SPATIO-TEMPORAL EVOLUTION AND DRIVING FACTORS OF ENERGY ECO-EFFICIENCY IN THE YANGTZE RIVER DELTA DURING 2016-2020: NOVEL EVIDENCE FROM EXPLAINABLE MACHINE LEARNING

TIAN, Y. C.[1*] – LI, M. Z.[2] – TANG, L.[3] – ZHAO, S.[1] – MA, T.[1]

[1]*School of Law, Anhui Normal University, Wuhu, China*

[2]*Physical Fitness and Scientific Exercise Research Centre, China Institute of Sport Science, Beijing, China*

[3]*School of Marxism, Anhui Agricultural University, Hefei, China*

**Corresponding author*
e-mail: tianyichuan@ahnu.edu.cn

(Received 12th Sep 2023; accepted 30th Oct 2023)

Abstract. Improving energy eco-efficiency is key to promote high-quality development in the Yangtze River Delta. In this study, we measured the energy eco-efficiency of the Yangtze River Delta region from 2016 to 2020 using a slacks-based measure model and analyzed the spatio-temporal evolution and the driving factors with the help of data visualization techniques and explainable machine learning algorithms. The results showed that:(1) Energy eco-efficiency in the Yangtze River Delta between 2016 and 2020 showed a U-shaped curve trend. (2) The energy eco-efficiency of cities in the Yangtze River Delta was gradually polarizing, with high-efficiency cities having a more even regional distribution in 2016, but showing a tendency to cluster in the north-central of the Yangtze River Delta in 2020. (3) Among the driving factors, urbanization, human capital, industrial structure, and transport infrastructure had a positive correlation with energy eco-efficiency, with only industrial dependence showing a significant negative correlation with energy eco-efficiency. (4) The coordinated promotion of transport infrastructure development, industrial structure upgrading and human capital enhancement had led to the continuous improvement of energy eco-efficiency in Chuzhou City.
Keywords: *energy eco-efficiency, case study, SHapley Additive exPlanations, slacks-based measure, Yangtze River Delta*

Introduction

Industrialization has played an important role in the economic development of developing countries, but the increase in industrial output also means that more energy is needed to maintain a consistent pace of economic growth (Shahbaz and Lean, 2012). Meanwhile, in China’s energy consumption structure, fossil energy sources such as coal still dominate (Zhao and Luo, 2018), while the combustion of fossil energy sources produces greenhouse gases such as carbon dioxide with air pollutants such as sulphur dioxide (Shindell and Smith, 2019). Over the past few decades, the development pattern at the expense of massive consumption of fossil energy has posed considerable ecological challenges, with the Yangtze River Delta (YRD) being particularly challenged. The YRD is one of the most active, open, and innovative regions in China, but it is also one of the biggest energy users with relatively low energy resource endowments (Wu et al., 2021). From 1990 to 2016, total energy consumption in YRD increased from 112.64 million tons of standard coal to 630.42 million tons of standard coal, and its share of total national energy consumption increased from 11.81% to

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

15.56% (Yu et al., 2020). Therefore, economic development in YRD has been accompanied by heavy inputs of energy factors and is characterized by crude growth. As a demonstration area for the construction of a modernized economic system, YRD urgently needs to explore a sustainable development path that balances energy use and ecological construction.

On the one hand, the concept of eco-efficiency has been widely used since its conception to measure the level of coordinated development of economic activities and eco-efficiency (Schaltegger and Sturm, 1990). On the other hand, energy efficiency connotes the contribution of energy consumption to the maintenance and promotion of sustainable human development. The core idea of both is to produce higher social value with less energy consumption and smaller environmental impact (Guan and Xu, 2016), which has provided a theoretical basis for the rise of energy eco-efficiency (EEE) research. Currently, relevant research in EEE is characterized by the refinement of the study area and the depth of the study field. Since Guan and Xu (2016) pioneered the use of SBM models to measure EEE in Chinese provinces during 1997-2012, research in this area has begun to flourish. Overall, the EEE study area is becoming increasingly refined. For instance, Chen et al. (2021) explored the evolutionary trends and drivers of EEE in eight comprehensive economic zones in China from a regional research perspective using the SFA model based on Shephard energy distance function. You et al. (2022) used the Super-SBM model with spatial analysis to deeply explore the evolutionary characteristics and spatio-temporal non-stationarity of the EEE of 108 cities in the Yangtze River Economic Belt during 2011-2020. Peng et al. (2020) confirmed the existence of a significant spatial correlation network of EEE among different regions based on panel data from 13 cities in Jiangsu. These studies provided more accurate data and more comprehensive perspectives, which contributed to the development of regionally targeted environmental policies. The research field of EEE is also progressively deepening, and scholars have already explored EEE in subsectors such as the manufacturing (Li et al., 2019), the logistics (Yu et al., 2023), and the shipping industry (Cui et al., 2022b).

With the massive adoption of machine learning, scholars are becoming dissatisfied with its “black box” properties, which has led to the rise of interpretable machine learning. Existing methods for explaining machine learning models can be broadly categorized into model-based analysis and post-hoc analysis (Murdoch et al., 2019). The former focuses on explaining the behavior and predictions of the model through the structure and parameters of the model itself. For instance, the regression coefficients of linear regression can be used to determine the degree of influence of the independent variable on the dependent variable, and the branching conditions of the decision tree can be used to understand the internal operating mechanism of the model. The latter is relatively “independent” of the model to be explained and does not require the simplicity of model structure (Wen et al., 2022). The SHapley Additive exPlanations (SHAP) method, proposed by Lundberg and Lee (2017), is one of the most used post-hoc analyses. In the field of energy sciences, Białek et al. (2022) used the ANN-SHAP model to predict the total heat demand of a district heating network and provided easily interpretable insights about the inner workings of the model, which is pioneering for the generalization of the SHAP method. However, most of the current studies identifying the driving factors of EEE are based on the Tobit model and spatial econometric models, and almost no scholars have analyzed them using machine learning algorithms.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

Therefore, this study has three outstanding contributions: (1) The application of machine learning algorithms to explore the driving factors of EEE provided novel evidence while improving the performance of the models; (2) The SHAP methods were used to endow the results derived from the machine learning algorithms with interpretability; (3) The refinement of the study sample to the Yangtze River Delta region facilitated the provision of targeted policy recommendations.

