Nat Hazards
DOI 10.1007/s11069-017-2986-0

ORIGINAL PAPER

An approach to quality validation of large-scale data from the Chinese Flash Flood Survey and Evaluation (CFFSE)

Ximin Yuan[1] · Yesen Liu[1] · Yaohuan Huang[2] · Fuchang Tian[1]

Received: 14 February 2017 / Accepted: 7 July 2017
© Springer Science+Business Media B.V. 2017

Abstract Quality control of large-scale flash flood survey and evaluation data is vital and refers to various social and natural factors. In this study, we present a quality validation approach that uses a data model, Anselin Local Moran's I (DM-Moran), which is based on a model of the flash flood data and a spatial data mining algorithm. The approach of the DM-Moran model involves examining logical relationships and detecting anomalous survey units, which effectively integrates the advantages of certainty rules and checking for reasonableness. It resolves the inconsistencies in massive amounts of flash flood survey data that result from inconsistencies. We used the DM-Moran model to validate the quality of the data of the Chinese Flash Flood Survey and Evaluation (CFFSE) project. The kappa coefficients of the two steps of this approach were 0.95 and 0.99, which meet the requirements of the CFFSE project. We consider the DM-Moran model an effective approach to checking the quality of various other large-scale disaster datasets.

Keywords Flash flood · Survey and evaluation · Data quality validation · DM-Moran model · Spatial data mining

✉ Yesen Liu
ysliu@lreis.ac.cn

Ximin Yuan
yxm@tju.edu.cn

Yaohuan Huang
huangyh@igsnrr.ac.cn

Fuchang Tian
tianfuchang@tju.edu.cn

[1] State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China

[2] Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Published online: 13 July 2017

1 Introduction

Flash floods are among the main disasters in China (Sun et al. 2012). Statistics show that the number of deaths caused by flash floods accounted for two-third of the total number of deaths caused by flood disasters in China each year before the 1990s (Sun et al. 2012). Since 2000, the percentage of flood deaths caused by flash floods has risen to approximately 80% (Sun et al. 2012). In November 2010, the Ministry of Water Resources (MWR) of the People's Republic of China and the Ministry of Finance (MF) of the People's Republic of China started to implement the National County-Level Non-Engineering Measures for Preventing and Controlling Flash floods project, which has significantly reduced casualties and property losses (MWR and MF 2013). However, due to limits imposed by investments and technological methods, most of this foundational work did not meet the requirements for disaster prevention. The threshold values of the warning indices were determined based only on empirical and fragmentary research results, whose reasonableness needs further verification. Therefore, in 2013, the MWR and the MF started the Chinese Flash flood Survey and Evaluation (CFFSE) project to determine the disasters' distribution, magnitude, main causes and warning indices with their threshold values (IWHR 2014a, b). The CFFSE dataset is huge. It is estimated that the total data storage exceeds 100 TB and that there are over 100 million individual records. Because of the numerous survey differences, data errors are frequent. Therefore, an effective data quality validation approach is significant for flash flood prevention.

Currently, a number of methods use numerical validation of field attributes, such as numerical and temporal accuracy, for checking data quality (Chang et al. 2005; Batini et al. 2009; Denev et al. 2011; Ringler et al. 2015). With the assistance of GIS, topological relationships are used for checking the spatial accuracy of spatial data (Piprani and Ernst 2008; Parmar and Goyal 2012; Fan et al. 2014). Furthermore, a data model is established to integrate numerical validation and topological validation (Cai et al. 2015) by applying expert knowledge to quality checking. Recent efforts to validate data quality have used a data warehouse to verify the quality of massive multidimensional data (Berrahou et al. 2015; Huang et al. 2015). The drawback of these methods is that they do not consider the possibility of unknown abnormal distribution patterns in the data, particularly abnormal distribution patterns in large-scale survey data with unknown results due to the inconsistent survey methods and specifications used by the survey personnel. Therefore, existing data validation methods are not sufficient for achieving the goal of validating the quality of large-scale flash flood survey and evaluation data.

