Sankhyā: The Indian Journal of Statistics
Special Issue on Sample Survey
1999, Volume 61, Series B, Pt. 1, pp.

NONPARAMETRIC REGRESSION FOR ANALYSIS OF COMPLEX SURVEYS AND GEOPGRAPHIC VISUALIZATION*

By GERALD WHITTAKER
Economic Research Service, Washington
and
DAVID W. SCOTT
Rice University, Houston

SUMMARY. The Averaged Shifted Histogram (ASH) provides a very fast, robust method of estimating multivariate nonparametric regressions. The ASH algorithm is fast enough that large (500,000 observations) georeferenced economic and physical data sets can be analyzed interactively. Applications include linking economic models to physical environmental models, analyzing effects of alternative policies on water quality, and examining the effects of agricultural policy on agricultural land values. The results of ASH estimation are conveniently visualized with three-dimensional slices from $\Re^d$ using a geographic information system (GIS).

1. Introduction

This paper presents the results of application of economic and statistical methods to the study of environmental problems. In these applications, the averaged shifted histogram (ASH) approach to density estimation and nonparametric regression was used to visualize the results from economic models, explore the multivariate relationships, and link economic and physical environmental models. In these studies, we faced large data sets (more than 500,000 observations), as well as data from complex stratified surveys.

The application of smoothing techniques to spatial data has attracted some criticism, on the basis that "real" detail is lost. In an application where data points were very densely and accurately sampled throughout the study region

---

AMS (1991) *subject classification.*
Key words and phrases. Mapping; spatial smoothing; averaged shifted histogram; environmental economics.
* Research was supported in part by the National Science Foundation under grant DMS-9626187, the Department of Defense under contract MDA 904-95-C-2203, and USDA cooperative agreement number 43-3AEL-5-80119.

(with digital elevation models, for example), this argument would have some strength. However, survey data by its nature is gathered at widely dispersed or imprecisely located geographic points. The usual practice is to assign the value of each sample point to the area from which the observation was taken. This practice, which results in chloropleth maps, is commonly applied to census divisions, county, and state level data. Brillinger (1990) deplores the use of maps which use a constant value for the whole geographic division. He argues that smoothed data provide a more accurate model of the underlying spatial process, and suggests applying a distance dependent local weighting scheme. In a similar spirit, Cowling, et al. (1996) construct nonparametric regression using a nearest-neighbors algorithm to geographically smooth Australian farm survey data. In a study comparing the more traditional method of analysis of spatial data, kriging, with smoothing techniques, Altman (1997) found comparable performance of the methods. In our experience, even if the data are oversmoothed, large-scale geographic relationships in the data are easily identified. In many applications, such as policy analysis at a national level, a simple interpretation without a high level of unnecessary detail is desirable.

In section 2, we describe the ASH density and regression estimation algorithms. Three real current applications of the ASH are described in section 3: the use of the ASH to explore the determinants of crop-land values; a description of an economic model and the application of the ASH density estimator to results from that model; and, finally, the use of the ASH to link data sets and models in an analysis of environmental policy concerning non-point source pollution. Most of our results computed via the ASH are visualized using powerful geographic information system (GIS) software.

2. **The Averaged Shifted Histogram**

In the ASH algorithm, we operate on binned data. In the simplest application, we are dealing with (x, y, z) data, where (x, y) represents the center of one of our bivariate bins containing one or more primary sampling units. The variables (x, y) are usually some geographic projection of longitude and latitude, respectively. The variable z represents the quantity of interest which has been observed at a known geographic location. For example, z might measure crop land value or nutrient application dollars in farm income. We seek to estimate $\mathrm{E}\left[Z(x,y)\right]$ or $\bar{z}(x, y)$ in areas where the density of observations, $f(x, y)$, is greater than zero.

