

1 **Revisions and responses on HESS-2015-539 (“Case-based**  
2 **knowledge formalization and reasoning method for digital**  
3 **terrain analysis — Application to determining the**  
4 **catchment area threshold for extracting drainage**  
5 **networks”)**  
6

7 The authors thank Dr. H. Mitsova for the comments on our manuscript. We make a point-by-  
8 point reply to these comments as below. A marked-up manuscript version showing the  
9 changes made is attached at the end of this document.  
10

11 *Comment 1: The revised manuscript now states clearly that when the DEM used in the case*  
12 *study was not available the authors used SRTM or ASTER DEM instead. It is important to*  
13 *show this in the Table 3 by including the resolution of the original DEM used in the*  
14 *articles (e.g. in the brackets along with the CA threshold) where this information is*  
15 *available..*

16 Response: Considering that the Appendix lists all cases used in the manuscript and Table  
17 3 in the manuscript lists 50 evaluation cases among them, we revised the Appendix to add the  
18 resolution of the DEM originally used in every case in the table of Appendix. We also  
19 revised the manuscript to declare that this information was included in the Appendix (the  
20 second paragraph of Section 5.1.1).  
21

22 *Comment 2: Throughout the article the term "relieve" is not used in the right context*  
23 *(please see <http://www.dictionary.com/browse/relieve?s=t>) leading to misunderstandings.*  
24 *Perhaps you wanted to say reduce, minimize or eliminate?*

25 Response: Thanks for the suggestion. We have revised the manuscript to use “reduce“ and  
26 “minimize“ instead of the “relieve“ used at several places in the original manuscript.

1

2 *Comment 3: Another case of misleading terminology is "magnitude of cell size" noted by*  
3 *two reviewers. The sentence where it is used "The difference in magnitude of cell size can*  
4 *better reflect the level of similarity between DTA 26 applications than the numerical*  
5 *difference in cell size." really does not make much sense. Perhaps you wanted to say that*  
6 *you used "absolute value of the difference in cell size"?*

7 Response: The difference in the logarithmic value of cell size can better reflect the level of  
8 similarity between DTA applications than the numerical difference in cell size. The greater  
9 the difference in the logarithm of cell size, the lower is the similarity. We have revised the  
10 manuscript to make it clear (the second paragraph of Section 4.2.2).

11

12 *Comment 4: One of the reviewers suggested that the title is too complex, however, in the*  
13 *revised paper the authors made it even longer and more complex - shorter title would be*  
14 *desirable.*

15 Response: We changed the title of the manuscript to be “Case-based knowledge  
16 formalization and reasoning method for digital terrain analysis — Application to extracting  
17 drainage networks“, which is six words shorter than the title of the former version.

18

19

# Case-based knowledge formalization and reasoning method for digital terrain analysis — Application to extracting drainage networks

删除的内容: determining the  
catchment area threshold for

C.-Z. Qin<sup>1,2,\*</sup> X.-W. Wu<sup>1,3</sup> J.-C. Jiang<sup>4</sup> A.-X. Zhu<sup>1,2,5,6</sup>

[1]{State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China}

[2]{Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China}

[3] {College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China}

[4] {Smart City Research Center, Hangzhou Dianzi University, Hangzhou, 310012, China}

[5] {Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA}

[6] {Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing 210023, China}

\*Correspondence to: C.-Z. Qin (qincz@lreis.ac.cn)

## Abstract

Application of digital terrain analysis (DTA), which is typically a modeling process involving workflow building, relies heavily on DTA domain knowledge of the match between the algorithm (and its parameter settings) and the application context (including the target task, the terrain in the study area, the DEM resolution, etc.), which is 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 situation makes the DTA workflow-building process difficult for users, especially non-expert users. This paper proposes a case-based formalization for DTA application-context knowledge and a corresponding case-based reasoning method. A case in this context consists of a series of indices that formalize the DTA

1 application-context knowledge and the corresponding similarity calculation methods for case-  
2 based reasoning. A preliminary experiment to determine the catchment area threshold for  
3 extracting drainage networks has been conducted to evaluate the performance of the proposed  
4 method. In the experiment, 124 cases of drainage network extraction (50 for evaluation and  
5 74 for reasoning) were prepared from peer-reviewed journal articles. Preliminary evaluation  
6 shows that the proposed case-based method is a suitable way to use DTA application-context  
7 knowledge to achieve a marked reduction in the modeling burden for users.

8

## 9 **1 Introduction**

10 Digital terrain analysis (DTA) is a useful approach to extracting topographic attributes and  
11 features from digital elevation model (DEM) and has been widely used in geography and  
12 related fields (Wilson, 2012). More and more users, including many with little knowledge of  
13 DTA, are becoming involved in DTA applications. Use of DTA is typically a non-trivial  
14 workflow-building process consisting of organizing the various DTA tasks and specifying the  
15 algorithm (including parameter settings) for each task (Hengl and Reuter, 2009). This process  
16 relies heavily on knowledge of DTA workflow building.

17 Knowledge used during DTA workflow building can be classified into three types (Qin et al.,  
18 2011): 1) task knowledge, which describes the relationship between DTA tasks and their  
19 input/output; 2) algorithm knowledge, which is the metadata of a DTA algorithm (including  
20 its parameters), such as the data type of input/output file, the number of parameters, and the  
21 valid range for each parameter; and 3) the so-called application-context knowledge consisting  
22 of how to specify the suitable algorithm and its parameter settings for a DTA task according  
23 to the application context (such as application goals, study area characteristics, and DEM  
24 resolution) (Qin et al., 2013). This knowledge is called application-matching knowledge in Lu  
25 et al. (2012). The best way to determine the optimal algorithm and its parameter-settings for a  
26 specific application should be the evaluation based on the field data. However, those field  
27 data might be not easy to be obtained at the beginning of the modeling and the evaluation  
28 process is often complicated for those non-expert users. Thus the application-context  
29 knowledge is crucial for building a reasonable DTA model for a specific application.

30 Among the three types of DTA knowledge, both task knowledge and algorithm knowledge  
31 have been formalized by means of rule or semantic networks (Russell and Norvig, 2009) and  
32 hence can be used in existing DTA-assisted tools, which include general purpose GIS

1 packages with DTA functionality (e.g., “Spatial Analyst“ toolbar in ArcGIS, r.\* modules in  
2 GRASS, “Terrain Analysis“ menu in SAGA, etc.) and domain-specific software (e.g.,  
3 Whitebox, TauDEM, etc.) (Hengl and Reuter, 2009). For example, by using these two types  
4 of DTA knowledge, the ModelBuilder module in ArcGIS can aid connecting a set of DTA  
5 algorithms to be an executable DTA workflow in an interactive visual way.

6 The application-context knowledge, which is crucial for building a suitable DTA model for a  
7 specific application, is more difficult to acquire than the other two types of knowledge.  
8 Currently, there is no well-established formalization method for application-context  
9 knowledge. Existing DTA-assisted tools consequently cannot use this type of knowledge to  
10 provide more effective support to DTA application modeling process (Qin et al., 2011). It is  
11 therefore difficult for users, especially those with little knowledge of DTA, to use DTA  
12 correctly and effectively. This situation exists mainly because this type of DTA knowledge is  
13 largely non-systematic and tacit knowledge, and often exists only in documents for specific  
14 case studies (DTA application instances) or even just in the experience of domain experts.

15 To solve this problem, this paper proposes a case-based formalization for DTA case studies  
16 involving DTA application-context knowledge and a corresponding case-based reasoning  
17 method. A DTA-assisted tool can then use this type of knowledge to reduce the difficulty of  
18 DTA application modeling.

19

## 20 **2 Basic idea**

21 Cases are a commonly used way of formalizing non-systematic knowledge in artificial  
22 intelligence. A case is a record of an existing problem-solving instance and its contextual  
23 information, which has two requisite parts: the problem and the solution (Kaster et al., 2005).

24 The problem describes the application purpose of the case and its contextual information. The  
25 solution is a set of methods (including their parameter settings) for achieving this purpose.

26 Note that the case is not the same as the concept of a prototype (Minda and Smith, 2001),  
27 which can also use existing instances to describe empirical knowledge and has been applied in  
28 the geographical domain (e.g., Qi et al., 2006; Qin et al., 2009). The prototype highlights the  
29 representativeness of the instances, whereas the case does not. Currently, most DTA  
30 application-context knowledge is empirical knowledge that often exists in application  
31 instances and is difficult to formalize as explicit rules or mathematical equations. In this

1 situation, the case is a suitable way to formalize DTA application-context knowledge (Lu et  
2 al., 2012).

3 Case-based reasoning (CBR) (Schank, 1983) is a method of solving problems by referring the  
4 solution of a new problem to the solutions of existing similar cases (Aamodt et al., 1994;  
5 Watson and Marir, 1994). Compared with traditional rule-based knowledge representation  
6 and reasoning methods, the case-based method transforms knowledge acquisition into case  
7 acquisition, with no need for an explicit expression of domain knowledge (Watson and Marir,  
8 1994). Therefore, the case-based method is suitable for application domains that lack a  
9 systematic expression of empirical domain knowledge. A case-based reasoning method could  
10 be designed to use DTA application cases to reduce the difficulty of DTA application  
11 modeling for users.

12

### 13 **3 Methodology**

14 According to the basic idea presented above, a case-based formalization methodology is  
15 designed for DTA application instances containing application-context knowledge and the  
16 corresponding inferences (Fig. 1). Case formalization and the corresponding case-based  
17 reasoning method are the two main stages in the methodology.

#### 18 **3.1 Case formalization**

19 Case formalization is the process of extracting and describing each individual case in a formal  
20 way, so that the case can be retrieved by a corresponding case-based reasoning method.  
21 Among the parts of a case, the case problem consists of a set of factors describing the  
22 contextual information associated with the case. This set of factors is quantified using a set of  
23 quantitative attributes that are directly involved in case-based reasoning. It is of crucial  
24 importance to design and quantify these factors properly for case-based reasoning. The  
25 solution part of a case records the candidate problem-solving result of the case-based  
26 reasoning and does not participate in the reasoning procedure. The case output is an optional  
27 part of the description that is used to record the status of factors describing the case problem  
28 after the case occurred (Kolodner, 1993). Therefore, the key to designing a case-based  
29 formalization of DTA application-context knowledge is how to choose and quantify a set of  
30 factors influencing DTA algorithm selection and parameter setting to describe the case  
31 problem appropriately.

1 According to the characteristics of DTA application modeling, the case problem can be  
2 described based on three groups of factors that influence DTA algorithm selection and  
3 parameter setting (Table 1): application purpose, data characteristics, and study area  
4 characteristics. For example, a single flow-direction algorithm (e.g., the classic D8 algorithm)  
5 is suitable for deriving flow accumulation from a SRTM DEM (with a resolution of 90 m) for  
6 drainage network extraction in high-relief areas, whereas a multiple flow-direction algorithm  
7 should be used with a 10-m DEM created from a contour map for estimating detailed spatial  
8 distribution of flow accumulation and other related regional topographic attributes (such as  
9 topographic wetness index) in a low-relief area. In this example, the choice between a single  
10 flow-direction algorithm and a multiple flow-direction algorithm is influenced by the  
11 application purpose (i.e., the DTA task of drainage network extraction or deriving the spatial  
12 distribution of regional topographic attributes), data characteristics (i.e., a SRTM DEM with  
13 90-m resolution or a contour-originated DEM with fine resolution), and study area  
14 characteristics (mainly terrain condition, e.g., high or low relief). This example shows the  
15 typical content of application-context knowledge in DTA application modeling.

