Ecological Footprint Prediction based on Global Macro Indicators in G-20 Countries using Machine Learning Approaches

Ahmad Roumiani (roumiani.ah@mail.um.ac.ir)
Ferdowsi University of Mashhad https://orcid.org/0000-0002-2818-9277

Abbas Mofidi
Ferdowsi University of Mashhad

Research Article

Keywords: Ecological footprint, prediction, variable selection, machine learning, G-20 countries

Posted Date: May 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-489246/v1

Ref.: Ms. No. ESPR-D-21-04799
Ecological footprint prediction based on global macro indicators in G-20 countries using machine learning approaches
Environmental Science and Pollution Research

Dear Dr Roumiani,

Thank you for submitting your manuscript.
After a careful editing examination of your paper, I am sorry to inform you that your manuscript is not acceptable for publication by ESPR in the present format however I encourage you to resubmit as a new manuscript after the appropriate changes.

Indeed, we have identified that several parts of the manuscript are copied/pasted from previous published publications. Your manuscript exhibits a similarity percentage higher than 25% with already published papers. The manuscript contains a number of phrases taken from other published papers, most notably in the Introduction section: (paragraph 5th), literature review section: (paragraph 1st, line no. 5-17, paragraph 2nd, line no. 11-50), result and discussion section: (paragraph 1st, 8th, 15th), but also conclusion section: (line no. 4-12). We recommend these are reworked before peer review

Please be informed that self-plagiarism isn't allowed due to copyright rules. Due to high degree of overlap, we are not able to consider your manuscript for publication in ESPR in its present form. I suggest you carefully rewrite the text to ensure that it is original and all the sources that you have based your work upon are appropriately and transparently acknowledged and cited.

I recommend also that you check the publication ethics at : http://publicationethics.org/

Thank you for your interest in publishing in ESPR journal.

Best regards,

Philippe Garrigues
Editor-in-chief
ESPR journal

Answer: Dear Editor-in-chief , Philippe Garrigues

I hope this email finds you well.

Thank you very much for giving us the opportunity to review the manuscript. It should be noted that the manuscript has been thoroughly and completely rewritten. We are proud to submit our paper to your great journal. ESPR-D-21-05761 and upload it as supplementary file.

All the best,

Ahmad Roumiani

Predicting ecological footprint based on global macro indicators in G-20 countries using machine learning approaches

Ahmad Roumiani*, Abbas Mofidi

Department of Geography, Ferdowsi University of Mashhad, Mashhad, 91735, Iran

Ahmad Roumiani, roumiani.ah@mail.um.ac.ir

Abbas Mofidi, abbasmofidi@um.ac.ir

Corresponding Author:

Department of Geography, Ferdowsi University of Mashhad, Mashhad, P.O: Box 91735, Iran.

Tel.: +98 912 422 5419. Email address: roumiani.ah@mail.um.ac.ir (Ahmad Roumiani)

Predicting ecological footprint based on global macro indicators in G-20 countries using machine learning approaches

Abstract

Paying attention to human activities in terms of land grazing infrastructure, crops, forest products and carbon impact, the so-called ecological impact (EF) is one of the most important economic issues in the world. In the present study, data from global databases were used. The ability of the penalized regression approach (PR including Ridge, Lasso and Elastic Net) and artificial neural network (ANN) to predict EF indices in the G-20 over the past two decades (1999-2018) was depicted and compared. For this purpose, 10-fold cross-validation was used to assess predictive performance and to specify a penalty parameter for PR models. Based on the results, a slight improvement in prediction performance was observed over linear regression. Using the Elastic Net model, more global macro indices were selected than Lasso. Although Lasso included only some indicators, it still had better predictive performance among PR models. Although the findings using PR methods were only slightly better than linear regression, their interest in selecting a subset of controllable indicators by shrinking the coefficients and creating a parsimonious model was apparent. As a result, penalized regression methods would be preferred, using feature selectivity and interpretive considerations rather than predictive performance alone. On the other hand, neural network-based models with higher values of coefficients of determination (R2) and values lower of RMSE than PR and OLS had significant performance and showed that they are more accurate in predicting EF. The results showed that the ANN network could provide considerable and appropriate predictions for EF indicators in the G-20 countries. predictions

Keywords: Ecological footprint, prediction, variable selection, machine learning, G-20 countries

1.Introduction

The presence of at least ten predictor variables in a linear regression model can theoretically cause multi-collinearity and over-fitting and increase the out-of-sample error. Such a model would produce unbiased estimates with significant variance, which may be completely inaccurate (Efron et al. 2004); in addition, the predictability of this model on testing set will be weak. (Cherlin et al. 2018). To deal

with these shortcomings of linear regression and reduce the variance of estimates and improve predictions, regulatory techniques (so-called penalized regression (PR)) have been proposed by introducing a slight bias into model estimation (Tibshirani 1996; Zou 2006; Hastie et al. 2009). This method creates reliable degrees of reliable and stable estimates by penalizing high coefficients and shrinking them to zero and adding a grade of bias when model training. When using predictive models, bias and variance play a key role in determining the model. Obtaining the best point between the error due to bias and variance and balancing it is called bias-variance trade-off. In this trade-off between complicated and straightforward models, medium complexity is the best possible. Bias is related to a failure to fit the training set, and variance is related to a failure to fit the testing set. A simple model has high bias and small variance, and in the case of a complex model, its opposite is true. Bias and variance can be reduced, which can minimize the mean squared error (MSE). $\mathrm{MSE} = \mathrm{Bias}^2 + \mathrm{variance} + \mathrm{Irreducible\ error}$, $\mathrm{Bias} = \mathrm{E}[\hat{\beta}x] - \beta^* x$ and $\mathrm{Variance} = \mathrm{E}[(\hat{\beta}x)^2] - \mathrm{E}[\hat{\beta}x]^2$. Here, β^* is a "true" parameter.

In general, three PR techniques have been widely welcomed in studies: Ridge (Hoerl and Kennard 1970; Le Cessie and Van Houwelingen 1992), Lasso (Least Absolute Shrinkage and Selection Operator) (Tibshirani 1996, 1997) as well as the Elastic Net (EN) (Zou and Hastie 2005). Ridge regression (RR) shrinks the regression coefficients toward zero using $\ell 2$- norm regularization (Le Cessie and Van Houwelingen 1992). To fit the training set less than ordinary squares (OLS), the coefficient values in this method are regularized by hyper-parameter λ. And because it is less sensitive to extreme variances, such as outliers, this technique can be better generalized. . Also, since RR does not impose a confidence limit, normal distribution errors are not assumed. The Lasso penalty value is adjusted using the $\ell 1$-norm setting (Tibshirani 1996), leading to variable selection and creating a sparse model.

The combination of RR and Lasso's properties is considered a particular case, i.e., EN, by setting the rule of $\ell 2$ and $\ell 1$-norm regularization (Zou and Hastie 2005). In general, these sparse models confirm bias coefficient estimates with less variance, thus improving prediction accuracy with lower MSE. By using these techniques, accuracy and simplicity are balanced. These strategies have yielded promising results for dealing with high-dimensional correlated datasets (e.g., DNA-

microarray/genomic studies) (Hastie et al. 2009; Waldron et al. 2011; Waldmann et al. 2013). Recently, good results have been reported using PR techniques in examining low-dimensional data in studies (e.g.,Porzelius et al. 2010; Ambler et al. 2012; Göbl et al. 2015).

Non-selective variable and engineering techniques, namely artificial neural networks (ANN), proposed by McCulloch and Pitts (1943), are processing systems inspired by human brain neural networks (Van Gerven and Bohte 2017). Relatively untreated data is managed directly using neural networks and the network is trained with potentially acceptable results (Mignan and Broccardo 2020). So far, several ANN forms have been released, but they all have an analogous framework. (Devillers 1996). Thus, a basic ANN is created by connecting multiple layers such as input, output, and hidden layer (s) to transmit information from one artificial neuron to another (Sözen et al. 2005). A three-layer ANN is illustrated in Figure 1 , and each layer contains its weight matrix, bias input, and output vectors. A set of input predictions is x_i received by the network, their average weight is calculated by w using the summation function, then output y is generated using some activation functions.

Figure 1: An ANN with an input layer (consisting of five predictors) connected to the hidden layer (with three neurons)

In other words, the inputs are received by a node, combined and amplified by weight and bias coefficients, and then adjusted through a non-linear activation function '$\mathcal{K}$' (Khan and Roy 2018).

$$f_j(x) = \mathcal{K} \sum (x_i \, w_{ij})$$

Here, f_j is the input to the jth neuron of the next layer. The ultimate output is passed through a transfer function that shows the summed input as the output value. The logistic (sigmoid) function is the most widely used transfer function, which uses the feature of non-linearity in the mapping method (Dunn et al. 2004):

$$g(x) = \frac{1}{1 + e^{-x}}$$

In fact, statistical optimization approaches in this method are performed as an iterative method of forward and backward paths (epochs) to minimize some loss functions (error) and learn weights (regression coefficients) and biases (intercepts) of inputs.

The activation functions are used to link weights to find the output at each layer in the forward pass. The result of a forward pass is considered as the new output of the prediction. After evaluating the derivatives of the error function between the predicted outputs and the actual outputs, the backward pass starts. Derivative findings are propagated backward, weights are updated, and new error terms are processed for that layer. For each layer, this method is repeated until the input layer is obtained again. The amount of training is indicated by the epochs number and the learning rate and should be considered in front of the validation or test set to avoid over-fitting (Waldmann 2018). Variable selection using ANN is made so that the neural network is trained with all the predictor variables in the input layer. To remove the predictors from the input layer, a backward propagation algorithm is applied. Relevant input nodes and their connection weights are deleted. Subsequently, the input weight for model training is corrected by reducing the error in subsequent iterations. Other suitable settings may be required to remove unusual input nodes (Tetko et al. 1996). ANN learns from experience and recognizes relationships between variables even when there is no tangible relationship (Bandyopadhyay and Chattopadhyay 2007). In modeling with ANN, the number of hidden layers and hidden nodes is important. Improper selection of these parameters affects the model's ability to generalize. Two issues can occur in these cases: 1) underfitting: When the network has poor performance in the training set and cannot match performance with the data and cannot perform well in the test set. 2) overfitting: When the network has remarkable performance in the training set and adapts the performance very close to the data, but is unable to generalize unseen data (Janković et al. 2020).

2. Literature review

The inhumane effects of global warming and environmental degradation threaten the global economy and a challenging issue against humanity in the 21st century. The world is trying to prevent the spread

of ecological degradation by enacting laws and regulations at the national and international agreements. Improper consumption patterns and lifestyles have destroyed everything that has formed in nature over millions of years. There are conflicting arguments about the destruction of the environment and natural resources. Human needs have changed ecosystems through ecological pressures such as land use change, resource extraction and deforestation, and overfishing and pollution (Rudolph and Figge 2017). Economic and industrial growth leads to increased extraction and consumption of national resources, which causes environmental unsustainability. On the other hand, the abundance of natural resources can discourage fossil fuel consumption by reducing imports (Ahmed et al. 2019). Ecological degradation through increasing desertification and its devastating effects on agriculture, especially in developing countries, is a serious concern(Nathaniel et al. 2019). The effect of human activities on the earth's available resources is called ecological footprint (EF) (Alola et al. 2019), which is used to survey the quality of the environment. The EF indicator measured by global hectares of land (gha) determines how much biologically productive land space is needed to provide the resources needed and absorb the population's waste(Ahmed et al. 2020). .

From 1961 to 2010, demand for renewable resources increased by about 140% (from 7.6 billion to 18.1 billion gha), so that the planet's bio-production will not be adequate to meet human needs. (Galli et al. 2015). We have been living in ecological conditions since the 1970s, which shows that the use of the planet's resources is more than its ability to recreate resources, according to the EF Atlas (Network 2010).It takes one and a half times as much time to reproduce the resources we consume in a year (Ahmed et al. 2019). (Rees 1992; Wackernagel et al. 1999). This global issue developed in the 1990s (Rees 1992; Wackernagel et al. 1999) and has attracted the attention of many researchers in various fields, policymakers and so far (Including studies by Ogata 2002; Curutchet et al. 2012; Hynes and Wang 2012; Ciarreta et al. 2014; Enfedaque and Martínez 2014; Robert and Castañeda 2014; Kattumuri 2018; Sarkodie 2018; Daigle and Vasseur 2019; Kotcher et al. 2019; Wang et al. 2019). Over the past ten years, the EF in the European Union has increased further, reaching 4.6 gha/person in 2016. The EU ranks second after North America in regions with the highest EF (6.6 gha/capita) (Gogonea et al. 2020). The growing gap between EF and biological capacity reduces land production capacity, leading

to climate change, food shortages, and biodiversity loss (Rashid et al. 2018). In recent literature, the EF has been studied as a compressive and comprehensive indicator for understanding the effects of human activities on the natural environment. (Ahmed et al. 2019). Ahmed et al. (2020) defined EF as a means of tracking the impact of human activities on ecosystems in terms of six land categories: arable land, fishery ground, grazing land, forests (wood and fuelwood), forest required to absorb CO2 emissions (carbon footprint), and built-up land (infrastructure footprint). Janković et al. (2020) stated that the EF could be considered a sustainability metric as every human activity leaves a footprint on the planet. In addition, EF can be considered as an element of demand and use of natural resources. The consequences are famine, water shortages, and competition over resources, which mainly affect the vulnerable populations of developing countries. (Rudolph and Figge 2017). Figge et al. (2017) stated that the current form of globalization increases the pressures on the environment and its degradation, and leads to the incompatibility of human demand with the land carrying capacity. Ahmed et al. (2020) also cited human activities, including mining and deforestation and chain saw operations, as the most important causes of natural habitat destruction and water, soil, and air pollution. According to these arguments, the empirical study of the role of natural resources in environmental sustainability cannot be agreed. Although natural resources have increased EF, some studies have reported the opposite(Hassan et al. 2019; Zafar et al. 2019). In this regard, the relationship among gross domestic product (GDP), primary and renewable energy utilization (REC), trade openness and environment-related indicators such as CO2 and EF has been investigated. (Usman et al. 2020). For example, a negative relationship between GDP (mainly from tourism), total energy consumption (TEC), trade openness and urbanization population (UP) and EF was reported for 144 countries (Ozturk et al. 2016). More attention has been paid to environmental responses to population dynamics, energy consumption, economic growth, carbon emissions (mainly CO_2) and several other significant factors (Alola et al. 2019). Energy demand is expected to increase by 80% between 2013 and 2035 (Nathaniel and Khan 2020). In short, global warming has been attributed to many factors, such as CO_2 emissions (Nathaniel et al. 2019). The European Union (EU) policies are to guide member countries in achieving the sustainable development goals (SDGs) towards sustainable development and environmental quality. For example, climate

change in 2030 contains a proposal to minimize greenhouse gas emissions by at least 40% compared to 1990 emissions. In general, with the development of large economies such as the United States, China, and some European countries, there are still environmental concerns that have attracted the attention of governments, environmentalists, and policymakers. (Alola et al. 2019). The potential effects of environmental degradation are so severe that a solution must be found to reduce it and requires large investments (Chai et al. 2016; Ozturk et al. 2016; Gergel et al. 2017).

To date, most studies of ecological footprint have examined the relationship between indicators. Undoubtedly, studies on the relationship between natural resources and comprehensive environmental indicators such as EF are not enough, and more research is needed to move towards a sustainable environment. Therefore, a statistical look at the ecological footprint and its factors and its prediction using different models seems necessary. Compared to classical prediction models, the efficiency and performance of ANNs have been the subject of many studies (Crossa et al. 2017; Van Gerven and Bohte 2017; Cherlin et al. 2018; Cabaneros et al. 2019; Cai et al. 2019; Biesbroek et al. 2020; He and Chalise 2020; Mignan and Broccardo 2020). A group of 20 emerging and industrialized countries with strong economies called the G-20 account for about 80% of the world's total primary energy consumption and 82% of CO2 emissions in international energy management (Röhrkasten et al. 2016). The decisions and actions of the G-20 countries have had a significant impact on global energy systems due to their innovative activities and the use of renewable energy. More details are available at https://www.g20.org. To reduce the deterioration of the environment, studying EF using machine learning techniques is attractive and not without merit. To our knowledge, there is no such study. Therefore, this paper uses machine learning (ML) in EF research and evaluates their potential benefits in the G-20 countries over the last two decades. In the current study, these models were trained, validated, and then tested to evaluate their comprehensive statistical analysis performance.

3.Materials and methods

3.1. Descriptive statistics and data preparation

The data needed to examine ecological footprint indicators over the last two decades (from 1999 to 2018) were extracted from https://databank.worldbank.org, http://www.fao.org/,

https://freedomhouse.org/, and https://data.footprintnetwork.org. To help make decisions more effective in the G20, we considered three dependent variables (outputs), i.e., Ecological Footprint (EF), Ecological Footprint vs. Biocapacity (EFB), Ecological Footprint vs. Biocapacity (TEFB), and fifteen global macro indicators (predictors): Total Population (TPOP), Total Fertility Rate (TFR), Agriculture Products (AP), Gross Domestic Product (GDP), Gross Fixed Capital Formation (GFC), Renewable Energy Consumption (REC), Carbon Dioxide Emissions (CO2E), Renewable Consumption (RE), Total Energy Consumption (TEC), Urbanization Population (UP), PMP (Particulate Matter Pollution), International Covenant on Civil and Political Rights (ICRL) and Political Rights of Public Relations, Civil Liberties (PRPRCL). The studied predictor variables and their sources are introduced in Table 1, and some descriptive information of the data set during the last two decades is presented in Table 2.

Table 1: Introduction and definition of studied factors

Table 2: Overall descriptive statistics of variables

Table 3 shows the descriptive statistics of the predictor variables for each country. The normality of the residuals (errors) of the linear regression model and their autocorrelation were investigated using Shapiro-Wilk and Durbin-Watson tests. This statistic test was defined to consider the correlation between the residues as follows:

$$\mathrm{D} = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

Where D in the range 1.5-2.5 indicates no correlation between the residues. After fitting in the initial model, multicollinearity was examined using VIF (Variance Inflation Factor (Myers and Myers 1990)) as follows:

$$\mathrm{VIF} = \frac{1}{1 - \mathcal{R}_i^2}$$

Finally, the initial fitted model's validation was depicted (for instance, Figure 2).

