Arab J Geosci
DOI 10.1007/s12517-013-1043-x

ORIGINAL PAPER

Exploring the association between the occurrence of earthquakes and the geologic-tectonic variables in the Red Sea using logistic regression and GIS

Khalid Al-Ahmadi · Sharaf Al-Ahmadi · Abdullah Al-Amri

Received: 23 February 2013 / Accepted: 19 July 2013

Abstract Al-Ahmadi et al. (Arab J Geosci doi:10.1007/s12517-013-0974-6, 2013) applied spatial pattern analysis techniques to a seismic data catalogue of earthquakes beneath the Red Sea in order to explore and detect global and local spatial patterns in the occurrence of earthquakes over the years from 1900 to 2009 using a geographical information system (GIS). They found that the techniques of spatial pattern analysis that they applied could detect global and local clusters and broader spatial patterns in the occurrence of earthquakes and concluded that earthquakes with higher magnitudes were notably concentrated beneath the central and southern areas of the Red Sea, while earthquakes with low and moderate magnitudes were concentrated beneath the northern area of the Red Sea. The aim of this paper is to report on the application of logistic regression models to explore the associations between the likelihood of the occurrence of an earthquake beneath the Red Sea and four selected variables, namely: (1) proximity to the boundary of the African–Arabian plates, (2) proximity to transform faults, (3) proximity to the mid-Red Sea ridge and (4) the stage of rift evolution. The study was undertaken to evaluate the potential of logistic regression modelling for research exploring potential associations between earthquakes and geological and tectonic variables. The results revealed that none of the assumptions underpinning the logistic regression models had been violated for the three logistic regression models that were used in this research. The authors inferred that the occurrence of the earthquakes beneath the Red Sea was statistically significantly associated with the proximity to the African–Arabian plate boundary. We concluded that earthquakes of moderate magnitudes occurred in the zone which represents the late evolutionary stage of the Red Sea rift, including the transition zone beneath the central area and the late-stage continental rift zone beneath the northern area of the Red Sea. In contrast, earthquakes with high magnitudes tended to occur in close proximity to the mid-ridges of the Red Sea.

Keywords Earthquakes · GIS · Logistic regression · Red Sea

Introduction

Many earthquakes are caused by rupture of geological faults, volcanic activity or landslides, but some are caused by human activities, including mining blasts and nuclear tests. An earthquake is any seismic event which generates seismic waves (Bolt 2003). Tectonics concerns the formation and structure of the earth's lithosphere. The theory of plate tectonics is that the outer rigid layer of the Earth—its lithosphere—is divided into plates which move over the Earth's surface relative to each other and interact in diverse ways (Kusky 2008). A tectonic plate boundary is a place where tectonic plates touch. The boundaries between these plates form zones where plates converge, diverge and slip past one another (Tarbuck et al. 2005). Earthquakes are the result of movement of the Earth's tectonic plates and are driven by convection currents. Although one might imagine that earthquakes would be randomly distributed over the Earth's crust, they are mostly found in the zones at the boundaries of tectonic plates where tension, compression and shearing forces change continually and the earth's rigid outer lithosphere is relatively thin.

K. Al-Ahmadi (✉)
King Abdulaziz City for Science and Technology,
Riyadh, Saudi Arabia
e-mail: alahmadi@kacst.edu.sa

S. Al-Ahmadi
Taibah University, Madinah, Saudi Arabia

A. Al-Amri
King Saud University, Riyadh, Saudi Arabia

When two tectonic plates separate, a new sea-floor forms at divergent boundaries. Mid-ocean ridges are the most common type of divergent boundary (Tarbuck et al. 2005). The focus of the current study is the area beneath the Red Sea in the Middle East, where the boundaries between the African and Arabian tectonic plates are divergent and the plates are relatively thin and weak. The boundaries between the plates lie along the long axis of the Red Sea, where there are mid-Red Sea spreading ridges. The Red Sea rift is a spreading centre between the African and the Arabian plates (Campbell et al. 2011). Continual dispersal further separates the continents until they split away from each other and a narrow way progresses. The Red Sea, which splits the Arabian plate from the Africa plate, is a good illustration of this development phase (Kusky 2008). The rock has not cooled completely, so it is still somewhat flexible; thus, large strains and stresses cannot develop and most earthquakes near spreading ridges are relatively shallow. Beneath the Red Sea, there are active faults which are related to earthquakes and associated seismic waves resulting in strong ground motion, surface faulting, tectonic deformation, landslides, liquefaction, tsunamis and seiches. Earthquakes are located along the faults which form the sides of the rift or beneath the floor of the rift. Divergent faults and rift valleys within the continental plates also experience shallow earthquakes. Plate tectonic models for the Earth's crust show that most active faults occur near the boundaries of tectonic plates. Fault length, displacement and geographical extent with the magnitude of earthquakes can be correlated to assess the hazards of faults and form the principal inputs for some types of risk analysis (Slemmons and Depolo 1986).