16 Among these three groups of factors, the application purpose can be formalized by an  
17 enumeration-type variable. Data characteristics can be mainly described by the spatial  
18 resolution of the DEM, the type of data source, etc. In particular, the spatial resolution, which  
19 is often indicated by the grid cell size for the widely used grid-based DTA, is the most  
20 important factor among the data characteristics. The group of factors describing the study area  
21 characteristics related to DTA application-context knowledge could include location, area,  
22 terrain condition, and other environmental conditions (such as climate, geology, etc.).  
23 Generally, terrain condition in a study area comprehensively reflects the influence of all  
24 geographical processes on the landforms in the area. This means that terrain condition might  
25 be one of the most important factors influencing the DTA algorithm selection and parameter  
26 settings. Because of its comprehensiveness, the terrain condition factor should be quantified  
27 by multiple attributes during case-based formalization of DTA application-context knowledge.  
28 Different designs of the quantitative attributes will result in different case-based methods.

29 In a case-based formalization of DTA application-context knowledge, the solution part of a  
30 case can be formalized by recording the name of the DTA algorithm and the corresponding  
31 parameter values used in this case, which is much simpler than describing the case problem.  
32 The output part of a case, which is optional in the case-based formalization (Kolodner, 1993),

1 is set to be null because normally there is no change in the application context of a DTA  
2 application problem when the solution of this case is applied to the application problem.

### 3 **3.2 Case-based reasoning method**

4 Case-based reasoning is based on the principle that solutions for similar problems are often  
5 similar, even identical. Therefore, a new DTA application problem can be formalized in the  
6 same way as the case problem part in a prepared DTA case base and then be used in case-  
7 based reasoning by calculating the similarity between this new application problem and the  
8 problem part of each case in the case base. The solution of the case with the highest similarity  
9 (i.e., the most similar application context considered) is retrieved as the solution for the new  
10 DTA application problem. Note that in the conceptual framework of a case-based reasoning  
11 method, the solution of the retrieved case with the highest similarity might be further revised  
12 to adapt to the new application problem when the final solution for the new application  
13 problem is retained in the case base (Watson and Marir, 1994). However, the method  
14 developed in this preliminary study currently considers neither the revision nor the retention  
15 process.

16 Calculating the similarity between a new DTA application problem in case format and the  
17 problem part of each case in the DTA case base consists of the following two steps:

18 Step 1. Calculate the similarity of each individual attribute between the new application  
19 problem and the problem description of an existing case. As usual the range of the similarity  
20 value is [0, 1]; the larger the value, the more similar are the two cases. As mentioned above,  
21 the attributes used to formalize the problem part of a DTA application case may have different  
22 value types, such as enumeration type (e.g., application purpose), single-value type (e.g.,  
23 spatial resolution and area), or even a frequency distribution (e.g., hypsometric curve). For  
24 each attribute, a similarity function should be designed correspondingly to quantify the  
25 deviation on this attribute between the new application problem and an existing case. The  
26 design is generated in an empirical way and should match the domain knowledge.

27 Step 2. Synthesize the similarity values for every individual attribute to calculate the overall  
28 similarity between the new application problem and the problem description of an existing  
29 case. In the geographical domain, a minimum operator based on the limiting factor principle  
30 is often used to synthesize similarity values on multiple attributes (Zhu and Band, 1994; Qin  
31 et al., 2009). Other synthesis means such as weighted average could also be considered.

1

## 2 **4 Design of a detailed method**

3 In this section, the methodology presented in the previous section is concretized by designing  
4 a detailed case-based formalization method for DTA application instances containing  
5 application-context knowledge and the corresponding inferences. The key issue in method  
6 design is designing a set of quantitative attributes describing the case problem and the  
7 similarity function on each individual attribute. Because the gridded DEM is widely used in  
8 practical applications, this method is designed mainly for grid-based DTA, although the  
9 methodology is available for both grid- and vector-based DTA.

### 10 **4.1 Selection of attributes**

11 The set of quantitative attributes should be designed to effectively reflect the contextual  
12 information related to DTA application modeling, and be fit for the case-based reasoning to  
13 follow. The purpose of a DTA application case is naturally described by an enumeration-type  
14 attribute, i.e., the name of the target task. Here, cell size has been chosen as the attribute to  
15 quantify the data characteristics of a DTA application case (Table 2); other potential factors  
16 (such as type of data source) for describing data characteristics are not currently considered.

17 To describe the study area characteristics of a DTA application case, the area and the terrain  
18 condition of the case are considered in the current method (Table 2). Like cell size, area is an  
19 attribute with a single numeric value. Terrain condition is an important and comprehensive  
20 factor indicating the difference in study area characteristics between a new DTA application  
21 problem and an existing case.

22 In this study, the three following attributes were designed to describe the terrain condition  
23 factor empirically (Table 2):

24 1) Total relief. The total relief attribute, which is calculated as the maximum minus minimum  
25 elevation within the study area, is a commonly used value to describe the overall terrain  
26 condition of a study area.

27 2) Slope distribution. The slope distribution provides information on the proportions of  
28 different intensities of local relief in the area, which cannot be described by the total relief in  
29 the overall area and is useful for judging the reasonableness of a DTA algorithm selection and  
30 its parameter settings. To describe in detail the slope distribution in a study area, we

1 quantified it by an elevation-slope frequency distribution. For this purpose, the slope gradient  
2 was divided into seven classes:  $0^{\circ}$ – $3^{\circ}$ ;  $3^{\circ}$ – $8^{\circ}$ ;  $8^{\circ}$ – $15^{\circ}$ ;  $15^{\circ}$ – $25^{\circ}$ ;  $25^{\circ}$ – $35^{\circ}$ ;  $35^{\circ}$ – $45^{\circ}$ ; and  $45^{\circ}$ – $90^{\circ}$   
3 (Tang et al., 2006). According to the total relief within the study area, the elevation within the  
4 study area was classified into one of ten elevation classes with equal elevation step. The  
5 elevation-slope frequency distribution obtained in this way is a two-dimensional table with 10  
6 elevation class  $\times$  7 slope class data items. Considering that the DEM resolution has a strong  
7 influence on calculating the slope gradient and its frequency distribution (Chang and Tsai,  
8 1991; Grohmann, 2015), an elevation-slope cumulative frequency distribution were used here  
9 instead of the elevation-slope frequency distribution to provide a quantitative description that  
10 reduces the DEM resolution effect. The elevation-slope cumulative frequency in each  
11 elevation class is calculated by accumulating the number of cells within each slope gradient  
12 class from low to high class in this elevation class. Note that the 10-class division of elevation  
13 considers only the relative relationship among the elevation classes inside the study area. The  
14 elevation class might consist of a distinct elevation step for a study area, in which case the  
15 total relief of the study area would be ignored for this attribute. This proposed design appears  
16 to be not only a convenient way to automate similarity calculations in case-based reasoning,  
17 but also reasonable because the total relief attribute reflects the total relief information  
18 throughout the study area.

删除的内容: relieve

19 3) Landscape development stage for the study area, which can provide information on the  
20 geomorphic processes (mainly hydrological erosion process) affecting terrain conditions in a  
21 study area (often a watershed). This information is useful for judging the reasonableness of a  
22 choice of DTA algorithm and its parameter settings related to hydrological and erosion  
23 processes. In this study, the hypsometric curve (Strahler, 1952), which is normally used to  
24 analyze the landscape development stage of river basins, was used as an attribute to quantify  
25 this information.

26

27 In the proposed method, location is not used as a study area characteristics. This decision was  
28 made because the influence of the study area location in DTA application-context knowledge  
29 could be reflected by the terrain condition of the study area, which directly impacts the choice  
30 of DTA algorithm and parameter settings and has already been considered in the method. For  
31 similar reasons and for the sake of brevity, in the proposed method, environmental conditions  
32 other than terrain condition are not considered.

1 Table 2 lists the attributes used to formalize a case problem in this method.

## 2 **4.2 Similarity function on each individual attribute**

3 The design of the similarity function for an individual attribute should be compatible with the  
4 value type of the attribute and in accord with domain knowledge regarding the level of  
5 similarity due to the difference in the attribute value between the new application problem and  
6 an existing case. Curently the similarity function on individual attribute is designed to be with  
7 a simpler form before more detailed research could be conducted to improve it. For an  
8 attribute of the enumeration type, its similarity value between a new application problem and  
9 an existing case can be calculated by a Boolean function (Fig. 2a). When the attribute values  
10 are matched, the similarity value is 1, otherwise it is 0.

11 For an attribute of the single numeric value type, two commonly used kinds of basic similarity  
12 function are considered in this study: the linear function and the bell-shaped function (Fig. 2).  
13 Both kinds of similarity function accord with common sense in that the similarity is 1 for the  
14 minimum difference (i.e., zero) of attribute value, and the greater the difference in attribute  
15 value, the lower is the similarity. With the linear function, the similarity value is set to 0 or 1  
16 when the absolute difference of the attribute between a new application problem and an  
17 existing case reaches its maximum or minimum value. The similarity can be calculated for  
18 other difference values by linear interpolation (Fig. 2b). The similarity function based on a  
19 linear function fits the specification that the maximum difference in attribute values can be  
20 preset.

21 With the bell-shaped function, the maximum difference in attribute values is not easy to  
22 preset and does not need to be. A simplified version of the commonly used bell-shaped  
23 function (Shi et al., 2005; Qin et al., 2009; Fig. 2c) is:

$$24 \quad S = e^{-0.693 \times (|v_{new} - v_{case}|/w)^{0.5}}. \quad (1)$$

25 where  $S$  is the similarity between a new application problem and an existing case;  
26  $v_{new}$  and  $v_{case}$  are attribute values of the new application problem and the existing case  
27 respectively; and  $w$  is the shape-adjusting parameter of the function. When the difference  
28 between  $v_{new}$  and  $v_{case}$  is equal to  $w$ , the similarity  $S = 0.5$  (Fig. 2c). Some sort of numerical  
29 transformation on the attribute value could be necessary for the similarity calculation to yield  
30 a reasonable reflection of the similarity level due to differences in the attribute.

1 For an attribute of more complex type (such as a frequency distribution), a quantitative index  
2 should be designed to quantify the difference in an attribute between a new application  
3 problem and an existing case. Then the similarity on this attribute can be calculated based on  
4 this index, similarly to the single numeric-value type.

5 Based on these kinds of basic similarity function, similarity functions for each individual  
6 attribute used for case-based reasoning in this paper were designed as shown in Table 2. The  
7 following discussion introduces them one by one.

#### 8 **4.2.1 Name of target task**

9 The name of the target task is an attribute of the enumeration type. The similarity value for  
10 this attribute between a new application problem and an existing case can be calculated by a  
11 Boolean function. When the names of two target tasks match, the similarity value is 1,  
12 otherwise it is 0. This is a strict limit which prevents the proposed method from determining a  
13 case to be the solution case for a new application problem with a totally different task.  
14 Although this limit could be relaxed by developing more complicated classification of DTA  
15 target task (such as hierarchical classification or fuzzy classification), currently the boolean  
16 function is applied in a cautious manner.

#### 17 **4.2.2 Cell size**

18 Note that the numerical difference in cell size cannot well reflect the level of similarity  
19 between DTA applications. Taking an application with 10-m resolution as example, another  
20 application with a coarser resolution of 25 m is comparable to it from a cell size perspective,  
21 while a finer resolution with same numerical difference does not exist because it cannot be  
22 with less than or equal to 0 m.

23 The difference in the logarithmic value of cell size can better reflect the level of similarity  
24 between DTA applications than the numerical difference in cell size. The greater the  
25 difference in the logarithm of cell size, the lower is the similarity. According to this  
26 knowledge, a base-10 logarithmic transformation was applied to the cell size during the  
27 similarity calculations for balancing the decrease of similarity value for those situations with a  
28 coarser resolution or a finer resolution. Because it is not easy to preset the maximum of the  
29 attribute value after logarithmic transformation, the bell-shaped function based on Eq. (1) was  
30 used to calculate similarity for cell size. Furthermore,  $w$  in Eq. (1) is set to 0.5, which means

删除的内容: magnitude

删除的内容: magnitude

1 that the similarity in cell size between a new application problem and an existing case will  
2 decrease to 0.5 when their difference in cell size reaches one order of magnitude (e.g., 1 m vs.  
3 10 m, or vice versa). The similarity function used in the proposed method for cell size is  
4 shown in Table 2.

