Crop Yield Estimation in the North China Plain from 2001 to 2016 using Multi-source Remote Sensing Data and Process-based FGM Model

Qiaoli Wu^{1,2}, Xinyao Wang^{1,2}, Jie Jiang^{1,2*}, Shaoyuan Chen³

¹ School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture(BUCEA), Beijing, 100044, China; wuqiaoli @bucea.edu.cn (Q.W.), 2108570021123@stu.bucea.edu.cn(X.W.), chensy@pku.edu.cn(S. C.)

² Key Laboratory of Urban Spatial Information, Ministry of Natural Resources of the People's Republic of China, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

³ Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing, 100091,

China

* Correspondence:jiangjie@bucea.edu.cn

KEY WORDS: Gross primary productivity (GPP), crop yield, process-based model, North China Plain

ABSTRACT:

Gross primary productivity (GPP) is an essential indicator of vegetation growth that reflects ecosystem function. GPP is the original source of energy entering cropland ecosystem and thus could serve as a direct indicator of crop yield. In the context of increasing population, changing climate, and decreasing available resources, accurate monitoring and forecasting of food and crop yields play an essential role in sustainable human development. In this study, the process-based Farquhar GPP model (FGM) driven by multisource remote sensing data was implemented to estimate the spatial and temporal dynamics of GPP in crop-growing areas of the North China Plain from 2001 to 2016. We found that the GPP of crops in the North China Plain is relatively high in the southern provinces while lower in the northern part. The GPP values showed a significant increasing trend from 2001 to 2016 (+2.19 Mt C yr⁻¹, P<0.05). Based on crop yield statistical yearbook, we found that GPP is well correlated with crop yield (R²=0.98, RMSE = 10.4 Mt yr⁻¹). Thus, we constructed an empirical regression model between GPP and crop yield (i.e., 'GPP-yield' empirical model). Finally, time-series GPP data and the 'GPP-yield' model were applied the crop yield in the North China Plain with spatial and temporal continuity. We found that the crop yield in the North China Plain changed in accordance with GPP, and also showed a significant increasing trend from 2001 to 2016, with a mean increasing rate of +2.84 Mt yr⁻¹ (P<0.05, R²=0.16, RMSE = 31.73 Mt yr⁻¹). This study proved an example of large-scale crop yield estimation using multi-source remote sensing data.

1. INTRODUCTION

Gross Primary Productivity (GPP) is defined as the total amount of organic carbon fixed by green plants through photosynthesis per unit of time and land area (Zhang et al, 2021). GPP plays a pivotal role in the global carbon cycle (Chen et al, 2021). GPP indicates the total amount of energy entering the terrestrial ecosystem to support animal consumption and human lives. Thus, crop productivity is a crucial indicator of food production, while the production of crops is highly dependent on climatic conditions. Therefore, quantitative estimation of primary crop productivity and driving force analysis are crucial for understanding the growth state of crops and their changes.

China's grain production is mainly concentrated in four regions: East China, Northeast China, Central China, and the North China Plain. The North China Plain has deep and fertile soils, and the main grain crops are wheat, rice, and corn. The North China Plain accounts for 13.66% of Chinese grain production and is one of the significant grain-growing areas in China. Therefore, it is important to estimate the productivity of crops in the North China Plain.

With the development of remote sensing technology, numerous vegetation productivity models have been proposed by scholars and continuously improved in order to stimulate regional or global crop productivity. These models can be broadly classified into three categories: empirical models, light use efficiency (LUE) models, and photosynthetic process-based models.

First, the empirical models are generally based on regression relationships between vegetation index and crop biomass, which is simple in form and widely adopted. For example, Anup K. Prasad et al. (2006) conducted crop yield assessment and prediction based on 19 years records of Normalized Difference Vegetation Index (NDVI), soil moisture, surface temperature, and rainfall data in Iowa, USA. Zhao et al. (2011) used the validated and corrected the MODIS GPP product to estimate the total GPP of winter wheat in the North China Plain during the 2010 growing season. Wang (2020) used MODIS enhanced vegetation index (EVI) data and auxiliary data for estimating winter wheat yield. However, the common drawback of these studies is that empirical models are often too empirical and lack mechanisms behind photosynthesis.

