# Case-based formalization of knowledge of digital terrain analysis

Cheng-Zhi Qin
State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS
Beijing, China
qincz@lreis.ac.cn

Xue-Wei Wu
State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS
Beijing, China
wuxw@lreis.ac.cn

Yan-Jun Lu
State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS
Beijing, China
luyj@lreis.ac.cn

Jing-Chao Jiang
State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS
Beijing, China
jiangjc@lreis.ac.cn

A-Xing Zhu
Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application and School of Geography
Nanjing Normal University
Nanjing, China

Department of Geography
University of Wisconsin-Madison
Madison, Wisconsin, USA
azhu@wisc.edu

*Abstract*—**Digital terrain analysis (DTA) in practical application is typically a workflow-building process which needs to organize the various DTA tasks properly and assign the algorithm (and its parameter settings) for each task. During this process it is crucial to use knowledge on specifying the proper algorithm and parameter settings for each DTA task according to the application context (such as the target task, the terrain condition of the study area, the DEM resolution, etc.), referred to as application-context knowledge. However existing DTA-assisted tools often cannot use application-context knowledge because this type of DTA knowledge has not been formalized to be available for inference in these tools. This is mainly because this type of DTA knowledge often exists in the minds of domain experts and is implicit in the text of case studies published in academic papers. This situation makes the DTA workflow-building process difficult for users, especially for non-experts. This study proposes a case-based formalization for application-context knowledge in the DTA domain and a corresponding case-based reasoning method. A preliminary experiment demonstrates the usability of the proposed case-based method.**

## I. Background

Digital terrain analysis (DTA) in practical application is typically a workflow-building process which needs to organize the various DTA tasks properly and assign the algorithm and its parameter settings for each task [1]. Tools to assist DTA have been developed to lighten the burden of this process on users (especially non-expert users). For this purpose, DTA-assisted tools not only need to integrate existing DTA algorithms, but also to use the formalized DTA knowledge [2].

The knowledge involved in DTA workflow-building can be classified into three types [2]:

1) task knowledge which describes the relationship between DTA tasks and their input/output;

2) algorithm knowledge which is the meta-data of a DTA algorithm and its parameters;

3) so-called application-context knowledge on how to specify the proper algorithm and its parameter settings for a DTA task according to the application context (such as application goals, characteristics of the study area, and DEM resolution) [3,4].

In: Geomorphometry for Geosciences, Jasiewicz J., Zwoliński Zb., Mitasova H., Hengl T. (eds), 2015. Adam Mickiewicz University in Poznań - Institute of Geoecology and Geoinformation, International Society for Geomorphometry, Poznań

Among the three types of DTA knowledge, both task and algorithm knowledge have been formalized by means of rule or semantic networks [2,4] and then used in exiting DTA tools. However, application-context knowledge, which is crucial for building a proper DTA model for specific applications and is more difficult for users to acquire than the other two types of knowledge, has no well-established formalization method for DTA tools. This study therefore focuses on formalizing application-context knowledge to aid in DTA workflow-building.

## II. BASIC IDEA

Unlike task and algorithm knowledge which is explicit, application-context knowledge is often implicit in the case studies documented in articles about applying DTA to specific study areas. The case method is a suitable way to formalize this type of knowledge, after which a case-based reasoning method can use this knowledge to solve a new similar problem [5].

## III. METHOD

### A. Case representation

In this study, the case is initially designed to use the features shown in Table 1 to describe the DTA application context.

