Co-occurrence correlations of heavy metals in sediments revealed using network analysis

Lili Liu a,b, Zhiping Wang a,c, Feng Ju a, Tong Zhang a,*

a Environmental Biotechnology Lab, The University of Hong Kong, Hong Kong, China
b State Environmental Protection Key Laboratory of Environmental Risk Assessment and Control on Chemical Process, East China University of Science and Technology, 200237 Shanghai, China
c School of Environment Science and Technology, Shanghai Jiao Tong University, 200240 Shanghai, China

ARTICLE INFO

Article history:
Received 23 October 2013
Received in revised form 20 January 2014
Accepted 28 January 2014
Available online 20 February 2014

Keywords:
Network
Co-occurrence correlations
Normalization
Sediment
Heavy metals

ABSTRACT

In this study, the correlation-based study was used to identify the co-occurrence correlations among metals in marine sediment of Hong Kong, based on the long-term (from 1991 to 2011) temporal and spatial monitoring data. 14 stations out of the total 45 marine sediment monitoring stations were selected from three representative areas, including Deep Bay, Victoria Harbour and Mirs Bay. Firstly, Spearman’s rank correlation-based network analysis was conducted as the first step to identify the co-occurrence correlations of metals from raw metadata, and then for further analysis using the normalized metadata. The correlations patterns obtained by network were consistent with those obtained by the other statistic normalization methods, including annual ratios, R-squared coefficient and Pearson correlation coefficient. Both Deep Bay and Victoria Harbour have been polluted by heavy metals, especially for Pb and Cu, which showed strong co-occurrence with other heavy metals (e.g. Cr, Ni, Zn and etc.) and little correlations with the reference parameters (Fe or Al). For Mirs Bay, which has better marine sediment quality compared with Deep Bay and Victoria Harbour, the co-occurrence patterns revealed by network analysis indicated that the metals in sediment dominantly followed the natural geography process. Besides the wide applications in biology, sociology and informatics, it is the first time to apply network analysis in the researches of environment pollutions. This study demonstrated its powerful application for revealing the co-occurrence correlations among heavy metals in marine sediments, which could be further applied for other pollutants in various environment systems.

1. Introduction

In general, sediment is considered as a major ‘sink’ for the pollutants, especially for heavy metals and some persistent organic pollutants (POPs) (Zheng et al., 2002). The metals may come from both natural and anthropogenic processes. Researchers have tried a variety of normalization techniques to determine natural mineral-
and Huh, 1989), and total organic carbon (Dashkalakis and O’Conner, 1995). Although normalization has been used to study the distribution of elements in some cases, different mathematical normalization techniques may interpret data distribution patterns in significantly different ways and may not reflect the true distribution (Woods et al., 2012). Moreover, normalization techniques might overestimate the human-induced signature or effects.

Network has been widely used as a powerful tool in biology, mathematics, social science and computer science, to explore the interactions between entities or parameters (Albert et al., 2000; Proulx et al., 2005; Barberán et al., 2012) and understand the behavior and function of the networked system, even insight into a vast array of complex and previously poorly understood phenomena (Newman, 2003). Network analysis is the mapping and measuring of relationships and flows (edges) between entities (nodes), according to the mathematical, statistical and structural properties. For nodes, they are the fundamental units of a network, and for edges, they are the lines connecting the interacting nodes. According to Newman (2003) the theory of network primarily includes three parts: finding out the statistical properties to suggest appropriate ways to measure the structure properties, creating models to understand the meaning of these properties (how they came to be as they are, and how they interact with one another), as well as predicting the behavior of networked systems with the basis of measured structural properties. Network analysis facilitates to explore and identify the co-occurring patterns of large and complex data that may be more difficult to detect or analyzed using traditional normalization methods. Without pre-required mathematics normalization, it could be used to explore the potentially spatial and temporal correlations and co-occurrence of items from the raw data. Therefore, in principle, network analysis could also be used in the environmental actuality investigation to reflect the relationships between parameters/elements, especially for the metadata obtained in long-term monitoring and research programs.