The second type is the LUE model. LUE models are most widely used because of the simple structure and physical basis. For example, Tao et al. (2005) used the CASA model and GLO-PEM2 model to estimate maize yield in China. Huang (2020) used the CASA light energy utilization model to simulate the spatial and temporal variation of total primary productivity of vegetation in the North China Plain over the past 19 years. However, the mechanistic expression of LUE models is still lacking, and the input variables for LUE models are generally limited to solar radiation, leaf area index, and environmental factors (such as temperature, and humidity). However, in the context of continuous rising atmospheric CO₂ concentration and global climate change, LUE model rarely considers changes in ambient CO₂ concentration, which could lead to substantial uncertainties in GPP estimations.

The third category are the process-based models, such as BEPS and FGM. Wang et al. (2009) improved the BEPS (Boreal Ecosystem Productivity Simulator) model to estimate winter wheat yield using remote sensing mechanistic models. Ji et al. (2021) simulated the gross primary productivity (GPP) and net primary productivity (NPP) of Chinese Moso bamboo forests from 2001-2018 using the BEPS model, a northern hemisphere ecosystem productivity simulator. Chen et al. (2021) implemented a large-scale canopy photosynthetic capacity model, the remote sensing-driving Farquhar GPP Model (FGM). The FGM combines the Farquhar model, the two-leaf model, and the radiative transfer process model to estimate GPP using multi-source remote sensing data. The FGM model overcomes the shortcomings of the current photosynthesis rate estimation process model with a complex structure, numerous parameters, and great computational effort. Process-based models provide an avenue to improve GPP estimation and thus crop yield estimation at a large-scale. However, the current domestic and international productivity studies are mainly focused on the effects of land use change and vegetation cover change on GPP. Studies of continuous crop yield estimation over long periods in North China Plain using process-based models are still lacking.

In this study, the process-based GPP model (i.e., FGM) was used to estimate the total primary productivity (GPP) of croplands in the North China Plain by combining long time series of multi-source remote sensing data, meteorological data and atmospheric CO_2 concentration observations. We hypothesized that GPP is well correlated with field measurements of crop production. Specifically, we address two scientific questions: (1) Can we build a reliable 'GPP-yield' model for crop yield estimation based on crop yield data obtained from the statistical yearbook? (2) How did GPP and crop yield change in the North China Plain from 2001 to 2016? This study helps us better understand the relationship between GPP and crop yield.

2. MATERIALS AND METHODS

2.1 Study area

The study area from northeast to southwest includes the Brigade area in the south of Liaodong Peninsula, Tianjin, Beijing, and all of Shandong Province; most of Henan Province, Hebei Province, Shanxi Province, and Shaanxi Province; and eastern and southern Gansu Province and northern Jiangsu and northern Anhui (Figure 1).

The area has a continental climate except for the coastal areas. The average elevation of the study area is about 200m with an average annual temperature between 9 °C and 15 °C. The average temperature of the coldest month ranges from 0.7 to 10.7 °C. The annual precipitation in the region ranges from 440 to 980 mm, with most areas around 200 mm. The distribution of precipitation between different seasons is uneven. The precipitation mostly happened in summer, especially in July and August. The major crop types grown in our study area mainly include winter wheat, summer corn, summer soybean, and spring corn.

2.2 Dataset

2.2.1 Remote sensing data

This study used the yearly MODIS land use land cover (LULC) data (i.e., MCD12Q1) with a temporal resolution of 1 year and a spatial resolution of 500 m. The data were resampled to a target spatial resolution of 1km using the nearest neighbor interpolation method. The GLASS leaf area index (LAI) and downward radiation (DSR) shortwave products were downloaded the from website (http://glassproduct.bnu.edu.cn/index.html). The GLASS LAI product is produced based on the generalized regression neural networks (GRNNs) method using pre-processed long-time series MODIS/AVHRR reflectance data, with a spatial resolution of 1 km and a temporal resolution of 8 days. The DSR data is retrieved from the spectral reflectance of the MODIS top-ofatmosphere (TOA), with a temporal resolution of 1 day and a



Figure 1. The distribution of land cover types in the study areas.

spatial resolution of 5 km. The DSR data was eventually sampled to a spatial resolution of 1 km using a bilinear interpolation method.