In this study, an approach for validating the quality of the CFFSE dataset based on Anselin Local Moran's 1 (DM-Moran) is proposed to address the problem of validating the quality of massive survey and evaluation data. Section 2 describes the study data and the DM-Moran data quality validation method, including the examination of logical relationships and the detection of anomalous survey units. Section 3 verifies the accuracy of the DM-Moran approach by applying it to the CFFSE project on the national scale. Section 4 summarizes the advantages and disadvantages of the DM-Moran model, and the reasons for the error distribution are also discussed.

2 Methodology

2.1 Data

The CFFSE dataset includes 40 types of tabular data, 20 types of spatial data, 14 types of multimedia data and 5 types of text data. The data from Chinese mountain villages include social and economic data (the population, the area, the number of houses, the locations of danger zones, etc.) from statistical data from 2014, a historical flash flood dataset for the period from 1949 to 2015, a hydrological and meteorological dataset (which includes historical maximum flood levels, water levels and precipitation since the 1950s), a physical geography dataset (which includes information on elevation, watershed, rivers networks and river sections) and an analysis and evaluation dataset (which includes flood warning indices and a table relating the water level to the population). The scale of the spatial data is 1:50,000. The dataset of the CFFSE project is shown in Table 1. (Data marked with * are used in this study.)

2.2 Study area

Figure 1 shows the CFFSE project area, which covers the mountainous regions (slope $> 2°$) of China. The mountainous regions of China cover an area of approximately 7,000,000 km^2, which account for approximately 70% if land area and one-third of the population of China. The population and property are distributed in the limited lowlands of these mountainous regions, which span 30 provinces and 2058 counties that are mainly located in the East Asian monsoon climate zone. Consequently, these regions suffer frequent rainstorms and flash floods during the summer, i.e., between May and September. Furthermore, because of the terrain and human activities, the percentages of casualties and economic losses resulting from flash floods are rapidly increasing. Flash floods in the mountainous regions are the main disasters with casualties in China.

2.3 Methods

To validate the quality of the large-scale massive CFFSE dataset, this study establishes a validation method based on DM-Moran by integrating a data model that validates the data based on rules and the DM-Moran outlier detection model. The DM-Moran method guarantees topological accuracy, associated accuracy, numerical accuracy and spatial reasonableness. A flowchart for the DM-Moran method is shown in Fig. 2.

2.3.1 Logic quality check

This study designs a flash flood survey and evaluation data model using an object-oriented method (Zeiler 2000; Maidment 2002; Kumar et al. 2010; Mandel et al. 2015). With natural villages and small watersheds as the core objects, objects in the CFFSE dataset are sorted by their hierarchical and confluent relationships. Social and economic objects are related to administrative divisions based on their hierarchical relationships. Natural objects are related to watersheds based on their confluent relationships. On this basis, the relationships between the entity objects are analyzed. Finally, the data are validated based on the attribute domain and various correlations, as shown in Table 2.

1. Topological relationship validation

The accuracy of the spatial data is verified based on the topological relationships between the spatial objects.

