ENVIRONMETRICS, VOL. 7, 401–416 (1996)

LOCALLY WEIGHTED LEAST SQUARES KERNEL REGRESSION AND STATISTICAL EVALUATION OF LIDAR MEASUREMENTS

ULLA HOLST, OLA HÖSSJER, CLAES BJÖRKLUND
Department of Mathematical Statistics, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden

PÄR RAGNARSON AND HANS EDNER
Department of Atomic Physics, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden

SUMMARY

The LIDAR technique is an efficient tool in monitoring the distribution of atmospheric species of importance. We study the concentration of atmospheric atomic mercury in an Italian geothermal field and discuss the possibility of using recent results from local polynomial kernel regression theory for the evaluation of the derivative of the DIAL curve. A MISE-optimal bandwidth selector, which takes account of the heteroscedasticity in the regression is suggested. Further, we estimate the integrated amount of mercury in a certain area.

KEY WORDS: locally weighted least squares regression; LIDAR measurements; air pollution; atmospheric atomic mercury; geothermal field

1. INTRODUCTION

Optical remote-sensing techniques are of increasing interest in the area of air pollution monitoring due to their many advantages compared to conventional point-monitoring techniques (Killinger and Mooradian 1983; Grant and Menzies 1983; Sigrist 1994; Ragnarson 1994).

In this paper we estimate the emission of mercury from the geothermal power plant Bella Vista in Italy using recently developed statistical techniques. These air pollution measurements have already been presented and extensively discussed in Edner *et al.* (1992) from a physical point of view.

Today the concentration is often straightforwardly estimated by differentiating the DIAL curve with no corresponding estimates of bias and variance for the concentration curve. Further, often the integrated amount of air pollution in a certain area is of special interest, an integral which of course also needs estimates of bias and variance.

Nonparametric regression has become a very important research field as researchers realized that parametric regression is not suitable when fitting curves to data sets in many applications; see the books by Wahba (1990) and Härdle (1990). The local polynomial kernel regression methods are extensively discussed in the book by Wand and Jones (1995).

A different and interesting approach would be to model the regression function stochastically. A physically based model for the concentration field, a random field in time and space, where also the wind field is considered, is a real challenge for future research.

CCC 1180–4009/96/040401–16

Accepted 26 March 1996

However, in this paper we assume that the regression function is unknown and deterministic and we want to point out the advantages of using local polynomial kernel regression theory for the evaluation of LIDAR measurements (LIght Detection And Ranging). In two papers, Ruppert *et al.* (1994, 1995a) have developed the theory, and the proposed statistical technique in this paper is much inspired by these two theoretical papers. However, there are a couple of differences. We have to deal with bandwidth selection in a heteroscedastic regression situation and we also consider bandwidth selectors that are optimal for estimating the derivative of the regression function. Further, our design is fixed and regular, while the basic papers treat a random design.

We really want to emphasize the benefits of the proposed statistical method of evaluating LIDAR measurements. Below follows a summary of the advantages:

(i) The bandwidth selection procedure gives an automatic adaptation to the curvature of the concentration profile and thus calculates the optimal amount of smoothing.
(ii) The variance of the concentration curve is easily calculated since the concentration estimator is a weighted least squares estimator.
(iii) An approximate estimate of bias of the concentration curve can be appropriately calculated based on the estimate of the third derivative of the regression function.
(iv) The concentration curve is easily estimated in arbitrarily chosen points, not necessarily the points of measurements. The method automatically handles non-equispaced data, due to missing values etc.
(v) The variance and the approximate bias of the integrated concentration surface is possible to estimate through straightforward summation. Further, it is possible to prove that the method renders a bias which is negligible, a fact that undoubtedly is of great importance for the application.
(vi) The procedure estimates the variance function in the regression model.

Section 2 deals shortly with the physical background and in Section 3 the statistical model is specified. Section 4 gives a short summary of the local least squares kernel regression method and the used bandwidth selection procedures are presented. Section 5 contains the results for the environmental application.

2. PHYSICAL BACKGROUND

The LIDAR (LIght Detection And Ranging) technique has proven to be an efficient tool in monitoring the distribution of meteorological parameters and several atmospheric species of importance (Measures 1988; Zanzottera 1990; Ragnarson 1994).

The received signal power $P(\lambda, r)$ as function of range r and wavelength λ is described by the single-scattering LIDAR equation

$$P(\lambda, r) = \frac{k(\lambda, r)}{r^2} \beta(\lambda, r) \exp\left\{ -2 \int_0^r (\sigma(\lambda) N(s) + \alpha(\lambda, s)) \, \mathrm{d}s \right\} \tag{1}$$

where $k(\lambda, r)$ is an instrument factor, $\beta(\lambda, r)$ the backscattering coefficient, $N(s)$ the concentration of the studied species at distance s, $\sigma(\lambda)$ the absorption cross-section for wavelength λ and $\alpha(\lambda, s)$ the attenuation due to general scattering and absorption of the aerosol.

The system factor $k(\lambda, r)$ includes parameters such as power and duration of the transmitted laser pulse, telescope area, spectral efficiency, overlap function etc. The temporal profile of the laser pulse is not considered and the influence of the noise has been omitted in the equation.

Furthermore, the effective range resolution is assumed to be small compared with the measurement range. A more detailed physical description is given in Measures (1984).

The DIAL technique (DIfferential Absorption Lidar) employs two different wavelengths, one in resonance with an absorption line of the species of interest and the other off resonance, denoted λ_{on} and λ_{off}, respectively. After differentiating the logarithm of the ratio between the two signals we get

$$\int_0^r N(s)\,ds = -\frac{1}{2\Delta\sigma}\ln\frac{P(\lambda_{on},r)}{P(\lambda_{off},r)} + \frac{1}{2\Delta\sigma}\ln\frac{\beta(\lambda_{on},r)}{\beta(\lambda_{off},r)} - \frac{\Delta\alpha}{\Delta\sigma} \tag{2}$$

where $\Delta\alpha = \int_0^r(\alpha(\lambda_{on},s) - \alpha(\lambda_{off},s))\,ds$ is the differential attenuation due to the general scattering and absorption in the atmosphere and $\Delta\sigma = \sigma(\lambda_{on}) - \sigma(\lambda_{off})$ is the differential absorption cross-section. The instrument factors for the two wavelengths are assumed to be the same.

The basic idea of the DIAL concept is that the second and third term can be neglected if the two wavelengths are close together. In its simplest form the DIAL equation is reduced to

$$N(r) = -\frac{1}{2\Delta\sigma}\frac{\partial}{\partial r}\ln\frac{P(\lambda_{on},r)}{P(\lambda_{off},r)}. \tag{3}$$

The concentration is usually evaluated by differentiating the DIAL curve

$$N_d(r) = -\frac{1}{2\Delta\sigma}\left(\ln\frac{P(\lambda_{on},r+\Delta r)}{P(\lambda_{off},r+\Delta r)} - \ln\frac{P(\lambda_{on},r)}{P(\lambda_{off},r)}\right). \tag{4}$$

One restriction with the DIAL technique is that the return signals are quite often weak, which makes a concentration profile from a single-pulse pair very noisy. Temporal averaging with up to several hundred shots is therefore generally used to improve the signal-to-noise ratio.

