

## Statistical approaches and hydrochemical modelling of groundwater system in a small tropical island

A. Z. Aris, S. M. Praveena, M. H. Abdullah and M. Radojevic

### ABSTRACT

The aquifer of Manukan Island of Borneo, Malaysia had been found to be affected by seawater intrusion associated with excessive groundwater exploitation. This research attempted to characterize the chemistry of an impacted zone in the island using factor analysis (FA), cluster analysis (CA) and a hydrochemical model package (PHREEQC). The factor scores were used to plot the spatial map and to group the relationships among the monitoring wells using CA. The results of FA analysis revealed that the three main processes associated with the seawater intrusion event are aquifer salinization, cation exchange process and redox sequences. Output from the PHREEQC simulation was used to support the findings from the multivariate analysis.

**Key words** | geochemistry, groundwater, island, modelling, multivariate analysis, seawater intrusion

**A. Z. Aris** (corresponding author)  
Department of Environmental Sciences, Faculty of  
Environmental Studies,  
Universiti Putra Malaysia,  
43400 UPM Serdang, Selangor, Malaysia  
E-mail: zaharin@env.upm.edu.my

**S. M. Praveena**  
**M. H. Abdullah**  
Environmental Science Programme, School of  
Science and Technology,  
Universiti Malaysia Sabah,  
88999 Kota Kinabalu, Sabah, Malaysia

**M. Radojevic**  
Faculty of Engineering,  
University of Nottingham Malaysia Campus,  
43500 Semenyih, Selangor, Malaysia

### INTRODUCTION

Various multivariate statistical methods, such as factor analysis (FA) and cluster analysis (CA), help in the interpretation of complex datasets, such as those having large numbers of parameters (variables), allowing a better understanding of the processes involved. Multivariate analysis is an explanatory tool to discover and interpret relationships between variables. An ever-increasing literature on multivariate analysis suggests that the technique is useful in evaluating multivariate geochemical data of various types (Saager & Sinclair 1974; Ashley & Lloyd 1978; Davis 1986; Helena *et al.* 2000; Xie *et al.* 2005; Choi *et al.* 2008). Commonly, it is possible to interpret the resulting multivariate (factors) in geological terms, such as processes or types, as it portrays a dynamic and complex environment.

The interpretation of hydrochemical data of groundwater samples collected from different sampling locations is hampered due to a few reasons. The reasons include the difficulty to visualize, consider only the major and minor ions with equal emphasis to interpret the group of variables to evaluate the chemical nature of

groundwater and evaluate the geochemical relationships that may exist using normal histograms, scatter plots and trilinear diagrams (Yammani *et al.* 2008). The multivariate statistical technique is deemed to be the best approach to avoid misinterpretation of large amounts of complex environmental data (Simeonov *et al.* 2002). Multivariate statistical methods have been widely used in drawing meaningful information from masses of environmental data and have often been used in exploratory data analysis tools for the classification of samples (observations) or sampling stations and the identification of pollution sources (Alberto *et al.* 2001; Reghunath *et al.* 2002; Bierman *et al.* 2011). These techniques have also been applied to characterize and evaluate the surface and freshwater quality, groundwater chemistry and quality as well as verifying spatial and temporal variations caused by natural and anthropogenic factors. Multivariate statistical techniques, FA and CA are effective means of manipulating, interpreting and representing data concerning groundwater pollutants and geochemistry (Lu *et al.* 2008). In the last decade,

Multivariate statistical techniques have been applied to characterize and evaluate the quality of groundwater. The effectiveness of this technique in groundwater chemistry discrimination over the traditional Piper and Stiffs diagrams stems from its ability to reveal hidden inter-variable relationships. The use of this technique also allows the use of virtually limitless numbers of variables, thus trace elements and physicochemical parameters can be part of the classification parameters. By its use of raw data as variable inputs, errors arising from close number systems are avoided. Since elements are treated as independent variables, the masking effect of chemically similar elements that are often grouped together is avoided (Dalton & Upchurch 1978).

The use of FA and CA has several advantages that overcome these limitations and, in the present study, this technique is an adequate tool to investigate the principles of interaction of components and their integration into a system. Since the aquifer systems are characterized by inherent variability (natural, anthropogenic, spatial/temporal), a combination of FA and CA is used to reveal their relations, the limits and hierarchy of the interaction between the components. Such applied statistical analyses to the environmental data were demonstrated clearly as a method in identifying the processes that could be affecting the chemistries of the groundwater (Davis 1986; Helena *et al.* 2000; Kim *et al.* 2003; Xie *et al.* 2005; Chae *et al.* 2006; Zhou *et al.* 2007).

When the groundwater has been severely salinized, it implies that the groundwater salinization may have been caused by heavy pumping rather than natural processes, as the natural process may take a longer time to salinize the aquifer. In this case, other important factors should also be considered. Therefore, this study is to identify and evaluate the main factors affecting the groundwater quality as well as other minor processes. A hydrochemical model was used to analyze and validate the groundwater chemistry process in the aquifer and this information indeed is very useful for understanding the dynamic and threatened small tropical island's environment. This study may contribute much to the limited data and information on relationships between the status and trends of physicochemical parameters as evidence of seawater

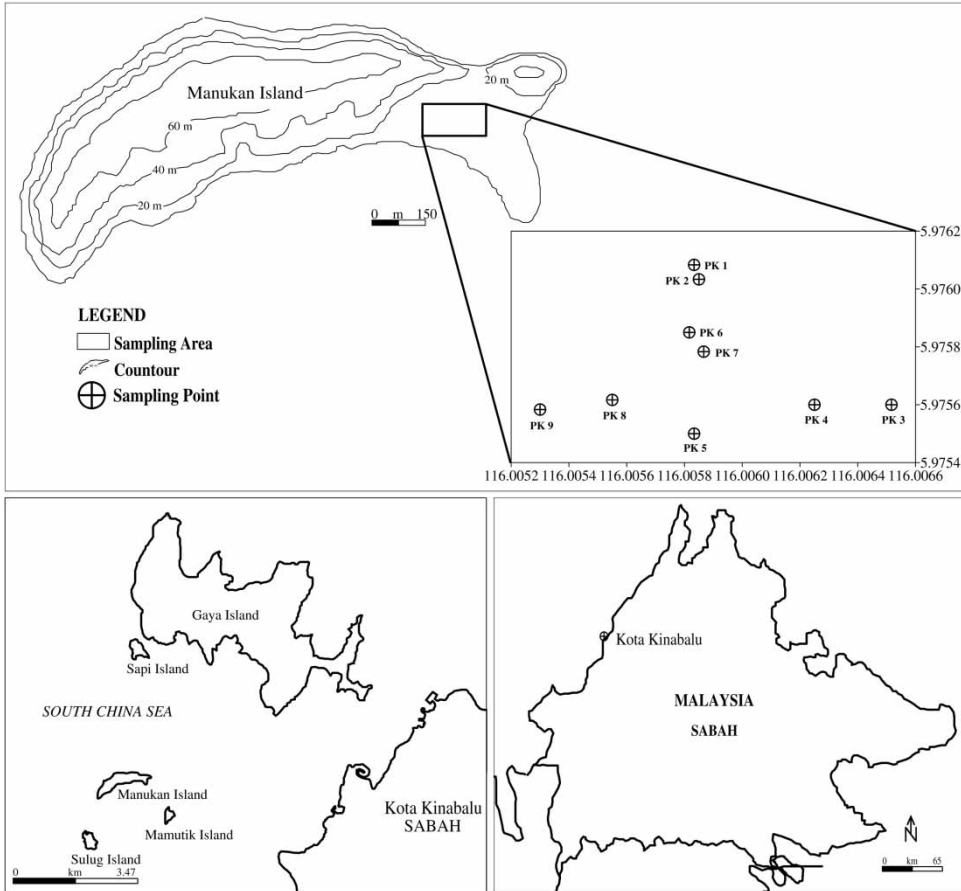
intrusion into shallow aquifers of small tropical islands, especially in Malaysia.

