# CONSISTENCY MATCHING IN THE INTEGRATION OF CONTOUR AND RIVER DATA BY SPATIAL KNOWLEDGE

Tinghua Ai \*, Min Yang

School of Resource and Environment Sciences, Wuhan University,Wuhan 430072, China, tinghua\_ai@tom.com

Commission II, WG II/4

KEY WORDS: Spatial data matching, Delaunay triangulation, Spatial data quality

# **ABSTRACT:**

As the representation of terrain surface and height information, the contour data has the strict constraint relationship with the distribution of river network. In spatial data integration and matching, the inconsistency usually occurs between the river network and contour generating "river climbing uphill". This study presents a method to build the matching relationship and to correct the inconsistency between river network and contour data. Based on Delaunay triangulation, the terrain landform features are extracted and by bend analysis build the matching relations between river network and contour. According to different inconsistency situations, we offer two correction approaches depending on which data is precise, which includes the river network displacement referenced to the contour and the opposite.

## 1. INTRODUCTION

After more and more spatial data infrastructures are completed, the integration of heterogeneous data sources becomes an important issue. In such fields as data update, interoperation and dissimilation (via web), the integration of spatial data plays a significant role not only to derive combined abstracted information but also to extract differences between various data. The up-to-date data is usually integrated to the existed database to update the database. Different scale data are often integrated, e.g. the large scale data to integrate and update small scale data by map generalization. In geographic analysis, the integration of spatial data is also an important process to prepare data since a complex analysis usually involves multiple spatial data from different spatial reference systems, levels of accuracy, thematic categoriess and other characteristics.

Database integration results in the matching question requiring some works at the schema level and some works at the data level, especially in the context of geographic data. The inconsistency detection and adjustment is such an important work when integrating and matching different data. During the data integration, the data from different sources may contradict to each other in geometric representation, topological relationship or semantic description. The same object in the real word may be represented in quite different ways, such as geometric dimensions, abstraction levels, semantic hierarchies and other properties. The data combination usually results in conflict, e.g. the shallow polygon represented river meets the collapsed axis represented river. The inconsistencies can occur among both homogeneous feature and heterogeneous features. For example, under the same hydrographic feature, the polygon lake and the line river may be inconsistent; On the other hand, the river feature may inconsistent with terrain data, e.g. contour line.

The inconsistency results from different reasons, e.g. the different representations (the river is represented by a shallow

polygon or a collapsed axis line), the database construction at different time by different agencies, or the cognition from different viewpoints. Correspondingly the data matching has to be performed aiming at different data integration situations, including the matching of (1) different scale data in same region, (2) different semantic description under same environment, (3) data from different domains, (4) associated data of different features, and so on.

As logical consistency is one of five aspects of spatial data quality (Goodchild 1991), to preserve the consistency becomes an important maintenance in database construction. A few methods have been developed to detect and adjust inconsistency when data matching in the field of spatial data handling. Among them the matching of different scale spatial data attracts more interests to build the associations between less detailed data and more detailed data (Devogele 1998, Walter 1999, Mustiere 2008, Birgit 2009, Revell 2009, Huh 2010). The road network is especially an active feature in this domain (Mustiere 2008). However, for the matching of heterogeneous features, namely the matching of different features with some associations to each other, such as contour and river data matching, road and bridge matching, vegetation class and terrain level data matching, there is few concerns on matching methods or inconsistency detection. For associated heterogeneous features, there exists some spatial distribution knowledge that can act as matching rules. We can use this kind of spatial knowledge to detect the inconsistency between the integrated data and further by some methods corrects them to be consistent. The spatial knowledge could be topological consistency, semantic consistence and spatial association relationship supported by geoscience, for example the first law of geography (Tobler 1965).