Materials and methods

Research area

The YRD is one of the fastest-growing regions in East Asia and an important intersection of the Belt and Road and the Yangtze River Economic Belt (Liang et al., 2022). As shown in *Figure 1*, in terms of administrative division, the YRD consists of one municipality (Shanghai) and three provinces (Zhejiang, Anhui, and Jiangsu) (NDRC, 2016). As the economy of the YRD continues to grow, its energy consumption also shows an increasing trend. From 1990 to 2016, total energy consumption in YRD increased from 112.64 million tons of standard coal to 630.42 million tons of standard coal, and its share of total national energy consumption increased from 11.81% to 15.56% (Yu et al., 2020). Increasing total energy consumption poses the dual challenge of energy conservation and emission reduction, which makes it relevant to explore the spatio-temporal evolution and driving factors of EEE in the YRD region.

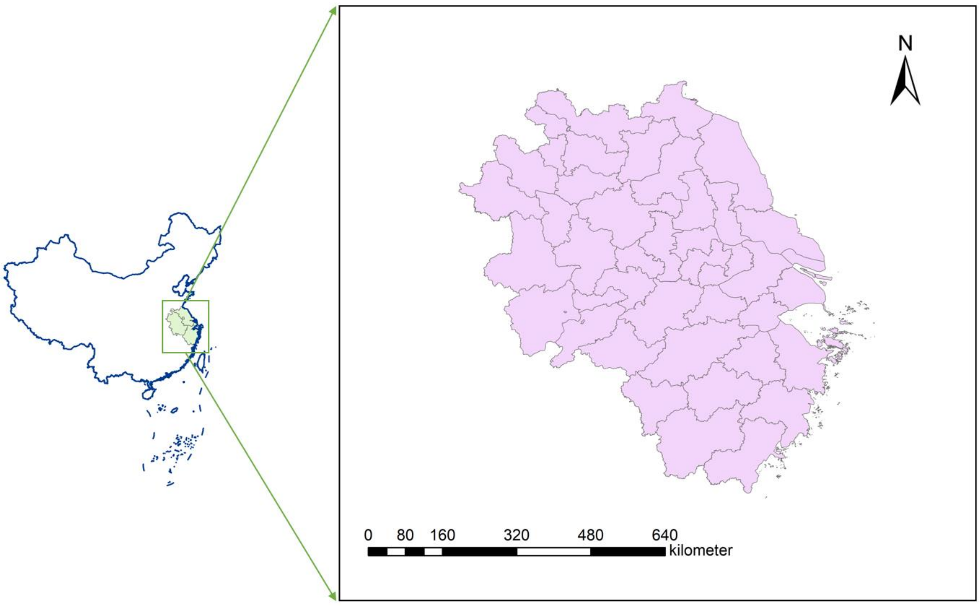

Figure 1. *Location map of the Yangtze River Delta*

Research methods

Traditional DEA models do not consider the effect of slack variables on efficiency values, nor do they take undesirable outputs into account, and the resulting efficiency values are subject to estimation bias. To address these issues, Tone (2001) proposed the Slacks-based measure (SBM) model. By maximizing the values of the slack variables,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

the SBM model can more accurately assess the efficiency of the decision unit. On this basis, Tone (2004) further constructed the SBM model containing undesirable outputs, thus solving the problem of efficiency evaluation in the presence of undesirable outputs. The SBM model with undesirable outputs based on variable returns to scale can be expressed as:

$$\rho = min \frac{1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{io}}}{1 + \frac{1}{s_1 + s_2}\left(\sum_{r=1}^{s_1}\frac{s_r^g}{y_{ro}^g} + \sum_{r=1}^{s_2}\frac{s_r^b}{y_{ro}^b}\right)} \qquad \text{(Eq.1)}$$

$$s.t.\begin{cases} x_o = X\lambda + s^- \\ y_o^g = Y^g\lambda - s^g \\ y_o^b = Y^b\lambda + s^b \\ s^- \geq 0, s^g \geq 0, s^b \geq 0, \lambda \geq 0 \end{cases} \qquad \text{(Eq.2)}$$

In the above equation, x, y^g, y^b represent the input variables, desired output variables, and undesirable output variables of the decision units, respectively. s^-, s^g, s^b denote slack variables for inputs, desired outputs, and undesired outputs, respectively. λ is the weight vector and subscript o is the evaluated unit. The objective function value ρ is EEE, which is strictly monotonically decreasing with respect to s^-, s^g, s^b, and when $s^- = s^g = s^b = 0$, there exists an optimal solution to the function, that is, $\rho = 1$, at which point the decision units can be considered sufficiently effective. If $0 < \rho < 1$, it means that there is a loss of efficiency in the decision units, and improvements can be made in inputs and outputs accordingly.

Random forest is a classical machine learning algorithm proposed by Breiman (2001), its core idea lies in classification and regression tree (CART) with bootstrap aggregation (bagging). CART is a dichotomous recursive segmentation technique that can split the sample set into 2 subsets at each node except the leaf nodes. In regression problems, the division is generally based on the minimum mean square error (Wei et al., 2021). In *Equation 3*, T, s respectively represent the division features and their corresponding division points. After division, the current dataset can be divided into the left dataset D_1 and the right dataset D_2, and c_1, c_2 respectively denote the sample output mean values of D_1, D_2. On this basis, the values of each sample of each feature are traversed, the sum of the squared errors of the predicted mean y_i and the output mean is calculated, and the resulting optimal value is used as the next division point. The principle is continuously used to divide at each branch node until the stopping condition is satisfied. The specific application process of Bagging in Random Forest is to have put back several data selected from the training set as the data for constructing several CART trees, and the rest of the unsampled samples are called out of bag (OOB). As shown in *Equation 4*, the probability that a sample is drawn at a particular random sampling session is $\frac{1}{m}$, then the probability that it is not drawn is $1 - \frac{1}{m}$, and the joint probability that it is not drawn at all m times is $\left(1 - \frac{1}{m}\right)^m$. The limit can be found to be $\frac{1}{e}$ as m tends to infinity, so about 36.8% of the samples in the training set become OOBs in each round of random sampling in bagging. Since each tree makes a prediction and

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

gets a regression value based on its training data, a voting process is needed to get the prediction result of the random forest regression (RFR), which is usually done by taking the average value.

$$min_{T,S} = \left\{ \sum_{x_i \in D_1(T,s)} (y_i - c_1)^2 + \sum_{x_i \in D_2(T,s)} (y_i - c_2)^2 \right\} \tag{Eq.3}$$

$$\lim_{m \to \infty} \left(1 - \frac{1}{m}\right)^m \to \frac{1}{e} \tag{Eq.4}$$

This study also employed three of the most commonly used regression algorithm evaluation metrics to assess the model, specifically including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2), as shown in the following formulas:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{Eq.5}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|(y_i - \hat{y}_i)| \tag{Eq.6}$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \tag{Eq.7}$$

