SCIENTIFIC REP**O**RTS

OPEN

Effects of land use/cover on surface water pollution based on remote sensing and 3D-EEM fluorescence data in the Jinghe Oasis

Xiaoping Wang[1,2] & Fei Zhang[1,2,3]

The key problem in the reasonable management of water is identifying the effective radius of surface water pollution. Remote sensing and three-dimensional fluorescence technologies were used to evaluate the effects of land use/cover on surface water pollution. The PARAFAC model and self-organizing map (SOM) neural network model were selected for this study. The results showed that four fluorescence components, microbial humic-like (C1), terrestrial humic-like organic (C2, C4), and protein-like organic (C3) substances, were successfully extracted by the PARAFAC factor analysis. Thirty water sampling points were selected to build 5 buffer zones. We found that the most significant relationships between land use and fluorescence components were within a 200 m buffer, and the maximum contributions to pollution were mainly from urban and salinized land sources. The clustering of land-use types and three-dimensional fluorescence peaks by the SOM neural network method demonstrated that the three-dimensional fluorescence peaks and land-use types could be grouped into 4 clusters. Principal factor analysis was selected to extract the two main fluorescence peaks from the four clustered fluorescence peaks; this study found that the relationships between salinized land, cropland and the fluorescence peaks of C1, W2, and W7 were significant by the stepwise multiple regression method.

Water quality plays pivotal roles in habitat protection, agriculture, industry, and public health[1]. The sources of water quality pollution not only come from rivers and lakes of but also from the land use/cover and the production activities of humans[2,3]. Land use/cover is the carrier of human activity in watersheds. Landscape patterns control various biogeochemical and physical processes of watersheds. Therefore, one of the most significant consequences of landscape pattern change is the deterioration of river water quality[1]. Water pollution not only leads to the imbalance of river ecosystems but also threatens public health security and socio-economic sustainability[4]. To take effective measures for the protection of river water safety, a scientific interpretation of the relationship between land use/cover and river water quality is greatly needed.

Many previous studies have shown that land use/cover is a key factor affecting water quality pollution[5,6]. Some researchers reported relationships at catchment or watershed scales, while other researchers conducted studies at the riparian buffer scale[7,8]. Previous results have shown that land-use types are closely related to human activities, and there is a positive relationship between cropland and urban areas and water quality pollution indicators (e.g., nitrogen, phosphorus, ammonia) and a negative relationship between forests, sandy areas, and grasslands and water quality pollution indicators (e.g., nitrogen, phosphorus, ammonia)[9]. These previous analyses and approaches were all based on the assumption that relationships between water quality indicators and land-use patterns were constant over the entire study area. In recent years, optical tools such as fluorescence measurements have been developed as quick and general monitoring technologies for water quality[10–12]. In particular, excitation-emission matrix (EEM) spectroscopy can be used to extract the excitation and emission wavelengths

[1]Key Laboratory of Smart City and Environmental Modeling of Higher Education Institute, College of Resources and Environment Sciences, Xinjiang University, Urumqi, 830046, People's Republic of China. [2]Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi, 830046, Xinjiang, People's Republic of China. [3]Engineering Research Center of Central Asia Geoinformation Development and Utilization, National Administration of Surveying, Mapping and Geoinformation, Urumqi, 830002, People's Republic of China. Correspondence and requests for materials should be addressed to F.Z. (email: zhangfei3s@163.com)

1

of a wide range of water samples containing a variety of fluorescence peaks[13]. In fluorescence excitation-emission, there is a certain relationship between protein-like fluorescence peaks, including tryptophan and tyrosine, and the concentration of microbes, while humic-like fluorescence peaks can identify condensed humified organic substances[14,15]. The EEM technique with parallel factor analysis (PARAFAC), a three-way decomposition method, has been found to produce a useful diagnosis of fluorescence-independent spectral overlap components[16–18]. PARAFAC can identify different types of independent fluorescence spectra, and these can be separated by the high-resolution fluorescence imaging of the overlapping parts of several independent EEM components; thus, PARAFAC is capable of identifying even small changes.

The use of remote sensing images coupled with spatial analysis techniques to investigate the influence of land use/cover on surface water pollution has become a critical research topic. However, previous studies have focused on the correlation between water quality and land use/cover under fuzzy boundaries at catchment or watershed scales[19]. It is difficult to subdivide the watershed unit. Therefore, the studies on the correlation between water quality and land use/cover with quantitative boundaries are scare. In addition, studies on the response relationship between water quality and land use/cover at a small scale by radius methods are scare.

In this paper, we studied the arid area of the Jinghe Oasis, Central Asia, using three-dimensional fluorescence spectral and Gao Fen-1 (GF-1) satellite images, and then we combined the data with EEM-PARAFAC methods and self-organizing feature map neural network (SOM) methods to determine the connection between land use/cover and the fluorescence peaks. The objectives were: (1) identify the fluorescence peak of water bodies in the arid area; (2) analyze the relationship between land use/cover and the fluorescence peaks at multiple spatial scales; (3) examine the radius of action on the effect of land use/cover on surface water pollution; and (4) provide more information and references for the management and control of water pollution.

Result and Analysis

Land use/cover characteristics at various spatial scales.
To understand the differences in the land use/cover types between the river scale and the buffer zones, this study used the river scale and 100 m, 200 m, 300 m, 400 m and 500 m buffers (Figure 1) in the Jinghe Oasis, which includes the Bortala River (B-R), Jinghe River (J-R), Aqikesu River (AQKS-R) and Kuitun River (K–R). The percentage of urban land within the buffer zones declined slightly compared to the value at the river scale, with high values for sample 28 and sample 29, which were 56.2% and 48.2% under the 100 m and 200 m scale, respectively. The lowest values were for sample 25 and sample 34, which were 0% at the 100 m, 200 m, 300 m, 400 m and 500 m buffers. The proportion of other land in sample 16 and sample 17, however, reached up to 74.93% in the 100 m buffer scale, reflecting the existence of a large amount of undeveloped land within this area. This is the main feature of land use in arid areas. The proportion of salinized land in sample 3, sample 4 and sample 10 was as high as 60% at the 100 m buffer scale, and with an increase in the scale of the buffer, the value declined, reflecting the existence of a large amount of undeveloped land within this area.

PARAFAC model components from EEM.
Overall, all fluorescent EEM data were resolved into a successful PARAFAC model analysis. Figure 2 reveals each contour profile of the three PARAFAC components. Seven peaks were extracted, decomposed from water samples of the Jinghe Oasis, as shown in Table 1. However, W1 and W2 represent the peak of Raman scattering; therefore, they are not considered in further calculations.

Figure 3 shows the differences in C1, T2, T3, T5 and T7 at the Jinghe Oasis. C1, T2, T3, T5 and T7 represent photodegradation products, humic-like substances, photodegradation products, humic substances + recent materials and tyrosine-like substances, respectively. The value of C1 ranges between 179.2 and 2884, and the value of W1 ranges between 1013 and 9998. The high values of C1 and W1 suggest the photodegradation product content is different in the watershed. The values of W2 were the highest in sample 17, ranging from 791 to 9991. The values of W5 were the highest in sample 34, sample 35 and sample 36, ranging from 291 to 8991.

Relationships between the fluorescence peak values and the land use and cover types.
At small scales, a positive correlation indicates the influence of built-up land on the peak value of fluorescence, showing that, with the increase in the composition of urban land, the concentrations of these water quality parameters increase and make the contamination more serious (Table 2). It should be especially noted that C1 and W7 are influenced differently by urban land at 200 m scales, with R^2 of 0.95. Petroleum is not affected by the composition of each land use type. C1 and W7 are influenced by salinized land at all scales, and the R^2 increases as the scale increases. The R^2 is 0.58 for the 200 m radius of operation. Therefore, relationships between the peak value of fluorescence and the land use/cover were explored within a 200 m radius of operation.

