ANZJS Final - Jan. 2007

PATTERN RECOGNITION VIA ROBUST SMOOTHING WITH APPLICATION TO LASER DATA

Carlo Grillenzoni*

University IUAV of Venice

Summary

Nowadays airborne laser scanning is used in many territorial studies, providing point data which may contain strong discontinuities. Motivated by the need of interpolating such data and preserving their edges, this paper considers robust nonparametric smoothers. These estimators, when implemented with bounded loss functions, have suitable jump-preserving properties. We develop iterative algorithms which are equivalent to nonlinear M-smoothers, but have the advantage of resembling the linear Kernel regression. The selection of their coefficients is carried out by combining cross-validation and robust-tuning techniques. Two real case studies and a simulation experiment confirm the validity of the method; in particular, the performance in building recognition is excellent.

Key words: Adaptive estimates; cross validation; discontinuous surfaces; laser scanning; kernel M-estimates; pseudolinear smoothers; redescending scores.

*Dipartimento di Pianificazione, Universita' IUAV di Venezia, 30135 Venezia, Italy.
e-mail: carlog@iuav.it

Acknowledgments. The author is very grateful to the Editors and the Referees for their helpful suggestions. A special thank is for Prof. Y.-H. Tseng, National Cheng Kung University, Taiwan, who provided a sample of laser data.

1. Introduction

Airborne light detection and ranging (LiDAR) is a relatively new technology for obtaining earth surface data having high density and high positional accuracy (e.g. Wang & Tseng, 2004). The system is placed on airmobiles and includes two measurement instruments: a laser scanner and a global positioning system (GPS). The laser sends to the ground an infrared signal which comes back to a sensor. The returning time allows the earth point elevation to be computed, and the reflectance value gives information on the physical nature of the ground. The GPS provides the corresponding spatial coordinates (latitude and longitude).

Unlike aerial and satellite images, the recorded data have a *punctual* nature. Their density is high because, taking into account flight conditions and sensor characteristics (e.g. scan angle 20 degrees, emission rate 10000 pulses per second), it may reach one point every 0.5 meters. With respect to optical systems, the fundamental advantage of LiDAR is that it can work by night, it is insensitive to shadows and also provides the buildings' height. On the other hand, observations are subject to several random effects and the two instruments may mismatch.

Airborne LiDAR data are mostly used in topography, geology and architecture: 1) On a large scale, they are utilized to develop digital elevation models (DEM) of the earth surface. Nowadays, these models are used in geographical information systems (GIS) for integrating the conventional cartography. 2) On a medium scale, they are employed for obtaining hazard maps of zones which may be subject to river floods, tidal inundations, landslides and other environmental risks. 3) On a small scale, laser data can be used for the detection and recognition of buildings in urban areas. Their representation is useful in architectural reliefs, volumetric computations and enables the construction of 3D city models for computer graphics.

In all of these cases there is need for data-*interpolation*, especially at the point 3 if the resolution required is less than one meter. Geostatistical smoothers, such as kriging, triangularizations and splines, have problems in urban areas because they are not able to preserve the discontinuities which are present on the ground (e.g. Morgan & Habib, 2002). Generally speaking, there are two major approaches for

estimating regression surfaces which contain jumps. In the first one, the jump locations are tentatively identified with change-point tests, then conventional smoothers are applied in each continuous subregion. In the second, jump-preserving smoothers are applied overall without taking account of the possible location of discontinuity points. The last solution is nearly automatic, but involves more bias at the jumps and lower efficiency in smooth regions.

The first approach was fundamentally used in the one-dimensional design space; e.g. Hall & Titterington (1992) considered three nearest neighbor estimates and proposed various diagnostics to decide whether the regression function was continuous at each point. For multi-dimensional design spaces, the second method is preferable because jump points are difficult to detect and their number may be infinite. In this context, nonparametric smoothers are the natural estimators in view of their flexibility and independence of a-priori assumptions (e.g. Härdle, 1991). Moreover, their *robust* versions, by treating the observations beyond discontinuities as outliers, are potentially jump-preserving.

Robust smoothing was mainly developed in the context of the M-estimation theory of Huber (1981). Generally speaking, there are two major approaches, which depend on the fact that the underlying loss functions are bounded or unbounded. Kernel M-type regression with unbounded loss (or monotone score) was considered by Härdle & Gasser (1984), Hall & Jones (1990) and Wang & Scott (1994). It was designed to resisting additive outliers and to allow consistency in the case of heavy-tailed distributions. Its extension to the local polynomial regression (LPR) was discussed by Tsybakov (1986), Fan *et al.* (1994) and Welsh (1996). This solution performs better than the kernel one at the borders.

Bivariate M-smoothers with bounded loss (non-monotone score) are widely employed in image processing, where they provide denoising filters. In particular, Chu *et al.* (1998) used *redescending* score functions and showed their good edge-preserving capability. Extension of this framework to the local polynomial regression was discussed by Rue *et al.* (2002) and Hwang (2004), and Hillebrand & Müller (2006) have recently derived the conditions of consistency at the edges. Finally,

Polzehl & Spokoiny (2000) developed an adaptive method which is intermediate to the kernel M-estimator and the denoising filter of Saint Mark *et al.* (1990).

In image processing, M-smoothers exploit the fact that data are available on regular lattices, which simplifies computation and analysis. Application of these estimators to point data involves some structural changes. In fact, interpolation deals with missing values and spatial coordinates of LiDAR data are entirely stochastic. By using the weighted average form of M-estimates (e.g. Hampel *et al.*, 1986 p.115), this paper derives robust *pseudolinear* smoothers. They have a structure similar to the linear kernel regression and can be implemented in a sequential way. Their smoothing coefficients can be selected with mixed cross-validation and robust tuning techniques, which satisfy relative efficiency requirements.

The plan of the work is as follows: Section 2 introduces the case study and applies nonrobust smoothers; Section 3 discusses kernel M-type estimators and tests their performance on the airborne laser data; Section 4 provides further applications and discusses the statistical properties.

2. Preliminary data analysis

To present the methods in an effective manner, we introduce the case study here. We consider a subset of the LiDAR data discussed in Wang & Tseng (2004), concerning the city of Hsinchu (Taiwan) and generated by Leica ALS40 instrument on April 2002. The measurement has a mean density of 2.3 point per meter and a declared accuracy of 30 cm; original dataset covers a square area of 0.5 Km^2, while our subset regards a zone of 75×50 m, which contains N=8369 points. For each point, measurements for the time (t), the latitude (y), the longitude (x), the height (Z) and the reflectance (w), are recorded; the latter describes the physical nature of the ground and is useful for classification purposes. In the urban context the main interest is on building recognition and the reflectance may detect the presence of objects which are extraneous, such as vehicles and trees. Figure 1(a,b) provides 2D and 3D representations of spatial coordinates $\{x_i, y_i, Z_i\}$. One can note that data

are very dense and the spatial coordinates form irregular and overlapping stripes which have variable density. However, only few points are placed on the building walls and this increases the need for interpolation.

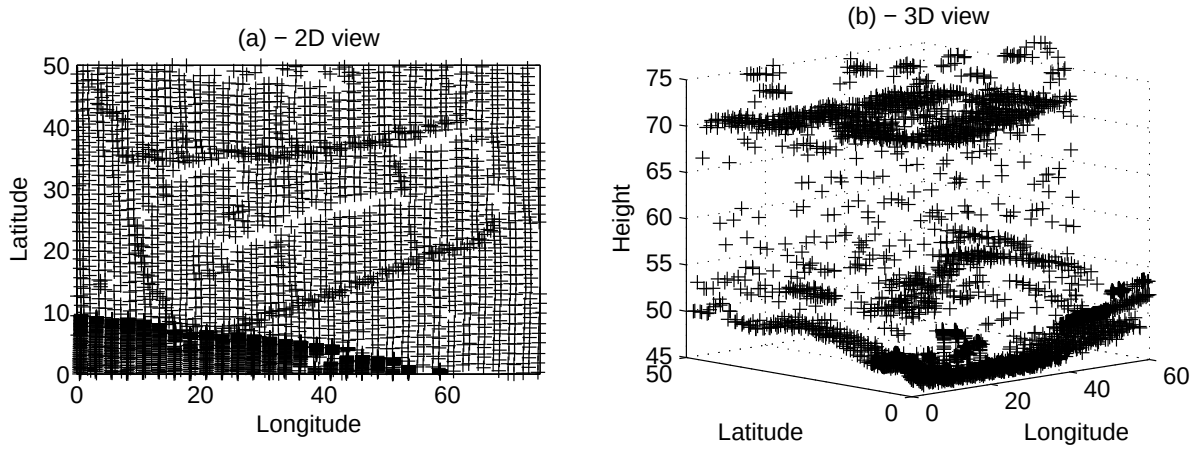

Figure 1. Spatial representation of LiDAR data; sample size N=8369.

