

# Use of Gestalt Theory and Random Sets for Automatic Detection of Linear Geological Features

Dafni Sidiropoulou Velidou · Valentyn A. Tolpekin · Alfred Stein · Tsehaie Woldai

Received: 23 July 2014 / Accepted: 31 January 2015 / Published online: 28 February 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** This paper presents the calibration and application of a Gestalt-based line segment method for automatic geological lineament detection from remote sensing images. This method involves estimation of the scale factor, the angle tolerance and a threshold on the false alarm rate. It identifies major lineaments as objects characterized by two edges on the image, which appear as transitions from dark to bright and vice versa. These objects were modelled as random sets with parameters drawn from their distributions. Following the geometry of detected segments, a novel validation method assesses the accuracy with respect to a linear vector reference. The methodology was applied to a study area in Kenya where lineaments are prominent in the landscape and are well identifiable from an ASTER image. Error rates were based on distance and local orientation, and the study showed that the existence and size of the objects were sensitive to parameter variation. False detection rate and missing detection rate were both equal to 0.50, which is better than values equal to 0.65 and 0.63, observed using the Canny edge detection. Modelling the uncertainty of geological lineaments with random sets further showed that no core set is formed, indicating that there is an

---

D. Sidiropoulou Velidou (✉) · V. A. Tolpekin · A. Stein  
Faculty of Geoinformation Science and Earth Observation ITC,  
Hengelosestraat 99, 7514 AE Enschede, The Netherlands  
e-mail: d.v.sidiropoulou@gmail.com

V. A. Tolpekin  
e-mail: v.a.tolpekin@utwente.nl

A. Stein  
e-mail: a.stein@utwente.nl

T. Woldai  
School of Geosciences, University of the Witwatersrand, P O WITS,  
Johannesburg 2050, South Africa  
e-mail: t.woldai01@gmail.com

inherent uncertainty in their existence and position, and that the variance is relatively high. Comparing the test area with four areas in the same region showed similar results. Despite some shortcomings in identifying full lineaments from partially observed lineaments, it is concluded that the procedure in this paper is well able to automatically extract lineaments from a remote sensing image and validate their existence.

**Keywords** Lineaments · FAR · Segments · Image analysis · Statistics

## 1 Introduction

In image processing and computer vision, edges are represented by discontinuities in pixel intensities. Edges are of interest, as they may show linear features or discontinuities on the Earth surface. Edge detection deals with their tracking and identification on an image (Ziou and Tabbone 1998), resulting into an edge map as an intermediate product during automatic object extraction. Linear features on the Earth surface can be man-made, like roads (Heipke et al. 1995; Mena and Malpica 2005), natural, such as rivers, mountain ridges, or man-made consisting of natural objects like field boundaries (Turker and Kok 2013), paths and hedgerows (Thornton et al. 2007). Natural edges as such could for example refer to borders between distinct surface compositions (Juneja and Sandhu 2009) or natural linear objects like dykes. In this study, an edge is defined as a set of pixels in a linear configuration, with similar intensities, notably different from the background.

Several edge detection methods have been developed, mainly using a linear model to identify changes in brightness values (Khomyakov 2012). Gradient operators, such as the Sobel operator are distinguished from second-order derivative operators, like the Laplacian or Gaussian filter. Other methods, based on the Hough transform, have been proposed (Guru et al. 2004). A widely used method, proposed by Burns et al. (1986), explores the gradient magnitude and orientation of the pixels for edge detection. A recent method is based on Gestalt theory and the Helmholtz principle, where Gestalt theory describes the laws of human visual reconstruction and the Helmholtz principle states that structures are perceived when they largely deviate from randomness (Desolneux et al. 2001). Application of this theory on edge detection discriminates structures that would hardly ever occur in a random image and are perceptually meaningful, as specified by the false alarm rate. The detection method requires no a priori knowledge of the scene. A recent implementation of line detection with the integration of the Helmholtz principle is the Line Segment Detector by Grompone von Gioi et al. (2012), based on the method of Burns et al. (1986).

In this paper, an extension of this line segment detection method for the detection of geological lineaments is presented. A geological lineament is a mappable, simple or composite linear feature of a surface, with parts that are aligned in a rectilinear or slightly curvilinear relationship and indicates a phenomenon below the surface (O'Leary et al. 1976). Lineament mapping can be used for tectonic studies, the inference of past tectonic trends and the analysis of structural deformations (Hashim et al. 2013; Mostafa and Bishta 2005; Solomon and Ghebream 2006), whereas lineaments can be proofs of faults (Pal et al. 2006). Lineament characterization can be

exploited for seismic models (Papazachos et al. 2004), earthquake precursors (Soto-Pinto et al. 2013), hazard assessment and identification of geological hazard zones (Chandrasiri Ekneligoda and Henkel 2010), hydrology studies and to manifest discontinuities in minerals or rock mass (Chandrasiri Ekneligoda and Henkel 2010; Hashim et al. 2013). Lineament mapping methods are not only surveys in the field but also interpretation from images. Remote sensing is preferred for lineament mapping over in situ surveys, since it is less costly and time consuming and allows for the development of semi-automatic and automatic detection methods. It can rapidly provide a full picture of an extended area. For automatic lineament mapping several methods are encountered in the literature, including the Hough transform (Lee and Moon 2002; Soto-Pinto et al. 2013; Wang and Howarth 1990), a Segment Tracing Algorithm (STA) (Koike et al. 1995) and the Canny algorithm (Hashim et al. 2013; Marghany et al. 2009; Marghany and Hashim 2010), from remotely sensed images.

Lineament mapping is still extensively carried out by manual interpretation, despite attempts of establishing automatic methods. The former relates to expert knowledge and experience. Subjectivity, however, is unavoidable and affects both the extent and the existence of the lineaments (Gómez and Kavzoglu 2005; Ramli et al. 2010). In this paper, such uncertainty is modelled by random sets (Molchanov 2005). Random sets for image objects with vague boundaries have been recently presented by Zhao et al. (2011), for the quantification of extensional uncertainty of objects with fuzzy boundaries. Here, the sensitivity of the detected segments in the algorithm parameters is exploited for the generation of a random set. In this study, a new method for the accuracy assessment of automatically detected lineaments is used. So far, to our knowledge, no appropriate method is available from the literature.

The objective of this research is to explore the potential of a line segment detection method, based on Gestalt theory, for the automatic detection of geological lineaments from satellite images. The uncertainty in the existence and extent of geological lineaments is modelled with random sets. A novel method for the accuracy assessment of automatically detected lineaments is implemented. The method is tested on an ASTER scene of the Magadi region, in Kenya.

This paper is structured as follows: Section 2 introduces the developed methodology on line segment detection (2.1, 2.2), circular noise estimation (2.3), random sets modelling (2.4), accuracy assessment (2.5) and the order of method application (2.6). Section 3 describes the study area and the data used in this research. The results are presented in Sect. 4. Discussion of the key points of this research is raised in Sect. 5. The paper concludes in Sect. 6.

## 2 Methods

### 2.1 Segment Detection

First, a raster image of size  $N \times M$  and pixel values  $i(x, y)$  is considered and the set of pixels  $\mathbb{F} = \{i.x, y \in \mathbb{Z} : 1 \leq x \leq M, 1 \leq y \leq N\}$  is defined. The goal is to bring the image to such a scale where small lineaments are still detectable and major lineaments appear as narrow linear features. For this purpose the image is downscaled

to a coarser spatial resolution, using a scale factor  $S < 1$ . It is performed by applying a Gaussian  $5 \times 5$  of filter for smoothing, with  $\sigma = 0.8/S$ , followed by sub-sampling. The image gradient is computed, using a  $2 \times 2$  neighbourhood. The pixel gradient  $(g_x, g_y)$  at  $(x, y)$ , in the  $x$  and  $y$  directions, is calculated as

$$g_x(x, y) = \frac{i(x+1, y) + i(x+1, y+1) - i(x, y) - i(x, y+1)}{2}, \quad (1)$$

$$g_y(x, y) = \frac{i(x, y+1) + i(x+1, y+1) - i(x, y) - i(x+1, y)}{2}, \quad (2)$$

with a magnitude

$$G(x, y) = \sqrt{g_x^2(x, y) + g_y^2(x, y)}. \quad (3)$$

The gradient corresponds to the location  $(x+0.5, y+0.5)$  and the gradient image is stored in a new raster grid to account for this offset. The level-line angle of the pixel  $(x, y)$  equals

$$\phi = \arctan 2(g_x(x, y), -g_y(x, y)). \quad (4)$$

The  $\arctan 2(Y, X)$  is a function that returns the angle between the  $x$ -axis and the vector from the origin to  $(X, Y)$ , measured counterclockwise. It encodes the direction of the maximum change in pixel values in the local neighbourhood of the pixel. Level-line angles are transformed to the range  $[0^\circ, 360^\circ)$ . Gradient orientations of  $\phi = \alpha$  and  $\phi = \alpha + 180^\circ$  are distinguished, to account for the direction of the transitions among dark and bright intensities. The higher the value of  $G(x, y)$ , the more likely it is that the pixel is a member of an edge. Homogeneous zones are characterized by lower gradient magnitude. A threshold  $\omega$  is applied to exempt pixels of low  $G(x, y)$ .

The image of level-line angles  $\phi(x, y)$  is segmented, to obtain regions of pixels with similar orientations that correspond to distinct objects (Glasbey and Horgan 1995). A region growing algorithm is applied on the image of  $\phi$  to group pixels in line-support regions  $A$ . A region  $A$  grows starting from a seed point and continuing over all  $A$  points, with an 8-connected neighbourhood system, under the condition that the difference in the pixel  $\phi$  and the  $\theta_A$  is up to an angle tolerance  $\tau$ . A pixel cannot be the member of more than one  $A$ . In total, these pixels describe a potential linear segment with an associated segment orientation  $\theta_A$ . The steps are presented in Algorithm 1. Regions smaller than or equal to two pixels are removed, being considered as noise.