5 Note that the similarity value on cell size by such a similarity function will rapidly decrease to  
6 be about 0.58 when the resolution is coarsened to be double the resolution of a case or is  
7 refined to be a half of the case's resolution. The lower similarity value will deny the  
8 corresponding case to be a credible solution provider for the new application problem. This  
9 means that the proposed method does not suggest a large-step downscaling and upscaling  
10 application of existing cases.

### 11 **4.2.3 Area**

12 Like cell size, area of a study site is also an attribute of the single numeric value type. The  
13 greater the difference in magnitude between two areas, the lower is their similarity on area.  
14 Similarly to the design for the cell size attribute, a base-10 logarithmic transformation is  
15 applied to the area attribute and then the similarity function for this attribute is designed based  
16 on the bell-shaped function. The  $w$  in Eq. (1) has been set to 1.5 for the area attribute by trial  
17 and error (see Table 2).

### 18 **4.2.4 Total relief**

19 The greater the difference in total relief value between a new application problem and an  
20 existing case, the lower is the similarity. The maximum difference in total relief between two  
21 DTA application areas can be preset due to the geometric nature of the Earth. Hence, the  
22 similarity function for the total relief attribute was designed as a linear function using the  
23 absolute difference between the total relief of the new DTA application problem and that of  
24 existing case. Corresponding to a zero similarity value, the maximum difference between two  
25 total relief values is the larger of the total relief differences between the new application  
26 problem values and each of two extreme cases (a flat area with a total relief of zero, and an  
27 area with relief from the 8848 m of Mount Everest to sea level). The similarity function used  
28 in this method for the total relief attribute is shown in Table 2.

1 **4.2.5 Elevation-slope cumulative frequency distribution (describing the**  
2 **slope distribution)**

3 The elevation-slope cumulative frequency distribution is a two-dimensional table with 10  
4 class  $\times$  7 class data items. This two-dimensional table can be viewed as a DEM having a  
5 volume with a constant projected area. The greater the overlap in volume between the  
6 distribution of a new application problem and that of an existing case, the higher is the  
7 similarity. Therefore, the similarity function for the elevation-slope cumulative frequency  
8 distribution was designed as the ratio of the intersection volume to the union volume between  
9 two distributions (Table 2).

10 **4.2.6 Hypsometric curve (describing the landscape development stage)**

11 The hypsometric curve is often summarized as a single numeric value, the hypsometric  
12 integral (HI, with a value range of [0,1]), which can be used to classify landscape  
13 development into three stages: youth ( $HI > 0.6$ ), maturity ( $0.35 < HI < 0.6$ ), and old age ( $HI <$   
14  $0.35$ ) (Strahler, 1952). The HI was used to design a similarity function for the hypsometric  
15 curve between a new application problem and an existing case. Similarly to that of the total  
16 relief attribute, it is a linear function using the absolute difference of their HI values. When  
17 the absolute difference in HI is 0, the corresponding similarity is 1. The similarity is 0 for the  
18 maximum possible deviation from the HI of the new application problem (see Table 2).

19 **4.3 Calculation of the overall similarity**

20 The overall similarity between a new application problem and an existing case is calculated as  
21 the minimum of all similarity values for every individual attribute between the new  
22 application problem and the existing case. The use of a minimum operator means synthesizing  
23 the similarity values on every attributes in a cautious manner. On the one hand, the overall  
24 similarity result by this means is lower (i.e., higher uncertainty of reasoning result) than those  
25 from other synthesis means such as weighted average. On the other hand, a case with a low  
26 similarity value for any individual attribute will not get a higher overall similarity result by  
27 the minimum operator. This can prevent the proposed method from some unreasonable  
28 performance. For example, two cases with similar values of total relief and very different area  
29 sizes will have a low overall similarity, because of their low similarity on the area attribute  
30 and the overall similarity calculation by the minimum operator. This means that these two

1 cases would not be credible solution provider for each other, which is reasonable. Another  
2 example is that because of using the minimum operator, a low similarity on cell size between  
3 two cases will prevent that a fake high similarity on an attribute due to the DEM resolution  
4 effect (such as the attribute of elevation-slope cumulative frequency distribution) drives the  
5 overall similarity up. Therefore, the overall similarity calculation by a minimum operator  
6 should be more effective than that by a weighted-average operator.

7

## 8 **5 Experiment**

### 9 **5.1 Experimental design**

10 The extraction of a drainage network, one of the most important DTA applications, was taken  
11 as an example to evaluate the proposed method. The commonly used workflow of river  
12 network extraction based on a gridded DEM includes the following three DTA tasks in  
13 sequence: 1) preparing a DEM by filling in the artificial pits and removing absolutely flat  
14 areas; 2) using a flow direction algorithm to derive the spatial distribution of flow  
15 accumulation; and 3) setting a catchment area (CA) threshold to extract those positions with a  
16 flow accumulation larger than the CA threshold to be the drainage network. Although there  
17 are some variants of this workflow based on new algorithms (e.g., Metz et al., 2011), it does  
18 not influence the following experimental design for evaluating the proposed method.

19 In this DTA workflow, proper selection of the DTA algorithms (such as the DEM preparation  
20 algorithm and the flow direction algorithm) and of parameter values (e.g., the CA threshold)  
21 is based on DTA application-context knowledge. In many geographical information systems  
22 (such as ArcGIS), the DTA algorithm used for drainage network extraction has often been set  
23 to a default selection (e.g., the D8 algorithm as the default flow direction algorithm) in such a  
24 way that the user cannot choose the DTA algorithm. The CA threshold is an empirical  
25 parameter which varies with the study area characteristics and affects the extraction results  
26 directly. Current DTA-assisted tools often leave the choice of CA threshold for drainage  
27 network extraction to the user. However, it is difficult for users, especially non-expert users,  
28 to determine the appropriate threshold for their applications.

29 Therefore, this experiment was designed to focus on using the proposed method to determine  
30 the CA threshold for drainage network extraction. This means that the cases used in this  
31 experiment have the same name as the target task, i.e., drainage network extraction. The core

1 of the solution part of the cases is the parameter value, i.e., the CA threshold. Although this  
2 experiment is somewhat simplified, we believe that it can evaluate the proposed method as  
3 effectively as an experiment with a more complex design.

#### 4 **5.1.1 Preparation of a case base**

5 The case base prepared for this experiment includes 124 cases of drainage network extraction  
6 (Fig. 3). Each case originated from a peer-reviewed article related to the target task that was  
7 recently published in mainstream journals of related domains (such as *Water Resources*  
8 *Research, Hydrology and Earth System Sciences, Hydrological Processes, Computers &*  
9 *Geosciences, and Advances in Water Resources*; see the Appendix for the list of the articles  
10 used for cases). These articles were manually selected to be as reliable as possible. They are  
11 supposed to provide good solutions (might not be optimal) for their specific study areas based  
12 on experts' experience and knowledge of the target task. When a single flow direction  
13 algorithm (such as D8 algorithm) was adopted by most of these articles (a few articles did not  
14 state clearly the flow direction algorithm used), the CA threshold values adopted in these  
15 articles were highly varied (about  $10^{-3}$ – $10^3$  km<sup>2</sup>).

16 Each case was manually prepared from a journal article. The main work involved in preparing  
17 the case problem was to specify each attribute of the study area, whereas the work involved in  
18 preparing the case solution focused on recording the CA threshold used in the article.  
19 Normally, the cell size used is clearly stated in the article and can be filled in as the  
20 corresponding case attribute. However, this is often not true for other attributes. Given the  
21 study area of a case, an automatic program was applied to a free DEM dataset of the study  
22 area (mainly an SRTM DEM with a resolution of 90 m and an ASTER GDEM with a  
23 resolution of 30 m) to derive the other attributes (such as area, total relief, elevation-slope  
24 cumulative frequency distribution, and hypsometric curve) for each case. Original DEM  
25 adopted in some articles has a finer resolution than that of ASTER GDEM (i.e., 30 m; see the  
26 Appendix). However, those DEMs are often not easy to collect. This experiment used open  
27 DEM data to derive above case attributes and to make each of these attributes comparable  
28 between different cases.

删除的内容: SRTM DEM or

29 For the solution part of each case, the CA threshold given explicitly in each article was  
30 recorded directly. If the CA threshold was shown only implicitly in the drainage network  
31 figure in an article, it was determined based on visual comparison between the drainage

1 network given in the article and those extracted from the DEMs used to prepare other  
2 attributes of this case, using trial and error.

### 3 **5.1.2 Evaluation method**

4 Among the 124 cases in the case base, 50 cases randomly selected were used as independent  
5 evaluation cases, which were assumed to be new application problems without a solution and  
6 were solved by the reasoning method proposed. The other 74 cases were set aside as the case  
7 base to be used by the proposed case-based reasoning method.

8 To perform a quantitative evaluation of the highly varied CA threshold results from the  
9 proposed method on the 50 evaluation cases, an index was used, specifically the relative  
10 deviation of river density ( $E$ ):

$$11 \quad E = \frac{|RiverDensity^{reason} - RiverDensity^{origin}|}{RiverDensity^{origin}}. \quad (2)$$

12 where  $RiverDensity^{origin}$  and  $RiverDensity^{reason}$  are the river density values of a new  
13 application problem (i.e., an evaluation case), obtained respectively from the original CA  
14 threshold and the CA threshold solution obtained from the 74-case base by the proposed  
15 reasoning method.  $E$  is the relative deviation in river density for the evaluation case. The  
16 smaller the value of  $E$ , the more reasonable is the result obtained for the evaluation case using  
17 the proposed method. Four deviation levels of  $E$  were established empirically, i.e.,  $E \in [0,0.1]$ ,  
18  $E \in (0.1,0.25]$ ,  $E \in (0.25,0.5]$ , and  $E \in (0.5,+\infty)$ . Then the relationship between  $E$  and the  
19 similarity value of the solution case to the evaluation case was analyzed to discuss the  
20 performance of the proposed method. Representative cases were also selected to discuss the  
21 reasonableness of its similarity result obtained using the proposed method.

22 In this experiment, we also tested the effect of calculating the overall similarity by a simple  
23 average operator instead of the minimum operator used in the proposed method. The simple  
24 average was selected for comparison because it is the common representative of weighted  
25 average, and currently it is difficult to suggest a more complex weighted average for  
26 synthesizing similarity values on multiple attributes.

## 1 5.2 Experimental results and discussion

2 Table 3 lists the results of 50 evaluation cases solved by the proposed method using the case  
3 base presented in the previous section. For six evaluation cases, the proposed method arrived  
4 at the CA threshold result same as that originally recorded in the evaluation case. The counts  
5 of evaluation cases which got shorter and longer drainage networks (i.e., larger and smaller  
6 CA threshold respectively) from the proposed method are 16 and 28, respectively. The  
7 similarities between every evaluation case and its most similar case as reasoned by the  
8 proposed method were found in this experiment to lie within a value range from 0.47 to 0.9. A  
9 larger overall similarity value from the proposed method often corresponds to a smaller  
10 relative deviation of river density ( $E$ ) (Table 3). Note that the higher the similarity, the lower  
11 is the uncertainty of the result from the proposed method. This shows that the proposed  
12 method performs reasonably.