Let K be a symmetric kernel function with support on $(-1, 1)$ satisfying $\int_{-1}^{1} K(t)\,dt = 1$. Given a positive smoothing parameter h, define the scaled kernel function by $K_h(t) = h^{-1}\,K\left(h^{-1}t\right)$. The trivariate product kernel density estimator (with possibly different smoothing parameters for each dimension) is given by

$$\hat{f}(x,y,z) = \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x-x_i)\, K_{h_y}(y-y_i)\, K_{h_z}(z-z_i)\}, \qquad \ldots(1)$$

and for regression, we mimic the well-known Nadaraya-Watson estimator,

$$\hat{m}(x,y) = \frac{\sum_{i=1}^{n} z_i K_{h_x}(x-x_i) K_{h_y}(y-y_i)}{\sum_{i=1}^{n} K_{h_x}(x-x_i) K_{h_y}(y-y_i)}, \qquad \ldots(2)$$

using the ASH.

Now let us change our notation to reflect the use of the ASH to operate on binned data; then

$$x_1, x_2, \ldots, x_{n_x} \qquad y_1, y_2, \ldots, y_{n_y} \qquad z_1, z_2, \ldots, z_{n_z}$$

are the midpoints along each axis of a trivariate mesh of size $n_x \times n_y \times n_z$ with spacings $\delta_x, \delta_y, \delta_z$. Thus

$$\Delta x_i = \delta_x = \frac{h_x}{m_x} \quad \Delta y_i = \delta_y = \frac{h_y}{m_y} \quad \Delta z_i = \delta_z = \frac{h_z}{m_z}$$

for some integers m_x, m_y, m_z and smoothing parameters h_x, h_y, h_z.

Let $\nu_{jk\ell}$ denote the number of data points $(x,y,z)_i$ falling in bin $B_{jk\ell}$. Note that $\sum \nu_{jk\ell} = n$, and we expect many of the $\nu_{jk\ell}$ to be 0.

The "naive ASH" is constructed by "computing" $m_x \times m_y \times m_z$ (different) trivariate histograms, each with rectangular bin size $h_x \times h_y \times h_z$, but with origins shifted by multiples of $\delta_x, \delta_y, \delta_z$ along the coordinate axes. To be specific, one bin is anchored at the point $(j\delta_x, k\delta_y, \ell\delta_z)$, as j,k,ℓ each range from 0 to $m_x - 1, m_y - 1, m_z - 1$. The ASH density estimator in bin (j,k,ℓ) is a weighted average of the neighboring bin counts, where the weights are given by

$$w_{abc} = \frac{K\left(\frac{a}{m_x}\right) K\left(\frac{b}{m_y}\right) K\left(\frac{c}{m_z}\right)}{\sum_a \sum_b \sum_c K\left(\frac{a}{m_x}\right) K\left(\frac{b}{m_y}\right) K\left(\frac{c}{m_z}\right)}\,; \qquad \ldots(3)$$

here, K is supported on $(-1,1)$ as before, and the indices a,b,c range over $-m_x < a < m_x$, $-m_y < b < m_y$, and $-m_z < c < m_z$. Note that $w_{abc} = w_a w_b w_c$. Furthermore, the weights $\{w_a, w_b, w_c\}$ need only be computed once.

Following the Nadaraya-Watson motivation, the ASH regression estimator is:

$$\hat{m}_{jk} = \frac{\sum_a \sum_b w_{ab} \nu_{j+a,k+b} \bar{z}_{j+a,k+b}}{\sum_a \sum_b w_{ab} \nu_{j+a,k+b}}\,, \qquad \ldots(4)$$

where $w_{ab} = w_a w_b$.

For the survey sampled data, each data point takes the extended form

$$\{(x, y, z, \alpha)_i\,, \quad i = 1, \ldots, n\}\ ,$$

where α_i is the effective sampling weight. Previously, we have assumed that $\alpha_i = 1$ for all cases. In this setting, the frequency counts $\nu_{jk\ell}$ are replaced by the sum of these α_i weights rather than 1's. Furthermore, it is often of interest to include other covariates in the regression analysis, of the form

$$\{(x, y, z, t, \alpha)_i\,, \quad i = 1, \ldots, n\}\ ,$$

where t is some covariate of interest. Then we compute the ASH regression estimator $\hat{m}(x, y, t)$ by simply adding another loop to the numerator and denominator of the $\hat{m}_{jk}$ equation above. The sampling weights are the same of course. Typically, we will map the estimate at several levels of t, for example, $\hat{m}(x, y, t = t_0)$. A fuller derivation of this material may be found in Scott and Whittaker (1996).

