A ROBUST METHOD FOR FILTERING NON-GROUND MEASUREMENTS FROM AIRBORNE LIDAR DATA

Fabio Crosilla, Domenico Visintini, Guido Prearo

Department of Geo-Resources & Territory, University of Udine, via Cotonificio, 114 I-33100 Udine, Italy
crosilla@dgt.uniud.it

KEY WORDS: LIDAR, DEM/DTM, Data, Surface, Detection, Algorithms

ABSTRACT:

This paper proposes a new filtering method of non-ground measurements from airborne LIDAR data through a Simultaneous AutoRegressive (SAR) analytical model and exploiting a Forward Search (FS) algorithm (Atkinson and Riani, 2000, Cerioli and Riani, 2003), a newly developed tool for robust regression analysis and robust estimation of location and shape.

In SAR models, with respect to classical spatial regression models, the correlation among adjacent measured points is taken into account, by considering two quantities for the measured dataset: a coefficient of spatial interaction and a matrix of point adjacency (binary digits for regular grids or real numbers for irregular ones).

FS approach allows a robust iterative estimation of SAR unknowns, starting from a subset of outlier-free LIDAR data, suitably selected. The method proceeds in its iterative computations, by extending such a subset with one or more points according to their level of agreement with the postulated surface model. In this way, worse LIDAR points are included only at the ending iterations. SAR unknowns and diagnostic statistical values are continuously estimated and monitored: an inferentially significant variation of the surface coefficients reveals as points included from now on can be classified as outliers or "non-ground" points.

The method has been implemented using Matlab® language and applied either to differently simulated LIDAR datasets or really measured points, these last acquired with an Optech® ALTM 3033 system in the city of Gorizia (North-East Italy). For both kinds of datasets the proposed method has very well modeled the ground surface and detect the non-ground (outliers) LIDAR points.

1. INTRODUCTION

Airborne Laser Scanning technique is extremely efficient to fulfil increasing demand of high accuracy Digital Terrain or Surface Models (DTM or DSM) for civil engineering, environment protection, planning purposes, etc. But, if standard procedures for acquiring Airborne Laser Scanning data have already come nowadays a long way, on the other hand, the choice of appropriate data processing techniques for different particular applications is still being investigated. For this last essential topic of research, several algorithms have been developed for semi automatically/automatically extracting of objects from bare terrain. But in general, their filtering efficiency seems to vary very much with local conditions. In fact, the quality of nearly all procedures too often depends on an appropriate setting or determination of thresholds and control values (Jacobsen *et al.*, 2002, Kraus, 1997, Voelz, 2001). Moreover, another important task not yet completely solved is to simultaneously proceed to both filtering and generation of DTM. For this last requirement, the filtering algorithm presented throughout this paper manages not only to "remove" additional features on ground such as buildings, vegetation etc., but even to generate DTMs with points classified as "ground".

Looking thought the recent literature in LIDAR data filtering, a significant number of techniques has been developed to remove man-made "artefacts" on the territory, in order to obtain the true Digital Terrain Model. Unfortunately, in order to completely remove non-terrain data points, these techniques often require interactive editing. This leads to increasing the production times. Thus, there is yet great interest in developing effective and reliable tools and algorithms on this topic.

Our research starts from the analysis of the most significant techniques and algorithms present in literature; that is:

- Least squares interpolation (Kraus e Pfeifer, 1997): filter out trees in forested areas by fitting an interpolating surface to the data and using a weighted ground iterative least squares scheme to bring down the contribution of points above the surface, so that it gets closer and closer to the lowest data points. A similar approach is used to filter out also buildings (Rottensteiner *et al.*, 2002).

- Erosion/dilatation functions in mathematical morphology (Zhang *et al.*, 2003): starting from an initial subset of points and by gradually increasing the window size of the filter using elevation difference thresholds, data of vehicles, vegetation, and buildings are removed, while ground data are preserved. Such points are then included in a DTM.

- Slope based functions (Vosselman, 2003): slope based filtering operates using mathematical morphology, and fixing a slope threshold. This, being the maximum allowed height difference between two points, is expressed as a function of the distance between different terrain points.

- TIN densification (Axelsson, 2000): an adaptive TIN model born to find ground points in urban areas. Initially seed ground points within a user defined grid of a size greater than the largest non ground features are selected to compose an initial ground dataset. Then, one point above each TIN facet is added to the ground dataset at each iteration if its parameters are below specific threshold values. Different thresholds have to be given for various land cover types.

- Application of Spline functions (Brovelli *et al.*, 2002): through a least squares approach with Tikhonov regularization, non-terrain points are filtered out by analyzing residuals from a spline interpolation.