Table 1 CFFSE dataset

Data type	Data list	Data description (data content or area)
Basic data	Small watersheds	Surface slope $\geq 2°$ with areas ranging from 10 to 50 km^2
	Basic geography	1:50,000 digital line graph data (administrative division, place names, river networks, roads)
Survey data	*Basic situation	Tables of the administrative divisions, tables of the prevention and control zones, tables of the danger zones, tables of the enterprises and public institutions in the prevention and control zones
	*Social and economic data	Tables on the social and economic situation, tables on the classification of residents' property, tables on the rural housing situation (typical cases), tables on the types of housing, etc.
	*Historical flash floods	Time, location, number of deaths, number of people evacuated, economic losses
	Mountain flood gullies	Length, prevention and control measures, numbers of people affected, etc.
	*Facilities for monitoring and providing warnings of flash floods	Automatic precipitation gauging stations, simple water level gauging stations, simple precipitation gauging stations, over-the-air warning broadcasting stations, etc.
	*Water-related engineering projects	Bridges, culverts, reservoirs, dikes and sluices
	*Riverside villages, cities and towns	Elevations of the foundations of the houses in the villages, populations of the villages, numbers of houses in the villages, etc.
	*Measurement data	Measurements of the topography of the river channels around the villages and disaster-causing water levels
Hydrological and meteorological data	Historical hydrological data	Flood data extracted from the hydrological stations as well as the relevant precipitation data extracted from the precipitation gauging stations upstream
	*Historical flood data	Collection of typical historical flood data
Analysis and evaluation data	*Flash flood analysis and evaluation list	List of riverside villages
	*Characteristics of storm floods in small watersheds	Every small watershed
	*Current flood prevention capacity	Every riverside village
	*Classes of danger zone	Every riverside village
	*Early warning indices	Every riverside village

