



Morphometric analysis of sinkholes using a semi-automatic approach in Zhijin County, China

Hao Chen^{1,2} · Takashi Oguchi^{3,4} · Pan Wu⁵

Received: 12 April 2018 / Accepted: 19 July 2018 / Published online: 31 July 2018
© The Author(s) 2018

Abstract

Efficient detection and analysis of natural sinkholes is essential for human sustainability in karst area. In this work, two methodologies were applied to detect and quantify natural sinkholes in southwest China, using of DEMs derived from ASTER and SRTM remote sensing images and topographic maps. The methodologies were a semi-automatic approach and a field-based approach. The first identifies the sinkholes based on DEM. The latter integrates the field survey, visual interpretation of high-resolution aerial photographs and topographic maps. The comparison between the semi-automatic approach and field-based approach generated the threshold values of the sinkhole shape, area, and TPI (topographic position index), which were used to distinguish true sinkholes and improve the identification accuracy. The result indicated that the semi-automatic approach produced the best accuracy by using the thresholds area > 60 m², ellipticity > 0.2, and TPI ≤ 0 with the DEM from the topographic maps. With these realistic thresholds, the semi-automatic model shows well performance that model accuracy ranges from 0.78 to 0.95 for DEM resolutions from 3 to 75 m. The morphometric analysis was then obtained by using parameters derived from the semi-automatic approach and field-based approach. The sinkhole morphometric parameters are area, perimeter, diameter, shape, orientation, and volume. The results demonstrate (1) morphometric characteristics of sinkholes derived from the semi-automatic approach are coincident with those from the field-based approach. (2) Sinkholes in the region are relatively old and skewed to irregular or elliptical shape. (3) Sinkhole diameter and spatial alignments are highly controlled by faults. (4) Sinkholes argue about solution-only origin in the region.

Keywords Karst sinkhole · DEM analysis · Remote sensing · GIS

Introduction

Sinkholes, which are also termed as dolines by European geomorphologists (Sauro 2003), are the most common depres-

sions in karst terrain. Sinkholes are identified as direct conduits to underground aquifers, and are regarded as the diagnostic surface feature of karst terrains (Ford and Williams 2013). Sinkholes are extensively distributed in karst areas worldwide and affect engineering structures, agriculture, natural resources and environment, and human sustainability (Gutiérrez et al. 2014; Witze 2014). Sinkhole collapses are one of the most serious geological hazards in the karst environment.

Many field-based studies, such as formation mechanisms (Tharp 1999; Salvati and Sasowsky 2002), terminology and classification (White 1988; Gunn 2004; Waltham et al. 2007; Gutiérrez et al. 2008), interrelations with the anthropogenic environment (Delle Rose et al. 2004; Gutiérrez et al. 2014), and hazard assessment (Van Schoor 2002; Galve et al. 2009; Taheri et al. 2015) were carried out to investigated sinkholes. The morphometric analysis of sinkholes is one of the topics addressed in these scientific literatures (Hyatt and Jacobs 1996; Bruno et al. 2008; Basso et al. 2013). The importance of morphometric analyses in karst had been well expounded

✉ Hao Chen
chenhao@hqu.edu.cn

¹ School of Political Science and Public Administration, Huaqiao University, Quanzhou 362021, China

² Center for Political Development and Public Governance, Huaqiao University, Quanzhou 362021, China

³ Department of Natural Environment Studies, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa 277-8568, Japan

⁴ Center for Spatial Information Science, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa 277-8568, Japan

⁵ College of Resources and Environmental Engineering, Guizhou University, Guiyang 550025, China

by La Valle (1968), Williams (1972) and Drake and Ford (1972) since the last century. Recently, there are some studies of sinkholes and their hazard assessment (Youssef et al. 2012; Elmahdy and Mostafa 2013). These studies showed that high-resolution satellite images or aerial photographs combined with geographic information systems (GIS) facilitated researches in large-scale studies. However, it is still a challenge to delineate sinkholes in finer scales accurately. In the past, it was dominant that sinkhole delineation were based on visual interpretation of remote sensing images and field work with the disadvantages of subjective, time-consuming, minimally reproducible, and unsuitable for studying large areas (Doctor and Young 2013). Compared to these manual work (Gutiérrez et al. 2014), the use of topographic maps, DEMs, LIDAR, InSAR, and aerial/satellite images offer advantages in various situations, such as in inaccessible areas, large coverage, and places covered with water or vegetation. A study in this regard has identified sinkhole formation and risk using radar data by Chang and Hanssen (2014). However, few case studies of sinkhole identification and morphometry with the application of multi-source DEMs have been done.

Doctor and Young (2013) has expounded that a questionable result was led by the manual interpretation methodology based on digital data (including LiDAR DEMs) because of the complicated karst features. Therefore, this paper identifies sinkholes by a semi-automatic model, which based on DEMs from the different data sources. The semi-automatic model is applied by a karst landscape in southwest China at the center of the Southeast Asian karst region. Since unreal sinkholes may be identified from a DEM due to artifacts, especially un-auto-correlated errors from different data sources, the objectives of this study are (1) to investigate the use of various DEMs derived from remote sensing data (ASTER and SRTM) and topographic maps to appropriately detect and quantify the natural sinkholes; (2) to conduct morphometric analysis of sinkholes using the semi-automatic approach with the DEM of best performance and compare the result with that from the traditional field work and visual interpretation of aerial photos.

Study area and data

Study area

This study area is based on one of the most famous karst caves in Zhijin County of Guizhou Province, China. The area is 171.47 km² and over 60% carbonate geology with limestones and dolostones of the Triassic, Permian, Carboniferous, and Cambrian. The non-karst area is composed of Permian basalt, shale silicalite, and mudstone. The study area has a well-developed faulted structure with a NE-SW direction and an elevation of 917–1693 m. It has a subtropical humid monsoon

climate that has 1172 h of mean annual sunshine time and 14.1 °C of mean annual temperature and 1436 mm of mean annual precipitation.

Data

The semi-automatic sinkhole identification is supported by several datasets. The first dataset includes 2.5-m resolution aerial photos captured during the winter of 2005, 1:200,000 hydrogeological maps, and 1:10,000 topographic maps. The aerial photos generated a 1:10,000 land use map. The topographic maps produced a 1-m DEM, which was resampled to DEMs of 3, 5, 10, 25, 30, 50, 60, 75, and 90 m grid sizes. The second dataset includes a 30-m spatial resolution ASTER-DEM (<http://gdem.ersdac.jspacesystems.or.jp/>) and 30- and 90-m spatial resolution SRTM-DEMs (<http://srtm.csi.cgiar.org>). Three-, 60-, and 90-m ASTER-DEMs were resampled from the original 30-m ASTER-DEM. Three and 60-m SRTM-DEMs were resampled from the original 30-m SRTM-DEM. The third dataset was generated from extensive fieldwork in July 2010. The field work data were used to replenish the unidentified features, sinkholes recognized on the aerial photographs and to verify the land use classifications.

Methodology

For hydrological terrain analysis, a sink-free DEM could be obtained from the process and recondition of a raw elevation data (Anderson 1988; Jenson and Domingue 1988; Grimaldi et al. 2007). A classic hydrological correction uses the surrounding pixels elevation to fill up closed depressions, which is difficult to keep the true sinkholes during the procedure. Thus, we propose a semi-automatic approach based on algorithms of Jenson and Domingue (1988) and Maidment (2002). This approach is preserving true sinkholes, while removing artificial ones with fieldwork and morphometric analysis using GIS. In order to discuss the suitability of DEMs, we compared the ASTER GDEM, SRTM-DEM, and DEMs from the topographic maps (hereafter, referred to as topographic DEMs) in this paper. The approach could be concluded as (1) sinkhole identification from the DEMs; (2) removing water area and obvious false sinkhole; (3) exclusion of false sinkholes by general threshold values of sinkhole area, ellipticity (E), and the topographic position index (TPI); (4) generation of true sinkhole map by using aerial imagery and field work; (5) adaptation of the best DEM with the highest accuracy.

The procedure of step 3 is described here. The true sinkhole map shows that the sinkhole area ranges from 0.1 to 60 m². Thus, the value of 60 m² is set as the general area threshold. The sharp eccentricity of a sinkhole (E) is calculated as follows:

$$\sqrt{1 - \frac{b^2}{a^2}} \quad (1)$$

where a and b are the one half of the major and minor axis lengths of a sinkhole, respectively. The E value ranges from 0 (perfect circle) to 1. Field observation in the Zhijin karst area indicates that the sinkholes there tend to be elliptical and elongated, and that E value of 0.2 worked as a good threshold for determining true sinkholes.

