

Assimilating MODIS-LAI into crop growth model with EnKF to predict regional crop yield

Sijie Wu¹, Jiangxi Huang^{2*}, Xingquan Liu¹, Jinlong Fan³, Guannan Ma², Jinqiu Zou⁴

¹ School of Geosciences and Info-Physics, Central South University, 410083, Changsha, China

² College of Information and Electrical Engineering, China Agricultural University, 100083, Beijing, China

³ National Satellite Meteorological Center, China Meteorological Administration, 100083, Beijing, China

⁴ Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, 100081, Beijing, China

Abstract: Regional crop yield prediction is a vital component of national food security assessment. Data assimilation method which combines crop growth model and remotely sensed data has been proven the most potential method in regional crop production estimation. This paper takes Hengshui district as study area, WOFOST as crop model, MODIS-LAI as observation data to test and verify the efficiency of EnKF assimilation method. The results show that the precision of crop yield estimation are obviously improved with EnKF assimilation, in the WOFOST potential level the R^2 improved from 0.10 to 0.38 and RMSE was reduced from 2480 kg/ha to 880kg/ha. Our study indicates that EnKF assimilation method has great potential in regional crop production forecasting. **Keywords:** Multi-temporal; forest cover; decision tree.

Keywords: Remote sensing, WOFOST, EnKF, Data Assimilation

1 Introduction

Regional crop yield estimation plays an important role in the agriculture monitoring and food security. Traditional crop yield estimating methods such as investigation statistics, agro-meteorological model forecasting etc. These methods are hard to accurately predict the regional crop production due to costly and time-consuming. The satellite remote-sensing is characterized as fast, macroscopic, comprehensive, objective, dynamic, but restricted to spatio-temporal resolution, most of remotely sensed data only reflect the soil surface and crop canopy radiation information, cannot reveal the inheritance mechanism of crop growth and yield formation. Crop growth models based on the crop photosynthesis, transpiration, respiration, nutrition are successfully applicable for yield forecasting in simple point scale. Because the model input parameters are high variable in the region scale, the simulated results are with high uncertainty. As a method that can integrate the crop growth model and remotely

sensed data, data assimilation provides an approach to estimate the crop yield with high-accuracy and becomes the most promising method to predict the regional crop yield.

There are two category data assimilation algorithms: global assimilation such as variation algorithm and sequential assimilation like Kalman Filter. Evensen [2] firstly applied the EnKF data assimilation method in marine area. Vazifiedoust [16] assimilated LAI and evapotranspiration (ET) retrieval from the remote sensing data into the SWAP model to explore whether data assimilation of LAI and ET can be used to forecast total wheat production as an indicator of agricultural drought, the results showed that predictions for in advance 1 month with assimilation at a regional scale were very promising with respect to the statistical data. However, longer-term predictions i.e. in advance 2 months, resulted in a higher bias between the simulated and statistical data. De Wit [1] used an EnKF to assimilate coarse resolution satellite microwave sensor derived soil moisture estimates (SWI) for correcting errors in the water balance of WOFOST model, the results showed that assimilation of SWI has clearly improved the relationship with crop yield statistics for winter wheat for the majority of regions where a relationship could be established.

In the study, EnKF was used to assimilate MODIS-LAI and WOFOST model to forecast winter wheat yield and assess the accuracy and applicability of the method. Since the MODIS - LAI data of study area are lower than the actual values, we corrected the MODIS-LAI by combining the field measured LAI data of winter wheat and the MODIS-LAI trend data. Time-series LAI was assimilated through combined corrected MODIS-LAI and WOFOST simulated LAI. The assimilated optimal LAI was used to drive the WOFOST model in each pixel to obtain the regional yield.