This study attempts to investigate the matching and integration of contour and river data by spatial knowledge. The spatial knowledge (constraint) used in the integration process concerns the semantic relationship between contour and river distribution that 'river should flow into its talweg'. This implies for instance

that a river should go cross contour lines only at the valley points which are hidden in the contour representation. According to this, we are able to detect whether or not an inconsistency occurs between a river and a contour line by comparing the river and the contour. Violations to the constraint lead to logical inconsistencies and a poor visualization. An example of such inconsistencies is when a river deviates far away from its river bed in the valley.

The remainder of paper is organized as follows. Section 2 discusses the integrating and matching of contour and river data and analyzes the possible inconsistency between them. Section 3 presents a method based on Delaunay triangulation to detect such consistency through the bend analysis. Two adjustment methods, namely the river adjustment referenced to the contour and the contour adjustment referenced to the river is respectively developed in section 4. Section 5 concludes with the future works.

### 2. INTEGRATING AND MATCHING CONTOUR AND RIVER DATA

Terrain and hydrographic features are two important natural objects in spatial representation of real world. They usually act as reference (background) themes for man-made features (e.g. buildings, roads and infrastructures). Their change and evolution rely on the natural forces resulting in special associations between them. From the perspective of geomorphology, the topography is formed as a result of the interplay of internal forces (geological processes) and external forces (e.g. glaciers, wind, fluvial erosions) (Leopold et al., 1964). The internal force constructs the main geomorphologic characteristic (e.g. main valley), and the external force generates the specific terrain characteristics (e.g. minor valley). As abstract representations, contour and river lines should reflect their relationships in reality. Specifically, the relationships (constraints) between the two representations include:

(  $\rm i$  ) Intersection between a river and a contour should coincide with the valley point that is implicit in the contour line;

(ii) River should overlap its talweg, i.e., a virtual line connecting the lowest points along the entire length of a valley in its downward slope; and

(iii) River flows into a direction where height decreases most.

For various reasons, the hydrographic and terrain data may have been collected, maintained and generalized in separation, and the two may have independent update cycles. Consequently, violations to the above constraints may occur between river and contour data during their integration process, thus leading to inconsistencies such as rivers 'climbing uphill' and contours 'falling into watercourses'. The key to address the problems is to find the correct matching between rivers and fragments of groups of contours that form related valleys and talwegs, which in turn identifies the inconsistencies and corrects (repairs) the data. Liu et al. (2008) proposed a distance-based approach to detect inconsistencies between rivers and valleys. Their approach was designed specifically to cases where rivers are close and approximately parallel to their talwegs, which limits its applicability in practical uses. Furthermore, they did not address how to correct those detected inconsistencies.

This paper focuses on the integration of river and contour data, and presents a method for identifying inconsistencies and two strategies for adjusting the integrated data to correct the inconsistencies. First, we detect bottom points of valleys by characterizing individual contour lines with Delaunay triangulation. Second, the detected valley points are matched with the intersections between rivers and contours. The inconsistencies can be determined by the relationship between rivers and their talwegs in terms of distance and direction. Finally, identified inconsistencies are corrected either by displacing river geometries to their talwegs (keep contours fixed), or by moving fragments of contour groups so that they are correctly aligned to the rivers (keep rivers fixed).

# **3. INCONSISTENCY DETECTION**

This section describes approaches to extracting valley fragments and characteristic points (bottom points of valley fragments) on contours and to identifying inconsistencies.

#### 3.1 Extraction of valley fragments and characteristic points

We first detect the bottom points of the valley fragments and then connecting them to a path which stands for the talweg of one valley based on the method described in Ai et al. (2003). After constructing the Delaunay triangulation on an individual contour, we only consider the triangles on one side of the contour (the side facing to areas with a lower height and thus containing the valley fragments). Using the method described in Ai et al. (2000), valley fragments (bends) on the contour line can be identified and skeletons of the bends can be calculated. The bend detection relies on a criterion called bend depth (BendDepth) and its parameterization is discussed in Section 4.3. The generation of a skeleton is shortly explained in Fig. 1: a tree starts from the open mouth of a bend (defined by a special triangle that satisfies certain conditions as described in Ai et al. (2000)) and branches to different terminals of sub-bends; a skeleton is the longest path of the tree (e.g. the path from O to C in Fig. 1) and the bottom point of the valley fragment is the leaf node of the tree that define the path (e.g. C in Fig. 1). After applying the bend extraction and skeleton generation to all contours, we get groups of valley fragments and bottom points as shown in Fig. 2. The bottom points of valley fragments can be connected to form talwegs considering the connection distance, connection angle and intersection criteria.





