STATISTICS IN MEDICINE
Statist. Med. **18**, 3201–3209 (1999)

APPLICATION OF A WEIGHTED HEAD-BANGING ALGORITHM TO MORTALITY DATA MAPS†

MICHAEL MUNGIOLE[1]*, LINDA W. PICKLE[1] AND KATHERINE HANSEN SIMONSON[2]

[1] *Centers for Disease Control and Prevention, National Center for Health Statistics, 6525 Belcrest Rd., Rm. 915, Hyattsville, MD 20782-2003, U.S.A.*
[2] *Sandia National Laboratories, PO Box 5800, Albuquerque, NM 87185-0844, U.S.A.*

SUMMARY

Smoothed data maps permit the reader to identify general spatial trends by removing the background noise of random variability often present in raw data. To smooth mortality data from 798 small areas comprising the contiguous United States, we extended the head-banging algorithm to allow for differential weighting of the values to be smoothed. Actual and simulated data sets were used to determine how head-banging smoothed spike and edge features in the data, and to observe the degree to which weighting affected the results. As expected, spikes were generally removed while edges and clusters of high rates near the U.S. borders were maintained by the unweighted head-banging algorithm. Incorporating weights inversely proportional to standard errors had a substantial effect on smoothed data, for example determining whether observed spikes were retained or removed. The process used to obtain the smoothed data, including the choice of head-banging parameters, is discussed. Results are considered in the context of general spatial trends. Published in 1999 by John Wiley & Sons, Ltd. This article is a U.S. Government work and is in the public domain in the United States.

1. INTRODUCTION

When looking at mortality data maps, the reader is often unable to discern any spatial trends, either because no particular patterns exist in the data, or because a substantial amount of noise (that is, spatial variability) masks the patterns that are present. For example, Lewandowsky and Behrens[1] found that the consistency of detection of high rate clusters on a simulated map was hampered as an increasing amount of noise was added to the data. Statistical smoothing algorithms can be used to enhance hidden patterns in noisy data. This is accomplished by removing local small-scale variations while maintaining larger-scale regional trends.

Two types of features that may influence one's ability to recognize broad trends in spatial data are spikes and edges. A spike represents an isolated extreme value (peak or depression) that contrasts greatly with other nearby data points. Spikes may be caused by measurement error and represent inherent noise in the mapped variable or they may be legitimate values, for example, a high rate for a city that is surrounded by suburbs with lower rates. Edges are zones of rapid transition in the variable of interest. They may appear as ramps, valleys, or ridges and generally represent legitimate structure in the data. In fact, edges can be particularly valuable in helping the

* Correspondence to: Michael Mungiole, Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD 20783-1197, U.S.A. E-mail: mmungiole@mail.arl.mil
† This article is a U.S. Government work and is in the public domain in the United States.

Published in 1999 by John Wiley & Sons, Ltd.

reader to identify spatial regions that are distinct from one another in terms of the mapped variable.

In applications such as the various geologic examples discussed in Hansen[2], spikes are physically unlikely and it is thus desirable for smoothers to remove all spikes while retaining the edges. When working with mapped statistical data, however, we can often distinguish between spikes that are credible (for example, rates in cities with large populations) and those that are of suspect origin (likely due to random noise). We would like our smoother to retain the former (and the edges as well) while removing the latter.

Many authors have observed that the use of mean-based smoothers results in the blurring of both the spikes and edges for one-dimensional time series data[3] and for spatial data.[2,4,5] Median-based approaches often represent a reasonable alternative, removing spikes while preserving the sharpness of edges. Tukey[6] and Velleman[3] propose a variety of median-based smoothers for time series data, while Jain[4] discusses non-linear filters for two-dimensional lattices.

The problem of smoothing non-gridded spatial data has received attention from a variety of authors. Kafadar[5] discusses a number of the techniques that have been proposed, and evaluates their performance on several constructed data sets using an overall mean squared error statistic. The results of this comparative study suggest that the weighted average (based on inverse squared distance) and the median-based head-banging smoother, proposed by Tukey and Tukey,[7] and implemented as in Hansen,[2] perform well. However, the weighted average and most other smoothers included in the comparison tended to oversmooth in the presence of edges, whereas head-banging was able to preserve their steep structure.

To utilize the best features of these smoothers, we decided to use a modified head-banging approach in smoothing mortality data for an atlas of maps of numerous small areas in the contiguous United States.[8] Because the original implementation of head-banging[2,7] does not allow for the use of reliability estimates during the smoothing process, we have developed an enhanced version, in which the user may provide a set of weights along with the mapped data. We refer to this new technique as *weighted head-banging* (a copy of the weighted head-banging code, written in the C language, is available from K. Simonson), and demonstrate its application to mortality data.

2. METHODS

2.1. Notation: Weighted Medians

Suppose that a mapped data set contains a total of n data points. Each point, i, consists of a location in the plane, $\mathbf{x}_i = (x_{i1}, x_{i2})$, which is assumed to be known without error, along with a (possibly noisy) observation, y_i, and a weight, w_i, which measures the reliability of y_i. The goal is to smooth the mapped values $\{y_i, i = 1, \ldots, n\}$, accounting for both location and reliability. Our approach assigns a smoothed value at location $\mathbf{x}_i$ that is based on weighted medians of the observations at neighbouring locations and the observation y_i itself.