TABLE I. CASE REPRESENTATION FOR DTA APPLICATION

| DTA application context | Feature | Formalized index |
|---|---|---|
| Application goal | DTA task type | DTA task enumeration |
| Data | DEM resolution | DEM grid size |
| | Data quality[a] | |
| Area characteristics | Position[a] | |
| | Area | Area (km$^2$) |
| | Terrain condition | Relief (m) |
| | | Hypsometric curve |
| | | Slope-relief histogram |
| | Other environmental conditions (such as climate, soil, and landuse) [a] | |

[a.] Not used in current study

### B. Case indexing

To enable case retrieval and comparison, indices were designed to formalize the features of a case (Table 1). Because the terrain condition of an area is crucial for choosing the DTA algorithm and its parameter settings, three indices were designed to describe the terrain condition of a study area:

- Relief.
- Hypsometric curve [6], which is widely used to characterize the stage of geomorphic development of a basin.
- Slope-relief histogram (the distribution of slope gradient for various relative elevation level), which is used to describe the configuration of the slope gradient as relief levels in a watershed. This is currently quantified by a two-dimensional frequency histogram with seven categories of slope gradient (0°–3°, 3°–8°, 8°–15°, 15°–25°, 25°–35°, 35°–45°, and 45°–90°) and ten categories of relief. Note that the effect of DEM resolution on slope gradient impacts the slope-relief histogram. However, this effect will be removed in later case-based reasoning because the index is taken into account when determining the DEM resolution. The equal classification of relief makes the resulting slope-relief histograms of different areas mathematically comparable. The design of this index ignores the relief difference between areas because the relief information is characterized by the "relief" index.

### C. Case-based reasoning

Case-based reasoning for solving new DTA application problems is designed to compute the similarity between the new application problem and each case involving the same DTA task. The similarity computation is designed as follows:

Step 1. Compute the similarity for each index between each case ($i$) and the new problem (Table 2).

TABLE II. SIMILARITY COMPUTATION FOR EACH INDEX BETWEEN A CASE AND THE NEW APPLICATION PROBLEM

| Index | Similarity computation on single index[a] |
|---|---|
| DEM grid size | $s_i = 2^{-\left(\left\|\lg r_{new} - \lg r_i\right\|/0.5\right)^{0.5}}$ |
| Area (km$^2$) | $s_i = 1 - s'_i/\max(s'_i)$, $s'_i = \left\|\lg(Area_{new}) - \lg(Area_i)\right\|$ |
| Relief (m) | $s_i = 1 - s'_i/\max(s'_i)$, $s'_i = \left\|Relief_{new} - Relief_i\right\|$ |
| Hypsometric curve | $s_i = 1 - s'_i/\max(1 - HypsoIntegral_{new}, HypsoIntegral_{new})$<br>$s'_i = \left\|HypsoIntegral_{new} - HypsoIntegral_i\right\|$ |
| Slope-relief histogram | $s_i = \frac{2\sum \min(SlpRlfHistogram_{new}, SlpRlfHistogram_i)}{\sum (SlpRlfHistogram_{new} + SlpRlfHistogram_i)}$ |

[a.] The subscript *i* means the *i*-th case; the subscript *new* means the application problem. *r* – DEM grid size; *HypsoIntegral* - integral value of the hypsometric curve; *SlpRlfHistgram* – slope-relief histogram.

Step 2. The similarity between a case and the new problem is calculated to be the minimum of the similarity on every index for this case.

Step 3. Among all cases the one with the highest similarity is retrieved as the solution case. The specific DTA algorithms and corresponding parameter-settings used in the solution case are then recommended for the new application. Currently, case adaptation is not included.

## IV. EXPERIMENT

### A. *Experimental design*

Taking as an example the determination of the catchment area (CA) threshold for extracting a drainage network, the authors prepared 12 cases from randomly-selected articles related to this task from journal in Chinese or English (Fig. 1).

