Assessing environmental interference in northern China using a spatial distance model: From the perspective of geographic detection

Wei Wei a, Zecheng Guo a,⁎, Liang Zhou b,c,⁎⁎, Binbin Xie d, Junju Zhou a

a College of Geography and Environmental Science, Northwest Normal University, Lanzhou 730070, Gansu, China
b Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China
c Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
d School of Urban Economics and Tourism Culture, Lanzhou City University, Lanzhou 730070, Gansu, China

HIGHLIGHTS
• The spatial distance model (SDM) was proposed to evaluate the natural environmental factors at large scale.
• Human interference factors were quantified and calculated the environmental interference index as input variable.
• The spatial visual expression is combined with GIS to comprehensively express the environmental interference.

GRAPHICAL ABSTRACT

Abstract

The rapid development of society and the expansion of human activities have resulted in interference with the natural environment. Assessing the environmental interference (EI) caused by human activities is highly important for socio-economic sustainable development. In this study, the spatial distance model (SDM) and resource endowment index (REI)-human activity index (HAI) ratio model were developed to calculate the environmental interference index (EII) in northern China (NC). The current spatial distribution and patterns of EII in NC were analyzed based on geographic information system (GIS) technology. In addition, the factors that influence the level of EI were examined through a geographical detector method. The results showed that the EII value in the eastern region was significantly higher than that in the western region and that differences in EI were spatial heterogeneity. The spatial distribution of EI was analyzed at the provincial, municipal and county scales, respectively. It was found that its distribution was closely related to urban development. The spatial distribution of EI displayed longitudinal zonality. East of 104.987°E, there were many large cities, such as Beijing, Tianjin, Qingdao and Zhengzhou, with high population densities and developed economies. Thus, these areas had high EI values. To the west of 104.987°E, such as in the Qinghai, Gansu, Xinjiang and Inner Mongolia regions, the EI values were generally low, with low environmental quality and fewer human activities. The level of EI in the Huang-Huai-Hai...
1. Introduction

Human activities affect changes in the natural environment through direct or indirect mechanisms (Nguyen and Liou, 2019). For a long time, the changes in ecosystem structure and function caused by various environmental and human factors have been the focus of ecological science (Kozlov et al., 2017). With disorderly population growth and the uncontrollable development of the economy, the global environment has undergone considerable changes recently, and serious environmental problems, such as increased temperatures, flooding, drought and desertification, have frequently occurred (Halpern et al., 2008; Hu and Xu, 2018). Simultaneously, human activities have exerted great pressure on the regional environment, leading to the degradation or extinction of native forests and grassland vegetation, the decline of biodiversity, soil and water losses, and the destruction of the physical and chemical properties of the soil (Dearborn and Kark, 2010). Therefore, it is of great significance for environmental protection and management to identify the degree of influence of and disturbance from human activities on the regional environment.

It is very important to evaluate disturbance from human activities in a real and objective way while considering the influence of the regional natural environment. The conditions that constitute the natural environment vary from region to region, and the interactions of various natural conditions constitute the environmental backgrounds of different regions (Zhao et al., 2019). In the relationship between human activities and the regional natural environment, the phenomenon of the “barrel effect” occurs (Fig. 1). The capacities of different regions to bear the pressures of human activities are closely related to various conditions that constitute the regional environments (e.g., the vegetation, soil, precipitation, and terrain) and mainly depend on the environmental weaknesses. With the increase of the intensity of human activities, the associated effects become closer to the bearing limitation of the regional environment. When environmental weaknesses break through, a spill-over effect occurs, which is the main cause of regional environment deterioration. In this paper, we propose the concept of “environmental interference (EI)” for the quantitative disclosure of the states and responses of regional natural environmental changes caused by human activity interference based on this principle. EI can directly reflect the coordination between social development and the natural environment, which is based on various comprehensive factors. The results of the quantitative assessment of EI can provide a decision-making basis for zoning management and planning, governance and restoration measures based on different levels of EI.

In recent years, the severity of all types of environmental problems caused by human activities has been cognized by researchers in different fields (Calvo and Rousseau, 2019). Scholars have extensively studied the footprints of human activities and the corresponding impacts on various aspects of the environment (Zhang et al., 2018; Plas et al., 2019), Xu et al. (2016) quantitatively evaluated the effects of human beings on the land surface by building an algorithm model of human activity intensity and establishing a method for converting different land use/land cover types into construction land equivalents. Zhang et al. (2019) explored the temporal patterns of human activities using a nuclear density estimation method and a two-dimensional temporal plane based on a one-week smart card dataset generated in the Shanghai metro system. Based on data collected at hydrological sites, Zhao et al. (2019) used a new quantitative assessment model to evaluate the impact of human activities on the phytoplankton community in Spring City, Jinan. Many scholars have attempted to identify the influence of human activities on natural elements (e.g., water, soil, vegetation and animals) through quantitative methods (Olabode, 2019) and determine the corresponding interference locations and levels through geographic mapping. However, due to the complexity of the natural environment and the difficulty of data acquisition, the comprehensive assessment of EI is difficult, as there is no universal assessment system for reference. In addition, it is difficult to evaluate EI over large areas because of regional environmental differences and complexities. Therefore, regional-scale assessment is necessary to accurately understand the local EI.

Currently, quantitative assessments typically adopt the multi-index weighted superposition method (Gao and Wu, 2010), which generally includes the establishment of an assessment index system, determination of index weights, and establishment of assessment models (Korpinen et al., 2013). Determining how to scientifically and objectively calculate the weight of each index is one of the most important tasks in the assessment of EI. Many methods can be used for this task, such as the analytic hierarchy process (AHP) (Nguyen et al., 2016), Delphi (Diakoulaki et al., 1995), fuzzy evaluation (Akter et al., 2019) and expert marking (Zhong et al., 2015) methods. However, most of these methods are not objective in terms of determining the weights of indexes, and the results are greatly affected by knowledge levels of the experts, which can lead to inaccurate results in comprehensive environmental assessments (Aryafar et al., 2013).

On this basis, many objective methods, such as the gray relative analysis method (You et al., 2017), the rough set method (Bai et al.,
the entropy weighting method (Xu et al., 2018) and principal component analysis (Sánchez-Navarro et al., 2015), can directly calculate the weight of each index. However, some variables and parameters are not easy to obtain in the calculation process, and these methods are highly reliant on the data and may ignore the local environmental state. Therefore, the spatial distance model (SDM) and ratio model are used in this study. The SDM is based on the theory of Euclidean distance (Amani et al., 2017; Mesquita et al., 2017; Wei et al., 2019) and can be used to calculate comprehensive values of assessment indexes because of its simple theory, objective data support, fast algorithms, and visualization of results (Zhang, 2016). In addition, a ratio model is used to calculate the coordinated changes resulting from EI. The calculation process is simple, efficient and objective. The ratio model can be used to determine the interference effects of real human activities hidden by natural environmental conditions in different regions to support EI assessments (Liu et al., 2008).