Although plate tectonics is an unrefined approach, it is a robust tool in earthquake prediction. It advises where the large proportion of the Earth's major earthquakes are likely to occur, although it cannot tell us much about exactly when they will occur, and thus it is a guide to the use of finer techniques for earthquake prediction (Spall 2003). Potential explanatory variables that can serve as additional information (covariates) in modelling and understanding seismic activities include plate tectonic boundaries and geological faults—either normal faulting or strike-slip faulting (Sykes 1967; Tatham and Savino 1974; Cinti et al. 2004; Okamura 2009; Stein et al. 2010). These two variables can be regarded as significant determinants for fixing the location of an earthquake epicentre. Empirical estimation models that apply statistical techniques can be applied to model the associations between the probability of occurrence of earthquakes and factors (covariates) such as plate tectonic boundaries and geological faults based on historical data.

Empirical models use historical statistics of seismic activities to model the associations between the probability of an earthquake and factors or covariates such as proximity to plate tectonic boundaries, proximity to faults and other geological phenomena. The logistic regression model belongs to a family of generalized linear models. Using historical earthquake data catalogues, García-Rodríguez et al. (2008) applied a logistic regression model to assess the probability of earthquake-triggered landslides in El Salvador and Honkura and Tanaka (1996) applied a similar model to study seismicity and estimate the probability of an earthquake in Greece. Amindan and Hagedorn (1998) applied logistic regression models to predict the source of a seismic event such as an earthquake or man-made explosion. Using a cross-validation test, their results showed that the model correctly predicted 99.7 % of earthquakes and 98.0 % of explosions for the chosen set of historical earthquake data.

The aim of this paper is to report on the application of logistic regression models to explore the associations between the likelihood of an earthquake beneath the Red Sea and four selected variables, namely: (1) proximity to the boundary of the African–Arabian plates, (2) proximity to transform faults, (3) proximity to the mid-Red Sea ridge and (4) the stage of rift evolution. The study was undertaken to evaluate the potential of logistic regression modelling for research into understanding the location, magnitude and other characteristics of earthquakes beneath the Red Sea using historical earthquake data catalogues.

Methodology

Data processing

Seismic data for the Red Sea area were acquired from the Seismic Studies Centre at King Saud University, King Abdulaziz City for Science and Technology and the Saudi Geological Survey. The seismic data were divided into categories by magnitude of earthquakes and by time from the years 1900 to 2009. A detailed seismic data catalogue and a description of the data processing can be found in Al-Ahmadi et al. (2013). The seismic and other spatial data were processed using a Geographical Information System (GIS), namely ESRI ArcGIS 10.1. Logistic regression analysis was carried out using the statistical package SPSS version 20 because the software has functions for handling dichotomous dependent variables. The Fishnet function in the GIS was used to divide the area covered by the Red Sea into a net of square cells, each with a dimension of 25×25 km. A total of 823 squares covered the Red Sea. The proximity of the explanatory variables in this study was calculated using the Euclidean distance function in the GIS.

Figure 1a shows the plate boundary, the mid-ocean ridge of the Red Sea and the transform faults. The Red Sea is a divergent-type boundary between the African and Arabian plates; the mid-ocean ridge of the Red Sea demarcates the boundary between the African and Arabian tectonic plates and is categorized as a divergent plate boundary.

