



Spatially balanced sampling designs for environmental surveys

Claire Kermorvant  · Frank D'Amico  ·
Noëlle Bru · Nathalie Caill-Milly  ·
Blair Robertson 

Received: 5 February 2019 / Accepted: 15 July 2019 / Published online: 30 July 2019
© Springer Nature Switzerland AG 2019

Abstract Some environmental studies use non-probabilistic sampling designs to draw samples from spatially distributed populations. Unfortunately, these samples can be difficult to analyse statistically and can give biased estimates of population characteristics. Spatially balanced sampling designs are probabilistic designs that spread the sampling effort evenly over the resource. These designs are particularly useful for environmental sampling because they produce good-sample coverage over the resource, they have precise design-based estimators and they can potentially reduce the sampling cost. The most popular spatially balanced design is Generalized Random Tessellation Stratified (GRTS), which has many desirable features including a spatially balanced sample, design-based estimators and the ability to select spatially balanced oversamples. This article considers the popularity of spatially balanced sampling, reviews several spatially

balanced sampling designs and shows how these designs can be implemented in the statistical programming language R. We hope to increase the visibility of spatially balanced sampling and encourage environmental scientists to use these designs.

Keywords BAS · GRTS · LPM · Probabilistic sampling · Spatially balanced

When sampling an environmental resource it is important to randomly choose the sampling locations (Conn et al. 2016) over the study area to provide formal statistical inference from the sample to the population. Smith et al. (2017) established that in ecology 12% of field studies selected samples using simple random sampling (SRS) and 9% used systematic sampling. These methods are probabilistic sampling designs (meaning there is an element of randomness in selecting their samples) and have well established statistical properties (MacKenzie 2006; Sica 2006; Stehman 2009). However most of the ecological studies in Smith et al.'s (2017) review were not probabilistic sampling designs. Some studies used haphazard or subjective judgement sampling methods and some studies did not specify how their samples were drawn (Smith et al. 2017). This is troubling because data gathered in a haphazard or subjective way can produce unrepresentative samples and biased estimates of population characteristics (Albert et al. 2010; Levy and Lemeshow 2013)

Choosing an appropriate sampling design for a particular study can be difficult and there is no *best* design for all research questions (Kenkel et al. 1990; Stehman

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10661-019-7666-y>) contains supplementary material, which is available to authorized users.

C. Kermorvant (✉) · F. D'Amico · N. Bru
Laboratoire de Mathématiques et de leurs Applications de Pau –
MIRA, CNRS/Univ Pau & Pays Adour/E2S UPPA, UMR 5142,
64600 Anglet, France
e-mail: claire.kermorvant@univ-pau.fr

N. Caill-Milly
Ifremer - Laboratoire Environnement Ressources d'Arcachon, 1
Allée du Parc Montaury, 64600 Anglet, France

B. Robertson
School of Mathematics and Statistics, University of Canterbury,
Christchurch, New Zealand

and Overton 1994). This choice depends on many things including the study objectives, available sampling frames and known auxiliary variables. This paper focuses on making an inference from a sample to the entire population using a specific class of probabilistic sampling designs called spatially balanced sampling designs. These designs were chosen because they are particularly useful for sampling natural resources (Stevens and Olsen 2004). For a full treatment on the subject, the reader is referred to Benedetti et al. (2015).

What is spatially balanced sampling?

To achieve good estimates of population characteristics, the spatial pattern of the sample should be similar to the spatial pattern of the population. However, the spatial pattern of the response variable may not be known before the sample is drawn. Fortunately, a common spatial feature in environmental sampling is that nearby locations tend to be more similar because they interact with one another and are influenced by the same set of factors (Stevens and Olsen 2004). Therefore, an effective strategy is to spatially spread the sample evenly over the resource. A sample that is evenly spread over the resource is called a spatially balanced sample. Stevens and Olsen (2004) introduced the phrase *spatially balanced sampling* and proposed a statistic that measures the spatial balance or regularity of a sample using Voronoi polygons.

Why should environmental scientists use spatially balanced designs?

The potential advantages of using spatially balanced sampling have been demonstrated in the field of environmental science (Stevens and Olsen 2004; Christianson and Kaufman 2016; McGarvey et al. 2016). The first advantage is that spatially balanced samples are evenly spread over the resource. Covering the resource avoids under-coverage and over-coverage, which can happen with probabilistic sampling designs with poor spatial balance (Stevens and Olsen 2004; Christianson and Kaufman 2016). If a researcher's analysis requires clusters of nearby observations, spatially balanced cluster sampling could be useful (Robertson et al. 2017).

