Causal complexity of environmental pollution in China: A province-level fuzzy-set qualitative comparative analysis

Yang Chen (992903250@qq.com)
Chongqing University
Jingke Hong
Chongqing University
Miaohan Tang
Chongqing University
Yuxi Zheng
Chongqing University
Maoyue Qiu
Chongqing University
Danfei Ni
Xiamen University

Research Article

Keywords: Fuzzy set, configuration, causal complexity, environmental pollution, environmental governance, QCA

Posted Date: April 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1529869/v1

Abstract

Environmental problems are endowed with the causal complexity of multiple factors. Traditional quantitative research on the influencing mechanism of environmental pollution has tended to focus on the marginal effects of specific influencing factors but generally neglected the multiple interaction effects between factors (especially three or more). Based on the panel data of 30 Chinese provinces between 2011 and 2017, this study employs fuzzy set qualitative comparative analysis (fsQCA) - which can provide a fine-grained insight into causal complexity of environmental issues - to shed light on the influencing mechanism of environmental pollution. The results show that there are several different configurations of pollution drivers which lead to high pollution or low pollution in provinces, confirming the multiple causality, causal asymmetry, and equifinality of environmental pollution. Further, the combination effect of advanced industrial structure, small population size, and technological advance is significant in achieving a state of green environment compared to environmental regulation factors. In addition, spatiotemporal analysis of the configurations indicates that strong path dependencies and spatial agglomeration exist in current local environmental governance patterns. Finally, according to our findings, targeted policy recommendations are provided.

1. Introduction

The rapid development of the Chinese economy has brought about serious environmental consequences (Huang et al., 2020). During the period from 2010 to 2015, China contributed nearly 20% of global emissions of nitrogen oxides (NO_x) and 30% of sulfur dioxide (SO_2) (Zhang et al., 2018a). In 2016, the concentrations of $PM_{2.5}$ in three-quarters of monitored cities in China were below the national grade II standard (≤35μg/m^3) and the WHO standard (≤10μg/m^3) (MEP, 2017). Environmental harm poses a major threat to sustainable development, and currently attracts extensive attention from the Chinese government (Guan et al., 2014). Moreover, the multifactorial complexity of environmental problems poses a great challenge to the traditional instruments of environmental governance patterns (Tan and Fan, 2019). From a panoramic point of view, therefore, identifying and measuring multiple synergy effects induced by the environmental complexity will be of great significance to the environmental governance.

A considerable amount of existent research has attempted to analyze the main factors that influence environmental pollution from a socioeconomic perspective. Ehrlich and Holdren (1971) firstly attributed the anthropocentric impact (I) on the environment to population growth (P), affluence development (A), and technological progress (T) in his IPAT model. Specifically, population dynamics have been at the center of arguments pertaining to environmental deterioration (Pham et al., 2020). Based on neoclassical growth theory, a rapid expansion of population, the consumption of limited natural resources by the existing population, and a compound of both could give rise to environmental pollution from the consumption side (Li et al., 2019). Similarly, the intertwined connections between economic activities and environmental impacts are undeniably significant (Guan et al., 2014). Economic growth undoubtedly stimulates demand for natural resource extraction and consumption, and leads to environmental unsustainability (Krueger, 1995). In addition, technological factors play a crucial part in maintaining or altering the balance between population, economy, and the environment. Unlike population growth and economic development, technological advances through declining natural resource consumption per unit output seem to have positive effects on environmental sustainability (Pham et al., 2020).

A further strand of literature has focused on the environmental impacts of other socioeconomic factors, including industrial structure (Zheng, 2020), urbanization (Zhang, 2018), industrial agglomeration (Shen and Peng, 2021), foreign direct investment (FDI) (Cheng et al., 2020), foreign trade (Chen et al., 2019), and energy prices (Li et al., 2020). For instance, Zheng (2020) argued that industrial structure determines the allocation of production factors (such as capital, labor, technology, and energy) among different sectors, and this, consequently, significantly affects resource consumption and pollutant emissions. As an important bond between environment and economy, the industrial structure is an indispensable element to realize integrative development. Zhang (2018) proved that urbanization had a significantly positive spatial spillover effect on CO_2 emissions in China by utilizing the Spatial Durbin Panel model. Shen and Peng (2021) conducted a spatial panel analysis of China's environmental efficiency and found an apparent U-curved relationship between industrial agglomeration and environmental efficiency. Of further note is the work of Cheng et al. (2020), who used the generalized moments method (GMM) to explore the influence of FDI on the environment based on the panel data of 285 Chinese cities; they asserted that FDI significantly intensified China's urban PM2.5 pollution.

Aside from socioeconomic factors, governmental intervention, which holds a significant role through institutional and regulatory aspects, has also attracted considerable attention in the environmental fields. On the one hand, the government has attempted to facilitate corporate green innovation through carrot-and-stick policies including assistances that are in the shape of green R&D subsidies (Bai, 2019), tax preferences for low-emission and high-tech enterprises (Zheng and Shi, 2017), and punitive taxes imposed upon technologies or actions that are environmentally undesirable (Hunt and Fund, 2016). These stimulus policies have contributed to pushing enterprises to achieve "innovation compensation" by reducing compliance costs and promoting production efficiency. For example, Bai (2019) argued that government R&D subsidies stimulated green innovations of energy-intensive firms. On the other hand, the government endeavors to control and eliminate environmental pollution by means of environmental regulations, including source-oriented treatments (Laplante and Rilstone, 2004) and end-of-pipe based treatments (Wang, 2019). For instance, Zhao et al. (2020) employed the GMM estimation method to investigate the effects of environmental regulation on greenhouse gas emissions. They found that apart from the direct effect on CO_2 emission reduction, environmental regulations also indirectly restrain CO_2 emissions by adjusting the structure of energy consumption.

However, it is difficult to reach a consensus about the influence of driving factors on environmental pollution in the existing literature. For instance, some researchers have revealed a positive relationship between population and energy-related pollution (Li et al., 2019; Liddle and Lung, 2010) while others have reported that population impact on the environment is an inverted U-shaped curve (Zhang et al., 2018b). Similarly, some studies have argued that economic growth tends to increase pollutant emissions (Dong et al., 2018), whereas substantial evidence also states that economic growth could arise from structural transformation and advance production technology, and that these factors may thence offset the negative effects of growing economic activities on the environment (Krueger, 1995). We summarize the divergent views of several representative factors and their impacts on pollutant emissions (See Table 1).

It can be observed that the effect of factors on pollution emission are diverse and may even be mutually contradictory. The failure of the adopted symmetrical analytical instruments to describe the practical asymmetric causal relationship may be a cause of the inconsistent findings. This mismatch of analytical instruments may lead to the attributes proved to be causally related in one situation to be unrelated or adversely related in another situation (Meyer et al., 1993). In detail, the estimation of the marginal effect in quantitative methods may neglect some samples with weak significance while valuing the samples with large variances. For this reason, at the individual level, cases opposite to the observed net effects often appeared, that is, not every case in the sample supported a fixed relationship between the dependent and independent factors (Woodside, 2013).

In addition, few studies have investigated the interaction effects and combined impacts of multiple factors (especially three or more) on environmental pollution at present. Far from being mutually exclusive, the pollution drivers not only coexist, but also prominently impact one another's operation because of causal complexity (a consequence of the mix of both the individual marginal effects and conjunction impacts induced by multiple causes). Traditional quantitative approaches which aim to examine marginal effects - such as multiple regression analysis - may, to some extent, interpret the multiple conjunctural causation between several variables. However, capturing interaction effects for in excess of three variables is arduous (Woodside, 2013). Moreover, these methods are weak in their ability to both handle the causal complexity from a holistic level and uncover individual heterogeneity observed in reality. For these reasons, this paper attempts to utilize the fsQCA method, which focuses on correlations between combinations of factors and the outcome, also making explicit the impact of the context and the interaction effects between factors, to overcome these limitations.

Based on configurational theory, the QCA method is a set-theoretic approach by applying Boolean algebra to explore the combinations of organizational attributes leading to the outcome at issue (Ragin, 2000). This method aims to combine approaches from quantitative and qualitative techniques; taking the best attributes from both (Pappas and Woodside, 2021). Although there are other comparative evaluation approaches in the configurational theory such as cluster analysis (Lim et al., 2006), deviation scores (Delery and Doty, 1996), and the interaction effects method (Dess et al., 1997), the QCA method is superior in grasping causal complexity at a fine-grained degree and can enable scholars to glean statuses of equifinality, substitution, and complementary effects among variables (Greckhamer et al., 2018). In brief, QCA's advantages in employing systematic comparisons over cases enable researchers to understand complex interactions across multiple causal situations (Lobe, 2010).To date, QCA has witnessed extensive use in different fields such as business strategies (Douglas et al., 2020), information systems (Park et al., 2020), and social networks (Rutten, 2020). However, it is less common in research upon environmental pollution. The method will be suitable for the study of environmental governance considering that environmental problems are endowed with the attribute of multi-factor causal complexity.

