A Numerical Experiment in Assimilating Agricultural Practices in a Mixed Pixel Environment using Genetic Algorithms

Kyoshi Honda
Associate Professor, Space Technology Applications and Research, School of Advanced Technologies, Asian Institute of Technology. P.O. Box 4 Klong Luang 12120 Pathumthani, Thailand.
Tel: +66-2-524-6149; Fax: +66-2-524-5597. Email: honda@ait.ac.th

Amor V.M. Ines
Senior Research Associate, Space Technology Applications and Research, School of Advanced Technologies, Asian Institute of Technology. P.O. Box 4 Klong Luang 12120 Pathumthani, Thailand.
Tel: +66-2-524-5579; Fax: +66-2-524-5597. Email: avmines@ait.ac.th

Abstract. Low spatial resolution remote sensing (RS) data (LSRD) are promising in agricultural monitoring activities due to their high temporal resolution, but under such a spatial resolution, mixing in a pixel is a common problem. In this study, a numerical experiment was conducted to explore a mixed pixel problem in agriculture using a combined RS-simulation model SWAP (Soil-Water-Atmosphere-Plant) and a Genetic Algorithm (GA) approach. Results of the experiments showed that it is highly possible to address the mixed pixel problem with LSRD.
Keywords. data assimilation, remote sensing, genetic algorithms

1. Introduction

Food security studies at the regional scale require a large amount of spatial and temporal information. RS and GIS have been used in the past to account for these spatial data requirements e.g. [1][2]. However, temporal integration, which is considered important in such studies, is difficult to achieve with high spatial resolution RS data (HSRD) e.g. Landsat. Lower spatial resolution RS data (LSRD) e.g. NOAA, MODIS, SPOT-VI are therefore a promising alternative. Aside from their wider area of swath, greater number of temporal data could be possible because of their high frequency of data acquisition.

A tradeoff in using LSRD is the question of how to extract sub-pixel information needed for further analysis e.g. in modeling crop growth at the regional level, unlike with HSRD, distinction among agricultural activities at the pixel level is almost impossible.

These sub-pixel information with LSRD is of high importance in food security studies at the regional scale. Sub-pixel information such as water management practices could include: when the farmers irrigated their crops, how much they irrigate, their irrigation scheduling criteria, frequency of irrigation etc., and agricultural practices may include: what crops they have planted, sowing dates and their distribution in the cropping season, etc. These data are needed for better monitoring of agricultural activities and for more accurate predictions of crop yields in a given time. Coupled with the high temporal information from LSRD, near-real time monitoring of agricultural activities in a regional scale can be improved significantly.

The objective of this paper is to present a methodology to quantify agricultural practices at the sub-pixel level using a combined RS-agro-hydrological modeling and Genetic Algorithm (GA) approach.

2. Methodology

Simulation model

A physically based simulation model SWAP [3] was used to simulate the processes in the soil-water-atmosphere-plant system. SWAP solves the Richards' equation to determine the soil water dynamics in the soil profile. It also considers the fate and transport of solutes in the soil. SWAP is equipped with crop models and water management modules where the growth and development of a crop can be simulated under different climatic and environmental conditions. In general, it is robust in simulating the processes in the soil-water-atmosphere-plant system. See [3] for more details.

Genetic Algorithms

Genetic Algorithms (GAs) are mathematical models of natural genetics where the processes of nature has been abstracted for search and optimization problems [4]. In practical applications, the unknown variables are coded as genes to form a string of variables called chromosome, which is a possible combination of parameters and a possible solution to the problem. First, these chromosomes are initially generated, then evaluated based on a fitness function. After this, they undergo through the process of reproduction. In this process, they have to compete and find their fate in time through the process of selection. The fitter wins, the weaker dies. The winners are then selected to participate in reproducing offspring for the next generation. These selected chromosomes will randomly unite and exchange genetic information, through the process of crossover to produce offspring. The new set of chromosomes are also subjected to mutation to randomly infuse new genetic materials to restore some which are loss along the process. The process of crossover and mutation are controlled by the probabilities of crossover and mutation. Then, the process of selection, crossover and mutation are repeated for many generations to come up with the solution of the problem. In this study, GA was used to unmix the mixed pixel problem in agriculture by estimating the unmixing variables as will be discussed in the next section.

