A method for quantifying artifacts in mapping methods, illustrated by application to headbanging*

Andrew Gelman, Phillip N. Price, and Chia-yu Lin

July 15, 1999

Abstract

Maps of disease rates (and other quantities) often must contend with variance associated with variable population sizes and low incidence within spatial units. These characteristics can lead to substantial statistical noise that can mask underlying spatial variation. As Gelman and Price[1] illustrated, most conventional mapping methods fail to address this problem, and in fact can introduce statistical artifacts: mapped quantities can show spatial patterns even when there are no spatial patterns in the underlying parameter of interest.

Kafadar[2] evaluated the performance of the *headbanging* algorithm for spatial smoothing (Tukey and Tukey[3], Hansen[4]) for eliminating small scale variation and preserving edge structure. Here we perform a simulation study to investigate the artifacts of maps smoothed by unweighted and weighted headbanging. We find substantial artifacts that depend on the spatial structure of the statistical variation (e.g., the spatial pattern of sample sizes) and on the details of the spatial distribution of geographic units.

The methods used here could readily be adapted to study other spatial smoothers; we choose headbanging because (1) it is an important method used in practice, and (2) its heavily computational nature is naturally studied using simulation (in contrast to the analytical methods used by Gelman and Price[1]).

Keywords: headbanging, mapping artifacts, simulation study, spatial statistics, weighted headbanging

1 Introduction

Maps are frequently used to display spatially varying quantities such as disease rates and pollutant concentrations. A general difficulty occurs when sparse sampling within geographical regions masks underlying spatial trends by introducing large uncertainties. Various smoothing techniques have been proposed to reduce this uncertainty and thus to decipher possible spatial trends by taking some kind of average (linear or nonlinear) of geographically neighboring data values.

Kafadar[2] compared some two-dimensional smoothers (disk and weighted average, empirical Bayes, loess, headbanging, resmoothed medians and median polish) by using statistics related to mean squared error and the amount of putative structure captured by the smoothers, and concluded

*To appear in *Statistics in Medicine*. Address all correspondence to Andrew Gelman, Department of Statistics, Columbia University, New York, New York 10027, gelman@stat.columbia.edu. We are grateful to K. Simonson of Sandia National Labs for providing the headbanging algorithm in C code. We thank the U.S. National Science Foundation for support through grant SBR-9708424 and Young Investigator Award DMS-9796129.

that weighted local mean and median polish tended to over-smooth edge structure whereas the median-based *headbanging* algorithm preserved steep spatial structure.

It is obviously desirable for a statistical method to preserve real spatial structure. In addition, however, we do not want a method to introduce spatial structure where none is actually present. Gelman and Price[1] examined this problem by considering the statistical behavior of estimates by geographic area when the underlying parameter has *no* spatial structure. They found that essentially all mapping methods generate statistical artifacts, in that the probability that a given county's parameter estimate will be in the top $x\%$ of all counties is generally a function of sample size, and thus the mapped estimates show patterns due to spatial structure in the sample sizes even if the underlying parameter of interest is patternless. (See Conlon and Louis[5], for related work.)

The present paper extends this work by examining the artifacts in maps smoothed using the unweighted and weighted headbanging algorithms: namely, in what way do the sample size and geographic location of a district affect the probability that its local estimate is in the top $x\%$ of all districts? The methods used in this paper are quite straightforward—setting up a simple spatial model and running simulations to study the probabilities of different sorts of outcomes for the smoothed maps. The purpose of this paper is both to study headbanging and to demonstrate how a simulation approach can be used to study mapping artifacts in a setting in which analytical studies are limited.

2 The headbanging algorithm

The headbanging algorithm is defined in terms of data y_i observed on points i on the plane, with the goal of estimating underlying quantities θ_i. When working with data in geographical districts, we identify the points with the centroids of the districts. Headbanging proceeds in two steps: first, the identification of *triples* of points that are used in a median smoother, and, second, an iterative smoothing algorithm that starts with the data points y_i and produces successive estimates $\hat{\theta}_i$, sequentially updating these estimates to convergence. The smoothing makes use of *weights* w_i, which might be the sample size of the data at point i. Before describing our study, we briefly outline the motivation and steps of the headbanging algorithm.

2.1 Background and motivation

Median-based smoothers were designed to remove spikes and prevent erosion in time series studies (see Goodall[6]). Tukey and Tukey[3] proposed headbanging as an extension of these algorithms to non-gridded spatial data. Hansen[4] implemented that idea on oil field and thermal gradient data. Mungiole et al.[7] expanded this idea to the weighted headbanging algorithm (which we use in this paper) to allow for differential weighting of the values to be smoothed.

