

基于自组织神经网络的污染场地多监测指标相关性分析

马春龙, 施小清, 许伟伟, 任静华, 王 佩, 吴吉春

Correlation analysis of multiple monitoring indicators of contaminated site based on self-organizing map

MA Chunlong, SHI Xiaoqing, XU Weiwei, REN Jinghua, WANG Pei, and WU Jichun

在线阅读 View online: <https://doi.org/10.16030/j.cnki.issn.1000-3665.202008001>

您可能感兴趣的其他文章

Articles you may be interested in

中试尺度下可渗透反应墙位置优化模拟——以铬污染地下水场地为例

A numerical simulation study of the position optimization of a pilot-scale permeable reactive barrier: a case study of the hexavalent chromium contaminated site

吕永高, 蔡五田, 杨骊, 边超, 李敬杰 水文地质工程地质. 2020, 47(5): 189–195

基于主成分分析法的Q2黄土湿陷特性研究

A study of the collapsibility of Q2 loess based on principal component analysis

王玉涛, 刘小平, 曹晓毅 水文地质工程地质. 2020, 47(4): 141–148

承德中部土壤-黄芩系统重金属生物有效性及迁移累积特征

Bioaccumulation and translocation characteristics of heavy metals in a soil and *Scutellaria baicalensis* system in Chengde Central Region

贾凤超, 卫晓峰, 孙厚云, 李多杰, 李健 水文地质工程地质. 2020, 47(6): 142–153

非饱和带二氧化碳入侵对表生植物及土壤的影响

Influence of carbon dioxide invasion in the unsaturated zone on vegetation and soil

潘颖, 赵晓红, 王文科, 邓红章, 韩枫, 罗平平, 杨雨萌, 张徽 水文地质工程地质. 2021, 48(4): 180–189

承德市土壤重金属空间结构与分布特征

Spatial structure and distribution characteristics of heavy metals in the soil in Chengde

安永龙, 万利勤, 李霞, 殷志强, 卫晓峰, 何泽新, 贾凤超 水文地质工程地质. 2020, 47(6): 119–131

区域地下水污染调查取样点布设量化分配方法

Quantitative sample distribution in regional groundwater contamination investigation

李颖智, 蔡五田, 耿婷婷, 张涛, 张智印 水文地质工程地质. 2019, 46(5): 24–30



关注微信公众号, 获得更多资讯信息

DOI: 10.16030/j.cnki.issn.1000-3665.202008001

基于自组织神经网络的污染场地多监测指标相关性分析

马春龙¹, 施小清¹, 许伟伟², 任静华², 王 佩³, 吴吉春¹

(1. 表生地球化学教育部重点实验室/南京大学地球科学与工程学院, 江苏 南京 210023; 2. 自然资源部国土(耕地)生态监测与修复工程技术创新中心/江苏省地质调查研究院, 江苏 南京 210018;
3. 常州市环境科学研究院, 江苏 常州 213022)

摘要: 为查明场地污染分布特征, 需对场地土壤和地下水进行钻探取样, 按规范的检测指标进行逐一测试。在初查和详查阶段将获得大量的土壤和地下水污染数据, 数据样本数量大、监测指标多, 数据结构复杂, 如何从场地大数据中提取价值信息已成为研究热点。以某有机污染场地为例, 基于自组织映射神经网络(SOM)和K均值算法开展大数据分析, 深入探讨地下水和土壤中各污染指标间的相关性。结果表明: (1) 基于自组织映射神经网络的大数据分析可快速挖掘复杂多维的污染场地监测数据, 有效完成关键信息的提取; (2) 地下水中污染检出指标存在显著的聚类特征, 同一聚类中的污染指标具备相似的空间分布特征。对场地污染物检测采取先分类后分级的优化筛选策略, 减少污染物检测指标数目, 从而有效降低场地检测费用; (3) 土壤和地下水中污染检出指标存在良好的空间相关性, 这与该污染场地地下水渗流速度缓慢有关。土壤和地下水污染检出指标空间分布的相关性, 有助于场地污染源的追溯。

关键词: 自组织映射神经网络; 污染场地; 多监测指标; 相关性分析; 土壤; 地下水

中图分类号: P641.69; X508

文献标志码: A

文章编号: 1000-3665(2021)03-0191-12

Correlation analysis of multiple monitoring indicators of contaminated site based on self-organizing map

MA Chunlong¹, SHI Xiaoqing¹, XU Weiwei², REN Jinghua², WANG Pei³, WU Jichun¹

(1. Key Laboratory of Surficial Geochemistry, Ministry of Education/School of Earth Sciences and Engineering, Nanjing University, Nanjing, Jiangsu 210023, China; 2. Engineering Technology Innovation Center for Ecological Monitoring and Restoration in Arable Land, Ministry of Natural Resources/Geological Survey of Jiangsu Province, Nanjing, Jiangsu 210018, China; 3. Changzhou Research Academy of Environment Sciences, Changzhou, Jiangsu 213022, China)

Abstract: In order to investigate the distribution characteristics of pollutants at contaminated sites, it is necessary to collect soil and groundwater samples by drilling and test them by the standard procedure. In the preliminary and detailed investigation, a large amount of data of soil and groundwater pollution will be obtained. These data are often characterized by large sample size, multiple monitoring indicators and complex data structures, and how to extract valuable information from the big data has become an important research issue. This study takes an organic contaminated site as an example, and carries out big data analytics by using self-organizing map (SOM) and k-means algorithm to explore the correlation between each organic pollution indicator of groundwater and soil. The results show that (1) the big data analytics based on self-organizing map can rapidly mine the complicated multi-

收稿日期: 2020-08-01; 修订日期: 2020-10-14

基金项目: 自然资源部国土(耕地)生态监测与修复工程技术创新中心开放课题; 国家自然科学基金项目(41672229)

第一作者: 马春龙(1996-), 男, 硕士, 主要从事地下水数值模拟研究。E-mail: machunlong@smail.nju.edu.cn

通讯作者: 施小清(1979-), 男, 教授, 博士生导师, 主要从事地下水数值模拟研究。E-mail: shixq@nju.edu.cn

dimensional monitoring data of contaminated site, and extract key information effectively. (2) The pollution indicators in groundwater are characterized by significant clustering, and the indicators in the same cluster are of similar spatial distribution characteristics. In view of this, a screening strategy may classify the indicators first and then rank them, and can be adopted at contaminated site to reduce the number of pollution indicators detected and finally save the cost of site detection. (3) The pollution indicators in soil and groundwater also have strong spatial correlation, which is mainly due to the slow seepage velocity of groundwater. According to the correlation of the spatial distribution of pollution indicators in soil and groundwater, it is helpful to trace the pollution sources at contaminated sites.

