[AM-03-026] Spatial Sampling for Spatial Analysis

Abstract

Spatial sampling is a key estimation method in spatial analysis, where known sample points are used to predict values at unknown locations. The goal is to improve prediction accuracy by ensuring high-quality samples. This requires controlling two main factors: the location of the samples and the sample size. Sample locations must be carefully chosen within the study area to balance distribution, cost, efficiency, and prediction accuracy. The sample size should ideally be optimized, as collecting too many samples may not provide additional useful information and can increase costs. The challenge is finding the right balance between sample quantity and quality.

Keywords: spatial statistics

Author & citation

Mwenda, K. and Şalap-Ayça, S. (2025). Spatial Sampling for Spatial Analysis. The Geographic Information Science & Technology Body of Knowledge (2025 Edition). John P. Wilson (Ed.). DOI: 10.22224/gistbok/2025.1.3.

Explanation

1. Definitions
2. Introduction
3. Systematic Sampling
4. Simple Random Sampling
5. Stratified Random Sampling
6. Quasi-Random Sampling
7. Cluster Sampling
8. Adaptive Sampling

1. Definitions

Systematic sampling: Creation of sample points in a uniform, gridded non-random pattern in the study area.

Simple Random sampling: Creation of sample points randomly with each location in study area standing an equal chance of being selected as a sample location.

Stratified Random sampling: Dividing study area into distinct sections then performing simple random sampling in each section.

Quasi-Random sampling: As a low-discrepancy sequence, this sampling algorithm 'remembers' the distribution of previously sampled points, effectively preventing clusters or gaps in the sample space.

Cluster sampling: Creation of cluster centers by first systematically or randomly sampling

locations then grouping them.

Adaptive sampling: Creation of sample points by weighting variable areas over uniform areas, such that more sample locations are created in areas of higher spatial variation.

Spatial prediction: Estimation of values at unknown locations.

Sample size: Number of observations or known locations in a region/study area.

2. Introduction

Spatial sampling is a fundamental estimation technique in spatial analysis. It involves using a sample of known points to estimate values for a variable at unknown locations. The objective is to enhance the accuracy of predictions by improving the quality of the samples. Achieving this requires controlling two key aspects of the sampling process. First, it is important to manage the locations of the samples. While the samples must fall within the study area, their distribution can vary based on several criteria, which will be explored in each sampling method. The decision regarding where to place these sample locations affects both the cost and efficiency of the study, as well as the accuracy of the predictions (Stevens and Olsen, 2004; Theobald et al, 2007). Secondly, selecting the appropriate sample size is crucial. Ideally, one would have unlimited time and resources to collect samples throughout the entire study area. However, in practice, the goal is to determine an optimal number of samples, beyond which collecting additional samples may not provide any more valuable information and could lead to increased costs. For optimal results, Theobald et al. (2007) suggest limiting the number of samples to less than 1 percent of all potential sample locations within the study area.

3. Systematic Sampling

Systematic sampling is one of the most common sampling patterns. It involves the creation of sample points in a uniform, gridded non-random pattern in the study area (Figure 1). Samples are evenly distributed at consistent X and Y intervals, usually appearing as systematically arranged points along parallel lines. Systematic sampling is relatively simple due to its consistent and uniform sampling intensity across locations. For instance, if a researcher aims to collect soil samples from a field, this approach guarantees that all areas within the field are sampled with the same intensity. However, this may not be the most statistically efficient sampling method in cases where the variable of interest varies in importance. For instance, if a researcher needs to collect water samples from a pond within a larger study area that includes land, systematic sampling might allocate some points on land, undermining the purpose of the study.

