ANALYSIS

Analysis of the major drivers of the ecological footprint using the STIRPAT model and the PLS method—A case study in Henan Province, China

Junsong Jia, Hongbing Deng, Jing Duan, Jingzhu Zhao

State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

ARTICLE INFO

Article history:
Received 26 November 2008
Received in revised form 20 May 2009
Accepted 20 May 2009
Available online 18 June 2009

Keywords:
EF
Drivers
STIRPAT
PLS
Henan Province

ABSTRACT

Taking Henan Province of China as an example, we computed and analyzed the ecological footprint (EF) in 1983–2006. The results showed that the EF in Henan Province quadrupled in the 23 years, but its ecological carrying capacity (EC) was rather low and was in a state of slow decline, indicating that Henan's ecological deficit (ED) had become a remarkable social problem. Therefore, the major drivers of the EF's change were analyzed. According to the simulations with STIRPAT model, the major drivers of Henan's EF were human population (P), GDP per capita (A1), quadratic term of GDP per capita (A2), percent of economy excluded in service sector (Ta1) and percent of urban population (Tb1). However, these drivers themselves had strong collinearity, which might produce some uncertain impact to the final results. In order to avoid the impact of collinearity, the method of partial least squares (PLS) was used. The results showed that the major drivers of EF were P, A1, A2 and Tb1, Ta1 was excluded. Compared with the results by the STIRPAT model, which showed that P is the most dominant driver and the effect of the other drivers could almost be ignored, the results by PLS method were considered as more reasonable and acceptable because the impacts of the A (Affluence) and T (Technology) conditions to the regional EF were still too important to be ignored. In addition, the results acquired by both methods showed that the curvilinear relationship between economic development and ecological impact (EF) or the classical EKC hypothesis didn't exist in Henan Province.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Sustainable development has become an important goal around the world since the Rio Earth Summit (Wackernagel et al., 2004; Chen et al., 2006). So far, a variety of models, methods and indicators have been put forward to evaluate sustainable development quantitatively at regional scales. As one of approaches, the ecological footprint (EF) indicator, developed by Wackernagel and Rees (1996), has received considerable attention and been broadly used (Erb, 2004; Holmberg et al., 1999). Although the EF has some inherent drawbacks and thus are criticized by some scholars (Ayres, 2000; Van den Bergh and Verbruggen, 1999), we cannot deny the fact that it is an effective evaluation tool to estimate the overall ecological impact caused by human activities (Rosa et al., 2004; Xu et al., 2005; Dietz et al., 2007) and provide some reasonable clues to help achieve regional sustainable development.

The STIRPAT model, which was originated by Dietz and Rosa (1994, 1997), was converted from the widely recognized IPAT that has been used to analyze environmental impacts (Chertow, 2001; York et al., 2002). STIRPAT was first applied to analyze EF, CO2 emissions and energy footprint by York, Rosa, and Dietz (2003a,b). The indicators above were used to describe ecological impact of human activities and STIRPAT model was used to analyze the driving forces of the impacts. Later on, Rosa et al. (2004) utilized EF to denote ecological impact and STIRPAT model to analyze the impact's driving forces. Dietz et al. (2007) still used EF and STIRPAT model to analyze anthropogenic ecological impact and its drivers. Taking the EFs of most provinces as an example, Xu and Cheng (2005) conducted some similar studies in China. Meaningful conclusions were found from these studies and some significant suggestions were provided for local governments. Our study is different in several ways from the previous studies (Table 1). Particularly, the cross-section data were used at a global scale in the previous studies but the time-series data in a certain region have hardly been used. More importantly, STIRPAT model can essentially be transformed into a random form of the Ordinary Least Squares (OLS) regression model, which allows for determining variables subjectively before selecting independent variables (X) and has uncertainties that cannot totally exclude the impact of collinearity among the independent variables.

The phenomenon of the existence of prefect correlation or a high degree of correlation between the explanatory (independent) variables in the linear regression model can be considered as collinearity (Wang, 1999, 2006). When there are more than two covariates that are highly correlated, this is called multicollinearity (Wang, 1999, 2006). O'Brien
Table 1
A comparison between the previous studies and this paper.