Keywords: self-organizing map; contaminated site; multiple monitoring indicators; correlation analysis; soil; groundwater

随着城市化进程加快以及产业结构调整,我国大批企业关停或搬迁,遗留下来的工业污染场地(即棕地)超过50多万块^[1]。由于棕地具有污染程度重、污染物组成复杂、土壤和地下水均受到污染等特点^[2],对居民食品安全、饮用水安全、生态环境、人居环境健康、经济社会可持续发展造成了严重威胁与挑战^[1,3-4]。因此,亟待开展污染场地的调查评估、风险管控和修复。

为进行场地污染风险管控和修复,首先需要开展场地土壤和地下水的污染状况调查。完成第一阶段场地污染状况调查后,在第二阶段调查中需采集场地样本,初步采样分析和详细采样分析中将获得大量包含土壤和地下水的污染数据。数据样本数量大,监测项目多,数据结构复杂,其中隐含着大量的特征信息、关系信息以及分类信息,例如污染物在地下水和土壤两种介质中存在何种关联性,能否通过大数据分析进行场地污染源追溯。如何采取有效方法从大数据中提取价值信息成为一个重要的研究问题^[5-6]。

受人力、物力以及财力的限制,无法对污染场地取大量样品并对所有指标逐一测试分析,如何在不影响污染场地客观评价的条件下,尽可能减少需检测的污染指标数目,降低场地检测费用,也是一个值得研究的问题。

目前,众多多变量方法已被应用于污染场地数据分析,如主成分分析(PCA)、因子分析(FA)、判别分析(DA)、层次聚类分析(HCA)等^[7-11],以此减小或消除数据中的冗余。然而由于污染数据本身的复杂性,应用传统数据分析方法存在较大局限性^[12]。传统数据分析方法,如主成分分析,无法处理复杂的大数据集^[13],在污染样本数和检测指标数较多的情况下,可能无法满足使用该方法的前提条件,即前三个主成分累积方差贡献达不到70%^[14]。另一方面,传统多变量

分析方法一般是线性方法,当数据结构不满足线性条件时,分析结果具有迷惑性,可能不准确^[15]。

随着计算能力的提高,神经网络算法开始被应用于各种数据挖掘任务,自组织映射神经网络(self-organizing map, SOM)便是其中之一。SOM是一种无监督人工神经网络,由于对大数据集有优秀的处理和解释能力^[16-17],近些年被逐渐应用于各类环境监测项目产生的多变量数据处理,如水文过程的预测、地表和地下水水质评估、地表水和地下水的时空相关关系、生态群落的研究等^[12,18-25],但在场地污染数据分析中的应用很少。SOM与传统多变量分析方法都可用于数据降维,从数据中提取主要特征。但对于缺失数据、离散或复杂的数据类型,传统多变量分析方法具有局限性,例如只能通过线性插值或须剔除存在缺失值的变量,处理文本标签数据时存在低效的问题。而SOM是一种基于无监督的机器学习算法,一方面由于具备自学习特性,可以更好地构建输入和输出之间的关联模式^[26]。SOM允许缺失值的存在,无需删除或插值,可最大程度保留原始数据特征。另一方面,由于SOM是一种非线性映射方法,可直接处理非线性数据^[13,27-28]。

本文采用自组织映射神经网络以及机器学习和模式识别领域常用的K均值聚类算法,通过对地下水和土壤中污染指标的相关性分析和聚类分析,尝试减少数据冗余,削减污染场地后续检测中的污染指标数目,以此降低场地检测费用。同时,本文还对污染物在地下水-土壤系统中的相关性进行了分析探讨。本文提出的数据挖掘框架可为污染场地数据分析和决策管理提供技术支持。

1 研究区概况

研究区污染地块位于江苏省某工业区(图1),面

积约 1 km²。20 世纪 70 年代开始有多家化工企业在此设厂, 至 2009 年全部停产搬迁。地块内企业主要产品有农药原药及制剂、农药中间体、化工中间体、甲萘胺、工业级甲萘酚、染料及染料中间体等。由于长期以来粗放的环境安全管理模式、无序的工业废水排放或泄漏导致污染物进入土壤和地下水中, 成为有机污染场地, 该场地污染类型多样, 污染状况复杂。

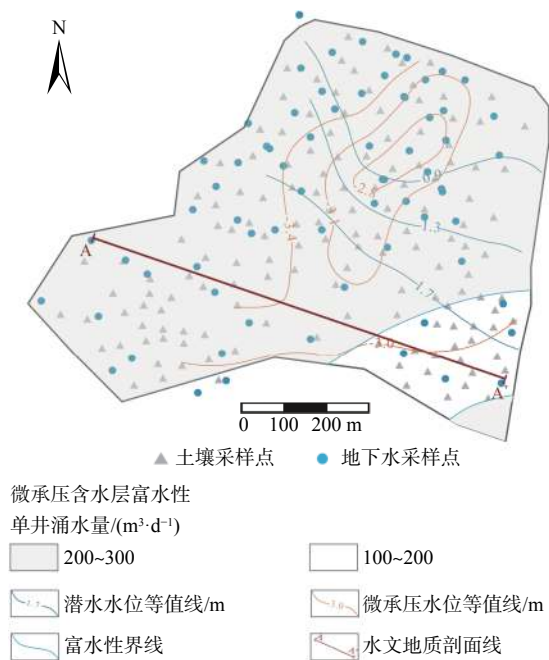


图 1 研究区及采样点分布
Fig. 1 Study area and distribution of sampling sites

场地地貌类型属长江三角洲冲积平原, 整体地势平坦, 地表高程约 5 m, 出露第四系冲积层, 自上而下主要存在四个含水层, 即孔隙潜水含水层和 I、II、III 承压含水层。潜水含水层主要由浅部填土和黏土、亚黏土组成。底板埋深 6.0~8.0 m, 潜水水位埋深 1.50 m 左右, 富水性差。I 承压含水层分为上下两段, 上段由粉砂、亚砂土组成, 顶板埋深 4.0~15.0 m, 厚度 2~20 m, 水位埋深 3~5 m, 具有微承压性质, 富水性较好; 下段由粉细砂组成, 顶板埋深 25~35 m, 厚度 2~13 m, 富水性一般。图 2 为图 1 中 A-A' 水文地质剖面, 研究区地块 30 m 内自上而下主要组成为填土、黏土、亚黏土、粉砂、亚黏土。

2 数据及方法

2.1 样本采集

污染地块内的土壤和地下水样品采集点见图 1。土壤和地下水采样按规范和技术导则^[29-32]执行。

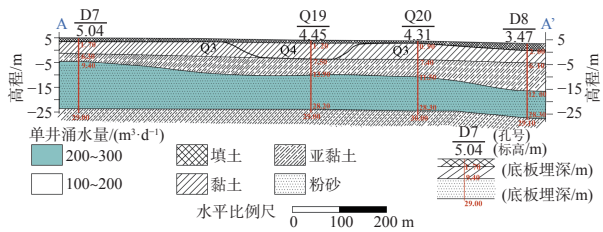


图 2 A-A'水文地质剖面
Fig. 2 Hydrogeological profile of long Line A-A'

场地的土壤样品采集, 分为表层土壤和下层土壤采集。表层土壤样品采集时, 用取样铲适当刨去裸露在空气中的表面土后, 再用取样铲取土, 装入专用密实袋。下层土壤采用钻机钻取土样, 达到规定深度后, 拔出钻杆取出土样, 采集人员戴一次性的无污染橡胶手套, 根据取样深度和个数要求取得所需深度的土样, 装入密实袋。表层土壤, 5 m 以内每隔 0.5 m 采集一个土壤样品, 5~10 m 每隔 1 m 采集一个土壤样品, 下层土壤, 10~30 m 每隔 2 m 采一个土壤样品。不同深度采集土壤样本总计 753 个。

在地下水监测井疏浚稳定 24 h 后, 进行地下水采样。水样采集后, 迅速装入带有保护剂的专用样品瓶中, 并保存在装有冰袋的冷藏箱中, 最大程度地避免样品间交叉污染。在距地表 15 m 处 (I 承压含水层上部, 微承压) 共采集地下水样本 167 个。

2.2 自组织神经网络方法

本文采用自组织映射神经网络结合 K 均值聚类算法对污染指标进行相关性分析和聚类分析。

自组织映射 (self-organizing map, SOM) 是一种竞争学习型的无监督神经网络 (图 3), 由芬兰学者 Kohonen^[23] 提出。其使用无监督训练将高维输入数据映射到低维空间, 同时保留输入数据在高维空间的拓扑结构, 即将高维空间中相似的样本点映射到二维输出层中的邻近神经元, 因此被广泛应用于数据降维^[26]。