The central idea behind headbanging is that certain kinds of spatial variation are likely to be due to small-sample variation or other statistical noise, while others are likely to represent genuine spatial variation in the underlying parameters. Specifically, a "spike"—a single elevated value surrounded by lower values—is assumed to be unlikely to represent a real spatial feature and thus is smoothed out, whereas a ridge or clump of contiguous high values is assumed more likely to be real and thus is preserved.

2.2 Identification of triples of points

To implement the idea that contiguous spatial features should be preserved, each point for which a parameter estimate is to be obtained is determined to be at the center of some number of triples of points. A triple of points is defined as a set of three points that (a) are nearly collinear, as defined by the angle α at the center, which must exceed a pre-specified α^*, and (b) have two endpoints that are among the N nearest neighbors of the center point. The smoothing algorithm (described below) generates a predicted parameter value for each point by using a median smoother that operates on some or all of the triples of which the point is in the center. The maximum number of triples used is determined by a pre-specified parameter, NTRIP: if the point is at the center of more than NTRIP triples, then only the NTRIP "thinnest" triples are used, where thinness is defined in terms of the perpendicular distance from the center point of a triple to the line segment joining the two endpoints.

A point on the plane is classified as an inner point if it is the center of two or more triples, as an edge if it is a center point exactly once, and as a corner if it is never a center point.

So-called edge or boundary effects are a fundamental difficulty in smoothing due mostly to the fact that fewer data are available near the boundaries, but also because of properties of the particular smoother being used. The headbanging algorithm creates "artificial" triples for edge or corner points i by linearly extrapolating the trend from points j and k, both among the N nearest neighbors of i, to a point lying along the line determined by points j and k, and such that point i is equidistant from point j and the extrapolated point, e. For this triple centered at i to be formed, the angle α_j at point j should exceed $90° + \alpha^*/2$. Given the current estimates $\hat{\theta}_i$, the extrapolated value at point e is $\hat{\theta}_e = \hat{\theta}_j + 2(\hat{\theta}_k - \hat{\theta}_j)\frac{d_{ij}}{d_{jk}} \cos(\alpha_j)$, where the notation d indicates the distance between two points.

2.3 Smoothing procedure

Weighted headbanging proceeds as follows:

1. Initialize by setting the estimate $\hat{\theta}$ to the data vector y.

2. At each point i, apply the following smoothing procedure:

   (a) Convert each triple with point i as the center into a pair (low_l, high_l) where low_l and high_l are the lower and higher of the present $\hat{\theta}$ estimates at the end points of the l-th triple.

   (b) Compute (high screen)$_i$, the weighted median of the high_l values, and (low screen)$_i$, the weighted median of the low_l values, for the triples l that include point i as center.

   (c) Take, for the new estimate $\hat{\theta}_i^{\text{new}}$ at point i the weighted median of the following three numbers: (low screen)$_i$, (high screen)$_i$, and the current estimate $\hat{\theta}_i$, with weights equal to the average weight of the units in the low screen, the average weights of the units in the high screen, and the weight of point i.

3. Repeat the above steps 2–4 for a fixed number of iterations, or until no further change takes place. (We know of no theoretical results on the convergence of this algorithm, but convergence has not been a problem in our experience.)

Unweighted headbanging is equivalent to weighted headbanging with all weights set to 1.

To avoid introducing a dependence on the order in which the data points are smoothed, the vector $\hat{\theta}$ is replaced by the updated $\hat{\theta}^{\text{new}}$ only at the end of each iteration. Parameters N, the number of nearest neighbors of a point that may be used in smoothing that point, and NTRIP, the maximum number of triples that may be used for smoothing at any point, must be set. Hansen[4] suggests setting $\alpha = 135°$.

3 Simulation study

3.1 Simulated data

To examine the artifacts that can appear in maps smoothed by headbanging, we perform simulation studies based on a Gaussian process with either independent or spatially correlated variances. For these simulations we use the locations of the 95 counties in Tennessee; see Figure 1.

We assume that each county i has an underlying parameter θ_i (e.g., a disease rate, or the average concentration of a pollutant) drawn from a Gaussian distribution, and that the data from a county allow estimation of this underlying parameter with an uncertainty that varies with sample size. We model θ as either independent ($\text{var}(\theta) = \tau^2 I$) or spatially correlated with a simple distance-dependent covariance ($\text{cov}(\theta_i, \theta_j) = \tau^2 \exp[-(d_{ij}/(50 \text{ km}))^2]$).

The data vector y is modeled as independent, conditional on θ, with Gaussian errors:

$$y = \theta + \epsilon, \tag{1}$$

with $\text{var}(\epsilon_i) = \sigma^2/n_i$. For each county i, we set the n_i to the county population divided by 5000, rounded to the nearest integer; these n_i's vary from 1 to 174, with a median of 6 and a mean of 11.3 (see Figure 1). We set the variance parameters to $\tau = 1.0$ and $\sigma = 0.7$, which were chosen to be consistent with the levels of variation in logarithms of measurements of home radon levels within counties, a topic that motivated some of this research (see Gelman and Price[1]).