2.2.2 Crop yield data

The yield of crops in the North China Plain from 2001-2016 was obtained from the grain production data statistical yearbook platform (https://www.yearbookchina.com/index.aspx).

Crop productivity is a crucial indicator of grain production, and quantitative estimation of primary crop yield under climate change and land use change, which is essential for human surviving. Based on the Origin platform, we conducted a linear regression of crop GPP and crop yield of each province in the study area from 2001 to 2006. Then, we built a "GPP-yield" regression model for estimating crop yield from GPP across the North China Plain and analyzed the correlation between GPP and yield using Pearson's correlation coefficient. The regression model and Pearson correlation coefficient was calculated as follows:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

And

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$
(2)

where β_0 and β_1 are coefficients, ε_i is the random error. *n* is the total number of samples sampled at the time of analysis, x_i and y_i are the data values of two variables, \bar{x} and \bar{y} are the mean of the sample data for each of the two variables.

2.3 Farquhar GPP Model (FGM)

The Farquhar GPP Model (FGM) is based on the Farquhar photosynthetic model, which is the key component to estimate leaf photosynthesis rate in most terrestrial ecosystem models. In order to upscale leaf photosynthesis rate to canopy scale, FGM synthesizes the two-leaf model with the canopy radiative transfer model. By improving the basic framework of the model of Song et al. (2009), Chen et al., (2021) introduces the equation of the relationship between leaf cell CO₂ concentration and atmospheric CO₂ concentration proposed by Medlyn et al. (2012) to improve model calculation efficiency in solving leaf photosynthesis rate. A new model for GPP estimation based on photosynthetic processes, namely FGM, is proposed, which is more suitable for large-scale GPP estimation and dramatically improves the model's computational efficiency.

The two-leaf model (Chen et al., 1999; Pury and Farquhar, 1997) divides canopy leaves into sunlit and shaded leaves The photosynthetic active radiation (PAR) absorbed by sunlit and shaded leaves and the photosynthetic rate of sunlit and shaded leaves were estimated separately. The total GPP of the entire canopy is the weighted summation of GPP using the respective LAI ratios for sunlit and shaded leaves:

$$A_n = A_{sunlit} LAI_{sunlit} + A_{shaded} LAI_{shaded}$$
(3)

where A_n is the net photosynthetic rate of the entire canopy (i.e., GPP); A_{sunlit} and A_{shaded} are the net photosynthetic rates of a unit light and shaded leaf, respectively; LAI_{sunlit} is the leaf area index of sunlit leaves, and LAI_{shaded} is the leaf area index of shaded leaves, respectively, which can be calculated as:

$$LAI_{sunlit} = \frac{1 - exp \left(-K_b(\theta_z) \times L \right)}{K_b(\theta_z)}$$
(4)

And

$$LAI_{shaded} = L - LAI_{sunlit} \tag{5}$$

where L is the total area of the canopy. The light extinction coefficient ($K_b(\theta_z)$) is a function of solar zenith angle and expressed as follow:

$$K_b(\theta_z) = \frac{\sqrt{x^2 + tan^2 \theta_z}}{x + 1.774(x + 1.182)^{-0.733}}$$
(6)

where θ_z is the solar zenith angle. We assumed x equals one with a spherical leaf angle distribution assumption.

Solar radiation is the ultimate energy source for photosynthesis. Thus, simulation of the radiation transport process within the canopy is the basis for GPP estimation. The total radiation absorbed by sunlit leaves comes from three components. The first part includes the direct radiation that is absorbed when it first reaches the sunlit canopies. The second part is the diffuse sky light collides with the leaves. The third part is the scattered light that is absorbed after the first hit. In contrast, the total radiation absorbed by the shaded leaves is only composed of the latter two scattered radiation mentioned above. The following formula can be used to estimate solar radiation hitting sunlit and shaded leaves:

$$I_{sunlit}(\theta_z) = K_b(\theta_z)I^0 + J_{dif} + I_{sca}$$
(7)

And

$$I_{shaded}(\theta_z) = J_{dif} + I_{sca} \tag{8}$$

where $K_b(\theta_z)$ is the extinction coefficient for beam light; I^0 is the direct solar light received at the top of the canopy, J_{dif} is the mean flux density of diffuse radiation, and I_{sca} is the average direct light downward scattered radiant flux density of the canopy.

