SEMANTICALLY CORRECT INTEGRATION OF A DIGITAL TERRAIN MODEL AND A 2D TOPOGRAPHIC VECTOR DATA SET

A. Koch

Institute of Photogrammetry and GeoInformation (IPI), University of Hannover
Nienburger Strasse 1, 30167 Hannover
koch@ipi.uni-hannover.de

KEY WORDS: GIS, Adjustment, Integration, Visualization, Modelling, DTM

ABSTRACT:

The most commonly used topographic vector data, the core data of a geographic information system (GIS), are currently two-dimensional. The topographic objects are modelled as points, lines or areas with additional attributes containing different information about function and dimension of the object, possibly including height values. Height values may also be added to the objects by using a digital terrain model (DTM), i.e. by integrating vector data and DTM. In general, however, discrepancies exist between the two different data sets. These discrepancies are caused by different object modelling and different surveying and production methods. In the vector data set, for example, a road is often modelled as a line or a polyline. The attributes contain information about the road width, the road type, etc. If the road is located on a dam, the corresponding part of the DTM can be modelled as an elongated, horizontal plane of some width with two slopes at the sides. Differences between the attribute road width of the vector data set, and the width of the horizontal plane in the DTM may cause discrepancies when integrating these data sets. Additionally, the data are often produced independently. The DTM may be generated using laser scanning or aerial photogrammetry. The topographic vector data may be based on digitized topographic maps or orthophotos resulting in differences in 2D location.
As already mentioned, the topographic objects of the vector data are two-dimensional. But there exist various objects which contain implicit height information. A lake, for example, represents a horizontal plane, and all height values of the lakes's area must be the same, the neighbouring banks, however, have to be higher than the water. To give another example, the slope of a road along and across the road direction must not exceed a certain maximum value.
This paper describes an algorithm which uses this semantic information for a semantically correct integration of DTM and 2D topographic vector data. It is based on a inequality constrained least squares adjustment formulated as the linear complementary problem (LCP). The algorithm results in improved height values of the DTM. First investigations were carried out by using two different data sets representing two classes of objects. The first simulated data set represents a tilted plane, the second data set is a real data set of the German ATKIS®. First results using these data sets are satisfying.

1. INTRODUCTION

The most commonly used topographic vector data, the core data of a geographic information system (GIS), are currently two-dimensional. The topographic objects are modelled as points, lines or areas with additional attributes which may contain different information on function and dimension of the object. In contrast a digital terrain model (DTM) in most cases is a 2.5D representation of the earth's surface. Each position in planimetry contains one height value. The points are distributed irregularly and/or in a grid, only points representing structure elements contain semantic meaning. They represent, for example, break lines or single significant height values.
Integration of these two data sets may be understood in terms of an integrated data modelling for consideration of height information in a GIS. The 2D topographic objects are completed by height information of the DTM, and after the integration both data sets represent an integrated 2.5D topographic data set.
There are different reasons for integration of the two data sets. First, data acquisition is mostly the main cost factor for generation of geographic information systems (Bill & Fritsch, 1994). Integration of DTM or other specific data sets with respect to the geographical data base may reduce costs because there is no need for new data acquisition in 2.5 or 3D.
Secondly, integration will result in consistent data sets. Inconsistencies are caused by different object modelling and different surveying and production methods. Vector data sets often contain roads modelled as lines or polylines. The attributes contain information on road width, road type etc. If the road is located on a dam, the corresponding part of the DTM can be modelled as an elongated, horizontal plane of some width with two slopes at its sides. Differences between the attribute road width of the vector data set, and the width of the horizontal plane in the DTM may cause discrepancies when integrating these data sets. Additionally, data are often produced independently. The DTM may be generated by using laser scanning or aerial photogrammetry. Topographic vector data may be based on digitized topographic maps or orthophotos.
A third reason for the integration is verification of the DTM. In many cases topographic vector data are almost up to date because objects as streets and railways posses major priority in GIS. A DTM, however, may be older than one decade. It is true that height changes appear less frequently than changes in the horizontal position of objects. Nevertheless, integration of both data sets considering the semantic meaning of objects will show discrepancies, and will allow to draw conclusions on the quality of the DTM.
This paper presents an algorithm for a semantically correct integration of a DTM and a 2D topographic vector data set. The following section describes the meaning of semantic correctness using some examples. Section 3 represents the algorithm and section 4 shows first results using simulated and real data sets. The paper concludes with a short summary and some future work.