Nonparametric smoothers (e.g. Härdle, 1991) can be very useful to deal with LiDAR data because they do not assume mathematical models for the underlying surfaces. Focusing on spatial coordinates and the model $Z = g(x, y) + \varepsilon$, where $g(\cdot)$ is an unknown function and ε is a noise process, the typical structure of the kernel (K) regression estimator is given by

$$\begin{aligned} \hat{g}_{\mathrm{K}}(x, y) &= \sum_{i=1}^{N} v_i(x, y)\, \mathrm{Z}_i \\ v_i(x, y) &= \frac{K_1\Big[(\mathrm{x}_i - x)/h_1\Big]\, K_2\Big[(\mathrm{y}_i - y)/h_2\Big]}{\sum_{i=1}^{N} K_1\Big[(\mathrm{x}_i - x)/h_1\Big]\, K_2\Big[(\mathrm{y}_i - y)/h_2\Big]} \end{aligned} \tag{1}$$

where $(x, y) \in \Re$ are continuous variables, $\{\mathrm{x}_i, \mathrm{y}_i, \mathrm{Z}_i\}$ are punctual observations, $K_{1,2}(\cdot)$ are symmetric densities and $0 < h_{1,2} < \infty$ are smoothing coefficients.

Optimal selection of such coefficients can be obtained with the cross-validation technique, which minimizes the sum of squared prediction errors

$$Q_N(h_1, h_2) = \sum_{j=1}^{N} \Big[\mathrm{Z}_j - \hat{g}_{\mathrm{K}-j}(\mathrm{x}_j, \mathrm{y}_j)\Big]^2 \tag{2}$$

where $\hat{g}_{\mathrm{K}-j}(\cdot)$ are estimates as in (1) obtained by omitting the j-th observation. Applying this method to the data of Figure 1, under the choice of Gaussian kernels,

we obtained $\hat{h}_1$=.047 and $\hat{h}_2$=1.32. However, these coefficients produce a surface which tends to follow the "stripes" of the aircraft (see Figure 1(a)). Instead, by constraining h_1=h_2, the method (2) yields $\hat{h}_{1,2}$=0.59. With this value we generated the surface in Figure 2 which has resolution 1 m^2 and size 75×50.

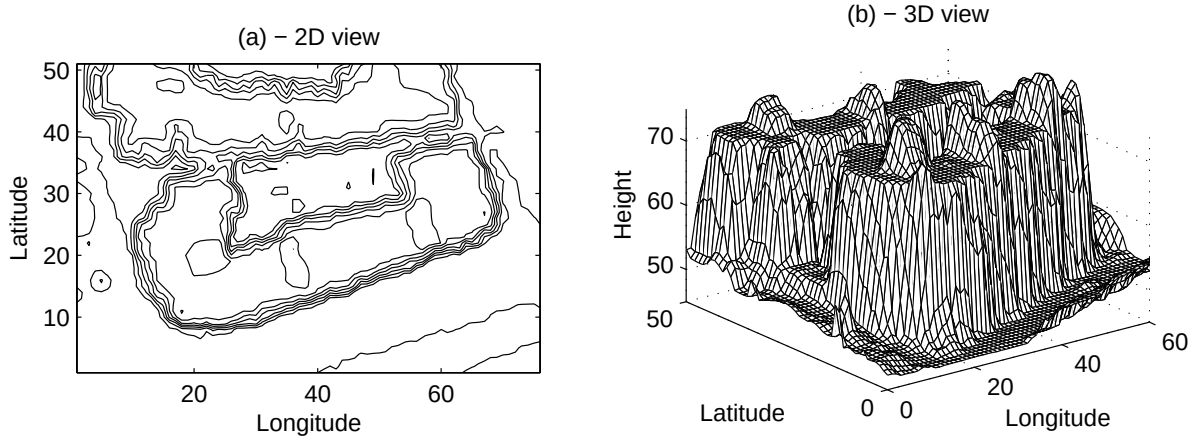

Figure 2. Kernel regression estimate of the data in Figure 1. It is generated with the algorithm (1) with Gaussian kernels and bandwidth $h_1 = h_2 = 0.6$.

Despite their flexibility, classical nonparametric estimators have major problems when the surfaces present discontinuities. This can be checked in Figure 2 by noting that building walls are not as sharp as should be expected in the reality, see Figure 1(b). A better *visual* performance could be obtained with a smaller bandwidth, but this only slightly improves the situation.

Analysis of the disturbances is an important diagnostic step. Here, one must distinguish between prediction errors $\hat{\varepsilon}_j = [\,\mathrm{Z}_j - \hat{g}_{\mathrm{K}-j}(\mathrm{x}_j, \mathrm{y}_j)\,]$ (spatial innovations) and residuals of regression $\check{\varepsilon}_i = [\,\mathrm{Z}_i - \hat{g}_{\mathrm{K}}(\mathrm{x}_i, \mathrm{y}_i)\,]$. While the variance of the first has a well-defined minimum with respect to the bandwidths, the variance of the latter converges to zero as $h_{1,2} \to 0$. A natural estimator for the noise variance is then given by $\hat{\sigma}^2_\varepsilon = Q_N/N$, but is sensitive to outliers. A robust alternative can be obtained from the median (med) absolute deviation (MAD) as

$$\hat{\sigma}_{\mathrm{M}} = \mathrm{med}_i\Big\{|\,\hat{\varepsilon}_i - \mathrm{med}_j(\hat{\varepsilon}_j)|\Big\}/.6745 \tag{3}$$

where .6745 allows consistency in the Gaussian case (Huber, 1981 p.107). Using

these formulas we obtained $\hat{\sigma}_\varepsilon$=3.24, $\check{\sigma}_\varepsilon$=2.57, and $\hat{\sigma}_{\mathrm{M}}$=0.25 which is very different from the others and reveals presence of non-normality. It is worth noting that this situation is extraneous to the original series Z_i, where both estimators provide a similar value; namely $\hat{\sigma}_Z = 10.5$ and $\hat{\sigma}_{\mathrm{M}} = 10.8$.

Another useful diagnostic tool is the kernel density estimation $\hat{f}_{\mathrm{K}}(\varepsilon)$. This can be generated with the heuristic bandwidth $\hat{\sigma}_\varepsilon/N^{1/5}$=.53 (see Härdle, 1991 p.91), which is close to the cross-validation estimate of $h_{1,2}$. Kernel density was computed both for innovations and residuals and is displayed in Figure 3(a); Panel (b) provides the normal QQ-plot of residuals. Both graphs diagnose the presence of a marked non-Gaussianity in the form of heavy tails; these are produced by the large errors at the building edges. A solution to the poor fitting of the kernel regression (1) can be achieved by filtering large residuals; indeed, these are the effect and the cause of the surface oversmoothing at the jumps. This task is typically pursued in *robust* estimation (e.g. Huber, 1981), where residuals are controlled by modifying the loss function.

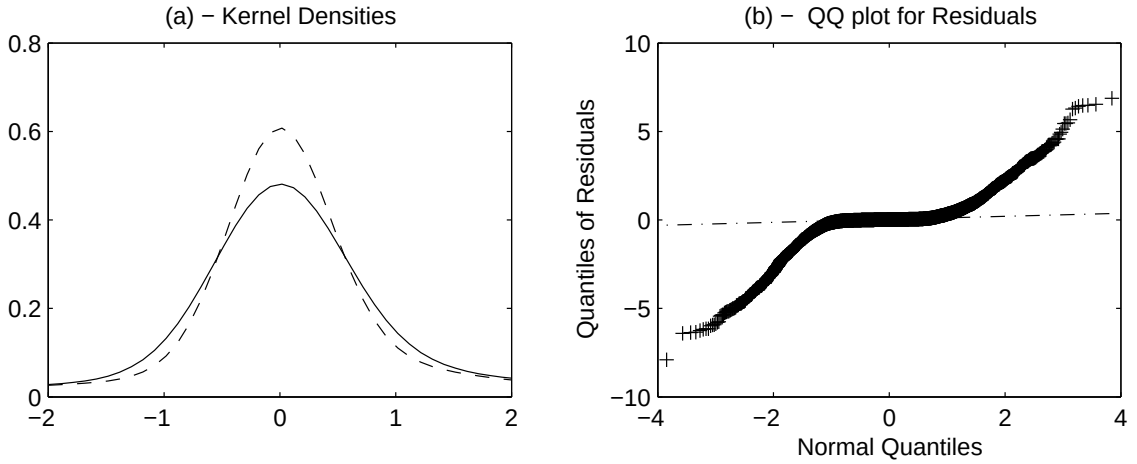

Figure 3. Diagnostic analysis for (1): (a) Kernel densities of prediction errors (solid line) and residuals (dashed line); (b) Normal quantiles plot for residuals.