Based on the Shapley value, Lundberg and Lee (2017) proposed the SHapley Additive exPlanations (SHAP) and applied it to improve the interpretability of machine learning algorithms. The explanatory model f is a linear function in the SHAP method, as shown in the following equation (Zheng et al., 2023):

$$f(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i' \tag{Eq.8}$$

where z' indicates whether input parameter i exists in the input parameter set or not, $z_i' = 1$ when i is in the input parameter set; and $z_i' = 0$ when it is not in the input parameter set. M and ϕ respectively denote the number and attributes of the input parameters. The specific SHAP values of the input features in the model are derived from *Equation 9*:

$$\phi_i = \sum_{T \subseteq N_{\{i\}}} \frac{|T|!\,(N - |T| - 1)!}{N!} \left(g_{T \cup \{i\}}(T \cup \{i\}) - g_T(T)\right) \tag{Eq.9}$$

where N is the set of all input parameters; T is the set of all input parameters except i; g is the prediction model, and $!$ is the factorial.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

Variable description and data sources

In terms of EEE measurement, this study used 41 cities in the 2016-2020 YRD as the study sample and drew on Peng et al. (2020) to select input indicators from energy, labor, and capital, and output indicators from both desired and undesirable aspects. The details of each indicator are shown in *Table 1*. Since the total energy consumption of each city is not directly given in the statistical yearbook, this study estimated it through the total gas supply (artificial, natural gas, LPG) and the electricity consumption of the whole society. The capital stock was estimated by using the perpetual inventory method with a base period of 2006 and a depreciation rate of 9.6%. CO2 emissions for each city were obtained from the CEADs database (Shan et al., 2022). And the remaining variables were mainly obtained from *China City Statistical Yearbook*.

Table 1. *The input-output indicators of the SBM model*

Target	Criteria	Indicator	Unit
Input	Energy input	Total energy consumption in the region	Ten thousand tons of standard coal
	Labor input	Industrial employment	Ten thousand people
	Capital input	Capital stock	Billion yuan
Output	Expected output	Regional GDP	Billion yuan
		Per capita urban green space area	Hectares
	Unexpected output	Carbon dioxide emissions	Million tons
		Industrial smoke and dust emissions	Tons
		Industrial wastewater discharge	Ten thousand tons
		Industrial sulfur dioxide emissions	Tons

EEE is a measure of the strong relationship between energy, economy, environment, and society, so there are many factors affecting EEE (Cui et al., 2022a). The enhancement of human capital can promote green technological innovation and thus increase EEE; industrial structure and the degree of industrial dependence have a direct impact on EEE, and the development of high value-added and low-energy industries can increase EEE; the government can guide enterprises towards environmental protection and energy conservation through taxation and subsidies; urbanization and increased population density can promote the intensive use of energy and help to control the undesirable output of energy consumption; the improvement of transport infrastructure can reduce transport energy consumption and transport pollutant emissions; and the introduction of foreign investment can bring in advanced energy technology and energy management experience. In summary, factors such as human capital, industrial structure, industrial dependence, fiscal intervention, urbanization, population density, transport base, and foreign investment all theoretically affect EEE to varying degrees, so this study included these factors as independent variables in the model. The primary source of the data used is the *China City Statistical Yearbook*, and missing values for a small portion were supplemented by consulting statistical yearbooks of various provinces and cities. The descriptive statistics for each variable are shown in *Table 2*, the raw data had fewer missing values and had been filled with linear interpolation.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

Table 2. *Descriptive statistics*

Variables	Measurement	Min	Max	Std	Mean
EFF (Energy eco-efficiency)	The results are estimated using the SBM-DEA model	0.096	1	0.351	0.519
HC (Human capital)	Number of students enrolled in higher education institutions	0.218	10.321	1.695	1.902
IS (Industrial structure)	Proportion of tertiary industry in GDP	34.440	73.150	6.964	48.564
FI (Fiscal intervention)	Proportion of general budgetary expenditure of local finance in GDP	2.052	25.556	4.019	8.712
UR (Urbanization)	Proportion of urban residents in total population	38.280	89.309	11.031	63.890
PD (Population density)	Ratio of permanent population to administrative area	152.924	2327.709	348.438	658.655
TI (Transportation infrastructure)	Per capita road mileage	4.370	45.530	7.205	24.092
FO (Foreign investment)	Proportion of actual utilization of foreign investment in GDP	0.269	9.317	1.784	2.541
IN (Industrial dependence)	Proportion of industrial value added in GDP	20.628	56.456	6.665	37.268

Results and discussion

Temporal evolution

Figure 2 presents the temporal changes of EEE in the YRD region between 2016 and 2020. The bar chart represents the regional average for the YRD. Due to Shanghai consistently maintaining DEA efficiency, the line chart section only displays the temporal changes for Anhui, Jiangsu, and Zhejiang.

The EEE of YRD showed a U-shaped curve trend during 2016-2020. This may be because the Air Pollution Prevention and Control Action Plan (APPCAP) and Carbon Emissions Trading Scheme (ETS), which have been implemented since 2013, respectively played an important role in improving air quality and promoting carbon emission reduction (Feng et al., 2019; Zhang et al., 2020), which in turn pushed the EEE to reach a staged peak in 2016. And from 2017 onwards, the proportion of clean energy in the structure of energy consumption has been gradually increasing, along with the increasing financial support and public awareness of environmental protection. In general, the combustion of fossil fuels produces greenhouse gases such as carbon dioxide along with air pollutants such as sulphur dioxide (Shindell and Smith, 2019), whereas the promotion of cleaner energy sources eases the dependence on traditional fossil fuels, promotes synergistic control of air pollution and carbon emission and ultimately contributes to the rebound of EEE. However, it should be noted that as of 2020, the EEE of YRD had still not yet rebounded to the peak of the phase reached in 2016, and the overall level of efficiency remains low. Furthermore, the rebound in 2019 may have been influenced to some extent by COVID-19. During the COVID-19 pandemic, global carbon dioxide emissions were reduced by over 8% due to a significant decrease in economic activities (Tollefson, 2020). In other words, while COVID-19 had a negative impact on economic development, it also brought about unexpected environmental benefits (*Fig. 2*).

