

1 **Assessment of groundwater evolution and its driving factors in an intensified**
2 **agricultural area: An integrated hydrochemical and machine learning approach**

3

4 Yao Wu^{a,b}, Zexin Wu^{a,b*}, Yexiang Yu^a, Lei Wang^c, Yingna Sun^b, Guangxin Zhang^{a*}

5

6 ^a State Key Laboratory of Black Soils Conservation and Utilization, Northeast Institute
7 of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102,
8 China

9 ^b School of Hydraulic and Electric-Power, Heilongjiang University, Harbin 150080,
10 China

11 ^c British Geological Survey, Keyworth, Nottingham NG12 5GG, UK

12

13

14

15

16

17

18

19

20

21

22

23

24

25 * Correspondence: wuzexin@iga.ac.cn

26 gxzhang@iga.ac.cn

27

28

29 **Abstract:**

30 In recent years, due to alterations in the natural environment and intensified
31 anthropogenic activities, there has been a widespread degradation in groundwater
32 quality in China, with the harmful effects of this decline most evident in agricultural
33 areas. However, it remains challenging to accurately distinguish the impacts of different
34 factors on groundwater quality changes, complicating efforts to effectively manage and
35 prevent groundwater pollution. This study integrates hydrochemical analysis, the
36 Entropy-Weighted Water Quality Index (EWQI), Self-Organizing Maps (SOM), and
37 the Boruta algorithm to explore the chemical evolution of groundwater and its driving
38 factors in the Sanjiang Plain, an important grain-producing region in China. Results
39 showed that compared to 2012, the quality of deep groundwater in the Sangjiang Plain
40 improved in 2022, while the quality of shallow groundwater significantly deteriorated.
41 The shallow groundwater classification transitioned from a predominantly $\text{HCO}_3\text{-Ca-}$
42 Mg type, influenced chiefly by water-rock interactions, to a mixed type comprising
43 $\text{HCO}_3\text{-Ca-Mg}$ and $\text{SO}_4\text{-Cl-Ca-Mg}$. Results of hydrochemical analysis and EWQI
44 evaluations indicated that elevations in chloride, sulfate, and nitrate levels due to
45 anthropogenic activities are the principal drivers of the decline in shallow groundwater
46 quality. The SOM results suggested that land use types significantly affect the quality
47 of shallow groundwater. Further analysis with the Boruta algorithm indicated that the
48 primary factors contributing to the deterioration of shallow groundwater quality were
49 the increased emissions of sewage and manure from expanding livestock operations
50 and heightened pollutant leakage due to the expansion of paddy fields. This study
51 provides new insights into the mechanisms driving groundwater evolution in intensive
52 agricultural areas affected by human activities and hydrogeochemical processes. The
53 findings could contribute to the prevention and control of groundwater pollution in the
54 Sanjiang Plain and may also serve as a reference for managing groundwater
55 contamination in other similar agricultural regions.

56 **Keywords:** Groundwater; Hydrochemistry; Water quality; Machine learning;

57 Sanjiang Plain

58

59 **1 Introduction**

60 Groundwater, the largest freshwater reservoir on earth, plays a critical role in
61 sustaining various human activities, including drinking, agricultural irrigation, and
62 industrial operations (Lapworth et al., 2022). However, the overexploitation and
63 excessive use of groundwater resources have resulted in a widespread decrease in the
64 water levels of numerous aquifers around the world (Jasechko et al., 2021; Konikow et
65 al., 2005; Rodell et al., 2018). In addition, increasingly intense human activities have
66 led to widespread deterioration of groundwater quality, ranging from localized
67 contamination caused by discharges from septic tanks or pit latrines (Graham et al.,
68 2013), to extensive nitrate pollution and increased salinity in large aquifer systems due
69 to the application of agricultural fertilizers and irrigation practices (Ascott et al., 2017;
70 Pulido-Bosch et al., 2018). The degradation of groundwater quality, in turn, poses
71 potential threats to human health and the growth of crops, emerging as a crucial
72 determinant in constraining the availability of groundwater resources. Indeed, even in
73 some heavily exploited aquifer systems, the constraints imposed by deteriorating
74 groundwater quality have already deteriorating groundwater quality have now
75 exceeded those caused by depletions in groundwater resulting from over-pumping
76 (MacDonald et al., 2016). Therefore, investigating the long-term evolution of the water
77 quality of aquifers impacted by human activities and identifying the main driving
78 factors are crucial for effective groundwater resource protection and pollution control.

79 In general, the physicochemical properties and quality of groundwater in an area
80 are significantly influenced by both natural processes and anthropogenic activities (Yan
81 et al., 2021). Under natural conditions, the interaction between water and rock leads to
82 the dissolution of minerals from the bedrock into the groundwater, resulting in the
83 progressive accumulation of ions and determining the basic hydrochemical properties
84 of regional groundwater (Elango et al., 2007). Concurrently, anthropogenic activities

85 also introduce contaminants that strongly affect the chemical composition of
86 groundwater (Huang et al., 2013). Given the multitude of natural and anthropogenic
87 factors affecting groundwater chemistry, coupled with the complex interplay among
88 these influences, it is very challenging to distinguish the impact of different factors. In
89 recent years, owing to its strong capabilities in result visualization and capturing the
90 spatiotemporal characteristics of nonlinear complex systems, machine learning
91 approaches have been increasingly introduced into the study of groundwater quality
92 evolution and have achieved inspired results (Deng et al., 2023; Kim et al., 2020;
93 Nguyen et al., 2015; Yin et al., 2021; Zhong et al., 2022). Self-organizing maps (SOM)
94 is now one of the most widely used machine learning approaches in groundwater quality
95 study. It is a powerful neural network model based on unsupervised learning algorithms,
96 capable of transforming complex nonlinear statistical relationships between high-
97 dimensional data into simple geometric relationships (Kohonen, 2001). The SOM
98 method is typically used in conjunction with clustering methods, such as hierarchical
99 clustering, to analyze the hydrochemical parameters of groundwater samples and group
100 the samples into different clusters (Kim et al., 2020; Lee et al., 2019; Nguyen et al.,
101 2015; Zhong et al., 2022). Then, the potential mechanisms underlying the formation of
102 these groundwater clusters can be explored by examining the correlations among the
103 hydrochemical parameters derived from SOM analysis, in conjunction with the local
104 hydrogeological conditions. Recently, some feature selection algorithms, such as the
105 Boruta algorithm, have been introduced to identify the driving factors of water quality
106 evolution. They can display the importance of different factors in water quality
107 evolution more intuitively than the SOM method, which can support and validate the
108 results obtained by the SOM method. Wang et al. (2023) successfully analyzed the
109 impact of land use on the water quality parameters of the Huai River in China combined
110 using the SOM method and Boruta algorithm (Wang et al., 2023). However, research
111 on the joint application of the SOM method and feature selection algorithms for
112 analyzing the driving factors in groundwater quality evolution remains scarce.

113 In China, groundwater accounts for one-third of the total available water resources
114 and is a major source of water supply. Especially in northern China, groundwater
115 provides about two-thirds of drinking water, half industrial water, and one-third of
116 irrigation water (MEP, 2011). However, in recent years, there has been a nationwide
117 degradation of groundwater quality due to the influences of human activities and
118 hydrogeochemical processes (Gu et al., 2013; Guo et al., 2014). The detrimental effects
119 of groundwater quality degradation are most pronounced in agricultural areas,
120 attributable to: (1) the challenge of controlling non-point source pollution from
121 agricultural activities, leading to increased vulnerability of groundwater to
122 contamination in these areas (Gu et al., 2013), and (2) the reliance on private wells as
123 the principal drinking water sources in most rural areas of China , coupled with a lack
124 of adequate purification systems, significantly heightens the health risks associated
125 with contaminated groundwater (Jia et al., 2018). Therefore, there is an urgent need to
126 conduct research on the patterns of groundwater quality evolution and its influencing
127 factors in agricultural areas of China.

128 The Sanjiang Plain, one of the most important grain-producing regions in China,
129 is a large-scale agricultural irrigation area where rice is the main crop. In this region,
130 groundwater serves as the primary source for both agricultural and residential drinking
131 water. In recent years, the area has been facing relatively serious groundwater pollution
132 issues (Deng et al., 2023). At present, research on groundwater in the Sanjiang Plain
133 predominantly focuses on water quantity (Ling et al., 2023; Liu et al., 2016; Sun et al.,
134 2022). There are only a few studies addressing groundwater quality, and these are
135 limited to basic analyses of hydrochemical indicators (Cao et al., 2014; Ye et al., 2022).
136 Systematic research on the evolution of groundwater quality and the driving factors in
137 this area is still lacking. Thus, the overall goal of this research is to identify the
138 hydrochemical evolution and governing mechanisms of groundwater in the Sanjiang
139 Plain. The corresponding research focuses include (1) analyzing the spatiotemporal
140 variations of regional groundwater chemistry based on water quality data in 2012 and

141 2022, (2) exploring the primary governing mechanisms for the hydrochemical
142 evolution of groundwater using the SOM method and Boruta algorithm. The findings
143 of this study contribute to a deeper understanding of the groundwater quality evolution
144 in agricultural areas and provide valuable information for groundwater pollution
145 protection and management.