## MATERIALS AND METHODS

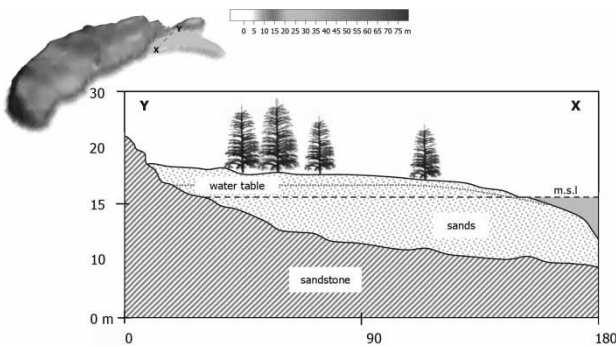
### Study site

Located approximately 7.5 km from Kota Kinabalu, the capital city of Sabah, Malaysia (Figure 1), Manukan (5°57'–5°58' N and 115°59'–116°01' E) is one of the protected islands of the Tunku Abdul Rahman Park. Covering an area of 206,000 m<sup>2</sup> (Abdullah 2001), the island falls into a very small island category. It is crescent-shaped, being 1.5 km long and 3 km wide in the middle. Almost 80% of the island is high relief and covered with forest (western side of the island), whereas the low lying area of the island is developed for tourism activities (eastern part of the island). Groundwater pumping has been practised over the years to provide the water supply.

Geologically, Manukan Island was isolated from the mainland about one million years ago (Basir *et al.* 1991). The island is underlain by interbedded sandstone and shales of the Middle Miocene Crocker Range Formation. The main aquifer of Manukan Island comprises quaternary carbonates which originated from coral deposits and is overlain by quaternary coarse sandy alluvium. In the lowland area, the thickness may reach 12 m with equal amounts of carbonate, sands and finer materials (Abdullah 2001; Figure 2). In general, the profiles in the low lying area are thinner compared to those of hilly areas as reported by Abdullah *et al.* (1997). Overall, the limited extent of the aquifer and low elevation of the island lead to very limited water storage (Aris *et al.* 2009). The aquifer capacity is primarily dependent on the aquifer thickness and distribution of the sands that are often found at sites near the coast. The groundwater recharge to the Manukan Island aquifer depends entirely on precipitation. Sabah has a warm and humid climate with annual rainfall between 2000–2700 mm, humidity between 80–90% all year round and temperatures from 21–32 °C. The climate is affected by the northeast and southwest monsoons, which are dominant from November to March and May to September,



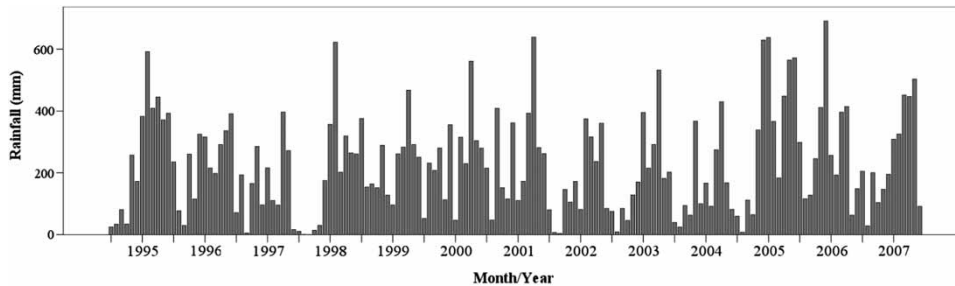
**Figure 1** | Schematic map showing the geographical locality of Manukan Island.



**Figure 2** | 3D elevation and the aquifer cross-section Y-X of Manukan Island.

respectively. The period between the monsoons is marked by heavy rainfall. The monthly average rainfall distribution from 1995 to 2007 is shown in Figure 3. At present, Manukan Island relies on this shallow aquifer for its groundwater supply. Dug wells are used for

extracting groundwater from its sandy aquifer. The wells have a diameter of 150 cm and depths between 55–78 cm from ground surface level (g.s.l). The wells are operated and controlled from a pump house and managed by the resort's operators. The water from the wells is pumped automatically using a water pump that was installed with a water level meter. The water from the wells is pumped into a storage compartment located near to the pump house before it is drained to the main water tanks situated on the hill. However, an explanation of the groundwater flow system, including reduced groundwater levels and pumping rates, is not significant since this study focuses on the geochemical processes in small-island groundwater. However, according to Abdullah (2001), the high relief area acts as a recharge area, with the groundwater flow directed downwards to low lying areas.



**Figure 3** | Average monthly rainfall distribution of the study area from 1995 to 2007.

### Sampling and analysis

Water samples were collected from the existing wells every two months between March 2006 and January 2007. A total of 162 samples were collected during the sampling period from nine wells located in the low lying area of the island (Figure 1). This study focused on the low lying area of Manukan Island so as to provide important information and implications of the hydrochemical facies of groundwater changes in response to salinization processes in the island's aquifer. The analyses of water samples were carried out to assess the major cations and anions including sodium (Na), potassium (K), calcium (Ca), magnesium (Mg), bicarbonate ( $\text{HCO}_3$ ), chloride (Cl) and sulfate ( $\text{SO}_4$ ). The water pH, temperature, electrical conductivity (EC), salinity, redox potential (Eh) and total dissolved solids (TDS) were measured in the field. Water samples were filtered at the time of collection using a 25 mm puradisc syringe filtration unit of 0.45  $\mu\text{m}$  pore size Whatman Milipore<sup>®</sup>, acidified to pH 2 with concentrated  $\text{HNO}_3$  acid and analysed using flame (air-acetylene burner) atomic absorption spectrometry (FAAS – Zeeman AAS Z-5000, Hitachi, Japan).  $\text{SO}_4$  was measured using a HACH (DR/2040 – Loveland, CO, USA) meter, while Cl and  $\text{HCO}_3$  were analysed using argentometric and titration methods (APHA 1995), respectively.

The pH, water temperature, EC, salinity, Eh and TDS were determined by means of a single meter probe for each parameter. The instruments were a WTW pH 315i for pH and temperature, a WTW Cond. 315i for EC, an Oakton 100 series for TDS and salinity, and an ORP Scan, Eutech for redox potential determinations. The instruments were calibrated using specific calibrating solutions. A mean value was calculated for each parameter, with the standard

deviation (SD) being used as an indication of the precision of each parameter measured in triplicate samples. Geochemical checks on the quality and correctness of the analyses were made for all 162 groundwater samples collected. The check for accuracy of the analysis employed the measured TDS to calculated TDS ratio, with the accepted ratio of the measured and the calculated TDS value being  $1.0 < \text{measured TDS}/\text{calculated TDS} < 1.2$ . Based on these criteria, all the groundwater samples were found to be within the ratio of 1.0 to 1.2. The excellent agreement between calculated and measured TDS means that the data collected were internally consistent and correct. Thus, the data were then processed with further statistical analyses. Overall, the analyses procedures adopted in this study were based on the methods described in APHA (1995).

### Data analysis

Available data obtained in this study consists of a relatively large number of analysed samples ( $n = 162$ ) and involved 13 parameters. Preliminary work was done on the data matrix which included the assembling and transformation of the data. Data which were below the detection limit were complemented with values equal to half the detection limit. Normal distribution tests were carried out with the help of the  $W$  (Shapiro–Wilk) test; the agreement of the distribution of the analysed ions and physicochemical parameters of water with the normal distribution were checked. Variables with a distribution other than normal were subjected to a transformation. In the case of variables, where their post-transformation distribution differed significantly from the normal one, those parameters were not taken into consideration when making the statistical analyses. In this study, the correlations between pairs of groundwater quality

parameters were primarily evaluated using Spearman's  $R$  coefficient, a non-parametric test that is often used to evaluate the correlation structure between water quality observed parameters with non-normal distributions, as suggested by Wunderlin *et al.* (2001).