Till now, network analysis has not been applied to exploring co-occurrence patterns between chemicals and parameters in complex environment, such as heavy metals or POPs. The rapid industrialization and urbanization in Pearl River Delta area of south China during the last few decades have led to widespread contamination of heavy metals in the coastal and estuarine sediments of Hong Kong (Lau and Chu, 2000), especially for the Deep Bay area, a semi-enclosed bay that close to Shenzhen Bay. Also, Victoria Harbour is a major port of Hong Kong, and lies between the most urbanized areas of the Kowloon Peninsula and the northern shore of Hong Kong Island. These two zones have been under great pressure of environmental pollution, due to the increasing discharges of industrial effluents, domestic sewage and livestock wastewater via runoffs along the bank (Hills et al., 1998; Lau and Chu, 2000; Liu et al., 2004). The Mirs Bay area, lies at the southeast corner of Hong Kong in the open sea, is relatively far away from local islands, thus less influenced by the urbanization and industrialization. Considering with their position specialty and environment quality, metals in marine sediments of these three distinct areas were selected to explore their co-occurrences and associations with the network analysis tools.

It is the first time for network analysis to be applied in the environmental pollution studies, which deciphers the structure of complex data among the various parameters of environment. With the analysis of network, it makes sense of co-occurring correlations of metals in sediments. Moreover, the co-occurrence results between metals and reference items (i.e. Fe, Al, etc.) proposed by network analysis might help to distinguish the natural or anthropogenic source of metals in these sediments. Besides the network analysis of the selected metals, the co-occurrence correlations of metals were further compared through mathematic normalization methods to verify the feasibility for network in environmental field. Therefore, this work provides a new method to analyze raw metadata across spatial and temporal gradients without initial normalization, and directly visualize the correlations among co-occurring items from complex systems.

2. Study site

Hong Kong lies in the east of Pearl River Estuary (PRE) and north of the South China Sea, 22.90–22.37°N and 113.52–114.30°E. As one of the highest populated areas in the world, Hong Kong is re-sided with about 7.13 million people in the 1170 square kilometers. The Environmental Protection Department (EPD) of Hong Kong has a comprehensive marine water/sediment quality monitoring program since 1986 (HKEPD, 2011). Water/sediment samples are collected and analyzed at 60 stations twice a year, covering the 17 typhoon shelters, marinas and dockyard, for analyses of over 80 physical, chemical and biological parameters.

According to the characteristics of geographical position and pollution type, three areas were selected for the analysis of metals in sediments, including the Deep Bay, Mirs Bay and Victoria Harbour. For Deep Bay area, a semi-enclosed bay, it is bordered on the Shenzhen Bay and received the input of inland, thus the quality of marine water and sediment are supposed to be highly affected by the anthropogenic pollutants. Victoria Harbour area is a transportation junction in Hong Kong, which receives variety of pollutants from transportation, sewage and streams. Meanwhile, it is strongly affected by tidal flush from open sea, thus facilitate the dispersion of pollutants and reduce the concentration of pollutants. On the other hand, Mirs Bay is far away from the lands and approaches South China Sea, and less influenced by the industrial and urban activities than the above two areas, and the sediments of which is supposed to be well protected and close to the natural condition.

Fourteen EPD sampling stations were selected in this study to provide good coverage of each area. Their locations were summarized in Table S1 and labeled in Fig. 1 (HKEPD, 2011), including DS1 to DS4 of Deep Bay area, MS13 to MS16 of Mirs Bay area, and VS3, VS5, VS6, VS9, ES1 and ES4 of Victoria Harbour.

3. Methodology

3.1. Data collection

Concentrations of metals and other physico-chemical parameters were downloaded from the open database of marine water and sediment at EPD website (http://www.epd.gov.hk/epd/indext.html). Eight metals, including Fe, Al, Zn, Pb, Ni, Cu, Mn, and Cr, were selected to investigate the co-occurrence correlations of these metals at the three areas. Moreover, a metalloid element B was selected since it naturally occurs in marine water and sediment. Additionally, two important parameters of sediment quality, i.e. Total Carbon (TC, %w/w) and Particle Size Fraction <63 micrometer (PSF, %w/w), were concerned as reference parameters (Dashkalakis and O’Conner, 1995; Woods et al., 2012).