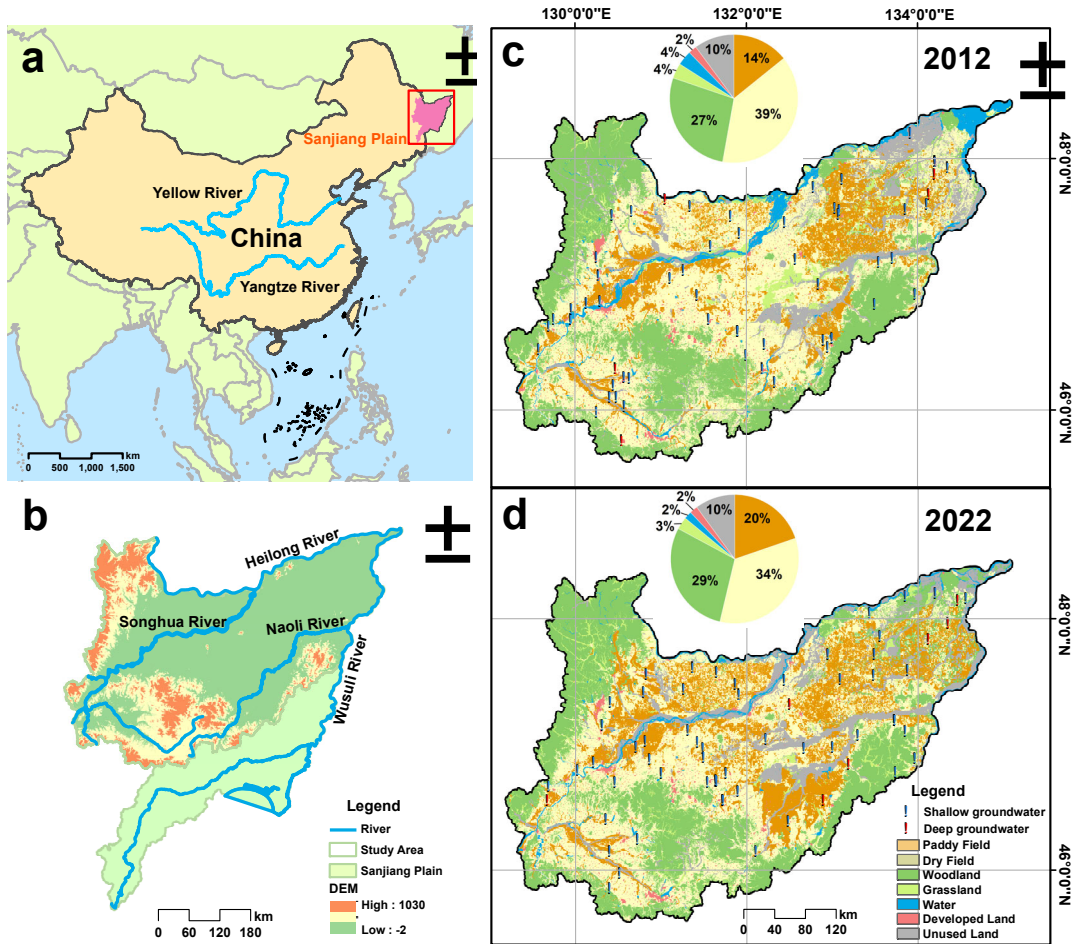
146 **2 Data and methods**

147 **2.1 Description of the study area**

148 The Sanjiang Plain is situated in northeastern China and is an alluvial lowland
149 formed by the diverging flows of the Songhua River, Heilong River, and Wusuli River.
150 It spans approximately 108,900 square kilometers (**Fig.1a**). In this paper, our focus is
151 on the basins of the Heilong, Songhua and Naoli Rivers (**Fig.1b**), which cover most of
152 the agricultural irrigation areas on the Sanjiang Plain. This area has a temperate humid
153 and semi-humid continental monsoon climate, characterized by warm and humid
154 summers, and cold and dry winters. The average annual precipitation in this area is
155 between 500 and 700 mm, the average annual temperature varies between 1.6°C and
156 3.9°C, and the annual sunshine ranges from 2400 to 2500 hours. The unique climatic
157 conditions is suitable for agricultural development, especially the cultivation of crops
158 such as rice, soybeans, wheat and corn (Jin et al., 2016).

159 The terrain of the study area is low and flat, with an average elevation of 50 to 60
160 meters. In the hilly regions of the southwest and northwest, the geological strata mainly
161 consist of granite, metamorphic rock, and volcanic rock. The groundwater in these areas
162 is bedrock fracture water. In the plains of the northeastern region, shallow groundwater
163 is primarily hosted within the porous aquifers of sand and gravel from the Quaternary
164 and Tertiary periods. Given the flat topography, the rate of groundwater flow is slow in
165 this region. The groundwater is mainly recharged by infiltration from atmospheric
166 precipitation and seepage from irrigation, while it is primarily discharged through
167 artificial pumping and evaporation. The flows from the southwest to the northeast
168 (Zhang et al., 2017). Agricultural land, including paddy fields and dry fields, is the

169 predominant type of land use in the region, accounting for more than half of the total
 170 area (Fig.1c, d). This is followed by woodland, which occupies over one quarter of the
 171 total area.



172
 173 **Fig.1.** Locations of the Sanjiang Plain (a) and the study area (b). Land use patterns
 174 and sampling sites in 2012 (c) and 2022 (d).

175 2.2 Data collection and processing

176 Sampling campaigns were conducted in November 2012 and July 2022. The
 177 samples were mainly gathered from existing irrigation wells near the cropland and
 178 household drinking wells. The samples are categorized into shallow ($\leq 20\text{m}$) and deep
 179 groundwater ($> 20\text{m}$) based on the depth of the groundwater. A total of 53 shallow
 180 groundwater and 6 deep groundwater samples were collected in 2012, and 56 shallow
 181 groundwater and 7 deep groundwater samples were collected in 2022 (Fig.1c, d).
 182 Before sampling, at least 5 L of water were drained from the well to ensure the sample's

183 freshness and representativeness. The physicochemical parameters such as pH, EC and
184 TDS were measured in situ using a portable pH/EC meter. The samples were collected
185 in pre-washed polyethylene (PE) bottles and carefully labeled. Subsequently, these
186 samples were brought back to the laboratory, stored in refrigeration at 4°C and tested
187 as soon as possible. The concentrations of cations (Na^+ , K^+ , Ca^{2+} , Mg^{2+}) were
188 determined using an atomic absorption spectrophotometer. Meanwhile, the
189 concentrations of anions (Cl^- , SO_4^{2-} , HCO_3^- , NO_3^-) were measured using ion
190 chromatography. Repeated measurements of water samples were taken to ensure that
191 the data were valid. In addition, the percentage error in ionic charge balance was less
192 than 5% for each sample, which further ensured the accuracy of the physicochemical
193 analysis.

194 **2.3 Entropy-weighted water quality index**

195 The EWQI emerged gradually after Li et al. (2010) integrated the concept of
196 entropy weighting into the traditional water quality index (WQI), utilizing entropy
197 theory to compute the weights for a method of water quality evaluation (Ji et al., 2020;
198 Kumar et al., 2021; Li et al., 2010; Wang et al., 2019). All the physical and chemical
199 parameters from water samples are consolidated into a single representative value that
200 can reflect water quality. This approach has been widely used by many scholars to
201 evaluate the overall quality of groundwater (Adimalla, 2021; Li et al., 2021). The
202 method consists of four main steps, first construct the eigenvalue matrix X , which can
203 be associated with all physicochemical parameters; then standardize the matrix X by
204 y_{ij} to form the standard matrix Y ; then calculate the information entropy e_j and
205 entropy weights w_j ; and finally calculate the EWQI values based on the quantitative
206 grading gradient q_j of the water chemistry indexes (Eq. (1)(2)) (Adimalla et al., 2020;
207 Krishna et al., 2023; Sheng et al., 2023).

$$208 \quad q_j = \frac{c_j}{s_j} \times 100\% \quad (1)$$

$$209 \quad EWQI = \sum_{j=1}^m (w_j \times q_j) \quad (2)$$

210 All physical and chemical parameters from water samples are consolidated into a

211 single representative value that mirrors water quality. This approach has been
212 extensively adopted by scholars for assessing groundwater's overall quality.

213 **2.4 Self-organizing maps**

214 In 1982, Kohonen proposed a method for reducing dimensionality and reflecting
215 data similarity by partitioning similar data items, the self-organizing mapping algorithm
216 (SOM) (Kohonen, 1998). SOM is a neural network model based on an unsupervised
217 learning algorithm that is capable of visualizing and interpreting linear and nonlinear
218 relationships in high-dimensional datasets, and is constructed from an input layer and
219 an output layer, and is connected to each neuron in the output layer by a weight vector
220 on a hexagonal grid (Liang et al., 2015; Olkowska et al., 2014; Zhang et al., 2023;
221 Zhong et al., 2022). The SOM neural network is mainly composed of four training steps:
222 initialization, competition, self-organization and iteration, which can be mainly
223 outlined as follows (Günter et al., 2002; Yin et al., 2021):

224 (1) Select the appropriate learning rate and the number of network nodes to
225 initialize the SOM neural network, where the learning rate determines the rate of
226 convergence, and the network nodes affect the fit between the trained network
227 distribution and the training data distribution.

228 (2) For different input patterns, each neuron in the network neuron calculates its
229 own discriminant function value by the following formula, in which the specific neuron
230 with the smallest discriminant function value is the winning neuron (BMU). The
231 discriminant function (squared euclidean distance) is calculated as follows (Eq. (3)):

$$232 d_j(x) = \sum_{i=1}^D (x_i - \omega_{ji})^2 \quad (3)$$

233 Where D is the input latitude; x_i is the input data; ω_{ji} is the weight vector between
234 neuron j and input data i .

235 (3) after obtaining the winning neuron (BMU) in the previous step, a topological
236 neighborhood is defined with the neuron as the center, and all the neurons in the
237 neighborhood will be updated to varying degrees according to their own weights in
238 order to adapt to different input vectors, in which the winning neuron The weights are

239 updated to the largest extent (Eq. (4)).

$$240 \quad T_{j,I(x)}(t) = \exp\left(-\frac{S_{j,I(x)}^2}{2\sigma(t)^2}\right) \quad (4)$$

241 Where $I(x)$ is the index of the BMU; $S_{j,I(x)}$ is the distance between neuron j and
242 the winning neuron; $\sigma(t)$ decreases linearly with time.

243 (4) Optimize the search process and converge to the best self-organizing mapping
244 (Eq. (5)).

$$245 \quad \Delta\omega_{ji} = \eta(t) \cdot T_{j,I(x)}(t) \cdot (x_i - \omega_{ji}) \quad (5)$$

246 Where: $\eta(t)$ is the learning rate; x_i is the input data.

247 The trained SOM model can be represented by two-dimensional planar graphs of
248 different variables, and the weight values of the variables in each neuron are visualized
249 with different color gradients, with high and low values represented in red and blue,
250 respectively, and comparative analyses of the color gradients of individual variables
251 can clearly identify the correlations between different variables. Therefore, in recent
252 years, SOM has been widely used to assess water quality and can effectively analyze
253 spatial and temporal variations in groundwater samples (Deng et al., 2023; Olkowska
254 et al., 2014; Zhang et al., 2023).