Specifically, the trend of EEE fluctuations in Jiangsu Province was basically in line with the overall trend, but Zhejiang Province and Anhui Province showed two different trends. On the one hand, the rebound process of EEE in Zhejiang Province is not ideal, and even appeared to fall back again during 2019-2020, which suggests that Zhejiang Province should continue to increase the financial support for EEE improvement, and

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

realize sustainable development through the development of clean energy, the adjustment of the energy structure, as well as the strengthening of governmental regulation. On the other hand, the rebound effect of EEE in Anhui Province was much higher than the average, and its EEE value in 2020 was already higher than the value in 2016, reflecting the fact that Anhui Province had balanced economic and ecological benefits while undertaking industrial transfers from the rest of the YRD to realize the EEE rebound within the region.

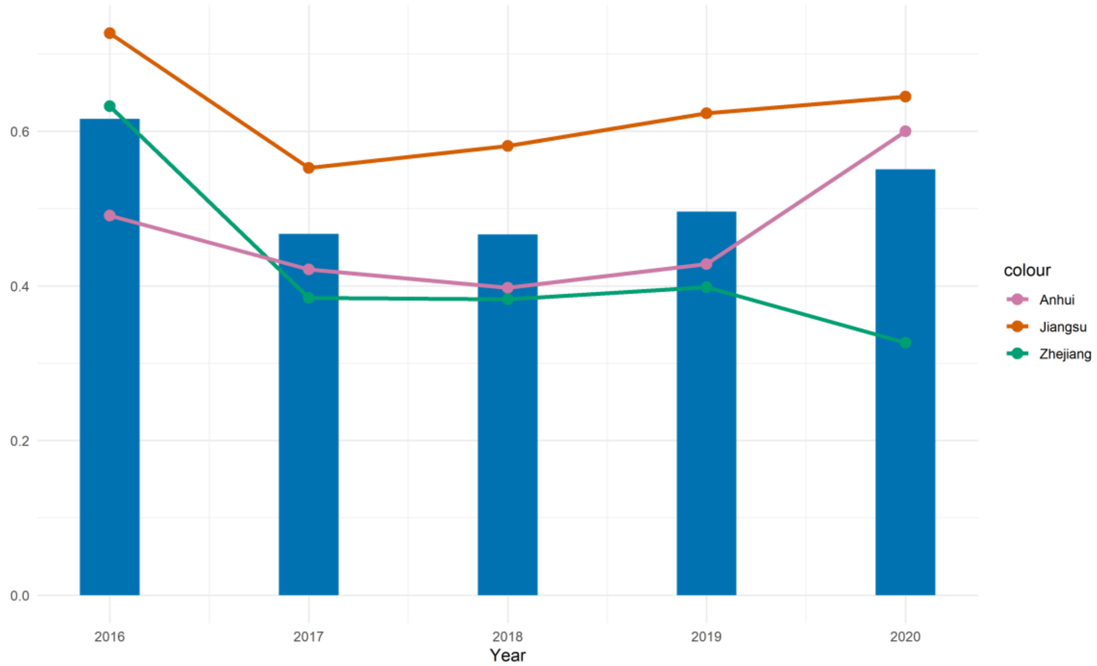

Figure 2. *Temporal evolution of energy eco-efficiency in the Yangtze River Delta from 2016 to 2020*

Spatial evolution

In the spatial dimension, 15 cities had high EEE (efficiency value of not less than 0.8) and 15 cities had low EEE (efficiency value of less than 0.4) in 2016. 16 cities had high EEE (efficiency value of not less than 0.8) and 22 cities had low EEE (efficiency value of less than 0.4) in 2020. While the number of cities with high EEE had slightly increased, the number of cities with low EEE had grown at a faster rate compared to that of cities with low EEE. This means that the EEE of YRD cities is gradually polarizing, with low EEE cities facing greater green development challenges. The increase in high EEE cities is also a positive trend, indicating that a subset of latecomer cities has made progress in efficiency improvements and that the development patterns of these cities have provided useful lessons for low EEE cities. In addition, the regional distribution of high-efficiency cities was more even in 2016, whereas in 2020 there was a trend towards clustering in the north-central part of the YRD. It is worth noting that cities such as Fuyang, Chuzhou, Xuancheng, Huaian, Yangzhou, and Changzhou had successively stepped into the high efficiency stage during 2016-2020, which had an average level of economic development but faster development compared to other cities in the YRD. This reflects the fact that YRD's latecomer cities are realizing the ecological benefits of unlocking the potential of

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

economic development. This reflects the fact that YRD’s latecomer cities have achieved a balance between unlocking the potential for economic development and ecological benefits (*Figs. 3* and *4*).