The principal objective of this applied statistical method is to characterize the groundwater environment through discretization of the groundwater hydrochemical system into a limited number of hydrochemical regimes; and the factor scores are also estimated for use in the CA. In this study, factor scores were obtained by the regression method ( $R$  mode) for interpreting commonly collected groundwater quality data in terms of specific hydrochemistry processes. The CA was performed to understand the groundwater hydrochemical system through discretization of the study area into groundwater zones using the factor scores computed from the FA approach. Since the inclusion of unnecessary and mutually dependent variables often makes the understanding of the principal background hydrochemical process difficult (Suk & Lee 1999; Liu *et al.* 2008), the use of factor scores computed from FA is more effective when all the major ion data are used than when cation and anion concentrations are separated (Ashley & Llyod 1978). With this procedure, there is a fair chance that classification error due to outliers (data error) or multicollinearity will be reduced, since factors are independent of each other and have simple relations that provide insight into the underlying structure of the variates (Suk & Lee 1999) compared to straightforward spatial plots of individual determinants.

Groundwater in a small tropical island aquifer is a complex and dynamic system and fluctuations of groundwater constituents are dynamic responses of a groundwater system to its input (i.e. freshwater recharge) and output (i.e. groundwater abstraction). Both input and output may affect the groundwater chemistry and constituents, which processes vary at different spatial and temporal scales. As a result, groundwater chemistry composition may vary over multiple spatial and temporal scales with no single characteristic spatial and temporal scale (Li & Zhang 2007). As pointed out by Deutsch & Hewert (1996) and Liu *et al.* (2004), realistic simulation of subsurface system flow requires a detailed description of hydraulic properties, especially for transport simulations that are sensitive to small-scale heterogeneity. Spatial variability is complex

and dynamic because (1) spatial patterns are difficult to predict, (2) the variance of observed values is high (especially for permeability) and (3) acquiring exhaustive measurements of these properties is expensive, as outlined by de Vries *et al.* (2009). It is well known that geophysical and physicochemical fields are generally characterized by a wide-range scaling (implying strong power law resolution dependencies) effect. This scaling property stems from the fact that most of the quantities measured obey non-linear laws, which usually generate non-stationarities. For example, a study conducted by Zhang & Schilling (2004) concluded that physical parameters of the groundwater (groundwater levels and base flow) of an aquifer may fluctuate as temporal fractals with an effect from physical characteristics of the aquifer such as specific yield and transmissivity. A recent study conducted by Praveena *et al.* (2010) found that decreasing freshwater storage and mixing of seawater and freshwater were due to irregularities in the patterns of recharge rate and overpumping. Aris *et al.* (2010) found that the mixing between freshwater-seawater has created diversity in the geochemical processes of Manukan Island's aquifer and altered the freshwater and seawater mixture away from the theoretical composition line, whereas the results from reactive transport modelling confirmed that the migration of seawater into the fresher parts of the aquifer were also due to the low topography of the study area in Manukan Island, making this aquifer prone to frequent inundations of seawater aggravated by continuous groundwater abstraction from its sandy aquifer. These non-stationarities often conceal existing correlations in the examined data and therefore, instead of the application of conventional methods, new analytical techniques capable of eliminating the non-stationarities in the data should be utilized. The study by de Vries *et al.* (2009) found that non-stationary training images can be used to generate suitable non-stationary facies distributions. For instance, nowadays, the wavelet technique and the detrended fluctuation analysis (DFA) are among the most often used tools along these lines. DFA was first introduced by Peng *et al.* (1994) and represents an extension of the fluctuation analysis, which is affected by non-stationarities. Very recently, DFA and multi-DFA have been applied to the data observed in several complex systems (Kantelhardt *et al.* 2002; Li & Zhang 2007) where the problem of environmental data reduction and

interpretation can be more easily handled through application of robust multivariate methods and exploratory data analysis (Wenning & Erickson 1994). DFA are capable of allowing the detection of long-range correlations that are artificial non-stationarities compared to conventional methods (Peng et al. 1994). Conventionally, the usual technique of the interpretation of groundwater quality is only a univariate procedure which is inadequate to characterize simultaneous similarities and differences between samples or variables in a complex environment (Dixon & Chiswell 1996).

The hydrogeochemical saturation indices (SI) were computed using the PHREEQC geochemical modelling program in order to assess the state of equilibrium between groundwater and the minerals present, in particular calcite, aragonite and dolomite (Parkhurst & Appelo 1999). The

output from the PHREEQC simulation was then used to validate the geochemical processes that occurred in the aquifer. SI is defined as  $\log(IAP/K_{sp})$ , where  $IAP$  is the ion activity product and  $K_{sp}$  is the equilibrium solubility product. Table 1 summarizes the techniques used as well as its application.

## RESULTS AND DISCUSSION

### Factor analysis

Table 2 shows the descriptive statistics for all analysed parameters and the distribution of major ions of the studied area is shown in Figure 4. Eigenvalues of three factors were extracted from the obtained analytical data as they

**Table 1** | Summary of the adopted techniques and their applications

Type	Technique	Purpose	Application	Reference
Summary	Factor analysis	Explanatory	Determine underlying factors responsible for variations	Ashley & Llyod (1978); Bierman et al. (2011)
Classification	Cluster analysis	Exploratory and confirmatory	Classification of objects into groups displaying similar properties	Bierman et al. (2011)
Modelling	Geochemical modelling (PHREEQC)	Exploratory, simulation and validation	Capable of simulating a wide range of geochemical reactions	Parkhurst & Appelo (1999)

**Table 2** | The physico-chemical properties and major ions of groundwater in the study area ( $n = 162$ )

	Unit	Range	Mean	SD	CV%
pH		6.59–7.97	7.27	0.19	3
Eh	mV	–55.3–22.2	–17.14	15.24	–89
Temp.	°C	26.3–29.4	27.89	0.74	3
TDS	mg L <sup>-1</sup>	1133–8294	4535	1506	33
EC	mS cm <sup>-1</sup>	0.30–12.26	4.79	1.86	39
Salinity	ppt	0.29–7.40	2.71	1.25	46
Ca	mg L <sup>-1</sup>	60–866	390	93.14	24
Mg	mg L <sup>-1</sup>	3–298	110	51.41	47
Na	mg L <sup>-1</sup>	104–2780	1181	484.16	41
K	mg L <sup>-1</sup>	4–94	37	17.30	47
Cl	mg L <sup>-1</sup>	340–4099	2118	788.72	37
HCO <sub>3</sub>	mg L <sup>-1</sup>	195–524	328	52.03	16
SO <sub>4</sub>	mg L <sup>-1</sup>	25–660	256	121.85	48