Spatially balanced samples can be very efficient when the response variable has a strong spatial trend (Stevens

and Olsen 2004; Barabesi and Franceschi 2011; Grafström and Lundström 2013; Robertson et al. 2013; Benedetti et al. 2017), because their design-based estimators take into account spatial heterogeneity (Wang et al. 2012) and spatial auto-correlation (Haining 2003). When estimating a population total or mean using the Horvitz-Thompson estimator, the local mean variance estimator (Stevens and Olsen 2003) is a popular variance estimator. There have been many studies showing the effectiveness of spatially balanced sampling with this estimator on a variety of populations with different spatial structures (c.f. Stevens and Olsen 2004; Grafström et al. 2012; Grafström and Lundström 2013; Robertson et al. 2013, 2018; Grafström & Matei 2018). If the spatial trend is weak or if there is no trend at all, there is no statistical advantage in choosing spatially balanced designs over other probabilistic designs (Robertson et al. 2013).

Another potential advantage of spatially balanced sampling is reduced sample cost. As mentioned above, spatially balanced designs can produce precise design-based estimators when there is a spatial trend in the response variable. Hence, to achieve a desired level of precision, fewer observations may be required when spatially balanced sampling is used. This was illustrated by Kermorvant et al. (2017, 2019) for a clam monitoring program in Arcachon Bay, France. They showed that when spatially balanced designs were used, the total survey cost was reduced by 30% when compared with simple random sampling.

Another useful feature of some spatially balanced designs is that they can draw spatially balanced oversamples (replacement units). This is particularly useful for environmental sampling because when sample units cannot be observed (private property, inaccessible, too dangerous, etc.), replacement units are often required to achieve the desired sample size (Stevens and Olsen 2004; Robertson et al. 2018; Theobald et al. 2007). For example, in the Oklahoma statewide stream and river monitoring program, only 130 of the 177 randomly chosen sites could be observed (Oklahoma Water Resources Board 2013). Of the unobservable sites, eight were on private land and 39 had dry channels or were not accessible. To achieve the desired sample size, the researchers selected replacement sites from an oversample drawn using simple random sampling. The potential advantage of using spatially balanced oversamples is that the observed sample maintains some degree of spatial balance over the observable resource (Stevens and Olsen 2004; Robertson et al. 2018). Although oversampling is of

practical importance, it does not eliminate the non-response from unobservable units or the bias of an inference (Robertson et al. 2018).

Generalized Random Tessellation Stratified

Generalized Random Tessellation Stratified (GRTS) (Stevens and Olsen 2004) is the most popular spatially balanced sampling design for environmental studies (Grafström and Tillé 2013; Foster 2016). It was developed by the U.S. Environmental Protection Agency (EPA) for the National Environmental Monitoring and Assessment Program (Messer et al. 1991; Stevens and Olsen 2004). The GRTS uses a complex algorithm to draw its sample, and we briefly discuss its main steps here. The reader is referred to Olsen et al. (2012) for a full, non-technical description of the GRTS. Initially, a grid is superimposed over the study area and each grid cell is hierarchically numbered using a base four numbering system. The numbered grid cells are then randomly permuted using reverse hierarchical ordering and mapped (in order) to the real line. A systematic sample from the ordered grid cells is then drawn, and one sampling unit is randomly selected from each of these grid cells. The selected units are then mapped back to their respective locations in the study area, to yield the sample locations. The systematic sampling and hierarchical ordering that the GRTS uses ensures that the sample is spatially balanced (Stevens and Olsen 2004).

To investigate GRTS's popularity, the Google Scholar search engine was used because it provided access to a wide range of publication types. It indexes journal papers (published online or in paper format), conference proceedings, posters and technical reports from research organizations in both the public and private sectors. The keyword "GRTS" was used for the search, and we did not include citations from 2018. All the documents found were categorized as either "publications" or "reports". Publications included peer-reviewed journal articles and refereed conference proceedings, and reports included all other publication types. Results are displayed in Fig. 1.

Our analysis found 600 documents citing the GRTS throughout the world. The citation and/or use of the GRTS showed a steady increase until 2013, after which it flattened out. At the beginning, there were more reports than publications, but this trend

appears to have reversed since 2014. Most of the documents found were in the fields of environmental science (mostly ecology but also environmental chemistry) and statistics (new designs, tests and comparisons). There were only two publications from other fields. The first was in economics, where the GRTS was compared with existing sampling designs for business surveys (Dickson et al. 2014) and the second was a thesis on sampling standards for maintenance management quality assurance (Liu and Chen 2018).

