Bayesian Analysis (2015) **10**, Number 3, pp. 749–752

Comment on Article by Ferreira and Gamerman[*]

James V. Zidek[†]

This paper concerns a very topical issue, namely the effect of preferential sampling the locations at which to measure a spatial process. The topic was highlighted at and studied by a research group at the Statistical and Applied Mathematical Sciences Institute (SAMSI) during its 2009–10 thematic year on spatial statistics, and a number papers came out of that initiative.

To put this paper in context, some background seems worthwhile. Selection bias in one form or another has always been an issue in statistical science, and it has been studied since at least the time when Horvitz and Thompson proposed their simple but ingenious approach to unbiasing estimates of finite population averages when sample items are preferentially selected (Horvitz and Thompson, 1952). Survey statisticians have long since recognized the adverse effect of such bias and the need to adjust for it when computing their estimates. Biostatisticians have also been concerned with this issue in the form of response biased sampling in estimating the relationship between a response Y and a covariate vector Z when instead of sites human subjects are the units (Scott and Wild, 2011). There, inter-subject dependence is ignored due to its complexity and the work of Liang and Zeger (1986) which allows that simplification to be made. The responses Y are assumed to be observed (although that assumption can be relaxed by modeling it) and subject selection is biased by these responses. In contrast, the present paper follows Diggle et al. (2010) and assumes instead that the role of Y is implicit and seen through the point process model that "knows" $Y = Y(x)$, or rather the latent process that generates it, through the intensity function $\exp\{\alpha + \beta S(x)\}$, quite a strong assumption. The just cited work in biostatistics would be of potential relevance in spatial regression, an important topic in environmental epidemiology, but the effect of preferential sampling in that domain, especially on the effect on optimal design as seen in this paper, has not been studied as far as we know.

In geostatistics, spatial dependence can often be of central importance especially when spatial prediction is of primary interest. The paper by Diggle et al. (2010) has awakened interest in a topic that has been conveniently ignored even by those charged with setting regulatory standards—where sites may be deliberately sited to detect the non-compliers and placed where response levels are expected to be high (Guttorp and Sampson, 2010).

The present paper shares with Diggle and Ribeiro (2007), Pati et al. (2011) and Gelfand et al. (2012) the goal of determining the effect of preferential sampling on statistical inference, specifically spatial prediction and parameter estimation. The more

[*]Main article DOI: 10.1214/15-BA944.

[†]Department of Statistics, University of British Columbia, 2207 Main Mall, Vancouver, BC, Canada V6T 1Z4, jim@stat.ubc.ca

 DOI: 10.1214/15-BA944C