2 Material and experiments

2.1 Study area

We chose Hengshui of Hebei province as the study area, since Hengshui is one of dominant winter wheat planting regions (Fig1). Hengshui is located in southeast Hebei and between $115^{\circ} 10' - 116^{\circ} 34'$ E longitude and $37^{\circ} 03' - 38^{\circ} 23'$ N latitude. Annual mean temperature in Hengshui is 9.0-15.3C and annual rainfall is 356.1-707.7mm. The winter wheat sow in the middle of October, seeding emerge 8-10 days later, grow into over-wintering stage in October, return green in late February, joint in late march, grow into anthesis stage in April, head out and fill in May every year. Winter wheat of Hengshui is usually harvested in early June. Because Hengshui is smaller, the growths of winter wheat of different areas are almost at the same time although growth stage in south is slightly earlier than in north.

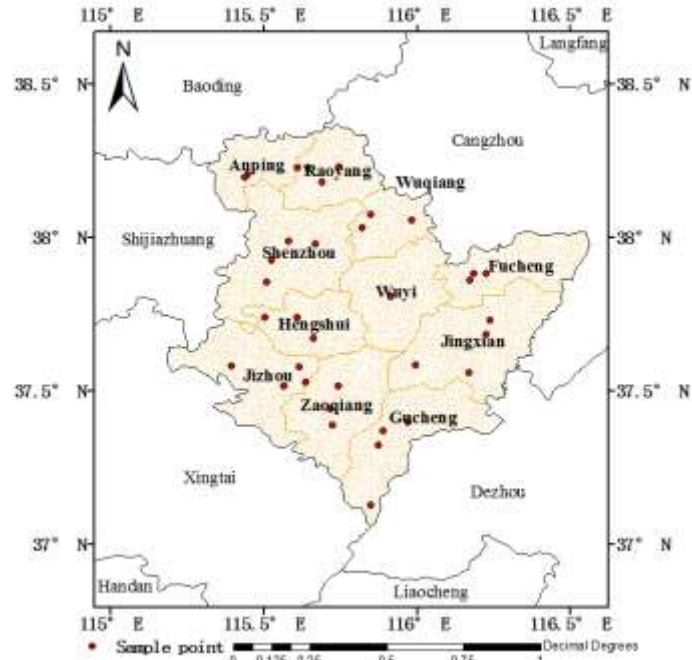


Figure1 Study Area(red points represent the filed measured LAI sites)

2.2 Crop Model

We used the world food studies (WOFOST) crop model of Wageningen University. WOFOST is a mechanistic crop growth model that describes plant growth with light interception, takes CO_2 assimilation as growth driving processes and takes crop phenological development as growth controlling [1]. WOFOST model can be used in three different ways: a potential mode that crop growth is purely driven by temperature and solar radiation and no limiting factors are taken into count; a water-limited mode that is characterized by the aforementioned factors as well as water availability derived from root characteristics, soil physical properties, rainfall and evapotranspiration; a nutrient-limited mode that nutrient availability depends on the supply of nutrients to the plant roots and is estimated from soil characteristics. The potential mode was used in this study.

To drive the WOFOST model, we collected regional meteorological, soil, crop data. Meteorological data include daily maximum and minimum temperature, radiation, rainfall, wind speed, water pressure between January 2007 and December 2008 in the study area were collected. These meteorological data were interpolated by distance inverse weight method. Regional crop data (including Temperature sum from sowing to emergence, temperature sum from emergence to anthesis, temperature sum from anthesis to maturity, sowing date) and soil data(including field moisture capacity, wilting point, saturated water of soil) were also got with the same method. In the study 1:1000000 China Soil Database were also used.

2.3 Remotely sensed data

In this study, we chose MODIS 250m 8-day LAI data which were downloaded from the Earth Observation System data (EOS) gateway (<http://eos.nasa.gov>) as observation data. The original data were converted to Albers Equal-area projections at 1 km resolution and with a size of 136 * 162 pixels by MRT tool. The pixels without the planted winter wheat were excluded by overlaying the land use data extracted from the 30m TM data and MODIS-LAI data in winter wheat growth stage of from October 2007 to June 2008.