Figure 1. The end point C is regarded as the valley bottom point according to the longest skeleton branch principle.

Figure 2. The extraction of terrain valley points from contour group

#### 3.2 Matching contour and river data

Rivers and streams should be within the scope of their containing valleys. That means that river segments should be logically linked to correct valley fragments to be meaningful. To

fulfil this goal, an algorithm is designed as follows to match rivers and talwegs extracted from the contours:

For each river segment:

I. Calculate the intersections between the river segment and contours, and we get a set of intersection points  $R = \{r1, r2...rn\}$ ;

II. For each  $ri \in R$ , a set of valley bottom points on the contour that intersect the river at ri can be collected as matching candidates set C i = {c1, c2, ..., cn}; by comparing elements in C with ri, we get a bottom point cj such that the length of the path between cj and ri along the contour is the shortest; cj is then the potential matching candidate and we denote the matching pair as mi (ri, cj);

III. Repeat step II, we get a set of matching pairs between the above-mentioned intersections and bottom points of valleys  $M = \{m1, m2...mn\}.$ 

Note that the cardinality of set Ci is determined by a userdefined number (SearchNum) which controls how many bottom points to search on each side of the intersection point. Fig. 3 demonstrates this matching process, where {c1, c2, c3, c4} is the set of matching candidates with respect to ri and c2 is the selected candidate based on the shortest path length principle. The matching pairs are highlighted in Fig. 4 as dashed lines between the intersections (solid triangles) and the bottom points (solid squares).



Figure 3. Building the matching relation



Figure 4. The matching relation between terrain valley points and intersection points of river network and contour

However, the above-mention algorithm is over-simplified and may fail to identify correct matching pairs in ambiguous situations (e.g. at places where branch rivers and valleys join). Fig. 5 demonstrates two typical types of incorrect matching. In Fig. 5(a), the correct matching pair m(R, B) was not found by the algorithm since the patch length between R and C is the shortest. The incorrect matching, m(R, A), in Fig. 5(b), on the other hand, was because the bottom point corresponding to R had not be successfully detected (in the contour representation the characteristics of the valley is not significant). To alleviate the above matching problems, we further propose a postprocessing to improve the correct rate of matching. The postprocessing takes into account water flow directions and talweg directions (in downward slope) and is formulated as follows:

I. Sort the elements of  $M = \{m1, m2..., mn\}$  in an order where the intersections between river segments and contours appears from upstream section to downstream section of the river network;

II. For each matching pair mi (ri, ci)  $\in$  M, calculate the angle value ai between directed line segments  $\overline{r_i r_{j-1}}$  and  $\overline{r_i r_{j+1}}$  and the angle value bi between directed line segments  $\overline{c_i c_{j-1}}$  and  $\overline{c_i c_{j+1}}$  and the angle value bi between directed line segments  $\overline{c_i c_{j+1}}$  and  $\overline{c_i c_{j+1}}$  along the clockwise direction; if (i = 1), ri-1 and ci-1 are replaced by the begin point of river segment, if (i = n), ri+1 and ci+1 are replaced by the end point of river segment; calculate the angle difference  $\theta_i = |a_i - b_i|$ ;