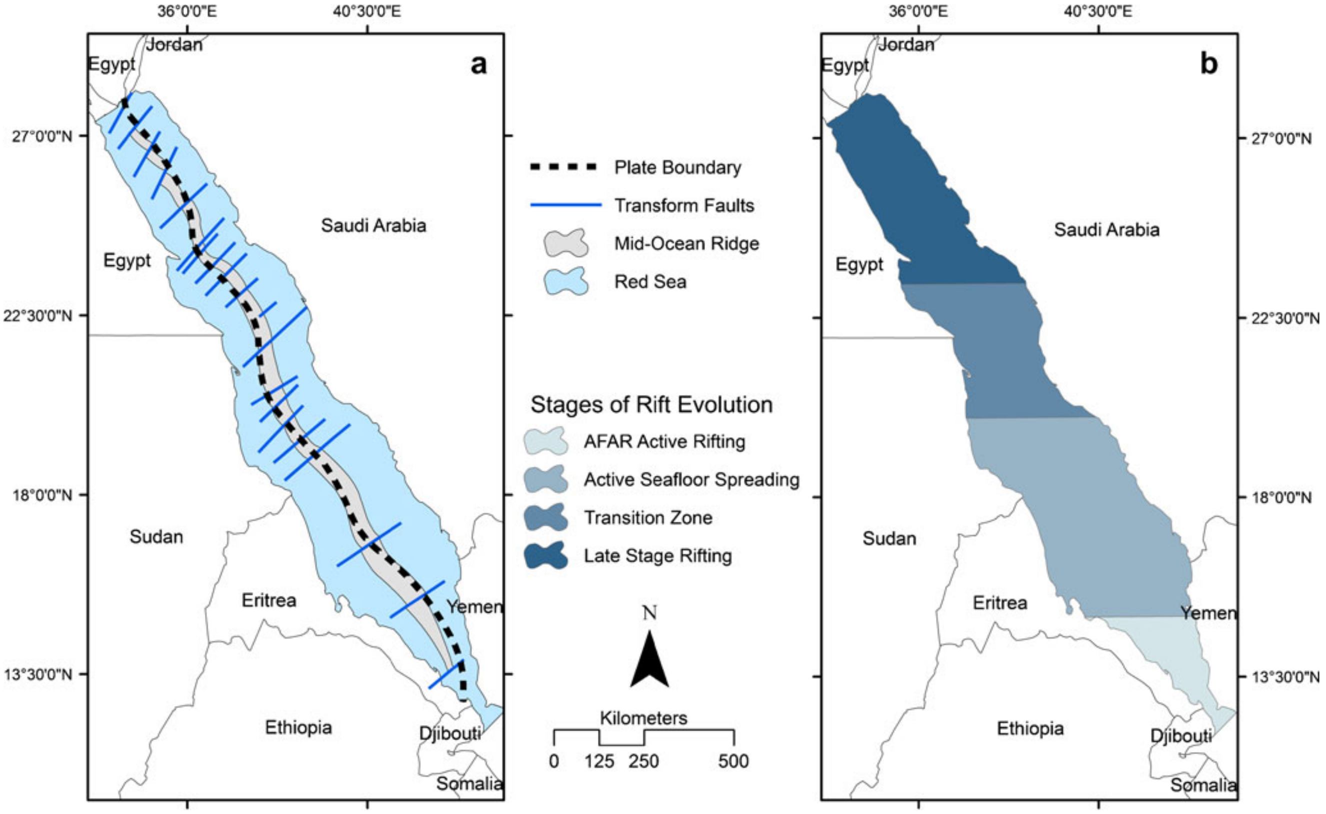

Fig. 1 **a** Plate boundary, transform faults and mid-ocean ridge and **b** stages of rift evolution

The Red Sea can be classified into four distinct zones with different morphological, structural and geophysical characteristics which represent different stages in the evolution of the rift (Fig. 1b). The four zones are (1) Active sea-floor spreading (southern Red Sea) located between 15° and 20 °N and characterized by a well-developed axial trough which has developed through normal sea-floor spreading during the last 5 Ma. (2) Transition zone (central Red Sea) located between 20° and 23°50′N, which contains isolated cells of sea-floor spreading and where the axial trough becomes discontinuous. (3) Late-stage continental rifting (northern Red Sea) located between 23°50′ and 28 °N, which is comprised of a wide trough without a recognizable spreading centre, although there are a number of small isolated deep troughs. This zone is currently in the late stages of continental rifting. (4) In the Afar depression in northern Ethiopia, the Earth's outermost shell, usually a relatively rigid, 150-km-thick plate, has been stretched, thinned and heated to the point of breaking. Hot, partially-melted rocks are rising up from the Earth's mantle and are either erupting at the Earth's surface or cooling just beneath it. Afar active rifting located between 12° and 15 °N (the line along which the southern Red Sea is expected to propagate through the Afar region in future) may be included in the second zone or considered separately (Cochran 1981; Cochran et al. 1986; Martinez and Cochran 1988; Ghebreab 1998; Hagos et al. 2006).

Logistic regression

Logistic regression belongs to a group of statistical models called generalized linear models. These models use independent variables to create a mathematical model which predicts the probability of an event occurring in a particular geographical area. Logistic regression predicts the probability that an observation will fall into one of the two categories of a dichotomous dependent variable based on single or multiple independent variables, which can be continuous or categorical. The basis of logistic regression is that the dependent variable is dichotomous. This means that it can take only the value 1 or 0 (Hair et al. 1998; García-Rodríguez et al. 2008), representing in this study the presence or absence of an earthquake. In many ways, logistic regression is similar to linear regression except that the dependent variable is dichotomous. Linear regression predicts the value of the dependent variable. In contrast, logistic regression estimates the probability of the dependent variable being in category 1 or 0 given the independent variables. An observation is assigned to whichever category is predicted as being most likely.

In the case of predicting the occurrence of an earthquake, the aim of applying logistic regression will be to discover the model which best fits for exploring the association of the presence or absence of an earthquake Y, the dependent variable that is 0 or 1 for its two likely groups, with a group of independent variables $X_1, X_2, \ldots, X_n$, such as proximity to the African–Arabian plate, proximity to transform faults and proximity to ridge and stage-rifting zones. The mathematical modelling approach known as logistic regression defines the expected value of Y in terms of the following expression (Field 2000):

$$P(Y) = \frac{1}{1 + e^{-z}}$$
$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon_i \quad (1)$$

Where $P(Y)$ is the probability that a case is in a particular category, e is the base of natural logarithms (about 2.72), Z is the linear equation of the model, B_0 is the constant of the equation, B is the coefficient of the predictor variables, X is the independent variables and ε is the random error.

The logits (log odds) are the B coefficients (the slope values) of the regression equation. The slope can be interpreted as the change in the average value of Y from one unit of change in X. Logistic regression calculates changes in the log odds of the dependent value, not changes in the dependent value as ordinary least square (OLS) regression does. For a dichotomous variable, the odds of membership of the target group are equal to the probability of membership in the target group divided by the probability of membership in the other group. The odds value can range from 0 to infinity and tells you how much more likely it is that an observation is a member of the target group rather than a member of the other group. For example, if the probability is 0.80, the odds are 4 to 1 or 0.80/0.20 (Burns and Burns 2008).