Due to the influence of clouds, vapor and aerosol, MODIS-LAI data are contaminated and needed to eliminate the abnormal data. In this study we used the Savitzky-Golay(S-G) filter method to smooth the time series of MODIS-LAI, restored the spatial-temporal continuum LAI data of fertility period of winter wheat. Fig shows LAI profile with S-G Filter.

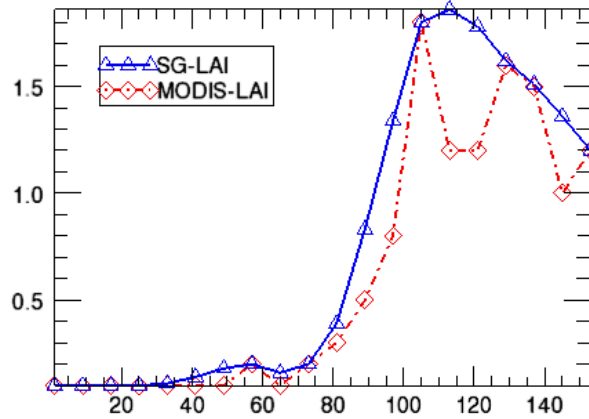


Fig2 (a) LAI profile with SG-Filter(Pixel[25,32])

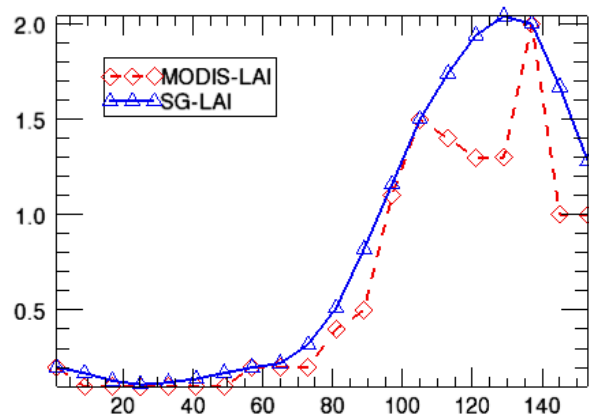


Fig2 (b) LAI profile with SG-Filter(Pixel[80,105])

2.4 Assimilating MODIS-LAI into WOFOST

Ensemble Kalman Filter

The Ensemble Kalman Filter is an optional recursive data assimilation method. The basic principle of EnKF is to construct a Monte Carlo ensemble such that the mean of the ensemble is the best estimate, and the ensemble error covariance is the estimate of the forecast covariance[3]. At the current assimilation time t , It assume that we have an ensemble of forecast $A_t^f=(a_{1,t}^f, a_{2,t}^f, \dots, a_{N,t}^f) \in \mathbb{R}(n \times N)$ (n : dimension, N :ensemble size) and an ensemble of observation $B_t=(b_{1,t}, b_{2,t}, \dots, b_{N,t}) \in \mathbb{R}(n \times N)$. $A_t^a=(a_{1,t}^a, a_{2,t}^a, \dots, a_{N,t}^a) \in \mathbb{R}(n \times N)$ is the optional estimate Ensemble, the mean of A_t^a is the best estimate of time t . The relationship between A_t^f , B_t and A_t^a is shows in following equations:

$$B_t = HA_t + vt \quad (1)$$

$$A_t^f = MA_{t-1}^a + wt \quad (2)$$

$$A_t^a = A_t^f + Kt(B - HA_t^f) \quad (3)$$

where H is the observation operator that model states to the observations, M is the equation which transform state form time t to time $t+1$, vt is the noise of observation and wt the error which is produced when state transforms. Kt is the Kalman gain matrix; it is computed with following formulas:

$$K_t = P_t^f H^T (HP_t^f H^T + R_t)^{-1} \quad (4)$$

where P_t^f is the variance of forecast ensemble, R_t is the variance of observation ensemble.