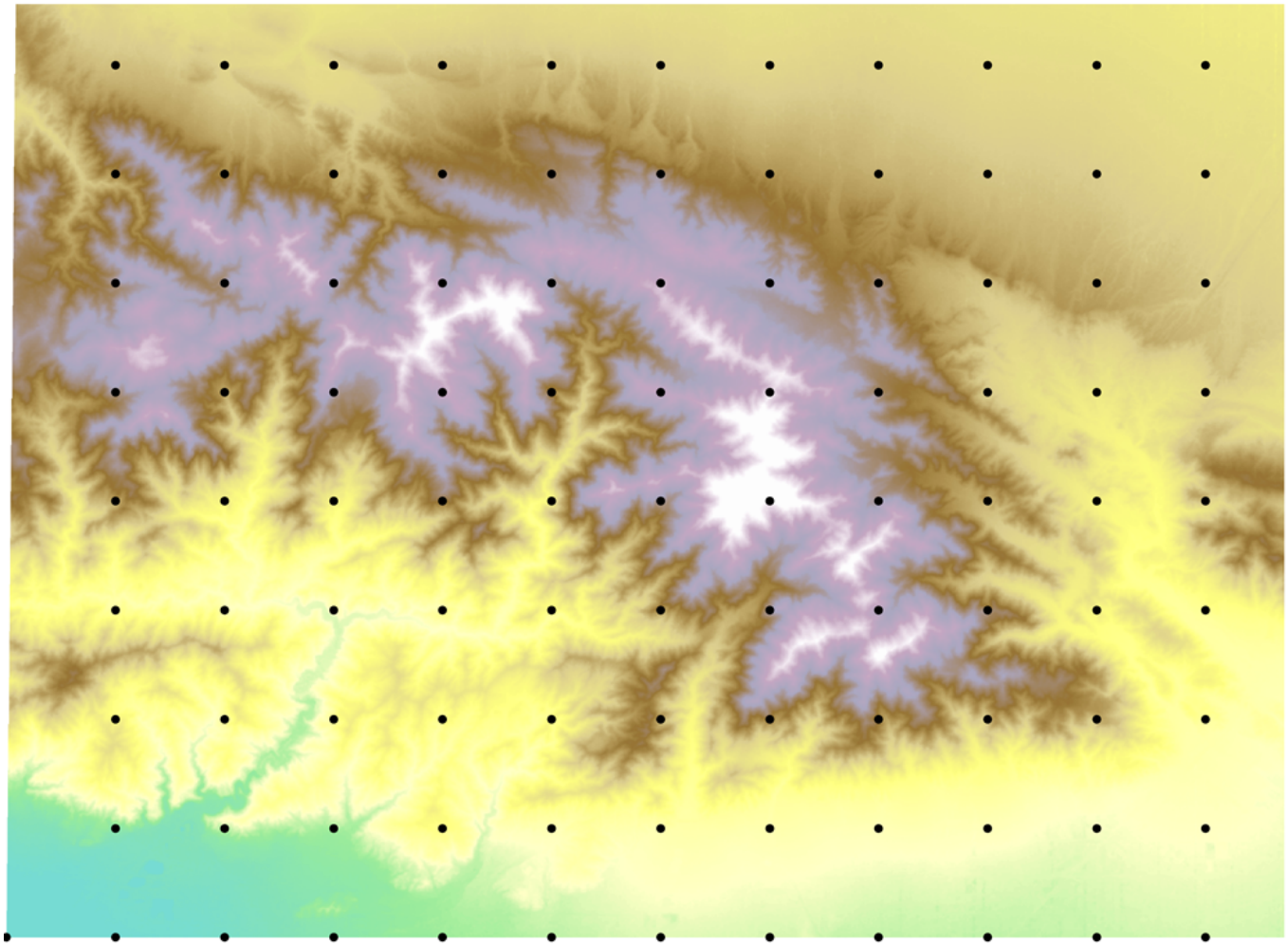


Figure 1. Example of systematic sampling pattern of 100 locations across a study area depicted by the underlying raster image. Source: authors.

