Estimation of Gross Primary Production using Numerical Weather Prediction Data with Deep Neural Network: Case Study of the Korean Peninsula

Nari Kim (1), Yang-Won Lee (2)

[1] Geomatics Research Institute, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan, 48513, Korea
[2] Department of Spatial Information Engineering, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan, 48513, Korea
Email: kimnari13@pukyong.ac.kr; modconfi@pknu.ac.kr

KEY WORDS: gross primary production; deep neural network; moderate resolution imaging spectroradiometer; local data assimilation and prediction system

ABSTRACT: The gross primary production (GPP) is a significant parameter for the carbon cycle and climate change research and is considered to be an indicator of the photosynthesis in relation to the growth of vegetation. Monitoring the GPP can analyze the change of terrestrial ecosystem and can provide important information such as vegetation health, crop yield, food security, and others. The GPP has been estimated quantitatively using the statistical models, ecosystem models, and others. Recently, Artificial Intelligence (AI) models have been used in research to improve the estimation accuracy for environmental variables. Particularly, the deep neural network (DNN) is evaluated as a more advanced technique that combines with the advantages of traditional neural networks and machine learning methods through an intensive learning process in a deep network. Therefore, we expect that DNN can contribute to improving the estimation accuracy of the GPP. The objective of this study is to develop an optimized GPP estimation model based on DNN and to evaluate its accuracy. The GPP was obtained from Terra Moderate Resolution Imaging Spectroradiometer (MODIS) and input data was acquired from the Local Data Assimilation and Prediction System (LDAPS) from the Korea Meteorological Administration (KMA). Through this study, the optimized DNN model built in this study can estimate the GPP in the Korean Peninsula accurately, and it will be able to be applicable to predict the crop yield and to evaluate the drought.

1. INTRODUCTION

The gross primary production (GPP) is a significant parameter for the carbon cycle and climate change research and plays an important role as an indicator of the photosynthesis in relation to the growth of vegetation. It depends on continual change in environmental factors such as temperature, humidity, precipitation, soil property, and others.

For measurement of GPP, the eddy covariance technique provides the best approach, but it is difficult to partition respiration of ecosystems into autotrophic respiration and heterotrophic respiration (Li et al., 2007). To overcome this difficulty, satellite remote sensing can provide consistent and systematic observations of vegetation and ecosystems, and also can overcome the lack of extensive flux tower observations over large areas (Running et al., 2000, Behrenfeld et al., 2001). In addition, artificial intelligence (AI) approaches have become necessary to solve nonlinear and complex problems, and it can estimate the environment variables more accurately than previous method such as statistical and empirical models.

The objective of this study is to develop an optimized GPP estimation model based on DNN and to evaluate its accuracy. The GPP was obtained from Terra Moderate Resolution Imaging Spectroradiometer (MODIS) and input data was acquired from the Local Data Assimilation and Prediction System (LDAPS) from the Korea Meteorological Administration (KMA).

2. DATA

2.1 Study Area and Period

This study focused on the cropland of South Korea, and the study period was set from May to September to capture the growing season of crop and between 2015 and 2017.

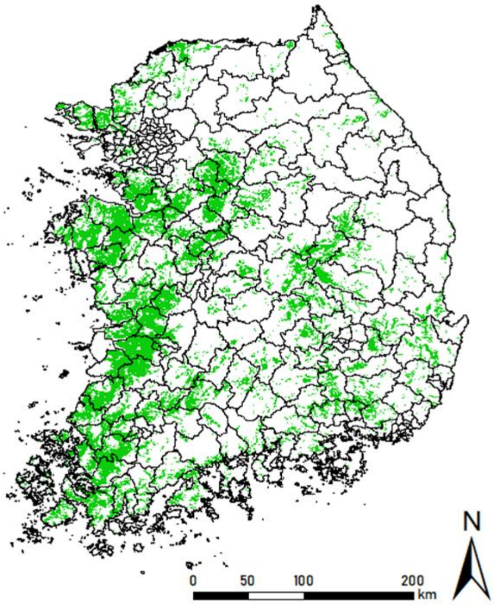

Figure 1. Distribution of croplands in South Korea.

2.2 Materials

MODIS GPP (MOD17A2) is composited over an 8-day interval with 500 m spatial resolution. For this study, it used as reference data. For masking the croplands, we extracted the pixels that were recorded as cropland (landcover ID = 12, 14) from the MODIS land-cover product (MCD12Q1).

The LDAPS based on the unified model (UM) of the United Kingdom has been improved to meet the Korean circumstances and it is being used as the operational forecast model. It is a high-resolution model with the spatial resolution of 1.5 km and the time resolution of 3-hour and it covers over the entire Korean peninsula. Table 1 showed the input variables used for this study and the correlation coefficient of input variables against GPP.

Table 1. Correlation coefficients of input variables against GPP.