Figure 5. The wrong matching relation

III. Get the matching pair mp(rp, cp) that the angle difference  $\theta p$  is largest, and If ( $\theta p > \Delta \theta$ ), clear cp from the matching candidates set Cp, and search the new potential matching candidate for rp (if Cp is empty, clear the matching pair from M), then go to step II; If ( $\theta p \le \Delta \theta$ ), go to step IV;

IV. Get the set of matching pair after post- processing;

# 3.3 Inconsistency detection

The inconsistencies can be detected based on the distance between river lines and extracted talwegs. To be specific, the matching pairs m (R1, C1) and m (R2, C2) are identified in Fig. 6. The average distance between segments R1R2 and C1C2 can be estimated using the following equation:

AverageDis 
$$(\overline{R_1R_2}, \overline{C_1C_2}) = \frac{S}{(Length(R_1R_2) + Length(C_1C_2))/2}$$

where S is the area of the quadrilateral face R1R2C2C1. If the distance is larger than a given precision threshold, the river (segment) is regarded to be inconsistent with the contour fragments that define the valley and talweg, thus indicating an error that river 'climbing uphill'. Note that this precision threshold can be varied depending on specific applications and scales.

# 4. ADJUSTMENT METHODS

After detecting the inconsistencies between rivers and contours, this section presents two adjustment methods to correct the inconsistencies. The basic idea of the methods is to move (part of) the inconsistent data to align them to the reference data. The first method (Section 4.1) is to move inconsistent the river segment to a new position such that it lies properly in its (extracted) valley (or aligns to its talweg). The second method (Section 4.2) is to move the contour bend groups (that represent the valley) and make them consistent with the river. Both methods can be valuable considering the fact that there are cases where river data is more accurate that contours and where contour data is more accurate than rivers. Section 4.3 discusses and evaluates results after data adjustment.

#### 4.1 Moving rivers (contours fixed)

A straightforward solution to correct the identified inconsistencies is to replace the inconsistent parts (segments) of a river by corresponding talwegs (approximated by polylines connected from bottom points of valleys). This however may lead to new problems. For example, the approximate talweg may fail to capture the characteristics (e.g. sinuosity) of (part of) the original river. Further, extra intersections may be introduced between corrected river segments and contours (such as shown in Fig. 7(a)).

As a result, we displace (parts of) rivers using linear transformation. The transformation process is explained as follows with Fig. 6. For example, the river segment between intersection vertices R1 and R2 is to be moved to align to the talweg between C1 and C2. Formally, R1 and R2 are transformed to C1 and C2, respectively. Next, the other vertices

on the river segment are transformed by linear transformation such that the river characteristics are preserved as much as possible. Assume that the position of a vertex P on the segment to be moved is known (P.x and P.y), the new position Q can be determined using the following linear equation system:

$$\Delta x_{l} = C_{l}.x - R_{l}.x \qquad \Delta y_{1} = C_{1}.y - R_{1}.y$$
  

$$\Delta x_{2} = C_{2}.x - R_{2}.x \qquad \Delta y_{2} = C_{2}.y - R_{2}.y$$
  

$$Q.x = P.x + \Delta x_{l} \times \frac{L_{2}}{L_{l} + L_{2}} + \Delta x_{2} \times \frac{L_{l}}{L_{l} + L_{2}}$$
  

$$Q.y = P.y + \Delta y_{l} \times \frac{L_{2}}{L_{l} + L_{2}} + \Delta y_{2} \times \frac{L_{l}}{L_{l} + L_{2}}$$



Figure 6. The displacement of river network to match contour

where L1 and L2 are distances between R1 and P and between R2 and P, respectively (see also Fig. 6); C1 and C2 are used as control points. This simple transformation works fine in usual cases as shown in Fig. 6. However, when a valley represented by some contours has a more complex form, e.g., more sinuous (Fig. 7), the transformation controlled by the two intersection points does not suffice, leading probably to new intersections other than in the locations of bottom points of the valley (see Fig. 7(a)). In such a case, more control points are needed to reshape the original river segment. In this paper, the point lies in the middle of the river segment between R1 and R2 is chosen for this extra control point R3. To determine the new location of R3 after displacement, the Delaunay skeleton of the corresponding valley fragment (represented by the contour) is generated (Fig. 7(b)), after which R3 is snapped to its nearest point on the generated skeleton. Fig. 7(c) shows the transformation after R3 is added as control point. This improved adjustment requires that the shape of the original river segments should be similar to that of the generated skeleton.