## 2.2 Gestalt Theory

The hypothesis that a linear segment is approximated by a narrow rectangle and the corresponding geometrical shape is estimated for all  $A$  is formed. The mass centre  $(c_x, c_y)$  of  $A$  equals

$$c_x = \frac{\sum_{j \in A} G(x_j, y_j) \times x_j}{\sum_{j \in A} G(x_j, y_j)}, \quad (5)$$

$$c_y = \frac{\sum_{j \in A} G(x_j, y_j) \times y_j}{\sum_{j \in A} G(x_j, y_j)}. \quad (6)$$

**Algorithm 1** Region growing

```

1:  $B = \begin{cases} G(x, y) & \text{if } G(x, y) > \omega \\ 0 & \text{otherwise} \end{cases}$ 
2:  $C = \{b \in B : b > 0\}$ 
3:  $seed = \{(x, y) : G(x, y) = \max(C)\}$ 
4:  $A = seed, \theta_A = \phi_{seed}$ 
5: identify 8-connected neighbourhood  $N_A$  of  $A$ 
6:  $N_{A^*} = \{h \in N_A : |\theta_A - \phi_h| \leq \tau\}$ 
7: if  $N_{A^*} \neq \emptyset$  then
8:    $A^* \leftarrow A \cup N_{A^*}$ 
9:    $\theta_{A^*} = \arctan 2(\sum_{j \in A^*} \sin \phi_j, \sum_{j \in A^*} \cos \phi_j)$ 
10:   $A \leftarrow A^*$ 
11:   $\theta_A \leftarrow \theta_{A^*}$ 
12:   $C^* \leftarrow \{b \in C : b \notin A^*\}$ 
13:   $C \leftarrow C^*$ 
14:  goto 5
15: if  $C \neq \emptyset$  then
16:  goto 3
    
```

The rectangle direction is the first inertia axis of  $A$ . It is the eigenvector angle corresponding to the smallest eigenvalue of the matrix

$$M = \begin{pmatrix} m_{xx} & m_{xy} \\ m_{xy} & m_{yy} \end{pmatrix},$$

where

$$m_{xx} = \frac{\sum_{j \in A} G(x_j, y_j) \times (x_j - c_x)^2}{\sum_{j \in A} G(x_j, y_j)}, \tag{7}$$

$$m_{yy} = \frac{\sum_{j \in A} G(x_j, y_j) \times (y_j - c_y)^2}{\sum_{j \in A} G(x_j, y_j)}, \tag{8}$$

$$m_{xy} = \frac{\sum_{j \in A} G(x_j, y_j) \times (x_j - c_x) \times (y_j - c_y)}{\sum_{j \in A} G(x_j, y_j)}. \tag{9}$$

The result is the smallest rectangle that contains all pixels of  $A$  (Fig. 1).

To check the likeliness of the occurrence of  $A$  in an a contrario model, the false alarm rate (FAR) is introduced. An a contrario model assumes that the gradient angles are realizations of a uniformly distributed variable. The FAR associated with  $A$  is defined as

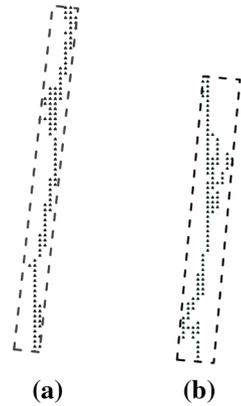
$$FAR = (N \times M)^{\frac{5}{2}} \times B(n, k, p), \tag{10}$$

with

$$B(n, k, p) = \sum_{j=k}^n \binom{n}{j} \times p^j \times (1 - p)^{n-j}, \tag{11}$$

where  $n$  equals the total number of pixels in the rectangle and  $k$  the number of pixels oriented similarly to  $A$ . The precision  $p$  is the probability that a pixel has a specific orientation in the a contrario model, and is set to  $p = \tau/180^\circ$ . Clearly, this value is

**Fig. 1** Examples of associated rectangles: the minimum rectangle that contains all the pixels of the line-support region. **a** High density of aligned points, low FAR, **b** lower density of aligned points, higher FAR



related to the angle tolerance  $\tau$  and the precision with which angles are considered. The value  $(N \times M)^{\frac{5}{2}}$  is used as an approximation of the total number of rectangles that could appear on the image, given the image dimensions and the precision  $p$ . This choice follows Grompone von Gioi et al. (2012), whereas the FAR is equivalent to the number of false alarms (NFA) but does not suggest an integer number. The FAR is an estimate of the probability of the occurrence of  $A$  in the a contrario model. A lower probability corresponds to a high degree of certainty that the given configuration represents an actual object on the image and that the detection is not an artefact of noise. A threshold  $\varepsilon$  is applied to FAR, to remove segments that are most likely created by noise. The variable  $\varepsilon$  is related to the significance level as used in statistical inference, and in the literature, it is usually set to a value  $\varepsilon \ll 1$  (Desolneux et al. 2000). This condition is tested below. Line-support regions with  $\text{FAR} < \varepsilon$  are termed  $\varepsilon$ -meaningful segments. The detection result is a set  $\mathbb{O}$  of  $\varepsilon$  meaningful segments.

### 2.3 Noise Estimation

After the application of the threshold to FAR, only structures that are not noise artefacts have to be accepted as meaningful segments. Hence, it is expected that the remaining part of the image (background) will accommodate pixels with uncorrelated gradient orientation. For the estimation of noise fraction, the image is split into two subsets. The first is the set of pixels  $\mathbb{O}$  and the second is  $\mathbb{F}^* = \mathbb{F} \setminus \mathbb{O}$ . Circular variance is calculated for both subsets and is compared to the variance of the pixels of the entire image. For circular variance, the mean resultant length of a vector  $\bar{R}$  is used.  $\bar{R}$  is an indicator of the concentration of the angular data set. It takes values close to 1 for clustered data and values close to 0 in case of dispersion (Mardia and Jupp 2009). A measure of the circular variance is

$$V = 2(1 - \bar{R})/w, \quad (12)$$

where  $w$  is the length of  $\bar{R}$  (Batschelet 1981). Calculation and interpretation of the variance are based on the assumption that all lineaments within the area follow a single

main direction. At this stage, orientations in the range  $[0^\circ, 180^\circ)$  are considered. Hence, lineament orientations of  $a$  and  $a + 180^\circ$  are regarded equivalent, as a convention to avoid having misleading variance of the data (Davis and Sampson 2002).

### 2.4 Random Sets by Parameter Variation

When segmenting an image containing uncertain objects, the extent of the detected segments is sensitive to the variation in the segmentation parameters (Zhao et al. 2011). In this study, experiments show that the detection of linear features is sensitive to the internal parameters of the detection algorithm. To address this issue, random sets are used that are defined on the basis of variation in the parameters of the detection algorithm. A random set is a spatial equivalent of a random variable and contains parameters that need to be estimated. If  $i(x, y)$  is the image pixel and  $\mathbb{O}_r$  is the sample  $r$ , that is the random realization of the object in iteration  $r$ , the indicator function equals

$$I_{\mathbb{O}_r(i(x,y))} = \begin{cases} 1 & \text{if } i(x, y) \in \mathbb{O}_r, \\ 0 & \text{if } i(x, y) \notin \mathbb{O}_r. \end{cases} \tag{13}$$

The covering function and set-theoretic variance of the random set are estimated as

$$\hat{P}_{\Gamma(i(x,y))} = \frac{1}{r} \times \sum_{j=1}^r I_{\mathbb{O}_r(i(x,y))}, \tag{14}$$

$$\Gamma_{\text{var}}(i(x, y)) = E[I_{\mathbb{O}_r(i(x,y))} - \hat{P}_{\Gamma(i(x,y))}]^2. \tag{15}$$

The covering function depicts the probability of a particular pixel to be an element of the random set. The set-theoretic variance is a measure of the existential and extensional uncertainty of each detected object and informs about the stability of the coverage of a particular pixel on all realizations (Fig. 13b). Low values imply either absence or presence of the lineament on the ground, as represented by the pixels. On the contrary, high variance shows that the membership of the pixel on the object fluctuates between iterations.

For our study, a random set is implemented by drawing random values of  $\omega$ ,  $S$  and  $\tau$  from their distributions. Random set generation is implemented as follows:

1. The number of set realizations ( $r$ ) is specified.
2. Distributions of the parameters  $S$ ,  $\omega$  and  $\tau$  are specified. Prior knowledge of the distributions is exploited for the probability density function (pdf), if available. Otherwise, the uniform distribution is used.
3. Parameter values for the two parameters are drawn at random from the distributions.
4. After segment detection the indicator function is calculated for every parameter configuration.
5. The covering function and set-theoretic variance of the random set are estimated (Eqs. 14, 15).

6. The median-level set and Vorob'ev expectation of the mean are estimated. Here, the median-level set includes only pixels with a covering function value above 0.5, whereas the Vorob'ev expectation of the mean of the random set identifies the pixels with covering values that sum up to half of the total support set area.

## 2.5 Accuracy Assessment

A novel accuracy assessment method is developed in this study. It compares the detection result  $\mathbb{O}$  and a set of points on reference lines  $\mathbb{E}$ . This scheme assesses the relative positional and directional accuracy between the sets  $\mathbb{O}$  and  $\mathbb{E}$ . It is based on Goodchild and Jeansoulin (1998) for the comparison of two sets of lines. It considers differences in local orientations and distance measures between reference and detection. The distinctiveness of the method is that it handles two datasets that represent the same feature (lineaments) under a different geometrical type: lines and points.

For the comparison between the two sets of features, measures of local orientation are defined. On the reference lines, points  $\mathbb{E}$  are sampled on intervals equal to the pixel spacing. For each of the sample points, the local line orientation ( $\alpha$ ) is calculated, considering a short linear segment, no longer than twice the pixel spacing, extending on both sides. For the points of the detected segments, a  $3 \times 3$  Gaussian kernel is applied on the neighbourhood of each pixel, considering only pixels that belong to the same segment. Doing so, a measure of local level-line angle ( $\bar{\phi}$ ) is produced. Next, corresponding points between reference and detection are determined. Two points are marked as corresponding when their distance and difference in local orientation are below some respective thresholds ( $d_E, d_O, \beta$ ). In other words, a pair of points is matching when the two points are close together and share similar orientations. The following sets are defined

$$\mathbb{E}^* = \{e \in \mathbb{E} : \exists o \in \mathbb{O}, (|\bar{\phi}_o - \alpha_e| < \beta), (d_{o-e} < d_E)\}, \quad (16)$$

$$\mathbb{O}^* = \{o \in \mathbb{O} : \exists e \in \mathbb{E}, (|\bar{\phi}_o - \alpha_e| < \beta), (d_{o-e} < d_O)\}, \quad (17)$$

where  $d_{o-e}$  is the distance between a point of the detection and a point on the reference. The following error rates of missing (Eq. 18) and false detections (Eq. 20) are used

$$\text{MR} = \frac{N_E - v_E}{N_E}, \quad (18)$$

$$\text{SR}_M = \frac{v_E}{N_E} = 1 - \text{MR}, \quad (19)$$

$$\text{FR} = \frac{N_O - v_O}{N_O}, \quad (20)$$

$$\text{SR}_F = \frac{v_O}{N_O} = 1 - \text{FR}, \quad (21)$$

where  $N_E = |\mathbb{E}|$ ,  $N_O = |\mathbb{O}|$ ,  $v_E = |\mathbb{E}^*|$  and  $v_O = |\mathbb{O}^*|$ . MR and FR correspond to errors of omission and commission, respectively.  $\text{SR}_M$  is the rate of successfully detected interval points and  $\text{SR}_F$  the rate of correctly detected points.

## 2.6 Order of Method Application

Methods presented in Sects. 2.1 and 2.2 are consecutively applied for the detection and selection of  $\varepsilon$ -meaningful segments. Noise estimation method of Sect. 2.3 is independently applied on the subset that is used for the tuning of the detection parameters. Random sets, in Sect. 2.4, are created from the segments detected after methods of Sects. 2.1 and 2.2 are applied. The accuracy assessment method of Sect. 2.5 is first exploited for the tuning of the internal algorithm parameters  $S$ ,  $\tau$  and  $\varepsilon$ . The tuning criterion is minimization of the sum of MR and FR. Each parameter is estimated individually by varying its value in a specified range and at specific intervals, keeping the other two parameters fixed. The sequence in which the parameters are tuned is based on their order of appearance in the algorithm. Second, accuracy assessment is applied on different image subsets for the validation of the detection result.