Logistic regression estimates the probability of an event (in this case, an earthquake) occurring. If the estimated probability of the event occurring is greater than 0.5 (better than an even chance), then the event is classified as occurring (e.g. earthquake being present). If the probability is less than 0.5, then the event is classified as not occurring (e.g. no earthquake). The overall percentage of the classification measure is usually used to assess the effectiveness of the predicted classification against the actual classification. The odds ratio presents the extent to which raising the corresponding measure by one unit influences the odds ratio. If the value of the odds ratio exceeds 1, then the odds of an outcome occurring increase; if the figure is less than 1, any increase in the predictor leads to a decrease in the odds of the outcome occurring. Odds ratios in logistic regression can be interpreted as the effect of one unit of change in X of the predicted odds ratio with the other variables in the model held constant (Burns and Burns 2008).

Logistic regression does not have as many assumptions as linear regression but still carries some, such as (1) Independence of cases/errors, which can be assessed by calculating the *Durbin–Watson statistic* where the acceptable values fall between 1 and 3. (2) A linear relationship between the continuous independent variables and the logit transformation of the dependent variable, which can be examined using the Box-Tidwell procedure (Box and Tidwell 1962), which adds a new interaction term for each continuous independent variable. These interaction terms are between the existing continuous independent variable and their natural log transformations. After performing Bonferroni correction, if the interaction term is statistically significant, then the original independent variable is not linearly related to the logit of the dependent variable. (3) No multi-collinearity: this was assessed in the present study using three measures: (a) correlation coefficient matrix, where a correlation of 0.9 or higher indicates multi-collinearity; (b) the Variance Inflation Factor (VIF), which measures how much the variance of the estimated coefficients are increased over the case of no correlation among the independent variables, with values less than 10 being acceptable and (c) the Tolerance, which is the reciprocal of the VIF where values larger than 0.1 indicate no multi-collinearity. (4) No significant outliers or influential points, the former assessed by examining the cases' standardized residuals and by measuring the Leverages value, which was in turn evaluated using Cook's Distance. Cases' standardized residuals less than ±3 and Leverages value above the safe value of 0.2 indicate no significant outliers. Cook's Distance values above 1 indicate no highly influential points (Field 2000).

One method of measuring the goodness of fit of the logistic regression model is to examine how poor the model is at predicting the categorical outcomes. This can be tested using the Hosmer and Lemeshow goodness of fit test. For this test, statistically significant results indicate a poorly fitting model (Hosmer and Lemeshow 2000). The Hosmer–Lemeshow test statistic is given by:

$$H = \sum_{g=1}^{n} \frac{\left(O_g + E_g\right)^2}{N_g \pi_g \left(1 - \pi_g\right)} \quad (2)$$

Here, O_g, E_g, N_g and π_g denote the observed events, expected events, observations and predicted risk for the g^{th} risk decile group, and n is the number of groups. The test statistic asymptotically follows an χ^2 distribution with $n=2$ degrees of freedom. The number of groups may be adjusted depending on how many are determined by the model.

Logistic regression does not have an equivalent to the R^2 which is found in OLS regression. Other methods are suggested by scientists for calculating the explained variation, referred to as pseudo R^2 values, such as the Cox and Snell R^2 and Nagelkerke R^2 values. These two measures normally have lower values than the conventional R^2 in OLS regression but are interpreted in the same manner but with more caution. Nagelkerke R^2 is a modification of the Cox and Snell R^2, which cannot have a value of 1. For this reason, it is better to report the Nagelkerke R^2 value (Field 2000). The Nagelkerke R^2 can be calculated as follows:

$$f(x) = \frac{1-\exp\left\{\frac{-2(LL_m - LL_o)}{N}\right\}}{1-\exp\left\{\frac{2LL_o}{N}\right\}} \tag{3}$$

where LL_m and LLo are the log-likelihood of the model and intercept (the model without any explaining variable), respectively, and N is the sample size. Pseudo R^2 values such as the Nagelkerke R^2 that fall between 20 and 40 % can be considered a good fit (Clark and Hosking 1986; Domencich and McFadden 1975).

Results

Logistic regression analysis

A logistic regression was performed using the *Enter* method to ascertain the effects of proximity to the African–Arabian plate boundary (Plate: Fig. 2a), proximity to transform faults (FaultS: Fig. 2b), proximity to mid-ocean ridges (Ridges: Fig. 2c) and the stages of the Red Sea evolution rift (Evolution: Fig. 2d) on the likelihood that a location beneath the Red Sea has a potential for earthquake occurrence. The first three factors were computed using the Euclidian function in a GIS and are continuous variables. As for the Evolution factor, the four zones of the Red Sea evolution rift which were described earlier and are shown in Fig. 2b were aggregated into two zones, namely the early evolution stage (which includes active sea-floor spreading and Afar active rifting) and the late evolution stage (which includes the transition zone and late-stage continental rifting) as shown in Fig. 1d. This aggregation reduces the complexity of interpretation when using multiple categorical data in logistic regression.