3.2. Network analysis

The concentrations of the selected metals in sediments, as well as physico-chemical parameters (i.e. dissolved oxygen, total kjeldahl nitrogen, total phosphorus, dry wet ratio) were sorted following the spatial and temporal sequence. Seven of the most concerned metals were selected and filtered from the metadata to be further analyzed for their co-occurrence in sediment. This filtering step excluded most of the trace pollutants and non-relevant
parameters from EPD database, thus reduced the complexity of network analysis and facilitated the exploration of those important indexes and parameters.

The Spearman’s correlation coefficient was set >0.6 and statistically significant ($P$-value) was <0.01 to reach the robust correlation among entities in a valid co-occurrence event (Junker and Schreiber, 2008; Barberán et al., 2012). Using pre-filtered data, selected metals were represented as the nodes, and the strong and significant correlation between nodes were represented as the edges in the network results. The corresponding network topological parameters were further calculated including network diameter, cumulative degree distribution, clustering coefficient and modularity (Newman, 2003). All statistical analyses of network were performed in the R platform (Ihaka and Gentleman, 1996) using vegan, igraph and Hmisc packages (Csardi and Nepusz, 2006; Oksanen et al., 2007; Harrell, 2008). Networks were explored and visualized with the interactive platform gephi (Bastian et al., 2009).

3.3 Normalization

Mathematical normalization models and tools have been used to interpret the human-induced contribution to surface sediment. Among them, the direct linear co-variability between elements and conservative materials, and a spatial and temporal homogeneity in the controlling process of the reference parameters has been widely concerned (Woods et al., 2012). The normalization methods were introduced to serve as a comprehensive and direct comparison of approaches of network to the literatures. According to the EPD data, Fe, Al, TC and PSF were selected to serve as the conservative items (Fe, Al, TC and PSF) in the sediments. Pearson correlation coefficients ($r$) are calculated with statistical significance levels to evaluate the relationship between selected metals and the chosen reference parameters.

To cluster the co-occurrence relationship of selected metals, another normalization method was undertaken, which was the analyzed from the average distribution concentration of metals among years. The statistical analysis for annual ratio by heat map was carried out in the R platform using RColorBrewer, pheatmap and gplots packages (http://cran.rstudio.com/bin/windows/contrib/3.0). And the analysis results of heat map with cluster correlations among metals were visualized in the R platform.

4. Results and discussion

4.1 Spatial distributions of raw data

The average concentrations of selected metals and items were calculated and summarized in Table S2. The average concentrations of Cu, Pb, Cr and Zn in Deep Bay and Victoria Harbour areas were significantly higher than those in Mirs Bay. The average concentrations of metals indicated the overall level of metal contamination in sediment. Compared with the National Standards of Marine Sediment Quality of China (Table S3), the average concentrations of metals in Mirs Bay area were at Level I, which was suitable for marine fishery, bathing beach, marine sports or public entertainment that direct exposure to sediment, and water supply for food industries. For Deep Bay and Victoria Harbour, the average concentrations of Zn and Cu in sediments were at Level I, and Cr and Pb were at Level I. Moreover, the average concentrations of Zn, Pb, Cu and Cr of the Deep Bay and Victoria Harbour were significantly higher than those in Mirs Bay. It indicated that the environmental quality of sediment in Mirs Bay area was better than that of the other two areas and basically at the natural level with little contamination.
4.2. Co-occurrence of metals based on network analysis

The co-occurrence correlations of marine sediment parameters were further explored using network inference based on strong and significant correlations through using non-parametric Spearman's rank coefficient (Junker and Schreiber, 2008). Moreover, network correlation structural properties could visually summarize lots of information from metadata (Chaffron et al., 2010), which might offer the potential for quick and easy comparisons among complex items (chemicals, parameters and physicochemical properties of environment). Therefore, the correlations of network were determined for the co-occurrence analysis of selected metals in sediment, based on the filtered EPD data in the three areas.