See references MWR and MF (2013) and IWHR (2014a) for details of the data

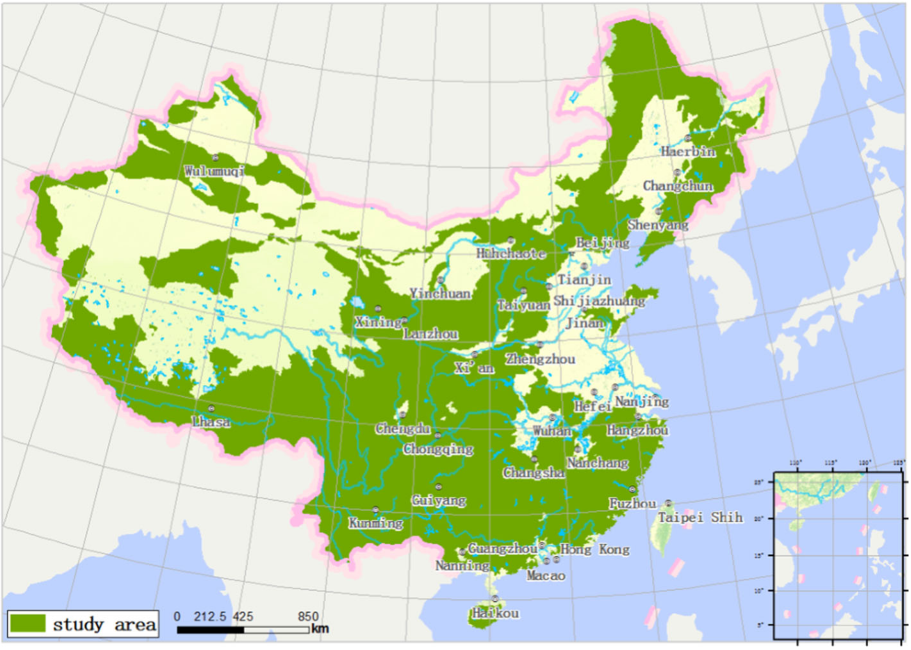

Fig. 1 CFFSE project area

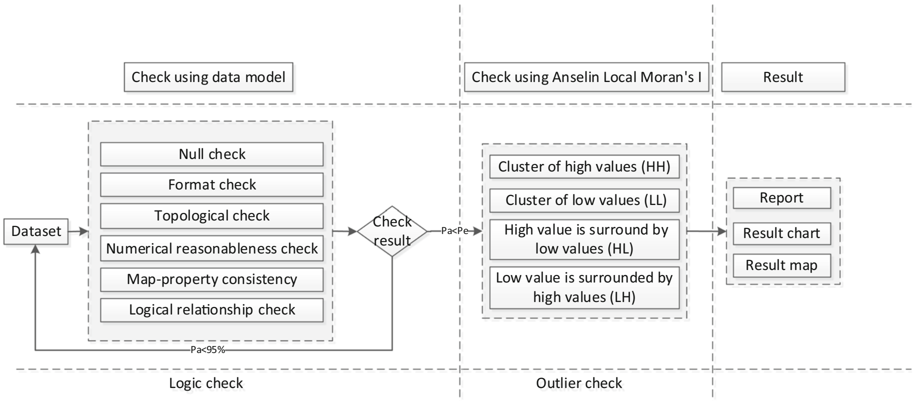

Fig. 2 Flowchart of the validation process of the DM-Moran method

$$\mathrm{TopoC} = \begin{cases} \mathrm{poly}, \mathrm{inc} \,\&\, \mathrm{dsj} \ \,\&\, \mathrm{dep} \\ \mathrm{line}, \mathrm{join} \,\&\, \mathrm{dsj} \,\&\, \mathrm{dep} \\ \mathrm{point}, \mathrm{inc} \,\&\, \mathrm{dep}, \end{cases} \tag{1}$$

where poly, line and point represent the object types in the spatial layers and inc, dsj, dep and join represent checking for inclusion, disjunction, dependency and joining, respectively.

2. Associated relationship validation

Table 2 List of object relations

Relationship type	Relationship content	Description (examples)
Topological	Inclusion relationship (inc), disjoint relationship (dsj), joint relationship (join), dependency relationship (dep), etc.	Resident households are included in residential areas; mountain flood gullies do not cross one another
Associated	Flow (upstream and downstream) relationship (flow), indication relationship (indi), simple association (rela)	Precipitation stations and gauging stations need to be in the same watershed
Numerical	Comparative relationship (comp), summarized relationship (sum), normativity (norm), range (ran) and null	The population of a prevention and control zone is less than the total population; the sum of the populations of the villages and towns equals the population of the county

The logical reasonableness of the data is validated based on its associated relationships.

$$RelaC = \begin{cases} \text{hydro, flow} \\ \text{natu,indi} \\ \text{fac,indi\&rela,} \end{cases} \tag{2}$$

where hydro, natu and fac represent hydrological objects (e.g., watersheds, water systems and stations), natural objects (e.g., villages and sections) and artificial engineering structure objects (e.g., dikes, dams and bridges), respectively, and flow, indi and rela represent flow relations, indications and associations, respectively.

3. Numerical relationship validation

Numerical logic errors in the data are identified based on numerical relationships.

$$\text{ValueC} = \begin{cases} \text{pop\&area, sum \&comp} \\ \text{ele\&len, ran} \\ \text{all, null\&norm,} \end{cases} \tag{3}$$

where pop and area represent the population and area fields, respectively, ele and len represent the elevation and length fields, respectively, all represents all the fields and sum, comp, ran, null and norm represent the validation methods for the fields.

The logic check (LC) includes the three previously mentioned types of check.

$$\text{LC} = \bigcup(\text{TopoC}, \text{RelaC}, \text{ValueC}). \tag{4}$$

The data are validated field by field using software. The data error rate can be calculated using the following equation:

$$Pi = \frac{\sum_{j=1}^{A_i} E_{ij}}{A_i \times \text{n}} \times 100\%, \tag{5}$$

where Pi represents the pass rate of the ith type of table, Ai represents the number of records contained in the ith type of survey table, n represents the number of fields in the ith type of survey table and E_{ij} represents the number of erroneous fields in the jth record.

The overall data error rate can be calculated using the following equation:

$$Pa = \frac{\sum_{i=1}^{z}\sum_{j=1}^{A_i} E_{ij}}{\sum_{i=1}^{z} A_i \times \mathrm{n}} \times 100\%, \tag{6}$$

where Pa represents the overall pass rate of the dataset and z represents the number of survey tables.

