DATA ASSIMILATION OF DSSAT MODEL WITH REMOTE SENSING FOR YIELD ESTIMATION IN RAINFED RICE FIELD AREA

Chudech Losiri[1], Sarawut Ninsawat[1] and Kiyoshi Honda[2]

[1] Remote Sensing and Geographic Information Systems FoS,
School of Engineering and Technology, Asian Institute of Technology,
P.O. Box: 4, Klong Luang, Pathumthani 12120, Thailand;
Telephone: +66- 2524-5577, Fax: +66-2524-5597
Email: Chudech.Losiri@ait.ac.th, sarawutn@ait.ac.th
[2] Int'l Digital Earth Applied Science Research Center (IDEAS),
Chubu Institute for Advanced Studies, Chubu University,
1200 Matsumoto-cho, Kasugai, Aichi, 487-8501, Japan; Telephone: +81-568-51-9787
E-mail: hondak@isc.chubu.ac.jp

ABSTRACT

The Crop Simulation Model (CSM) is a curtail tool to identify, estimate, acquire, and predict the agricultural production. It can provide the information to decision makers for agricultural activities and prediction of crop yield under environmental data; climate, fertilizer, cultivars and management, which relate to the farm. Using the CSM is always limited by the uncertainties of input parameters which can affect to the yield estimation. To overcome relevant problem, this study tries to develop a new methodology to improve the performance of the CSM through data assimilation with remote sensing data. The Decision Support System for Agrotechnology Transfer (DSSAT) is selected as the CSM. Genetic Algorithm (GA) is used as an optimizer to obtain the best set of parameters for the DSSAT. Data assimilation, in this approach, is used to calibrate the DSSAT model based on GA using remotely sensed imagery such as MODIS data. Comparison between estimated result from the model and observation data at rainfed rice field in the test site of Trakan phutphon district, Ubon ratchathani province was performed to evaluate the model performance. The results show evidently that the estimated yield can be retrieved with moderately satisfactory accuracy from the simulated output of the DSSAT model.

1. INTRODUCTION

The Crop Simulation Model (CSM) plays an important tool to identify, estimate, acquire, and predict growth and development of the plants. There are lots of crop simulation models that can be used for studying agricultural practices especially rice plantation. Decision Support System for Agrotechnology Transfer (DSSAT) is one of CSMs that facilitates an application of mathematical crop growth models in research, teaching, extension, outreach, and policy decision making (Jones et al., 2003). Huge parameters are acquired for applying the CSM. The *in-situ* sampling and laboratory testing are reasonable method to get all crop parameters at field scale; however, they are always expensive, laborious and time consuming to get all data at regional scale. Therefore, developing data assimilation technique can be a key challenge to determine a set of unknown parameters by using remote sensing data.

The Data assimilation is an advance technique that incorporates between the available data and model output to obtain a better estimation of parameters. There are several sources of available data such as satellite imagery data, derived product, management data and land surface model output. The main advantage of data assimilation technique is to reduce the impact of measurement data or observation data when it is unclear, unavailable or insufficient. SWAP-GA is an example of its application. This model uses the performance of Genetic Algorithm (GA) to determine the unknown parameters for the SWAP model. The GA is an optimizer which mimics a natural selection in order to provide a robust search in the complex solution (Goldberg, 1989).

The Satellite remote sensing supplies observations over large areas at even times and therefore provides other crop monitoring techniques with possible spatial extension. Moreover, it also gives help to extend crop models to a regional scale without ground field data (Moulint, Boudeau and Delécolle, 1998). In order to improve the model's performance and to find optimum crop parameters, many studies have tried to combine crop simulation model with remote sensing through data assimilation technique (Moulin, Bondeau and Delécolle, 1998; Fang et al., 2008 and Vazifedoust *et al.*, 2009)

