HEIGHT TEXTURE OF LOW VEGETATION IN AIRBORNE LASER SCANNER DATA AND ITS POTENTIAL FOR DTM CORRECTION

Ben Gorte[a], Norbert Pfeifer[a], Sander Oude Elberink[b]

[a]Delft Insitute of Earth Observation and Space systems, TU Delft,
Kluyverweg 1, 2629HS Delft, The Netherlands - b.g.h.gorte@lr.tudelft.nl, n.pfeifer@lr.tudelft.nl
[b]Adviesdienst Geo-Informatie en ICT, Ministerie voor Rijkswaterstad,
Derde Werelddreef 1, 2629HA Delft, The Netherlands – S.J.Oude-Elberink@agi.rws.minvenw.nl

KEY WORDS: airborne laser scanning, vegetation, texture, DEM/DTM, correction

ABSTRACT:
The influence of low vegetation on laser scanning and its disturbing effect (a systematic positive height shift) during DTM generation is well recognized. Based on our experience in a previous study of estimating the effect of medium-height vegetation (shrubs and bushes) in rough terrain by using point-cloud based co-occurrence texture methods, we now investigate the effect of lower vegetation, such as grass, in very flat marshland conditions. When vegetation is very low, and thus the effect very small, it appears difficult to separate the systematic shift from random measurement noise. When the shift becomes larger, however, it shows significant correlation with texture measures such as slope texture and standard deviation. These measures are derived from the laser data itself and do not require any additional information.

1. INTRODUCTION

Airborne laser scanning is a method to measure points on the earth surface from a flying platform (plane or helicopter). These points can be used to generate a digital terrain model (DTM), which is then used in various applications. Especially for hydrological applications it is necessary that the DTM de-scribes the bare earth, as this is the level where water runs off superficially or seeps away. Also volume determinations or fixing ground water gauge levels require precise DTMs. From many investigations (Sec. 2), it is known that laser scanning does not always measure the distance between the laser scanner device and a point the bare earth, but vegetation influences the measurement, and a range that is too short is recorded. Especially low vegetation (heights below 20cm, e.g. grass or branches lying on the earth) is causing problems in determining the bare earth elevation. The bare earth elevation is considered here as the transit surface from dry soil (rock) to the air.

This paper attempts to quantify and correct the disturbing influences of low vegetation on DTM determination. Earlier investigations have been carried out by AGI, the geo-information section of the Dutch Ministry for Transport, Public Works and Water Management, and the TU Delft, Section of Photogrammetry and Remote Sensing, see Sec. 2. During this research the possibility of using texture measures for correcting the data was studied. The laser data used in this prior research was not typical for the laser data used for country-wide generation of DTMs, resulting in the following re-search question: Determine if statistical measures and texture measures can be used for reliably estimating the disturbing influence of low vegetation on the DTM height obtained from airborne laser scanning within the DTM production process.

1.1 Previous study

A previous study (Pfeifer et al, 2004) investigated the influence of vegetation on laser altimetry measurements. In vegetated areas, laser points generally appear too high; in other words a positive height shift occurs in the measurements under the influence of vegetation. The study investigated the interaction of laser beams with vegetated terrain and attempted to estimate shift on the basis of:

- Additional terrain data (vegetation classes)
- Features derived from the laser data (texture)

It was found that there is positive correlation between certain measures of texture and height shift. Very dense laser data (10 points/m^2) was analyzed, and shifts of +7.3cm were found for long dense grass, +9.4cm for the ground of a young forest, and +11.6cm for old willow forest.

However, the data that were used in that study were somewhat uncontrolled, containing very diverse vegetation classes on sometimes very rough terrain. Also the control points were sparse compare to the laser points and the selection criteria for control points were not explicit. Finally, there was too much time between the acquisition of laser and control points.

The current study repeats the experiments under more controlled circumstances: flat terrain, low vegetation, control point measurements at the same day as laser measurements. For terrain with higher and varying inclinations detrending, i.e. subtraction of a DTM featuring the landforms, is necessary first. This extension of the topic is not studied here.