2. SEMANTIC CORRECTNESS

There are several approaches for integration of a DTM and a topographic vector data set. Some methods are based on the use of height values as attribute, i.e. the Z-coordinate is linked to the object. In case of area based objects polynomial faces may be added. Other approaches use triangles or grids for the combination of topographic objects and DTM and for completion of the structure of a triangulated or gridded DTM, whereas the latter method locally densifies the gridded structure of the DTM by also using triangulations. An overview about existing methods is given by (Lenk, 2001).

All these methods do not consider any semantic meaning of the topographic objects. For example, if a street crosses a ridge, the integration of a street into an existing triangulated DTM may cause an incorrect terrain morphology, if no additional point is added on the ridge, which will result in a terrain insection. Or, if a lake or a subset of a lake is situated on a slope, the water of a simulated standing water body flows. Figure 1 shows a DTM and a topographic data set with three lakes. The heights at the border of the lakes are not the same as inside the lakes, there is no constant height level, the semantics are incorrect.

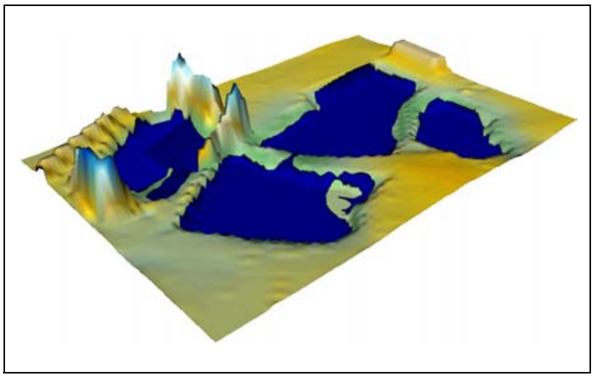

Figure 1: A visualization of a DTM (ATKIS® DGM5) and a topographic vector data set with three lakes (ATKIS® Basis-DLM)

The topographic objects are two-dimensonal. Nevertheless, objects exist which contain implicit relative height information on the terrain. Table 1 shows some of these objects and their representation in the terrain.

The table contains three object classes. The first class contains objects which represent horizontal planes. For example, sports fields (football or soccer fields) are mainly horizontal. Furthermore, runways or lakes represent almost horizontal planes, in addition, lakes have banks at their sides, and the terrain is ascending.

The second class describes objects representing tilted planes. Street or roads represent tilted planes with some slope in driving direction and across. Perpendicular to the road direction the inclination of the street is necessary because of water drainage. Additionally, in curves the slope will reduce the forces which affect the vehicles. In topographic vector data sets streets and railways are mostly modelled by lines. Therefore, these objects have to be buffered using additional information on the width of the object when integrating them with a DTM.

Representation	Object
Horizontal plane	Sports field Race track Runway Dock Canal Lake, pool
Tilted plane with maximum slopes	Street, Road Path Railway, tramway River
Height relation	Bridge Undercrossing, crossover

Table 1. Topographic objects and their representation in the terrain from the ATKIS® Basis-DLM feature catalogue

The last class shown in table 1 describes objects which have a height relation to other objects. Bridges, undercrossings and crossovers may contain a certain relation to their neighbouring terrain. These relations mainly are components of the 2D vector data set.

3. AN ALGORITHM FOR THE SEMANTICALLY CORRECT INTEGRATION

A DTM is composed of regularly or irregularly distributed points with its coordinates X,Y,Z and an interpolation function to derive Z values at arbitrary positions X,Y. Additionally structure elements are included which contain information on the morphology of the terrain (break lines, significant points, etc.).

The topographic vector data currently are two-dimensional. The topographic objects are modelled by points, lines or areas. In this paper the planimetric coordinates of the topographic objects will be introduced as error-free. Any kind of systematic or random errors will be neglected, i.e. errors of the interpolated height value caused by inaccurate planimetric coordinates of the vector data are not considered up to now.

The aim of integrating both data sets is an integrated data modelling for considering height information in GIS. The data sets must be integrated in such a way that the topographic objects have to fulfill certain constraints which arise from the semantics of the objects.