As mentioned earlier, the Red Sea was divided into 823 squares where each square grid has a dimension of 25×25 km. Each grid was coded as 0 or 1, representing the presence or absence of earthquakes over the years from 1900 to 2009. A requirement of logistic regression is that the dependent variable must be dichotomous. This means that the variable can only take a value of 1 or 0. Such a variable can be used to represent the presence or absence of the occurrence of earthquakes in a particular square. It cannot, however, be used to represent the associated magnitude of an earthquake, at least not in a straightforward way. To circumvent this problem, the association between the four factors of Plate, Faults, Ridges and Evolution and the likelihood of the occurrence of an earthquake beneath the Red Sea was explored by considering three scenarios:

1. 'All earthquakes' will apply the presence or absence of all earthquakes which occurred beneath the Red Sea during the years from 1900 to 2009 regardless of their magnitude. Each of the 823 squares covering the Red Sea was coded as 1 or 0 for the presence or absence of an earthquake.
2. 'Moderate earthquakes' will apply the presence or absence of occurrences of all earthquakes beneath the Red Sea that had associated magnitudes of $3 \leq M \leq 4$, which account for 71 % of the total earthquakes that occurred during the years from 1900 to 2009.
3. 'Major earthquakes' will apply the presence or absence of occurrence of all earthquakes beneath the Red Sea that had associated magnitudes of $4 < M \leq 7$, which account for 29 % of the total earthquakes that occurred during the years from 1900 to 2009.

These three scenarios comply with the fundamental requirement of having a dichotomous dependent variable for logistic regression and were considered to be a neat way of discovering the effects of the four factors on the occurrence of all earthquakes (regardless of their magnitudes), earthquakes with moderate magnitudes and earthquakes with major magnitudes.

Examining the assumptions of the logistic regression models

The assumptions of the three logistic regression models above were studied and assessed, including the independence of cases/errors, a linear relationship between the continuous independent variables and the logit transformation of the dependent variable, the absence of multi-collinearity and the absence of significant outliers or influential points.

The results indicated that there was independence of errors in the three models (scenarios), as assessed by the Durbin–Watson statistic, with values of 1.55 (all earthquakes), 1.52 (moderate earthquakes) and 1.55 (major earthquakes). By applying the Bonferroni correction, the interaction terms were not statistically significant at 0.18–0.904 (all earthquakes), 0.034–0.552 (moderate earthquakes) and 0.35–0.63 (major earthquakes), indicating that the independent variables were linearly related to the logit of the dependent variable. Also, there was no multi-collinearity among the independent variables in the three models, as the VIF values were 1.27–4.02 (all earthquakes), 1.27–4.02 (moderate earthquakes) and 1.27–4.02 (major earthquakes); the Tolerance values were 0.26–0.78 (all earthquakes), 0.25–0.78 (moderate earthquakes) and 0.26–