The LAI is an important vegetation biophysical parameter which describes a ratio of leaf area to per unit ground surface area. LAI can provide an understanding of dynamic changes in productivity and climate impacts on agricultural. Furthermore, LAI can be applied as an indicator of stress in rice field, thus, it can be used to explain relationships between environmental stress factors and insect damage (Zheng and Moskal, 2009). Therefore, this study attempts to utilize a data assimilation technique for improving the capability of DSSAT CERES Rice model in case of yield estimation at rainfed rice field in the test site of Trakan phutphon, Ubon ratchathani (15^o 42' 42''N, 105^o 00' 21''E). The cropping cycle in this area starts with the growing of the seedlings around April to May. Transplanting of rice seedlings into the field occurs around one month later, normally in late June. Rice is usually harvested around October to November.

2. DATA ASSIMULATION TECHNIQUE

The Data assimilation is a mathematical approach that uses of all accessible information within a given specific time window to estimate various unknown variables (Liang and Qin, 2008). The general methodology of the data assimilation is illustrated in Figure 1. The detail of each part is explained in next section.

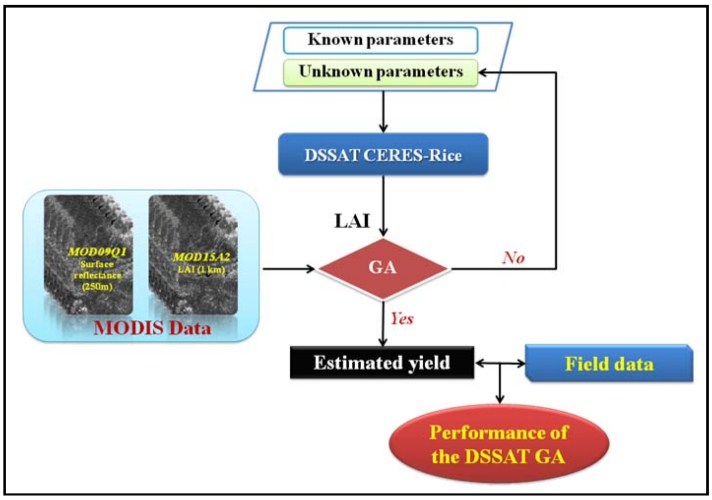


Figure 1. The general methodology of the data assimilation

2.1 DSSAT CERES-Rice

The CERES-Rice model under DSSAT is developed from CERES-Maize and CERES-Wheat models. It was develop by Ritchie *et al*. (1986) and modified for transplanted rice by researchers at the International Fertilizer Development Center (IFDC) (Godwin and Singh, 1989). It is still undergoing, being tested and being refined refining by the IBSNAT project scientists. The various processes are simulated by this model such as phenological development of the drop; growth of leaves, stems and roots; biomass accumulation and partitioning among leaves, stem, panicle, grains and roots; soil water balance and water use by the crop; and soil nitrogen transformations uptake by the crop. The phenological stages including sowing or transplanting, germination, emergence, juvenile phase, panicle imitation, heading, beginning of grain filling, end of grain filling, and physiological maturity are simulated by the model are. The model simulates the total biomass of the crop as the product of level involved the prediction of these two important processes. The yield of the crop will be the fraction of total biomass partitioned to grain.

2.2 DSSAT-GA

Data assimilation, nowadays, becomes the extensively approach that several studies have adapted for improving the ability of the model performance to obtain the better estimation of the crop model. The SWAP-GA-RS (Ines, 2002 and Charoenhirunyingyos, 2009) is an example of the data assimilation that took an advantage to improve an estimation of crop production and find the best set of parameters for the SWAP model. The DSSAT-GA, in this study, is a data assimilation technique that adopted Genetic Algorithm (GA) to determine the best set of DSSAT parameters. GA mimics the process of natural selection and evolutionary genetics. The detailed information of GA can be found in Goldberg (1989). The general process of the DSSAT-GA in this study is to find the input parameters in order to simulate the DSSAT model with the corresponding MODIS LAI data. The process runs until reach the maximum generation and the set of optimum parameters is obtained. The advantage of using GA is that it avoids the initial guess selection problem and provides a systematic scanning of the whole reasonable solution as a global optimum solution can be achieve.