The first step of the algorithm is a constrained Delaunay triangulation of all points of the DTM (section 3.1.). Then, different objects containing implicit height information are introduced, and their heights are interpolated using the height values of the triangulated DTM. Equality and inequality constraints are then considered in an optimization process, resulting in improved height values and in integrated data sets which fulfill the predefined constraints. A precondition of the algorithm is that the terrain morphology in the neighbourhood of the objects has to be considered but the improvements of the heights have to be small.

The objects are finally introduced as constraints in a new Delaunay triangulation, producing two consistent data sets.

3.1. Constrained Delaunay triangulation

The hybrid DTM will be triangulated using a constrained Delaunay triangulation. The structure elements are introduced as constraints, and the result is a triangulated irregular network (TIN). There are many Delaunay triangulation algorithms. In our approach the "divide and conquer" algorithm presented by Guibas and Stolfi, 1985, was used because it performs very well (Shewchuk, 1997). The structure elements were introduced in such a way that the points of the structure elements intersecting the edges of the triangles were added as new points of the triangulation.

3.2. Equality and inequality constraints

For the topographic objects which contain implicit height information (see table 1), equality and inequality constraints are formulated. Each class contained in table 1 has its own constraints, which will be derived in the following.

Horizontal plane
Height values of objects representing a horizontal plane must be identical, i.e. in case of area based objects the points with planimetric coordinates inside the object polygon must have the same Z-coordinate Z_{HP}. In ideal cases the following equation must be fulfilled:

$$Z_i = Z_{HP} \tag{1}$$

where $i = 1,\ldots,n_{ins}$
n_{ins} Number of points with planimetric coordinates lying inside the polygon representing the object
Z_{HP} Height value of the horizontal plane

The height values of the object boundary polygon points must in general be interpolated from the neighbouring DTM heights. Again, these heights must have the same height level Z_{HP}:

$$Z_j(X_j, Y_j, Z_u, Z_v, Z_w) = Z_{HP} \tag{2}$$

X_j, Y_j are the planimetric coordinates of an object boundary point, Z_u, Z_v, Z_w are the height values of the triangle of the TIN which contains the planimetric coordinates of the object point. The height values Z_j are computed using the equation of a tilted plane.

The neighbouring terrain of the object outside the polygon also has to be considered in the algorithm. The first reason is that some of the objects have banks or slopes at their sides which have a certain value with respect to the horizontal plane. For example, the height values of a bank of a lake shore have to increase as otherwise the water would flow out. Secondly, the object is related to the neighbouring terrain: The further a DTM point lies away from an object, the smaller should be the influence of the object on the height change of these points.
Figure 2 shows some points of a lake and its neighbouring terrain. The different colours of the points indicate the equality or inequality constraint which must be fulfilled after applying the algorithm. The gray and the dark blue points must have the same height values as they represent the water body. The Z-coordinates of the orange points have to be higher than the height level of the lake. The orange points are all points outside the object polygon which belong to triangles of the TIN whose edges intersect the boundary polygon of the lake (for example point 214 intersects the boundary polygon, see the fat dark blue lines). The corresponding inequation is:

$$Z_k > Z_{HP} \tag{3}$$

where $k = 1,\ldots,n_{obj}$
n_{obj} Number of points of the topographic object
Z_{HP} Height value of the horizontal plane

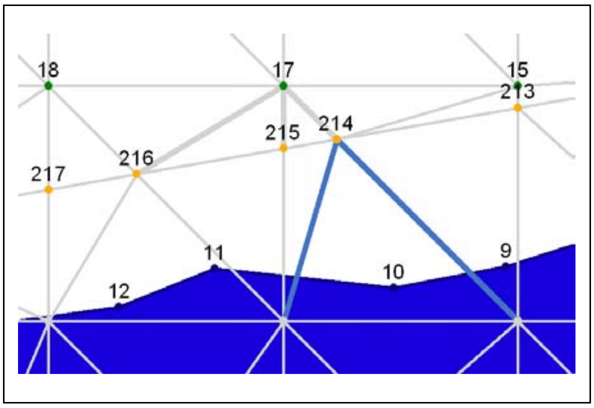

Figure 2: DTM points and boundary polygon of a topographic object and the neighbouring terrain

The green DTM points are all points outside the object polygon which are connected to orange DTM points of the bank. In figure 2, the DTM point 17 is connected to three orange points (214, 215, 216). The connections will be added by using the height differences between the points:

$$\Delta Z_{kl} = Z_k - Z_l \tag{4}$$

Z_k ist the height value of a DTM point of the bank, Z_l is the corresponding height value of the green DTM points. These height differences lead to a terrain morphology which is nearly the same as before.