4. Simple Random Sampling

Simple random sampling involves generating sample points through independent random processes, ensuring that every location within the study area has an equal probability of being selected (Figure 2). This method addresses some limitations of systematic sampling by minimizing the likelihood of aligning with existing patterns in the study area, thereby reducing bias and improving the accuracy of predictions. However, like systematic sampling, simple random sampling does nothing to distribute samples in areas of high variation or interest. In the example provided for systematic sampling, simple random sampling might lead to collecting more soil samples than necessary from the field while obtaining fewer water samples than required from the pond.

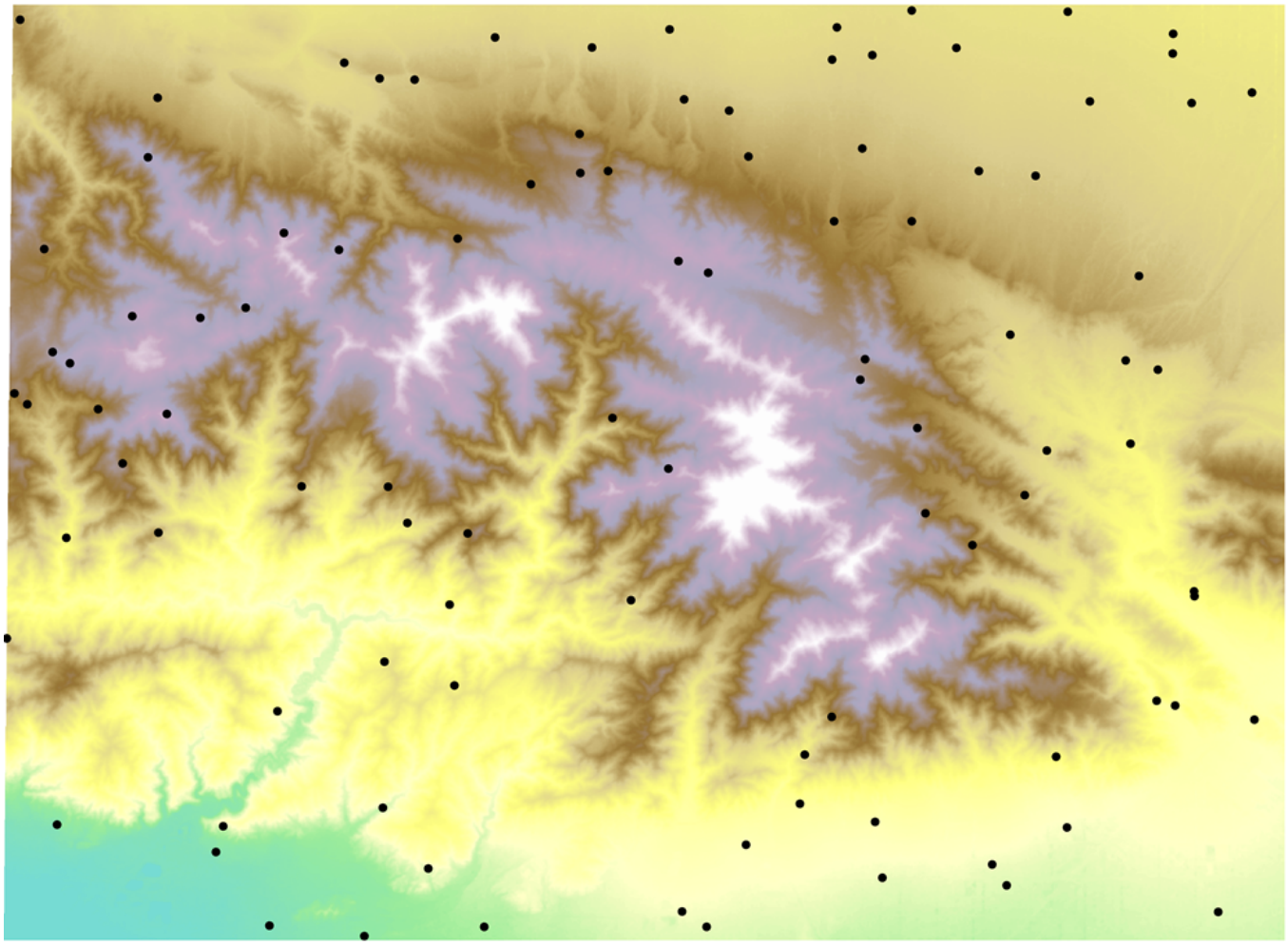


Figure 2. Example of simple random sampling pattern of 100 locations across a study area depicted by the underlying raster image. Source: authors.