The contribution of land use/cover to water quality pollution in a 200 m radius.
Spatial framework of land use/cover in a 200 m radius. With regard to network structure selection, the neural network with a more complicated structure will generally have a better capability to address complicated non-linear problems. However, more complicated neural networks require a longer training time. Using a greater land use and cover type area can provide more abundant information; however, the correlation among indices will increase. The topological values were selected to determine grid size in this study, and the *k*-means clustering method was adopted to obtain the results. Overall, after the standard processing of the water quality data, the best network training effect was obtained from 36 (6 × 6) nerve cells (Figure 4), with QE and TE values of 1.033 and 0.001, respectively.

Figures 4–5 show that the distribution of the land use/cover type area varies in different clustering layers. Among the six clusters, the land use/cover type area is generally relatively better in Cluster 4. However, Cluster 2 and has 5 sampling points, 11, 12, 24, 25 and 28, which contain a water body. In addition, the water body, forest-grass land, cropland, salinized land, desert and other areas are low in Cluster 1, which indicates relatively
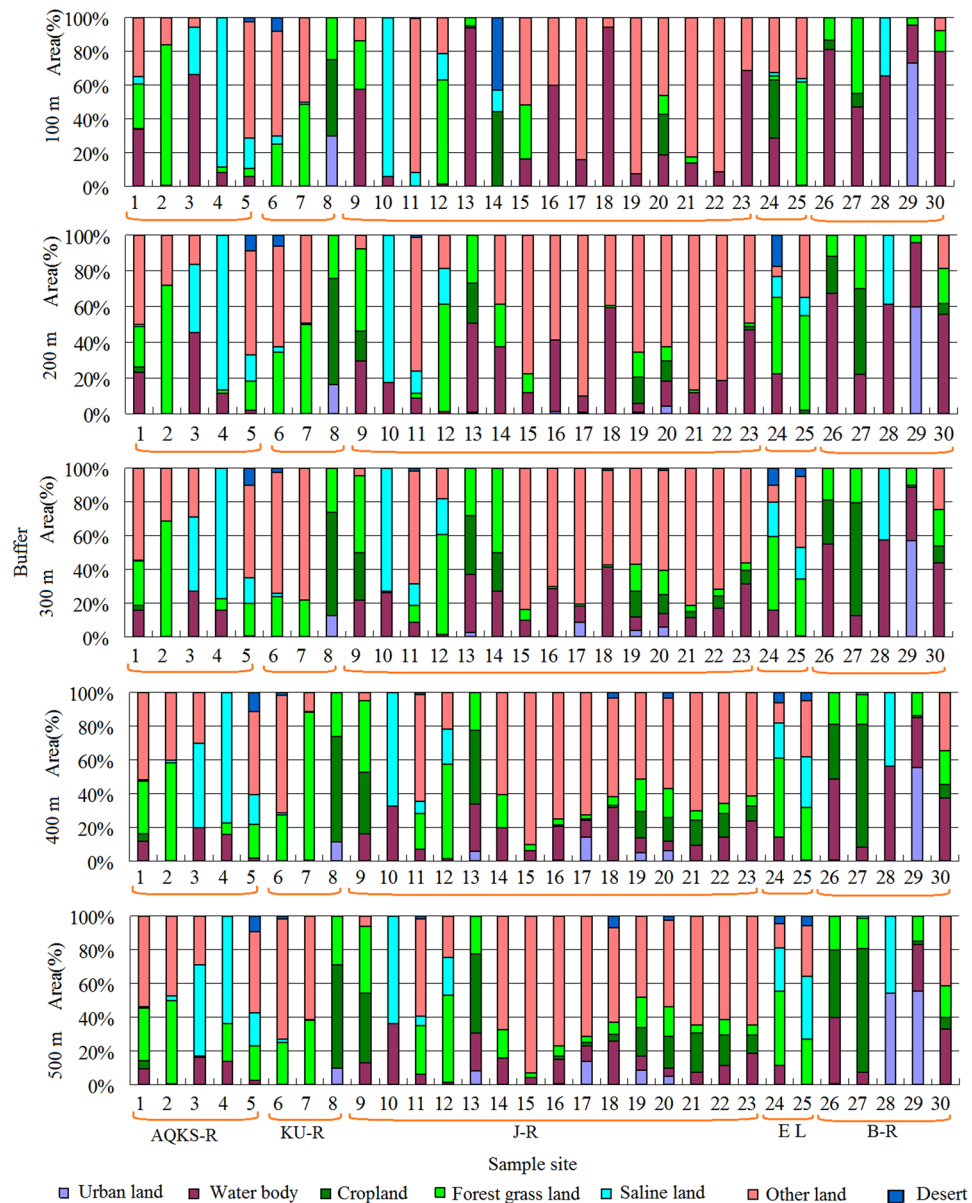
Figure 1. Statistical analysis of land use/covers area under different Radius (AQKS-R, represents Akeqisu River; KU-R represents Kuitun River; EL represents Ebinur Lake; B-R represents Bortala River) (Map by EXCEL (https://www.microsoft.com/software)).

concentrated human activity. Cluster 1 has 5 sampling points, 1, 17, 8, 19, 29, and 30. Cluster 3 has 6 sampling points, 2, 3, 4, 5, 6 and 7. Cluster 4 has 10 sampling points, 1, 17, 15, 16, 9, 10, 14, 18, 26 and 29.

Figure 5 shows the relations among the distribution of different variable classifications and water quality parameters. The land use and cover pattern and distribution of sampling points in Cluster1 are as follows: the town of TuoTuo is west of the sampling points, point 8 is in the middle of Jing River, points 19, 21, 22 and 23 are in the middle of Bortala River, and points 29 and 30 are in the eastern part of Bortala River. There is a large urban area in this region, so the cluster is therefore defined as urban land-oriented. The main samples of Cluster 2 were from the Bortala River and Jing River estuary, where forest-grassland and salinized land were the main land use types; thus, this cluster is defined as having grassland- and salinized land-oriented land use patterns. The rest of the samples were located near the farmland areas and were defined by a cropland-oriented utilization pattern. To further visualize the four components, the results are shown in Figure 6.