In this study, the research objectives are as follows: (1) provide an assessment system/framework for EI at the macro-scale and (2) identify the interference area and influence degree of human activities, reveal the spatial distribution characteristics of EI in north China using various spatial analysis methods, and put forward suggestions for the harmonious development of society and nature.

2. Study areas

Northern China (NC) has a vast territory and a large span located at 73°40′-135°2′E, 31°9′-53°33′N (Hu et al., 2019). The total area is about 5.642 × 106 km², accounting for 59% of China’s total land area. The administrative region covers 15 provinces (autonomous regions and municipalities directly under the central government) including Qinghai (QH), Gansu (GS), Ningxia (NX), Shaanxi (SN), Inner Mongolia (IM), Xinjiang (XJ), Hebei (HE), Shandong (SD), Henan (HA), Shanxi (SX), Heilongjiang (HL), Jilin (JL), Liaoning (LN), Beijing (BJ) and Tianjin (TJ) (Chen et al., 1998). NC is a relatively fragile area in terms of the ecological environment. In recent years, with the successive implementation of the national ecological management project, the vegetation coverage has been significantly improved in NC. The relief is rugged, spanning three steps from west to east. The landform types are complex and diverse, including deserts, mountains, hills, plateaus, plains and basins. The climate types are diverse, including the monsoon climate, temperate continental climate and plateau mountain climate. The annual precipitation decreases from east to west from >1000 mm to below 100 mm (Li et al., 2016b). Simultaneously, the distribution of the average annual temperature is obviously regional (Li et al., 2016a). In NC, the southeast region compared with the northwest region presents the phenomenon of rapid economic development, rapid population growth and a large amount of land being reclaimed (Fig. 2).

3. Methodology

3.1. Data sources and processing

Eight basic categories of data are used in this study, which are meteorological data (including the average temperature (°C) and precipitation (mm)), vegetation coverage data (including the normalized differential vegetation index (NDVI) and vegetation type), soil data (including the organic carbon content (%), soil type and soil erosion type), digital elevation model (DEM), land use data, 1:100,000 China basic geographic databases (including rivers, ditches, lakes, reservoirs, settlements, roads, highways, railways and high-speed railways), NPP VIIRS night-time light data and statistical data (population (persons) and gross domestic product (GDP/ten thousand yuan)). The detailed descriptions are summarized in Table 1.

The pre-processing of the collected data mainly consists of the following two steps. (1) Projection transition is used for all spatial data, and Albers equal area projection is regarded as the target projection. (2) All data are converted to the 1 km resolution raster. In addition, all spatial data are processed and mapped using the software ArcGIS 10.6 and ENVI 5.3. The statistical graphs are drawn using the Origin 2017 software.

3.2. Index selection

Environmental changes are the products of interactions among various natural and man-made factors (Zhao et al., 2018). In terms of the natural environment, vegetation cover is a decisive factor that reflects the quality of the natural environment, plays an important role in the anti-interference and buffering ability of the regional environment, and maintains the soil and water conservation ability of regions (Nguyen et al., 2016). Therefore, the normalized difference vegetation index (NDVI) reflects the status of vegetation growth and the degrees of aggregation (Li et al., 2017; Munavar et al., 2018), and different vegetation types have different abilities to prevent water and soil erosion (Xu et al., 2019).

Soil is the material basis for the circulation and development of terrestrial ecosystems, as well as the basis for the survival and

Fig. 2. Location of the study area.
The geomorphologic types in NC are mainly plains and hills in the east, and plateaus, basins and mountains occur in the west. The elevation, slope degree and slope aspect are the main indices of landform change (Tang et al., 2018; Kang et al., 2018). The topographic elements create spatial differences in the surface cover, climatic conditions and surface runoff. In NC, the climate is mainly a temperate continental climate with four distinct seasons. Precipitation and temperature are the driving forces of all-natural factors and the energy foundation of the ecosystem, and they experience obvious seasonal variations (Zhao et al., 2018). The hydrological environment plays a key role in the survival and development of plants and animals in the region (Zhou et al., 2019). Surface water and groundwater are two important types of regional water resources. However, the characteristics of the groundwater units of various geomorphic types are quite different in this study area, partly due to the rugged terrain. As a result, the representativeness of indexes reflecting groundwater, such as the groundwater depth, is not clear, and some data are difficult to obtain. Therefore, the indexes of groundwater are not considered in this paper. The distributions of rivers, ditches, lakes and reservoirs affect the environmental evolution in different geographical units. Generally, the distance from rivers, canals, lakes and reservoirs can be used as a proxy for the abundance of surface runoff (Wei, 2018).

In addition, from the perspective of human interference, this paper simulates the distribution of human activities from five aspects: population economy, settlement aggregation, traffic infrastructure, land utilization and urbanization level. Socio-economic development has a significant effect on the regional natural environment. In NC, the population distribution and economic level occupy a large proportion in the process of socio-economic development (Wei et al., 2014). On the other hand, a settlement refers to a place where human beings gather and live under the comprehensive influence of many factors (e.g., nature, socio-economy and culture) (Clark et al., 2009). Human habitation is closely related to the surrounding geographical environment. Typically, the denser is the distribution of settlements, the greater is the interference of humans on the region and the surrounding natural environment.

During the construction and use of roads, various disturbances, including the pollution of water resources, occupation, the destruction and loss of land resources, damage to original vegetation growth and biodiversity, and atmospheric pollution caused by automobile exhaust, affect the surrounding natural environment (Mostafa and Lei, 2014). It is necessary to obtain the distribution information for various roads to study the human interference level in NC.

Different land use types have diverse influences on the regional natural environment (Verburg et al., 2013). Changes in land use patterns affect biological, hydrological and energy processes on the Earth’s surface. As a result, some environmental and socio-economic problems, such as soil erosion, soil nutrient transfer, biodiversity reduction, increased river runoff and food shortages, are becoming increasingly common (Yan et al., 2017). The population increase, changes in the spatial extents of different land covers and intensified economic activity are the main reasons for the continuous changes that occur during the urbanization process (Ma et al., 2012). As a habitat becomes dominated by human beings, man-made environmental changes often result in an urban agglomeration effect related to a high population density, high energy consumption intensity and social and extensive economic activities (Montgomery, 2008). It is very important to explore the relation between the urbanization development level and human disturbance activities in NC (Ma et al., 2014).

