A Comparison of Design-Based and Model-Based Approaches for Spatial Data

(in alphabetical order) Michael Dumelle and Matthew Higham and Others

Overview

Why This Paper?

We believe the distinction between these approaches is often misunderstood and there are several ways we could enchance the literature surrounding this topic:

1. Spatial Design-Based vs Spatial Model-Based: There are no comparisons in the literature between spatial model-based approaches and *spatially balanced* design approaches. From what we have seen, these comparisons are between spatial model-based approaches and independent random sample designs. While important to study this behavior, this comparison is no longer fair nor modern. Spatially balanced sampling has exploded in popularity throughout the last decade, and the design-based vs model-based literature needs to reflect this trend.

2. A "fair" comparison: We feel that literature in this area has considered scenarios that are more well-suited for the model-based scenario. For example, if you simulate a Gaussian erorr with an exponential covariance and then then compare design-based estimates to model-based estimates assuming an exponential covariance, of course the model-based approach will outperform design-based approaches. And these are the types of comparisons in the literature, which find model-based approaches generally yield more precise variance estiamtes. A challenge lies in creating a comparison scenario that is reasonable and intuitive. One this to consider would be exploring the comparison after estimating a misspecified covariance function using model-based approaches.

3. Finite AND infinite populations: Literature in the area focuses specifically on finite populations or infinite populations; we want to discuss both in detail.

4. Pragmatic Focus: We see papers in this area tend to be fairly technical. We want the focus to be less on details, more on discussing the pragmatic questions practitioners

26   will be faced with. For example, a thorough discussion of benefits and drawbacks of

27   each method written for practicioners is warranted.

28   5. Provide reliable software

Initial Literature

30   • Design-Based Overview (Särndal et al., 2003; Lohr, 2009)

31   • Model-Based Overview (Cressie, 2015; Schabenberger and Gotway, 2017)

32   • Design-Based and Model-Based Comparisons (Hansen et al., 1983; Brus and De Grui-

33   jter, 1997; Ver Hoef, 2002; Cooper, 2006; Sterba, 2009; Brus, 2020; Chan-Golston et al.,

34   2020)

35   • Spatially Balanced Design and Analysis (Stevens Jr and Olsen, 2003, 2004)

36   • Finite Population Block Kriging (Ver Hoef, 2002, 2008; Higham et al., 2020)

Potential Journals

38   • Ecological Applications

39   • Methods in Ecology and Evolution

40   • Journal of Applied Ecology

41   • Environmetrics

42   • Environmental and Ecological Statistics

OUTLINE

1. INTRODUCTION

2. BACKGROUND

3. NUMERICAL ANALYSIS

3.1 Simulation-Based

We would like to keep this section manageable. Perhaps we start with the following examples

- simulate via correct model (model outperforms sampling)

- simulate via slightly misspecified model (model still outperforms sampling)

- simulate via very misspecified model (sampling outperforms model)

- simulate via extremely misspecified model (e.g. counts with lots of zeroes and a lot of overdispersion) (neither does well)

3.2 Data-Based

3.3 Software

4. DISCUSSION

References

Brus, D. and De Gruijter, J. (1997). Random sampling or geostatistical modelling? choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, 80(1-2):1–44.

Brus, D. J. (2020). Statistical approaches for spatial sample survey: Persistent misconceptions and new developments. *European Journal of Soil Science*.

Chan-Golston, A. M., Banerjee, S., and Handcock, M. S. (2020). Bayesian inference for finite populations under spatial process settings. *Environmetrics*, 31(3):e2606.

Cooper, C. (2006). Sampling and variance estimation on continuous domains. *Environmetrics: The official journal of the International Environmetrics Society*, 17(6):539–553.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78(384):776–793.

Higham, M., Ver Hoef, J., Madsen, L., and Aderman, A. (2020). Adjusting a finite population block kriging estimator for imperfect detection. *Environmetrics*, page e2654.

Lohr, S. L. (2009). *Sampling: design and analysis*. Nelson Education.

Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.

Schabenberger, O. and Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. CRC press.

Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration. *Multivariate behavioral research*, 44(6):711–740.

Stevens Jr, D. L. and Olsen, A. R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, 14(6):593–610.

Stevens Jr, D. L. and Olsen, A. R. (2004). Spatially balanced sampling of natural resources.

85   *Journal of the american Statistical association*, 99(465):262–278.

86   Ver Hoef, J. (2002). Sampling and geostatistics for spatial data. *Ecoscience*, 9(2):152–161.

87   Ver Hoef, J. M. (2008). Spatial methods for plot-based sampling of wildlife populations.

88   *Environmental and Ecological Statistics*, 15(1):3–13.