Other spatially balanced sampling designs

Many spatially balanced designs have been proposed in the literature. In this section, we mention several approaches.

The Local Pivotal Method (LPM) (Grafström et al. 2012) is a flexible spatially balanced design that can draw equal and unequal probability samples in multiple dimensions. Unequal probability sampling can be more efficient than equal probability sampling if there is a positive correlation between the inclusion probabilities and the response values (Robertson et al. 2013). Additional dimensions could include auxiliary information such as ecological threats, time intervals, species population structure or environmental data (Brown et al. 2015). The Swedish national forest inventory, for example, has implemented the LPM with five auxiliary variables (Grafström et al. 2017). The LPM is a popular method with 100 citations (using Google Scholar), where most of its applications were related to forestry.

Grafström (2012) also presented spatially correlated Poisson sampling (SCPS). This design is a modification of correlated Poisson sampling (Bondesson and Thorburn 2008) that draws spatially balanced samples. The LPM is algorithmically easier than the SCPS, but the SCPS may produce better results for some populations (Grafström and Schelin 2014).

Balanced acceptance sampling (BAS) (Robertson et al. 2013, 2017) is another spatially balanced design. The BAS uses the Halton sequence (Halton 1960) to spread its sample across multiple dimensions. The BAS is conceptually simple, computationally efficient and is particularly useful for drawing spatially balanced oversamples (Robertson et al. 2018). We found 34 publications citing the BAS,

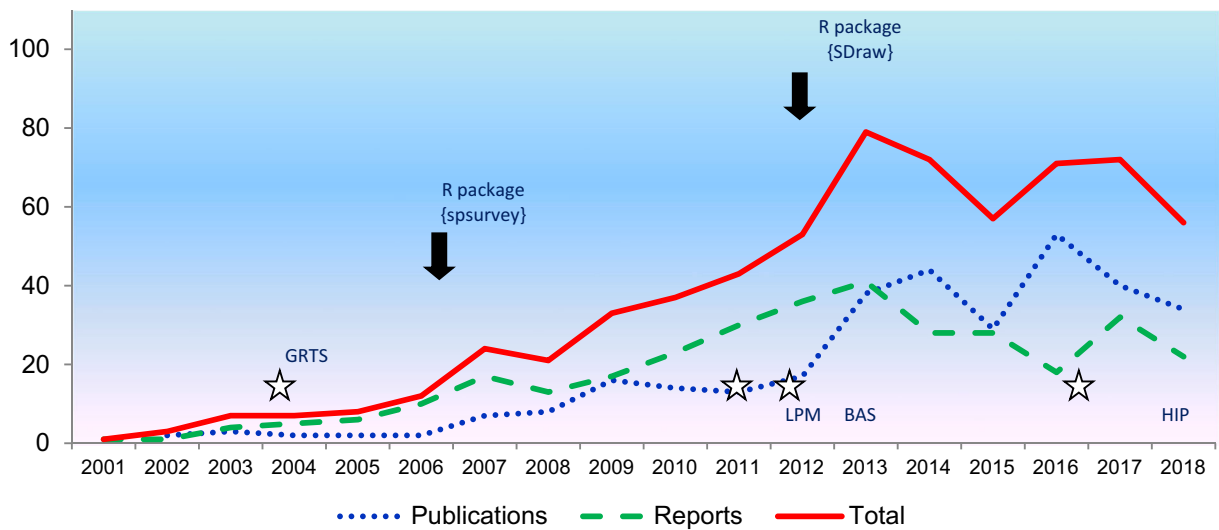


Fig. 1 Flow representation of use and/or citations of GRTS in the literature, publication date (stars) of several spatially balanced designs and R packages (arrows) are shown

where most of the papers were methodological rather than applied. The BAS has been used to survey bats in Bighorn Canyon National Recreation Area (Keinath and NRA 2016) and is being used for New Zealand's national monitoring program (van Dam-Bates et al. 2018). The BAS is well suited for areal resources (geographic areas), but it can be inefficient on some point resources (Robertson et al. 2018). To improve the performance of the BAS on point resources, Robertson et al. (2018) presented Halton iterative partitioning (HIP). This spatially balanced design uses properties of the Halton sequence to partition a point resource into nested boxes to draw its sample, rather than using the sequence itself.