The estimates are linear so that variability can be quantified (although difficult to display). We have experimented with a bootstrap method using a resampling procedure for complex surveys (Sitter, 1992). We drew a large number of samples, fit an ASH estimate for each sample, then calculated probability levels at each bin. In geographic applications, the results of this procedure were generally that variance was lower where major features existed. This probably reflects the nature of the survey data used. As a practical matter, our clients, who are primarily policy analysts, staff, or policy makers, have shown no interest in variability of the estimates. It is recognized that all nonparametric methods also have estimation bias, and this is an active and continuing area of research.

We are fortunate in that a great deal is known about these data, and the U.S. Department of Agriculture is a large, easily accessible source of knowledge available to make expert judgments about smoothing choices. This is an important problem, and we discuss it in our applications, and expect further improvements will evolve.

3. **Applications and Data**

Three current applications using the ASH are summarized in this section. A multivariate analysis of agricultural land values demonstrates ASH regression estimation in section 3.1. In section 3.2, the predicted effect of an environmental input tax on farm profits is displayed spatially, by estimation and visualization utilizing the ASH density estimator. Finally, a third application illustrates the use of the ASH to link data sets and to link economic models with physical models of the environment in section 3.3.

The use of georeferenced data and geographic information systems (GIS) is discussed in the multivariate and linking applications. The policy applications (density estimation and linking) use the ASH in conjunction with an economic model of agricultural production. The geographical areas the three applications are concerned with are shown in Figure 1.

Data on the physical environment comprise the largest data sets we used, with one set consisting of over 13 million observations on three variables. From the U.S. Geological Survey (USGS), we used digital elevation models (DEM), records of terrain elevations for ground positions at regularly spaced horizontal intervals; hydrologic unit maps of the United States, which show the boundaries of river basins; and lcc159, a one kilometer grid of land use-land cover for the conterminous United States based on satellite imagery (Loveland, et al., 1991). These data are available on the USGS website.

The STATSGO data set is a digital general soil association map developed by the National Cooperative Soil Survey and distributed by the Natural Resources Conservation Service (formerly Soil Conservation Service) of the U.S. Department of Agriculture. It consists of a broad-based inventory of soils and non-soil areas that occur in a repeatable pattern on the landscape. STATSGO is available on the NRCS website.

Data on population are from the 1990 census at the block group level. These data are available in ARC/INFO export format on the USGS website.

Economic data on farm operations were obtained from the Farm Costs and Returns Survey (FCRS), a survey jointly administered by the National Agricultural Statistics Service (NASS) and the Economic Research Service (ERS) of the U.S. Department of Agriculture. The target population of the FCRS is the official USDA farm population, defined as all establishments that sold or would have sold at least $1,000 of agricultural products during the year. The FCRS is multiple frame survey, with a list frame and an area frame. The list frame is stratified into groups with respect to expenses or commodities. The area frame consists of resident farm operator tracts, which are post-stratified based on value of sales. Both list and area frames are a simple post-stratified sample in all states. The response rate varies by type of farm, *e.g.*, in 1987, farms with control data for potatoes had a response rate of 51, compared to a rate of 85 for farms with control data for tobacco (Dillard, 1987). To deal with non-response, it is assumed that no refusals and inaccessibles were out of business, and all would have reported positive data. A nonresponse adjustment factor based on based on farm type and size is used to overcome the downward bias introduced by nonresponse (Turner, 1992).

The land values were taken from the June Agricultural Survey, a stratified survey representing farmer's estimates of the current market value of specific tracts they operate (Barnard and Westenbarger, 1995). This survey is conducted by NASS. All of the survey data are subject to federal law on confidentiality, and are restricted in availability.

3.1 *Multivariate analysis of land values.* Land values are important to agriculture because land provides the major asset value which supports agricultural production. Visualization and analysis of land values is consequently important in agricultural policy analysis. By linking nonparametric regression to GIS mapping software, we are able to provide a unique perspective on the spatial distribution of farmland values.