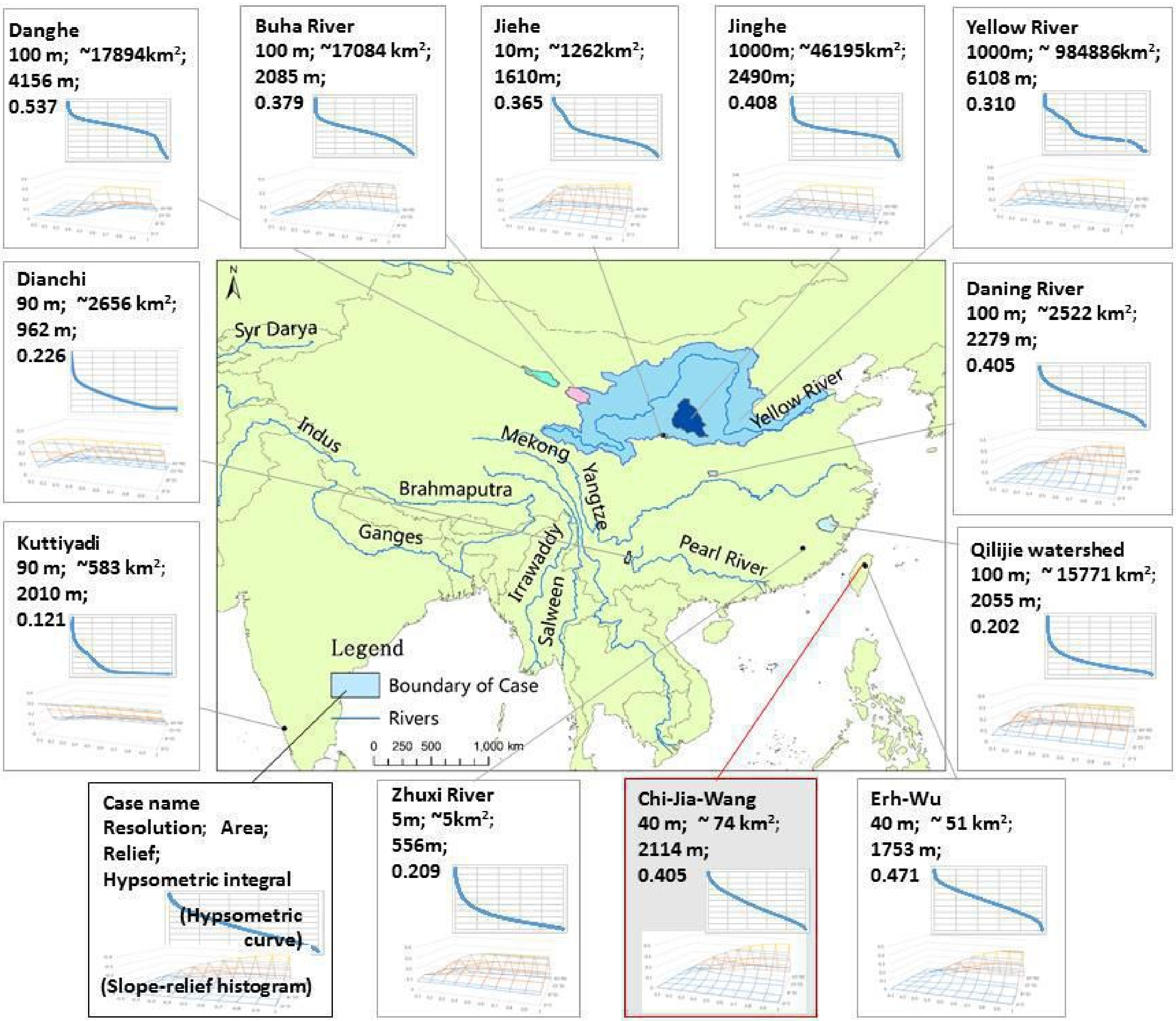


Figure 1. Case base in this study

It was assumed that the author(s) of each article set the CA threshold properly to match the study area. The 12 cases were manually prepared, whereas the reasoning process was automatic.

The case of the Chi-Jia-Wang watershed [7] was chosen as the new problem without determining the CA threshold. Then the proposed method was applied to the other cases to determine the CA threshold for the Chi-Jia-Wang watershed.

### *B. Experimental results and discussion*

The similarity values between the Chi-Jia-Wang watershed and each case are shown in Table 3. The solution came from the case of the Erh-Wu watershed, in which the CA threshold was 0.232 $km^2$. Compared to the threshold values from other cases, this was the closest to that proposed in the case of the Chi-Jia-Wang watershed (0.344 $km^2$). Table 3 further shows that in general, the lower the similarity of a case, the larger will be difference between the CA threshold value of the case in question and that of the Chi-Jia-Wang watershed. This indicates that the proposed case-based method is reasonable for use in this application.