TPI introduced by Weiss (2001) depicts the difference between elevation at the central point (Z_c) and the average elevation at the surrounding points within a certain radius (r).

$$TPI = z_c - \frac{1}{n_r} \sum_{i \in r} z_i \quad (2)$$

where n_r is the number of the raster cells of the predetermined area and i stands for the i th cell. Due to its ability of dividing morphological classes, $TPI < 0$ representing negative topographic position is chosen to identify true sinkholes.

Furthermore, sinkholes detected within 90 m from a major river centerline are classified as artificial sinkholes. Sinkholes in karst landscape were also identified based on their underlying geology. The sinkholes recognized outside the limestone areas were assigned as artificial and removed from the dataset.

Based on the result of the field-based and automatic approaches, a large number of sinkholes were found in the study area. The accuracy of the DEM was assessed by comparing the sinkholes extracted from the remote sensing images and field work with those automatically identified from the DEMs using the spatial join function of ArcMap. For a quantitative comparison, the numbers of three classes of sinkholes were counted: (1) the number of linked records of the reference map (identified true sinkholes, true positive—TP); (2) the number of non-linked records of the reference map (non-identified true sinkholes, false negative—FN); (3) the number of non-linked records of auto-classified map (artificial sinkholes, false positive—FP). Note that another commonly used class, true negative, was found to be almost null in this case. Accordingly, the accuracy statistic was calculated as follows:

$$\text{Accuracy} = \frac{TP}{TP + FN + FP} * 100\% \quad (3)$$

Results

The true sinkhole map digitalized by using aerial imagery and field inventory shows that there were 531 sinkholes in the study area (Fig. 1). The comparison between the investigated sinkholes and those identified through the semi-automatic method enables evaluation of the different data sources for

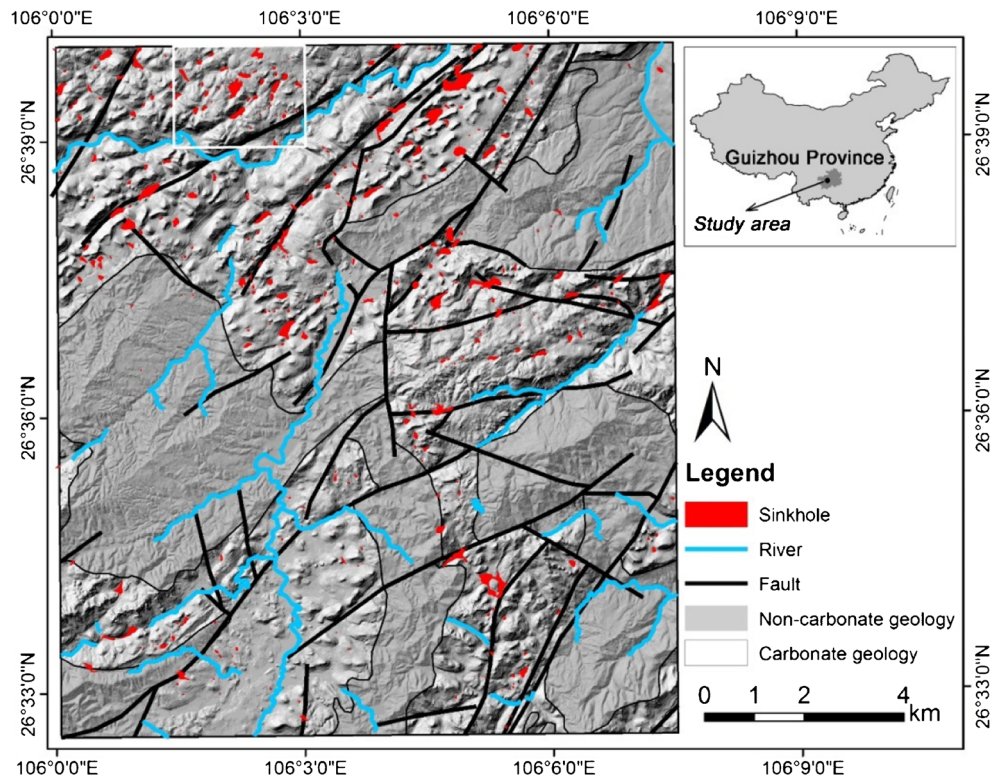
delineating karst sinkholes and sinkhole morphometric analysis in similar regions. The resultant difference between the two approaches reveals the uncertainties of the semi-automatic method.

Sinkhole identification from different DEM data by using the semi-automatic approach

As noted, the thresholds for the semi-automatic model are area = 60 m², $E = 0.2$, and $TPI = 0$. In order to assess the model, the automatically delineated sinkholes were compared with those identified by the field-based model with the given thresholds. Figure 2 shows the evaluated accuracy of the model, and Table 1 shows the detailed performance of the model with the 3-, 30-, 60-, 90-m DEMs from the different data sources. The performance of the semi-automatic approach differs from each other with different data sources. The model accuracy with the SRTM-DEM and GDEM increases with the grid size (Table 1, Fig. 2). The model accuracy with the SRTM-DEM and GDEM ranges from 7 to 55% and from 4 to 30%, respectively. As we can see from Fig. 2, the model performance with the SRTM-DEM is better than that of the GDEM. The performance of the semi-automatic approach with the topographic DEMs can be divided into two sections: grid sizes of 3–10 m and secondly 10–90 m. At the first section, the accuracy of the model with the thresholds decreases from 95 to 80%. At the second section, the model accuracy remains within a range of 75–80% with grid size of 10–75 m and significantly increased to 99% at the grid size of 90 m. However, since the true positive value is low (Table 1), the result of 90-m grid size was not considered as a proper grid size for further morphometric analysis.

In general, if more sinkholes are delineated by the semi-automatic approach, the chance of identifying true sinkholes becomes higher (high TP value). However, the chance of detecting artificial sinkholes also increases (high FN value). In other words, there is a positive correlation between TP and FP values (Table 1). For example, for the SRTM-DEM and GDEM, the TP value decreases corresponding to the decrease in the FP value with the coarsening of the grid size. Especially for the GDEM, the number of artificial sinkholes (FN) remains large. At the grid size of 3 m, the number of identified true sinkholes is smaller than that from any other resolutions. In this paper, the 90-m SRTM data was considered as a pragmatic implementation of large sinkhole identification with the proposed thresholds due to their characteristic of fewer artifacts and an acceptable amount of identified true sinkholes. However, the ASTER GDEM is susceptible to noise, leading to a significant number of visual artifacts with small areas that do not correspond to the field inventory. This anomaly of the ASTER GDEM makes it unsuitable for the detection of sinkholes in the study area.

Fig. 1 Maps of the study site in southwest China, geology, and hillshaded DEM showing the terrain. The sinkholes are identified by aerial photos and the field inventory. The area in a white box is shown in Fig. 3, Fig. 8, and Fig. 10



Spatial distribution of sinkholes

Sinkhole density is the number of sinkholes per square kilometer in karst area and describes the spatial distribution of

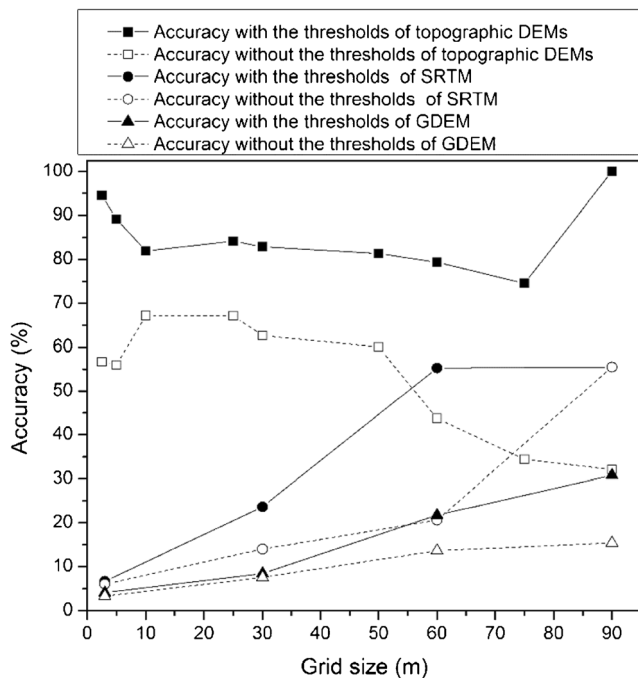


Fig. 2 Accuracy of sinkhole identification for the semi-automatic approach with/without the thresholds of area = 60 m², ellipticity = 0.2 and TPI = 0

sinkholes. Related to that the area of karst in the study area is 104.87 km², the mean densities of field-based and semi-automatic approach are 5.06 and 4.03 sinkholes/km², respectively (Table 2). Compared with the density in plateau karst region in Australia and temperate karst in Austria, these values are low. The sinkhole density is 122 km⁻² in Hochschwab Plateau, Australia (Plan and Decker 2006) and ranges from 91 to 146 in Styria basin based on the calculation of multiply models (Bauer 2015). The density values of this study area are comparable with those of Florida karst, which is known to be very flat, with broad, shallow sinkholes. In Florida, the sinkhole density values range from 2.6 to 15.8 in the Suwannee River basin (Denizman 2003) and obtain a highest value of 7.94 in the investigated regions of Troester et al. (1984).