4.2.1. Network description

With network, it is possible to work on multivariate data among complex dataset and cluster the important items without mathematical normalization. Some topological properties, including cluster coefficient and modularity index, commonly used in network analysis were calculated to describe the complex pattern of inter-relationships between items quickly and simply (Newman, 2003). According to the spatial and temporal monitoring data from 1991 to 2011 (HKEPD, 1991–2011) (Fig. 2), the co-occurrence correlations between metals were determined with the Spearman's coefficient >0.6, which indicated that the co-occurring metals have good correlations. The setup of a Spearman's coefficient cutoff could efficiently reduce the interference factors and highlight the strong correlations between items. The overall topological properties of the network in each area were summarized in Table S4. The average clustering coefficients were high (being at 0.58–0.81), which suggested the strong internal correlations among metals (Watts and Strogatz, 1998). These parameters for network initially indicated correlations of raw data, and further supported the co-occurring and clustering among metals in each distinct area respectively.

4.2.2. Deep Bay area

According to the EPD report in 2011, the Deep Bay area has the poorest water quality in the territory with high concentrations of organic and inorganic pollutants and low levels of dissolved oxygen (HKEPD, 2011). In general, lower concentration of dissolved oxygen indicated the poorer water quality and heavier water pollution. It receives discharges from Shenzhen River all year round, as well as Kam Tin River, Yuen Long Creek and Tin Shui Wai Nullah from the Hong Kong side. Moreover, the water quality in Deep Bay area has been also influenced by Pearl River during the wet season. Therefore, the quality of sediment in Deep Bay area might be greatly affected by the discharge from Shenzhen River and Pearl River, especially for some heavy metals and organic pollutants. The strong co-occurrence correlations of selected metals and parameters in Deep Bay area were shown in Fig. 2a with the corresponding Spearman's coefficients (>0.6) summarized in Table 1. It has been reported that Fe, Al, TC and PSF could be used as the reference parameters to distinguish natural background or anthropogenic pollution sources (Woods et al., 2012). But there was no apparently positive relationship between the selected metals in

---

**Fig. 2.** Correlation network of ocean pollutant indexes (selected metals and parameters) variation profiles at spatial and temporal scales in sediment of the three areas. The size of each node is proportional to the number of connections (i.e. degree), and the thickness of each connection between two nodes (i.e. edge) is proportional to the value of Spearman’s coefficient coefficient >0.60 and P-value < 0.01. (a) Deep Bay, 4 monitoring stations from 1991 to 2011, the number of spatial and temporal data (n) amounted to 162 for each item; (b) Mirs Bay, 4 monitoring stations from 1995 to 2011, n = 126; and (c) Victoria Harbour, 4 monitoring stations from 1991 to 2011, n = 242.
4.2.4. Victoria Harbour area

Since the implementation of the marine monitoring program in 1986, elevated levels of selected heavy metals, in particular Cu and Ag, could often be detected in the sediments of Victoria Harbour (HKPDP, 2011). This should be attributed to the previous industrial pollution sources in the 1960s to 1980s of last century, before pollution control legislation was introduced. The elevated levels of heavy metals were found in marine sediments at some localized “hot spots” areas, which was associated with historical pollution in sediment near the discharge points of the preliminary treatment plants in central Victoria Harbour (HKPDP, 2011).

Based on the analysis of network, it could be found the high correlations between some metals and Fe, Al and PSF (Fig. 2b). Compared with the network analysis results of Deep Bay and Mirs Bay, the correlations among metals were much more complicated (the corresponding edges being at 30, higher than that in the other areas being at 12 and 17 respectively), and the co-occurring metals with reference items (Fe, Al and PSF) were also significantly higher than the other metals (Table 3). The low correlations with reference items indicated that Pb and Cu in this area should be derived from the human-induced source, including the discharges of the electroplating, dying, printing, metal finishing and chemical industries, which was consistent with the report of Sin et al. (2001). Thus the existence of Pb in the Victoria Harbour sediments most likely originated from mixed sources, including the leaded gasoline used in the past and other anthropogenic sources (Tang et al., 2008). And the contamination of Cu might be due to the historical discharges from the nearby industrial activities and untreated sewage. Therefore, the network analysis results with Spearman’s coefficients indicated that the contamination of Cu and Pb in Victoria Harbour might be initiated from human-induced source, which was consistent with the previous reports (Choi et al., 2006; Tang et al., 2008).