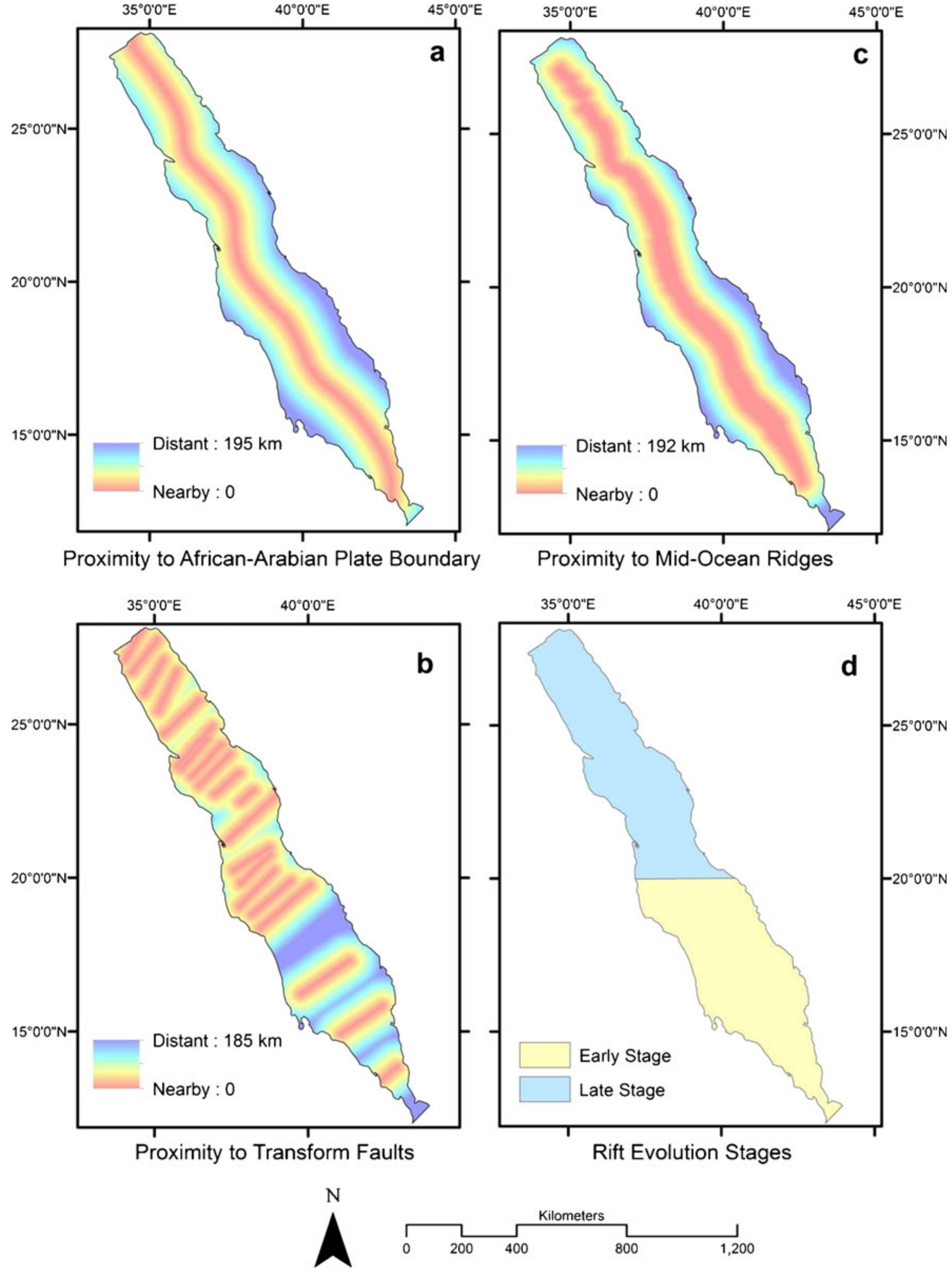

Fig. 2 **a** Proximity to the African-Arabian plate boundary, **b** proximity to transform faults, **c** proximity to mid-ocean ridges and **d** rift evolution stages

0.78 (major earthquakes) and the correlation coefficients r were 0.018–0.45 (all earthquakes), 00.010–0.543 (moderate earthquakes) and 0.018–0.45 (major earthquakes). The only exception was the correlation found between the Plate and Ridges variables, where r=0.85 in all scenarios. There were no significant outliers or influential points. In the three

Table 1 Logistic regression predicting likelihood of 'all earthquakes' scenario

Variables	*B*	SE	Wald	DF	Sig	Odds ratio	95 % CI for odd ratio	
							Lower	Upper
Plate	−0.000015	0.000	18.964	1	0.000	0.999985	0.999978	0.999992
Faults	−0.000005	0.000	3.862	1	0.049	0.999995	0.99998	1.000
Ridges	−0.000007	0.000	3.060	1	0.080	0.999993	0.999986	1.000
Evolution	−0.490	0.176	7.764	1	0.005	0.613	0.434	0.865
Constant	1.39	0.207	45.350	1	0.00	4.034		

Evolution is for early stage compared to late stage

scenarios, all cases had standardized residuals of less than ±3 and there were no Leverage values above the safe value of 0.2, indicating no significant outliers. Further, there were no Cook's Distance values above 1, indicating no highly influential points. Therefore, none of the assumptions had been violated for the three logistic regression models.

All earthquakes scenario

For the 'all earthquakes' scenario $3 \leq M \leq 7$, the logistic regression model was statistically significant, where the model chi-square statistic was $\chi^2(4)=148.703$, $p<0.05$ and −2LL (log-likelihood)=969.306. The model explained 22.2 % (Nagelkerke R^2) of the variance in earthquake occurrence, indicating a good fit. The model correctly classified 70.05 % of the cases. Hosmer and Lemeshow's goodness of fit test was not statistically significant ($\chi^2(4)=11.27$, $p=0.187$), indicating that the model is not a poor fit. The Wald test was used to determine the statistical significance for each of the independent factors (Table 1) and results shows that Plate ($p<0.05$), Faults ($p<0.05$) and Evolution ($p<0.05$) added significantly to the model/prediction but Ridges ($p=0.080$) did not.

Table 1 shows the estimated coefficients and odds ratios for the logistic regression model of the 'all earthquakes' scenario. The odds ratio for the Plate variable was 0.999993 (95 % confidence interval (CI), 0.999978–0.999992) and for the Faults variable was 0.999995 (95 % CI, 0.999989–1.0), indicating that an increase of 1 km in the proximity to the African–Arabian plates or proximity to transform faults was associated with a decreased likelihood of the occurrence of the 'all earthquakes' scenario. The odds ratio for the Evolution variable was 0.613 (95 % CI, 0.434–0.865), indicating that the occurrence of all earthquakes in the late-stage evolution zone was 0.613 times more likely than in the early stage evolution zone.