In conclusion, there are some deficiencies in existent understanding. First, it is difficult to reach a consensus about the influence of driving factors on environmental pollution from a single factor perspective. Second, traditional quantitative analysis instruments, such as regression analysis, have advantages in estimating the net effect of a single factor on the outcome ceteris paribus, whereas are difficult to elaborate multiple interaction effects (more than three factors) considering the complicated statistical interpretations and the multicollinearity problems. Third, fsQCA method, which is suitable for handling the causal complexity of multiple factors, is seldom used in environmental pollution research. To fill these gaps, this paper adopts fsQCA to investigate the casual complexity of environmental pollution at the provincial level and provides environmental improvement paths for high-pollution provinces. This study contributes to extant works in the following ways: 1) By introducing the fsQCA method creatively, this article assesses the multiple causation and asymmetric causality of environmental pollution at the individual level, which fills the gaps of previous studies on influencing mechanism of pollution factors. 2) We compare the empirical results of econometric models and fsQCA method, and thus offer a fine-grained and comprehensive insight into the interactive mechanism of environmental drivers; 3) This study depicts a strategic map for green development by considering the spatiotemporal characteristics and environmental configuration changes, and provides the corresponding improvement paths for high-pollution provinces.

In the following section, we illustrate the specification of the fsQCA method. Then, we report the results and undertake a discussion of the implications of our findings for policymakers. The final part of the paper presents the major conclusions.

2. Method And Data

There are three main variations: crisp-set QCA (csQCA), multi-value QCA (mvQCA), and fuzzy-set QCA (fsQCA). CsQCA is suitable to handle complex sets of binary data (Ragin, 1987), while mvQCA, which regards variables as multivalued rather than dichotomous, is an extension of csQCA. Both csQCA and mvQCA require their selected data to be classified according to explicit distinguish criteria; as a result, it is hard to grasp complexity in cases that naturally change by level or degree (Rihoux and Ragin). FsQCA integrates fuzzy-sets and fuzzy-logic manners to break through this limitation; it offers a more fine-grained insight into data by providing a more realistic approach in which variables can capture all values from 0 to 1. Therefore, fsQCA was applied to our research subject as our variables had no clear classification criteria. The basic steps in fsQCA method are shown in Fig.1.

2.1. Selection of variables

The first step in performing fsQCA analysis is specifying the configural model; identifying what antecedent conditions should be involved in estimation accounting for the outcome (Douglas et al., 2020). On the basis of IPAT model (population, economy, and technology), two indispensable dimensions of industry and government are extended in our research framework. Within these aspects, technological factors include technological innovation capacity and R&D subsidy; governmental factors containing the environmental regulations on source treatment and end-of-pipe treatment; economic, demographic and industrial factors include economic growth, population scale, and industry structure respectively. Within this framework, the configuration of seven environmental driving factors that may lead to high or low environmental pollution is shown in Fig.2.

This paper used panel data from 30 provinces in China between 2011 and 2017. In view of the policy effects of the Five-Year Plan (2011-2015), we divided the time spans into two sections (2011-2015 and 2016-2017) to provide a time-variant perspective. To avoid abnormal values in some provinces during a specific

year, we calculated the average value of 2011-2015 and 2016-2017 respectively. In addition, the price took the deflator coefficient into account and 2011 was considered as the base period (year 2011=100). The detailed data source of variables is presented in Table 2.

Specifically, industrial pollution was adopted to represent environmental pollution, and includes, given available data, industrial wastewater and industrial waste gas emissions (e.g., sulfur dioxide, smoke, dust, and nitrogen oxides). To circumvent subjective bias, the entropy method was used to calculate the comprehensive index of environmental pollution as this method maximizes the overall situation of pollution (Jianqin et al., 2010).

The indicators representing technological innovation were R&D expenditure (Shen et al., 2021), the number of patents (Linares et al., 2019), the number of researchers (Wen et al., 2020), full-time equivalents (Li et al., 2018), and new product sales (Bruno et al., 2006). Similarly, we used the entropy method to depict the overall picture of technological innovation ability.

2.2. Data calibration

Following the stages outlined above, the original data should be calibrated to fuzzy-set membership scores (range from 0 to 1) that represent the membership of a variable (Ragin, 2008). For example, 1 means the high level or presence of a defined set, whereas 0 means it is low level or absent. Based on Fiss (2011), we used the first quartile, the third quartile, and their average as the three qualitative anchors of fully out, fully in, and crossover point respectively. To these, we then applied the direct calibration method in the fsQCA3.0 to transform the data into fuzzy-set memberships. Table 3 summarizes the descriptive statistics and the calibration thresholds of the variables.

2.3. Necessity and sufficiency analyses

In the following, a truth table, which is a data matrix, should be constructed to provide all logically possible configurations of variables in 2^k rows (k=number of variables), where each row represents a specific configuration. Based on the memberships in the fuzzy sets, every case in our study is associated with a row of the truth table. To identify whether a variable was necessary or sufficient for an outcome, we then analyzed whether the conditions were always present (or absent) in each case when the outcome was present. It follows, that if an innovative advantage is necessary to lead to low pollution in all regions, low pollution will not happen if one region lacks such an advantage. Likewise, if an advantage in technological innovation is a sufficient condition that leads to low pollution, all regions with such an advantage will have low pollution.

The interpretive tools for both the necessary and sufficient conditions are consistency and coverage. Consistency establishes the extent to which the cases that share a configuration of variables agree in their outcome and is analogous to a correlation, while coverage displays the proportion of cases representing an outcome in a certain configuration and is comparable with the coefficient of determination (e.g., R^2).

The consistency and coverage of configuration *i* is

$$Consistency_i = \sum_{r=1}^{30} \min(X_{i,r}, Y_{i,r}) \Big/ \sum_{r=1}^{30} X_{i,r} \qquad (1)$$

$$Coverage_i = \sum_{r=1}^{30} \min(X_{i,r}, Y_{i,r}) \Big/ \sum_{r=1}^{30} Y_{i,r} \qquad (2)$$

Where $X_{i,r}$ and $Y_{i,r}$ is the membership of region *r* in the set of solution *i* and outcome, respectively.

To identify the necessary condition, the necessity analyses of all conditions and their negation were conducted with a consistency criterion of ≥0.9 (Wagemann, 2012). Thereafter, sufficiency analyses were performed using the truth table algorithm to recognize configurations that were constantly related to an outcome. To avoid “simultaneous subset” relations of configurations in both the outcome and its absence, the raw consistency benchmark of sufficiency analysis must be more than 0.8 accompanied by a benchmark for PRI (proportional reduction in inconsistency) score of over 0.65 – the higher the value, the more robust the solution (Misangyi and Acharya, 2014). In our study, we set the thresholds for raw consistency and PRI as 0.8 and 0.75, respectively. Then we set the frequency threshold of one strong case for a configuration’s inclusion to ensure 100% of the studied sample in the sufficiency analysis (≥80% is recommended proportion) (Greckhamer and Gur, 2021).

In the next step, and based on the truth table algorithm, the truth table rows should be logically simplified to classify causal conditions into core and peripheral conditions. In this progress, there may be no practical cases of any particular configuration, which is a common problem named “limited diversity”. For this, counterfactual analysis based on theoretical and substantive knowledge supports solving the limitations, and obtains parsimonious, intermediate, and complex solutions (Wagemann, 2012). The core conditions appear in parsimonious solutions, while the peripheral conditions appear in the intermediate or complex solutions. In general, core conditions are more convincing than peripheral conditions; the latter are relatively complementary. As a result, we placed interpretative emphasis on parsimonious and complex solutions.

2.4. Robustness analyses

Wagemann (2012) proposes two set-theoretic-method-specific dimensions of robustness: the set-relational states of the different principles and the variations in the coefficients of fit. Where different decisions bring about different solution terms but retain a subset relation between each other, the results can be interpreted as robust. In like manner, where different decisions result in differences in the coefficients of fit that are too insignificant to ensure a meaningfully different substantive interpretation, the results can also be considered robust. Given this, the calibration threshold of the crossover point in this paper was changed from the average value of the first and third quartile to the median (Greckhamer and Gur, 2021). Thereafter, we recalculated the explained variable (environmental pollution) by adding the production of solid waste. Next, we improved the threshold of raw consistency from 0.8 to 0.9, and the PRI from 0.75

to 0.85. Finally, predictive validity was performed to test the accuracy of the models in the first time window (2011-2015) using the data for the second time window (2016-2017).

It should be noted that the first robustness check with the alternative calibration may transform an empirically observed row into a logical remainder (or vice versa), or from a consistent row into an inconsistent one (or vice versa), drawing different conclusions (Wagemann, 2012). However, minor changes are observed in the results (including the specific number of solutions, solution consistency, solution coverage, and the characters of configurations), that is to say, the explanations of the main conclusions remain substantively unchanged (see Supplementary Table S2-S3). As for the second and third robust tests, we found that substitution of the explained variable and tighter threshold requirements produced substantially similar solutions with expected minor changes including subset relations of configurations, solution consistency, and solution coverage (see Supplementary Table S4-S7). At last, the finding of the predictive validity test shows that the configurations between 2011-2015 and 2016-2017 are highly consistent (see Table 6). In short, a series of robustness analyses confirmed the credibility of the results presented in this paper.

3. Results

3.1. The regression analysis

To demonstrate the potential contribution of fsQCA in understanding environmental pollution compared to the quantitative analysis, we first conducted the traditional regression models, considering including all the same antecedent conditions and three technology-related interaction terms as independent variables. As Table 4 shown, models 1-5 introduce different interaction terms when controlling individual fixed effect and time fixed effect. Several findings are summarized below.

