

# On quantifying agricultural and water management practices from low spatial resolution RS data using genetic algorithms: A numerical study for mixed-pixel environment

Amor V.M. Ines<sup>a,\*</sup>, Kiyoshi Honda<sup>b,1</sup>

<sup>a</sup> International Research Institute for Climate Prediction, The Earth Institute at Columbia University, Palisades, NY 10964, USA

<sup>b</sup> Space Technology Applications and Research, School of Advanced Technologies, Asian Institute of Technology (AIT), P.O. Box 4, Klong Luang, 12120 Pathumthani, Thailand

Received 8 June 2004; received in revised form 15 October 2004; accepted 5 November 2004

Available online 14 March 2005

## Abstract

In this paper, we present a genetic algorithm-based methodology to quantify agricultural and water management practices from remote sensing (RS) data in a mixed-pixel environment. First, we formulated a linear mixture model for low spatial resolution RS data where we considered three agricultural land uses as dominant inside the pixel—rainfed, irrigated with two, and three croppings a year; the mixing parameters we considered were the sowing dates, area fractions of agricultural land uses in the pixel, and their corresponding water management practices. Then, we carried out numerical experiments to evaluate the feasibility of the proposed approach. In the process, the mixing parameters were parameterized by data assimilation using evapotranspiration and leaf area index as conditioning criteria. The soil–water–atmosphere–plant system model SWAP was used to simulate the dynamics of these two biophysical variables in the pixel. The results of our numerical experiments showed that it is possible to derive some sub-pixel information from low spatial resolution data e.g. the existing agricultural and water management practices in a region, which are relevant for regional agricultural monitoring programs.

© 2005 Elsevier Ltd. All rights reserved.

**Keywords:** Genetic algorithms; SWAP simulation model; Remote sensing; Low spatial resolution; Mixed-pixel; Data assimilation

## 1. Introduction

Remote sensing (RS) is very promising in monitoring agricultural and water management activities because both the spatial and temporal characteristics of a region can be easily accounted for by satellite imageries [14]. As a result, seasonal change of vegetation activities at the regional level can be monitored reasonably well helping

policy makers and farm/water managers make better operational decisions.

Generally, crop monitoring with RS is a two-step approach in practice—first, deriving simplified relational models for spatial analysis by relating vegetation indices (e.g. the normalized difference vegetation index, soil adjusted vegetation index etc.) and observed crop parameters (e.g. crop age, yield, leaf area index etc.) from selected training areas, then applying these models to the RS image to determine the spatial conditions of the crops in the region. This is rather a static approach and not readily adoptable to environments different from where the relational models are developed (e.g. [40,41,30]). There is also a temporal scale limitation unless a multi-temporal model is developed for this

\* Corresponding author. Present address: Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX 77843. Tel.: +1 845 680 4501; fax: +1 845 680 4864.

E-mail addresses: [ines@iri.columbia.edu](mailto:ines@iri.columbia.edu), [avmines@tamu.edu](mailto:avmines@tamu.edu) (A.V.M. Ines), [honda@ait.ac.th](mailto:honda@ait.ac.th) (K. Honda).

<sup>1</sup> Tel.: +66 2 524 6149; fax: +66 2 524 5597.

purpose, e.g. per crop development stage. These models commonly use high spatial resolution RS data for crop monitoring. High spatial resolution data applications, however, are limited for larger domains because of several reasons—their spatial coverage is limited, they have low frequency of data acquisition, the uncertainty of cloud-free images during a cropping season and their high cost [34,35].

Several studies have developed methodologies to improve the temporal resolution of data when using high spatial resolution data for agricultural and water management applications. A dynamic approach called RS-simulation modeling, which explores the synergy of RS and simulation models, has been proposed. In this approach, RS derived information, e.g. evapotranspiration (ET) or leaf area index (LAI), are used to steer the simulation models to determine the site specific initial and boundary conditions for their applications in the field or region (e.g. [7,8,11,14,19], among others). Using this approach, a temporal resolution as fine as one day is possible for in-depth analysis. Moreover, the RS-simulation modeling is not only limited to crop monitoring but also has potentials for crop forecasting [20].

Although the combined RS-simulation model approach appears to be very promising for monitoring agricultural and water management activities, there is still a need to address the problem of high spatial resolution data availability in a cropping season. Is there enough data available for reasonable RS-simulation modeling? Some studies have tackled this question by exploring the sensitivity of the physical properties of a hydrologic system with remotely sensed data to infer if timing of data acquisition could help resolve the problem on data limitation. Jhorar et al. [22] found that soil hydraulic parameters are highly sensitive to ET during periods of water stress. Ines and Droogers [20] applied this idea in deciding which Landsat7 ETM+ images to use in their study and based on their results, corroborated the finding of Jhorar et al. [22].