## 3 Study Area and Data

Lineaments under well-observable conditions occur in the southern Kenyan Rift, situated North of the lake Magadi in Kenya. This area was chosen as the study area. It is part of the East African Rift System (EARS) that extends from the Red Sea down to Mozambique. The Kenyan rift is regarded as one of the first occurrences of tectonic activity in the continent (Achauer and Masson 2002). In particular, the Magadi area is a region where rifting is still observed. The joints of the area are mainly straight and follow parallel tracks. The study area has almost no other linear structures and has limited vegetation. This facilitates detection, since lineaments are not occluded by other objects and there is no confusion between edge types.

### 3.1 Data

The image used for the study is an ASTER scene of the Magadi area, acquired in 2007. Scene area covered by the scene is approximately 1,230 km<sup>2</sup>. Bands in the visible and near-infrared (VNIR) spectrum have a spatial resolution of 15 m. The coordinate system is the Universal Transverse Mercator (UTM), Zone 37 South, and the datum World Geodetic System (WGS) 1984. The image is a Level 1B ASTER product. Level 1B products are radiometrically calibrated to radiance and geometrically co-registered data for all image bands (ERSDAC 2005).

The segment detection algorithm takes as input a single band image. Principal component analysis (PCA) is applied in the three bands of the VNIR spectrum of the used image and the first component is used. A subset of the image was selected for the calibration of the algorithm. The parameters were tuned according to the respective part of the reference dataset. The subset was selected such that it represented the surrounding area, containing most of the characteristic lineament forms in the region. Major and long lineaments as well as minor ones were encountered. Four additional subsets were selected from the same image, for the purpose of validation. All subsets have an extent of 60 km<sup>2</sup>.

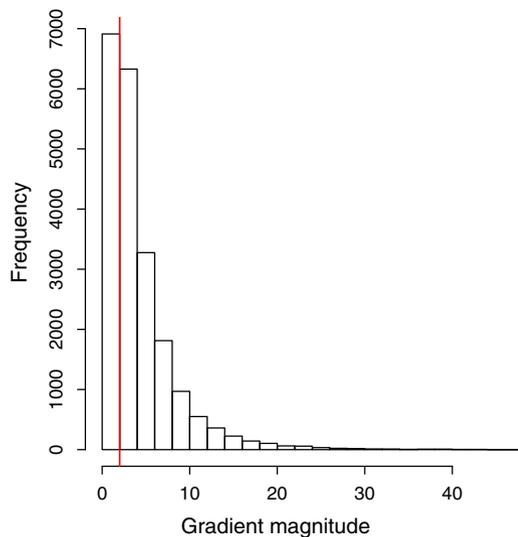
### 3.2 Reference Data

Reference data were obtained as vector lineaments, extracted from the geological map using a manual interpretation by an expert geologist, over an area of 1,230 km<sup>2</sup>. Respective parts of the dataset were used for tuning and validation.

## 4 Results

First, the different thresholds on the different parameters are specified. For the threshold of the gradient magnitude the histogram of the gradient magnitude image after resampling, shown for  $S = 0.3$  in Fig. 2, is considered. It was found that the shape of the histograms is relatively stable for different  $S$  values. To remove the smallest gradients, a uniform  $\omega$  threshold for all values of  $S$  was applied. For this purpose the threshold  $\omega = 2$  was selected. For lineament orientations that are commonly classified in intervals of 10° (Singhal and Gupta 2010), an additional error in the orientation estimation due to image smoothing is taken into account. To allow for the noise, the angle threshold is set equal to  $\beta = 12.5^\circ$  (Eqs. 16, 17). The distance thresholds  $d_E$  and  $d_O$  are different due to the difference in the geometry of the elements represented by  $\mathbb{E}$  and  $\mathbb{O}$ . As region growing is a pixel-based method, the shape and size of the resulting segments are scale dependent. For this reason, both  $d_E$  and  $d_O$  varied with scale. The distance threshold  $d_E$  of Eq. (16) reflects the relative positional accuracy of the detected segments with respect to the reference, being dependent on varying values of  $S$ ; it was set equal to  $d_E = 2 \times \text{pixel size}$ . The threshold  $d_O$  of Eq. 17 reflects the width of the detected segments, also varying with varying values of  $S$ ; it was set equal to  $d_O = 1\text{st quantile of the segment widths}$ .

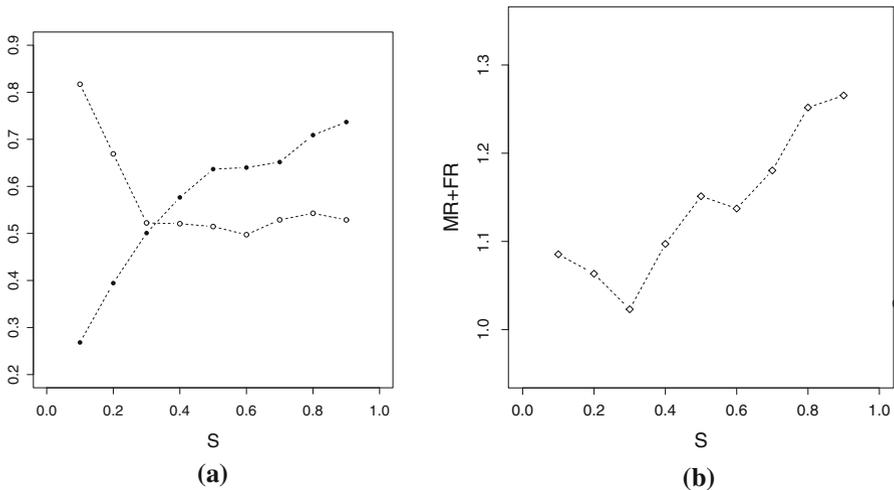
**Fig. 2** Histogram of gradient magnitude for  $S = 0.3$ . The *red line* indicates the threshold  $\omega = 2$



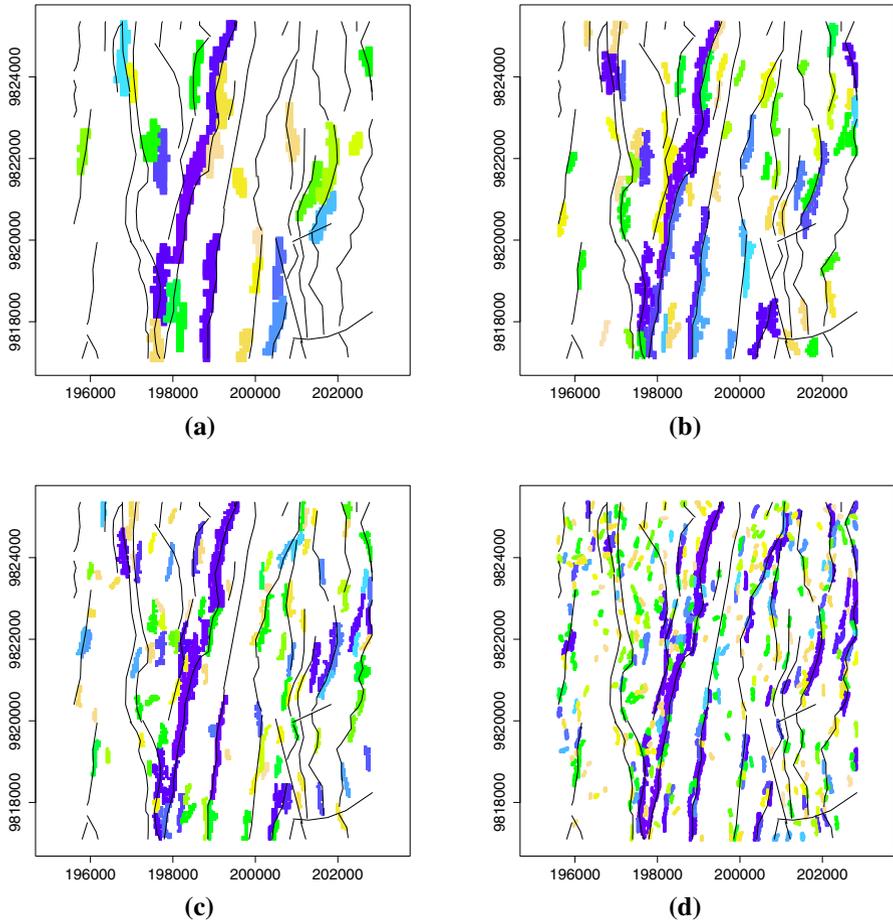
## 4.1 Parameter Tuning

### 4.1.1 Tuning of $S$

The first parameter to be tuned was the scale factor  $S$ . For the tuning of  $S$ , angle tolerance  $\tau = 22.5^\circ$  and  $\varepsilon = 1$  are kept constant (Grompone von Gioi et al. 2012). Scale factor values  $S$  between 0.1 and 0.9 were tested, at intervals of 0.1. The computation time of the region growing algorithm varies with  $S$ . Indicatively, for coarse resolutions corresponding to  $S = 0.1$  and  $S = 0.2$ , computation time was 0.27 and 1.39 s, respectively. For a scale factor of 0.9, producing a finer resolution image, the processing time is 1.40 min, on a PC with a 2.93 GHz processor. Resulting error ratios are presented in Fig. 3a. MR varied in the range of [0.50, 0.82], achieving its maximum value at  $S = 0.1$ , followed by  $MR = 0.67$  at  $S = 0.2$ . For  $S \in [0.3, 0.9]$ , MR was approximately equal to 0.5, having small variations (standard deviation = 0.01). FR followed an ascending trend moving towards high  $S$  values, in the range of [0.27, 0.74], being a wider range than that of MR. Its minimum was observed for  $S = 0.1$  and its maximum for  $S = 0.9$ . Figure 3b shows  $MR + FR$  for all the tested  $S$  values. A minimum  $MR + FR = 1.01$  was observed for  $S = 0.3$ , while  $MR + FR$  increased for  $S < 0.3$  and  $S > 0.3$ . It reached a maximum of  $MR + FR = 1.27$  at  $S = 0.9$ . The difference between the minimum and maximum value of the sum showed that the result is sensitive to the variation of  $S$  and that coarse resolutions are more suitable for the automatic detection. The value where the minimal sum of errors was observed was selected, fixing  $S$  at 0.3. Figure 4 presents the detection results, whereas the two error ratios and their sum are presented in Table 1. At coarse resolutions, the number of  $\varepsilon$ -meaningful segments, for example, was equal to 27 for  $S = 0.2$  and 69 for  $S = 0.3$  (Fig. 4a, b), whereas at a fine resolution (Fig. 4d), the number of  $\varepsilon$ -meaningful segments increased to 475 for  $S = 0.9$ .