Spatial framework of the peak fluorescence value in the 200 m radius.   Regarding network structure selection, the neural network with a more complicated structure will generally have better capability to address complicated non-linear problems. However, more complex neural networks require a longer training time. Using a greater peak value of fluorescence can provide more abundant information; however, the correlation among indices will increase. The topological values were selected to determine the grid size in this study, and the *k*-means clustering
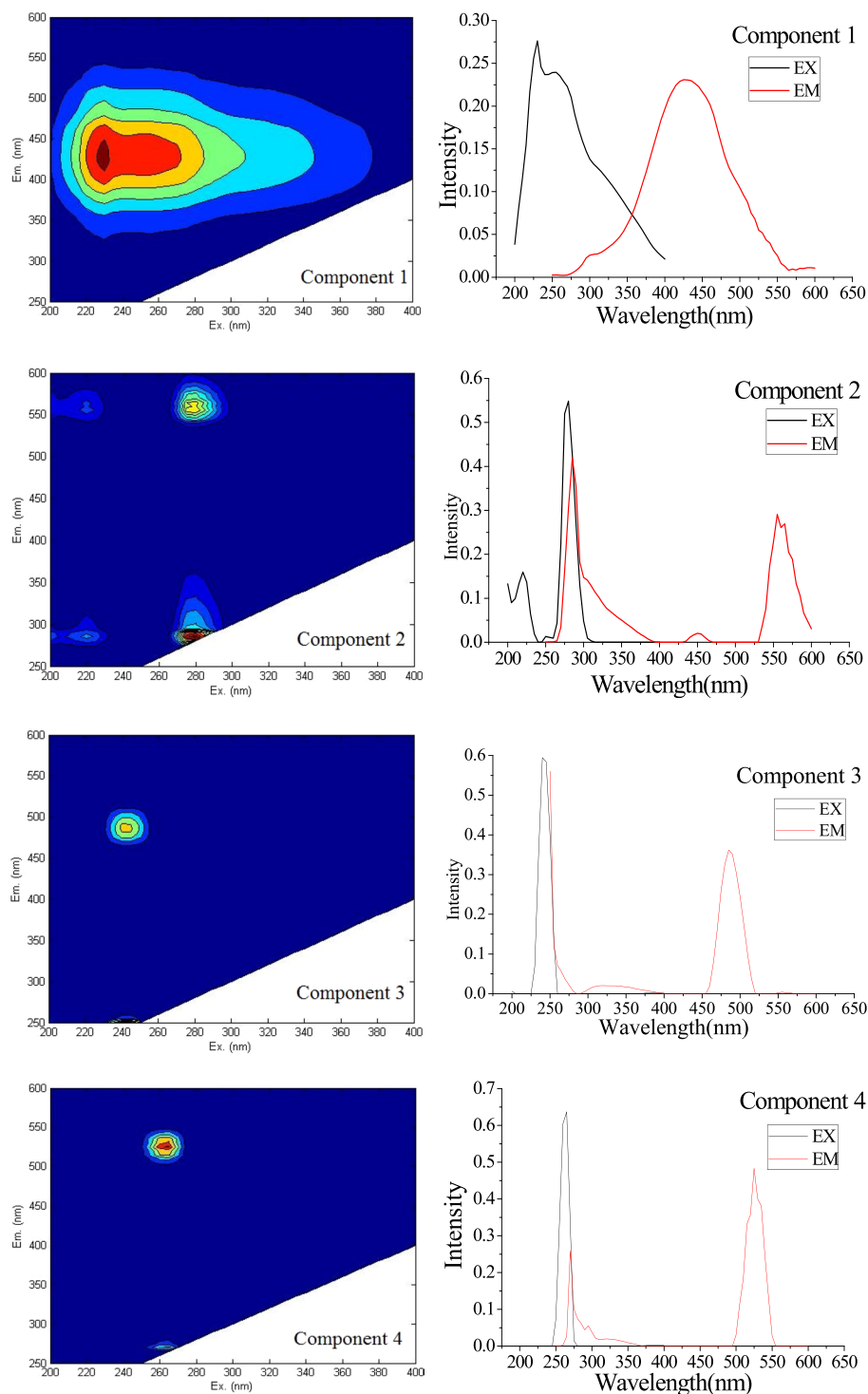
Figure 2. Contour plots of the four PARAFAC components decomposed from water samples of the Jinghe Oasis. (Map by Matlab (https://www.mathworks.com/software)).

method was adopted to obtain results. Overall, after the standard processing of the water quality data, the best network training effect was obtained from 36 (6 × 6) nerve cells (Figure 7), and the QE and TE values were 1.043 and 0.001, respectively.

Figures 7–8 show that the distribution of the peak fluorescence value varies in different clustering layers. Among the four clusters, the peak fluorescence value distribution is generally relatively better in cluster 4. However, Cluster 2 only has 4 sampling points, 11, 12, 24, 25 and 28, which contain a water body. Cluster 1 has 5 sampling points, 1, 17, 8, 19, 29, and 30. Cluster 3 has 6 sampling points, 2, 3, 4, 5, 6 and 7. Cluster 4 has 10 sampling points, 1, 17, 15, 16, 9, 10, 14, 18, 26 and 29.

Components	Peaks	EX(nm)	EM(nm)	Substance	Reference
C1	(C1)	255	425	Photodegradation product	[46]
C2	W1	215	280	—	—
	W2	275	550	humic-like	[47]
	W3	265	280	Photodegradation product	[48]
	W4	515	550	—	—
C3	W5	240	485	Humic substances + recent materials	[49]
C4	W6	260	265	—	—
	W7	260	485	tyrosine-like	[50]

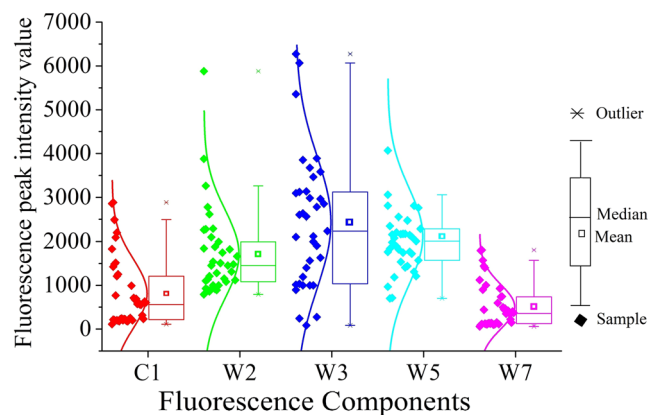**Table 1.** Peaks of the four PARAFAC components decomposed from water samples of the Jinghe Oasis.

Figure 3. Fluorescence peak intensity value statistics (map by Origin 9.1 (http://www.originlab.com/software)).

Figure 8 shows the distribution relation among the various variables and water quality parameters of different clusters. For example, C1, W2 and W7 are recorded in the right corner of the SOM network, thereby indicating a declining trend in the southern part of Ebinur Lake and the surrounding Kuitun River. The main samples of cluster 2 were from the Bortala River and the Jing River, which flow into the lake. Cluster 3 is mainly a class of six samples from the Akeqisu River and the Kuitun River. The rest of cluster 4 is located near farmland. To further visualize the four components, the results are shown in Figure 9.

Principal components of fluorescence spectra under different clusters. Data redundancy in fluorescence data is often important for spectral analysis. The principal component analysis method was used to reduce the dimension of data in this study. Three principal components were selected that accounted for a cumulative variance greater than 85%. The main components of the band and the main score matrix are shown in Table 3.

The contribution of land use/cover to water quality pollution in a 200 m radius. Using the multiple regression method, the contribution of land use/cover type to the fluorescence peak of the water body area was analyzed. R^2 and F values were used to test their correlations, as shown in Table 4. At cluster 1, the regression analysis between the land use/cover area and the principal components of the fluorescence spectra shown that there is no significant relationship between W7, W2 and the land type; similarly, there is no significant relationship between C1 and the corresponding clustering component of the land use, the contribution of salinized land and forest-grassland types ($R^2 = 0.91$). At cluster 2, the regression analysis between the land use/cover area and the principal components of the fluorescence spectra shows there is a significant relationship between W2 and the urban land and cropland, with R^2 of 0.49 and 0.67, respectively. At cluster 3, the regression analysis between the land use/cover area and the principal components of the fluorescence spectra shows that there is no significant relationship between W2 and the desert or salinized land, with R^2 of 0.69. At cluster 4, the regression analysis between the land use/cover area and the principal components of the fluorescence spectra shows there is no significant relationship between W2 and the salinized land and "other" area (R^2 is 0.66). Salinized and construction land types are major contributors to the dissolved organic matter affecting the surface water quality in the Jinghe Oasis.

