

This is really well done. I think we will have better luck with an ecological journal, and it will get read more with more citations. Ecological applications and MEE would be my top choices. Because MEE is marginally more methods focused, I'd probably make that my first choice.

For our discussion next week, the bigger points below are numbers 27, 33, and 39-40, 50, and 52. I think these are pretty easily resolved. All of the rest of the comments are editorial.

1. line 22: Old-school Bayesians (and some today, although the borders have blurred) had a 3rd general approach to statistical inference based on personal uncertainty. We could add the qualifier, "...to make frequentist statistical inferences..." but not sure it is really necessary. A real Bayesian might get wound up being left out of "two general approaches".
2. line 40: Should be "Cooper (2006) reviews ..."
3. line 53-55: I think it should be "point locations" and these are infinite populations, and "polygon that create aerial data" are finite. It is important to stress the spatial support here, and not the response variable.
4. line 56: Perhaps ... "we introduce and compare several sampling and estimation procedures for design-based..."
5. line 59: I think it would be better to give the application, "we use both approaches to analyze real data consisting of mercury concentration from lakes in the contiguous United States."
6. line 93: I think that "...realization of a data-generating stochastic process..." sounds better.
7. line 97: missing a period to end sentence.
8. line 101: I think we should clarify, "...fixed but unknown parameters, such as the mean (as in..."
9. line 106: "...then derived using assumptions..."
10. line 107: It might be less confusing to say "...from the same spatial process..."
11. line 108: "...repeatedly obtained samples, then ..."
12. line 112: "... in Figure 1. Ver Hoef (2002) and Brus (2020) provide similar figures."
13. lines 131-137: I think a few sentences would be better than this long list separated by semicolons.
14. line 160: "... tends to reduce..."

15. line 166: “Instead of developing inference based on a specific sampling design...”
16. line 169: “...be an  $N \times 1$  response vector at locations  $s_1, s_2, \dots, s_n$  that can be measured at the  $N$  population units. [delete rest of sentence].
17. line 177: “... with subscript  $u$ , let ”
18. line 179: bold beta
19. line 179: “...fixed effects. [[new paragraph]]  
Let  $\boldsymbol{\delta} \equiv [\boldsymbol{\delta}_s, \boldsymbol{\delta}_u]'$ , where  $\boldsymbol{\delta}_s$  and  $\boldsymbol{\delta}_u$  are random errors for the sampled and unsampled population units, respectively. We assume  $E(\boldsymbol{\delta}) = \mathbf{0}$ , and also that there is spatial correlation in  $\boldsymbol{\delta}$  that can ...”
20. line 181 (just before equation 3): “... the  $i, j$ th element of the matrix  $\text{cov}(\boldsymbol{\delta})$  is ...”
21. line 186: remove “however”
22. line 192: “... meaning that they do not rely ...”
23. lines 195-198: “... model-based approaches. Other methods, such as k-nearest-neighbors (Fix and Hodges, 1951; Ver Hoef and Temesgen, 2013), random forest (Breiman, 2001), and Bayesian models (Chan-Golston et al., 2020), could also ...”
24. lines 199-202: “We choose to use FPBK because it is faster than a Bayesian approach and it was developed with theoretically-based variance estimators of means and totals for spatial data, whereas random forests and k-nearest-neighbors use ad hoc variance estimators in most cases (Ver Hoef and Temesgen, 2013). Additionally, FPBK outperformed the other methods in most scenarios.”
25. lines 204-205 and Table 1. I think we need to be clearer on each of these labels, and say exactly what we mean. For example, is IRS-Design using the classical simple random sampling estimators for mean and variance, or the local variance one? We only say on lines 157-158 that the local variance estimator does not rely on  $\pi_{ij}$ , and it is not clear whether there are model assumptions in here or is it completely inclusion-probability based? I assume that “model” always means FPBK, but less certain what “design” means.  
  