According to the results of network analysis, the correlations between metals and TC showed no significance in all three areas.

### Table 1
The Spearman’s coefficients of metals in the sediment of Deep Bay.

<table>
<thead>
<tr>
<th></th>
<th>Fe</th>
<th>Al</th>
<th>Mn</th>
<th>B</th>
<th>Cu</th>
<th>Pb</th>
<th>Cr</th>
<th>Zn</th>
<th>Ni</th>
<th>Node degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

* The Spearman’s coefficient between metals is lower than 0.60. The node degree means the number of edges that connect the focal node to other nodes.

### Table 2
The Spearman’s coefficients of metals in the sediment of Mirs Bay.

<table>
<thead>
<tr>
<th></th>
<th>Fe</th>
<th>Al</th>
<th>Mn</th>
<th>B</th>
<th>Cu</th>
<th>Pb</th>
<th>Cr</th>
<th>Zn</th>
<th>Ni</th>
<th>Node degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

* The Spearman’s coefficient between metals is lower than 0.60. The node degree means the number of edges that connect the focal node to other nodes.
which indicated that TC was not a suitable reference parameter in this study. For Fe, Al and PSF, none of them could be used as universal reference item in all the three areas, but they were still significantly correlated with some metals in a certain area. Moreover, it helped to figure out the key point and multi-items correlations among items from raw metadata quickly and directly, avoiding the biased amplification of human-induced effects, especially for the complicated environmental survey researches.

4.3. Mathematic normalization

Normalization is an important method that has been used to identify the source of pollutants, natural or anthropogenic accumulation. In this study, three different normalizations were performed and discussed with the selected items. Before normalization, reference items should be chosen to interpret the human-induced contribution. As a consequence of the diversity of environmental systems in space and time among the three areas, it was normalized with a generic normalizer for each particular dataset, resulting in datasets being split and normalized to the same parameter based on an apparent linear relationship with a proposed reference parameter.

4.3.1. Normalization with annual trends of ratios and concentrations

The heat map is a visual alternative for data matrix, through looking at the raw data or hierarchically clustering samples and variables based on their similarity or differences. Heat map color and organization can be used to encode information about the metadata, if it is complex and hard to read. To further investigate the correlations among metals, heat map was used to exhibit the normalized data. The average concentrations of selected metals and items were firstly calculated within each year in the three areas respectively, based on the raw data from EPD. The annual ratio of average concentration for each item was then calculated as the average concentration in one year divided by the average concentration of the whole determination period. This annual ratio could represent the contribution of each year to the entire phase. Thus this normalization process could also demonstrate the overall changing trends of each metal with the temporal sequence in the three areas.

The heat maps were operated at the R platform, based on the annual ratio of average concentration for each item. It could be found that the metals concentrations in sediment changed much greater in Deep Bay and Victoria Harbour than that in Mirs Bay (Fig. 3). For the cluster results in Deep Bay (Fig. 3a), Zn and Cu, Ni and Cr, as well as Fe and Mn were closely related with each other respectively. According to the network analysis with raw data, Zn–Cu and Ni–Cr were also significantly correlated in this area, with Spearman’s coefficients being at 0.95 and 0.88 respectively (Table 1), while the correlations between most metals and reference items (Fe, Al, TC and PSF) were very poor. It showed the same clustering results of metals co-occurrence in Deep Bay by network with raw data and normalized data with heat map analysis. The high concentrations of Zn, Cu, Ni and Cr in Deep Bay area were supposed to be due to the inputs from Pearl River, Shenzhen River and local villages, especially for the discharge of wastewaters from the printed circuit board industries and the application herbicides in agriculture in the Pearl River Delta region, thus leading to the comparatively higher inputs of anthropogenic metals (Liu et al., 2003; Ip et al., 2007).