<table>
<thead>
<tr>
<th>Items</th>
<th>York et al. (2003a,b)</th>
<th>Rosa et al. (2004)</th>
<th>Xu et al. (2005)</th>
<th>Dietz et al. (2007)</th>
<th>This paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study scales</td>
<td>World</td>
<td>World</td>
<td>China</td>
<td>World</td>
<td>A province in China</td>
</tr>
<tr>
<td>Data features</td>
<td>Cross-section</td>
<td>Cross-section</td>
<td>Cross-section</td>
<td>Cross-section</td>
<td>Time-series</td>
</tr>
<tr>
<td>Methods</td>
<td>IMPACT &amp; STIRPAT</td>
<td>STIRPAT</td>
<td>STIRPAT</td>
<td>STIRPAT</td>
<td>STIRPAT &amp; PLS</td>
</tr>
<tr>
<td>Ecological impact’s indicators</td>
<td>CO₂ emissions and energy footprint, EF</td>
<td>O₃ and CO₂ emissions and EF</td>
<td>EF</td>
<td>EF</td>
<td>EF</td>
</tr>
</tbody>
</table>

(2007) believes that the values of the Variance Inflation Factor (VIF, used in measuring the degree of multicollinearity) of 10, 20, 40, or even higher do not, by themselves, discount the results of regression analyses, and suggests to eliminate one or more independent variables from the analysis by using the ridge regression. From this point of view, multicollinearity (multicollinearity) may not be a serious problem. However, other scientists (Wang, 1999, 2006; Naes and Mevik, 2001; Tu et al., 2004; Shacham and Brauner, 1997) consider that collinearity is a problem that should receive more attention. For example, if there is multicollinearity in the regression model, regression coefficients will no longer have the meaning for general interpretation (Wang, 1999); further, collinearity can sometimes lead to serious instability of the variables' coefficients (Naes and Mevik, 2001).

Partial least squares is a class of regression estimation methods initially developed by Wold (1966, 1973), which has recently been increasingly popular among scientists as a technique for treating highly collinear data. The method can exclude the impact of the independent variables’ collinearity problem successfully and integrates the multiple regression (MR), principal component analysis (PCA) and canonical correlation analysis (CCA) into a model at the same time. It can choose accurately (not stochastic) dominant latent (integrated) component (t) from independent variables (X) and determine quantitatively the size of the contribution from X variables to Y variables (dependent variables) (Wang, 1999, 2006) and, thus, appears to be more appropriate in analyzing EF’s drivers.

Therefore, we tried to utilize this innovative PLS method in analyzing EF’s drivers. Firstly, we took Henan Province of China as a study site and used Henan’s historical (time-series) data to calculate its EF and EC and analyzed the condition of natural resources per capita (dependent variables) (Wang, 1999, 2006) and, thus, appears to be more appropriate in analyzing EF’s drivers.

2. Study area

Henan Province, geographically ranging from 31°23’ to 36°22’N and from 110°21’ to 116°39’E, is located in central China (Fig. 1). Its total land area is 1,670,000 km², accounting for 1.73% of entire land area in China. Henan Province has the largest population in China and its regional resources per capita are becoming less as the population keeps increasing continuously. For example, the arable land area per capita is 0.087 ha in 2003, which is much lower than the national average (0.112 ha, Xing et al., 2005). At the same time, its economic or technical development lags behind relatively developed regions in China. Henan’s ecological conditions are getting worse (Li and Fan, 2005). In particular, in recent years, a lot of forests are destroyed, which leads to more serious regional soil erosion and environmental degradation.