3.2 Smoothed estimates

For both the spatially independent and autocorrelated models, we simulate 1000 data vectors y and, to each simulation, apply the unweighted and weighted headbanging algorithms with $\alpha^* = 135°$ and the following three choices of $(\text{N}, \text{NTRIP}) = (8, 10), (6, 8), (4, 6)$. (These values appeared reasonable in the sense that they performed well in recovering the mean structure in some exploratory simulations.) For the weighted algorithm, we use the county sample sizes as weights.

We evaluate headbanging in terms of the properties of the simulated smoothed estimates. Of particular interest is whether counties with certain characteristics are more likely to be highlighted than others, in a map in which the top 10% of counties are highlighted. For each simulation study and each of the three sets of headbanging parameters (N, NTRIP), we examine the probability that a county will be highlighted—that is, the fraction of the 1000 simulations for which its parameter estimate is in the highest 10%—as a function of the number of triples used for smoothing and the county location (in particular, if it is an inner, edge, or corner point).

Even before simulating any data, we can examine the spatial information that will be used by the headbanging algorithm. Figure 2 displays the number of triples used to estimate each county's parameter value, for the three choices of the smoothing parameters N and NTRIP. Inner points tend to have most triples whereas corner points have the least, but some corner points are associated with large numbers of triples when the extrapolation conditions work out just right; for example, one of the counties in the northeast corner of the state is a corner point that has 9 triples when N=8 and NTRIP=10.

In examining artifacts of mapping, we also compare headbanging to three other estimates: (a) the raw data; (b) a nonspatial Bayes estimate (shrinking each county's data y_i toward the mean by multiplying by $\tau^2/(\tau^2 + \sigma^2/n_i)$); and (c) a spatial Bayes estimate (shrinking the data vector y toward the mean by multiplying by the matrix $(T^{-1}+\Sigma^{-1})^{-1}\Sigma^{-1}$, where T is the spatially-correlated covariance matrix given in Section 3.1 and Σ is the diagonal matrix with elements $\Sigma_{ii} = \sigma^2/n_i$). The non-spatial and spatial Bayes models exactly match the simulation methods for the non-spatial and spatial simulated data, so they represent a sort of best case: there is no model misspecification, so the variation between underlying parameters and their estimates is purely due to sampling variation.

(Even when the model is known, however, we should expect artifacts, as in Gelman and Price[1].)

We might expect headbanging to perform poorly with respect to the non-spatial simulated data. For instance, in the non-spatial data, high parameter values have no tendency to clump together spatially, and thus will often occur as "spikes" that headbanging will tend to smooth out. One would hope, though, that headbanging would perform better with the spatially correlated data, since that is at least the sort of data for which the method was designed, although still not optimal in that headbanging is designed to preserve linear features of elevated (or depressed) values, which will not be preferentially produced by our assumed isotropic Gaussian process.

3.3 Results

As an illustration of potential spatial artifacts, Figure 3 displays, for each county in Tennessee, the probability that it is highlighted (that is, in the top 9 of Tennessee's 95 counties) after unweighted headbanging (for one choice of the headbanging parameters N, NTRIP), based on simulations from the independent and correlated models. Edge and corner counties are more likely to be highlighted than are inner counties, with this effect being larger for the independent data than for the spatially correlated data. Similar plots for other settings of the headbanging parameters and for weighted headbanging show other patterns, which we shall explore systematically by examining what factors affect the probability that a county is highlighted under the different mapping methods.

We begin by plotting the probability of each county being highlighted vs. the logarithm of the sample size in the county, as displayed in Figures 4, 5, and 6 for unweighted headbanging, weighted headbanging, and the comparison methods (described at the end of the previous section), respectively. In each figure, the top row of plots shows results based on simulations from the spatially independent model, while the bottom row concerns the spatially correlated model.

Figure 4 shows that, for unweighted headbanging, for all three settings of the headbanging parameters, there is quite a bit of variation in the probability of being highlighted (by comparison, if there were no artifacts, the probability of being highlighted would be 10% for all counties). This is true for both the spatially uncorrelated and spatially correlated simulated data. There is a slight negative correlation with sample size in the plots, meaning that counties with smaller samples are more likely to be highlighted, but this effect is weak.

In Figure 5, we see that the pattern is reversed for weighted headbanging, where counties with high sample sizes (and thus, high weights, since we set weights equal to sample sizes) are more likely to be highlighted. For both the weighted and unweighted algorithms, corner and edge points are more likely to be shaded than inner points, which makes sense since inner points tend to have more neighbors and thus are smoothed more in the headbanging algorithm.

By comparison, Figure 6 shows the probability of being highlighted vs. log sample size for the raw data and the Bayes algorithms. Just as using an intrinsically spatial method such as headbanging to analyze data without spatial correlation might be expected to yield poor results, we might also expect that fitting the spatial Bayes model to the non-spatial data (or vice-versa) would be unsatisfactory.