3. STATISTICAL MODEL

Laser pulses are emitted and the shots alternate between the two wavelengths λ_{on} and λ_{off}. The available observations corresponding to the different measuring distances, $r_1,\ldots,r_n$, are averages of a number of shots M

$$P(\lambda_{on},r_i) = \frac{1}{M}\sum_{j=1}^{M} P^{(j)}(\lambda_{on},r_i)$$

$$P(\lambda_{off},r_i) = \frac{1}{M}\sum_{j=1}^{M} P^{(j)}(\lambda_{off},r_i)$$

where M is approximately 100. Let

$$Y(r_i) = \ln\frac{P(\lambda_{on},r_i)}{P(\lambda_{off},r_i)}.$$

In Figure 1 the variance function $\nu(r)$ seems to be increasing with the range r. We have chosen to work with the following stochastic model:

$$Y(r_i) = m(r_i) + \nu(r_i)^{1/2}\varepsilon(r_i), \quad i = 1,\ldots,n \tag{5}$$

where $\nu(r_i) = V(Y(r_i))$ and $\varepsilon(r_i)$ is a sequence of independent stochastic variables with expectation zero and variance one.

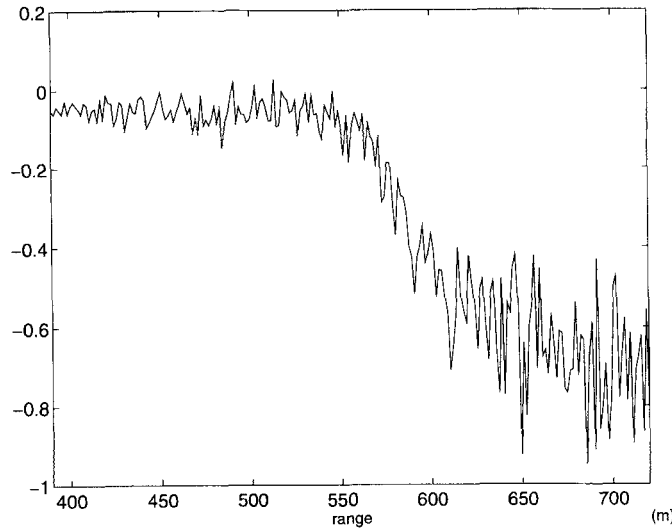

Figure 1. The observations, Vista 186 (direction 2), $y(r_i) = \ln P(\lambda_{on}, r_i)/P(\lambda_{off}, r_i)$

The regression function, $m(r) = E(Y(r))$, is

$$m(r) = -2\Delta\sigma \int_0^r n(s)\,\mathrm{d}s \tag{6}$$

where $\Delta\sigma = \sigma(\lambda_{on}) - \sigma(\lambda_{off})$ and $n(s)$ is the concentration at distance s. (Note that we have chosen to work with different notations for the concentration in the deterministic and the stochastic model, $N(r)$ and $n(r)$, respectively.) Hence we get the derivative

$$\frac{\partial m(r)}{\partial r} = -2\Delta\sigma n(r) \tag{7}$$

and finally the concentration curve

$$n(r) = -\frac{1}{2\Delta\sigma}\frac{\partial m(r)}{\partial r}. \tag{8}$$

As a parametric model for the variance we will use

$$\nu(r) = \exp\{\alpha_0 + \alpha_1 r + \alpha_2 r^2\}. \tag{9}$$

4. LOCALLY-WEIGHTED LEAST-SQUARES KERNEL REGRESSION

This section is closely related to the papers by Ruppert *et al.* (1994, 1995a). They treat the case when the r_i are random variables with a common density $f(r)$, while the LIDAR measurements follow a regular fixed design with 1·5 metres between the measurements. However, it is possible to treat $f(r)$ as a regular distribution over the measurement interval, and for the expression $nf(r)$ in the asymptotic formulas in Ruppert *et al.* (1994, 1995a) to use $n(1{\cdot}5(n-1))^{-1} \approx 2/3$.

Much of our attention will be devoted to the local polynomial kernel regression estimators of $m(r)$, and its derivative $\partial m(r)/\partial r$. Minimize

$$\sum_{i=1}^{n}\{Y(r_i) - \beta_0 - \beta_1(r_i - r) - \cdots - \beta_p(r_i - r)^p\}^2 K_h(r_i - r) \tag{10}$$

with respect to $(\beta_0, \ldots, \beta_p)$ to get estimates of the regression function and its derivatives in the point r. Here $K(u)$ is a kernel function and $K_h(u) = (1/h)K(u/h)$, where $h > 0$ is a bandwidth. Further $\int_{-\infty}^{\infty} K(u)\,du = 1$ and $\int_{-\infty}^{\infty} uK(u)\,du = 0$. This may be repeated over a grid of r-values, not necessarily the measuring points r_i. However, in all computations and illustrations below, we will use the grid of equally spaced measuring points r_i as the grid of r-values.

In the literature several choices of kernel functions are proposed; see the books by Härdle (1990) or Wand and Jones (1995). The properties of the regression estimates depend critically on the bandwidth selector and less on the actual choice of the kernel function as long as it satisfies some regularity conditions. Computational aspects are more important and throughout this paper we will use the standard normal density function as kernel function.

The minimization above is a straightforward weighted least squares problem with solution

$$\begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = (X_{p,r}^{\mathrm{T}} W_r X_{p,r})^{-1} X_{p,r}^{\mathrm{T}} W_r Y, \quad \text{where} \quad X_{p,r} = \begin{bmatrix} 1 & (r_1 - r) & \cdots & (r_1 - r)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (r_n - r) & \cdots & (r_n - r)^p \end{bmatrix}$$

$Y = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ and $W_r = \mathrm{diag}\{K_h(r_1 - r), \ldots, K_h(r_n - r)\}$.

The local least squares estimator of the regression function is

$$\hat{m}(r, h, p) = \hat{\beta}_0 = e_1^{\mathrm{T}} (X_{p,r}^{\mathrm{T}} W_r X_{p,r})^{-1} X_{p,r}^{\mathrm{T}} W_r Y \tag{11}$$

where e_j is a $(p+1) \times 1$ vector having 1 in the jth entry and all other entries 0. In the same way the derivative is estimated with

$$\widehat{\frac{\partial m(r)}{\partial r}} = \hat{m}_1(r, h, p) = \hat{\beta}_1 = e_2^{\mathrm{T}} (X_{p,r}^{\mathrm{T}} W_r X_{p,r})^{-1} X_{p,r}^{\mathrm{T}} W_r Y. \tag{12}$$

Below we use $p = 1$ when estimating the regression function and $p = 2$ for the derivative. More generally, a polynomial degree one unit greater than the order of the derivative gives a simple expression for the bias (Ruppert and Wand 1994).