Figure 2: Normally distributed residuals for EF

Statistical procedures and model configuration

If we consider a standard multiple linear regression model as follows: $\mathcal{Y} = 1\beta_0 + \mathcal{X}\beta + e$ that $\mathcal{Y}$ is a response variable vector, $\mathcal{X} = x_{i1},\ldots, x_{ip}$ is a predictor variables matrix, β_0 is the intercept, $\beta = \beta_1, \ldots, \beta_p$ is a regression coefficients vector, and e is an error terms vector, assuming normal distribution $e \sim \mathcal{N}(0, \sigma_e^2)$. In this case, β_0 and βs coefficient values be estimated by minimizing the residual sum of squares (RSS) (Waldmann et al. 2013)as:

$$\hat{\beta}_0, \hat{\beta}_{(\mathrm{OLS})} \stackrel{\text{def}}{=} \arg \min_{\beta_0, \beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$$

Anda PR coefficients is expressed as:

$$\hat{\beta}_0, \hat{\beta}_{(\mathrm{PR})} \stackrel{\text{def}}{=} \arg \min_{\beta_0, \beta \in \mathbb{R}^p} \underbrace{\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2}_{\text{Loss function}} + \underbrace{\lambda P(\lambda, \beta)}_{\text{Penalty function}}$$

Here, to prevent overfitting and control the penalty function's shrinkage amount of contraction, the hyper-parameter λ tunes the equation. In fact, bias-variance trade-off is set by this hyperparameter. Its amount is directly related to bias and inversely associated with variance, i.e., with increasing lambda, bias increases and variance decreases.

By applying an ℓ_2-norm penalized least squares criterion [i.e., $P(\lambda, \beta) = \lambda \|\beta\|\ell_2$] on the linear regression coefficients(Hoerl and Kennard 1970) , the RR estimates are obtained as follows:

$$\hat{\beta}_0, \hat{\beta}_{(\mathrm{RR})} \stackrel{\text{def}}{=} \arg \min_{\beta_0, \beta \in \mathbb{R}^p} \left[\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \|\beta_j\|_2^2 \right]$$

In RR, the shrinkage value is tuned so that no variable is exactly zero and only reduces their variance, so the estimates are biased.

In another case of PR, the values of the coefficients are obtained by applying the Lasso constraint (i.e., an ℓ_1-norm penalized least-squares as $P(\lambda, \beta) = \lambda \|\beta\|\ell_1$)(Tibshirani 1996). An important feature of Lasso is that it allows the coefficients to be exactly zero, thus selecting the variable. If we

consider x_{ij} to be standardized so that $\frac{\sum_i x_{ij}}{\mathcal{N}} = 0$ and $\frac{\sum_i {x_{ij}}^2}{\mathcal{N}} = 1$,then the lasso coefficients are estimated as follows:

$$\hat{\beta}_0, \hat{\beta}_{(\mathrm{LASSO})} \stackrel{\mathrm{def}}{=} \arg\min_{\beta_0, \beta \in \mathbb{R}^p} \left[\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \|\beta_j\|_1 \right]$$

The EN method is another mode of PR that uses a combination of two penalties applied in RR and Lasso on the coefficients (Zou and Hastie 2005):

$$\hat{\beta}_0, \hat{\beta}_{(\mathrm{EN})} \stackrel{\mathrm{def}}{=} \arg\min_{\beta_0, \beta \in \mathbb{R}^p} \left[\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \left(\frac{1}{2}(1-\alpha)\|\beta_j\|_2^2 + \alpha\|\beta_j\|_1 \right) \right]$$

where $0 \leq \alpha \leq 1$ is a penalty weight. If α is equal to 1, EN functions like Lasso but modifies how it deals with high correlated variables (Waldmann et al. 2013).

Cross-validation and parameter optimization

To maintain the original distribution of variables, the data in both train and test sets were scaled through a min-max normalization method. 70% of the data was used to training the models, and the rest was used for testing. In this study, the performance of the models was evaluated using the 10-fold cross-validation method. The whole data was randomly divided into ten equal subsets. A subset of the validation set was considered for testing the model, and the remaining K-1 subset was used to train the model. This method reduces the dependence of performance on the test-training set and reduces the variance of the performance criteria and confirms that the results are free from any sampling bias (James et al. 2013). The lambda value, which minimizes the cross-validation prediction error rate in the training set, is considered the optimal value, automatically determined using the *cv.glmnet* function. By default 10-fold cross-validation, the *cv.glmnet* function sets the optimal lambda value to provide the simplest model. In fact, the proper model lies within one standard error of optimal lambda, i.e., *lambda.1se*. If the lambda is equal to *lambda.1se* we will have a simpler model than the lambda equal to *lambda.min*, but it may be slightly less accurate. Choosing lambda values for each fold Fi was performed by the following cross-validation technique:

$$\epsilon_{n,\lambda} \stackrel{\text{def}}{=} (y_n - \hat{\beta}_{\mathcal{PR}}^{\mathcal{F}_i,\lambda} x_n)^2 \quad \forall_n \epsilon \mathcal{F}_i$$

that $\hat{\beta}_{\mathcal{PR}}^{\mathcal{F}_i,\lambda}$ was estimated on $\mathcal{D} - \mathcal{F}_i$ (here, $\mathcal{D} = \{x_n, y_n\}$ and $\mathcal{F}_i$ is the data not including the ith fold). i^{th} fold was used as a test set and the rest of the data as the training set. So, λ was chosen as:

$$\lambda^{x-val} = \arg\min_{\lambda} \frac{1}{\mathcal{N}} \sum\nolimits_{n=1} \mathcal{N}_{\epsilon \mathcal{F}_i,\lambda}$$

3.2. Performance evaluation

The behavior of the models was evaluated when introducing new data using the following criteria: MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAD (mean absolute deviation), MAE (Mean absolute error), (lower values indicate suitable fit) and R-squared values (Determination coefficient). By considering n as the total number of observations:

$$MAE = \frac{\sum_{i=1}^{n} |e_i|}{n}$$

$$MAD = \frac{\sum_{i=1}^{n} |y_i - \bar{y}|}{n}$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$$

$$RMSE = \sqrt{MSE}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

It should be noted that the mentioned analyses were performed in Linux Ubuntu by R 4.0. (R Core Team, 2020) program using different packages.

4. Results and discussion

Ecological footprint as a measure of human demand for natural capital can help people understand consumption and its impact on the planet and convince local leaders to improve people's well-being by investing in it. According to Table 3, the United States, Canada, and Australia's EF indicator is higher than the other G-20 countries.

Table 3: Descriptive statistics of the predictor variables in G-20 countries

According to Table 3, in Canada, the average EF was 4.89, which means that the country's people need 4.89 gha per person to meet their needs for resources and ecological waste. The value of this index is lower for Australia and is reported at about 4.6 hectares. This result highlights the unsustainable global nature of the Australian lifestyle, particularly energy and animal products, with significant inequities between western countries such as Australia and developing countries (Verones et al. 2017). India, Indonesia, and Korea. India, Indonesia can be mentioned as countries with the lowest EF values. The average EF in the United States is about 50% higher than in most European countries, accounting for more than 20 percent of the world's environmental degradation and 13 percent of the world's greenhouse gas emissions (Dietz et al. 2007).

Obviously, correlation does not indicate causality; Although finding causation is not the goal of the present study, it can be helpful to to evaluate the potential relationships between parameters, especially between independent and dependent variables. The corresponding P-value for variables whose correlation was not significant (P>0.05), has been specified in Figure 3.

Figure 3: Correlation plot of the variables

According to Figure 3, there was a significant correlation between some predictors and the response variable. The existence of multi-collinearity could cause high variability in the regression model, which confuses. For example, there is a significant correlation of 0.70 between "ICRL" and "GDP" and they are also correlated with the response variable (EF). We fitted the regression model based on these two variables, the "GDP" coefficient was positive and the "ICRL" coefficient was negative, meaning that "GDP" has a positive effect on "EF" and the other has a negative effect. Once again, we fit the model based on each of these variables separately, the coefficient of each variable was unexpectedly positive. This means that the regression coefficient for each variable depends on what other variables are included in the model, in addition, the coefficients of these correlated variables become over-inflated. These coefficients fluctuated sharply, indicating that they are more prone to over-training sets (i.e., high

variance in the bias-variance trade-off space). VIF can be used to find correlated variables, but this index does not always specify which variables should be removed (Myers and Myers 1990). A better option to solve the problem is to use PR to monitor the estimation of coefficients. PR models impose constraints on the value of the coefficient and shrink them towards zero. These constraints reduce the amount and fluctuations of the coefficient and variance of the model. Feature importance is shown in Figure 4.

Figure 4: Shrinkage and variable selection using PR

According to Figure 4-A, some coefficients obtained from the linear regression model (OLS) have a significant variance. Therefore, to deal with this problem and multi-collinearity, using PR models seems reasonable (Kuhn and Johnson 2013). All features (predictors) have been retained in the model using RR but have been severely shrunk in weight compared to OLS (Figure 4- B vs. 4- A). PR models with less fixed parameters than OLS are less exposed to the over-fitting training set. Although OLS may perform better than PR on training data, the PR model is better generalized to new data in the presence of extreme variance. In general, a predictive model aims to increase outcome prediction rather than considering underlying cause structures (Yarkoni and Westfall 2017). However, both viewpoints of predictive and causal share some resemblances. Statistically, the model that has the best estimate is not necessarily the most effective in predicting real-world results (Shmueli 2010). Lasso allows some coefficients to be exactly zero and, in the case of correlated variables, holds only one variable and sets the rest to zero (Wei et al. 2015). Therefore, this approach leads to variable selection and presents the final model with fewer parameters, as seen in Figure 4-C. Based on this, the coefficients of "Year", "TOP", "UP" and "PROTOCOL" have been shrunk to zero, creating a parsimonious model. In this way, Lasso allows us to focus on the strongest predictions to understand how the TEFB will change. In other words, in the variable selection process, "CO_2E", "GFC" and "GDP" were identified as the most important variables. Not surprisingly, EF is strongly associated with these macro-global indicators. It is noteworthy that "Year", "TPOP" and "UP" are of little importance, indicating that they are not the

main drivers of EF prediction in the PR model. Figure 4 confirms the results of Table 4, with the most accurate prediction obtained by the Lasso followed by the EN and the RR.

Table 4: Comparison of validation of prediction for ecological footprint indicators in G-20 countries using linear and penalized regression models

In fact, the reality may be much more complex than a proposed model. Hence, there is no assurance that the events that are routinely reviewed are so simple that they can be approximated by models that are understandable to humans (Yarkoni and Westfall 2017). "*Everything should be made simple as possible, but not simpler – ALBERT EINSTEIN*". In Figure 4-D, EN behaves by shrinking and selecting variables as a combination of RR and Lasso. Totally, PR models performed better than OLS, with decent R-squared and stable RMSE values.

Plots of test MSE by λ value and the PR models' practical aspects are demonstrated in Figures 5, 6 and 7 for EF, EFB and TEFB.

Figure 5: 10-fold CV mean squared error (MSE) across the λ values for EF

Figure 6: 10-fold CV mean squared error (MSE) across the λ values for EFB

Figure 7: 10-fold CV mean squared error (MSE) across the λ values for TEFB

The upper part of the plots illustrates the 10-fold CV mean squared error (MSE) across the λ values, and the lower part depicts coefficient values of predictors versus shrinkage parameters. The first and second vertical dashed lines correspond to the λ value with the minimum MSE and the maximum λ value within one standard error of the minimum MSE. T This shows how much the coefficients can be limited while still predicting the utmost accuracy. In the figures shown, the x-axis above shows the number of model variables. RR reduced the coefficient of the variables to almost zero, but did not reduce any to zero, and all variables remained in the model. Indeed, RR has shrunk correlated variables relative to each other rather than allowing one to be positive and the other to be negative. Therefore, we reduced the data noise, which makes the model more accurate in identifying real signals. As shown in Figure 6, the model improvement and the automatic variable selection are represented by increasing the λ values via the Lasso method. Our output shows the decrease in the number of predictor variables in

the Lasso model when log (λ) → -5. These variables are likely to be strongly correlated with other data, causing their coefficients to be inflated. The advantage of EN is that it allows adjustment via RR with the Lasso variable selection feature.

The studied models are assessed based on the highest determination coefficient (R^2) and the least RMSE values. Besides, MSE, MAD, and MAE values are also available in Tables 4 and 5. Table 4 indicates the comparison of validation of EF indicators prediction using OLS and PR. According to Table 4, PR methods have a relative improvement in predictive performance comparable to linear regression. In this study, Lasso selected fewer variables; however, the predictive performance of this method was favorable for all three indicators. The superiority of PR models is not only because of their predictive ability but also because of their ability to select variables and interpretive considerations. For EF, the RMSE and R^2 values of the Lasso model were 0.642 and 0.752, respectively, while these values were 1.910 and 0.663 for EFB. In the end, the best performance for the index TEFB was achieved by the Lasso model with RMSE of 1.467 and R^2 of 0.617.

As shown in Figures 8, 9 and 10, the best neural networks achieve superior performance with the least error obtained for EF indicators (EF, EFB and TEFB) were obtained by a single hidden layer containing six neurons, two hidden layers containing 4 and 2 And 5 and 3, respectively. For each indicator, the most suitable neural networks are presented in two plots. In the upper plots, the weights are calculated using the backpropagation algorithm.

The blue line represents the term bias. In the lower plots, the weight values of the variables are marked using the line thickness as positive coefficients in black and negative coefficients in gray. The degree of the color of the variables in the input layer is also commensurate with their importance.

Figure 8: The best neural networks for EF

Figure 9: The best neural networks for EFB

Figure 10: The best neural networks for TEFB

To compare neural network performance criteria with the OLS model, we performed rescaling on variables. Compared to the linear model (LM), the proper performance of ANN models on the test set for visual comparison is shown in Figure 11. According to Figure 11, the ANN predictions are centered around the line more than those made by LM.

Figure 11: The appropriate ANN models' performance compared to the linear model (LM) on the test set

Figure 12 shows the change in each predictor while keeping the other predictors at quantiles values for models suitable for EF indicators. The profiling method uses the generated / learned model to predict new values, which sets a prediction through its full range while having all but one predictor level. This was repeated in several static predictors (e.g., Quan standards) and repeated for all predictors.

Figure 12: Changing each predictor while holding other predictors at quantiles for the suitable models

The resulting graphs have an exciting insight into the performance of the model. A significant side effect in estimating all the different predictor effects in this way is that the amplitude of the response values provides a new measure of the importance of the variable. Some of them have not changed with the change in other predictors. Some of them are very depressed or have significantly different curves than the maximum. Predict validation for EF indicators using ANN compared to OLS is presented in Table 5.

Table 5: Comparison of validation of prediction for ecological footprint indicators in G-20 countries using linear regression and artificial neural networks

As shown in Table 5, the MSE, RMSE, MAD, and MAE for ANN models are lower than those of the LM. In addition, the performance of ANN neural network models in predicting new data was better than the OLS model. For EF, the ANN model with six neutrons had better performance. But, for the other two indicators, the two-layer model performed better. This was expected due to its remarkable flexibility / nonlinearity. Obtained RMSE and R^2 for EF in the appropriate ANN model were 0.413 and

0.9082, respectively; corresponding values were 0.350, 0.991, and 0.695 and 0.941 for EFB and TEFB. Overall, the results obtained are consistent with studies conducted and literature in this area. By studying environmental impact prediction using neural network modeling, Spitz and Lek (1999) concluded that ANNs can learn the complex relationships between ecological variables and evaluate the impact and produce operational predictions. They suggested that reasonable predictions could help managers take preventive, protective and compensatory measures.

The use of modified ML models for climate change raises forecasts on a time scale. It delivers outstanding results to support climate adaptation, emphasizing drought warnings and helping to change general climate policy. (Huntingford et al. 2019). Using ML models, in particular, K-nearest neighbor regression, random forest regression and artificial neural networks to predict EF based on energy parameters (natural gas resources, coal resources, oil resources, wind resources, solar photovoltaic sources), hydropower, nuclear and other renewable resources) led to impressive results (Janković et al. 2020). Ma et al. (2012) developed an EF model to calculate the per capita EF of 24 countries. Their results show that the EF model based on support vector machine performs well. According to (Tsanas and Xifara 2012; Robinson et al. 2017), ML models also understand energy system functionality in the context of complex human interactions. Furthermore, using ML models for conventional energy systems and alternative and renewable energy systems was promising and valuable in the study of Voyant et al. (2017). Saleh et al. (2015) reported that their findings could even be practical and helpful in industrial performance, especially for executives to make effective decisions for business performance by considering the costs of CO_2 monitoring.

5. Conclusions

In this study, we present the application of PR and ANN to EF to predict it based on global macro indicators in G-20 countries. The results show that EF values are at the top of the G20 in the United States, Canada, Australia, the European Union and Germany. In this, It was important to use PR approaches to identify important global macro indicators. Among the PR methods, Lasso was parsimonious in selecting variables but performed better. Still, the number of indicators selected by AN was higher, meaning that this model was generous in selecting variables.

But overall, there was no significant difference in variable selection and therefore interpretation between Lasso and EN. Understanding both the characteristics of variable selection and predicting each model is essential for informed decision-making. For example, translating a well-predicting forecasting model with many indicators may be difficult in a brief tool. In contrast, an interpretable model with poor predictive performance may call into question the usefulness of such indicators (Greenwood et al. 2020). The results showed better predictive performance of the PR compared to the OLS, albeit slightly. The ANNs have shown the highest prediction accuracy compared to the PR models evaluated. In general, the results showed that both types of models could predict EF indicators. In conclusion, the relatively simple ANN architecture performed better than the PR models in predicting EF using global macro indicators in the G-20.

Ethical Approval: Not applicable

Consent to Participate: Not applicable

Consent to Publish: Not applicable

Authors Contributions: AR Investigation, Methodology, Writing - original draft; AM, Supervision; Writing - review & editing and HM, Data curation; Visualization. All authors have read and approved the final manuscript

Funding: The authors received no financial support for the research

Competing Interests: The authors declare that they have no conflict of interest

Availability of data and materials: The datasets used and/or analyses during the current study are available from the corresponding author on reasonable request.