EnKF Assimilation

In our study, LAI is the state variable of assimilation, observation data is corrected MODIS-LAI, $H=1(B_t=A_t+vt)$, M is defined as the LAI part of WOFOST model

$$\text{Exponential growth stage: } LAI_t = LAI_{t-1} + Lexp, t * \Delta T \quad (5)$$

$$\text{Source limited growth stage: } LAI_t = LAI_{t-1} + Lsc, t * \Delta T \quad (6)$$

where $Lexp, t$: LAI growth rate at time t during exponential growth stage[ha ha⁻¹d⁻¹]

Lsc, t : LAI growth rate at time t during source limited growth stage

LAI_t : LAI value at time t [ha ha⁻¹]

In this study, the EnKF assimilation was conducted for each pixel in the study area independently. WOFOST began running at sowing date. At first day of 2008 (DOY=1), a white Gaussian noise was added to shift the simulated LAI and generate the first ensemble of forecasted $MLAI\{mLAI_1, mLAI_2, \dots, mLAI_N\}$ (mean value is the simulated LAI). If there is observation data at time t , we add a Gaussian perturbation-ensemble to it and generate an observation ensemble $OLAI\{OLAI_1, OLAI_2, \dots, OLAI_N\}$, then assimilate the forecast ensemble and observation ensemble with EnKF, obtain the optimal estimate ensemble ($MOLAI\{LAI_1, LAI_2, \dots, LAI_N\}$) which will be put into WOFOST model to get the forecast ensemble at time $t+1$. If there isn't observation LAI, the forecast ensemble is put into the model directly. Repeat the process until

crop mature. The average of $MOLAI$ is the best estimate at time t .

Observation ensemble perturbation: after correction, the relations between MODIS-LAI data its size and actual values has been stable, in this research, the error of observation LAI is seted between 10% and 20%. EnKF Ensemble size: In EnKF assimilation, the ensemble size has great influence on the result and efficiency. In this study we test different size (10, 30, 50, 90) respectively and analyzed the relationship between the ensemble size and the assimilation results. EnKF Assimilation step: The ensemble size has great influence on the result. We set different size (1,4,8,16) to analyzed the impact and chose the best one to derive the finally result.

Finally, we put the optimal LAI into the WOFOST model and estimated the yield of each pixel. For getting convenient viewing of comparison results, we conducted spatial statistical analysis with the Geographic Information System (GIS) technology and calculated the forecast yield of each county and compared the estimated yield with statistical yield.