Given m values $\{y_1, y_2, \ldots, y_m\}$, with corresponding weights $\{w_1, w_2, \ldots, w_m\}$, the *weighted median* of the y's is found as follows. First, sort the observations, so that $y_{(1)} \leqslant y_{(2)} \leqslant \ldots \leqslant y_{(m)}$. The weights, which remain with their original observations, are re-ordered such that $w_{(j)}$ corresponds to the sorted observation $y_{(j)}$. Now compute the cumulative sums of the re-ordered weights:

$$S_k = \sum_{j=1}^{k} w_{(j)}, \quad k = 1, \ldots, m. \tag{1}$$

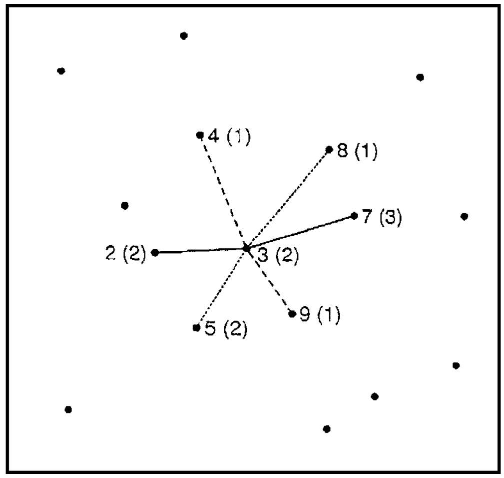

Figure 1. Three nearly collinear triples are used for smoothing at this centre point. Observed values are printed at each data location, with corresponding weights given in parentheses. The low screen is *weighted median* (2, 4, 5) = 4. The high screen is *weighted median* (7, 8, 9) = 7. The observed value at the centre point lies below the low screen, so the weights are used to determine whether adjustment is needed. Because the sum of endpoint weights (10) exceeds the number of triples times the centre point weight ($3 \times 2 = 6$), the smoothed value at the centre point is set to the low screen

The index location, r, of the weighted median is the smallest value of k such that S_k is at least half of the total sum of the weights:

$$r = \min \{k | S_k \geqslant S_m/2\}. \tag{2}$$

If S_r is strictly greater than $S_m/2$, then the weighted median value is given by $y_{(r)}$. If S_r is equal to $S_m/2$, then the weighted median value is taken to be the average of $y_{(r)}$ and $y_{(r+1)}$.

2.2. The Weighted Head-Banging Smoother

In one-dimensional running median smoothers, the output value at a given point is the median of the observed value at that point, and a specified number of neighbouring points on both the left and right sides. Head-banging attempts to bring the simplicity of this approach to spatial data, where left and right neighbours are not well defined. Triples of nearby data points are used in their place.

For each point in the data set, a collection of nearly collinear triples is defined. Every triple in the collection consists of three data points in the same region of the plane, and is centred at the location of the point to be smoothed. A method for identifying such collections is discussed in the next section; for now assume that at least one triple has been chosen for each point in the data set. The iterative weighted head-banging smoother is carried out as illustrated in Figure 1.

For each i, set the initial estimate for location $\mathbf{x}_i$ at its observed value. Let T_i represent the collection of triples centred at $\mathbf{x}_i$. Now for the jth triple in T_i, let low_{ij} represent the smaller of the two endpoint values, and let high_{ij} represent the larger. Define the following two quantities:

$$\text{high screen } (i) = \text{weighted median}_j \ (\text{high}_{ij})$$

$$\text{low screen } (i) = \text{weighted median}_j \ (\text{low}_{ij}).$$

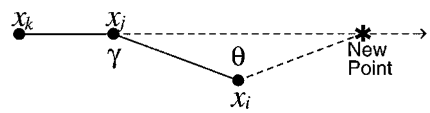

Figure 2. Extrapolating triples for an edge or corner point. If no qualifying triples are centred at point $\mathbf{x}_i$, new triples are formed by extrapolating the y-values from two neighbouring locations, $\mathbf{x}_j$ and $\mathbf{x}_k$. These values are linearly extrapolated to a point lying along the line determined by $\mathbf{x}_j$ and $\mathbf{x}_k$, such that $\mathbf{x}_i$ is equidistant from $\mathbf{x}_j$ and the new point [after Hansen[2]]

The low screen is an estimate of the lowest value that is consistent with the data in the neighbourhood of the centre point, and the high screen is an estimate of the highest such value.

If the current centre value lies between these screens, it is left unchanged at this iteration. If the current centre value is smaller than the low screen, the weights are used to determine whether it will be adjusted. Specifically, if the sum of the weights of all endpoints of the triples exceeds the number of triples times the weight of the centre point, then the smoothed value at the centre point takes the value of the low screen. Similarly, if the current centre value is larger than the high screen, and the summed endpoint weights exceed the number of triples times the centre weight, the smoothed centre value is set equal to the high screen. Even if the current centre value is changed to the low or high screen, it still retains its original weight.

This process is carried out for each point in the data set and all data points are updated simultaneously at the end of each iteration. Thus, the order in which points are smoothed does not influence the values resulting from a particular iteration. The procedure is repeated for a fixed number of iterations, or until no further changes take place. When equal weights are assigned to each data point, this procedure is identical to that outlined in Tukey and Tukey.[7]

2.3. Selection of Triples

Two general criteria for selecting triples are as follows: the three points should be located close to one another, so that smoothing takes place on a local or regional rather than a global scale; and the three points should be roughly collinear, so that directional trends can be recognized and enhanced. An algorithm for identifying collections of triples meeting these criteria is described in Hansen[2] and has been adopted here.