**Fig. 3** **a** Ratio of missing (MR: open circles) and false (FR: closed circles) detections, **b**  $MR + FR$  for varying scale factor  $S$ . Lines are added to assist interpretation. Low  $S$  values imply coarse and large  $S$  values, fine spatial resolution, respectively



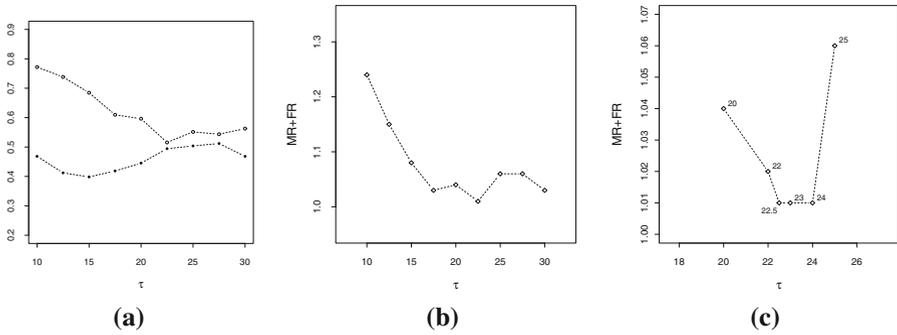
**Fig. 4** Detection results for different values of the scale factor  $S$ . **a**  $S = 0.2$ , **b**  $S = 0.3$ , **c**  $S = 0.4$ , **d**  $S = 0.9$ . Different colours represent different segments. Reference lines are overlaid in black

**Table 1** Error ratios for some  $S$  values

	$S$	MR	FR	MR + FR
	0.2	0.67	0.39	1.06
	0.3	0.52	0.49	1.01
	0.4	0.52	0.58	1.10
The sum of errors is minimized for $S = 0.3$	0.9	0.53	0.74	1.27

#### 4.1.2 Tuning of $\tau$

The second parameter to be tuned was the angle tolerance  $\tau$ . For tuning of  $\tau$ ,  $S$  was fixed at  $S = 0.3$  and  $\varepsilon$  was kept constant at 1. Angle tolerance values in the range of  $[10^\circ, 30^\circ]$  at intervals of  $2.5^\circ$  were examined. Low  $\tau$  implies strict threshold for region

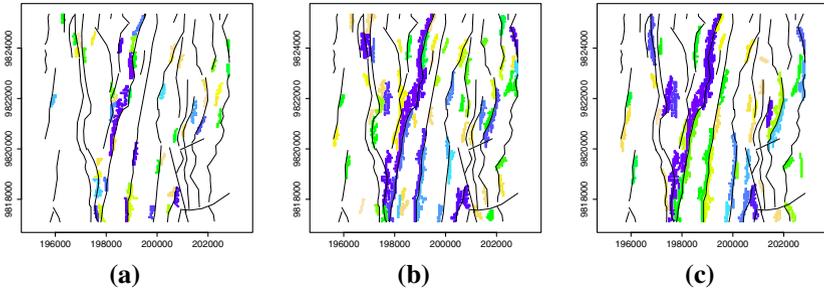


**Fig. 5** **a** Missing (MR: open circles) and false detections’ ratio (FR: closed circles), **b** MR + FR for varying angle tolerance ( $\tau^\circ$ ), **c** MR and FR for some additional  $\tau$  values. Lines are added to assist interpretation

growing. Higher  $\tau$  values are less restrictive for region growing, allowing pixels with greater differences in gradient orientation to be included into the same segment. Error ratios for the tested  $\tau$  values are plotted in Fig. 5a. MR, in the range of the tested values, showed a maximum equal to 0.77 at  $\tau = 10^\circ$ . It decreased for larger values, until it reached a minimum of 0.52 at  $\tau = 22.5^\circ$ . For  $\tau > 22.5^\circ$  it showed a small increase, up to MR = 0.56, for  $\tau = 30^\circ$ . FR is lower than MR for all values of  $\tau$  and ranges in [0.40, 0.51], being a smaller range than that of MR. It reached a minimum of 0.40 at  $\tau = 15^\circ$  and a maximum of 0.51 at  $\tau = 27.5^\circ$ . The sum MR + FR (Fig. 5b) started at a maximum of 1.24 at  $\tau = 10^\circ$ . It decreased for increasing values of  $\tau$  in the range of  $[10^\circ, 17.5^\circ]$ . A minimum of MR + FR = 1.01 was observed for  $\tau = 22.5^\circ$ . In general, MR + FR was lower for  $\tau \geq 17.5^\circ$  than for  $\tau < 17.5^\circ$ . The variation of the sum in the range of  $[17.5^\circ, 25^\circ]$ , where also the minimum was observed, was small with a maximum deviation of 0.05, between  $\tau = 22.5^\circ$  and  $\tau = 25^\circ$ . Other  $\tau$  values were tested in this range, to confirm the location of the minimum for (MR + FR). For values  $\tau = 22^\circ$ ,  $\tau = 23^\circ$  and  $\tau = 24^\circ$ , the respective sums are plotted in Fig. 5c. It was observed that MR + FR is equal to 1.01 for all tested  $\tau$  values in the range of  $[22.5^\circ, 24^\circ]$ . Hence, it is concluded that the sum is minimized for all the tested  $\tau$  values in the range of  $[22.5^\circ, 24^\circ]$ . The value  $\tau = 22.5^\circ$  was thus selected. To visualize the sensitivity, the detection result for three different values of  $\tau$  is presented in Fig. 6: the selected value  $\tau = 22.5^\circ$  (Fig. 6a), a much smaller  $\tau = 12.5^\circ$  (Fig. 6b), and a much larger  $\tau = 30^\circ$  (Fig. 6c). Using the narrow angle a substantial number of segments were missed, whereas little differences were noted between the selected and wider angle. The respective errors are given in Table 2.

### 4.1.3 Tuning of $\varepsilon$

The third parameter to be tuned was the FAR threshold  $\varepsilon$ . To do so, the values of  $S = 0.3$  and  $\tau = 22.5^\circ$  were fixed. Several values above and below 1 were tested with the intention to check whether segments with high probability of occurrence in a white noise image would be allowed to pass the FAR check, i.e.  $\varepsilon > 1$ . Imposing the threshold  $\varepsilon < 1$  would result in the acceptance of only segments with a low probability

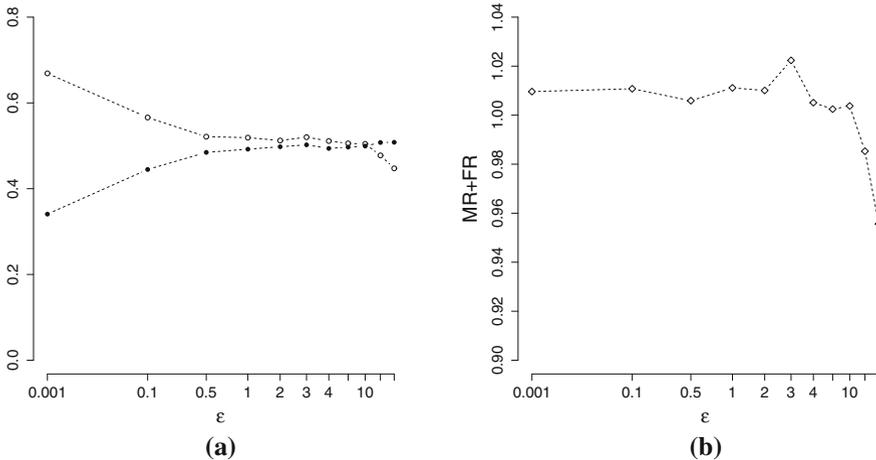


**Fig. 6** Detection results for different values of the angle tolerance ( $\tau$ ). **a**  $\tau = 12.5^\circ$ , **b**  $\tau = 22.5^\circ$ , **c**  $\tau = 30^\circ$

**Table 2** Error ratios for some  $\tau$  values

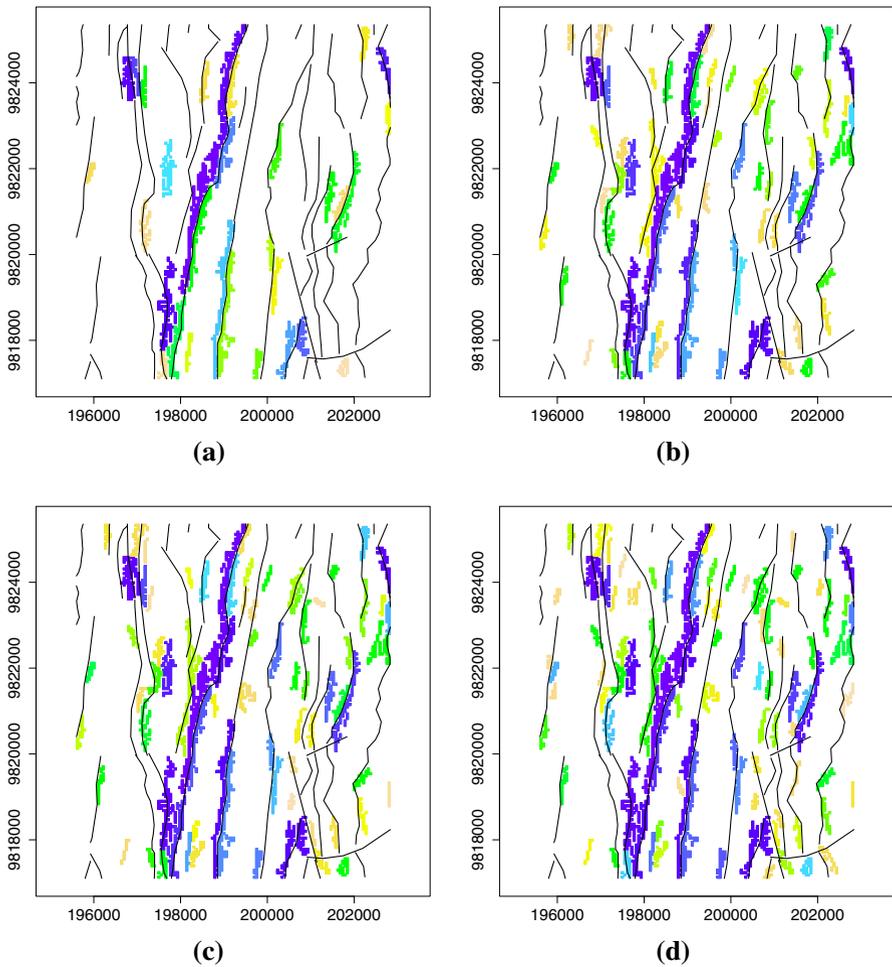
$\tau$	MR	FR	MR + FR
$12.5^\circ$	0.74	0.41	1.15
$22.5^\circ$	0.52	0.49	1.01
$30^\circ$	0.56	0.47	1.03