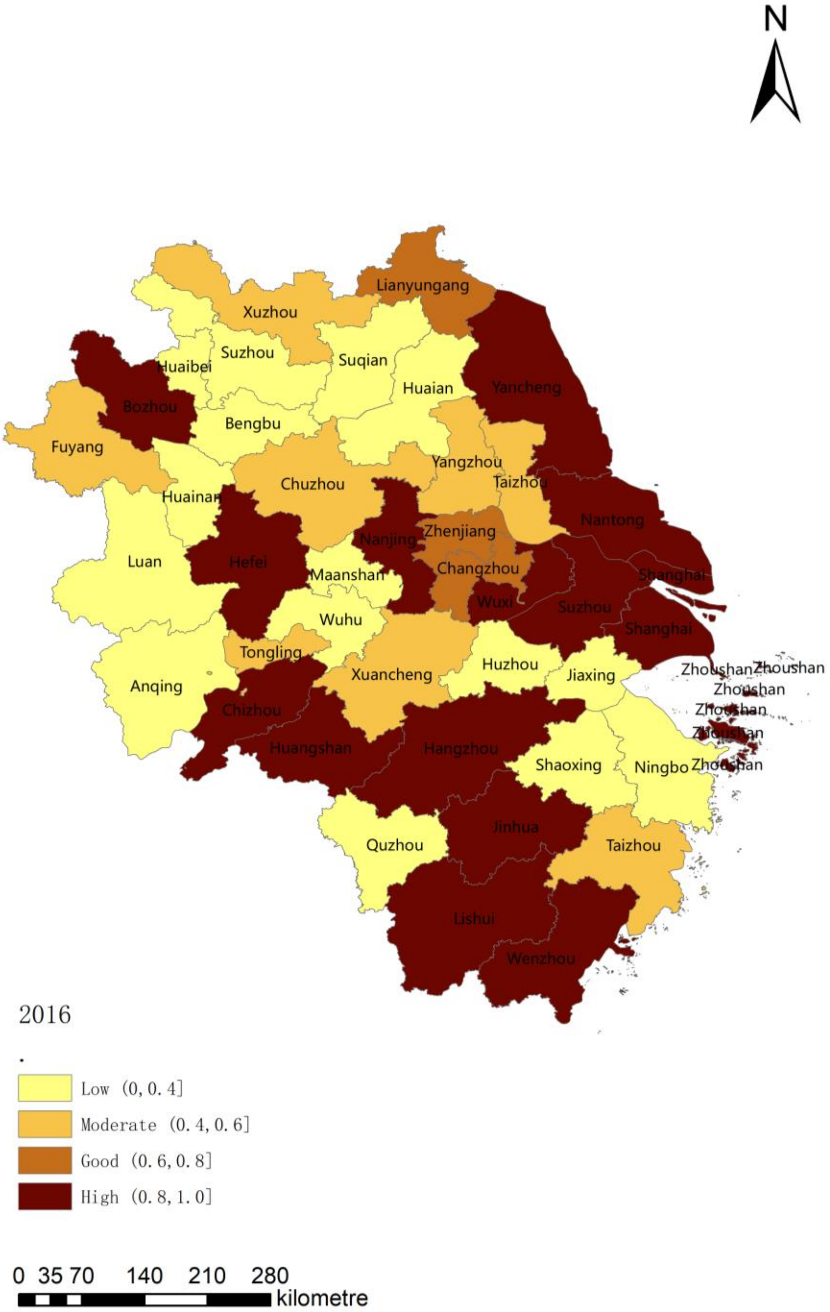

***Figure 3**. Spatial distribution of energy eco-efficiency in the Yangtze River Delta in 2016*

Algorithm and parameter

In this study, the algorithm was trained on 75% of randomly selected data and the remaining 25% was used to test the model. Secondly, to match the data with a suitable model, this study first selected three commonly used regression algorithms, namely RFR, SVR, and Lasso, and screened the optimal parameters of each algorithm with the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

help of 10-fold cross-validated GridSearchCV, and the specific results are shown in *Table 3*. On this basis, by comparing the R^2, RMSE, and MAE metrics, RFR was identified as the base algorithm, and the specific results are shown in *Table 4*. When machine learning algorithms are used to compare with traditional econometric regression, the commonly used comparison metric is R^2, and the results of this study were improved by 0.1225 and 0.07 compared to the GTWR of You et al. (2022) and the SEM of Guan and Xu (2016), respectively. The superiority of machine learning algorithms in identifying the driving factors of EEE is also confirmed to some extent, although the selected study samples are different.

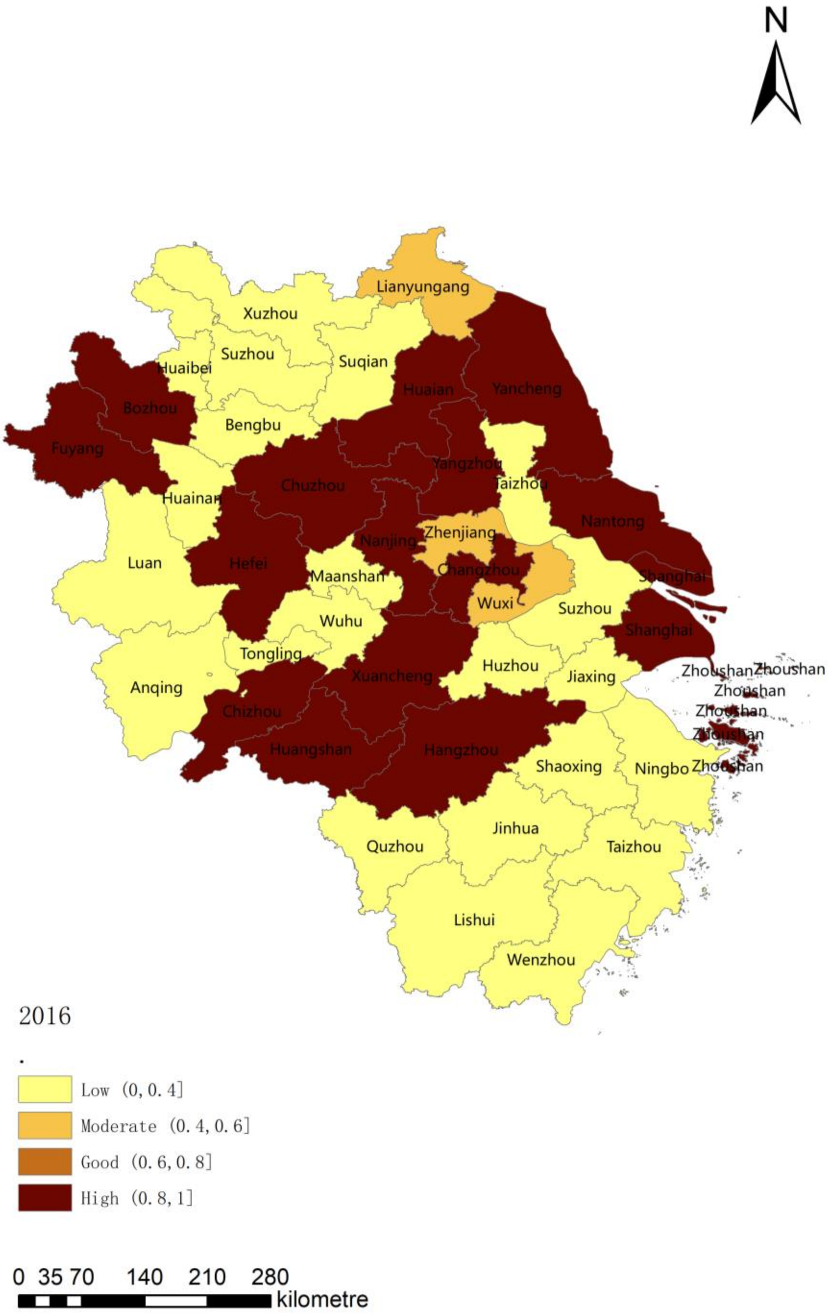

***Figure 4.** Spatial distribution of energy eco-efficiency in the Yangtze River Delta in 2020*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

Table 3. *The process of GridSearchCV optimization*

Model	Parameter	Search range	Optimal parameter
RFR	max_depth	[3,5,10,20,30]	20
	min_samples_leaf	[1,2,3,4,5]	1
	min_samples_split	[1,2,3,4,5]	2
	n_estimators	[50,100,150,200,250,300]	50
SVR	C	[0.1,1,10]	10
	gamma	[0.01,0.1,1]	0.01
	epsilon	[0.1,0.2,0.3]	0.1
LASSO	alpha	(0,1.001,0.001)	0.001
	max_iter	(100,10000,100)	100