The regional criterion is satisfied by requiring that each of the endpoints of a candidate triple be among the NN nearest neighbours of the centre point. The scalar NN is a parameter chosen by the user; increasing NN will increase the degree of smoothness in the final map. The directional criterion is specified in terms of θ, the centre angle. The user selects a threshold, θ^*, and those triples for which $\theta \geqslant \theta^*$ are considered candidates. We have always set $\theta^* = 135°$; this gives good results over a variety of applications.

For many centre points, the collection of triples eligible under these two criteria is sizable, and can be reduced by placing a limit, NTRIP, on the total number of selected triples. From the eligible set, the NTRIP thinnest (defined in terms of the perpendicular distance from the centre point to the line passing through the two endpoints) are retained. The use of NTRIP imposes an additional collinearity constraint on triples, which becomes more restrictive as the endpoints move away from the centre point. It also reduces the run-time of the full smoothing procedure.

The method discussed above may not identify any eligible triples for data points located at the edges and corners (perimeter) of a map. For such points, triples can be formed by extrapolation from pairs of neighbouring points, as illustrated in Figure 2. Suppose that the observation at $\mathbf{x}_i$ is to be smoothed, and that locations $\mathbf{x}_j$ and $\mathbf{x}_k$ (with observed values y_j and y_k, respectively) are among its NN nearest neighbours. Consider the triangle with vertices at $\mathbf{x}_i$, $\mathbf{x}_j$ and $\mathbf{x}_k$, and let γ be the angle at $\mathbf{x}_j$. If $\gamma \geqslant 90^\circ + \theta^*/2$, then a new triple is formed by linearly extrapolating the trend from y_k through y_j to a new location such that $\mathbf{x}_i$ is equidistant from the new location and $\mathbf{x}_j$. This new location is one endpoint of the created triple, and $\mathbf{x}_j$ is the other. The restriction on γ ensures that the created triple has a centre angle that is greater than or equal to θ^*.

The weight assigned to the extrapolated point is equal to the smaller of the two weights, w_j and w_k. Once a collection of extrapolated triples has been identified for each edge and corner location, smoothing proceeds in the same manner as previously described.

2.4. Data Sets

The data used to illustrate the technique are mortality for lung and prostate cancer for 1988–1992 among White males from the recently-published NCHS *Atlas of United States Mortality*.[8] These causes were chosen because lung cancer has a pronounced ridge whereas prostate cancer rates exhibited a high degree of scatter across the small areas. These data consisted of directly age-adjusted death rates[9] and their variances for 798 geographic units in the contiguous U.S. The geographic unit that was mapped, health service areas (HSA), represents aggregates of counties based on where residents obtain their routine hospital care.[10] All data were mapped as classed choropleth (area shaded) maps using a monochromatic colour scheme; rates were classified into five categories according to quintiles of the presmoothed distribution. We considered the inverse of standard error of the rates as weights when smoothing the mortality data. In addition, simulated data which included edge and spike effects were provided by K. Kafadar for comparison.[5]

3. RESULTS

3.1. Lung Cancer Mortality Data

Plate 1 illustrates how weighted and unweighted head-banging differ in their smoothing characteristics. Each of the maps shown contain the same ranges (cutpoints) for all five classes to allow valid map comparisons. Plate 1(*a*) represents the original lung cancer mortality rate data (prior to smoothing); Plates 1(*b*) and 1(*c*) show the results of smoothing with the unweighted and weighted algorithms, respectively, using parameter values of NN = 12 and NTRIP = 8. An important requirement in smoothing mortality maps, as previously mentioned, is retaining the edges that represent valid data. The pronounced ridge running along the Ohio valley in a northeast-to-southwest direction on the original map (Plate 1(*a*)), is maintained as the map data is smoothed using both the weighted and unweighted head-banging algorithms.

Another feature of both algorithms is a narrowing of the distribution of mortality rates, due to the attenuation of outliers which are typically based on sparse data. Evidence of this distributional change in the lung cancer maps is the smaller range of rates seen for the smoothed maps as compared to the raw data map. Because none of the lung cancer outlier rates is based on large populations, the resulting ranges from both smoothing algorithms are very similar.

While it may not be apparent when viewing a small map of the contiguous U.S., one region that exhibits a difference between weighted and unweighted smoothing is north central Tennessee (Plate 2), which includes the Nashville metropolitan area. This densely-populated region has a small standard error and, hence, a large weight. Without weighting, the Nashville rate is smoothed from the second into the highest (darkest) class because of the influence of high rates in the surrounding areas. With weighting, the Nashville rates remain closer to the original data values because of their relatively high weights as compared to surrounding areas.

A comparison of the perimeter of the U.S. on the original map (Plate 1(*a*)) with either of the smoothed maps (Plates 1(*b*) or 1(*c*)) indicates the moderate changes in the values for each HSA after being smoothed. Thus, unlike many other smoothers that may cause undesirable effects on the edges, head-banging results in areas near the edges retaining their original values to a large extent.

