Environmental Modelling & Software xxx (2010) 1–3

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Software data and modelling news

A spatial sampling optimization package using MSN theory

Mao-Gui Hu, Jin-Feng Wang*

State Key Laboratory of Resources & Environmental Information System, Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences, A11, Datun Road, Beijing 100101, China

A R T I C L E I N F O

Article history:
Received 10 April 2010
Received in revised form
21 August 2010
Accepted 6 October 2010
Available online xxx

Keywords:
Spatial sampling
Nonhomogeneous surface
Variance of estimation
Particle swarm optimization

A B S T R A C T

The density and distribution of spatial samples heavily affect the precision and reliability of estimated population attributes. An optimization method based on Mean of Surface with Nonhomogeneity (MSN) theory has been developed into a computer package with the purpose of improving accuracy in the global estimation of some spatial properties, given a spatial sample distributed over a heterogeneous surface; and in return, for a given variance of estimation, the program can export both the optimal number of sample units needed and their appropriate distribution within a specified research area.

Software availability

Software name: MSN Spatial Sampling Optimization
Developers: Mao-Gui Hu and Jin-Feng Wang
Contact: humg@lreis.ac.cn
Available since: December 2009
Operating system: 32-bit Windows
Program language: C++, C#
Availability: free from http://www.sssampling.org/msn

1. Introduction

Spatial sampling is one of the most basic methods of collecting information in spatial investigation and research, such as studies on natural resources, the environment, and ecosystems (Haining, 2003). Dense samples usually bring high precision, while sparse ones cannot make as precise an estimate of the population (Cochran, 1977). However, cost is another important factor to be considered carefully in practice. An optimized spatial sampling scheme would make a tradeoff between precision and cost. Meteorological network design is a typical example. We will introduce and apply a spatial sampling optimization method using meteorological network design, but the method is not restricted to this application alone.

A meteorological network collects data on a variety of meteorological elements that are commonly used in weather forecasting, environmental evaluation, and global climate change research, such as temperature, pressure, wind, humidity, and rainfall (Yu and Neil, 1990; Pardo-Iguzquiza, 1998; Hansen et al., 2006). Designing a network that provides a representative picture of the properties of the whole region under study is an important but difficult problem. Generally, a good meteorological network would consist of a minimal number of stations that can provide as much information as possible. Depending on the spatial correlation of observation posts throughout a region, criteria based on Kriging variance are adopted by many researchers to optimize a meteorological network (Bastin et al., 1984; Gao et al., 1996; Ahmed, 2004; Barca et al., 2008). To improve the accuracy of areal averages of meteorological elements within a covered region, the total number and distribution of observing stations can be optimized by minimizing the global estimation variance. This optimization is dynamic in that the process usually consists of adding, removing, and moving stations, and it performs well in many instances (Barca et al., 2008). When the goal is to estimate the population mean of some spatial property rather than to predict the value at unsampled locations, however, the minimization of Kriging variance methods cannot be used directly. On one hand, the Kriging method is based on the modeling assumption of surface homogeneity, which in practice is often not satisfied for extensive regions due to financial reasons or physical geography (Li et al., 2008). On the other hand, there is no quantitative relation between the precision of estimated

* Corresponding author.
E-mail address: wangjf@lreis.ac.cn (J.-F. Wang).

1364-8152/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2010.10.006

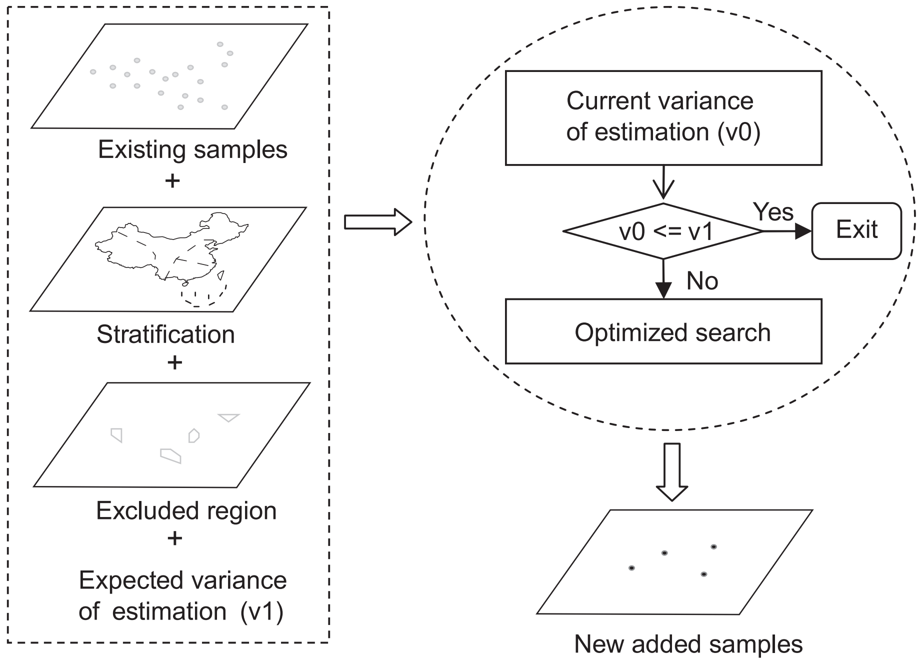

Fig. 1. Architecture of the MSN spatial sampling optimization package.

population mean and the best predicted value at some location, although the former could be affected by the latter (van Groenigen et al., 1999; Stein and Ettema, 2003).