Table 4. *Algorithm performance evaluation*

Model	R^2	RMSE	MAE
RFR	0.724	0.184	0.147
SVR	0.538	0.238	0.204
LASSO	0.330	0.286	0.246

Variable importance analysis

SHAP feature importance refers to the extent to which each feature (independent variable) contributes to improving the predictive power of the overall model, and is expressed as the absolute value of the extent to which the independent variable influences the dependent variable. The higher the importance of the feature, the greater the effect of that independent variable on EEE. As shown in *Figure 5*, the degree of influence of independent variables on EEE is PD, UR, HC, IN, IS, TI, FO, and FI in descending order. Meanwhile, it is also possible to find out the positive and negative direction of independent variables' influence on EEE through *Figure 5*. If SHAP value > 0, it means that the independent variable has a boosting effect on EEE; if SHAP value < 0, it means that the independent variable has an unfavorable effect on EEE. Among the respective variables, PD, FO, and FI had no significant positive or negative relationship with EEE, UR, HC, IS, and TI had a positive correlation with EEE, and only IN showed a significant negative correlation with EEE.

On this basis, this study plotted the dependence plots of the top four independent variables in terms of importance against EEE as a way of exploring the non-linear relationship between the independent variables and EEE, and the results are shown in *Figure 6*. Combining *Figures 3* and *4*, it can be found that PD shows a left-skewed W-shaped curve, which means that PD cannot simply be considered to have a positive or negative unidirectional effect on EEE. In other words, the effect of PD on EEE has a non-linear character. It should be noted that the points with PD > 2000 are all mapped for Shanghai in the figure, and the Shap value is all greater than 0, indicating that PD positively influenced the EEE of Shanghai in all the years between 2016 and 2020. Secondly, HC and UR showed a positive effect on EEE, with points with SHAP > 0 basically concentrated in the interval between HC > 2 and UR > 70. On the one hand, through education and training, people can learn and acquire knowledge of green

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

technologies, and high-quality human capital can also contribute to the continuous advancement of green technologies. This knowledge dissemination and technological advancement helps to promote the application and innovation of green technologies and ultimately contributes to the enhancement of EEE. On the other hand, urbanization leads to a greater clustering of economic and social activities, and this clustering effect can be an effective way to increase EEE; for instance, centralized heating systems can be used to distribute energy more efficiently to urban residents and businesses through efficient pipeline networks and equipment. In addition, IN showed an overall negative effect on EEE, with points with SHAP < 0 largely concentrated in the IN < 35 interval. When the share of industrial GDP is high, the structure of the economy may be overly dependent on the industrial sector, which in turn makes it difficult to advance the structural transformation of the economy and ultimately limits the sustainable upgrading of the electrical and electronics industry.

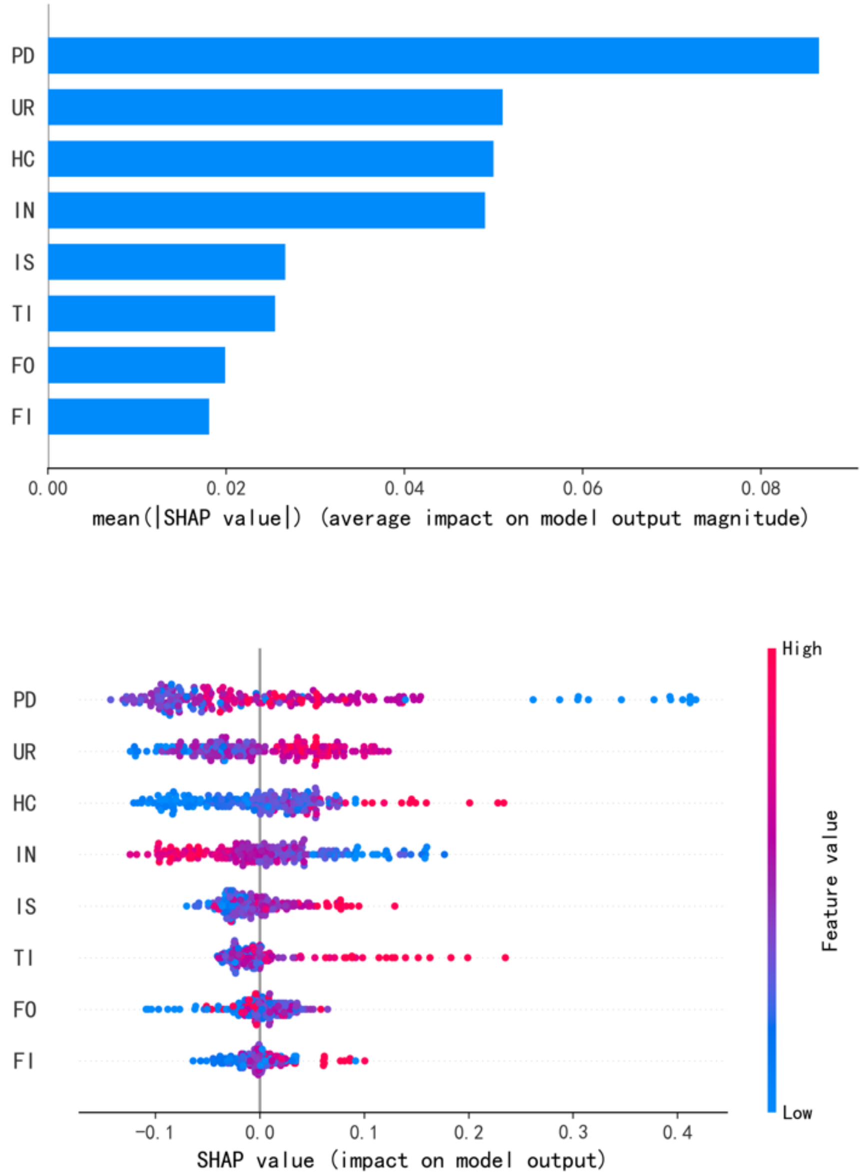

Figure 5. *Driving factor analysis of energy eco-efficiency based on SHapley Additive exPlanations*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