Tilted plane
The objects representing a tilted plane are mainly elongated objects (see table 1). Along the main direction these objects are not allowed to exceed a predefined slope value s_{Max}:

$$\left|\frac{Z_m - Z_n}{D_{mn}}\right| \le s_{Max} \tag{5}$$

Additionally the slope difference ds_{Max} between neighbouring object sections, comparable to the curvature of the object in longitudinal direction, is restricted:

$$\left|\frac{Z_m - Z_n}{D_{mn}} - \frac{Z_n - Z_o}{D_{no}}\right| \le ds_{Max} \tag{6}$$

D_{mn} and D_{no} are the horizontal distances between the object boundary points P_m, P_n and P_n, P_o, respectively. In case of objects which are modelled by lines, these points belong to the object polyline.
The corresponding height values in perpendicular direction must have the same height level, because the slope of streets or railways will be neglected (equation 7). Objects modelled by lines have to be buffered using attributes representing the width of the object. Then, Z_p represents the height value of the buffered left or right side and Z_m is the height of the centre axis.

$$Z_p = Z_m \tag{7}$$

Finally, the height values of points with their planimetric coordinates inside the object have to be situated on the tilted plane. The distance between the point P_q and the plane must be zero:

$$\frac{\left|\underline{n} \cdot \underline{q} - d\right|}{\left|\underline{n}\right|} = 0 \tag{8}$$

d is the distance of the plane to the origin of the coordinate system, $\underline{n}$ is the normal vector of the tilted plane and $\underline{q}$ contains the coordinates of the point P_q.
Again the connection to the DTM points outside the object polygon will be introduced using the height differences between these points and points representing the object (see equation 4).

Height relation
Bridges, undercrossings and crossovers have a certain height relation to other objects (for example street, railway, river, etc.). The height values which belong to these objects must be higher or lower than the related objects:

$$\begin{aligned} Z_r &< Z_s \\ Z_r &> Z_s \end{aligned} \tag{9}$$

3.3. Inequality constrained least squares adjustment

The defined constraints (equations 1 to 9) have to be introduced in an optimization process. The algorithm used is based on an inequality constrained least squares adjustment which is formulated as the linear complementary problem (LCP). In the following the basic principle of the algorithm will be described, for details see Lawson & Hanson, 1995, Fritsch, 1985 and Schaffrin, 1981.

The aim of a least squares adjustment is to minimize the square sum of the residuals $\underline{v}$ (L_2 approximation):

$$\begin{aligned} &\underline{v}^T \underline{P} \underline{v} \rightarrow Min \Leftrightarrow \\ &\left(\underline{A}\underline{x} - \underline{l}\right)^T \underline{P} \left(\underline{A}\underline{x} - \underline{l}\right) \rightarrow Min \end{aligned} \tag{10.1}$$

where
- $\underline{v}$ vector of residuals
- $\underline{P}$ weight matrix
- $\underline{A}$ Jacobean matrix of the partial derivations of the observation equations with respect to the unknown parameters $\underline{x}$
- $\underline{x}$ vector of unknown parameters
- $\underline{l}$ vector of reduced observations

The equality constraints 1, 2, 4, 7 and 8 can be used to derive observation equations of this adjustment procedure:

$$0 + \hat{v}_i = \hat{Z}_{HP} - Z_i \tag{1.1}$$

$$0 + \hat{v}_j = \hat{Z}_{HP} - Z_j\left(X_j, Y_j, Z_u, Z_v, Z_w\right) \tag{2.1}$$

$$\Delta Z_{kl} + \hat{v}_{kl} = \hat{Z}_k - \hat{Z}_l \tag{4.1}$$

$$0 + \hat{v}_{pm} = \hat{Z}_p - \hat{Z}_m \tag{7.1}$$

$$0 + \hat{v}_{distance} = \frac{\left|\underline{n} \cdot \underline{q} - d\right|}{\left|\underline{n}\right|} \tag{8.1}$$

Here, a Gauss-Markoff adjustment model was used because the defined equality constraints are not strictly adhered to. The height values $\hat{Z}$ are the unknowns, the observations are pseudo-observations. By weighting them using the weight matrix $\underline{P}$, the fulfilment of the constraints can be controlled.

