Understanding of groundwater salinity using statistical modeling in a small tropical island, East Malaysia

Sarva Mangala Praveena · Mohd Harun Abdullah ·
Kawi Bidin · Ahmad Zaharin Aris

Abstract This paper presents an understanding of groundwater salinity by identification of effective factors using chemometric methods (cluster analysis and multiple linear regressions) in Manukan Island, Sabah. Local groundwater and environmental properties were used to explore the effective factors of groundwater salinity. Cluster analysis showed salinity and chloride illustrated the highest similarities. Electrical conductivity and total dissolved solids were also grouped in the same cluster. Seawater is the only chloride source in groundwater of Manukan Island demonstrated an indication of seawater mixing in freshwater. It is an effect of upward movement of the seawater by pumping activities. Precipitation and evapotranspiration (environmental condition) with hydraulic heads were clustered together to show that they also influence salinity concentration in groundwater. Multiple linear regressions showed descending order of the factors from chloride (the largest contribution) to evapotranspiration (the smallest contribution) and illustrated the contribution to groundwater salinity in Manukan Island. The integrated results using chemometric methods have provided a way to identify the effective factors on groundwater salinity. This similar approach and resulting equation can be applied in other small tropical islands with alike hydrogeological condition and limited information available for a better understanding of its groundwater salinity.

Keywords Groundwater · Salinity · Factors · Chemometric methods

S. M. Praveena (✉) · K. Bidin
School of Science and Technology, Universiti Malaysia Sabah,
Locked Bag 2073, 88999 Kota Kinabalu, Sabah, Malaysia
e-mail: smpraveena@gmail.com

M. H. Abdullah
Water Research Unit, School of Science and Technology,
Universiti Malaysia, Sabah, 88999 Kota Kinabalu, Sabah,
Malaysia

A. Z. Aris
Faculty of Environmental Studies, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia

1 Introduction

Groundwater is an important water resource especially in small tropical islands. Small tropical islands have limited sources of freshwater, no surface water and fully reliant on rainfall and groundwater recharge. The demand of freshwater in small islands has been increasing due to tourism sector development. However, uncontrolled exploitation and management of groundwater will lead to environmental degradation. Overexploitation of freshwater by pumping well distorts the natural recharge–discharge equilibrium and causes drawdown of the water table or upcoming of the saltwater interface. Greater withdrawals may not only create localized upcoming but also result in reduction of freshwater lens (Rejani et al. 2008; Bonell et al. 1993). Similarly, groundwater resources of small tropical Manukan Island are limited and face seawater intrusion problem. Tourism impacts on Manukan Island made more severe to their limited water resources (Wong 1993). A number tourist to Manukan Island has increased about 46% from 1999 to 2008 and resulted in overextraction to meet the groundwater demand supply (Aris et al. 2009).

Groundwater quality is controlled by factors such as precipitation, evapotranspiration, mineralogy, type of aquifers, climate, topography, and also pumping rates. The

combinations of these factors create diverse water types that change groundwater composition spatially and temporally (Cheni and Khemiri 2009; Gholami et al. 2009). Groundwater salinity is a crucial component in groundwater quality. Salinization results from the concentration and precipitation of water-soluble salts such as chlorides in the groundwater (Metternicht 2001). According to Woszczyk et al. (2010) and Gholami et al. (2009), salinity changes more significant locally than temporally. In small islands, seawater intrusion is the main contributor to the increasing groundwater salinity (Rejani et al. 2008). Seawater intrusion causes irreversible deteriorations of groundwater quality in terms of drinking water quality. Drinking water safety is a primary consideration especially in small islands that solely reliable on groundwater. Moreover, high salinity concentration can affect the island ecological balance, plant growth, and environmental degradation. Thus, identification of factors related to groundwater salinity is important for maintaining a good supply of groundwater resource in small islands (Burns et al. 2005).

Different techniques have been used to evaluate groundwater quality such as numerical modeling, hydrochemical analysis, and graphical method (Praveena et al. 2010). Numerical modeling requires large data sets in order to simulate the study area whereas hydrochemical analysis requires more variables to identify the geochemical composition. Graphical methods (Piper and Schoeller and Collins diagrams) only consider major ion constituents hence these methods are not suitable for many environmental studies. Chemometric analysis is also one of the ways to evaluate groundwater quality such as evaluation of salinity. Chemometric analysis is also very useful to investigate spatial and temporal variations (Singh et al. 2005). Chemometric analysis performs quantitatively and qualitatively using powerful methods such as cluster analysis and multiple linear regressions to analyze groundwater quality. The most important point is these methods have been widely applied as unbiased technique in groundwater quality analysis (Kumar and Riyazuddin 2008).

The objectives of present research were to define the effective factors on groundwater salinity using chemometric methods namely cluster analysis (CA) and multiple linear regressions (MLR). These methods act as an accurate tool to fully integrate and better interpret effective factors influence groundwater salinity. Identification of effective factors on groundwater salinity is crucial to understand the groundwater chemistry processes took in the aquifer. This information indeed very useful especially in threatened small tropical islands environment.