Discussion

The water environment is reflected by the fluorescence peak.
Intensive land use in river watersheds and the rapid response of organic pollutants from different sources may cause the substantial deterioration of water quality, posing a direct or indirect threat to the quality of life of local people and the health of aquatic ecosystems[14,20]. Excitation-emission matrix (EEM) spectroscopy can be used to interpret a wide range of excitation and emission wavelengths contained within a variety of fluorescing water samples. In excitation-emission

Radius	components	Urban land	Water body	Cropland	Forest-grassland	Salinized land	Other land	Desert
100 m	C1	−0.088	−0.108	−0.204	−0.125	0.555**	−0.075	−0.062
	W2	0.019	0.177	0.03	−0.05	−0.413	0.008	−0.057
	W3	−0.06	0.262	−0.07	−0.003	−0.281	0.001	0.006
	W5	0.001	0.038	−0.063	−0.004	−0.187	0.147	−0.087
	W7	−0.087	−0.111	−0.211	−0.118	0.548**	−0.071	−0.053
200 m	C1	0.959**	−0.1	−0.246	−0.108	0.582**	−0.169	0.272
	W2	−0.345*	0.124	0.224	−0.068	−0.214	0.011	−0.044
	W3	−0.307*	0.194	0.15	−0.042	−0.3	0.068	−0.11
	W5	−0.194	−0.054	0.135	−0.039	−0.185	0.147	−0.073
	W7	0.932**	−0.104	−0.25	−0.099	0.571**	−0.164	0.275
300 m	C1	−0.085	−0.075	−0.287	−0.15	0.587**	−0.039	0.257
	W2	0.1	0.131	0.237	−0.098	−0.214	−0.049	−0.118
	W3	0.026	0.152	0.224	−0.064	−0.317	0.017	−0.159
	W5	0.111	−0.066	0.124	−0.065	−0.193	0.103	−0.116
	W7	−0.082	−0.079	−0.289	−0.146	0.575**	−0.032	0.255
400 m	C1	−0.108	−0.063	−0.265	0.127	0.535**	−0.146	0.238
	W2	0.182	0.145	0.201	−0.163	−0.209	−0.042	−0.17
	W3	0.084	0.11	0.191	−0.177	−0.339	0.109	−0.194
	W5	0.206	−0.061	0.107	−0.071	−0.209	0.081	−0.143
	W7	−0.103	−0.066	−0.267	0.129	0.524**	−0.138	0.237
500 m	C1	0.771**	−0.461**	−0.303	0.566**	0.510*	−0.549*	−0.109
	W2	0.482	0.336	0.666**	−0.125	−0.094	0.536*	0.208
	W3	−0.226	0.237	0.319	−0.133	−0.145	0.143	0.288
	W5	−0.023	0.012	0.338	−0.081	−0.198	0.041	−0.095
	W7	0.771**	−0.164	0.021	−0.017	−0.205	−0.242	−0.527**

Table 2. Relationship between the peak values of fluorescence and the land use/cover types under different water radii of action. **Correlation is significant at the 0.01 level. *Correlation is significant at the 0.05 level.
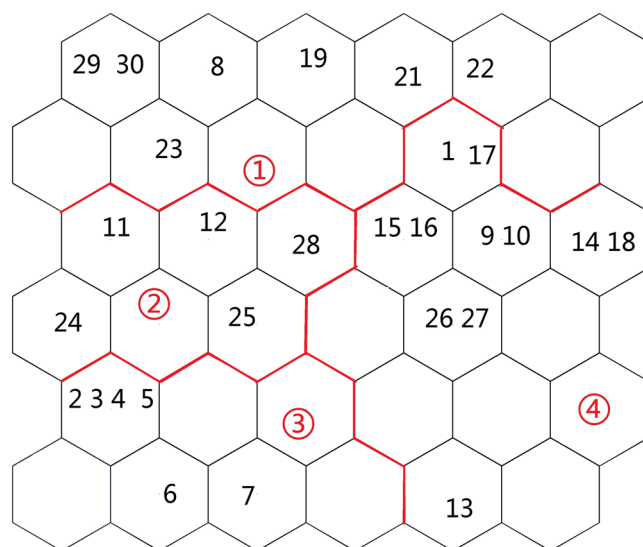

Figure 4. Sample distribution map of SOM analysis (Map by Matlab (https://www.mathworks.com/software)).

fluorescence, there is a relationship between fluorescence peaks and the number of water quality parameter[21]. The fluorescence peak of a water body reflects its environmental conditions. Research by Wang *et al.*, indicates that 3D-fluorescence techniques are capable of estimating and monitoring surface water pollution in the Jinghe Oasis[14]. Fluorescence spectroscopy has received much attention in recent years due to its potential application in monitoring the water of rivers and lakes. This technique is attractive for monitoring the water quality in inland water bodies, as it is a rapid technique that requires no reagents and no sample preparation for analysis. The relationship between water quality and land use/cover is equivalent to the relationship between the peak of
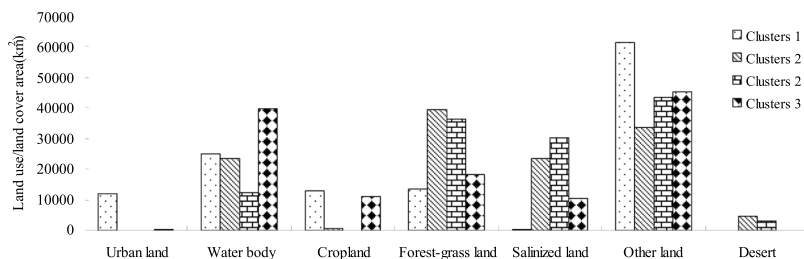
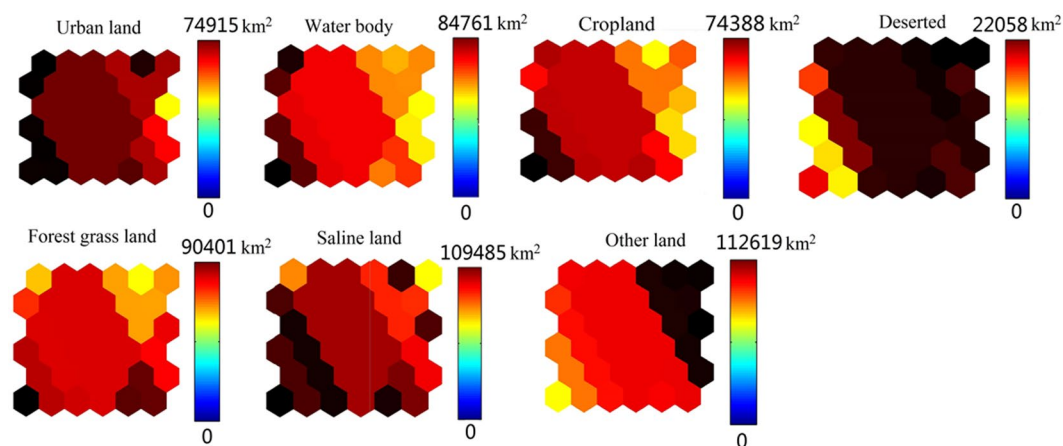**Figure 5.** Average values for the land use/cover type area (map by Excel (https://www.microsoft.com/software)).

Figure 6. Visualization map of fluorescence values of components. (Map by Matlab (https://www.mathworks.com/software)).
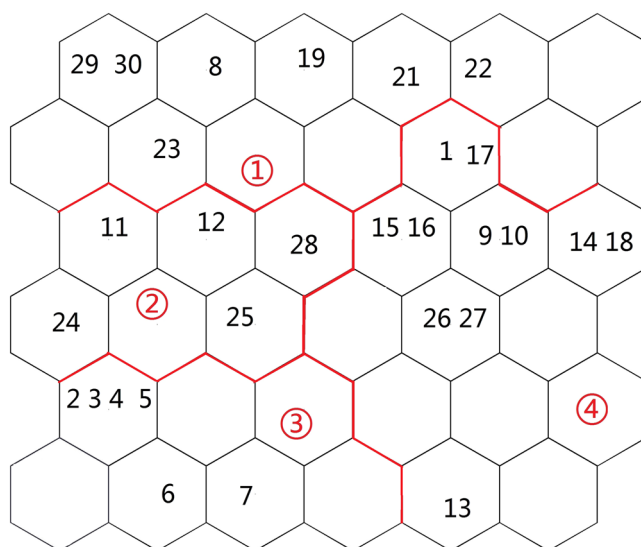
Figure 7. Sample distribution map of SOM analysis (map by Matlab (https://www.mathworks.com/software)).

fluorescence and land use/cover. Kiedrzyńska *et al*. found that cropland areas were found to influence nitrogen, and forests areas were negatively related to loads of both nitrogen and phosphorus[19], important organic matter factors in this study. Therefore, the results of this study are consistent with previous results.