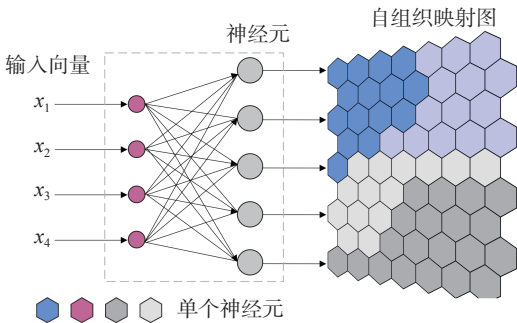


图 3 自组织神经网络结构图^[26]
Fig. 3 Structure of the self-organizing map

SOM 网络中的输出层神经元以矩阵方式排列在二维空间中, 每个神经元具有与之相连接的权向量, 在接收到一个训练样本后, 每个输出层神经元会计算该样本与自身携带的权向量间的距离, 距离最近的神元成为竞争获胜者, 称为最佳匹配单元(best matching unit, BMU)。然后, 最佳匹配单元及其邻近神经元的权向量将被重新调整, 以缩小权向量与当前样本的距离。此过程不断迭代, 直至收敛。

在获得自组织映射结果之前, 需要设置输出神经元的数量, 根据文献 [12, 21], 神经元数量最优设置为 $5\sqrt{n}$, n 为样本数量。

本研究采用赫尔辛基理工大学 Vesanto 等 [29] 开发的 SOM 工具箱在 MATLAB 中计算。

K 均值算法 (k-means clustering) 是流行于数据挖掘领域的聚类算法。其步骤是预先随机选取 K 个对象作为初始聚类中心, 随后计算每个对象与各个聚类中心的距离, 将每个对象分配到最近的聚类中心, 聚类中心及其分配的对象代表一个聚类。一旦全部对象完成分配, 每个聚类的聚类中心会根据聚类中现有的对象被重新计算。此过程不断重复直至聚类中心不再变化 [33]。K 均值聚类算法的具体实现过程详见文献 [34], 此处不再赘述。

3 结果与讨论

3.1 地下水中污染物相关性分析

表 1 为 SOM 输入数据统计特征 (已剔除浓度低于检测限的指标和无机监测指标), 根据地下水质量标准 (GB/T 14 848—2017) 中 IV 类水标准, 该场地地下水中主要有机污染物为邻二甲苯、氯苯、四氯化碳、1,2-二氯乙烷、间二甲苯和对二甲苯、三氯甲烷、乙苯以及二氯苯。

按相似度对 SOM 输出图重排序 (图 4), 每种污染物对应一种映射图, 其中的颜色梯度可用来识别污染指标间的相关性, 相同或相似颜色梯度指示正相关性, 相似程度越高, 相关性越强 [12, 21—22]。由此可知, 1,3-二氯苯、1,4-二氯苯、1,2-二氯苯、1,2,3-三氯苯、2,4-二氯酚、2,6-二氯酚, 溴苯间存在较强的相关性; 三氯甲烷 (氯仿)、锰、甲苯、乙苯、二甲苯、砷、二硫化碳、1,3,5-三甲苯、异丙基苯、苯、苯酚、4-氯甲苯、丙酮、4-甲基-2-戊酮、1-萘胺存在良好相关性; 以及四氯化碳、三氯乙烯、四氯乙烯相关性极好, 氯苯、1,2-二氯乙烷、2-氯甲苯间也存在较好相关性。

表 1 地下水中污染物数据统计特征

Table 1 Statistical characteristics of pollutant data in groundwater

污染物	均值/ ($\mu\text{g}\cdot\text{L}^{-1}$)	极大值/ ($\mu\text{g}\cdot\text{L}^{-1}$)	标准差	极大值 高于IV类水 上限倍数
邻-二甲苯	8 702.00	1 010 000.00	80 780.00	1 010.00
氯苯	4 386.00	453 000.00	36 400.00	755.00
四氯化碳	363.30	33 900.00	2 691.00	678.00
1,2-二氯乙烷	108.60	10 200.00	818.10	255.00
间&对二甲苯	763.20	109 000.00	8 463.00	109.00
三氯甲烷	411.30	28 900.00	2 614.00	96.33
锰	2 056.00	61 600.00	5 840.00	41.06
1,4-二氯苯	93.02	4 200.00	463.80	7.00
1,2-二氯苯	122.50	6 000.00	622.40	3.00
1,2,4-三氯苯	2.32	206.00	18.35	1.14
三氯乙烯	0.68	57.00	4.72	0.27
四氯乙烯	3.65	288.00	23.52	0.96
砷	9.07	268.00	26.53	5.36
乙苯	90.62	4 840.00	546.10	8.06
甲苯	144.30	5 150.00	684.60	3.67
苯	5.64	157.00	18.85	1.31
二硫化碳	1.16	102.00	8.79	1.02
2,4-二氯酚	1.25	88.30	9.33	/
2,6-二氯酚	0.17	20.10	1.59	/
1,2,3-三氯苯	0.82	59.30	5.61	0.33
1,3-二氯苯	17.90	911.00	95.78	/
溴苯	1.23	30.50	4.07	/
2-氯甲苯	7.21	239.00	29.90	/
1-萘胺	629.40	58 600.00	4 816.00	/
4-氯甲苯	68.27	6 550.00	604.80	/
异丙基苯	3.81	485.00	37.63	/
1,3,5-三甲苯	0.06	4.80	0.51	/
丙酮	8.08	1 350.00	104.50	/
4-甲基-2-戊酮	0.68	62.00	6.24	/

注: “/”表示非《地下水质量标准》(GB/T 14848—2017)要求控制指标。

在 SOM 映射结果基础上, 进一步结合 K 均值算法, 得到如图 5 所示的聚类结果。污染指标被分为 4 类, 根据图 3 污染指标映射图, 可将聚类与具体污染指标一一对应。

第一类 (Cluster-1) 包含绝大部分样本点, 相关指标包括氯苯、1,2-二氯乙烷等, 说明第一类中的污染物空间分布范围最为广泛。第二类 (Cluster-2) 包含 1,2,4-三氯苯、1,4-二氯苯、1,2-二氯苯等 8 个指标, 第三类 (Cluster-3) 包含邻二甲苯、间二甲苯和对二甲苯、甲苯、乙苯、三氯甲烷等。第四类 (Cluster-4) 为三氯乙烯、四氯乙烯和四氯化碳。