## 2 Materials and methods

2.1 Study area

Manukan Island (5°57′–5°58′N and 115°59′–116°01′E) covers an area of 206,000 m^2, with a crescent shape, (Fig. 1). The study area is located in Tunku Abdul Rahman Park, offshore of Kota Kinabalu, Sabah, East Malaysia, on the island of Borneo cover by South China Sea (Fig. 1). The island is enacted under Parks Enactment 1978 and managed by The Sabah Parks Trustees. Almost 80% of the area is covered by dense vegetation in high relief area while the rest of the area in the low lying is developed for tourism activities. Manukan Island has limited water resources options. It has no surface water and fully reliant on rainfall and groundwater recharge. Tourism impacts on small islands such as Manukan Island made more severe to their limited water resources and have resulted in tremendous increased of groundwater extraction to meet the groundwater supply and domestic needs (Aris et al. 2009; Wong 1993). Moreover, all the wells for groundwater extraction are located in the low-lying area of the island (8 wells have been shut down and only 1 well is currently operating). The small area and low elevations lead to very limited water storage in the low-lying area of Manukan Island (Abdullah et al. 2002).

2.2 Climate

Manukan Island experiences a tropical climate, which has high temperatures and annual rainfall. The mean annual temperature in Manukan Island is 26.5°C. The area receives annual rainfall between 2,000 and 2,500 mm (Fig. 2) with humidity between 80 and 90%. Annual
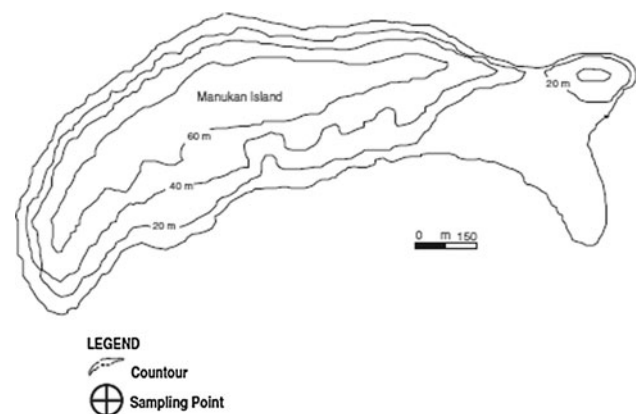
Fig. 1 Manukan Island with a crescent shape
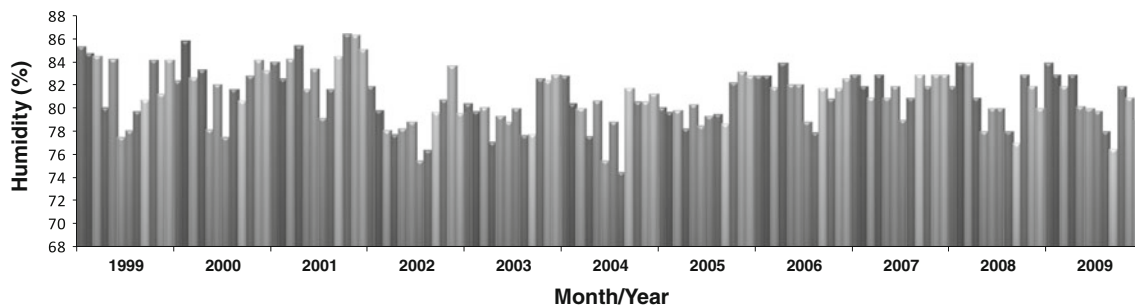
Fig. 2 Monthly average rainfall distribution from 1999 to 2009

average evapotranspiration in Manukan Island is about 11 mm/year.

2.3 Geology

The topography of the island is relatively hilly land with maximum elevations of approximately 60 m in the western and decreasing elevation toward the eastern coast. The island is about one and half a kilometer long and three kilometer wide in the middle. Generally, the profiles at low-lying area are flatter and thinner than hilly area. The regional groundwater flow is mainly toward the South China Sea. The island consists of unconfined sandy aquifer, underlain sedimentary rock (sandstone and shale), and part of the Crocker range rock formation of the western coast of Sabah deposited during Late Eocene to Middle Miocene (Basir et al.1991; Abdullah et al. 2002; Tunku Abdual Rahman 2007). The medium of the aquifer consists of fine to coarse sand mixed with some fine gravel. The parent materials at hilly area of the island are sandstone and mudstone whereas in the low lying of the island is alluvium. Alluvium acts as a sufficient water storage, which solely depends on its thickness. The sedimentary rock forms a slight symmetrical syncline in the low area and can be observed in several locations around the island (Abdullah et al. 2002). The sedimentary rock of Manukan Island dips toward the low-lying area (east–northeast) with dipping angles of 15–45°. Abdullah et al. (2002) conducted a study on the morphological of the island found that the thickness of the aquifer from the ground surface to bedrock are approximately, 5.7 m (northern part), 11 m (southern part) and 12 m (at the middle). The geological structures indicate that the freshwater of Manukan Island mainly depends on recharge water from precipitation.

2.4 Research methodology

In this research, low-lying area of Manukan Island was selected for this study (Fig. 3) since it has been developed for tourism activities, and all the wells for groundwater extraction are located in this area (8 wells have been shut down and only 1 well is currently operating). Local groundwater properties and environmental conditions were used in the identification of effective factors of groundwater salinity. Local groundwater properties consist of in situ parameters namely pH, temperature, oxygen reduction potential (ORP), electrical conductivity (EC), salinity (Sa), and total dissolved solids (TDS) together with hydraulic head and chloride. Moreover, environmental conditions are represented by meteorological data from meteorological station established in the low-lying area of Manukan Island namely precipitation, evapotranspiration, and humidity.