To compare the effects of the degree of smoothing, we applied the unweighted algorithm to the lung cancer mortality data (Plate 3), varying the number of nearest neighbours (NN) and the maximum number of triples (NTRIP) while the minimum triple angle (135°) and the number of iterations (10) were kept constant. Use of the unweighted algorithm allows one to observe the effects of smoothing parameters without the potential confounding effects of different weights. Although there are not major differences between these two smoothed maps, closer inspection in the southeastern U.S. indicates at least three areas where distinctions do exist: east-central Texas, the eastern border between Florida and Georgia, and part of the border between North and South Carolina. These areas remain in the highest class (darkest saturation) when a small number of nearest neighbours (8) and maximum triples (5) are used but move into the second highest class when NN and NTRIP are increased to values of 16 and 11, respectively. Because a majority of the additional HSAs used for this greater degree of smoothing also happen to be in the second highest class, they reduce the three indicated areas into this class.

3.2. Prostate Cancer Mortality Data

In an effort to determine how weighted head-banging smooths data containing a significant amount of noise, we applied the weighted algorithm to prostate cancer data (Plate 4). The raw data indicates the large number of spikes (peaks), which are primarily located in the central part of the U.S. These isolated high rates (spikes) in rural HSAs that are surrounded by lower rate HSAs are often smoothed out when weighted head-banging is applied. This is a good example of the head-banging algorithm drawing a more apparent spatial pattern from scattered raw data.

3.3. Simulated Structure Data

We also examined simulated data to extend the results of Kafadar[5] using the weighted algorithm. The height of a predetermined structure (Figure 3), representing the measured value, y, was generated and shown on a three-dimensional perspective graph. This structure includes a ridge, peak, and depression in known locations of an area representing 86 western U.S. counties. Gaussian error of 10 per cent (normalized by the weight for each respective county) was added to the structure (Plate 5(*a*)) and the result was smoothed using parameter values of 10 and 7 for NN and NTRIP, respectively, for the unweighted (Plate 5(*b*)) and weighted (Plate 5(*c*)) head-banging algorithms. Because standard errors were not available, the weights (Plate 5(*d*)) were approximated by the square root of county population. The unweighted algorithm (Plate 5(*b*)) retains the height of the region towards the back of the ridge, a ‘compromise’ between values at the front of the ridge and values on either side of the back of the ridge. However, weighting reduces the height of this low-weighted region to a level similar to nearby low-weighted areas (Plate 5(*c*)) because the low weights (Plate 5(*d*)) are nearly equal for counties along and on either side of the back portion of the ridge (Figure 3). Another difference between the weighted and unweighted head-banging

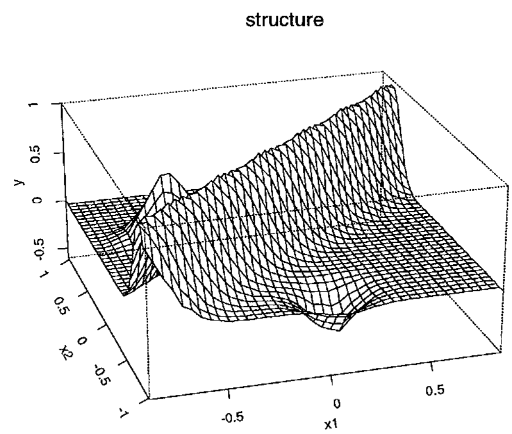

Figure 3. Simulated structure containing a ridge, peak and depression

Table I. Mean squared errors (and SE) between structure and structure plus noise (unsmoothed) and between structure and smoothed structure plus noise for unweighted and weighted head-banging. Results are given at three different noise levels

Condition	Noise level		
	10%	25%	50%
Unsmoothed	12.13 (0.412)	75.80 (2.573)	303.2 (10.29)
Unweighted head-banging	2.994 (0.0536)	6.687 (0.178)	17.66 (0.715)
Weighted head-banging	2.931 (0.0551)	6.129 (0.159)	13.74 (0.584)

results is that the peak is retained to a greater degree (that is, less height reduction) for the weighted algorithm due to the influence of the weights in this region.

The perspective graphs (Plates 5(*a*), (*b*), (*c*) and (*d*)) are also shown as choropleth maps (Plates 5(*e*), (*f*), (*g*) and (*h*), respectively), with the same orientation as the graphs. For comparison, the three maps of structure height have the same colours and cutpoint values for each of the five classes. Categorization of the structure height into five classes masks differences between the unweighted and weighted smoothed maps that are apparent in perspective graphs.

The performance of the unweighted and weighted head-banging smoothing of the structure plus noise were compared by calculating a mean squared error (MSE) between the original structure and the smoothed results, similar to the methods employed by Kafadar.[5] These results, for three different levels of noise, are given in Table I. For comparison, the MSE between the structure and structure plus noise are also included in this table. The weighted algorithm captures