From the other side, the point density is much lower now (1 point / m^2 instead of 10 p/m^2) and the vegetation type is more uniform throughout the study area. Therefore, texture is not expected to be such a characteristic feature. The goal of this study is to investigate the correlation between texture and shift in these conditions. If this correlation is strong enough and consistent over the various test sites, than we may conclude that shift can be predicted from texture, and we would be able to improve heights by subtracting the predicted shift from the raw measurements. The research aim is, therefore, to determine to which extent statistical measures, especially texture measures, are reliable for estimating the influence of low vegetation on laser scanning measurements. The emphasis is on relatively open areas with gras, reed, and similar vegetation. For these types of area the water level decisions are very critical.

1.2 Related work

Many research groups around the world realized the influence of low vegetation on laser scanning and its disturbing influence for DTM generation. Below, a short overview on the findings of those groups is given.

In (Pfeifer et al 2001) DTM accuracy has been investigated, which can be higher than laser measurement accuracy due to elimination of random errors in the modelling process. Within one flight and using roughly 800 terrestrial check points, the results obtained were in a street without cars: ±1.0cm, street height below parking cars: ±3.7cm, in an open park area (grass): ±4.5cm, park with few trees: ±7.8cm, and park with dense tree stocking: ±11.1cm. Additionally, there is a systematic shift of the DTM above the check points, growing in the same manner as the accuracies.

In (Ahokas et al 2003) the land cover classes investigated are asphalt, gravel, and grass (flying height 550m), and forest ground (flying height 400m), and comparison to roughly 3500 ground points lead to ±10cm, ±4cm, ±11cm, and ±17cm, respectively, but a less systematic behavior in the upward shift of the laser points compared to the ground points.

(Bollweg and de Lange 2003) investigated horizontal areas and found an upward shift of 8cm for long dense grass, and 4cm on solid ground, and the standard deviation of the laser measurements increased from ±7cm on solid ground to ±14cm on low vegetation, whereas standard deviations of the terrestrial control measurements showed an increase from ±3cm to ±7cm.

In (Oude Elberink and Crombaghs 2004) it is shown that upward shifts occurred up to 15cm on low vegetation areas (creeping red fescue, thrift). A relation could be seen between the density of the vegetation coverage (and height) and the systematic error: 0% coverage meant no upward shift, 100% coverage showed a 15cm shift.

(Hodgson and Bresnahan 2004) investigate a laser dataset with 650 checkpoints, for the land cover classes pavement, low grass, high grass, brush/low trees, evergreen, and deciduous, and conclude that the accuracy of the elevation measurement is between 17cm (evergreen) and 26cm (deciduous). The observed shifts are in their study all very small, between +6cm (overprediction for pavement) and -6cm (underprediction for brush/low tree).

The above contributions worked according to the following principle. Laser scanner data is compared to field data measured by tacheometry or GPS, which is considered to be of superior quality. During the field measurements the characteristics of the vegetation or landcover were recorded. For correcting laser measurements within these approaches it is necessary to know the vegetation class, something which normally not available without fieldwork. Additionally, the question arises, what are good land cover class descriptions from the “laser scanner point of view”, and if two instances of one class (e.g. “grass”) always have the same influence on the measurements.

Another approach has been taken by (Hopkinson et al, 2004) were after filtering for DTM generation, the laser scanner first and last pulse measurements were detrended with the terrain model (i.e. the ground elevation was subtracted). Then the standard deviation of the heights was computed for different fields and compared to measurements of average vegetation height per field. For 14 plots of low vegetation (in this case between 0.2m and 1.3m) a linear regression of the type

Vegetation height = 2.7 * standard deviation of detrended laser heights

was found. Assuming that this relationship holds for all types of low vegetation, this method can – in principle – be used to correct laser measurements. The r.m.s.e. of the regression height is ±15cm.

In (Pfeifer et al 2004) texture measures have been exploited in a similar way. Vegetation height was lower (also below 0.2m) and texture measures have been used in place of the standard deviation. The regression results show a relation between the texture measure contrast and the average shift of the laser points above the ground, it is, however, less strong than in the findings of Hopkinson.