The heights of the green DTM points outside the object (see figure 2) are used to form an additional observation equation:

$$0 + \hat{v}_l = \hat{Z}_l - Z_l \tag{11}$$

If a height value has nearly to be unchanged, this observation gets a high weight.

The adjustment procedure is completed by introducing the inequations 3, 5, 6 and 9 in the optimization process:

$$0 > \hat{Z}_{HP} - \hat{Z}_k \tag{3.1}$$

$$s_{Max} \geq \left|\frac{\hat{Z}_m - \hat{Z}_n}{D_{mn}}\right| \tag{5.1}$$

$$ds_{Max} \geq \left|\frac{\hat{Z}_m - \hat{Z}_n}{D_{mn}} - \frac{\hat{Z}_n - \hat{Z}_o}{D_{no}}\right| \tag{6.1}$$

$$\hat{Z}_r < \hat{Z}_s, \quad \hat{Z}_r > \hat{Z}_s \tag{9.1}$$

These inequations form the following inequation system the unknown parameters also have to satisfy:

$$\underline{B} \cdot \underline{x} \leq \underline{b} \tag{10.2}$$

where
- $\underline{B}$ matrix of partial derivations of the inequations with respect to the unknown parameters $\underline{x}$
- $\underline{b}$ right side of the inequation system

The LCP is formulated as follows:

$$\underline{\hat{z}} = \underline{M} \cdot \underline{\hat{u}} + \underline{k} \tag{12}$$

where

$$\begin{aligned} \underline{M} &:= \underline{B} \cdot \left(\underline{A}^T \underline{P} \underline{A}\right)^{-1} \cdot \underline{B}^T \\ \underline{k} &:= \underline{b} - \underline{B} \cdot \left(\underline{A}^T \underline{P} \underline{A}\right)^{-1} \cdot \left(\underline{A}^T \underline{P} \underline{l}\right) \end{aligned} \tag{13}$$

The solution (12) is called a complementary solution, because the vectors $\underline{\hat{z}}$ and $\underline{\hat{u}}$ are complementary to each other, i.e. the following complementary condition is fulfilled:

$$\begin{aligned} &\underline{\hat{z}} \geq \underline{0}, \quad \underline{\hat{u}} \geq \underline{0} \\ &\underline{\hat{u}}^T \cdot \underline{\hat{z}} = 0 \end{aligned}$$

The unknown parameter vector $\underline{x}$ of the initial least squares adjustment with inequality constraints (see 10) are computed using the result vector $\underline{\hat{u}}$ of equation 12:

$$\underline{\hat{x}} = -\left(\underline{A}^T \underline{P} \underline{A}\right)^{-1} \cdot \left(\underline{B}^T \underline{\hat{u}} - \underline{A}^T \underline{P} \underline{l}\right) \tag{14}$$

In our approach the Linear Complementary Problem is solved using the Lemke algorithm (Lemke, 1968).

4. RESULTS

The first results were determined by means of two different test data sets. The first data set is a simulated one and consists of a small DTM strip and an object representing a tilted plane. The second data set is the area from figure 1. The topographic vector data are objects of the German ATKIS® Basis-DLM. The data set consists of three lakes bordered by polygons with planimetric coordinates. The DTM, the ATKIS® DGM5, is a hyrid data set, containing regularly distributed points with a grid size of 12,5 m and structure elements.

4.1. Tilted plane

The simulated regular DTM consists of 6 rows and 31 columns with a grid size of 10 m. The heights were calculated using a sine function in X-direction (east-west) with an additional random correction. The object is a straight road with random X- and constant Y-coordinates, it consists of 54 points. The line has been buffered by using a road width of 4 m on the left and 4 m on the right side of the centre axis. Figure 3 shows the triangulated DTM.