3. Kernel M-type smoothers

In this section we discuss robust smoothers by linking Kernel regression and M-type estimation. This approach was developed by Härdle & Gasser (1984), with regard to univariate models contaminated by outliers. Subsequently, Hall & Jones (1990) extended the method to the random design, and Chu *et al.* (1998) and Rue *et al.* (2002) applied it to image processing (i.e. bivariate fixed design).

Assume that the data follow a non-linear model with stochastic regressors

$$\mathrm{Z}_i = g(\mathrm{x}_i, \mathrm{y}_i) + \varepsilon_i\,, \quad \varepsilon_i \sim \mathrm{IID}(0, \sigma_\varepsilon^2)\,; \qquad i = 1, 2 \ldots N \tag{4}$$

where $g(x, y) = \mathrm{E}(Z \,|\, x, y)$ is a discontinuous function, with jumps located at unknown points, and $\{\varepsilon_i\}$ is an independent and identically distributed (IID) sequence with symmetric density $f(\varepsilon)$. As an example one could have

$$g(x, y) = \gamma(x, y) + \delta_1 \cdot I_1\Big\{(x, y) :\; y \geq [\,\phi(x) + \delta_2 \cdot I_2(x \geq \mathrm{x}_0)\,]\Big\}$$

where $\gamma(\cdot)$ is a continuous function, $\delta_{1,2}$ are jumps and $I_{1,2}(\cdot)$ are indicator functions. Note that in the above scheme, the discontinuity edge of $g(\cdot)$ follows the relationship $y = \phi(x)$, which also has a jump at the point x_0.

The connection between discontinuous models and models with outliers can be shown by including the jump component of $g(\cdot)$ in the noise component of (4). Relaxing the IID assumption, it turns out that $\{\varepsilon_i\}$ have a mixture density of the type $f_\varepsilon^* = f_0 \cdot I[\,y < \varphi(x)] \; + f_\delta \cdot I[\,y \geq \varphi(x)]$, where f_0 is centered on 0 and f_δ is centered on δ. This remark renders robust estimators suitable for the model (4).

3.1 Redescending smoothers

Because the estimator (1) minimizes the functional $\sum_{i=1}^N v_i(x, y)\,(\mathrm{Z}_i - g)^2$, its "robustization" can be achieved by replacing the quadratic loss with a convex function $\rho(\cdot)$ which is less sensitive to extreme values. Specifically, the kernel M-smoother is the solution of the locally weighted maximum likelihood type problem

$$\hat{g}_{\mathrm{M}}(x, y) = \arg\min_g \; \Big[P_N(g) = \frac{1}{N}\sum_{i=1}^N v_i(x, y)\,\rho_\alpha\Big(\mathrm{Z}_i - g\,\Big)\Big] \tag{5}$$

where the local weights $\{v_i\}$ are defined as in (1), and α is a tuning constant which is related to the scale parameter σ_ε. In the parametric literature, the solution to the optimization problem (5) is usually viewed as the root of the normal equation $\sum_i v_i(x,y)\,\psi_\alpha(\mathrm{Z}_i - g\,) = 0$, where $\psi_\alpha(\varepsilon) = \partial\rho_\alpha(\varepsilon)/\partial\varepsilon$.

To enable robustness, the function $\rho(\varepsilon)$ must not grow too rapidly as $|\varepsilon| \to \infty$; or, more precisely, the score function $\psi(\varepsilon)$ must be uniformly bounded. In this context there are two alternative philosophies: Huber (1981) states that $\psi(\cdot)$ must be monotone and must achieve its maximum value asymptotically, because outliers may contain useful information. On the contrary, Hampel *et al.* (1986) claim that it should tend to zero because outliers are usually extraneous to the models. These approaches have opposite effects on the properties of consistency and robustness of estimates; in fact, the second one is insensitive to outliers, but may not converge to the global minimum point because it admits multiple local roots.

Following Huber's and Hampel's philosophies, the loss function can be designed as unbounded or bounded, respectively. Some important examples are:

$$
\begin{array}{lrcl}
\text{a)} & \rho_a(\varepsilon) & = & \Big|\,\varepsilon\,\Big|, \qquad \psi_a(0) \equiv 0 \\
\text{b)} & \rho_b(\varepsilon) & = & \begin{cases} \varepsilon^2/2\,, & |\varepsilon| \le \alpha \\ \alpha\,|\varepsilon| - \alpha^2/2\,, & |\varepsilon| > \alpha \end{cases} \\
\text{c)} & \rho_c(\varepsilon) & = & \begin{cases} \varepsilon^2/2\,, & |\varepsilon| \le \alpha \\ \alpha^2/2\,, & |\varepsilon| > \alpha \end{cases} \\
\text{d)} & \rho_d(\varepsilon) & = & -L\Big(\varepsilon/\alpha)/\alpha
\end{array}
\tag{6}
$$

where $L(\cdot)$ is a density/kernel function and $0 < \alpha < \infty$ is a tuning constant that controls the degree of robustness and must be selected according to the rate of outlier contamination. The loss function (6,a) was stressed by Wang & Scott (1994) and is independent of α; (6,b) is the one preferred by Huber, and has a monotone derivative; (6,c) corresponds to the *trimmed* method and approximates the bisquare one of Tukey; finally, (6,d) is a smoothed solution which provides a *redescending* ψ-function (see Hampel *et al.*, 1986 p.149). Graphical behavior of these functions and of their transformations is shown in Figure 4.

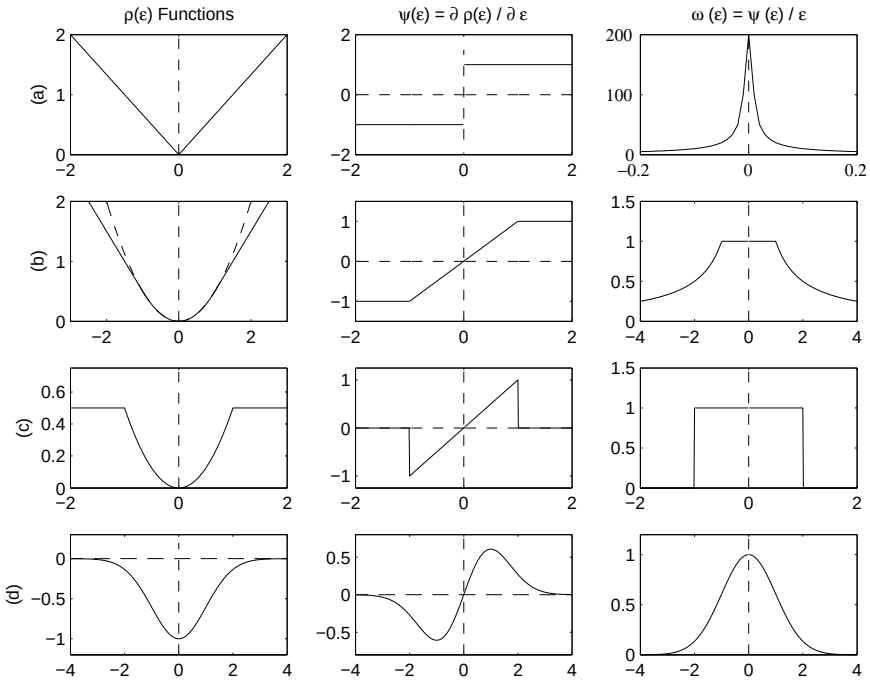

Figure 4. Graphs of loss functions in (6) with $L(\cdot)$ Gaussian and $\alpha = 1$. The figure also provides the score functions $\psi = \partial\rho/\partial\varepsilon$ and weight functions $\omega = \psi/\varepsilon$.

The utility of the M-smoother (5) in fitting discontinuous surfaces arises from the fact that its local properties are better than those of conventional estimators (LPR included). To be specific, jump-preserving is related to outlier-resistance because the observations which are placed near a jump point typically yield anomalous residuals. Since the estimator (5) is the solution of the equation $\sum_i v_i\,\psi(\mathrm{Z_i}-g)=0$, it follows that data on the edges tend to be censored by the functions $\psi(\cdot)$. However, such scores typically behave as threshold functions, so that discontinuities in the estimated surface are finally generated.