References:

Ahmed Z, Wang Z, Mahmood F, Hafeez M, Ali N. 2019. Does globalization increase the ecological footprint? Empirical evidence from Malaysia. Environmental Science and Pollution Research. 26(18):18565-18582.

Ahmed Z, Zafar MW, Ali S. 2020. Linking urbanization, human capital, and the ecological footprint in G7 countries: an empirical analysis. Sustainable Cities and Society. 55:102064.

Alola AA, Bekun FV, Sarkodie SA. 2019. Dynamic impact of trade policy, economic growth, fertility rate, renewable and non-renewable energy consumption on ecological footprint in Europe. Science of the Total Environment. 685:702-709.

Ambler G, Seaman S, Omar R. 2012. An evaluation of penalised survival methods for developing prognostic models with rare events. Statistics in medicine. 31(11-12):1150-1161.

Bandyopadhyay G, Chattopadhyay S. 2007. Single hidden layer artificial neural network models versus multiple linear regression model in forecasting the time series of total ozone. International Journal of Environmental Science & Technology. 4(1):141-149.

Biesbroek R, Badloe S, Athanasiadis IN. 2020. Machine learning for research on climate change adaptation policy integration: an exploratory UK case study. Regional Environmental Change. 20(3):1-13.

Cabaneros SM, Calautit JK, Hughes BR. 2019. A review of artificial neural network models for ambient air pollution prediction. Environmental Modelling & Software. 119:285-304.

Cai M, Pipattanasomporn M, Rahman S. 2019. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Applied Energy. 236:1078-1088.

Chai X, Tonjes DJ, Mahajan D. 2016. Methane emissions as energy reservoir: context, scope, causes and mitigation strategies. Progress in Energy and Combustion Science. 56:33-70.

Using penalized regression to predict phenotype from SNP data. BMC proceedings; 2018: BioMed Central.

Ciarreta A, Espinosa MP, Pizarro-Irizar C. 2014. Is green energy expensive? Empirical evidence from the Spanish electricity market. Energy Policy. 69:205-215.

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y. 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends in plant science. 22(11):961-975.

Curutchet G, Grinberg S, Gutiérrez RA. 2012. Degradación ambiental y periferia urbana: un estudio transdiciplinario sobre la contaminación en la región metropolitana de Buenos Aires. Ambiente & sociedade. 15(2):173-194.

Daigle C, Vasseur L. 2019. Is it time to shift our environmental thinking? A perspective on barriers and opportunities to change. Sustainability. 11(18):5010.

Devillers J. 1996. Neural networks in QSAR and drug design. Academic Press.

Dietz T, Rosa EA, York R. 2007. Driving the human ecological footprint. Frontiers in Ecology and the Environment. 5(1):13-18.

Dunn NA, Lockery SR, Pierce-Shimomura JT, Conery JS. 2004. A neural network model of chemotaxis predicts functions of synaptic connections in the nematode Caenorhabditis elegans. Journal of computational neuroscience. 17(2):137-147.

Efron B, Hastie T, Johnstone I, Tibshirani R. 2004. Least angle regression. Annals of statistics. 32(2):407-499.

Enfedaque ND, Martínez JMD. 2014. El impacto económico de los desastres naturales. eXtoikos.(15):99-101.

Figge L, Oebels K, Offermans A. 2017. The effects of globalization on Ecological Footprints: an empirical analysis. Environment, Development and Sustainability. 19(3):863-876.

Galli A, Lin D, Wackernagel M, Gressot M, Winkler S. 2015. Humanity's growing Ecological Footprint: sustainable development implications. Glob Footpr Netw.

Gergel DR, Nijssen B, Abatzoglou JT, Lettenmaier DP, Stumbaugh MR. 2017. Effects of climate change on snowpack and fire potential in the western USA. Climatic Change. 141(2):287-299.

Göbl CS, Bozkurt L, Tura A, Pacini G, Kautzky-Willer A, Mittlböck M. 2015. Application of penalized regression techniques in modelling insulin sensitivity by correlated metabolic parameters. PloS one. 10(11):e0141524.

Gogonea R-M, Ghita SI, Saseanu AS. 2020. Biocapacity—Premise of Sustainable Development in the European Space. Sustainability. 12(3):1037.

Greenwood CJ, Youssef GJ, Letcher P, Macdonald JA, Hagg LJ, Sanson A, Mcintosh J, Hutchinson DM, Toumbourou JW, Fuller-Tyszkiewicz M. 2020. A comparison of penalised regression methods for informing the selection of predictive markers. PloS one. 15(11):e0242730.

Hassan ST, Xia E, Khan NH, Shah SMA. 2019. Economic growth, natural resources, and ecological footprints: evidence from Pakistan. Environmental Science and Pollution Research. 26(3):2929-2938.

Hastie T, Tibshirani R, Friedman J. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

He J, Chalise P. 2020. Nested and Repeated Cross Validation for Classification Model With High-dimensional Data. Revista Colombiana de Estadística. 43(1):103-125.

Hoerl AE, Kennard RW. 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 12(1):55-67.

Huntingford C, Jeffers ES, Bonsall MB, Christensen HM, Lees T, Yang H. 2019. Machine learning and artificial intelligence to aid climate change research and preparedness. Environmental Research Letters. 14(12):124007.

Hynes W, Wang S. 2012. Green growth and developing countries: A summary for policymakers. Document DCD/DAC (2012). 27.

James G, Witten D, Hastie T, Tibshirani R. 2013. An introduction to statistical learning. Vol. 112. Springer.

Janković R, Mihajlović I, Štrbac N, Amelio A. 2020. Machine learning models for ecological footprint prediction based on energy parameters. Neural Computing and Applications.1-15.

mKattumuri R. 2018. Sustaining natural resources in a changing environment: evidence, policy and impact. Taylor & Francis.

Khan PM, Roy K. 2018. Current approaches for choosing feature selection and learning algorithms in quantitative structure–activity relationships (QSAR). Expert opinion on drug discovery. 13(12):1075-1089.

Kotcher J, Maibach E, Choi W-T. 2019. Fossil fuels are harming our brains: identifying key messages about the health effects of air pollution from fossil fuels. BMC public health. 19(1):1-12.

Kuhn M, Johnson K. 2013. Applied predictive modeling. Vol. 26. Springer.

Le Cessie S, Van Houwelingen JC. 1992. Ridge estimators in logistic regression. Journal of the Royal Statistical Society: Series C (Applied Statistics). 41(1):191-201.

Ma H, Chang W, Cui G. 2012. Ecological footprint model using the support vector machine technique. PloS one. 7(1):e30396.

McCulloch WS, Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics. 5(4):115-133.

Mignan A, Broccardo M. 2020. Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations. Seismological Research Letters. 91(4):2330-2342.

Myers RH, Myers RH. 1990. Classical and modern regression with applications. Vol. 2. Duxbury press Belmont, CA.

Nathaniel S, Khan SAR. 2020. The nexus between urbanization, renewable energy, trade, and ecological footprint in ASEAN countries. Journal of Cleaner Production. 272:122709.

Nathaniel S, Nwodo O, Adediran A, Sharma G, Shah M, Adeleye N. 2019. Ecological footprint, urbanization, and energy consumption in South Africa: including the excluded. Environmental Science and Pollution Research. 26(26):27168-27179.

Network GF. 2010. Ecological footprint atlas 2010. Retrieved May. 25:2014.

Ogata S. 2002. State security-human security. United Nations University, Public Affairs Section.

Ozturk I, Al-Mulali U, Saboori B. 2016. Investigating the environmental Kuznets curve hypothesis: the role of tourism and ecological footprint. Environmental Science and Pollution Research. 23(2):1916-1928.

Porzelius C, Schumacher M, Binder H. 2010. Sparse regression techniques in low-dimensional survival data settings. Statistics and Computing. 20(2):151-163.

Rashid A, Irum A, Malik IA, Ashraf A, Rongqiong L, Liu G, Ullah H, Ali MU, Yousaf B. 2018. Ecological footprint of Rawalpindi; Pakistan's first footprint analysis from urbanization perspective. Journal of Cleaner Production. 170:362-368.

Rees WE. 1992. Ecological footprints and appropriated carrying capacity: what urban economics leaves out. Environment and urbanization. 4(2):121-130.

Robert J, Castañeda K. 2014. Las ciencias sociales frente a las problemáticas medio-ambientales: apuntes de un primer encuentro. Lima, 12 de noviembre de 2014. Bulletin de l'Institut français d'études andines.(43 (3)):668-674.

Robinson C, Dilkina B, Hubbs J, Zhang W, Guhathakurta S, Brown MA, Pendyala RM. 2017. Machine learning approaches for estimating commercial building energy consumption. Applied energy. 208:889-904.

Röhrkasten S, Thielges S, Quitzow R. 2016. Sustainable energy in the G20: Prospects for a global energy transition. IASS Study, December 2016 https://doiorg/102312/iass2016030.

Rudolph A, Figge L. 2017. Determinants of ecological footprints: what is the role of globalization? Ecological Indicators. 81:348-361.

Saleh C, Chairdino Leuveano RA, Ab Rahman MN, Md Deros B, Dzakiyullah NR. 2015. Prediction of CO2 emissions using an artificial neural network: The case of the sugar industry. Advanced Science Letters. 21(10):3079-3083.

Sarkodie SA. 2018. The invisible hand and EKC hypothesis: what are the drivers of environmental degradation and pollution in Africa? Environmental Science and Pollution Research. 25(22):21993-22022.

Shmueli G. 2010. To explain or to predict? Statistical science. 25(3):289-310.

Sözen A, Arcaklıoğlu E, Özalp M, Çağlar N. 2005. Forecasting based on neural network approach of solar potential in Turkey. Renewable Energy. 30(7):1075-1090.

Spitz F, Lek S. 1999. Environmental impact prediction using neural network modelling. An example in wildlife damage. Journal of Applied Ecology. 36(2):317-326.

Tetko IV, Villa AE, Livingstone DJ. 1996. Neural network studies. 2. Variable selection. Journal of chemical information and computer sciences. 36(4):794-803.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 58(1):267-288.

Tibshirani R. 1997. The lasso method for variable selection in the Cox model. Statistics in medicine. 16(4):385-395.

Tsanas A, Xifara A. 2012. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings. 49:560-567.

Usman M, Kousar R, Yaseen MR, Makhdum MSA. 2020. An empirical nexus between economic growth, energy utilization, trade policy, and ecological footprint: a continent-wise comparison in upper-middle-income countries. Environmental Science and Pollution Research. 27(31):38995-39018.

Van Gerven M, Bohte S. 2017. Artificial neural networks as models of neural information processing. Frontiers in Computational Neuroscience. 11:114.

Verones F, Moran D, Stadler K, Kanemoto K, Wood R. 2017. Resource footprints and their ecosystem consequences. Scientific Reports. 7(1):1-12.

Voyant C, Notton G, Kalogirou S, Nivet M-L, Paoli C, Motte F, Fouilloy A. 2017. Machine learning methods for solar radiation forecasting: A review. Renewable Energy. 105:569-582.

Wackernagel M, Onisto L, Bello P, Linares AC, Falfán ISL, García JM, Guerrero AIS, Guerrero MGS. 1999. National natural capital accounting with the ecological footprint concept. Ecological economics. 29(3):375-390.

Waldmann P. 2018. Approximate Bayesian neural networks in genomic prediction. Genetics Selection Evolution. 50(1):1-9.

Waldmann P, Mészáros G, Gredler B, Fuerst C, Sölkner J. 2013. Evaluation of the lasso and the elastic net in genome-wide association studies. Frontiers in genetics. 4:270.

Waldron L, Pintilie M, Tsao M-S, Shepherd FA, Huttenhower C, Jurisica I. 2011. Optimized application of penalized regression methods to diverse genomic data. Bioinformatics. 27(24):3399-3406.

Wang Z, Rasool Y, Asghar MM, Wang B. 2019. Dynamic linkages among CO 2 emissions, human development, financial development, and globalization: empirical evidence based on PMG long-run panel estimation. Environmental Science and Pollution Research. 26(36):36248-36263.

Wei L, Tian W, Silva EA, Choudhary R, Meng Q, Yang S. 2015. Comparative study on machine learning for urban building energy analysis. Procedia Engineering. 121:285-292.

Yarkoni T, Westfall J. 2017. Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science. 12(6):1100-1122.

Zafar MW, Zaidi SAH, Khan NR, Mirza FM, Hou F, Kirmani SAA. 2019. The impact of natural resources, human capital, and foreign direct investment on the ecological footprint: the case of the United States. Resources Policy. 63:101428.

Zou H. 2006. The adaptive lasso and its oracle properties. Journal of the American statistical association. 101(476):1418-1429.

Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology). 67(2):301-320.

685 **Table 1:** Introduction and definition of studied factors

Variable	Definition	Measurement units	Sources link
TPOP	Total Population	10^4people	WDI
TFR	Total Fertility Rate	Births per woman	WDI
AP	Agriculture Products	Gross Production Value (contestant=2010 US$)	FAO
GDP	Gross Domestic Product	per capita (constant 2010 US$)	WDI
GFC	Gross Fixed Capital Formation	% of GDP	WDI
REC	Renewable Energy Consumption	Million tones oil equivalents	WDI
CO2E	Carbon Dioxide Emissions	Metric tons per capita	WDI
RE	Renewable Consumption	Million tons oil equivalent per capita	WDI
TEC	Total Energy Consumption	Calculated by Divided by GDP to total energy consumption	WDI
UP	Urbanization Population	% of total population	WDI
PMP	Particulate Matter Pollution	(mg/m^3)	WDI

ICRL	International Covenant on Civil and Political Rights	Average for Political Rights and Civil Liberties (1-7)	Freedom House
PRPRCL	Political Rights of Public Relations, Civil Liberties	Average for Political Rights and Civil Liberties (1-7)	Freedom House
EF	Ecological Footprint	Number of Earths	GFN
EFB	Ecological Footprint vs Biocapacity	gha per person	GFN
TEFB	Ecological Footprint vs Biocapacity	gha	GFN

686

687 **Table 2:** Overall descriptive statistics of variables

Variables	Overall mean ± SE	CV (%)	Range
TPOP	232324647.52±4512334.08	155.38	1.37E+09
TFR	1.95±0.01	28.3	3.12
AP	66610.55±1505.43	180.8	651188.8
GDP	24129.54±213.71	70.86	53848.98
GFC	23.3±0.08	26.37	33.73
REC	3.15±0.05	138	20.52
CO2E	1094.67±22.49	164.39	9154.59
RE	14.62±0.17	91.38	51.78
TEC	525.98±8.53	129.74	3107.17
UP	72.77±0.19	20.72	64.42
PMP	26.5±0.29	88.05	96.18
ICRL	4.15±0.02	31.06	4.5
PRPRCL	2.32±0.02	83.8	6
EF	2.51±0.02	55.29	5.39
EFB	3.53±0.06	125.02	19.47
TEFB	3.56±0.04	91.03	24.86

688 The variables were defined in Table 1. CV: Coefficient of variation, known as relative standard deviation (RSD),

689 defines as $\mathrm{CV\%} = \frac{\sigma}{\mu} \times 100$

690

691

692

693 **Table 3:** Descriptive statistics of the predictor variables in G-20 countries

G-20 Countries	EF				EFB				TEFB			
	Mean±SE	CV(%)	Skewness	Kurtosis	Mean±SE	CV(%)	Skewness	Kurtosis	Mean±SE	CV(%)	Skewness	Kurtosis
Argentina	1.9±0.05	10.64	0.25	-1.44	1.72±0.02	4.41	0.31	-1.33	6.62±0.04	2.95	-1.17	2.02
Australia	4.6±0.1	9.63	-0.09	-1.58	14.58±0.41	12.68	0.20	-1.43	14.76±0.7	21.35	1.81	3.46
Brazil	1.65±0.03	7.50	0.07	-1.67	9.43±0.13	5.94	0.28	-1.36	3.44±0.32	42.04	3.12	9.67
Canada	4.89±0.05	4.27	0.45	-0.63	16.37±0.27	7.29	1.25	1.67	5.51±0.06	4.85	0.24	-1.11
China	1.73±0.11	27.42	-0.26	-1.61	1.73±0.02	4.40	0.20	-1.47	2.93±0.16	23.77	-0.40	-1.38
European Union	3.03±0.06	9.51	1.79	3.34	3.01±0.01	1.51	0.13	-0.29	1.93±0.17	39.89	0.00	-2.09
France	2.95±0.04	6.20	-0.93	1.07	2.74±0.04	5.83	-0.81	0.47	2.67±0.03	5.30	-0.41	0.39
Germany	3.62±0.25	31.27	-1.53	0.95	1.65±0.01	3.57	-0.10	-0.91	3.02±0.02	3.16	0.27	-1.29
India	0.59±0.02	17.36	0.00	-1.73	1.01±0.03	13.17	-0.04	-1.75	3.68±0.44	53.14	2.18	3.41
Indonesia	0.87±0.03	13.80	0.28	-1.35	1.69±0.04	9.73	-2.66	7.84	1.49±0.03	9.04	0.14	-1.35
Italy	2.94±0.06	9.13	-0.32	-1.14	0.99±0.03	12.35	1.64	2.08	1.55±0.06	17.74	-0.98	-0.64
Japan	1.08±0.17	69.72	1.15	-0.32	2.8±0.02	3.04	-1.06	0.55	6.22±0.09	6.76	-0.30	-0.62
México	1.62±0.04	10.88	-0.99	1.09	1.33±0.04	12.30	1.01	0.33	2.8±0.06	9.22	0.09	-1.07