4.1. Bandwidth selection

Remember that

$$\hat{m}(r, h, 1) = \hat{\beta}_0 = e_1^{\mathrm{T}} (X_{1,r}^{\mathrm{T}} W_r X_{1,r})^{-1} X_{1,r}^{\mathrm{T}} W_r Y \tag{13}$$

$$\hat{m}_1(r, h, 2) = \hat{\beta}_1 = e_2^{\mathrm{T}} (X_{2,r}^{\mathrm{T}} W_r X_{2,r})^{-1} X_2^{\mathrm{T}} W_r Y. \tag{14}$$

The variances follow straightforwardly as

$$V(\hat{m}(r, h, 1)) = e_1^{\mathrm{T}} (X_{1,r}^{\mathrm{T}} W_r X_{1,r})^{-1} X_{1,r}^{\mathrm{T}} W_r V_y W_r X_{1,r} (X_{1,r}^{\mathrm{T}} W_r X_{1,r})^{-1} e_1 \tag{15}$$

$$V(\hat{m}_1(r, h, 2)) = e_2^{\mathrm{T}} (X_{2,r}^{\mathrm{T}} W_r X_{2,r})^{-1} X_{2,r}^{\mathrm{T}} W_r V_y W_r X_{2,r} (X_{2,r}^{\mathrm{T}} W_r X_{2,r})^{-2} e_2 \tag{16}$$

where $V_y = \mathrm{diag}\{\nu(r_1), \ldots, \nu(r_n)\}$.

Ruppert and Wand (1994) (Theorem 4.1 and 4.2) have derived formulas for the asymptotic bias and variance. When $h \to 0$ and $nh \to \infty$ we get for the estimate of the regression function

$$E(\hat{m}(r, h, 1)) - m(r) \approx \frac{m^{(2)}(r)}{2} h^2 \int_{-\infty}^{\infty} u^2 K(u)\,du \tag{17}$$

$$V(\hat{m}(r, h, 1)) \approx \frac{\nu(r)}{nhf(r)} \int_{-\infty}^{\infty} K^2(u)\,du. \tag{18}$$

Further, for the estimate of the derivate of the regression function, we have when $h \to 0$ and $nh^3 \to \infty$

$$E(\hat{m}_1(r,h,2)) - m(r) \approx \frac{m^{(3)}(r)}{6} h^2 \int_{-\infty}^{\infty} u^4 K(u)\,\mathrm{d}u \cdot \left(\int_{-\infty}^{\infty} u^2 K(u)\,\mathrm{d}u \right)^{-1} \tag{19}$$

$$V(\hat{m}_1(r,h,2)) \approx \frac{\nu(r)}{nh^3 f(r)} \int_{-\infty}^{\infty} u^2 K^2(u)\,\mathrm{d}u \cdot \left(\int_{-\infty}^{\infty} u^2 K(u)\,\mathrm{d}u \right)^{-2}. \tag{20}$$

Hence, with the standard normal kernel function, MISE (mean integrated square error) with weight function $f(r)$ can be approximated by

$$\mathrm{MISE}(\hat{m}(r,h,1)) \approx \frac{1}{nh} \frac{1}{2\sqrt{\pi}} \int_S \nu(r)\,\mathrm{d}r + \frac{h^4}{4} \int_{-\infty}^{\infty} \left(m^{(2)}(r)\right)^2 f(r)\,\mathrm{d}r \tag{21}$$

$$\mathrm{MISE}(\hat{m}_1(r,h,2)) \approx \frac{1}{nh^3} \frac{1}{4\sqrt{\pi}} \int_S \nu(r)\,\mathrm{d}r + \frac{h^4}{4} \int_{-\infty}^{\infty} \left(m^{(3)}(r)\right)^2 f(r)\,\mathrm{d}r \tag{22}$$

where S is the measuring interval. The two bandwidths, which we are using, h_{MISE} and h_{MISE_1}, are chosen to minimize the approximative mean integrated square errors

$$h_{\mathrm{MISE}} = \left[\frac{\int_S \nu(r)\,\mathrm{d}r}{2\sqrt{\pi} \int_{-\infty}^{\infty} \left(m^{(2)}(r)\right)^2 n f(r)\,\mathrm{d}r} \right]^{1/5} \tag{23}$$

$$h_{\mathrm{MISE}_1} = \left[\frac{3 \int_S \nu(r)\,\mathrm{d}r}{4\sqrt{\pi} \int_{-\infty}^{\infty} \left(m^{(3)}(r)\right)^2 n f(r)\,\mathrm{d}r} \right]^{1/7}. \tag{24}$$

However, the integrals $\int_{-\infty}^{\infty} \left(m^{(2)}(r)\right)^2 f(r)\,\mathrm{d}r$, $\int_{-\infty}^{\infty} \left(m^{(3)}(r)\right)^2 f(r)\,\mathrm{d}r$ and $\int_S \nu(r)\,\mathrm{d}r$ are unknown and need to be estimated. The first two integrals can be expressed in a general form, see Ruppert *et al.* (1995a)

$$\theta_{q,s} = \int_{-\infty}^{\infty} m^{(q)}(r) m^{(s)}(r) f(r)\,\mathrm{d}r, q, s > 0 \text{ and } q + s \text{ even} \tag{25}$$

which can be estimated straightforwardly as

$$\hat{\theta}_{q,s}(g) = \frac{1}{n} \sum_{i=1}^{n} \hat{m}_q(r_i, g, p) \hat{m}_s(r_i, g, p) \tag{26}$$

where

$$\hat{m}_q(r,g,p) = q! e_{q+1}^{\mathrm{T}} (X_{p,r}^{\mathrm{T}} W_r X_{p,r})^{-1} X_{p,r}^{\mathrm{T}} W_r Y \tag{27}$$

and $W_r = \mathrm{diag}\{K_g(r_1 - r), \ldots, K_g(r_n - r)\}$. In our cases $q = s = 2$ or $q = s = 3$ and we will use $p = 3$ and $p = 4$, respectively. The MSE-optimal bandwidth for $\hat{\theta}_{22}(g)$ and $\hat{\theta}_{33}(g)$ are found by minimizing the squared bias, see Ruppert *et al.* (1995a) for details, to obtain

$$g_{\mathrm{MSE}} = C_2(K) \left[\frac{\int_S \nu(r)\,\mathrm{d}r}{|\theta_{24}| n} \right]^{1/7} \tag{28}$$

and

$$g_{\mathrm{MSE}_1} = C_{23}(K)\left[\frac{\int_S \nu(r)\,\mathrm{d}r}{|\theta_{35}|n}\right]^{1/9} \tag{29}$$

where

$$C_2(K) = \begin{cases} C_2^I(K) = \left(\dfrac{3}{8\sqrt{\pi}}\right)^{1/7} & \text{if } \theta_{24} < 0 \\ C_2^{II}(K) = \left(\dfrac{15}{16\sqrt{\pi}}\right)^{1/7} & \text{if } \theta_{24} > 0 \end{cases}$$

and

$$C_{23}(K) = \begin{cases} C_{23}^I(K) = \left(\dfrac{15}{16\sqrt{\pi}}\right)^{1/9} & \text{if } \theta_{35} < 0 \\ C_{23}^{II}(K) = \left(\dfrac{105}{32\sqrt{\pi}}\right)^{1/9} & \text{if } \theta_{35} > 0. \end{cases}$$