13 Table 4 summarizes the distribution of the similarity results of the evaluation cases from the  
14 proposed method among the deviation levels of the drainage network results using the solved  
15 CA thresholds. The counts of evaluation results with  $E \in [0,0.1]$ ,  $E \in (0.1,0.25]$ ,  $E \in (0.25,0.5]$ ,  
16 and  $E \in (0.5,+\infty)$  are 26, 16, 3, and 5 respectively (Table 4). For most of the evaluation cases,  
17 the results from the proposed method are with lower deviation level of  $E$ , which means that  
18 the proposed method performs effectively. All solution cases with higher similarity (above 0.7)  
19 to the evaluation cases produced drainage network results with smaller  $E$  values, whereas  
20 solution cases with lower similarity (below 0.7) often produced the drainage network results  
21 with larger  $E$  values. This shows the effectiveness with which similarity reflects uncertainty in  
22 the proposed method.

23 Taking the results on two evaluation cases, Godavari [1053] (the “[1053]” means that the  
24 original CA threshold recorded in the Godavari case was 1053 km<sup>2</sup>) and Burdekin [502]  
25 (“[502]” defined similarly) as examples, their most similar cases in the case base as reasoned  
26 by the proposed method were KrishnaRiver [908.08] and MahanadiRiver [891] respectively  
27 (Table 3). The CA threshold values from the solution of the most similar cases (908.08 km<sup>2</sup>  
28 and 891 km<sup>2</sup>) were applied respectively to the Godavari and Burdekin evaluation cases. The  
29 extracted drainage networks are with close spatial distribution as those extracted with the  
30 original CA thresholds of the evaluation cases (Fig. 5). Their values of relative deviation of  
31 river density are smaller (i.e., 0.07 and 0.24 respectively).

1 The evaluation results with larger  $E$  values also have lower similarities. This means that there  
2 is no case in the current case base that has an application context highly similar to that of the  
3 evaluation case. Hence, the solution from the proposed method has higher uncertainty and  
4 might lead to questionable or even unreasonable application results for new application  
5 problems. Taking the result for the YbbsRiver [1.01] evaluation case ( $E=0.43$ ) as an example,  
6 the similarities between this evaluation case and other cases in the case base depend mostly  
7 on the similarities on the cell size attribute during the case-based reasoning process proposed  
8 in this paper (Table 5). Because the cell size of the YbbsRiver case is 10 m, which is  
9 relatively unlike cell size (30 m or 90 m) of most other cases in the case base, the overall  
10 similarities between this evaluation case and these cases in the case base are mainly limited  
11 by the individual similarity on cell size when synthesizing the similarities on individual  
12 attributes by the proposed method. Furthermore, Table 5 shows that the CA threshold values  
13 of the cases with the top 10 highest similarity values to the YbbsRiver evaluation case would  
14 make large  $E$  value of the application result for the evaluation case ( $E$ : 0.33–21.73). The  
15 solution selected by the proposed method achieved a relatively better application result.

16 As for the reasoning results on the Kasilian [0.08] evaluation case ( $E=0.63$ ) using the  
17 proposed method, no individual attribute has a controlling effect on the overall similarity  
18 between the Kasilian evaluation case and the other cases in the case base (Table 6). The CA  
19 threshold values of the cases with the top 10 highest similarity values to the Kasilian  
20 evaluation case would almost always lead to a larger  $E$  value of the application result for the  
21 evaluation case ( $E$ : 0.48–0.92). The similarities between this evaluation case and the cases in  
22 the case base are lower (Table 6). This problem could be mitigated by extending the case base  
23 to contain cases with more combinations of data characteristics and study area characteristics.

24 The effect of calculating the overall similarity by a simple average operator instead of the  
25 minimum operator used in the proposed method was also evaluated (Table 3). When the  
26 minimum operator was replaced by the simple average operator, the overall similarity for  
27 every case increased and the lowest overall similarity among results for 50 evaluation cases  
28 increased from 0.47 to 0.68. Among 50 evaluation cases, the solutions for 13 evaluation cases  
29 from the proposed method changed because the cases with the highest similarity resulted by  
30 the simple average operator were different from those resulted by the minimum operator. Due  
31 to the synthesis by the simple average operator instead of the minimum operator, the relative  
32 deviation of river density ( $E$ ) increased for 10 of these 13 evaluation cases with different

1 solutions, when  $E$  slightly decreased for other 3 evaluation cases. The increase of  $E$  even  
2 reached 20–80 times for some cases (e.g., the evaluation cases YbbsRiver [1.01] and  
3 Batchawana [0.75]) with the overall similarity values larger than 0.8 (see Table 3). Because  
4 the overall similarity values by the simple average operator were larger than 0.8 for most of  
5 evaluation cases, there is no such a reasonable relationship between the overall similarity  
6 value and the  $E$  as the proposed method with the minimum operator achieved. This shows that  
7 the proposed method performed poorly when the simple average operator was used instead of  
8 the minimum operator. Therefore the synthesis by a minimum operator is proper for the  
9 proposed method.

10

## 11 6 Summary

12 Although DTA application-context knowledge is of key importance in building an appropriate  
13 DTA application, currently this type of knowledge has not been formalized to be available for  
14 DTA-assisted tools to minimize the modeling burden of DTA users (especially non-expert  
15 users). This paper has proposed a case-based methodology for formalizing DTA application-  
16 context knowledge and corresponding case-based reasoning. A detailed method based on this  
17 methodology has been developed. Taking drainage network extraction from a gridded DEM  
18 as an application example, 124 cases (50 for evaluation and 74 for reasoning) of drainage  
19 network extraction from peer-reviewed journal articles were used to evaluate the performance  
20 of the proposed method. Preliminary evaluation shows the reasonableness of the proposed  
21 case-based method. Combining the propose method with existing methods for using other two  
22 types of DTA knowledge (i.e., task and algorithm knowledge), automated DTA modeling  
23 could be implemented to make DTA easy to use for users and ensure that the result model is  
24 reasonable comparatively. This is valuable especially for non-expert users at the beginning of  
25 the modeling when field data for evaluation might be not easy to obtain.

26 Additional research is needed to enhance the proposed method. In this paper the proposed  
27 methodology is implemented as a primary method which focuses on DTA domain and  
28 considers the area and the terrain condition through a few simple attributes for describing the  
29 study area characteristics of a DTA application case. The design for the individual attributes  
30 and their quantification in each case could be improved to describe the domain-specific  
31 application-context knowledge in a more adaptive and efficient manner for various DTA  
32 application targets. Another possible improvement to the method would be to consider the

删除的内容: relieve

1 reliability of the case and revise the solution part of the case as suggested by case-based  
2 reasoning before applying the solution to the new application problem. The possibility of  
3 synthesizing the solutions of the cases in the base with higher similarity to build a solution to  
4 the new application problem could also be explored.

5 The size of the case base does matter. An expanded case base containing as many cases as  
6 possible with more combinations of all kinds of characteristics would improve the application  
7 effectiveness of the proposed method. The expansion of the case base (not only for the current  
8 target task, but also for other DTA application tasks) is valuable for evaluating the  
9 effectiveness of the case-based reasoning method and its successive versions. If case base is  
10 with a large size, machine learning algorithms (such as multidimensional regression) might be  
11 available for automatically calibrating the similarity functions and their shape-adjusting  
12 parameters used in the proposed method. Currently the size of current case base is still  
13 comparatively limited because current cases used in the experiment were mainly manually  
14 prepared from journal articles, except for certain attribute calculations (e.g., total relief,  
15 hypsometric curve), for which an automatic computer program was used. This inefficient way  
16 of preparing cases needs to be improved through developing automatic or semi-automatic  
17 case-creation methods.

18 In other geographical modeling domains, the task and algorithm knowledge have been used  
19 by formalization and inference methods and corresponding tools, such as Gregersen et al.  
20 (2007) and Škerjanec et al. (2014) in automated watershed modeling domain. For those  
21 domains in which the application-context knowledge is also largely non-systematic and tacit  
22 knowledge, the case-based idea proposed in this paper could also be available to combining  
23 with the existing automated modeling methods of using the task and algorithm knowledge in  
24 those domains, towards new geographical analysis tools which is easy to use for non-expert  
25 participants (Lin et al., 2013).

26

## 27 **Acknowledgements**

28 This study was supported by the National Natural Science Foundation of China (No.  
29 41422109, 41431177), and the National Science & Technology Pillar Program of China (No.  
30 2013BAC08B03-4).

31

## 1 **References**

- 2 Aamodt, A. and Plaza, E.: Case-based reasoning: foundational issues, methodological  
3 variations, and system approaches, *AI Commun.*, 7, 39-59, 1994.
- 4 Chang, K. and Tsai, B.: The effect of DEM resolution on slope and aspect mapping, *Cartogr.*  
5 *Geogr. Inf. Syst.*, 18, 69-77, 1991.
- 6 Gregersen, J. B., Gijssbers, P. J. A., and Westen, S. J. P.: OpenMI: open modelling interface, *J.*  
7 *Hydroinfo.*, 9(3), 175-191, 2007.
- 8 Grohmann, C. H.: Effects of spatial resolution on slope and aspect derivation for regional-  
9 scale analysis, *Comput. Geosci.*, 77, 111-117, 2015.
- 10 Hengl, T. and Reuter, H. I.: *Geomorphometry: Concepts, Software, Applications*, Elsevier,  
11 Amsterdam, 2009.
- 12 Kaster, D. S., Medeiros, C. B., and Rocha, H. V.: Supporting modeling and problem solving  
13 from precedent experiences: the role of workflows and case-based reasoning, *Environ. Modell.*  
14 *Softw.*, 20, 689-704, 2005.
- 15 Kolodner, J.: *Case-based Reasoning*, Morgan Kaufmann Publishers, San Mateo, 1993.
- 16 Lin, H., Chen, M., Lu, G., Zhu, Q., Gong, J., You, X., Wen, Y., Xu, B., and Hu, M.: Virtual  
17 geographic environments (VGEs): a new generation of geographic analysis tool, *Earth-Sci.*  
18 *Rev.*, 126, 74-84, 2013.
- 19 Lu, Y., Qin, C. Z., Zhu, A. X., and Qiu, W. L.: Application-matching knowledge based  
20 engine for a modelling environment for digital terrain analysis, in: *GeoInformatics*, The  
21 Chinese University of Hong Kong, China, 15-17 June 2012.
- 22 Metz, M., Mitasova, H., and Harmon, R. S.: Efficient extraction of drainage networks from  
23 massive, radar-based elevation models with least cost path search, *Hydrol. Earth Syst. Sci.*, 15,  
24 667-678, 2011.
- 25 Minda, J. P. and Smith, J. D.: Prototypes in category learning: The effects of category size,  
26 category structure, and stimulus complexity, *J. Exp. Psychol. Learn. Mem. Cogn.*, 27, 775-  
27 799, 2001.
- 28 Qi, F., Zhu, A-X., Harrower, M., and Burt, J. E.: Fuzzy soil mapping based on prototype  
29 category theory, *Geoderma*, 136, 774-787, 2006.

1 Qin, C.-Z., Zhu, A.-X., Shi, X., Li, B.-L., Pei, T., and Zhou, C.-H.: Quantification of spatial  
2 gradation of slope positions, *Geomorphology*, 110, 152-161, 2009.

3 Qin, C.-Z., Lu, Y.-J., Zhu, A.-X., and Qiu, W.-L.: Software prototyping of a heuristic and  
4 visualized modeling environment for digital terrain analysis, in: 11th International Conference  
5 on GeoComputation, University College London, UK, 20-22 July 2011.

6 Qin, C.-Z., Jiang, J.-C., Zhan, L.-J., Lu, Y.-J., and Zhu, A.-X.: A browser/server-based  
7 prototype of heuristic modelling environment for digital terrain analysis, in: *Geomorphometry*,  
8 Nanjing Normal University, China, 15-20 October 2013.