5. Stratified Random Sampling

Stratified random sampling is similar to simple random sampling, but it is preceded by dividing the study area into distinct sections, or strata. These sections can be represented as polygons, rasters, or other variable identifiers. In the example provided for systematic sampling, stratified random sampling might entail first sectioning the field into various land use types, then conducting simple random sampling to collect soil samples.

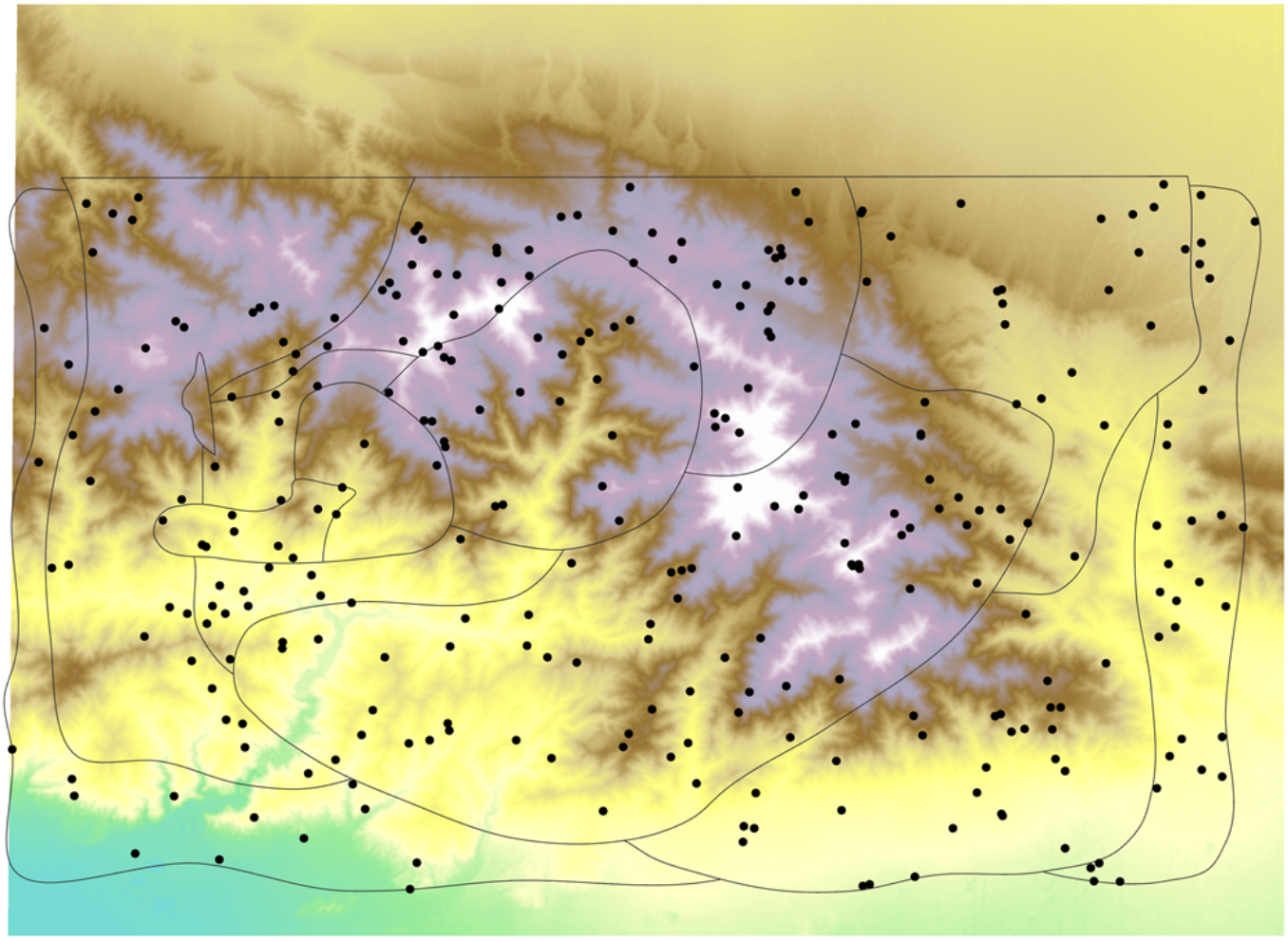


Figure 3. Example of stratified random sampling pattern of 100 locations within each stratum (polygon) across a study area depicted by the underlying raster image. Source: authors.

6. Quasi-Random Sampling

This process involves deterministic methods that are designed to approximate randomness by using a low-discrepancy sequence such as the Sobol sequence (Sobol' et al. 2011). One key advantage of this technique is that the samples are often more evenly spread across the study area than simple random sampling. This technique is commonly used in Monte Carlo simulation or spatial optimization problems when a uniform spread of points is important. In environmental studies, quasi-random sampling can be used to monitor variables like temperature, soil quality, or vegetation, ensuring that all areas of the study site are adequately represented without introducing inefficiencies often associated with simple random methods (Lilburne and Tarantola 2009; Şalap-Ayça et al, 2018; Saltelli et al. 2010). However, the drawback of quasi-random sampling is that it requires a predefined sample size and may not be easily adaptable to dynamic or field-driven sampling scenarios, as seen with adaptive methods. Furthermore, its deterministic nature may limit flexibility in studies where true randomness is a priority or where data collection logistics prevent a uniform distribution of samples.

Quasi-random sampling patterns can be visualized as evenly distributed points that appear

almost grid-like but incorporate subtle randomness. Figure 4 illustrates a quasi-random sampling pattern (right) generated using the Sobol sequence in comparison with simple random sampling (left), highlighting the former's uniformity across the study area.