Wang et al. proposed the so-called Mean of Surface with Nonhomogeneity (MSN) method for those complicated surfaces so as to estimate the mean of some property and its variance (Wang et al., 2009). The method combines the merits of spatial stratification and Kriging variance. It assumes that a nonhomogeneous surface could be converted to several small homogeneous surfaces by stratification (Wang et al., 2010), which is possible in most cases. It can estimate the population mean directly. The estimated result of MSN is proven to be the best linear unbiased estimator (BLUE) of the true value (Wang et al., 2009). In practice, a meteorological network often covers very large areas. It is very hard to guarantee that the distribution of a meteorological element follows the homogeneity assumption in the whole region. A spatial sampling optimization method based on the MSN theory would make global estimations and statistical inferences in complicated regions much more efficient and reliable.

2. Software features

A software package that performs spatial sampling optimizations using MSN theory has been developed to run under the Microsoft Windows operation system. A flow chart outlining the algorithmic procedures involved is presented in Fig. 1. The main goal of the package is to measure whether currently installed spatial samples within a region can meet the overall accuracy (variance of mean estimation) required by users. If not, the program determines the number of additional samples needed and where they should be located. The optimization operator is wrapped into a COM component using C++, and can be easily integrated into other systems. The main GUI is a basic GIS platform implemented using C# that makes it convenient for users to view the spatial data. The data format supported is ESRI Shapefiles, a popular format shared by most GIS software.

The optimization process is completed by a wizard that consists of three pages, viz., 'data source', 'set parameters', and 'show results' (Fig. 1). The package is user friendly and can be understood by non-experts. The required input data include the number and location of current spatial samples (meteorological stations) and various stratified regions. The 'forbidden region' file records those selected regions that are to be excluded from consideration as potential sites for samples. This option is of great practical use when regions are inaccessible for one reason or other. Variogram parameters for each stratum can either be imported from a text file or calculated according to observational data from stations. When a stratum's variogram parameters are known, it is then possible to allocate additional stations in the region after performing the optimization, and this can be done even if there were initially no stations.

Another distinguishing feature of the package is its implementation of a combined Monte Carlo and Particle Swarm Optimization (MC-PSO) algorithm to accelerate the optimization process. Finding the best sample locations from thousands of candidates is a combinational problem. It is infeasible and quite impracticable to trial each possible combination and compare them due to the number of stations involved. Given an expected estimated global mean standard deviation that is smaller than the current value, we can first estimate the possible maximum stations m needed with simple sampling theory. Second, a binary search method is used in the range [0, m]. Finally, for a given search number, we then apply the MC-PSO algorithm to find the best candidate sites subject to minimum estimated mean variance criteria.

3. Concluding remarks

Because many different factors are involved, meteorological network optimization is a difficult problem to resolve. The freely downloadable package we have developed for this task is the first version and addresses only the issue of improving data reliability and accuracy by 'adding stations' in an optimally distributional

Please cite this article in press as: Hu, M.-G., Wang, J.-F., A spatial sampling optimization package using MSN theory, Environmental Modelling & Software (2010), doi:10.1016/j.envsoft.2010.10.006

manner. It is not only useful for meteorological network optimization, but also for other similar spatial sampling optimization. A newer version is being developed that will include other features with more complicated functions (Christakos, 2005), such as optimally deleting and moving spatial samples. It will also be freely available to download from the website.

Acknowledgments

This study was supported by CAS (KZCX2-YW-308), MOST (2009ZX10004-201; 2008BAI56B00; 2007DFC20180; 2007AA12Z233; 2006BAK01A13), and NSFC (40471111, 70571076) grants. The authors would like to thank to Luke Driskell for his very kind help on English language polishing of the article.

References

Ahmed, S., 2004. Geostatistical estimation variance approach to optimizing an air temperature monitoring network. Water, Air, and Soil Pollution 158, 387–399.

Barca, E., Passarella, G., et al., 2008. Optimal extension of the rain gauge monitoring network of the Apulian Regional Consortium for Crop Protection. Environmental Monitoring and Assessment 145, 375–386.

Bastin, G., Lorent, B., et al., 1984. Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. Water Resources Research 20 (4), 463–470.

Cochran, W.G., 1977. Sampling Techniques. John Wiley & Sons, Inc.

Christakos, G., 2005. Random Field Models in Earth Sciences. Dover Publ., NY.

Gao, H., Wang, J., et al., 1996. The updated Kriging variance and optimal sample design. Mathematical Geology 28 (3), 295–313.

Haining, R., 2003. Spatial Data Analysis: Theory and Practice. Cambridge University Press, Cambridge.

Hansen, J., Sato, M., et al., 2006. Global temperature change. Proceedings of the National Academy of Sciences 103 (39), 14288–14293.

Li, L., Wang, J., et al., 2008. An information-fusion method to regionalize spatial heterogeneity for improving the accuracy of spatial sampling estimation. Stochastic Environmental Research and Risk Assessment 22 (6), 689–704.

Pardo-Iguzquiza, E., 1998. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. Journal of Hydrology 210 (1–4), 206–220.

Stein, A., Ettema, C., 2003. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. Agriculture, Ecosystems and Environment 94 (1), 31–47.

van Groenigen, J.W., Siderius, W., et al., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87 (3–4), 239–259.

Wang, J.F., Christakos, G., Hu, M.G., 2009. Modeling spatial means of surfaces with stratified nonhomogeneity. IEEE Transactions on Geoscience and Remote Sensing 47 (12), 4167–4174.

Wang, J.F., Haining, R., Cao, Z.D., 2010. Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. International Journal of Geographical Information Science 24 (4), 523–543.

Yu, B., Neil, D.T., 1990. Global warming and regional rainfall: the difference between average and high intensity rainfalls. International Journal of Climatology 11 (6), 653–661.