2. DATA

A set of terrestrically surveyed control areas and a laser scanner dataset were given. Between the “wad” (salt marsh) and “dike” area of Noordpolder 426, 8 control fields are situated, numbered consecutively as vegetation field 1 to 8. Field 1 is closest to the sea and field 8 closest to the dike. The points have been measured with RTK-GPS, approximately in a regular 1m grid. The number of points is between 115 and 149 per “vegetation field”, and between 130 and 161 in the “wad” and “dijk” areas. The height accuracy of these points is ±2cm; the measurements were performed with RTK-GPS on 15. and 18.October 2004. The areas covered by these control measurements are therefore between 100m^2 and 160m^2.

Fig. 1 shows vegetation fields 2 (left, closer to the sea) and 6 (closer to the dike). The laser scanner data set was measured with an Optech ALTM system on 11.October 2004. It is given in the area of the 8 vegetation fields, separated into four overlapping strips. The overall point density is higher than 1 point/m^2, but not homogeneous within each strip. At the outer ends of the strips the points are organized in lines perpendicular to the flying direction, which is the normal pattern of oscillating mirror scanners.

Figure 1: photographs of the control fields 2 and 6

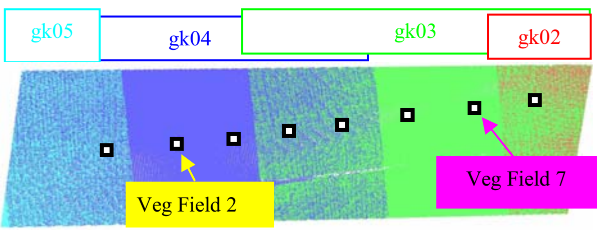

Figure 2: the point cloud laser scanner data and the location of the control fields

Fig. 2 gives an overview of the given laser data. The strips gk02, gk03, gk04, and gk05 are shown in different colors to make the areas of overlap visible. The location of the vegetation fields is also indicated.

3. ANALYSIS

The point cloud data is given in 4 laser strips, numbered 2,3,4 and 5. It covers 8 fields with terrestrial measurements (control points), named Field1, Field2, etc. in the following. Some of the fields are covered by only one laser strips, others by two strips (Fig. 2). There are 12 strip-field combination, listed in Table 1.

For a control point field that is covered by two strips, points from both strips could be combined to increase the point density. We did not use this possibility, since it might introduce artefacts in texture.

Generally, each of the supplied strips covers an area that is much larger than the control point fields. For each strip-field combination we extracted subsets from the strips around the control points using a nearest-neighbor criterion: the 100 laser points that are nearest to each control point are contained in the subset. For each strip-field combination, Table 1 shows:

- The name, containing strip and field number
- Control points: number, height mean and standard deviation; for the fields that are contained in two strips, duplications occur (*Field* 1, 4, 5 and 8).
- Laser points in an area around control points (100 nearest neighbors, subset A): number, mean height and standard deviation.
- Laser points that are inside a TIN of control points (subset B): number of points, the average height of the (x,y) positions of these points according to the TIN and standard deviation, the average height of these points, with standard deviation, and the average difference between these two heights, with standard deviation.

Table 1 shows that most of the fields are really flat, with height variations of a few centimeters only. The sigma in the laser points within these fields is in the order of the measurement noise. Only fields 5 and 7 show a slightly higher variation; strangely, in field 7 the sigma of the control points is larger than that of the laser points.

The average shifts per field, which may be caused by a systematic vegetation effect, are not very large. In 4 from 12 fields the average shift has a value between 10 and 15 cm, in 3 fields the average shift is between 5 and 10 cm, in 4 fields it is between 0 and 5 cm, and in one field it is negative (-2.2 cm).

3.1 Systematic shifts and noise

We made a TIN of strip 3 and projected the points of strip 4 into this TIN. For the strip 4 points located within inside the strip 3 TIN we obtain a projected (strip 3) height, in addition to the original strip 4 height, and we can regard the difference between these two heights. This leads to following statistics:

```
 npts   avg.diff  sigma
49146   -0.0163  0.1011
```

The average difference of -16.3mm indicates a systematic shift between the strips, which is not caused by measurement noise or vegetation since these would average out over the large number of points. The existence of such a systematic effect should be considered when judging the average shifts per field of the previous section.