The bandwidths g_{MSE} and g_{MSE_1} depend on θ_{24} and θ_{35}, which, if estimated in the same way as θ_{22}, require a new bandwidth etc., and theoretically the procedure never ends. Hence we use an *ad hoc* procedure when estimating θ_{24} and θ_{35}. A fifth-order polynomial is fitted, i.e.

$$m^Q(r) = q_0 + q_1 r + q_2 r^2 + q_3 r^3 + q_4 r^4 + q_5 r^5. \tag{30}$$

Assuming that $\{\varepsilon(r_i)\}$ is a sequence of Gaussian distributed stochastic variables, the maximum likelihood method provides estimates of $q_0, \ldots, q_5$ and $\alpha_0, \alpha_1, \alpha_2$, so that

$$\begin{aligned} \hat{\theta}_{24}^Q &= \frac{1}{n}\sum_{i=1}^{n} (\hat{m}^Q)^{(2)}(r_i)(\hat{m}^Q)^{(4)}(r_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} (2\hat{q}_2 + 6\hat{q}_3 r_i + 12\hat{q}_4 r_i^2 + 20\hat{q}_5 r_i^3)(24\hat{q}_4 + 120\hat{q}_5 r_i) \end{aligned} \tag{31}$$

$$\begin{aligned} \hat{\theta}_{35}^Q &= \frac{1}{n}\sum_{i=1}^{n} (\hat{m}^Q)^{(3)}(r_i)(\hat{m}^Q)^{(5)}(r_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} (6\hat{q}_3 + 24\hat{q}_4 r_i + 60\hat{q}_5 r_i^2)120\hat{q}_5 \end{aligned} \tag{32}$$

and $\hat{\nu}(r) = \exp\{\hat{\alpha}_0 + \hat{\alpha}_1 r + \hat{\alpha}_2 r^2\}$.

The other unknown quantity in (28) and (29) is $\int_S \nu(r)\,\mathrm{d}r$, which now can be estimated as $\int_S \hat{\nu}(r)\,\mathrm{d}r$ and finally the estimates of g_{MSE} and g_{MSE_1} are

$$\hat{g}_{\mathrm{MSE}} = C_2(K)\left[\frac{\int_S \hat{\nu}(r)\,\mathrm{d}r}{|\hat{\theta}_{24}^Q|n}\right]^{1/7} \tag{33}$$

$$\hat{g}_{\mathrm{MSE}_1} = C_{23}(K)\left[\frac{\int_S \hat{\nu}(r)\,\mathrm{d}r}{|\hat{\theta}_{35}^Q|n}\right]^{1/9}. \tag{34}$$

Summarizing, the bandwidths we are using for the estimate of the regression function and its derivative are

$$\hat{h}_{\mathrm{MISE}} = \left[\frac{\int_S \hat{\nu}(r)\,\mathrm{d}r}{2\sqrt{\pi}\hat{\theta}_{22}(\hat{g}_{\mathrm{MSE}})} \right]^{1/5} \tag{35}$$

$$\hat{h}_{\mathrm{MISE}_1} = \left[\frac{3\int_S \hat{\nu}(r)\,\mathrm{d}r}{4\sqrt{\pi}\hat{\theta}_{33}(\hat{g}_{\mathrm{MSE}_1})} \right]^{1/7}. \tag{36}$$

Finally, for classification of the fairly involved bandwidth selection technique above, we summarize the procedure in a couple of steps below. The theoretical background is extensively treated in the paper by Ruppert *et al.* (1995a):

(i) Compute preliminary maximum likelihood estimates of $q_0, \ldots, q_5$ and $\alpha_0, \alpha_1, \alpha_2$ followed by $\hat{\theta}_{24}^{Q}$, $\hat{\theta}_{35}^{Q}$ and $\hat{\nu}(r)$ (equations (31) and (32)).
(ii) Compute $\hat{g}_{\mathrm{MSE}}$ and $\hat{g}_{\mathrm{MSE}_1}$ (equations (33) and (34)).
(iii) Compute $\hat{\theta}_{22}(\hat{g}_{\mathrm{MSE}})$ and $\hat{\theta}_{33}(\hat{g}_{\mathrm{MSE}_1})$ (equation (26) with $q = s = 2$, $p = 3$ and $q = s = 3$, $p = 4$, respectively).
(iv) Compute $\hat{h}_{\mathrm{MISE}}$ and $\hat{h}_{\mathrm{MISE}_1}$ (equations (35) and (36)).

5. RESULTS

5.1. The measurements

The observations we are dealing with come from measurements of atmospheric atomic mercury at the geothermal power plant Bella Vista in Italy. The measurements are obtained for different angles in a vertical plane through the mercury plume from the cooling tower. The number of observations in each direction is 1000 which corresponds to 1500 m. The mercury plume is located between 300 m and 700 m from the measuring equipment.

Here we will study the same vertical plane at three timepoints, Vista 186–188, see Table I. A more extensive presentation of the results is given in Björklund (1994).

5.2. The concentration curve

In this section we discuss the estimate of the concentration curve and its properties. As an example we use the second line in the first plane, see Figure 10 and Table I. The estimate of the concentration is, using (8) and (12)

$$\hat{n}(r) = C\hat{m}_1(r, \hat{h}_{\mathrm{MISE}_1}, 2)$$

where for mercury

$$C = -\frac{1}{2\Delta\sigma} = -\tfrac{1}{16}10^6\ \mathrm{ng/m^2}.$$

As an estimate of bias and variance for the concentration curve we will use

$$\hat{E}(\hat{n}(r)) - n(r) \approx \frac{C}{2}\hat{m}_3(r, \hat{h}_{\mathrm{MISE}_3}, 4) \tag{37}$$