The FGM model introduces the clumping index (Ω) to ameliorate the effect of overestimating GPP using the Beer's law. The FGM model also introduces VPD and a fitting parameter to express the relationship between internal leaf CO₂ concentration and atmospheric CO₂ concentration. The model framework of Song et al. (2009) is improved to allow direct calculation of photosynthetic rates in vegetation canopies when combined with the appealed two-leaf model and the radiation transfer model. In this study, we uesd the FGM process model to estimate the GPP of crops in the North China Plain by combining long time series of multi-source remote sensing data, meteorological data, and atmospheric CO₂ concentration monitoring data.

3. RESULTS

3.1 Changes in land cover and vegetation structure

Land use and land cover change (LULCC) could cause changes in LAI simultaneously. Thus, the impacts of land cover change on GPP are generally combined with LAI change effect on GPP. We refer to the total effect of LULCC and changes in LAI on GPP as the vegetation structure change effect, which is also considered as one of the essential drivers of the FGM model. In this study, we calculated land use transition matrix for the study area using MODIS LULC data in 2001 and 2016 (Figure 2a). The results showed that the primary vegetation type in the North China Plain was cropland from 2001 to 2016, but the cropland area in this study area decreased from 2001 to 2016. Of the area reduced by crops, 1.64% of croplands were converted to shrubs, 1.38% of croplands were converted to grassland, 0.65% of croplands were converted to urban land, and 0.11% of croplands were converted to forests.



Figure 2. (a) is Land cover transition matrix over the North China Plain (b) is interannual trends of LAI in the North China Plain and (c) is a spatial distribution of average LAI in the North China Plain from 2001 to 2016 (The redder part of the figure indicates a lower LAI value. The greener part indicates a higher LAI value).

LAI is a biophysical parameter indicating leaf abundance in canopies. In this study, we statistically analyzed the GLASS LAI products. The results showed that the annual average LAI increased by 0.0039 m⁻² yr⁻¹ (R²=0.20, p<0.05) during 2001 to 2016 (Figure 2b). The LAI of the North China Plain cropland area decreased continuously from 2007 to 2011. The spatial trend distribution of LAI was shown in Figure 2c. The LAI of the province in the southern part is higher than that in the north. For example, the Henan Province, Anhui Province, Jiangsu Province, and Shandong Province are greener than other provinces.

3.2 Spatial and temporal distribution of crop GPP in the North China Plain

In this study, we estimated the GPP in the North China Plain crop growing areas from 2001 to 2016 using the FGM process model. The annual total GPP showed a significant increasing trend from 2001-2016, with a mean increasing rate of \pm 2.19 Mt C yr⁻¹ (P<0.05). The mean annual total GPP for the cropland was 670.57 Mt C yr⁻¹ during 2001 to 2016 (Figure 3). The maximum value of crop GPP occurred in 2015 (703.06 Mt C yr⁻¹), and the minimum value occurred in 2003 (618.62 Mt C yr⁻¹). The total GPP of crops in the North China Plain from 2001 to 2007 showed an increasing trend, while the total GPP of crops from 2007 to 2011 showed a significant(P<0.05) decreasing trend.



Figure 3. The interannual variations in GPP from 2001 to 2016.

The spatial distribution of GPP followed the same spatial pattern of LAI with large spatial variability (Figure 4). The GPP for southern part of the North China Plain was larger than that of the northern province. Low LAI mainly occurred in Beijing, Tianjin, and the southern coastal area of Liaoning, with GPP mean values of $3.65 \text{ Mt C yr}^{-1}$, $5.59 \text{ Mt C yr}^{-1}$, and $8.19 \text{ Mt C yr}^{-1}$, respectively. High LAI mainly occurred in the south part of Henan Province, most of Shandong Province, and northern Anhui Province; and the corresponding mean annual total GPP was 154.78 Mt C yr⁻¹, 140.17 Mt C yr⁻¹, and 91.58 Mt C yr⁻¹, respectively.

ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-3/W2-2022 Urban Geoinformatics 2022, 1–4 November 2022, Beijing, China



Figure 4. The spatial distribution of cropland GPP in the North China Plain.

3.3 Crop yield estimation in the North China Plain

3.3.1 Empirical 'GPP-Yield' regression model

The results showed that changes in annual total GPP could explain 99.2% of the variation in grain production (RMSE = $10.4 \text{ Mt C yr}^{-1}$) (Figure 5). GPP and crop yield showed a strong positive linear regression relationship, which means that crop yield increases with GPP since the original source of grain production is GPP.



Figure 5. Regression plot of GPP and Grain yield.

3.3.2 Spatial and temporal variation of crop yield in the North China Plain

Finally, we calculated the crop yields from FGM GPP using the "GPP-yield" model (Figure 5). The results showed that the crop yields in the North China Plain fluctuate in accordance with GPP, and also showed a significant increasing trend from 2001 to 2016, with a mean increasing rate of +2.84 Mt yr⁻¹ (Figure 6a). The maximum value of crop yield occurred in 2015 (913.07 Mt), and the minimum value occurred in 2003 (803.4 Mt). High crop yields occurred in Henan Province (>2000 Mt yr⁻¹),



Figure 6. (a) is inter-annual variation trend of crop yield in the North China Plain from 2001 to 2016. (b) shows the spatial distribution of mean annual total crop yields in the North China Plain from 2001 to 2016.

Shandong Province (1000 to 2000 Mt yr⁻¹), and northern

Jiangsu Province (1000 to 2000 Mt yr⁻¹). The areas with lower crop yields were mainly located in Beijing (<100 Mt yr⁻¹), Tianjin (<100 Mt yr⁻¹), and some coastal areas of Liaoning (<100 Mt yr⁻¹) (Figure 6b), due to rapidly urbanizing progress in the last twenty years and a relatively small share of arable land. The spatial and temporal pattern of crop yield agreed with that of GPP estimated by the FGM model.

4. DISCUSSION AND CONCLUSIONS

By applying a process-based GPP model, multi-source remote sensing data, and the grain production data statistical yearbook, this study achieves the goal of estimating the spatial and temporal dynamics of the crop productivity and crop yield in the North China Plain from 2001 to 2016, with a spatial resolution of 1 km and temporal resolution of 1 day.

The FGM model effectively estimated crop GPP in the North China Plain. Crop GPP in the North China Plain shows a fluctuating increasing trend from 2001 to 2016 (+2.19 Mt C yr⁻¹, P<0.05), and the total annual average GPP is about 670.75 \pm 24.98 Mt C yr⁻¹. The GPP of the North China Plain shows an overall increasing trend from the north to the south, with large spatial variability. The crop productivity is low in Beijing-Tianjin and coastal areas, while high in the southern part of Henan Province, Shandong Province, and Jiangsu Province.

The correlation between crop productivity and crop yield in the North China Plain is robust, with a determination coefficient of $R^2 = 0.98$ (RMSE = 10.4 Mt yr⁻¹). Therefore, GPP can be used to estimate the crop yield effectively and to analyze its spatial and temporal distribution scientifically. The spatio-temporal distribution pattern of crop yield and GPP in the North China Plain is highly consistent. Meanwhile, the crop yield in the North China Plain also showed a significant increasing trend during 2001 to 2016 (+2.84 Mt yr⁻¹ P<0.05), with a mean annual total crop yield of about 870.87 \pm 32.44 Mt yr⁻¹. Crop yield is low in Beijing and Tianjin, mainly because of the rapid economic development and high urbanization level, with limited portion of agricultural land. The southern regions had relative high yields due to larger proportions of agricultural lands, with slower urbanization levels. In addition, the southern regions have favorable climatic and soil conditions for crop growth.