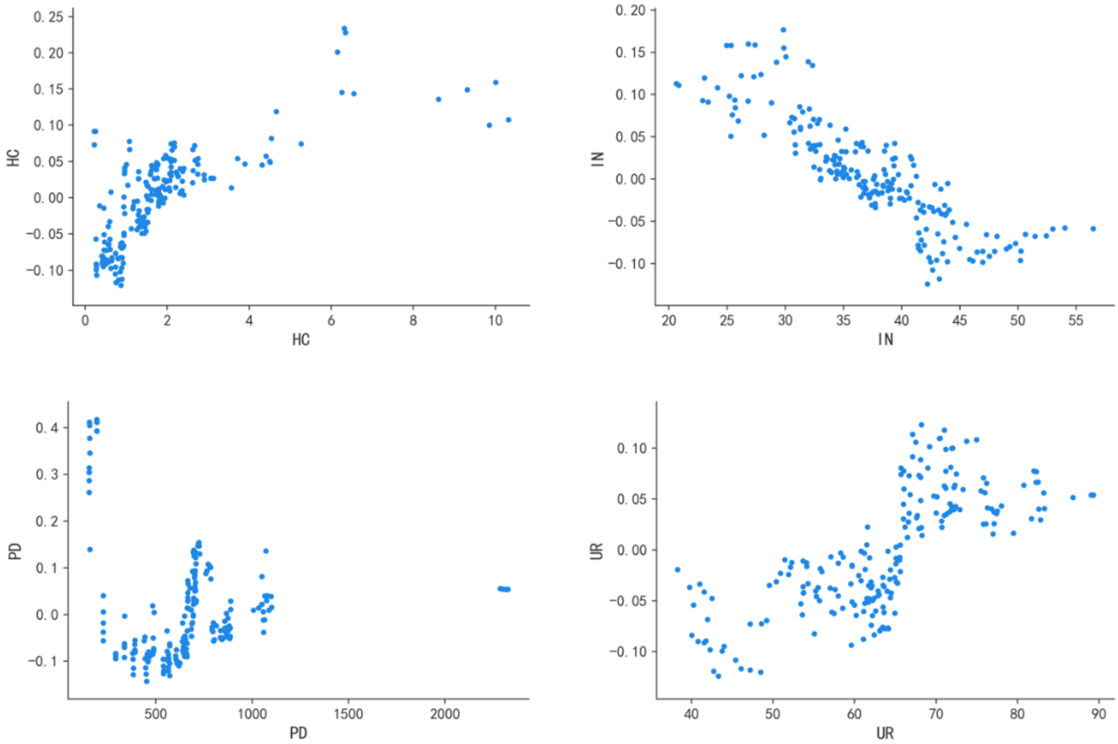

Figure 6. *Feature dependency plot based on SHapley Additive exPlanations*

Case study

In order to better summarize the experience of EEE from low to high levels, this study chose to discuss Chuzhou as a case study. The EEE value of Chuzhou City is only 0.536 in 2016, but it reached DEA validity in 2020. This trend has attracted the attention of this study because Chuzhou City, as a typical latecomer, has empirical evidence that is important for the development of other similar cities. Therefore, this study dissected the driving factors in the process of EEE value from low level to DEA effective in Chuzhou with the help of force plots of SHAP, and the results are shown in *Figures 7* and *8*. Comparing the force plots for Chuzhou City in 2016 and 2020, TI is still the most important positive driving factor, even though the rapid rise in total population has caused TI to decline. Secondly, the decline in IN turned its impact on EEE from negative to positive. Over time, the industrial structure of Chuzhou City changed, starting from a traditional industrial dominance to transition to the dominance of high-technology industries and services. This structural shift reduced overdependence on energy, thereby increasing energy eco-efficiency. In addition, HC and FO boost provide an important role. Improvement in the level of human capital can effectively promote technological progress and green innovation, while the increase in foreign investment brings more advanced technology and management experience, both of which contribute to the improvement of energy eco-efficiency, and HC’s promotional effect is stronger. In summary, relying on the coordinated promotion of transport infrastructure construction, industrial structure upgrading and human capital enhancement, Chuzhou City achieved the progress of EEE from low level to high efficiency.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

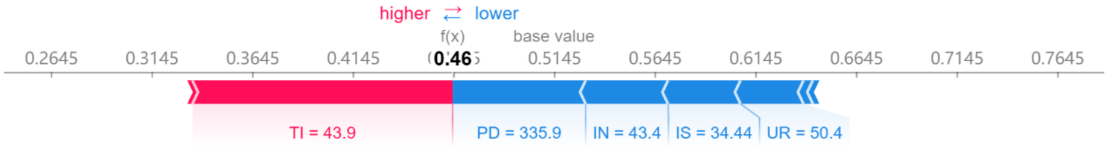

Figure 7. *Force plot of Chuzhou City in 2016*

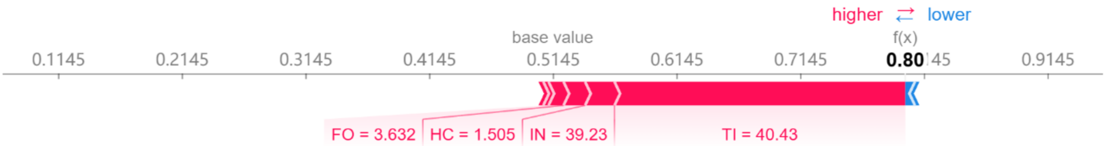

Figure 8. *Force plot of Chuzhou City in 2020*

Conclusions

This study measured the energy eco-efficiency of the Yangtze River Delta from 2016 to 2020 based on the SBM model, analyzed the spatio-temporal evolution and driving factors with the help of data visualization techniques and explainable machine learning algorithms, and finally obtained the following conclusions. (1) During 2016-2020, energy eco-efficiency in the Yangtze River Delta showed a U-shaped curve trend. Specifically, the fluctuation trend of energy eco-efficiency in Jiangsu Province is basically consistent with the overall trend, the rebound process of energy eco-efficiency in Zhejiang Province is not satisfactory, and the rebound effect of energy eco-efficiency in Anhui Province is much higher than the average. (2) From the spatial dimension, the energy eco-efficiency of the cities in the Yangtze River Delta is gradually polarizing, with high-efficiency cities having a more even regional distribution in 2016, while in 2020 there is a tendency to cluster towards the north-central part of the Yangtze River Delta. (3) The degree of influence of the drivers on energy eco-efficiency is, in descending order, population density, urbanization, human capital, industrial dependence, industrial structure, transport infrastructure, foreign investment, and fiscal intervention. Among them, population density, foreign investment, and fiscal intervention have no significant positive or negative relationship with energy eco-efficiency, urbanization, human capital, industrial structure, and transport infrastructure have a positive relationship with energy eco-efficiency, and only industrial dependence shows a significant negative relationship with energy eco-efficiency. (4) As a typical representative of a latecomer city, Chuzhou City achieved energy eco-efficiency from low level to high efficiency by relying on the coordinated advancement of transport infrastructure construction, industrial structure upgrading, and human capital enhancement.

