9 – I’d mention Dail-Madsen models too…

17-18 – I’d love to discuss the use of “nonrandom” here. My understanding is that its biased when its random but not randomized a priori. Seems like subsequent developments treat the sampling as if it arises from a stochastic process with an unknown intensity function, so the term “nonrandom” seems ill-advised to me.

62 – perhaps state here that Z ~ f(mu) is the terminology for the density process model, and Y ~ f(nu) is for the sampling process model (I noticed myself getting confused between Z and Y and mu and nu)

71-77 – I would say that something more about scale, i.e., that the distinction between fine and course is implicitly made by the scale of sampling units U, and that multiscale models might eventually eliminate this distinction. Also perhaps other “fine-scale” processes, e.g., seasonal or daily timing, sampler attunement to numerically dominant species in a multispecies survey, etc.

95 – I think a good point for establishing intuition is that, in the design-based world, unequal sampling intensity in a random-sampling design is accommodated via stratification. Stratification (i.e., estimating a fixed effect for every stratum) also arises as the limit of our proposal. A though experiment shows that if there’s different categories of sampling intensity for U’s, such that U(1) through U(n1) having sampling intensity lambda(1), U(n1+1) through U(n2) have sampling intensity lambda(2), etc., then our method could result in:

p \_proportional to\_ exp( beta1\*lambda + beta0 )

where lambda is the vector of {lambda(1), lambda(2) …}, and where lambda is then a covariate in the function predicting density too. However, this is a highly parametric assumption for PS. Instead, the nonparametric form would be:

p \_proportional to\_ exp( beta \* Design(lambda) }

where Design(lambda) is a design matrix with number of columns equal to the number of unique levels of lambda, etc (the notation sucks but hopefully you see my point), and where Design(lambda) is then the design-matrix for stratification used in the function for predicting density. Anyway, stratification arises as the limit of PS given categorical sampling intensity, and a nonparametric treatment of the link between p and sampling intensity.

126 – “for the process model [of sampling locations]” (we’ll have a separate process model for density later)

127 – is delta a vector (single spatial process) or matrix (multiple processes)? perhaps note that, if delta is a matrix, the use of B in Eq. 9 means that Eq. 7 could include a spatial process (column of B) that is not shared by the density process (Eq. 9). This speaks to our original discussiong, where we had a spatial process in each of the density and sampling processes, and a third that was shared between the two.

173 – This bias-correction accounts for the nonlinear transformation of random effects when calculating the derived variable of interest, where the MLE provides an optimal estimate of random effects:

Epsilon\_hat = E[ elpsilon | theta\_hat ]

Where theta\_hat is the MLE of fixed effects, epsilon is the random effects, and Epsilon\_hat is the empirical Bayes estimate of random effects, but where

f( E[ epsilon | theta\_hat] ) != E[ f(epsilon) | theta\_hat]

i.e., a nonlinear transformation of the Empirical Bayes estimator is no longer identical to the optimal estimator for the nonlinear transformation. Anyway, our paper on the topic is accepted in Fish Res and I’ll send it along.