In summary, twelve indexes related to natural environmental factors were selected, including the normalized difference vegetation index (NDVI), erosion protection (EP), soil organic matter content (SOMC), soil depth (SD), soil erosion intensity (SEI), elevation (ELE), slope degree (SDE), slope aspect (SA), annual precipitation (AP), annual average temperature (AAT), river network density (RND), and distance from lakes and reservoirs (DLR). In addition, seven indexes related to human interference factors were selected, including the population density (PD), economic density (ED), settlement density (SED), road network density (ROND), railway network density (RAND), land use intensity (LUI), and night-time light intensity (NTL). The calculation methods of these indexes are described in detail in Section A and illustrated in Figs. A.1 and A.2 in Appendix A. Meanwhile, in this study, multicollinearity diagnosis is adopted to avoid the problem of collinearity. In Table 2, the results show that all the variance inflation factors (VIF) for the indexes are lower than 10, while the tolerance (TOL) values

<table>
<thead>
<tr>
<th>Data</th>
<th>Year</th>
<th>Resolution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meteorological data</td>
<td>2017</td>
<td>–</td>
<td>Chinese Meteorological Data Sharing Service System (<a href="http://cdc.cma.gov.cn">http://cdc.cma.gov.cn</a>)</td>
</tr>
<tr>
<td>Vegetation coverage data</td>
<td>2017</td>
<td>1 km</td>
<td>NASA data system (<a href="https://earthdata.nasa.gov">https://earthdata.nasa.gov</a>)</td>
</tr>
<tr>
<td>MOD13A3</td>
<td>2001</td>
<td>1 km</td>
<td>Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (<a href="http://www.resdc.cn">http://www.resdc.cn</a>)</td>
</tr>
<tr>
<td>Vegetation type</td>
<td></td>
<td></td>
<td>Harmonized World Soil Database (HWSD 1.2 version) (<a href="http://www.fao.org/land-water/en">http://www.fao.org/land-water/en</a>)</td>
</tr>
<tr>
<td>Soil data</td>
<td></td>
<td></td>
<td>Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (<a href="http://www.resdc.cn">http://www.resdc.cn</a>)</td>
</tr>
<tr>
<td>Soil erosion type</td>
<td>2010</td>
<td>1 km</td>
<td>Harmonized World Soil Database (HWSD 1.2 version) (<a href="http://www.fao.org/land-water/en">http://www.fao.org/land-water/en</a>)</td>
</tr>
<tr>
<td>Digital elevation model (DEM)</td>
<td>2003</td>
<td>90 m</td>
<td>Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (<a href="http://www.resdc.cn">http://www.resdc.cn</a>)</td>
</tr>
<tr>
<td>Land use data</td>
<td>2015</td>
<td>1 km</td>
<td>Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (<a href="http://www.resdc.cn">http://www.resdc.cn</a>)</td>
</tr>
<tr>
<td>China basic geographic databases</td>
<td>2017</td>
<td>1:1000000</td>
<td>National Geomatics Center of China (<a href="http://www.ngcc.cn">http://www.ngcc.cn</a>)</td>
</tr>
<tr>
<td>NPP VIIRS night-time light data</td>
<td>2017</td>
<td>0.00417°</td>
<td>National Oceanic and Atmospheric Administration (<a href="https://www.ngdc.noaa.gov/eog/viirs">https://www.ngdc.noaa.gov/eog/viirs</a>)</td>
</tr>
<tr>
<td>Statistical data</td>
<td>2018</td>
<td>–</td>
<td>China Statistical Yearbooks Database (<a href="http://tongji.cnki.net">http://tongji.cnki.net</a>)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Statistical yearbooks of provinces, districts and municipalities</td>
</tr>
</tbody>
</table>

Table 1: Description of the data used in the study.
Here, $X$ and $Y$ are individuals with multidimensional characteristics, where $X = (x_1, x_2, x_3, ..., x_n)$ and $Y = (y_1, y_2, y_3, ..., y_n)$. $d(X, Y)$ is the Euclidean distance between $X$ and $Y$, and $x_i$ and $y_i$ are different characteristics.

Based on the theory of Euclidean distance (Mesquita et al., 2017), the SDM is introduced to calculate the comprehensive index (Zhang, 2016; Wei et al., 2019). The basic principle of this model involves constructing a multidimensional space, with each index representing each dimension and $n$ indexes constituting an $n$-dimensional space. In this space, the spatial distribution of the natural environment or human interference is quantitatively simulated by calculating the distances from other points to the datum point (this point represents the lowest level of natural environment or human interference) in the space. The longer is the distance, the larger is the value calculated by the SDM and the better is the natural environment or the higher is the level of human interference. A schematic diagram of the SDM in three-dimensional space ($n = 3$) is shown in Fig. 3 (Amani et al., 2017). In three-dimensional space, point A represents the worst natural environment or the lowest level of human interference, point B represents the best natural environment or the highest level of human interference, and distance $d$ represents the Euclidean distance between B and A in the space, which is the longest distance in the space. In addition, the REI or HAI can be calculated by Eqs. (4) and (5), and the results are shown in Fig. 4a and b.

$$REI = \sum_{i=1}^{n} (E_i - E_{i\_\text{min}})^2$$

$$HAI = \sum_{i=1}^{n} (H_i - H_{i\_\text{min}})^2$$

Here, $E_i/H_i$ is the ith index of the natural environmental (total of 12 indexes)/human interference (total of 7 indexes) factors and $E_{i\_\text{min}}/H_{i\_\text{min}}$ is the minimum value of the ith index of natural environmental/human interference factors.

### 3.3. Construction of EI model

#### 3.3.1. Standardization of indexes

To eliminate the influence of different units among the different indexes, a dimensionless assessment data set was obtained by the range standardization method (Wei et al., 2019). According to the different effects of different indexes on the natural environment and human activities, the indexes were divided into positive indexes and negative indexes (Sun et al., 2013; Zhao et al., 2018). For natural environmental factors, a positive index is defined as an index that causes the natural environment to become better upon the increase of the index value, while a negative index (e.g., SEI, DLR) refers to the opposite case. For human interference factors, the definition of a positive index is an index that will be more affected by human activities with the increase of the index value, while a negative index refers to the opposite case. In this study, each index of human interference factors is a positive index. The original value calculation formulas of the indexes are standardized as follows:

**Positive index**:

$$NI = \frac{I - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}}$$  \hspace{1cm} (1)

**Negative index**:

$$NI = \frac{I_{\text{max}} - I}{I_{\text{max}} - I_{\text{min}}}$$  \hspace{1cm} (2)

Here, $NI$ is the standardized value of the index, $I$ is the original index value, and $I_{\text{max}}$ and $I_{\text{min}}$ are the maximum and minimum original index value, respectively.