255 **2.5 Boruta algorithm**

256 The Boruta feature selection algorithm was introduced by Stoppiglia et al. (2003)
257 based on the development of the Random Forest algorithm, which is mainly used to
258 identify important input parameters from multiple related features to match the
259 attributes of independent features (Breiman, 2001; Stoppiglia et al., 2003). The Boruta
260 algorithm is a variable selection algorithm for detecting relevant explanatory variables
261 that identifies predictor variable results that are critical to explaining the modeled
262 phenomenon, and it finds relevant predictor variables by comparing the importance of
263 the original variable to the achievable importance estimated using randomly arranged
264 copies of the predictor variables (Laakso et al., 2018).

265 The importance of a feature variable is quantified by the loss of accuracy of the
266 classification model due to the features randomly assigned to the variable, and then the

267 mean and standard deviation of the computed loss of accuracy are both divided to obtain
 268 the *Z*-score of the measure of importance in the Boruta algorithm (Ahmadpour et al.,
 269 2021; Habibi et al., 2023). The steps of the Boruta algorithm consist of creating shadow
 270 features, obtaining importance, setting thresholds, feature filtering and iteration (Lawal
 271 et al., 2023). In this study, the importance of land use on ion concentrations in
 272 groundwater was quantitatively distinguished based on the Boruta package in R
 273 software.

274 3 Results and discussion

275 3.1 General hydrochemical characteristics

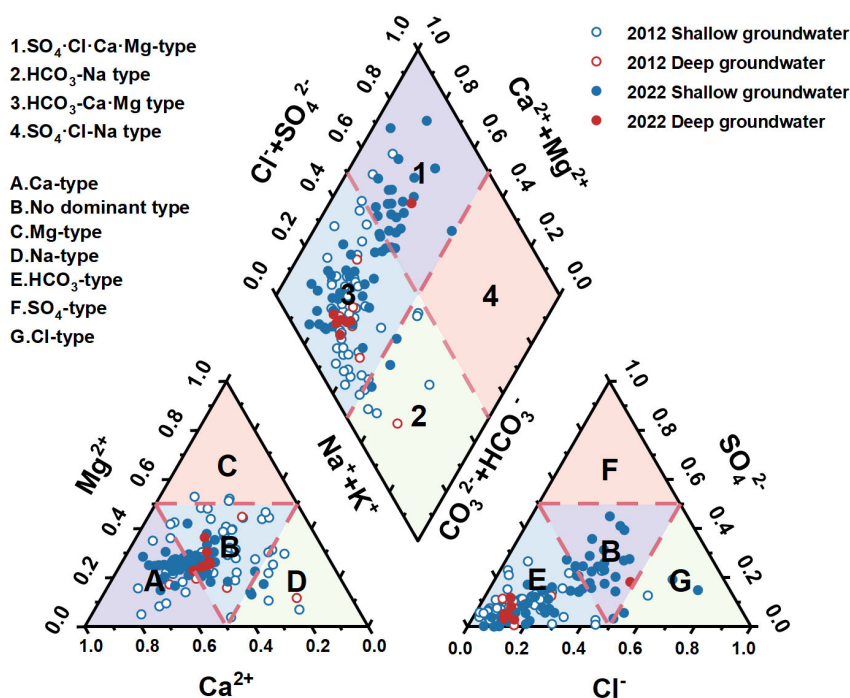
276 The statistical results of hydrochemical parameters groundwater samples are
 277 shown in Table 1. Compared to 2012, there has been little change in the pH of
 278 groundwater in 2022. The concentration of major ions in deep groundwater has
 279 decreased, whereas in shallow groundwater, there is a noticeable increase in ion
 280 concentrations, except for K^+ and HCO_3^- . The major cations and anions in both deep
 281 and shallow groundwater exhibited similar abundance patterns. The predominant cation
 282 and anion in groundwater samples was Ca^{2+} and HCO_3^- , respectively. The average
 283 concentrations of cations and anions were in the following order: $Ca^{2+} > Na^+ > Mg^{2+} > K^+$
 284 and $HCO_3^- > Cl^- > SO_4^{2-} > NO_3^-$, respectively.

285 **Table 1** Statistics on basic hydrochemical parameters in groundwater samples (concentration in
 286 mg/L, pH in unit, EC in $\mu\text{s/cm}$).

Parameters	2012						2022					
	Shallow Groundwater			Deep Groundwater			Shallow Groundwater			Deep Groundwater		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
pH	6.00	8.00	7.00	6.10	7.40	6.60	6.40	8.00	7.00	6.60	7.70	7.00
TDS	39.8	1266	264	107	660	231	72.1	744	248	113	140	125
EC	79.4	1849	463	214	1319	462	144	1487	497	225	279	251
TH	5.80	193	66.7	10.9	101	53.1	12.3	213	57.9	22.2	40.1	27.9

K ⁺	0.60	13.4	3.42	1.86	4.50	3.16	0.06	15.2	2.71	0.48	1.74	1.17
Na ⁺	3.65	65.1	22.2	12.9	31.7	22.7	4.50	103	24.1	12.3	18.8	14.2
Ca ²⁺	4.81	125	27.8	8.02	43.3	26.1	10.0	188	44.9	16.2	31.5	20.4
Mg ²⁺	0.97	33.1	12.2	2.92	19.4	8.84	2.35	42.3	13.0	5.70	12.1	7.55
Cl ⁻	2.13	95.0	16.5	9.93	28.4	15.1	9.94	146	39.8	7.10	35.5	13.7
SO ₄ ²⁻	0.61	41.7	15.1	0.77	23.5	11.9	0.15	163	32.5	3.46	18.4	8.63
HCO ₃ ⁻	45.0	532	156	89.3	214	146	32.2	389	143	42.9	164	115
NO ₃ ⁻	0.64	10.2	3.87	1.10	8.50	3.83	0.01	60.5	7.81	0.05	6.45	1.07

287 The Piper diagram provides a comprehensive view of the chemical compositions
288 of natural water (Egbueri et al., 2020), which is widely used to identify the
289 hydrogeochemical types and determine the origins of the constituents in groundwater
290 (Zhang et al., 2021). As shown in Fig. 2, the cation plot indicates that most groundwater
291 samples fall within the Zone A (Ca-type) and B (No dominant type), while the anion
292 plot reveals that almost all the samples are situated in Zone B (No dominant type) and
293 E (HCO₃-type). The general characteristics of groundwater chemistry are shown in the
294 central diamond. The deep groundwater chemistry remains relatively constant, with
295 almost all samples being the HCO₃-Ca-Mg type. In contrast, the hydrochemical
296 characteristics of shallow groundwater have undergone significant changes, shifting
297 from mostly (84.9%) the HCO₃-Ca-Mg type in 2012 to a mix of HCO₃-Ca-Mg (53.6%)
298 and SO₄-Cl-Ca-Mg (41.1%) types by 2022. This observed evolution trend in
299 hydrochemical characteristics might be related to changes in the factors affecting
300 groundwater quality. The HCO₃-Ca-Mg type represents natural groundwater quality
301 predominantly affected by water-rock interactions, such as the weathering of carbonate
302 minerals in sandstone, silicates, and clay minerals in alluvial layers (Edmunds et al.,
303 2008; Lee et al., 2019). The composition change from HCO₃-Ca-Mg type to SO₄-Cl-
304 Ca-Mg type suggests the influence of anthropogenic pollution due to chloride and
305 sulfate are typical inorganic pollutants originating from human activities, such as
306 domestic sewage, livestock breeding wastewater, and agricultural fertilizers (Cao et al.,



308

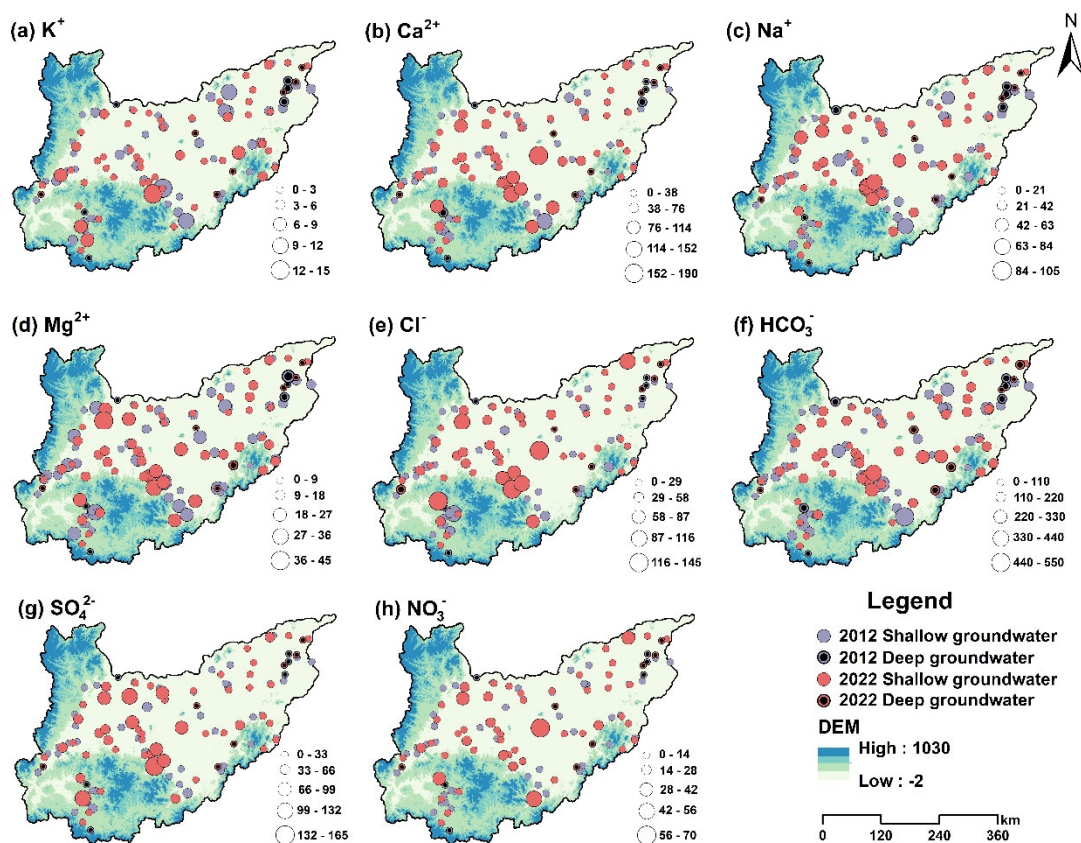
309 **Fig.2.** Piper diagram of groundwater samples in the Sanjiang Plain.