Table 1. The mercury flux at three adjacent time points

	390–720 m (221 observations)					Interval 405–675 m		
	(j)	φ_j	$\hat{h}_{\mathrm{MISE}}$, m	$\hat{h}_{\mathrm{MISE}_1}$, m	$\hat{h}_{\mathrm{MISE}_3}$, m	$\hat{I}$, mg/m	$\hat{D}(\hat{I})$, mg/m	Bias, mg/m
Vista 186	1	0	25·5	28·5	48·9	2·47	0·38	−0·011
(14:22)	2	1·06	15·7	19·6	37·9			
	3	2·12	10·2	12·6	36·3			
	4	3·18	6·1	8·9	34·2			
	5	4·24	18·5	18·3	40·5			
	6	5·30	16·9	16·9	40·3			
	7	6·36	8·5	12·3	44·2			
	8	7·42	20·4	17·4	48·7			
	9	8·48	10·3	13·2	38·4			
	10	9·54	21·5	26·6	64·7			
Vista 187	1	0	35·6	42·7	85·6	2·43	0·51	0·053
(14:54)	2	1·06	12·5	15·5	36·1			
	3	2·12	6·7	9·6	31·6			
	4	3·18	16·2	22·6	64·7			
	5	4·24	13·3	14·8	44·3			
	6	5·30	21·3	21·6	53·4			
	7	6·36	9·4	12·9	42·8			
	8	7·42	12·0	14·4	38·5			
	9	8·48	7·0	9·8	40·6			
	10	9·54	6·6	9·8	40·7			
Vista 188	1	0	15·1	16·5	44·9	2·55	0·41	0·007
(15:22)	2	0·742	16·5	23·8	55·2			
	3	1·484	13·1	15·7	42·3			
	4	2·226	16·8	19·5	47·4			
	5	2·968	15·2	18·7	40·6			
	6	3·710	19·5	20·5	40·7			
	7	4·452	10·8	14·4	44·1			
	8	5·194	14·6	16·7	47·1			
	9	5·936	14·1	14·5	40·4			
	10	6·678	21·1	25·8	59·8			

$$\hat{V}(\hat{n}(r)) = C^2 \hat{V}(\hat{m}_1(r, \hat{h}_{\mathrm{MISE}_1}, 2))$$
$$= C^2 e_2^{\mathrm{T}} (X_{2,r}^{\mathrm{T}} W_r X_{2,r})^{-1} X_{2,r}^{\mathrm{T}} W_r \hat{V}_y W_r X_{2,r} (X_{2,r}^{\mathrm{T}} W_r X_{2,r})^{-1} e_2 \tag{38}$$

where $\hat{V}_y = \mathrm{diag}\{\hat{\nu}(r_1), \ldots, \hat{\nu}(r_n)\}$. The estimate of bias of the concentration curve depends on the estimate of the third derivative of the regression function, $\hat{m}_3(r, h, 4)$, where the bandwidth h has to be determined. We have chosen to use the bandwidth which minimizes $\mathrm{MISE}(\hat{m}_3(r, h, 4))$, which in turn can be approximated by

$$\mathrm{MISE}(\hat{m}_3(r, h, 4)) \approx \frac{1}{nh^7} \frac{15}{16\sqrt{\pi}} \int_S \nu(r)\, \mathrm{d}r + \frac{h^4}{4} \int_{-\infty}^{\infty} \left(m^{(5)}(r)\right)^2 f(r)\, \mathrm{d}r \tag{39}$$

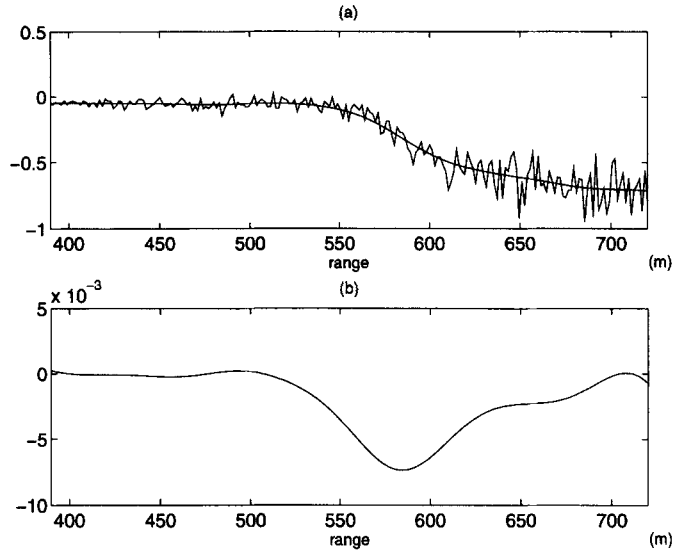

Figure 2. (*a*) The estimated regression function and the observations. Optimal bandwidth 15·7 m. (*b*) The estimated derivative of the regression function. Optimal bandwidth 19·6 m

which is minimized by

$$\hat{h}_{\mathrm{MISE}_3} = \left[\frac{105 \int_S \hat{v}(r)\,\mathrm{d}r}{16\sqrt{\pi} \int_{-\infty}^{\infty} \left((\hat{m}^{\mathrm{Q}})^{(5)}(r)\right)^2 n f(r)\,\mathrm{d}r} \right]^{1/11} \tag{40}$$

where as above $(\hat{m}^{\mathrm{Q}})^{(5)}(r) = 120\hat{q}_5$.

Figure 2 shows the regression function ($p = 1, \hat{h}_{\mathrm{MISE}} = 15{\cdot}7$) and the derivative of the regression function ($p = 2, \hat{h}_{\mathrm{MISE}_1} = 19{\cdot}6$) and Figure 3 the fitted fifth-order polynomial and the estimated variance $\hat{v}(r)$. The normalized residuals

$$(y_i - \hat{m}(r_i, \hat{h}_{\mathrm{MISE}}, 1))/\sqrt{\{\hat{v}(r_i)\}}$$

and the corresponding estimated correlation function are shown in Figure 4. Conveniently, the normalized residuals seem to behave almost as a sequence of white noise.

The concentration curve with its estimates of bias and variance is shown in Figure 5. As expected, bias is negligible when the concentration curve is almost linear, while the variance is increasing with range.

The estimate of the concentration curve is dependent on the bandwidth selection, which determines the optimal amount of smoothing. For comparison the concentration curves for a smaller and a larger value of h, $h = 10$ and $h = 40$, respectively, are calculated and shown in Figures 6 and 7. The corresponding bias and variance curves are also displayed. *Note that too small bandwidths give rise to an extremely large variance curve, while too large bandwidths render a disastrous bias.*

Further, in Figure 8 $\mathrm{MISE}((\hat{m}_1(r, h, 2))$ is calculated as a function of h and the minimum is not especially distinct, which shows that the bandwidth selection procedure is fairly insensitive as long as the bandwidth is kept inside the interval 15–25.