#### 3.3.2. Calculation of the resource endowment index (REI) and human activity index (HAI)

In mathematics, the Euclidean distance is one of the most important distance metrics used to measure the absolute distance between points in multidimensional space (Mesquita et al., 2017). The calculation formula is shown in Eq. (3).

$$d(X, Y) = \sum_{i=1}^{n} (x_i - y_i)^2$$

Note: NDVI: normalized difference vegetation index; EP: erosion protection; SOMC: soil organic matter content; SD: soil depth; SEI: soil erosion intensity; ELE: elevation; SDE: slope degree; SA: slope aspect; AP: annual precipitation; AAT: annual average temperature; RND: railway network density; DLR: distance from lakes and reservoirs; PD: population density; ED: economic density; SED: settlement density; ROND: road network density; RAND: railway network density; LUI: land use intensity; NTL: night-time light intensity.

<table>
<thead>
<tr>
<th>Index</th>
<th>VIF</th>
<th>TOL</th>
<th>Index</th>
<th>VIF</th>
<th>TOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI</td>
<td>4.726</td>
<td>0.212</td>
<td>RND</td>
<td>1.321</td>
<td>0.757</td>
</tr>
<tr>
<td>EP</td>
<td>2.294</td>
<td>0.436</td>
<td>DLR</td>
<td>1.221</td>
<td>0.819</td>
</tr>
<tr>
<td>SOMC</td>
<td>1.121</td>
<td>0.892</td>
<td>PD</td>
<td>7.193</td>
<td>0.139</td>
</tr>
<tr>
<td>SD</td>
<td>1.754</td>
<td>0.57</td>
<td>ED</td>
<td>4.292</td>
<td>0.233</td>
</tr>
<tr>
<td>SEI</td>
<td>2.203</td>
<td>0.454</td>
<td>SED</td>
<td>1.816</td>
<td>0.551</td>
</tr>
<tr>
<td>ELE</td>
<td>2.145</td>
<td>0.466</td>
<td>ROND</td>
<td>2.063</td>
<td>0.485</td>
</tr>
<tr>
<td>SDE</td>
<td>1.581</td>
<td>0.632</td>
<td>RAND</td>
<td>1.281</td>
<td>0.781</td>
</tr>
<tr>
<td>SA</td>
<td>1.009</td>
<td>0.991</td>
<td>LUI</td>
<td>2.213</td>
<td>0.452</td>
</tr>
<tr>
<td>AP</td>
<td>3.766</td>
<td>0.266</td>
<td>NTL</td>
<td>1.199</td>
<td>0.834</td>
</tr>
<tr>
<td>AAT</td>
<td>2.164</td>
<td>0.462</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results of multicollinearity diagnostics.
3.3.3. Calculation of the environmental interference index (EII)

The natural environment and human activities are interrelated and co-restrictive (Kreuter et al., 2001). How to accurately express the real interference effect of human activities in the context of the regional environment is very important. The ratio model has been used widely to extract some of the specific information required, such as the ratio vegetation index (RVI) (Major et al., 1990), normalized difference water index (NDWI) (Campos et al., 2012), and index-based built-up index (IBI) (Xu, 2008). The model identifies the needed information by suppressing other information (Liu et al., 2008). Therefore, the ratio model is introduced in this paper. The spatial distribution of the EII is shown in Fig. 4c. The calculation formula of the HAI-REI ratio model is as follows.

\[
EII = \frac{HAI}{REI} \tag{6}
\]

3.4. Classification of the assessment results and spatial scale analysis

To compare and summarize the spatial distribution of EII in NC, the research results need to be classified. Here, combined with regional geographical characteristics, the results are classified by the natural breaks (Jenks) method. The classification standard is shown in Table 3.

In addition, for the purpose of further exploring the spatial heterogeneity characteristics of EI at different scales (e.g., the administrative scale and longitude and latitude scale), this study adopt the EI classification index (EIIC) to change the resolution of the assessment results in a more accurate way by calculating the EI of each unit area at different scales (Li et al., 2006). The calculation formula is shown in Eq. (7).

\[
EIIC = \sum_{i=1}^{n} \frac{P_i \times A_i}{S} \tag{7}
\]

Here, \( P_i \) is the class value (Table 3), \( A_i \) represents the areas of different interference classes, and \( S \) is the total area of the evaluation unit.

3.5. Spatial autocorrelation analysis

Spatial dependence means that a phenomenon is not completely isolated in space but has a certain correlation with adjacent spatial units. Spatial autocorrelation is a quantitative analysis reflecting this correlation. According to different emphases, there are global and local spatial autocorrelation methods (Zawadzki et al., 2005; Sikder et al., 2019). The global measure mainly includes Moran’s伊拉 and it can measure the spatial characteristics of EI in NC. The calculation formula of Moran's伊拉 is shown in Eq. (8).

\[
l_g = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}} \tag{8}
\]

Here, \( l_g \) is the global Moran’s ila, n is the total number of grids in the study area, \( W_{ij} \) is the row-standardized contiguity matrix, \( x_i \) and \( x_j \) are the EII at grids i and j, respectively, and \( \bar{x} \) is the average level of the EII. The value of Moran’s伊拉 ranges from 1 (positive spatial autocorrelation) to -1 (negative spatial autocorrelation) (Hu and Xu, 2018).

Global autocorrelation can only grasp the spatial clustering characteristics of EI in NC as a whole. However, to accurately reflect the changes of spatial differences of regional EI, the local analysis method of exploratory spatial data analysis (ESDA) is needed. Local indicators of spatial association (LISA) are used in the local spatial autocorrelation method, which is used to look for the distribution characteristics of local unbalanced EI that are covered or may be inconsistent with the global spatial autocorrelation. The LISA value describes the spatial clustering degrees of EI between the spatial unit and its surrounding area of similar values (Hu et al., 2015). The local Moran’s ila (Eq. (9)) is employed to show the LISA in this study.

\[
l_i = \frac{n(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sum_{j=1}^{n} W_{ij}(x_j - \bar{x}) \tag{9}
\]

Here, \( l_i \) is the local Moran’s ila.

The global Moran’s伊拉 and the local Moran’s ila were calculated in ArcGIS 10.6 (Chen and Zhu, 2012). In the program, the spatial relationships between evaluation units were determined by the method of inverse distance weighting (Łukaszyk, 2004). In the method, the neighborhood size is set automatically using a default value of its radius, which is a Euclidean distance, and a weighted average is taken of the observation values within this neighborhood. For the local Moran’s ila, a cluster map incorporating information about the significance of the local spatial patterns was created. In particular, the map resulted in a spatial typology consisting of five categories in terms of the EI, including “High-High Cluster”, “Low-Low Cluster”, “Low-High Outlier”, “High-Low Outlier” and “Not significant” (Hu et al., 2015).

3.6. Geographic detector

The geographic detector is a novel statistical method proposed by Wang et al. (2010), which can detect spatial variations and reveal the driving factors behind them. The detector model contains four modules.
including the risk detector, factor detector, ecological detector and interaction detector (Bai et al., 2019). In this study, we use the factor detector to examine the explanatory powers of different influencing factors for environmental disturbances in NC (Zhou et al., 2018).