2.3.2 Spatial distribution outlier check

After logic check, DM-Moran is used to determine whether there are outliers in the spatial distribution of the data objects. The output of this method consists of the Local Moran's I index, the z-score, the p value and the cluster/outlier type. The z-score and the p-value are mainly used to measure the significance of the statistic based on the calculated similarity (the spatial clustering of high or low values) or dissimilarity (spatial outliers) between the study objects. If the z-score of an element is relatively high and positive, it signifies that the elements surrounding it have similar values (high or low values); if the z-score of an element is relatively low and negative, it signifies that there is a statistically significant outlier in the spatial data (Anselin 1995; Ng and Han 1994; Wang et al. 2010; Wang and Hu 2012).

The Local Moran's I statistic for spatial association is

$$I_i = \frac{x_i - X}{S_i^2} \sum_{j=1, j\neq i}^{n} w_{i,j}\left(x_j - X\right), \tag{7}$$

where x_i represents the density or quantity of survey objects in a flash flood survey unit i, for example, the number of flash floods in each county unit, which sum all the flash floods in the county by overlaying county boundary and flash flood point layer.$\bar{X}$ represents the corresponding average density or quantity, and $w_{i,j}$ represents the spatial weight between survey units i and j. In addition,

$$S_i^2 = \frac{\sum_{j=1, j\neq i}^{n}\left(x_j - X\right)^2}{n-1} - X^2, \tag{8}$$

where n is the total number of units investigated.

The z_{I_i}-score of the statistical data is calculated using the following equation:

$$z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}}, \tag{9}$$

where

$$\mathrm{E}[I_i] = -\frac{\sum_{j=1, j\neq i}^{m} w_{ij}}{n-1} \tag{10}$$

$$[I_i] = E\left[I_i^2\right] - E[I_i]^2. \tag{11}$$

The content of the spatial distribution outlier check includes

$$\mathrm{LC} = \cup(\mathrm{Moran}, p, Z), \tag{12}$$

where Moran represents the Local Moran's I, p represents the p-value and Z represents the z-score.

2.3.3 DM-Moran check

Based on this study's design, the results of the logical check have an error rate that is lower than 5% (or a pass rate that is greater than 95%), which is the requirement for passing the first step of validation. Therefore, the outlier check can be performed to determine whether there are anomalous distribution patterns. A combination of the two validation methods can ensure data quality based on logic and reasonableness.

$$\mathrm{DT\text{ - }Moran} = \cup(\mathrm{LC}, \mathrm{OC}). \tag{13}$$

3 Results and discussion

3.1 Results

The CFFSE dataset was validated using the DM-Moran method. Errors were detected by the logic quality check. Table 3 shows the results of the logical check of some data tables. The overall error rate is 21.19%. Tables T01, T02, T07 and T14 have the highest error rates of 37.06, 38.46, 86.93 and 37.70%, respectively. The main errors in Tables T01 and T02 are related to numerical relationships. Table T07 contains a relatively large number of null values. The main errors in Table T14 are related to topological relationships.

Anomalous survey units were detected by DM-Moran. The results showed three types of anomalous survey unit: HH (a cluster of high values), HL (an outlying high value is surrounded primarily by low values) and LH (an outlying low value is surrounded primarily by high values). Historical flash flood survey data are used as an example to illustrate the outlier check (Fig. 3). In data from 1736 counties, 16 counties include low-value outliers, 13 counties include high-value outliers and 94 counties include high-value clusters. There are two reasons for these anomalous survey units. One reason is the inaccurate coordinate of the disasters, because there was no accurate record when disaster occurs; therefore, only according to the memory of the villagers, the other reason is the repeated records, due to the operational mistakes. We deal with the anomalous units by feeding back to the investigators.

The method can be applied to the quality validation of other large-scale natural hazards investigation, such as the investigation on disaster losses of historic floods, whose accuracy is related to respondents' memory, age and education. For example, the logical relationships of water level mostly should lower than altitude of dike and river mostly should not crossover can be confirmed by logic quality check. Furthermore, the extremely high loss units in the investigating area can be detected by outlier check.

3.2 Accuracy validation

To verify the accuracy of the validation results, sampling field surveys were conducted to revalidate the data found to contain errors as well as the data found to contain no errors during the validation process.