This paper proposes instead a new stochastic approach for filtering, based on the following spatial regression model.

2. SIMULTANEOUS AUTOREGRESSIVE (SAR) MODELS FOR SPATIAL FILTERING

The analytical models called as SAR (*Simultaneous Auto Regressive,* Whittle, 1954) belong to a class of algorithms largely used in many fields for describing spatial variations.

Their nature is rather different from models usually implemented in time-series analysis. This is mainly due to the fact that, while the natural flow of time from past to present to future imposes a natural ordering or direction on patterns of interaction, a two-dimensional model generally possesses no such equivalent ordering.

Hence, our filtering algorithm works under the hypothesis that LIDAR measures of terrain/objects height can be rightfully represented by means of SAR models. At this step of research, first order isotropic SAR models have been employed.

To introduce SAR models, one can start from the very simple expression of a n-dimensional measurement (observation):

$$\mathbf{z} = \mathbf{\mu} + \mathbf{\varepsilon} \tag{1}$$

where (specialising it for LIDAR data):

- $\mathbf{z}$ is the [n x 1] vector of LIDAR height values (being n the total number of points to be filtered),
- $\mathbf{\mu}$ is the [n x 1] vector of "true" terrain height values,
- $\mathbf{\varepsilon}$ is the [n x 1] vector of independent and normally distributed errors (noise) with mean 0 and variance σ_ε^2.

Considering now for height errors, the effect of a global interaction and the spatial dependence among points, height values (1) can be rewritten as:

$$\mathbf{z} = \mathbf{\mu} + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{\varepsilon} \tag{2}$$

where:

- $\mathbf{I}$ is the [n x n] identity matrix,
- ρ is a value (constant for the whole dataset) that measures the mean spatial interaction between neighbouring points,
- $\mathbf{W}$ is a [n x n] spatial weight (binary) matrix defined as:

$$\mathbf{W} = \begin{cases} w_{ij} = 0 \\ w_{ij} = 1 \end{cases}$$

where $w_{ij} = 1$ if the points are neighbours, $w_{ij} = 0$ otherwise.

In a regular grid scheme (lattice), a common definition of a neighbourhood structure is that $w_{ij} = 1$ if the *j-th* point is immediately North, South, East or West of the *i-th* point.

But, since to grid LIDAR data for a lattice scheme leads to a loss of information, we preferred operate with raw data, that is non-lattice displacement points. In such a non-regular case, $\mathbf{W}$ can be computed by means of one of the following methods:

- Distance functions: each LIDAR height measure is linked to all others within a specified distance;
- Nearest neighbor approach: each LIDAR measure is linked to its *k* (*k* = 1, 2, 3, …n) nearest neighborhoods;
- Gabriel's graph: two generic LIDAR measures *i* and *j* are said to be contiguous if and only if all other measures are outside the *i-j* circle, that is, the circle whose circumference *i* and *j* are at opposite points;
- Delaunay triangulation: all LIDAR measures which share a common border in a Dirichlet partitioning of the area are joined by an edge.

This last option was chosen, so obtaining a matrix $\mathbf{W}$ with no more binary digits but real numbers (Pace, Barry and Sirmans, 1988); furthermore, a row standardization of $\mathbf{W}$ is applied by:

$$\mathbf{W} = w_{ij} \Big/ \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij}$$

For (2) to be meaningful, it is assumed that $(\mathbf{I} - \rho\mathbf{W})^{-1}$ exists; this condition leads to restrain the range of values of ρ in the interval $0 \le \rho < 1$.

An important task is given from the modelling of $\mathbf{\mu}$, containing the "true" height terrain values of (1) by means of some polynomial function of East-North coordinates, in such a way to analytically define the so-called "trend surface":

$$\mathbf{\mu} = \mathbf{A}\mathbf{\theta} \tag{3}$$

where:

- $\mathbf{A}$ is a [n x r] matrix, with $\mathbf{A}_i = \begin{bmatrix} 1 & E_i & N_i & ... & E_i^s & N_i^s \end{bmatrix}$ as rows and where s = (r-1)/2,
- $\mathbf{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & ... & \theta_{r-1} \end{bmatrix}^T$ is a [r x 1] vector of coefficients.

Equation (3) represents the classical autoregressive problem: the estimation of trend coefficients $\mathbf{\theta}$ by the measured points. Substituting (3) in (2), the general SAR model finally arises:

$$\mathbf{z} - \mathbf{A}\mathbf{\theta} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{\varepsilon} \tag{4}$$

This form shows the main characteristic of SAR models: they require/permit the *simultaneous autoregressive* estimation of both trend $\mathbf{\theta}$ and interaction ρ process parameters.