Moderate earthquakes scenario

For the 'moderate earthquakes' scenario $3 \leq M \leq 4$, the logistic regression model was statistically significant, where the model chi-square statistic was $\chi^2(4)=148.703$, $p<0.05$ and −2LL= 969.306. The model explained 15 % (Nagelkerke R^2) of the variance in earthquake occurrence, indicating a weak fit because 20–40 % indicates a good fit. The model correctly classified 69.4 % of the cases. The Hosmer and Lemeshow goodness of fit test was statistically significant ($\chi^2(4)=30.86$, $p<0.05$), indicating a weak fit of the model. By applying the statistical significance of the Wald test (Table 2), the authors found that only PLATE ($p<0.05$) added significantly to the model/prediction, while Evolution ($p=0.777$), Faults ($p=0.101$) and Ridges ($p=0.330$) did not.

Table 2 shows the estimated coefficients and odds ratios for the logistic regression model of the 'moderate earthquakes' scenario. The odds ratio for the Plate variable was 0.999987 (95 % CI, 0.999980–0.999994), indicating that an increase of 1 km in proximity to the African–Arabian plates was related to a decreased likelihood of the occurrence of the 'moderate earthquakes' scenario.

Major earthquakes scenario

For the 'major earthquake' scenario of $4<M \leq 7$, the logistic regression model was statistically significant, where the model

Table 2 Logistic regression predicting likelihood of 'moderate earthquakes' scenario

Variables	*B*	S.E.	Wald	DF	Sig.	Odd Ratio	95 % CI for Odd Ratio	
							Lower	Upper
Plate	−0.000013	0.000	14.400	1	0.000	0.999987	0.999980	0.999994
Faults	−0.000005	0.000	2.696	1	0.101	0.999995	0.999990	1.000
Ridges	−0.000004	0.000	0.947	1	0.330	0.999996	0.999989	1.000
Evolution	−0.049420	0.174	0.080	1	0.777	0.952	0.676	1.340
Constant	0.549495	0.198	7.669	1	0.006	1.732		

Evolution is for early stage compared to late stage

Table 3 Logistic regression predicting likelihood of the ‘major earthquakes’ scenario

Variables	*B*	SE	Wald	DF	Sig	Odd ratio	95 % CI for odd ratio	
							Lower	Upper
Plate	−0.000019	0.000	20.449	1	0.000	0.999981	0.999972	0.999989
Faults	0.000	0.000	0.011	1	0.917	1.000	0.999994	1.000000
Ridges	−0.000010	0.000	4.747	1	0.029	0.999990	0.999981	0.999999
Evolution	−1.134428	0.216	27.481	1	0.000	0.322	0.210	0.492
Constant	0.680	0.226	9.037	1	0.003	1.974		

Evolution is for early stage compared to late stage

chi-square statistic was $\chi^2(4)=160, p<0.05$ and $-2LL=701.485$. The model explained 27.30 % (Nagelkerke R^2) of the variance in earthquake occurrence, indicating a good fit. The model correctly classified 82.6 % of the cases. The Hosmer and Lemeshow goodness of fit test was statistically significant ($\chi^2(4)=20.50$, $p<0.05$), indicating a weak fit of the model. By applying the statistical significance of the Wald test (Table 3), the authors found that Plate ($p<0.05$), Evolution ($p<0.05$) and Ridges ($p<0.05$) added significantly to the model/prediction but Faults ($p=0.917$) did not.

Table 3 shows the estimated coefficients and odds ratios for the logistic regression model of the ‘major earthquakes’ scenario. The odds ratio for the Plate variable was 0.999981 (95 % CI, 0.9999720.999989) and for Ridges was 0.999990 (95 % CI, 0.9999810.999999), indicating that a 1 km increase in proximity to the African–Arabian plates or proximity to ridges was associated with a reduction in the likelihood of the occurrence of a major earthquake. The odds ratio for the Evolution variable was 0.322 (95 % CI, 0.210–0.492), indicating that the occurrence of a major earthquake in the late stage evolution zone was 0.322 times more likely than in the early stage evolution zone.

Discussion and conclusions

The aims of this research was to explore the associations between the likelihood of the occurrence of an earthquake and four explanatory variables, namely proximity to the African–Arabian plate boundary, proximity to transform faults, proximity to mid-ocean ridges and stages of Red Sea rifting evolution using a logistic regression model.

According to Al-Ahmadi et al. (2013), both global and local spatial statistics indicated that earthquakes were clustered beneath the Red Sea. In general, earthquakes with higher magnitudes were concentrated notably in the central and southernmost parts of the Red Sea, the most active zone in terms of seismic activities. Unlike the central and southern parts, the northern part of the Red Sea has had a remarkably large number of earthquakes with low and moderate magnitudes.

The results revealed that none of the assumptions had been violated for the three logistic regression models that were used in this research. Therefore, the results are valid statistically. From the results of the logistic regression analysis, the authors inferred that the occurrence of the earthquakes beneath the Red Sea was statistically significantly associated with the proximity to the African–Arabian plate boundary. The association was distinguished only when the geographical locations and magnitudes of earthquakes were taken into account. This confirms the theory that the probability of occurrence of an earthquake is high in the area closest to the plate boundary. In terms of the morphological, structural and geophysical characteristics of the Red Sea and its relationship with the seismic activity, the results showed that earthquakes with moderate magnitudes occurred in the zone which represents the late evolutionary stage of the Red Sea rift. This zone includes the transition zone (central Red Sea) and late-stage continental rifting (northern Red Sea). In contrast to the earthquakes with moderate magnitudes, earthquakes with high magnitudes tended to occur in close proximity to the mid-ocean ridges of the Red Sea. This indicates that the sea-floor is spreading, that the spreading is associated with volcanic and seismic activities which occur at the mid-ocean ridges and that the spreading is probably caused by continental drift and plate tectonics, which typically give rise to earthquakes with high magnitudes. Earthquakes with moderate magnitudes tended to occur in close proximity to the transform faults more than earthquakes of higher magnitudes. This may explain the high frequency of occurrence of moderate earthquakes in the northern Red Sea, where there are more transform faults than in the southern Red Sea.