3. Methods and data

3.1. EF computation

The EF is the total area of productive land and water required continuously to produce all the resources consumed and to assimilate all the wastes produced, by a defined population, wherever on Earth that land is located (Wackernagel et al., 1997). Eq. (1) was used to calculate EF in Henan Province in this paper (Wackernagel et al., 1997, 1999):

\[
EF = N(EF_i) = N \sum r_i(a_{ij}) = N \sum r_i(C_i / Y_i)
\]

where, EF is the total regional EF (gha); N is the population; EF_i is the per capita EF (gha/cap); a_{ij} is the consumption item number (i = 1, 2, ..., n); r_i is the equivalent factor of different types of land, obtained from the literatures (Wackernagel et al., 2004); C_i is the per capita consumption of i item (kg/cap); Y_i is the average productivity of producing i item in a certain biological productive area in a certain year (kg/gha). The per capita EF is classified into six types, corresponding to arable land, pastureland, forestland, fossil energy land, built-up land and water land, respectively.

The EC reflects the ability of regional available land resources. Eq. (2) was used to calculate EC (Wackernagel et al., 1997, 1999):

\[
EC = N(EC_i) = N \sum (a_j \times r_j \times y_j)
\]

where, EC is the total regional EC (gha); EC_i is the per capita EC (gha/cap); a_j is the per capita biological productive area of j type land in a
region; and $y_i$ is the yield factor of different types of land, got from the literatures (Wackernagel et al., 1997).


3.2. Driver analysis methods

3.2.1. STIRPAT

The STIRPAT is calculated with a non-linear regression formula and the coefficients in it represent the hypotheses (Eq. (3)) which need to be tested:

$$ l = aP^bA^cT^d e $$

where, the constant $a$ scales the model, $b$, $c$ and $d$ are the exponents of $P$, $A$ and $T$, respectively and $e$ is the error item.

In order to estimate the parameters, Eq. (3) is converted to the linear logarithmic form as follows:

$$ \ln(l) = \ln a + b \ln(P) + c \ln(A) + d \ln(T) + \ln e $$

3.2.2. PLS

Partial least squares (PLS) is particularly useful when independent variables have a strong collinearity (Wang, 1999, 2006).

The method will extract the latent variables (integrated or latent variables extracted from the observed variables and the predicted variables, can express the importance of independent variables) as much as possible and $X$ variables as far as possible

$$ Y_{\text{obs}} = \sum_{i=1}^{m} t_i u_i + \epsilon $$

where $Y_{\text{obs}}$ is the observed variables, $t_i$ is the integrated or latent variables extracted from the $X$ variables, $u_i$ are the exponents of $X$ variables as much as possible and $\epsilon$ is the error item.

In this study, we adopted directly the software Simca-P11.5, developed by the Umetrics company, to perform PLS calculations. The input data were the same as those in STIRPAT model.

There are two important tables or plots used to explain the applicability of the PLS Method: the $t_1/t_2$ scatter plot (also called to be $T^2$ oval plot) and $t_1/u_1$ scatter plot. In the $T^2$ oval plot, $t_1$ and $t_2$ are the integrated or latent variables extracted from the $X$ variables. They can represent the key information of $X$ variables as much as possible and have strong capacity to explain the $Y$ variables as far as possible (Wang, 1999, 2006). If the $t_1/t_2$ relationship of the sample data is all included in the oval (Fig. 6), these sample data are homogeneous and good and can be accepted perfectly (Wang, 1999, 2006). In the $t_1/u_1$ scatter plot, $u_1$ is the integrated or latent variables extracted from the $Y$ variables. If the $t_1/u_1$ relationship of the sample data is near linear, the PLS linear regression model built is appropriate to the study problem (Wang, 1999, 2006).

Similarly, there are three important tables or plots used to explain or determine the fitting effect of PLS model: the Observed vs. Predicted Plot, Variable Importance Plot (VIP) and the Coefficients Table. The Observed vs. Predicted Plot, which presented the relationship between the observed variables and the predicted variables, can express intuitively the fitting effect of the model. An almost linear relationship presented in the Observed vs. Predicted Plot meant the explanation or fitting effect of the results made by the PLS method is excellent (Wang, 1999, 2006). The VIP plot reveal quantitatively the statistical importance of independent variables ($X$) to all dependent variables ($Y$) in the model. The sum of squares of all VIP value is equal to the number of $X$ items in the model hence the average VIP should be equal to 1 (Wang, 1999). VIP values more than 1 indicate $X$ variables are “important”, and values less than 0.5 indicate $X$ variables are “unimportant”. The interval between 1 and 0.5 is a gray zone, where the importance level depends on the values of VIP (Wang, 2006). The Coefficients Table showed the specific correlation coefficient between $X$ and $Y$ variables and the cumulative explanatory capacity from $X$ variables to $Y$ variables after the $t_1$, $t_2$ or $t_1$ were extracted.