Oh, I see that we explain it in lines 221 - 227. Table 1 should be introduced here, and come after Table 2.
26. line 231-232: “... estimating covariance parameters with ...”
27. following line 232: While we have defined our term EFF, Figure 2 strikes me as a bit backward. The y-axis, “efficiency,” makes one think that larger is better (more efficient). We could label the axis relative RMS(P)E instead, where the reader has

come to expect that smaller is better. I do like having approach/IRS-design, as that generally keeps things bounded between 0 and 1. But if we think of efficiency, then something gets more “efficient” as  $1/\text{rMS(P)E}$  grows, which would flip IRS-Design into the numerator. I could go either way: 1) re-label y-axis, or 2) flip  $\text{rMS(P)E}$  but then most points other than IRS-Design will be  $> 1$ .

28. line 233: no indentation
29. line 235: “... (Figure 2, top row), ...”
30. line 236: How are the assumptions of the model-based approaches violated? The estimated autocorrelation can be zero, yielding spatial independence.
31. line 239: “... correlation (Figure 2, bottom row) ...”
32. line 241: I think we should label subfigures with (a), (b), ..., (f). “Bottom-right facet)” is awkward.
33. lines 241-244: This is incorrect. In the bottom-right facet, Model IRS and Design GRTS are almost identical. Throughout the plots, Model IRS is only slightly worse than Design GRTS, showing that the model largely mitigates the poor design properties of IRS.
34. lines 253-255: Change in tense from past to present is awkward.
35. lines 255-258: I think the Central Limit Theorem works in both cases.
36. line 259: Not sure what is meant by “random location layout?”
37. line 262: This sentence contradicts the previous one (last phrase in lines 260-261).
38. line 263: Missing period
39. line 263: I am not sure what is meant that the intervals are based on the normal distribution. The 95% coverage depends on 1) the validity (accuracy) of the variance estimates, and 2) the rate of convergence of the estimator to a normal distribution.
40. line 264-267: IRS-Design has an unbiased variance estimator, so, building on the comments above, the reason for lack of coverage is slow convergence of the Central Limit Theorem due to underlying skewness in the (fixed) population. For the other models, it is less clear about overall variance estimation, but we will have a slower convergence of Central Limit Theorem, no matter what. Also, I don’t really see that interval coverage improved with increasing  $\phi$ , but it clearly improves with increasing sample size.
41. line 280: Not sure what is meant by “to obtain an interval”? Don’t we want to estimate a mean, along with a confidence/prediction interval?

42. line 285: Figure 4 is a map, and doesn't really show skew. We would probably need a histogram for that.
43. line 290: It would be nice to give the reader some idea of the estimated autocorrelation. Something like, "autocorrelation using FPBK with REML was estimated to be 0.9 at  $x$  distance, dropped to 0.5 at  $y$  distance, and decayed to 0.05 by  $z$  distance."
44. line 291: "... of a single IRS sample, ..."
45. line 292: Is the model-based analysis for the same IRS sample, or did we draw another one?
46. line 293: Same question: Was a new GRTS sample drawn for model-based analysis?
47. line 295: "Realization" is most often used for a process, but I suppose it could also be used for a sample. Still, it is not clear whether we are speaking of mercury concentrations as a realization of its generating process, or a realization of our sampling designs.
48. line 307: "... by which samples are ..."
49. line 310-311: "... the realized values of a random process. ..."
50. line 315-316: This is not so clear, because our simulations showed that if IRS sampling was used, the model-based decision to analyze with a spatial model mitigates most of the inefficiency of the IRS design. This is a realistic scenario, as often the statistical analysis is conducted later and independently of how the sample was drawn.
51. lines 315-323: Again, I am not a big fan of the long list separated by semicolons. I would turn them into sentences, or create a bulleted list.
52. line 323: Again, I am not convinced that we have demonstrated clear that coverage improves with increased strength of correlation (although I have not looked at all simulation results in the Appendix). If there is additional support in the Appendix, then we should say so.
53. lines 336-337: Add cross-validation to the list.
54. line 343: "heavily" seems like the wrong word. Maybe "seriously".
55. lines 354-436: First letters of the main words for books (and journals) should be capital. Only first word of journal article should have first letter capitalized. The publisher for a book should have a location, and number of pages for books included. These rules are inconsistent throughout the References.
56. line 368: Brus reference looks incomplete.

- 57. line 389: Fix reference looks incomplete.
- 58. line 418: Is the article really a single page?
- 59. line 431: I think it is now PLOS ONE (all caps)