It is notable that the cluster results in Victoria Harbour were separated into two main groups (Fig. 3b). In the first group, Zn and Mn were clustered together with the 4 reference elements, which indicated to be originated from the natural sources; the second group including Cu, Cr, Ni and Pb, which was supposed to be from the human-induced pollutions. In the results of network with raw data, the Spearman’s coefficients of Cr and Ni as well as Fe and Mn were 0.91 and 0.85 respectively, which indicated the high correlations between metals (Table 3). The normalized data (heat map) showed similar results with raw data (network) in the analysis of co-occurrence of metals. In Mirs Bay, the correlations between metals and reference items were relatively complicated (Fig. 3c). Compared with the network results (Table 2), the selected metals were correlated with each other, and did not show significant clustering patterns. This result could reflect the co-occurrence correlations among metals with normalized data, and be applied to compare with the network results based on raw data in each area. In brief, the network analysis results with raw data were consistent with the normalization results, in terms of the co-occurring properties among metals.

To further identify the reliability of network analysis, the correlations of regional and temporal trends were determined among the three distinct sub-areas based on normalized annually ratios data (Fig. 4), and the correlation result was significant with strong connection (Spearman’s coefficients > 0.996). The resulting network consisted of 51 nodes (spatial and temporal data) and 140 edges (high correlations between nodes). The corresponding clustering coefficient was 1, which indicated that these nodes were embedded with their neighborhood tightly and tended to cluster together (Barberán et al., 2012). In network analysis, “modularity” is used to quantify the true community structure in a network corresponding to a statistically arrangement of edges. And the modularity index of 0.78 (significantly large than 0.4) strongly suggested that the network had a modular structure (Newman, 2006). The co-occurrence correlations in the three areas were significantly separated from each other, and varied with the time according to the annually average concentrations of selected items (Fig. 4). For Mirs Bay area, it was highly correlated with each other during the entire researching period of 1995–2011, indicating the quality of sediment were relatively stable without significant input and disturbance of pollutants. In Victoria Harbour area, the co-occurrence of selected metals was mostly centralized from 2000 to 2008.
which might be due to the effects of environment protection on the discharges control and water quality improvement. It indicated the quality of sediment fluctuated with time intervals, depending on the input of pollutants. However, for Deep Bay area, it has not shown remarkable trends with time in terms of the average concentration of metals, which might because the exogenous input of pollutants were not continuous or the pollution type was changed during the survey period (1991–2011). The correlations in this network co-occurrence are similar with the results of heat map based on the normalized data (Fig. 3). Therefore, it confirmed that network could be applied for the analysis of co-occurrence correlations with both raw metadata and normalized metadata.

4.3.2. Normalization to references items by R-squared coefficients

The regression of each bulk metal concentration versus the bulk concentrations of reference items and the 95% confidence intervals were calculated (van Alsenoy et al., 1993). The R-squared coefficients ($R^2$) between every two items to identify the linear relationships are summarized in Tables S5–S7. In Deep Bay, the selected metals showed no significant correlations with the reference items (Fe, Al, TC and PSF), except a relatively high correlations between Mn and Fe (Table S5). This result indicated that anthropogenic inputs were a significant control on the distribution of selected metals in Deep Bay sediments. The high correlations of Mn and Fe were probably due to the geochemical process of Mn and Fe oxides in the sediment. For Mirs Bay, the $R^2$ coefficients between metals and Fe/Al showed higher correlations than that in Deep Bay area (Table S5), which implied sediment in Mirs Bay is quite clear and close to the natural status. The results of $R^2$ coefficient in Victoria Harbour area showed that Cu and Pb were no significant correlations with Fe, Al and PSF, which meant that the human-induced pollution of Cu and Pb was the dominant source other than natural geochemical process in this area (Table S7 and Fig. S1). Moreover, the high correlations between other metals with PSF showed that PSF might be a suitable reference item for Victoria Harbour area, which was consistent with the previous study (Tang et al., 2008). However, there were no significant correlations between metals and TC in all of the three areas, which indicated that TC was not suitable to be the reference items for these survey areas and selected metals.