4. Results and discussion

4.1. Analysis of ecological conditions in Henan Province

Fig. 2 shows the computation results of different types of EF during 1983–2006. It can be found that all types of EF had generally increased in the whole period. Fig. 2a shows that although the EFs of built-up land, forest land and water land increased, they took small shares in the total EF and thus could almost be ignored (0.03 < 3 ha/cap, Fig. 2). However, it should be noted that the increase in EF for built-up land changed significantly. It could explain the fact that the area used for construction such as housing and road, was increased rapidly in the
regional human activities. Regional resources’ supply capacity increases inadequately to support the per capita EF. From 1983 to 2006, the Henan Province’s EF decreased from 0.20 ha/cap to 0.15 ha/cap, indicating that the Henan’s anthropogenic ecological impact has become a more important social problem in China.

4.2. Major drivers

The potential drivers included the total population \( (P) \), affluence (denoted by GDP per capita, \( A_1 \)), quadratic term of affluence (\( A_2 \)), technology (denoted by the percent of the economy not in service sector or percent of the urban population, \( T_1 \) or \( T_2 \)), quadratic term of technology (\( T_1^2 \) or \( T_2^2 \)), fixed investment and percentage of all types of student, etc. The data-processing method adopted the commonly used procedure based on works by York et al. (2003b), Rosa et al. (2004) and Xu and Cheng (2005). In these hypothesized drivers, \( T_1 \), \( T_2 \), fixed investment and percentage of all types of student, etc. are ultimately excluded through computations due to the low correlations with the dependent variable (EF).

The regression results of STIRPAT models are shown in Table 2. Overall, it can be seen that \( P \) had an elasticity of about 2.3–3.7, which means that a 1% change in population may induce about 2.3–3.7 percentages change in EF. Compared with the \( P \) drivers, the other drivers may have almost no effects on Henan’s EF because their elasticities were so low (<<2.3–3.7).

Model 1 is the basic equation including variables \( P, A_1 \) and \( A_2 \), and they can explain almost all ecological impacts (Adjusted \( R^2 = 0.995 \), Durbin–Watson = 1.636, \( \text{Sig} < 0.001 \)). Compared with the \( P \) drivers, \( A_1, A_2 \) may have relatively small effects to Henan’s EF because their elasticities were low (0.092–0.210<<2.3). Then, the impacts

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P )</td>
<td>0.851***</td>
<td>0.769***</td>
<td>0.873***</td>
<td>0.848***</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>1.000</td>
<td>0.982***</td>
<td>0.952***</td>
<td>0.983***</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>1.000</td>
<td>0.488***</td>
<td>0.645***</td>
<td>0.630***</td>
</tr>
<tr>
<td>( T_1 )</td>
<td>1.000</td>
<td>-0.645***</td>
<td>-0.088</td>
<td>-0.088</td>
</tr>
<tr>
<td>( T_2 )</td>
<td>1.000</td>
<td>-0.645***</td>
<td>-0.088</td>
<td>-0.088</td>
</tr>
<tr>
<td>( T_1^2 )</td>
<td>1.000</td>
<td>0.552***</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>( T_2^2 )</td>
<td>1.000</td>
<td>0.552***</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Notes: *** \( \text{Sig} < 0.001 \); ** \( \text{Sig} < 0.01 \).

Table 3: Collinearity correlation of independent variables.