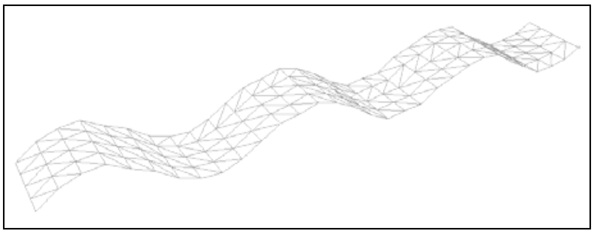

Figure 3: TIN of the simulated data set

In all computations the values s_{Max} and ds_{Max} of the inequality constraints were set to 0.06 and 0.05, respectively. The first investigations were done using the same weight for all observations, i.e. the matrix $\underline{P}$ from equation (10.1) is a unit matrix.

Figure 4 shows the profile of the interpolated height values of the road before the algorithm was applied (gray line). The blue line represents the resulting profile after the integration of both data sets.

The height range of the new profile is smaller than the height range of the original data, the mean height of the centre axis of the road presented in figure 4 is nearly the same as before. The defined constraints are fulfilled, all slope values are smaller than 0.05 and all curvature values are smaller than 0.06.

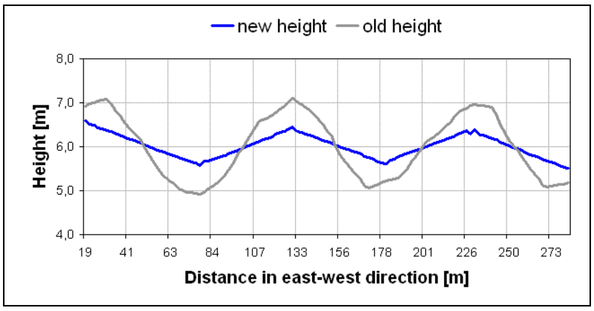

Figure 4: East-west profile before and after the algorithm was applied

Perpendicular to the driving direction of the road the height differences between points of the centre axis and the corresponding heights of the border are not zero, because the inequality constraints (inequation 5.1 and 6.1) refer to the centre axis. Using a higher weight for these kind of observation results in horizontal cross sections.

A higher weight of the green DTM points and of the height differences between the green and the orange points outside the object results in increasing inclined cross sections. The road adjusts to the terrain, the improvements of the point heights outside the object become smaller.

4.2. Horizontal plane

The second example deals with ATKIS® data. The test data set consists of about 3400 DTM points, three lakes are represented by total of 300 planimetric coordinates. Figure 5 shows the DTM after the constrained Delaunay triangulation. The blue points are the interpolated height values of the topographic vector data set. Obviously, the blue points do not all have the same height level. Most of them are situated on the bank of the lakes. Thus, the interpolated heights are higher than the points inside the lakes. Additionally, the height values inside the object polygon also do not have the same height level.

Again, the first investigations were carried out by using the same weight for all observations. The height values of the bank have to be at least 1 cm higher than the height level of the lakes (see inequation 3.1).

The algorithm results in lake heights which are nearly identical to the mean values of all height values representing a lake. The inequality constraints are fulfilled.
A higher weight for the green DTM points outside the lake (equation 11) and for the height differences between these points and the points of the bank (equation 4.1) results in lake heights which differ from the mean values. The terrain outside the objects is nearly the same as before. Only the height values which are to low were improved significantly.

Figure 6 represents the TIN of the integrated data sets. The blue areas are the lakes: All height values inside the lake and on the bounding polygon of the lake are the same.