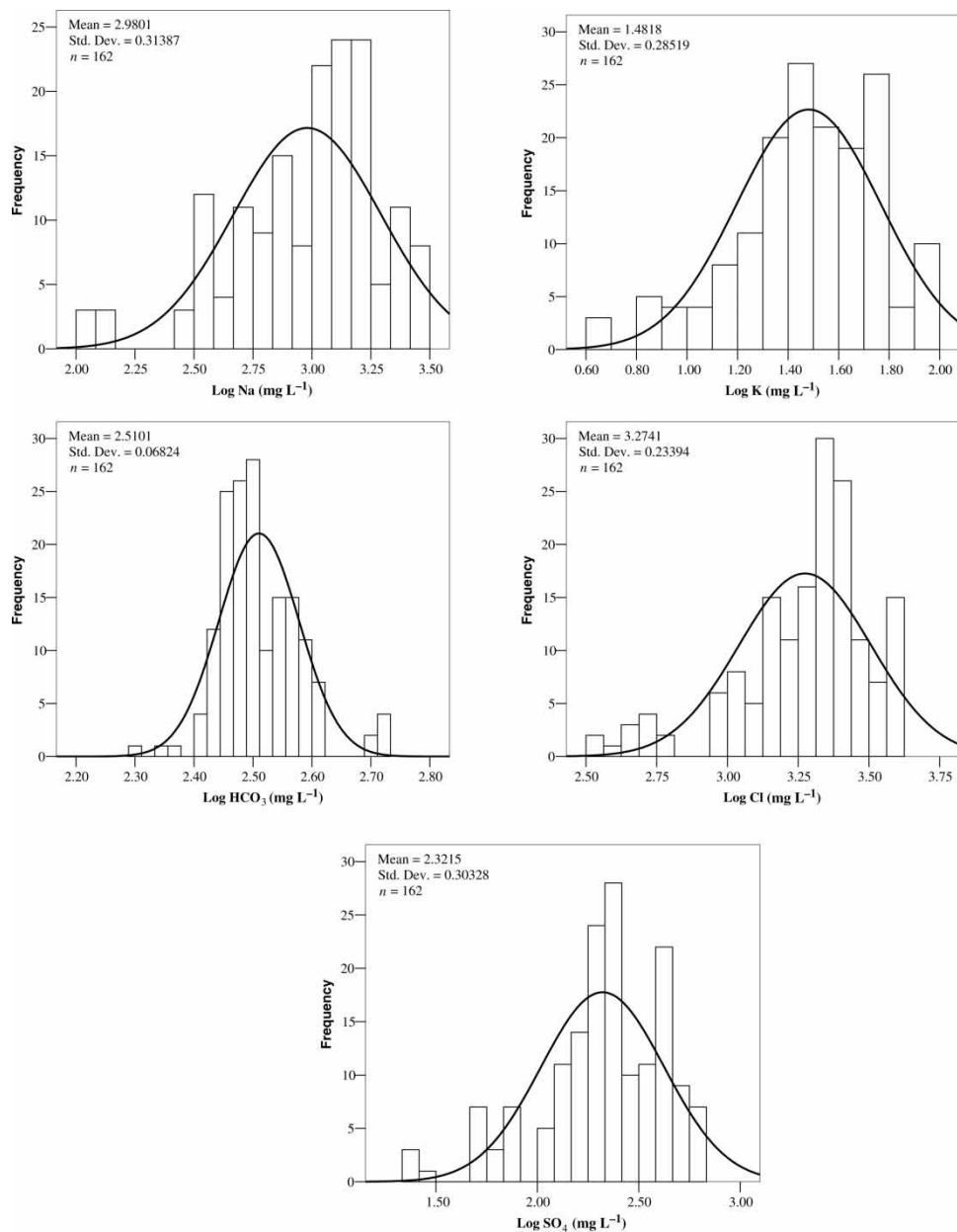


Figure 4 | Distribution of the major ions.

have eigenvalues greater than 1 (Table 3). It reveals that the eigenvalues of the three factors explained 80.64% of the total variance (Table 3). Complex variables may have loadings on more than one factor, and they make the interpretation of the output difficult and more complex. Therefore, rotation may be necessary (Praveena 2008). According to Davis (1986) and Praveena (2008), varimax rotation is to move each of the components to the positions

so that projections from each variable onto the component axes are either near the extremities or near the origin. Varimax rotation changes the factor loadings so the original variables have either a high positive or negative correlation (near  $\pm 1$ ) with a factor, or correlation, near to zero. The absolute values of factor loadings of over 0.6 were considered as strong correlation and shaded in Table 3. This is to elucidate the relationships between the factors and the

**Table 3** | Loadings for varimax-rotated factor matrix for three factor model

Parameters	Factors		
	1	2	3
EC	0.931	0.034	-0.031
Salinity	0.926	0.059	-0.055
K	0.913	0.302	-0.114
Mg	0.912	0.077	-0.047
SO <sub>4</sub>	0.885	0.374	-0.084
Cl	0.696	0.543	0.134
Ca	-0.228	0.853	0.325
TDS	0.636	0.741	0.025
Na	0.540	0.668	-0.178
HCO <sub>3</sub>	-0.163	-0.364	0.231
Eh	-0.143	-0.059	0.941
pH	0.078	-0.093	-0.918
Temperature	0.530	-0.018	0.611
Eigenvalues	6.56	2.39	1.54
% of variance	50.43	18.38	11.83
Cumulative % of variance	50.43	68.81	80.64

hydrochemical data. A factor score associated with each sampling station was determined and plotted to illustrate the spatial characteristics of groundwater chemistry in Manukan Island (Figure 5). The spatial factor loading maps were illustrated by the interpolation from point-to-point groundwater sampled wells' location. The contours were not extrapolated into the hilly terrain area because no sampling station (well) exists in that area.

The three factors (Factor 1, Factor 2 and Factor 3) can be used to explain the background hydrochemical processes without losing any significant characteristics. Factor 1 (F1), which explains 50.43% of the total variance, has strong positive loadings on EC, salinity, K, Mg, SO<sub>4</sub> and Cl. The hydrochemical items in F1 are the dominant solutes in seawater, especially SO<sub>4</sub> and Cl, and are significantly correlated with EC and salinity. This particular pattern, in conjunction with high positive loadings of EC, salinity, K, Mg, SO<sub>4</sub> and Cl, reflects the intrusion of seawater into the shallow unconfined aquifer of Manukan Island. The main reason for serious groundwater table drawdown, which leads to intrusion of seawater into the island's aquifer, is over-pumping. According to a study done by Mao *et al.* (2005) in Ardeer (Scotland), seawater intrusion occurs in two ways: one is

the advancement of the seawater at the bottom of the aquifer resulting from the density difference between seawater and freshwater. The other was the infiltration of seawater through the beach. As the results of this paper showed, the seawater intrusion occurs in this study area. The highest observed salinity and Cl concentration among the sampling stations were 7.10 ppt and 4099 mg L<sup>-1</sup>, respectively. From the spatial factor loading map (Figure 5(a)), the high F1 score values are indicated by the dark green, representing high salinization areas, and the low score values are shown by the light colour, denoting low salinization areas. F1 has positive factor loadings in PK 4, PK 6 and PK 7 while they are negative in PK 1, PK 2, PK 3, PK 8 and PK 9 (Figure 6). Groundwater in the south-east of Manukan (PK 4, PK 6 and PK 7) is more saline than that in other sampling locations. Groundwater in the north-central and western areas of Manukan is less saline since the groundwater in those particular areas receives direct fresh groundwater recharge from the hilly terrain area. The salinization pattern in all wells agrees with the findings in its salinity trends over F1 scores (Figure 7). Figure 8 shows the overall plot of water composition on a Piper diagram, including seawater signature. The plots deduced that the water type found in the study area within the sampling episodes were Na-Cl, Ca-Cl and Mg-Cl. The influence of marine chemistry on the groundwater samples is clearly shown in Figure 8. From the Piper diagram, it is deduced that the most important hydrochemical process in the Manukan Island aquifer is the salinization process through mixing of fresh groundwater and seawater, though processes within the aquifer that augment the mineral content of the fresh groundwater might also be important.