First, for models 2-5, the coefficients of technology-related interaction terms are significant, and the significance and value of interaction terms increase when considering more interaction terms. The finding indicates that interaction effects between environmental drivers exist significantly;

Second, by adding more interaction terms, the number of significant parameters and the goodness of fit (R^2) increase from model 1 to model 5, which implies that the inclusion of interaction terms enhance the explanatory power of models;

Third, the parameter signs of technology between model 2 and 5 are in opposite directions when three technology-related interaction terms are incorporated into model 5, which illustrates that the impact of technology on environmental pollution is asymmetrical due to the existence of interaction effects;

Fourth, the parameters of other variables without corresponding interaction terms are insignificant. Combined with the conclusions above, it can be inferred that the effect of a single factor is not significant as that of a combination of multiple factors.

To sum up, the regression analysis confirmed the existence of interaction effects between pollution drivers. However, the construction of regression models faces a dilemma. On the one hand, adding too many interaction terms into models will make results complicated and redundant. For example, the economic implications of multiple interaction terms (especially for over three factors) will be difficult to interpret. More seriously, multiple interaction terms will result in serious multicollinearity problems in the model, which then leads to a biased estimation; On the other hand, the absence of interaction terms will be inconsistent with reality and lose partial explanatory power of the model.

Therefore, we then performed fsQCA analysis that can reveal the multiple interaction effects of environmental drivers while avoiding multicollinearity problems.

We firstly identified the necessary condition of high pollution or low pollution. According to the results (see Supplementary Table S1), no variable strictly meets the criteria of necessary conditions. This finding echoes the theory of complementarities that no organizational elements are best practices alone but will affect positively only when they occur in conjunction with other elements.

3.1. Sufficiency analyses between 2011 and 2015

During the 12th Five-Year Plan Period (2011-2015), there were 10 configurations that led to either high pollution or low pollution in regions, as Table 5 presents. These configurations illustrated that there were varied strategic paths that led to equifinal outcomes, and this in turn, verifies the existence of multiple causal relationships in environmental issues. Further, these 10 pathways can be grouped into five distinct pairs of neutral permutations (C1-C5). Pathways in each pair represented the same core conditions and only varied in their complementary conditions.

The solution coverages of high pollution and low pollution were 0.587 and 0.755, respectively, which exhibits a strong explanatory power, whilst all the configurations maintained very high consistencies (0.977 in high pollution, and 0.962 in low pollution); suggesting that these configurations are persuasive for the outcomes.

3.1.1. Configurations for high pollution

There were five configurations (C1a-C3) that illustrated the possible causal relationships that led to high pollution between 2011 and 2015. It is worth noting that the first four configurations (C1a-C2) contained the same core condition of possessing a backward industrial structure; this illustrates that structural imbalance was the leading factor causing high pollution in the involved regions. Therefore, we label these four configurations as *structural imbalance* type.

Specifically, configuration 1a and 1b (C1a and C1b) featured technical lag, small R&D subsidies, large end-of-pipe treatment costs, small populations, and backward industrial structures. These features signified that even though local government spent a large amount of money on end-of-pipe treatment,

backward technological development and industrial structure were still harmful to the environment. In addition, the features of large source treatment costs and backward industrial structure in C2 further revealed that the source treatment measure was ineffective for mitigating environmental burden when the industrial structure was backward. In other words, no matter how much the government had spent on environmental regulation, a backward industrial structure hindered environmental improvements to a greater extent.

In C3, the core conditions included both intensive cost on end-of-pipe treatment and large populations, with the peripheral conditions including advanced innovation ability, strict environmental regulations, and the possession of highly developed economies. The representative regions of C3 are Guangdong and Jiangsu, both are well-developed and densely populated. Specifically, Guangdong and Jiangsu have topped China's provinces for the past decades in terms of their recorded levels of GDP. These regions have coupled their possession of massive natural resources with rapid economic and social development. Apart from this, the growth polar effect that has arisen as a consequence of economic agglomeration has undoubtedly attracted the inward migration of people from surrounding regions. Rapid population growth and the possession of a large population will, therefore, not only lead to population agglomeration, but also accelerate the consumption of limited resources, and bring about enormous population pressures. In turn, these effects stimulate further, economic-social activities and may give rise to either predatory or disruptive use of resources. Increasing populations also give rise to huge levels of consumptive pollution (Ehrlich and Holdren, 1971). Given these assorted facets, we labelled C3 as the *extensive population* type.

3.1.2. Configurations for low pollution

According to Table 5, there were five alternative configurations (C4a-C5b) that led to low pollution. C4a-C4c shared the same core conditions: advanced industrial structure and low inputs for source treatment. In addition, they exhibited a comparatively ideal path for pollution governance in which local government paid more attention to structural optimization than the cost of pollution treatments. Consequently, this configuration was labelled as *green development* type.

Specifically, C4a featured smaller populations, advanced innovation abilities, substantial R&D support, advanced industrial structures, and low source treatment costs. C4a included two municipalities, Beijing and Shanghai; these two cities have realized win-win situations that have balanced economic development with environmental protection.

C4b and C4c were inferior in terms of technological innovations, governmental regulations, and economic development, but were superior with regard to their industrial structures. The typical cases in this configuration included Yunnan and Heilongjiang, where tertiary industries make up more than half of the region's GDP. Yunnan, for example, records that its tourist industry accounted for 51.5% of its GDP in 2020. To respond to the national strategy "clear water and green mountains are as valuable as mountains of gold and silver", the local government attempted, in addition to continuing to promote the transformation and upgrading of its tourism and cultural industry, to adopt a series of ecological measures, including developing green finance, implementing coal substitution, and increasing forest carbon sinks; all of which are beneficial to maintaining a low-pollution status.

C5a and C5b both possessed core conditions of low inputs for end-of-pipe treatment, less developed economies, and small populations with the peripheric condition of low source treatment costs. These configurations indicated that these less developed areas could ensure low pollution whilst expending (comparatively) less on environmental governance. One explanation for this is that a small population size alleviates the contradiction between humans and nature; i.e. consumption-based pollution is reduced. Therefore, we labelled these pathways as the *scarce population* type.

3.2. Spatiotemporal variations of configurations

So that we could elaborate further on the evolutionary patterns of configurations, we further studied the sufficient conditions between 2016 and 2017. As evidenced in Table 6, we found seven configurations that led to high pollution whilst five configurations resulted in low pollution. It can be observed that the solution coverage and the solution consistency of these configurations were high; indicating a strong explanatory power for outcomes. The diverse pathways indicated that multiple solutions existed for achieving the equifinality of outcomes. In addition, several impressive findings were obtained by comparing the configurations in two time spans from temporal and spatial perspectives.

Path dependencies existed in regional development patterns. In the period between 2016 and 2017, the levels of the conditions (e.g., pollution levels, environmental regulation inputs, and others) in most configurations were parallel with the former time span (2011-2015); indicating that there were strong path dependencies in most provinces where few changes had been undertaken about their development patterns. For instance, for high polluted regions, the states of seven conditions in C7b, C8a, and C8c were the same as C2, C1c, and C3, respectively. Similarly, for low polluted regions, C9a, C9b, C9c, and C10a were similar to C4a, C4b, C4c, and C5a, separately. Therefore, we label these paths as being the same as for the previous period (see Table 6). To further explore the extent of path dependency, predictive validity (using the second data set from 2016-2017 to compute the fuzzy scores for each of the ten configurations in Table 5) was performed as presented in Table 7. Taking C1a for example, the second data set is largely consistent (98.2%) with the argument that C1a is a subset of high pollution, and C1a accounts for 3.9% of the total memberships in high pollution. It is found that the consistency in most of the configurations exceeds 0.95, which indicates that the configurations between 2011-2015 and 2016-2017 are highly consistent. Further, the high raw coverage of each configuration, especially for C3 (0.360), C4b (0.304), C5a (0.469), confirms the strong path dependencies in regional development patterns.

There was a crowding-out effect in Shaanxi province. During the whole of the investigated period, Shaanxi spent hugely on R&D subsidies, but its technological innovation capacity changed from high (C5b in 2011-2015) to low (C10b in 2016-2017). In other words, governmental R&D subsidies failed to achieve the desired effect and instead eliminated regional technological innovation. This suggests that the crowding-out effect was more dominant in local environmental governance. Specifically, corporate R&D strategies stressed short-term benefits while the subsidies offered by the local governments sought to achieve long-term technical progress. Such a conflict in the direction of R&D initiatives weakened the driving forces of R&D investments. It should also be

noted that China currently faces challenges in supervising governmental R&D investments and that this may result in the misuse of R&D subsidies. This regulatory defect can be seen to lead to inefficient capital utilization.

To intuitively investigate the spatial distribution of configuration types and their transformations over time, this study applied the regional pollution labelling to a map of China (see Fig.3). During the two periods, most regions in the involved types exhibited geographical adjacency. As Fig.3 illustrates, most of the eastern regions belonged to the high-polluted group (e.g., *extensive population* type) whilst western areas were mainly belonging to low-pollution clustering (e.g., *green development* type).

It may also be noted that the number of regions with *extensive population* and *technical laggard* types has grown and that most of them changed from the *structural imbalance* type. This transformation may be a consequence of continuous improvements to the quality and efficiency of supply-side structural reforms, a consequence of the constraints induced by the imbalanced structure being weakened, and technological and demographic factors becoming the main drivers of environmental pollution. It can be noted, for instance, that the technical lag factor became the main driver that led to high environmental pollution in northern regions while the possession of a large population base identified as the driving factor leading to severe pollution in central areas.