Moreover, writing (4) explicitly, we obtain:

$$z_i = \mu_i + \varepsilon_i + \rho \sum_j^{N_i} w_{ij}(z_j - \mu_j) \tag{5}$$

with N_i number of sites joined to *i* by an edge (its neighbours). Following equation (5), the *i-th* measured height value can be understood as the algebraic sum of two terms: the stochastic surface modelling on *i-th* point ($\mu_i + \varepsilon_i$) and the global effect of errors of such a stochastic modelling on its surrounding points ($z_j - \mu_j$) via the spatial interaction ρ.

3. ESTIMATION OF SAR UNKNOWNS AND OUTLIERS

Estimation of a spatial autoregressive process with dependent variables can be taken over through different approaches.

A first problem related to this task is due to the computational dimension: dealing with millions of LIDAR measures, requires great amount of memory storage. For our method, computational counts for operation, such as determinants and inverses seen in (4), grow with the cube of n. For this computational problem, the acquired strip laser data has to be suitably shared in sub-zones of about 15.000 points each.

Moreover, as an analytical problem, traditional maximum likelihood techniques require non-linear optimization processes using either analytic derivatives or finite difference approximations. Unfortunately these can fail to find the global optimum and do so without informing the user of their failure. Hence summarising, an ideal spatial estimator would:

- handle large datasets,
- handle point estimation and inference quickly,
- not rely on local non-linear optimization algorithms.

In order to deal with the last requirement, in the following section, the so-called Maximum Likelihood (ML) method for estimate unknown parameters within SAR model is presented.

3.1 Maximum Likelihood computations

For our purposes, a maximum likelihood approach for the estimate of unknown SAR parameters has been chosen. Let us start from the general likelihood function:

$$l(\boldsymbol{\theta},\rho,\sigma^2) = (2\pi\sigma^2)^{-n/2}|\mathbf{I}-\rho\mathbf{W}|\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{z}-\mathbf{A}\boldsymbol{\theta})^T\boldsymbol{\Sigma}(\mathbf{z}-\mathbf{A}\boldsymbol{\theta})\right\} \quad (6)$$

where:

$$\boldsymbol{\Sigma} = (\mathbf{I}-\rho\mathbf{W})^T(\mathbf{I}-\rho\mathbf{W})$$

is the weight matrix, symmetric and positive definite; unfortunately and differently with respect to ordinary estimations, here the weight matrix contains an unknown term as ρ. It is then necessary to maximize (6) not only with respect to $\boldsymbol{\theta}$ and σ^2, but also with respect to ρ.

It can be performed in stages (Pace, Barry and Sirmans, 1998) by selecting a vector of length f of values over [0,1] labelled as:

$$\rho_v = \begin{bmatrix} \rho_1 & \rho_2 & & \rho_f \end{bmatrix}$$

and considering the log-likelihood function:

$$L(\boldsymbol{\theta},\rho,\sigma^2) = \ln|\mathbf{I}-\rho\mathbf{W}| - \left(\frac{n}{2}\right)\ln(\mathbf{e}_0{}^T\mathbf{e}_0 - 2\rho\mathbf{e}_d{}^T\mathbf{e}_0 + \rho^2\mathbf{e}_d{}^T\mathbf{e}_d)$$

where:

- $\mathbf{e}_0$ are the residuals from an Ordinary Least Squares (OLS) regression of $\mathbf{z}$ on $\mathbf{A}$,
- $\mathbf{e}_d$ are the residuals from an OLS regression of $\mathbf{Wz}$ on $\mathbf{A}$.

Thus, to maximize (6), the following m terms are evaluated:

$$\begin{bmatrix} L(\boldsymbol{\theta},\rho_1,\sigma^2) \\ L(\boldsymbol{\theta},\rho_2,\sigma^2) \\ ... \\ L(\boldsymbol{\theta},\rho_f,\sigma^2) \end{bmatrix} = \begin{bmatrix} \ln|\mathbf{I}-\rho_1\mathbf{W}| \\ \ln|\mathbf{I}-\rho_2\mathbf{W}| \\ ... \\ \ln|\mathbf{I}-\rho_f\mathbf{W}| \end{bmatrix} - \left(\frac{n}{2}\right)\begin{bmatrix} \ln(\mathbf{e}_0{}^T\mathbf{e}_0 - 2\rho_1\mathbf{e}_d{}^T\mathbf{e}_0 + \rho_1^2\mathbf{e}_d{}^T\mathbf{e}_d) \\ \ln(\mathbf{e}_0{}^T\mathbf{e}_0 - 2\rho_2\mathbf{e}_d{}^T\mathbf{e}_0 + \rho_2^2\mathbf{e}_d{}^T\mathbf{e}_d) \\ ... \\ \ln(\mathbf{e}_0{}^T\mathbf{e}_0 - 2\rho_f\mathbf{e}_d{}^T\mathbf{e}_0 + \rho_f^2\mathbf{e}_d{}^T\mathbf{e}_d) \end{bmatrix} \quad (7)$$

and the value ρ_{ML} giving the maximum log-likelihood value L among those in (7) is assumed as the ML estimation $\hat{\rho}$ of ρ.