310 **3.2 Spatio-temporal variations of chemical components**

311 Through a decade of evolution, the concentration of cations in deep groundwater
 312 has generally decreased (Fig.3). While, for the shallow groundwater, the concentrations
 313 of cations showed an increasing trend, except for K^+ . The rise in cation levels might be
 314 due to more intense water-rock interactions which is related with the elevated acidic
 315 anions. In addition, in 2022, the spatial distribution of cation concentrations in shallow
 316 groundwater showed a different pattern compared to 2012, mainly characterized by a
 317 notable increase in concentration levels within the central region of the study area.

318 In general, HCO_3^- originated from natural sources, such as chemical weathering
 319 of minerals. In 2012, the mean HCO_3^- concentrations in shallow groundwater and deep
 320 groundwater were 156 mg/L and 146 mg/L, respectively. In 2022, these values were
 321 143 mg/L for shallow groundwater and 115 mg/L for deep groundwater (Table 1).
 322 HCO_3^- concentrations in both shallow and deep groundwater decreased throughout the
 323 study area, with no significant evolution in spatial or temporal distribution, and high
 324 concentrations were mainly found at the interface between the southeastern mountains

325 and the plains (Fig.3(f)). Cl^- , SO_4^{2-} and NO_3^- are commonly derived from human
 326 activities, and are often used as indicators of anthropogenic pollution (Cao et al., 2022;
 327 Huang et al., 2013; Xiajing et al., 2018). The acceptable limits for Cl^- and SO_4^{2-} are
 328 250mg/L (WHO, 2021). In this study, significant increases in Cl^- and SO_4^{2-}
 329 concentrations were observed in the shallow groundwater in 2022, particularly in the
 330 central and western parts of the study area (Fig.3e, g). However, these concentrations
 331 remained within permissible limits, as detailed in Table 1. NO_3^- in shallow groundwater
 332 also exhibited the same increasing trend, with an average concentration of 7.81mg/L in
 333 2022 (Table 1). Seven shallow groundwater samples had NO_3^- levels exceeding the
 334 groundwater quality standard (20 mg/L) (WHO, 2021). The high NO_3^- concentration
 335 values were scattered at multiple sites within the study area (Fig.3h).

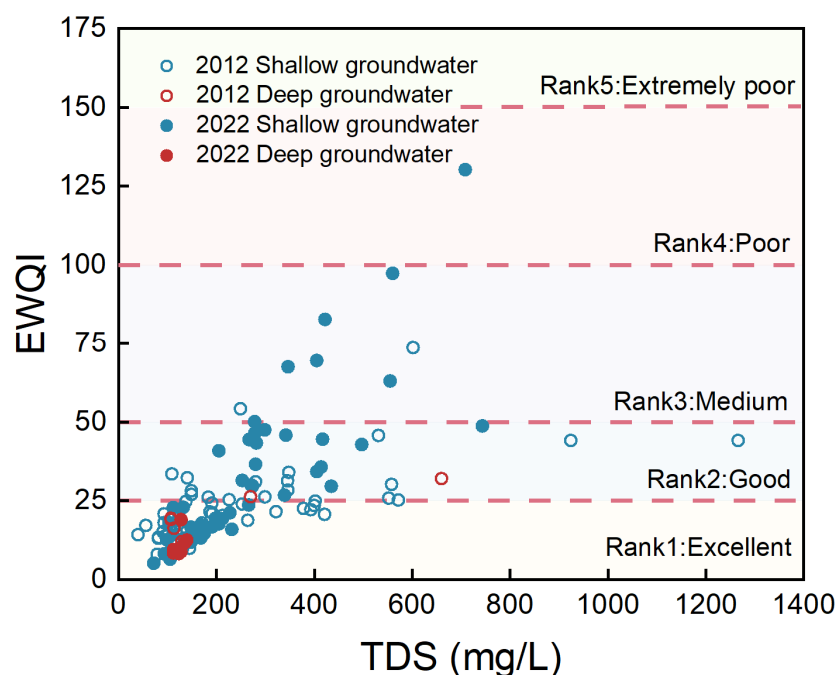


336
 337 **Fig.3.** Spatial and temporal distribution of major ion concentrations of groundwater
 338 samples from the Sanjiang Plain.

339 **3.3 Spatio-temporal variations of groundwater quality**

340 In this work, K^+ , Na^+ , Ca^{2+} , Mg^{2+} , Cl^- , SO_4^{2-} , HCO_3^- , NO_3^- and TDS were used to

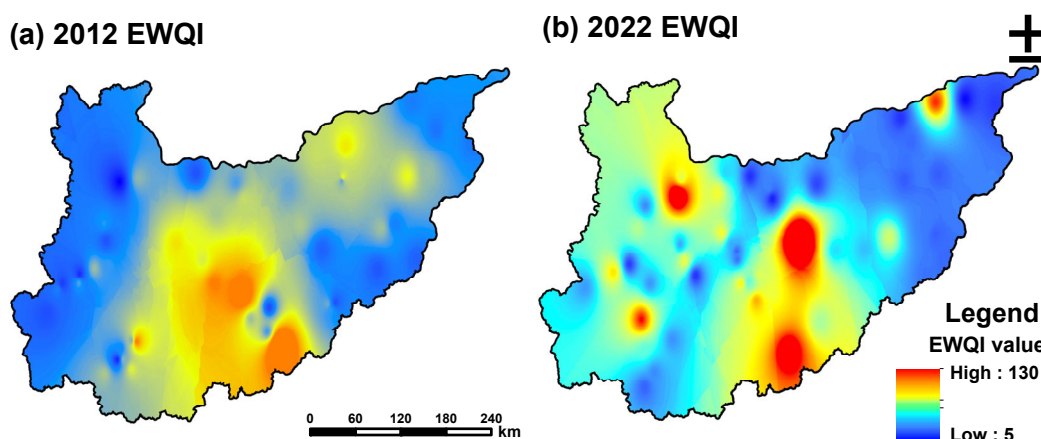
341 determine the EWQI values for groundwater samples in the Sanjiang Plain. Based on
 342 the EWQI value, the groundwater can be categorized as excellent (<25), good (25~50),
 343 medium (50~100), poor (100~150) and extremely poor (>150) (Wang et al., 2019). The
 344 calculation results show that the EWQI values for shallow groundwater and deep
 345 groundwater in 2012 ranged from 8.02 to 73.59 and 12.10 to 32.05, with mean values
 346 of 24.33 and 20.89, respectively; that the EWQI values for 2022 ranged from 5.18 to
 347 130.26 and 8.7 to 19.01, with mean values of 29.56 and 11.19. In 2012, among 53
 348 shallow groundwater samples, 62.26% were classified as excellent (Rank I), 33.96% as
 349 good (Rank II), and 3.78% as medium (Rank III). Similarly, of the 6 deep groundwater
 350 samples, 66.67% were rated excellent (Rank I) and 33.33% as good (Rank II). By 2022,
 351 out of the 56 shallow groundwater samples, 58.93% were excellent (Rank I), 28.57%
 352 were good (Rank II), 10.71% were medium (Rank III), and 1.79% were poor (Rank IV).
 353 All seven deep groundwater samples tested in 2022 showed excellent (Rank I) quality
 354 (Fig.4).



355
 356 **Fig.4.** Plot of the entropy weighted water quality index (EWQI) versus TDS.

357 The above results indicate that the quality of deep groundwater in the study area
 358 has improved, while the quality of shallow groundwater has declined. In general,
 359 groundwater classified as Rank I, II and III is suitable for drinking, while water

360 classified as Rank IV is of poor quality. In 2012, all shallow groundwater samples in
 361 the study area conformed to at least Rank III, satisfying the criteria for potable use.
 362 However, by 2022, the emergence of Rank IV water quality was recorded, accompanied
 363 by a significantly increased number of Rank III water samples. In response to the
 364 deterioration of shallow groundwater quality, spatial interpolation maps of EWQI
 365 values of shallow groundwater in 2012 and 2022 were plotted, and used to analyze the
 366 spatial and temporal evolutionary characteristics of the water quality. As shown in Fig.5,
 367 higher values of EWQI of groundwater in 2012 were distributed in the central and
 368 southern parts of the study area, whereas, after a decade of evolution, high values of
 369 EWQI of groundwater in 2022 were distributed in the central, southern, and
 370 northwestern parts of the area, as well as in the small area in the northeast. When the
 371 calculation of 2022 EWQI values for shallow groundwater was carried out, Cl^- , SO_4^{2-}
 372 and NO_3^- had higher w_j values (Eq.6) than the other parameters, with NO_3^- having the
 373 highest w_j value of 0.25 (Table S1), indicating that the elevated EWQI values for
 374 shallow groundwater were primarily driven by fluctuations in the concentrations of Cl^- ,
 375 SO_4^{2-} and NO_3^- . Meanwhile, the spatial and temporal distribution plots of Cl^- , SO_4^{2-} and
 376 NO_3^- concentrations (Fig.5 (e)(g)(h)) and the spatial and temporal evolution plots of the
 377 EWQI (Fig.5(b)) reflected striking similarities, which further illustrated that Cl^- , SO_4^{2-} ,
 378 and NO_3^- were the dominant factors contributing to the degradation of shallow
 379 groundwater quality.



