Spatial Statistics 21 (2017) 226–240

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

A study of variance estimation methods for systematic spatial sampling

Geir-Harald Strand

Survey and statistics division, NIBIO, P.O.Box 115, N-1431 Ås, Norway

HIGHLIGHTS

- Three estimators for variance in systematic spatial samples were compared.
- A correction factor based on the autocorrelation often underestimated the variance.
- A local stratified estimator and a model-based prediction both gave good estimates.

ARTICLE INFO

Article history:
Received 6 April 2017
Accepted 26 June 2017
Available online 8 July 2017

Keywords:
Spatial sampling
Systematic sample
Uncertainty
Variance
Land cover
Area frame survey

ABSTRACT

An undesirable property of systematic spatial sampling is that there is no known method allowing unbiased estimation of the uncertainty of statistical estimates from these surveys. A number of alternative variance estimation methods have been tested and reported by various authors. Studies comparing these estimators are inconclusive, partly because the studies compare different sets of estimators. In this paper, three estimators recommended in recent studies are compared using a single test dataset with known properties.

The first estimator compared in this study (ST4) is based on post-stratification of the data. The second estimator (V08) is using a predetermined correction factor calculated from the spatial autocorrelation. The third estimator (MB) is a model based prediction calculated using values from the semivariogram. MB and ST4 were both found to be fairly accurate, while V08 consistently underestimated the variance in this study. V08 relies on the assumption that the autocorrelation structure in the dataset can be described using a particular exponential function. The most likely explanation of the weak result for V08 is that this assumption is violated by the empirical data used in the experiment. A better correction factor can be calculated, but the safe approach is to use MB or ST4.

E-mail address: ghs@nibio.no.

http://dx.doi.org/10.1016/j.spasta.2017.06.008

1. Introduction

Spatial sampling is a cost-efficient way to conduct surveys for ecological and environmental studies and monitoring projects. Various sampling techniques are used when the study area is restricted in size, but national and other wide ranging surveys will often rely on systematic sampling. Many examples can be found. National forest inventories are habitually carried out using field-based systematic sampling surveys (Tomter et al., 2010; Tomppo and Tuomainen, 2010). Systematic spatial sampling provides the basis for landscape monitoring programs in Norway (Dramstad et al., 2002) and Sweden (Ståhl et al., 2011) and is used in land cover and land use surveys on a national (Strand, 2013; Aune-Lundberg and Strand, 2017) as well as a continental scale (Eurostat, 2003; Martino and Fritz, 2008). Many soil surveys also employ systematic sampling (Morvan et al., 2008).

Systematic spatial sampling is a sampling strategy with a number of favorable properties (Wang et al., 2012). It is easy to implement and there is no risk of finding sample units clustered in a few regions while other regions are left with few or no samples. In order to draw the systematic spatial sample, the population of locations must be organized as a regular frame. A starting point is drawn randomly and the rest of the sample is collected at regular intervals from this starting point. The systematic sample will result in more precise estimates than a simple random sample, in the spatial context and under commonly occurring conditions, because the sampling units are distributed more evenly across the sampled area (Bellhouse and Sutradhar, 1988; Dunn and Harrison, 1993; D'Orazio, 2003; Ambrosio et al., 2004).

The systematic sample is in particular preferable as a sampling method when nearby sampling units show a high degree of positive correlation (Cochran, 1977). This was demonstrated by Flores et al. (2003) who compared the relative efficiency of systematic sampling to simple random sampling from populations with known properties. The study demonstrated that systematic sampling was more efficient than simple random sampling and showed that the improvement in efficiency was related to sampling distance. The relative efficiency of systematic sampling was higher when the sampling distance was short and lower when the sampling distance increased. The change in relative efficiency was closely related to the spatial autocorrelation.

An undesirable property of systematic sampling is that there is no known method allowing unbiased estimation of the uncertainty in these surveys. The higher precision achieved by systematic sampling may therefore go unnoticed. The reason for this shortcoming is found in the systematic sample design, where the population – at least in theory – is divided into a number of partitions. Each partition consists of the population elements included in the sample when a particular starting point is selected. There is a finite set of starting points representing a finite set of partitions (Madow and Madow, 1944). Each and every population element is assigned to one (and only one) partition. When a partition is included in the sample, then every population element in this partition is included (Thompson, 2002 pp. 129–131). A simple example is illustrated in Fig. 1 where a population of grid cells is divided into four partitions labeled A, B, C and D.

Systematic sampling is (usually) limited to drawing a single partition by choosing a single starting point. This is equivalent to a sample size of $n = 1$ partitions. Ordinary variance estimation methods require a denominator of $n - 1$ and can therefore not be applied (Thompson, 2002).

The conservative approach for handling uncertainty in a systematic sample is to calculate the variance using the estimators intended for simple random sampling (Milne, 1959; Cochran, 1977; Wolter, 1984, 2007). This is usually a safe approach and will in certain situations be both acceptable and commendable, but has a tendency to overestimate the variance (McRoberts et al., 2016). A large number of alternative, more or less biased, variance estimation methods have been tested and reported by various authors (Matèrn, 1947, 1960; Wolter, 2007; Gallego and Delincé, 2010; Aubry and Debouzie, 2000; Dunn and Harrison, 1993; D'Orazio, 2003; Opsomer et al., 2012).