The variability in the difference, on the other hand, *is* caused by measurement noise and/or texture, where the latter can be subdivided into terrain texture (undulating terrain) and vegetation texture (non-uniform vegetation cover). Between strip 3 and 4 we obtain a shift of –1.6cm with a standard deviation of 10cm (Fig. 3).

Terrain texture (undulating terrain) seems to be a minor effect, since the standard deviation in the control points is smaller than the one in the laser points (Table 1). Since the two measurements are independent, the variance of the difference is the sum of the variances of the strips. In each strip we obtain a standard deviation of 7cm: the combined effect of noise and texture.

We conclude here that the spatial frequency of the texture (caused by vegetation) is too high to be captured by the laser point density. The plants are much smaller than the point spacing. Therefore, the influence of plants on measured heights is stochastic. Vegetation causes noise rather than spatial patterns.

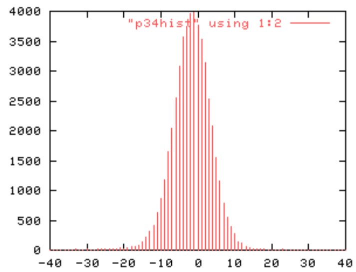

Figure 3: Histogram of difference between strips 3 and 4: points of strip 4 projected into TIN of strip 3.

3.2 Texture measures

The following measures were studied.

Slope texture. The Optech ALTM scanner measures laser points in scan lines, such that the spacing of points inside these lines is much smaller than between the lines. At the side of a strip, lines appear in groups of two, due to the zig-zag scanning pattern that is caused by the oscillating movement of the scanning mirror. We made an attempt to handle this anisotropy by taking the horizontal distance between points into account, together with height differences, in a *slope texture* measure, which computes the absolute slope for each pair of points (xi,yi,zi) and (xj,yj,zj) in a neighborhood as

$$\text{slope} = \text{abs}(zi-zj) \;/\; \text{sqrt}((xi-xj)^2 + (yi-yj)^2) \;,$$

and then computes slope texture as the average of all pairwise slopes (there are, for example, 435 pairs in a 30-point neighborhood).

Neighborhood variance and **neighborhood standard deviation.** Under the assumption that the influence of vegetation on laser point heights is noisy (see above) the variance (or the standard deviation) within a neighborhood of laser points around a certain position may be appropriate to quantify the noise (rather than texture) at that position (see also Hopkinson et al 2004 mentioned in Sec. 2).

3.3 Laser point based vs. control point based method

The previous study distinguished between two methods to find correlation between texture and shift, depending on where these quantities are computed: in each laser point or in each control point.

The laser point based method computes for each laser point of subset B (*i.e.* located within a triangle of the control point TIN):

- texture within neighborhood of *k* nearest laser points (including the current one); neighbors are selected for subset A, which also contains laser points *around* the control points
- shift as difference between the point height and the projection (interpolation) into the TIN of control points

Now, we can proceed in different ways:

- compute correlation between texture and shift over all laser points in one field
- compute average texture and average shift per field, and correlate these averages over the entire data set
- compute correlation between texture and shift over all laser points in the entire data set.

The **control point based** method computes at each control point:

- texture within a neighborhood of *k* nearest laser points
- average height *h* of *m* nearest laser points
- shift as difference between *h* and *z* (the control point height)

Subsequently, the same possibilities exist as above:

- compute correlation between texture and shift over all control points in one field
- compute average texture and average shift per field, and correlate these averages over the entire data set
- compute correlation between texture and shift over all control points in the entire data set.

(In the current study we use the same value, 30, for *k* and *m*).

In the previous study, the laser points were much denser than the control points, whereas the terrain (under the vegetation) seemed to be quite rough. Therefore, it was difficult to estimate the true terrain height (and therefore the shift) of each laser point. In fact, this was the motivation to develop the control point based method: the “laser height” of each control point could be estimated quite well, given the large number of laser points in the neighbourhood.