380

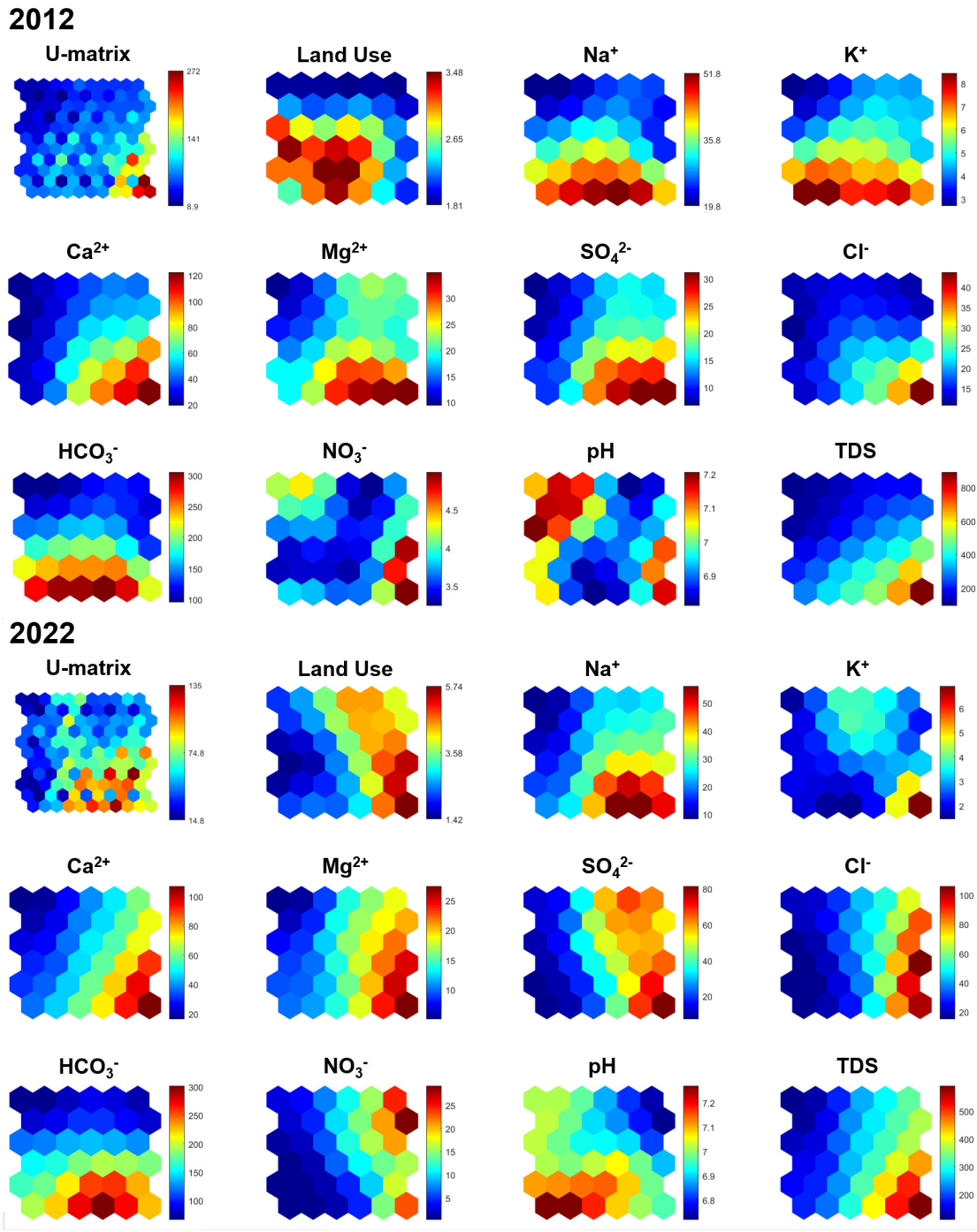
381 **Fig.5.** Spatial and temporal evolution of shallow groundwater quality (EWQI) in the

382

383 **3.4 SOM output results**

384 Since deep groundwater quality has improved, while shallow groundwater quality
385 has deteriorated, accompanied by notable changes in its hydrochemical characteristics,
386 our analysis focused on the evolution of shallow groundwater quality and its driving
387 factors. We used Na^+ , K^+ , Ca^{2+} , Mg^{2+} , SO_4^{2-} , Cl^- , HCO_3^- , NO_3^- , pH, and TDS in shallow
388 groundwater as input parameters, and trained the SOM on 53 samples from 2012 and
389 56 samples from 2022, respectively. The results of the SOM training are visualized
390 through the generation of U-matrices for the eleven parameters, along with their
391 component planar topologies. The high and low weight vector values of the neurons are
392 shown in red and blue, respectively (Fig.6). Based on comparing the color gradient of
393 each parameter, the correlation between them can be obtained (Lee et al., 2019).

394 As shown in Fig. 6, there are similar correlations between the hydrochemical
395 parameters of shallow groundwater in 2012 and 2022. Specially, Na^+ , K^+ , and HCO_3^-
396 exhibited a similar color gradient distribution pattern, and the remaining five ions
397 displayed analogous distribution patterns. In addition, in both 2012 and 2022, pH did
398 not show similar color distribution patterns with any ions, while the distribution of TDS
399 was similar to that of several ions, indicating that both are the results of combined
400 effects of multiple ions. It is noteworthy that in 2012, land use showed a color
401 distribution pattern similar to that of HO_3^- , while in 2022, it was similar to SO_4^{2-} , Cl^- ,
402 and NO_3^- . Since these three ions are major contributors to water quality deterioration,
403 this suggests that land use is an important factor affecting the quality of shallow
404 groundwater.



405

406

407

Fig.6. SOM-trained U-matrix and component planes for shallow groundwater.

408

3.5 Driving factors of shallow groundwater evolution

409

3.5.1 Water-rock interactions

410

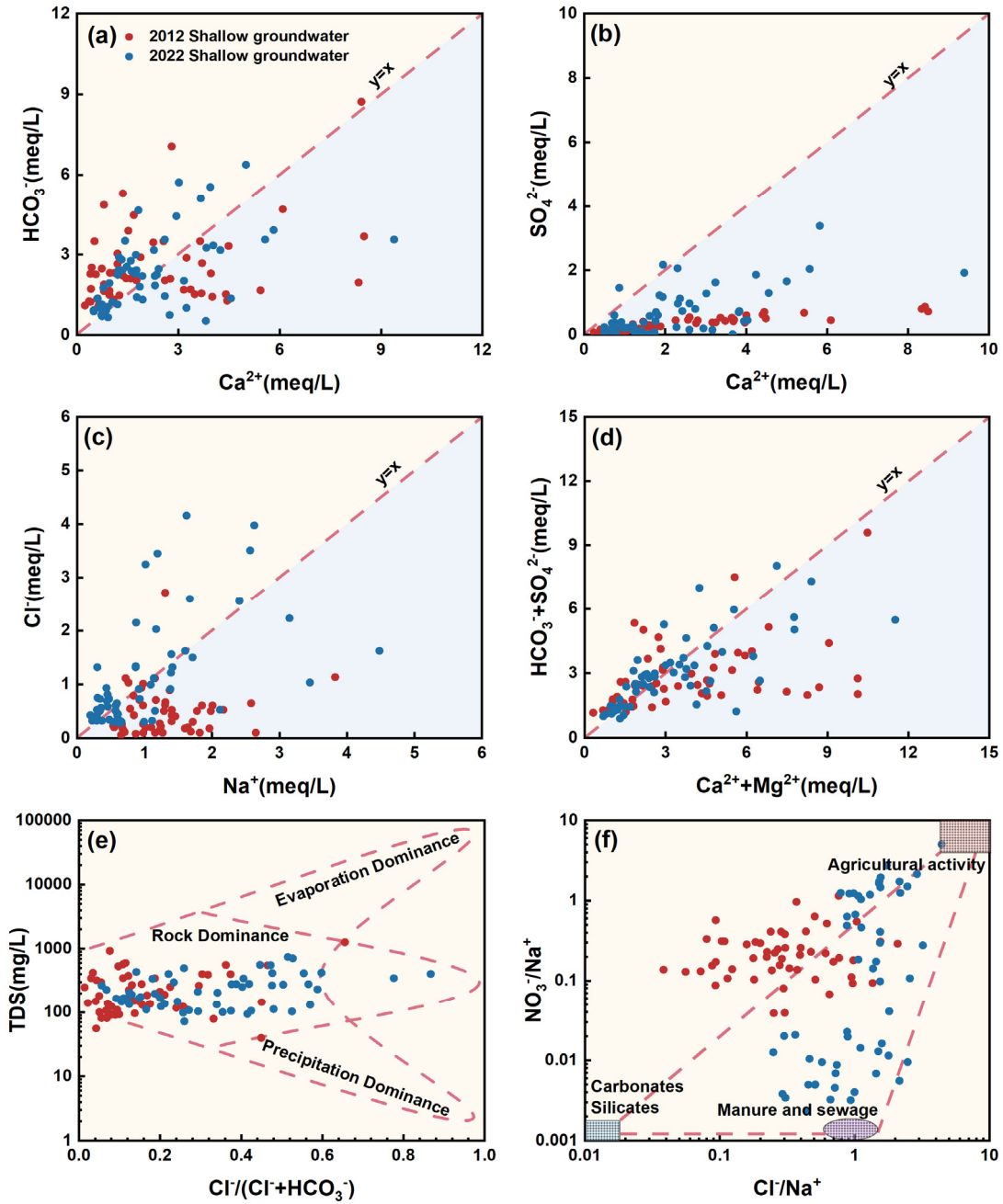
411

412

413

The ratio coefficients of major anions and cations in the water samples can reflect the main hydrogeochemical processes affecting the chemical characteristics. As shown in Fig.7a, the shallow groundwater samples are distributed on both sides of $y=x$, indicating that the shallow groundwater is affected by the weathering of carbonate and

414 silicate minerals together (Liu et al., 2021). The milliequivalent ratio of Ca^{2+} to SO_4^{2-}
415 not equaling one suggests that gypsum dissolution has a minor influence on water
416 chemistry (Fig. 7b). In 2012, most shallow groundwater had higher Na^+ levels than Cl^-
417 (Fig. 7c). By 2022, a marked increase in Cl^- concentrations was observed, attributed to
418 heightened anthropogenic emissions, resulting in levels even surpassing those of Na^+ .
419 As shown in Fig. 7d, in both 2012 and 2022, many sample points are distributed near
420 the $y=x$ line, suggesting possible dissolution of minerals such as calcite, dolomite, and
421 gypsum within the aquifer system. In addition, some sample points deviate from the
422 $y=x$ line, indicating the participation of other anions and cations in the reaction process.
423 Gibbs plots are instrumental in showcasing the ionic composition and trends in natural
424 water, thereby being widely used to identify the mechanisms of water chemistry
425 formation (Gibbs, 1970; Li et al., 2021; Liu et al., 2023). In both 2012 and 2022,
426 shallow groundwater samples were mainly located within the rock-dominant zone (Fig.
427 7e), indicating that the anions and cations are mainly affected by the weathering of
428 rocks, and the water-rock interaction is the dominant factor determining the
429 concentration of water chemistry. The correlation diagram between NO_3^- and Cl^- is
430 often used to discern the origins of nitrate in water bodies (Ke et al., 2021; Wu et al.,
431 2023). The sources of nitrate in shallow groundwater exhibited significant changes
432 between 2012 and 2022. Figure 7f illustrates that in 2012, NO_3^- was primarily derived
433 from agricultural activities and rock weathering, while in 2022, in addition to these two
434 sources, the levels of NO_3^- were also significantly affected by the discharge of manure
435 and sewage.