9 Qin, C.-Z., Wu, X.-W., Lu, Y.-J., Jiang, J.-C., and Zhu, A.-X.: Case-based formalization of  
10 knowledge of digital terrain analysis, in: *Geomorphometry for Geosciences (Proceedings of*  
11 *Geomorphometry'2015)*, edited by: Jasiewicz, J., Zwoliński, Zb., Mitasova, H., and Hengl, T.,  
12 Adam Mickiewicz University in Poznań, 209-212, 2015.

13 Russell, S. and Norvig, P.: *Artificial Intelligence: a Modern Approach (3rd Edition)*, Prentice  
14 Hall, 2009.

15 Schank, R. C.: *Dynamic Memory: a Theory of Reminding and Learning in Computers and*  
16 *People*, Cambridge University Press, New York, USA, 1983.

17 Shi, X., Zhu, A.-X., and Wang, R.: Fuzzy representation of special terrain features using a  
18 similarity-based approach, in: *Fuzzy Modeling with Spatial Information for Geographic*  
19 *Problems*, edited by: Petry, F. E., Robinson, V. B., and Cobb, M. A., Springer, Berlin  
20 Heidelberg, 233-251, 2005.

21 Škerjanec, M., Atanasova, N., Cerepnalkoski, D, Dzeroski, S., and Kompare, B.:  
22 Development of a knowledge library for automated watershed modeling, *Environ. Modell.*  
23 *Softw.*, 54, 60-72, 2014.

24 Strahler, A. N.: Hypsometric (area-altitude) analysis of erosional topography, *Bull. Geol. Soc.*  
25 *Am.*, 63, 1117-1142, 1952.

26 Tang, G. A. and Song, J.: Comparison of slope classification methods in slope mapping from  
27 DEMs, *J. Soil Water Conserv.*, 20, 157-160, 2006.

28 Watson, I. and Abdullah, S.: Developing case-based reasoning systems: a case study in  
29 diagnosing building defects, in: *Case Based Reasoning: Prospects for Applications (Digest No.*  
30 *1994/057)*, IEE Colloquium on. IET: 1/1-1/3, 1994.

- 1 Watson, I. and Marir, F.: Case-based reasoning: a review, *Knowl. Eng. Rev.*, 9, 327-354,
- 2 1994.
- 3 Wilson, J. P.: Digital terrain modelling, *Geomorphology*, 137, 107-121, 2012.
- 4 Zhu, A-X. and Band, L.: A knowledge-based approach to data integration for soil mapping,
- 5 *Can. J. Remote Sens.*, 20, 408-418, 1994.
- 6

1 Table 1. General composition of DTA application-context knowledge in a case-based  
2 formalization.

Part of case	Composition of DTA application-context knowledge
Case problem	Application purpose
	Data characteristics (spatial resolution, data source, etc.)
	Study area characteristics (location, area, terrain condition, other environmental conditions)
Case solution	DTA algorithm used and its parameter settings
Case output (optional)	(not considered in the current DTA application)

3

1 Table 2. Attributes used in this study to formalize the case problem and the corresponding  
 2 similarity functions for case-based reasoning using DTA application-context knowledge.

DTA application context			Similarity function
Factor group	Factor	Attribute	
Application purpose	Target task type	Name of target task	Boolean function
Data characteristics	Spatial resolution	Cell size (m)	$S_i = 2^{-(2 lgR_{new}-lgR_i )^{0.5}}$
	Area	Area (km <sup>2</sup> )	$S_i = 2^{-( lgArea_{new}-lgArea_i /1.5)^{0.5}}$
Characteristics of study area	Terrain condition	Total relief (m)	$S_i = 1 - S_i' / \max(8848 - Relief_{new}, Relief_{new})$  $S_i' =  Relief_{new} - Relief_i $
		Elevation-slope cumulative frequency distribution (describing slope distribution)	$S_i = \frac{Intersect(RlfSlp_{new}, RlfSlp_i)}{Union(RlfSlp_{new}, RlfSlp_i)}$
		Hypsometric curve (quantifying the landscape development stage)	$S_i = 1 - S_i' / \max(1 - HI_{new}, HI_{new})$  $S_i' =  HI_{new} - HI_i $

3 Note:  $S_i$  is the similarity (value range: [0, 1]) of an individual attribute between a new  
 4 application problem and the  $i$ -th case;  $R_{new}, R_i$  are the DEM resolutions (m) of the new  
 5 application problem and the  $i$ -th case respectively;  $Area_{new}, Area_i$  are the areas (km<sup>2</sup>) of the  
 6 new application problem and the  $i$ -th case respectively;  $Relief_{new}, Relief_i$  are the total relief  
 7 (m) of the new application problem and the  $i$ -th case respectively;  $RlfSlp_{new}, RlfSlp_i$  are the  
 8 histograms of the elevation-slope cumulative frequency distributions of the new application  
 9 problem and the  $i$ -th case respectively; and  $HI_{new}, HI_i$  are the hypsometric integrals of the  
 10 new application problem and the  $i$ -th case respectively.

11

- 1 Table 3. Evaluation results of the proposed method (in order of  $E$ ) and the corresponding results when a simple average operator was used  
 2 instead of the minimum operator.

Evaluation case [original CA threshold (km <sup>2</sup> )]	The proposed method (using a minimum operator)			Using a simple average operator instead of the minimum operator		
	Most similar case [CA threshold (km <sup>2</sup> )]	Overall similarity	$E$	Most similar case [CA threshold (km <sup>2</sup> )]	Overall similarity	$E$
UpperRhone [81]	KernRiver [81]	0.83	0	KernRiver [81]	0.92	0
MicaCreek1 [0.03]	MicaCreek2 [0.03]	0.85	0	MicaCreek2 [0.03]	0.95	0
WillowRiver [40.5]	Bowron [40.5]	0.89	0	Bowron [40.5]	0.94	0
YamzhogYumCo [12.15]	CedoCaka [12.15]	0.75	0	CedoCaka [12.15]	0.86	0
Stanley [0.2]	Pettit [0.2]	0.73	0	Pettit [0.2]	0.86	0
Alturas [0.2]	Pettit [0.2]	0.68	0	Pettit [0.2]	0.85	0
WarregoSC2 [4.42]	WarregoSC4 [4.33]	0.83	0.01	WarregoSC4 [4.33]	0.94	0.01
Toachi [3.13]	SanPabloLaMana [3.07]	0.76	0.01	SanPabloLaMana [3.07]	0.88	0.01
FuRiver [0.009]	CameronHighlands [0.0093]	0.64	0.02	CameronHighlands [0.0093]	0.84	0.02
Davidson [0.48]	UpperMcKenzie [0.5]	0.59	0.02	Haeon [0.55]	0.8	0.05
Komati [36.64]	Bowron [40.5]	0.60	0.04	Bowron [40.5]	0.79	0.04
UpperTananim [0.52]	Bellever [0.59]	0.81	0.05	Bellever [0.59]	0.91	0.05
Crocodile [36.30]	Bowron [40.5]	0.74	0.05	Bowron [40.5]	0.87	0.05
Cheakamus [8.1]	LiWuRiver [9]	0.80	0.05	LiWuRiver [9]	0.87	0.05

Susquehanna [810]	DoloresR_Cisco [763.17]	0.71	0.05	DoloresR_Cisco [763.17]	0.86	0.05
RoubachPlaten [0.32]	HJA [0.27]	0.80	0.06	HJA [0.27]	0.9	0.06
Godavari [1053]	KrishnaRiver [908.08]	0.80	0.07	KrishnaRiver [908.08]	0.92	0.07
Gard [8.09]	JuniataRiver [6.98]	0.69	0.07	Babaohe [18]	0.82	0.3
Urola [5.22]	OitaRiver [6.48]	0.79	0.07	OitaRiver [6.48]	0.91	0.07
UpperDalya [0.45]	Bellever [0.59]	0.82	0.08	Bellever [0.59]	0.94	0.08
WarregoSC3 [5.05]	WarregoSC4 [4.33]	0.77	0.08	WarregoSC4 [4.33]	0.89	0.08
SanJuanR_Bluff [708.35]	ColoradoR_Cameron [794]	0.87	0.08	ColoradoR_Cameron [794]	0.93	0.08
Monastir [3.47]	Baba [4.19]	0.80	0.08	OitaRiver [6.48]	0.9	0.25
SouthPark [24.3]	CooperRiver [29.34]	0.78	0.09	CooperRiver [29.34]	0.9	0.09
Rhone [398.97]	PoRiver [486]	0.86	0.1	PoRiver [486]	0.94	0.1
Bishop_Hull [0.86]	Brue [0.70]	0.78	0.1	Brue [0.70]	0.91	0.1
AlzetteEttel [0.23]	Bellebeek [0.31]	0.76	0.12	SouthForkNew[2.7]	0.87	0.7
PedlerCreek [0.41]	Bellever [0.59]	0.70	0.12	Bellever [0.59]	0.83	0.12
Fengman [243]	UpperGuadiana [324]	0.66	0.14	CedoCaka[12.15]	0.79	3.21
Cauvery [1053]	ColoradoR_Cameron [794]	0.77	0.15	ColoradoR_Cameron [794]	0.93	0.15
MiddleColorado [5.93]	WarregoSC4 [4.33]	0.85	0.15	WarregoSC4 [4.33]	0.94	0.15
LuckyHills [6.3]	SouthForkNew [2.7]	0.71	0.15	SouthForkNew [2.7]	0.88	0.15
Limpopo [987.22]	DoloresR_Cisco [763.17]	0.61	0.16	DoloresR_Cisco [763.17]	0.85	0.16
LittlePiney [2.84]	Blackwater [4.35]	0.86	0.17	Blackwater [4.35]	0.94	0.17

ChiJiaWang [0.34]	ErhWu [0.23]	0.80	0.17	ErhWu [0.23]	0.89	0.17
Hailogou [2.03]	SanPabloLaMana [3.07]	0.68	0.18	HunzaRiver[56.7]	0.79	0.79
Batchawana [0.75]	ClearCreek [1.22]	0.58	0.2	XianNanGou[0.004]	0.81	17.16
Liene [5.37]	LiWuRiver [9]	0.74	0.2	LiWuRiver [9]	0.85	0.2
Zwalm [0.36]	Haean [0.55]	0.73	0.2	Haean [0.55]	0.87	0.2
TapajosRiver [2720]	SaoFrancisco [5160]	0.67	0.23	SaoFrancisco [5160]	0.84	0.23
Burdekin [502]	MahanadiRiver [891]	0.90	0.24	MahanadiRiver [891]	0.95	0.24
Garonne [247.68]	PoRiver [486]	0.71	0.24	PoRiver [486]	0.87	0.24
NorthEsk [1.22]	SanPabloLaMana [3.07]	0.63	0.33	UpperGuadiana[324]	0.82	0.98
YbbsRiver [1.01]	Davidson [0.48]	0.69	0.43	CameronHighlands[0.0093]	0.84	11.44
Cordevole [0.68]	SouthForkNew [2.7]	0.69	0.46	HJA[0.27]	0.83	0.67
NarayaniRiver [130]	Durance [51.21]	0.51	0.52	HunzaRiver[56.7]	0.75	0.45
YaluTsangpo [81.56]	SalmonRiver [486]	0.47	0.55	RhoneRiver[40.5]	0.68	0.41
Kasilian [0.08]	Haean [0.55]	0.63	0.63	Haean [0.55]	0.83	0.63
UpstreamGarza [0.2]	NorsmindeFjord [4.05]	0.69	0.74	Haean [0.55]	0.83	0.37
Zhanghe [33.11]	Lonquen [7.29]	0.69	1.06	Lonquen [7.29]	0.89	1.06

1 Table 4. Relationship between  $E$  and the similarity value ( $S$ ) of the solution case to the  
 2 evaluation case.

	$S \in [0.8, 1]$	$S \in [0.7, 0.8)$	$S \in [0.6, 0.7)$	$S \in [0, 0.6)$	Total count of cases
$E \in [0, 0.1]$	10	11	3	2	26
$E \in (0.1, 0.25]$	3	8	4	1	16
$E \in (0.25, 0.5]$	0	0	3	0	3
$E \in (0.5, +\infty)$	0	0	3	2	5

3

1 Table 5. Top 10 similarity values between the YbbsRiver evaluation case and existing cases  
 2 as reasoned by the proposed method.