Looking at Figure 4, one can see that the thresholding effect is particularly hard in the case of bounded loss functions (6;c,d). From the latter it can be noted that the minimization (5) coincides with the maximization of the kernel density

$$\hat{f}_{\mathrm{K}}(x,y,Z) = \frac{1}{h_1 h_2\,\alpha\,N}\sum_{i=1}^{N} K_1\!\left(\frac{\mathrm{x}_i - x}{h_1}\right) K_2\!\left(\frac{\mathrm{y}_i - y}{h_2}\right) L\!\left(\frac{\mathrm{Z}_i - Z}{\alpha}\right) \qquad (7)$$

This shows the close connection between redescending M-smoothing and the *modal* regression approach discussed in Scott (1992, sec. 8.3.2).

3.2 Pseudolinear smoothers

The computation of (5), for every point (x, y), typically proceeds by non-linear algorithms, such as the Gauss-Newton one

$$\hat{g}_{\mathrm{M}}^{(k+1)}(x,y) = \hat{g}_{\mathrm{M}}^{(k)}(x,y) + \Big[\sum_{i=1}^{N} v_i(x,y)\,\rho''\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{M}}^{(k)}(x,y)\Big)\Big]^{-1} \sum_{i=1}^{N} v_i(x,y)\,\psi\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{M}}^{(k)}(x,y)\Big) \tag{8}$$

where (k) is a generic iteration and the initial value may be $\hat{g}_{\mathrm{M}}^{(0)} = \hat{g}_{\mathrm{K}}$. The direct minimization of (5) is computationally demanding, and is suitable only if the grid of values for (x, y) and/or the sample size N are small.

An alternative solution can be obtained from the *weighted average* form of M-estimates introduced by Tukey (see Hampel *et al.* 1986, p.115). Using the residual weight function $\omega(\varepsilon) = \psi(\varepsilon)/\varepsilon$, one can obtain the equation

$$P'_N(g) = \sum_{i=1}^{N} v_i(x,y)\,\psi(\mathrm{Z}_i - g) = \sum_{i=1}^{N} v_i(x,y)\,\omega(\mathrm{Z}_i - g)\,(\mathrm{Z}_i - g) = 0$$

and solving for g (in iterative form), provides the weighted (W) smoother

$$\hat{g}_{\mathrm{W}}^{(k+1)}(x,y) = \Big[\sum_{i=1}^{N} v_i(x,y)\,\omega\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{W}}^{(k)}(x,y)\Big)\Big]^{-1} \sum_{i=1}^{N} v_i(x,y)\,\omega\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{W}}^{(k)}(x,y)\Big)\,\mathrm{Z}_i \tag{9}$$

In parametric models it can be shown that W-estimators have the same influence function and asymptotic variance as M-estimates (e.g. Hampel *et al.* 1986, p.116). If the weights $\{v_i\}$ are non-negative, the same property can be extended to robust smoothers and it can be concluded that (8) and (9) are equivalent.

Because the ω-functions have a kernel nature (see Figure 4), it follows that (9) has a structure similar to the kernel regression in the 3D design space. As an example, consider the loss (6,d) with $L(\cdot)$ Gaussian; one can check that also $\omega(\cdot)$ is Gaussian, and inserting it into (9) provides the redescending (R) smoother

$$\hat{g}_{\mathrm{R}}^{(k+1)}(x,y) = \frac{\sum_{i=1}^{N} K_1\Big[(\mathrm{x}_i - x)/h_1\Big]\,K_2\Big[(\mathrm{y}_i - y)/h_2\Big]\,L\Big[\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{R}}^{(k)}(x,y)\Big)/\alpha\Big]\,\mathrm{Z}_i}{\sum_{i=1}^{N} K_1\Big[(\mathrm{x}_i - x)/h_1\Big]\,K_2\Big[(\mathrm{y}_i - y)/h_2\Big]\,L\Big[\Big(\mathrm{Z}_i - \hat{g}_{\mathrm{R}}^{(k)}(x,y)\Big)/\alpha\Big]} \tag{10}$$

Apart from the iterative nature, the above has the same structure as a kernel regression which performs local weighting also in the direction of the dependent variable

Z. It also has a certain connection with the *sigma* filter used in image denoising (see Chu *et al.*, 1998); the relationship becomes evident if one replaces $\hat{g}_{\mathrm{R}}$, within $L(\cdot)$, with the observation Z_j which is spatially closer to the point (x, y). However, this modification significantly worsens the performance of (10). Finally, one can interpret (9)-(10) as the solution of the nonlinear problem (5) by reweighted least squares; these solve iteratively the normal equation associated with (5) without computing the gradient (e.g. Hall & Jones, 1990 p.1716).

In the presence of a large amount of data, it is useful to combine the iterations of algorithms (8)-(10) with *sequential* processing of the data (e.g. Grillenzoni, 1997). In practice, this can be realized by splitting the data-set into $m \geq 10$ disjoint random subsets of size $n = N/m$, and then averaging the resulting estimates. With respect to the trimmed solution (6,c), whose ω-function is the indicator $I(\cdot)$, the sequential algorithm is given by

$$\begin{aligned} \bar{g}_{\mathrm{R}}^{(k)}(x,y) &= \frac{k-1}{k}\,\bar{g}_{\mathrm{R}}^{(k-1)}(x,y) + \frac{1}{k}\,\hat{g}_{\mathrm{R}}^{(k)}(x,y) \\ \hat{g}_{\mathrm{R}}^{(k+1)}(x,y) &\propto \sum_{i=1}^{n} K_1\!\left(\frac{\mathrm{x}_{ki}-x}{h_1}\right) K_2\!\left(\frac{\mathrm{y}_{ki}-y}{h_2}\right) I\!\left(\left|\mathrm{Z}_{ki} - \bar{g}_{\mathrm{R}}^{(k)}(x,y)\right| < \alpha\right) \mathrm{Z}_{ki} \end{aligned} \tag{11}$$

where the first equation is the recursive version of the mean $\bar{g}_{\mathrm{R}}^{(k)} = k^{-1}\sum_{h=1}^{k}\hat{g}_{\mathrm{R}}^{(h)}$ and $\{\mathrm{x}_{ki}, \mathrm{y}_{ki}, \mathrm{Z}_{ki}\}$ is the k-th sub-sample, $k = 1 \ldots m$. Notice that $\bar{g}_{\mathrm{R}}^{(k)}$ is nested in the smoother through the kernel $I(\cdot)$, and also provides the final estimate.

The above procedure is particularly useful where the data density is much greater than the surface resolution, and not just for computational reasons. Algorithms which process all data together implicitly average observations within the same "pixels" and this tends to blur discontinuities of the surface. In fact, pixels which are placed on the edges usually include observations with different height. Now, in the case of data sub-sampling, at each iteration the probability of having pixels with non-homogeneous observations is drastically reduced; hence, the edge-preserving ability of (11) is better than that of (10). Experimentally, we have checked that a suitable sub-sample size is nearly 1/10 of the surface resolution, namely $n^* = (n_i \cdot n_j)/10$, which yields the number of iterations $m^* = N/n^*$.

3.3 Coefficient Selection

Selection of the coefficients h, α of M-smoothers could still be carried out with the cross-validation criterion (2), by computing $\hat{g}_{\mathrm{R}-j}(\mathrm{x}_j, \mathrm{y}_j)$ iteratively. However, in the application to LiDAR data, we encountered problems of convergence of the type $\hat{\alpha} \to \infty$. A similar drawback was observed by Hall & Jones (1990, p.1717) with respect to the coefficient of the Huber function applied to models with outliers, and also Chiu *et al.* (1998, p.536) encountered difficulties in applying cross-validation. As a consequence, they recommended using "visual evaluation" for selecting α, and Leung (2005) simply assigns subjective values to it.

As pointed out by a referee, the attempt to select α with the cross-validation is wrong because it is *not* a bandwidth. It is a tuning constant that controls robustness and therefore should be treated as a scale parameter. From a parametric viewpoint, α should be sensitive to the discontinuity edges of a surface, but these have area zero. This raises a problem of identifiability with respect to the criterion (2). A more technical argument comes from the analysis of the asymptotic integrated MSE of redescending M-smoothers. For the model $y = g(x) + \varepsilon$, it is given by

$$\int_{\Re} \mathrm{MSE}[\,\hat{g}_{\mathrm{M}}(x)\,]\,\mathrm{d}\,x \;\approx\; B_1\,h^4 + B_2\,/(Nh\,\alpha^3)$$

where the constants $B_{1,2}$ depend on integrals of the functions g'', f_ε, K, L; see Rue *et al.* (2002) and Hwang (2004). Now, differentiating the above with respect to α and equating to zero, the optimal solution becomes $\alpha \to \infty$.