由相关性和聚类结果可知, 除个别污染物外, 该场地中绝大部分污染物均存在与之相关的污染指

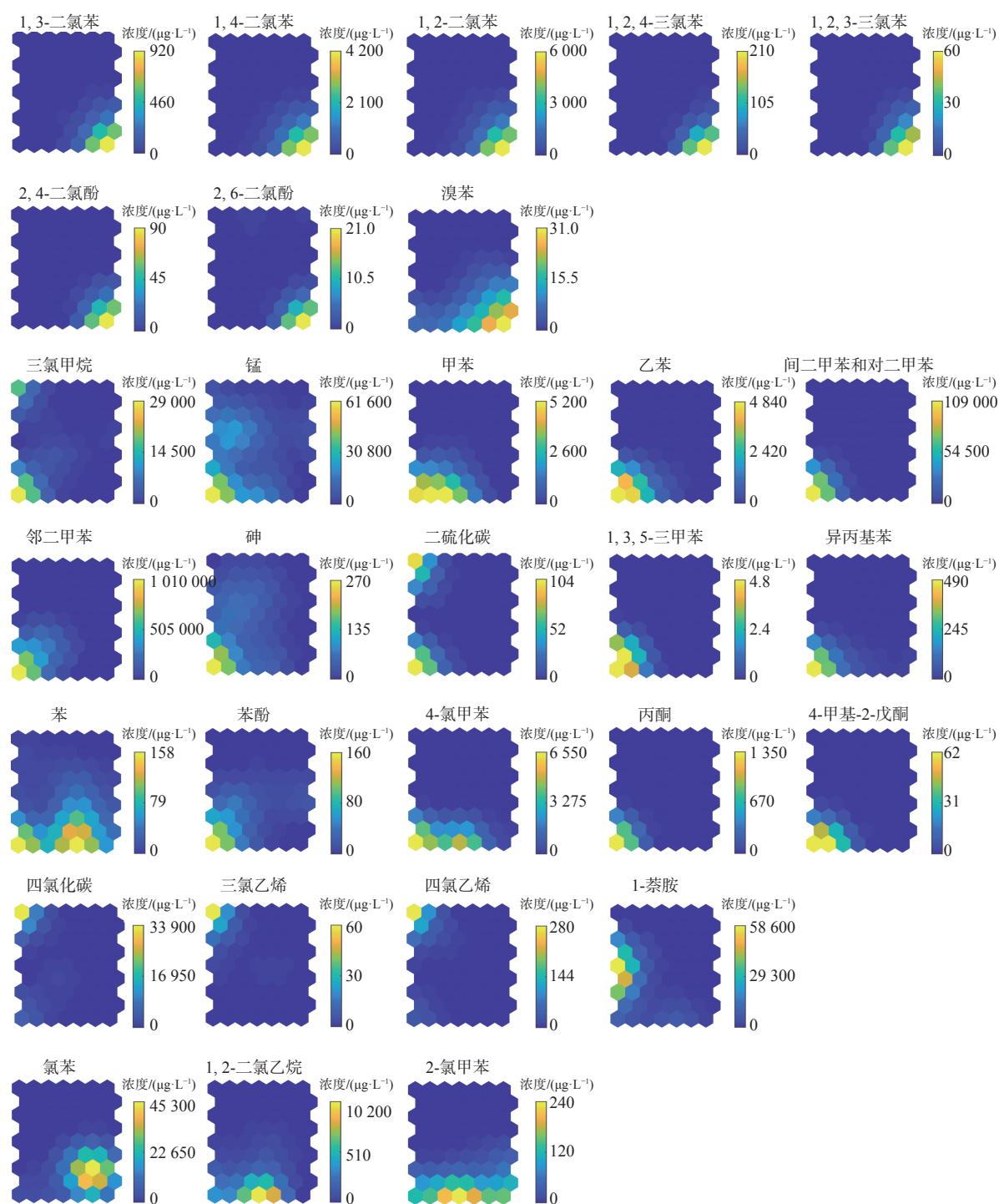


图 4 地下水中污染物指标 SOM 映射图

Fig. 4 Component planes for the pollution indicators analyzed in the SOM of groundwater

标。相当的正相关性说明场地污染物可能具有相似的来源特征,如工业中常用三氯乙烯制备四氯乙烯和四氯化碳,故三者有良好的相关性,同时其在环境介质中可能存在相似的生物化学降解以及迁移扩散途径等行为特征。

进一步分析,发现其相关性主要表现在空间分布的相似性,图 6 为不同聚类污染物空间分布特征。从中可知,从属于同一聚类的污染物,其浓度空间分布具有高度相似性,有基本一致的高值区。据此本文提出在场地污染物后续检测中可根据相关性的不同,对

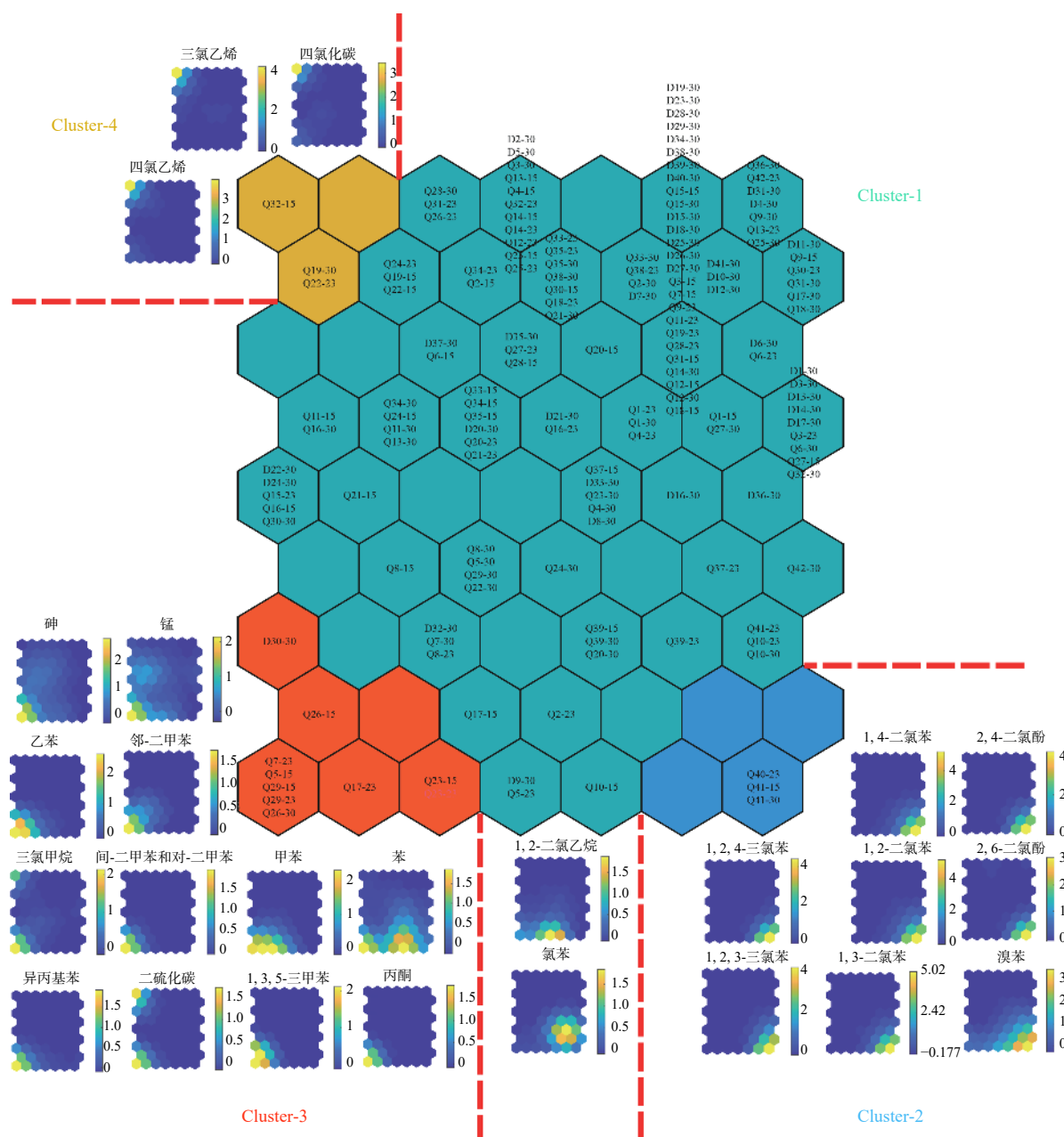


图 5 SOM-K 均值聚类结果 (根据 SOM 映射图可将污染物与聚类一一对应)