The use of a finite set of ρ will cause some small granularity in the chosen values ρ_{ML}, but it should not prove difficult to make the granularity small relative to the statistical precision of estimated ρ_{ML}. While this approach may suffer a small loss of precision relative to non-linear maximization, the evaluation of log-likelihood function over an interval ensures robustness, the main property of this approach (Griffith and Layne, 1999).

Once so obtained $\hat{\rho}$, it is possible to ML estimate SAR unknowns and also a new weight (optimised) matrix $\hat{\boldsymbol{\Sigma}}$:

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}\mathbf{z} \quad (8.1)$$

$$\hat{\sigma}^2 = n^{-1}(\mathbf{z}-\mathbf{A}\hat{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}(\mathbf{z}-\mathbf{A}\hat{\boldsymbol{\theta}}) \quad (8.2)$$

$$\hat{\boldsymbol{\Sigma}} = (\mathbf{I}-\hat{\rho}\mathbf{W})^T(\mathbf{I}-\hat{\rho}\mathbf{W}) \quad (8.3)$$

3.2 Spatial outliers searching

A spatial outlier is defined as "an observation which is unusual with respect to its neighbouring values" (Haining, 1990). For our purposes, the way to assess spatial outlyiness is to compute individual departure from the fitted polynomial trend surface.

To accomplish this goal, starting from (4), the vector $\mathbf{e} = \sigma^{-1}\boldsymbol{\varepsilon}$ of standardised residuals is estimated as:

$$\mathbf{e} = \hat{\sigma}^{-1}(\mathbf{I}-\hat{\rho}\mathbf{W})(\mathbf{z}-\mathbf{A}\hat{\boldsymbol{\theta}}) \quad (9)$$

where $\hat{\rho}$, $\hat{\boldsymbol{\theta}}$, $\hat{\sigma}$ are the unknown parameters of the SAR model just simultaneously estimated by means of (7) and (8). Afterwards, its n components are inferentially evaluated to find which measures do not fit the estimated surface: vector $\mathbf{e}$ defines in fact the lack of fit statistic $\mathbf{e}^T\mathbf{e}$.

Standardised residuals $\mathbf{e}$ over residuals $\boldsymbol{\varepsilon}$ have been thus preferred, since they allow a robust spatial autocorrelation estimation, which we believe is a sensible property for the purpose of detecting spatial outliers.

From the methodological point of view, the main property of our algorithm allowing to detect LIDAR outliers is to perform estimations (7), (8) and (9) on different subset of the whole dataset. In particular, we start from an initial subset of LIDAR data up to include all the dataset of the sub-zone to be filtered.

4. IMPLEMENTATION OF AN ITERATIVE SEARCHING PROCEDURE

4.1 Block Forward Search for SAR models

An interesting algorithm to perform iterative SAR estimations on increasing datasets is the so-called "Block Forward Search" (BFS) proposed by Atkinson and Riani (2000) and Cerioli and Riani (2003) for econometric purposes.

It makes possible to proceed to the joint robust estimation $\hat{\rho}$ and $\hat{\boldsymbol{\theta}}$ at each step of the search, starting from a partition of datasets in n blocks of contiguous spatial location, and considers these blocks as its elementary unit. In the general case of grid data, each block is a set of cells, while handling raw data is difficult to univocally create the blocks and so the block dimension is merely unitary (UFS, Unitary Forward Search).

The basic idea of the BFS approach is to repeatedly fit the postulated model $\boldsymbol{\mu} = \mathbf{A}\boldsymbol{\theta}$ to increasing subsets size, selecting for any new iteration the observations $\mathbf{z}$ best fitting the previous subset, that is having the minimum standardised residuals $\mathbf{e}$ computed by (9). It must to be stressed as in equation (9), $\hat{\rho}$ and $\hat{\boldsymbol{\theta}}$ are estimated on the subset outlier free only, while $\mathbf{z}$, $\mathbf{A}$, $\mathbf{W}$ and σ terms are referred to the whole dataset with outliers.

Thanks to the strategy of block growing, the outliers present into $\mathbf{z}$ values are included only at the end of the BFS procedure.