Time is not obvious, but the EC of pastureland, built-up land (Fig. 4a) and the arable land (Fig. 4b) decreased obviously. Water EC (Fig. 4b) dropped during the whole period, except a rapid increase from 1995 to 1996 that could be resulted from the construction of regional water conservancy projects in 1995. Fig. 5 shows that Henan’s EF is higher than China’s average (World Wildlife Fund, 2006). Furthermore, Henan’s EF is only 0.44 ha/cap, which is less than the national average level (0.80 ha/cap). Therefore, the ED of Henan Province (1.43 ha/cap) is much more than the national average level (0.80 ha/cap), which means that “human–land” relationship in Henan Province is tenser than China’s average and Henan’s anthropogenic ecological impact has become a more important social problem in China.

![Fig. 4. Computation results of EC for different land types in Henan Province in 1983–2006.](image)

![Fig. 5. Differences in EF, EC and ED between Henan and the average level in China in 2003.](image)
from $T_a1$ and $T_b1$ are added in Model 2 and Model 3, respectively. The explanation capacity of the two model are improved slightly (Adjusted $R^2 = 0.996-0.995$), which means the two drivers have about the same net effects on EF. The effect from $T_a1$ is positive but the effect from $T_b1$ is negative ($-0.333 < 0$). We also can get the similar result from Model 4, where the cumulative effect of the two drivers ($T_a1$ and $T_b1$) are considered but all of their contributions to the EF's change are not important (Sig $>0.05$).

Moreover, there is no curvilinear relationship between economic development and ecological impact (EF) in Table 2, because the coefficients of $A_2$ are positive in Model 1 to Model 4, which indicates that the classical EKC hypothesis does not exist in the Henan's EF. This result is consistent with the study results of Xu and Cheng (2005) and Dietz et al. (2007).

In summary, the results of STIRPAT model show that the major driving forces of EF are $P$, $A_1$, $A_2$, $T_a1$ and $T_b1$. However, there is a strong collinearity among these independent variables, thus the accuracy and credibility of the STIRPAT model are not very high according to Wang (1999). The testing methods of collinearity contain correlation matrix method and VIF method (Wang, 1999). The correlation matrix method is utilized in this paper and the correlation coefficients of independent variables are showed in Table 3. It is obvious that the collinearity among these independent variables is strong because there are many correlation coefficients that are near 1 (very high).

The PLS method was used in this paper in order to reduce the collinearity impact of these variables themselves. First of all, the applicability of PLS Method should be tested by checking the $T^2$ oval plot and $t_1/u_1$ scatter plot. The sample data in this paper are acceptable because there is only one inhomogeneous point in the $T^2$ oval plot of sample data (Fig. 6). It is obvious that the $t_1/u_1$ relationship of the sample data is nearly linear (Fig. 7), thus PLS linear regression model built is reasonable to the study problem of this paper.

The Observed vs. Predicted Plot showed a perfect linear relationship between the predictive value (YPred) and the actual value (YVar) (Fig. 8). This means that explanation effect of the results made by the PLS method is acceptable. The R2VY (cum) in the Coef fi cients Table (Table 4) denotes the cumulative fraction of the variation of the Y variables explained after the selected component. R2VY (cum) reached 0.960 after the $t_1$ component extracted, which means that the extracted components $t_1$ from the X variables can explain 96.0% information of the Y variables. When $t_1$ and $t_2$ were extracted, they can explain 98.1% information of the Y variables. When the three components $t_1$, $t_2$ and $t_3$ were all extracted from the X variables, they can explain 99.5% information of the Y variables. These R2VY (cum) also showed a good explanation effect of the PLS method.

Moreover, Table 4 also shows the regression coefficients of the PLS method. It can be seen that the coefficient of $P$ was 0.177 after the $t_1$ component extracted. When $t_1$, $t_2$ were extracted, the coefficient had changed into 0.169 and when $t_1$, $t_2$ and $t_3$ were all extracted, the coefficient had changed into 0.231 again. These showed that a 1% change in population ($P$) may induce about 0.169–0.231 percentages change in EF or $P$ had an elasticity of about 0.169–0.231. Similarly, $A_1$, $A_2$, and $T_b1$ had an elasticity of 0.195–0.301, 0.188–0.256 and 0.192–0.278. This suggests that $P$, $A_1$, $A_2$ and $T_b1$ may be the major drivers of Henan’s EF. However, the coefficients of $T_a1$, $T_a2$ and $T_b2$ had changed from negative (positive) to positive (negative) number and they were all small ($\leq 0.131$), which means that their impact to EF are not important or they are not the major drivers of Henan’s EF. About this point, the same results could be easily concluded from the VIP plot, in which the statistical importance of these independent variables ($X$) relative to the dependent variables $Y$ (EF) is revealed in a graphic.