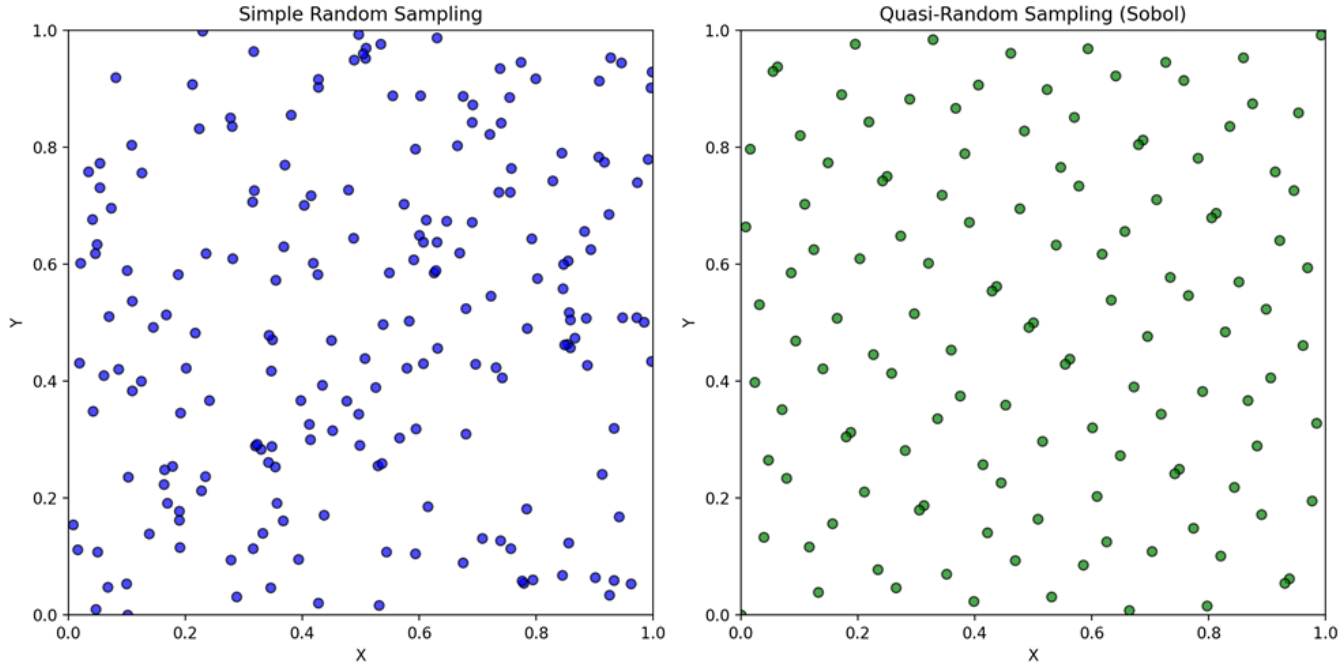


Figure 4. Example of quasi - random sampling pattern of 100 locations (right, as green points) compared to a simple random sampling pattern (left, as blue points). Source: authors.

7. Cluster Sampling

Cluster sampling involves first selecting locations either systematically or randomly, and then grouping them based on a defined criterion, such as the maximum distance between sample points. These groupings, or clusters, are user-defined and typically created for the purpose of intensive studies or sampling in parts of the study area that are of particular interest, for example prospecting of mineral deposits. Note that this method may result in certain large sections of the study area being left without any samples.