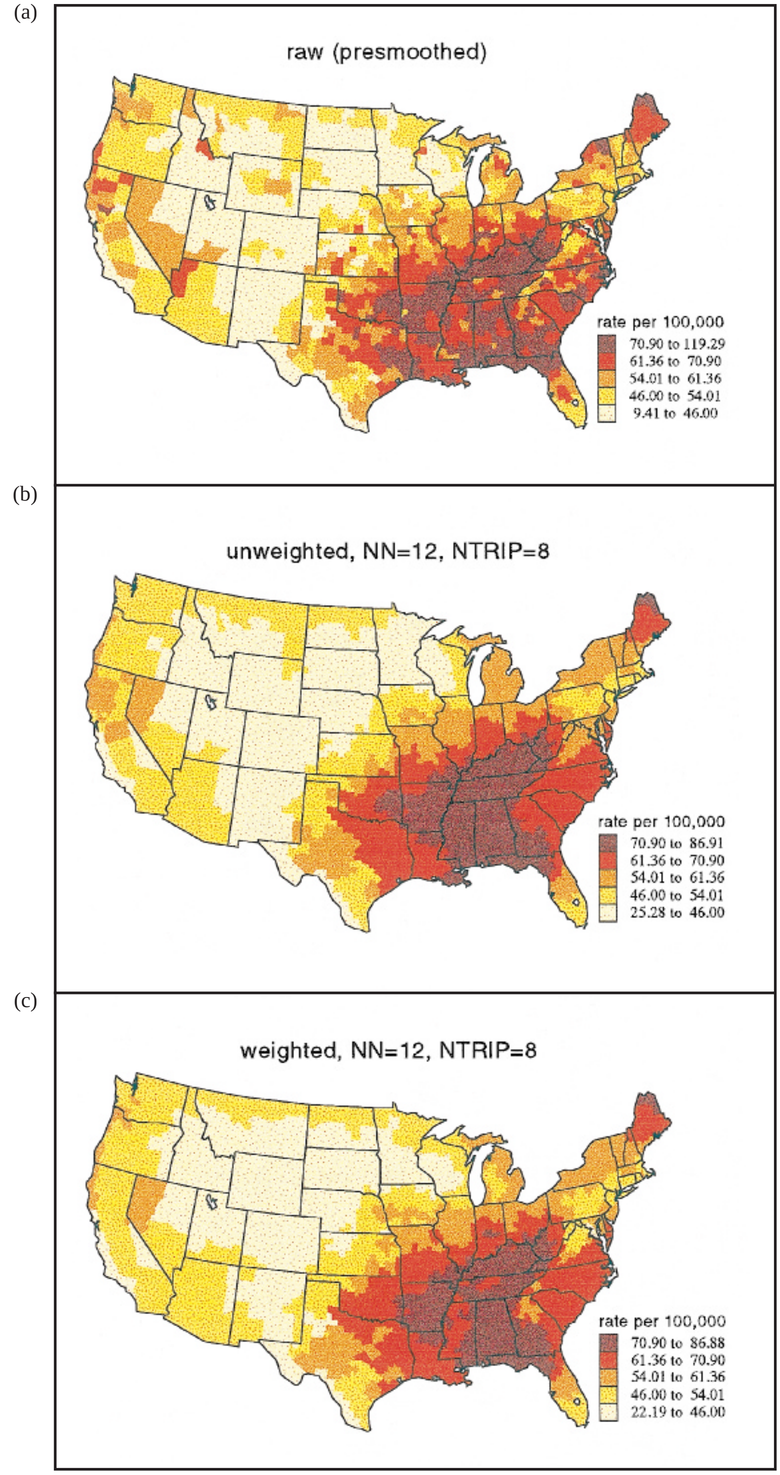

Plate 1. Comparison of weighted and unweighted algorithms for lung cancer mortality data