Fig. 5 Sampling sites clustering patterns in SOM-K means (according to the SOM map, the pollutants can be one-to-one corresponded to the cluster)

污染指标施行先分类后分级的筛选策略,即将同一类(相关性强)中的污染指标根据超标倍数或环境风险进行分级检测,在同一聚类中筛选少数污染指标,总体上达到削减检测费用的目标。表 2 为根据 SOM-K 均值聚类以及《地下水质量标准》(GB/T 14848—2017)进行的污染物优化筛选结果,本文根据超标倍数在 Cluster-1—Cluster-4 类中分别选择氯苯, 1,4-二氯苯, 邻二甲苯和四氯化碳。

由于污染场地通常需要采样分析大量指标,本文提出的方法可能有助于减少场地检测费用。

3.2 地下水和土壤中污染物相关性分析

为探究污染物在地下水和土壤中的相关性,本文采取相同及邻近监测点,考虑到污染物先进入土壤后进入地下水,将 15 m 处的地下水水质数据和 8~15 m 处的土壤数据(土壤采样点高于地下水)进行综合分析。表 3 为土壤和地下水中部分污染物的统计特

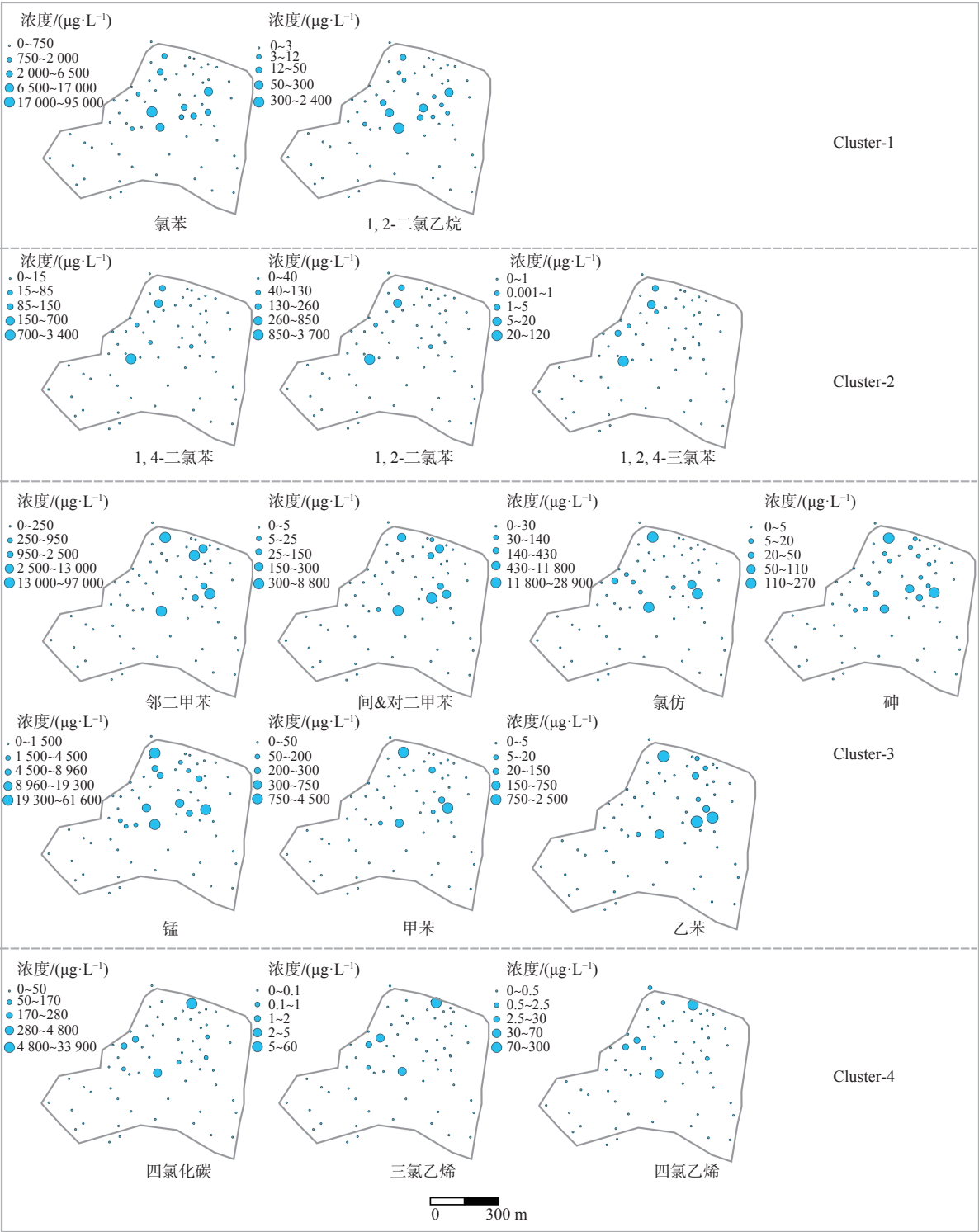


图 6 同一聚类的污染物表现出相似的空间分布特征 (取样深度 15 m)

Fig. 6 The pollutants in the same cluster showed similar spatial distribution characteristics (groundwater samples at depth 15 m)

征。图 7 为经 SOM 训练输出的自组织映射图, 根据前文, 映射图颜色梯度相似度指示污染指标相关性。由图 7 及土壤和地下水中污染物相关系数矩阵 (图 8) 可知, 除个别污染物如二甲苯外, 同种有机污染

物在土壤和地下水中具有较高相关性, 不同污染物如苯、甲苯、三氯甲烷、异丙基苯在土壤和地下水两种介质中也存在较好相关性。地下水和土壤中污染物的相关性表现在空间分

布特征的相似性,以四氯化碳、氯苯、三氯甲烷(氯仿)为例,由图 9 可知,3 种污染物在地下水和土壤中浓度高值区一致,地下水和土壤中污染物关系密切,

其原因可能是该地块低渗介质分布广泛,同时场地地下水水力坡度很小导致地下水渗流速度缓慢,污染物未能发生较大规模的侧向迁移,地下水中污染物主要来自于化工厂污废水排泄以及污染物在土壤中的持续释放。少数监测位点差别较大,三氯甲烷和氯苯在个别监测点土壤中浓度高,而地下水中浓度低(图 9),可能因为该区域黏土层较厚(图 2),大部分污染物仍滞留在土壤中。

4 结论

(1)对于有机污染场地,基于自组织映射神经网络结合聚类算法的大数据分析框架,发现地下水中除个别污染物以外,多个污染指标之间存在良好关联性,表现显著聚类特征,同一聚类中污染物浓度空间分布具有高度相似性。对于该污染场地的后续检测中可根据相关性在同一聚类中只检测少量关键指标,即氯苯,1,4-二氯苯,邻二甲苯和四氯化碳,既使场地检测有的放矢,又减少检测费用。

(2)对于该污染场地,由于地下水流速缓慢,相同有机污染物在两种不同环境介质中存在较强相关性,不同污染物如苯、甲苯、三氯甲烷、异丙基苯在土壤和地下水两种介质中存在较好相关性,表现为空间分布的一致性,该结果可能有助于场地污染源的追溯。