During the periods investigated, the large inputs of end-of-pipe treatments and the possession of backward industrial structures were common characteristics shared by most high-polluted regions. In contrast, the low costs expended on source treatment and end-of-pipe treatment, the possession of smaller populations, and less developed economies were common features for most low-pollution regions. To a great extent, it contributes to their unique geographical advantages (e.g., climate, terrain, and vegetation resources) on the ecological environment and their original status of low pollution level (e.g., Yunnan). Further, local governments are confronted with multiple tasks from the central government, including economic growth and environmental protection. Thereinto, environmental targets are obligatory in China's performance evaluation system, but there are no incentives for local government to surpass these targets (Zhang, 2020). In other words, the local governments in low-pollution areas just needed to input small environmental treatment costs to surpass the cut-off score of environmental requirements decided by the central government.

4. Discussion And Implications

4.1. Implications from fsQCA

An extensive number of studies have explored the net effects induced by single factors on environmental issues from a symmetric perspective. These works have tended to construct a linear or curvilinear relationship among theoretical elements of interest, and have failed to explore practical asymmetric causal relationships. However, the fsQCA method enables heterogeneity to be revealed; something that traditional symmetric analytical methods challenge to manage. The approach also results in a more fine-grained taxonomy of development types. This study explores what conditions of configurations are relevant for pollution and how these conditions unite to work. The fsQCA method used in this paper provided new insights into these situations by developing a multifaceted comprehension of the conditions' dynamics. The results exposed subtle details of heterogeneity among the regions of China and identified sub-groups for which various pathways lead to the same outcome. In addition, it was shown that the configurations of high pollution and low pollution disclosed multiple causal relationships of variables, which demonstrates that the factors leading to environmental pollution at a provincial level are complex and asymmetrical.

The causal complexity of environmental pollution in our findings provides direct implications for local government when it comes to their addressing of environmental issues. For example, whilst it is widely accepted that technological progress aids environmental protection (Pham et al., 2020), we found that, for some regions (e.g., Henan, Hebei, and Anhui) of the *structural imbalance* type, their backward industrial structures still caused high pollution despite their possession of high-level technological innovation capabilities. Therefore, the priority of local governments of the *structural imbalance* type is to optimize their industrial structures.

This study also confirmed that the causal asymmetry derived from causal complexity may resolve a long-standing dispute about whether the relationship between environmental pollution and its driving factor is positive or negative. For example, the effect of economic growth on environmental quality in *extensive population* type is negative while in *green development* type it is positive. This result is exactly because of the conjunction of the multiple conditions (e.g., population, technology, economy, and governmental intervention), instead of being only (as advanced in previous studies) a consequence of economic growth. Given this, our findings can inspire future research to consider the causal complexity of environmental issues.

4.2. Path dependence

According to our results, regional environmental governance presented strong path dependence during the investigated period. As Fiss (2011) argues, configurations and types appear to impact future configural states by influencing the tracks of subsequent development models, thereby making certain tracks more possible while decreasing the probability of others. In the present study, most regions showed strong path dependencies in their development models; especially high-polluted regions. This leads to a significant issue; namely, why do these regions become locked into development models that lack dynamism, whilst other (rare) regions evade the lock-in effect and renovate themselves via consecutive new pathways. The lock-in effect works on a self-reinforcing logic that prefers continuity and replication (David, 1985). Specifically, the first-mover advantage of one built-in development pattern decreases operating costs through the scale effect, and the popularity of such patterns further leads to the improvement of the learning effect. Then, the synergetic effect of both contributes to achieving a virtuous circle of self-reinforcing, thus maintaining the original pattern in a state of persistence or lock-in, unless with the help of exogenous shock (e.g., policy reform). It follows, that the originally reasonable pattern and correcting errors in time are crucial to environmental governance. The local government needs to consider the long-term impacts of policy implementation instead of just the short-term effects. In addition, the government should take corrective actions as soon as possible once deviation between the practical effects of reform and its goal is detected.

4.3. Crowding-out effect

In our findings, the crowding-out effect that occurred in some regions, such as Shaanxi province, showed that government subsidies for technological innovation reduced or eliminated the improvement of technological performances. This phenomenon may have occurred because government R&D contracts were designed to bring social benefits or long-run efficacy, whereas grantees such as private firms tend to pursue economic interests or improvements to short-run performance. The divergence of original intentions between the two sides might crowd out the corporate R&D investment of the private and originally planned research agenda. Another possible explanation from a firm's point of view is that R&D subsidies released possible liquidity constraint—as a cheaper cost to apply for government subsidies than to raise funds in the capital market, enterprises considered the innovation subsidy an alternative source of financing instead of an actual R&D incentive. It follows that the government should lay greater stress on the improvement of the regulatory regime for firms' capital flows.

4.4. Policy implications for high-polluted regions

In our results, C9a (*green development* type) represented the best practice for achieving the dual targets of environmental protection and economic development. Its core conditions included possessing an advanced industrial structure, high technological innovation capacity, and a relatively small population. This is worth learning for high-polluted regions which were divided into *structural imbalance*, *extensive population*, and *technical laggard* types. Fig.4 shows the advised paths to achieve a green environment for these high-polluted types.

Possession of a backward industrial structure was a common and significant condition that led to high pollution in these polluted types. To address this, local governments should first optimize industrial structure. For instance, Beijing and Shanghai (*green development* type) which were low-pollution regions, began their industrial transitions alongside declining dependence on pollution-extensive sectors. Their low-emission industries (e.g., telecommunication equipment and transport equipment manufacturing) dominate their secondary industry sectors, and their tertiary industries have become pillars of their sustainable economies. In direct contrast, the pillar industries of Shanxi (*structural imbalance* type), which is the high-polluted region, are predominantly concentrated in high-emission fields, such as the smelting and pressing of ferrous metals, petroleum processing, and coke refining. Within such provinces, well-designed and well-enforced industrial policies should be established to accelerate the independent elimination of pollution-intensive enterprises and the emergence of eco-friendly industries; a process that will, in turn, minimize the negative by-product pollutants of economic development. In addition, inter-provincial cooperation between the provinces should be encouraged.

For high-polluted regions of the *technical laggard* type (including Shanxi and Inner Mongolia), poor R&D subsidies and technological innovation abilities, and high end-of-pipe treatment inputs indicated that their local governments tended to solve issues pertaining to current pollution rather than improving technological innovation to ensure longer-term benefits. To address this deficiency and the problems associated with short-term localized vanity projects (and the problems that they may create for successor administrations), the central government should establish a retroactive investigation mechanism for environmental governance decisions. Former officials who leave a legacy of environmental destruction should pay a penalty for their actions even if they have left office or retired. In addition, and especially when making major administrative decisions for environmental governance, local governments should be required to be open and transparent, and to implement sound supervision and assessment mechanisms.

Local governments in these regions should also increase R&D subsidies and tax preferences for high-tech and low-emission enterprises. Concurrently, to improve the utilization efficiency of R&D subsidies, the government should perfect the supervision system to deter adverse selection problems. For example, cooperative innovation organizations such as national engineering laboratories or industrial R&D centers are encouraged to be set up by enterprises, universities, and research institutes. The government should also undertake external supervision of innovative organizations by designing a post-evaluation system for R&D achievements.

Regarding the high-polluted regions with *extensive population* type, it is concluded that their large populations were the core factor that led to the huge levels of their consumption-related pollution. Large population size will result in overconcentration of the population which may accelerate the superfluous consumption of natural resources, further stimulate economic-social activities, and even lead to predatory or disruptive use of resources. Finally, an increasing population gives rise to huge consumptive pollution. Plus, the population concentrations and inflows into the developed central megacities have caused additional population pressures which have further enhanced resource waste. The government should complete the system for integrating urban and rural development to optimize the spatial distribution of population among cities of different sizes. The alleviation of population pressure in large cities will assist in the realization of a win-win situation between economic development, population density, and environmental protection.

5. Conclusion

This study creatively introduced a fine-grained analysis tool – the fsQCA method – to explore the interactive mechanism and causal complexity of environmental drivers by using the panel data of 30 Chinese provinces in the period between 2011 and 2017. The findings will provide targeted policy recommendations for local governments at the provincial level. The main conclusions are as follows.

1) The regression analyses and necessity analyses confirm that no single factor is necessary condition for high or low pollution, and the interaction effects of pollution drivers exist significantly.

2) There are several different configurations of environmental drivers which lead to high pollution or low pollution in regions. This confirms the multiple causality, asymmetry, and equifinality of environmental issues;

3) The factors of industrial structure, population, and technological innovation are more significant to achieving a state of 'green environment' compared to environmental regulation factors;

4) By testing predictive validity and analyzing the spatiotemporal variations of configurations in the two periods of 2011-2015 and 2016-2017, we found that most regions showed strong path dependencies and spatial agglomeration for their development patterns.

Although this study filled some of the gaps within existing literature pertaining to the environmental area, the scope of this research is not broad enough due to a lack of available data for some variables at a more fine-grained scale, such as city-level or prefecture-level. As the heterogeneity between cities is also prominent, more new conclusions may be found in the future research at the city level.