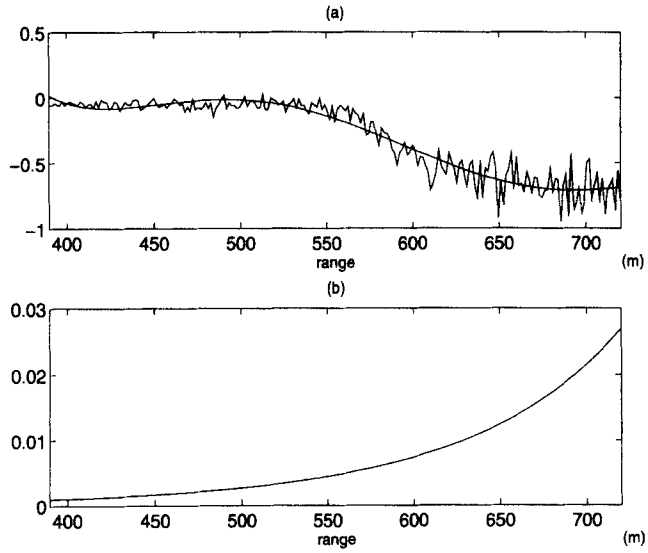

Figure 3. (*a*) The fitted fifth-order polynomial and the observations. (*b*) The estimated variance function, $\hat{v}(r)$

5.3. The mercury flux

So far in Section 5 we have separately considered measurements along single lines. More interesting of course is to estimate the concentration in different planes through the mercury plume; see Edner *et al.* (1992) for a detailed description of the LIDAR system and the geothermal power plant.

Here the measurements $n(r_i, \varphi_j)$ are obtained for ten different angles φ_j through the mercury plume, see Table I, and hence the grid of measurement points in the vertical plane can be

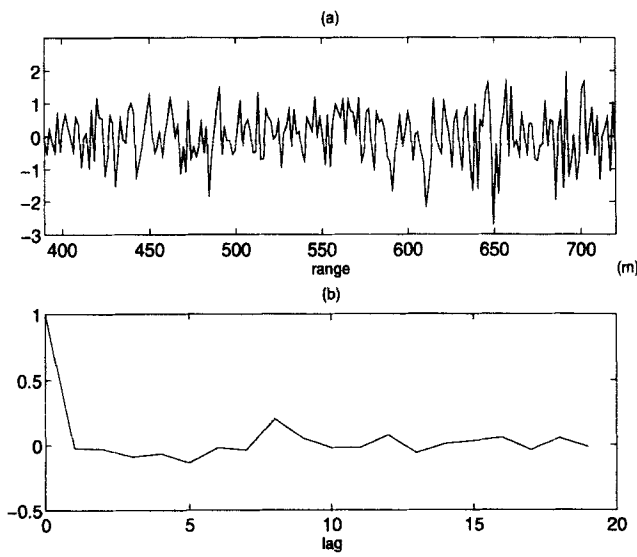

Figure 4. (*a*) The normalized residuals. (*b*) The estimated correlation function

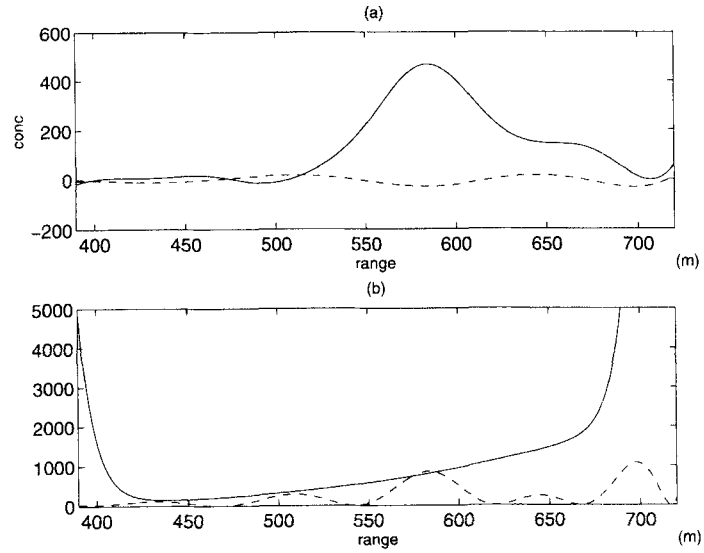

Figure 5. (*a*) The concentration (ng/m^3) with bias (– – –) (*b*) Variance and squared bias (– – –) (ng^2/m^6). Optimal bandwidth 19·6 m

described by polar co-ordinates, i.e.

$$\begin{cases} x_{ij} = r_i \cos \varphi_j \\ y_{ij} = r_i \sin \varphi_j. \end{cases}$$

The directions along which measurements are made are shown in Figure 9 and the estimated and interpolated concentration surface in Figure 10.

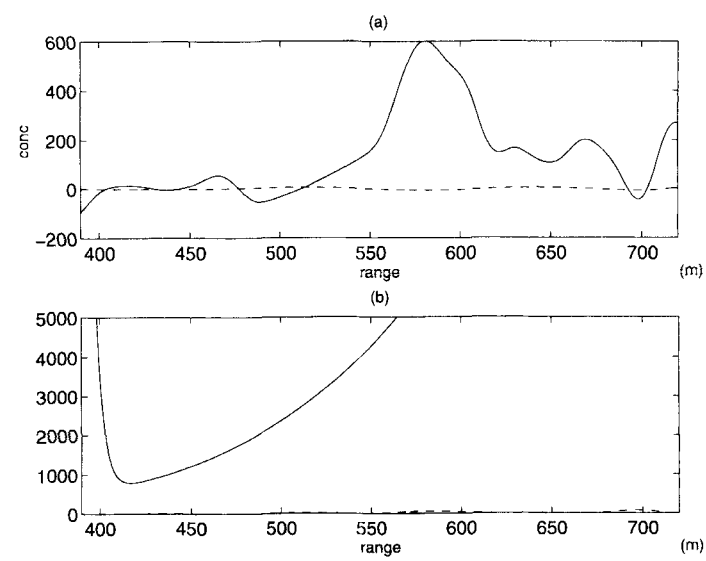

Figure 6. (*a*) The concentration (ng/m^3) with bias (– – –) (*b*) Variance and squared bias (– – –) (ng^2/m^6). Bandwidth too small (10 m)

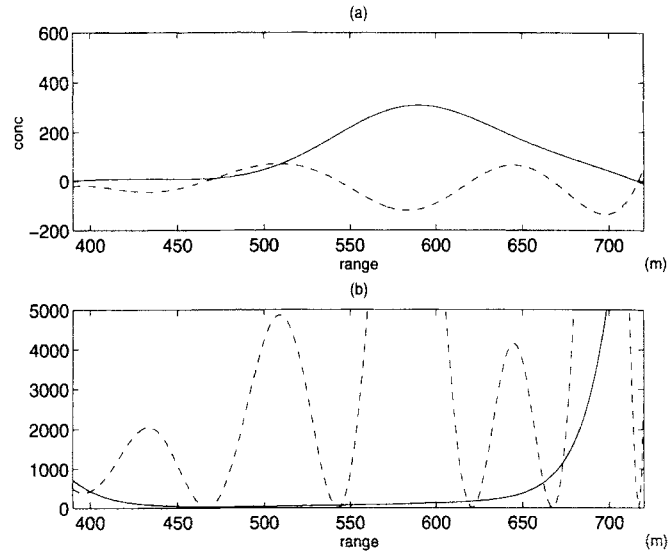

Figure 7. (*a*) The concentration (ng/m^3) with bias (– – –) (*b*) Variance and squared bias (– – –) (ng^2/m^6). Bandwidth too large (40 m)

5.4. The mercury flux

Of great interest is the volume under the concentration surface, since, after multiplication by the wind speed, it is possible to estimate the total flux of mercury passing the plane of measurements.