4.2 SFS implemented algorithm

The proposed algorithm, called simply SFS (Spatial Forward Search), implements the Forward Search approach, but, since raw LIDAR data are irregularly located, unitary blocks to increase the subset size were chosen: in other words, only one measured point enlarges the subset at each iteration.

The SFS algorithm has been implemented as a software tool using Matlab® language. Its main steps are (see Figure 1):
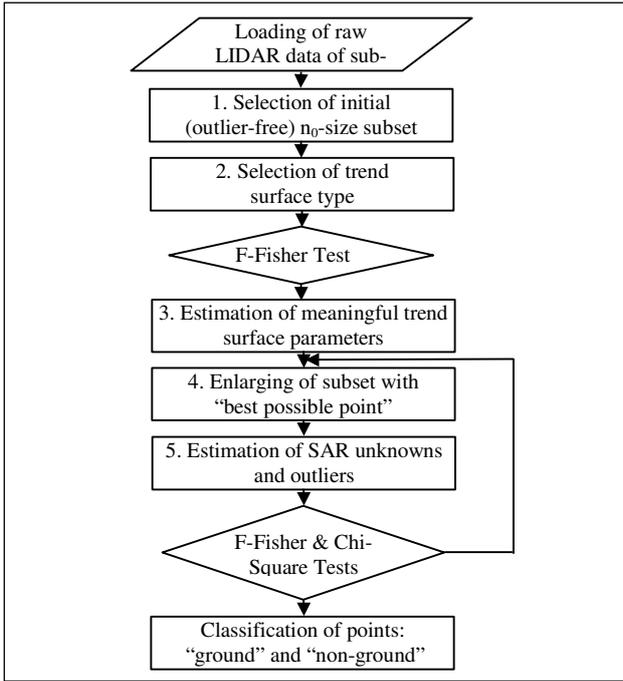
Figure 1: Work-flow of the SFS filtering algorithm.

1. Selection of the initial subset of $(n_0 << n)-size$: this is meant to be outlier free, containing then terrain (ground) points only. Many automatic criteria could be implemented for this purpose, mostly evaluating data variations statistics (e.g. least median), but as a general statement, a user defined graphic selection has to be preferred.

2. Selection of trend surface type: for modelling the subset ground surface by (3), the user chooses a redundant k-degree polynomial (e.g. cubic k = 3) (see Figure 2).
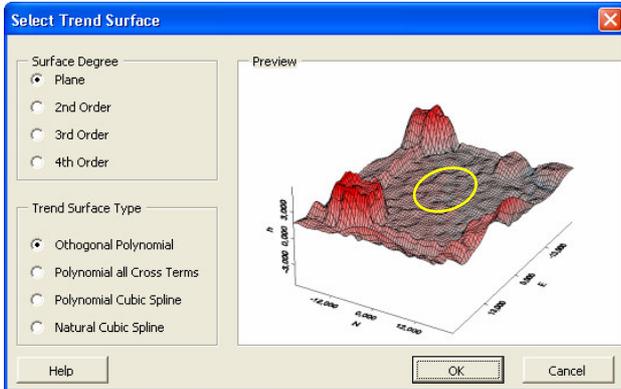
Figure 2: Selection of trend surface type in SFS: 3D-view of (simulated and noisy) LIDAR points and initial subset.

3. Estimation of meaningful subset trend surface parameters:
Once estimated $\hat{\theta}$ by (8), the assessment of a reduced s < k degree, describing with plenty sensitivity the trend, is performed by an inferential F-Fisher Test, so skipping not meaningful (k-s) parameters in $\hat{\theta}$. In such a way, r = 2s+1 is the size of the engaged polynomial coefficient vector.

Once steps 1.÷3. are accomplished, the program goes on iteratively enlarging the initial n_0-size subset up to the n-size dataset. For each *m-th* iteration, $\hat{\rho}, \hat{\theta}, \hat{\sigma}_{dataset}, \hat{\sigma}_{subset}, \hat{e}$ are re-estimated by (7), (8) and (9). Furthermore, also other statistical

quantities are computed, allowing to diagnostically monitor either the trend surface modelling (see Figures 3. & 4.) or the outlier searching (see Figure 5).

4. Enlarging of subset: its size grows from m to (m+1) adding the point with smallest absolute standardised residual e_b given by (9). This point is called "best possible point", since it best fits the trend surface although it does not (yet) belong to the subset; anyway, it can be classified as "ground" point.

5. Estimation of SAR unknowns and the best point detection is iteratively computed by (7), (8) and (9), on the *(m+1)-th* subset of (m+1)-size composed with ground points only.