Declarations

Acknowledgments

The authors wish to express their sincere gratitude to the Fundamental Research Funds for the Central Universities (No. 2021CDJSKJC20 and 2021CDJSKCG28), the Natural Science Foundation of China (Grant No. 72071022 and 71801023), Chongqing Science & Technology Commission (No. cstc2020jscx-msxmX0036) for funding this research project.

Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

Compliance with ethical standards

Conflict of interest All the authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

Author's contributions Yang Chen: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. Jingke Hong: Conceptualization, Writing – review & editing. Miaohan Tang: Data analysis, Resources. Yuxi Zheng: Visualization. Maoyue Qiu: Software. Danfei Ni: Writing – review & editing.

References

1. Bai, Y.S., Siyi; Jiao, Jianling; Yang, Ranran, 2019. The impacts of government R&D subsidies on green innovation: Evidence from Chinese energy-intensive firms. Journal of Cleaner Production 233, 819-829.
2. Bruno, Cassiman, Reinhilde, Veugelers, 2006. In search of complementarity in innovation strategy: Internal R&D and external knowledge acquisition. Operations Research 46, 489-492.
3. Chen, Y., Wang, Z., Zhong, Z., 2019. CO2 emissions, economic growth, renewable and non-renewable energy production and foreign trade in China. Renewable Energy 131, 208-216.
4. Cheng, Z., Li, L., Liu, J., 2020. The impact of foreign direct investment on urban PM2.5 pollution in China. Journal of Environmental Management 265, 110532.
5. David, P.A., 1985. Clio and the economics of QWERTY. American Economic Review 75, 332-337.
6. Delery, J.E., Doty, D.H., 1996. Modes of Theorizing in Strategic Human Resource Management: Tests of Universalistic, Contingency, and Configurational Performance Predictions. Acad Manage J 39, 802-835.
7. Dess, G.G., Lumpkin, G.T., Covin, J.G., 1997. Entrepreneurial Strategy Making and Firm Performance: Tests of Contingency and Configurational Models, Strategic Management Journal, Jg. 18 (9), S. 677-695. Strategic Management Journal 18, 677-695.
8. Dong, K., Hochman, G., Zhang, Y., Sun, R., Li, H., Liao, H., 2018. CO2 emissions, economic and population growth, and renewable energy: Empirical evidence across regions. Energy Economics 75, 180-192.
9. Douglas, E.J., Shepherd, D.A., Prentice, C., 2020. Using fuzzy-set qualitative comparative analysis for a finer-grained understanding of entrepreneurship. Journal of Business Venturing 35.
10. Du, K., Cheng, Y., Yao, X., 2021. Environmental regulation, green technology innovation, and industrial structure upgrading: The road to the green transformation of Chinese cities. Energy Economics 98.
11. Ehrlich, P.R., Holdren, J.P., 1971. Impact of population growth. Science 171, 1212-1217.
12. Fiss, P.C., 2011. Building Better Causal Theories: A Fuzzy Set Approach To Typologies In Organization Research. Acad Manage J 54, 393-420.
13. Greckhamer, T., Furnari, S., Fiss, P.C., Aguilera, R.V., 2018. Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. Strategic Organization 16, 482-495.
14. Greckhamer, T., Gur, F.A., 2021. Disentangling combinations and contingencies of generic strategies: A set-theoretic configurational approach. Long Range Planning 54.
15. Guan, D., Su, X., Zhang, Q., Peters, G.P., Liu, Z., Lei, Y., He, K., 2014. The socioeconomic drivers of China's primary PM 2.5 emissions. Environmental Research Letters 9.

16. Hanif, I., Gago-de-Santos, P., 2017. The importance of population control and macroeconomic stability to reducing environmental degradation: An empirical test of the environmental Kuznets curve for developing countries. Environmental Development 23, 1-9.
17. Huang, G., Zhang, J., Yu, J., Shi, X., 2020. Impact of transportation infrastructure on industrial pollution in Chinese cities: A spatial econometric analysis. Energy Economics 92.
18. Hunt, R.A., Fund, B.R., 2016. Intergenerational Fairness and the Crowding Out Effects of Well-Intended Environmental Policies. Journal of Management Studies 53, 878-910.
19. Jianqin, M.A., Zhengzhou, China, CAS, Beijing, China, 2010. Water Quality Evaluation Model Based on Principal Component Analysis and Information Entropy:Application in Jinshui River. Journal of Resources and Ecology.
20. Krueger, G.M.G.a.A.B., 1995. Economic Growth and the Environment. The Quarterly Journal of Economics 110, 353-377.
21. Laplante, B., Rilstone, P., 2004. Environmental Inspections and Emissions of the Pulp and Paper Industry in Quebec. Journal of Environmental Economics & Management 31, 19-36.
22. Li, H., Zhang, J., Wang, C., Wang, Y., Coffey, V., 2018. An evaluation of the impact of environmental regulation on the efficiency of technology innovation using the combined DEA model: A case study of Xi'an, China. Sustainable Cities and Society 42, 355-369.
23. Li, K., Fang, L., He, L., 2019. How population and energy price affect China's environmental pollution? Energy Policy 129, 386-396.
24. Li, K., Fang, L., He, L., 2020. The impact of energy price on CO2 emissions in China: A spatial econometric analysis. Science of The Total Environment 706, 135942.
25. Liddle, B., Lung, S., 2010. Age-structure, urbanization, and climate change in developed countries: revisiting STIRPAT for disaggregated population and consumption-related environmental impacts. Population and Environment 31, 317-343.
26. Lim, L., Acito, F., Rusetski, A., 2006. Development of archetypes of international marketing strategy," Journal of International Business Studies, Vol. , pp. Management International Review 47, 453-475.
27. Linares, I., Paulo, A., Porto, G.S., 2019. Patent-based network analysis to understand technological innovation pathways and trends. Technology in Society 59, 101134-.
28. Lobe, B.R.B., 2010. The Case for Qualitative Comparative Analysis (QCA) : Adding Leverage for Thick Cross-Case Comparison. The SAGE Handbook of Case-Based Methods.
29. MEP, M.o.E.P., 2017. The City Air Quality of China in 2016.
30. Meyer, A.D., Tsui, A.S., Hinings, C.R., 1993. Configurational Approaches to Organizational Analysis. The Academy of Management Journal 36, 1175-1195.
31. Misangyi, V.F., Acharya, A.G., 2014. Substitutes or Complements? A Configurational Examination of Corporate Governance Mechanisms. Acad Manage J 57, 1681-1705.
32. Pappas, I.O., Woodside, A.G., 2021. Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for research practice in Information Systems and marketing. International Journal of Information Management 58.
33. Park, Y., Fiss, P.C., El Sawy, O.A., 2020. Theorizing the Multiplicity Of Digital Phenomena: The Ecology Of Configurations, Causal Recipes, And Guidelines for Applying Qca. Mis Quart 44, 1493-1520.
34. Pham, N.M., Huynh, T.L.D., Nasir, M.A., 2020. Environmental consequences of population, affluence and technological progress for European countries: A Malthusian view. J Environ Manage 260, 110143.
35. Ragin, C., 2008. Redesigning Social Inquiry: Fuzzy Sets and Beyond. University of Chicago Press.
36. Ragin, C.C., 1987. The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. Social Forces 67.
37. Ragin, C.C., 2000. Fuzzy-Set Social Science. University of Chicago Press Economics Books.
38. Rihoux, B., Ragin, C., Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques. Configurational comparative methods.
39. Ritter, H., Schopf, M., 2014. Unilateral Climate Policy: Harmful or Even Disastrous? Environmental & Resource Economics 58, 155-178.
40. Rutten, R., 2020. Applying and Assessing Large-N QCA: Causality and Robustness From a Critical Realist Perspective. Sociological Methods & Research published online.
41. Shahbaz, M., Loganathan, N., Muzaffar, A.T., Ahmed, K., Ali Jabran, M., 2016. How urbanization affects CO2 emissions in Malaysia? The application of STIRPAT model. Renewable and Sustainable Energy Reviews 57, 83-93.
42. Shen, F., Liu, B., Luo, F., Wu, C., Chen, H., Wei, W., 2021. The effect of economic growth target constraints on green technology innovation. Journal of Environmental Management 292, 112765.
43. Shen, N., Peng, H., 2021. Can industrial agglomeration achieve the emission-reduction effect? Socio-Economic Planning Sciences 75, 100867.
44. Tan, H., Fan, Z., 2019. Technical management ability, attention allocation and local government website construction-A configuration analysis based on TOE frame [in Chinese]. Management World (Monthly) 9, 81-94.
45. Wagemann, S., 2012. Set-Theoretic Methods for the Social Sciences. Cambridge University Press.
46. Wang, H., Wei, W., 2020. Coordinating technological progress and environmental regulation in CO2 mitigation: The optimal levels for OECD countries & emerging economies. Energy Economics 87, 104510.
47. Wang, P., Wu, W.S., Zhu, B.Z., Wei, Y.M., 2013. Examining the impact factors of energy-related CO2 emissions using the STIRPAT model in Guangdong Province, China. Applied Energy 106, 65-71.