(3)本次研究未对深层地下水采样分析,也未在不同时间段进行采样,因此土壤和地下水检测数据的相关性分析时,未考虑在污染物浓度垂向不同深度的变异性以及随时间的变化过程,这是本研究的不足之处。另外,污染物在土壤和地下水两种环境介质中的良好相关性,后续需要更多地球化学和生物证据佐证说明污染物相似的来源特征。

表 2 地下水中污染物聚类分级优化筛选结果
Table 2 Clustering optimization results of pollutants in groundwater

聚类	污染物	极大值高于Ⅳ类水上限倍数
Cluster-1	氯苯	755.00
	1,2-二氯乙烷	255.00
	2-氯甲苯	/
	1-萘胺	/
Cluster-2	1,4-二氯苯	7.00
	1,2-二氯苯	3.00
	1,2,4-三氯苯	1.14
	2,4-二氯酚	/
	2,6-二氯酚	/
	1,2,3-三氯苯	0.33
	1,3-二氯苯	/
	溴苯	/
Cluster-3	邻-二甲苯	1 010.00
	间&对二甲苯	109.00
	三氯甲烷	96.33
	锰	41.06
	砷	5.36
	乙苯	8.06
	甲苯	3.67
	苯	1.31
	二硫化碳	1.02
	4-氯甲苯	/
	异丙基苯	/
	1,3,5-三甲苯	/
	丙酮	/
	4-甲基-2-戊酮	/
Cluster-4	四氯化碳	678.00
	三氯乙烯	0.27
	四氯乙烯	0.96

表 3 地下水和土壤数据统计特征
Table 3 Statistical characteristics of groundwater and soil data

污染物	地下水			土壤		
	均值/($\mu\text{g}\cdot\text{L}^{-1}$)	极大值/($\mu\text{g}\cdot\text{L}^{-1}$)	标准差	均值/($\text{mg}\cdot\text{kg}^{-1}$)	极大值/($\text{mg}\cdot\text{kg}^{-1}$)	标准差
二甲苯	7 790.33	995 700.00	78 790.33	6.28	236.00	38.27
氯苯	5 433.29	453 000.00	32 256.93	208.70	7 890.00	1 279.75
四氯化碳	311.25	35 450.00	2 077.31	0.79	12.50	2.65
三氯甲烷	527.11	27 680.00	1 701.52	0.50	8.15	1.49
1,2-二氯乙烷	548.30	10 200.00	902.40	0.01	0.34	0.06
1,2-二氯苯	145.09	6 000.00	599.40	25.87	615.00	110.86
甲苯	214.00	4 970.00	409.60	1.73	48.20	8.05
苯	9.08	157.00	121.20	0.08	1.21	0.24
异丙基苯	8.06	100.00	16.05	0.07	1.05	0.22

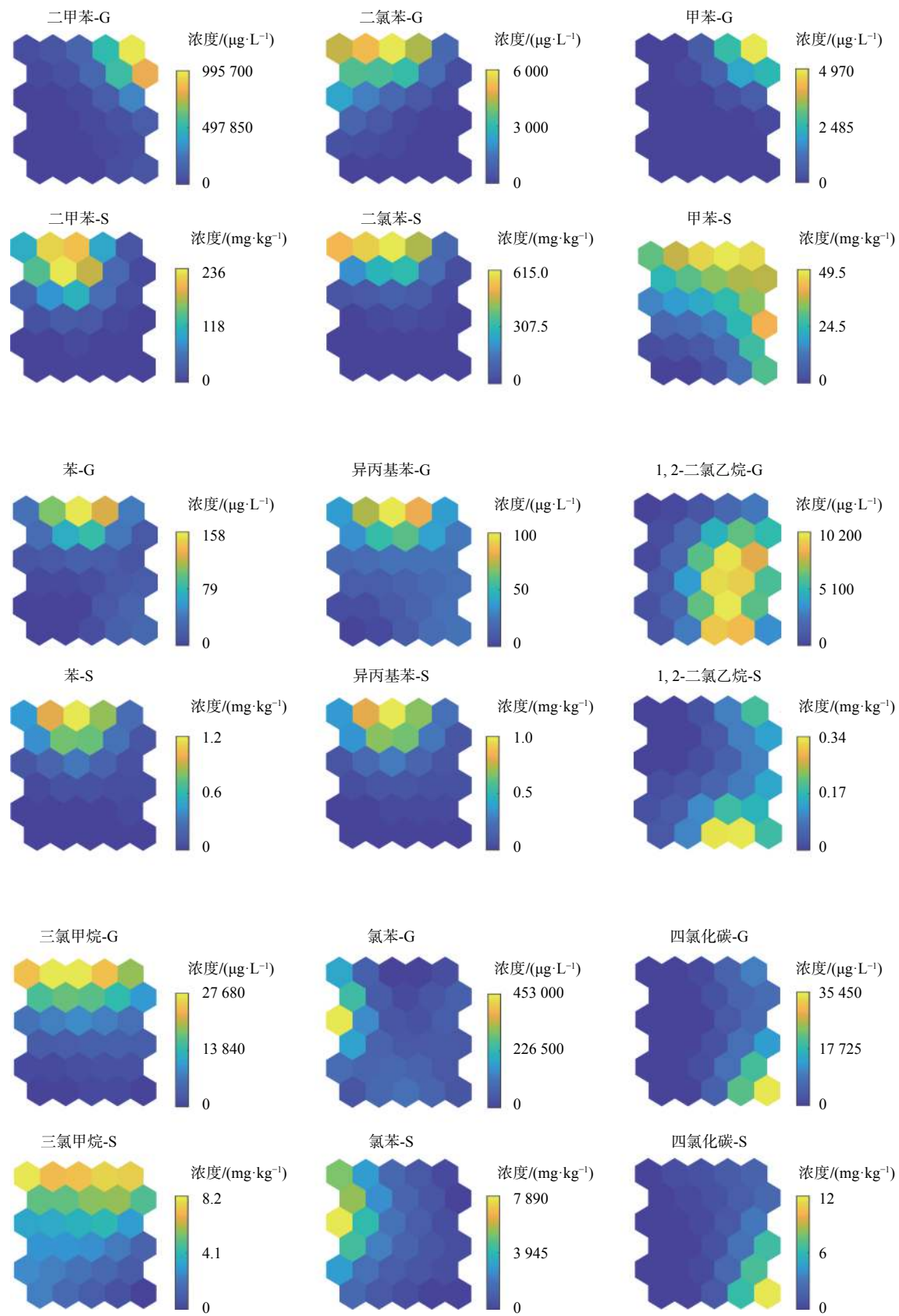


图 7 土壤和地下水中的污染物 SOM 映射结果

Fig. 7 Component planes for the pollution indicators analyzed in the SOM of groundwater and soil