In a study of land values in the upper Midwest, we were interested in the effect of population and direct government payments on agricultural land values. Figure 2 displays an ASH estimate of crop land values per acre as a function of longitude and latitude. Even a cursory look at the spatial distribution of land values confirms the common knowledge that land values increase as the urban core is approached (note particularly the area by Chicago). Figure 3 shows an ASH surface of the population density in the study area. The bin size is quite small, and is clearly visible in the figure (since we performed relatively little smoothing). This figure is offered to show the correlation between land value and population density, although a closer examination suggests the relationship differs in different areas.

The first part of the analysis was to re-estimate the value of land conditioned on a rural population density. A four-dimensional ASH estimate of land value as a function of longitude, latitude, and population density was used for this estimation. Given that the data are binned, we are in effect calculating a catagorical regression. In the longitude and latitude dimensions, where greater detail was needed, we used 300 and 250 bins, respectively. For population density, only two bins were used, less than 1,500 and ge 1,500 persons per square mile. The dependent variable, crop land values, is continuous. A three-dimensional slice from the four-dimensional estimation gave land value as a function of longitude ,latitude, and a population density which eliminates or significantly downweights the influence of urban data in our analysis (1,500 persons or less per square mile). This slice is presented in Figure 4, which provides an estimate of crop land values if no urban areas were present. This estimate retains the largest features of the estimate of crop land values. however, highest value drops from \$10,000 to \$4,000 per acre; the large, local peaks due to urban influence have been effectively diminished. The Chicago area remains the highest value, but it turns out that Chicago is also located on potentially the most productive land in the Midwest.

To complete the analysis, we re-estimated land values as a function of longitude, latitude, population density, and direct government payments per acre. Direct government payments are made to farmers under government programs designed to support and stabilize commodity prices. An ASH estimate of direct government payments per acre as a function of longitude and latitude is shown in Figure 5. To calculate the fraction of agricultural land value due to direct government payments, we took a four-dimensional slice of the five-dimensional estimation space where population density was limited to less than 1,500 persons per square mile. Next, we extracted two three-dimensional slices from this subset, where the slices are limited in the government payment dimension to ranges of 9 to 15 dollars per acre and 15 to 21 dollars per acre. These slices represent low and high levels of direct government payments per acre, and their difference is shown in Figure 6. Clearly, land values are positively correlated with the value of government payments. The difference in these two slices divided by land value given a low level of direct government payments results in an estimate of the fraction of land value due to high direct government payments per acre (Figure 7).

The ASH estimation readily lends itself to visualization with GIS software. The results of the estimation provide a value for the center of each bin. These values are easily imported in software such as ARC/INFO or ArcView, and they tile a map. Boundaries and other features available in geographic data bases can then be added to help the interpretation. Our estimates usually extend beyond the boundaries of the study area, as we have chosen not to employ boundary correction techniques (see Scott and Whittaker, 1996). The GIS software is used to clip the edges of the estimate to match the boundaries, on the assumption that the level of the estimate shows no unusual behavior near the edges.

3.2 *Change in distribution of profits.* Policy analysis in production economics is commonly approached by building a model of the production technology, then comparing the effects of different policies on the model. In the applications discussed here, we chose to apply data envelopment examples of application of DEA to agricultural production). analysis (DEA) to model the production technology of farms. DEA is a nonparametric approach in the sense that no assumptions are made about the functional relationships among variables. The solution to a DEA model provides a measure of the "best-practice" technology, *i.e.*, the highest profit that each firm could possibly achieve. The effect of policies can be modeled by additional contraints on the basic DEA model of technology. Examples of application of DEA to agricultural production include evaluation of the decision to participate in government commodity programs (Whittaker, 1994), and interaction of multiple production processes in dairy production (Färe and Whittaker, 1995).