436

437 **Fig.7.** Correlation diagrams of (a) HCO_3^- vs. Ca^{2+} , (b) SO_4^{2-} vs. Ca^{2+} , (c) Cl^- vs. Na^+ ,

438 (d) $\text{HCO}_3^- + \text{SO}_4$ vs. $\text{Ca}^{2+} + \text{Mg}^{2+}$, (e) TDS vs. $\text{Cl}^-/(\text{Cl}^- + \text{HCO}_3^-)$, (f) $\text{NO}_3^-/\text{Na}^+$ vs. Cl^-

439

$/\text{Na}^+$ in shallow groundwater.

440

3.5.2 Anthropogenic pollution

441

As discussed in §3.3, the elevation in concentrations of SO_4^{2-} , Cl^- , and NO_3^- is the primary reason for the degradation of shallow groundwater quality in the Sanjiang Plain.

442

443

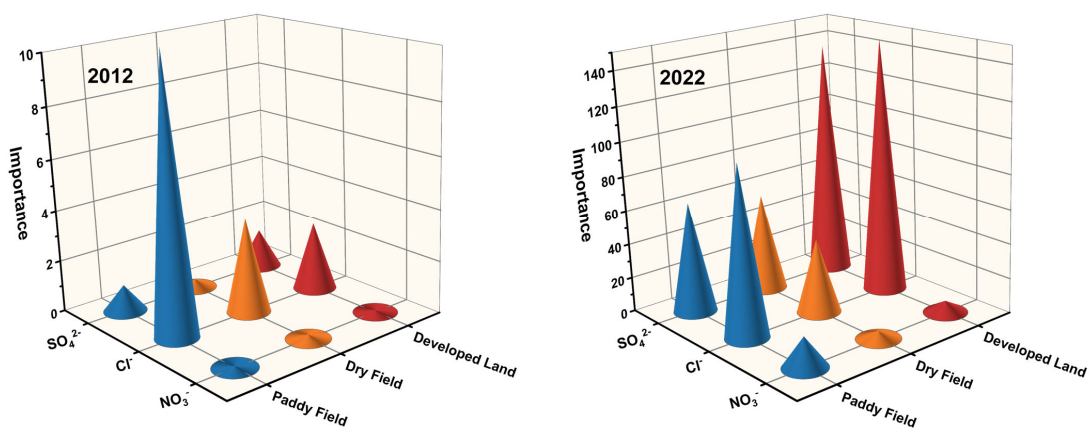
Additionally, the SOM analysis findings indicate a considerable influence of land use

444

on these ions' levels in shallow groundwater. Therefore, we used the land use type to

445 which the locations of the shallow groundwater sample sites in 2012 and 2022 belong
446 as a starting point to further quantify the impacts. We quantitatively discriminated the
447 degree of importance of paddy field, dry field, and developed land for SO_4^{2-} , Cl^- , and
448 NO_3^- concentrations in 2012 and 2022 based on the Boruta algorithm. As shown in
449 [Fig.8](#), the influence of land use on these three ions in 2022 was markedly different from
450 that in 2012. In 2012, the influence of developed land on the SO_4^{2-} in shallow
451 groundwater was the greatest, which is consistent with the fact that domestic sewage
452 typically contains high amounts of SO_4^{2-} ([Zhang et al., 2020](#)). Cl^- was predominantly
453 affected by paddy fields and dry land, which might be associated with the widespread
454 use of chloride-containing fertilizers (e.g., KCl) in this region ([Wang et al., 1994](#)). The
455 three land use types had similar importance for the distribution of nitrate in shallow
456 groundwater. In 2022, the significance of the three land use types for SO_4^{2-} remains
457 similar to the rankings seen in 2012. However, for Cl^- , the impact of developed land
458 exceeds that of paddy fields and dry land, becoming the most critical influencing factor.
459 In general, domestic sewage and animal manure contain high levels of Cl^- , positioning
460 them as one of the primary sources of Cl^- in water bodies ([Khazaei et al., 2017](#)). Data
461 from Heilongjiang Province Statistical Bulletin reveals a decrease in the population of
462 the study area from 2.58 million in 2012 to 2.20 million in 2022, indicating that
463 domestic sewage discharge might not be the primary factor behind the elevated impacts
464 of developed lands on Cl^- . Over the past decade, there was a significant rise in the
465 number of livestock and poultry, increasing from 35.7 million in 2012 to 113.5 million
466 by 2022. Therefore, the dramatic increase in livestock and poultry production, and the
467 subsequent boost in wastewater and manure emissions, was likely the foremost
468 contributor to the increased impact of developed land on the Cl^- content in shallow
469 groundwater. As shown in [Fig.1](#), during the span of a decade, there was a marked
470 increase in the extent of paddy fields in the study area, alongside a reduction in drylands.
471 Specially, the area of paddy fields expanded from 23999 km^2 in 2012 to 33453 km^2 in
472 2022. Previous studies have shown that paddy fields, relative to an equivalent spatial

473 extent of drylands, are associated with higher nitrogen emissions (Liu et al., 2015; Tong
 474 et al., 2022). Thus, relative to an equivalent spatial extent of drylands, are associated
 475 with high nitrogen emissions. Therefore, the increase in paddy field area is the primary
 476 reason for the heightened impact of paddy fields on the nitrate content in shallow
 477 groundwater. Additionally, the rise in livestock and poultry production, which generates
 478 manure and wastewater rich in nitrates (Ke et al., 2021), might be the reason for the
 479 increased impact of developed land on nitrate levels. These findings are consistent with
 480 the conclusions drawn from Figure 7f.



481
 482 **Fig.8.** Importance of land use for water quality parameters.

483 3.6 Recommendations for sustainable groundwater development

484 Our findings provide essential insights for managers in the Sanjiang Plain on
 485 developing strategies for sustainable groundwater management. Given the
 486 improvement in deep groundwater quality and the significant deterioration in shallow
 487 groundwater quality in the Sanjiang Plain over the past decade, prioritizing the control
 488 of shallow groundwater pollution is essential for sustainable development. The primary
 489 causes of the decline in shallow groundwater quality in this region are the increased
 490 emissions of wastewater and manure from rising livestock and poultry farming, along
 491 with greater pollutant seepage due to the expansion of paddy fields. Therefore, we make
 492 the following suggestions: (1) promote the expansion of large-scale livestock and
 493 poultry farming, along with the establishment of associated manure and wastewater
 494 treatment facilities (Qian et al., 2018); (2) regulate the use of chemical and organic
 495 fertilizers and promote water-saving irrigation practices to mitigate pollutant leaching

496 (Yin et al., 2021); (3) enhance groundwater quality monitoring in this region, promptly
497 track changes in groundwater conditions, and implement measures to prevent
498 contamination. Moreover, it should be noted that there is an increasing trend of nitrates
499 in the shallow groundwater of this area, which poses a serious threat to the health of
500 residents, as many rural families rely on it for drinking water. Therefore, private wells
501 in residential areas are recommended to be closed, and the water supply be switched to
502 a centralized deep groundwater source.

503 **4 Conclusions**

504 This study investigates the spatiotemporal of groundwater hydrochemistry in the
505 Sanjiang Plain using data from 2012 and 2022. Through hydrochemical analysis, the
506 Entropy Weighted Water Quality Index, Self-Organizing Maps, and the Boruta
507 algorithm, the primary driving factors influencing the chemical evolution of
508 groundwater were identified. Our findings indicate that the concentrations of major ions
509 in deep groundwater showed a declining trend, with the chemical properties remaining
510 relatively stable, and nearly all samples were of the $\text{HCO}_3\text{-Ca-Mg}$ type. In contrast, in
511 shallow groundwater, the concentrations of all ions, except for K^+ and HCO_3^- , have
512 increased. The hydrochemical types of the samples have shifted from predominantly
513 $\text{HCO}_3\text{-Ca-Mg}$ (84.9%) type to a mixed type, with 53.6% being $\text{HCO}_3\text{-Ca-Mg}$ and 41.1%
514 being $\text{SO}_4\text{-Cl-Ca-Mg}$. The EWQI results indicate that, compared to 2012, the water
515 quality of deep groundwater improved in 2022, while the quality of shallow
516 groundwater deteriorated. When analyzed together with the spatial distribution maps of
517 ion concentrations in shallow groundwater, it is evident that the increased levels of Cl^- ,
518 SO_4^{2-} , and NO_3^- are the main reasons for the decline in water quality. Notably, the nitrate
519 concentration in shallow groundwater showed a significant increase, with the average
520 concentration rising from 3.87 mg/L to 7.81 mg/L, and 12.5% of the sampling points
521 exceeding WHO drinking water standards. The results of the SOM and Boruta
522 algorithm indicate that the increase in wastewater and manure discharge from livestock
523 and poultry farming, as well as the intensification of pollutant leaching due to the

524 expansion of paddy fields, are the primary factors contributing to the decline in
525 groundwater quality. Therefore, it is imperative to manage pollutant discharge from
526 livestock and poultry farming and the application of fertilizers in agricultural fields to
527 ensure the sustainable use of groundwater resources in the Sanjiang Plain. Additionally,
528 this case study demonstrates that Self-Organizing Maps and the Boruta algorithm can
529 serve as valuable tools for exploring the driving mechanisms behind groundwater
530 evolution.

531 **Acknowledgement**

532 This work was financially supported by the the Strategic Priority Research
533 Program of the Chinese Academy of Sciences, China (XDA28100105 and
534 XDA28020502).

535

536 **References**

537 Adimalla, N., 2021. Application of the Entropy Weighted Water Quality Index
538 (EWQI) and the Pollution Index of Groundwater (PIG) to Assess Groundwater Quality
539 for Drinking Purposes: A Case Study in a Rural Area of Telangana State, India. Arch
540 Environ Contam Toxicol 80(1), 31-40. <https://doi.org/10.1007/s00244-020-00800-4>.

541 Adimalla, N., Qian, H. and Li, P., 2020. Entropy water quality index and
542 probabilistic health risk assessment from geochemistry of groundwaters in hard rock
543 terrain of Nanganur County, South India. Geochemistry 80(4).
544 <https://doi.org/10.1016/j.chemer.2019.125544>.

545 Ahmadpour, H., Bazrafshan, O., Rafiei-Sardooi, E., Zamani, H. and Panagopoulos,
546 T., 2021. Gully Erosion Susceptibility Assessment in the Kondoran Watershed Using
547 Machine Learning Algorithms and the Boruta Feature Selection. Sustainability 13(18).
548 <https://doi.org/10.3390/su131810110>.