The factor detector is measured according to the following q-statistic:

\[ q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N \sigma^2} = 1 - \frac{SSW}{SST} \]  

\[ SSW = \sum_{h=1}^{L} N_h \sigma_h^2 \]  

\[ SST = N \sigma^2 \]

Here, \( h = 1, 2, ..., L \) (the number of strata for variable \( Y \) or factor \( X \)). \( N \) represents the total number of spatial units over the entire study area. \( N_h \) denotes the number of units in strata \( h \). \( \sigma^2 \) and \( \sigma_h^2 \) denote the variance of the \( Y \) value in the entire study area and strata \( h \), respectively. \( SSW \) and \( SST \) are the within sum of squares and total sum of squares, respectively. The value range of \( q \) is \([0, 1]\). The larger the value, the greater is the influence of the index on the EI.

4. Results

4.1. Spatial distribution and differentiation

Fig. 5 shows the overall and local spatial distributions of the classified results for EI. The distribution of EI showed significant differentiation. In NC, the regions with high levels of EI (including high interference and extreme interference) were mainly concentrated in the east, and the level of EI in the west was relatively low. From the natural environmental distribution (Fig. 4a), the regions with high levels of EI were mainly plain and hilly regions with flat terrain and abundant natural resources. In regions with harsh natural conditions, such as plateaus, mountains and deserts, the interference degrees were relatively low. From the human activity trajectory (Fig. 4b), regions with developed economies, convenient transportation and dense population had high EI. In addition, from the area statistics of different EI classes (Fig. 6), NC was dominated by low levels of EI (including no interference, light interference and low interference) in 76.134% of the total area, and the areas with high levels of EI accounted for only 11.570% of the total area.

We obtained the local distribution of EI by amplification region analysis (Fig. 5). Eight sample regions were selected from different agricultural districts for comparison and analysis. The EI had various

---

![Fig. 5. Spatial distribution map of EI in NC.](image-url)
distribution characteristics in different sample regions. For sample regions A, B and C located in the northern arid and semi-arid region, the natural conditions were relatively harsh, making the region susceptible to human activities. The distribution boundaries of EI were clear, showing a trend of high EI surrounded by low EI. The interference class of high EI regions was mainly high interference, with the main land use type in these regions being cultivated land. Sample regions H and I were located in the farming area of the northeast plain, with better natural conditions. As a result, the interference class of cultivated land was mainly medium interference, which was significantly lower than that in sample regions A, B and C. For sample regions D, E, F and G with high levels of socio-economic development, EI displayed a certain “concentric circle” distribution. The interference degree in the city center was the highest, and the interference level gradually decreased outward.

### 4.2. Spatial heterogeneity of EI

#### 4.2.1. Spatial heterogeneity of different administrative scales

Regions of intense human activity tend to be concentrated in large or medium-sized cities with developed economies and good infrastructure. By calculating the distribution of EI at the provincial, municipal and county levels (Eq. (7)), we explored the distribution of human interference at different administrative scales (Fig. 7). The results showed that the spatial distribution of EI at different administrative scales displayed certain directivity and aggregation trends. At the provincial scale, the degrees of EI gradually increased from west to east. Tianjin had the highest interference degree, and Qinghai and Xinjiang had the lowest interference degree. At the municipal scale, two large urban clusters with relatively high levels of EI (including medium interference and high interference) appeared in the Beijing-Tianjin-Hebei region and Henan Province. At the county scale, the distribution of EI had certain clustering characteristics; that is, both high and low interference showed a pattern of continuous distribution. There were 31 counties with extreme interference, which were mainly distributed in the downtown areas of cities in Henan, Hebei, Beijing, Tianjin, Shandong, Liaoning and Shaanxi provinces. In general, these phenomena showed that the EI distribution in NC was closely related to urban development.

#### 4.2.2. Spatial heterogeneity of latitude and longitude zones

The spatial heterogeneity of EI can be revealed by exploring regional changes in different latitudinal and longitudinal zones in NC. Fig. 8 shows the statistical distribution of EI in different latitude and longitude zones. It indicates that NC was characterized by significant longitudinal zonality and that the distribution of EI showed a polarization phenomenon. With 104.987° E as the dividing line, the EI in the west was relatively low, ranging between 1.5 and 2.4. The western longitudinal zone of 78.987°–79.987° had the lowest EI, with a value of 1.591, which basically reflected no interference. The EI in the east was relatively high, with a range of 2.4 to 4.1. The highest EI was located in the eastern longitudinal zone of 114.987°–115.987°, with a value of 4.051, which reflected medium interference. However, there was no obvious latitudinal zonality in NC, and the distribution displayed a fluctuating trend, with two troughs in the northern latitude zones of 38.884°–39.884° and 51.884°–52.884°.

Note: Different longitude and latitude zones are divided based on intervals of 1°, and the interference class value is calculated by the EICI.

#### 4.3. Spatial dependence of EI

Based on the results of global autocorrelation analysis for EI, the p-value was 0, and the z-score was 737.185, which indicated that the results passed the significance test at 1%. Additionally, the Moran’s index value was 0.778. Overall, the spatial distribution of EI in NC displayed positive spatial autocorrelation and high spatial dependency. Fig. 9 shows the results of local spatial autocorrelation. The results indicated that the two clustering patterns of high-high interference and low-low interference were distributed in NC and showed obvious characteristics of minimal clustering with low interference and considerable clustering with high interference. Among them, the low-low cluster was mainly distributed in northwest Gansu, western Inner Mongolia, and many areas of Xinjiang and Qinghai. The high-high cluster was distributed in the east of NC, namely, the economically developed and densely populated provinces of Beijing, Tianjin, Shandong, Henan, Hebei, Shaanxi, Shanxi, Liaoning, Jilin and Heilongjiang.

Note: Not Significant means not statistically significant. High-High/Low-Low Cluster signifies that the statistically significant regions of high/low interference are surrounded by their homogeneous regions, respectively. High-Low/High-Outlier indicates the statistically significant...
significant regions with high/low interference surrounded by their heterogeneous regions, respectively (Hu and Xu, 2018).

4.4. Influencing mechanism of EI

The impacts (q-value) of 19 indexes on EI were obtained by a geographical detector method, and the results are reported in Table 4. As shown in Table 4, all 19 indexes passed the significance test at 1%, which indicated that the selected indicators were statistically significant for the quantitative simulation of EI in NC. Among the natural environmental factors, the NDVI (0.400) played the largest role, which was the most direct factor reflecting the quality of the natural environment. The SA (0.005) had the smallest influence. Among the human interference factors, the impacts of 7 indexes were all >0.1. Among them, the LUI had a significant relation to EI, and the q-value reached 0.851, indicating that the index was closely related to the intensity of human activities in different land use types. In general, different land use behaviors affected the distribution and changes of the land use cover structure to a large extent and restricted the track and influence range of human activities on the whole. The lowest impact was associated with the RAND, potentially because the railway network was not as dense as the road network; thus, the associated interference effect on the surrounding environment was limited. However, the dense road network influenced the environment, with a q-value reaching 0.403.