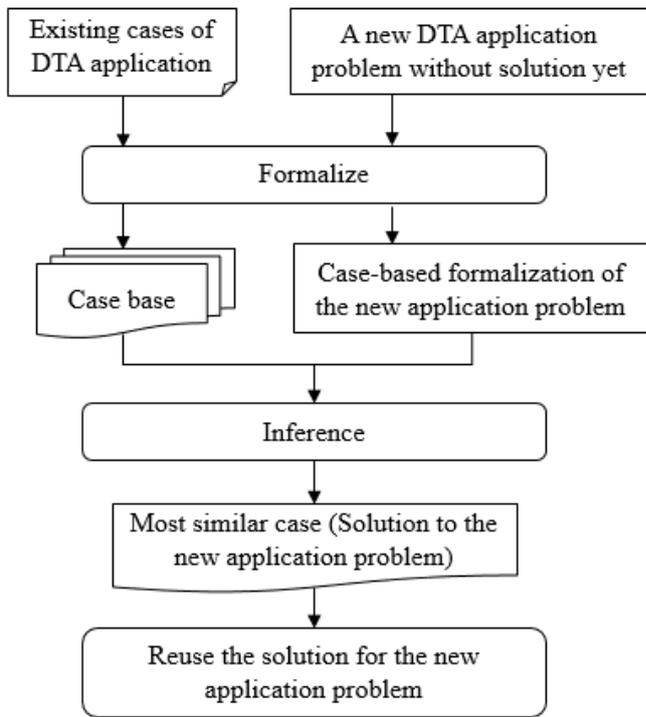
Case name	Similarity value on individual attribute					Overall similarity	<i>E</i>
	Cell size	Area	Total relief	Elevation-slope distribution	Hypso metric curve		
UpperMcKenzie	1	0.73	0.90	0.62	0.92	0.62	0.43
XianNanGou	0.58	0.61	0.88	0.59	0.76	0.58	21.73
NorsmindeFjord	0.58	0.74	0.84	0.64	0.91	0.58	0.44
Pettit	1	0.56	0.96	0.62	0.76	0.56	1.19
Bellebeek	0.54	0.69	0.83	0.54	0.81	0.54	0.73
Haean	0.51	0.65	0.94	0.78	0.93	0.51	0.33
MicaCreek2	0.51	0.53	0.89	0.62	0.75	0.51	5.23
SouthForkNew	0.51	0.69	0.89	0.76	0.52	0.51	0.35
Babaohe	0.51	0.57	0.88	0.73	0.90	0.51	0.73
ClintonRiver	0.51	0.59	0.85	0.56	0.55	0.51	0.79

3

1 Table 6. Top 10 similarity values between the Kasilian evaluation case and existing cases as  
 2 reasoned by the proposed method.

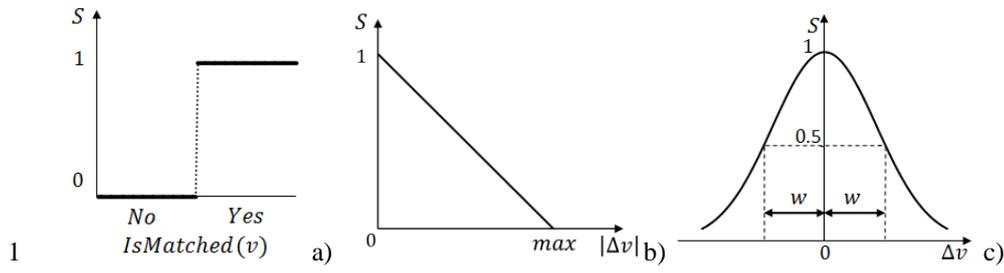
Case name	Similarity value on individual attribute					Overall similarity	<i>E</i>
	Cell size	Area	Total relief	Elevation-slope distribution	Hypso metric curve		
Haean	0.63	0.92	0.83	0.83	0.93	0.63	0.63
SanPabloLaMana	0.61	0.61	0.74	0.60	0.76	0.60	0.84
Brue	0.61	0.67	0.73	0.59	0.88	0.59	0.66
OitaRiver	0.61	0.57	0.95	0.73	0.96	0.57	0.91
Baba	0.61	0.55	0.98	0.83	0.97	0.55	0.87
JuniataRiver	0.63	0.55	0.78	0.64	0.86	0.55	0.92
NorsmindeFjord	0.54	0.74	0.71	0.72	0.95	0.54	0.87
Lonquen	0.61	0.52	0.82	0.73	0.93	0.52	0.92
HJA	0.63	0.90	0.86	0.51	0.64	0.51	0.48
Bellever	0.61	0.78	0.74	0.50	0.68	0.50	0.63

3

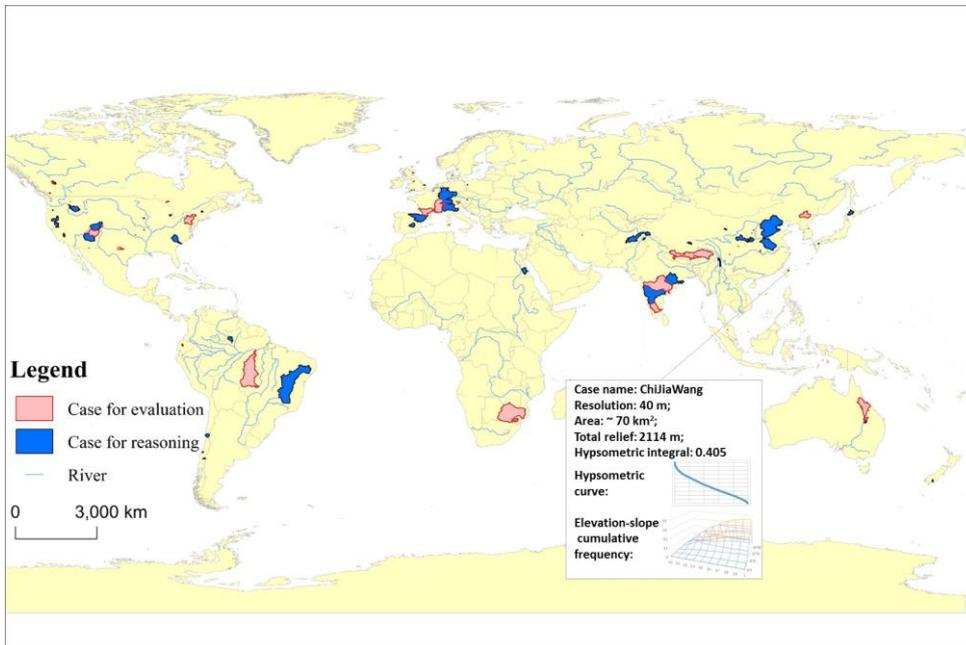


1  
2  
3  
4  
5

Figure 1. Structure of the case-based formalization and reasoning method for DTA application-context knowledge.



1  
2  
3 Figure 2. Basic kinds of similarity function: a) Boolean function; b) linear function; c) bell-  
4 shaped function.  
5



1  
 2 Figure 3. Spatial distribution of the cases used in this study (the box in the map shows an  
 3 example of a formalized case).  
 4

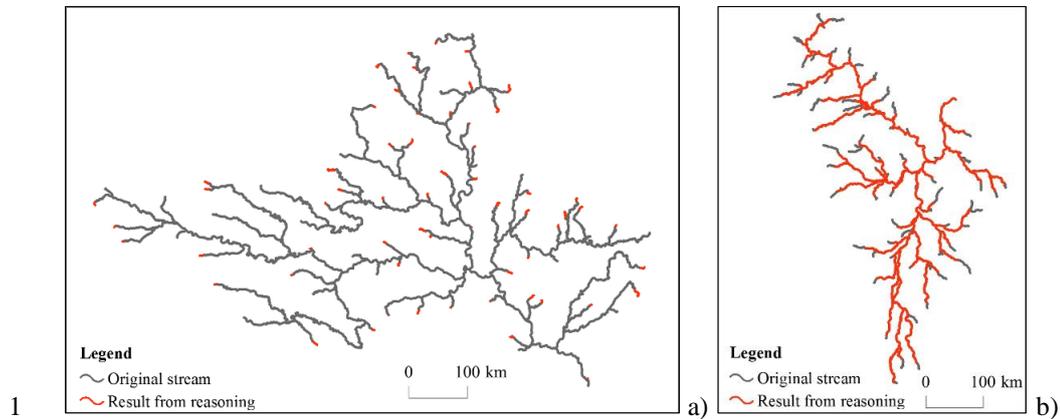


Figure 4. Comparison between the original drainage network of an individual evaluation case and its extraction result using case-based reasoning: a) Godavari case with an underestimated CA threshold; and b) Burdekin case with an overestimated CA threshold.

1 Appendix. List of cases

带格式表格

Case name <u>(cell size used originally)</u>	Source paper
LittlePiney <u>(90 m)</u>	Botter G. Flow regime shifts in the Little Piney creek (US)[J]. <i>Advances in Water Resources</i> , 2014, 71: 44-54.
PoRiver <u>(90 m)</u>	Lanzoni S, Luchi R, Pittaluga M B. Modeling the morphodynamic equilibrium of an intermediate reach of the Po River (Italy)[J]. <i>Advances in Water Resources</i> , 2015, 81: 95-102.
UpperMcKenzie <u>(10 m)</u>	Di Lazzaro M, Zarlenga A, Volpi E. Hydrological effects of within-catchment heterogeneity of drainage density[J]. <i>Advances in Water Resources</i> , 2015, 76: 157-167.
Babaohe <u>(30 m)</u>	Lei F, Huang C, Shen H, et al. Improving the estimation of hydrological states in the SWAT model via the ensemble Kalman smoother: Synthetic experiments for the Heihe River Basin in northwest China[J]. <i>Advances in Water Resources</i> , 2014, 67: 32-45.
OldMansCreek <u>(30 m)</u>	Ayalew T B, Krajewski W F, Mantilla R, et al. Exploring the effects of hillslope-channel link dynamics and excess rainfall properties on the scaling structure of peak-discharge[J]. <i>Advances in Water Resources</i> , 2014, 64: 9-20.
UpstreamGarza <u>(20 m)</u>	Balistrocchi M, Grossi G, Bacchi B. Deriving a practical analytical-probabilistic method to size flood routing reservoirs[J]. <i>Advances in Water Resources</i> , 2013, 62: 37-46.
Peach eater <u>(50 m)</u>	Kim J, Warnock A, Ivanov V Y, et al. Coupled modeling of hydrologic and hydrodynamic processes including overland and channel flow[J]. <i>Advances in Water Resources</i> , 2012, 37: 104-126.
Cauvery <u>(90 m)</u>	Konar M, Todd M J, Muneeppeerakul R, et al. Hydrology as a driver of biodiversity: Controls on carrying capacity, niche formation, and dispersal[J]. <i>Advances in Water Resources</i> , 2013, 51: 317-325.
Krishna <u>(90 m)</u>	
Krishna1 <u>(90 m)</u>	
Godavari <u>(90 m)</u>	
Klodawka <u>(15 m)</u>	Jasiewicz J Ł, Metz M. A new GRASS GIS toolkit for Hortonian analysis of drainage networks[J]. <i>Computers &amp; Geosciences</i> , 2011,