A Kernel function was applied to calculate the density of the sinkhole (deepest point) in a predefined neighborhood. We set the neighborhood radius as 400 m, corresponding to the area classes. The resultant sinkhole density of semi-automatic approach has lower value range (max. 21 sinkholes/km²) than that of field-based approach (max. 50 sinkholes/km²). Figure 6 shows the sinkhole density of a part of the northern study area. The situation that dispersed spatial distribution occurs from the semi-automatic approach is also reflected by the comparison between Fig. 3a, b. Many sinkhole clusters with close distance shown in Fig. 3a had not been identified by semi-automatic approach shown in Fig. 3b.

Table 1 Numbers of the (1) true positive (TP), (2) false positive (FP), and (3) false negative (FN) identification of DEMs generated from different resources by the running of semi-automatic model with/without the thresholds of area = 60 m², ellipticity = 0.2, and TPI = 0

DEM sources	Spatial resolution (m)	Thresholds	No. of TP	No. of FP	No. of FN		
Topographic	3	N	423	108	216		
		Y	415	0	27		
	30	N	391	140	93		
		Y	379	80	78		
		N	262	269	68		
		Y	262	0	68		
	90	N	187	344	51		
		Y	172	0	0		
		SRTM	3	N	92	439	996
			30	Y	78	0	952
N	138			393	455		
60	Y		131	0	423		
	N		140	393	144		
90	Y	113	0	118			
	N	102	431	82			
	Y	102	0	82			
	GDEM	3	N	189	342	5179	
Y			78	0	1830		
30		N	169	262	1705		
		Y	157	0	1702		
60		N	143	388	515		
		Y	143	0	515		
90		N	125	406	280		
		Y	125	0	280		

Morphometry and statistics using different approaches

We analyzed the morphometric characteristics in the study area and compared the results of the manually delineated sinkholes by field-based approach and those identified by the semi-automatic approach. Because the 3-m topographic DEM-derived sinkholes from the semi-automatic approach were found to be accurate, they were used in further analysis. The statistics associated with the geometric characteristics of

the sinkholes delineated by the different approaches are shown in Table 2, Fig. 4, Fig. 5, and Fig. 6.

The area of the sinkholes is in a large span with big standard deviation (Table 2). The area of the sinkholes of field-based approach and semi-automatic approach skews to area < 17,000 m² with 93.79 and 97.21%, respectively. There are several small sinkholes and fewer large sinkholes, which is the same as that of temperate karst areas (Brinkmann et al. 2008; Bauer 2015). With the exclusion of the unusual large sinkholes, a log transformation was adopted to normalize the area data

Table 2 Summary statistics of sinkholes delineated according to the different approaches

	Field-based approach (n = 531)				Semi-automatic approach (n = 423)			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Area (m ²)	60	147,985	5423	11,761	6.25	52,100	3342.10	4562.25
Perimeter (m)	31.64	2329.89	283.58	295.99	10	3200	278.88	241.76
Diameter (m)	6.15	283.60	46.68	41.69	1.61	199.72	32.51	30.65
Ellipticity	0	0.99	0.78	0.15	0	0.99	0.78	0.15
Orientation (°)	0	179.89	82.67	48.97	1.84	178.86	79.66	50.13
Depth (m)	1	110	15.70	16.41	1	110	17.59	17.41
Volume (m ³)	125	2,663,730	60,367.96	223,139.80	2.08	303,555.4	20,543.83	35,901.49
Sinkhole density (km ²)			5.06				4.03	

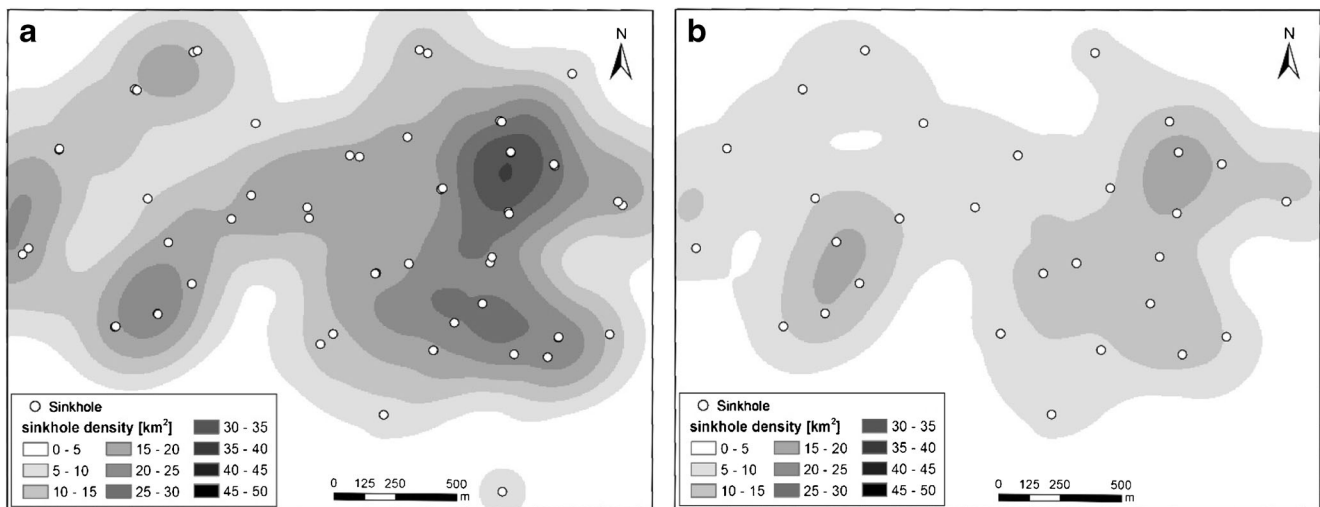


Fig. 3 Spatially distributed sinkhole density of a portion of the study area (highlighted in Fig. 1) by different delineation processes based on Kernel function: **a** field-based approach; **b** semi-automatic approach

(Figs. 4a and 5a). We divided the sinkholes based on geometrical interval classification into five classes: very small (0–400 m²), small (400–1900 m²), medium (1900–8300 m²), large (8300–35,000 m²), and very large (> 35,000 m²) (Fig. 6a).

Perimeter is generally positively related to area, except for crenulated sinkholes. The perimeter of the sinkholes is also not normally distributed with extreme outliers (> 900 m). The normalization is completed with a log transformation (Figs. 4b and 5b). See Fig. 6b, the sinkholes were classified to four perimeter categories based on two-step cluster analysis: very small (< 46 m), small (46–71 m), medium (71–134 m), and large (> 134 m).

The diameter of the sinkholes is defined as the line drawn with the most two distant point of the perimeter. Figures 4c and 5c show the skewed distribution of diameter < 100 m and normal distribution without the outliers. The sinkholes were divided to three classes using a two-step clustering analysis: short (< 41 m), middle (41–98 m), and long (> 98 m) diameter (Fig. 6c).

Based on the field work, most of the sinkholes in the study area tend to be elliptical. A way to discriminate between ellipse and non-ellipse depression is the use of ellipticity. The ellipticity value in this study area ranges from 0 to 0.99 with a mean value of 0.78 and a standard deviation of 0.15 (Table 2). See Figs. 4d and 5d, the mathematical distribution of the ellipticity calculated by different approaches is normally distributed and skewed to the right that the population is more ellipse than non-ellipse. As shown in Fig. 6d, two classes were separated by means of a two-step cluster analysis: 0–0.9, 0.9–0.99.