In the DEA model used here, there are $k = 1, \ldots, K$ observations of farms. Each farm uses $x = (x_1, \ldots, x_M) \in \Re^M_+$ inputs to produce $u = (u_1, \ldots, u_N) \in \Re^N_+$ outputs. The observed inputs $x^k = (x^k_1, \ldots, x^k_M)$ and the observed outputs $u^k = (u^k_1, \ldots, u^k_N)$ are used together with the intensity variables $z^k \geq 0$, $k = 1, \ldots, K$, to form the reference technologies. Our basic model is then

$$T = \{(x, u) : u_n \leq \sum_{k=1}^{K} z^k u^k_n, \ n = 1, \ldots, N, \qquad \ldots(5.1)$$

$$x_m \geq \sum_{k=1}^{K} z^k x^k_m, \ n = 1, \ldots, M, \qquad \ldots(5.2)$$

$$\sum_{k=1}^{K} z_k = 1, \; z_k \geq 0, \; k = 1, \ldots, K\}. \qquad \ldots(5.3)$$

Denote input prices by $p^k \in \Re_+^M$ and output prices by $r^k \in \Re_+^N$. Then the profit of farm k is a linear programming problem (subject to constraints (5.1-5.3) with objective function:

$$\pi(r^k, p^k) \;=\; max \sum_{n=1}^{N} r_n^k u_n - \sum_{m=1}^{M} p_m^k x_m \qquad \ldots(6.1)$$

$$s.t. \sum_{k=1}^{K} z^k u_n^k, \; n = 1, \ldots, N, \qquad \ldots(6.2)$$

$$\sum_{k=1}^{K} z^k x_m^k, \; m = 1, \ldots, M, \qquad \ldots(6.3)$$

$$\sum_{k=1}^{K} z_k = 1, \qquad \ldots(6.4)$$

$$z_k \geq 0, \; k = 1, \ldots, K. \qquad \ldots(6.5)$$

To apply an input tax on the Mth input, for example, the objective function (6.1) becomes

$$\pi(r, p, tax) = max \sum_{n=1}^{N} r_n u_n - \sum_{m=1}^{M-1} p_m x_m - x_M(p_M + tax), \qquad \ldots(7)$$

where tax is the tax rate. Taxes on multiple inputs are represented by additional terms in the objective function.

The motivation of this application of the ASH is to investigate the economic effects of mitigation of nitrate pollution in water on the Columbia Plateau. Elevated levels of nitrate have been observed in ground and surface water in that area, and are attributed to agricultural fertilizer (Bortleson, 1991).

Taxes of various sorts have been proposed to mitigate the effects of non-point source pollution from agriculture, including taxes on emissions, inputs, and outputs. Taxes on inputs are by far the easiest to enforce and monitor (Schmutzler and Goulder, 1997). Therefore we chose to analyze an input tax on nitrogen. The effects on farm profits of an imposition of multiple levels of an input tax was modeled in the DEA profit maximizing framework described above. The ASH version of density estimation (Scott, 1992) was used to visualize the distribution of profits under different tax schemes. We used the bivariate product kernel density estimator $\hat{f}(x, y)$ from (1), where x is the percent reduction in profit resulting from an input tax on nitrogen fertilizer and y is farm size (as measured by gross cash farm income) . The distribution of profit reduction from three different levels of taxation are displayed in Figure 8. The density estimates show that all farms are affected, but profits of small and very large farms are reduced proportionately more than medium to large size farms. We conclude from these results that what appears to be an income neutral environmental tax is in fact regressive by farm size.

3.3 *Linking data sets and models.* It is a common situation to find that data sets which ostensibly refer to the same geographic region, are in fact measured at different levels of aggregation and different geographical sub-regions. For example, in the crop land values study discussed above, population density is known at the census block group level, direct government payments at the county level, and crop land values at scattered points. To link these data sets for analysis, we need find a set of points where a value can be calculated for each variable. This process is relatively easy using the ASH regression estimator. One simply bins each data set to the same geographic grid, then operates on the binned data as usual. As noted above, the survey sampling weights are used in the binning procedure, so the associated weights are used in binning variables from different surveys.

In our final application, we take advantage of this ability to link data sets to link economic and physical models to effect a comprehensive environmental policy analysis. Non-point pollution sources from agriculture have been identified as contributing large amounts of nitrogen to both ground and surface water in the Columbia River Basin of Oregon, Washington, and Idaho (Greene et al., 1994). A tax on nitrogen fertilizer application is one of the so-called incentive measures for the reduction of non-point source pollution. In this application, we linked the DEA models of nitrogen input taxation to an environmental model to estimate physical and economic effects of the tax regime.