Note that in the application of the above adjustment to the whole river network, nodes (where more than two segments join) are kept fixed in order to ensure the connectivity of the network.



(a) Generating new inconsistency (b) Adding new control points  $(C_3, R_3)$  (c) The transformation after adding new control points Figure 7. Adding new control points to avoid new inconsistency after river network

# 4.3 Moving contours (rivers fixed)

This method displaces the bend fragments in groups of contours, and aims to align the implicit talwegs to corresponding river segments. This is explained by moving the bottom point C and

the related valley fragment (light gray area) in Fig. 8 to new locations. The contour segment to be moved is from P1 through C to P2. The vertices on the segment are moved by applying affine transformation, where the control point C is moved to its nearest point C' on the river segment and P1 and P2 are kept fixed. Affine transformation is defined as follows:

$$P'.x = a \times P.x - b \times P.y + c$$

$$P.y = d \times P.x + e \times P.y + f$$

Note that when there are large discrepancies between river segments and their corresponding talwegs, the displacement of contours may change their topographic forms dramatically and may also leads to intersections between neighbouring contours.



Figure 8. The displacement of contour to match river



Figure 9. Contours intersected after displacement

### 4.4 Integration results and evaluation

The test data set of YUNNAN province (1:10k) characterized primarily by mountainous terrain was used for the experiment. A generalized data set at 1:50k was derived from this initial data, where a height interval of 40 meters was used for the generalization of contours. The purpose of the experiment is to integrate (match) river and contour class and check and correct inconsistencies

The parameters used are described as follows. BendDepth = 465 meters (ground unit) was used to ensure that all valley



Figure 10. Displacement of river network to match contour



Figure 11. Displacement of contour to match river

fragments (bends) in the test data are identified. An empirical value (SearchNum = 4) was set for searching matching candidates on each side of the intersection point. To improve the initial matched candidates, an angular deviation ( $\Delta\theta = 30$  %) was used. Finally, a precision threshold (Precision = 2.5 meter) was set to detect the inconsistencies in the integration.

After integration, we evaluated the matching result by a human subject working with visual methods. The experiment shows that 94.6% of the intersections between river segments and contours were correctly linked to their corresponding talwegs (bottom points of valleys). The inconsistency detection reports that 88.1% of the matched pairs were inconsistent (distances between river segments and their talwegs are larger than Precision). For both adjustment methods, approximately 90% and higher of the inconsistencies was successfully corrected

An in-depth analysis indicates that quality of inconsistency correction was influenced by the following factors. The first factor is the ratio (Ratio) between the distance between the river segment and its talweg (e.g. Fig. 6) and the length of the open mouth of the corresponding valley fragment (e.g. P1 and P2 in

Fig. 8). The second factor is the angular deviation ( $\Delta\theta$ ). It suggests further that for the inconsistent situations that are characterized by Ratio < 0.25 and  $\Delta\theta$  < 30 °, the first adjustment method (i.e. moving river fixing contour) worked reasonably well; whereas for the same situations the second adjustment method (moving contour fixing river) introduced more undesirable intersections, which needs to be improved in the future. Fig. 10 and 11 demonstrates results of integration after correction based on the two adjustment methods.