A total of nine boreholes were constructed and installed by hand auger manually to a depth of between 1 and 3.5 m in the aquifer. A total of nine boreholes and eight observations wells were selected in the sampling network, which reasonably represent the groundwater quality in the low lying area of study region. Samplings network were done in 2009 from the observation wells and boreholes. Samplings network in 2009 were done in the middle of each month from January to December 2009. A total of 204 groundwater samples (108 boreholes and 96 observation wells groundwater samples) were collected throughout samplings network done in 2009. Prior to water sample collection, hydraulic heads of each borehole and observation well were measured using water level meter using Solinst 101 water table detector. Water samples were collected by using modified dry vacuum pump technique and water sampler for borehole and well sample collection, respectively. First, water sample from borehole/well was collected in measuring bottle for determination of in situ parameters namely pH, temperature, ORP, EC, Sa, and TDS. It is to acquire representative values of current aquifer condition. The selected parameters were measured using devices as showed in Table 1. Next, water samples were collected by a precleaned polyethylene bottles. Polyethylene bottles were cleaned by rising it with tap water followed by soaking with acid wash water (1:10 HNO_3) overnight and rinsed with distilled water to eliminate traces precipitated at the surface of the sampling bottles. Three replicates from each borehole were taken during each sampling period to obtain an average value.

Fig. 3 Low-lying area of
Manukan Island
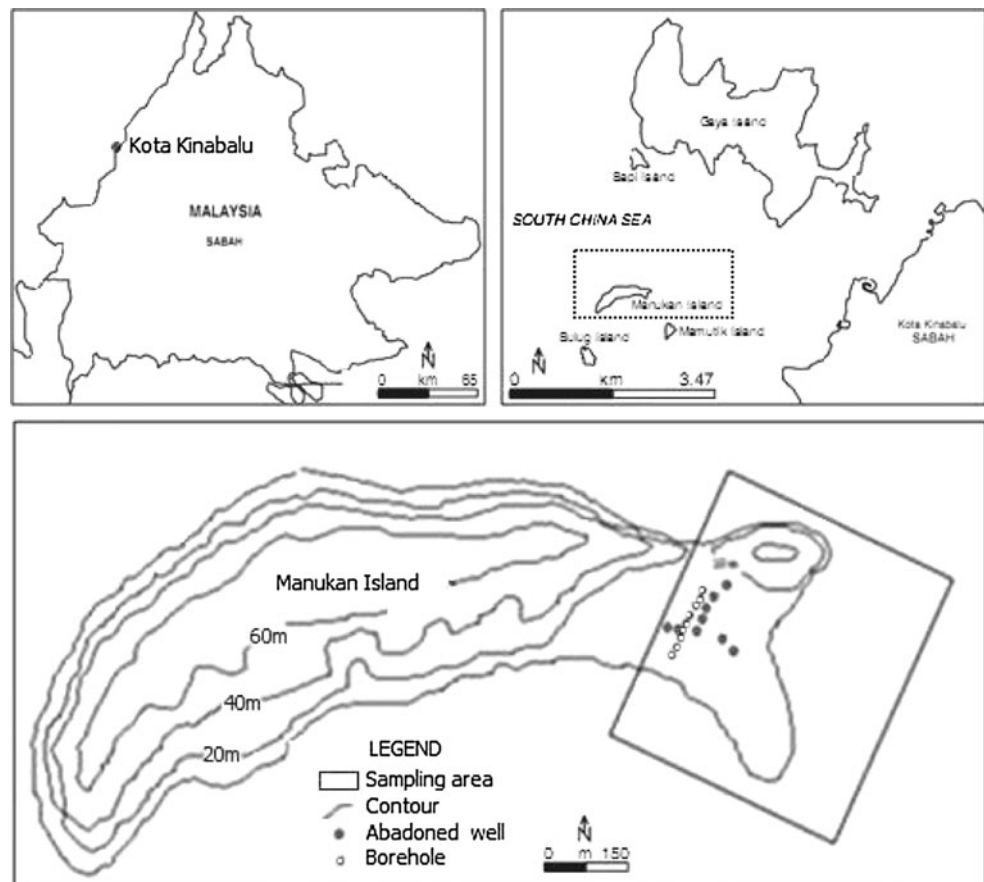
Table 1 List of parameters for in situ detection

Parameters	Unit	Devices
pH	–	Ecoscan pH 6, WTW pH 3400i
Temperature	°C	Ecoscan pH 6, WTW pH 3400i
Oxygen reduction potential	Rmv	ORP Testr 10 Oakton
Electrical conductivity	µS/cm	Oakton Cond. 100 series
Salinity	ppt	Salttestr
Total dissolved solids	mg/L	Oakton Cond. 100 series

The polyethylene bottles were transported to the laboratory
under low-temperature conditions in icebox and analyzed
immediately for chloride determination. Analysis for
chloride was carried out using argentometric method
(APHA 1995). All the devices were calibrated with buffer
solutions before and after the field samplings to ensure that
they functioned properly and accurately. Moreover, the
devices probes sensors were soaked in distilled water to
protect the electrodes before and after field samplings and
calibrations. Personal field log book was brought to field
samplings to record all the readings as soon as possible to
avoid any misplacement of data.