The orientation is the direction of the diameter. The orientation in the study area ranges from 0 to 180° (Table 2). Based on the analysis of the orientation of the two approaches, the orientation population is normally distributed. The bin size of the rose diagrams in Figs. 4e and 5e is 10°. The orientation slightly skewed to the value < 80°. A two-step clustering process was used to examine the population. The orientation value of 80°

splits the data into two cluster: one is sinkholes with orientation of 0–80° and the other is sinkholes with orientation of 80–180°. The peak values in the separate groups are both around 40° and 140° for both field-based and semi-automatic approach.

The volume of a sinkhole for the semi-automatic approach is obtained by calculating the difference of the filled-DEM and the original DEM. The volume of a sinkhole for the field-based approach is calculated by assuming a cone shape with the fieldwork sinkhole depth (Plan and Decker 2006). The resultant volume holds a large range of values as with big standard deviation as well as the mean value (Table 2). The population is not normally distributed and there are extreme outliers. We normalized the data when these outliers are removed and a log transformation was performed (Figs. 4f and 5f). Figure 6f shows the classes of sinkhole volume: very small (0–200 m³), small (200–2000 m³), medium (2000–20,000 m³), large (20000–200,000 m³), and very large (> 200,000 m³).

In addition, we calculated the depth of a sinkhole by semi-automatic approach as the maximum difference of the filled-DEM and the original DEM. As the depth statistics in Table 2, the depth ranges from 1 to 110 m with mean and SD value of ~20 m for both approaches. Figure 7 depicts statistics related with sinkhole depth. As shown in Fig. 7a) > 60% of the sinkhole depth derived from both approaches are > 20 m, demonstrating a predominantly deep morphology. In Fig. 7b), the area-to-volume ratio shows a similar and dispersed pattern for both approaches.

Discussion

The impact of DEM resolution

DEM resolution affects the ability to describe true sinkholes, most of which are small in the study area. It is noticed that

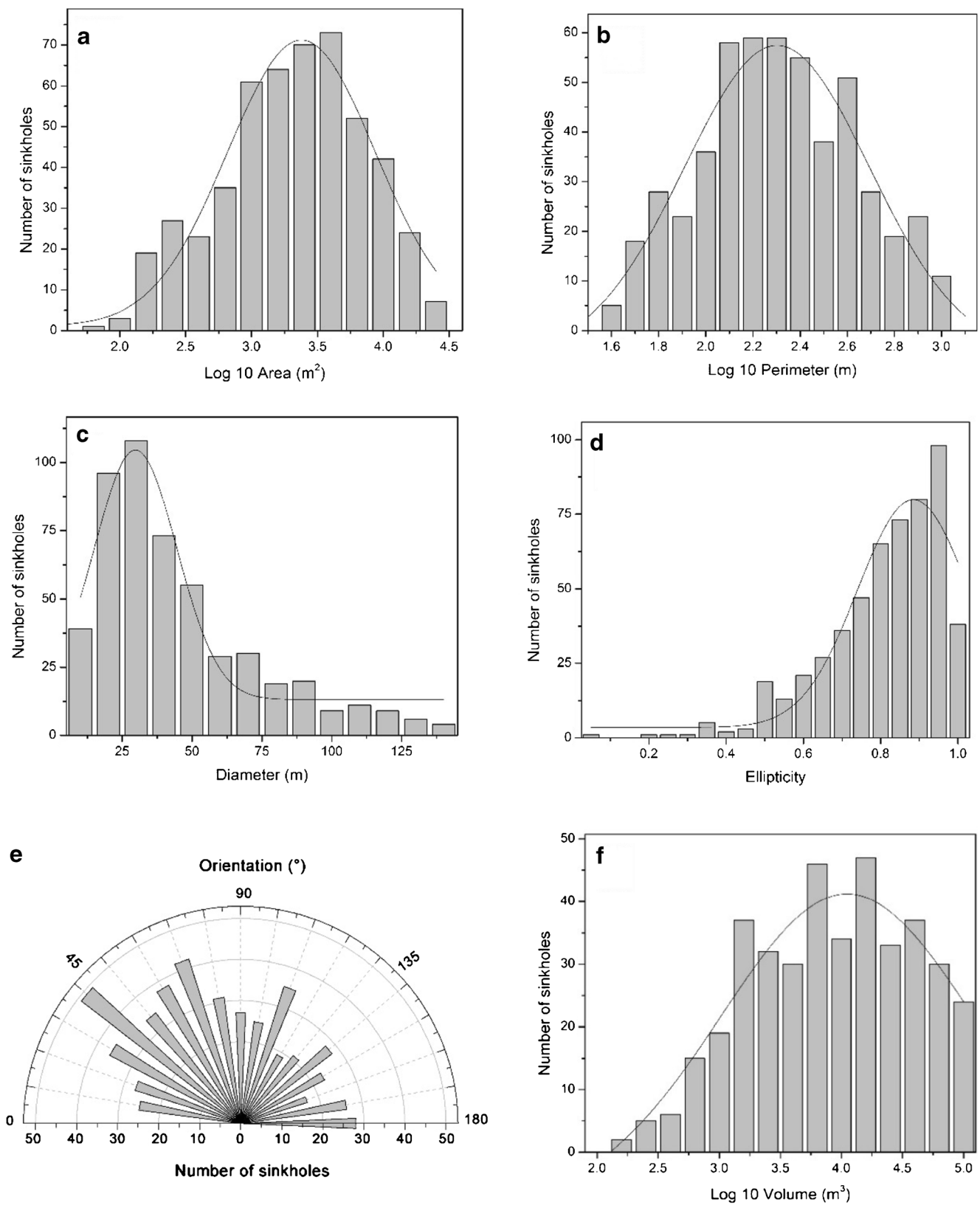


Fig. 4 The sinkhole **a** area, **b** perimeter, **c** diameter, **d** ellipticity, **e** orientation, and **f** volume distribution of sinkholes delineated by field-based approach: (1) all sinkholes with skewed nature; (2) sinkholes exclude extreme outliers; (3) normalized and log transformed sinkholes