The Soil and Water Analysis Tool (SWAT) was used to model the nitrogen loadings in water based on results from the economic models. SWAT is a simulation model for large basins of tens of thousands of square kilometers. SWAT is physically based, since calibration is not possible in large, ungaged basins. The components of SWAT include weather, surface runoff, return flow, percolation, crop growth, irrigation, groundwater flow, reach routing, and nutrient and pesticide loading, among other features (Srinivasan and Arnold, 1994; Srinivasan, 1995). It is a multiple equation, deterministic model based on principles of hydrology, plant physisiology, and soil science. SWAT runs in a GIS system and uses raster maps as input data sets. SWAT also provides a very complete set of variables to simulate different management practices. SWAT requires three data sets describing the physical geography of a study area in order to model that area. The data sets are entered into the model in the form of raster maps, which are used by a GIS interface to SWAT. The data used in this application (consisting of several million observations) are land use (Figure 10), soil (Figure 11), and topography (Figure 12). All variables in SWAT are physical, *e.g.* nutrient application, planting date, harvest date, type of machinery, crop variety, and so forth. All policy is exogenous to SWAT, and is modeled by effects on physical inputs.

In order to analyze the environmental effects of policy, we link the economic model (6) with the environmental model SWAT. Note that in the economic model described above in (5) and (6), all variables other than prices are physical: input and output quantities. In fact, equation (5) provides a model of the same technology that SWAT does, in considerably less detail. To analyze a policy, the effect of the policy on profit maximizing behavior is modeled using (6). The economic optimization establishes the level of the input and output variables in response to a given policy, then those levels are used in SWAT to examine the environmental effects.

The innovation supported by the use of the ASH is to link human economic behavior to the model of the physical environment provided by SWAT. SWAT is a spatially distributed model, while practically all economic data are referenced to point locations. A model of the economic effect of a policy using (6) results in point estimates of inputs of physical variables corresponding to variables in SWAT. On the assumption that these points are observations from a surface describing each variable, we estimate an ASH nonparametric regression surface for each variable as a funtion of longitude and latitude. Each ASH surface is then used as an input to SWAT.

Thus, the ASH surface provides a way to geographically distribute point data and link the economic models and apatially distributed physical model. (see for example, Figure ??) for each physical variable. In this way human economic behavior drives the physical model through the ASH surfaces. The input and output data for the economic model is a subset of 1992 expenditure version of the FCRS. The subset consists of 108 observations which are located in the Columbia Basin, have dry-land wheat and/or barley, and received no income from livestock. The soils input map is from the STATSGO data set, the land use and digital elevation input maps are from the U.S. Geological Survey. The state boundaries are a data set included with the ArcView software package, and the hydrologic unit boundaries are from the U.S. Geological Survey. Details about these data sets are in section 3.

The values of the parameters h and δ were chosen using expert opinion. The ASH estimator is very fast, so interactive estimation and visualization with agricultural experts judging the results was quite efficient. The typical estimation process quickly established values where the data was clearly undersmoothed, and clearly oversmoothed. Major features would remain constant in the range between the extremes, and most secondary features varied little. The difference in values between independent choices by experts was always small.

To analyze the nitrate tax policy, we solved the models (6) and (7) for comparison of different levels of taxation. The results from (6) and (7) provided points to estimate surfaces of nutrient applications and yields. The surfaces were used to estimate the mode of each variable in each watershed. The final step was to run SWAT and compare the costs to producers and environmental benefits of a nutrient tax. A tax rate of 300% was about the lowest level of taxation which produced a reduction in nutrient application.

One result from this analysis is that there is a large variation in effective tax burden within the study area. The amount per acre profits are reduced by the imposition of a 300% tax vary from $0 to $142.03 throughout the region (Figure 13). As a measure of the efficiency of the 300% tax in nitrogen reduction, we calculated the cost of a 1 kilogram per hectare reduction in nitrogen reaching the mouth of the watershed (Figure 14). Again, there is large variability, although there seems to some correlation with nitrogen application rate (compare Figure 9 and Figure 14).