	37(8): 1162-1173.
Chabagou <a href="#">(50 m)</a>	Li T, Wang G, Chen J. A modified binary tree codification of drainage networks to support complex hydrological models[J]. Computers & Geosciences, 2010, 36(11): 1427-1435.
SaoFrancisco <a href="#">(200 m)</a>	Saraiva A G S, Paz A R. Multi-step change of scale approach for deriving coarse-resolution flow directions[J]. Computers & Geosciences, 2014, 68: 53-63.
TapajosRiver <a href="#">(200 m)</a>	
CooperRiver <a href="#">(30 m)</a>	Castronova A M, Goodall J L. A hierarchical network-based algorithm for multi-scale watershed delineation[J]. Computers & Geosciences, 2014, 72: 156-166.
MiddleColorado <a href="#">(90 m)</a>	Karimipour F, Ghandehari M, Ledoux H. Watershed delineation from the medial axis of river networks[J]. Computers & Geosciences, 2013, 59: 132-147.
FuRiver <a href="#">(10 m)</a>	Xu C, Xu X, Dai F, et al. Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China[J]. Computers & Geosciences, 2012, 46: 317-329.
JuniataRiver <a href="#">(30 m)</a>	Yu X, Bhatt G, Duffy C, et al. Parameterization for distributed watershed modeling using national data and evolutionary algorithm[J]. Computers & Geosciences, 2013, 58: 80-90.
YoungWomansCreek <a href="#">(30 m)</a>	
YaluTsangpo <a href="#">(85 m)</a>	Wang H, Fu X, Wang G. Multi-tree Coding Method (MCM) for drainage networks supporting high-efficient search[J]. Computers & Geosciences, 2013, 52: 300-306.
KaghanValley <a href="#">(90 m)</a>	Dehvari A, Heck R J. Removing non-ground points from automated photo-based DEM and evaluation of its accuracy with LiDAR DEM[J]. Computers & Geosciences, 2012, 43: 108-117.
CameronHighlands <a href="#">(10 m)</a>	Lim S L, Sagar B S D, Koo V C, et al. Morphological convexity measures for terrestrial basins derived from digital elevation models[J]. Computers & Geosciences, 2011, 37(9): 1285-1294.
W_Kharit <a href="#">(30 m)</a>	Milewski A, Sultan M, Yan E, et al. A remote sensing solution for estimating runoff and recharge in arid environments[J]. Journal of Hydrology, 2009, 373(1): 1-14.

ChiJiaWang <a href="#">(40 m)</a>	Lin W T, Chou W C, Lin C Y, et al. Automated suitable drainage network extraction from digital elevation models in Taiwan's upstream watersheds[J]. Hydrological Processes, 2006, 20(2): 289-306.
ErhWu <a href="#">(40 m)</a>	
Demeni <a href="#">(90 m)</a>	Getirana A C V, Bonnet M P, Rotunno Filho O C, et al. Improving hydrological information acquisition from DEM processing in floodplains[J]. Hydrological Processes, 2009, 23(3): 502-514.
Batchawana <a href="#">(5 m)</a>	Creed I F, Hwang T, Lutz B, et al. Climate warming causes intensification of the hydrological cycle, resulting in changes to the vernal and autumnal windows in a northern temperate forest[J]. Hydrological Processes, 2015, 29: 3519–3534.
Hailougou <a href="#">(90 m)</a>	Xing B, Liu Z, Liu G, et al. Determination of runoff components using path analysis and isotopic measurements in a glacier-covered alpine catchment (upper Hailuogou Valley) in southwest China[J]. Hydrological Processes, 2015, 29, 3065–3073.
Bellebeek <a href="#">(25 m)</a>	Loosvelt L, Pauwels V, Verhoest N E C. On the significance of crop-type information for the simulation of catchment hydrology[J]. Hydrological Processes, 2015, 29(6): 915-926.
WeiRiver <a href="#">(90 m)</a>	Zuo D, Xu Z, Peng D, et al. Simulating spatiotemporal variability of blue and green water resources availability with uncertainty analysis[J]. Hydrological Processes, 2015, 29(8): 1942-1955.
HunzaRiver <a href="#">(90 m)</a>	Biber K, Khan S D, Shah M T. The source and fate of sediment and mercury in Hunza River basin, Northern Areas, Pakistan[J]. Hydrological Processes, 2015, 29(4): 579-587.
Kasilian <a href="#">(50 m)</a>	Saghafian B, Meghdadi A R, Sima S. Application of the WEPP model to determine sources of run-off and sediment in a forested watershed[J]. Hydrological Processes, 2015, 29(4): 481-497.
Lonquen <a href="#">(90 m)</a>	Stewart R D, Abou Najm M R, Rupp D E, et al. Hillslope run-off thresholds with shrink–swell clay soils[J]. Hydrological Processes, 2015, 29(4): 557-571.
MicaCreek1 <a href="#">(30 m)</a>	Du E, Link T E, Gravelle J A, et al. Validation and sensitivity test of the distributed hydrology soil-vegetation model (DHSVM) in a forested mountain watershed[J]. Hydrological Processes, 2014, 28(26): 6196-
MicaCreek2 <a href="#">(30 m)</a>	

	6210.
NarayaniRiver <a href="#">(30 m)</a>	Neupane R P, Yao J, White J D. Estimating the effects of climate change on the intensification of monsoonal-driven stream discharge in a Himalayan watershed[J]. Hydrological Processes, 2014, 28(26): 6236-6250.
WillowRiver <a href="#">(90 m)</a>	Zhang M, Wei X. Contrasted hydrological responses to forest harvesting in two large neighbouring watersheds in snow hydrology dominant environment: implications for forest management and future forest hydrology studies[J]. Hydrological Processes, 2014, 28(26): 6183-6195.
Bowron <a href="#">(90 m)</a>	
UpperDalya <a href="#">(90 m)</a>	Peleg N, Shamir E, Georgakakos K P, et al. A framework for assessing hydrological regime sensitivity to climate change in a convective rainfall environment: a case study of two medium-sized eastern Mediterranean catchments, Israel[J]. Hydrology and Earth System Sciences, 2015, 19(1): 567-581.
UpperTananim <a href="#">(90 m)</a>	
SanFrancisco <a href="#">(90 m)</a>	Timbe E, Windhorst D, Crespo P, et al. Understanding uncertainties when inferring mean transit times of water trough tracer-based lumped-parameter models in Andean tropical montane cloud forest catchments[J]. Hydrology and Earth System Sciences, 2014, 18: 1503-1523.
HuaiRiver <a href="#">(90 m)</a>	Chen X, Hao Z, Devineni N, et al. Climate information based streamflow and rainfall forecasts for Huai River basin using hierarchical Bayesian modeling[J]. Hydrology and Earth System Sciences, 2014, 18(4): 1539-1548.
WarregoSC2 <a href="#">(90 m)</a>	Alvarez-Garreton C, Ryu D, Western A W, et al. Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes[J]. Hydrology and Earth System Sciences, 2015, 19(4): 1659-1676.
WarregoSC3 <a href="#">(90 m)</a>	
WarregoSC4 <a href="#">(90 m)</a>	
Ishikari <a href="#">(50 m)</a>	Duan W L, He B, Takara K, et al. Modeling suspended sediment sources and transport in the Ishikari River basin, Japan, using SPARROW[J]. Hydrology and Earth System Sciences, 2015, 19(3): 1293-1306.
Limari <a href="#">(90 m)</a>	Scott C A, Vicuña S, Blanco-Gutiérrez I, et al. Irrigation efficiency and water-policy implications for river basin resilience[J]. Hydrology and Earth System Sciences, 2014, 18(4): 1339-1348.

Limpopo <a href="#">(90 m)</a>	Trambauer P, Werner M, Winsemius H C, et al. Hydrological drought forecasting and skill assessment for the Limpopo River basin, southern Africa[J]. Hydrology and Earth System Sciences, 2015, 19(4): 1695-1711.
Crocodile <a href="#">(90 m)</a>	Saraiva Okello A M L, Masih I, Uhlenbrook S, et al. Drivers of spatial and temporal variability of streamflow in the Incomati River basin[J]. Hydrology and Earth System Sciences, 2015, 19(2): 657-673.
Komati <a href="#">(90 m)</a>	
Haean <a href="#">(30 m)</a>	Shope C L, Maharjan G R, Tenhunen J, et al. Using the SWAT model to improve process descriptions and define hydrologic partitioning in South Korea[J]. Hydrology and Earth System Sciences, 2014, 18(2): 539-557.
Durance <a href="#">(90 m)</a>	Kuentz A, Mathevet T, Gailhard J, et al. Building long-term and high spatio-temporal resolution precipitation and air temperature reanalyses by mixing local observations and global atmospheric reanalyses: the ANATEM method[J]. Hydrology and Earth System Sciences, 2015, 19: 2717–2736.
Kabul <a href="#">(90 m)</a>	Wi S, Yang Y C E, Steinschneider S, et al. Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change[J]. Hydrology and Earth System Sciences, 2015, 19(2): 857-876.
Garonne <a href="#">(90 m)</a>	Habets F, Philippe E, Martin E, et al. Small farm dams: impact on river flows and sustainability in a context of climate change[J]. Hydrology and Earth System Sciences, 2014, 18(10): 4207–4222.
Rhone <a href="#">(90 m)</a>	
Ebro <a href="#">(25 m)</a>	Peñas F J, Barquín J, Snelder T H, et al. The influence of methodological procedures on hydrological classification performance[J]. Hydrology and Earth System Sciences, 2014, 18(9): 3393-3409.
Olifants <a href="#">(30 m)</a>	Dabrowski J M. Applying SWAT to predict orthophosphate loads and trophic status in four reservoirs in the upper Olifants catchment, South Africa[J]. Hydrology and Earth System Sciences, 2014, 18: 2629–2643.
WeiRiver <a href="#">(90 m)</a>	Zhan C S, Jiang S S, Sun F B, et al. Quantitative contribution of climate change and human activities to runoff changes in the Wei River basin, China[J]. Hydrology and Earth System Sciences, 2014, 18(8): 3069-3077.

Bellever <a href="#">(90 m)</a>	Liu J, Han D. On selection of the optimal data time interval for real-time hydrological forecasting[J]. Hydrology and Earth System Sciences, 2013, 17(9): 3639-3659.
Brue <a href="#">(90 m)</a>	
Bishop_Hull <a href="#">(90 m)</a>	
Pomahaka <a href="#">(90 m)</a>	McMillan H K, Hreinsson E Ö, Clark M P, et al. Operational hydrological data assimilation with the recursive ensemble Kalman filter[J]. Hydrology and Earth System Sciences, 2013, 17(1): 21-38.
ColoradoR_Cameron <a href="#">(90 m)</a>	Rosenberg E A, Clark E A, Steinemann A C, et al. On the contribution of groundwater storage to interannual streamflow anomalies in the Colorado River basin[J]. Hydrology and Earth System Sciences, 2013, 17(4): 1475-1491.
SanJuanR_Bluff <a href="#">(90 m)</a>	
DoloresR_Cisco <a href="#">(90 m)</a>	
RioSanFrancisco <a href="#">(90 m)</a>	Windhorst D, Waltz T, Timbe E, et al. Impact of elevation and weather patterns on the isotopic composition of precipitation in a tropical montane rainforest[J]. Hydrology and Earth System Sciences, 2013, 17(1): 409-419.
RioSanFrancisco1 <a href="#">(90 m)</a>	
Rhine <a href="#">(90 m)</a>	Vorogushyn S, Merz B. Flood trends along the Rhine: the role of river training[J]. Hydrology and Earth System Sciences, 2013, 17(10): 3871-3884.
Urola <a href="#">(90 m)</a>	Cowpertwait P, Ocio D, Collazos G, et al. Regionalised spatiotemporal rainfall and temperature models for flood studies in the Basque Country, Spain[J]. Hydrology and Earth System Sciences, 2013, 17: 479-494.
KrishnaRiver <a href="#">(10 m)</a>	Surinaidu L, Bacon C G D, Pavelic P. Agricultural groundwater management in the Upper Bhima Basin, India: current status and future scenarios[J]. Hydrology and Earth System Sciences, 2013, 17(2): 507-517.
ClearCreek <a href="#">(90 m)</a>	Zhang H L, Wang Y J, Wang Y Q, et al. The effect of watershed scale on HEC-HMS calibrated parameters: a case study in the Clear Creek watershed in Iowa, US[J]. Hydrology and Earth System Sciences, 2013, 17(7): 2735-2745.
Baba <a href="#">(90 m)</a>	Arias-Hidalgo M, Bhattacharya B, Mynett A E, et al. Experiences in using the TMPA-3B42R satellite data to complement rain gauge
Toachi <a href="#">(90 m)</a>	