注: 后缀-G 表示地下水, -S 表示土壤。

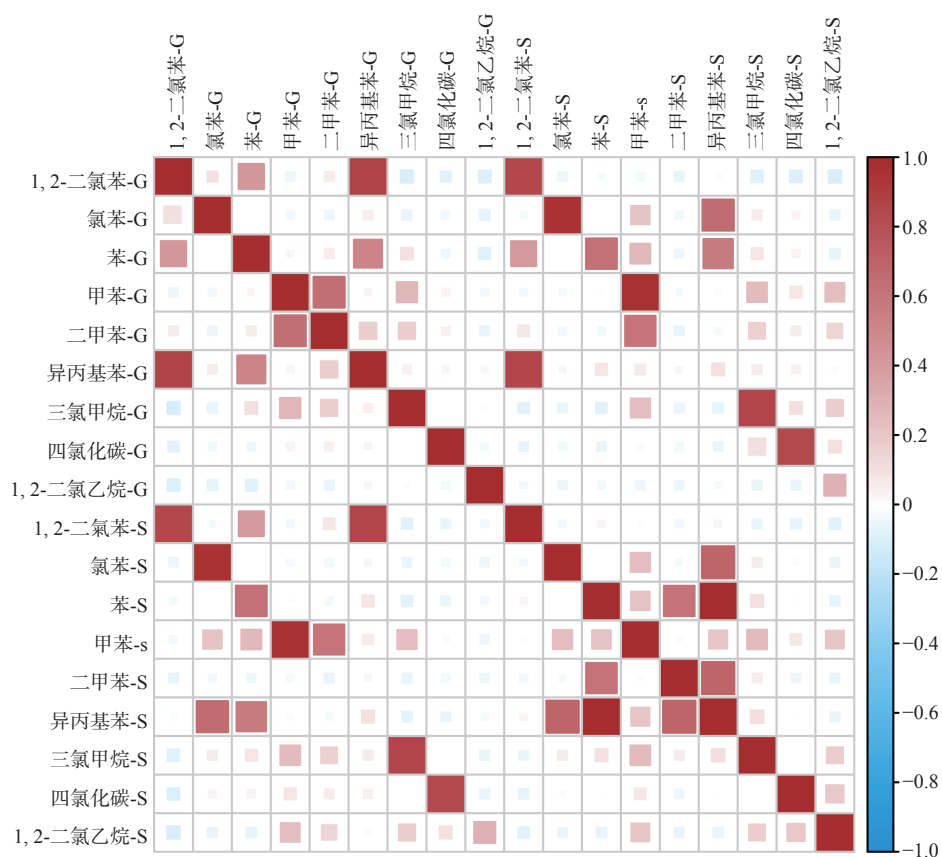


图 8 土壤和地下水中污染物相关系数矩阵

Fig. 8 Correlation coefficient matrix of pollutants in soil and groundwater

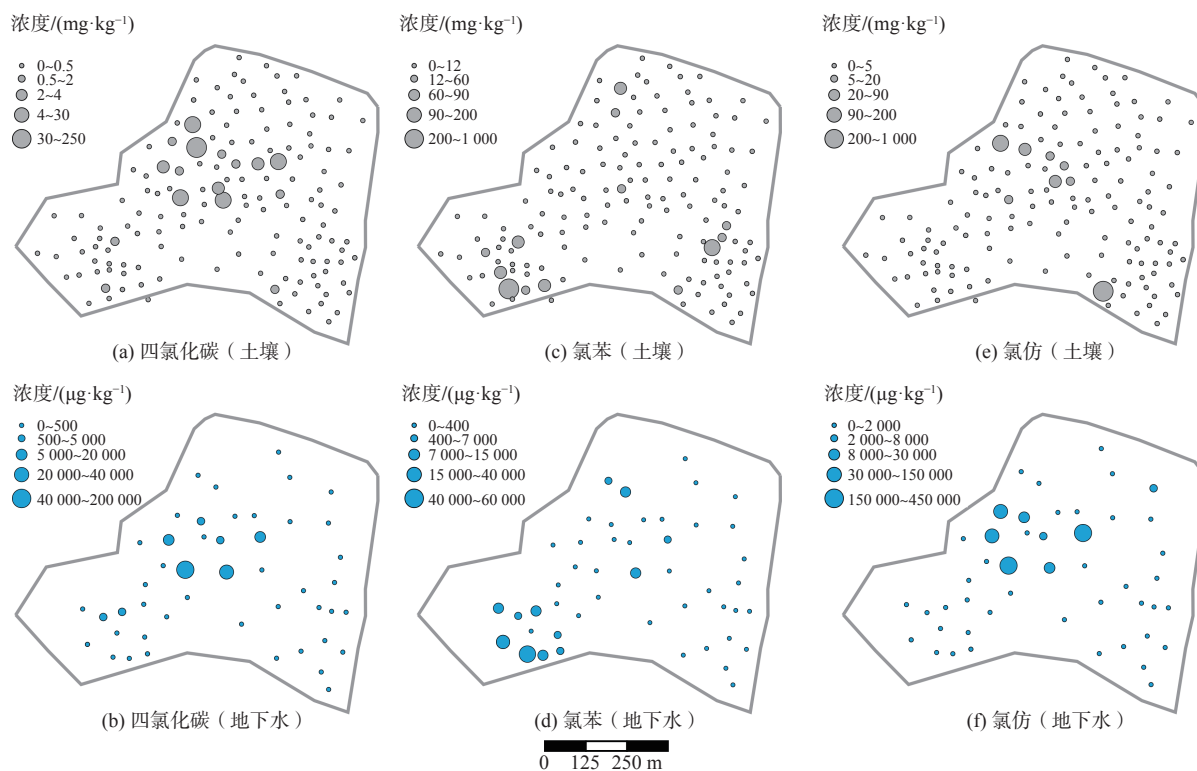


图 9 四氯化碳、氯苯、三氯甲烷在地下水和土壤中的空间分布情况对比

Fig. 9 Comparison of the spatial distribution of carbon tetrachloride, chlorobenzene and chloroform in groundwater and soil

参考文献 (References):