Factor	Variable name	Corr.
RNET	Net Radiation	0.194
TEMP	Temperature (K)	0.321
RH	Relative Humidity (%)	0.079
PPT	Precipitation (kg m-2)	-0.326
LE	Latent Heat Net Flux (W m-2)	0.243
SoilTemp	Soil Temperature (K)	0.316

3. DEEP NEURAL NETWORK

The classic ANN has a local minima problem in which an optimization process often stops at a locally, rather than globally, optimized state. In addition, generic machine learning models sometimes have problems with overfitting, in which they cannot handle data with outliers due to excessive learning from the given dataset. Such problems can be resolved by DNNs through an intensive optimization process in a deep network structure. To handle local minima and issues with overfitting, L1/L2 regularization can be employed to ensure sparsity (L1) and simplicity (L2) of the DNN model. Also, backward and forward optimization is conducted in the back-propagation algorithm to improve accuracy. The problem of vanishing gradients of loss functions, which may occur during the back-propagation process, can be managed by applying appropriate activation functions, such as sigmoid and rectified linear units (ReLU). The drop-out method deals with unexpected outliers via a learning mechanism in which the DNN model becomes more robust to extreme cases through iterations of a type of handicapped training with randomly deleted links and nodes (Pham et al., 2014). In addition, a weight and bias set built in an existing DNN model can be imported as an initial value of a new DNN model for more custom-tailored training. This is called pre-training and transfer learning, which can improve the optimization of a DNN model. Fine-tuning can also be incorporated into the optimization process to adjust the weight and bias set in more detail, by including additional training data (Erhan et al., 2010). We set up the configurations of our DNN model through the parameter optimization procedure presented in Figure 3.

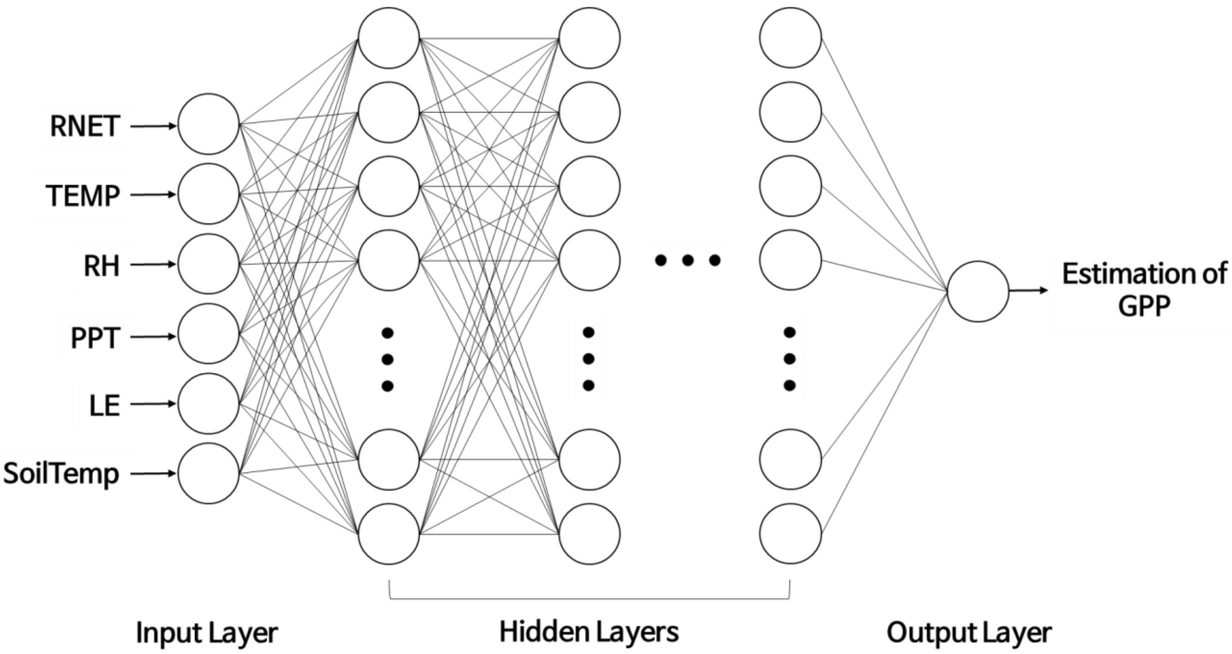

Figure 2. The diagram of DNN in this study.