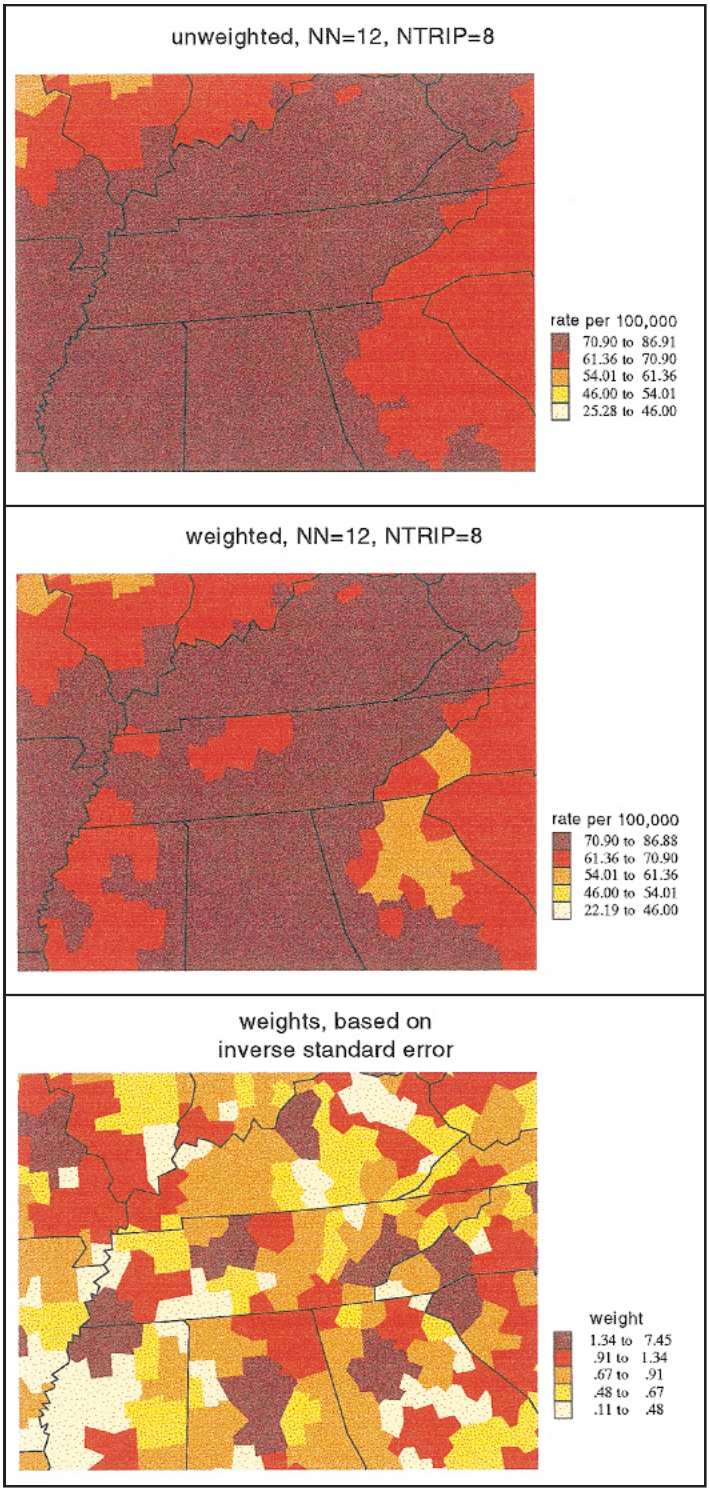

Plate 2. Comparison of the degree of smoothing between weighted and unweighted algorithms for a highly populated area in the vicinity of less populated areas

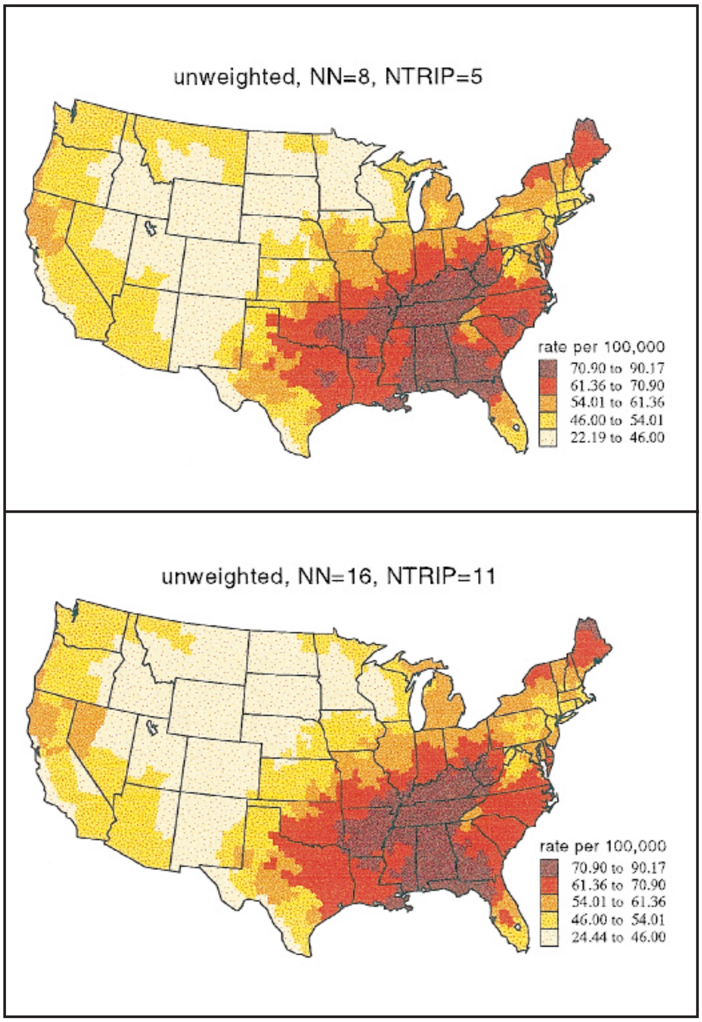

Plate 3. Influence of smoothing parameters for lung cancer mortality data, using unweighted head-banging