But if we develop further the approach for operational use, a careful consideration of cost is necessary. In view of this, using low spatial resolution RS data may be a practical alternative for RS-simulation modeling. Low spatial resolution data have attributes that clearly outweigh the limitations associated with the use of high spatial resolution data for larger scale agricultural monitoring—they have higher temporal resolution and wider spatial coverage, and minimal cost [34]. In fact, there exist public domain standard data products on the Internet e.g. MODIS-LAI, which are available at both daily and composite time scales. Moreover, the scan line correction problem that afflicted Landsat7 in providing high quality RS data has resulted to an increasing interest on the use of moderate or low spatial resolution data for agro-hydrological studies, thus warrants further their exploration in this regard. One big

challenge, however, is the problem of mixing inside a pixel (the basic unit in an RS image). Since the spatial resolution is coarser, several agricultural land uses could be embedded in one pixel.

This opens a research avenue for low spatial resolution RS data in agricultural monitoring. In the literature, the mixed-pixel problem with low spatial resolution data has been extensively studied, but not primarily for agricultural and water management applications. Shimabukuro and Smith [32] earlier proposed a least squares mixing model to generate fraction images from RS multi-spectral data. This method has been applied and improved significantly yielding promising results (e.g. [17,29,2,4,32], among others). Kerdiles and Grondona [23] used the method to produce fraction images of crops and pastures, and concluded that its application to crop monitoring is highly promising. But in agricultural monitoring programs, crop area fractions are only part of the required information. Other data such as agricultural and water management practices are vital and needed to devise strategies that could improve the overall performance of the agricultural system. The question is: Is it possible to derive from low spatial resolution data sowing dates, irrigation dates and frequency, expected yields, etc.?

The objective of this study is to develop a method to quantify agricultural and water management practices from low spatial resolution data using the combined RS-simulation model approach (see [20]). It is based upon the hypothesis that with the wealth of information embedded in low spatial resolution data, it is possible to derive such information at the sub-pixel level.

## 2. Materials and methods

### 2.1. Mixed-pixel model formulation

The most practical way to solve mixing in the pixel is by assuming it as a linear mixture problem. In RS, linear mixture models are generally formulated at the spectral level. It is assumed that the average reflectance in a spectral band is equal to the sum of the products of the area fractions of each component and their corresponding standard reflectance in that part of the spectrum, plus an error term. The sum of the area fractions should be equal to 1; the standard reflectance can be derived either from spectral libraries, endmember determination, or directly from higher resolution data. The linear mixture model is usually solved using a constrained least squares optimization (see [32], [17,2,4]). Other methods used direct information from higher resolution data to derive linear downscaling models for extrapolation from coarser resolution data; the vegetation–soil–water index works well with this approach [34,35]. Norman et al. [28] successfully disaggregated energy fluxes from coarser

scale data based on the concept of blending height and downscaling scheme using information from high spatial resolution data. In this section, we present a linear mixture model based on temporal integration of derived RS data e.g. [12]. The formulation is limited to agricultural applications.

At the pixel level, the RS data can be mixed, it could be composed of signatures coming from several land features such as a variety of crops, bare soil etc. It is well known in the literature that the observed signatures from RS are somewhat directly related to the activities on the ground ([19,14], among others). Exploring this dependency could give light to the features of the pixel or group of pixels under study.

Let us denote the spatial features in a pixel as  $i$ . Assuming that there are three significant agricultural land uses (classes) in a region of interest, i.e. rainfed ( $i = 1$ ), irrigated with two croppings a year ( $i = 2$ ) and irrigated with three croppings a year ( $i = 3$ ). Then a time series of a system variable e.g. evapotranspiration (ET) in a mixed-pixel environment can be defined as:

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

$$\mathbf{k}_p = \{\mathbf{sd}_{j=1,\eta}, \mathbf{a}_{i=1,m}\}_p \quad \forall p \quad (2)$$

where,  $\mathbf{k}_p$  is a vector of sensitive parameters significantly influencing the temporal behavior of ET, e.g. date of sowing ( $sd_j$ ) and area fraction ( $a_i$ ) of an agricultural class in pixel  $p$  (see Eq. (2));  $m$  stands for the maximum possible number of agricultural classes in a pixel (here,  $m = 3$ );  $ET_{tip}$  is an instantaneous component of ET at time  $t$ , under class  $i$ , in pixel  $p$ ;  $j$  is the index for date of sowing, which has a maximum possible value of  $\eta$  in this case six (6), i.e. one date for rainfed ( $j = 1$ ); two, for irrigated with two croppings ( $j = 2, 3$ ); and three dates for irrigated with three croppings in a year ( $j = 4, 5, 6$ ).