48. Wang, X.Z., C.; Zhang, Z., 2019. Pollution haven or porter? The impact of environmental regulation on location choices of pollution-intensive firms in China. J Environ Manage 248, 109248.
49. Wen, Q., Chen, Y., Hong, J., Chen, Y., Shen, Q., 2020. Spillover effect of technological innovation on CO2 emissions in China's construction industry. Building and Environment 171, 106653.
50. Woodside, A.G., 2013. Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. Journal of Business Research 66, 463-472.
51. Yeh, J.-C., Liao, C.-H., 2017. Impact of population and economic growth on carbon emissions in Taiwan using an analytic tool STIRPAT. Sustainable Environment Research 27, 41-48.
52. Zhang, G.Z., Ning; Liao, Wenmei, 2018. How do population and land urbanization affect CO2 emissions under gravity center change? A spatial econometric analysis. Journal of Cleaner Production 202, 510-523.
53. Zhang, P., 2020. Target Interactions and Target Aspiration Level Adaptation: How Do Government Leaders Tackle the “Environment-Economy” Nexus? Public Administration Review 81, 220-230.
54. Zhang, W., Liu, Y., Feng, K., Hubacek, K., Wang, J., Liu, M., Jiang, L., Jiang, H., Liu, N., Zhang, P., Zhou, Y., Bi, J., 2018a. Revealing Environmental Inequality Hidden in China's Inter-regional Trade. Environ Sci Technol 52, 7171-7181.
55. Zhang, W., Xu, H., 2017. Effects of land urbanization and land finance on carbon emissions: A panel data analysis for Chinese provinces. Land Use Policy 63, 493-500.
56. Zhang, Z., Hao, Y., Lu, Z.-N., Deng, Y., 2018b. How does demographic structure affect environmental quality? Empirical evidence from China. Resources, Conservation and Recycling 133, 242-249.
57. Zhao, J., Jiang, Q., Dong, X., Dong, K., 2020. Would environmental regulation improve the greenhouse gas benefits of natural gas use? A Chinese case study. Energy Economics 87, 104712.
58. Zheng, D., Shi, M., 2017. Multiple environmental policies and pollution haven hypothesis: Evidence from China's polluting industries. Journal of Cleaner Production 141, 295-304.
59. Zheng, Y.P., J. C., 2020. Industrial structure transformation and provincial heterogeneity characteristics evolution of air pollution: Evidence of a threshold effect from China. Atmospheric Pollution Research 11, 598-609.

Tables

Table 1. The relationships between pollutant emissions (e.g., CO_2, SO_2, NO_x) and their several influencing factors.

Influencing factor	Relationship	Method	Reference
Economic growth	● Positive	• Ridge regression model • Panel regression model	Dong et al. (2018); Wang et al. (2013)
	● Negative	• Time series analysis	Yeh and Liao (2017)
	● Inverted U-shaped curve	• GMM method • Instrumental variable method	Hanif and Gago-de-Santos (2017); Zhang et al. (2018b)
Population growth	● Positive	• Ridge regression • Panel regression model	Dong et al. (2018); Li et al. (2019); Liddle and Lung (2010); Wang et al. (2013)
	● Inverted U-shaped curve	• GMM method	Hanif and Gago-de-Santos (2017); Zhang et al. (2018b)
Urbanization	● Positive	• Ridge regression • Spatial Durbin Panel model	Wang et al. (2013); Zhang (2018)
	● Negative	• Spatial econometric model	Zhang and Xu (2017)
	● Inverted U-shaped curve	• Vector error correction model	Shahbaz et al. (2016)
Technological progress	● Positive	• Panel regression model	Wang and Wei (2020)
	● Negative	• Panel regression model	Pham et al. (2020)
Environmental regulation	● Positive	• General equilibrium carbon leakage model	Ritter and Schopf (2014)
	● Negative	• GMM method	Zhao et al. (2020)

Table 2. The basic profile of variables.

Factor	Variable	Explanation	Reference
Explained variable	● Environmental pollution	Calculated by the entropy method including the following indicators: • Industrial wastewater • Industrial sulfur dioxide • Industrial smoke • Industrial dust • Industrial nitrogen oxides	Du et al. (2021); Huang et al. (2020); Jianqin et al. (2010)
Technology	● Technological innovation (TI)	Calculated by the entropy method including the following indicators: • R&D expenditure • Number of patents • Number of researchers • Full-time equivalents • New product sales	Bruno et al. (2006); Li et al. (2018); Linares et al. (2019); Shen et al. (2021); Wen et al. (2020)
	● R&D subsidy (RS)	• Government subsidy on R&D	Du et al. (2021)
Government	● Source treatment (ST)	• The number of people in environmental protection system	Laplante and Rilstone (2004)
	● End-of-pipe treatment (ET)	• The investment in pollution abatement	Wang (2019)
Economy	● Economic growth (EG)	• GDP per capita	Wang (2019)
Population	● Population scale (PS)	• The number of people in the province	Wang (2019)
Industry	● Industrial structure (IS)	• The ratio of the added value of the tertiary industry to that of the secondary industry	Du et al. (2021); Shen et al. (2021)

Note: All data of selected variables come from *China Statistical Yearbook, China Statistical Yearbook on Environment,* and *China Statistical Yearbook on Science and Technology.*

Table 3. Descriptive statistics and calibration thresholds for the variables.

Period	Variable	Mean	SD	Fully out, crossover, and fully in anchors for calibration
2011-2015	Pollution (Dimensionless)	28.6	22.0	13.0, 24.7, 36.4
	TI (Dimensionless)	21.5	24.6	4.93, 13.7, 22.5
	RS (Billion RMB)	6.94	10.3	2.19, 5.71, 9.22
	ST (Thousand persons)	6.99	5.06	4.01, 6.85, 9.68
	ET (Billion RMB)	23.5	16.2	15.0, 23.5, 31.0
	EG (Thousand RMB per person)	40.4	17.5	29.1, 39.5, 49.9
	PS (Million persons)	45.1	27.0	25.2, 42.5, 59.8
	IS (%)	99.0	0.57	72.0, 86.0, 99.0
2016-2017	Pollution (Dimensionless)	26.4	22.0	11.1, 22.0, 33.0
	TI (Dimensionless)	21.8	25.5	5.28, 14.7, 24.1
	RS (Billion RMB)	7.45	10.7	2.78, 7.08, 11.4
	ST (Thousand persons)	4.94	3.60	3.11, 4.62, 6.12
	ET (Billion RMB)	21.0	14.7	10.6, 20.4, 30.2
	EG (Thousand RMB per person)	40.1	17.9	28.3, 37.8, 47.2
	PS (Million persons)	46.0	27.7	25.5, 43.5, 61.4
	IS (%)	132	0.67	97.0, 113, 129

Notes: $Obs_{2011\text{-}2015}$=150, $Obs_{2016\text{-}2017}$=60; SD=Standard deviation.

Table 4. Configurations for high pollution and low pollution between 2011 and 2015.

2011-2015	High pollution					Low pollution				
	C1a	C1b	C1c	C2	C3	C4a	C4b	C4c	C5a	C5b
TI	⊗	⊗	●	●	●	●	⊗	⊗	⊗	●
RS	⊗	⊗	●	⊗	●	●	⊗		⊗	●
ST	●	⊗	⊗	●	●	⊗	⊗	⊗	⊗	⊗
ET	●	●	●		●		⊗	⊗	⊗	⊗
EG	⊗	●	⊗	⊗	●	●	⊗	⊗	⊗	⊗
PS	⊗	⊗	●	●	●	⊗		⊗	⊗	⊗
IS	⊗	⊗	⊗	⊗		●	●	●		⊗
Raw coverage	0.072	0.074	0.104	0.174	0.326	0.203	0.388	0.362	0.491	0.070
Unique coverage	0.038	0.050	0.040	0.108	0.256	0.174	0.040	0.012	0.135	0.035
Consistency	0.972	0.908	0.981	0.966	0.988	0.931	0.992	1.000	0.977	1.000
Typical regions	*Shanxi*	*Inner Mongolia*	*Anhui*	*Henan, Hebei*	*Jiangsu, Guangdong*	*Beijing, Shanghai*	*Yunnan*	*Heilongjiang*	*Qinghai, Jilin*	*Shaanxi*
Type	*Structural Imbalance*				*Extensive Population*	*Green Development*			*Scarce Population*	
Solution coverage	0.587					0.755				
Solution consistency	0.977					0.962				

Notes: Black circles indicate the high level (or presence) of a condition; circles with crosses indicate the low level (or absence) of a condition; large circles indicate core conditions; small ones, peripheral conditions; and blank spaces indicate "don't care".

Table 5. Configurations for high pollution and low pollution between 2016 and 2017.

2016-2017	High pollution							Low pollution				
	C6a	C6b	C7a	C7b	C8a	C8b	C8c	C9a	C9b	C9c	C10a	C10b
TI	⊗	⊗	⊗	●	●		●	●	⊗	⊗	⊗	⊗
RS	⊗	⊗	⊗	⊗	●	●	●	●	⊗	⊗	⊗	●
ST	●	●	⊗	●	⊗	●	●	⊗	⊗	⊗	⊗	⊗
ET	●	●	●	●	●	⊗	●		⊗		⊗	●
EG	⊗	●	⊗	⊗	⊗	⊗	●	●	⊗	⊗		⊗
PS	⊗	⊗	●	●	●	●	●	⊗		⊗	⊗	⊗
IS	●	⊗	⊗	⊗	⊗	●		●	●	●	⊗	⊗
Raw coverage	0.085	0.076	0.072	0.156	0.111	0.139	0.360	0.191	0.304	0.296	0.328	0.066
Unique coverage	0.046	0.040	0.034	0.087	0.031	0.086	0.259	0.159	0.035	0.028	0.229	0.036
Consistency	0.961	0.973	0.963	1.000	1.000	0.944	1.000	0.887	0.967	0.979	0.981	0.990
Typical regions	*Shanxi*	*Inner Mongolia*	*Jiangxi*	*Henan, Hebei*	*Anhui*	*Sichuan, Liaoning*	*Jiangsu, Guangdong*	*Beijing, Shanghai*	*Yunnan*	*Heilongjiang*	*Qinghai, Jilin*	*Shaanxi*
Type	*Technical Laggard*		*Extensive Population*					*Green Development*			*Scarce Population*	
Solution coverage	0.716							0.763				
Solution consistency	0.979							0.953				

Notes: Black circles indicate the high level (or presence) of a condition; circles with crosses indicate a low level (or absence); large circles indicate core conditions; small ones, peripheral conditions; and blank spaces indicate "don't care".