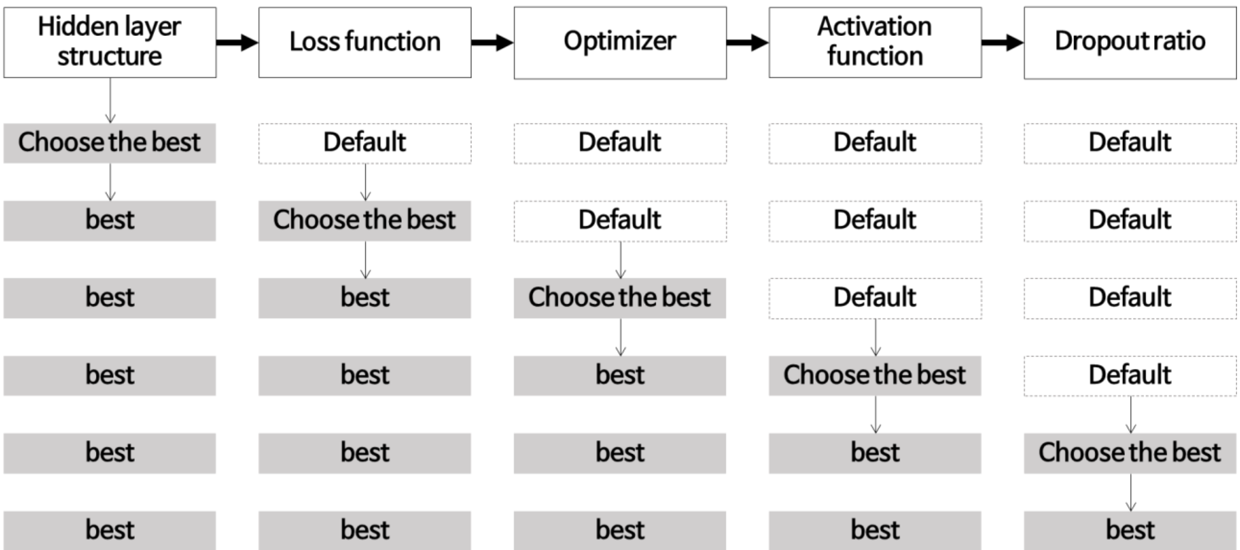

Figure 3. Parameter optimization process for a deep neural network (DNN) model in this study (Kim et al., 2019).

4. RESULTS

We conducted a fine-tuning process for the DNN model for GPP estimation by adjusting detailed configurations. The best structure of hidden layers was determined to be 500 × 500 × 500 × 500 nodes, which was derived from an optimization experiment using several layers of structures such as 50 to 500 as hidden nodes and 2 to 4 as hidden layers (Table 2).

Table 2. Optimal configuration for the DNN model.

Option	Setting
Hidden Layer	500 × 500 × 500 × 500
Loss function	Sum of squared errors (SSE)
Optimizer	Adaptive delta (AdaDelta)
Activation function	Rectified linear unit (ReLU)

We used the random-sampling method to train and validate the DNN model. It was trained by randomly extracting 80 % of the data, and the remaining 20 % was validated. The mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and the correlation coefficient (Corr.) between the observed and the estimated GPP during the period of 2015 and 2017 were calculated for the validation statistics.

For comparison's sake, we conducted not only the DNN model but random forest (RF) and support vector machine (SVM) which are known for machine learning models (Table 3). The correlation coefficient (Corr.) and MAPE of the optimized DNN model showed 0.741 and 15.5 %, respectively. The validation statistics of the DNN model showed the highest accuracy compared to the RF and SVM model. It means the optimized DNN model in this study can estimate the GPP.

Table 3. Summary of validation statistics for GPP.

Model	MAE (kg C m^{-2})	RMSE (kg C m^{-2})	MAPE (%)	Corr.
RF	0.006	0.008	21.6	0.535
SVM	0.006	0.008	21.0	0.587
DNN	0.004	0.006	15.5	0.741

5. CONCLUSION

This study described the development of the optimized DNN model for GPP estimation in South Korea, 2015-2017. We optimized DNN model by adjusting hidden layer, activation function, optimizer, dropout ratio, and others to improve the estimation accuracy of GPP. The result of optimized DNN model showed the highest accuracies in terms of correlation coefficient, RMSE, and MAPE compared with machine learning methods. Through this study, the optimized DNN model built in this study can estimate the GPP in South Korea accurately, and it will be able to be applicable to North Korea.

REFERENCE

Behrenfeld, M.J., Randerson, J.T., McClain, C.R., Feldman, G.C., Los, S.O., Tucker, C.J., Falkowski, P.G., Field, C.B., Frouin, R., Esaias, W.E., Kolber, D.D, and Pollack, N.H., 2001, Biospheric Primary Production during an ENSO Transition, Science, 291, pp. 2594-2597.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., and Vincent, P., 2010, Why Does Unsupervised Pre-training Help Deep Learning? J. Mach. Learn. Res., 11, pp. 625-660.

Li, Z.Q., You, G.R., Xiao, X.M., Li, Y.N., Zhao, X.Q., Ren, C.Y., Zhang, L.M., and Fu, Y.L., 2007, Modeling Gross Primary Production of Alpine Ecosystems in the Tibetan Plateau using MODIS Images and Climate Data, Remote Sensing of Environment, 107, pp. 510-519.

Kim, N., Ha, K.-J., Park, N.-W., Cho, J., Hong, S., and Lee, Y.-W., 2019. A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006-2015, ISPRS International Journal of Geo-Information, 8(5), 240.

Pham, V., Bluche, T., Kernorvant, C., and Louradour, J., 2014. Drop-out Improves Recurrent Neural Netowkrs for Handwriting Recognition, In Proceedings of the 2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), Crete, Greece, 1-4 September 2014.

Running, S.W., Thornton, P.E., Nemani, R., Glassy, and J.M., 2000, Global Terrestrial Gross and Net Primary Productivity from the Earth Observing System, O.E. Sala, R.B. Jackson, H.A. Mooney (Eds.), Methods in Ecosystem Science, Springer, New York, pp. 44-57.