The influence of spatial and temporal scales. The influence of land use/cover patterns on water quality fluorescence is scale dependent[22]. The results showed that the 200 m action radius was better than the other action radii in explaining the overall fluorescence variations. The 200 m action radius acts as a filter by reducing surface
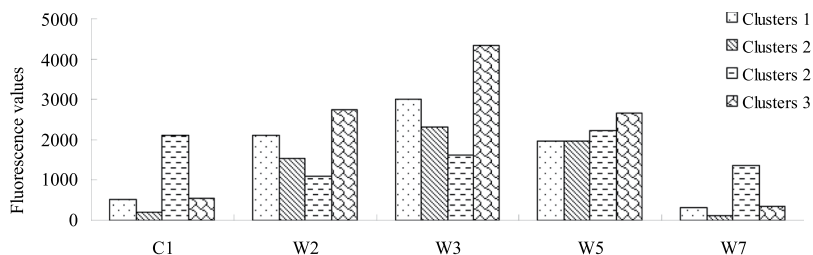
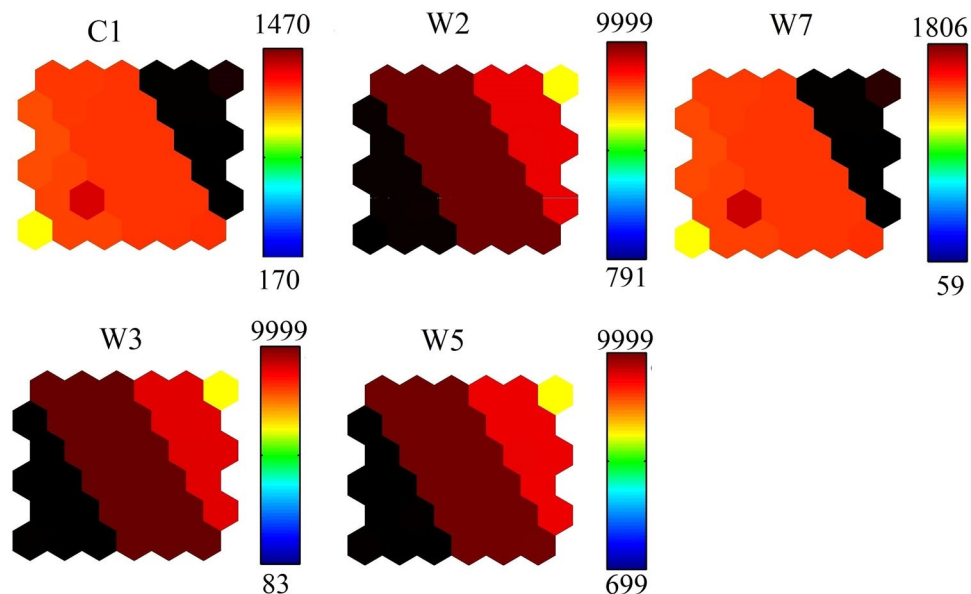**Figure 8.** Average values for the peak value of fluorescence (map by EXCEL (https://www.microsoft.com/software)).

Figure 9. Visualization map of the peak values of fluorescence (map by Matlab (https://www.mathworks.com/software)).

	Cluster 1			Cluster 2			Cluster 3			Cluster 4		
Types	1	2	3	1	2	3	1	2	3	1	2	3
C1	−0.262	**0.960**	−0.096	0.693	0.517	0.458	−0.964	0.180	0.166	−0.701	0.703	−0.024
W2	0.879	0.467	−0.088	**0.952**	−0.161	0.129	**0.995**	0.073	−0.061	**0.861**	0.481	0.000
W3	**0.976**	0.008	0.186	−0.504	0.828	−0.212	0.883	0.099	0.454	**0.902**	0.311	−0.092
W5	0.903	0.148	0.203	0.908	0.11	−0.083	0.824	0.562	−0.052	0.816	0.508	0.084
W7	−0.232	0.961	−0.133	0.028	**0.993**	−0.052	−0.977	0.115	0.153	−0.695	0.706	−0.043

Table 3. Surface water fluorescent components score matrix in the Jinghe Oasis.

Cluster	y	X (Area)	Model	R²	F	SD
1	C1	Salinized land, Forest-grassland	$y = 0.002x_{Salinized\ land} - 0.008x_{Forest-grassland} + 734.46$	0.91	3.02	59.8
2	W2	Urban land, Cropland	$y = 0.006x_{Urban\ land} - 0.008x_{Cropland} + 408.69$	0.67	3.11	57.2
2	W7	Urban land, Cropland	$y = -0.001x_{Urban\ land} - 0.005x_{Cropland} + 121.28$	0.49	1.46	11.8
3	W2	Desert, Salinized land	$y = 00.063x_{Desert} - 0.007x_{Salinized\ land} + 1591.4$	0.69	3.39	269.1
4	W3	Salinized land, Other land	$y = 0.006x_{Other\ land} - 0.03x_{Salinized\ land} + 4402.79$	0.66	2.98	70.2

Table 4. Contribution of land use/cover to the fluorescence peaks of dissolved organic matter in surface water.

runoff, processing nutrients to improve the water quality of rivers[23–26]. The radius of action was first introduced to explore the effect of land use on water quality in arid areas. Although it is essentially a case of buffer analysis, it also provides a new way of thinking about the effect of land use on surface water quality. The radius of action

is based on the analysis of geospatial data, which lacks a strong theoretical basis, the primary problem to be addressed by future research.

Management suggestions. The urban areas and salinized land in this watershed were mainly dispersed along the river, which had a negative effect on the river water quality. Consistent with a typical continental climate, this region is extremely dry and windy, and it has little rainfall and frequent dust storms in the Jinghe Oasis. Therefore, it is important to control urban runoff and ensure that the water quality meets national standards. Salinized land had a strong impact on water quality at a large scale. Reasonable irrigation and soil salt improvements are important measures. Forest-grassland areas have strong contributions to water quality variations at the river scale. Landscape pattern planning should be used to improve the water quality of watersheds in the arid region of central Asia.

Conclusion

In the scope of applied conservation, understanding the impact of the surrounding land use/cover and human activities on the water quality at multiple scales is essential to adapt scale appropriate strategies to protect and rehabilitate in basin scale. The influence of the landscape on the water quality is scale dependent; this scale is beneficial to the management of water quality, convenience, economy and safety. Therefore, the key problem of the effective management of surface water is identifying the effective radius of the surface water pollution and blocking the pollution source. The Jinghe Oasis is located in the China-Kazakhstan border in the Xinjiang Uyghur Autonomous Region of China; we demonstrated the potential of integrated remote sensing and three-dimensional fluorescence technologies to investigate the effect of land use/cover types on surface water. The PARAFAC model and self-organizing map (SOM) neural network model were used to determine the effects of land use/cover types on water quality.

(1) The four fluorescence components that were successfully extracted by the PARAFAC factor analysis modeling from the fluorescence EEM data are as follows: microbial humic-like (C1), terrestrial humic-like organic substances (C2, C4), and protein-like organic substances (C3).

(2) Taking 30 water sampling points to build 5 buffer zones (100 m, 200 m, 300 m, 400 m, and 500 m), we found that "the most significant relationship between land use type and fluorescence components is found with 200 m radios, and the maximum contribution is from buildup land and salinized land.

(3) Typical three-dimensional fluorescence peak and land use type were classified by the SOM neural network method, which demonstrated that four different types exist between three-dimensional fluorescence peak and land use types.

(4) A principal factor analysis method applied to four fluorescence peaks and the stepwise multiple regression method showed that the clustering type contributes mostly to the surface water organic pollution in Ebinur lake were salinized land and cropland land, which was the contributed source of C1, W2, W7 fluorescence peak; salinized land was the most contributed source of W2, W3 fluorescence peak.

The relationship between the land use/cover and the water quality is complicated and can be influenced by numerous factors. From the perspective of the landscape, the design of effective radius of the surface water pollution is not simply related to the land use types, but it also depends on the spatial structure of the land use types. Therefore, the effective radius can better solve the water quality management in the watershed.