The sum of errors is minimized for  $\tau = 22.5^\circ$



**Fig. 7** **a** Missing (MR: *open circles*) and false detections' ratio (FR: *closed circles*), **b** MR + FR for varying  $\epsilon$  (plotted in log scale). Lines are added to assist interpretation

of occurrence in the a contrario model. Figure 7a shows the behaviour of MR and FR for the tested  $\epsilon$  values. As expected, MR increased and FR decreased for low  $\epsilon$  values. For  $\epsilon = 0.001$ , MR = 0.67, being its maximum value, and FR = 0.34, being its minimum value. As  $\epsilon$  increases, MR decreased until it reached a minimum of 0.45 for  $\epsilon = 100$ , whereas FR increased up to a maximum of 0.51 for the same  $\epsilon$  value. In Fig. 7b the sum MR + FR is presented for the tested  $\epsilon$  values. A minimum of 0.96 was observed for  $\epsilon = 100$ . The sum MR + FR increased for  $\epsilon < 100$ , with a maximum of 1.02 at  $\epsilon = 3$ . The maximum difference in MR + FR was equal to 0.06, occurring



**Fig. 8** Detection results for different values of the FAR threshold ( $\epsilon$ ). **a**  $\epsilon = 0.001$ , **b**  $\epsilon = 1$ , **c**  $\epsilon = 10$ , **d**  $\epsilon = 100$ . The respective values of MR and FR are found in Table 3. The respective numbers of  $\epsilon$ -meaningful segments are 31, 69, 78 and 96

between the values  $\epsilon = 3$  and  $\epsilon = 100$ . The order of magnitude of this difference did not lead to the rejection of any value or to the selection of an optimal value for  $\epsilon$ . Thus,  $\epsilon$  is set equal to 10, the value for which  $MR = FR$ . Detection results for  $\epsilon = 1$  and  $\epsilon = 10$  are shown in Fig. 8b and c. The number of  $\epsilon$ -meaningful segments was equal to 69 and 78, respectively. As Fig. 8a shows, the number of detected segments reduced to 31. Finally, for  $\epsilon = 100$ , the number of accepted segments increased to 96.

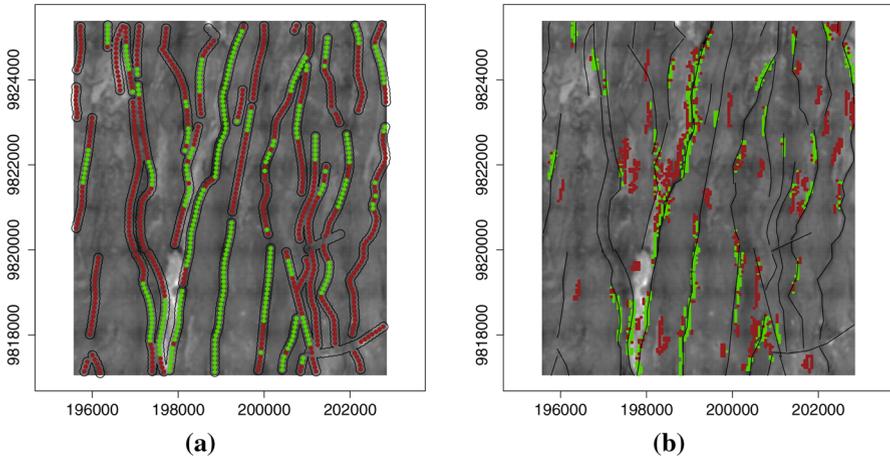
#### 4.2 Errors

The locations of false and missing detections are indicated in the maps of Fig. 9. In Fig. 9a it is seen that the majority of the reference lines were partially detected.

**Table 3** Error ratios for some  $\varepsilon$  values

$\varepsilon$	MR	FR	MR + FR
0.001	0.67	0.34	1.01
1	0.52	0.49	1.01
10	0.50	0.50	1.00
100	0.45	0.51	0.96

Deviation of the error sum for different  $\varepsilon$  values is relatively small. MR and FR are equal for  $\varepsilon = 10$

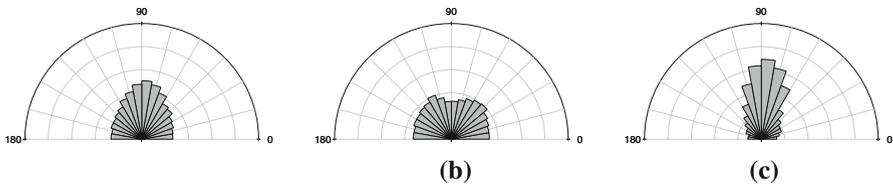


**Fig. 9** Error visualization. **a** Successful and Missing detections: *green* reference points that have been detected. *Red* reference points that have not been detected. The *black buffer line* illustrates the distance threshold. **b** True and False detections: *green* detected points that correspond to the reference; *red* false detections

These missing detections corresponded to parts of the lineaments that are covered by sediment and, consequently, do not appear as edges on the image. The delineation of these lineaments in the reference stems from experience in interpretation, that allows the connection of interrupted parts of the same structure. Missing detections were also observed on lineaments with curvature. In Fig. 9b it is shown that the majority of false detections corresponds to linear features of other type, for example drainage network or soil texture. Further, as the detected segments were wider than the reference lines, detected pixels on the periphery of wide segments were marked false. False detections were also observed when a detected segment extended further than the length of the reference line.

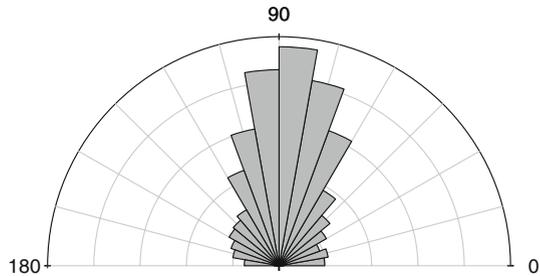
### 4.3 Noise Estimation

After selecting the detection parameters  $S = 0.3$ ,  $\tau = 22.5^\circ$  and  $\varepsilon = 10$ , the image was examined for noise estimation. The rose diagrams of Fig. 10 derived from the level-line angles ( $\phi$ ) of the image pixels portrayed the frequency of gradient directions. Gradient angles were transformed in the range  $[0^\circ, 180^\circ]$  and directions with a difference of  $180^\circ$  were considered parallel. The rose diagram of  $\mathbb{F}$  (Fig. 10a) showed



**Fig. 10** Rose diagrams of level-line angles. **a** Entire image, **b** background, **c** detection. Angles are transformed in the range  $[0^\circ, 180^\circ]$ . Angle of  $90^\circ$  corresponds to a N/S direction

**Fig. 11** Rose diagram of the lineaments in the reference dataset. Angles are transformed in the range  $[0^\circ, 180^\circ]$ . Angle of  $90^\circ$  corresponds to a N/S direction. The diagram shows a main direction N–NNE/S–SSW



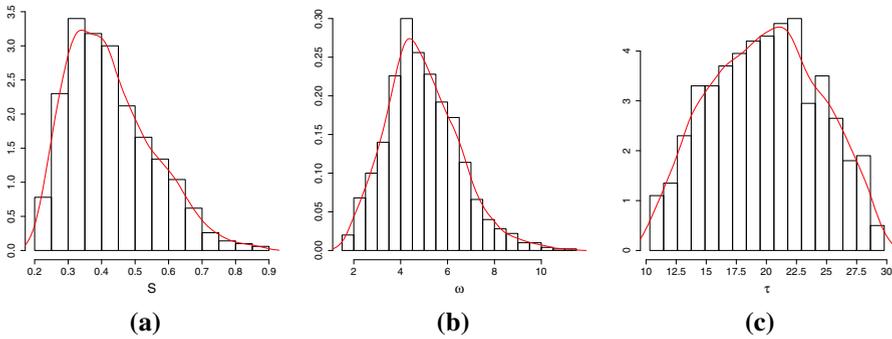
**Table 4** Circular variance of level-line angles of the entire image, background and detected pixels

	Circular variance
Entire image	0.484
Background	0.618
Detection	0.189

that the angles were homogeneously distributed over the range, with a higher frequency between the values  $80^\circ$  and  $90^\circ$ . In the rose diagram of  $\mathbb{F}^*$  (Fig. 10b) the clustering of values that was noticed before was now absent. Many pixels with an orientation between  $80^\circ$  and  $90^\circ$  were eliminated since they belong to  $\mathbb{O}$ . Gradient orientations again were spread over the whole range. The rose diagram of  $\mathbb{O}$  (Fig. 10c) presented the frequency of the detected lineaments in all orientations. It depicted a high concentration around  $80^\circ$ – $90^\circ$ , while the frequency in all other directions was significantly lower. This result was as expected, since almost all lineaments on the image have a N–NNE/S–SSW direction, as shown in Fig. 11. The circular variance (Eq. 12) was calculated for the three sets (Table 4). The circular variance of  $\mathbb{F}$  was higher than the one of  $\mathbb{F}^*$ . In the original image, the frequency of orientations around  $90^\circ$  was prevalent. Once pixels of  $\mathbb{O}$  were excluded, the remaining orientations had a more uniform distribution than the original. The variance in the detected pixels was much lower, since most of the orientations were clustered around  $80^\circ$ – $90^\circ$ .

#### 4.4 Random Sets by Parameter Variation

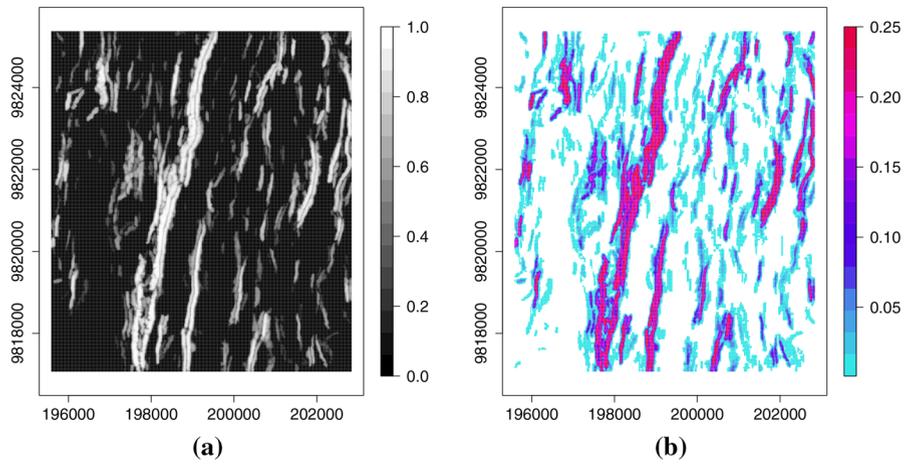
For the generation of a random set, parameter values of  $S$ ,  $\omega$  and  $\tau$  were sampled from their probability distributions. The threshold  $\varepsilon$  was kept constant at  $\varepsilon = 1$ , a