In the parametric literature, a sensible approach to tune α consists in finding a suitable trade-off between efficiency and robustness (e.g. Hampel *et al.*, 1986 p.399). Indeed, from (6) one can see that robustness of M-estimates is inversely proportional to α, but their efficiency in absence of outliers is directly proportional to it. Now, setting $\alpha = C\,\sigma_\varepsilon^2$, with $C > 0$, it can be shown that M-estimates of the location parameter of a Gaussian model maintain 95% asymptotic relative efficiency with respect to the least squares only if $1 < C < 3$. Specifically, in the case of the loss functions (6) the constant takes on the values C_b =1.345 (Huber), C_c=2.81 (Trimmed), C_d=2.11 (−Gauss); see Fox (2002).

In this framework the crucial point is the estimation of σ_ε. In Section 2 we have seen that the ordinary standard error is biased upward ($\hat{\sigma}_\varepsilon$=3.24), but the MAD statistic may lead to underestimation ($\hat{\sigma}_{\mathrm{M}}$=0.25). Now, following Huber (1981, p.180) an unbiased estimate can be obtained from Winsorized residuals $\varepsilon_i^* = \psi_b(\varepsilon_i/\sigma_\varepsilon)/\sigma_\varepsilon$. In particular, starting from $\hat{\varepsilon}_i = \mathrm{Z}_i - \hat{g}_{\mathrm{K}-i}(\mathrm{x}_i, \mathrm{y}_i)$ and iterating we have

$$\begin{aligned} \left|\hat{\varepsilon}_i^*(k)\right| &= \min\left\{ |\hat{\varepsilon}_i|,\ 1.345 \cdot \hat{\sigma}_\varepsilon^*(k-1) \right\} \\ \hat{\sigma}_\varepsilon^*(k) &= \left(\frac{N}{N_k}\right)\left[N^{-1}\sum_{i=1}^{N} |\hat{\varepsilon}_i^*(k)|^2\right]^{1/2} \end{aligned} \tag{12}$$

where N_k is the number of unmodified errors at the k-th iteration. In LiDAR data the estimator (12) converged to $\hat{\sigma}_\varepsilon^* = .64$, independently of the initial $\hat{\sigma}_\varepsilon^*(0)$.

An improvement of the selection strategy can be gained from the *robust* cross-validation estimation of the bandwidths. This is obtained by replacing the quadratic loss in (2) with one of the ρ-functions in (6), and provides optimal MSE estimates $\hat{h}_{1,2}$ (see Leung, 2005). On the basis of these values, one can select α as follows:

1. *Robust*. Under the assumption of $f(\varepsilon)$, $L(\varepsilon)$ Gaussian, and the condition 95% asymptotic relative efficiency with respect to the kernel smoother, the robust solution is $\alpha = 2\,\sigma_\varepsilon$, where σ_ε must be estimated with (3) or (12).

2. *Constrained*. In order to solve the non-identifiability problem of α, one may impose the constraint $\alpha = D\,h_{1,2}$, with $D > 0$, and then proceed to the cross-validation estimation. From the robust solution the constant can be defined as $\hat{D} = 2\hat{\sigma}_\varepsilon^*/\hat{h}_{1,2}$, where the estimates come from the kernel regression.

3. *Graphical*. The approach of "visual evaluation" of Chu *et al.* (1998) can be made less subjective by defining the set of *admissible* values. Running the M-smoother with $\hat{h}_{1,2}$, one may find the set $S_1 = \{\alpha < \alpha_1^*\}$ for which it is unstable (i.e. $\hat{g}_{\mathrm{M}} \to \infty$), and $S_2 = \{\alpha > \alpha_2^*\}$ for which it is oversmoothed (i.e. $\hat{g}_{\mathrm{M}} \to \hat{g}_{\mathrm{K}}$). A sensible choice is given by the midpoint $\alpha^* = (\alpha_1^* + \alpha_2^*)/2$.

In LiDAR data, robust cross-validation (with the criterion $|\cdot|$) applied to the kernel regression provided $\hat{h}_{1,2} = .51$. Using the robust solution $\alpha = 2\sigma_\varepsilon$, the MAD statistic and the unbiased variance (12) gave $\hat{\alpha} = (0.44,\ 1.16)$ respectively. Validity of these

values is confirmed by the graphical approach. Indeed, conditionally on $\hat{h}_{1,2} = .51$ and $L(\cdot)$ Gaussian, the admissible set is $\alpha^*_{1,2}$=(.2, 1.5).

Surfaces generated by various robust smoothers are displayed in Figure 5. Panels (a,c) show the results of the nonlinear method (5) with Huber loss, and Panels (b,d) show the results of the pseudolinear method (11) with Hampel loss. Detailed description of their coefficients is reported in the heading of the figure. It is apparent that the method (11) with $L(\cdot)$ Gaussian enjoys the best jump-preserving property, both in situations of large and small scale variability. This performance worsens by using the trimmed solution $L = I(\cdot)$, and in the case of unbounded loss functions (6;a,b) it moves closer to the kernel one in Figure 2.

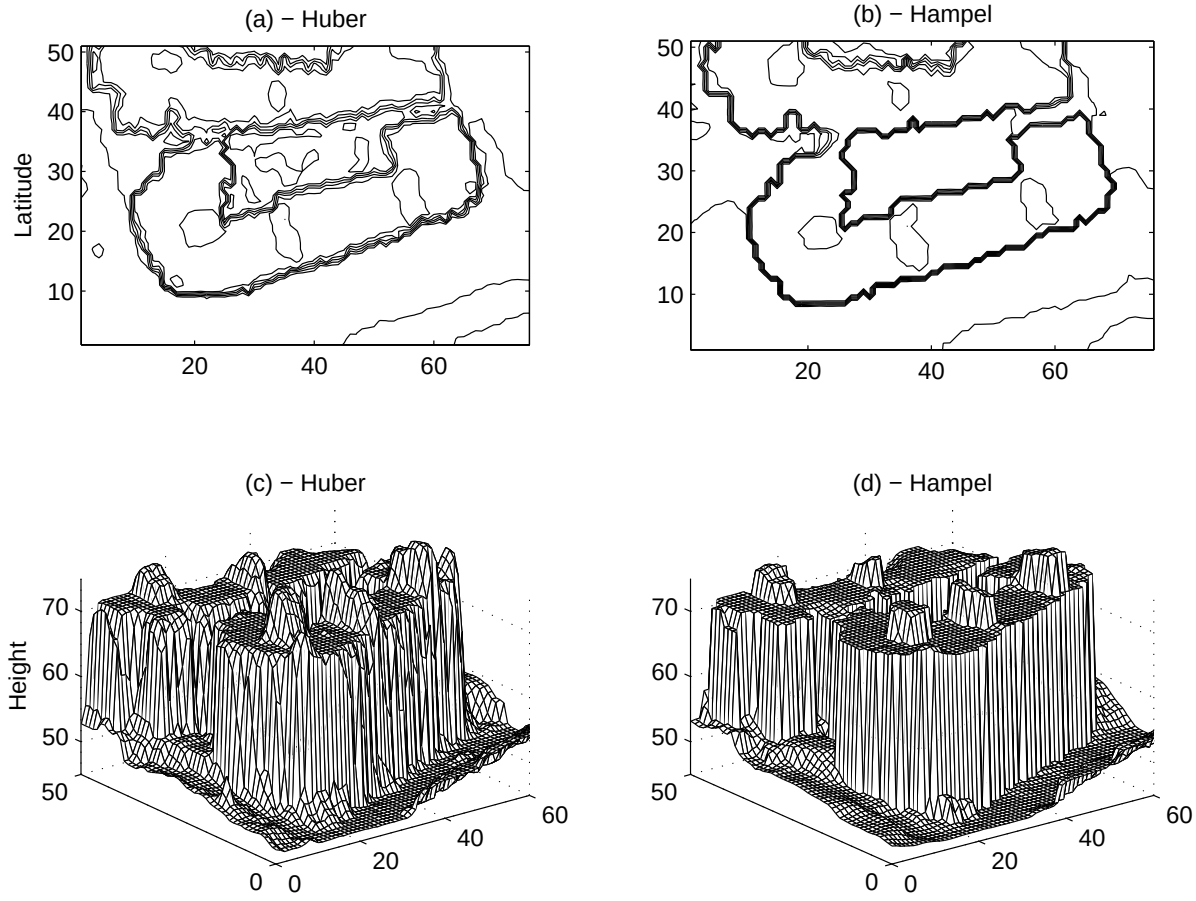

Figure 5. Robust regression estimates obtained with $K_{1,2}$ Gaussian and $h_{1,2}$=0.5: (a,c) Method (5) with Huber loss function (6,b) and α=1.16; (b,d) Method (11) with Hampel loss function (6,d), $L(\cdot)$ Gaussian, α=.44 and m=30.