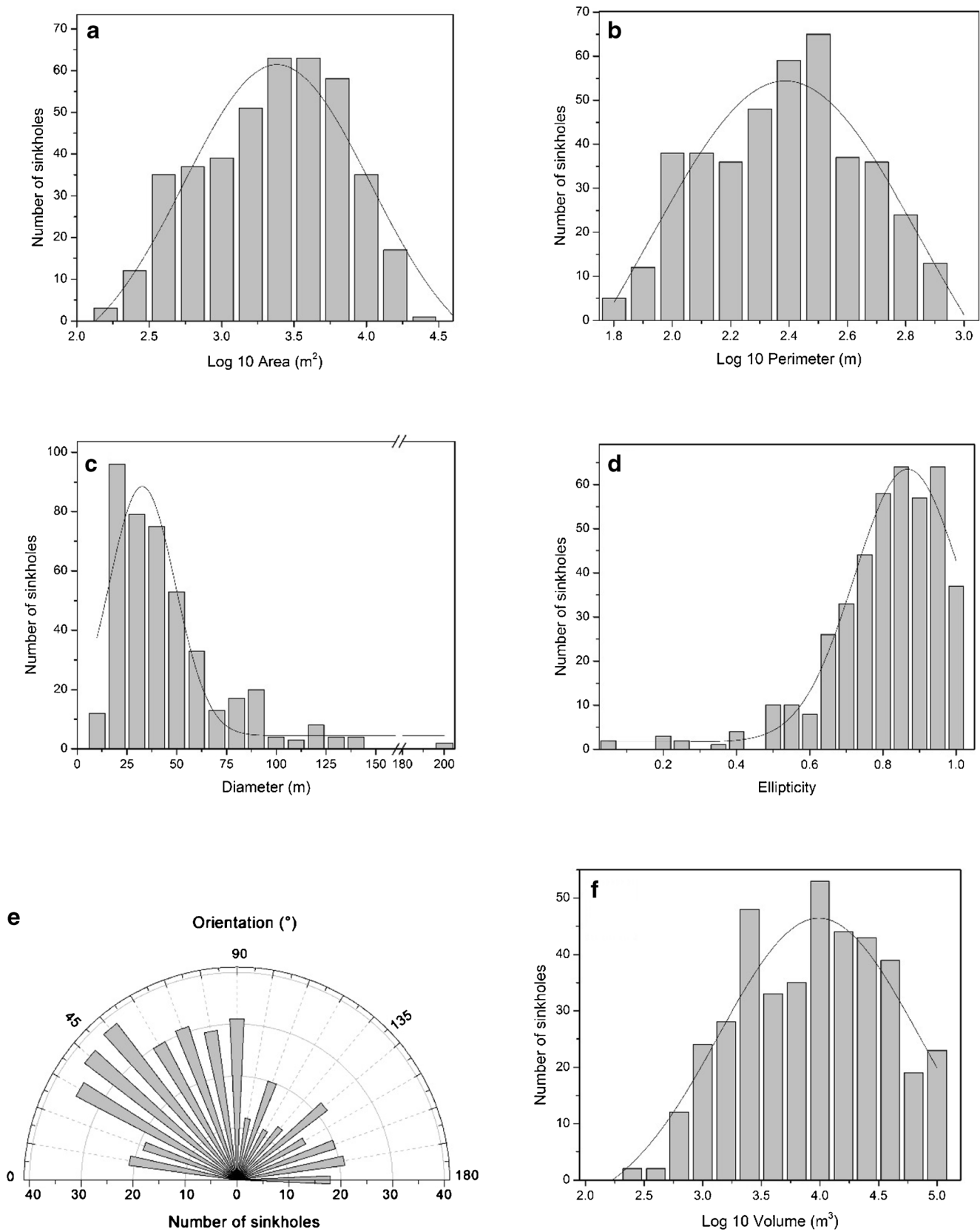


Fig. 5 The **a** area, **b** perimeter, **c** diameter, **d** ellipticity, **e** orientation, and **f** volume distribution of sinkholes distribution of sinkholes delineated by semi-automatic approach: (1) all sinkholes with skewed nature; (2) sinkholes exclude extreme outliers; (3) normalized and log transformed sinkholes

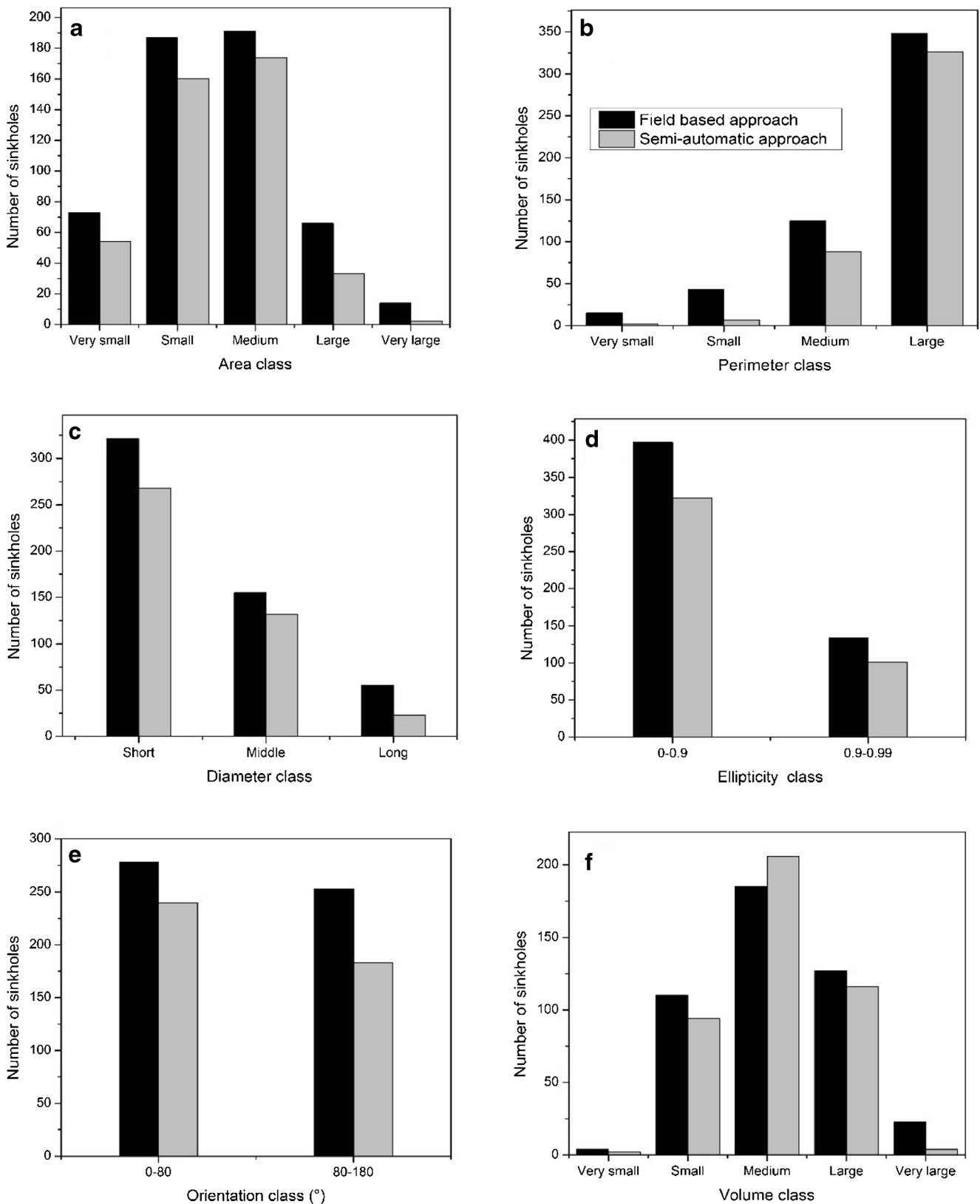


Fig. 6 The numbers of sinkholes delineated by different approaches in classes.

artificial sinkholes could be generated even from highly accurate elevation models (Li et al. 2011). In the present study, the

cell size of the original DEM from the topographic maps is 1 m, while the smallest sinkhole in the manual dataset is 60 m²

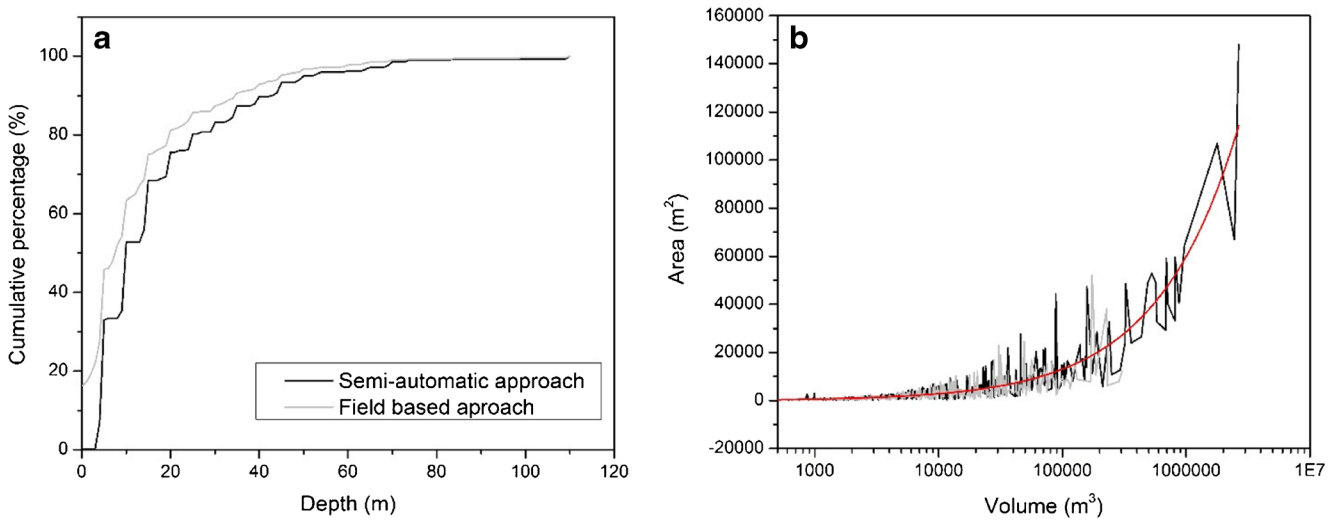


Fig. 7 The statistics related with sinkhole depth: **a** sinkhole depth frequency distribution, **b** distribution of sinkhole volume as a function of area

in area. It may be unnecessary to use such a detailed DEM, which might affect computing speed as well as producing many small artificial sinkholes due to local data noises.