Moreover, meteorological data (precipitation and
humidity) from the meteorological station were down-
loaded to a Twinhead laptop each month during samplings

in 2009. Evapotranspiration (ET) was calculated by the
Penman method, a combination of energy balance and
aerodynamic approaches (Stewart and Howell 2003). This
formula has a few advantages over the other numerous ET
formulas, such as Thornthwaite, The Blaney–Criddle and
the van Banel methods. The advantages of the Penman
method are that it provides the most comprehensive
approach to the estimation of ET, more widely used than
any of other alternatives formulas, acceptable for a wide
range of climates, and has given satisfactory results in ET
estimation elsewhere (Wallacae et al. 1990). According to
Bello et al. (2010), various empirical methods of ET esti-
mation have their own limitations. Limitation in Penman
method is the radiation measurement for sunshine hours.
Penman method uses approximation for radiation mea-
surement for sunshine hours. However, radiation value of
sunshine hours varies with latitudes and climates.

2.5 Data analysis

In chemometric methods, normal distribution is preferred
for optimal results and reliable interpretations of the results.
In order to analyze the effective factors on groundwater
salinity, the groundwater samples data were summarized
through chemometric methods (CA and MLR) using SPSS

Version 17.0 for Windows statistical program. Simenov et al. (2002) mentioned CA is carried out to further classify elements of different sources on the basis of their similarities between the chemical properties. Cluster analysis aims at sorting different objects into same groups based on degree of associations between objects. The purpose of cluster analysis is to discover a system of organizing observations into groups that share similar properties in common. Hierarchical cluster analysis (HCA) using dendograms identifies relatively homogeneous groups of variables in similar properties and combines clusters until only one is left (Bierman et al. 2009). A dendogram distinguishes groups of high similarities that have small distances between clusters whereas dissimilarity between groups is represented by the maximum of all possible distances between clusters. Dendogram is read from left to right whereby vertical lines show joined clusters. The line position on the scale shows distance at which clusters are joined. The observed distances are in the range of 1–25. The first vertical corresponds to the smallest rescaled distance. The next vertical line is at distances for other merges. In the last two steps, fairly dissimilar clusters are combined. Moreover, according to Cheni and Khemiri (2009) and Helena et al. (2000), this approach has been used to extract related variables and infer the processes that control water chemistry and effective in terms of manipulating, interpreting, and representing data concerning groundwater pollutants. Multiple linear regression (MLR) is an extension of correlation that provides equation linking independent variables, V_d (pH, temperature, ORP, EC, TDS, Cl, HH, precipitation, and ET) to the dependent variable, V_i (Salinity) using the following form as stated in Eq. 1. These parameters selected as local changes in salinity are more important than temporal changes in smaller-scale studies (Gholami et al. 2009). Thus, local changes in salinity can be represented by selected parameters (pH, temperature, ORP, EC, TDS, Cl, HH, precipitation, and ET).

$$V_d = \beta_o + \beta_o V_i + \cdots + \beta_n V_n \qquad (1)$$

when the intercept β_o and the regression coefficients of descriptors (β_i) are determined by least square method (Green and Carroll 1996), V_i descriptors are used to describe water quality and dependent. (n) is the number of water samples. The reduction in the number of descriptors (variables) was included in the study to minimize the information overlap in variables. The best equation is selected based on the highest multiple correlation coefficients (r), lowest standard deviation (SD), and F-ratio value. Relationships between variables were established using the stepwise regression method (Broto 2006; Bernstein 1988). Stepwise selection is a combination of the forward and backward procedures. It allows the later removal of variables that were previously entered (Coakes et al. 2010). CA is a multivariate analysis technique that seeks information about variables so that similar clusters can be formed. MLR selection process allows the researcher to obtain a reduced set of variables from a larger set of variables, eliminating unnecessary variables, simplifying data, and enhancing predictive accuracy. Thus, these two approaches were used in the search of effective factors on groundwater salinity in the low-lying area of Manukan Island.