We concluded that the results indicate statistically significant associations and that these associations provide new insights into the occurrence of earthquakes in relation to the chosen geological and tectonic variables used in this research. Logistic regression models provide a generic technique which can be readily applied to other geological and tectonic variables. Geologists should consider using the same seismic data catalogue for exploring the location of earthquakes and potential associations with other geological and tectonic variables. They should also consider how the results of this research can inform strategies for capturing further sets of seismic data beneath the Red Sea.

Acknowledgments We are thankful to King Abdulaziz City for Science and Technology (KACST) for the accomplishment of this research.

References

Al-Ahmadi K, Al-Amri A, See L (2013) A spatial statistical analysis of the occurrence of earthquakes along the Red Sea floor spreading: clusters of seismicity. Arab J Geosci. doi:10.1007/s12517-013-0974-6

Amindan BG, Hagedorn DN (1998) Logistic Regression Applied to Seismic Discrimination. Pacific Northwest National Laboratory, the US Department of Energy

Bolt BA (2003) Earthquakes, 5th edn. WH Freeman, Gordonsville

Box GEP, Tidwell PW (1962) Transformation of the independent variables. Technometrics 4:531–550

Burns RB, Burns RA (2008) Business research methods and statistics using SPSS. SAGE Publications, London

Campbell AJ, Waddington ED, Warren SG (2011) Refugium for surface life on Snowball Earth in a nearly-enclosed sea? A first simple model for sea-glacier invasion. Geophys Res Lett 38(19)

Cinti FR, Faenza L, Marzocchi W, Montone P (2004) Probability map of the next M≥5.5 earthquakes in Italy. Geochem Geophy Geosy 5(11)

Clark WAV, Hosking PL (1986) Statistical methods for geographers. Wiley, New York

Cochran JR (1981) The Gulf of Aden, structure and evolution of a young ocean basin and continental basin. J Geophys Res 86:263–288

Cochran JR, Martinez F, Steckler HMA (1986) Conrad Deep, a new northern Red Sea deep, origin and implications for continental rifting. Earth Planet Sci Lett 78:18–32

Domencich TA, McFadden D (1975) Urban travel demand: Behavioural analysis. North-Holland, Amsterdam

Field A (2000) Discovering statistics using SPSS for windows: Advanced techniques for beginners. Sage, London

García-Rodríguez MJ, Malpica JA, Benito B, Díaz M (2008) Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. Geomorphology 95:172–191

Ghebreab W (1998) Tectonics of the Red Sea region reassessed. Earth Sci Rev 45:1–44

Hagos L, Shomali H, Roberts R (2006) Re-evaluation offocal depths and source mechanisms of selected earthquakesin the Afar depression. Geophys J Int 167:297–308

Hair JF, Anderson RE, Tatham R, Black WC (1998) Multivariate data analysis, 5th edn. Prentice Hall, Upper Saddle River

Honkura Y, Tanaka N (1996) Probability of earthquake occurrence in Greece with special reference to the VAN predictions. Geophys Res Lett 23(11)

Hosmer DW, Lemeshow S (2000) Applied logistic regression. Wiley, New York

Kusky T (2008) Earthquakes: Plate Tectonics and Earthquake Hazards (Hazardous Earth). Facts on File; 1st Edition

Martinez F, Cochran JR (1988) Structure and tectonics of the northern Red Sea: catching a continental margin between rifting and drifting. Tectonophysics 150:1–32

Okamura Y (2009) Relationship between Earthquakes and Reverse Faults along the Eastern Margin of Japan Sea. International Conference in Commemoration of the 10th Anniversary of the 1999 Chi-Chi Earthquake, Taiwan

Slemmons DB, Depolo CM (1986) Evaluation of active faulting and related hazards, in Studies in Geophysics: Active Tectonics. National Academy Press, Washington, DC, pp 45–62

Spall H (2003) Earthquakes and plate tectonics. Abridged from Earthquake Information Bulletin 9 (6). Reston, USGS

Stein A, Tolpekin V, Spatenkova O (2010) The Use of Statistical Point Processes in Geoinformation Analysis. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (38) Part II

Sykes LR (1967) Mechanism of earthquakes and nature of faulting on the mid-oceanic ridges. J Geophys Res 72:2131–2153

Tarbuck EJ, Lutgens FK, Tasa D (2005) Earth Science, Prentice Hall, 11th edition

Tatham RH, Savino JM (1974) Faulting mechanisms for 2 oceanic earthquake swarms. J Geophys Res 79:2643–2652