Table 6. Configurations for high pollution and low pollution between 2016 and 2017.

2016-2017	High pollution							Low pollution				
	C6a	C6b	C7a	C7b	C8a	C8b	C8c	C9a	C9b	C9c	C10a	C10b
technology	⊗	⊗	⊗	●	●		●	●	⊗	⊗	⊗	⊗
subsidy	⊗	⊗	⊗	⊗	●	●	●	●	⊗	⊗	⊗	●
source	●	●	⊗	●	⊗	●	●	⊗	⊗	⊗	⊗	⊗
end-of-pipe	●	●	●	●	●	⊗	●		⊗		⊗	●
gdp	⊗	●	⊗	⊗	⊗	⊗	●	●	⊗	⊗		⊗
population	⊗	⊗	●	●	●	●	●	⊗		⊗	⊗	⊗
industry	●	⊗	⊗	⊗	⊗	●		●	●	●	⊗	⊗
Raw coverage	0.085	0.076	0.072	0.156	0.111	0.139	0.360	0.191	0.304	0.296	0.328	0.066
Unique coverage	0.046	0.040	0.034	0.087	0.031	0.086	0.259	0.159	0.035	0.028	0.229	0.036
Consistency	0.961	0.973	0.963	1.000	1.000	0.944	1.000	0.887	0.967	0.979	0.981	0.990
Typical regions	*Shanxi*	*Inner Mongolia*	*Jiangxi*	*Henan, Hebei*	*Anhui*	*Sichuan, Liaoning*	*Jiangsu, Guangdong*	*Beijing, Shanghai*	*Yunnan*	*Heilongjiang*	*Qinghai, Jilin*	*Shaanxi*
Type	*Technical Laggard*		*Extensive Population*					*Green Development*			*Scarce Population*	
Solution coverage	0.716							0.763				
Solution consistency	0.979							0.953				

Notes: Black circles indicate the high level (or presence) of a condition; circles with crosses indicate a low level (or absence); large circles indicate core conditions; small ones, peripheral conditions; and blank spaces indicate "don't care".

Table 7. The test for predictive validity.

Configuration for high pollution	Consistency	Raw coverage
C1a: *~tech*~subsidy*source*end*~gdp*~population*~industry*	0.982	0.039
C1b: *~tech*~subsidy*~source*end*gdp*~population*~industry*	0.953	0.042
C1c: *tech*subsidy*~source*end*~gdp*population*~industry*	1.000	0.111
C2: *tech*~subsidy*source*~gdp*population*~industry*	0.992	0.181
C3: *tech*subsidy*source*end*gdp*population*	1.000	0.360
Configuration for low pollution	Consistency	Raw coverage
C4a: *tech*subsidy*~source*gdp*~population*industry*	0.887	0.191
C4b: *~tech*~subsidy*~source*~end*~gdp*industry*	0.967	0.304
C4c: *~tech*~source*~end*~gdp*~population*industry*	0.966	0.271
C5a: *~tech*~subsidy*~source*~end*~gdp*~population*	0.985	0.469
C5b: *tech*subsidy*~source*~end*~gdp*~population*~industry*	0.990	0.066

Notes: ~; Negation (NOT), *; Logical conjunction (AND).

Figures

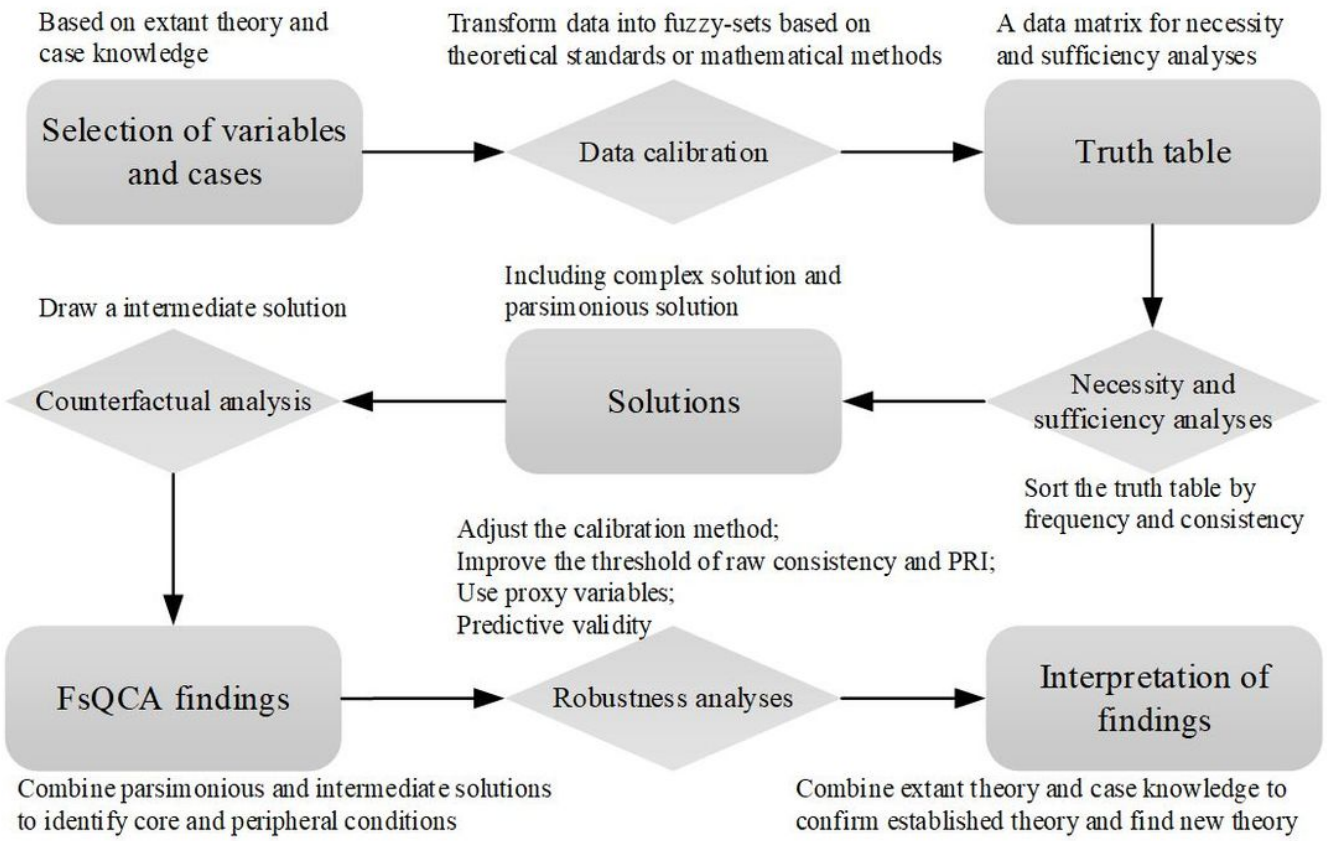

Figure 1

Basic steps in fsQCA method.