Materials and Methods

Study area. The Jinghe Oasis is located in the center of Eurasia in the northwest Xinjiang Uygur Autonomous Region at $44°02'\sim45°10'$N and $81°46'\sim83°51'$E. The Jinghe Oasis is composed of wetland and desert oasis vegetation and wildlife and is a national desert ecological reserve. The study area has a unique wetland ecological environment, and it has been listed as the Xinjiang Uygur Autonomous Region "Wetland Nature Reserve" (Figure 10). The Jinghe Oasis was once fed by 12 branch rivers belonging to three major river systems, and the major rivers were the Bortala River (B-R), Jinghe River (J-R), Aqikesu River (AQKS-R) and Kuitun River (K-R). Owing to natural environmental changes and human activities (i.e., modern oasis agricultural development), many rivers gradually lost their hydraulic connections with Ebinur Lake, and only the Bortala River and Jinghe River now supply water to Ebinur Lake. The western region of the Bortala River (B-R) valley is south of the Jing River (J-R) oases and the Dandagai Desert and east of the Mutetaer desert zone of the lower reaches of the Akeqisu–Kuitun River (AKQS-R). Ebinur Lake is located in the center of a watershed at the lowest elevation and represents a typical lake of the arid areas of Central Asia. The total watershed area is 50,621 km^2. It is surrounded by a mountainous region (24,317 km^2; Alatau Mountains) and plain areas (26,304 km^2) to the north, west and south[27,28]. The climate is a typical temperate arid continental climate, and the mountain–oasis–desert system has typical temperate arid ecological characteristics. The study region is located inland (2000 km away from the Pacific and Indian Ocean and 3000 km away from the Arctic Ocean); the moisture sources in the study area are derived from the Atlantic Ocean (7000 km), but overall, there is limited water vapor transport from maritime areas[29].

Data sources and Data processing. *Water sample acquisition and processing.* Water samples were collected on 5 July 2016 from 29 locations within the Jinghe Oasis, a typical arid oasis. The collected samples were kept in low-temperature cold storage (under 2 °C) during transport before the water quality measurements were carried out in the laboratory. Samples were transported in polyethylene plastic bottles, previously washed in 10% HCI and cleaned with deionized water, to minimize changes in the water chemical characteristics.

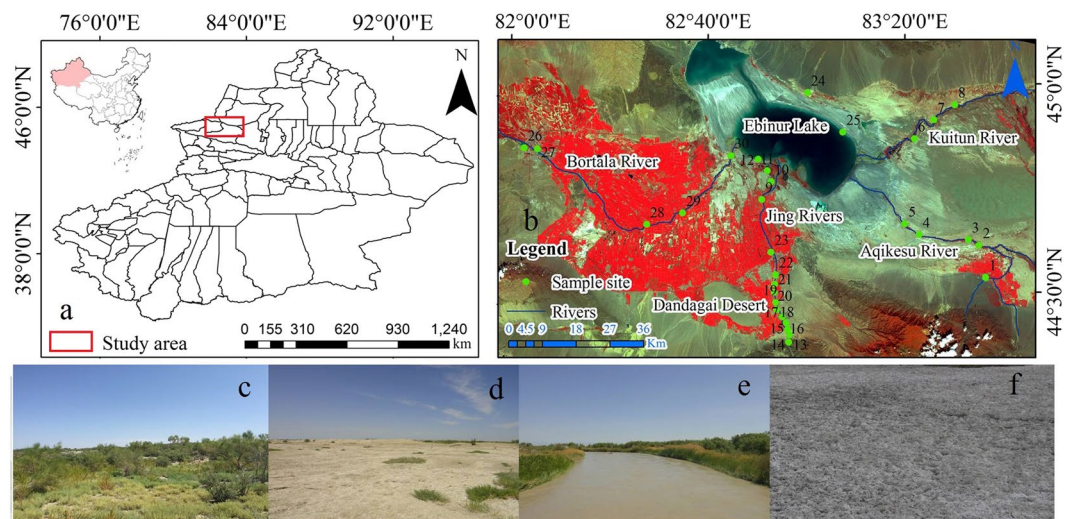| Band	Reflectance	Resolution
1	Blue (0.45–0.52) um	16
2	Green (0.52–0.59) um	16
3	Red (0.63–0.69) um	16
4	NIR (0.77–0.89) um	16
5	PAN (0.52–0.89) um	8

Table 5. Image information for GF-1.

Figure 10. (**a**) Map of the study area with an inset map showing the location of the Xinjiang Autonomous Region in China; (**b**) satellite map of the study area; (**c**) forest-grassland, (**d**) other land, (**e**) water body, (**f**) salinized land (photographed (**c**–**f**) by Xiaoping Wang (map of (**a**,**b**) by ArcGIS10.2.2 (http://www.esri.com/software/arcgis)).

The collected samples were filtered using a pre-washed GF/F filter. All fluorescence intensities were determined using a Cary Eclipse Fluorescence Spectrophotometer (F-7000, Hitachi High-Technology Corp, Tokyo, Japan). The spectrophotometer that measured the fluorescence spectra was equipped with a 150 W xenon arc lamp as the light source and two grating monochromators coupled with a slit as the EEM wavelength selectors. The scanning speed was set to 60000 nm/min. Therefore, each measurement period was 2 minutes in duration. The EEMs were measured every 10 nm over an excitation range of 200–450 nm, with an emission range of 200–550 nm by 10 nm. MilliQ water EEMs were used as blanks and subtracted from each sample EEM. The emission and excitation correction files generated by the FluoroMax manufacturer were applied to each MilliQ-subtracted sample EEM. Fluorescence intensities were standardized to a Raman peak at 395 nm emission, as suggested by Lawaetz[30].

Remote sensing data. Medium spatial resolution cloud-free GF-1 images were used in this study (Table 5) and were acquired near the actual water sampling date on April 12th, 2016. These data were obtained from CRESDA (China Resources Satellite Application Center, http://www.cresda.com/CN/). The images were acquired in clear and dry weather conditions during the dry season. These GF-1 satellite images have a ground sampling distance (GSD) of 16 m and a pan band GSD of 8 m. The GF-1 images contain 5 bands that record the reflected or emitted radiation from the Earth's surface in the B, G, R NIR and pan bands of the electromagnetic spectrum.

The sensor and atmosphere can cause changes to the spectral characteristics of a target in a multitemporal remote sensing image, which affects the extraction of image data[31]. Therefore, the remote sensing images must undergo an atmospheric correction. To accomplish this correction, the pre-processing of GF-1 images was first conducted using ENVI 5.1 (Environment for Visualizing Images 5.1) (ExelisVisual Information Solution Corporation, America) software. Universal Transverse Mercator (UTM) projection was selected to rectify the satellite images. The GF-1 images were also geometrically corrected with a previously corrected GF-1 image that had a geometric accuracy of <0.5 pixels[32]. Then, the ENVI 5.1 radiometric calibration tool and the gain and deviation ratio in the GF-1 image data head document were used for the radiometric calibration of GF-1 data. Finally, the ENVI5.1 FLAASH atmospheric correction model was used for the atmospheric correction of the remote sensing images.

Methods. The methodology is explained in the following section, with a conceptual flow chart describing the methodology. Figure S1 shows the workflow of the study detailed in the following sections.

Data Fusion. Data fusion is useful because it takes advantage of different spectral and/or resolution information for effective image interpretation. Pan bands and four multispectral band images were used for the image fusion[33]. We implemented selective principal component analysis (S-PCA) transformation, rather than the conventional standard PCA method.