Then, the value of the kappa coefficient ($K_{\mathrm{hat}} = 0.99, 0.96$) in the results of the DM-Moran check was calculated using Tables 4 and 5. Blackman and Landis (Cohen 1968; Landis and Koch 1977; Blackman and Koval 2000) assigned values between 0 and 1 to the kappa coefficient in their analysis of data agreement, which has become the standard reference for assessing data agreement in actual research (Table 6).

Table 3 Validation results

Table ID	Table name	Total fields	TopoC errors	RelaC errors	ValueC errors	Total errors	Error rate (%)
T01	General information about the administrative divisions	27,689	133	0	10,129	10,262	37.06
T02	County-level social and economic information	26	0	0	10	10	38.46
T03	Summary of the classification of residents' family property	120	0	0	0	0	0.00
T04	Summary of the rural housing situation (typical cases)	361	0	0	70	70	19.39
T05	Basic information on the prevention and control zones	12,834	3	358	1730	2091	16.29
T06	Basic information on the danger zones	354	56	1	2	59	16.67
T07	Summary of enterprises and public institutions in the prevention and control zones	9734	1	560	7901	8462	86.93
T08	Summary of historical flash floods	150	0	0	29	29	19.33
T09	Summary of riverside village residents	11,182	88	65	1157	1310	11.72
T10	Summary of automatic monitoring stations	1451	0	0	65	65	4.48
T11	Summary of over-the-air warning broadcasting stations	2926	0	188	549	737	25.19
T12	Summary of simple precipitation gauging stations	3168	0	325	627	952	30.05
T13	Summary of simple water level gauging stations	153	0	24	10	34	22.22
T14	Summary of pond dams (weirs)	382	97	12	35	144	37.70
T15	Summary of culverts	468	23	4	10	37	7.91
T16	Summary of bridges	594	17	4	139	160	26.94
T17	Results of the vertical sections of gullies	259	14	0	0	14	5.41
T18	Measurement points on the vertical sections of gullies	11,362	12	0	0	12	0.11
T19	Measurement points of historical flood marks in gullies	27	0	0	0	0	0.00
T20	Results of the transverse sections of gullies	1324	0	0	68	68	5.14
T21	Measurement points on the transverse sections of gullies	31,165	0	0	3	3	0.01
	Total	115,729	–		–	24,519	21.19

* Only some tables are presented in this table

The kappa coefficients for logic check result and outlier check result are 0.99 and 0.96, respectively, which means the two datasets are in good agreement. The approach we presented provides acceptable data validation results.