549 Ascott, M.J., Goody, D.C., Wang, L., Stuart, M.E., Lewis, M.A., Ward, R.S. and
550 Binley, A.M., 2017. Global patterns of nitrate storage in the vadose zone. Nature

551 Communications 8(1). <https://doi.org/10.1038/s41467-017-01321-w>.

552 Breiman, L., 2001. Random forests. *Machine Learning* 45(1), 5-32.

553 <https://doi.org/10.1023/a:1010933404324>.

554 Cao, M., Hu, A., Gad, M., Adyari, B., Qin, D., Zhang, L., Sun, Q. and Yu, C.-P.,

555 2022. Domestic wastewater causes nitrate pollution in an agricultural watershed, China.

556 *Science of The Total Environment* 823. <https://doi.org/10.1016/j.scitotenv.2022.153680>.

557 Cao, Y., Tang, C., Song, X., Liu, C. and Zhang, Y., 2014. Identifying the

558 hydrochemical characteristics of rivers and groundwater by multivariate statistical

559 analysis in the Sanjiang Plain, China. *Applied Water Science* 6(2), 169-178.

560 <https://doi.org/10.1007/s13201-014-0215-5>.

561 Deng, Y., Ye, X. and Du, X., 2023. Predictive modeling and analysis of key drivers

562 of groundwater nitrate pollution based on machine learning. *Journal of Hydrology* 624.

563 <https://doi.org/10.1016/j.jhydrol.2023.129934>.

564 Edmunds, W.M., Shand, P., 2008. *Natural Groundwater Quality*. John Wiley &

565 Sons, New Jersey. <https://doi.org/10.1002/9781444300345>.

566 Egbueri, J.C., Ezugwu, C.K., Ameh, P.D., Unigwe, C.O. and Ayejoto, D.A., 2020.

567 Appraising drinking water quality in Ikem rural area (Nigeria) based on chemometrics

568 and multiple indexical methods. *Environmental Monitoring and Assessment* 192(5).

569 <https://doi.org/10.1007/s10661-020-08277-3>.

570 Elango, L. and Kannan, R., 2007. *Developments in Environmental Science*. Sarkar,

571 D., Datta, R. and Hannigan, R. (eds), pp. 229-243, Elsevier.

572 <https://doi.org/10.1016/b978-0-08-098319-6.09996-9>.

573 Gibbs, R.J., 1970. Mechanisms Controlling World Water Chemistry. *Science*

574 170(3962), 1088-&. <https://doi.org/10.1126/science.170.3962.1088>.

575 Graham, J.P. and Polizzotto, M.L., 2013. Pit Latrines and Their Impacts on

576 Groundwater Quality: A Systematic Review. *Environmental Health Perspectives* 121(5),

577 521-530. <https://doi.org/10.1289/ehp.1206028>.

578 Gu, B., Ge, Y., Chang, S.X., Luo, W. and Chang, J., 2013. Nitrate in groundwater

579 of China: Sources and driving forces. *Global Environmental Change* 23(5), 1112-1121.

580 <https://doi.org/10.1016/j.gloenvcha.2013.05.004>.

581 Günter, S. and Bunke, H., 2002. Self-organizing map for clustering in the graph
582 domain. Pattern Recognition Letters 23(4), 405-417. [https://doi.org/10.1016/s0167-](https://doi.org/10.1016/s0167-8655(01)00173-8)
583 [8655\(01\)00173-8](https://doi.org/10.1016/s0167-8655(01)00173-8).

584 Guo, H., Wen, D., Liu, Z., Jia, Y. and Guo, Q., 2014. A review of high arsenic
585 groundwater in Mainland and Taiwan, China: Distribution, characteristics and
586 geochemical processes. Applied Geochemistry 41, 196-217.
587 <https://doi.org/10.1016/j.apgeochem.2013.12.016>.

588 Habibi, A., Delavar, M.R., Sadeghian, M.S., Nazari, B. and Pirasteh, S., 2023. A
589 hybrid of ensemble machine learning models with RFE and Boruta wrapper-based
590 algorithms for flash flood susceptibility assessment. International Journal of Applied
591 Earth Observation and Geoinformation 122. <https://doi.org/10.1016/j.jag.2023.103401>.

592 Huang, G., Sun, J., Zhang, Y., Chen, Z. and Liu, F., 2013. Impact of anthropogenic
593 and natural processes on the evolution of groundwater chemistry in a rapidly urbanized
594 coastal area, South China. Science of The Total Environment 463-464, 209-221.
595 <https://doi.org/10.1016/j.scitotenv.2013.05.078>.

596 Jasechko, S. and Perrone, D., 2021. Global groundwater wells at risk of running
597 dry. Science 372(6540), 418-+. <https://doi.org/10.1126/science.abc2755>.

598 Ji, Y., Wu, J., Wang, Y., Elumalai, V. and Subramani, T., 2020. Seasonal Variation
599 of Drinking Water Quality and Human Health Risk Assessment in Hancheng City of
600 Guanzhong Plain, China. Exposure And Health 12(3), 469-485.
601 <https://doi.org/10.1007/s12403-020-00357-6>.

602 Jia, Y., Xi, B., Jiang, Y., Guo, H., Yang, Y., Lian, X. and Han, S., 2018. Distribution,
603 formation and human-induced evolution of geogenic contaminated groundwater in
604 China: A review. Science of The Total Environment 643, 967-993.
605 <https://doi.org/10.1016/j.scitotenv.2018.06.201>.

606 Jin, X., Du, J., Liu, H., Wang, Z. and Song, K., 2016. Remote estimation of soil
607 organic matter content in the Sanjiang Plain, Northeast China: The optimal band
608 algorithm versus the GRA-ANN model. Agricultural and Forest Meteorology 218-219,

609 250-260. <https://doi.org/10.1016/j.agrformet.2015.12.062>.

610 Ke, S., Chen, J., Zheng, X. and Sun, X., 2021. Reference ion method: A simple
611 and fast method for quantitatively identifying the source of nitrate and denitrification
612 rate in groundwater. *Science of The Total Environment* 769.
613 <https://doi.org/10.1016/j.scitotenv.2020.144555>.

614 Khazaei, E. and Milne-Home, W., 2017. Applicability of geochemical techniques
615 and artificial sweeteners in discriminating the anthropogenic sources of chloride in
616 shallow groundwater north of Toronto, Canada. *Environmental Monitoring and
617 Assessment* 189(5). <https://doi.org/10.1007/s10661-017-5927-1>.

618 Kim, K.-H., Yun, S.-T., Yu, S., Choi, B.-Y., Kim, M.-J. and Lee, K.-J., 2020.
619 Geochemical pattern recognitions of deep thermal groundwater in South Korea using
620 self-organizing map: Identified pathways of geochemical reaction and mixing. *Journal
621 of Hydrology* 589. <https://doi.org/10.1016/j.jhydrol.2020.125202>.

622 Kohonen, T., 1998. The self-organizing map. *Neurocomputing* 21(1), 1-6.
623 [https://doi.org/https://doi.org/10.1016/S0925-2312\(98\)00030-7](https://doi.org/https://doi.org/10.1016/S0925-2312(98)00030-7).

624 Kohonen, T., 2001. Self-organizing maps of massive databases. *Engineering
625 Intelligent Systems for Electrical Engineering and Communications* 9(4), 179-185.

626 Konikow, L.F. and Kendy, E., 2005. Groundwater depletion: A global problem.
627 *Hydrogeology Journal* 13(1), 317-320. <https://doi.org/10.1007/s10040-004-0411-8>.

628 Krishna, B. and Achari, V.S., 2023. Groundwater chemistry and entropy weighted
629 water quality index of tsunami affected and ecologically sensitive coastal region of
630 India. *Heliyon* 9(10). <https://doi.org/10.1016/j.heliyon.2023.e20431>.

631 Kumar, P.J.S. and Augustine, C.M., 2021. Entropy-weighted water quality index
632 (EWQI) modeling of groundwater quality and spatial mapping in Uppar Odai Sub-
633 Basin, South India. *Modeling Earth Systems and Environment* 8(1), 911-924.
634 <https://doi.org/10.1007/s40808-021-01132-5>.

635 Laakso, T., Kokkonen, T., Mellin, I. and Vahala, R., 2018. Sewer Condition
636 Prediction and Analysis of Explanatory Factors. *Water* 10(9).

637 <https://doi.org/10.3390/w10091239>.

638 Lapworth, D.J., Boving, T.B., Kremer, D.K., Kebede, S. and Smedley, P.L., 2022.
639 Groundwater quality: Global threats, opportunities and realising the potential of
640 groundwater. *Science of The Total Environment* 811.
641 <https://doi.org/10.1016/j.scitotenv.2021.152471>.

642 Lawal, I.M., Bertram, D., White, C.J., Kutty, S.R.M., Hassan, I. and Jagaba, A.H.,
643 2023. Application of Boruta algorithms as a robust methodology for performance
644 evaluation of CMIP6 general circulation models for hydro-climatic studies. *Theoretical
645 and Applied Climatology* 153(1-2), 113-135. [https://doi.org/10.1007/s00704-023-
646 04466-5](https://doi.org/10.1007/s00704-023-04466-5).

647 Lee, K.-J., Yun, S.-T., Yu, S., Kim, K.-H., Lee, J.-H. and Lee, S.-H., 2019. The
648 combined use of self-organizing map technique and fuzzy c-means clustering to
649 evaluate urban groundwater quality in Seoul metropolitan city, South Korea. *Journal of
650 Hydrology* 569, 685-697. <https://doi.org/10.1016/j.jhydrol.2018.12.031>.

651 Li, P.-Y., Qian, H. and Wu, J.-H., 2010. Groundwater Quality Assessment Based
652 on Improved Water Quality Index in Pengyang County, Ningxia, Northwest China. *E-
653 Journal of Chemistry* 7, S209-S216.

654 Li, X., Huang, X. and Zhang, Y., 2021. Spatio-temporal analysis of groundwater
655 chemistry, quality and potential human health risks in the Pinggu basin of North China
656 Plain: Evidence from high-resolution monitoring dataset of 2015-2017. *Science Of The
657 Total Environment* 800, 149568. <https://doi.org/10.1016/j.scitotenv.2021.149568>.

658 Liang, W., Wang, Q., Wang, D., Wen, X., Yu, G., He, N. and Wang, R., 2015.
659 Differences in SOM Decomposition and Temperature Sensitivity among Soil Aggregate
660 Size Classes in a Temperate Grasslands. *Plos One* 10(2).
661 <https://doi.org/10.1371/journal.pone.0117033>.