The current situation is quite the opposite. “True” terrain heights of laser points can be estimated quite well from a TIN of control points, since the terrain is flat and the control points are dense. Both methods, control point and laser point based, yield therefore similar results. Only those for the control point based method will be presented.

3.4 Measurements with the control point based method

In the remainder of this section the results of regression analyses with the control point based method are presented. As a measure of the quality of the regression the value R^2 is given.

The following figures show the texture measures standard deviation and slope texture against the shift of the laser points above the terrestrial points TIN. Fig. 4 shows the measures based at each control point, and Fig. 5 shows the average values per field. Fig. 7 shows the slope texture against the shift for the points of each field separately.

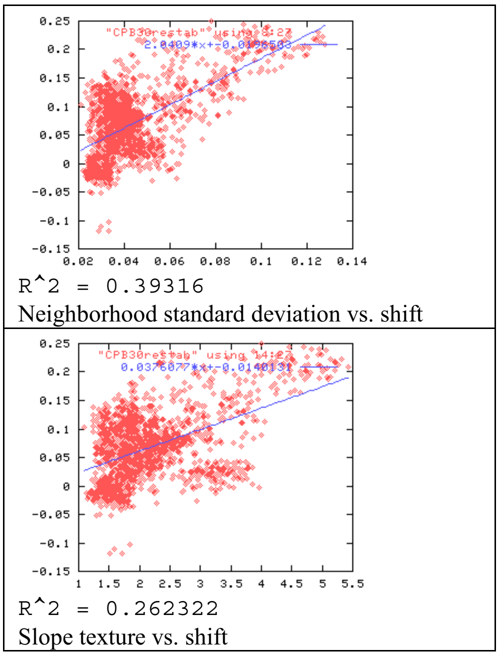

Figure 4: Correlation between different texture measures and shift, according to the control point based method applied to all control points of all fields.

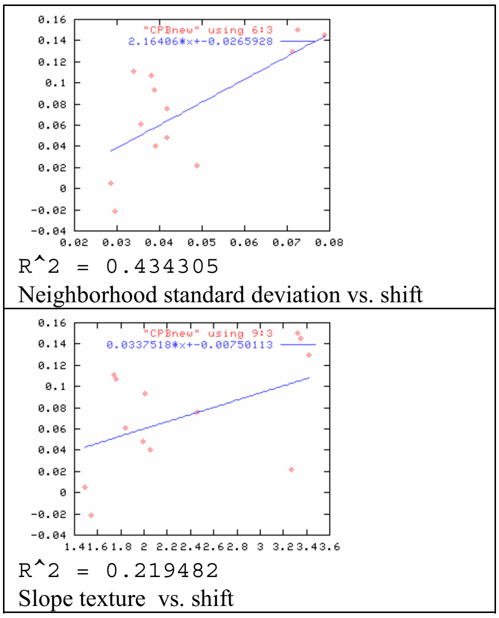

Figure 5: Correlation between different texture measures and shift, according to the control point based method applied to all control points of each field, after averaging per field.

3.5 Analysis of entire areas

The methods used above are point-based. For each point a neighborhood is used, but still, no knowledge of the homogeneity of the vegetation within one area is assumed. As it has also been shown in the previous study (Pfeifer et al 2004) it is also possible to analyze the data plot-wise. Fig. 6 shows the mean shift per field, i.e. the average height of the laser points above the terrestrial ground points TIN and also the standard deviations of the heights in each vegetation field (see also Table 1, area B). Using a TIN of the terrestrial points as reference surface, this method is independent of the overall terrain shape, e.g. its slope.

As it can be seen, for all fields but one, the standard deviation is between ±3cm and ±5cm. The corresponding shift values range from -2cm to +11cm, without a notable relation between shift and standard deviation. Only for field 5 the standard deviation and the shift are remarkably higher (shift=+14cm, std.dev. =±9cm). During the outdoor measurement of the control points no significantly higher vegetation or other special circumstances were noted in field 5. Currently we have no explanation for this phenomenon. The same analysis was also done for the extended laser areas (area A in Table 1), showing the same result.