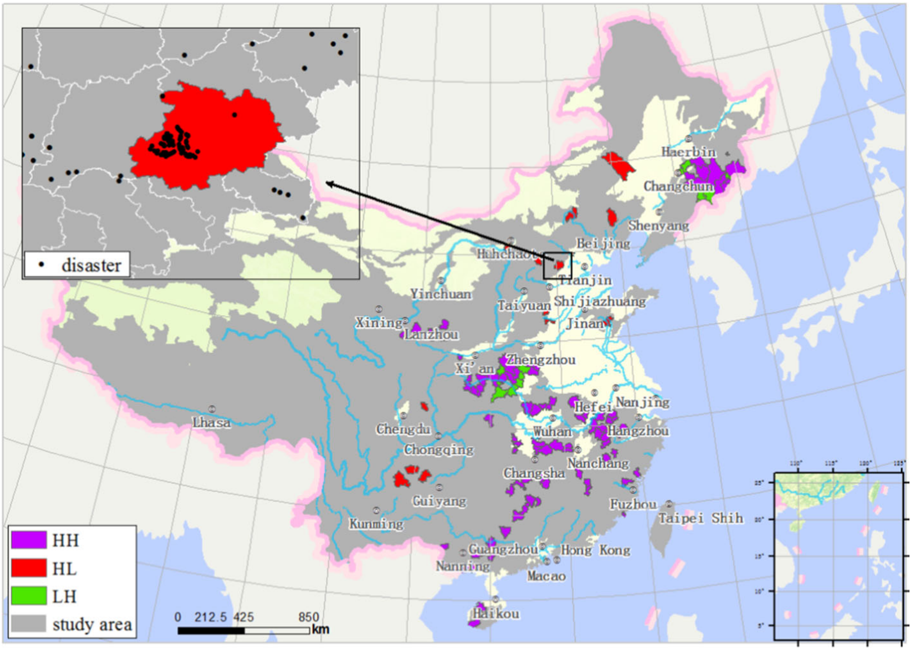

Fig. 3 Outlier analysis results

Table 4 Confusion matrix for the LC and the actual situation

Validation results	Actual situation		
	Errors	No errors	Total
Errors	24,218 (20.93%)	301 (0.26%)	24,519 (21.19%)
No errors	69 (0.06%)	91,141 (78.75%)	91,210 (78.81%)
Total	24,287 (20.99%)	91,442 (79.01%)	115,729 (100.00%)

The CFFSE dataset has the following problems: (1) It contains normativity errors. Data are not recorded in the standard formats required by the technical specifications. (2) Field values exceed the ranges. This is a result of using the wrong method or of data unit errors. (3) The accuracy of spatial layers does not meet the requirements. (4) There are erroneous topological relationships. For example, rivers cross or cover one another. (5) Different survey organizations and inconsistent survey methods and specifications result in abnormal differences in data from different survey units.

Based on these reasons for data errors, the following steps should be implemented during the survey work: (1) The format and logical relationships should be checked using software to discover and correct data errors as early as possible. (2) The frequency of training and communication should be improved during the survey process, and various survey organizations should be encouraged to audit the data of others to ensure consistent survey methods and specifications.

Table 5 Confusion matrix for the outlier check and the actual situation

Validation results	Actual situation		
	Errors	No errors	Total
Errors	114 (6.57%)	9 (0.52%)	123 (7.09%)
No errors	0	1613 (92.91%)	1613 (92.91%)
Total	114 (6.57%)	1622 (93.43%)	1736 (100.00%)

Table 6 Grades based on the kappa coefficient

Kappa	0–0.02	0.02–0.2	0.21–0.40	0.41–0.60	0.61–0.80	0.81–1
Agreement level	Very poor	Poor	Minor	Moderate	Good	Excellent

4 Conclusions

Data quality is one of the main issues of large-scale flash flood survey and evaluation data. The large number of organizations participating in the survey and evaluation process and the large spatial scale create challenges in data quality control and in validating the consistency of data from various survey units. In this study, a DM-Moran-based validation approach is established to perform quality control on the CFFSE dataset. This approach examines the logical relationships among the data at the field level using the ranges and relationship rules of the data model and detects anomalous survey units using the spatial outlier detection method.

It is found that the DM-Moran model can effectively integrate the advantages of control based on certainty rules and outlier detection based on spatial reasonableness and solve the inconsistency problem in massive data resulting from inconsistent survey methods or survey specifications. The validation results obtained using the DM-Moran model are very accurate. The results of the LC have a kappa coefficient of 0.99, and the results of the outlier check have a kappa coefficient of 0.96. The DM-Moran model provides a basis for promptly improving survey methods and performance.

Although the data validation approach used in this study is very effective, there are still some problems. The validation rules used in the LC do not necessarily represent the real situation completely. In addition, the outlier check only considers outliers between survey units based on the administrative divisions. Regions are not divided based on their economic development and natural conditions. Although the DM-Moran model designed in this study is for large-scale flash flood survey and evaluation data, it is also suitable for validating highly specialized survey data that cover a large area and are provided by a large number of organizations.

Acknowledgements This study was supported by National Key R&D Program of China (No. 2017YFC0405601), Fund for Key Research Area Innovation Groups of China Ministry of Science and Technology (No. 2014RA4031), Science Fund for Creative Research Groups of the National Natural Science Foundation of China (No. 51621092), Program of Introducing Talents of Discipline to Universities (No. B14012).