The resampling process changed the original DEMs to coarser ones and created a smoother surface by eliminating fine details (Fig. 8). The mean error (ME) and root mean

square error (RMSE) between a resampled DEM and the original one are shown in Fig. 9. As expected, ME between the original topographic DEM and its coarsened DEMs generally increases with the grid size. RMSE of the DEMs tends to vary except for the relatively similar values at the grid size of 60 m, which can also be seen from the similarly delineated

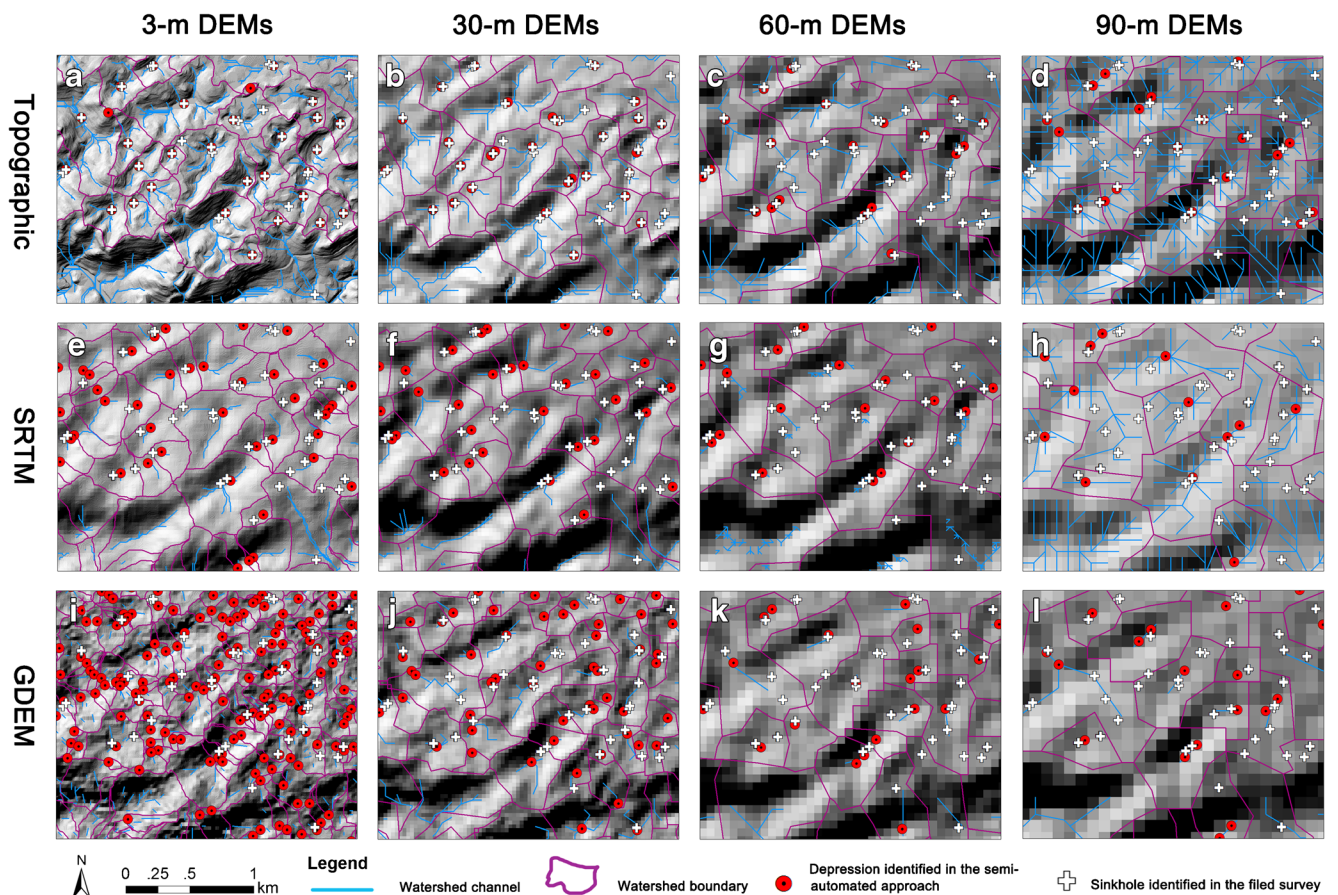
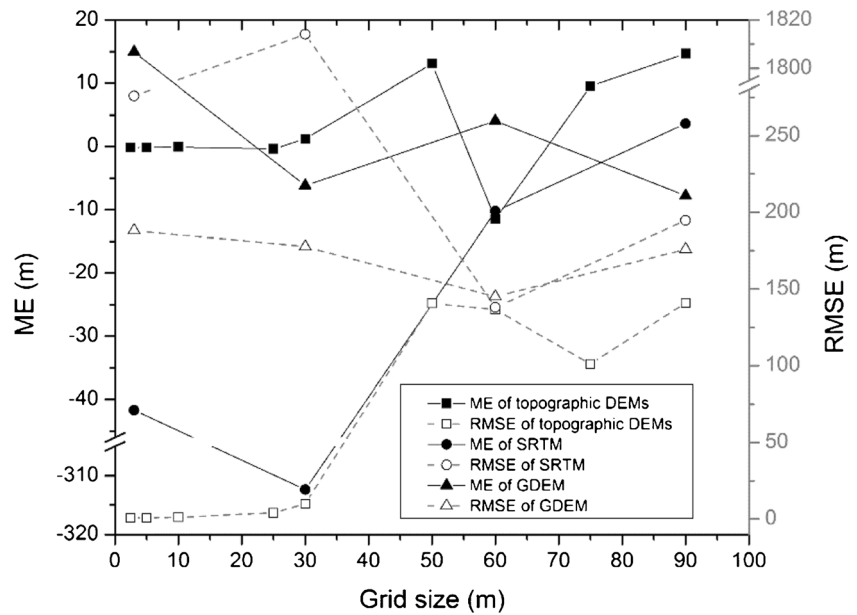


Fig. 8 Hillshaded maps of a portion of the study area (highlighted in Fig. 1) showing sinkhole connections and the effects of DEM resampling on the depression identification. Sinkhole locations are those determined by the semi-automated

Fig. 9 Effect of DEM resampling. ME and RSME are respectively the mean error and root mean square error of resampled DEMs (ASTER GDEM, SRTM-DEM, and DEMs from topographic maps) and the original topographic DEM with a resolution of 1 m



watershed boundaries and channels in Fig. 8. Also, ME and RMSE of the topographic DEMs show two patterns (Fig. 9): the values tend to be constant for the grid sizes < 30 m but more fluctuated at the grid sizes ≥ 30 m. The absolute values of ME and RMSE for the SRTM-DEM and GDEM tend to be larger than those for the topographic DEMs at all the grid sizes. However, the values for the SRTM-DEM change more drastically than those for the GDEM.

Moreover, the resampling probably restricts the shape of sinkholes. For instance, some extremely elongated small sinkholes were likely to be excluded in the resampling process to a coarser DEM. As mentioned in “Results,” we can see from Fig. 1 and Table 1 that the model with/without thresholds gets lower accuracy at coarser grid sizes (10–90 m in resolution) in the situation of DEMs derived from topographic maps, since the grid cells are square in shape with clustering area characteristics.

The application of morphometric thresholds

The application of the thresholds has improved the accuracy of the model for all the DEMs. Figure 2 and Table 1 evaluated accuracy and detailed performance of the model with DEMs from the different data sources in two scenarios: with or without thresholds. With the implementation of the thresholds, the model accuracy with the SRTM-DEM and GDEM improves (Fig. 2). As mentioned in “Results,” the performance of the semi-automatic approach with the topographic DEMs were separated to two sections. At the 3–10-m resolution section, in contrast to the decline with the thresholds, the accuracy of the model without the threshold increases from 55 to 65% with an increasing grid interval. This may reflect the morphometric characteristics of the sinkholes in the study area that are

skewed to the relatively small area class and an elongated shape; therefore, the number of the recognized true sinkholes increases with the grid size when there is no threshold. At the 10–90-m resolution section, the accuracy of the non-threshold model decreases apparently with the increasing of the grid size. However, the threshold model accuracy varies within a small range. This stable and good performance of the model and elimination of artifacts indicate that the thresholds we set are appropriate. In addition, at the grid size of 90 m, the model with thresholds shows markedly high accuracy, because the false positive and false negative sinkholes were removed by the application of the thresholds (Table 1).