Benedetti and Piersimoni (2017) presented a flexible class of spatially balanced designs that draw their samples based on a within-sample distance (Benedetti and Piersimoni 2017). The algorithm is simple to implement in multiple dimensions, and any distance or similarity measure can be used to define the within sample distance.

Spatially balanced sampling packages in R

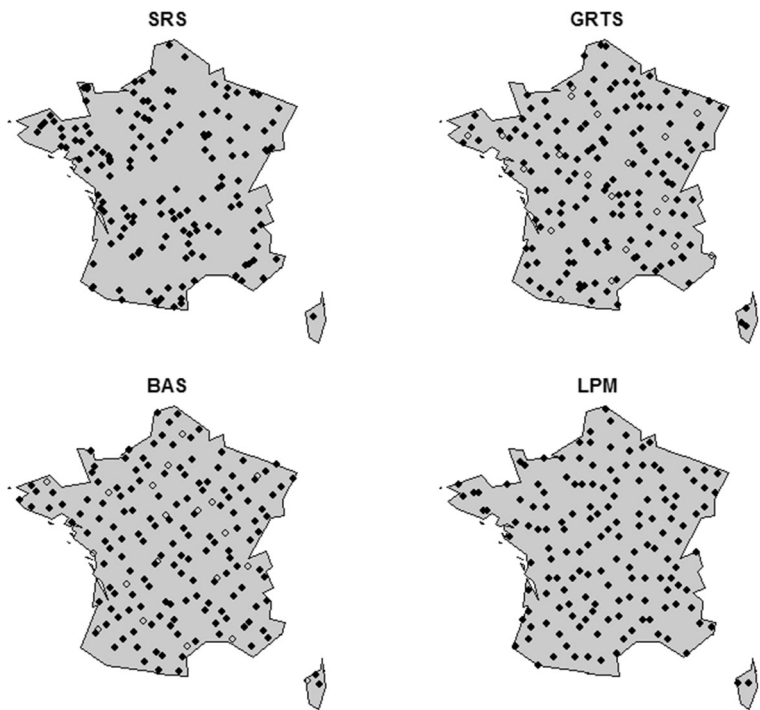
Several R software (R Core Team 2014) packages are freely available to draw spatially balanced samples. To draw the GRTS samples, `spsurvey` (Kincaid and Olsen 2015) or `SDraw` (McDonald 2016) can be

used. These packages can draw samples from point resources (geographic locations), linear resources (rivers) and areal resources (geographic areas), and can also draw spatially balanced oversamples. The `spsurvey` package can also draw stratified spatially balanced GRTS samples with user-defined strata.

The other spatially balanced designs mentioned in this article can be selected using the following packages. `BalancedSampling` (Grafström and Lisic 2016) draws equal and unequal probability LPM and SCPS samples from point resources. The BAS and HIP samples/oversamples can be selected from point, linear and areal resources using `SDraw` (McDonald 2016). Historical or legacy sites can also be incorporated into a BAS design (Foster et al. 2014) using the `MBHdesign` package (Foster 2016). Finally, the R package `Spbsampling` (Pantalone et al. 2019) can be used to draw the within-sample distance-based methods of Benedetti and Piersimoni (2017).

Figure 2 shows examples of equal probability spatially balanced samples of 150 points drawn from a point resource using the `SDraw` and `BalancedSampling`. An oversample of 20 points is also illustrated for the BAS and GRTS. Note how the oversample points are spatially balanced with respect to the primary sample. To illustrate the R syntax for these packages, an annotated R script that creates Fig. 2 is given in the [supplementary material](#) section.

Fig. 2 Several spatially balanced samples drawn using different designs, where open symbols denote oversamples sites



Conclusion

Environmental scientists are beginning to use more advanced sampling designs to achieve robust statistical results. Spatially balanced designs are particularly useful for environmental science because they can produce good-sample coverage over a resource, precise design-based estimators and potentially reduce sampling cost. The GRTS is the most popular spatially balanced sampling design, and it is easy to implement using freely available R packages like the *spsurvey* and *SDraw*. Another useful feature of the GRTS is that spatially balanced oversamples can be drawn. Although oversampling is of practical importance, it does not eliminate the non-response from unobservable units or the bias of an inference. Several other spatially balanced designs are also available, each with an accompanying R package. The LPM and SCPS samples can be drawn using the *BalancedSampling*, and the *SDraw* selects the BAS and HIP samples/oversamples. Although spatially balanced sampling has mostly been used in ecology, we encourage all environmental scientists to these designs when the research objective is to make an inference from a sample to the entire population.