3.4 A Simplified smoother

Although robust smoothers are satisfactory in preserving big jumps, they exhibit some weakness in small variability contexts, such as on the terrain and on the building roofs. One way to improve this aspect consists of simplifying previous algorithms by dropping unnecessary components. For example, in the estimator (9) one could drop the local weights $v_i(\cdot)$ by approximating $\mathrm{Z}_i \approx \hat{\mathrm{Z}}_i = \hat{g}_{\mathrm{K}}(\mathrm{x}_i, \mathrm{y}_i) -$ the rationale is that kernel estimates $\hat{g}_{\mathrm{K}}$ already include those weights. Moreover, specifying $\omega(z) = L(z)$ one can obtain the simplified (S) smoother

$$\hat{g}_{\mathrm{S}}^{(k+1)}(x,y) = \frac{\sum_{i=1}^{N} L\Big[\Big(\hat{g}_{\mathrm{K}}(\mathrm{x}_i, \mathrm{y}_i) - \hat{g}_{\mathrm{S}}^{(k)}(x,y)\Big)/\alpha\Big]\, \mathrm{Z}_i}{\sum_{i=1}^{N} L\Big[\Big(\hat{g}_{\mathrm{K}}(\mathrm{x}_i, \mathrm{y}_i) - \hat{g}_{\mathrm{S}}^{(k)}(x,y)\Big)/\alpha\Big]} \tag{13}$$

This looks like a parametric M-estimator in weighted average form (e.g. Hampel *et al.* 1986, p.115), whose weights $\omega(z)$ are designed for piecewise constant surfaces. In fact, by modeling $L(\cdot)$ as the indicator function $I(\cdot)$, the formula (13) just provides local means of the data Z_i. Moreover, as the value $\Big|\, \hat{g}_{\mathrm{K}}(\mathrm{x}_i, \mathrm{y}_i) - \hat{g}_{\mathrm{S}}(x,y) \,\Big|$ becomes large compared to α, the two points are almost classified in different regions. The simplified smoother has also some connections with gradient-based filters proposed by Saint-Mark *et al.* (1991) and Polzehl *et al.* (2000) for image denoising. These filters use weights $\omega(x,y)$ which are Gaussian kernels of the Laplacian gradient of the image. Now, the term $[\, \hat{g}_{\mathrm{K}} - \hat{g}_{\mathrm{S}} \,]$ of (13) has a role similar to the surface gradient evaluated at the empirical points.

At the computational level, the component $\hat{g}_{\mathrm{K}}(\cdot)$ of (13) could be replaced by more efficient estimates, such as (10) or (13) itself. However, in real applications we noted that they do not improve, or actually worsen, the jump-preserving ability. The above algorithm is very fast, is stable at the borders and is not very sensitive to the choice of α. In LiDAR data the admissible range for such a coefficient was $\alpha^*_{1,2} = (.2, 1)$, and Figure 6(a,b) exhibits estimates (13) obtained with α^*=.6 and k=15 iterations. As one can see, small objects and constant components of the surface are significantly enhanced; on the other hand, smooth regions on the terrain are segmented like a staircase. This problem cannot be solved by increasing α,

because this only causes erosion of small objects on the roofs. Compared with Figures 2 and 5, the simplified smoother (13) seems the best one for the recognition of the buildings shape, at least whenever they have a piece-wise constant form.

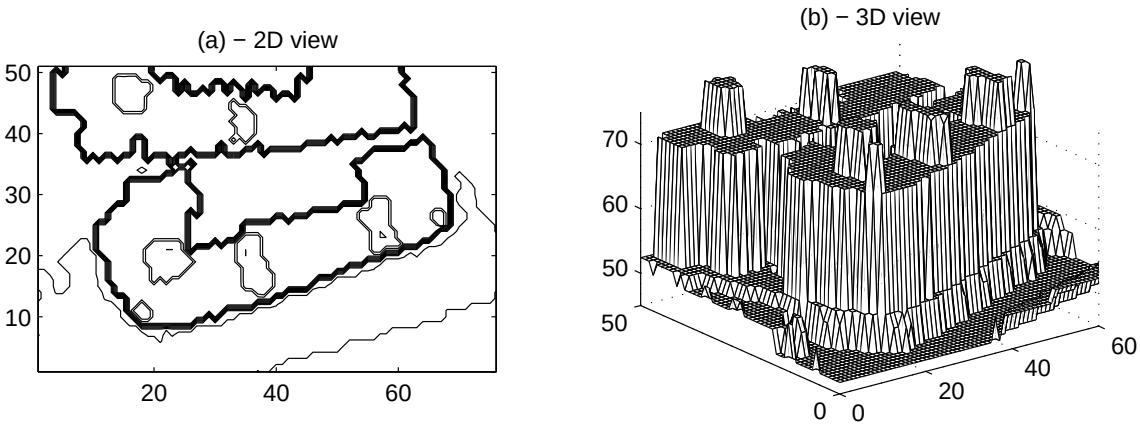

Figure 6. Weighted regression estimates of the data in Figure 1 obtained with the smoother (13), with Gaussian kernels, and the coefficient $\alpha = 0.6$.

3.4 Using the estimates

Fitting laser data of urban areas with robust smoothers mainly concerns building recognition. The resulting information can be directly employed in urban planning, military intelligence and civil protection (e.g. Wang & Tseng, 2004). Specifically, building extraction is useful in the following fields:

1) Architectural reliefs and computer graphics. They provide 3D digital surfaces which can represent the morphology of a whole city and can simulate virtual tours (see e.g. the recent software Google Earth$^{\mathrm{TM}}$). In general, the detected shapes are refined and made realistic by integrating them with digital cartography and ground images, typically by means of CAD systems (e.g. Rottensteiner, 2003).

2) Preliminary cartography of non-accessible zones. Obtaining maps from aerial or satellite images may be hindered by light conditions (clouds, fog and shadows). From laser data, fast cartography can be obtained merely on the basis of 2D representations like Figures 5(b) or 6(a), by merging zones which are external to the building contours. This is simply the complement of the building extraction.

3) Digital terrain models. Altimetry maps are used by civil protection for identifying zones which may be subject to environmental risks, such as water floods and terrain flows. In all of these cases the preliminary action is to discard data corresponding to the buildings. Once jump-preserving smoothers have provided the building contours, one can drop observations inside them by means of GIS softwares. Subsequently, remaining data are fitted with smoothers (1) or (10).

On the other side, LiDAR technology has few undesirable features that make it incapable of being a standalone reliable method. For example, laser data have no positional information along object break-lines, their planimetric accuracy is worse than the vertical one and they lack semantic information. This urges to integrate laser information with conventional photogrammetry.

4. Simulation experiments

This section provides further empirical evidence of the goodness of the proposed methodology. We consider an application to the density function of earthquake data and a simulation experiment. Both problems deal with jump-preserving.

4.1 Discontinuous densities

When discontinuities are present in probability functions $f(x, y)$, the classical method of kernel density estimation is inadequate. Following Fan & Gijbels (1996, p.50) an alternative solution consists in fitting the frequency histogram with a non-parametric smoother. If jumps are present at the borders, then the local polynomial regression can be used (e.g. Cheng *et al.* 1997). However, robust smoothing provides a more powerful approach.

We consider seismic data of the Northern California Earthquake Data Center in the area of San Francisco in the period 1968-2005. Figure 7(a) provides the scatterplot of the magnitude of events (y) versus their depth in Km (x). A frequency histogram f_{ij} of size 40×60 was constructed on these data. Figure 7(b) shows the kernel regression estimate of the density $f(x, y)$ obtained with the cross-validation

bandwidth $\hat{h}_{1,2} = 1.5$. The main experiment consists in fitting f_{ij} *without* the portion of data $\{\, k : (\mathrm{y}_k \leq y_{\mathrm{mod}}) \cap (\mathrm{x}_k \leq x_{\mathrm{mod}}) \,\}$, hence creating an artificial discontinuity. By applying the redescending M-smoother (10) with $L(\cdot)$ Gaussian and $\alpha = 2\,\hat{\sigma}_{\mathrm{M}} = 1$, we obtained the density in Figure 7(c). We can see that it preserves the introduced discontinuity well.