In general, in NC, various types of interference associated with human activities, including population expansion, economic development and infrastructure construction, affect the regional environment by directly and indirectly changing the land use structure. However, growing vegetation is the most important way to mitigate human interference.

5. Discussions

5.1. Understanding the EI effect to inform regional management

The purpose of EI assessment is to provide policy makers with recommendations for environmental improvement (Zou and Yoshino, 2017). We divided the maps of the natural environment and human interference into high and low classes based on the natural breaks (Jenks) method and then compiled the spatial divisions of environmental management and optimization in NC with the map superposition method. As shown in Fig. 10, the general programming was divided into four regions (i.e., the comprehensive control region, environmental migration region, environmental conservation region, and integrated development region).

(1) Comprehensive control region (CCR). The CCR encompassed “low-quality natural environment and low level of human interference” areas in NC. These areas mainly included sandy land, bare

---

**Table 4**

Results of the geographic detector analysis.

<table>
<thead>
<tr>
<th>Natural environmental factors</th>
<th>Natural environmental factors</th>
<th>Human interference factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index</td>
<td>q</td>
<td>Index</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.400**</td>
<td>SDE</td>
</tr>
<tr>
<td>EP</td>
<td>0.278**</td>
<td>SA</td>
</tr>
<tr>
<td>SOMC</td>
<td>0.090**</td>
<td>AP</td>
</tr>
<tr>
<td>SD</td>
<td>0.113**</td>
<td>ATT</td>
</tr>
<tr>
<td>SEI</td>
<td>0.162**</td>
<td>RND</td>
</tr>
<tr>
<td>ELE</td>
<td>0.284**</td>
<td>DLR</td>
</tr>
<tr>
<td>NTL</td>
<td>0.152**</td>
<td></td>
</tr>
</tbody>
</table>

**Indicates that the q-value is significant at the 0.01 level (p < 0.01).**

---

**Fig. 8.** Statistics of EI in longitude and latitude zones. a: the distribution of the longitude zone; b: the distribution of the latitude zone.

**Fig. 9.** Spatial distribution of the local spatial autocorrelation in NC.
land, bare rock, gobi, desert and saline-alkali land in the northern arid and semi-arid region and the Qinghai Tibet Plateau, accounting for 31.170% of the total area. The classes of EI in this region were mainly no interference (57.300%), light interference (14.516%) and low interference (24.500%). The main reason for the low level of interference was that the natural environment is fragile in these areas, and the areas are difficult to develop. However, if human activities are not strictly controlled, serious environmental problems such as salinization and desertification will occur. As this region in NC has an extremely fragile environment, ecological restoration projects should be strengthened, with an emphasis on protecting existing vegetation, and artificial windbreak and sand fixation forests should be constructed to prevent the further deterioration of environmental problems.

(2) **Environmental migration region (EMR).** The EMR encompassed “low-quality natural environment and high level of human interference” areas, which accounted for only 1.784% of the total area. These areas were mainly located in southern Ningxia, northern Shaanxi and Gansu. These areas are part of the interlaced farming-pastoral belt in the northern arid and semi-arid region and the Loess Plateau. The land use type was mainly cultivated land (79.156%), and the classes of EI were mainly high interference (75.903%) and extreme interference (18.877%). Overall, the interference effect of human activities was notable, and the natural environmental background was fragile, resulting in considerable pressure on the regional environment. In the future, we should continue to implement ecological restoration projects, take mandatory measures to return farmland to forests and grasslands with low yields, encourage the intensive production of agriculture and animal husbandry, and actively build green ecological barriers. Additionally, in regions with extremely fragile and sensitive environments, we should strengthen the implementation of ecological migration policies to fundamentally reduce the interference of human activities and alleviate regional environmental pressure.

(3) **Environmental conservation region (ECR).** The ECR encompassed “high-quality natural environment and low level of human interference” areas, accounting for 45.226% of the total area. These areas were less prevalent in the Huang-Huai-Hai Plain but were widely distributed in other regions. This region was mainly composed of forestland (26.507%) and grassland (49.124%), and the classes of EI were mainly no interference (16.115%), light interference (54.975%) and low interference (28.743%). Simultaneously, this region had good environmental background conditions (i.e., high vegetation coverage and abundant natural resources). Moreover, due to the rugged terrain, limited availability of transportation, sparse population and basic economy, the interference effect of human activities was low. To maintain the balance among ecosystems in NC, it is very important to focus on conservation in this region. For grassland resources, we should establish closed grassland breeding areas and strengthen the implementation of policies such as grazing bans, rest grazing and rotational grazing. In addition, for forest resources, which are known as the “lungs of the environment”, we need to strictly control deforestation and other practices and continue to protect and maintain the growth of forest areas. At the same time, at the junction of this region and other management regions, the implementation of ecological restoration projects should be strengthened, green barriers should be improved, and interference from human activities should be restrained.

(4) **Integrated development region (IDR).** The IDR region encompasses “high-quality natural environment and high level of human interference” areas, which were mainly distributed on the Huang-Huai-Hai Plain and Northeast China Plain, accounting for 21.820% of the total area. These areas were the most economically developed areas in NC, and they are also the main grain-producing regions and grain storage areas in China. The land use types were mainly cultivated land (75.438%) and construction land (9.689%). The classes of EI in this region were mainly medium interference (50.434%), high interference (33.172%) and extreme interference (11.990%) due to the frequent interference of various socio-economic activities (e.g., dense population and settlement, rapid economic development, high urbanization level, and developed road network). For such a region, the dual advantages of the natural environment and socio-economic development should be considered to promote green industrialization. Simultaneously, attention should be paid to controlling the development of various resources within the region, promoting the development of the ecological economy and a circular economy, and building sustainable cities based on a coordinated development model that considers the social, economic, and environmental sectors.

### 5.2. Performance assessment of the EI

The main contribution of this paper is to propose a quantitative EI assessment method to identify the interference of human activities in the natural environment. In the process of the continuous development of human society, the interference effects of various human activities will not only bring about a series of positive or negative impacts on the local environment (e.g., changes of the land cover, regional climate and biodiversity) but will also produce different degrees of radiation effects on the surrounding areas. When the interference intensity is within the controllable range of the regional environment, it will not only bring real economic benefits to human beings but will also maintain the balance between human society and the natural environment. However, when the degrees of interference increase without limit, the external pressure brought by human beings on the environment will become larger, and the environmental carrying capacity will keep approaching the critical point, which will lead to the collapse of the regional environment. This is not conducive to the survival and development of human beings.