Russia	2.98±0.08	12.73	-1.28	1.44	1.8±0.05	13.41	2.17	3.48	5.19±0.08	7.20	-0.41	0.17
Saudi Arabia	2.88±0.14	21.70	-0.13	-1.53	1.7±0.03	8.10	-1.58	2.64	0.57±0.03	21.52	-0.02	-1.85
South Africa	1.95±0.04	8.81	-0.48	-0.93	1.09±0.02	8.05	0.08	-0.89	3.38±0.05	7.23	0.21	-1.26
Korea	0.75±0.04	22.63	-0.16	-1.78	0.63±0.01	8.94	-0.41	-0.67	0.69±0.01	5.14	0.29	-1.20
Turkey	1.79±0.05	13.20	-0.51	-0.98	1.58±0.03	7.30	0.23	-1.50	1.53±0.03	7.95	-0.02	-1.25
United Kingdom	3.1±0.08	10.90	-0.47	-0.91	1.17±0.02	6.38	0.09	-1.20	1.17±0.02	6.37	0.03	-1.20
United States	5.27±0.08	6.60	0.05	-1.32	3.66±0.02	2.72	0.54	-0.43	1.96±0.26	59.70	1.13	0.40

694 Skewness $= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{(n-1)\times\sigma^3}$, Kurtosis $= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{(n-1)\times\sigma^4}$

Table 4: Comparison of validation of prediction for ecological footprint indicators in G-20 countries using linear and penalized regression models

	Method	MSE	RMSE	MAD	MAE	R^2
EF	OLS	0.469	0.685	0.444	0.536	0.722
	RR	0.438	0.662	0.336	0.517	0.737
	Lasso	0.418	0.642	0.318	0.507	0.752
	EN	0.433	0.658	0.329	0.523	0.749
EFB	OLS	5.765	2.401	1.770	1.949	0.604
	RR	4.456	2.111	1.590	1.711	0.643
	Lasso	3.648	1.910	1.493	1.607	0.663
	EN	4.601	2.145	1.704	1.719	0.641
TEFB	OLS	3.658	1.912	1.424	1.530	0.559
	RR	2.570	1.603	1.388	1.432	0.587
	Lasso	2.153	1.467	1.083	1.350	0.617
	EN	2.318	1.522	1.348	1.443	0.591

Table 5: Comparison of validation of prediction for ecological footprint indicators in G-20 countries using linear regression and artificial neural networks

	Method	MSE	RMSE	MAD	MAE	R^2
EF	OLS	0.469	0.685	0.444	0.536	0.722
	NN_2	0.239	0.489	0.214	0.349	0.866
	NN_4	0.195	0.442	0.156	0.274	0.889
	NN_6	0.170	0.413	0.117	0.230	0.908
	NN_8	0.247	0.497	0.151	0.273	0.864
	NN_4.2	0.175	0.418	0.167	0.252	0.902
	NN_5.2	0.190	0.436	0.118	0.243	0.898
	NN_5.3	0.191	0.437	0.106	0.246	0.891
	NN_6.3	0.195	0.442	0.115	0.242	0.893
EFB	OLS	5.765	2.401	1.770	1.949	0.604
	NN_2	0.346	0.589	0.463	0.494	0.976
	NN_4	0.227	0.476	0.206	0.308	0.983
	NN_6	0.285	0.534	0.212	0.326	0.979
	NN_8	0.469	0.685	0.182	0.410	0.965
	NN_4.2	0.123	0.350	0.177	0.241	0.991
	NN_5.2	0.128	0.357	0.172	0.251	0.991
	NN_5.3	0.284	0.533	0.201	0.318	0.980
	NN_6.3	0.134	0.367	0.161	0.228	0.990

TEFB	OLS	3.658	1.912	1.424	1.530	0.559
	NN_2	0.808	0.899	0.605	0.725	0.903
	NN_4	0.985	0.993	0.544	0.691	0.888
	NN_6	1.961	1.400	0.460	0.828	0.783
	NN_8	1.121	1.059	0.356	0.647	0.866
	NN_4.2	1.561	1.249	0.305	0.656	0.866
	NN_5.2	1.246	1.116	0.291	0.623	0.856
	NN_5.3	0.484	0.695	0.401	0.511	0.941
	NN_6.3	0.684	0.827	0.316	0.531	0.918

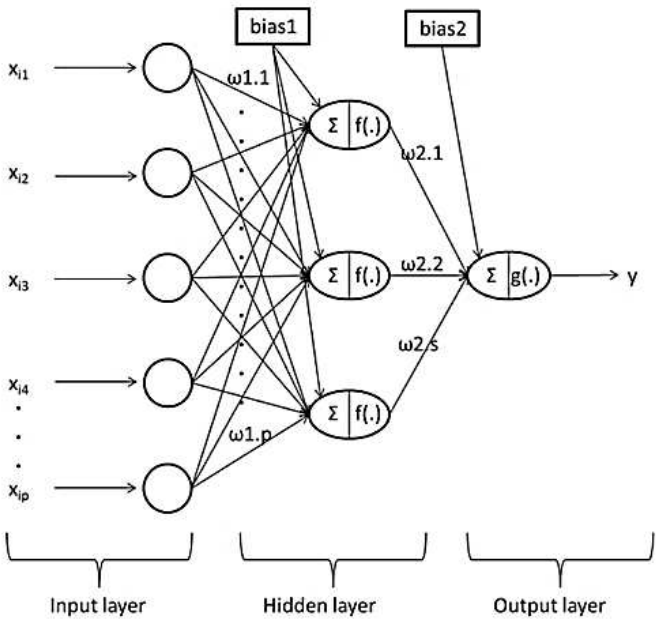


Figure 1: An ANN with one input layer (consisting of five predictors) connected to the hidden layer (with three neurons)