SanPabloLaMana <a href="#">(90 m)</a>	measurements in the Ecuadorian coastal foothills[J]. Hydrology and Earth System Sciences, 2013, 17(7): 2905
Monastir <a href="#">(25 m)</a>	Mascaro G, Piras M, Deidda R, et al. Distributed hydrologic modeling of a sparsely monitored basin in Sardinia, Italy, through hydrometeorological downscaling[J]. Hydrology and Earth System Sciences, 2013, 17(10): 4143-4158.
Gard <a href="#">(90 m)</a>	Braud I, Ayral P A, Bouvier C, et al. Multi-scale hydrometeorological observation and modelling for flash-flood understanding[J]. Hydrology and Earth System Sciences, 2014, 18(9): 3733-3761.
Zhanghe <a href="#">(9 m)</a>	Xie X, Meng S, Liang S, et al. Improving streamflow predictions at ungauged locations with real-time updating: application of an EnKF-based state-parameter estimation strategy[J]. Hydrology and Earth System Sciences, 2014, 18(10): 3923
Davidson <a href="#">(25 m)</a>	Yang J, Castelli F, Chen Y. Multiobjective sensitivity analysis and optimization of distributed hydrologic model MOBIDIC[J]. Hydrology and Earth System Sciences, 2014, 18(10): 4101-4112.
Lienz <a href="#">(25 m)</a>	He Z H, Parajka J, Tian F Q, et al. Estimating degree-day factors from MODIS for snowmelt runoff modeling[J]. Hydrology and Earth System Sciences, 2014, 18(12): 4773-4789.
Cheakamus <a href="#">(30 m)</a>	Bourdin D R, Nipen T N, Stull R B. Reliable probabilistic forecasts from an ensemble reservoir inflow forecasting system[J]. Water Resources Research, 2014, 50(4): 3108-3130.
YbbsRiver <a href="#">(10 m)</a>	Ceola S, Bertuzzo E, Singer G, et al. Hydrologic controls on basin-scale distribution of benthic invertebrates[J]. Water Resources Research, 2014, 50(4): 2903-2920.
Susquehanna <a href="#">(90 m)</a>	Giuliani M, Herman J D, Castelletti A, et al. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management[J]. Water Resources Research, 2014, 50(4): 3355-3377.
NorsmindeFjord <a href="#">(20 m)</a>	He X, Koch J, Sonnenborg T O, et al. Transition probability-based stochastic geological modeling using airborne geophysical data and borehole data[J]. Water Resources Research, 2014, 50(4): 3147-3169.

SouthPark <a href="#">(30 m)</a>	Ball L B, Caine J S, Ge S. Controls on groundwater flow in a semiarid folded and faulted intermountain basin[J]. Water Resources Research, 2014, 50(8): 6788-6809.
KernRiver <a href="#">(90 m)</a>	Giroto M, Cortés G, Margulis S A, et al. Examining spatial and temporal variability in snow water equivalent using a 27 year reanalysis: Kern River watershed, Sierra Nevada[J]. Water Resources Research, 2014, 50(8): 6713-6734
UpperRhône <a href="#">(90 m)</a>	Bordoy R, Burlando P. Stochastic downscaling of climate model precipitation outputs in orographically complex regions: 2. Downscaling methodology[J]. Water Resources Research, 2014, 50(1): 562-579.
Pettit <a href="#">(10 m)</a>	Mallard J, McGlynn B, Covino T. Lateral inflows, stream-groundwater exchange, and network geometry influence stream water composition[J]. Water Resources Research, 2014, 50(6): 4603-4623.
Stanley <a href="#">(10 m)</a>	
Alturas <a href="#">(10 m)</a>	
Burdekin <a href="#">(90 m)</a>	Bainbridge Z T, Lewis S E, Smithers S G, et al. Fine-suspended sediment and water budgets for a large, seasonally dry tropical catchment: Burdekin River catchment, Queensland, Australia[J]. Water Resources Research, 2014, 50(11): 9067-9087.
Blackwater <a href="#">(90 m)</a>	Cooper R J, Krueger T, Hiscock K M, et al. Sensitivity of fluvial sediment source apportionment to mixing model assumptions: A Bayesian model comparison[J]. Water Resources Research, 2014, 50(11): 9031-9047.
OitaRiver <a href="#">(90 m)</a>	Higashino M, Stefan H G. Modeling the effect of rainfall intensity on soil-water nutrient exchange in flooded rice paddies and implications for nitrate fertilizer runoff to the Oita River in Japan[J]. Water Resources Research, 2014, 50(11): 8611-8624.
Zwalm <a href="#">(30 m)</a>	Guingla P, Douglas A, Keyser R, et al. Improving particle filters in rainfall-runoff models: Application of the resample-move step and the ensemble Gaussian particle filter[J]. Water Resources Research, 2013, 49(7): 4005-4021.
XianNanGou <a href="#">(5 m)</a>	Ichoku C, Karnieli A, Verchovsky I. Application of fractal techniques to the comparative evaluation of two methods of extracting channel networks from digital elevation models[J]. Water Resources Research,

	1996, 32(2): 389-399.
Hodder <a href="#">(200 m)</a>	Bulygina N, Ballard C, McIntyre N, et al. Integrating different types of information into hydrological model parameter estimation: Application to ungauged catchments and land use scenario analysis[J]. Water Resources Research, 2012, 48(6), W06519.
NorthEsk <a href="#">(90 m)</a>	Capell R, Tetzlaff D, Soulsby C. Can time domain and source area tracers reduce uncertainty in rainfall-runoff models in larger heterogeneous catchments?[J]. Water Resources Research, 2012, 48(9), W09544.
SouthForkNew <a href="#">(30 m)</a>	Gu C, Anderson W, Maggi F. Riparian biogeochemical hot moments induced by stream fluctuations[J]. Water Resources Research, 2012, 48(9), W09546.
LiWuRiver <a href="#">(30 m)</a>	Huang Jr C, Yu C K, Lee J Y, et al. Linking typhoon tracks and spatial rainfall patterns for improving flood lead time predictions over a mesoscale mountainous watershed[J]. Water Resources Research, 2012, 48(9), W09540.
AlzetteEttel <a href="#">(30 m)</a>	Krier R, Matgen P, Goergen K, et al. Inferring catchment precipitation by doing hydrology backward: A test in 24 small and mesoscale catchments in Luxembourg[J]. Water Resources Research, 2012, 48(10), W10525.
MessPontpierre <a href="#">(30 m)</a>	
Colpach <a href="#">(30 m)</a>	
RoudbachPlaten <a href="#">(30 m)</a>	
Burdekin <a href="#">(90 m)</a>	Kuhnert P M, Henderson B L, Lewis S E, et al. Quantifying total suspended sediment export from the Burdekin River catchment using the loads regression estimator tool[J]. Water Resources Research, 2012, 48(4), W04533.
Cajon <a href="#">(30 m)</a>	Mendoza P A, McPhee J, Vargas X. Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment[J]. Water Resources Research, 2012, 48(9), W09532.
Tenderfoot <a href="#">(100 m)</a>	Payn R A, Gooseff M N, McGlynn B L, et al. Exploring changes in the spatial distribution of stream baseflow generation during a seasonal recession[J]. Water Resources Research, 2012, 48(4), W04519.
Wattenbach <a href="#">(200 m)</a>	Rogger M, Pirkl H, Viglione A, et al. Step changes in the flood

Weerbach <a href="#">(200 m)</a>	frequency curve: Process controls[J]. Water Resources Research, 2012, 48(5), W05544.
UpperRhône <a href="#">(90 m)</a>	Leite Ribeiro M, Blanckaert K, Roy A G, et al. Hydromorphological implications of local tributary widening for river rehabilitation[J]. Water Resources Research, 2012, 48(10), W10528.
WhiteRiver <a href="#">(30 m)</a>	Steinschneider S, Polebitski A, Brown C, et al. Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change[J]. Water Resources Research, 2012, 48(11), W11525.
AmericanRiver <a href="#">(30 m)</a>	Woldemichael A T, Hossain F, Pielke R, et al. Understanding the impact of dam-triggered land use/land cover change on the modification of extreme precipitation[J]. Water Resources Research, 2012, 48(9), W09547.
MahanadiRiver <a href="#">(90 m)</a>	Kannan S, Ghosh S. A nonparametric kernel regression model for downscaling multisite daily precipitation in the Mahanadi basin[J]. Water Resources Research, 2013, 49(3): 1360-1385.
Nujiang <a href="#">(90 m)</a>	Kibler K M, Tullos D D. Cumulative biophysical impact of small and large hydropower development in Nu River, China[J]. Water Resources Research, 2013, 49(6): 3104-3118.
LuckyHills <a href="#">(30 m)</a>	Sivandran G, Bras R L. Dynamic root distributions in ecohydrological modeling: A case study at Walnut Gulch Experimental Watershed[J]. Water Resources Research, 2013, 49(6): 3292-3305.
Sacramento <a href="#">(30 m)</a>	Ficklin D L, Stewart I T, Maurer E P. Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada in California[J]. Water Resources Research, 2013, 49(5): 2765-2782.
Feather <a href="#">(30 m)</a>	
ClintonRiver <a href="#">(30 m)</a>	Shen C, Niu J, Phanikumar M S. Evaluating controls on coupled hydrologic and vegetation dynamics in a humid continental climate watershed using a subsurface-land surface processes model[J]. Water Resources Research, 2013, 49(5): 2552-2572.
HJA <a href="#">(30 m)</a>	Garcia E S, Tague C L, Choate J S. Influence of spatial temperature estimation method in ecohydrologic modeling in the Western Oregon Cascades[J]. Water Resources Research, 2013, 49(3): 1611-1624.

UpperGuadiana <a href="#">(90 m)</a>	Loon A F, Lanen H A J. Making the distinction between water scarcity and drought using an observation-modeling framework[J]. Water Resources Research, 2013, 49(3): 1483-1502.
HaiRiver <a href="#">(1000 m)</a>	Jia Y, Ding X, Wang H, et al. Attribution of water resources evolution in the highly water-stressed Hai River Basin of China[J]. Water Resources Research, 2012, 48(2), W02513.
Cordevole <a href="#">(30 m)</a>	Rigon E, Comiti F, Lenzi M A. Large wood storage in streams of the Eastern Italian Alps and the relevance of hillslope processes[J]. Water Resources Research, 2012, 48(1), W01518.
SalmonRiver <a href="#">(90 m)</a>	Yearsley J. A grid-based approach for simulating stream temperature[J]. Water Resources Research, 2012, 48(3), W03506.
CedoCaka <a href="#">(90 m)</a>	Zhang G, Xie H, Yao T, et al. Snow cover dynamics of four lake basins over Tibetan Plateau using time series MODIS data (2001–2010)[J]. Water Resources Research, 2012, 48(10), W10529.
YamzhogYumCo <a href="#">(90 m)</a>	

1  
2