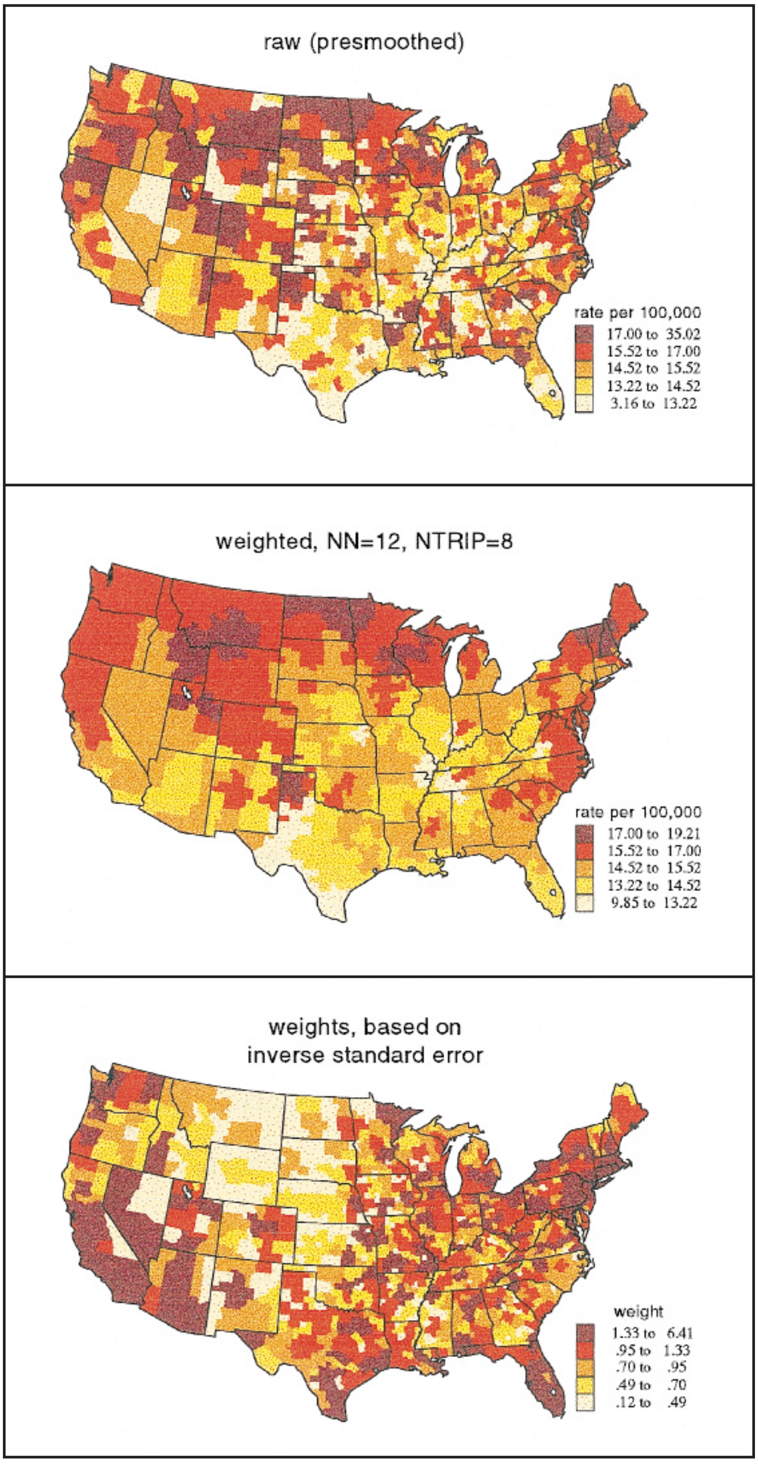

Plate 4. Effect of the weighted algorithm on ‘noisy’ prostate cancer mortality data