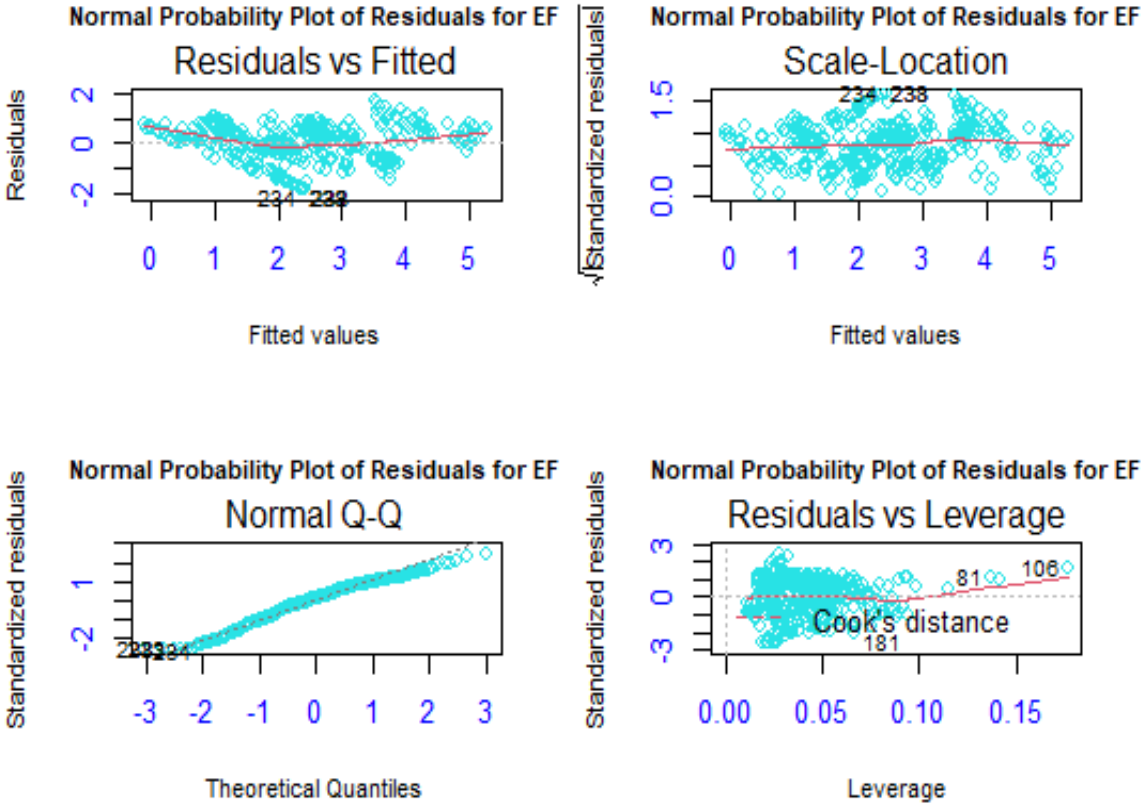


Figure 2: Normally distributed residuals for EF

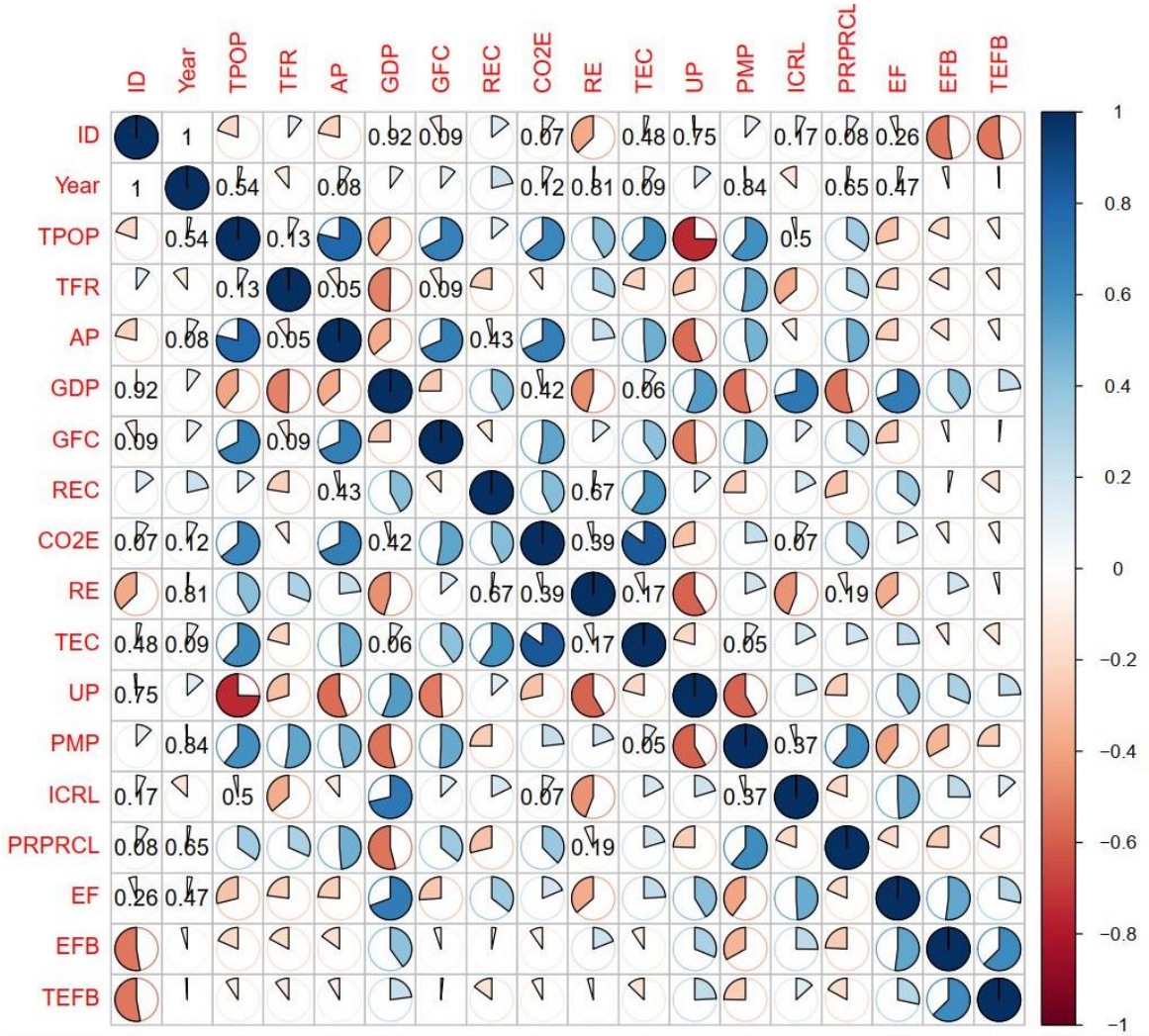


Figure 3: Correlation plot of the variables

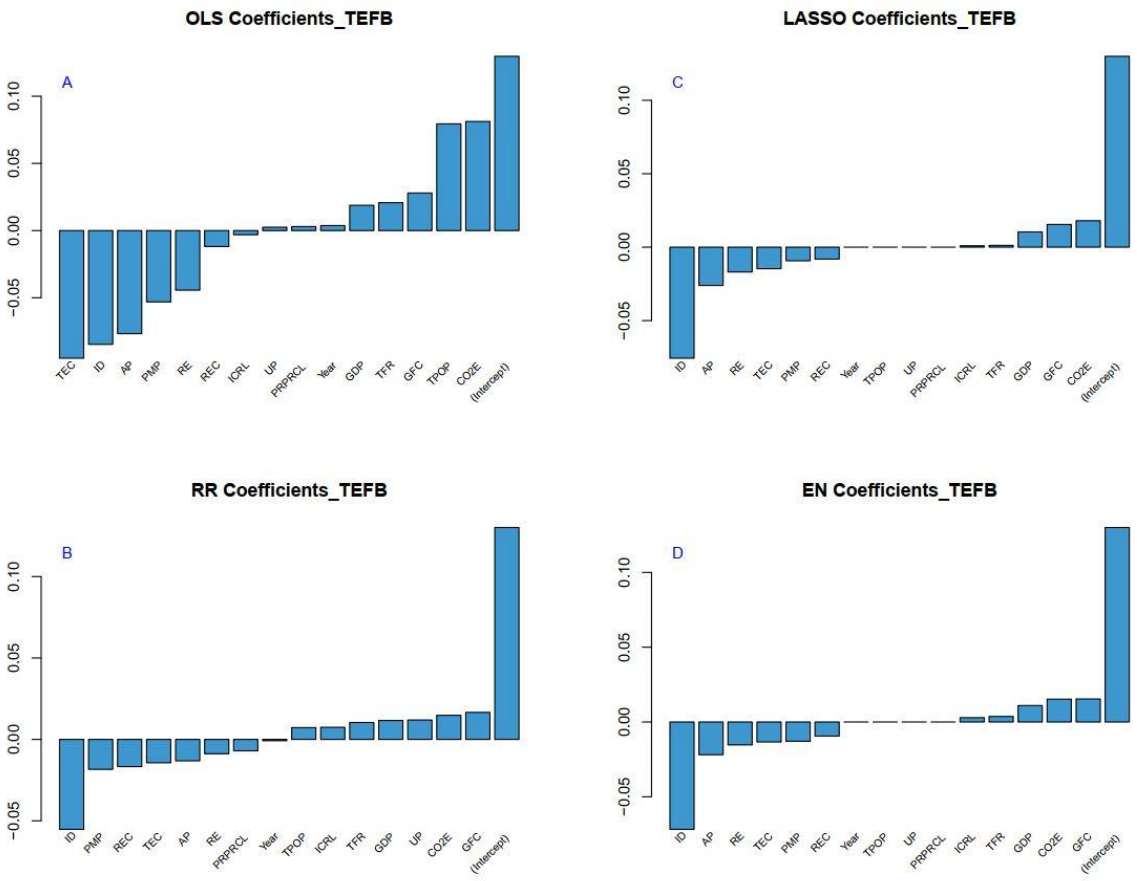


Figure 4: Shrinkage and variable selection using PR

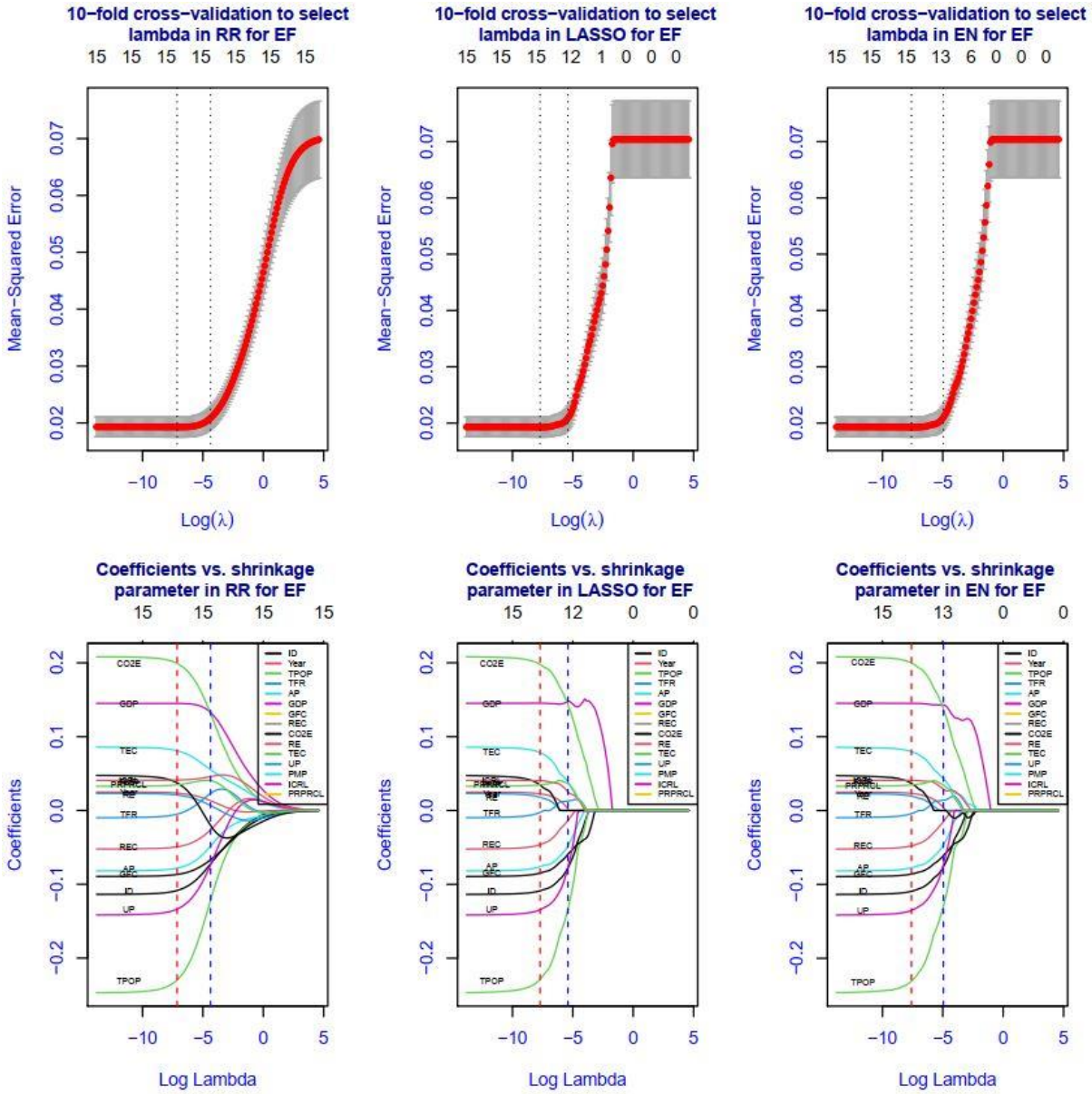


Figure 5: 10-fold CV mean squared error (MSE) across the λ values for EF

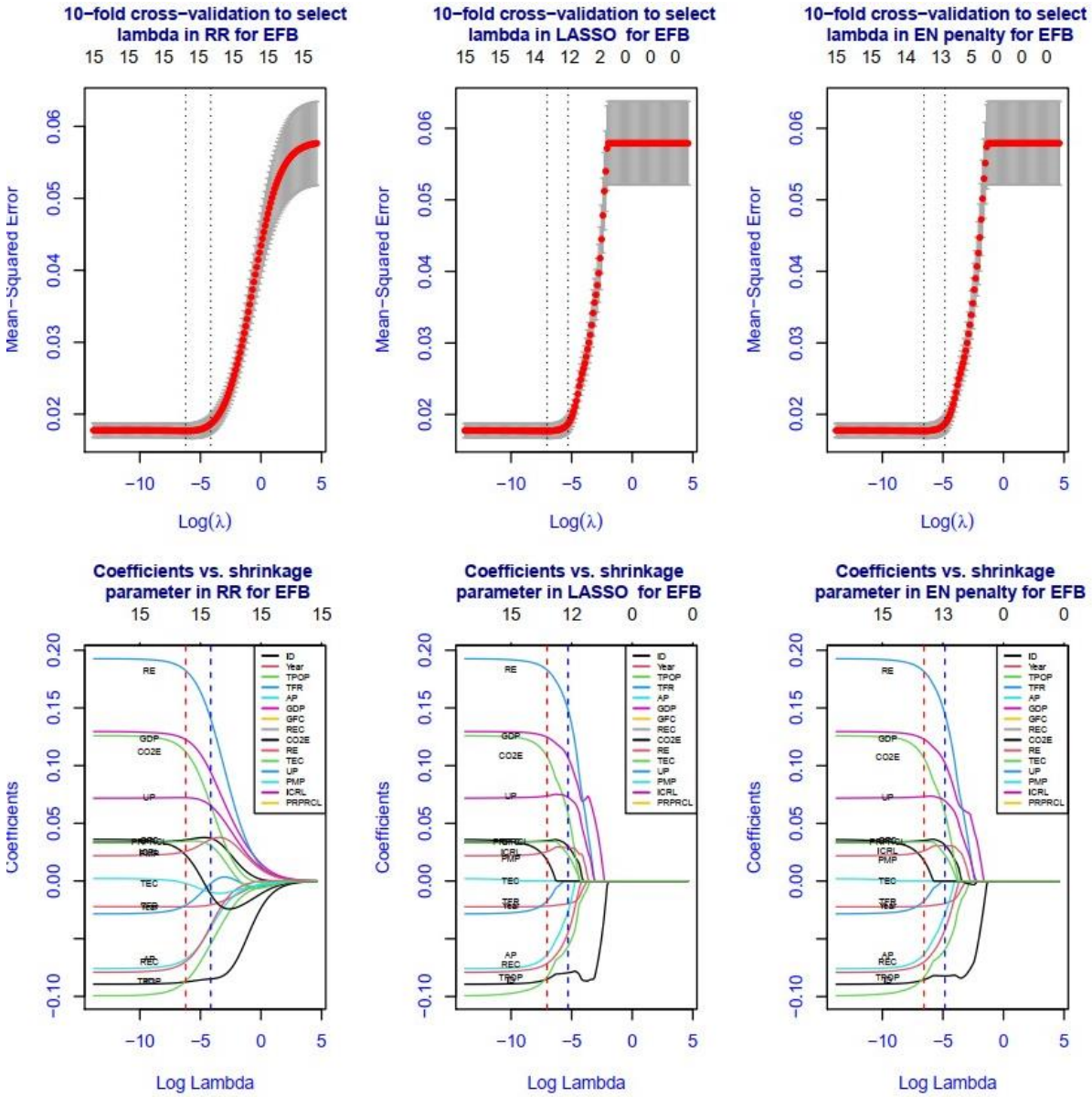


Figure 6: 10-fold CV mean squared error (MSE) across the λ values for EFB

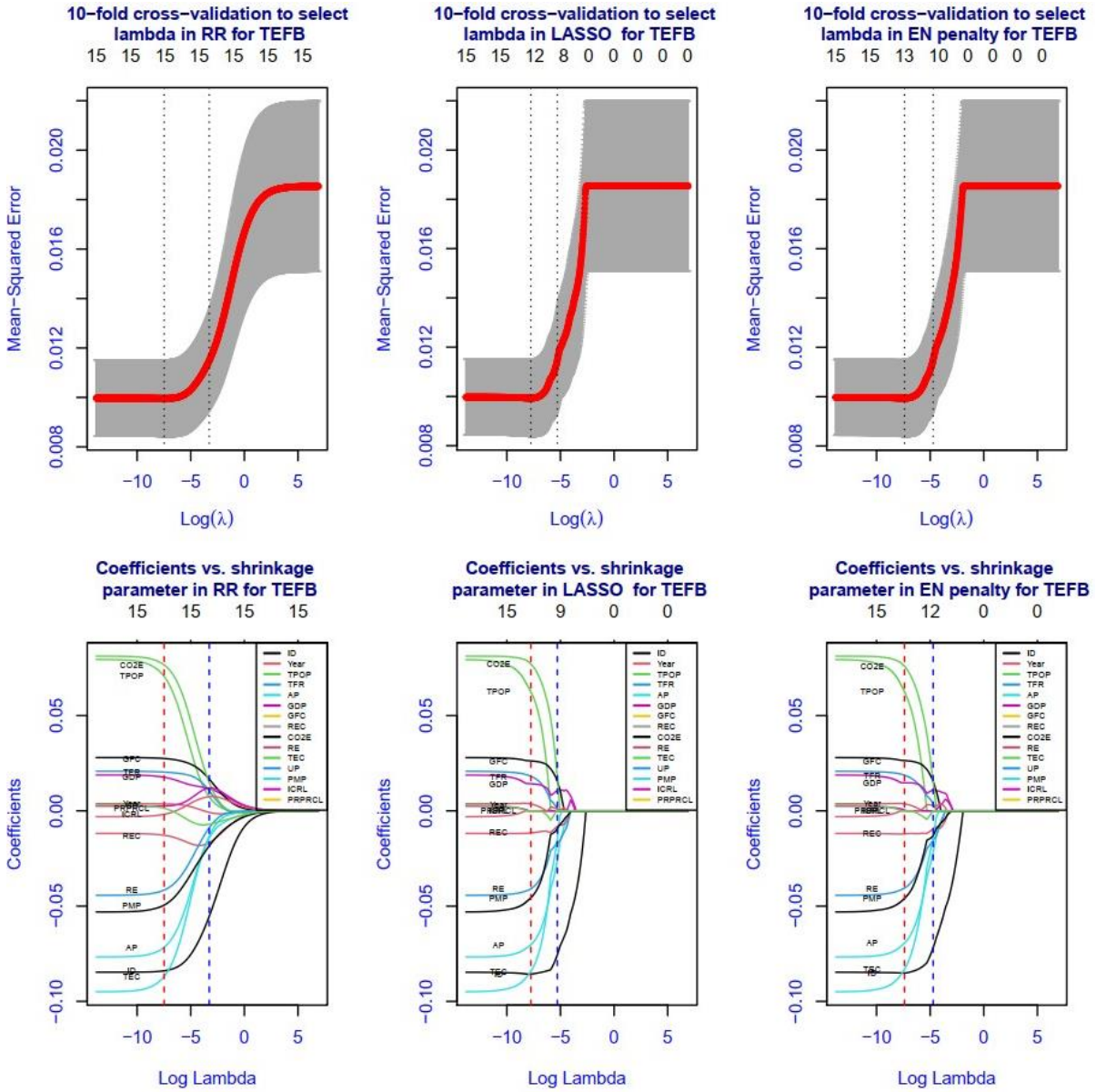


Figure 7: 10-fold CV mean squared error (MSE) across the λ values for TEFB

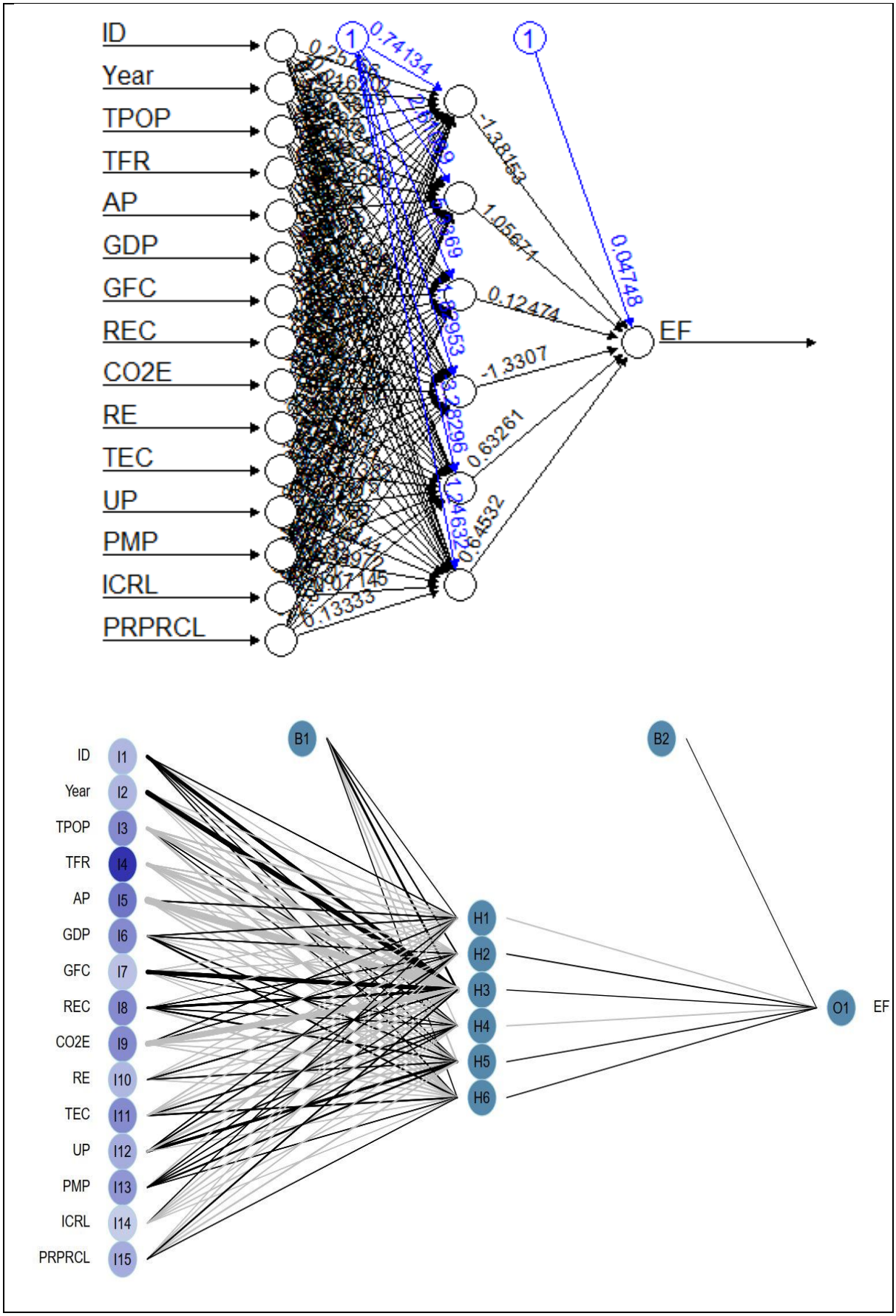


Figure 8: The best neural networks for EF

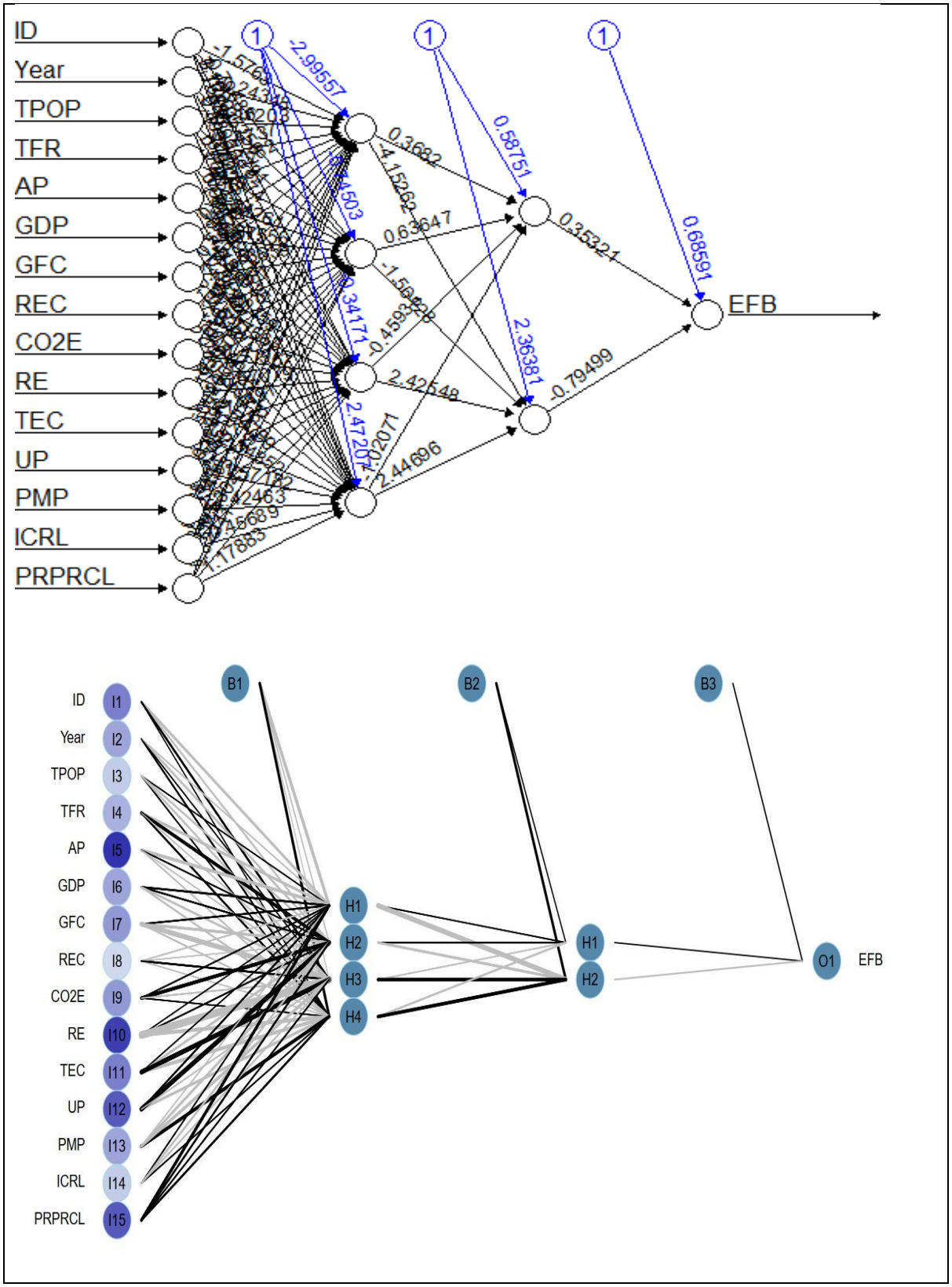
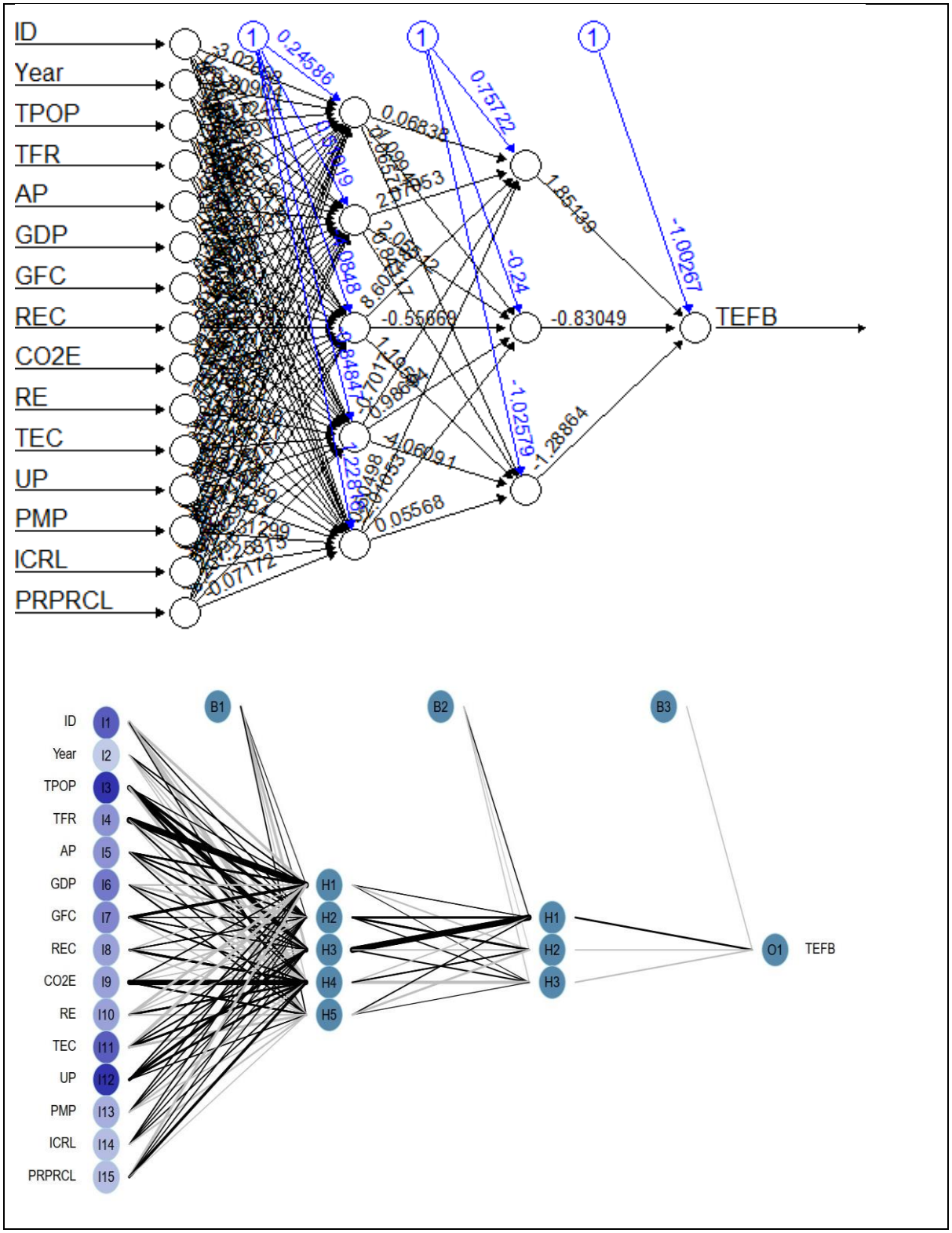