The difference between the fields-based approach and semi-automatic approach

The morphometric characteristics generally follow those of the field-based approach. However, compared with the result of field work, the semi-automatic approach is more effective in identifying the sinkholes with moderate morphometric characteristics (not very small nor very large). For big sinkholes, as shown in Fig. 6a, the majority of sinkholes identified by the semi-automatic approach are in small and medium class, and < 40% are in large and very large class. Similarly, the sinkholes identified by the semi-automatic approach are less than those by the field-based approach at small sinkholes. The difference of sinkholes perimeter with relatively short perimeter (< 134 m) in Fig. 6b and short diameter class in Fig. 6c is much higher than that of other classes.

The good performance of the semi-automatic approach on the moderate sinkholes can also be seen from the statistic characteristics in Table 2. For example, the mean and SD values of the semi-automatic sinkhole area are much smaller

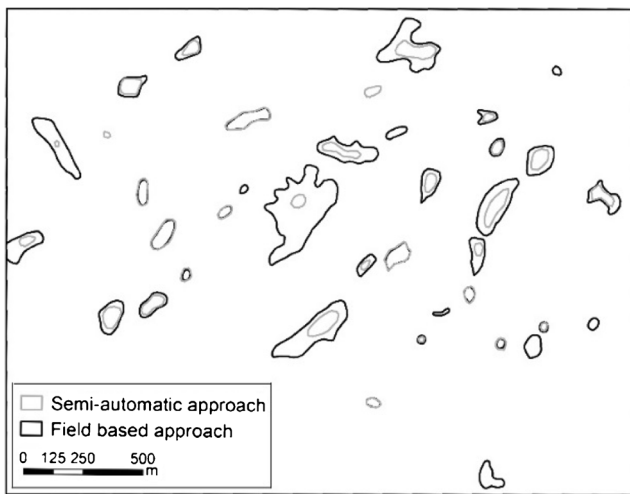


Fig. 10 Results of the two different approaches to delineate sinkhole boundaries (area highlighted in Fig. 1)

than those from the field-based approach. This is re-examined by the calculated sinkhole volume value. The mean value of sinkhole volume by field-based approach is three times as big as that of semi-automatic approach, while the statistical values of sinkhole depth by the two approaches are not significantly different with each other. In Fig. 6f, unlike the decreasing in other classes, the number of sinkholes identified by the semi-automatic approach in medium volumes class increased compared to that from the field-based approach.

The reason for the higher efficiency of the semi-automatic approach on the normal sinkholes identification might be related to the detailed delineation process. As shown in Fig. 10, the boundaries of the sinkholes with relatively small area are similarly depicted by the two approaches. For large sinkholes, the boundary from the semi-automatic approach is smaller than that from the field-based approach.

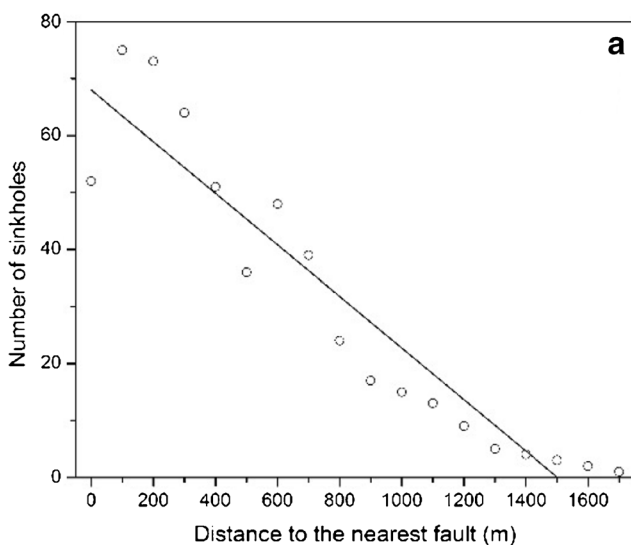


Fig. 11 The distribution of sinkholes and faults: **a** sinkhole frequency with the distance to the nearest fault. Linear correlation assigned: $R^2 = 0.88$; **b** sinkhole area percentage with the distance to the nearest fault. Linear correlation assigned: $R^2 = 0.89$

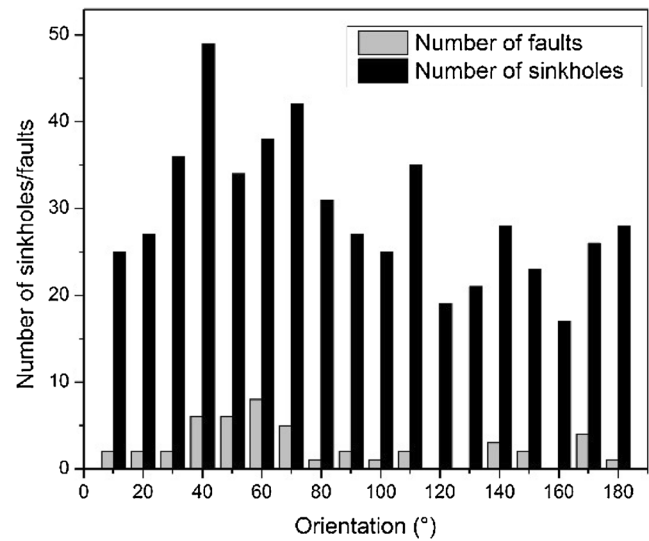
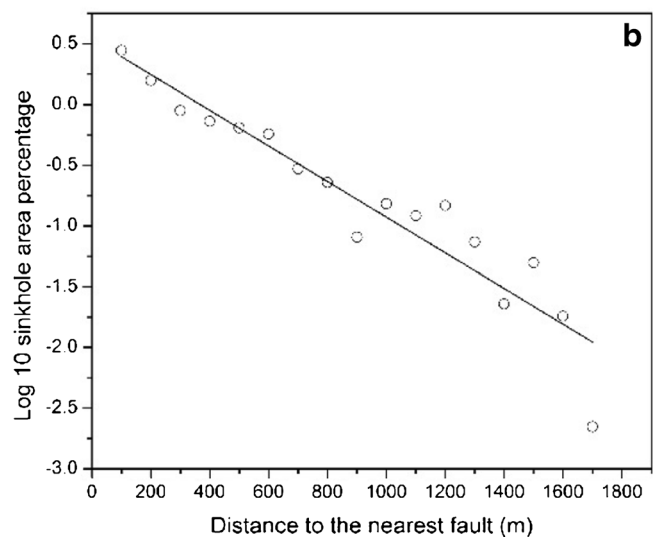


Fig. 12 Orientation frequency distribution of the sinkholes and the faults

The coincidence between sinkholes and geologic structures

As shown in Fig. 1, large sinkholes tend to occur in interfluvies, whereas smaller sinkholes are clustered closer to the river networks. The elongation of the large sinkhole and the alignment of small sinkholes generally extend along the NE direction. It is a striking trend similar to that of the NE trending faults due to tectonic deformation effect in the study area. This performance assumed that the spatial alignment and the elongation of sinkhole long axes are influenced by location of the faults.

To confirm this linkage between buried faults and sinkhole lines, we measured the distance between the sinkhole centroid and the nearest fault (Fig. 11) and the orientation between the sinkholes and faults (Fig. 12). It is noted that the number of



sinkholes (Fig. 10a) and the area of sinkholes (Fig. 10b) decrease with the distance to the neighborhood fault illustrates a potential structural control in the study area. As noted from Fig. 12, the bimodal distribution of sinkholes orientation mentioned in “Results” are similar with that of the fault orientation. The bimodal distribution is such that there is a separation at 80°. For the orientation curves of sinkhole diameter and faults, the high sinkhole frequency and peaks of each bimodal are at ranges of 40°–70° and 140°–170°.

In addition, it is interesting that more than 1/5 of the sinkholes are in extremely high ellipticity class (Fig. 6d). Intensively, in the ellipticity class of 0–0.9, there are only 34 sinkholes with the ellipticity value < 0.5. The high sinkhole ellipticity value reflected that the study area is dominated by the irregular or non-circular sinkholes. It infers a relatively old karst landscape in the study area (Brinkmann et al. 2008). This is not particularly surprising, since the landscape is higher than 900 m elevation and was inundated in the Triassic, Permian, and Carboniferous.

Conclusions

Our results suggest that the semi-automatic sinkhole identification approach using various DEMs provides an effective way to analyze sinkholes in broad and/or inaccessible areas. It reduces manual errors and processing time. The comparison of results from different datasets can be realized through the application of fast data acquisition at low cost. Although the ASTER GDEM is not suitable for research in the study area, it is not a general criticism of the data and it might perform better in other areas or for different objectives.