One of the obvious advantages of our proposed EI is that it can identify the real interference effects caused by human activities in regions (i.e., it can identify the actual responses of different regional environments to human activities), such as sample regions A, B and C located in the northern arid and semi-arid region and sample regions H and I.
located in the northeast China plain (Fig. 5), which are both farming areas. Due to the spatial heterogeneity of regional environments, there are differences in the interference intensity of the same human activities. Our method can reflect this phenomenon well, but there are still some defects. The proposed method is not sensitive in identifying and monitoring EI in regions of high-intensity human activities (e.g., in Fig. 5, sample regions E and F located in the Huang-Huai-Hai Plain). This is the same principle indicating that the NDVI is less sensitive to high-vegetation areas in vegetation monitoring (Lamchin et al., 2018).

Throughout the process, the research work helps provide a decision-making basis for coordinating the contradiction between the regional environment and the development of human society by identifying, exploring, and monitoring the characteristics and regularities of the distribution of EI and exploring the internal causes.

In addition, from the scientificity and feasibility of the proposed assessment method, we aim to provide a simple and convenient system/framework based on standardized processes. The assessment results are calculated following the standard processing procedure, and strict control of each process is essential to ensure the accuracy of the final results (Liu et al., 2017). Such controls include:

- selecting indexes appropriately and comprehensively;
- improving the accuracy of data by using real and reliable data;
- performing data processing scrupulously and accurately; and
- minimizing the uncertainties of index fitting and weight calculation by objective methods.

According to the EI analysis, the primary task of calculating assessment results is to construct an effective assessment index system (Dale and Beyeler, 2002; Keersmaecker et al., 2015). Here, we screen in the data and feasibility of the proposed assessment method, we aim to provide a simple and convenient system/framework based on standardized processes. The assessment results are calculated following the standard processing procedure, and strict control of each process is essential to ensure the accuracy of the final results (Liu et al., 2017).

Such controls include:

- selecting indexes appropriately and comprehensively;
- improving the accuracy of data by using real and reliable data;
- performing data processing scrupulously and accurately; and
- minimizing the uncertainties of index fitting and weight calculation by objective methods.

According to the EI analysis, the primary task of calculating assessment results is to construct an effective assessment index system (Dale and Beyeler, 2002; Keersmaecker et al., 2015). Here, we screen in the data and feasibility of the proposed assessment method, we aim to provide a simple and convenient system/framework based on standardized processes. The assessment results are calculated following the standard processing procedure, and strict control of each process is essential to ensure the accuracy of the final results (Liu et al., 2017).

In the assessment process of EI, 19 indexes were selected from the perspectives of the natural environment and human interference. The problem of collinearity among the indexes was eliminated by using multicollinearity diagnosis to reduce the influence of data redundancy on the accuracy of the assessment results. Then, the SDM and REI-HAI ratio model were constructed using the EI assessment model and spatial analysis methods to calculate and summarize the distribution characteristics (e.g., spatial heterogeneity and dependence) and influencing mechanisms of EI in NC. The results can be used to propose potential control measures and policies.

In this study, the raster of 1 km spatial resolution was used as the basic assessment unit to eliminate the influence of subjective factors caused by the artificial division of units and to perform the related assessment of EI in an objective manner. Multisource data and geographic information system (GIS) technology were used to conduct comprehensive assessments and fill the research gap associated with quantitative simulations of interference with the natural environment caused by human activities. In addition, all the data involved in calculating the indexes were derived from remote-sensing data and socio-economic statistical data, which could facilitate the preparation of long-term series of EI monitoring data and relevant studies on EI in other regions. The constructed assessment model avoids the effects of subjective factors and yields objective and applicable results.

Declaration of competing interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

We gratefully acknowledge the support by the National Natural Science Foundation of China (grant numbers 41861040, 41761047 and 41501176), Natural Science Foundation of Gansu Province (grant number 1506RJZA129).

Appendix A. Calculation of assessment indexes

A.1. Calculation of assessment indexes from natural factors

A.1.1. Calculation of the normalized difference vegetation index (NDVI)

MODIS NDVI data (MODIS V3 products, MOD13A3) has a 1 km spatial resolution and a 1-month temporal resolution. In this paper, thirteen images, with orbit numbers h23v04, h23v05, h24v04, h24v05, h25v03, h25v04, h25v05, h26v03, h26v04, h26v05, h27v04, h27v05 and h28v05 were used to cover the entire study area. In total, 156 scenes from January to December were downloaded for the period of 2017. The pretreatment process is as follows. After projection transformation, image splicing, grid resampling, the averaging method and other operations, the 2017 NDVI data are processed into a 1-year time resolution and a 1 km spatial resolution (Liu et al., 2017; Ge et al., 2018). The results are shown in Fig. A.1a.

A.1.2. Calculation of erosion protection (EP)

EP is quantified according to vegetation types (Xu et al., 2019). According to different types of vegetation with different soil and water conservation capacities, the order from low to high is as follows: swamp, broad-leaved forest, coniferous forest, mixed coniferous broad-leaved forest, alpine vegetation and water area < cropland, grasslands,
meadow, thick growth of grass < scrubland < desert, rock and no-vegetation area, with assigned values of 1–4, respectively (Maestre et al., 2009). The spatial distribution of EP is shown in Fig. A.1b.

A.1.3. Calculation of the soil organic matter content (SOMC)

The SOMC is calculated according to the organic carbon content attribute of the Chinese soil data set obtained by Eq. (1-A). The spatial distribution of the SOMC is shown in Fig. A.1c.

\[
\text{SOMC} = \frac{T_{OC}}{0.58}
\]  

(1 - A)

Here, SOMC is the soil organic matter content (%) and \(T_{OC}\) is the percentage of organic carbon in the topsoil (%).

A.1.4. Calculation of the soil depth (SD)

The SD is quantified according to the depth information of different soil types. The order of the depth from low to high was as follows: aeolian sandy soil, stony soil, skeleton soil < gray desert soil, gray-brown desert soil, brown desert soil, red clay, new deposit soil, cracked soil, saline soil, desert saline soil, cold plain saline soil, alkaline soil, irrigated soil, irrigated desert soil, cold desert soil, northwest salt crust, urban area < brown calcium soil, gray calcium soil, yellowish brown soil, forest irrigation meadow soil, moist soil, paddy soil, grass felt soil, black felt soil, frigid calcic soil, cold calcic soil < brown coniferous forest soil, brown soil, dark brown forest soil, cinnamon soil, gray cinnamonic soil, black soil, gray forest soil, chernozems, bog soil, chestnut soil, peat soil, water area, with assigned values of 1–4, respectively (Xu et al., 2019). The spatial distribution of the SD is shown in Fig. A.1d.