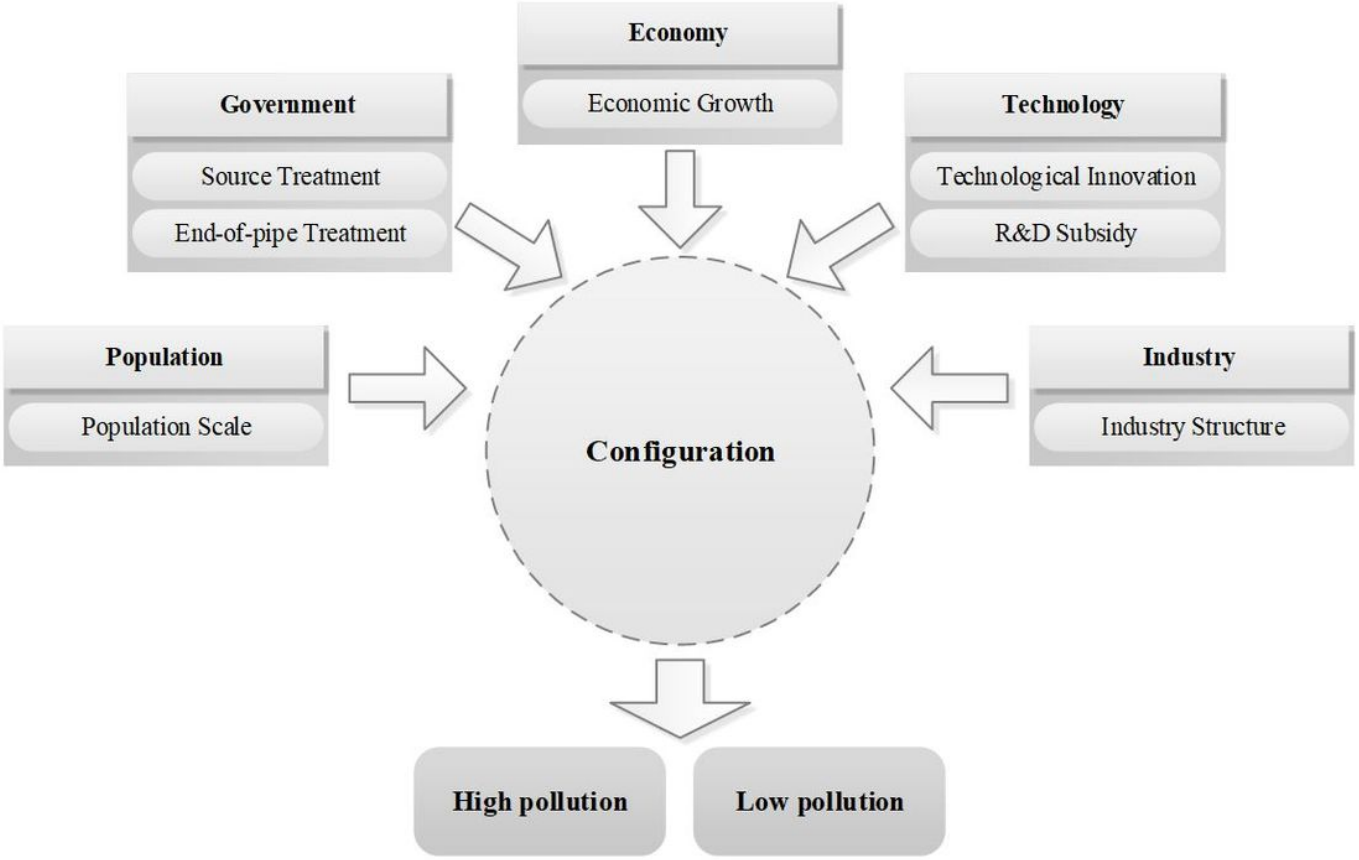

Figure 2

Configuration analysis framework of environmental driving factors for leading to pollution.

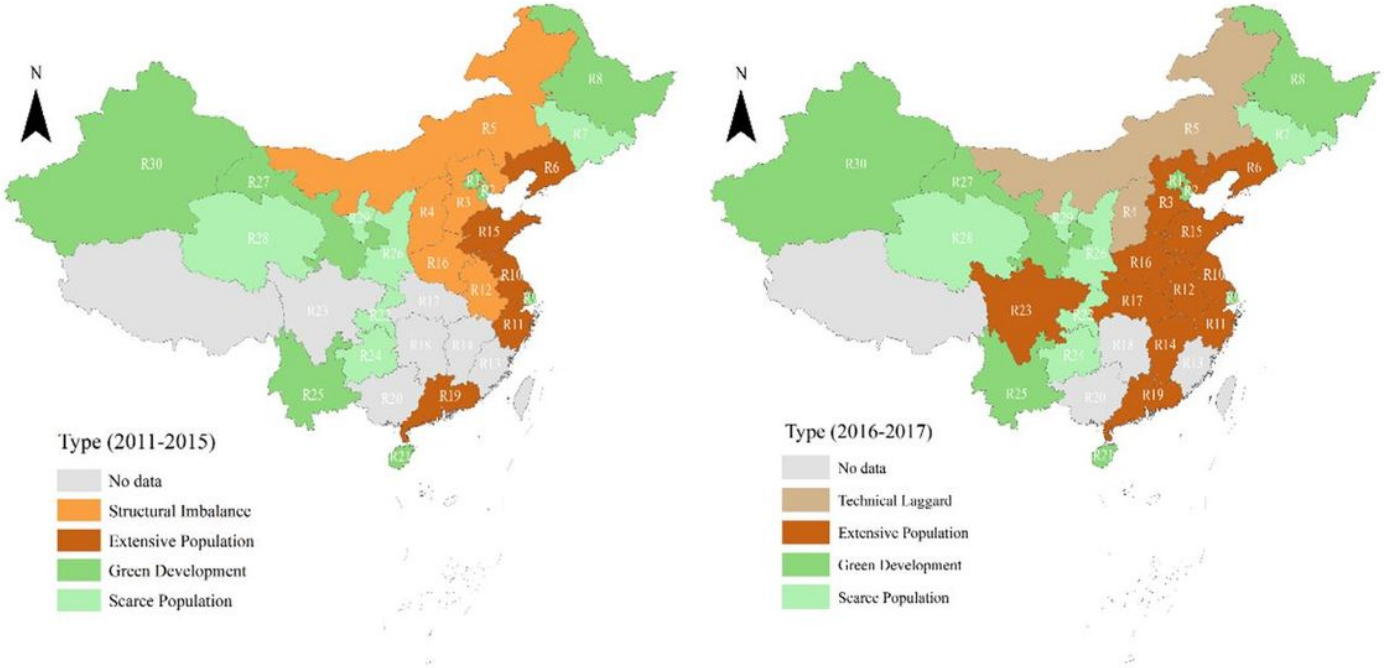

Figure 3

The maps of configural types in the two time periods.

Note: The abbreviations of regions detail in Appendix Table A2.

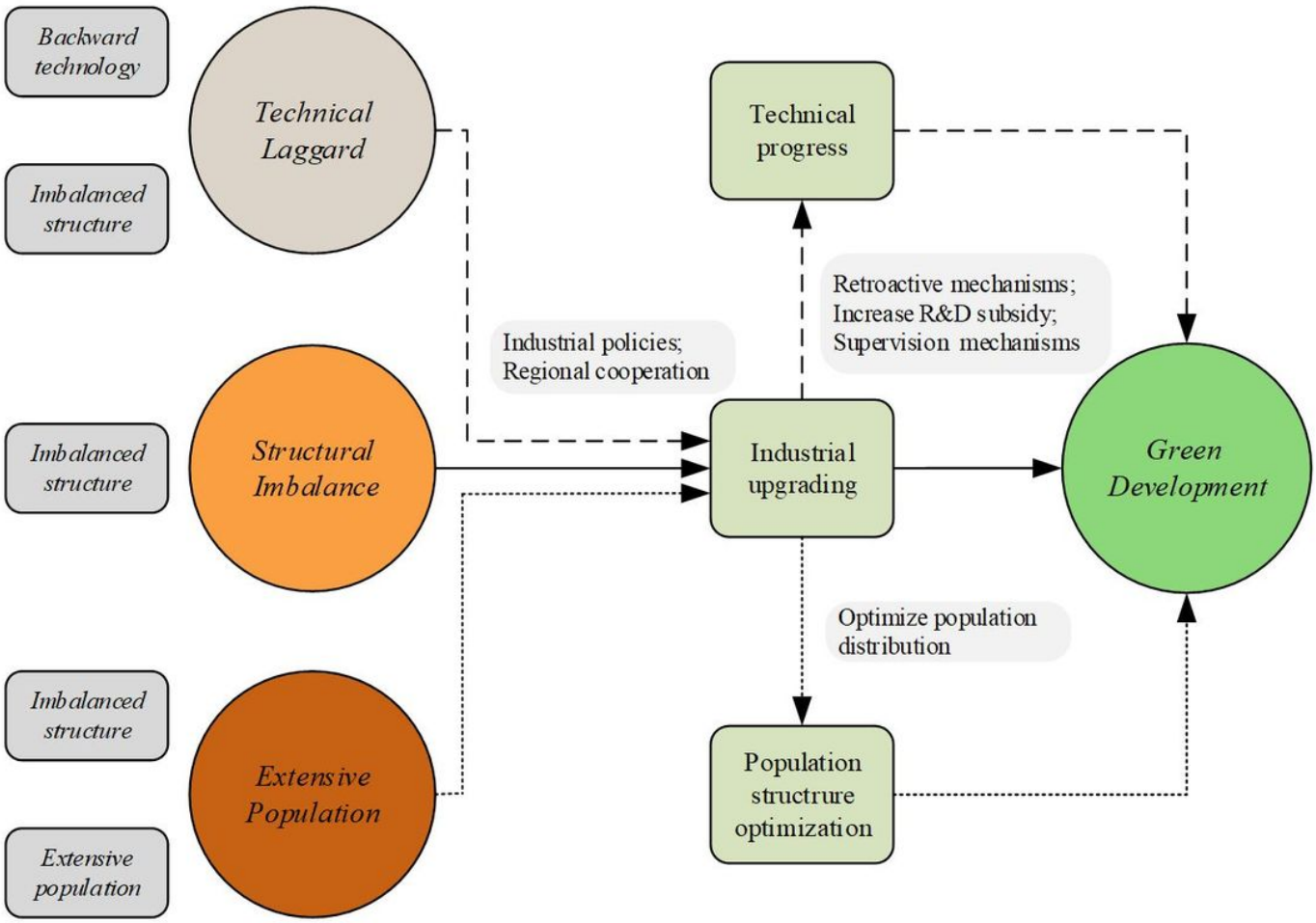

Figure 4

The advised paths to a green environment for three high-polluted types.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx
- Appendix.docx