Acknowledgements We thank Jennifer Brown, Trent McDonald, Anton Grafström and Roberto Benedetti and anonymous referees for valuable comments that improved this paper.

Funding information This work was supported by “Communauté d’Agglomération Pays Basque – Euskal Hirigune Elkargoa” through a thesis grant.

References

Albert, C. H., Yoccoz, N. G., Edwards, T. C., et al. (2010). Sampling in ecology and evolution—bridging the gap between theory and practice. *Ecography*, *33*, 1028–1037.

Barabesi, L., & Franceschi, S. (2011). Sampling properties of spatial total estimators under tessellation stratified designs. *Environmetrics*, *22*, 271–278.

Benedetti, R., & Piersimoni, F. (2017). A spatially balanced design with probability function proportional to the within sample distance. *Biometrical Journal*, *59*, 1067–1084.

Benedetti, R., Piersimoni, F., & Postiglione, P. (2015). *Sampling spatial units for agricultural surveys*. Springer.

Benedetti, R., Piersimoni, F., & Postiglione, P. (2017). Spatially balanced sampling: a review and a reappraisal. *International Statistical Review*, *85*, 439–454.

- Bondesson, L., & Thorburn, D. (2008). A list sequential sampling method suitable for real-time sampling. *Scandinavian Journal of Statistics*, 35, 466–483.
- Brown, J. A., Robertson, B. L., & McDonald, T. (2015). Spatially balanced sampling: application to environmental surveys. *Spatial Statistic Conference*, 27, 6–9. <https://doi.org/10.1016/j.proenv.2015.07.108>.
- Christianson, D. S., & Kaufman, C. G. (2016). Effects of sample design and landscape features on a measure of environmental heterogeneity. *Methods in Ecology and Evolution*, 7, 770–782. <https://doi.org/10.1111/2041-210X.12539>.
- Conn, P. B., Moreland, E. E., Regehr, E. V., Richmond, E. L., Cameron, M. F., & Boveng, P. L. (2016). Using simulation to evaluate wildlife survey designs: polar bears and seals in the Chukchi Sea. *Royal Society Open Science*, 3, 150561. <https://doi.org/10.1098/rsos.150561>.
- Dickson, M. M., Benedetti, R., Giuliani, D., & Espa, G. (2014). The use of spatial sampling designs in business surveys. *Open Journal of Statistics*, 4, 345–354. <https://doi.org/10.4236/ojs.2014.45034>.
- Foster, S. D. (2016). MBHdesign: spatial designs for ecological and environmental surveys. R package version, 1, 61.
- Foster, S. D., Hosack, G. R., Hill, N. A., Barrett, N. S., & Lucieer, V. L. (2014). Choosing between strategies for designing surveys: autonomous underwater vehicles. *Methods in Ecology and Evolution*, 5, 287–297.
- Grafström, A. (2012). Spatially correlated Poisson sampling. *Journal of Statistical Planning and Inference*, 142, 139–147. <https://doi.org/10.1016/j.jspi.2011.07.003>.
- Grafström, A., Lisic, J. (2016). BalancedSampling: balanced and spatially balanced sampling [online]. R package version 1.5.2
- Grafström, A., & Lundström, N. L. (2013). Why well spread probability samples are balanced. *Open Journal of Statistics*, 3, 36–41.
- Grafström, A., & Matei, A. (2018). Spatially balanced sampling of continuous populations. *Scandinavian Journal of Statistics* 45:792–805
- Grafström, A., & Schelin, L. (2014). How to select representative samples. *Scandinavian Journal of Statistics*, 41, 277–290.
- Grafström, A., & Tillé, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*, 24, 120–131.
- Grafström, A., Lundström, N. L., & Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68, 514–520.
- Grafström, A., Zhao, X., Nylander, M., & Petersson, H. (2017). A new sampling strategy for forest inventories applied to the temporary clusters of the Swedish national forest inventory. *Canadian Journal of Forest Research*, 47, 1161–1167.
- Haining, R. P. (2003). *Spatial data analysis: theory and practice*. United Kingdom: Cambridge University Press.
- Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2, 84–90.
- Keinath, D. A., & NRA B.C. (2016). Bat population monitoring of Bighorn Canyon National Recreation Area: 2015 progress report. 3. pp
- Kenkel, N., Juhász-Nagy, P., & Podani, J. (1990). On sampling procedures in population and community ecology. In: *Prog Theor Veg Sci* 83:195–207. <https://doi.org/10.1007/BF00031692>