Figure 9: The best neural networks for EFB



725 **Figure 10:** The best neural networks for TEFB

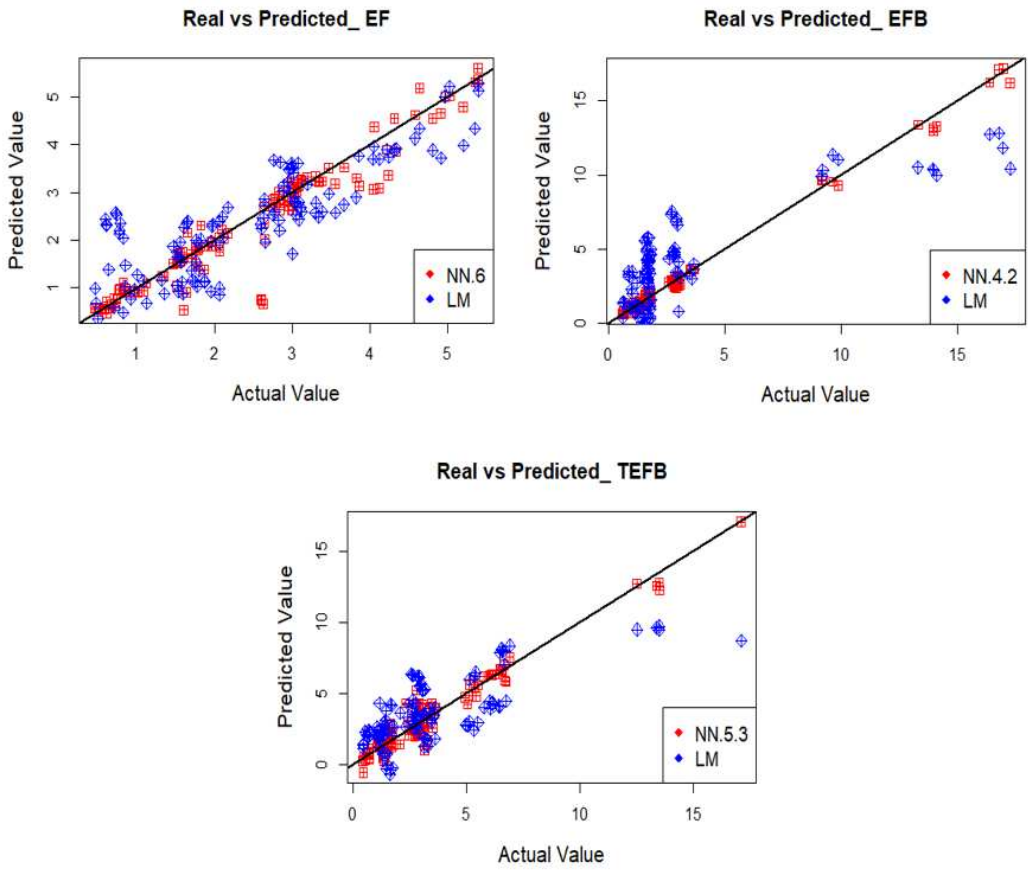


Figure 11: The appropriate ANN models' performance compared to the linear model (LM) on the test set

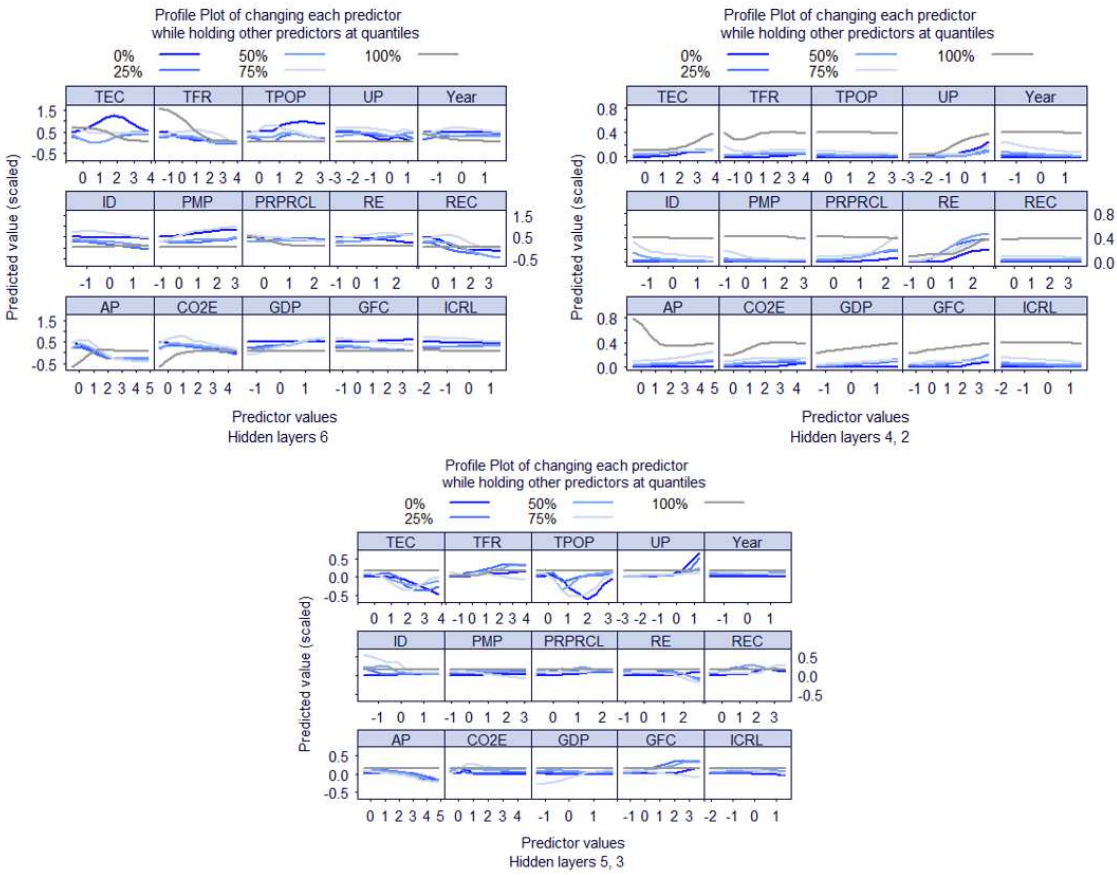


Figure 12: Changing each predictor while holding other predictors at quantiles for the suitable models

Figures

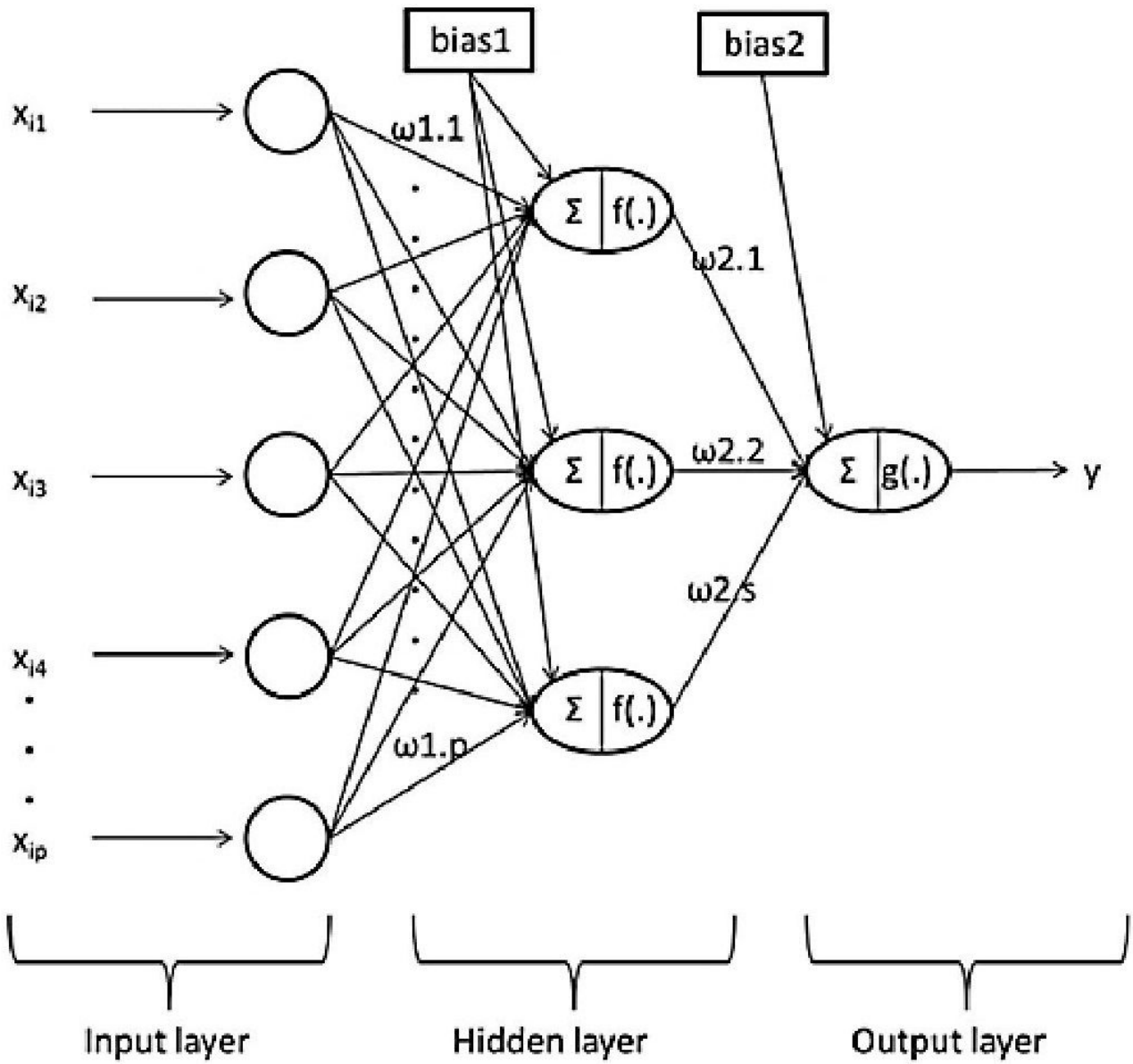


Figure 1

An ANN with one input layer (consisting of five predictors) connected to the hidden layer (with three neurons)