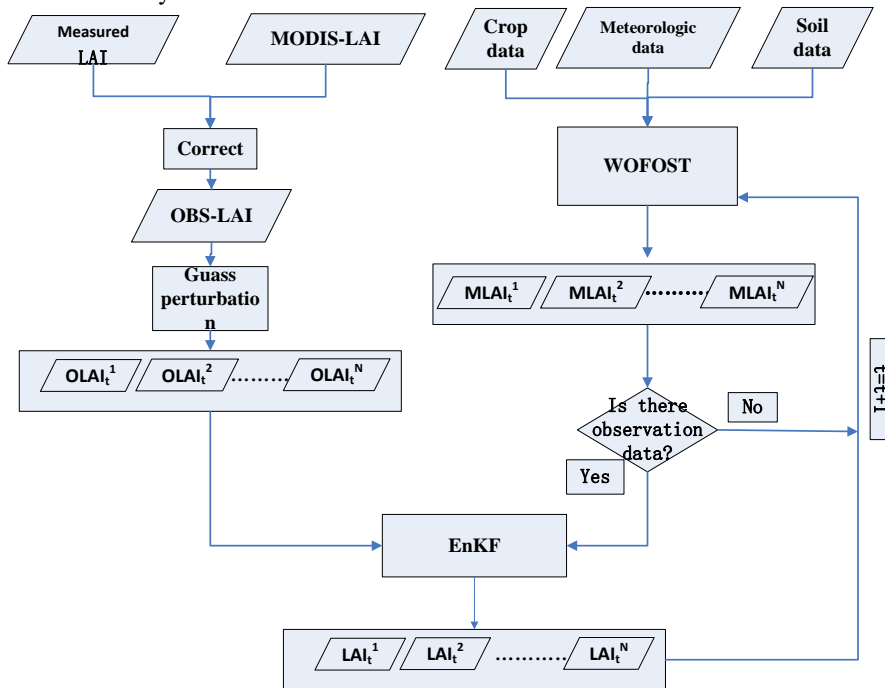


Figure 3 The flowchart of LAI EnKF assimilation

3 Results and discussion

Compared with the LAI simulated by WOFOST (WOFOST-LAI), LAI value with the assimilation (EnKF-LAI) are lower and indicates the trend that in over-wintering stage LAI decreased due to the death of leaves. Compared with the observation LAI, EnKF-LAI values are higher before 86th day and lower between 86th and 130th day. The EnKF adjust the time-series LAI trend and values effectively. In the research we

found that at single point, when the EnKF ensemble size increases, the difference between the profiles of EnKF-LAI and observation LAI is become larger. For the whole study area, the bigger ensemble size, the better the result is obtained. The confidence coefficient(R^2) increased from 0.25 to 0.38 and the root mean square error (RMSE) decreased from 1020 kg/ha to 880 kg/ha as the size increased.

It was found that the bigger the assimilation step size, the greater the fluctuation of LAI profile is; the smaller the assimilation step size is, the more smoother the LAI profile. In terms of the yield forecast for each county, in the potential level, RMSE decreased from 1240kg/ha to 880kg/ha as the assimilation step size decreased.

Figure 4(a) show the estimated yield of WOFOST potential mode. In potential mode, without assimilation most of pixels' estimated yield are higher than 7000 kg/ha, the spatial distribution of the yield has no obvious differences, the reason is probably that in potential level WOFOST assuming that crop growth conditions are the best, in the condition of model parameters are similar, estimated yield are similar and lead to the higher estimated yield.

Figure 4(b) indicate that the estimated yield of WOFOST potential level with EnKF assimilation(with ensemble size of 50 and assimilation step size of 1 day). Through data assimilation, estimated yield accuracy improve greatly at the whole area, most of the estimate yield in potential mode with EnKF assimilation meets the requirement of spatial distribution of winter wheat production statistics. For the yield estimate accuracy with EnKF assimilation the RMSE decreased form 2480 kg/ha to 880 kg/ha, the R^2 improved from 0.13 to 0.38.

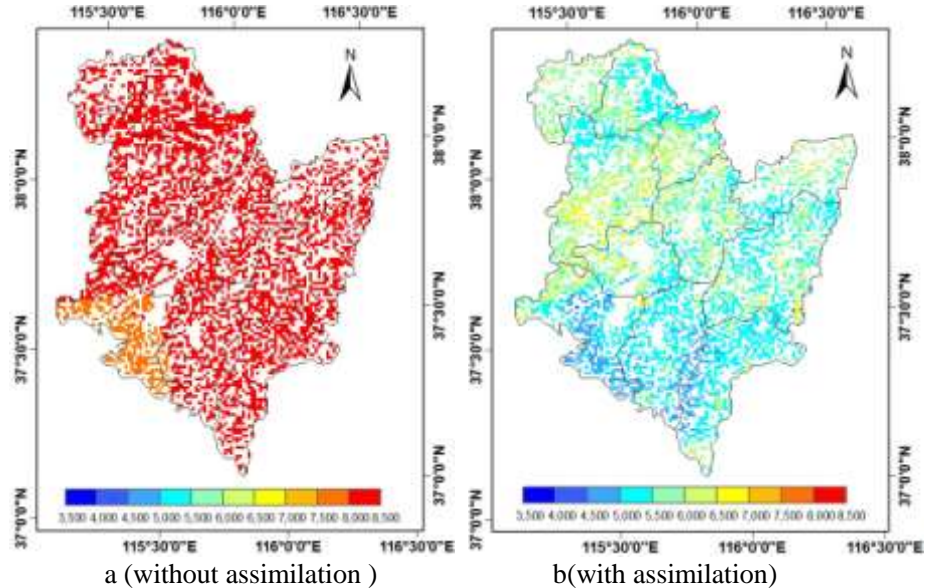


Figure 4 Estimated yield with and without assimilation (in potential level)