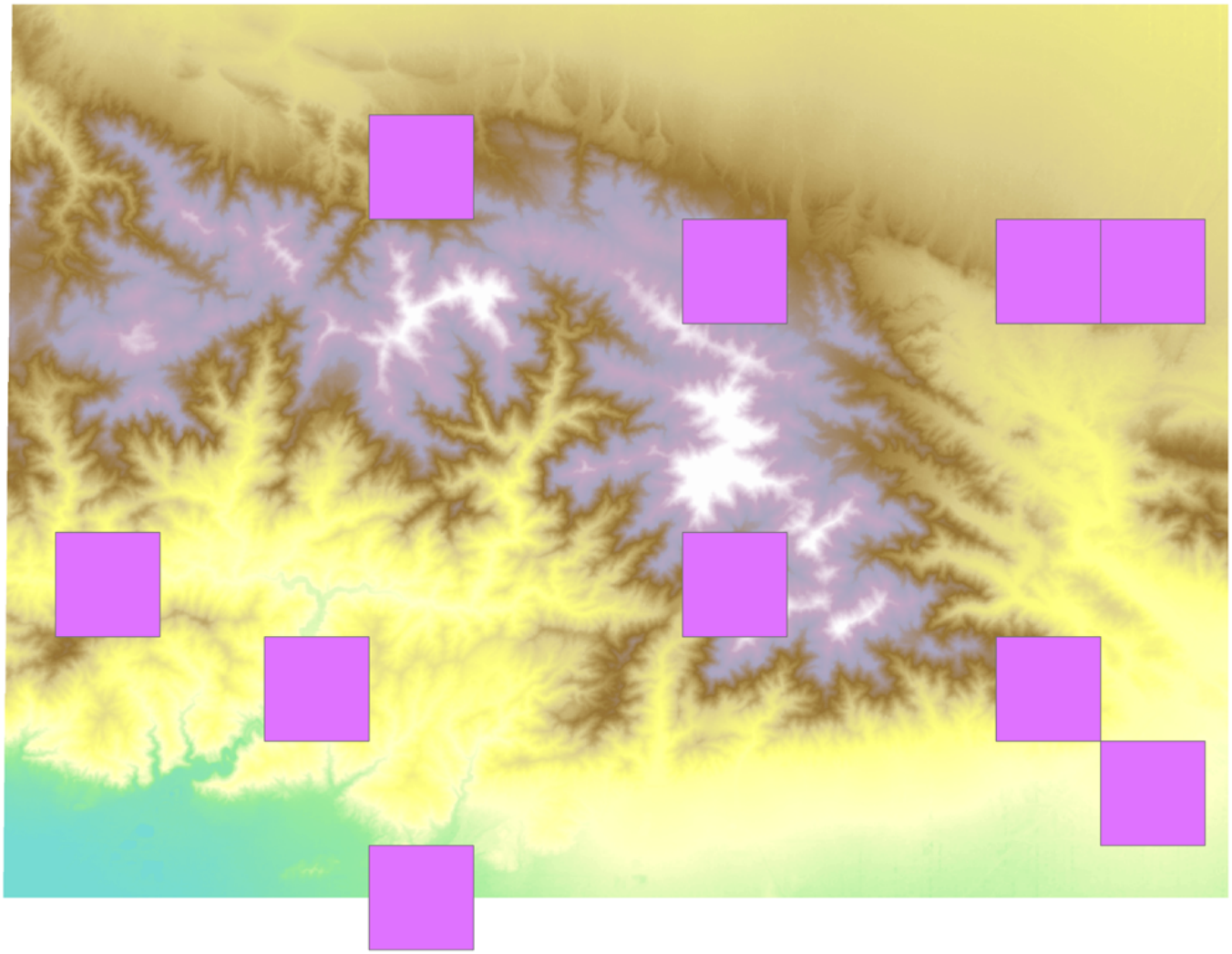