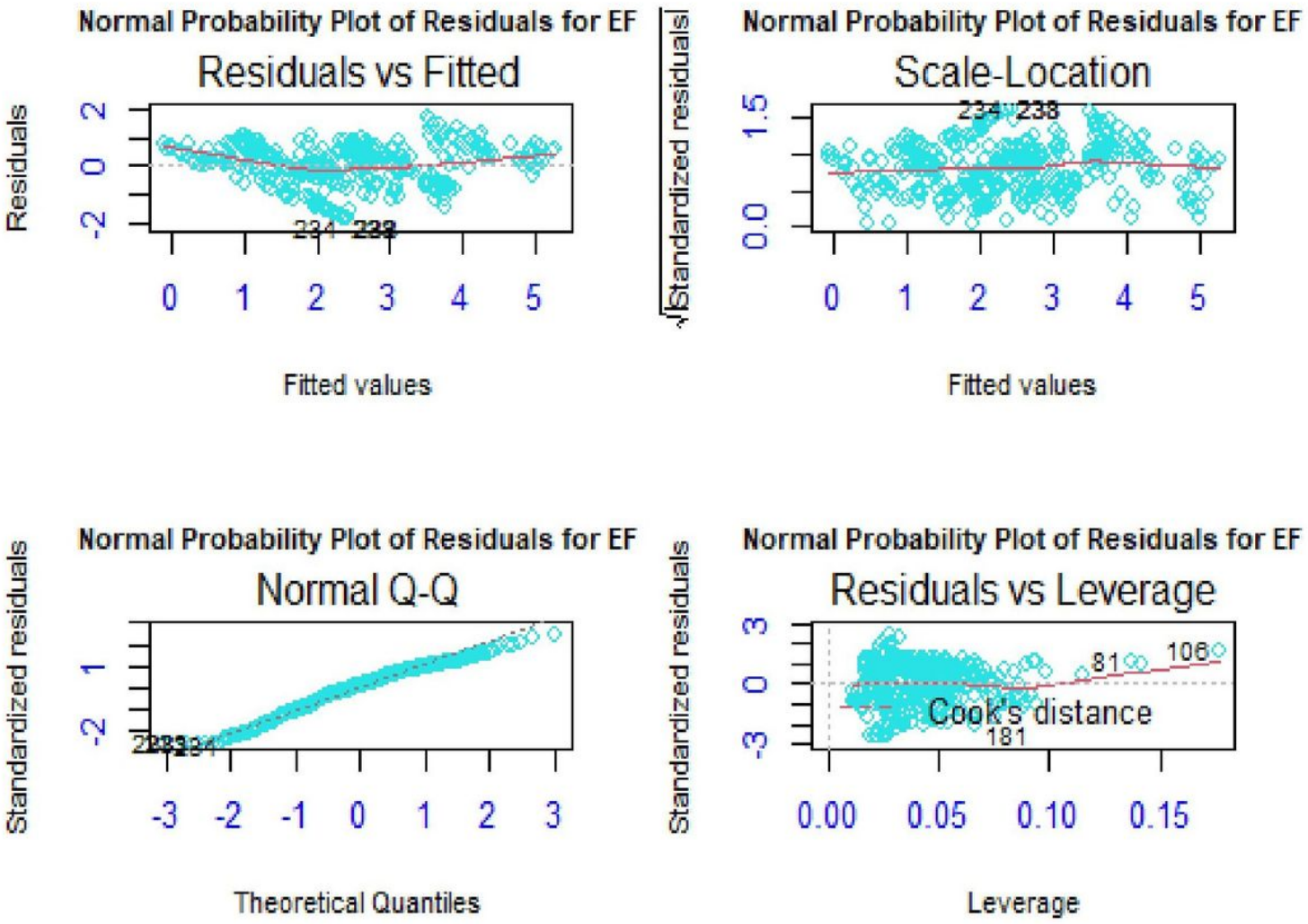


Figure 2

Normally distributed residuals for EF

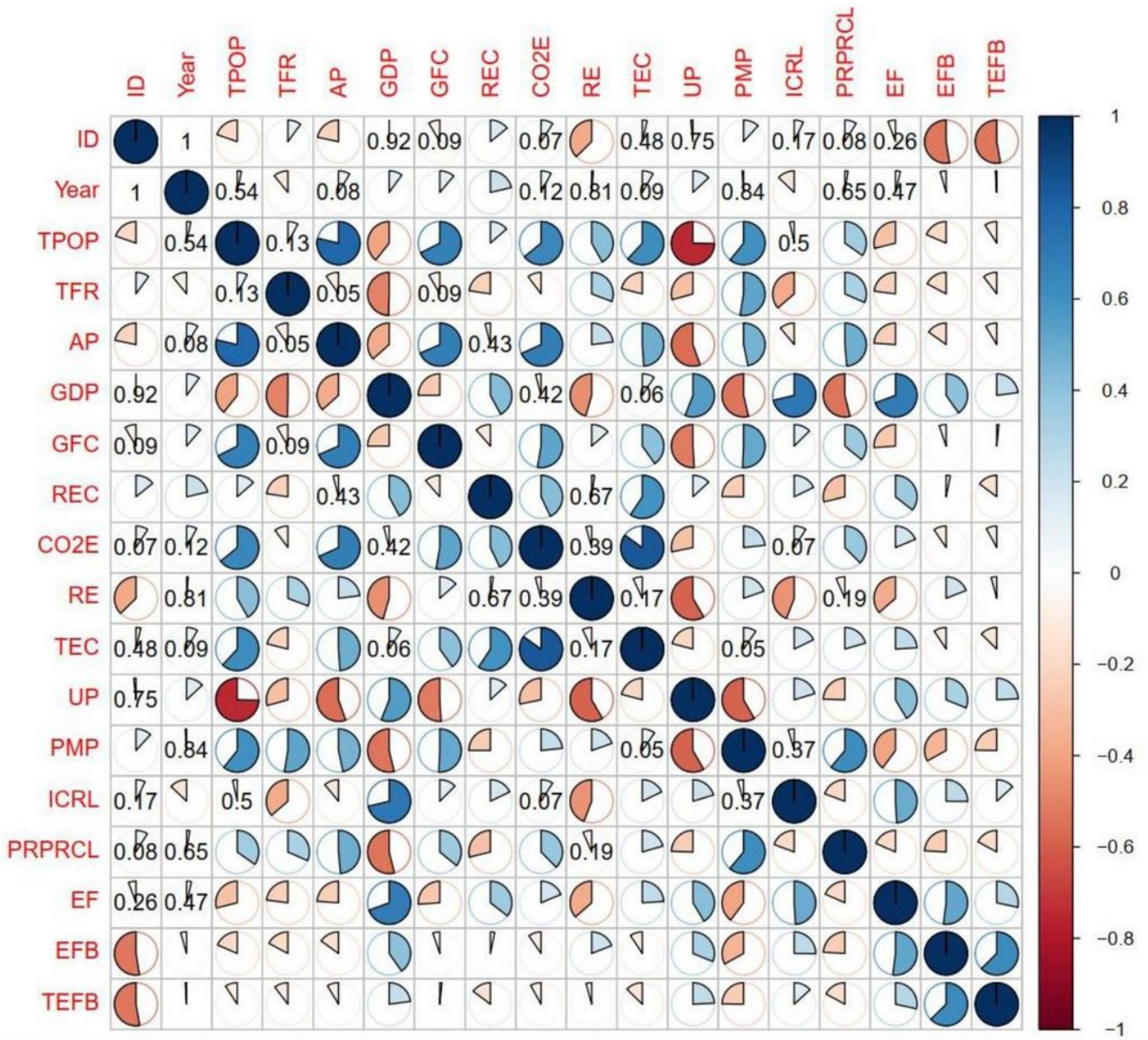


Figure 3

Correlation plot of the variables

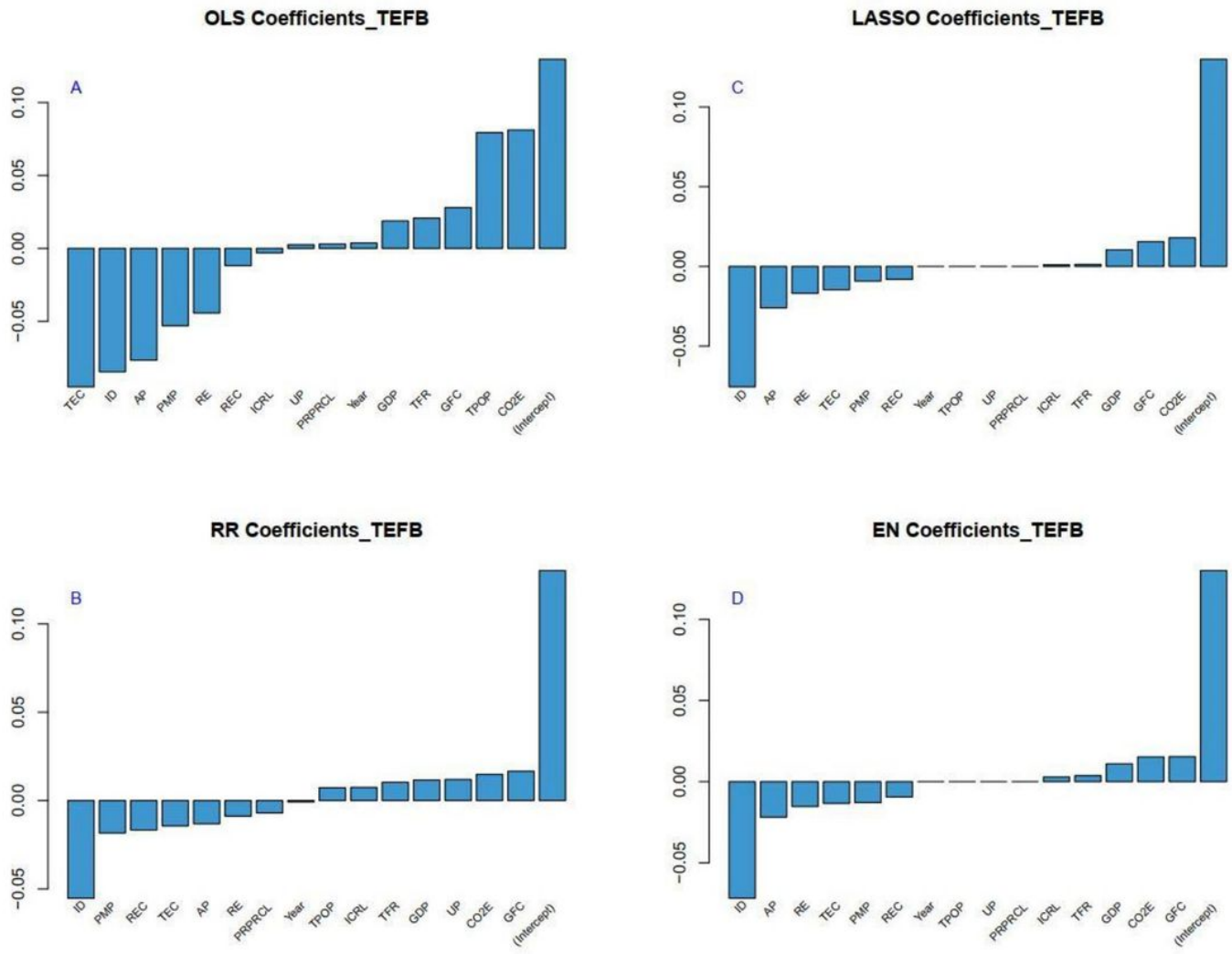


Figure 4

Shrinkage and variable selection using PR

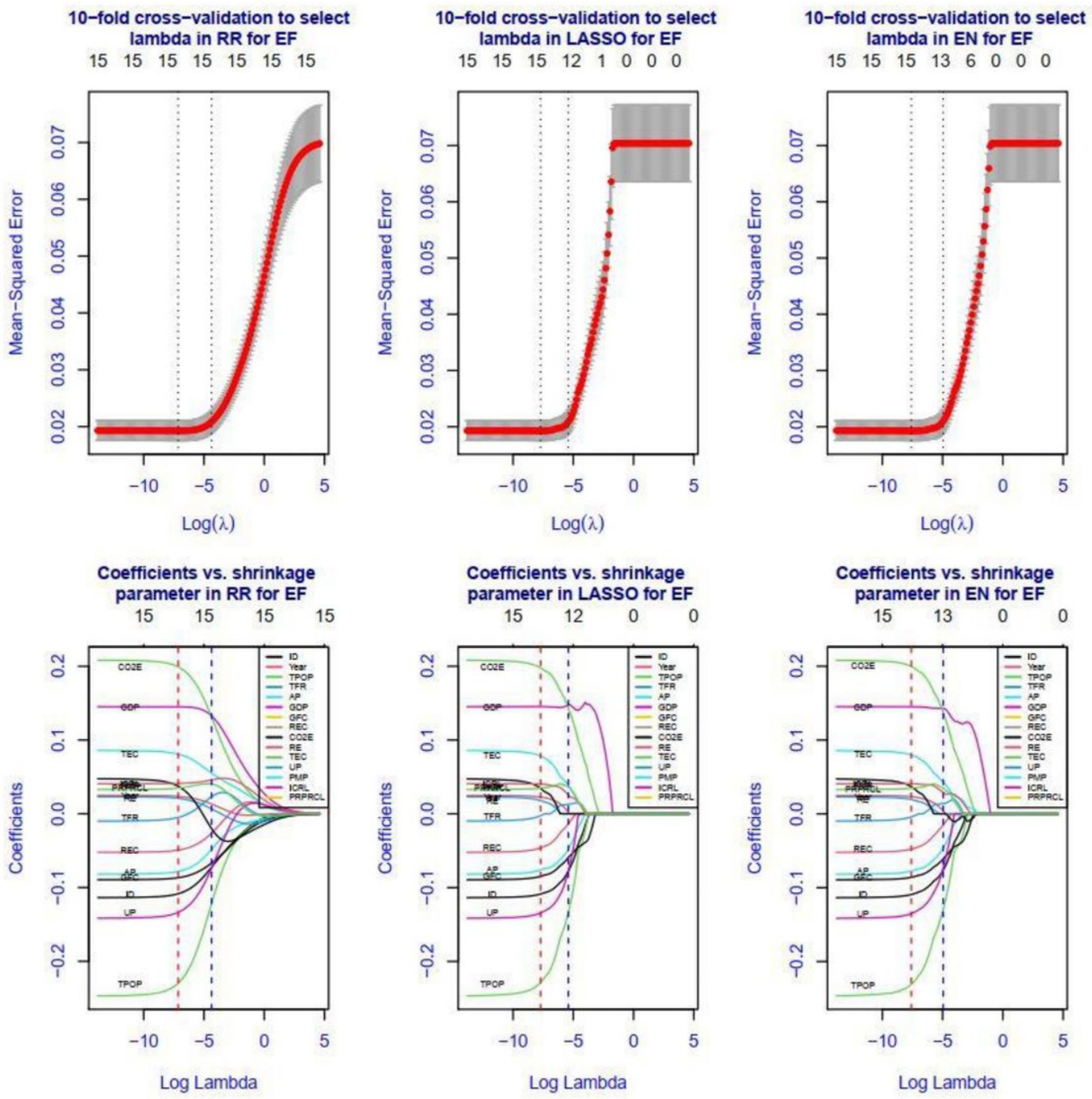


Figure 5

10-fold CV mean squared error (MSE) across the λ values for EF

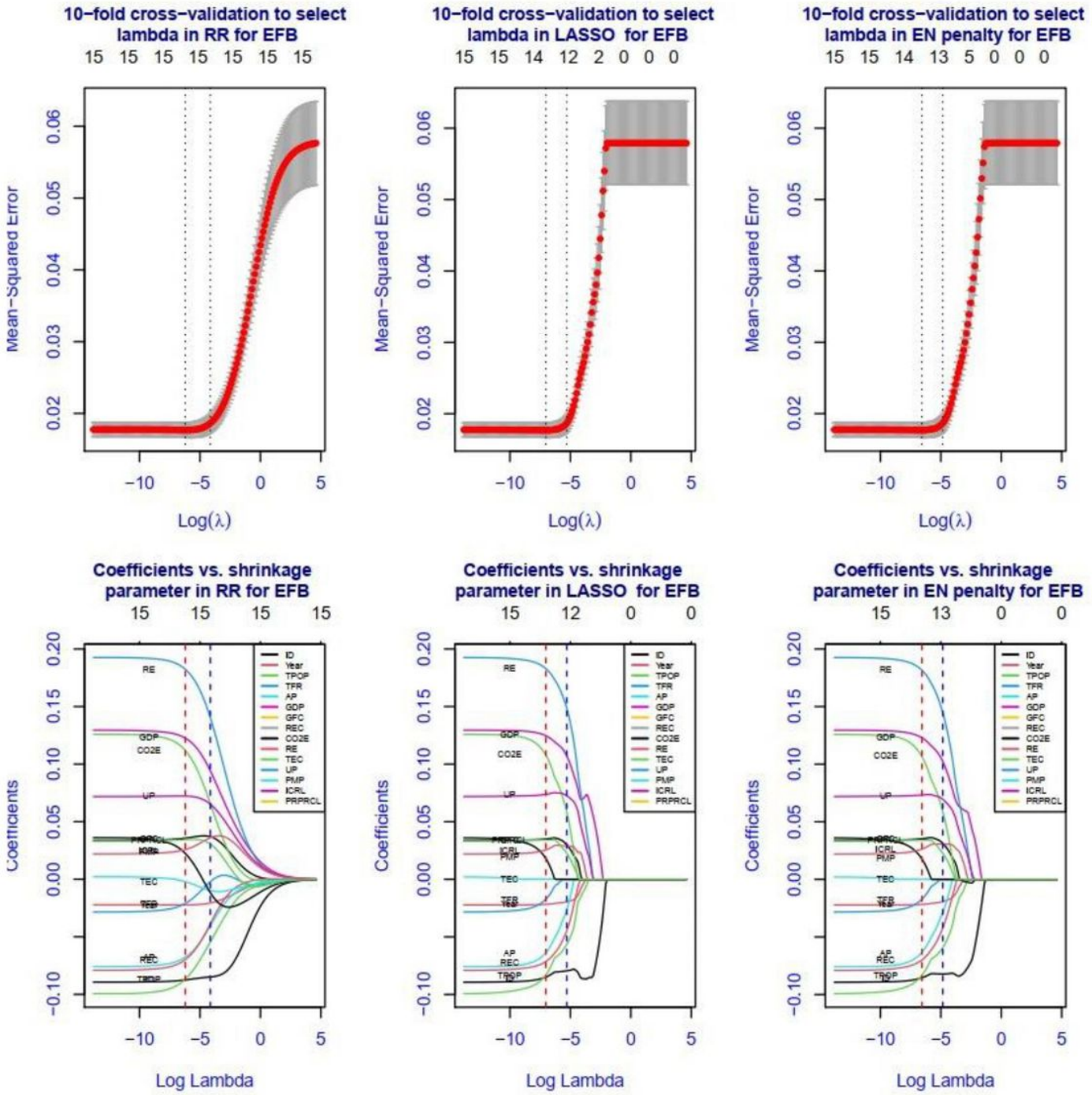


Figure 6

10-fold CV mean squared error (MSE) across the λ values for EFB

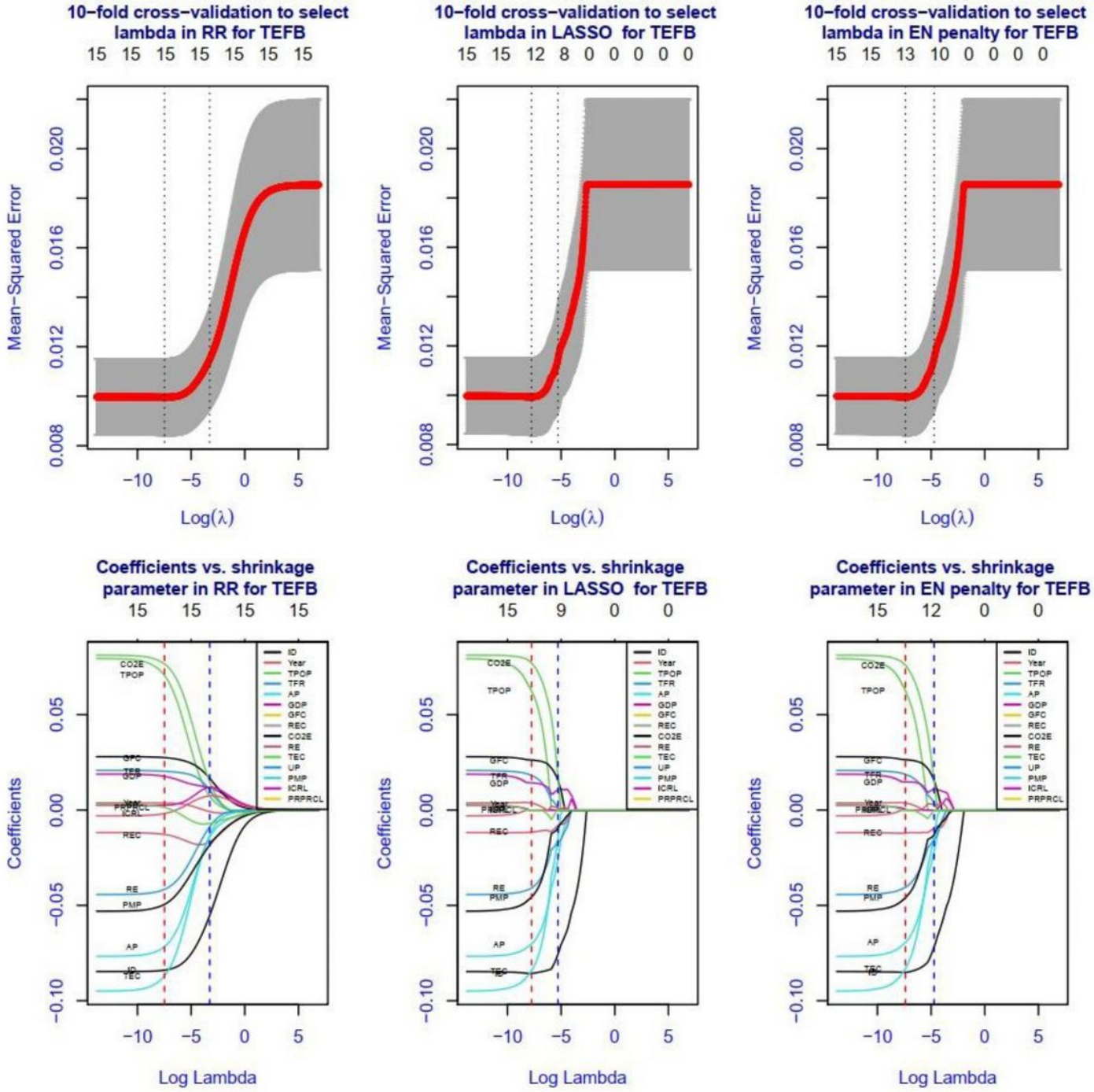


Figure 7