For both the raw data and the non-spatial Bayes estimates, the vertical scatter at a given sample size is due to the finite number of simulations (1000): there is nothing in the simulated data or in the statistical methods that systematically treats two counties differently if their sample sizes n_i are the same. For the spatial Bayes estimate, though, as for the headbanging estimate, this is no longer the case: the sample sizes of the surrounding counties (and, in the case of headbanging, their angular distribution) can also have an influence. Thus, if the number of simulations were greatly increased, the plots in the left and center columns of Figure 6 would approach smooth curves, while the rightmost would not.

For both the independent and spatially correlated models, we mostly find the patterns identified by Gelman and Price[1]: maps of raw data and Bayes estimates favor counties with small and large sample sizes, respectively. The only exception is when the correct (spatial) Bayes estimate is fit to the spatially correlated data: the correlation with sample size appears here too, but, on the whole, the spatial artifacts are nearly nonexistent, with the probability of being highlighted close to a constant 10%. This is interesting as a best-case scenario, though it is perhaps not so realistic a comparison to the other methods, since it assumes exact knowledge of the spatial autocorrelation matrix. (By comparison, the assumption that τ/σ is known in the nonspatial Bayes estimate is less controversial, since this scalar parameter can be readily estimated from data.)

Figures 4–6 have shown us that headbanging has substantial artifacts: when the parameter of interest has no underlying spatial structure (top row of each figure), some counties are about five times as likely to have estimates in the top 10% as are others. The situation is only slightly better when there is underlying spatial structure to the parameter (bottom row)—some counties are still much more likely to be highlighted than are others. However, these artifacts are only weakly associated with sample size. What other factors could be relevant here? The most natural place to look is in the headbanging algorithm itself. Figures 7 and 8 display, for the unweighted and weighted methods, the probability of being highlighted as a function of the number of triples used in smoothing a county's estimate. (The number of triples depends on the spatial location of a county relative to its neighbors, and also on the parameters (N, NTRIP), as illustrated in Figure 2.)

Figure 7 shows that, for unweighted headbanging, the probability of being highlighted is strongly negatively correlated with the number of triples. For counties with the same number of triples, corner points are favored over edge points, which in turn are favored over inner points. With the correlated model and the high (N, NTRIP) settings, the dependence on number of triples is weaker, indicating perhaps that the data from the spatial model are smooth enough that, once averaging is done over 4 or more triples, the estimates are relatively stable.

Figure 8 shows the artifacts for weighted headbanging as a function of number of triples. Once again, we see that counties with fewer triples are favored, but this effect is relatively weak, with the exception of the rightmost plots, where the number of triples is small. (In those counties with no triples, the headbanging estimate is simply the raw data, and so these are the most likely to be highlighted.)

For the weighted headbanging plots (Figures 5 and 8), we see a clustering of some outlying counties with high probabilities of being highlighted that are not explained by having particularly high sample sizes, few triples, or being corner or edge points. In seeking a factor to explain this, it is natural to look at the weights that go into the smoothing. In particular, for each of the settings of (N, NTRIP) in the weighted headbanging algorithm, we calculate the relative weight for each county: its weight divided by the total weights of itself and its neighbors. (For counties with no neighbors—that is, at the center of no triples—the relative weight is set to 1.) In Figure 9, clusters are apparent in the upper right of each plot, indicating a set of counties that are more likely to be highlighted because of their high weights relative to their neighbors. This is the factor missing in Figures 5 and 8.

Because these plots are based on 1000 simulation draws, we can expect each point to have simulation variability on the order of $\sqrt{(0.1)(0.9)/1000} = 0.01$; this is apparent in the left two columns of Figure 6 where, in fact, sample size is the only factor influencing the probability of being highlighted. We could, of course, reduce this variability by running more simulations, but we choose not to for three reasons. First, the computer package that we used for this work (S-Plus) ran into memory problems when we tried to do more simulations. Second, the simulation variability is minor and does not obscure the systematic variation we are studying. Third, and most important, one of the goals of this paper is to illustrate the simulation approach as a method of studying mapping artifacts; for this purpose, it is important to see that 1000 simulations are enough for us to clearly see the important patterns.

4 Discussion

What would happen if you naively look at a map smoothed by headbanging with the 10% highest locations highlighted? Our simulations show that, even if the underlying parameters θ_i come from a spatially stationary process, you should expect to see some artifacts. Corner points, edge points, and, with unweighted headbanging, points with fewer triples will be more likely to be highlighted.

These edge effects are particularly apparent in cases (as in the long, skinny state of Tennessee) in which a large fraction of spatial units are on the boundary of the space being considered. With weighted headbanging, points with high weights relative to their neighbors are more likely to appear on the highlighted map. By comparison, the spatial artifacts of raw data and Bayes estimates, although comparable in magnitude to those of headbanging, are determined by sample size rather than spatial location.