2.2.1 *Required input parameters preparation*

To estimate the LAI using CSM is always strongly influence by several data such as soil, weather, genetic coefficient of rice, and management data. The detail of each parameter is explained below.

Weather data: This data are required and have to collect in daily through a simulation period. The weather data from the tower at the field consist of minimum and maximum temperatures, humidity, solar radiation, wind speed, and rainfall.

Soil data: The basic information of physical and chemical properties of the soil can also be obtained from the field and the soil map which provides by Land Development Department (LDD). The major soil data used for rice cultivar in this study includes color, slope, runoff potential, drainage type along with layered classification of soil texture, pH, phosphorous, potassium, carbon, nitrogen and cation exchange capacity.

Management data: consisting of seeding date, transplanting date, harvest area and date, plant population, treatment and irrigation data.

Genetic coefficient: Jusmine Rice (KDML 105) is a famous rice variety in Thailand which was used to estimate the LAI in this simulation. The cultivar specific parameters were taken from the rice genetic coefficient which already proposed by Dr. Attachai Jintrawat from Chiangmai university.

2.2.2 *MODIS LAI preparation*

Time series of satellite data which relates to crop development can affect the assimilation performance (Fang *et al*., 2011). Eight day surface reflectance and LAI data with spatial resolution at 250 m (from MOD09Q1) and 1 km (from MOD15A2) respectively of MODIS data were prepared as shown in Figure 2.

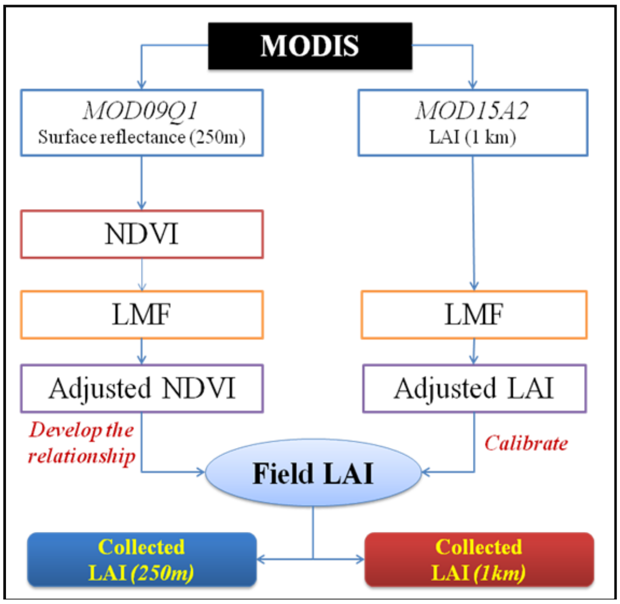


Figure 2. Remote sensing data preparation

MOD09Q1 is 8-day MODIS band 1 and 2 surface reflectance at with 250 m resolution. This time series data was calculated NDVI based on the basic function between near infrared and red bands. Then the NDVI time series data had to adopt Local Maximum Fitting (LMF) to remove the uncertainty from the cloud and atmosphere. After that the development between NDVI and LAI from the field was carried on to convert the NDVI to LAI data.

MOD15A2 is 8-day MODIS LAI data with 1 km resolution. This product describe canopy structure and relates to functional process relate to energy and mass exchange. MODIS LAI also was removed the effect of cloud by applying LMF. After that the adjusted LAI was calibrated with the field LAI to obtain the best quality of LAI at 1 km.

LAI from different spatial resolutions were used as a key variable in the data assimilation scheme to explore the best set of DSSAT parameters such as planting date, harvesting date, planting population, row space and soil hydraulic properties. The assimilation process was explained in detail as following section.