Steps 4. & 5. are then iteratively repeated until Chi-square Test on $\hat{\sigma}^2$ variation and F-Fisher Test on $\hat{\theta}$ variation (with respect to the initial ones) do not reveal that best possible point is really an outlier. In fact, as known, the presence of outliers among observations damages the estimation of $\hat{\sigma}^2$ and $\hat{\theta}$, as can be easily view in the right sides of Figures 3. & 4. Moreover, any new point included from now on up to the whole dataset, can be classified as outlier or "non-ground".

As last consideration, it has to stress how the same classification of points as ground/non-ground would be impossible considering instead the whole dataset for masking effect on components of **e** (see Figure 6 for last iteration/abscissa).
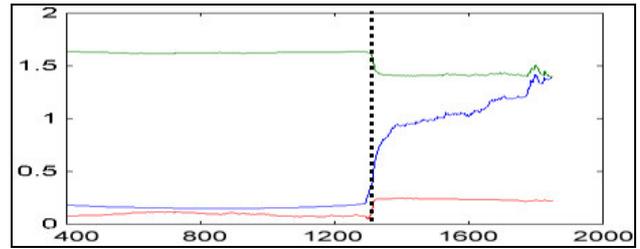
Figure 3: Values of $\hat{\sigma}_{dataset}$ (green), $\hat{\sigma}_{subset}$ (blue), $\hat{\rho}$ (red).
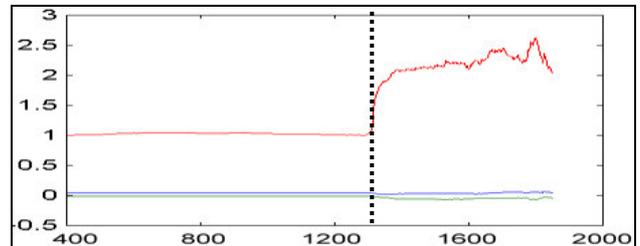
Figure 4: Values of θ_0 (red), θ_1 (blue), θ_2 (green).
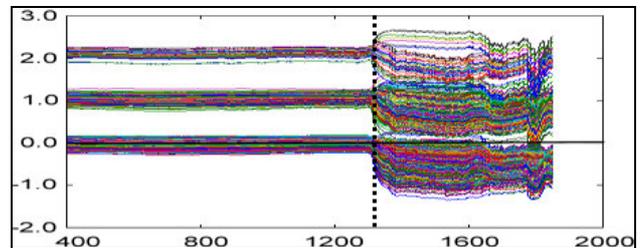
Figure 5: Values of n components of **e** along the iterations.

Once the SFS program has been carried out, the trend parameters of the ground are those relating to the maximum subset outlier-free and every point is binary classified as "ground" (0, green in Figure 11) or "non-ground" (1, red in same Figure 11). Starting now from this classification, could be possible to repeat whole SFS processing on outlier points only, to find other small surfaces, e.g. building roofs; the developing and implementation of this idea is currently in progress.

5. APPLICATIONS OF SFS FILTER ON LIDAR DATA

The SFS program has been tested both on differently simulated LIDAR datasets and really measured points acquired with an Optech[®] ALTM 3033 airborne system.

5.1 Testing on simulated data

As far as simulated datasets are concerned, a lot of experiments has been carried on, here reporting 6 tests differing for surface type (plane and quadratic), for value of spatial interaction ρ and for mean noise |ε|. In each dataset, with irregularly spaced points, the presence of some buildings (outliers of the ground surface) has been simulated. The number of ground and non-ground points is then exactly known, so that the efficiency of the algorithm could be easily verified.

General characteristics of these 6 examples, simulating real survey conditions, are reported in Table 6.

Surface type	Plane (r=3)	2nd Order (r=5)
Polynomial coefficients	θ_0=1,000 θ_1=0,050 θ_2=-0,010	θ_0=1,000 θ_1=+0,005 θ_2=-0,001 θ_3=+0,0015 θ_4=-0,002
Uncorrelated noise σ_ε over surface	Plan-1: 0,10 m Plan-2: 0,20 m Plan-3: 0,25 m	Quad-1: 0,10 m Quad-2: 0,20 m Quad-3: 0,25 m
Spatial interaction ρ	Plan-1: 0,0 Plan-2: 0,1 Plan-3: 0,2	Quad-1: 0,0 Quad-2: 0,1 Quad-3: 0,2
Number of points (n)	1.886	
Raw data (not grid)	Yes	
Points sampling	1 point/m^2 (mean)	
Dataset area	1.760 m^2	
Δz	13,6 m	
Number of "building points"	413 (mean)	
Courtyard closed areas	Yes	

Table 6: Summary of simulated LIDAR data.

Processing such datasets by SFS (Figures 3÷5 relate to Plan-3) has given very satisfactory results: ground trend surface and building/outlier have been well detected (see Table 7).