Assume that the concentration in x/y co-ordinates is given by $f(x, y)$ with the integral

$$I = \int_x \int_y f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int_r \int_\varphi f(r\cos\varphi, r\sin\varphi) r\,\mathrm{d}r\,\mathrm{d}\varphi = \int_r \int_\varphi n(r, \varphi) r\,\mathrm{d}r\,\mathrm{d}\varphi. \tag{41}$$

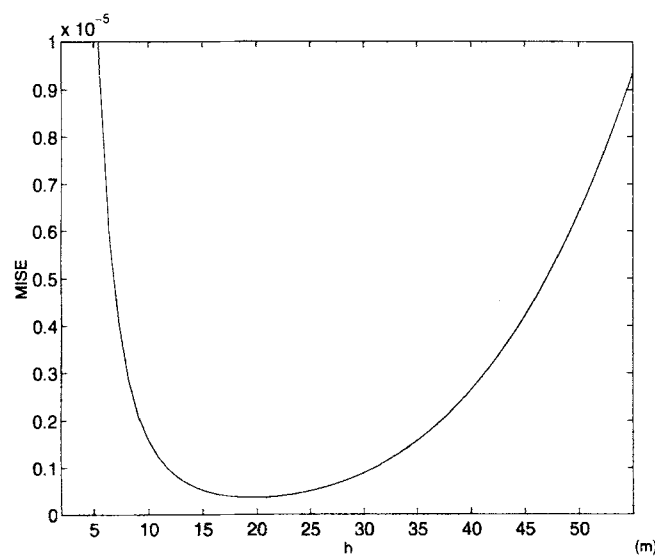

Figure 8. The mean integrated square error as a function of h

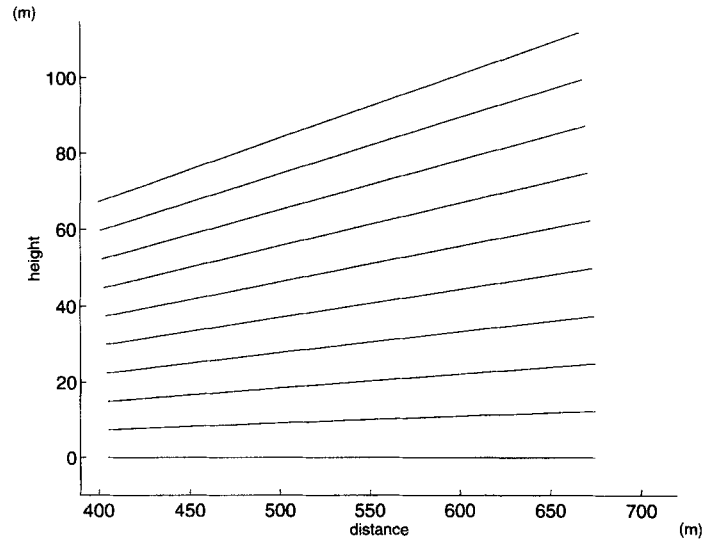

Figure 9. The area where measurements are obtained

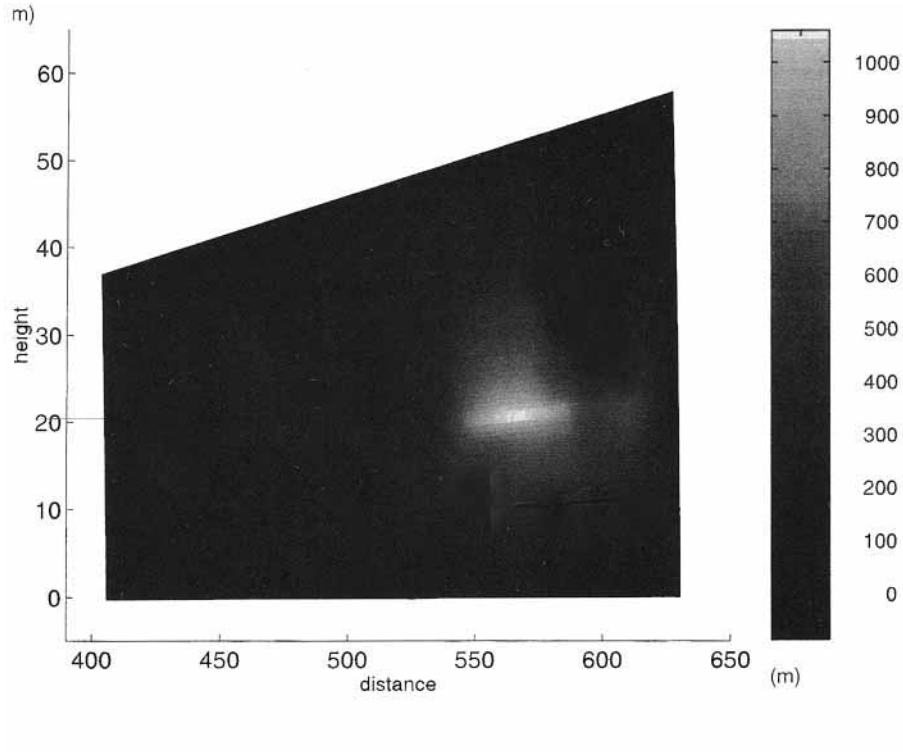

Figure 10. The concentration surface

As an approximation of the integral we will use

$$\hat{I} \approx \sum_{i=1}^{n_r} \sum_{j=1}^{n_\varphi} v_{ij} r_i \hat{n}(r_i, \varphi_j) \Delta r_i \Delta \varphi_j \tag{42}$$

where $\Delta r_i = r_{i+1} - r_i = \Delta r$, $\Delta \varphi_j = \varphi_{j+1} - \varphi_j = \Delta \varphi$ and v_{ij} is a weight function with values

$$v_{ij} = \begin{cases} \frac{1}{4} & \text{in corners} \\ \frac{1}{2} & \text{on boundaries (not corners)} \\ 1 & \text{otherwise} \end{cases}$$