4 Conclusion

In this study we use EnKF to assimilate corrected MODIS-LAI into WOFOST model to estimate winter wheat yield and assess the accuracy and applicability of the method. The EnKF assimilation results indicate that, in potential mode, production forecast precision by EnKF data assimilation is significantly higher than without assimilation. Future work will focus on how to integrate the light, temperature and water parameters information retrieval from the multi-source remote sensing data (visible-near infrared, thermal infrared, microwave radar) using EnKF multi-parameters algorithm to further improve the estimation accuracy.

Acknowledgements

This work is supported by the National Science Foundation of China (NSFC) project (NO.40901161), National High Technology Research and Development Program 863 of China (NO.2008AA10Z217) and Chinese Universities Scientific Fund (Project No. 2011JS142) .

References

1. A.J.W.de Wit, C.A.van Diepen: Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*.146:38-56(2007)
2. Bach,H, Mauser,W, Schneider,K: The use of radiative transfer models for remote sensing data assimilation in crop growth models. 4th European Conference on Precision Agriculture. Berlin, Germany (2003)
3. C.Lin, Z.Wang, J.zhu. An Ensemble Kalman Filter for severe dust storm data assimilation over China[J] *Atmospheric Chemistry and Physics Atmos. ChemPhys*, 8:2975–2983(2008)
- 4.Doraiswamy,P.C., Hatfield,J.L., Jackson,T.J., Akhmedov,B., Prueger,J., Stern,A: Crop condition and yield simulations using LANDSAT and MODIS. *Remote Sensing of Environment*.92: 548-559(2004)
5. Evensen G: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*. 99(CS):10143-10162 (1994)
6. Evensen G.: *Data Assimilation The Ensemble Kalman Filter (Second Edition)*. Springer, 2009.
7. Evensen G.: The Ensemble Kalman Filter:theoretical formulation and practical implementation. *Ocean Dynamics*.53:343-367(2003)
8. Gu érif M., Duke,C,L.: Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation. *AgricEcosyst Environ*. 81: 57-69(2000)
9. HongLi: Accounting for Model Error in Ensemble Data Assimilation. *Monthly Weather Review*.12: 3407-3419(2009)

10. Jurgen Komma, Gunter Bloschl, Christian Reszler. Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. *Journal of Hydrology* .357:228-242(2008)
11. Laura Dente, Giuseppe Satalino, Francesco Mattia, Michele Rinaldi: Assimilation of leaf area index derived from ASRA and MERIS data into CERES-Wheat model to map wheat yield. *Remote Sensing of Environment*. 112:1395-1407(2008)
12. Launay M., Guerif M.: Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *AgricEcosyst Environ*. 111: 321—339(2005)
13. Li Xin, Huang Chunlin, Che Tao, Jin Rui et al: Progress and foresight of China land surface data assimilation system research. *Progress in Natural Sciences (in Chinese)*.217(2):163-173(2007)
14. M.R.J.Turner, J.P.Walker, P.R.Oke: Ensemble member generation for sequential data assimilation. *Remote Sensing of Environment* .112: 1421–1433(2008)
15. Nie Su-ping, Zhu Jiang, Luo Yong. Comparison experiments of different model error schemes in ensemble Kalman filter soil moisture assimilation. *Chinese Journal of Atmospheric Sciences (in Chinese)*.34 (3): 580-590(2010)
16. M.Vazifedoust, J.C.van Dam, W.G.M.Bastiaanssen, R.A.Feddes: Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. *International Journal of Remote Sensing*.30 (10): 2523- 2545(2011)