Detection of surface type	Correct
Coefficient estimation	Correct within 5%
Statistical errors on classification: 1st kind (false outlier) 2nd kind (false ground)	0,0% 1,7%
ρ estimate	Correct within 10%

Table 7: SFS filtering of the simulated data: general results.

The performance of the SFS for classification can be significantly validate by applying onto same datasets the program TerraScan[®] (Soininen, 2003), a very well known software for LIDAR data processing developed by Terrasolid Ltd. A binary classification (ground/non-ground) was obtained by suitably exploiting the following routines:

1. "Classify ground": classifies ground points by iteratively building a triangulated surface model.
2. "Low points": classifies points that are lower than other points in the vicinity. It is often used to search for possible error points that are clearly below the ground.

3. "Below surface": classifies points that are lower than other neighbouring points in the source class. This routine was run after ground classification to locate points that are below the true ground surface.
4. "By height from ground": classifies points that are located within a given height range when compared with ground point surface model.

Comparison among true, SFS and TerraScan classification results is shown in Figure 8. As a general statement, we can say:

- SFS provides about 2% of errors of second statistical kind (false ground), so that some outlier has not been detected;
- TerraScan[®] seems to commit more than 10% of first kind errors (false outlier), so that many points were "rejected", although they belong to the ground (but noisy) surface.
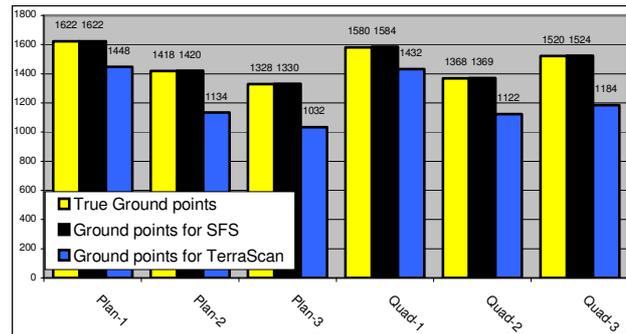
Figure 8: True vs. SFS vs. TerraScan classification of points.

5.2 Testing on really acquired data (city of Gorizia)

To evaluate LIDAR technology for DTM production, millions of points were acquired in October 2003 over the city of Gorizia with an airborne Optech[®] ALTM 3033 laser scanning system. Data strips have been split into different sub-zones, in order to avoid heavy computations with huge quantities of memory storage, but anyway still being capable to test the efficiency of the SFS method for real cases. General characteristics of sub-zones are reported in Table 9.

Surface type	Urban area
Data type	First & Last pulse
Number of points (n)	15.000 (mean for sub-zone)
Raw data (not grid)	Yes
Points sampling	1 point/m^2 (mean)
Dataset area	15.000 m^2 (mean)
Δz	44,3 m
Vegetation	Yes
Buildings	Yes
Courtyard closed areas	No

Table 9: Summary of Optech[®] LIDAR data on Gorizia.

The sub-zone submitted to test is the downtown square, mainly constituted of quasi-horizontal plane terrain; furthermore different types of building were present, together with high and low vegetation and a lot of parked cars. No power-lines or other structures were present.

LIDAR points were processed either by SFS or by TerraScan: with this last software, firstly objects are classified in two classes: ground and non-ground points. Successively, other classes such as buildings and vegetation were detected yet.

The difference among SFS/TerraScan classifications regards 679 points (4,5% on 14.953 total points), ranked as "ground"

with SFS and "non ground" with TerraScan (see Figure 10): this is in agree with results obtained for simulated data.
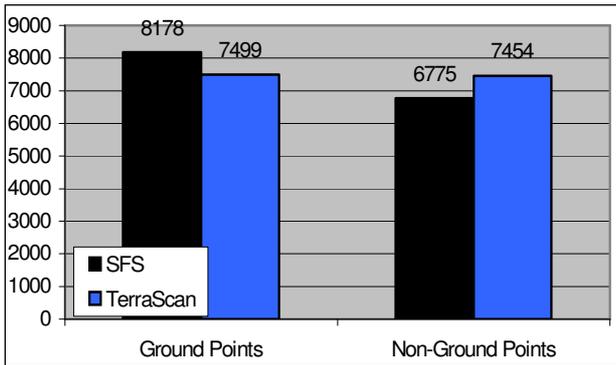
Figure 10: SFS vs. TerraScan classification of Gorizia points.