Typical mixing problem in agriculture

The data in an LSRD pixel could be mixed. A time series of this data can describe better the behavior of the features embedded in that large pixel. For example, in agricultural areas, we can monitor a time series of mixed evapotranspiration (ET). This ET, in this particular pixel could be coming from different land uses such as rainfed and irrigated areas, where the latter could composed of 2-croppings and 3-croppings in a year (Eq. 1). The unmixing variables then in this problem are the sowing dates of rainfed (sd_1), irrigated with 2-croppings(sd_2, sd_3), irrigated with 3-croppings (sd_4, sd_5, sd_6) and the area fractions of rainfed and irrigated with 2- and 3-croppings, a_1, a_2 and a_3, respectively (Eq. 2). Where j is an index for sowing dates, i an index for area fractions, p means pixel, m is the total number of land uses (m=3), t an index for time, and **k** is a vector of unmixing variables. Sowing dates are all in Day of Year (DOY).

$$ET_{t,p}(\mathbf{k}_p) = \sum_{i=1}^{m} a_{i,p} ET_{t,i,p} \qquad \forall t, \forall i, \forall p \qquad (1)$$

$$\mathbf{k}_p = \{ sd_{j,p}, a_{i,p} \} \qquad \forall j, \forall i, \forall p \qquad (2)$$

Unmixing model

Considering one pixel, the unmixing algorithm can be formulated then as follows:

$$Obj(\mathbf{k})_{min} = Min \left\{ \frac{1}{n} \sum_{t=1}^{n} \left| ET_t(\mathbf{k}) - \hat{ET}_t \right| \right\} \quad \forall t \qquad (3)$$

$$b_{min_j} \le sd_j \le b_{max_j} \qquad (j = 1,\ldots,6) \qquad (4)$$

$$sd_j - sd_{j-1} \ge 100 \qquad (j = 3,5,6) \qquad (5)$$

$$\sum_{i=1}^{m} a_i = 1.0 \qquad \forall i \qquad (6)$$

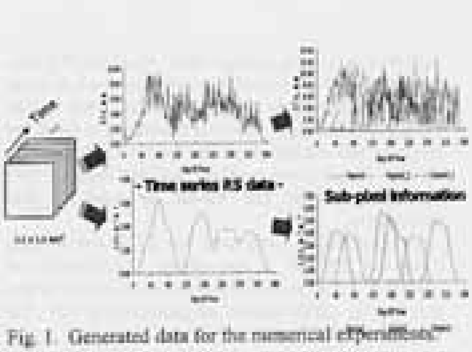

Fig. 1. Generated data for the numerical experiments.

$$0 \le a_i \le 1.0 \qquad \forall i \qquad (7)$$

where ET with hat is the mixed data from remote sensing which should be compared with the mixed signature produced by SWAP using the proposed solution of GA; b_{min} and b_{max} define the range of the search space for every sowing date and n is the total number of data used in the solution. This unmixing model is implemented using generated RS data as discussed below.

Numerical data

First, a numerical data was generated for a hypothetical pixel comprising three agricultural land classes: (i) rainfed area (ii) irrigated area with two croppings and (iii) irrigated area with three croppings in a year (Fig 1). These classes were individually simulated by SWAP then the results were mixed using their corresponding area fractions in the pixel to arrive at the mixed signature of the pixel in a year.

In this study, two biophysical data were used, evapotranspiration (ET) and Leaf Area Index (LAI) as they can be derived from remote sensing data.

Then, GA was used to estimate back the agricultural practices based on the temporal mixed signal (i.e. ET or LAI) of the pixel by solving Eqs. 3-7. The area fractions of each agricultural class and their corresponding sowing dates: one, for rainfed; two dates for irrigated with two croppings and three, for irrigated with three croppings, were coded as a chromosome in the GA. In the implementation, two major cases were analyzed, considering (i) a dataset without noise and (ii) data with noise, both ET and LAI. Likewise, appropriate sampling scheme was also evaluated where the mixed data were aggregated every 10 days (e.g. ET_{10d}) and a moving average of every 10 days (e.g. ET_{10dma}).