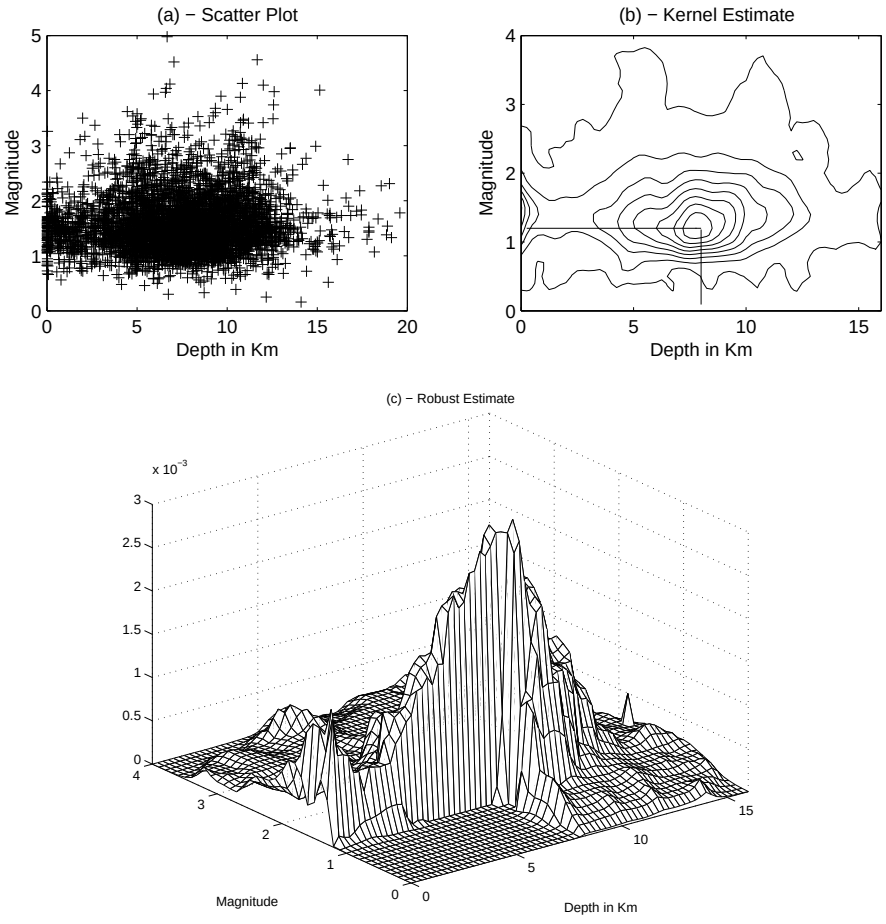

Figure 7. Regression estimation of a density function. (a) Earthquake data, x= depth, y=magnitude. (b) Kernel regression (1) on the whole histogram; (c) Robust regression (10) on the partial histogram; L, $K_{1,2}$ Gaussian, $h_{1,2} = 1.5$, $\alpha = 1$.

4.2 A simulation experiment

It is also useful to test the methods with a simulation experiment. We consider the deterministic surface

$$g(x, y) = .3(1 - x)\, y \;+\; \Big[1 + .5\sin(2\pi x)\Big] \cdot I\Big[y \geq .6\sin(\pi x) + .2\Big] \qquad (14)$$

which is displayed in Figure 8(a) for $0 \leq x, y \leq 1$. A random sample Z_i of size n= 100 was extracted from $g(x, y)$ by assuming $(x_i, y_i) \sim \mathrm{U}_2(0, 1)$, a bivariate uniform density. A typical realization is shown in Figure 8(b).

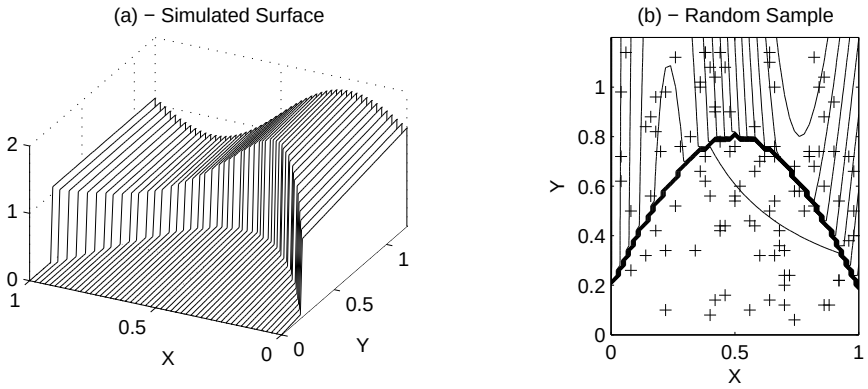

Figure 8. Simulation experiment: (a) Surface path; (b) Random sample.

The experiment consists of reconstructing the surface in Figure 8(a) by fitting the data in Figure 8(b) with various smoothers. Kernel regression (1) provided the cross-validation estimates $\hat{h}_{1,2} = .053$ and $\hat{\sigma}_{\mathrm{M}} = .074$. The resulting surface was oversmoothed and large prediction errors occurred at the jumps. Application of the robust smoother (10) has improved the situation. Using the robust selection rule $\alpha = 2\sigma_\varepsilon = .15$ we generated the estimates in Figure 9(a). Finally, the simplified filter (13) has presented serious difficulties in estimating the smooth component, even reducing its coefficient to critical values, see Figure 9(b).

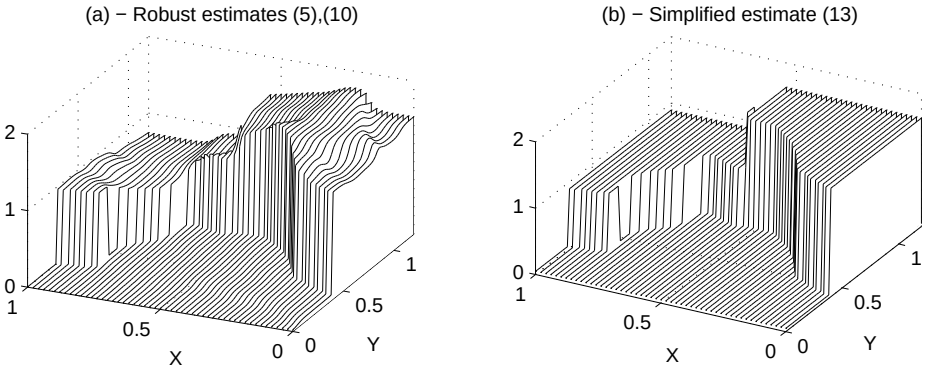

Figure 9. Robust smoothing of the data in Figure 8(b): (a) M-estimates (10) with $h_{1,2}$=.053 and α=.15; (b) Simplified estimates (13) obtained with α=.05.

Analysis of the integrated squared error ISE $=\sum_i \sum_j \left[g(x_i, y_j) - \hat{g}_{\mathrm{M}}(x_i, y_j)\right]^2$ sheds light on the difficulties to estimate α. Figure 10(a) shows the shape of the cross-validation function $Q_n(h_{1,2}, \alpha)$. As in LiDAR data, it has a well-defined minimum in the first coefficient, but not in the second one. Figure 10(b) shows the path of ISE(α) conditioned on the optimal value $h_{1,2}$=.05. One can see that it has a maximum at α=.15, which is the value that allows the best visual trade-off between sharpness at jumps and smoothness otherwise (see Figure 9(a)). This *seeming* contradiction arises from the fact that discontinuity edges have area zero and, in smooth regions, the estimator $\hat{g}_{\mathrm{M}}$ is less efficient than $\hat{g}_{\mathrm{K}}$. In other words, the adaptivity of robust smoothers at the jump points is largely paid for in continuous regions. These remarks confirm that α cannot be designed with common bandwidth selectors, and it is better to tune it as a robustness parameter.