Land use land cover clustering based on decision tree classification. The decision tree (DT) classifier is a simple and widely used classification technique. The DT classifier is an effective method to incorporate a variety of data types from multiple sources to find pixels that fulfill the criteria[34–36]. We conducted radiation and orthographic corrections for the remote sensing image data combined with 1:50,000 digital elevation model (DEM) data. We established five land use/cover types by using the Environment for Visualizing Images software (ENVI Version 5.0), urban land, cropland, forest-grassland, water body, salinized land, desert and others based on the actual conditions of the research zone. The final results showed that the producer's accuracy of the classified LCLUC maps was 85.29%. The user's accuracy of the classified LCLUC maps was 84.47%. The overall accuracy was 89%, and the kappa coefficient was 0.88.

PARAFAC Modeling. PARAFAC applications allow full use of the fluorescence EEM data samples. Fluorescence spectrum data multiplexer (three-way) as a sample of the fluorescent changes depends on the wavelength of light absorption (excitation) and fluorescence wavelength was observed (emission). PARAFAC decomposes the EEM dataset into a set of trilinear terms and a residual array[37] and fits an equation by minimizing the residual sum of squares of three linear models as follows:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{if} c_{kf} + \varepsilon i_{jk}, \qquad i = 1 \dots \text{I}; j = \dots. \text{J}; \quad \text{K} = 1 \dots \text{K} \tag{1}$$

where x_{ijk} is the intensity of the three-way data array for the ith sample at the emission wavelength j and excitation wavelength k; a_{if} is directly proportional to the concentration of the fth three-way data array in the ith sample (defined as scores); b_{if} and C_{kf} are the estimates of the three-way data; F represents the number of components in the model; and e_{ijk} is the residual element, representing the variability not accounted for by the model[38,39].

MATLAB 2014a (MathWorks, Natick, MA, USA) and the DOM Fluor toolbox (http://www.models.life.ku.dk) were selected to perform the complete PARAFAC modeling. Residual analysis, spectral scores at the core of the consistency and visual inspection of each group were selected to diagnose the correct number of components[37]. The fluorescence intensity of each component (I_i) was estimated using the following formula[40]:

$$I_i = Score_s \times Ex_i(\lambda \text{ max}) \times Em_i(\lambda \text{ max}) \tag{2}$$

where $Score_s$ represents the ith relative fluorescence intensity; $Ex_i(\lambda \text{ max})$ represents the first N components to stimulate the maximum load; and $Em_i(\lambda \text{ max})$ represents the first n components of the emission load of the maximum number.

Analysis of water pollution radius. The "radius of action" was first introduced in the engineering field to discuss the quantity of blasting and the accurate blasting scope in the multiboundary blasting system. The term "radius of water quality action" was introduced to understand the effective range of land use in watersheds. To summarize the results of previous studies, we explored the effect of land use at a range of scales, 100 m, 200 m, 300 m, 400 m to 500 m, on the water quality in rivers. The framework is designed as shown in Figure S2.

Recognition of land use/cover and fluorescent component spatial characteristics based on the self-organizing map SOM method. A self-organizing map (SOM) is one of the branches of artificial neural network algorithms. It is a self-organizing and self-learning network visual method that can express multidimensional spatial data in low-dimensional points through non-linear mapping[41]. The SOM is an all-purpose classification tool that can connect samples with variables[42]. In recent years, the SOM has become increasingly popular in environmental research because of its capacity to address non-linear relations. The idea motivating the development of the SOM method was to represent a large amount of data typical of samples. The map usually has a 2D structure with a map unit associated with a weight vector.

$$N_{ij} = \{W_{ij}; 1 \le i \le L, 1 \le J \le M\} \tag{3}$$

where N_{ij} is a 2D map grid (also called a neuron); W_{ij} is the weight vector assigned to (i, j), the unit of SOM architecture; and L and M are number of rows and columns, respectively[43–45]. The steps of the SOM algorithm are displayed as follows:

- **Step 1:** Data normalization and SOM network initialize, the weight vector w_{ij} (*i* = 1, 2, ……, S; *j* = 1, 2, 3, ……, R) is randomly set in the interval [0, 1], R is the sample dimension, and S is the number of output neurons.
- **Step 2:** An input vector $p_k = (p_1^k, p_2^k, \dots p_R^K)$ (*k* = 1, 2… M, where M is the number of samples) is presented to the SOM network, and distance is calculated.
- **Step 3:** The smallest distance is chosen, and the best-matching unit (BMU) is identified.

- **Step 4:** The weight vector w_{ij} in the neighbor ration is updated.
- **Step 5:** The process proceeds in an iterative way until the optimal number of iteration steps is satisfied, and then it returns to step 2.

The SOM technique has a distinct capability to represent the complex relationships of the fluorescence intensity data and the land use area data using component planes and U-matrix. All simulations were implemented in MATLAB R2014a using an SOM toolbox.

Statistical Analyses. A descriptive statistical analysis was applied to evaluate the water quality indices and EEM-PARAFAC and 3D Fluorescence spectral index. Linear regression and correlation analyses were constructed using Origin 8.0 (OriginLab Corporation, America). The significances of the correlations in the statistics were evaluated using P values and t values.