<table>
<thead>
<tr>
<th>Components extracted</th>
<th>$t_1$</th>
<th>$t_1$ and $t_2$</th>
<th>$t_1$, $t_2$, and $t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.443</td>
<td>2.443</td>
<td>2.443</td>
</tr>
<tr>
<td>$P$</td>
<td>0.177</td>
<td>0.169</td>
<td>0.231</td>
</tr>
<tr>
<td>$A_1$</td>
<td>0.195</td>
<td>0.240</td>
<td>0.301</td>
</tr>
<tr>
<td>$A_2$</td>
<td>0.188</td>
<td>0.235</td>
<td>0.256</td>
</tr>
<tr>
<td>$T_a1$</td>
<td>$-0.12$</td>
<td>$-0.04$</td>
<td>0.01</td>
</tr>
<tr>
<td>$T_a2$</td>
<td>$-0.12$</td>
<td>$-0.04$</td>
<td>0.01</td>
</tr>
<tr>
<td>$T_b1$</td>
<td>0.152</td>
<td>0.226</td>
<td>0.278</td>
</tr>
<tr>
<td>$T_b2$</td>
<td>0.134</td>
<td>0.331</td>
<td>$-0.02$</td>
</tr>
<tr>
<td>R2VY(cum)</td>
<td>0.960</td>
<td>0.981</td>
<td>0.995</td>
</tr>
</tbody>
</table>
5. Conclusions

(1) Henan’s EF doubled nearly two times during the past 1983–2006 years. But at the same time, the supporting capability of regional natural resources to human activities (EC) was rather low and in a state of slow decline. Thus, Henan’s ecological deficit (ED) was bigger and bigger and its “man–land” relationship was tenser and tenser over time. In particular, Henan’s EF was more than the national average level in recent years such as 2003, which means that ecological condition of Henan Province has become a remarkable social problem. The local governments need to pay much more attention to it.

(2) Compared with the STIRPAT model, the analysis results of the PLS method are more reasonable and acceptable because the collinearity among these drivers themselves cannot be avoided in the STIRPAT model, which can bring some negative impacts to the ultimate results, but the collinearity can hardly impact the PLS method. In addition, the analysis results of EF’s drivers by STIRPAT model showed that the major driving forces of Henan’s EF were P, A1, A2, T1, and T2. The most important driving forces is P (its elasticity is extremely high (over 2)) and the effect of the other drivers (their elasticities are low) almost can be ignored. The analysis results of the STIRPAT model are so extreme that it is inconsistent with the fact. In reality, the impacts of the A (Affluence) and T (Technology) condition to the regional EF were still large and important (cannot be ignored). However, the analysis results by PLS method showed that the major drivers were P, A1, A2 and T1, and T2, was excluded as the unimportant drivers. Moreover, the P was not among the extremely important driving forces and the impact of P to EF was almost the same as the other drivers. Thus, the results using the PLS method are more reasonable and acceptable.

(3) The results acquired by using the both methods showed that the curvilinear relationship between economic development and ecological impact (EF) or the classical EKC hypothesis does not stand in Henan’s EF.

Acknowledgements

This work was financially supported by 100 Talents Programme of Chinese Academy of Sciences: Comprehensive study on sustainable urban construction. We would like to appreciate Xiaobo Xiong, Rencai Dong, Bofu Zheng and Feng Liu for their kind help in data collection and valuable comments. We are grateful to Professor Guofan Shao from Purdue University and Dr. Helen Lee from Taylor and Francis publishing company in English revision for their help in English language embellishment. We also express our thanks to two anonymous reviewers and the editors for their helpful comments and constructive suggestions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolecon.2009.05.012.

References