Factor 2 (F2) accounts for 18.38% of the total variance. It shows high loadings on Ca (0.853), TDS (0.741) and Na (0.668). Although the Na concentration in seawater is relatively high, it is not included in F1. Thus, it can be concluded that this second factor is related to the cation exchange process that occurred simultaneously during seawater-freshwater mixing. In combination, the depletion and enrichment of cations (Ca and Na) in the groundwater and the composition of the exchanger suggest the development of a chromatographic ion exchange pattern in the Manukan Island aquifer. In seawater, the most dominant ions are Na and Cl, and sediment in direct contact with seawater due to seawater intrusion will have mostly Na in the



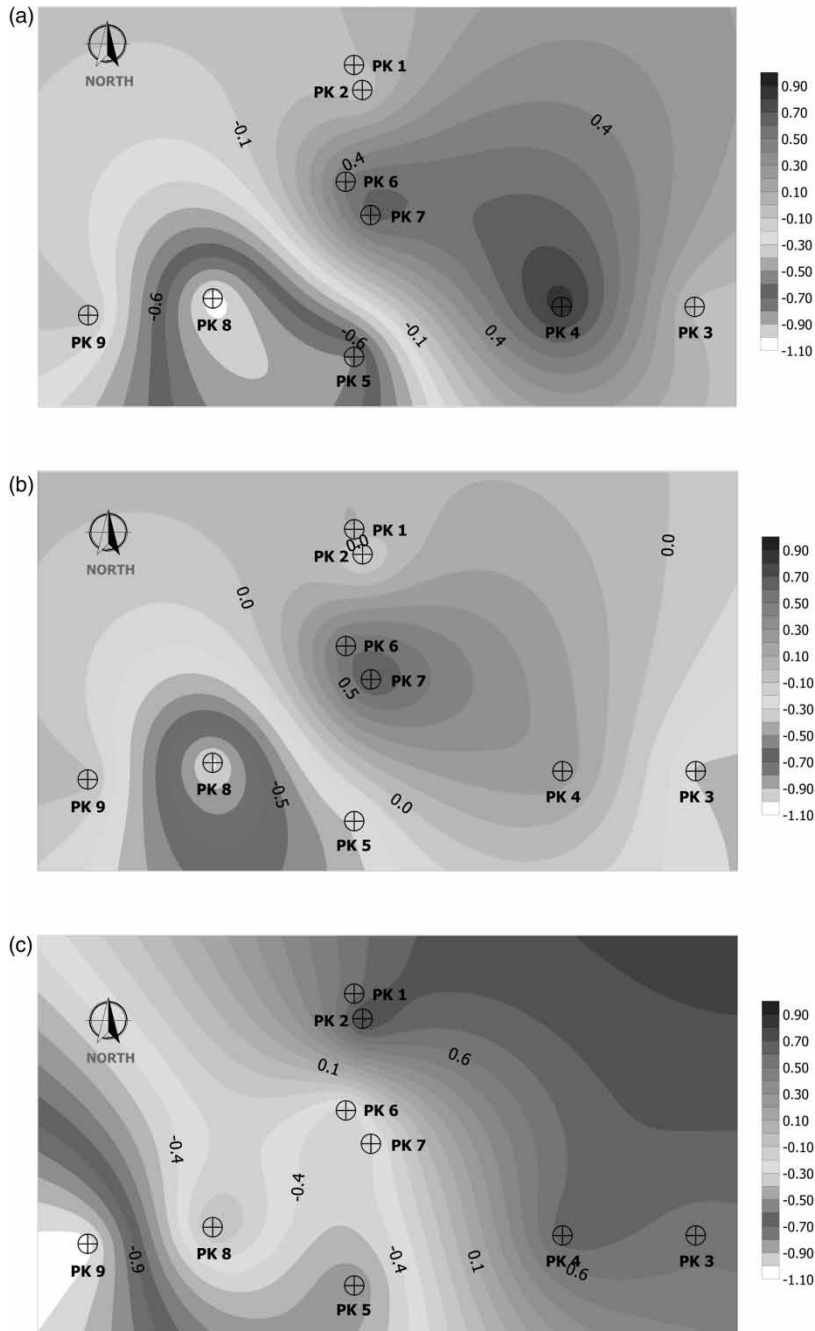
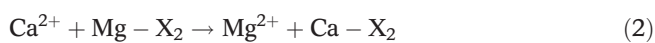
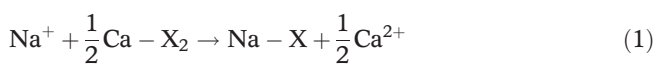


Figure 5 | Spatial distribution scores of (a) Factor 1, (b) Factor 2 and (c) Factor 3.

aquifer's matrix (Appelo & Postma 2005) and the inferred sequential reactions are as follows:



where X indicates the soil exchanger. Equation (1) was normally observed in the aquifer affected with seawater transition while Equation (2) normally indicates that the aquifer is experiencing freshening, as shown by its Ca-Cl (observed at PK 8) water type. Apart from the observed overall salinization pattern, a minor freshening process of the aquifer can be

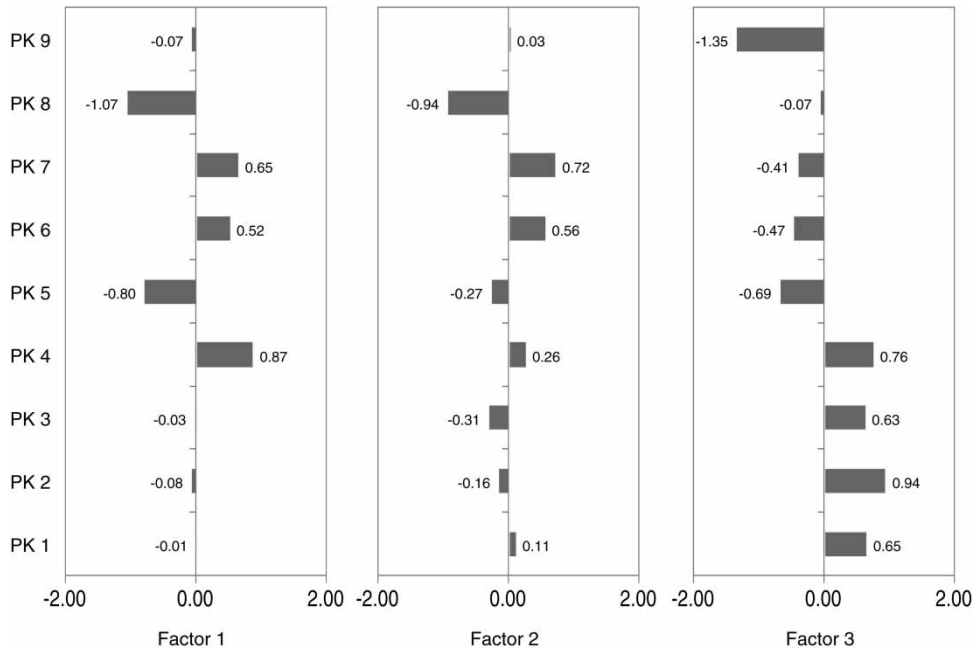


Figure 6 | Results of factor scores for Manukan Island's groundwater.