3 Results and discussion

3.1 Local groundwater and environmental conditions properties

Data of the local groundwater and environmental conditions properties appeared to follow normal distribution.
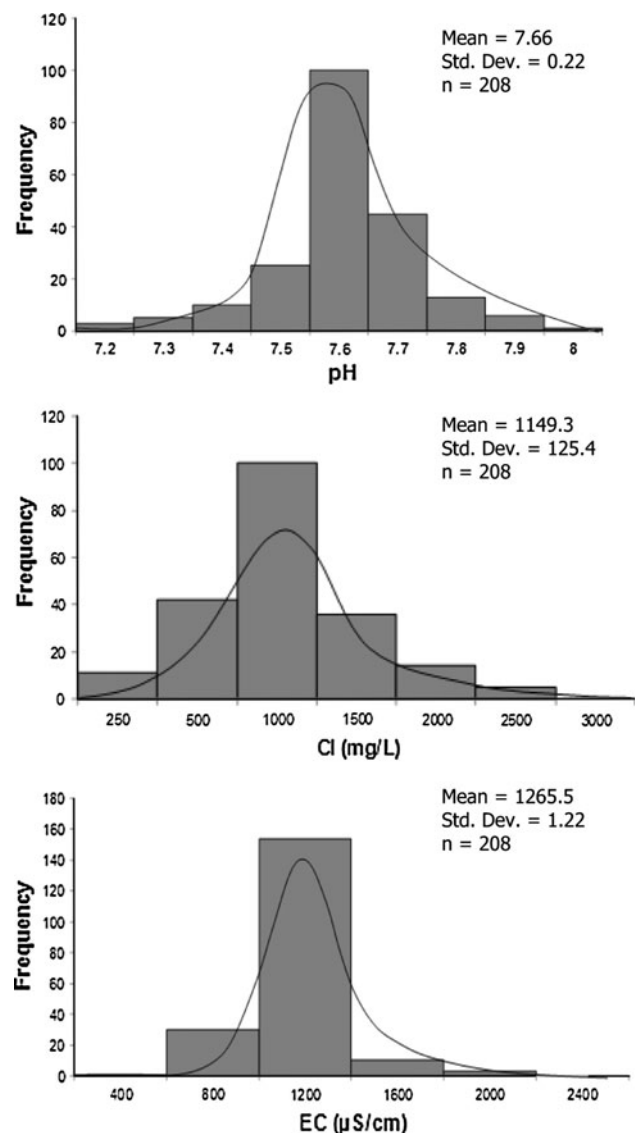
Fig. 4 Frequency distribution of several environmental factors (pH, EC, and Cl) (*SD* standard deviation)

Table 2 Mean values of local groundwater properties in the study area

	pH	Temp (°C)	ORP (mV)	EC (µS/cm)	Sa (ppt)	TDS (mg/L)	Hydraulic Head (m)	Cl (mg/L)
Observation well (n = 108)								
W1	7.4 ± 0.13	28.4 ± 0.55	−71.2 ± 4.58	2,286.8 ± 0.45	1.2 ± 0.11	1,122.6 ± 334.1	15.5 ± 0.3	1,862.4 ± 345.1
W2	7.5 ± 0.15	27.6 ± 0.60	−77 ± 3.23	970.7 ± 0.33	0.6 ± 0.23	523.8 ± 230.1	15.9 ± 0.1	2,563.8 ± 234.1
W3	7.6 ± 0.20	27.1 ± 0.51	−81 ± 2.13	1,071.6 ± 0.31	0.6 ± 0.14	534.8 ± 123.1	15.6 ± 1.3	1,354.9 ± 231.6
W4	8.0 ± 0.22	27.1 ± 0.49	−107.8 ± 2.30	1,902.1 ± 0.55	0.5 ± 0.22	491.6 ± 101.2	15.6 ± 0.3	2,399.5 ± 134.5
W5	7.9 ± 0.24	27.5 ± 0.44	−109 ± 2.34	1,035.6 ± 0.56	0.6 ± 0.26	546.5 ± 167.1	15.7 ± 0.1	2,679.3 ± 335.2
W7	7.8 ± 0.22	27.1 ± 0.51	−99.3 ± 3.44	1,005.1 ± 0.24	0.6 ± 0.17	486.5 ± 98.3	16.0 ± 0.4	1,032.3 ± 120.1
W8	7.6 ± 0.25	27.3 ± 0.55	−71.3 ± 3.12	808.6 ± 0.39	0.5 ± 0.35	445.5 ± 67.2	15.6 ± 1.1	1,071.6 ± 90.5
W9	7.7 ± 0.13	27.2 ± 0.42	−50.1 ± 4.22	1,064.2 ± 1.23	0.6 ± 0.33	564.8 ± 65.2	15.6 ± 0.3	2,281.2 ± 234.1
Borehole (n = 96)								
B1	7.2 ± 0.20	28.1 ± 0.34	−52.9 ± 2.33	1,659.1 ± 2.34	0.8 ± 0.24	673.5 ± 34.2	14.8 ± 1.1	457.5 ± 101.2
B2	7.3 ± 0.15	25.9 ± 0.32	−248 ± 1.31	540 ± 2.56	0.3 ± 0.43	270 ± 12.3	15.7 ± 0.4	470.0 ± 189.1
B3	7.9 ± 0.18	28.5 ± 0.25	−101.8 ± 3.11	953.6 ± 1.22	0.6 ± 0.33	438.2 ± 61.3	15.9 ± 0.5	474.6 ± 245.1
B4	7.5 ± 0.10	29.3 ± 0.34	−73.1 ± 2.44	1,509 ± 3.12	0.8 ± 0.12	757.3 ± 45.2	15.4 ± 0.7	545.5 ± 335.1
B5	7.3 ± 0.13	28.7 ± 0.36	−60.7 ± 1.12	1,229.2 ± 2.22	0.7 ± 0.20	1,204.5 ± 14.5	14.6 ± 0.8	557.6 ± 234.1
B6	7.5 ± 0.14	29.5 ± 0.38	−73.3 ± 2.89	1,378.5 ± 3.03	0.8 ± 0.13	691.9 ± 67.2	16.6 ± 0.3	405.3 ± 456.1
B8	7.4 ± 0.13	29.6 ± 0.33	−38.1 ± 2.45	1,303.6 ± 4.12	0.8 ± 0.45	660.2 ± 58.3	15.4 ± 0.1	504.4 ± 345.1
B9	7.4 ± 0.16	29.9 ± 0.32	−68.7 ± 2.33	1,445.2 ± 3.22	0.8 ± 0.23	710.4 ± 34.1	16.7 ± 0.4	587.0 ± 234.1
B10	7.7 ± 0.17	29.6 ± 0.34	−90 ± 3.03	1,346.9 ± 2.52	0.7 ± 0.28	620.4 ± 23.4	16.6 ± 0.7	292.3 ± 123.2