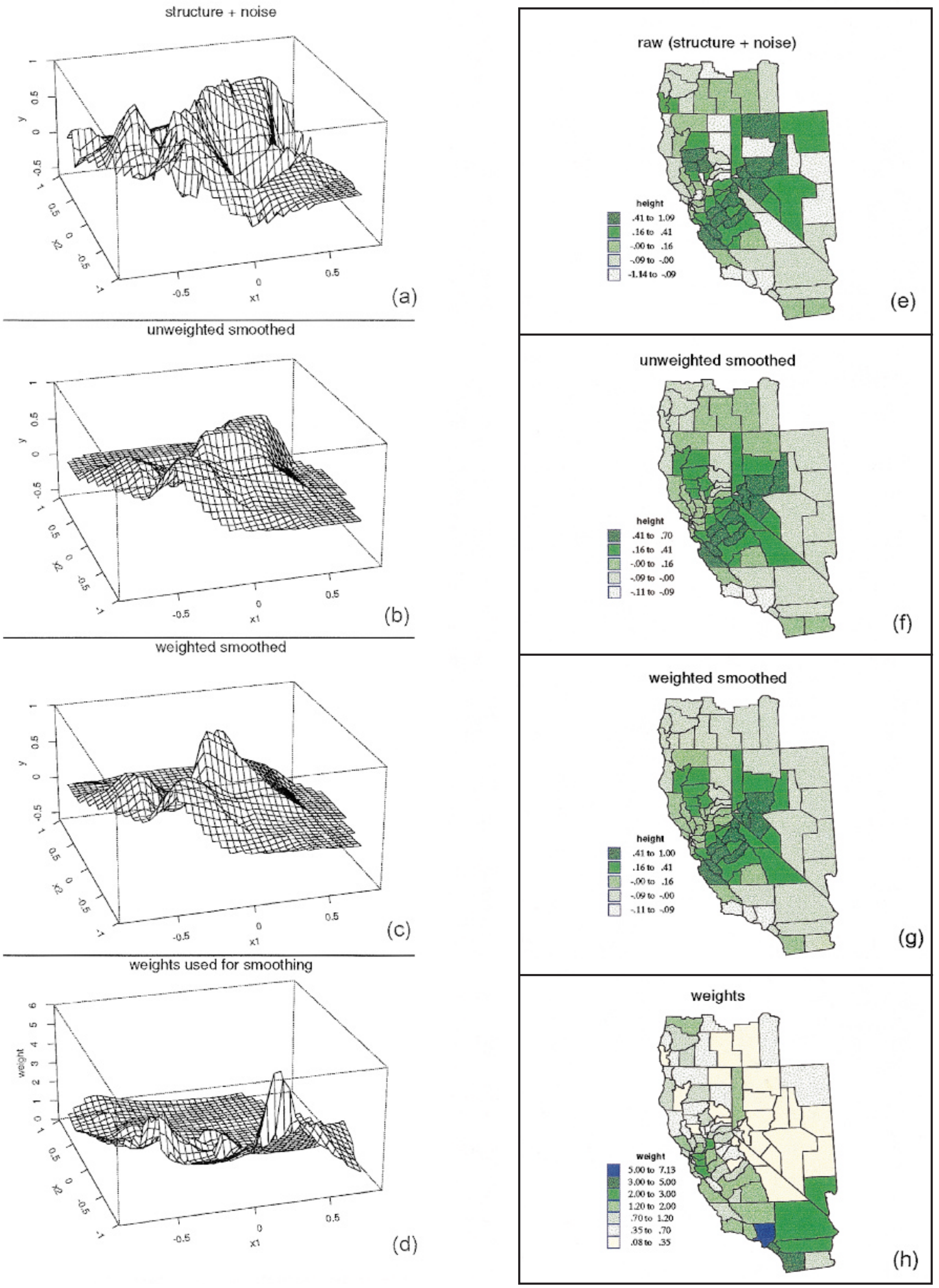

Plate 5. Effect of weighted and unweighted head-banging algorithms when used to smooth a simulated structure with added noise

the original structure better for the highest level of noise, while there is little difference between weighted and unweighted head-banging for the two lower levels of noise.

4. DISCUSSION

These results demonstrate the utility of using the head-banging smoothing algorithm for mortality rate maps, but the conclusions should apply to any choropleth maps of statistical data which are of varying reliability. Application of this algorithm to maps of lung cancer mortality and simulated data suggests that this technique retains edge effects both within the map and along its perimeter. Furthermore, the addition of weighting to the algorithm allows the degree of smoothing to depend on the variability of the mapped statistic. This feature produces greater smoothing of data that are less reliable, while retaining nearly original values of statistics based on large sample sizes.

For the prostate cancer mortality data, the head-banging smoother highlighted existing geographic patterns that were somewhat masked due to the high degree of variability of the rates. Both the weighted and unweighted algorithms removed the spikes of the many high rate HSAs in less populous areas. In contrast, high rates of HIV mortality nearly always occur in urban areas, where the population is large. Unpublished results indicate that these values are maintained with the weighted algorithm while they are often smoothed away for the unweighted case.

The weighted head-banging algorithm was better at recovering the structure from simulated data than the unweighted procedure, due to the reduced influence of the low-weighted counties when smoothing using the weighted algorithm. The improvement of the weighted algorithm ranged from a 22.2 per cent reduction in the MSE with 50 per cent noise, to slightly greater than a 2 per cent reduction for the 10 per cent noise level.

Selection of the appropriate parameters for smoothing these mortality and simulated data was subjective. It should be noted that the optimal parameters for other data sets may deviate from the values we selected and would be based on such factors as the number of data values to be smoothed, the noisiness in the data, and the distribution of the weights. Users should experiment with a number of parameter values to determine the effect of the weighted head-banging smoother on their data. Further research needs to be conducted to obtain general guidelines for choosing optimal parameter values.

In summary, results of applying the weighted head-banging algorithm to both actual and simulated statistical data for small geographic areas indicate its superiority for revealing underlying spatial patterns in the data. Unlike many other proposed smoothers, important features such as edges and spikes based on adequate data are retained, but unreliable data points are smoothed toward adjacent, more reliable values. The addition of weights extends the head-banging algorithm beyond applications where homogeneity of variance may be assumed, such as for maps of death rates. The modified algorithm should prove to be a useful tool for examining mapped data where statistical variability may mask important spatial structure.

REFERENCES

1. Lewandowsky, S. and Behrens, J. T. 'Accuracy of cluster detection in mortality maps', in Pickle, L. W. and Herrmann, D. J. (eds), *Cognitive Aspects of Statistical Mapping*, NCHS Working Paper Series, No. 18, National Center for Health Statistics, Hyattsville, MD, 1995.
2. Hansen, K. M. 'Head-banging: robust smoothing in the plane', *IEEE Transactions on Geoscience and Remote Sensing*, **29**, 369–378 (1991).

3. Velleman, P. F. 'Definition and comparison of robust nonlinear data smoothing algorithms', *Journal of the American Statistical Association*, **75**, 609–615 (1980).
4. Jain, A. K. *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.
5. Kafadar, K. 'Choosing among two-dimensional smoothers in practice', *Computational Statistics and Data Analysis*, **18**, 419–439 (1994).
6. Tukey, J. W. *Exploratory Data Analysis*, Addison-Wesley, Reading, MA, 1977.
7. Tukey, P. A. and Tukey, J. W. 'Graphical display of data sets in 3 or more dimensions', Barnett, V. (ed.), *Interpreting Multivariate Data*, Wiley, New York, 1981.
8. Pickle, L. W., Mungiole, M., Jones, G. K. and White, A. A. *Atlas of United States Mortality*, National Center for Health Statistics, Hyattsville, Maryland, 1996.
9. Fleiss, J. L. *Statistical Methods for Rates and Proportions*, Wiley, New York, 1981.
10. Makuc, D. M., Haglund, B., Ingram, D. D. *et al. Health Service Areas for the United States*, Vital Health Statistics 2(112), National Center for Health Statistics, Hyattsville, MD, 1991.