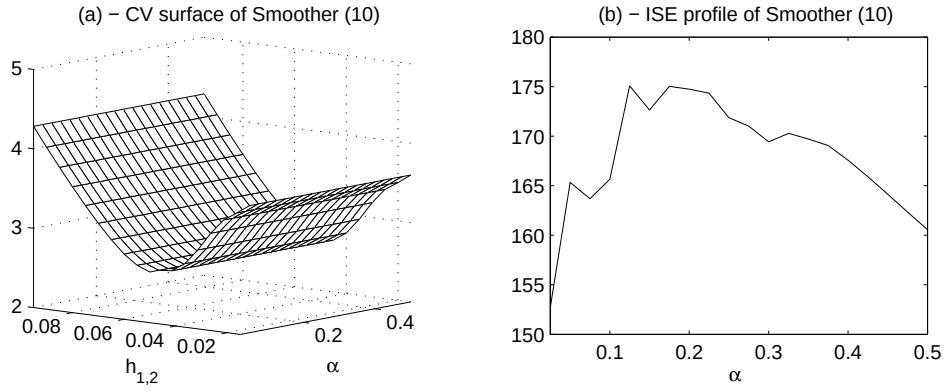

Figure 10. (a) Cross validation (CV) criterion for the smoother (10) with h_1=h_2. (b) Integrated squared error (ISE) conditioned on $h_{1,2}$=.05.

4.3 Statistical properties

This paper has discussed three kinds of robust smoothers, showing their algebraic connections. In particular, we have shown that pseudolinear (10) and nonlinear (5) estimators are equivalent when they are properly iterated. The bridge between the two classes is provided by the weighted algorithm (9), which can be simplified as in (13). Numerical applications have shown that smoothers with bounded loss

functions have a good jump-preserving ability. In smooth and continuous regions, however, they may lack convergence in probability.

This problem is well known in parametric M-estimation, where specific assumptions are needed to establish the consistency of algorithms with non-monotone score functions. For example, Freedman & Diaconis (1982) showed that uniqueness of the global minimum is not enough, and symmetry and monotonicity, on (-∞,0), of the underlying probability density are necessary. Other examples are provided in Jurečková & Sen (1996). What emerges in these studies is that the non-monotone nature of $\psi(z)$, by allowing for multiple local solutions, conflicts with the possible multi-modality of $f(Z)$. Now, if $v_i(x, y)$ are strictly positive weights, this conclusion can be extended to M-smoothers, yielding that their convergence depends on the shape of the noise density. In particular, if $f(\varepsilon)$ has saddle points, then consistency is not guaranteed (see Hillebrand & Müller, 2006).

This statement concerns M-smoothers in regions where $g(\cdot)$ is continuous. At the jump points the non-consistency should, in general, be expected. Indeed, Rue *et al.* (2002) showed that the asymptotic bias (AB) and variance (AV) of $\hat{g}_{\mathrm{M}}$ depend on the jump size δ, and resemble the moments of a Bernoulli function

$$|\mathrm{AB}| = \pi_\delta \delta\,, \quad \mathrm{AV} = \pi_\delta (1 - \pi_\delta)\, \delta^2\,, \quad \text{with} \quad \pi_\delta = \int_{\delta/2}^{\infty} f(\varepsilon)\, \mathrm{d}\varepsilon$$

where $f(\varepsilon)$ is assumed symmetric. However, these expressions also show that if the jump signal δ is large compared to the noise variance, then the statistical quality of M-estimates tends to improve. In particular, if $f(\varepsilon)$ has a bounded support, with range less than δ, then the consistency of $\hat{g}_{\mathrm{M}}(\cdot)$ at the jumps can be achieved.

From the computational viewpoint, problems of convergence of redescending M-smoothers are evident if one uses Newton-type algorithms as (8). In this case, the minimization of (5) would involve the second derivative of $\rho(z)$, i.e. first derivative of $\psi(z)$. Thus, if the latter is non-monotone, its derivative may be negative or may not exist, and the "Hessian" $\left[\sum_i v_i\, \rho''(\varepsilon_i)\right]$ is nonpositive definite. This leads the algorithm (8) in wrong directions and the numerical convergence toward local or global minima fails. To avoid these problems it is preferable to solve (5) with direct search methods, or with pseudolinear algorithms (10)-(11).

5. Conclusions

Motivated by airborne laser scanning, this paper has discussed robust nonparametric smoothers. We have proved their efficacy in fitting point data which contain various discontinuities. The jump-preserving ability of robust smoothers is due to the fact that they treat observations beyond the jumps as outliers. By ignoring such data, their local and adaptive properties are enhanced. The resulting surfaces can be used as 3D schemes for architectural reliefs, computer graphics and urban planning.

We have developed pseudolinear algorithms which are equivalent to nonlinear M-estimates, but have the advantage of resembling the linear kernel regression. This approach is derived through the weighted average form of M-estimates, and can be implemented in a sequential manner. We have shown that the best jump-preserving is provided by bounded loss functions, although they may lack consistency. Pseudolinear estimators can also be simplified as gradient-based filters, and this improves their performance in the case of piecewise constant surfaces.

The paper has also discussed practical methods to select smoothing coefficients. Those of the score components encounter problems of estimability due to the fact that discontinuity edges have area zero. A good tuning method consists in establishing a compromise between the efficiency and the robustness of the estimators.

References

Cheng, M.-Y., Fan, J. & Marron, J.S. (1997). On automatic boundary corrections, *Annals of Statistics*, **25**, 1691-1708.

Chu, C.-K., Glad, I., Godtliebsen F. & Marron, J.S. (1998). Edge-preserving smoothers for image processing, *J. of American Statistical Association*, **93**, 526-541.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, **74**, 829-836.

Fan, J., Hu, T.-C. & Truong, Y.K. (1994). Robust nonparametric function estimation, *Scandinavian Journal of Statistics*, **21**, 433-446.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modeling and its Applications*. London: Chapman & Hall.

Fox, J. (2002). *An R and S-PLUS Companion to Applied Regression*. Thousand Oaks (CA): Sage Publications.

Freedman, D.A. & Diaconis, P. (1982). On inconsistent M-estimators, *Annals of Statistics*, **10**, 454-461.

Grillenzoni, C. (1997). Recursive generalized M-estimators of system parameters, *Technometrics*, **39**, 211-224.

Hall, P. & Jones, M.C. (1990). Adaptive M-estimation in nonparametric regression, *Annals of Statistics*, **18**, 1712-17-28.

Hall, P. & Titterington, M. (1992). Edge-preserving and peak-preserving smoothing, *Technometrics*, **34**, 429-440.

Hampel, F., Ronchetti, E., Rousseeuw, P. & Stahel, W. (1986). *Robust Statistics: the Approach Based on Influence Functions*. New York: Wiley.

Härdle, W. (1991). *Smoothing Techniques: with Implementation in S*. Berlin: Springer.

Härdle, W. & Gasser, T. (1984), Robust non-parametric function fitting, *J. of Royal Statistical Society, B*, **46**, 42-51.

Hillebrand, M. & Müller, C.H. (2006). On consistency of redescending M-kernel smoothers, *Metrika*, **63**, 71-90.

Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.

Hwang, R.-C. (2004). Local polynomial M-smoothers in nonparametric regression. *J. of Statistical Planning and Inference*, **126**, 55-72.

Leung, D.H.-Y. (2005). Cross-validation in nonparametric regression with outliers, *Annals of Statistics*, **33**, 2291-2310.

Morgan, M. & Habib, A. (2002). Interpolation of Lidar data for automatic building extraction, *Proc. of ASPRS/ACSM Conference*, Washington D.C.

Jurečková, J. & Sen, P.K. (1996). *Robust Statistical Procedures*. New York: Wiley.

Polzehl, J. & Spokoiny, V.G. (2000). Adaptive weights smoothing with application to image restoration. *J. Royal Statist. Society, B*, **62**, 335-354.

Rottensteiner, F. (2003). Automatic generation of high-quality building models from Lidar data. *IEEE Computer Graphics and Applications*, **3**, 42-50.

Rue, H., Chu, C.-K., Godtliebsen, F. & Marron, J.S. (2002). M-smoother with local linear fit. *J. of Nonparametric Statistics*, **14**, 155-168.

Saint-Mark, P., Chen, J.-S. & Medioni, G. (1991). Adaptive smoothing: A general tool for early vision, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **13**, 514-529.

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization*. New York: Wiley.

Tsybakov, A.B. (1986). Robust reconstruction of functions by local-approximation method. *Problems of Information Transmission*, **22**, 133-146.

Wang, F. & Scott, D. (1994). The L1 method for robust nonparametric regression. *J. of the American Statistical Association*, **89**, 65-76.

Wang, M. & Tseng, Y.-H. (2004). LiDAR data segmentation and classification based on octree structure. `http://www.isprs.org/istanbul2004/comm3/papers/286.pdf`

Welsh, A.H. (1996). Robust estimation of smooth regression and spread functions and their derivatives, *Statistica Sinica*, **6**, 347-366.