Figure 4 shows the frequency distribution of several environmental factors (pH, EC and Cl). The properties of the local groundwater and environmental conditions were presented in Tables 2 and 3. Trend between local groundwater properties, hydraulic heads, and chloride shows the influence of saline water toward groundwater quality. Moreover, wide distributions of studied local groundwater properties indicated that the composition was affected by multiple processes, including seawater–freshwater mixing (Aris et al. 2007b). Next, CA was performed on the data set using average linkage between groups (rescaled distance cluster) to study the connections between local groundwater properties and environmental conditions toward the identification of effective factors on groundwater salinity.
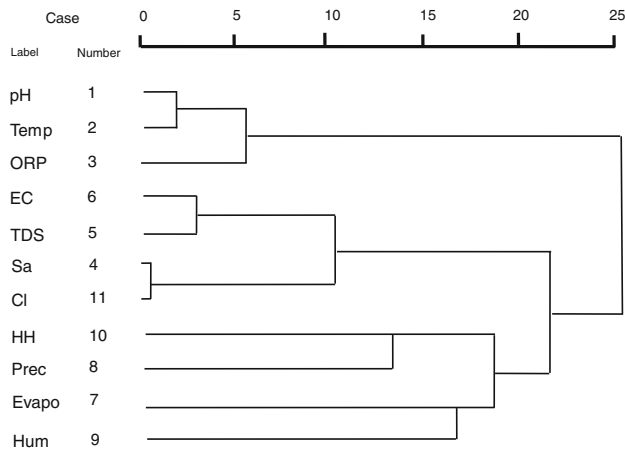
3.2 Cluster analysis output

Figure 5 shows dendrogram of studied local groundwater and environmental properties in the identification of effective factors on groundwater salinity. According to Banoeng-Yakubo et al. (2009) and Yidana et al. (2008), decision on the number of clusters to use in an HCA is a subjective manner. It is done based on the knowledge of the researcher and also on the subject matter being investigated. In this study, the differentiation was based on the knowledge of current groundwater data as well as on previous groundwater findings (Aris et al. 2009) of low lying area in Manukan Island.

Table 3 Mean values of environmental conditions in the study area

Months	Humidity (%)	Precipitation (mm)	Evapotranspiration (mm)
Jan-2009	75.5 ± 1.12	370.3 ± 112.1	5.82 ± 0.89
Feb-2009	74.5 ± 1.34	243.1 ± 101.3	5.28 ± 0.67
Mac-2009	73.5 ± 1.43	181.9 ± 99.3	6.48 ± 0.56
Apr-2009	69.5 ± 2.34	136.8 ± 88.3	6.4 ± 0.66
May-2009	73.5 ± 2.21	165.4 ± 77.4	5.27 ± 0.34
Jun-2009	47 ± 2.13	135 ± 55.2	4.01 ± 0.31
Jul-2009	58 ± 2.27	211 ± 110.1	4.92 ± 0.45
Aug-2009	72.5 ± 1.05	461.8 ± 125.2	5.31 ± 0.31
Sep-2009	70.5 ± 1.54	269.8 ± 96.5	5.4 ± 0.47
Oct-2009	66 ± 1.21	339 ± 102.3	4.94 ± 0.69
Nov-2009	74.5 ± 2.24	213.6 ± 99.2	5.2 ± 0.78
Dec-2009	74 ± 2.34	200 ± 88.3	5.1 ± 0.88

From groundwater samples' dendogram (Fig. 5), salinity and chloride exhibit the highest similarities and eventually assembled into the same cluster with electrical conductivity and total dissolved solids. The salinity and chloride association also give effects toward electrical conductivity and total dissolved solids. This is also supported by study carried out by Aris et al. (2007a, b) where positive correlations were found between salinity, chloride, electrical conductivity, and total dissolved solids. This cluster also explains that the salinity of groundwater is an

Fig. 5 Dendogram showing hierarchical cluster analysis between local groundwater and environmental condition properties (*Temp* temperature, *ORP* oxidation reduction potential, *EC* electrical conductivity, *TDS* total dissolved solids, *Cl* chloride, *HH* hydraulic head, *Prec* precipitation, *Evapo* Evapotranspiration, *Hum* humidity)

Table 4 Summary for the linear regressions predicting groundwater salinity

	ß	*t* value	Sig *t*
Cl	0.543	26.025	0.000
TDS	0.283	11.406	0.000
EC	0.281	11.403	0.000
Prec	−0.247	−10.023	0.000
HH	−0.141	−5.721	0.000
Evapo	0.143	5.725	0.000

Cl chloride, *TDS* total dissolved solids, *EC* electrical conductivity, *Prec* precipitation, *HH* hydraulic head, *evapo* evapotranspiration