The possible components of the mixed signature in Eq. (1) can be defined as in Eqs. (3)–(5) considering timing of sowing as the most dominant factor in defining the shape of the individual series:

$$ET_{i1p} = ET(sd_1)_p \quad \forall t, \forall p \quad (3)$$

$$ET_{i2p} = ET(sd_2, sd_3)_p \quad \forall t, \forall p \quad (4)$$

$$ET_{i3p} = ET(sd_4, sd_5, sd_6)_p \quad \forall t, \forall p \quad (5)$$

where  $ET(sd_j \dots)$  can be derived from a simulation model. These are the individual signatures in a pixel that are of prime interests in solving the mixing problem. Note further that the shape of Eq. (1) depicts the spatial features (both mixing and activities) in a pixel. We will exploit the shape of the curve later to determine vector  $\mathbf{k}_p$ .

## 2.2. Un-mixing algorithm

### 2.2.1. Formulation

The mixed-pixel problem in low spatial resolution RS data is a typical optimization problem where true values of the mixing parameters  $\mathbf{k}$  have to be determined to optimize an objective while satisfying a set of constraints. This section presents an un-mixing algorithm for one pixel considering agricultural practices as the dominant factor in determining the shape of Eq. (1). Later, in this paper, we will extend the problem to include water management practices as additional mixing parameters.

The objective function can be defined as:

$$\text{Obj}(\mathbf{k})_{\text{Min}} = \text{Min} \left\{ \frac{1}{n} \sum_{t=1}^n |ET_t(\mathbf{k}) - \hat{ET}_t| \right\} \quad (6)$$

Subject to these constraints:

(a) Range of sowing dates

$$b_{\text{min}_j} \leq sd_j \leq b_{\text{max}_j} \quad \forall j \quad (7)$$

(b) Constraints to avoid cropping overlaps

$$sd_j - sd_{j-1} \geq 100 \text{ days} \quad (j = 3, 5, 6) \quad (8)$$

(c) Sum of the area fractions

$$\sum_{i=1}^m a_i = 1.0 \quad (9)$$

(d) Range of area fractions

$$0 \leq a_i \leq 1.0 \quad \forall i \quad (10)$$

where  $\hat{ET}_t$  is the measured ET from RS at time  $t$  (mixed data);  $ET_t(\mathbf{k})$  is an estimated ET given  $\mathbf{k}$ ; Min means the objective function (Obj) should be minimized;  $n$  is the time domain. Eq. (8) forces the gap between the sowing dates to be greater than or equal to 100 days to allow full maturity of the previous crop before the start of the next cropping season. This constraint is applicable only to irrigated areas with two and three croppings a year.

### 2.2.2. Simulation model

The un-mixing algorithm we implemented here uses the combined RS-simulation model approach where in real situations the time series of ET or LAI can be simulated and compared with RS measurements. We used the soil–water–atmosphere–plant model SWAP to simulate the processes of the three agricultural land uses we considered in the pixel, i.e. rainfed, irrigated with two, and three croppings a year.

SWAP is a physically based, field scale, agro-hydrological model that simulates the processes occurring in the soil–water–crop–atmosphere–system [37]. The core of the model is the 1D-Richards' equation, where verti-

cal soil water movement is calculated from the spatial and temporal potential differences in the soil column. The soil hydraulic functions are defined by the Mualem–Van Genuchten equations [27,38], which describe the capacity of the soil to store, release and transmit water under different environmental and boundary conditions.

SWAP can account for several combinations of the top and bottom boundary conditions. It is equipped with crop models including WOFOST [33], which we used in modeling crop growth, and water management modules for irrigation and drainage studies. The model simulates both the soil–water quantity and quality with a temporal resolution of one day, along with other state variables.

The potential evapotranspiration ( $ET_{pot}$ ) is calculated by Penman–Montieth equation. The potential transpiration ( $T_{pot}$ ) and soil evaporation ( $E_{pot}$ ) are partitioned from  $ET_{pot}$  using LAI or the soil cover fraction. As the soil dries, the model reduces  $ET_{pot}$  into  $ET_{act}$  (actual ET) where the  $E_{pot}$  component is reduced to  $E_{act}$  according to some established empirical relationships e.g. [5,6] or by using Darcy’s law in the soil surface, and a water stress reduction factor is applied to reduce  $T_{pot}$  into  $T_{act}$ . The model is well validated under different climatic and environmental conditions (see [39,15,36,31], among others).

### 2.2.3. Genetic algorithms

The problem presented above is highly combinatorial in nature. A robust search and optimization procedure, not easily trapped to local optima in the search surface, is necessary for successful implementation. In this study, we have selected genetic algorithm (GA) to solve the mixed-pixel problem.