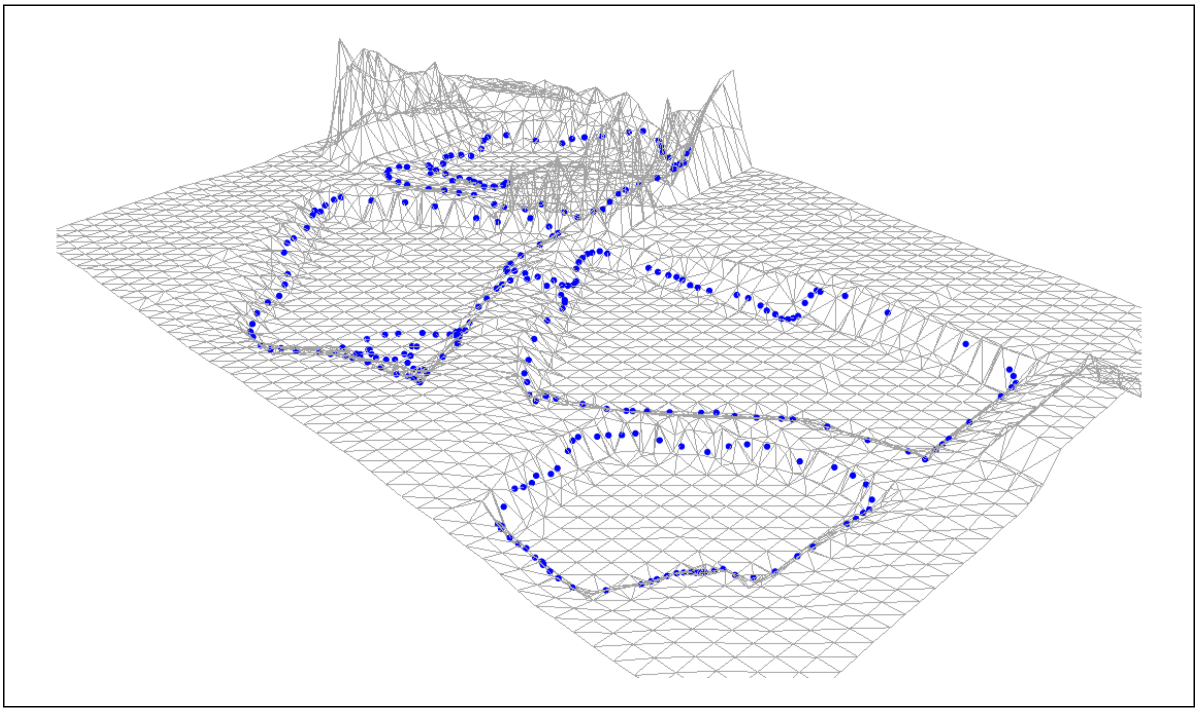

Figure 5: Triangular irregular network (TIN) of the DTM, topographic vector data set with interpolated heights