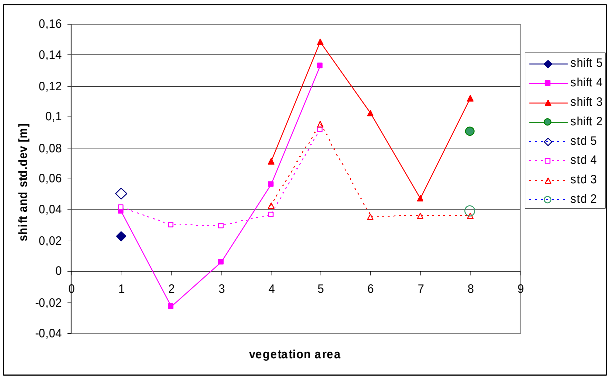

Figure 6: Average shift of laser points above the ground TIN and standard deviation of laser points per vegetation field. Shift values are drawn as solid lines and filled points, standard deviations as dashed lines and point symbols without color filling.

Fig. 6 also shows that the different offset values (e.g. high for field 5) are not caused by errors in the navigation data of the aircraft. Firstly, it has to be observed, that the offsets for each field are similar in all strips, whereas an error in the navigation data of one strip should stand out in one strip only, too. Secondly, an error in the navigation should affect the entire strip, something that cannot be concluded from Fig. 6 either. The offsets of the fields from one strip show no homogeneous value.

4. CONCLUSIONS

One important conclusion can be drawn from Fig. 4. Looking at the upper regression analysis between standard deviation of the laser point heights versus shift, we see that in the range of standard deviations between 2cm and 6cm and shift values between -5cm and +10cm no functional connection can be established. All possible combinations appear. It has to be mentioned, that the standard deviation was measured from the 30 nearest neighbors, which corresponds to an area of roughly 5m^2. Reducing this area would result in higher variations in the standard deviation.

Using the values averaged per field (Fig. 5), the correlation coefficients become larger by a factor of two. However, looking at the standard deviations in the lower range, i.e. up to ±5cm, the same pattern appears as in the single point based method: no correlation between shift and standard deviation. The same result can also be obtained from an area-wise analysis (Fig. 7). It is interesting to note, that the functional relationship between laser point height standard deviation and shift, shift = 2.2*std.dev.-0.03) is similar to the one obtained by (Hopkinson et al 2004), which is, vegetation height = 2.7*std.dev.

From the upper image of Fig. 4 another conclusion can be drawn. If the standard deviation of the measurements is between 3cm and 5cm, a shift of 5cm gives on average the best improvement. For higher vegetation a linear relation between std. dev. or another texture measure and the shift could be assumed.

Combination of different effects

The standard deviation, or a measure of texture, can be split up into 3 contributing components.

- the std.dev. of the measurement itself (2cm-4cm)
- the std.dev. caused by the terrain (a few cm)
- the std.dev. caused by the vegetation (vegetation dependent)

The standard deviation of the terrain always influences the measures, because a standard deviation (a texture) cannot be computed from one point alone, but always an area of a certain size has to be used.

For low vegetation also its standard deviation is very low. It is therefore difficult, if not impossible, to differentiate between terrain and vegetation influences. This is another argument for the impossibility to find a linear relation between standard deviation and a shift for very low vegetation.

While relations between vegetation height and standard deviation of the laser points (detrended for sloped areas) was successful for larger vegetation (30cm or higher), it is not possible with current laser scanning technology to quantify and remove the influence of very low vegetation. As this influence is depending strongly on the development state of the vegetation, a (static) land classification cannot be used to assign offset values to vegetation types either. This limits the possibility of using RGB and nIR imagery for tackeling the problem. It would, however, be interesting to compare our laser point/control point studies with studies on how much light (at the laser wavelength) reaches the ground, expressed e.g. as number of green pixels vs. number of brown pixels in a very high resolution image of the vegetation, or a more integral measure (e.g., NDVI).

