Causes and Impacts of Decreasing Chlorophyll-a in Tibet Plateau Lakes during 1986–2021 Based on Landsat Image Inversion

Shuyu Pang [1,2], Liping Zhu [1,2,*], Chong Liu [1,3] and Jianting Ju [1]

[1] State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources (TPESER), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China; psy@itpcas.ac.cn (S.P.); liuchong@itpcas.ac.cn (C.L.); jujianting@itpcas.ac.cn (J.J.)

[2] University of Chinese Academy of Sciences, Beijing 100049, China

[3] Piesat Information Technology Co., Ltd., Beijing 100195, China

* Correspondence: lpzhu@itpcas.ac.cn; Tel.: +86-10-84097093

Supplementary materials

Text: BP neutral net model

The BP neural network model comprises the input, hidden, and output layers; it automatically calculates the error according to the error formula and propagates the error back. The weight value connecting the neurons in each layer is continuously adjusted by the backpropagated error to improve the accuracy. Adjusting of network training parameters are performed in two steps. The first step is the forward-propagation of data, where the input data are transmitted from the input layer to the hidden layer via the activation function, weight calculation, and finally to the output layer to obtain the results, which are compared with the actual data, and the error is calculated. The second step is the backpropagation of error. When the error does not reach the standard requirements, the intermediate parameters have to be improved to promote the prediction accuracy. According to the calculated error, the gradient descent method adjusts the weight among each node, thus enhancing the model's accuracy. However, if one-time forward and back propagation is not able to yield the expected accuracy, the cyclic propagation will be operated in BP neural network and the model would achieve the optimal solution under the condition of setting the stop condition.

For obtaining the fitted intermediate network parameters in the BP neural net model, the input data need to be randomly divided into training set and testing set. The training data sets are applied to train the network model for achieving the expected training goal through forward and backward propagation. The testing data sets are used to test the model's adaptation derived from training results. This is required to avoid the occurrence of better fitting of the model for the training data and lower accuracy of the non-training data results.

Due to the different data dimension levels that exist in input and output layers, at times, the data normalization between 0 and 1 is needed before inputting the training data. The data normalization formula is:

$$X = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where X is the normalized data, and x_{min} and x_{max} are the minimum and maximum values of the data, respectively.

Generally, the input layer is composed of several sub-nodes, representing pre-selected relationships between actual data and predicted results. The output layer is a single neuron, reflecting the expected results. The number of nodes in the hidden layer needs to be selected artificially, and they critically affect the model's accuracy. Too many hidden layer nodes would cause the model to run too long, reduce the model's generalization, and render it prone to overfitting; too few hidden layer nodes render the model too simple to meet the accuracy requirements. The number of hidden layers is closely related to the number of input and output neurons. The formula for best node number determination is:

$$l = \sqrt{n + m} + a$$

Where l is the number of hidden layer nodes; n and m are the numbers of input layer neurons and output layer neurons, respectively; and a is a constant from 0–10