Some limitations of this study provide directions for future research. Firstly, the sample data is only 205 cases, which is not conducive to taking advantage of machine learning algorithms, and future research can overcome this limitation by expanding the sample area and sample year. Secondly, the relationship between the driving factors and energy eco-efficiency identified in this study is correlational rather than causal, and future research can be expanded by applying causal inference techniques. Finally, as with most studies of this type, the undesirable output indicators in this study do not reflect water and soil pollution to an adequate degree, largely due to limitations in data availability. We hope that the above limitations can be addressed by future research.

REFERENCES

[1] Białek, J., Bujalski, W., Wojdan, K., Guzek, M., Kurek, T. (2022): Dataset level explanation of heat demand forecasting ANN with SHAP. – Energy 261: 125075.

[2] Breiman, L. (2001): Random Forests. – Machine Learning 45: 5-32.

[3] Chen, J., Liu, N., Ma, X. (2021): Energy eco-efficiency measurement and driving factors of China's eight comprehensive economic zones (in Chinese). – China Environ. Sci 41: 2471-2480.

[4] Cui, S., Wang, Y., Zhu, Z., Zhu, Z., Yu, C. (2022a): The impact of heterogeneous environmental regulation on the energy eco-efficiency of China's energy-mineral cities. – Journal of Cleaner Production 350: 131553.

[5] Cui, Y., Qiu, K., Li, G., Jiang, H., Kong, L. (2022b): Spatiotemporal differentiation of energy eco-efficiency of shipbuilding industry in China. – Ocean & Coastal Management 230: 106347.

[6] Feng, Y., Ning, M., Lei, Y., Sun, Y., Liu, W., Wang, J. (2019): Defending blue sky in China: effectiveness of the "Air Pollution Prevention and Control Action Plan" on air quality improvements from 2013 to 2017. – Journal of Environmental Management 252: 109603.

[7] Guan, W., Xu, S. (2016): Study of spatial patterns and spatial effects of energy eco-efficiency in China. – Journal of Geographical Sciences 26: 1362-1376.

[8] Li, G., Liu, J., Li, T. (2019): Regional differences of energy eco-efficiency in manufacturing industry under consideration of undesirable outputs based on the SBM-TOBIT two-stage model (in Chinese). – Chin. J. Manag. Sci 27: 76-87.

[9] Liang, M. S., Huang, G. H., Chen, J. P., Li, Y. P. (2022): Energy-water-carbon nexus system planning: a case study of Yangtze River Delta urban agglomeration, China. – Applied Energy 308: 118144.

[10] Lundberg, S. M., Lee, S.-I. (2017): A unified approach to interpreting model predictions. – Advances in neural information processing systems 30. https://doi.org/10.48550/arXiv.1705.07874.

[11] Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B. (2019): Definitions, methods, and applications in interpretable machine learning. – Proceedings of the National Academy of Sciences 116: 22071-22080.

[12] NDRC (2016): Development Plan of the Yangtze River Delta City Urban agglomeration (2015-2030). – https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/201606/W020190905497826154295.pdf.

[13] Peng, B., Wang, Y., Wei, G. (2020): Energy eco-efficiency: Is there any spatial correlation between different regions? – Energy Policy 140: 111404.

[14] Schaltegger, S., Sturm, A. (1990): Ökologische Rationalität: Ansatzpunkte zur Ausgestaltung von ökologieorientierten Managementinstrumenten. – Die Unternehmung 44: 273-290.

[15] Shahbaz, M., Lean, H. H. (2012): Does financial development increase energy consumption? The role of industrialization and urbanization in Tunisia. – Energy Policy 40: 473-479.

[16] Shan, Y., Guan, Y., Hang, Y., Zheng, H., Li, Y., Guan, D., Li, J., Zhou, Y., Li, L., Hubacek, K. (2022): City-level emission peak and drivers in China. – Science Bulletin 67: 1910-1920.

[17] Shindell, D., Smith, C. J. (2019): Climate and air-quality benefits of a realistic phase-out of fossil fuels. – Nature 573: 408-411.

[18] Tollefson, J. (2020): How the coronavirus pandemic slashed carbon emissions--in five graphs. – Nature 582: 158-160.

[19] Tone, K. (2001): A slacks-based measure of efficiency in data envelopment analysis. – European Journal of Operational Research 130: 498-509.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144

[20] Tone, K. (2004): Dealing with undesirable outputs in DEA: a Slacks-Based Measure (SBM) approach. – Nippon Opereshonzu, Risachi Gakkai Shunki Kenkyu Happyokai Abusutorakutoshu 2004: 44-45.
[21] Wei, Q., Chen, S., Huang, W., Ma, G., Tao, C. (2021): Forecasting method of clearing price in spot market by random forest regression (in Chinese). – Proc. CSEE 41: 1360-1367.
[22] Wen, X., Xie, Y., Jiang, L., Li, Y., Ge, T. (2022): On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development. – Accident Analysis & Prevention 168: 106617.
[23] Wu, W., Zhang, T., Xie, X., Huang, Z. (2021): Regional low carbon development pathways for the Yangtze River Delta region in China. – Energy Policy 151: 112172.
[24] You, J., Dong, H., Jiang, B., Zhu, Y., Tao, J. (2022): Spatiotemporal non-stationarity of energy ecological efficiency and its driving factors in the Yangtze River Economic Belt (in Chinese). – Resources Science 44: 2207-2221.
[25] Yu, M., Li, Y., Chen, W., Xue, Y. (2020): Decomposition analysis of driving factors for energy consumption changes in the Yangtze River Delta region: an analysis based on data from 1995 to 2016 (in Chinese). – Journal of Southeast University (Philosophy and Social Science) 22: 69-79.
[26] Yu, X., Xu, H., Lou, W., Xu, X., Shi, V. (2023): Examining energy eco-efficiency in China’s logistics industry. – International Journal of Production Economics 258: 108797.
[27] Zhang, W., Li, J., Li, G., Guo, S. (2020): Emission reduction effect and carbon market efficiency of carbon emissions trading policy in China. – Energy 196: 117117.
[28] Zhao, X., Luo, D. (2018): Forecasting fossil energy consumption structure toward low-carbon and sustainable economy in China: evidence and policy responses. – Energy Strategy Reviews 22: 303-312.
[29] Zheng, G., Zhang, Y., Yue, X., Li, K. (2023): Interpretable prediction of thermal sensation for elderly people based on data sampling, machine learning and SHapley Additive exPlanations (SHAP). – Building and Environment 242: 110602.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2201_129144