One way to understand these results is in terms of random variation: plots (not displayed here) show that, for all the mapping methods, the counties that have the highest probability of being highlighted are those with the highest between-simulation standard deviation in their mapped values: they are thus the most likely to be highest or lowest, and thus most prominent, in a map. Spatial smoothing methods such as headbanging can equalize the variances somewhat, but there is ultimately no way to avoid unequal variances given that data come with unequal sample sizes. As a result, one must be aware of these sorts of artifacts when interpreting maps of raw or estimated values.

Once noted, all these patterns make sense and are no surprise. However, the simulation study is valuable, not merely to alert us to these source of artifacts, but to indicate the magnitude of the potential problems (in our simulations, the probability of a county being highlighted varies between about 5% to 20%). The simulation study also allows us to graphically investigate the factors that contribute to the artifacts.

The spatial artifacts we are examining can be important because, without knowledge of them, a map user can easily misinterpret patterns in a map (as discussed by Gelman and Price[1]). The simulation study presented here illustrates the kind of analysis that can be done to systematically study such artifacts in the context of a computational procedure such as headbanging.

We think an investigation of statistical artifacts should be a part of the development of any new mapping method, since such artifacts, if substantial, can easily lead to incorrect inference. For instance, if the Tennessee department of health made a map of a rare disease by county, smoothed by headbanging, and noticed that many of the highest estimated rates occurred on the borders of the state, we would suspect that they were seeing an artifact of the headbanging procedure rather than some environmental influence related to different pollution laws in adjacent states.

The simulation methods outlined in this paper allow artifacts to be investigated with minimal effort: one can simply define a statistical model, simulate from it, perform whatever smoothing method is being investigated, and summarize the results. This approach can be used in combination with studies such as Kafadar[2] that evaluate the performance of mapping methods based on accuracy of estimates and ability to capture underlying spatial structure.

References

[1] Gelman, A. and Price, P. N. (1999). All maps of parameter estimates are misleading. *Statistics in Medicine* **18**, 3221–3234.

[2] Kafadar, K. (1996). Smoothing geographical data, particularly rates of disease. *Statistics in Medicine* **15**, 2539–2560.

[3] Tukey, P. A., and Tukey, J. W. (1981). Graphical display of data sets in 3 or more dimensions. In *Interpreting Multivariate Data*, ed. Barnett, V. New York: Wiley.

[4] Hansen, K. M. (1991). Headbanging: robust smoothing in the plane. *IEEE Transactions on Geoscience and Remote Sensing* **29**, 369–378.

[5] Conlon, E. M., and Louis, T. A. (1999). Addressing multiple goals in evaluating region-specific risk using Bayesian methods. In *Disease Mapping and Risk Assessment for Public Health*, ed. A. Lawson et al., 31–47. New York: Wiley.

[6] Goodall, C. (1990). A survey of smoothing techniques. In *Modern Methods of Data Analysis*, ed. Fox, J., and Long, J. S.. Newbury Park: Sage Publications.

[7] Mungiole, M, Pickle, L. W., and Simonson, K. H. (1998). Application of a weighted headbanging algorithm to mortality data maps. *Statistics in Medicine*.

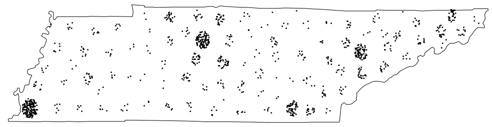

Figure 1: Tennessee with its 95 counties; the number of dots in each county indicates the hypothetical sample size, which is set to the county population divided by 5000.

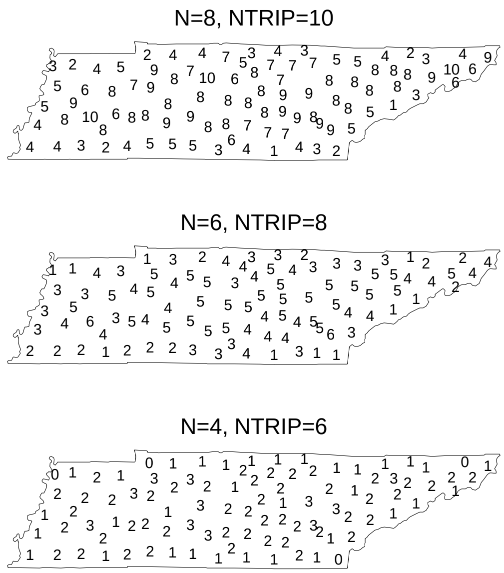

Figure 2: The counties of Tennessee with the number of triples used to smooth each county value under three different settings of the headbanging parameters. Inner points tend to have more sets of triples than edge points, and the number of triples possessed by a corner point depends on its neighboring condition for extrapolation.