3. Results and Discussions

Table 1 shows the solutions of GA to the mixed-pixel problem. It is clear that GA is able search for the near-optimum solution of the problem using either ET or LAI and either with aggregated 10-day data or with 10-day moving average data. However, it is also apparent that there are solutions that are stuck to some insensitive regions in the search surface such as that when using LAI_{10dave}. This has something to do on the sampling scheme that should be used in collecting data for the mixed pixel problem and the type of bio-physical variables that should be used in data assimilation.

The population sizes used for ET_{1dd} and ET_{10dave} cases are 10 and 5 respectively; probability of crossover is 0.5; probability of mutation (creep) is also 0.5; seed is -1000; number of generation is 150. For LAI, the population size used is 5 in both cases; genetic parameters are the same as ET. The good solutions were achieved after 1.8 hours with Pentium 4, 1.8 GHz.

In reality, RS data are corrupted with errors, hence simulating a situation like this is worth noting. Using data with a 10% random noise, GA was still able to produce results that are reasonable. An average error of 0.28 and 0.29 mm d^{-1} were achieved using ET_{1dd} and ET_{10dave}, respectively. The parameter values are relatively near the base values regardless of the added noise. See [5] for more details.

The results of the numerical experiments show that it is highly possible to solve a typical mixed pixel problem in agriculture using the proposed approach. Note, however, that this present approach (a dynamic linkage approach) requires a lot of computing power and might not be practical in analysing a regional data unless it is implemented in cluster computers. A more practical approach such as using Look Up tables [6] is needed.

4. Conclusion

A methodology that can possibly solve mixed pixel problem with LSRD in their agricultural applications is presented in this paper. It is shown that using a combined RS-agrohydrological modeling and Genetic Algorithms approach, it is possible to simulate a mixed pixel problem in agriculture. Consequently, this approach can also unmix such a mixed pixel problem.

In this paper, the unmixing model was implemented using simulated RS data. It was found that GA was powerful to unmix the mixed signature in an agricultural area. The methodology, however, needs still to be improved for practical application in the field.

5. References

[1] Droogers P., Bastiaanssen W.G.M., Beyazgül M., Kayam Y., Kite G.W. and Murray-Rust H. (2000). Distributed agro-hydrological modeling of an irrigation system in Western Turkey. Agric. Water Manage. 43, 183-202.

[2] Ines, A.V.M., Gupta A.D., Loof R. (2002). Application of GIS and crop growth models to estimate water productivity. Agric. Water Manage. 54, 205-225.

[3] Van Dam, J. C. (2000). Field-scale water flow and solute transport. SWAP model concepts, parameter estimation and case studies. Doctoral Thesis, Wageningen University, The Netherlands.

[4] Goldberg D. E. (1989). Genetic Algorithms in search and optimization and machine learning. Addison-Wesley Publ. Co. Inc, USA.

[5] Honda K., Ines A.V.M. and Kamthonkiat D. (2003). Quantifying agricultural practices in a mixed-pixel environment using Genetic Algorithms. (in preparation)

[6] Ines, A.V.M. and Honda, K. (2003). Development of a methodology to quantify water management and agricultural practices in a mixed pixel environment - Numerical experiment. 1st International Symposium on Southeast Asian Water Environment. October 23-25, 2003, AIT, Bangkok, Thailand (accepted).

Table 1. Solution of GA to the mixed pixel problem.

	Un-mixing Parameters									
RS variables	sd_1	sd_2	sd_3	sd_4	sd_5	sd_6	a_1	a_2	a_3	Error
ET										
ET_{1dd}	139	32	186	1	121	248	0.15	0.50	0.35	0.01
ET_{10dave}	138	32	186	1	121	248	0.16	0.50	0.34	0.03
LAI										
LAI_{1dd}	140	32	186	2	121	246	0.16	0.50	0.34	0.03
LAI_{10dave}	120	31	186	1	129	249	0.12	0.52	0.36	0.08
Base values	141	32	186	1	121	248	0.15	0.50	0.35	-