We resampled the DEMs and set thresholds for sinkhole identification, which aims to (1) exclude the false sinkholes due to data resources; (2) correspond sinkholes morphometric characteristics with different landscapes; (3) improve the accuracy of the model. DEM coarsening should be cautious because of the fact that small true sinkhole could not be captured in the DEMs with sizes > 30 m. The thresholds we set are area = 60 m², ellipticity = 0.2, and TPI = 0. With these thresholds, using DEMs derived from the topographic maps could produce the highest accuracy of the model. The accuracy of the semi-automatic model ranges from 0.78 to 0.95 for the DEM resolutions of 3 to 90 m. To sum up, appropriate combination of DEM resampling and thresholds allocation could achieve high model performance.

This study also demonstrates the sinkhole morphometry derived from different approaches. Some conclusions are made: (1) The morphometric characteristics of sinkholes derived from the semi-automatic approach are coincident with those from a field-based approach. The sinkhole morphometry (area, perimeter, diameter, orientation, ellipticity, and volume) in the study area covers a large span. (2) Sinkholes are skewed

with irregular or elliptical shape. This indicates that the sinkholes in the region are relatively old. (3) Sinkhole diameter and sinkhole alignments are generally parallel with that of the faults. Tensional faults provided the necessary conduits and structural conditions for the formation of sinkholes. (4) Area-to-volume ratio argues about solution-only origin for the sinkholes in the region. Following the description of Bauer (2015), the deepening of collapse sinkholes in this study area should be directly coupled with area widening, because the correlation between area and depth reflects the solutional process origin of sinkholes. However, as shown in Fig. 12b, the dispersed pattern of area-to-volume ratio (Fig. 12b) demonstrates that the morphogenesis and shape of the sinkholes in our study area are not only attributed by the solutional origin. Processes such as raveling and subsidence can also arise the generation of a sinkhole. Moreover, the sinkhole morphometry based on our field survey are similar with Caramanna et al. (2008) that some sinkholes are cylindrical shape with steep-sided walls.

Finally, it is important to mention that the mapping and examination of the morphometric features of sinkholes always represents a very difficult task as fieldwork. The semi-automatic approach is intended to provide a contribution toward an easier and deeper understanding of the local karst environment as a fundamental basis for the hazard associated with sinkholes.

Funding information This project is supported by Scientific Research Starting Foundation for High-level Talents of Huaqiao University (16SKBS305).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Anderson MG (1988) Modelling geomorphological systems. Modelling geomorphological systems. Wiley, Chichester
- Basso A, Bruno E, Parise M, Pepe M (2013) Morphometric analysis of sinkholes in a karst coastal area of southern Apulia (Italy). *Environ Earth Sci* 70(6):2545–2559
- Bauer C (2015) Analysis of dolines using multiple methods applied to airborne laser scanning data. *Geomorphology* 250:78–88
- Brinkmann R, Parise M, Dye D (2008) Sinkhole distribution in a rapidly developing urban environment: Hillsborough County, Tampa Bay area, Florida. *Eng Geol* 99(3):169–184
- Bruno E, Calcaterra D, Parise M (2008) Development and morphometry of sinkholes in coastal plains of Apulia, southern Italy. Preliminary sinkhole susceptibility assessment. *Eng Geol* 99(3):198–209
- Caramanna G, Ciotoli G, Nisio S (2008) A review of natural sinkhole phenomena in Italian plain areas. *Nat Hazards* 45(2):145–172

- Chang L, Hanssen RF (2014) Detection of cavity migration and sinkhole risk using radar interferometric time series. *Remote Sens Environ* 147:56–64
- Delle Rose M, Federico A, Parise M (2004) Sinkhole genesis and evolution in Apulia, and their interrelations with the anthropogenic environment. *Nat Hazards Earth Syst Sci* 4(5/6):747–755
- Denizman C (2003) Morphometric and spatial distribution parameters of karstic depressions, Lower Suwannee River Basin, Florida. *J Cave Karst Stud* 65(1):29–35
- Doctor DH, Young JA (2013) An evaluation of automated GIS tools for delineating karst sinkholes and closed depressions from 1-meter LiDAR-derived digital elevation data. 13th sinkhole conference
- Drake JJ, Ford DC (1972) The analysis of growth patterns of two-generation populations: the example of karst sinkholes.
- Elmahdy SI, Mostafa MM (2013) Natural hazards susceptibility mapping in Kuala Lumpur, Malaysia: an assessment using remote sensing and geographic information system (GIS). *Geomat Nat Haz Risk* 4(1):71–91
- Ford D, Williams PD (2013) *Karst hydrogeology and geomorphology*. Wiley, London
- Galve J, Gutiérrez F, Remondo J, Bonachea J, Lucha P, Cendrero A (2009) Evaluating and comparing methods of sinkhole susceptibility mapping in the Ebro Valley evaporite karst (NE Spain). *Geomorphology* 111(3):160–172
- Grimaldi S, Nardi F, Di Benedetto F, Istanbuluoglu E, Bras RL (2007) A physically-based method for removing pits in digital elevation models. *Adv Water Resour* 30(10):2151–2158
- Gunn J (2004) *Encyclopedia of caves and karst science*. Taylor & Francis, New York
- Gutiérrez F, Guerrero J, Lucha P (2008) A genetic classification of sinkholes illustrated from evaporite paleokarst exposures in Spain. *Environ Geol* 53(5):993–1006
- Gutiérrez F, Parise M, De Waele J, Jourde H (2014) A review on natural and human-induced geohazards and impacts in karst. *Earth Sci Rev* 138:61–88
- Hyatt JA, Jacobs PM (1996) Distribution and morphology of sinkholes triggered by flooding following Tropical Storm Alberto at Albany, Georgia, USA. *Geomorphology* 17(4):305–316
- Jenson SK, Domingue JO (1988) Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogramm Eng Remote Sens* 54(11):1593–1600
- La Valle P (1968) Karst depression morphology in south central Kentucky. *Geogr Ann Ser A Phys Geogr*:94–108
- Li S, MacMillan R, Lobb DA, McConkey BG, Moulin A, Fraser WR (2011) Lidar DEM error analyses and topographic depression identification in a hummocky landscape in the prairie region of Canada. *Geomorphology* 129(3):263–275
- Maidment DR (2002) *Arc Hydro: GIS for water resources*. ESRI, Inc
- Plan L, Decker K (2006) Quantitative karst morphology of the Hochschwab plateau, Eastern Alps, Austria. *Z Geomorphol Suppl* 147:29
- Salvati R, Sasowsky ID (2002) Development of collapse sinkholes in areas of groundwater discharge. *J Hydrol* 264(1):1–11
- Sauro U (2003) Dolines and sinkholes: aspects of evolution and problems of classification. *Acta Carsologica* 32(2):41–52
- Taheri K, Gutiérrez F, Mohseni H, Raeisi E, Taheri M (2015) Sinkhole susceptibility mapping using the analytical hierarchy process (AHP) and magnitude-frequency relationships: a case study in Hamadan province, Iran. *Geomorphology* 234:64–67
- Tharp TM (1999) Mechanics of upward propagation of cover-collapse sinkholes. *Eng Geol* 52(1):23–33
- Troester J, White EL, White WB (1984) A comparison of sinkhole depth frequency distributions in temperate and tropic karst regions. *Proceedings of the First Multidisciplinary Conference on sinkholes, Orlando Florida*. (ED. Beck, BF) Balkema, Rotterdam
- Van Schoor M (2002) Detection of sinkholes using 2D electrical resistivity imaging. *J Appl Geophys* 50(4):393–399
- Waltham T, Bell FG, Culshaw M (2007) *Sinkholes and subsidence: karst and cavernous rocks in engineering and construction*. Springer Science & Business Media, Chichester
- Weiss A (2001) Topographic position and landforms analysis. Poster presentation, ESRI User Conference, San Diego, CA
- White WB (1988) *Geomorphology and hydrology of karst terrains*. Oxford University Press, New York
- Williams PW (1972) Morphometric analysis of polygonal karst in New Guinea. *Geol Soc Am Bull* 83(3):761–796
- Witze A (2014) Florida forecasts sinkhole burden. *Nature* 504:196–197
- Youssef AM, El-Kaliouby HM, Zabramawi YA (2012) Integration of remote sensing and electrical resistivity methods in sinkhole investigation in Saudi Arabia. *J Appl Geophys* 87:28–39