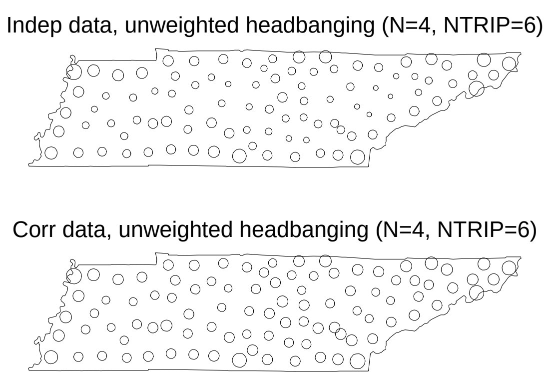

Figure 3: The probability that each county is highlighted (that is, in the top 10%) of headbanging-smoothed maps with data from the independent and correlated models. Area of each circle is proportional to the probability that the county is highlighted. (Results are shown for unweighted headbanging with N=4 and NTRIP=6.) Counties on the border of the state are more likely to be highlighted, especially with data from the independent model.

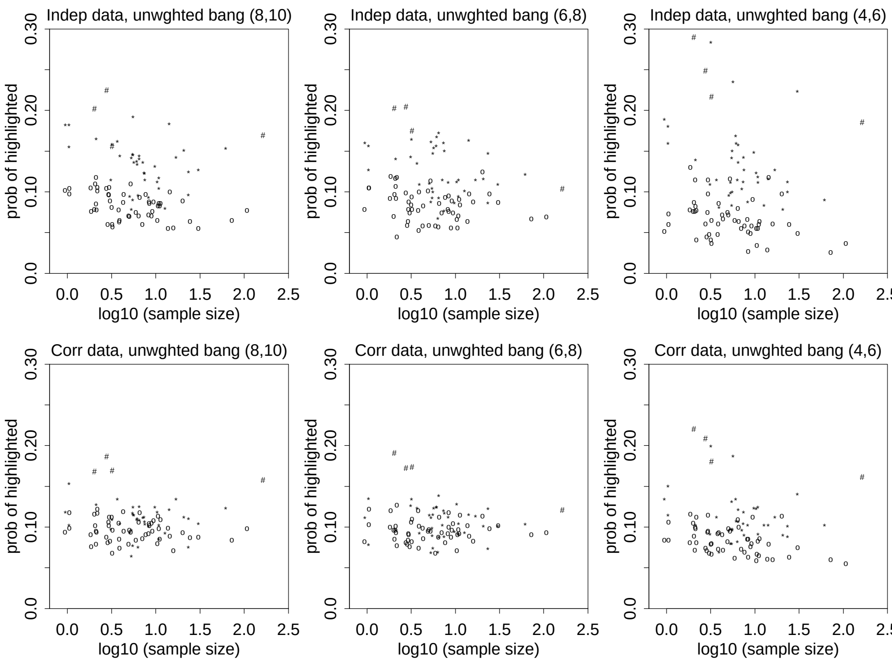

Figure 4: Probability of a county being highlighted vs. log sample size, for maps smoothed using *unweighted* headbanging, for three settings of the headbanging parameters: $(\mathrm{N}, \mathrm{NTRIP}) = (8, 10)$, $(6, 8)$, and $(4, 6)$. The top and bottom row are based on simulations from the spatially independent and autocorrelated models, respectively. Within each plot, each symbol represents a county in Tennessee, with o, *, and # indicating inner, edge, and corner points, respectively. Horizontal jitter has been added to separate the points.