REFERENCES

Ahokas, E., H. Kaartinen, and J. Hyyppä (2003). A quality assessment of airborne laser scanner data. In Internat. Archives of Photogrammetry and Remote Sensing, Vol. XXXIV, 3/W13.

Bollweg, A. and R. de Lange (2003). Wat ruist er door het struikgewas... Welke maaiveldhoogte wordt met laseraltimetrie gemeten bij aanwezigheid van vegetation? Technical Report AGI-GAR-2003-22 (in Dutch), AGI, RWS, The Netherlands.

van Heerd, R.M., E.A.C. Kuijlaars, M.P. Teeuw, R.J. van 't Zand (2000). Productspecificatie AHN 2000. Technical Report MDTGM 2000.13 (in Dutch), MD, RWS, The Netherlands.

Hodgson, M. and P. Bresnahan (2004). Accuracy of Airborne Lidar-Derived Elevation: Empirical Assessment and Error Budget. Photogrammetric Engineering & Remote Sensing (3).

Hopkinson, C. et al (2004). Wet-land grass to plantation forest – estimating vegetation height from the standard deviation of lidar frequency distributions. In International Archives of Photogrammetry and Remote Sens-ing, Vol. XXXVI, 8/W2.

Pfeifer, N., B. Gorte, and S. Oude Elberink (2004). Influences of vegetation on laser altimetry – analysis and correction approaches. In International Archives of Photogrammetry and Remote Sensing, Vol. XXXVI, 8/W2.

Pfeifer, N., P. Stadler, and C. Briese (2001). Derivation of digital terrain models in the SCOP++ environment. In Proceedings of OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Terrain Models.

Oude Elberink, S. and M. Crombaghs (2004). Laseraltimetrie voor de hoogtemetingen van de kwelders Waddenzee. Technical Report AGI-GAP-2003-50 (in Dutch), AGI, RWS, The Netherlands.

Table 1. Statistics of 12 strip-field combinations that occur in the supplied data set

field	control points			laser points (A)			laser points projected in TIN of control points (B)						
name	#	avgz	sigma	#	avgz	sigma	#	avg.int.z	sigma	avgz	sigma	avg.dif	sigma
Strip2Field8	141	1.2872	0.0195	366	1.3709	0.0411	70	1.2917	0.0160	1.3837	0.0391	0.0920	0.0451
Strip3Field4	129	1.0242	0.0106	423	1.0984	0.0458	101	1.0251	0.0105	1.0982	0.0429	0.0731	0.0430
Strip3Field5	139	1.1213	0.0374	392	1.2362	0.0802	76	1.1261	0.0362	1.2757	0.0954	0.1496	0.0886
Strip3Field6	136	1.2273	0.0276	410	1.3390	0.0459	95	1.2272	0.0208	1.3302	0.0353	0.1031	0.0455
Strip3Field7	149	1.3475	0.0644	437	1.3928	0.0495	107	1.3576	0.0399	1.4049	0.0362	0.0473	0.0339
Strip3Field8	141	1.2872	0.0195	435	1.3960	0.0358	98	1.2895	0.0165	1.4013	0.0362	0.1118	0.0443
Strip4Field1	115	0.8580	0.0120	439	0.8981	0.0374	112	0.8579	0.0106	0.8969	0.0416	0.0390	0.0404
Strip4Field2	131	0.8815	0.0190	442	0.8661	0.0321	113	0.8823	0.0201	0.8600	0.0302	-0.0223	0.0323
Strip4Field3	124	0.8876	0.0102	445	0.9015	0.0327	117	0.8862	0.0090	0.8926	0.0295	0.0065	0.0301
Strip4Field4	129	1.0242	0.0106	445	1.0913	0.0367	102	1.0261	0.0092	1.0834	0.0369	0.0574	0.0373
Strip4Field5	139	1.1213	0.0374	409	1.2188	0.0809	88	1.1283	0.0356	1.2615	0.0918	0.1332	0.0917
Strip5Field1	115	0.8580	0.0120	394	0.8768	0.0442	112	0.8573	0.0113	0.8811	0.0503	0.0238	0.0516