It is interesting to note that the productivity and yield of crops in North China have been rising during this study period, but the crop area appears to be decreasing. This may be due to the advancement of farming technologies that have led to the rise in food production, for example, irrigation and long-term fertilization (Zhang et al, 2021). In addition, climate change and the fertilization effect of CO_2 may also lead to an increase in crop productivity (Zhang et al, 2021). This needs to be improved in further research.

This research can provide research data for large-scale crop production estimation and help us better understand how crop yield changes under climate change.

ACKNOWLEDGEMENTS

The research is supported by the National Key Research and Development Program of China (Grant 2021YFE0117500).

REFERENCES

Zhang, Y.L., Song, C.H., Lawrence E. Band., Sun, G., 2021. No Proportional Increase of Terrestrial Gross Carbon Sequestration From the Greening Earth. *Journal of Geophysical Research: Biogeosciences*, 124(8), 2540-2553.

Chen, S.,Y., 2021. Spatio-temporal Dynamic Simulation and Driving Factor Analysis of Vegetation Productivity in China using a Process-based Model. *Doctor Thesis, Beijing Normal University.*

Prasad, A. K., Chai, L., Singh, R.P., Kafatos, M., 2005. Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observations & Geoinformation*, 8(1), 26-33.

Zhao, J., Liu, L., Xu, Z., Jiao, Q., Liu, S., 2011. Monitoring winter wheat gpp in huabei plain using remote sensing and flux tower. *Transactions of the Chinese Society of Agricultural Engineering*, 27(13), 346-351.

WANG, Z.Y., 2020. Remote Sensing Estimation of Winter Wheat Yield in the North China Plain Based on Phenological Adjustment. *Master Thesis, Henan University.*

Tao, F., Yokozawa, M., Zhao, Z., Xu, Y., Hayashi, Y., 2004. Remote sensing of crop production in china by production efficiency models: models comparisons, estimates and uncertainties. *Ecological Modelling*, 183(4), *385-396*.

HUANG, Y.Y, 2020. Study on the spatial and temporal distribution of vegetation net out-grade productivity and its drivers in the the North China Plain from 2000 to 2018. *Master Thesis, Northwest Normal University.*

Wang, P., Xie, D., Zhang, J., Rui, S., Chen, S., Zhu, Q., 2009. Application of BEPS model in estimating winter wheat yield in the North China Plain Plain. *Transactions of the Chinese Society of Agricultural Engineering*, 25(10), 148-153.

Ji, R., Tan, K., Wang, X., Pan, C., Xin, L., 2021. Spatiotemporal Monitoring of a Grassland Ecosystem and Its Net Primary Production Using Google Earth Engine: A Case Study of Inner Mongolia from 2000 to 2020. *Remote Sensing*, 13(21), 4480.

Song, C., Katul, G., Oren, R., Band, L.E., Tague, C.L., Stoy, P. C., 2009. Energy, water, and carbon fluxes in a loblolly pine stand: results from uniform and gappy canopy models with comparisons to eddy flux data. *Journal of Geophysical Research Biogeosciences*, 114(G4), 950-950.

Medlyn, BE., Duursma, RA., Eamus, Ellsworth., 2011. Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biol*, 17(6)(-), 2134-2144.

Chen, J.M., Liu, J., Cihlar, J., Goulden, M.L., 1999. Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. *Ecological Modelling*, 124(2-3), 99-119.

Pury, D., Farquhar, G.D., 2010. Simple scaling of photosynthesis from leaves to canopies without the errors of big -leaf models. *Plant Cell & Environment*, 20(5), 537-557.

Sun, H.Y., Liu, C.M., Zhang, X.Y., Shen, Y.J., Zhang, Y.Q., 2006. Effects of irrigation on water balance, yield and wue of winter wheat in the north China plain. *Agricultural Water Management*, 85(1-2), 211-218.

Men, M.X., Li, X.W., Xu, H., 2008. Effects of long-term fertilization on yield and stability of fluvo-aquic soil crops in the North China Plain. *Scientia Agricultura Sinica*.