Author contributions The study was designed by Ximin Yuan and Yesen Liu. The data were collected and analyzed by Yesen Liu. The paper was written by Yesen Liu, Yaohuan Huang and Fuchang Tian.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Anselin L (1995) Local indicators of spatial association—LISA. Geogr Anal 27(2):93–115

Batini C, Cappiello C, Francalanci C, Maurino A (2009) Methodologies for data quality assessment and improvement. ACM Comput Surv 41(3):75–79

Berrahou L, Lalande N, Serrano E, Molla G, Berti-équille L (2015) A quality-aware spatial data warehouse for querying hydrogeological data. Comput Geosci 85(85):126–135

Blackman NJ, Koval JJ (2000) Interval estimation for Cohen's Kappa as a measure of agreement. Stat Med 19(5):723–741

Cai Y, Xie WJ, Fu J, Cheng YL, Chen ZD, Cheng JG (2015) Some key technologies of geospatial information system for the China water census. Acta Geodaetica Cartogr Sin 44(5):585–589

Chang XW, Paige CC, Tiberius CCJM (2005) Computation of a test statistic in data quality control. Siam J Sci Comput 26(6):1916–1931

Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 70(4):213–220

Denev D, Mazeika A, Spaniol M, Weikum G (2011) The SHARC framework for data quality in Web archiving. VLDB J 20(2):183–207

Fan H, Zipf A, Fu Q, Neis P (2014) Quality assessment for building footprints data on OpenStreetMap. Int J Geogr Inf Sci 28(4):700–719

Huang YH, Xu C, Yang HJ, Wang JH, Jiang D, Zhao CP (2015) Temporal and spatial variability of droughts in southwest china from 1961 to 2012. Sustainability 7(10):13597–13609

Kumar M, Bhatt G, Duffy C (2010) An object-oriented shared data model for GIS and distributed hydrologic models. Int J Geogr Inf Sci 24(7):1061–1079

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174

Maidment D (2002) Arc hydro: GIS for water resources. ESRI Press, Redlands

Mandel P, Maurel M, Chenu D (2015) Better understanding of water quality evolution in water distribution networks using data clustering. Water Res 87:69–78

Ministry of Water Resources of the People's Republic of China and Ministry of Finance of the People's Republic of China (2013) Plan for the Implementation of the National Mountain Flood. Disaster Prevention and Control Project (2013–2015)

Ng RT, Han J (1994) Efficient and effective clustering methods for spatial data mining. In: International conference on very large data bases. Morgan Kaufmann Publishers Inc, pp 144–155

Office of State Flood Control and Drought Relief Headquarters of China and China Institute of Water Resources and Hydropower Research (2014a) Technical specifications for flash flood survey

Office of State Flood Control and Drought Relief Headquarters of China and China Institute of Water Resources and Hydropower Research (2014b) Technical specifications for flash flood analysis and evaluation

Parmar V, Goyal P (2012) Analysis on axioms of spatial data quality. In: IJCA proceedings on national workshop-cum-conference on recent trends in mathematics and computing (RTMC) 2011, vol RTMC 1

Piprani B, Ernst D (2008) A model for data quality assessment; OTM confederated international workshops and posters on the move to meaningful internet systems: 2008 workshops, pp 750–759

Ringler AT, Hagerty MT, Holland J, Gonzales A, Gee LS, Edwards JD (2015) The data quality analyzer: a quality control program for seismic data. Comput Geosci 76:96–111

Sun DY, Zhang DW, Cheng XT (2012) Framework of national non-structural measures for flash flood disaster prevention in China. Water 4(1):272–282

Wang J, Hu Y (2012) Environmental health risk detection with GeogDetector. Environ Model Softw 33:114–115

Wang JF, Li XH, Christakos G, Liao YL, Zhang T, Gu X (2010) Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. Int J Geogr Inf Sci 24(1):107–127

Zeiler M (2000) Modeling our world. ESRI Press, Redlands