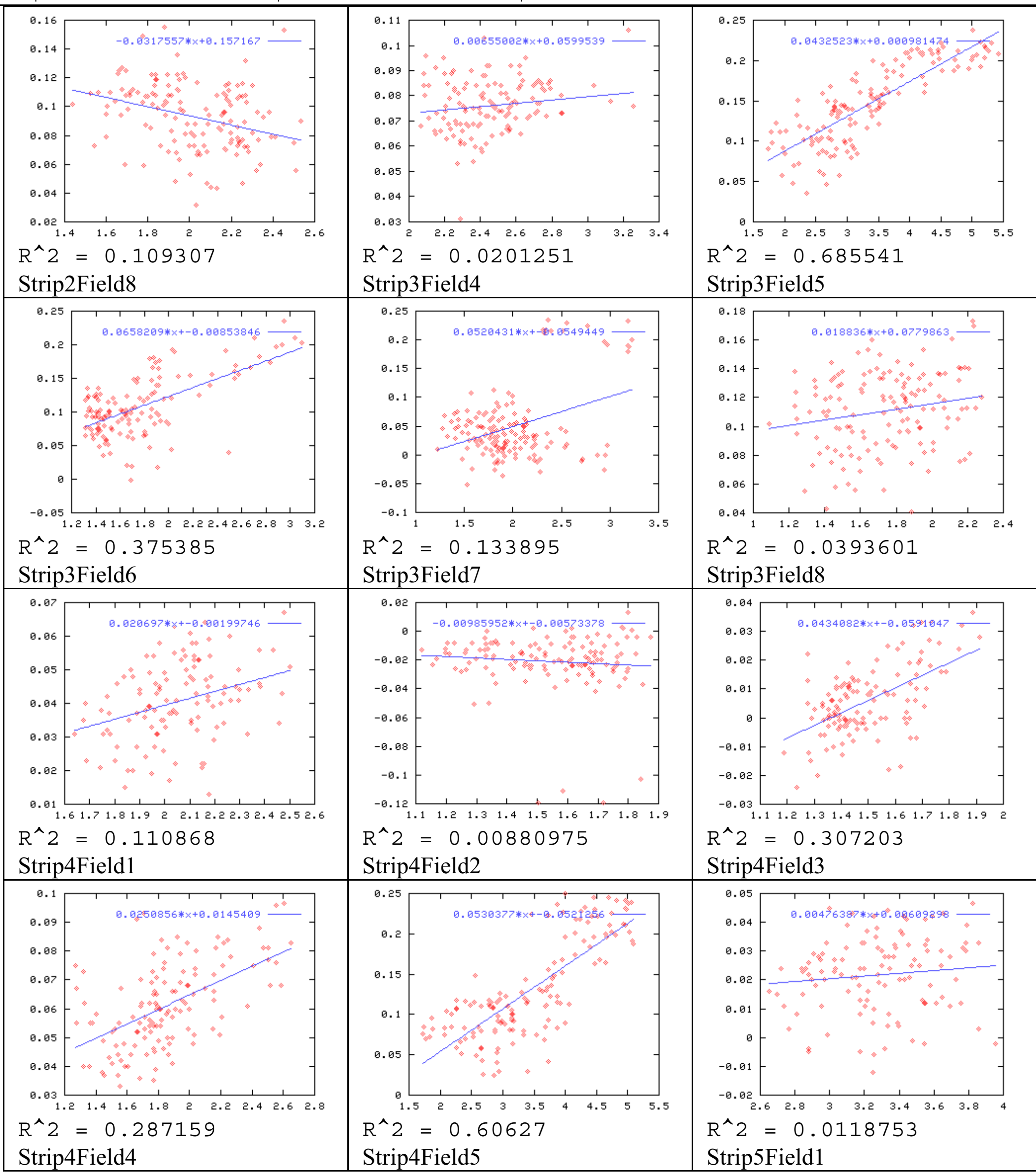

Figure 7: Correlation between slope texture and shift, according to the control point based method applied to all control points of each field.