662 Ling, Z., Shu, L., Wang, D., Lu, C. and Liu, B., 2023. Assessment and projection
663 of the groundwater drought vulnerability under different climate scenarios and land use
664 changes in the Sanjiang Plain, China. *Journal of Hydrology: Regional Studies* 49.

665 <https://doi.org/10.1016/j.ejrh.2023.101498>.

666 Liu, J., Peng, Y., Li, C., Gao, Z. and Chen, S., 2021. Characterization of the
667 hydrochemistry of water resources of the Weibei Plain, Northern China, as well as an
668 assessment of the risk of high groundwater nitrate levels to human health.
669 *Environmental Pollution* 268(Pt B), 115947.
670 <https://doi.org/10.1016/j.envpol.2020.115947>.

671 Liu, X., Li, D., Zhang, H., Cai, S., Li, X. and Ao, T., 2015. Research on Nonpoint
672 Source Pollution Assessment Method in Data Sparse Regions: A Case Study of Xichong
673 River Basin, China. *Advances in Meteorology* 2015, 1-10.
674 <https://doi.org/10.1155/2015/519671>.

675 Liu, Y., Jiang, X., Zhang, G., Xu, Y., Wang, X. and Qi, P., 2016. Assessment of
676 Shallow Groundwater Recharge from Extreme Rainfalls in the Sanjiang Plain,
677 Northeast China. *Water* 8(10). <https://doi.org/10.3390/w8100440>.

678 Liu, Z., Feng, S., Zhangsong, A., Han, Y. and Cao, R., 2023. Long-term evolution
679 of groundwater hydrochemistry and its influencing factors based on self-organizing
680 map (SOM). *Ecological Indicators* 154. <https://doi.org/10.1016/j.ecolind.2023.110697>.

681 MacDonald, A.M., Bonsor, H.C., Ahmed, K.M., Burgess, W.G., Basharat, M.,
682 Calow, R.C., Dixit, A., Foster, S.S.D., Gopal, K., Lapworth, D.J., Lark, R.M., Moench,
683 M., Mukherjee, A., Rao, M.S., Shamsudduha, M., Smith, L., Taylor, R.G., Tucker, J.,
684 van Steenbergen, F. and Yadav, S.K., 2016. Groundwater quality and depletion in the
685 Indo-Gangetic Basin mapped from in situ observations. *Nature Geoscience* 9(10), 762-
686 766. <https://doi.org/10.1038/ngeo2791>.

687 Ministry of Environmental Protection (MEP), 2011. National Plan for
688 Groundwater Pollution Prevention and Control (2011–2020), Beijing. (in Chinese).

689 Nguyen, T.T., Kawamura, A., Tong, T.N., Nakagawa, N., Amaguchi, H. and
690 Gilbuena, R., 2015. Clustering spatio–seasonal hydrogeochemical data using self-
691 organizing maps for groundwater quality assessment in the Red River Delta, Vietnam.
692 *Journal of Hydrology* 522, 661-673. <https://doi.org/10.1016/j.jhydrol.2015.01.023>.

693 Olkowska, E., Kudłak, B., Tsakovski, S., Ruman, M., Simeonov, V. and Polkowska,
694 Z., 2014. Assessment of the water quality of Kłodnica River catchment using self-
695 organizing maps. *Science of The Total Environment* 476-477, 477-484.
696 <https://doi.org/10.1016/j.scitotenv.2014.01.044>.

697 Pulido-Bosch, A., Rigol-Sanchez, J.P., Vallejos, A., Andreu, J.M., Ceron, J.C.,
698 Molina-Sanchez, L. and Sola, F., 2018. Impacts of agricultural irrigation on
699 groundwater salinity. *Environmental Earth Sciences* 77(5).
700 <https://doi.org/10.1007/s12665-018-7386-6>.

701 Qian, Y., Song, K., Hu, T. and Ying, T., 2018. Environmental status of livestock
702 and poultry sectors in China under current transformation stage. *Science of The Total*
703 *Environment* 622, 702-709. <https://doi.org/10.1016/j.scitotenv.2017.12.045>.

704 Rodell, M., Famiglietti, J.S., Wiese, D.N., Reager, J.T., Beaudoin, H.K., Landerer,
705 F.W. and Lo, M.H., 2018. Emerging trends in global freshwater availability. *Nature*
706 557(7707), 651-659. <https://doi.org/10.1038/s41586-018-0123-1>.

707 Sheng, D., Meng, X., Wen, X., Wu, J., Yu, H., Wu, M. and Zhou, T., 2023.
708 Hydrochemical characteristics, quality and health risk assessment of nitrate enriched
709 coastal groundwater in northern China. *Journal of Cleaner Production* 403.
710 <https://doi.org/10.1016/j.jclepro.2023.136872>.

711 Stoppiglia H, Dreyfus G, Dubois R, Oussar Y., 2003. Ranking a random feature
712 for variable and feature selection. *J Mach Lear Res*3:1399–1414.
713 <https://dl.acm.org/doi/10.5555/944919.944980>.

714 Sun, Q., Xu, C., Gao, X., Lu, C., Cao, B., Guo, H., Yan, L., Wu, C. and He, X.,
715 2022. Response of groundwater to different water resource allocation patterns in the
716 Sanjiang Plain, Northeast China. *Journal of Hydrology: Regional Studies* 42.
717 <https://doi.org/10.1016/j.ejrh.2022.101156>.

718 Tong, X., Zhou, Y., Liu, J., Qiu, P. and Shao, Y., 2022. Non-point source pollution
719 loads estimation in Three Gorges Reservoir Area based on improved observation
720 experiment and export coefficient model. *Water Science and Technology* 85(1), 27-38.

721 <https://doi.org/10.2166/wst.2021.508>.

722 Wang, D., Wu, J., Wang, Y. and Ji, Y., 2019. Finding High-Quality Groundwater
723 Resources to Reduce the Hydatidosis Incidence in the Shiqu County of Sichuan
724 Province, China: Analysis, Assessment, and Management. EXPOSURE AND
725 HEALTH 12(2), 307-322. <https://doi.org/10.1007/s12403-019-00314-y>.

726 Wang, F., Dong Y., Yang K. and Zhang A., 1994. A study on the application of
727 potassium fertilizer to maize on low humidity farmland in the Sanjiang Plain. Journal
728 of Heilongjiang Bayi Agricultural Reclamation University 1994,7(01): 13-16.

729 Wang, L., Han, X., Zhang, Y., Zhang, Q., Wan, X., Liang, T., Song, H., Bolan, N.,
730 Shaheen, S.M., White, J.R. and Rinklebe, J., 2023. Impacts of land uses on spatio-
731 temporal variations of seasonal water quality in a regulated river basin, Huai River,
732 China. Science of The Total Environment 857(Pt 2), 159584.
733 <https://doi.org/10.1016/j.scitotenv.2022.159584>.

734 WHO, 2021. A global overview of national regulations and standards for drinking-
735 water quality, 2th edition. World Health Organization.

736 Wu, Y., Ju, H., Jiang, H., Zhang, G., Qi, P. and Li, Z., 2023. Identifying nitrate
737 sources and transformations in an agricultural watershed in Northeast China: Insights
738 from multiple isotopes. Journal of Environmental Management 340.
739 <https://doi.org/10.1016/j.jenvman.2023.118023>.

740 Xiajing, L., Aili, Y. and Fuxiang, C., 2018. Measurement and Evaluation of
741 Residual Disinfection by Products in Tap Water from Xiamen. IOP Conference Series:
742 Earth and Environmental Science 146. [https://doi.org/10.1088/1755-
743 1315/146/1/012015](https://doi.org/10.1088/1755-1315/146/1/012015).

744 Yan, J., Chen, J. and Zhang, W., 2021. Study on the groundwater quality and its
745 influencing factor in Songyuan City, Northeast China, using integrated
746 hydrogeochemical method. Science of The Total Environment 773.
747 <https://doi.org/10.1016/j.scitotenv.2021.144958>.

748 Ye, X., Zhou, Y., Lu, Y. and Du, X., 2022. Hydrochemical Evolution and Quality

749 Assessment of Groundwater in the Sanjiang Plain, China. *Water* 14(8).
750 <https://doi.org/10.3390/w14081265>.

751 Yin, Z., Luo, Q., Wu, J., Xu, S. and Wu, J., 2021. Identification of the long-term
752 variations of groundwater and their governing factors based on hydrochemical and
753 isotopic data in a river basin. *Journal of Hydrology* 592.
754 <https://doi.org/10.1016/j.jhydrol.2020.125604>.

755 Zhang, B., Song, X.-f., Zhang, Y.-h., Han, D.-m., Tang, C.-y., Yang, L. and Wang,
756 Z., 2017. The renewability and quality of shallow groundwater in Sanjiang and Songnen
757 Plain, Northeast China. *Journal of Integrative Agriculture* 16(1), 229-238.
758 [https://doi.org/10.1016/s2095-3119\(16\)61349-7](https://doi.org/10.1016/s2095-3119(16)61349-7).

759 Zhang, H., Han, X., Wang, G., Mao, H., Chen, X., Zhou, L., Huang, D., Zhang, F.
760 and Yan, X., 2023. Spatial distribution and driving factors of groundwater chemistry
761 and pollution in an oil production region in the Northwest China. *Science of The Total*
762 *Environment* 875. <https://doi.org/10.1016/j.scitotenv.2023.162635>.

763 Zhang, Q., Wang, H. and Lu, C., 2020. Tracing sulfate origin and transformation
764 in an area with multiple sources of pollution in northern China by using environmental
765 isotopes and Bayesian isotope mixing model. *Environmental Pollution* 265.
766 <https://doi.org/10.1016/j.envpol.2020.115105>.

767 Zhang, X., Zhao, R., Wu, X. and Mu, W., 2021. Hydrogeochemistry, identification
768 of hydrogeochemical evolution mechanisms, and assessment of groundwater quality in
769 the southwestern Ordos Basin, China. *Environmental Science and Pollution Research*
770 29(1), 901-921. <https://doi.org/10.1007/s11356-021-15643-2>.

771 Zhong, C., Wang, H. and Yang, Q., 2022. Hydrochemical interpretation of
772 groundwater in Yinchuan basin using self-organizing maps and hierarchical clustering.
773 *Chemosphere* 309. <https://doi.org/10.1016/j.chemosphere.2022.136787>.

774