#### 5. CONCLUSIONS

The integration of river and contour (or features in general that are geographically related) is a common practice in establishing framework data. Inconsistencies between related features may quite open occur, e.g., the so-called river 'climbing uphill' issue, due to separate maintenance and processing of the data. This paper proposes an approach to automatically matching river segments and their talwegs (or valleys) and correcting identified inconsistencies. This approach enriches the original contour data set with valley fragments and talwegs of rivers using Delaunay triangulation based bend extraction and Delaunay skeletons. Further, it describes a distance-based method for inconsistency detection and two strategies for correcting the inconsistencies. The experiment shows that 94.6% of the intersection between rivers and contours were successfully linked to corresponding talwegs. In addition, 90% and higher of the inconsistencies were corrected by both strategies.

Unlike the data matching between identical objects of different scales, matching different features classes actually aims to maintain correct semantic relationships between them (e.g. river flows into its talweg, bridge over river and road, etc.). Such matching (integration) relies to a large extent on use of the spatial knowledge in specific domains. Therefore, the acquisition and formalization of the domain knowledge (rules) is the key to the integration of different thematic features.

The proposed approach can be improved in many ways. First, advanced techniques should be sought in order to address the distortion and intersection introduced by applying the second adjustment method (i.e. moving contours fixing rivers). Second, the presented approach can be extended for the integration and consistency maintenance between areal water bodies such as ponds and lakes and terrain represented by contours.

### REFERENCES

Ai, T., Zhu, G., Zhang, G., 2003. Extraction of Landform Features and Organization of Valley Tree Structure Based on Delaunay Triangulation Model. *Journal of Remote Sensing*, 7(4), pp.292-299.

Ai, T., Guo, R., Liu, Y., 2000. A Binary Tree Representation of Bend Hierarchical Structure based on Gestalt Principles. In: *Proceedings of the 9th International Symposium on spatial Data Handling*. Beijing.

Birgit, K., Huang, W., Haunert, J. et al., 2009. Matching River Datasets of different Scales. *Advances in GIScience*.

Ai, T., WU, H., 2000. Consistency Correction of Shared Boundary between Adjacent Polygons. *Journal of Wuhan Technical University of Surveying and Mapping*, 25(5), pp.426-431.

Chen, Y., Gong, J., Shi, W., 2007. A Distance-Based Matching Algorithm for Multi-scale Road Network. *Acta Geodaetica et Cartographica Sinica*, 36(1), pp. 84-90.

Devogele, T., Parent, C., Spaccapietra, S., 1998. On spatial database integration. *International Journal of Geographical Information Science*, 12(4), pp.335-352.

Goodchild, M F. 1991. Issues of Quality and Uncertainty. Advances in cartography, 113-140.

Huh, Y., Yu, K., Heo, J., 2010. Detecting conjugate-point pairs for map alignment between two polygon datasets. *Computer, Environment and Urban Systems*, 8(1), pp.1-13.

Hope, S., Kealy, A., 2008. Using Topological Relationships to Inform a data Integration Process, *Transactions in GIS*, 12(2), pp.267-283.

Liu, W., Chen, J., Deng, K., et al. 2008. Detecting the Spatial Inconsistency between the Updated Rivers and Valleys. *Journal of Image and Graphics*, 13(5), pp.1003-1008.

Mustiere, S., Devogele, T., 2008. Matching Networks with different Levels of detail. *GeoInformatica*, 12(4), pp. 435-453.

Revell, P., Antoine, B., 2009. Automated matching of building features of differing levels of detail: A case study. *ICC prpceedings*.

Safra, E., Kanza, Y., Sagiv, Y., 2010. Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets. *International Journal of Geographical Information Science*, 24 (1), pp.69-106.

Tobler, W. R., 1970. A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, 46, pp.234–240.

Walter, V., Fritsch, D., 1999. Matching spatial data sets: a statistical approach. *International Journal of Geographical Information Science*, 13(5), pp.445-473.

Zhao, D., Sheng, Y., 2010. Research on Automatic Matching of Vector Road Networks Based on Global Optimization. *Acta Geodaetica et Cartographica Sinica*, 39(4), pp.416-421.