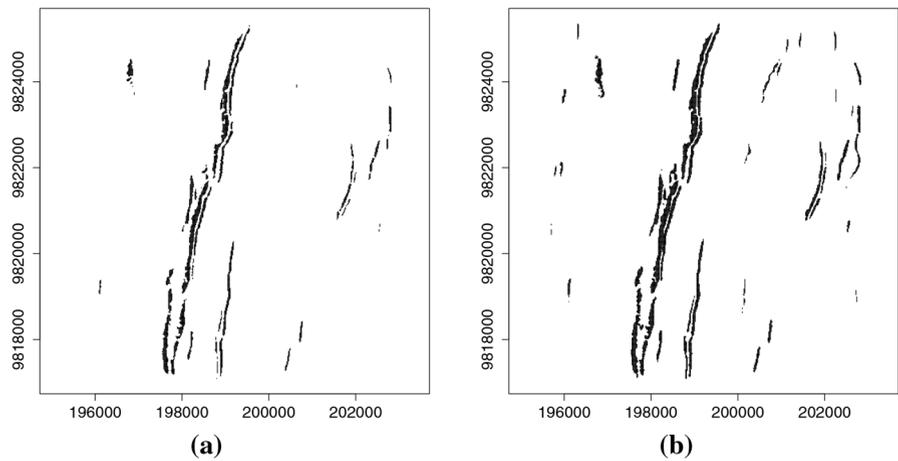
**Fig. 12** Parameter distributions: **a** generalized  $B$  distribution for  $S$ , **b**  $\Gamma$  distribution for  $\omega$ , **c** generalized  $B$  distribution for  $\tau$

value from the range that was obtained from parameter tuning. All three parameters take only positive values and have a finite range. For this reason, they were described by the  $B$  and  $\Gamma$  distributions. The scale factor  $S$  was expressed by a generalized  $B$  distribution with parameters  $\alpha = 2$  and  $\beta = 7$  (Fig. 12a). Values ranged from 0.2 to 0.9 with a peak around 0.3, which was the value selected after parameter tuning. The gradient magnitude threshold ( $\omega$ ) was described by a  $\Gamma$  distribution with parameters  $\kappa = 10$  and  $\theta = 0.5$ . The values for the  $\omega$  distribution were selected after visual inspection of thresholded magnitude images. Experiments showed that values close to 5 removed the pixels of the image background and maintained pixels on the edges. The resulting range was  $[1.3, 10.0]$ , with a peak around 5 (Fig. 12b). Angle tolerance ( $\tau$ ) was described by a generalized  $B$  distribution with parameters  $\alpha = 2$  and  $\beta = 2$  (Fig. 12c). The generalized  $B$  distribution was used for the transformation of the range  $[0, 1]$  to the range of tested  $\tau$  values. Also, the peak of the distribution of the angle tolerance was adjusted based on the experiments of parameter tuning (Sect. 4.1). Values were in the range  $[10^\circ, 30^\circ]$  with a peak around  $22.5^\circ$ . The number of realizations was set to  $r = 1,000$ , which gave similar results as to  $r = 100$  and took around 10 h to complete, on a PC with a 2.93-GHz processor. Random values were sampled from these distributions and 1,000 random combinations of values of the three parameters were created. Segments were detected for all 1,000 combinations and each result was considered a realization of a random set representing geological lineaments. From these realizations, the covering function, the set-theoretic variance and level sets were computed (Figs. 13, 14, 15).

Figure 13a showed that pixels with high values of the covering function point to major geological structures. These objects were detected in most of the realizations and were, therefore, less dependent on the parameter values. There were many smaller segments with low covering function values, implying that their existence is less certain. Few segments had higher values on their central pixels and lower values at the start and end vertices. In these cases, an extensional uncertainty was depicted. High variance on major structures was explained by the absence of a core set ( $\hat{P}_{\Gamma(x)} = 1$ ). For example, small segments with low values of the covering function and a low variance reflected a low certainty of their existence. From Fig. 13a it was observed that

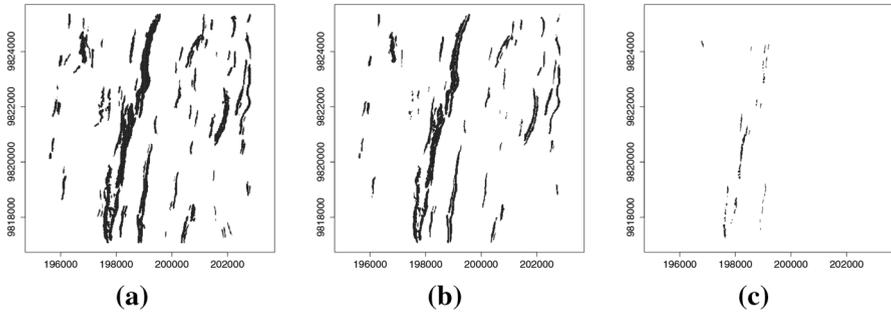


**Fig. 13** Random set statistics. **a** Covering function, in *black* pixels that are not set elements. **b** Set-theoretic variance, in *white* pixels with no variance



**Fig. 14** **a** Median-level set, **b** Vorob'ev expectation of the mean, in *black* pixels that belong to the respective sets

small segments occur with different covering function values, oriented in alignment. This implied that the parameter values also affected the segment connectivity. Their variation can assist decision making of whether individual segments are part of a single lineament. The median-level set includes pixels with high probability of existence. In our case, the Vorob'ev expectation set of the mean contained 8,105 detected pixels, in contrast to 5,150 of the median set, over a total number of 177,694 pixels. The relatively limited number of pixels in both the median-level set and the Vorob'ev expectation of the mean set was due to the overall low values of covering function (Fig. 15). Low covering function values, in turn, indicated a high degree of uncertainty of the existence of the detected segments.

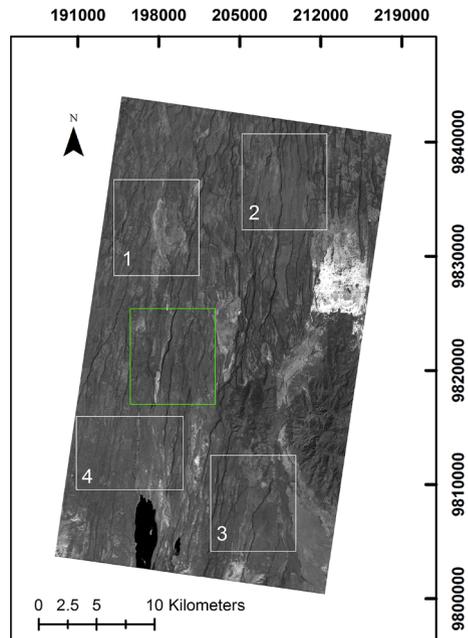


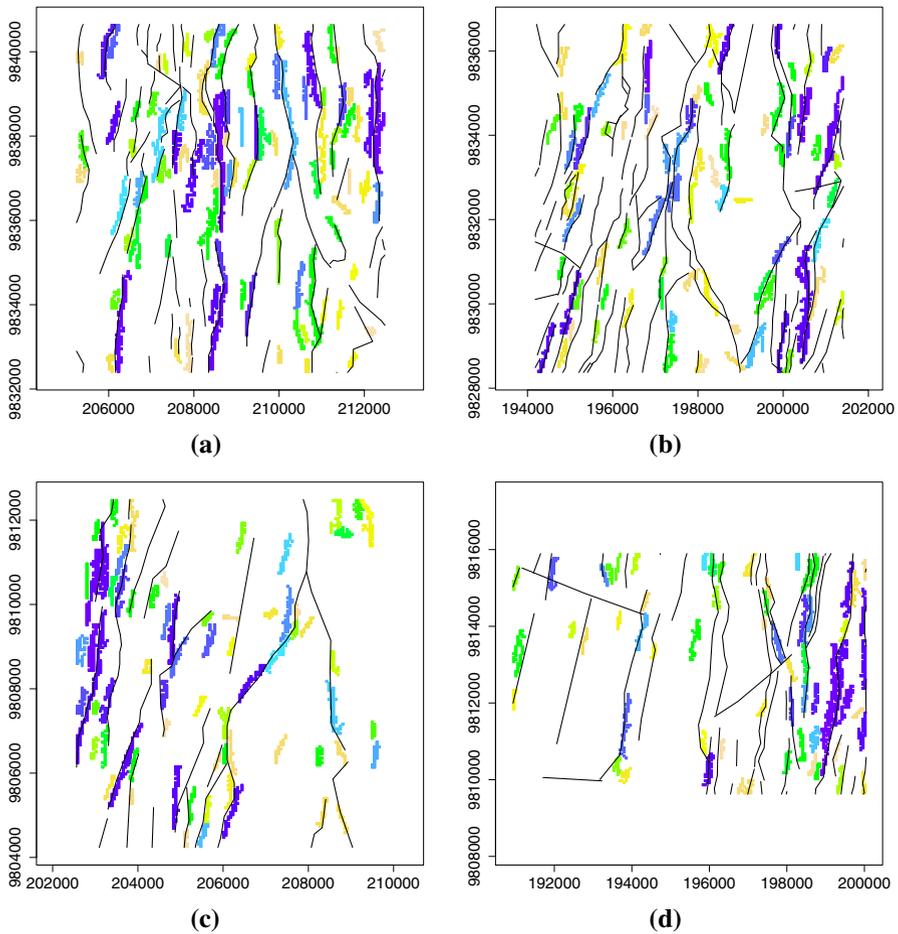
**Fig. 15** P-level sets. **a**  $p = 0.1$ , **b**  $p = 0.2$ , **c**  $p = 0.8$ . Very few pixels with high covering function value

#### 4.5 Validation

Once the parameter values were fixed, the algorithm was tested in four subsets of the image, labelled subsets 1–4, using the same parameter values. The distribution of the subsets around the area is found in Fig. 16. All subsets, including the one used for tuning, cover areas of approximately  $60 \text{ km}^2$ . The detection results are presented in Fig. 17 and the respective error ratios in Table 5. The magnitude of the error ratios for subset 1 was lower than on the tuning subset. Also,  $\text{FR} = 0.48$  was larger than  $\text{MR} = 0.43$ . False detections corresponded to short edges of other types, which might reflect geological boundaries between varying soil and rock. Missing detections were mainly noticed on noisy and weak edges. There were also lineaments of the