10-fold CV mean squared error (MSE) across the λ values for TEFB

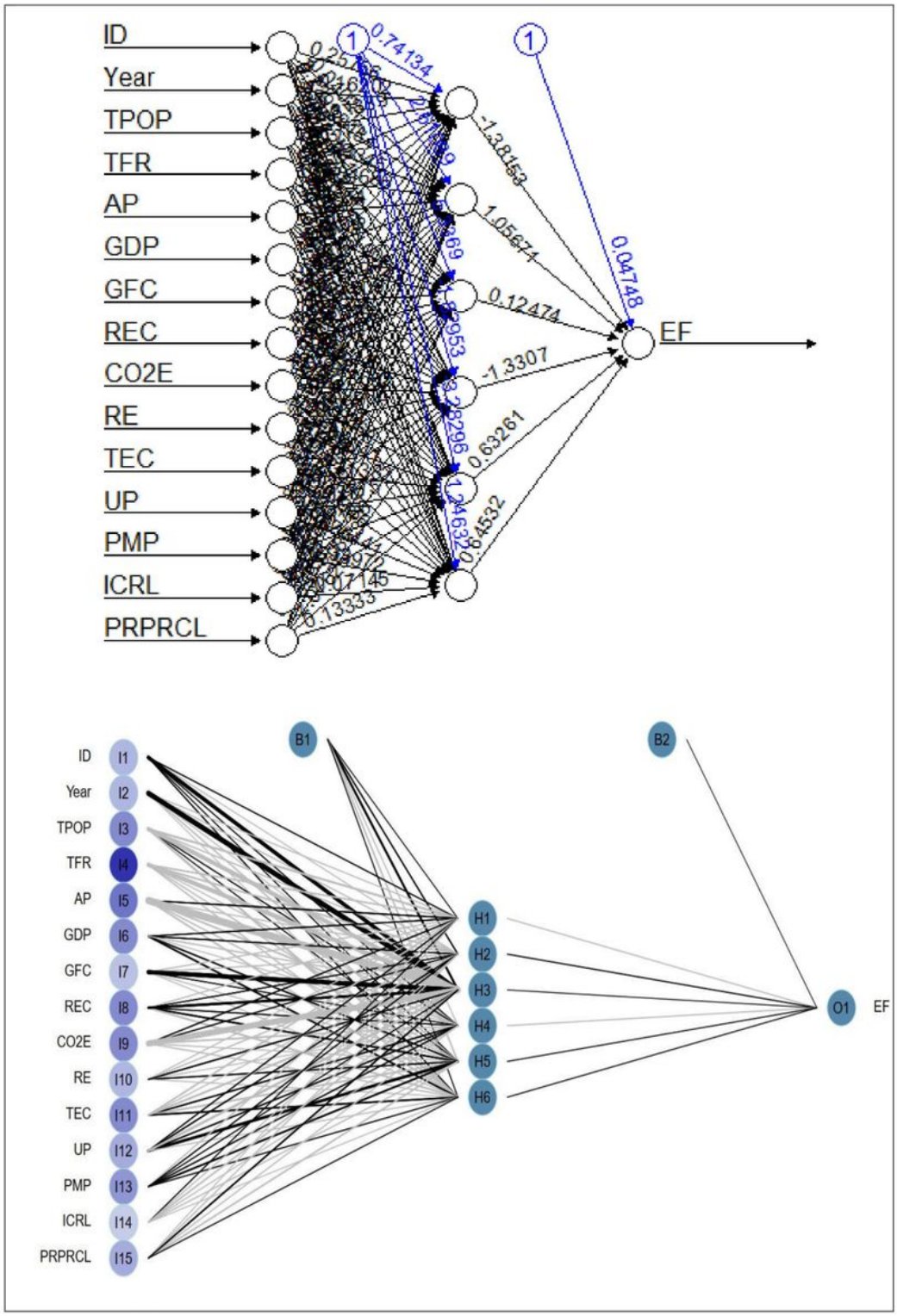


Figure 8

The best neural networks for EF

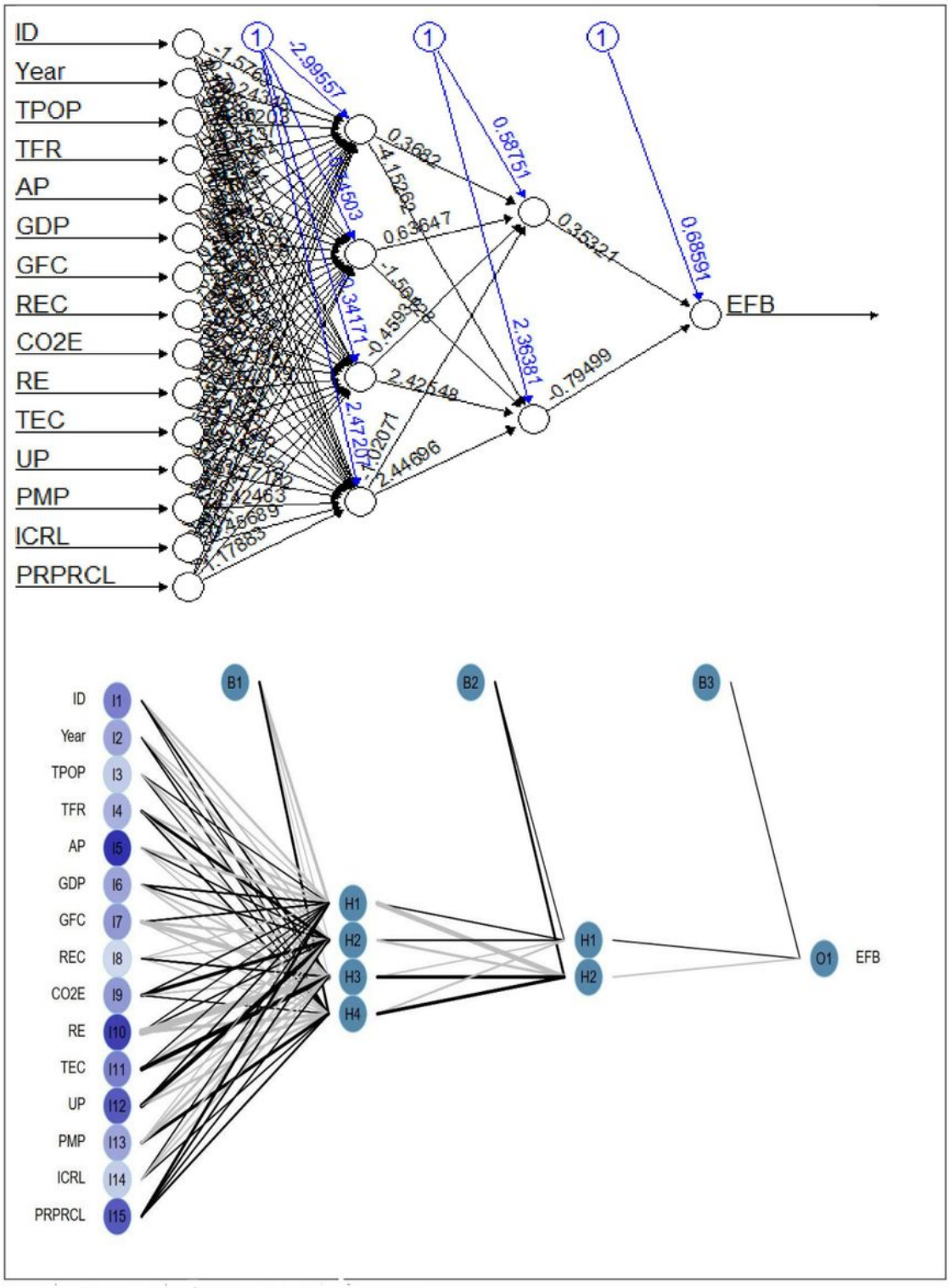


Figure 9

The best neural networks for EFB

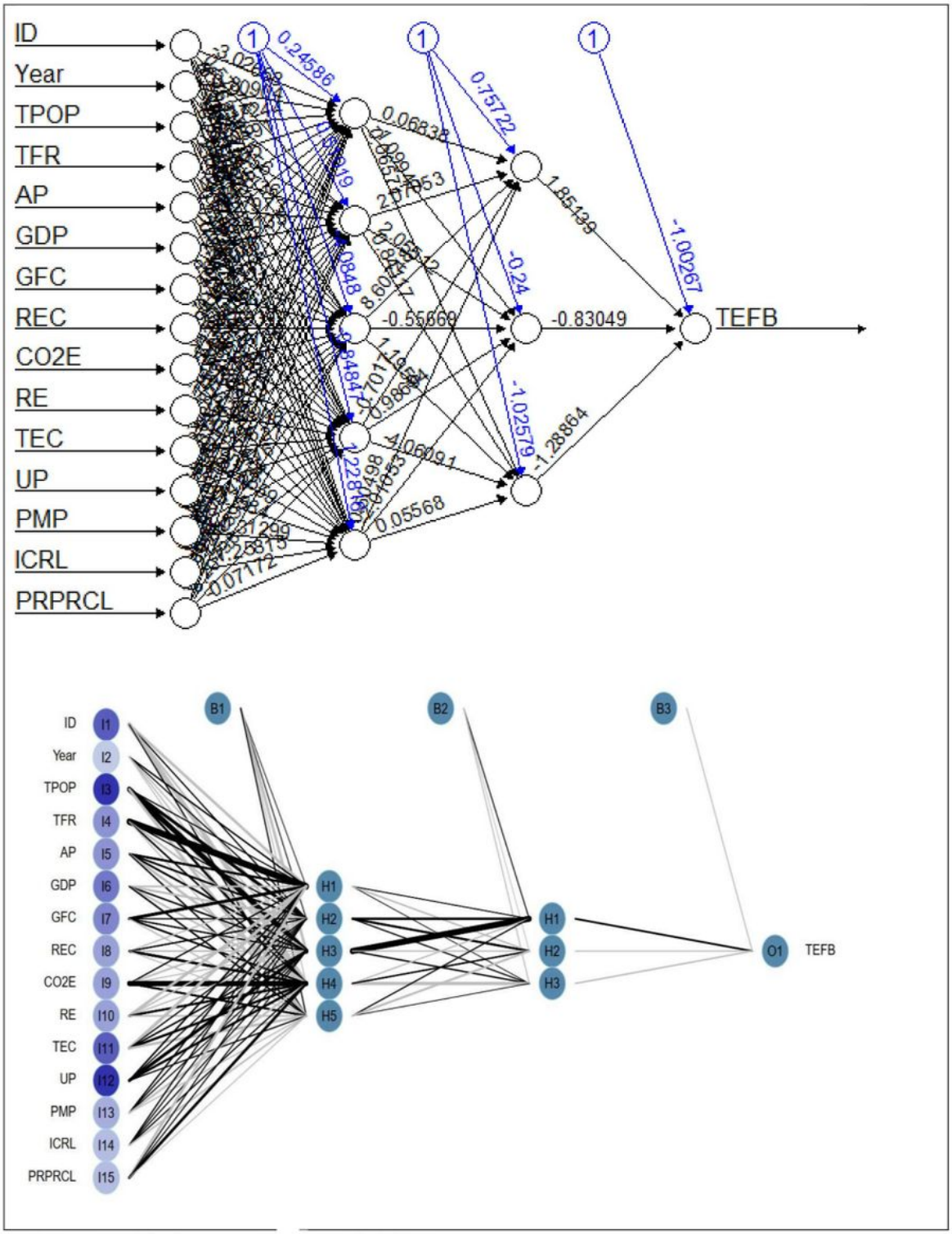


Figure 10

The best neural networks for TEFB

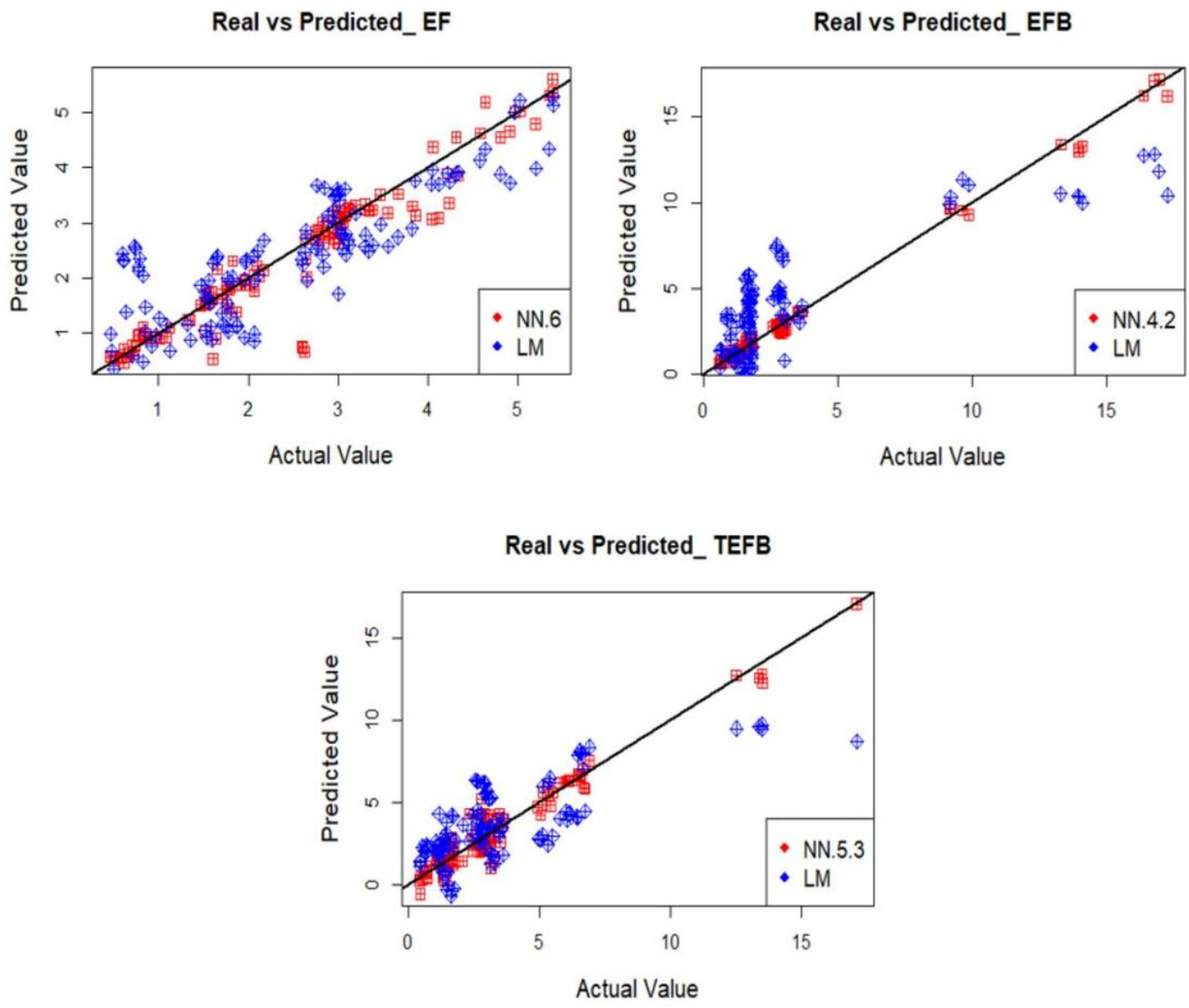


Figure 11

The appropriate ANN models' performance compared to the linear model (LM) on the test set

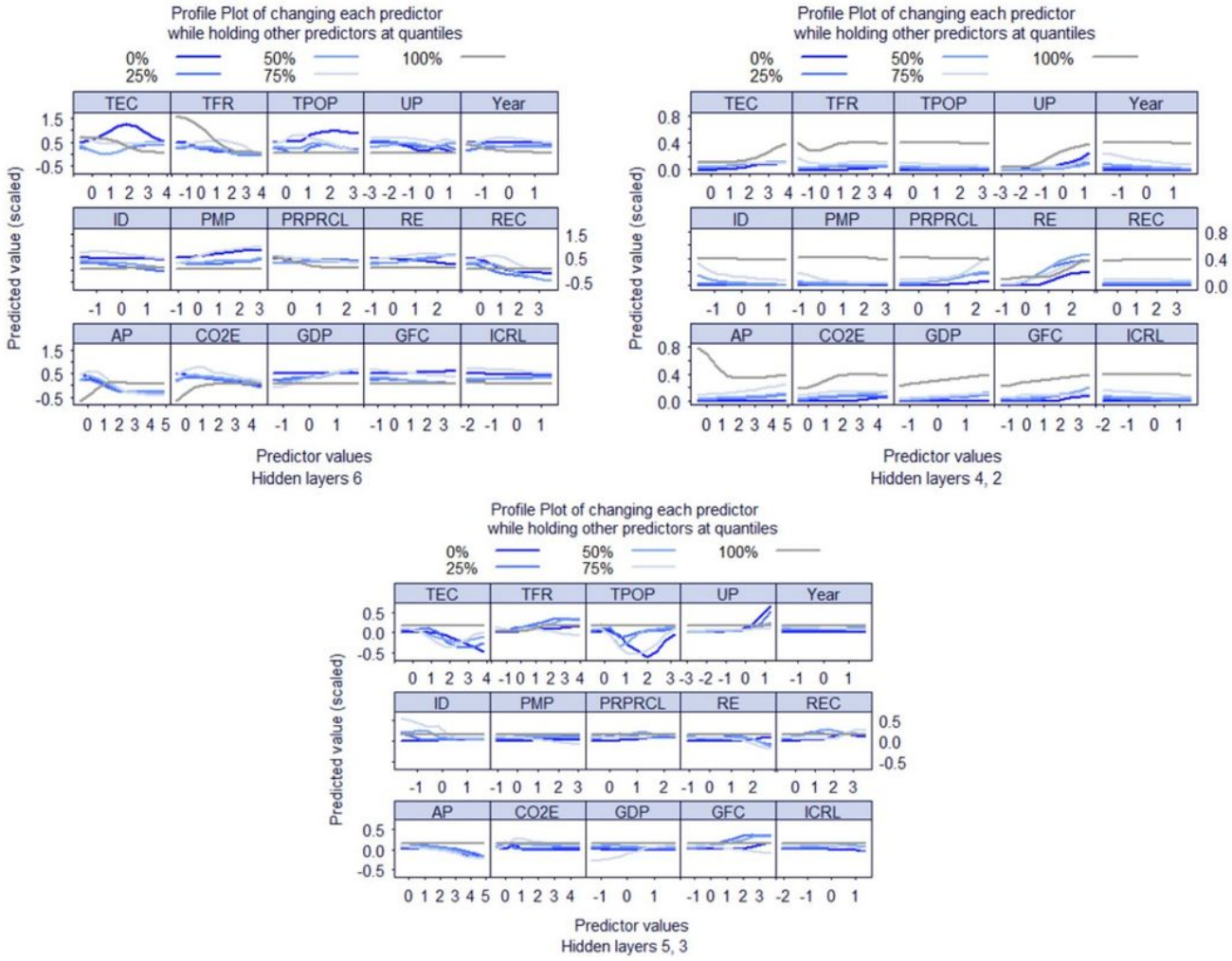


Figure 12

Changing each predictor while holding other predictors at quantiles for the suitable models