* ß values express the value of a predictand variable as a linear function of one or more predictor variables and an error term

indication of seawater mixing in freshwater. Seawater is the only chloride source in groundwater of Manukan Island. Thus, chloride concentration in groundwater samples is an effect of upward movement of the seawater by pumping activities. Environmental condition (precipitation and ET) was clusters together to show that they also influence the salinity concentration in groundwater. According to Habib et al. (2007), rainfall plays an important role in the salinity percentages in groundwater. Precipitation has a greater effect in diluting the salt in groundwater. Thus, high precipitation tends to decrease the salinity concentration in groundwater (Kresic 2006). Strong evaporation in areas with shallow water tables may also lead to salinization (Vigna et al. 2009). In terms of hydraulic head, Kim et al. (2000) illustrated that hydraulic heads showed a relatively higher and prompt response to rainfall. It also influences the hydrochemical characteristics of groundwater. The clusters showed that humidity plays a smaller role in the recognition of effective factors in groundwater salinity. This is because humidity is the only environmental factor that showed maximum distance compared with other environmental condition (precipitation and ET) clusters.

3.3 Multiple linear regressions output

MLR analysis was conducted to investigate the relationships between salinity, local groundwater, and environmental properties. MLR method was used to select the best model to identify the effective factors on groundwater salinity. In the selection of best model, an analysis of multiple *R* coefficients and r^2 values were studied. The multiple *R* coefficients indicated correlations between local groundwater properties, environmental conditions, and salinity ranged from 0.951 to 0.995. Among all the possible equations, equation where r^2 values closer to 1 was selected. The model was also checked for any multicollinearity effect via variance inflation factor (VIF). The VIF value was closer to one (VIF = 0.989) thus there is no any multicollinearity effect (Alin 2010). Table 4 presents the MLR output predicting the effective factors on groundwater salinity. The output showed that local groundwater properties and environmental properties were found significant ($p < 0.00$) as effective factors on groundwater salinity in Manukan Island. From the results obtained, descending order of factors from Cl to evapotranspiration illustrates the contribution of groundwater salinity in Manukan Island. Positive/negative sign of beta coefficients (ß) and *t*-values (*t* value) in Table 4 indicated that there is positive/negative relationships between salinity and elements of local groundwater and environmental properties. The relative importance of the estimation of a dependent variable, Cl, made the largest contribution toward groundwater salinity with the highest ß-value (Table 4). This is also supported by CA output where salinity and chloride exhibit the highest similarities. Thus, the selected equation for identification of effective factors on groundwater salinity in Manukan Island is as followed in Eq. 2. Regression coefficient value (0.035) can be interpreted as the predicted salinity value if all the other effective factors (chloride, total dissolved solids, electrical conductivity, precipitation, hydraulic head, and ET) are set to be zero value. Thus, 0.035 value represents the concentration of salinity in seawater. Thus, the obtained model can also be applied to classify groundwater in other small islands of similar hydrogeological condition with limited information.

$$Sa = 0.035 + 0.543Cl(mg/L) + 0.283TDS(mg/L)$$
$$+ 0.281EC(\mu S/cm) - 0.247Prec(mm)$$
$$- 0.141HH(m) + 0.143Evapo(mm) + \varepsilon \qquad (2)$$

where ε is the error of estimation in the regression model.

4 Conclusion

Groundwater of Manukan Island was investigated to determine the effective factors on groundwater salinity using chemometric methods. CA was used to analyze the similarities between homogeneous groups of variables in similar properties. The CA output showed salinity and chloride exhibit the highest similarities together with EC and TDS. Precipitation and ET in environmental condition properties were clustered together, which also influence the groundwater salinity. Evaluation of relationships between salinity, local groundwater and environmental properties was made using MLR. Estimation of effective factors on groundwater salinity showed Cl is the largest contributor. The only contributor of Cl in groundwater of Manukan Island is from seawater. The pumping activities are believed to be the contributor of seawater and freshwater mixing in groundwater of Manukan Island. The integration of chemometric methods (cluster analysis and multiple linear regression) to evaluate the effective factors of groundwater salinity showed that groundwater salinity is related to local groundwater properties (Cl, TDS, and EC) and environmental conditions (Prec, HH, and evapotranspiration). It is important to be noted that ET estimation (Penman method) uses approximation in radiation measurement for sunshine hours. Availability of new instruments for radiation measurement for sunshine hours will produce more reliable results in ET values in the study area. This will also influence CA or MLR results in the estimation of effective factors on groundwater salinity. Despite the limitation explained, proposed methodology allows the incorporation of local groundwater and environmental properties into a statistical model to determine the effective factors of groundwater salinity. It is an important part in order to understand the groundwater chemistry. This applied methodology and resulting equation is useful especially in similar small tropical islands hydrogeological condition with limited information available to estimate salinity of groundwater in those study sites.