4. Conclusion

The averaged shifted histogram has proven to be a versatile, computationally efficient technique in several different applications in environmental economics. The ASH density and nonparametric regression estimators were able to accommodate economic data sets from complex, stratified surveys, with unknown distributions. The ASH made the estimation and visualization of model results possible for previously intractable problems in linking economic behavior to environmental processes. Output from the ASH estimators is suitable for geographic information system software, which provides greatly enhanced visualization for geographically referenced data.

References

ALTMAN, N. (1997). Krige, smooth, both or neither? *American Statistical Association, Proceedings of the Statistical Computing Section.*

BARNARD, C. AND D. WESTENBARGER (1995). *Agricultural Land Values*, AREI UPDATES, No. 17., Economic Research Service, U.S. Department of Agriculture.

BORTLESON, G.C. (1991). *Water fact sheet, National Water Quality Assessment Program Mid-Columbia River Basin, Washington and Idaho.* U.S. Geological Survey Open-File Report 91-164.

BRILLINGER, DAVID R. (1990). Spatial-temporal modelling of spatially aggregate birth data. *Survey Methodology*, **16**, 255-269.

COWLING, ANN, RAY CHAMBERS, RAY LINDSAY, AND BHAMATHY PARAMESWARAN (1996). Applications of spatial smoothing to survey data. *Survey Methodology*, **22**, 175-183.

DILLARD, DAVE (1987). *Weighting Survey Data for Analysis.* National Agricultural Statistics Service, USDA.

FÄRE, ROLF AND GERALD WHITTAKER (1995). An intermediate input model of dairy production using complex survey data. *Journal of Agricultural Economics,* **46**, 201-213.

GREENE, K.E., J.C. EBBERT, AND M.D. MUNN (1994). *Nutrients, Suspended Sediment, and Pesticides in Streams and Irrigation Systems in the Central Columbia Plateau in Washington and Idaho, 1959-1991.* Water-Resources Investigations Report 94-4215, U.S. Geological Survey.

LOVELAND, THOMAS R., JAMES W. MERCHANT, DONALD O. OHLEN, AND JESSLYN F. BROWN (1991). Development of a Land-Cover Characteristics Database for the Conterminous U.S. *Photogrammetric Engineering & Remote Sensing*, **57**, 1453-1463.

RYKER, S.J., AND J.L. JONES (1995). *Nitrate Concentrations in Ground Water of the Central Columbia Plateau.* U.S. Geological Survey Open-File Report 95-445.

SCHMUTZLER, ARMIN, AND LAWRENCE H. GOULDER (1997). The Choice between Emission Taxes and Output Taxes under imperfect monitoring. *Jour. Environmental Economics and Management*, **32**, 51-64.

SCOTT, DAVID W. (1992), *Multivariate Density Estimation.* John Wiley & Sons, Inc., New York.

SCOTT, DAVID W. AND GERALD WHITTAKER (1996). Multivariate applications of the ASH in regression. *Communications in Statistics,* **25**, 2521-2530.

SITTER, R.R. (1992). A resampling procedure for complex survey data. *Jour. Amer. Statist. Assoc.*, **87**, 755-765.

SRINIVASAN, R. AND J.G. ARNOLD (1994). Integration of a Basin-Scale Water Quality Model with GIS. *Water Resources Bulletin*, **30**, 453-462.

SRINIVASAN, R., J.G. ARNOLD, R.S. MUTTIAH, AND P.T. DYKE (1995). Plant and hydrologic simulation for the conterminous U.S. using SWAT and GIS. *Hydrological Science and Technology*, **11**, 160-168.

TURNER, KAY (1992). *Modification of FCRS Nonresponse Adjustment Procedures.* SRB Research Report Number SRB-92-08, National Agricultural Statistics Service, USDA.

WHITTAKER, GERALD (1994). The relation of farm size and government programme benefits: an Application of data envelopment analysis to policy evaluation. *Applied Economics,* **26**, 469-478.

GERALD WHITTAKER
ECONOMIC RESEARCH SERVICE
1800 M ST. NW, #4177
WASHINGTON, DC 20036-5831
e-mail: gerryw2@ag.econ.gov

DAVID W. SCOTT
STATISTICS MS-138
RICE UNIVERSITY
6100 MAIN STREET
HOUSTON, TX 77005-1892
e-mail: scottdw@rice.edu