean ratio” or something to clarify the idea of this test abundantly

- around 710 punctuation is not quite right (e.g., would start sentence with cap)

- 712: tallying up

- 717: .... models is (not are)

- stuff on page 21 is great contentwise, but rough format wise (at least top of page)

- 726. in order to do a check of model fit in the presence of a binary response

- 728: bin into bins is not beautiful

- 729: comma after N\_g

- 731: tends to: this is always the case ? E(N\_g) = N/G and when we make very small bins, then N/G will become binary, right ?

-742: in typical SCR applications

- 743: might add “and presumably not much power”

- 749: point process on which inference is focused

- footnote on p. 21 is not necessary !

- 768: remind people what the indices are here

- 773: three instead of 3 (JWM style !)

- 775: we instead of you. I think you have mostly the we perspective

- around 772, where you say that one needs to aggregate over some dimension of the full data array. I found this interesting, because this is what I just did in a dynamic occupancy model for the European lynx in the Alps, where we added up the #detections over the three seasons per year and did a bayesian p-value for a chisquare-type discrepancy. Interestingly, the model did not fit (p was almost 0), but then, and I would be interested what you think about this idea, we expressed the degree of lack of fit by dividing the value of the sum of the test statistic for the actual data by that for the perfect data and interpreted it as a measure for lack of fit, in analogy to the c\_hat that is so widely used in capture recapture. We had a “lack of fit ratio” (i.e., pseudo c\_hat or something) of 1.14 I believe and so we argued that this is negligible, because it would make our SEs too short by a factor of 1/sqrt(1.14). If this reasoning is sensible, then I think this is an important thing, because it allows us to quantify the magnitude of the lack of fit and if this is small, we may feel justified to simply ignore it.

- 792: too disaggregated

- 795: explains part of ?

- 799: heterogeneity and M\_h-like statistic

- 805: do you need a ‘-’ at carry out ?

- 807: trap- or time-specific

- 817: I feel that the idea of using CV for model evaluation is so important that it should be introduced much earlier in the chapter, even if you don’t say much abot it

- start of page 24: the sequence of periods and semicolons does not make sense

- 839: relatively tight compared to ....

- 858: cool stuff !

- 868: aren’t close to 0 or 1 at all

- 890 and elsewhee: I would perhaps call X1 T to make it consistent with the algebraic expressions for these quantities

- around there: again, what about this idea of quantifying the degree of lack of fit ? For a chisquare discrepancy, the ratio of the value for the observed data and that for the expected (under the model) data could be used as something like a c\_hat value in caprecap. Or perhaps as exactly the same thing who knows ? This has the advantage that people know this. And really, we have learned that it is better to estimate a parameter than to simply report a significance test value for the question of whether the parameter could be zero, so in a similar vein, wouldn’t it be more informative also to have a quantity that measures the *magnitude of lack of fit* rather than simply telling us the model fits or not with a probability ?

- sentence starting on line 907 is not correct

- 909: one trap

- top of page 26: that’s extraordinary that with a simple move to a M\_b model you could get rid of all lack of fit !

- 962: after the dash: this implies an equivalence with both things mentioned before, which is not true

- sentence right after that is a prime example of where I feel you relativize way too much.

- 982: can two-step model assessment be done with bootstrapping ?

- also (general comment for this chapter): you might want to cite Link and Barker with the two-step assessment in a Bayesian p-value, they do that for their model of muskrat houses somewhere.

- 989: relevant instead of interesting ?

- 1000: can you back up this statement with some references or further explanation ? Also, I think it would be good to end on a more positive note.

- 1003: Aitkin cite is incomplete

- 1055: Richard Chandler got lost here !