4.3.3. Normalization to references items by Pearson correlation coefficients

To further verify the co-occurring patterns of metals obtained by network analysis, another normalization coefficient – Pearson correlation coefficient ($r$) was calculated according to the selected elements and reference items in the three areas respectively. The coefficients were summarized in detail in Tables S8–S10. The results showed no single reference element was strongly associated to all elements in the three areas. Similarly with the results of $R^2$, Pearson correlation coefficients between metals and TC showed no significant correlations, thus it was not suitable to be considered as a reference item in these research areas. And the selected items (Fe, Al and PSF) were not general reference items for the potential correlations determination among metals in these areas. The baseline relationship between metals and reference element generally represents the expectation or prediction of naturally occurring concentrations in sediments. For Mirs Bay,
the coefficients were significantly correlated with reference items (Fe, Al and PSF), which further demonstrated natural geochemical process might be the dominant source of metals in this area (Table S9). However, in Deep Bay, most of the metals did not show significant correlations with Fe and Al, including Cu, Pb, Zn and Ni (Table S8), and the correlations of Cu and Pb with reference items (Fe, Al and PSF) showed lower correlations than the other metals in Victoria Harbour sediments (Table S10). The lower correlation coefficients of metals with reference items might be due to the human-induced pollutions in these areas, and Pb and Cu in this area might originate from historically complex human-induced inputs.

On the contrary, positive correlation between Fe and Mn should be affiliated to the Fe and Mn oxides during geochemical process as well as the exogenous input in sediment.

According to the results of Heise et al. (2010), relatively homogenous regions could be identified, even if the elements distribution and pollution types in sediments are spatially different and complex. The natural occurrence of metals complicates assessments of potentially contaminated marine sediments, because measurable quantities of metals do not automatically infer anthropogenic enrichment (Schiff and Weisberg, 1999). In this study, three sub-regions were selected to provide suitable boundaries for network and normalization analysis, to normalize natural geochemical variability and attempt to better constrain human-induced pollution regimes in marine sediments of Hong Kong. It has been found the three areas showed different distribution patterns and changing trends of heavy metals respectively. As the normalization process could emphasize the anthropogenic effects, network analysis could avoid the initial assumptions and mathematic calculations with raw data. It is applicable for network to be used in the research fields of environmental survey and spatial distribution, especially for the spatial and temporal data.

5. Conclusions

Based on the EPD data of sediments in Hong Kong, the correlations of co-occurring metals with reference items were investigated in the three distinct areas. According to the network and normalization results, the sediment quality of Mirs Bay was good, and the metals were accumulated mainly with natural geography process. Nevertheless, the Deep Bay and Victoria Harbour areas might be polluted by heavy metals from anthropogenic pollutions. For Deep Bay area, it was supposed to be affected by the Shenzhen River and local discharges, especially for heavy metals discharged from industries, such as the information technology, electronic-waste disassembling factories and electroplating industries. In Victoria Harbour, it has been polluted by Pb and Cu since 1960s, which might be due to the discharges of untreated industrial waste and sewage. These results indicated that anthropogenic inputs significantly controlled the distribution of heavy metals in Deep Bay and Victoria Harbour sediments rather than natural geochemical process. Moreover, among the reference items, TC was not significantly correlated with the other metals in all of the three areas, PSF was only applicable for that of Victoria Harbour area and Al was proved to be a good predictor of natural or artificial occurrence for Cu, Pb, Cr and Ni in the three areas. As expected, the network analysis results of raw data and normalized data were consistent with the mathematic normalization results, especially for the correlations clustering among metals and parameters. In brief, the network analysis with raw data showed potential co-occurrence correlations of elements in different areas, which could be used for the first step of screening from complicated database, thus simplified the normalization process and decrease the effects of normalization on the practical distribution patterns. Network analysis could also cluster the normalized data directly among
complex parameters and demonstrated the results visually with the topological. This approach of network tools could further enhance the spatial and temporal distribution investigations and pollutants filtration from complex metadata in different environment conditions.

Acknowledgements

The authors would like to acknowledge the kind support of Ying Yang for the R script and operation. This research work is supported by the open database of Environmental Protection Department (EPD) of Hong Kong. And it is financed by projects of the National Natural Science Foundation of China (41001316) and the Fundamental Research Funds for the Central Universities (WB1214059). This study was also financially supported by General Research Fund (GRF) of Hong Kong (HKU7122/10E).

Appendix A. Supplementary materials

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chemosphere.2014.01.068.

References