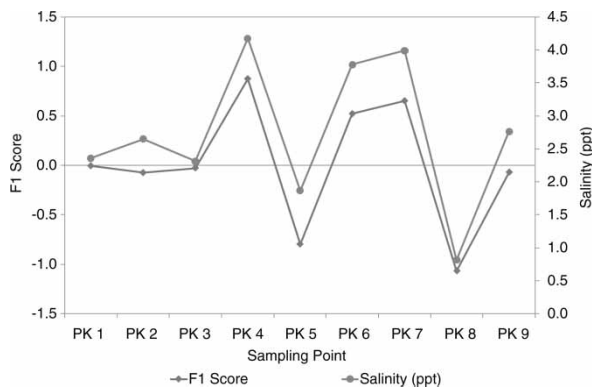


Figure 7 | Relationship between F1 score and salinity for the studied groundwater.

inferred from the depletion of Na and enrichment of Ca, which was observed at the monitoring wells near to the recharge area (western areas of Manukan Island). The chromatographic sequence is caused by the sequential displacement starting with the cations with the highest affinity for the exchanger. During the seawater intrusion, when seawater mixes with freshwater, the cation affinity order is normally  $\text{Na} > \text{K} > \text{Ca} > \text{Mg}$ , with Ca being displaced from the exchanger in the first order and Na eventually dominating the water and the exchanger. An adverse sequence will be observed in a freshening aquifer and it was found that some of the dissolved

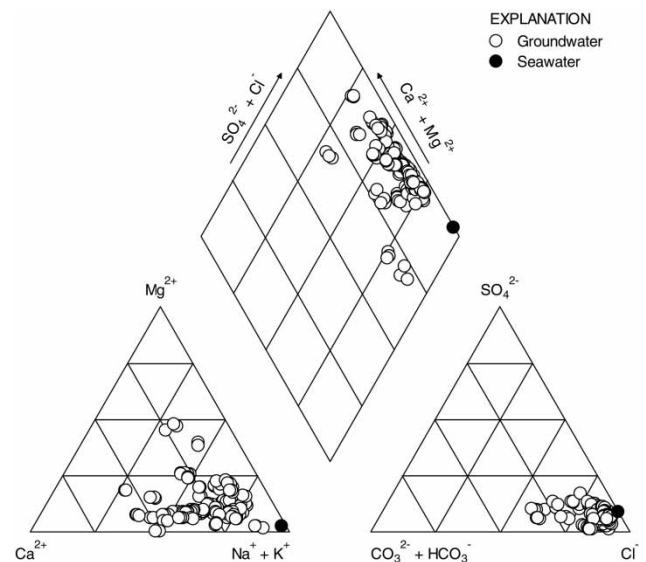


Figure 8 | Piper plot for studied groundwater of Manukan Island.

Na and K in the displacing seawater may exchange directly with adsorbed Ca on the exchanger (Aris et al. 2010). The effect of the cation exchange process may significantly control the TDS concentrations of the groundwater, as can be seen from the loadings of F2. As described previously in the case of F1, positive loadings appear mostly in the southern

area while negative loadings only appear in PK 8 (Figure 5(b)). Factor score studies reflect the stationwise variation of the geochemical factors controlling the water chemistry. Scores showing negative values denote areas essentially unaffected by the said factors while most affected areas are denoted by extreme positive scores and a near-zero score may be considered as areas affected to an average degree to unaffected. From that, it is clearly seen that all sampling locations were affected by the cation exchange process as shown by its water chemistry composition, with the exception for PK 8. This is attributed to the natural condition of PK 8 where it may not be significantly affected by seawater intrusion since it is an abandoned well. As the groundwater moves from elevated areas, the groundwater in PK 8 becomes more diluted, so the factor scores have gradually reached  $-0.1$ .

Factor 3 (F3) has high loadings on Eh (0.941), pH ( $-0.918$ ) and temperature (0.611) and accounts for 11.83% of the total variance. From the spatial distribution map (Figure 5(c)) and from the relationship plot between F3 scores and Eh values (Figures 6 and 8), PK 1, PK 2, PK 3 and PK 4 were characterized with positive F3 scores while PK 5, PK 6, PK 7, PK 8 and PK 9 were characterized with negative F3 scores. This factor clearly indicates the interactions of water with the surrounding chemical and biological processes that agree with the aquifer redox sequences. This reduction of groundwater pH values compared to seawater pH values might be due to the natural processes of biological nitrification that produced organic acids from decaying organic matter. The inverse relation between pH and Eh thus explains the redox reactions such as denitrification because these processes raise alkalinity and pH, as can be seen from the rise of pH in the study area which ranges from 6.59–7.97. This can be clearly seen from the negative loading on pH and high positive loading on Eh (Figure 9). Factors such as low permeability, the stagnant hydrodynamic regime of groundwater, the presence of natural organic matter and permanent leakage of organic pollutants cause the decrease of redox potential through a saturated soil section. In an isolated environment, organic matter forms a redox couple with the inorganic substances available. The organic matter becomes oxidized and the inorganic matter is reduced.

Accordingly, the redox factor (F3) includes the pH term with strong negative loading ( $-0.918$ ); the same observation

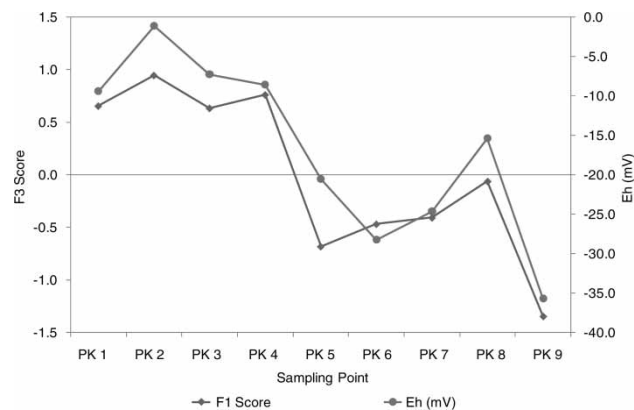


Figure 9 | The relationship between average F3 score and Eh of the studied groundwater.

was made by Liu *et al.* (2008) in Kinmen Island. Since the pH value depends on the carbonic species equilibrium, the  $\text{HCO}_3^-$  concentration may significantly control the pH value. As the equilibrium is significantly affected by temperature and pressure, it obviously changes the pH. However, since the  $\text{HCO}_3^-$  parameter is not included in the F3 factor and its correlation to pH, Eh and temperature is low and insignificant ( $r = -0.074$ ,  $r = 0.245$  and  $r = -0.006$ , respectively), it is concluded that seawater intrusion as the main process that controls the chemistry of Manukan Island's groundwater has an insignificant impact on the chemistry of  $\text{HCO}_3^-$  in the groundwater and therefore may only reflect its equilibrium conditions. This equilibrium condition may be controlled by the temperature of the groundwater. The temperature of groundwater recorded during this study ranged from 26.3–29.0 °C. A higher temperature of seawater compared to the groundwater of the island, with a small difference of between 3.2–4.1 °C, was observed during the study period. Such a high temperature reflected the impact of environmental disturbances on the groundwater system of the small island. The rise of the groundwater temperature could be attributed to the intrusion of seawater of a higher temperature from the sea into the aquifer or could be an indication of larger ground exposure to direct sunlight as a result of lesser vegetation on the island, especially on the low relief side (study focused).

### Cluster analysis

Three main clusters are identified from nine sampling stations using CA – Group 1, Group 2 and Group 3 (Figure 10) as well as the relationship between factor

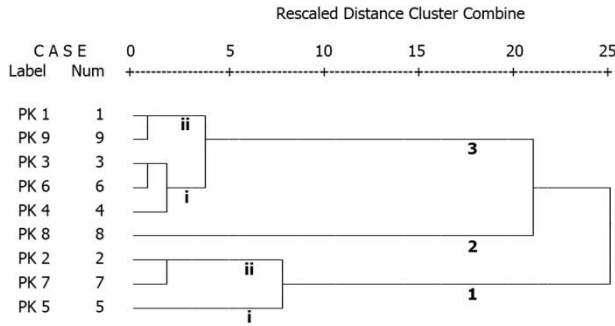


Figure 10 | Dendrogram groups of sampling stations determined in CA.

scores obtained from FA analysis and groups for each wells (Table 4). From the CA result, Group 1 includes sampling stations PK 2, PK 5 and PK 7, representing the groundwater quality having similar conditions. From the well locations, it clearly shows that the three sampling stations lie along the same groundwater flow line (Figure 11). Since PK 2 served as a major well for supplying water to the chalets nearby where pumping activity was carried out extensively over a decade, the movement of seawater may contaminate wells along the flow line (PK2, PK 5 and PK 7). This may explain why the three wells have similar groundwater chemistry conditions. Wells located near to the coast are not affected and belong to the same proximity group. It is because these wells are abandoned and no pumping activities were done which allows the seawater to intrude into the fresh groundwater wedge. Group 2 contains only sampling station PK 8 with a high score in F3, revealing the groundwater having a major impact from redox sequences as discussed

Table 4 | The relationship between factor scores determined by FA and groups for each well identified by CA

Sampling station	Score			Group <sup>a</sup>
	Factor 1 (F1)	Factor 2 (F2)	Factor 3 (F3)	
PK 1	-0.01	0.11	0.65	G3:ii
PK 2	-0.08	-0.16	0.94	G1:ii
PK 3	-0.03	-0.31	0.63	G3:i
PK 4	0.87	0.26	0.76	G3:i
PK 5	-0.80	-0.27	-0.69	G1:i
PK 6	0.52	0.56	-0.47	G3:i
PK 7	0.65	0.72	-0.41	G1:ii
PK 8	-1.07	-0.94	-0.07	G2
PK 9	-0.07	0.03	-1.35	G3:ii

<sup>a</sup> G1: Group 1, G2: Group 2, G3: Group 3.