- [1] 陈梦舫. 我国工业污染场地土壤与地下水重金属修复技术综述[J]. 中国科学院院刊, 2014, 29(3): 327 - 335. [CHEN Mengfang. Review on heavy metal remediation technology of soil and groundwater at industrially contaminated site in China[J]. Bulletin of Chinese Academy of Sciences, 2014, 29(3): 327 - 335. (in Chinese with English abstract)]
- [2] 吕永高, 蔡五田, 杨骊, 等. 中试尺度下可渗透反应墙位置优化模拟——以铬污染地下水场地为例[J]. 水文地质工程地质, 2020, 47(5): 189 - 195. [LYU Yonggao, CAI Wutian, YANG Li, et al. A numerical simulation study of the position optimization of a pilot-scale permeable reactive barrier: a case study of the hexavalent chromium contaminated site[J]. Hydrogeology & Engineering Geology, 2020, 47(5): 189 - 195. (in Chinese with English abstract)]
- [3] 郭琼泽, 施小清, 王慧婷, 等. 井间分溶示踪估计重非水相污染物残留量的影响因素数值分析[J]. 水文地质工程地质, 2019, 46(6): 165 - 172. [GUO Qiongzhe, SHI Xiaoqing, WANG Huiting, et al. Numerical analysis of the influencing factors for estimating DNAPL residual by the partitioning interwell tracer tests[J]. Hydrogeology & Engineering Geology, 2019, 46(6): 165 - 172. (in Chinese with English abstract)]
- [4] 骆永明. 中国污染场地修复的研究进展、问题与展望[J]. 环境监测管理与技术, 2011, 23(3): 1 - 6. [LUO Yongming. Contaminated site remediation in China: progresses, problems and prospects[J]. The Administration and Technique of Environmental Monitoring, 2011, 23(3): 1 - 6. (in Chinese with English abstract)]
- [5] 史健婷. 多维环境与地球化学数据分析与可视化研究[D]. 广州: 中国科学院广州地球化学研究所, 2017. [SHI Jianting. Statistical analysis and visualization for multidimensional environmental and geochemical data[D]. Guangzhou: Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, 2017. (in Chinese with English abstract)]
- [6] 姜光辉, 李红春, 郭芳. 地下水污染场地水质空间相关性分析[J]. 水文地质工程地质, 2017, 44(2): 137 - 143. [JIANG Guanghui, LI Hongchun, GUO Fang. Spatial variability of multi-tracers in groundwater contamination sites[J]. Hydrogeology & Engineering Geology, 2017, 44(2): 137 - 143. (in Chinese with English abstract)]
- [7] 王玉玲, 王蒙, 闫岩, 等. 基于聚类算法的ERT污染区域识别方法[J]. 中国环境科学, 2019, 39(3): 1315 - 1322. [WANG Yuling, WANG Meng, YAN Yan, et al. An ERT pollution area identification method based on clustering algorithm[J]. China Environmental Science, 2019, 39(3): 1315 - 1322. (in Chinese with English abstract)]
- [8] 安永龙, 黄勇, 孙朝, 等. 北京通州某改造区土壤中PAHs的来源分析及风险评价[J]. 水文地质工程地质, 2017, 44(5): 112 - 120. [AN Yonglong, HUANG Yong, SUN Zhao, et al. Source apportionment and risk assessment of PAHs in soil from a renewal area in the Tongzhou District of Beijing[J]. Hydrogeology & Engineering Geology, 2017, 44(5): 112 - 120. (in Chinese with English abstract)]
- [9] 陈洁, 施维林, 张一梅, 等. 电镀厂遗留场地污染分析及健康风险空间分布评价[J]. 环境工程, 2018, 36(4): 153 - 159. [CHEN Jie, SHI Weilin, ZHANG Yimei, et al. Pollution analysis and spatial distribution of health risk in electroplating abandoned site[J]. Environmental Engineering, 2018, 36(4): 153 - 159. (in Chinese with English abstract)]
- [10] TAO H, LIAO X Y, ZHAO D, et al. Delineation of soil contaminant plumes at a co-contaminated site using BP neural networks and geostatistics[J]. Geoderma, 2019, 354: 113878.
- [11] OLAWOYIN R, NIETO A, GRAYSON R L, et al. Application of artificial neural network (ANN)-self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions[J]. Expert Systems with Applications, 2013, 40(9): 3634 - 3648.
- [12] LEE K J, YUN S T, YU S, et al. The combined use of self-organizing map technique and fuzzy c-means clustering to evaluate urban groundwater quality in Seoul metropolitan City, South Korea[J]. Journal of Hydrology, 2019, 569: 685 - 697.
- [13] ASTEL A, TSAKOVSKI S, BARBIERI P, et al. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets[J]. Water Research, 2007, 41(19): 4566 - 4578.
- [14] 林海明, 杜子芳. 主成分分析综合评价应该注意的问题[J]. 统计研究, 2013, 30(8): 25 - 31. [LIN Haiming, DU Zifang. Some problems in comprehensive evaluation in the principal component analysis[J]. Statistical Research, 2013, 30(8): 25 - 31. (in Chinese with English abstract)]
- [15] LISCHIED G. Non-linear visualization and analysis of large water quality data sets: a model-free basis for efficient monitoring and risk assessment[J]. Stochastic Environmental Research and Risk Assessment, 2009, 23(7): 977 - 990.
- [16] KOHONEN T. Essentials of the self-organizing map[J]. Neural Networks, 2013, 37: 52 - 65.
- [17] KOHONEN T. Exploration of very large databases by self-

- organizing maps[C]//Proceedings of International Conference on Neural Networks (ICNN'97). June 12-12, 1997, Houston, TX, USA. IEEE, 1997: PL1-PL6.
- [18] KALTEH A M, HJORTH P, BERNDTSSON R. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application[J]. *Environmental Modelling & Software*, 2008, 23(7): 835 – 845.
- [19] CHEN I T, CHANG L C, CHANG F J. Exploring the spatio-temporal interrelation between groundwater and surface water by using the self-organizing maps[J]. *Journal of Hydrology*, 2018, 556: 131 – 142.
- [20] UNDERWOOD K L, RIZZO D M, SCHROTH A W, et al. Evaluating spatial variability in sediment and phosphorus concentration-discharge relationships using Bayesian inference and self-organizing maps[J]. *Water Resources Research*, 2017, 53(12): 10293 – 10316.
- [21] JAMPANI M, HUELSMANN S, LIEDL R, et al. Spatio-temporal distribution and chemical characterization of groundwater quality of a wastewater irrigated system: a case study[J]. *Science of the Total Environment*, 2018, 636: 1089 – 1098.
- [22] DAI L J, WANG L Q, LI L F, et al. Multivariate geostatistical analysis and source identification of heavy metals in the sediment of Poyang Lake in China[J]. *Science of the Total Environment*, 2018, 621: 1433 – 1444.
- [23] LIAO X Y, TAO H, GONG X G, et al. Exploring the database of a soil environmental survey using a geo-self-organizing map: a pilot study[J]. *Journal of Geographical Sciences*, 2019, 29(10): 1610 – 1624.
- [24] MELO D S, GONTIJO E S J, FRASCARELI D, et al. Self-organizing maps for evaluation of biogeochemical processes and temporal variations in water quality of subtropical reservoirs[J]. *Water Resources Research*, 2019, 55(12): 10268 – 10281.
- [25] KIM K H, YUN S T, YU S, et al. Geochemical pattern recognitions of deep thermal groundwater in South Korea using self-organizing map: Identified pathways of geochemical reaction and mixing[J]. *Journal of Hydrology*, 2020, 589: 125202.
- [26] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016. [ZHOU Zhihua. Machine learning[M]. Beijing: Tsinghua University Press, 2016.(in Chinese)]
- [27] LI T, SUN G H, YANG C P, et al. Using self-organizing map for coastal water quality classification: Towards a better understanding of patterns and processes[J]. *Science of the Total Environment*, 2018, 628/629: 1446 – 1459.
- [28] CHOI B Y, YUN S T, KIM K H, et al. Hydrogeochemical interpretation of South Korean groundwater monitoring data using Self-Organizing Maps[J]. *Journal of Geochemical Exploration*, 2014, 137: 73 – 84.
- [29] 中华人民共和国环境保护部. 场地环境调查技术导则: HJ 25.1—2014[S]. 北京: 中国环境科学出版社, 2014. [Ministry of Environmental Protection of the People's Republic of China. Technical guidelines for environmental site investigation: HJ 25.1—2014[S]. Beijing: China Environment Science Press, 2014. (in Chinese)]
- [30] 中华人民共和国环境保护部. 场地环境监测技术导则: HJ 25.2—2014[S]. 北京: 中国环境科学出版社, 2014. [Ministry of Environmental Protection of the People's Republic of China. Technical guidelines for environmental site monitoring: HJ 25.2—2014[S]. Beijing: China Environment Science Press, 2014. (in Chinese)]
- [31] 中华人民共和国环境保护部. 土壤环境监测技术规范: HJ/T 166—2004 [S]. 北京: 中国环境科学出版社, 2004. [Ministry of Environmental Protection the People's Republic of China. Technical specification for soil environmental monitoring: HJ/T 166-2004[S]. Beijing: China Environment Science Press, 2014.(in Chinese)]
- [32] 中华人民共和国环境保护部. 地下水环境监测技术规范: HJ/T 164—2004 [S]. 北京: 中国环境科学出版社, 2004. [Ministry of Environmental Protection the People's republic of China. Technical specifications for environmental monitoring of groundwater: HJ/T 166—2004 [S]. Beijing: China Environment Science Press, 2004.(in Chinese)]
- [33] VESANTO J, ALHONIEMI E. Clustering of the self-organizing map[J]. *IEEE Transactions on Neural Networks*, 2000, 11(3): 586 – 600.
- [34] LIKAS A, VLASSIS N, VERBEEK J J. The global k-means clustering algorithm[J]. *Pattern recognition*, 2003, 36(2): 451 – 461.

编辑: 张若琳