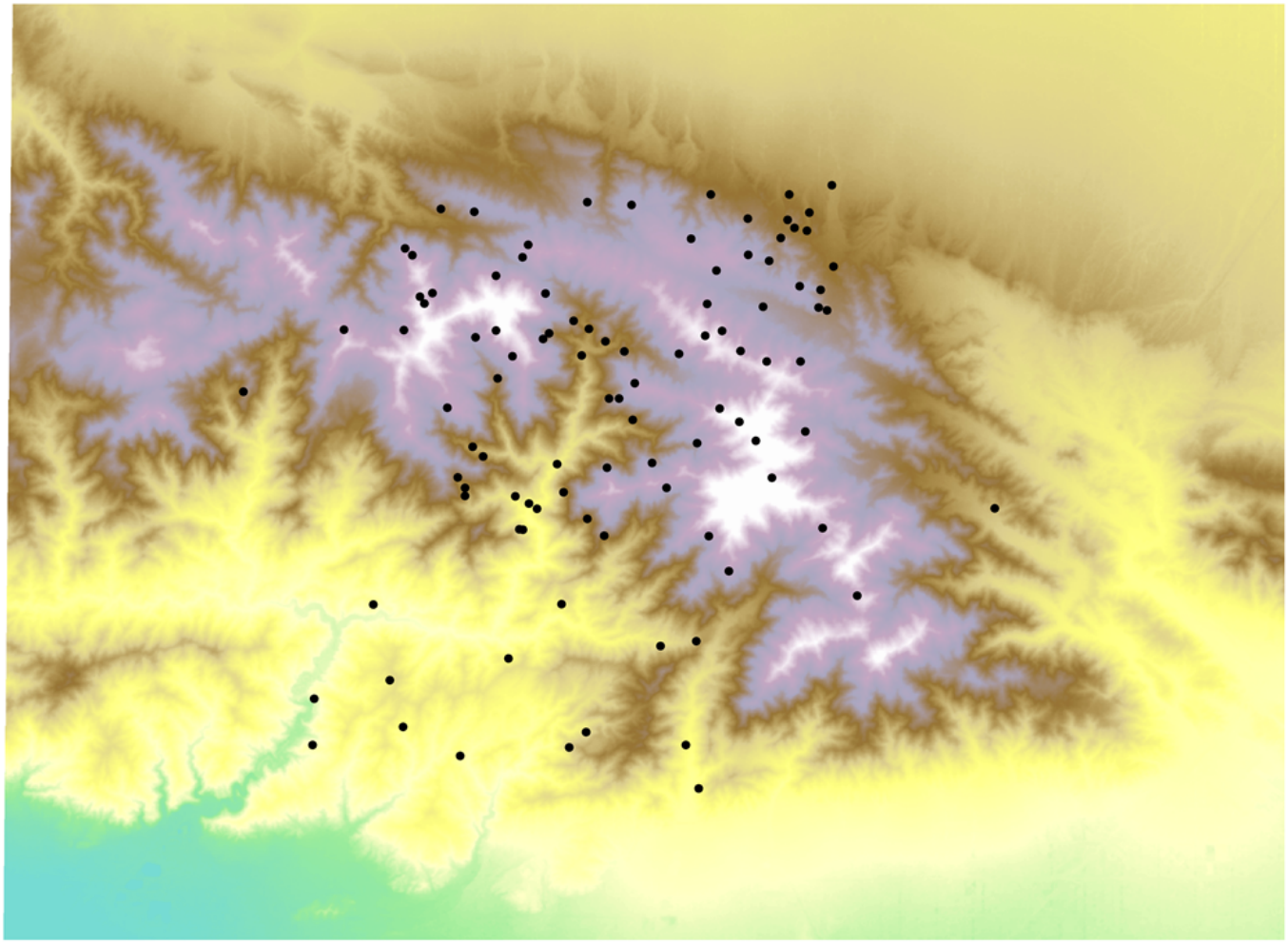