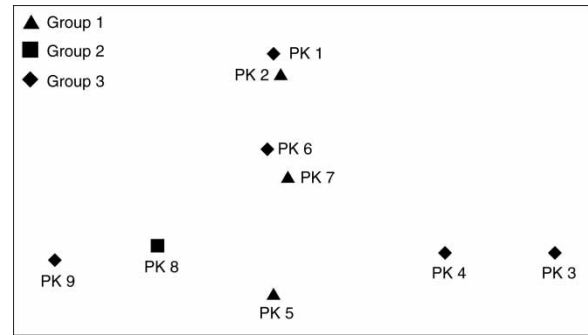


Figure 11 | Results of clustering shown by sampling points.

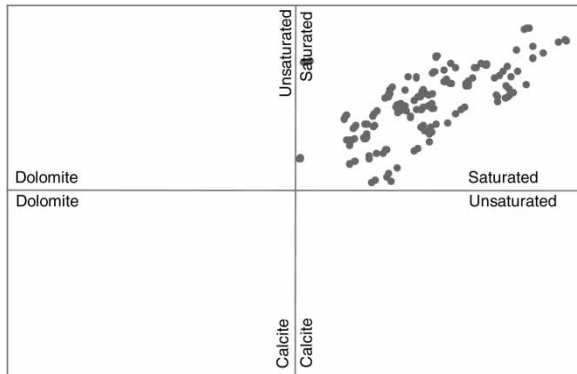
previously. However, Group 3 includes the remaining sampling stations (PK 1, PK 3, PK 4, PK 6 and PK 9) and exhibits more complicated structures in the groundwater chemistry. Group 3 was further divided into two groups – Group 3:i and Group 3:ii. Sampling stations in Group 3:i have high scores in F1 and F2, but are considerably different for F2, especially for the mild cation exchange process in sampling station PK 6 which is the last to be combined into this group. Group 3:ii possesses high scores in F3 but low scores in F1 (less saline intrusion).

### Hydrochemical modelling analysis

From the calculation, most of the groundwater samples are at or close to/above saturation with respect to calcite and dolomite (Table 5). The saturation index (SI) plot of calcite versus dolomite (Figure 12) for the groundwater demonstrates that the waters are supersaturated with respect to dolomite and calcite and the dolomite SI values are higher than the calcite SI values. The values are relatively high for the aquifer intruded by seawater as in this study. The over-saturated values were probably due to more carbonate minerals dissolved under closed conditions as the initial water enters from an intermediate to deep aquifer, thus

Table 5 | Range of values for selected SI of aragonite, calcite and dolomite for present studied groundwater (n = 162)

	Aragonite	Calcite	Dolomite
Mean	0.51	0.65	1.11
SD	0.26	0.26	0.51
Min	-0.08	0.06	0.03
Max	1.09	1.23	2.35



**Figure 12** | Saturation indices plot for calcite and dolomite.

simultaneously exchanged Ca and Mg with Na which led the water to remain at, or slightly above, saturation states with respect to the carbonate minerals. The pH of groundwater influenced the saturation state of calcite and aragonite, where Ca and Mg are usually transferred to a solid phase at high pH values, and therefore their concentrations are controlled by mineral precipitation.

The low significant correlation ( $r = -0.152$ ;  $p < 0.01$ ) between Ca and Mg is possibly attributed to the precipitation state of dolomite, calcite and high-Mg calcite. Precipitation took place during the cation exchange process as an extending effect from seawater intrusion as discussed previously. The cementing material of the Manukan Island aquifer mainly consists of carbonate mineral species, as indicated by an early study done by *Abdullah et al.* (1997). In calcite crystals, Mg substituted Ca owing to the similarities in ionic radius and charge. Such explanation is supported by the SI value of both calcite and aragonite which indicated supersaturation values (Table 5). The increase of Ca concentration in the groundwater samples that was also observed during the study period was probably not due to calcite dissolution since  $\text{HCO}_3^-$  did not increase much when compared to the overall data. Such a weak correlation ( $r = -0.103$ ) among major anions and cations reveals that dissolved salts from seawater had an insignificant impact on the  $\text{HCO}_3^-$  concentration in the groundwater.

Thus, the effect of seawater intrusion which was manifested by the PHREEQC output validated the factors controlling the groundwater chemistry of the study area, especially on the salinization process and cation exchange process. Ion exchange, related to aquifer salinization, primarily affects the saturation state for carbonate minerals

by the release of Ca from the exchanger, which may simulate further carbonate mineral precipitation. From the PHREEQC calculation, the most important geochemical processes are ion exchange where seawater solutes are displacing fresh groundwater ions from the sediment surface and carbonate precipitation (as extended effects from the aquifer salinisation process).

## CONCLUSIONS

The salinization process attributed to seawater intrusion, cation exchange process and redox processes are found to be the three main processes that control characteristics of the groundwater in the Manukan Island aquifer. The factor scores calculated for each monitoring well were spatially plotted to illustrate the characteristics of the groundwater and it was found that south-east Manukan is more saline while groundwater in the north-central and western areas of Manukan is less saline since the groundwater in those particular areas received direct fresh groundwater recharge from the hilly terrain area. Three main clusters (Group 1, Group 2 and Group 3) computed from the factor scores identified the interactions among these groups and explain spatially the similar patterns of the groundwater chemistry in the island. Results of common properties determined by the FA approach were further compared with those of similar properties identified by CA. The computation reveals that the variance of the seawater intrusion trait exceeds that of the cation exchange process trait, while the variance of cation exchange process trait exceeds that of the redox sequence trait. Therefore, factor properties with a large variance were identified among wells prior to grouping. Furthermore, based on the hydrochemical calculation, the effects of seawater intrusion on the groundwater chemistry were analysed and validated by the salinization and cation exchange processes manifested from the PHREEQC output.

## ACKNOWLEDGEMENTS

This work has been supported by the Ministry of Science, Technology and Innovation (MOSTI), Malaysia through

the ScienceFund research project SCF0039-SEA-1/2007. Permission from the Sabah Parks Trustees for the study site exploration is gratefully acknowledged.