2.3 Model integration

The combination of DSSAT and GA is used for identifying unknown crop parameters. The GA tries to minimize the different value between simulated and the satellite LAI and re-initials input parameters for running in the DSSAT again. The processes of optimization are; GA randomly creates a set of chromosome containing all unknown crop parameters which are some inputs for DSSAT CERES-Rice. Then, LAI outputs are compared with the satellite LAI based on the fitness function is shown as equation 1.

$$\text{Fitness function} = \frac{1}{\frac{1}{N}\sum_{t=1}^{N}\left|LAI_{SAT} - LAI_{DSSAT}\right|} \quad (1)$$

where LAI SAT is LAI estimated from MODIS; LAI DSSAT is simulated LAI from DSSAT CERES-Rice Model; and N is the number of LAI data

When the fitness value shows the highest value or simulated LAI from DSSAT CERES-Rice is matched or nearly matched with LAI from remote sensing data, the processing is kept that value as a best fitness. In contrast, if the LAI value between satellite image and simulated LAI is not good or not much more different, GA proposes the population as a parent for creating the new generation by draping genetic operation (selection, crossover and mutation) to simulate and compare the result again. The evaluation process is stopped, when the generation is a maximum.

2.4 Model evaluation

This step takes advantages of the DSSAT-GA and MODIS LAI to find the best set of input parameters for the DSSAT. After collecting all input parameters, they were put into the DSSAT and estimated the yield and LAI. The estimated yield and the LAI obtained from the different resolutions of MODIS data were compared with the field data. The evaluation of the model performance to estimate yield was done.

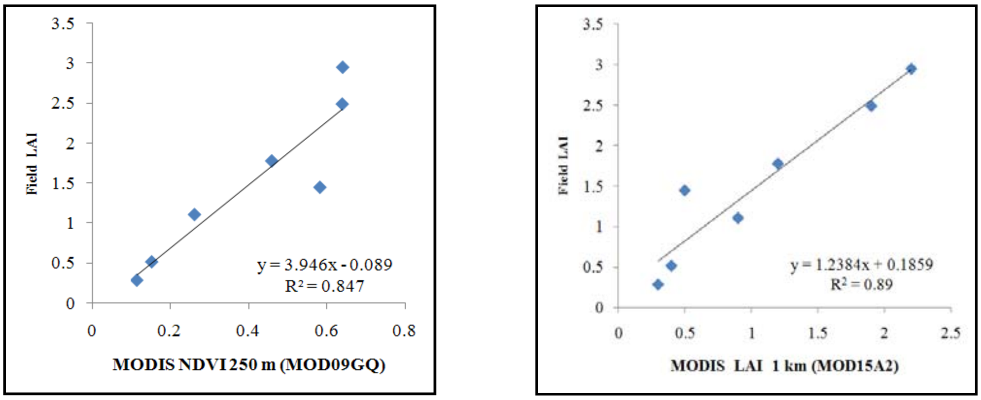
3. RESULT AND DISCUSSION

3.1 Retrieving LAI from MODIS

A time series of NDVI calculated from MOD09GQ was compared with LAI from the field data to develop the relationship for converting NDVI to LAI. From the Figure 3 (a), It can be illustrated that the correlation between MODIS NDVI (250 m) and field measurement showed strong coefficient of determination with R^2 = 0.847. The equation to convert LAI from the MODIS NDVI is defined as equation 2.

$$\textit{MODIS LAI (MOD09GQ)} = 3.946 \times \textit{MODIS NDVI (MOD09GQ)} - 0.089 \quad (2)$$

After applying the equation 2, the LAI from the MOD09GQ with 250 m was calculated. The time series of the LAI value was used as a reference data in the assimilation process (Figure 4).