Figure 5. Example of 10 cluster sampling areas representing 100 locations across a study area depicted by the underlying raster image. Source: authors.

8. Adaptive Sampling

Adaptive sampling entails frequent sampling in more relevant or variable areas and less sampling in uniform or less relevant areas. For example, when measuring arsenic levels in groundwater and drinking water supplies, it may be prudent to adaptively increase the number of samples around supply wells (Ayotte et al, 2017). In such a case, sample density may be increased based on field observations of human activities such as mining or farming practices.



Example of adaptive sampling pattern of 100 locations across a study area depicted by the underlying raster map. Source: authors.

References

Ayotte, J.D., Medalie, L., Qi, S.L., Backer, L.C., & Nolan B.T. (2017). Estimating the High-Arsenic Domestic-Well Population in the Conterminous United States. Environmental Science & Technology 51 (21) 12443 – 12454. DOI: 10.1021/acs.est.7b02881.

Şalap-Ayça, S., Jankowski, P., Clarke, K.C., Kyriakidis, P.C., & Atsushi, N. (2018). A Meta-Modeling Approach for Spatiotemporal Uncertainty and Sensitivity Analysis: An Application for a Cellular Automata-based Urban Growth and Land-use Change Model. International Journal of Geographic Information Science 32:1–26. DOI: 10.1080/13658816.2017.1406944

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance-based Sensitivity Analysis of Model Output: Design and Estimator for the Total Sensitivity Index. Computer Physics Communications, 181(2), 259–270. DOI: 10.1016/j.cpc.2009.09.018

Sobol, I.M. (2001) Global Sensitivity Indices for Nonlinear Mathematical Models and their

Monte Carlo estimates. Mathematics and Computers in Simulation 55:271–280.

Stevens, D.L., & Olsen, A.R. (2004). Spatially Balanced Sampling of Natural Resources. Journal of the American Statistical Association 99 (465), 262–278. DOI: 10.1198/016214504000000250

Tarantola, S., Giglioli, N., Jesinghaus, J., & Saltelli, A. (2002). Can Global Sensitivity Analysis steer the Implementation of Models for Environmental Assessments and Decision-Making? Stochastic Environmental Research and Risk Assessment, 16(1), 63–76. DOI:10.1007/s00477-001-0085-x

Theobald, D.M., Stevens Jr., D.L., 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. DOI: 10.1007/s00267-005-0199-x