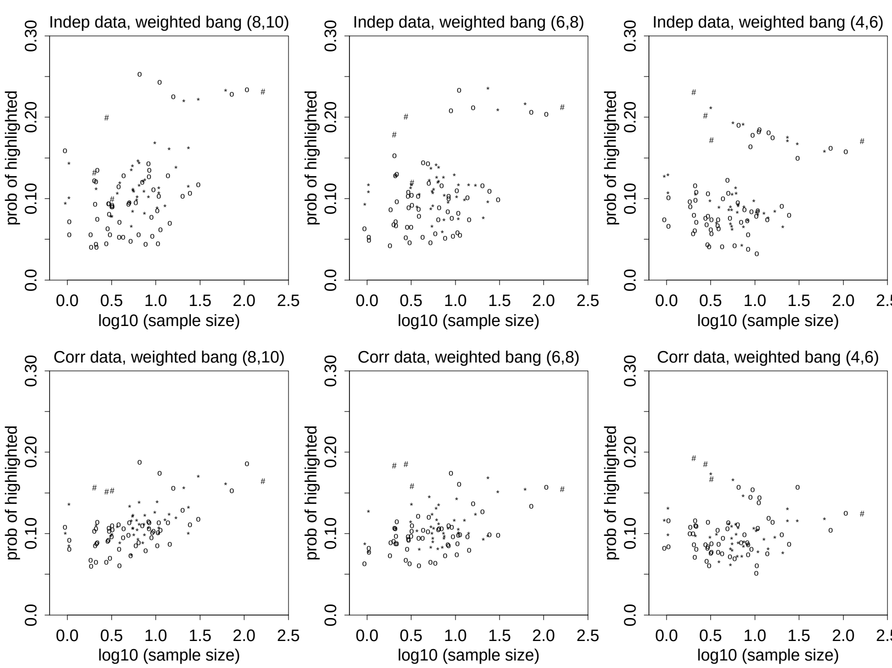

Figure 5: Probability of a county being highlighted vs. log sample size, for maps smoothed using *weighted* headbanging. See caption of Figure 3 for further explanation of these plots.

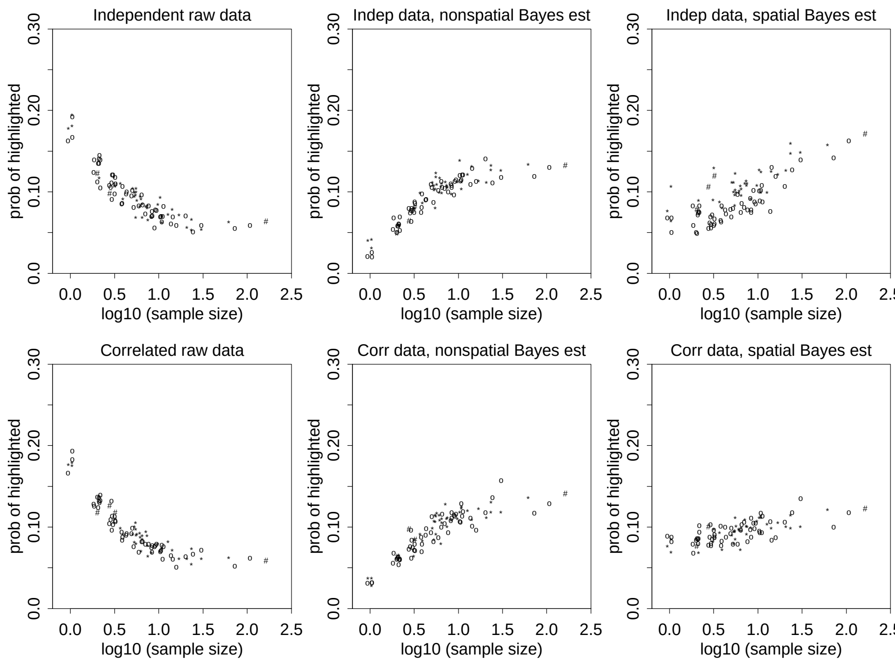

Figure 6: Probability of a county being highlighted vs. log sample size, for raw data and Bayes estimates based on spatial and nonspatial models. See caption of Figure 3 for further explanation of these plots. In the top row, the center plot corresponds to the Bayes estimate under the correct model, and in the bottom row, the rightmost plot is the correct Bayes estimate.

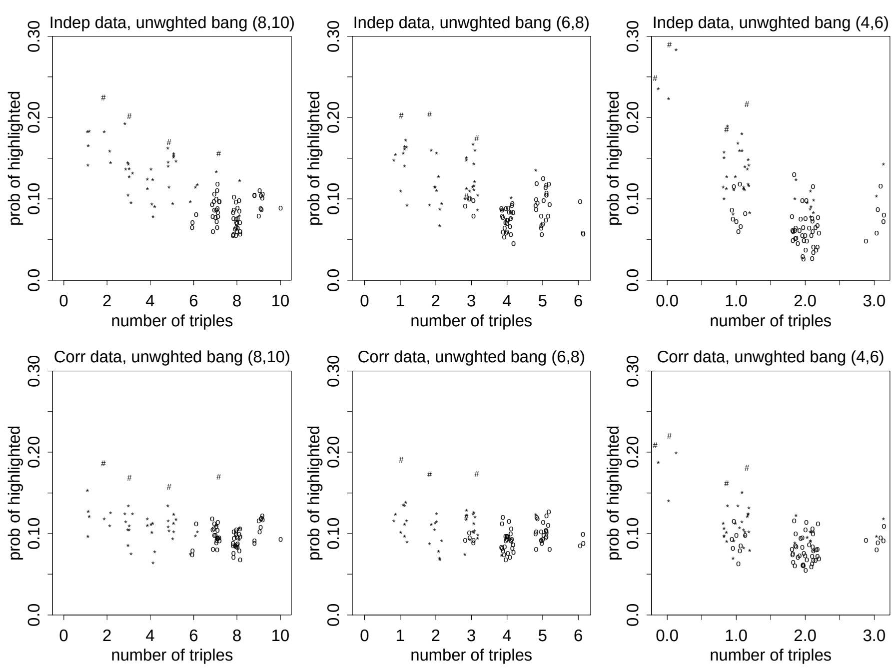

Figure 7: Probability of a county being highlighted vs. number of triples used in the smoothing, for maps smoothed using *unweighted* headbanging. See caption of Figure 3 for further explanation of these plots. Corner and edge points with fewer triples tend to be more likely to be highlighted.