(a) (b)

Figure 3. (a) Correlation between MODIS NDVI 250 m (MOD09GQ) and field measurement (b) Correlation between MODIS LAI 1 km (MOD15A2) and field measurement

In terms of the MODIS LAI from MOD15A2, this time series data (purple line) shows slightly lower value when comparing with the field data; however, they had the similar trend. Therefore, the calibration of the MODIS LAI (red line) from MOD15A2 was explored to remove the mismatch between both data sets. It can see from the Figure 3 (b) that there is a strong coefficient between MODIS LAI and field measurement with R^2 = 0.89. The formula for converting the LAI from MODIS 1 km to field data is showed as equation 3.

$$\textit{Calibrated MODIS LAI (MOD15A2)} = 1.2384 \times \textit{MODIS LAI (MOD15A2)} - 0.1859 \quad (3)$$

After applying the formula, the LAI from the MOD15As is nearly matched with LAI from the field (Figure 4) and this time series data was also used as a reference in the assimilation process.

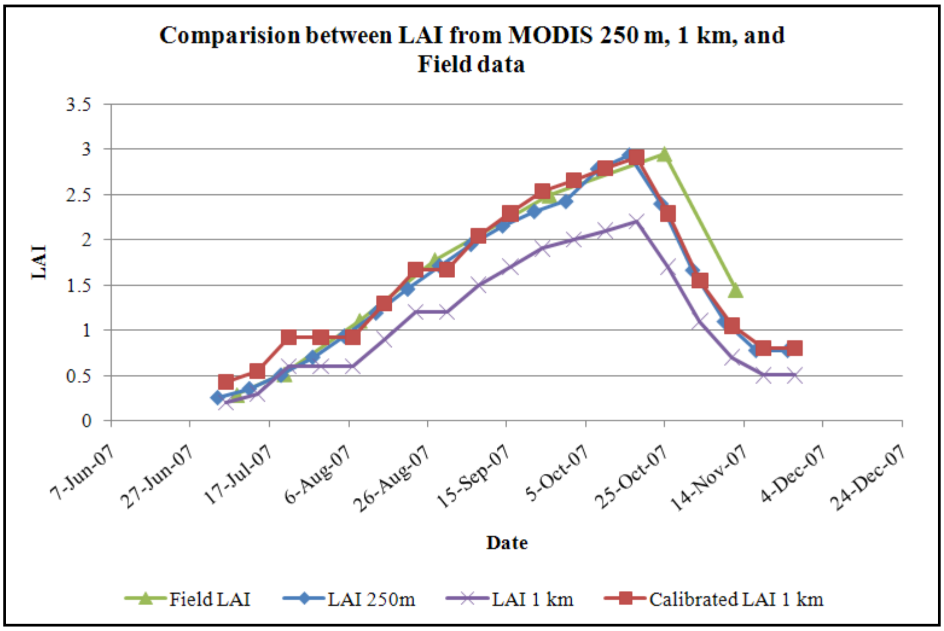


Figure 4. Comparison between LAI from MODIS 250 m, 1 km, and field measurement

3.2 Rice yield estimation

The DSSAT-GA coupled with MODIS LAI from different resolutions was explored unknown parameters based on the fitness function. The set of unknown parameters to be used for yield simulation could obtain from the set that has the highest fitness. After retrieving the unknown parameters from this assimilation scheme, all of these unknown parameters and others, which were explained in section 2.2.1, were taken into the DSSAT model to estimate the rice yield and LAI.

Table 1. Comparison between yield estimations from MODIS and field measurement

Field Data (kg/rai)	Estimated Yield (kg/rai)	
	MODIS 250 m	MODIS 1 km
350	322	312

The simulation of the crop yield was made for the rainfed rice field in the study site. According to Table 1, the actual average yield was 350 kg/rai. It was note that the simulated rice yield from the DSSAT model showed some sensitivities to the noted parameters obtained from the time series of MODIS LAI data. The estimated yield using unknown parameters from the MODIS LAI 250 m indicated the better estimation when comparing with the MODIS LAI 1 km. However, there was a quite difference in the simulated yield from MODIS LAI 250 m and 1 km with 8 % and 10 % respectively.

In addition to the rice yield estimation, the DSSAT model also simulated LAI itself. However, the simulated LAI from both time series of MODIS data had a great dissimilarity when comparing with the LAI data from the field. This situation could occur due to an inappropriate of the genetic coefficients of rice in which several studies have mention that these parameters are importance and have a high effect to growth and development processes particularly in daily LAI simulation of the DSSAT model (Pabico, Hoogenboom, and McClendo, 1999).