The estimates of the concentration from the different measurement directions are independent, which makes the variance easy to calculate:

$$\hat{V}(\hat{I}) = \hat{V}\left(\sum_{i=1}^{n_r} \sum_{j=1}^{n_\varphi} v_{ij} r_i \hat{n}(r_i, \varphi_j) \Delta r \Delta \varphi \right) = \sum_{j=1}^{n_\varphi} \hat{V}\left(\sum_{i=1}^{n_r} v_{ij} r_i \hat{n}(r_i, \varphi_j) \Delta r \Delta \varphi \right) \tag{43}$$

$$= \sum_{j=1}^{n_\varphi} \widehat{\mathrm{cov}}\left(\sum_{i=1}^{n_r} v_{ij} r_i \hat{n}(r_i, \varphi_j) \Delta r \Delta \varphi, \sum_{k=1}^{n_r} v_{kj} r_k \hat{n}(r_k, \varphi_j) \Delta r \Delta \varphi \right) \tag{44}$$

$$= \Delta r^2 \Delta \varphi^2 \sum_{j=1}^{n_\varphi} \sum_{i=1}^{n_r} \sum_{k=1}^{n_r} v_{ij} r_i v_{kj} r_k \widehat{\mathrm{cov}}(\hat{n}(r_i, \varphi_j), \hat{n}(r_k, \varphi_j)) \tag{45}$$

where

$$\widehat{\mathrm{cov}}(\hat{n}(r_i, \varphi_j), \hat{n}(r_k, \varphi_j)) = C^2 e_2^{\mathrm{T}} (X_{2,r_i}^{\mathrm{T}} W_{r_i} X_{2,r_i})^{-1} X_{2,r_i}^{\mathrm{T}} W_{r_i} \hat{V}_y W_{r_k} X_{2,r_k} (X_{2,r_k}^{\mathrm{T}} W_{r_k} X_{2,r_k})^{-1} e_2 \tag{46}$$

and $\hat{V}_y = \mathrm{diag}\{\hat{\nu}(r_1), \ldots, \hat{\nu}(r_n)\}$.

It is also important to estimate the bias of the integral. Using

$$\hat{E}(\hat{n}(r, \varphi)) - n(r, \varphi) \approx \frac{C}{2} \hat{m}^{(3)}(r, \varphi) h_{\mathrm{MISE}_1}^2$$

we get

$$\hat{E}(\hat{I}) - I \approx \frac{C}{2} \sum_{i=1}^{n_r} \sum_{j=1}^{n_\varphi} v_{ij} r_i \hat{m}^{(3)}(r_i, \varphi_j) h_{\mathrm{MISE}_1}^2 \Delta r \Delta \varphi \tag{47}$$

where $\hat{m}^{(3)}(r_i, \varphi_j) = \hat{m}_3(r_i, \hat{h}_{\mathrm{MISE}_3}, 4)$ is estimated as in the preceding section for the different directions φ_j. The numerical results are presented in Table I. *Note that bias seems to be negligible for the integrated concentration.*

This effect has been studied theoretically in a slightly different context in two papers by Härdle *et al.* (1992) and Härdle and Stoker (1989) and it is our conjecture that it can be thoroughly proved here that we can regard our estimate of the integrated concentration surface as approximately normally distributed with expectation zero and variance $V(\hat{I})$, i.e. confidence intervals can be straightforwardly constructed as $\left(\hat{I} \pm \lambda_{\alpha/2} \hat{D}(\hat{I})\right)$.

6. FUTURE RESEARCH

We strongly believe that the local least-squares kernel regression methods have a potential to improve the evaluation of LIDAR measurements. As far as the authors know, this is the first

such investigation and much research remains to be done:

(i) This paper concentrates on bandwidth selection procedures which are MISE-optimal. Other bandwidth selection procedures need investigation.

(ii) This paper treats the variance function parametrically. Next we plan to consider a generalization of the technique which handles the variance function non-parametrically. In fact, during the revising procedure of this paper, we have studied different methods to estimate also the variance function $\nu(r)$ through a similar local smoothing technique, see Ruppert *et al.* (1995b).

(iii) On-line (in time) efficient versions of the estimation procedure should be considered especially in situations with large numbers of rapid measurements. Also some robust modifications of the method might be useful. In a paper by Fan *et al.* (1994) robust versions of the local polynomial kernel regression methods are discussed.

ACKNOWLEDGEMENTS

U.H. is supported by the Swedish Natural Science Research Council, 9365-305 and 9365-308 and by the Swedish National Board for National and Technical Development 91-02637P.

REFERENCES

Björklund, C. (1994). 'Analysing LIDAR-measurements by the method of locally weighted least squares kernel regression (in Swedish)', Master's thesis, Department of Mathematical Statistics, Lund University, Lund, Sweden.

Edner, H., Ragnarsson, P., Svanberg, S., Wallinder, E., De Liso, A., Ferrara, R. and Maserti, B. (1992). 'Differential absorption lidar mapping of atmospheric atomic mercury in Italian geothermal fields', *Journal of Geophysical Research*, **97**, 3779–3786.

Fan, J., Hu, T.-C. and Truong, Y. (1994). 'Robust non-parametric function estimation', *Scandinavian Journal of Statistics*, **21**, 433–446.

Grant, W. and Menzies, R. (1983). 'A survey of laser and selected optical systems for remote measurements of pollutant gas concentration', *Journal of Air Pollution Control Association*, **33**, 187–194.

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press.

Härdle, W., Hart, J., Marron, J. and Tsybakov, A. (1992). 'Bandwidth choice for average derivative estimation', *Journal of the American Statistical Association*, **87**, 218–226.

Härdle, W. and Stoker, T. (1989). 'Investigating smooth multiple regression by the method of average derivatives', *Journal of the American Statistical Association*, **84**, 986–995.

Killinger, D. and Mooradian, A. (eds) (1983). *Optical and Laser Remote Sensing*, Springer.

Measures, R. (1984). *Laser Remote Sensing*, Wiley.

Measures, R. (1988). *Laser Remote Chemical Analysis*, Wiley.

Ragnarson, P. (1994). *Optical Techniques for Measurements of Atmospheric Trace Gases*, PhD thesis, Department of Physics, Lund Institute of Technology, Sweden.

Ruppert, D., Sheather, S. and Wand, M. (1995a). 'An effective bandwidth selector for local least squares regression', *Journal of the American Statistical Association*, **90**, 1257–1290.

Ruppert, D. and Wand, M. (1994). 'Multivariate locally weighted least squares regression', *Annals of Statistics*, **22**, 1346–1370.

Ruppert, D., Wand, M., Holst, U. and Hössjer, O. (1995b). 'Local polynomial variance function estimation', *Technometrics* (Submitted.)

Sigrist, M. (ed) (1994). *Air Monitoring by Spectroscopic Techniques (Chemical Analysis Series, Vol. 127)*, Wiley.

Wahba, G. (1990). *Spline Models for Observational data*, SIAM.

Wand, M. and Jones, C. (1995). *Kernel Smoothing*, Chapman and Hall.

Zanzottera, E. (1990). 'Differential absorption lidar techniques in the determination of trace pollutants and physical parameters of the atmosphere', *Analytical Chemistry*, **21**, 279–319.