References

1. Xiao, R., Wang, G., Zhang, Q. & Zhang, Z. Multi-scale analysis of relationship between landscape pattern and urban river water quality in different seasons. *Scientific Reports* https://doi.org/10.1038/srep25250 (2016).
2. Guo, Q. H., Ma, K. M., Liu, Y. & He, K. Testing a dynamic complex hypothesis in the analysis of land use impact on lake water quality. *Water Resources Management* **24**(7), 1313–1332 (2010).
3. Qiang, C. *et al*. Impacts of land use and population density on seasonal surface water quality using a modified geographically weighted regression. *Science of the total environment* **572**, 450–466 (2016).
4. Liu, W., Zhang, Q. & Liu, G. Influences of watershed landscape composition and configuration on lake-water quality in the Yangtze River basin of China. *Hydrological Processes* **26**(4), 570–578 (2012).
5. Jiang, M., Chen, H., Chen, Q. & Wu, H. Study of landscape patterns of variation and optimization based on non-point source pollution control in an estuary. *Marine Pollution Bulletin* **87**(1–2), 88 (2014).
6. Clough, Y. *et al*. Land-use choices follow profitability at the expense of ecological functions in Indonesian smallholder landscapes. *Nature Communications* **7**(65145), 13137, https://doi.org/10.1038/ncomms13137 (2016).
7. Li, S., Sheng, G., Xiang, T. & Zhang, Q. Water quality in the upper Han River basin, China: the impacts of land use/land cover in riparian buffer zone. *Journal of Hazardous Materials* **165**(1–3), 317–324 (2009).
8. Sahu, M. & Gu, R. R. Modeling the effects of riparian buffer zone and contour strips on stream water quality. *Ecological Engineering* **35**(8), 1167–1177 (2009).
9. Kang, J. H. *et al*. Linking land-use type and stream water quality using spatial data of fecal indicator bacteria and heavy metals in the Yeongsan river basin. *Water Research.* **44**, 4143–4157 (2010).
10. Dinç, E., Ertekin, Z. C. & Büker, E. Multiway analysis methods applied to the fluorescence excitation-emission dataset for the simultaneous quantification of valsartan and amlodipine in tablets. *Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy* **184**, 255–261 (2017).
11. Spagnuolo, M. L., Marini, F., Sarabia, L. A. & Ortiz, M. C. Migration test of bisphenol a from polycarbonate cups using Excitation-Emission fluorescence data with parallel factor analysis. *Talanta* **167**, 367–378 (2017).
12. Pan, H., Lei, H., Xin, L., Wei, H. & Liu, S. Assessment on the leakage hazard of landfill leachate using three-Dimensional Excitation-Emission fluorescence and parallel factor analysis method. *Waste Management.* **67**, 214–221 (2017).
13. Jacquin, C., Lesage, G., Traber, J., Pronk, W. & Heran, M. Three-dimensional Excitation and Emission matrix fluorescence (3DEEM) for quick and pseudo-quantitative determination of protein- and humic-like substances in full-scale membrane bioreactor (MBR). *Water Research* **118**, 82–92 (2017).
14. Wang, X. P. *et al*. Evaluation and estimation of surface water quality in an arid region based on EEM-parafac and 3d fluorescence spectral index: a case study of the Ebinur lake watershed, china. *Catena* **155**, 62–74 (2017).
15. Zeinalzadeh, K. & Rezaei, E. Determining spatial and temporal changes of surface water quality using principal component analysis. *Journal of Hydrology Regional Studies* **13**, 1–10 (2017).
16. Stedmon, C. A., Markager, S. & Bro, R. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry* **82**(3–4), 239–254 (2003).
17. Heibati, M. *et al*. Assessment of drinking water quality at the tap using fluorescence spectroscopy. *Water Research* **125**, 1–10 (2017).
18. Zhao, Y. *et al*. Evaluation of cdom sources and their links with water quality in the lakes of northeast china using fluorescence spectroscopy. *Journal of Hydrology* **550**, 80–91 (2017).
19. Kiedrzyńska, E., Jóźwik, A., Kiedrzyński, M. & Zalewski, M. Hierarchy of factors exerting an impact on nutrient load of the Baltic Sea and sustainable management of its drainage basin. *Marine Pollution Bulletin* **88**, 162–173 (2014).
20. Carey, R. O. *et al*. Land use disturbance indicators and water quality variability in the Biscayne Bay watershed, Florida. *Ecological Indicators* **11**(5), 1093–1104 (2011).
21. Hudson, N., Baker, A. & Reynolds, D. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters—a review. *River Research & Applications* **23**(6), 631–649 (2010).
22. Ding, J. *et al*. Influences of the land use pattern on water quality in low-order streams of the Dongjiang river Basin, China: a multi-scale analysis. *Science of the total environment* **551–552**, 205–216 (2016).
23. Miserendino, M. L. *et al*. Assessing land-use effects on water quality, in-stream habitat, riparian ecosystems and biodiversity in Patagonian northwest streams. *Science of the total environment.* **409**(3), 612–624 (2011).
24. Fernandes, J. D. F., de Souza, A. L. & Tanaka, M. O. Can the structure of a riparian forest remnant influence stream water quality? A tropical case study. *Hydrobiologia* **724**(1), 175–185 (2014).
25. Baker, A. & Inverarity, R. Protein-like fluorescence intensity as a possible tool for determining river water quality. *Hydrological Processes* **18**(15), 2927–2945 (2004).
26. Onyutha, C. & Willems, P. Influence of spatial and temporal scales on statistical analyses of rainfall variability in the river Nile basin. *Dynamics of Atmospheres & Oceans* **77**, 26–42 (2017).
27. He, X. *et al*. Effects of simulated nitrogen deposition on soil respiration in a populus euphratica community in the Ebinur lake area, a desert ecosystem of northwestern China. *Plos One* **10**(9), e0137827 (2015).
28. Yu, H. Y. *et al*. Analysis of land cover and landscape change patterns in Ebinur lake wetland national nature reserve, China from 1972 to 2013. *Wetlands Ecology & Management* **2017**, 1–19 (2017).
29. Jilili, A., Gabchenko, M. V. & Xu, J. Eolian transport of salts-a case study in the area of lake Ebinur (Xinjiang, northwest china). *Journal of Arid Environments* **72**(10), 1843–1852 (2008).
30. Lawaetz, A. J. & Stedmon, C. A. Fluorescence intensity calibration using the raman scatter peak of water. *Applied Spectroscopy* **63**(8), 936–940 (2009).
31. Hadjit, H., Oukebdane, A. & Belbachir, A. H. Atmospheric correction of earth-observation remote sensing images by monte carlo method. *Journal of Earth System Science* **122**(5), 1219–1235 (2013).
32. Pahlevan, N. *et al*. Landsat 8 remote sensing reflectance (RRS) products: evaluations, intercomparisons, and enhancements. *Remote sensing of environment* **190**, 289–301 (2017).

33. Chen, B., Huang, B. & Xu, B. Multi-source remotely sensed data fusion for improving land cover classification. *Isprs Journal of Photogrammetry & Remote Sensing* **124**, 27–39 (2017).
34. Dubovyk, O. *et al.* Spatial targeting of land rehabilitation: a relational analysis of cropland productivity decline in arid Uzbekistan. *Erdkunde* **67**(2), 167–181 (2013).
35. Guermazi, E., Bouaziz, M. & Zairi, M. Water irrigation management using remote sensing techniques: a case study in central tunisia. *Environmental Earth Sciences* **75**(3), 202–214 (2016).
36. Panigrahi, N. & Prashnani, M. Impact evaluation of feature reduction techniques on classification of hyper spectral imagery. *Journal of the Indian Society of Remote Sensing* **43**(1), 1–10 (2015).
37. Zeng, Z. *et al.* Source analysis of organic matter in swine wastewater after anaerobic digestion with EEM-PARAFAC. *Environmental Science & Pollution Research International* **24**(7), 6770–6778 (2017).
38. Vera, M. *et al.* Monitoring UF membrane performance treating surface-groundwater blends: Limitations of EEM-PARAFAC on the assessment of the organic matter role. *Chemical Engineering Journal* **317**, 961–971 (2017).
39. Lenhardt, L. *et al.* Characterization of cereal flours by fluorescence spectroscopy coupled with parafac. *Food Chemistry* **229**, 165–171 (2017).
40. Zhang, S., Chen, Z., Wen, Q. & Zheng, J. Assessing the stability in composting of penicillin mycelia dreg via parallel factor (PARAFAC) analysis of fluorescence excitation–emission matrix (EEM). *Chemical Engineering Journal* **299**, 167–176 (2016).
41. Kohonen, T. *Self-Organizing Maps.* (Springer, Berlin, 2001).
42. Kohonen, T. Essentials of the self-organizing map. *Neural Networks* **37**, 52–65 (2013).
43. Farzad, F. & El-Shafie, A. H. Performance enhancement of rainfall pattern–water level prediction model utilizing self-organizing-map clustering method. *Water Resources Management* **31**, 1–15 (2016).
44. Matić, F. *et al.* Oscillating adriatic temperature and salinity regimes mapped using the self-organizing maps method. *Continental Shelf Research* **132**, 11–18 (2017).
45. An, Y., Zou, Z. & Li, R. Descriptive characteristics of surface water quality in Hong Kong by a Self-Organizing map. *International Journal of Environmental Research & Public Health* **13**(1), 115 (2016).
46. Buck, O., Niyogi, D. K. & Townsend, C. R. Scale-dependence of land use effects on water quality of streams in agricultural catchments. *Environmental Pollution* **130**(2), 287–299 (2004).
47. Hur, J. & Cho, J. Prediction of BOD, COD, and total nitrogen concentrations in a typical urban river using a fluorescence excitation-emission matrix with PARAFAC and UV absorption indices. *Sensors* **12**(1), 972–986 (2002).
48. Yu, G. H., He, P. J. & Shao, L. M. Novel insights into sludge dewater ability by fluorescence excitation–emission matrix combined with parallel factor analysis. *Water Research* **44**(3), 797–806 (2010).
49. Vignudelli, S., Santinelli, C., Murru, E., Nannicini, L. & Seritti, A. Distributions of dissolved organic carbon (DOC) and chromophoric dissolved organic matter (CDOM) in coastal waters of the northern Tyrrhenian Sea (Italy). *Estuarine Coastal & Shelf Science* **60**(1), 133–149 (2004).
50. Henderson, R. K., Baker, A., Murphy, K. R. & Hambly, A. Fluorescence as a potential monitoring tool for recycled water systems: a review. *Water Research.* **43**, 863–881 (2009).

Acknowledgements

Author Contributions

Xiaoping Wang led the idea conceptualization, analysis, figure generation, and writing. Xiaoping Wang and Fei Zhang discussed idea conceptualization. Xiaoping Wang and Fei Zhang contributed to editing and organization of the revised paper. All co-authors discussed the results and commented on the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-31265-0.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.