4. CONCLUSION

The CSM simulates crop growth status and predicts crop yield and its associated uncertainties at harvest maturity. Remote sensing is a powerful tool for estimating crop biophysical parameters. This study used the GA, which adopts the natural selection, as an optimizer to determine the unknown parameter for the DSSAT model. Based on the concept of using data assimilation technique using DSSAT-RS-GA, this study was developed to improve the rice yield estimation in the rainfed area of Thailand. Deriving LAI from the different resolutions of MODIS data, MOD09Q1 (250 m) and MOD15A2 (1 km), was compared with the field measurement, the result showed the reasonable relationship. Therefore, these MODIS time series data were used as a reference in the assimilation scheme. The estimated rice yield and LAI from simulating the DSSAT model using unknown parameter was completed. However, it had a quite difference in the rice yield and LAI estimation when comparing with the field data. This research proposes that the future study will apply GA for other necessary parameters such as genetic coefficients of rice in specific cultivar. The genetic coefficients are importantly required parameters in the DSSAT model. They have a strongly effect to the growth and development processes in the simulation which also affect to yield and LAI estimation. The LAI is a crucial key variable to determine the phonological process of plant. It can be simulated to reflect the growth and development of the rice in each state by the model using the specific genetic coefficients.

5. REFERENCES

Charoenhirunyingyous, S., 2009. Estimation of soil hydraulic properties through data assimilation in the agro-hydrological model and its application to impact assessment of dry spell on rice yields. *Doctoral Dissertation.* Asian Institute of Technology, Pathum Thani, Thailand.

Fang, H., Liang, S., Hoogenboom, G., Teasdale, J. and Cavigelli, M., 2008. Corn-yield estimation through assimilation of remotely sensed data into the CMS-CERES-Maize model. *International Journal of Remote Sensing*, 29(10), 3011-3032.

Godwin, D.C. and Singh, U., 1989. Nitrogen dynamics in IBSNAT crop models. In *Agronomy Abstracts*. American Society of Agronomy, Madison, Wisconsin.

Goldberg, D.E., 1989. *Genetic Algorithms in search and optimization and machine learning*. Reading, Mass.: Addison-Wesley Pub.

Ines, A.V.M., 2002. Improve Crop Production Integrating GIS and Genetic Algorithm. *Doctoral Dissertation.* Asian Institute of Technology, Pathum Thani, Thailand.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J. and Ritchie, J.T., 2003. The DSSAT cropping system model. *European Journal of Agronomy,* 18, 235-265.

Liang, s. and Qin, J., 2008. Data Assimilation Methods for Land Surface Variable Estimation. In Liang, S. (Ed.), *Advances in Land Remote Sensing System, Modeling, Inversion and Application* (313-340). Springer Science.

Moulin, S., Boudeau, A., and Delécolle, R., 1998. Combining agricultural crop models and satellite observations: From field to regional scales. *International Journal of Remote Sensing*, 19(6), pp. 1021-1036.

Pabico, J.P., Hoogenboom, G., and McClendo, R.W. 1999. Determination of cultivar coefficients of crop model using genetic algorithm: a conceptual framework. *Transaction of the ASAE*, 42(1), 223-232.

Ritchie, J.T., Alocilja, B.C., Singh, V. amd Vehara, G., 1986. IBSNAT/CERES Rice Model. Agrotechnology Transfer, *Newsletter of the International Benchmark Site Network for Agrotechnology Transfer (IBSNAT) Project and The Soil Management Support Services (SMSS)*, 3, 1-5.

Vasifedoust, M., Van Dam, J.C., Bastiaanessen, W.G.M. and Feddes, R.A., 2009. Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. *International Journal of Remote Sensing*, 30(9-10), 2523-2545.

Zheng,G. and Moskal, L. M., 2009. Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors*, 9(4), 2719-2745.