**Fig. 16** PC1 of ASTER image. *Green* the subset used for parameter tuning. *White* subsets used for validation





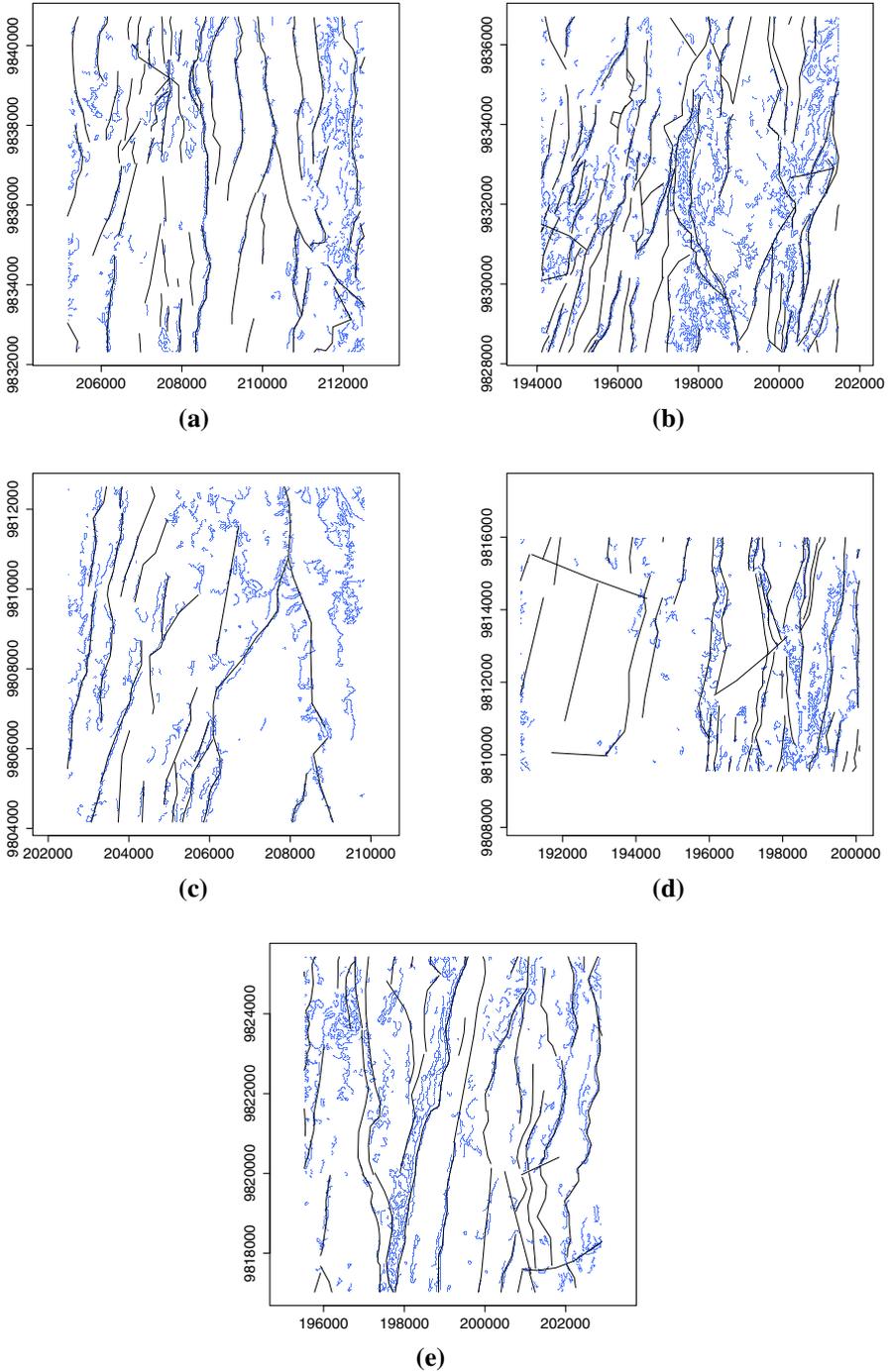
**Fig. 17** Detection results on the four validation subsets (1:4 from **a–d**). Different *colours* represent different segments. *Reference lines* are overlaid in *black*

**Table 5** Error ratios MR and FR for the tuning and four validation subsets

The last two columns present the respective error ratios produced by the application of the Canny algorithm

	Subset		Canny	
	MR	FR	MR	FR
Tuning	0.52	0.49	0.65	0.63
1	0.43	0.48	0.67	0.55
2	0.57	0.44	0.68	0.65
3	0.43	0.57	0.61	0.72
4	0.61	0.56	0.79	0.65

reference, which do not correspond to image edges, as explained in Sect. 4.2, leading to missing detections. For subset 2,  $MR > FR$  and  $MR = 0.57$ , being larger than 0.50 produced in the tuning subset. FR was smaller than its corresponding ratio in



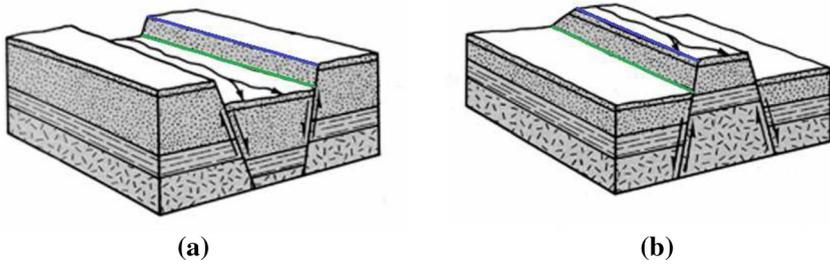
**Fig. 18** Detection results of the Canny algorithm on the four validation subsets (1:4 from **a–d**) and the tuning subset (**e**). *Reference lines* are overlaid in *black*

the tuning subset and was equal to  $FR = 0.44$ . This subset was characterized by the presence of many lineaments where discrimination was sometimes hindered by the presence of thick sediment cover at the surface, resulting in long breaks on the edge. False detections also occurred due to the presence of low vegetation. For subset 3,  $MR = 0.43$  and  $FR = 0.57$ . The relatively high ratio of false detections was explained by the presence of linear sequences of vegetation in the area. The presence of weak and noisy edges was responsible for missing detections. Some lineaments were crossing with other perpendicular lines of the landscape. For subset 4, both ratios were higher than their corresponding values in the tuning subset, being  $MR = 0.61$  and  $FR = 0.56$ , respectively. Missing detections corresponded to delineated lineaments that did not correspond to image edges. This phenomenon was also observed in lineaments of the secondary orientation WNW/ESE. It should be noted that these lineaments were not detected because they corresponded to weak edges of the image, with low gradient magnitude and not because the detection algorithm is restricted in specific orientations. These lineaments are more recent and were inferred only from the displacements they have caused in the NW lineaments. False detections reflected boundaries in the soil texture. Finally, the error rates were compared to the ones given by the application of the Canny algorithm (Canny 1986) on the tuning and four validation subsets. The results are shown in Fig. 18. The respective error ratios are presented in Table 5. For all five subsets both MR and FR were higher than the ones of our method.

## 5 Discussion

This paper explored the application of a line segment detection algorithm for the detection of geological linear forms, on a satellite image. Line detection algorithms based on the Helmholtz principle have mainly been tested so far in close-range images that contain man-made objects (Awrangjeb et al. 2010; Kit and Lüdeke 2013). The boundaries of such objects, however, consist mainly straight, clear and narrow lines. Natural objects appearing on images of moderate spatial resolution are different, containing often vague boundaries and lines. This makes the automatic detection of such features a challenging task.

The geometry of natural linear features on moderate-resolution images is distinctive. Geological lineament definitions encountered in the literature commonly speak of linear features (Clark and Wilson 1994; Hung et al. 2005; Hobbs 1904; O’Leary et al. 1976). When it comes to lineament interpretation from satellite images, this definition is not directly applicable. Depending upon the spatial resolution, the topography, the satellite viewing angle, the flight direction, the sun angle, as well as the type of lineament, identification of geological lineaments often appears adhered to a plane and not on only their presence (Fig. 19). For example, if the slope of the faulting is shallow and the viewing angle is almost perpendicular to the plane of the fault, the latter is projected and clearly visible on the image, whereas in the opposite case, where the fault scarp is not visible, that is on the other side of the graben, the slope may create a wide shadow that extends further from the fault trace. In both cases, an aerial feature is perceived from the image that stands out from the background landscape. It is then hard to distinguish the fault scarp from the fault trace, being the



**Fig. 19** **a** Graben structure, **b** Horst structure. *Green lines* indicate the lineament. The plane between the *green and blue lines* is often visible on the image and detected as such. Background image reprinted from [ERS ERI \(2014\)](#)

location where the crusting happens. In a manual interpretation, expert knowledge helps to overcome those ambiguities and to delineate the geological feature on its accurate position, although interpretation is still dependent on those conditions. In an automatic detection, however, detected segments deviate from lines and are closer to polygons. The method presented in this paper allowed to detect lineaments in the form of segments, thus capturing their geometry.

For the calibration of the segment detection algorithm a novel accuracy assessment method was applied. Scale factor  $S$  was tuned first, since it controls the spatial resolution of the image on which detection is carried out. Second came the angle tolerance  $\tau$ , being the main parameter of region growing. Threshold  $\varepsilon$  was tuned last, being the final parameter that distinguishes  $\varepsilon$ -meaningful segments from all detected segments. The defined ratios of missing and false detections allowed the tuning of the internal parameters. It was shown that image rescaling by a factor  $S$  is not only necessary to tackle quantization and aliasing artefacts. It is also related to the scale of the phenomenon to be mapped and helps to bring the image to such a scale that lineaments are detectable lines. The scale factor value  $S = 0.3$  was selected, in contrast to 0.8 in ([Grompone von Gioi et al. 2012](#)), showing that a coarser resolution than the original is most suitable for the automatic detection of the features of interest. The tuning of  $\tau$  pointed to the original value  $\tau = 22.5^\circ$  and showed that error ratios are stable in the range  $[22.5^\circ, 24^\circ]$ . For the parameter  $\varepsilon$  it has been claimed that the value  $\varepsilon = 1$  should be used universally ([Desolneux et al. 2001, 2003](#)). This study showed that deviations in the error ratios are relatively small for  $\varepsilon$  values in the range  $[1, 10]$ . Whereas the sum of the error ratios showed a clear minimum for  $S$  and  $\tau$ , fluctuations in the sum for different  $\varepsilon$  values were relatively small. Therefore, the value of  $\varepsilon$  was set at  $\varepsilon = 10$  by balancing the error ratios.

The restriction of using single parameter values was outreached by modelling the detected lineaments with random sets. The high values of set-theoretic variance showed that in the case of linear features, a stable core set is not formed. The highest values of the covering function were encountered on major geological structures. Segments whose parts close to the vertices had lower values of covering function, than the central pixels, indicated extensional uncertainty. Extensional uncertainty of aligned lineaments described by random sets could be potentially exploited for the connection of individual segments, and join parts that are covered by sediment or continued

underground, to form complete geological structures. The variance of major structures was high, while it was lower on smaller segments with low covering function values. For those segments, it is almost certain that they are not covered by the set. It was observed that the correctly detected lineaments correspond to high covering function values, showing that they are stable through the realizations. Further, the random set contained more segments than the result of fixed parameter values, including missing detections from the original result.

Accuracy assessment was done by applying the developed methodology to four subsets of the image. Missing detections occurred in lineaments identified by the expert, but without correspondence to an image edge. Most of these lineaments were connecting segments between apparent edges on the image, which were interpreted based upon experience, thus explaining the high error ratio  $MR = 0.5$ . The ratio of false detections mainly corresponded to the detection of other types of edges, such as soil texture change and vegetation. From the magnitude of the final error ratios it was concluded that the presented methodology does not produce a complete lineament detection. Yet, the final error ratios described the coincidence of the result with the given reference dataset and should not be considered as complete measures of the performance of the detection method. For the latter, the result should be validated against more interpretations or ground truth data. The presented methodology for lineament detection is a line segment detection method that does not incorporate elements of human geological experience, for the detection of lineaments that are obscured on the image. To avoid false detections, it is suggested that irrelevant linear features of the image are masked out before the application of the detection algorithm. To eliminate missing detections, it is suggested to further explore the random set result for the establishment of connectivity rules between individual segments.