References

Abdullah MH, Kassim MA, Hanapi MN (2002) Saltwater encroachment into the sandy aquifer of Manukan island. Borneo Island 12:1–22

Alin A (2010) Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics 2:370–374

APHA (1995) Standard methods for the examination of water and wastewater, 19th edn. American Water Works Association, Water Environment Federation, Washington

Aris AZ, Abdullah MH, Ahmed A, Woong KW (2007a) Controlling factors of groundwater hydrochemistry in a small island's aquifer. Int J Environ Sci Technol 4:441–450

Aris AZ, Abdullah MH, Kim KW (2007b) Hydrogeochemistry of groundwater in Manukan island, Sabah. The Malaysian Journal of Analytical Sciences 11:407–413

Aris AZ, Abdullah MH, Kim KW, Praveena SM (2009) Hydrochemical changes in a small tropical island's Aquifer, Manukan Island, Sabah, Malaysia. Environ Geo 56:1721–1732

Banoeng-Yakubo B, Yidana SM, Emmanuel N, Akabzaa T, Asiedu D (2009) Analysis of groundwater quality using water quality index and conventional graphical methods: the Volta region, Ghana. Environ Earth Sci 59:867–879

Basir J, Sanudin T, Tating FF (1991) Late Eocene planktonic foraminifera from the Crocker Formation, Pun Batu, Sabah. Warta Geologi 14:1–15

Bello NJ, Eruola AO, Ufeogbune GC, Awomeso JA (2010) Evaluation of some empirical methods of estimating potential evapotranspiration for determination of length of growing season in a tropical wet and dry climate. African J of Agricultural Res 5:2116–2123

Bernstein IH (1988) Applied multivariate analysis. Springer, New York

Bierman P, Lewis M, Ostendirf B, Tanner J (2009) A review of methods for analysing spatial and temporal patterns in coastal water quality. Ecol Indi doi:10.1016/j.ecolind.2009.11.001

Bonell M, Hufschmidt MM, Gladwell JS (1993) Hydrology and water management in the humid tropics: hydrological research issues and strategies for water management. Cambridge University Press, Cambridge, p 925

Broto AS (2006) Statistics made simple, 2nd edn. Mandaluyong, Philippines, p 299

Burns D, Vitvar T, Mcdonnell I, Hasserr J, Duncan J, Kendall C (2005) Effects of suburban development on runoff generation in the Croton River Basin, New York, USA. J Hydrol 311:266–281

Cheni II, Khemiri S (2009) Evaluation of groundwater quality using multiple linear regression and structural equation modeling. Int J Environ Sci Technol 6:509–519

Coakes SJ, Steed LG, Ong C (2010) SPSS : analysis without anguish : version 17.0 for windows. Wiley, New Jersey, p 300

Gholami V, Yousef Z, Rostami HZ (2009) Modeling of groundwater salinity on the Caspian Southern Coasts. Water Resour Manage 24:1415–1424

Green P, Carroll J (1996) Mathematical tools for applied multivariate analysis. Academic Press, New York, p 376

Habib E, Larson BF, Nuttle WK, Rivera-Monroy VH, Nelson BR, Meselhe EA, Twilley RR (2007) Effect of rainfall spatial variability and sampling on salinity prediction in an estuarine system. J Hydrol 350:56–67

Helena B, Pardo R, Vega M, Barrado E, Fernandez JM, Fernandez L (2000) Temporal evaluation of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water Res 34:807–816

Kim T, Lee KK, Ko KS, Chang HW (2000) Groundwater flow system inferred from hydraulic stresses and heads at an underground LPG storage cavern site. J Hydrol 236:165–184

Kresic N (2006) Hydrogeology and groundwater modeling. CRC Press, Boca Raton, p 807

Kumar RA, Riyazuddin P (2008) Application of chemometric techniques in the assessment of groundwater pollution in a suburban area of Chennai city, India. Current Sci 94:1012–1022

Metternicht G (2001) Assessing temporal and spatial changes of salinity using fuzzy logic, remote sensing and GIS. Foundations of an expert system. Ecol Model 144:163–179

Praveena SM, Abdullah MH, Aris AZ (2010) Groundwater solution techniques: environmental applications. J Water Resour Protection 2:8–13

Rejani R, Jha MK, Panda SN, Mull R (2008) Simulation modeling for efficient groundwater management in Balasore coastal basin, India. Water Resour Manage 22:23–50

Simenov V, Einax JW, Stanimirova I, Kraft J (2002) Environmetric modeling and interpretation of river water monitoring data. Anal Bioanal Chem 374:898–905

Singh KP, Malik A, Singh VK, Mohan D, Sinha S (2005) Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain, North India. Anal Chi. Acta 550:82–91

Stewart BA, Howell TA (2003) Encyclopedia of water science. Marcel Dekker, New York, p 1076

Tunku Abdual Rahman (2007) Tunku Abdul Rahman marine park. http://www.sabahtourism.com/sabah-malaysian-borneo. Accessed on 18 November 2009

Vigna FL, Ciadamidaro S, Mazza R, Mancini L (2009) Water quality and relationship between superficial and ground water in Rome (Aniene River basin, central Italy). Environ Earth Sci 60:1267–1279

Wallacae JS, Gash JHC, Sivakumar MVK (1990) Preliminary measurements of net radiation and evaporation over bare soil and fallow bushland in the Sahel. Int J Climatol 10:203–210

Wong PP (1993) Tourism vs. environment: the case for coastal areas. Dordrecht, p 225

Woszczyk, M, Spychalski, W, Lutynska, M, Cieslinski, R (2010) Temporal trend in the intensity of subsurface saltwater ingressions to coastal Lake Sarbsko (Northern Poland) during the last few decades. IOP Conference Series: Earth and Environ Sci 9: doi:10.1088/1755-1315/9/1/012013

Yidana SM, Ophori D, Banoeng-Yakubo B (2008) A multivariate statistical analysis of surface water chemistry–The Ankobra Basin, Ghana. J Environ Manage 86:80–87