## REFERENCES

- Abdullah, M. H. 2001 *Phreatic Water Extraction from Shallow Aquifer of a Small Island*. PhD thesis. Universiti Teknologi Malaysia, Malaysia.
- Abdullah, M. H., Musta, B. & Tan, M. M. 1997 A preliminary geochemical study on Manukan Island, Sabah. *Borneo Sci.* **5**, 43–51.
- Alberto, W. D., Pilar, D. M. D., Valeria, A. M., Fabiana, P. S., Cecilia, H. A. & Angeles, B. M. D. L. 2001 *Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Cordoba-Argentina)*. *Wat. Res.* **35** (12), 2881–2894.
- APHA 1995 *Standard Methods for the Examination of Water and Wastewater*, 19th edition. American Water Works Association, Water Environment Federation, Washington, District of Columbia.
- Appelo, C. A. J. & Postma, D. 2005 *Geochemistry, Groundwater and Pollution*, 2nd edition. A.A. Balkema, Rotterdam.
- Aris, A. Z., Abdullah, M. H. & Kim, K. W. 2009 Hydrogeochemistry of groundwater in Manukan island, Sabah. *Sains Malaysiana* **11** (2), 407–413.
- Aris, A. Z., Abdullah, M. H., Praveena, S. M., Yusoff, M. K. & Juahir, H. 2010 Extenuation of saline solutes in shallow aquifer of a small tropical island: a case study of Manukan Island, North Borneo. *EnvironmentAsia* **3**, 84–92.
- Ashley, R. P. & Lloyd, J. W. 1978 An example of the use of factor analysis and cluster analysis in groundwater chemistry interpretation. *J. Hydrol.* **39** (3–4), 355–364.
- Basir, J., Sanudin, T. & Tating, F. F. 1991 Late Eocene planktonic foraminifera from the Crocker Formation, Pun Batu, Sabah. *Warta Geologi* **14** (4), 1–15.
- Bierman, P., Lewis, M., Ostendorf, B. & Tanner, J. 2011 A review of methods for analysing spatial and temporal patterns in coastal water quality. *Ecol. Indicat.* **11** (1), 103–114.
- Chae, G.-T., Yun, S.-T., Kim, K. & Mayer, B. 2006 Hydrogeochemistry of sodium-bicarbonate type bedrock groundwater in the Pocheon spa area, South Korea: water-rock interaction and hydrologic mixing. *J. Hydrol.* **321**, 326–343.
- Choi, B. Y., Kim, H. J., Kim, K., Kim, S. H., Jeong, H. J., Park, E. & Yun, S. T. 2008 Evaluation of the processes affecting vertical water chemistry in an alluvial aquifer of Mankyong Watershed, Korea, using multivariate statistical analyses. *Environ. Geol.* **54** (2), 335–345.
- Dalton, M. G. & Upchurch, S. B. 1978 Interpretation of hydrochemical facies by factor analysis. *Groundwater* **16** (4), 228–233.
- Davis, J. C. 1986 *Statistics and Data Analysis in Geology*, 2nd edition. Wiley, New York.
- de Vries, L. M., Carrera, J., Falivene, O., Gratacos, O. & Slooten, L. J. 2009 Application of multiple point geostatistics to non-stationary images. *Math. Geosci.* **41**, 29–42.
- Deutsch, C. V. & Hewet, T. A. 1996 Challenges in reservoir forecasting. *Math. Geol.* **28**, 829–842.
- Dixon, W. & Chiswell, B. 1996 Review of aquatic monitoring program design. *Wat. Res.* **30**, 1935–1948.
- Helena, B., Pardo, B., Vega, M., Barrado, E., Fernandez, J. M. & Fernandez, L. 2000 Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Wat. Res.* **34** (3), 807–816.
- Kantelhardt, J. W., Zschiegner, S. A., Bunde, E. K., Havlin, S., Bunde, A. & Stanley, H. E. 2002 Multifractal detrended fluctuation analysis of nonstationary time series. *Phys. A: Statist. Mech. Appl.* **316** (1–4), 87–114.
- Kim, J. H., Yum, B. W., Kim, R. H., Koh, D. C., Cheong, T. J., Lee, J. & Chang, H. W. 2003 Application of cluster analysis for the hydrogeochemical factors of saline groundwater in Kimje, Korea. *Geosci. J.* **7**, 313–322.
- Li, Z. & Zhang, Y. K. 2007 Quantifying fractal dynamics of groundwater systems with detrended fluctuation analysis. *J. Hydrol.* **336** (1–2), 139–146.
- Liu, C. W., Jang, C. S., Chen, C. P., Lin, C. N. & Lou, K. L. 2008 Characterization of groundwater quality in Kinmen Island using multivariate analysis and geochemical modelling. *Hydrol. Process.* **22** (3), 376–383.
- Liu, G., Zheng, C. & Gorelick, S. M. 2004 Limits of applicability of the advection-dispersion model in aquifers containing connected high-conductivity channels. *Wat. Res. Res.* **40**, W08308. doi: 10.1029/2003WR002735.
- Lu, H.-Y., Peng, T.-R. & Liou, T.-S. 2008 Identification of the origin of salinisation in groundwater using multivariate statistical analysis and geochemical modelling: a case study of Kaoshiung, Southwest Taiwan. *Environ. Geol.* **55** (2), 339–352.
- Mao, X., Enot, P., Barry, D. A., Li, L., Binley, A. & Jeng, D. S. 2005 Tidal influence on behaviour of a coastal aquifer adjacent to a low relief estuary. *J. Hydrol.* **327**, 110–127.
- Parkhurst, D. L. & Appelo, C. A. J. 1999 *User's Guide to PHREEQC (version 2) – A Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculations*. US Geological Survey, Washington, District of Columbia.
- Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E. & Goldberger, A. L. 1994 Mosaic organization of DNA nucleotides. *Phys. Rev. E* **49** (2), 1685–1689.
- Praveena, S. M. 2008 *A Study on Heavy Metals Distribution in Mangrove Sediment of Mengkabong, Sabah*. MSc thesis. Universiti Malaysia Sabah, Malaysia.
- Praveena, S. M., Abdullah, M. H. & Aris, A. Z. 2010 Modeling for equitable groundwater management. *Int. J. Environ. Res.* **4** (3), 415–426.
- Reghunath, R., Murthy, T. R. S. & Raghavan, B. R. 2002 The utility of multivariate statistical techniques in hydrogeochemical

- Studies: an example from Karnataka, India. *Wat. Res.* **36**, 2437–2442.
- Saager, R. & Sinclair, A. J. 1974 Factor analysis of stream sediment geochemical data from the Mount Nausen area, Yukon Territory, Canada. *Miner. Depos.* **9**, 243–252.
- Simeonov, V., Einax, J. W., Stanimirova, I. & Kraft, J. 2002 Environmetric modeling and interpretation of river water monitoring data. *Anal. Bioanal. Chem.* **374** (5), 898–905.
- Suk, H. J. & Lee, K. K. 1999 Characterization of a ground water hydrochemical system through multivariate analysis: clustering into ground water zones. *Ground Wat.* **37** (3), 358–366.
- Wenning, R. J. & Erickson, G. A. 1994 Interpretation and analysis of complex environmental data using chemometric methods. *Trends Analyt. Chem.* **13**, 446–457.
- Wunderlin, D. A., Diaz, M. D. P., Ame, M. V., Pesce, S. F., Hued, A. C. & Bistoni, M. D. 2001 Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Cordoba Argentina). *Wat. Res.* **35** (12), 2881–2894.
- Xie, Z., Sun, L., Phengfei, Z., Sanping, Z., Xuebin, Y., Xiadong, L. & Bangbo, C. 2005 Preliminary geochemical evidence of groundwater contamination in coral islands of Xi-Sha, South China Sea. *Appl. Geochem.* **20** (10), 1848–1856.
- Yammani, S. R., Reddy, T. V. K. & Reddy, M. R. K. 2008 Identification of influencing factors for groundwater quality variation using multivariate analysis. *Environ. Geol.* **55** (1), 1–16.
- Zhang, Y. K. & Schilling, K. 2004 Temporal scaling of hydraulic head and river base flow and its implication for groundwater recharge. *Wat. Res. Res.* **40**, W03504. 2881–2894.
- Zhou, F., Liu, Y. & Guo, H. 2007 Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong. *Environ. Monit. Assess.* **132**, 1–13.

First received 11 September 2009; accepted in revised form 2 September 2010. Available online 12 April 2011