- Kermorvant, C., Caill-Milly, N., D'Amico, F., et al. (2017). Optimization of a survey using spatially balanced sampling: a single-year application of clam monitoring in the Arcachon Bay (SW France). *Aquatic Living Resources*, 30, 37.
- Kermorvant, C., Caill-Milly, N., Bru, N., & D'Amico, F. (2019). Optimizing cost-efficiency of long term monitoring programs by using spatially balanced sampling designs: the case of manila clams in Arcachon bay. *Ecological Informatics*, 49, 32–39. <https://doi.org/10.1016/j.ecoinf.2018.11.005>.
- Kincaid, T. M., Olsen, A. R. (2015). spsurvey: Spatial Survey Design and Analysis. R package version 3.1. <http://www.epa.gov/nheerl/arm/>.
- Levy, P. S., & Lemeshow, S. (2013). *Sampling of populations: methods and applications*. New Jersey: John Wiley & Sons.
- Liu, X. C., & Chen, Z. (2018). Statistical analysis and sampling standards for maintenance management quality assurance (MMQA). (No. MPC 18-350). Mountain Plains Consortium
- MacKenzie, D. I. (2006). *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press.
- McDonald, T. (2016). SDraw: spatially balanced sample draws for spatial objects. R package version 2.1.8. <https://CRAN.R-project.org/package=SDraw>
- McGarvey, R., Burch, P., & Matthews, J. M. (2016). Precision of systematic and random sampling in clustered populations: habitat patches and aggregating organisms. *Ecological Applications*, 26, 233–248.
- Messer, J. J., Linthurst, R. A., & Overton, W. S. (1991). An EPA program for monitoring ecological status and trends. *Environmental Monitoring and Assessment*, 17, 67–78.
- Oklahoma Water Resources Board. (2013). *The statewide stream/river probabilistic monitoring network-final report. 3800 N* (p. 73118). Oklahoma City: Classen.
- Olsen, A., Kincaid, T., & Payton, Q. (2012). Spatially balanced survey designs for natural resources. In: *Design and analysis of long-term ecological monitoring studies* (pp 126–150). New York: Cambridge University Press.
- Pantalone, Benedetti R., & Piersimoni, F. (2019). Spbsampling: spatially balanced sampling. [online]. R package version 1.2.0
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Robertson, B., Brown, J., McDonald, T., & Jaksons, P. (2013). BAS: balanced acceptance sampling of natural resources. *Biometrics*, 69, 776–784.
- Robertson, B., McDonald, T., Price, C., & Brown, J. (2017). A modification of balanced acceptance sampling. *Statistics & Probability Letters*, 129, 107–112.
- Robertson, B., McDonald, T., Price, C., & Brown, J. (2018). Halton iterative partitioning: spatially balanced sampling via partitioning. *Environmental and Ecological Statistics*, 25:1–19. <https://doi.org/10.1007/s10651-018-0406-6>
- Sica, G. T. (2006). Bias in research studies I. *Radiology*, 238, 780–789.
- Smith, A. N., Anderson, M. J., & Pawley, M. D. (2017). Could ecologists be more random? Straightforward alternatives to haphazard spatial sampling. *Ecography*, 40, 1251–1255. <https://doi.org/10.1111/ecog.02821>.

- Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, 30, 5243–5272.
- Stehman, S. V., & Overton, W. S. (1994). 9 Environmental sampling and monitoring. In: *Handbook of Statistics* (pp 263–306). United Kingdom: Elsevier.
- Stevens, D. L., & Olsen, A. R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, 14, 593–610.
- Stevens, D. L., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, 99, 262–278.
- Theobald, D. M., Stevens, D. L., Jr., White, D., Urquhart, N. S., Olsen, A. R., & Norman, J. B. (2007). Using GIS to generate spatially balanced random survey designs for natural resource applications. *Environmental Management*, 40, 134–146.
- van Dam-Bates, P., Gansell, O., & Robertson, B. (2018). Using balanced acceptance sampling as a master sample for environmental surveys. *Methods in Ecology and Evolution*, 9: 1718–1726. <https://doi.org/10.1111/2041-210X.13003>.
- Wang, J-F, Jiang, C-S, Hu, M-G, et al. (2012). Design-based spatial sampling: Theory and implementation. *Environmental Modelling and Software*, 40, 280–288.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.