A.1.5. Calculation of the soil erosion intensity (SEI)

The SEI is quantitatively obtained from soil erosion data (Fig. A.1e). Different soil erosion classes, including micro erosion, mild erosion, moderate erosion, intensity erosion, extreme erosion and severe erosion, are assigned values of 1–6, respectively.

A.1.6. Calculation of the elevation (ELE)

The ELE is generated directly from the DEM. As shown in Fig. A.1f, the elevation varies from -264 m to 7311 m.

A.1.7. Calculation of the slope degree (SDE) and slope aspect (SA)

The SDE is extracted from the DEM using the slope tool in ArcGIS 10.6, and the results are shown in Fig. A.1g. The value range of the SDE varies between 0° and 47.91°. The calculation process of the SA is as follows: first, the aspect tool in ArcGIS 10.6 is used to extract the slope aspect information based on DEM, and then the SA (Fig. A.1h) is obtained quantitatively according to different natural environments in different directions. The sequence from low to high is as follows: \(-1 < [0°, 45°], [315°, 360°] < [45°, 90°], [270°, 315°] < [90°, 135°], [225°, 270°] < [135°, 225°]\), with assigned values of 1–5, respectively.

A.1.8. Calculation of the annual precipitation (AP) and annual average temperature (AAT)

To obtain the AP and AAT, the influence of the elevation on precipitation and temperature is considered. The elevation data attached to 342 meteorological site data are used as auxiliary data, and the collocating Kriging technique in the Geo statistical wizard of ArcGIS 10.6 is used for interpolation (Croitoru et al., 2016; Cui et al., 2017; Shi et al., 2018). The interpolation results are smooth, and the spatial distribution is real and reliable. The spatial distributions of the AP and AAT are shown in Fig. A.1i and j.

A.1.9. Calculation of the river network density (RND)

The RND is calculated according to the ratio of the length of regional rivers to the regional area (Fig. A.1k). The calculation formula is Eq. (2-A).

\[
RND = \sum_{i} \frac{L_i}{A_i}
\]  

(2 - A)

Here, RND is the river network density (km/km²); \(L_i\) is the total length of rivers of the unit area (km); and \(A_i\) is the unit area (km²).

A.1.10. Calculation of the distance from lakes and reservoirs (DLR)

The DLR refers to the distance from lakes and reservoirs and indicates the extent of regional water resources (Wei, 2018). The DLR (Fig. A.1l) is calculated using Euclidean distance tools in ArcGIS 10.6.

A.2. Calculation of assessment indexes from human interference factors

A.2.1. Calculation of the population density (PD) and economic density (ED)

In this study, county-level units were used for statistical data. To facilitate analysis and mapping, China’s county-level administrative boundary data are used to convert population and economic data into a shapefile layer in ArcGIS 10.6. Then, by calculating the unit area of each county, the ratios of population and gross domestic product (GDP) data to the unit area of each county are used to obtain the PD and ED. The calculation formulas are Eqs. (3-A) and (4-A). To facilitate calculation with other indexes at the same spatial scale, the Kriging technique is used to interpolate the data of each county-level unit to obtain the spatial distribution data sets of the PD and ED (Wei et al., 2017), which are shown in Fig. A.2a and b.

\[
PD = \frac{POP}{A_c}
\]  

(3 - A)

\[
ED = \frac{GDP}{A_c}
\]  

(4 - A)

Here, PD is the population density (persons/km²); ED is the economic density (one hundred thousand yuan/km²); POP is the population (persons) of the unit area; GDP is the GDP (yuan) of the unit area; and \(A_c\) is the unit area of each county (km²).

A.2.2. Calculation of the settlement density (SED)

The SED refers to the quantity of settlements per unit area, and the calculation formula is Eq. (5-A). In this study, we use the kernel density estimation method to calculate the settlement density map (Fig. A.2c) in NC (the spatial resolution is 1 km).

\[
SED = \frac{n}{A_i}
\]  

(5 - A)

Here, SED is the settlement density (numbers/km²); \(n\) is the quantity of settlements (numbers) of the unit area; and \(A_i\) is the unit area (km²).

A.2.3. Calculation of the road network density (ROND) and railway network density (RAND)

The ROND and RAND show the extent of traffic construction in northern China, which refer to the length of roads/rails per unit area. The calculation formula is Eq. (6-A), and the spatial distribution maps of the ROND and RAND are shown in Fig. A.2d and e.

\[
RD = \frac{L_r}{A_i}
\]  

(6 - A)

Here, RD is the ROND or RAND (km/km²); \(L_r\) is the length of roads or railways (km) of the unit area; and \(A_i\) is the unit area (km²).

A.2.4. Calculation of the land use intensity (LUI)

The LUI reflects the degree of human interference in different land use patterns (Fig. A.2f). The inversion process of the LUI is as follows.
According to the theory that land use quantification is based on the limit of land use (Zhuang and Liu, 1997), the land use is defined as 4 levels of land use under the ideal state, and the 4 levels of land use are assigned. The order of the human interference degree from low to high is as follows: unutilized land, unable land, woodland, grassland, water area, cultivated land, garden land, artificial grassland, towns, habitation, industrial and mineral land, traffic land, with assigned values of 1–4, respectively.

A.2.5. Calculation of the night-time light intensity (NTL)

The NTL has been widely used in urban studies due to its low-light sensing characteristics at night without moonlight (Elvidge et al., 1999; Small and Elvidge, 2011). The night-time light data set herein was cloud-free NPP VIIRS data. The process is as follows. NPP VIIRS data is downloaded from https://www.ngdc.noaa.gov/eog/viirs. The spatial resolution is 0.00417°, and through projection transformation, grid resampling, and clipping, the NPP VIIRS monthly data set (2017) of the study area is obtained. Here, we only use March data, because this data is only available in monthly composites, with less cloud cover. March is the month with the fewest holidays in China, which not only reduces the migration in and out but also ensures relatively stable illumination. Meanwhile, other months have data loss problems in Xinjiang, Heilongjiang, Inner Mongolia and Jilin, and some months have image quality problems (Ma et al., 2014; Chen et al., 2019).
NPP VIIRS original data set is filtered to exclude data impacted by stray light, lightning, lunar illumination, and cloud-cover. In this study, the threshold method is used to remove outliers by forcing values \( N > 219.601 \) and lower than 0.191 to 219.601 and 0.191, respectively (Chen et al., 2019). The lake with the least artificial interference is selected as the sample area for threshold setting, Qinghai Lake is the largest natural lake in the study area, and the mean pixel value (0.191) of the lake area is taken as the minimum value of night-time light brightness in the study area. At the same time, the maximum value of night light data in the central first-tier cities in NC (Beijing) is selected as the maximum value of night light brightness in the whole research area (Ma et al., 2014). Fig. A.2g shows the revised results of the NPP VIIRS data (2017) in NC.

References