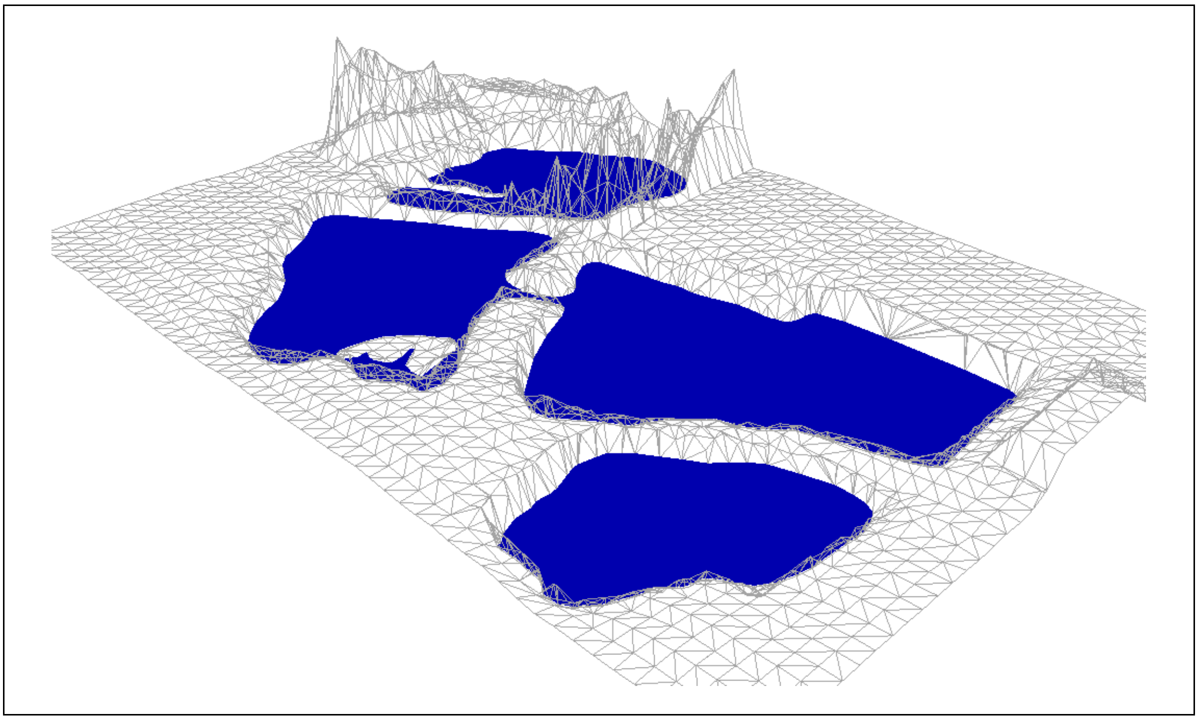

Figure 6: TIN of the integrated data sets, DTM and topographic vector data sets after applying the algorithm

5. SUMMARY AND FUTURE WORK

This paper presents the first results which have been derived using an algorithm for semantically correct integration of DTM and a topographic vector data set. Based on a constrained Delaunay triangulation the heights of the two-dimensional vector data set are calculated using the neighbouring heights of a triangular irregular network. Some of the objects of the topographic vector data contain implicit height information. For example, lakes describe a horizontal plane, roads or railways do not exceed maximum slope and curvature values. This information has been used to derive equality and inequality constraints. The algorithm is based on a inequality constrained least squares adjustment formulated as the linear complementary problem (LCP). The algorithm results in improved height values of the DTM.

First investigations were carried out by using two different data sets: A simulated one representing a tilted plane and a real data set of the German ATKIS®. The constraints have been fulfilled but big differences between the constraints and the DTM may cause a non-realistic improvement of the original height information. Thus, blunders or big differences have to be analyzed in the algorithm in the future.

Furthermore, the planimetric coordinates of the topographic vector data set were introduced as error-free. This may cause a reduced height level of the topographic object if the planimetric coordinates of the object representing a road on a dam are situated beneath the corresponding part of the DTM. Or, if the planimetric coordinates of the bounding polygon of a lake are situated on the banks, the lake height arises. That means, the accuracy of the planimetric coordinates of the topographic data have to be considered.

In addition, planimetric coordinates of structure elements have to be considered in the adjustment process. Otherwise, structure elements inside area based objects will be deleted and the morphology can be erroneous.

Nevertheless, the algorithm shows first satisfying results but further investigations have to be carry out.

REFERENCES

Bill, R., Fritsch, D., 1994. Grundlagen der Geo-Informationssysteme, Bd. 1, Hardware, Software und Daten, Wichmann Verlag, Heidelberg.

Fritsch, D., 1985. Some Additional Information on the Capacity of the Linear Complementary Algorithm, in: E. Grafarend & F. Sanso, Eds., “Optimization and Design of Geodetic Networks”, Springer, Berlin, pp. 169-184.

Guibas, L., Stolfi, J., 1985. Primitives for the Manipulation of General Subdivisions and the Computation of Voronoi Diagrams. ACM Transactions on Graphics 4(2), pp. 74-123.

Lawson, C. L., Hanson, R. J., 1995. Solving Least Squares Problems. Society for Industrial and Applied Mathematics, Philadelphia.

Lemke, C. E., 1968. On complementary pivot theory, in: G. B. Dantzig, A. F. Veinott, Eds., “Mathematics in the Decision Sciences”, Part 1, pp. 95-114.

Lenk, U., 2001. – 2.5D-GIS und Geobasisdaten – Integration von Höheninformation und Digitalen Situationsmodellen, Wiss. Arb. Fachr. Verm. Universität Hannover Nr. 244 und DGK bei der Bayer. Akad. D. Wiss., Reihe C, Nr. 546. Diss., Univ. Hannover.

Schaffrin, B., 1981. Ausgleichung mit Bedingungs-Ungleichungen. AVN, 6, pp. 227-238.

Shewchuk, J. R., 1997. Delaunay Refinement Mesh Generation. PhD Thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh.

ACKNOWLEDGEMENT

This research was supported by the surveying authority of Lower Saxony Landesvermessung und Geobasisinformation Niedersachsen (LGN). We are express our gratitude to LGN for providing the data.