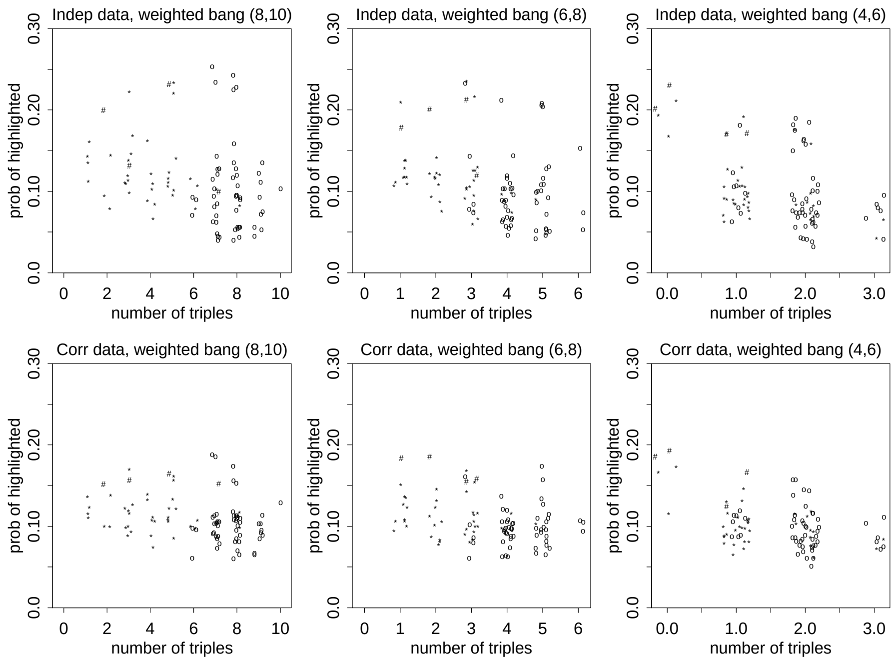

Figure 8: Probability of a county being highlighted vs. number of triples used in the smoothing, for maps smoothed using *weighted* headbanging. See caption of Figure 3 for further explanation of these plots. Corner and edge points with fewer triples tend to be more likely to be highlighted, but this pattern is less strong than in Figure 7.

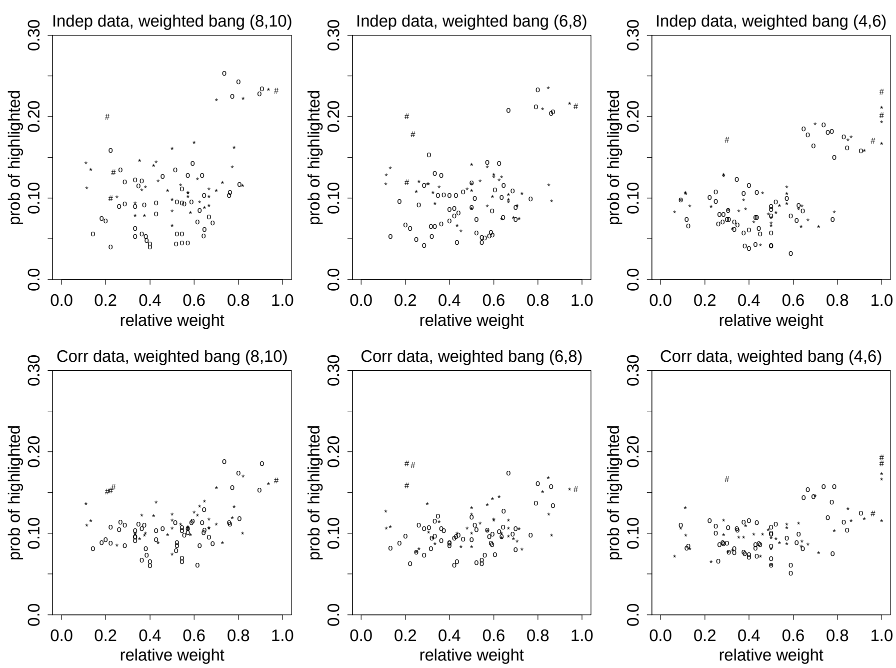

Figure 9: Probability of a county being highlighted vs. relative weight of a point (its weight divided by the total weights of that point and its neighbors), for maps smoothed using *weighted* headbanging. See caption of Figure 3 for further explanation of these plots. Points with very high relative weight are more likely to be highlighted, which explains the clusters of high points in Figures 4 and 7.