Assessment of automatic detection of lineaments, however, is usually based on visual examination of the result (Jordan and Schott 2005; Koike et al. 1995) or histogram of the lineament orientations (Wladis 1999). No solid validation method is encountered in literature for the assessment of lineament automatic detection. This study, for the first time, proposed a numerical method that quantified error rates, against a reference. The search of matching points between reference and detection, for the error ratio calculation, involved thresholds for distance and angular difference in orientation. Choice of the threshold values was based on the nature and the geometry of the detected segments. Threshold values can be modified, depending on the required positional and directional accuracy of the extracted segments. The threshold value for  $\omega$  is related to the radiometric calibration of the VNIR ASTER bands and is expected to be different if other bands of the same sensor or different sensor images are used.

The group of authors has explored various ways to upscale the method, so as to be applicable over the whole scene. To do so, availability of small-scale geological information, say in the form of a well-interpreted geological map, is indispensable, as it was found that some sub-areas (subset 1) are providing high precision output, whereas other sub-areas (subsets 2, 3 and 4) provide information that is far less precise. The process should then be that sub-areas are identified that likely contain a large set of lineament features, and separate such sub-areas from areas that do not contain many lineament features, where the focus of the identification will be on the first group of

sub-areas. Such a map, however, is apparently not available, and creating it would be outside the scope of the current paper that has a focus on identifying lineaments from remote sensing images and relating those to geological processes. Therefore, this is left for future research. The potential of the method should be explored for other remotely sensed datasets. Elevation data, such as ASTER DEM, or derivatives, such as slope and aspect images, can be exploited. Further, LIDAR and SAR images can also be considered as input for the geological lineament detection method.

## 6 Conclusions

In this paper, the calibration and application of a Gestalt theory-based line segment detection method for geological lineament detection have been presented. Its application to a study area in Kenya, i.e. a tuning subset, showed that it was successful as it partially captured existing lineaments in the area. Starting with estimating the scale parameter  $S$ , followed by the angle tolerance  $\tau$  and finally the threshold on the false alarm rate  $\varepsilon$ , it was found that values  $S = 0.3$ ,  $\tau = 22.5^\circ$  and  $\varepsilon = 10$  provided the best results. Using probability distributions of these parameters, lineaments were detected as two-dimensional objects including their uncertainties. Uncertainty remained high throughout, as a core set was missing for all the segmented objects. The relationship between a geological lineament and its observation from the remotely sensed image was explored using four subsets from the same image. Error rates of these subsets were in line with those of the tuning subsets and their interpretation explained the occurrence of missing and false detections. Thus, the study showed that the approach is interesting and novel, with lineaments and their uncertainties identified by means of random sets. Improvements of the methodology are still possible, in particular to reduce the false detection rate.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Achauer U, Masson F (2002) Seismic tomography of continental rifts revisited: from relative to absolute heterogeneities. *Tectonophysics* 358:17–37
- Awrangjeb M, Ravanbakhsh M, Fraser CS (2010) Automatic detection of residential buildings using LIDAR data and multispectral imagery. *ISPRS J Photogramm Remote Sens* 65(5):457–467
- Batschelet E (1981) Circular statistics in biology, vol 371. Academic Press, London
- Burns J, Hanson AR, Riseman EM (1986) Extracting straight lines. *IEEE Trans Pattern Anal Mach Intell* PAMI-8(4):425–455
- Canny J (1986) A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell* 6:679–698
- Chandrasiri Ekneligoda T, Henkel H (2010) Interactive spatial analysis of lineaments. *Comput Geosci* 36(8):1081–1090
- Clark C, Wilson C (1994) Spatial analysis of lineaments. *Comput Geosci* 20(7–8):1237–1258
- Davis JC, Sampson RJ (2002) Statistics and data analysis in geology, vol 3. Wiley, New York
- Desolneux A, Moisan L, Morel JM (2000) Meaningful alignments. *Int J Comput Vis* 40(1):7–23

- Desolneux A, Moisan L, Morel JM (2001) Edge detection by Helmholtz principle. *J Math Imaging Vis* 14(3):271–284
- Desolneux A, Moisan L, Morel JM (2003) Computational gestalts and perception thresholds. *J Physiol Paris* 97(2):311–324
- ERS ERI (2014) Sketch maps of extensional structures. <http://www.europlanet-ri.eu/>
- ERSDAC (2005) Aster user's guide ver.4.0
- Grompone von Gioi R, Jakubowicz J, Morel JM, Randall G (2012) LSD: a line segment detector. *Image Process On Line*. doi:[10.5201/ipol.2012.gjmr-lsd](https://doi.org/10.5201/ipol.2012.gjmr-lsd). <http://www.ipol.im/pub/art/2012/gjmr-lsd/>
- Glasbey CA, Horgan GW (1995) *Image analysis for the biological sciences*, vol 1. Wiley, Chichester
- Gómez H, Kavzoglu T (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Eng Geol* 78:11–27
- Goodchild MF, Jeansoulin R (1998) Statistical representation of relative positional uncertainty for geographical linear features. In: *Data quality in geographic information: from error to uncertainty*. Hermes, Paris pp 87–96
- Guru D, Shekar B, Nagabhushan P (2004) A simple and robust line detection algorithm based on small eigenvalue analysis. *Pattern Recognit Lett* 25(1):1–13. doi:[10.1016/j.patrec.2003.08.007](https://doi.org/10.1016/j.patrec.2003.08.007)
- Hashim M, Ahmad S, Johari MAM, Pour AB (2013) Automatic lineament extraction in a heavily vegetated region using Landsat Enhanced Thematic Mapper (ETM+) imagery. *Adv Space Res* 51(5):874–890
- Heipke C, Steger C, Multhammer R (1995) A hierarchical approach to automatic road extraction from aerial imagery. In: *Proceedings of the SPIE The International Society for Optical Engineering*, p 222
- Hobbs WH (1904) Lineaments of the Atlantic border region. *Geol Soc Am Bull* 15:480–506
- Hung L, Batelaan O, De Smedt F (2005) Lineament extraction and analysis, comparison of LANDSAT ETM and ASTER imagery. Case study: Suoimuoi tropical karst catchment, Vietnam, vol 5983
- Jordan G, Schott B (2005) Application of wavelet analysis to the study of spatial pattern of morphotectonic lineaments in digital terrain models. A case study. *Remote Sens Environ* 94(1):31–38
- Juneja M, Sandhu PS (2009) Performance evaluation of edge detection techniques for images in spatial domain. *Methodology* 1(5):614–621
- Khomyakov M (2012) Comparative evaluation of linear edge detection methods. *Pattern Recognit Image Anal* 22(2):291–302
- Kit O, Lüdeke M (2013) Automated detection of slum area change in Hyderabad, India using multitemporal satellite imagery. *ISPRS J Photogramm Remote Sens* 83:130–137
- Koike K, Nagano S, Ohmi M (1995) Lineament analysis of satellite images using a segment tracing algorithm (STA). *Comput Geosci* 21(9):1091–1104
- Lee T, Moon W (2002) Lineament extraction from Landsat TM, JERS-1 SAR, and DEM for geological applications, vol 6, pp 3276–3278
- Mardia KV, Jupp PE (2009) *Directional statistics*, vol 494. Wiley, New York
- Marghany M, Hashim M (2010) Lineament mapping using multispectral remote sensing satellite data. *Int J Phys Sci* 5(10):1501–1507
- Marghany M, Mansor S, Hashim M (2009) Geologic mapping of United Arab Emirates using multispectral remotely sensed data. *Am J Eng Appl Sci* 2(2):476
- Mena JB, Malpica JA (2005) An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recognit Lett* 26(9):1201–1220
- Molchanov I (2005) Expectations of random sets. In: *Theory of random sets, probability and its applications*. doi:[10.1007/1-84628-150-4\\_2](https://doi.org/10.1007/1-84628-150-4_2)
- Mostafa ME, Bishta AZ (2005) Significance of lineament patterns in rock unit classification and designation: a pilot study on the Gharib-Dara area, Northern Eastern Desert, Egypt. *Int J Remote Sens* 26(7):1463–1475
- O'Leary DW, Friedman JD, Pohn HA (1976) Lineament, linear, lineation: some proposed new standards for old terms. *Geol Soc Am Bull* 87(10):1463–1469. doi:[10.1130/0016-7606\(1976\)87<1463:LLSPN>2.0.CO;2](https://doi.org/10.1130/0016-7606(1976)87<1463:LLSPN>2.0.CO;2)
- Pal S, Majumdar T, Bhattacharya A (2006) Extraction of linear and anomalous features using ERS SAR data over Singhbhum Shear Zone, Jharkhand using fast Fourier transform. *Int J Remote Sens* 27(20):4513–4528
- Papazachos B, Scordilis E, Panagiotopoulos D, Papazachos C, Karakaisis G (2004) Global relations between seismic fault parameters and moment magnitude of earthquakes. *Bull Geol Soc Greece* 36:1482–1489
- Ramli M, Yusof N, Yusoff M, Juahir H, Shafri H (2010) Lineament mapping and its application in landslide hazard assessment: a review. *Bull Eng Geol Environ* 69(2):215–233. doi:[10.1007/s10064-009-0255-5](https://doi.org/10.1007/s10064-009-0255-5)

- Singhal B, Gupta R (2010) Fractures and discontinuities. In: Applied hydrogeology of fractured rocks. doi:[10.1007/978-90-481-8799-7-2](https://doi.org/10.1007/978-90-481-8799-7-2)
- Solomon S, Ghebreab W (2006) Lineament characterization and their tectonic significance using Landsat TM data and field studies in the central highlands of Eritrea. *J Afr Earth Sci* 46(4):371–378
- Soto-Pinto C, Arellano-Baeza A, Sánchez G (2013) A new code for automatic detection and analysis of the lineament patterns for geophysical and geological purposes (ADALGEO). *Comput Geosci* 57(0):93–103. doi:[10.1016/j.cageo.2013.03.019](https://doi.org/10.1016/j.cageo.2013.03.019)
- Thornton M, Atkinson PM, Holland D (2007) A linearised pixel-swapping method for mapping rural linear land cover features from fine spatial resolution remotely sensed imagery. *Comput Geosci* 33(10):1261–1272
- Turker M, Kok EH (2013) Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS J Photogramm Remote Sens* 79:106–121
- Wang J, Howarth P (1990) Use of the Hough transform in automated lineament. *IEEE Trans Geosci Remote Sens* 28(4):561–567. doi:[10.1109/TGRS.1990.572949](https://doi.org/10.1109/TGRS.1990.572949)
- Wladis D (1999) Automatic lineament detection using digital elevation models with second derivative filters. *Photogramm Eng Remote Sens* 65:453–458
- Zhao X, Stein A, Chen X, Zhang X (2011) Quantification of extensional uncertainty of segmented image objects by random sets. *IEEE Trans Geosci Remote Sens* 49(7):2548–2557. doi:[10.1109/TGRS.2011.2109064](https://doi.org/10.1109/TGRS.2011.2109064)
- Ziou D, Tabbone S (1998) Edge detection techniques—an overview. *Int J Pattern Recognit Image Anal* 8:537–559