About the location of the two classifications, while they substantially correspond for building and vegetation zones, the disparities mainly occur for roads and parking areas.
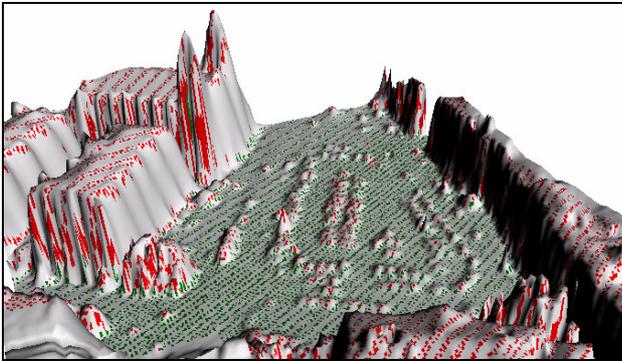
Figure 11: LIDAR DSM of the city of Gorizia (Italy).

Figure 11 shows the DSM obtained with such LIDAR data. The results of filtering/classification by SFS algorithm are painted over with green spots when ground classified, red spots otherwise (outliers). They seem to be very truth, so the SFS correctness is qualitatively proved, being very hard to exactly evaluate it quantitatively (cars positions are unknown/variable).

6. CONCLUSIONS AND FUTURE PERSPECTIVES

This paper illustrates a new robust technique for the filtering of non-ground measurements from airborne LIDAR data. The algorithm represents an efficient method for automatic classification of LIDAR data, mainly based on a newly developed tool for robust regression analysis and robust estimation of location and shape. The main advantage of using SAR models and BFS algorithm relies not only on its accuracy but also on its statistical robustness. It makes possible to efficiently and simultaneously detect either trend surface or outlier points by suitably enlarging a data subset.
Through a significant number of examples, the paper shows how the proposed SFS method is a valuable tool for the purpose of filtering LIDAR height measures and terrain modeling. Here, examples of rough terrain were processed, showing that the method can deal with dataset containing many break lines. Besides the Gorizia dataset shown in this paper, the method is currently being applied to other large urban datasets of increasing point density, for the goal of building extraction also. In a near future, to improve the efficiency of the SFS algorithm, other space interaction models will be tested, as second order SAR models and *Conditional AutoRegressive* models (CAR).

REFERENCES

Atkinson, A.C., Riani, M., 2000. Robust Diagnostic Regression Analysis. Springer, New York.

Axelsson, P., 2000. DEM Generation from laser scanner data using TIN adaptive models. International Archives of Photogrammetry & Remote Sensing , vol. XXXIII, part B3, pp. 85-92.

Brovelli, M.A., *et al.*, 2001. Modelli matematici del terreno per mezzo di interpolatori a spline. Bollettino SIFET, supplemento speciale al n. 2, pp. 55-80.

Cerioli, A., Riani, M., 2003. Robust methods for the analysis of spatially autocorrelated data. Journal of the Italian Statistical Society, 11, pp. 334-358.

Griffith, D.A., Layne L.J., 1999. A casebook for spatial statistical data analysis. Oxford University Press, New York.

Haining, R., 1990. Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge.

Jacobsen, K., 2001. New Developments in Digital Elevation Modelling. Geoinformatics, June, pp. 18-21.

Kraus, K., 1997. Eine neue Methode zur Interpolation und Filterung von Daten mit schiefer Fehlerverteilung. Österreichische Zeitschrift für Vermessung & Geoinformation 1, pp. 25-30.

Kraus, K., Pfeifer, N., 1997. A new method for surface reconstruction from laser scanner data. International Archives of Photogrammetry & Remote Sensing, vol. XXXII, part 3-2W3, pp. 80-86.

Pace, R.K., Barry, R., Sirmans, C.F., 1998. Quick computation of spatial autoregressive estimators. Journal of Real Estate Finance and Economics, pp. 1-12.

Rottensteiner, F., *et al.,* 2002. LIDAR Activities at the Viennese Institute of Photogrammetry and Remote Sensing. Proceedings of the 3rd Int. LIDAR Workshop, Columbus.

Soininen, A., 2003. TerraScan. User Guide.

Voelz, H., Mathematische Grauwertmorphologie zur Filterung Digitaler Oberflächenmodelle, Diplomarbeit, Universität Hannover, http://www.ipi.unihannover.de/html/lehre/ diplomarbeiten/2001/2001.htm

Vosselman, G., 2000. Slope based filtering of laser altimetry data. International Archives of Photogrammetry & Remote Sensing, vol. XXXIII, part B4, pp. 958-964.

Whittle, P., 1954. On stationary process in the plane. Biometrika, 41, pp. 434-449.

Zhangh, K., *et al.,* 2003. A Progressive Morphological Filter for Removing Non ground Measurements From Airborne LIDAR Data. IEEE Transactions on Geoscience and Remote Sensing, 41, 4, pp. 872-882.