TABLE III. SIMILARITY VALUES BETWEEN THE CHI-JIA-WANG WATERSHED AND EACH CASE USING THE PROPOSED CASE-BASED REASONING METHOD

| Case | CA threshold used in the case ($km^2$) | Similarity | Index with minimum similarity |
|---|---|---|---|
| Erh-Wu | 0.232 | 0.89 | Hypsometric curve |
| Daning River | 10 | 0.54 | Grid size |
| Qilijie Watershed | 0.61 | 0.40 | Slope-relief histogram |
| Jiehe | 4 | 0.40 | Slope-relief histogram |
| Zhuxi River | 0.03 | 0.39 | Grid size |
| Kuttiyadi | 0.1215 | 0.33 | Slope-relief histogram |
| Buha River | 6.5 | 0.29 | Slope-relief histogram |
| Danghe | 41 | 0.25 | Slope-relief histogram |
| Dianchi | 60 | 0.23 | Slope-relief histogram |
| Jinghe | 100 | 0.04 | Slope-relief histogram |
| Yellow River | 1000 | 0 | Area, Relief |

## V. SUMMARY

This study has proposed a case-based formalization for DTA application-context knowledge existing in journal papers. The corresponding case-based reasoning method was designed as an inference process for computing similarity. A preliminary experiment showed the usability of the proposed case-based method.

The proposed method can be implemented as an inference engine in a DTA modeling environment to provide the user with heuristic DTA modeling capabilities [4].

Ongoing research involves evaluating the feasibility of both the case indices and the similarity computation on each index in other DTA applications with different goals. Moreover, several questions raised by this preliminary research remain open. For example, if the solution case recommended by the case-based reasoning method based on the current case base has little similarity to the new application problem (which means that no case is similar enough to the new application problem), how should case adaptation be performed? To ensure an adequate number of cases in the case base, could an automatic method be developed to create relevant cases by crawling an article database?

## ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (No. 41422109), and the National Science & Technology Pillar Program of China (No. 2013BAC08B03-4).

## REFERENCES

[1] Hengl, T., and H.I. Reuter (Eds), 2008. “Geomorphometry: Concepts, Software, Application”. Developments in Soil Science - Volume 33, Elsevier, 765 p.

[2] Qin, C.-Z., Y.-J. Lu, A.-X. Zhu, and W.-L. Qiu, 2011. “Software prototyping of a heuristic and visualized modelling environment for digital terrain analysis”. In 11th International Conference on GeoComputation, July 20-22, University College London, UK.

[3] Lu, Y-.J., C.-Z. Qin, A.-X. Zhu, and W.-L. Qiu, 2012. “Application-matching knowledge based engine for a modelling environment for digital terrain analysis”. In 20th International Conference on Geoinformatics, June 15-17, Chinese University of Hong Kong, China.

[4] Qin, C.-Z., J.-C. Jiang, L.-J. Zhan, Y.-J. Lu, and A.-X. Zhu, 2013. “A Browser/Server-based prototype of heuristic modelling environment for digital terrain analysis”. In Geomorphometry’2013 Conference Proceedi-ngs, Edited by: Tang, G., Q. Zhou, T. Hengl, X. Liu, and F. Li, Nanjing Normal University.

[5] Watson, I., and F. Marir, 1994. “Case-based reasoning: a review”. The Knowledge Engineering Review, 9(4), 327-354.

[6] Strahler, A. N., 1952. “Hypsometric (area-altitude) analysis of erosional topography”. Geological Society of America Bulletin, 63, 1117-1142.

[7] Lin, W.-T., W.-C. Chou, C.-Y. Lin, P.-H. Huang, and J.-S. Tsai, 2006. “Automated suitable drainage network extraction from digital elevation models in Taiwan's upstream watersheds”. Hydrological Processes, 20(2), 289-306.