GAs are search algorithms based on the mechanics of nature, they combine survival of the fittest among string structures with a structured yet randomized information exchange to arrive at a solution [16,18]. The unknown variables are coded as a set of binary sub-strings (in binary GA) to form a string structure called a chromosome. The bits (0s and 1s) arrangement in the chromosome is a possible combination of the unknown variables and can be a solution to the problem. The process starts by initially generating a set of chromosomes (called a population), which are then evaluated individually to determine their suitability based on a fitness function. After this, they undergo through the process of selection. Based on their suitability, they compete to be selected for the mating pool. The fitter chromosomes survive, the weaker dies. The survivors participate in generating the new offspring for the next generation. The selected chromosomes (parents) randomly unite and exchange genetic information through the process of crossover to produce offspring. The resulting new chromosomes (now offspring) are subjected to mutation to randomly infuse

new genetic materials in the new generation. Random mutation is necessary to restore some genetic materials lost due to “genetic drift”. Crossover and mutation are controlled by their respective probability values (e.g.  $p_{cross}$ ,  $p_{mutate}$ ). If a random number generated is less than or equal to  $p_{cross}$ , crossover happens, otherwise, not. If no crossover occurs the parents enter into the new generation and participate further in the process. Mutation also occurs if a generated random number during the mutation process is less than or equal to  $p_{mutate}$ . The process of selection, crossover and mutation are repeated for many generations to arrive at the best possible solution.

In this study, we have used a modified- $\mu$ GA to solve the mixed-pixel problem. A conventional  $\mu$ GA uses micro-population to sample the search space and restarts when converges along the generations, no mutation is allowed [25]. The modified- $\mu$ GA is a variant that introduces creep mutation to randomly alter the sub-strings of a chromosome, and the criterion for restarting set higher to increase the rate of population restarts, relatively increasing the infusion rate of new genetic materials in the process. During restart, the elite chromosome is preserved and the rest of the population are randomly generated (see [21,9]). Some of the GA variant operators we used are described below:

*Selection.* The selection method we used is a binary tournament selection with shuffling. Before the selection process, the positions of the chromosomes in the population are randomly shuffled such that a chromosome at position 2, for instance, can be transferred to position 5, vice versa. The binary tournament proceeds by selecting two chromosomes (hence the term binary) from the shuffled population then the selected chromosomes compete for a position in the mating pool according to their fitness. The chromosome with a higher fitness value wins and joins the mating pool.

*Creep mutation.* Creep mutation ( $p_{creep}$ ) is a variant of mutation that occurs at the real space (base 10). The binary sub-strings are mutated between their minimum and maximum values, i.e. the extent of the search space for a particular variable. Unlike jump mutation ( $p_{mutate}$ ), which happens at the binary space (i.e. bit by bit random mutation), lesser perturbation is introduced to the micro-population compromising not from restarting along the generations.

*Restarting micro-population.* The idea behind the restarting micro-population is to increase the sampling of the global search space even with a smaller population. This happens when the criterion for restarting is achieved along the generations. Usually, this is quantified by the degree of dissimilarity of the bits positions in the population. For example, if all of the chromosomes are not similar by only 5% (i.e. 95% similarity) after a generation then the population restarts. Another implication is that, the solution is achieved faster than

with the conventional notion of a population in simple GA.

*Elitism.* The structure of the elite chromosome is important in the search process because it contains some of the primordial ingredients of the solution. Thus, in an elitist GA, the best chromosome in the past generation has a secured position for the next generation. Regardless if the micro-population is restarting or not, the elite chromosome is always generated in the new population with its position randomly selected. During restart, however, it always occupies the first position in the population.

*2.2.3.1. GA implementation.* We implemented two unconstrained forms of the optimization problem (Eqs. (6)–(10)) in GA, the classic penalty method (CPM) of Goldberg [16]—Eq. (11) and a modified one (MPM) by Chan-Hilton and Culver [10]—Eq. (12).

$$Z(\mathbf{k})_{\text{Min}} = \text{Min} \left\{ \frac{1}{n} \sum_{t=1}^n |\text{ET}_t(\mathbf{k}) - \hat{\text{ET}}_t| + \text{Penalty} \right\} \quad (11)$$

$$Z(\mathbf{k})_{\text{Min}} = \text{Min} \left\{ \frac{1}{n} \sum_{t=1}^n |\text{ET}_t(\mathbf{k}) - \hat{\text{ET}}_t| \times (1 + \text{Penalty}) \right\} \quad (12)$$

where penalty is expressed as:

$$\text{Penalty} = \sum_{h=1}^4 \lambda_h \Phi_h; \quad h \text{ is a running index for constraints} \quad (13)$$

where

$$\Phi_h = (\text{sd}_j - \text{sd}_{j-1} - 100)^2 \quad (h = 1, 2, 3; j = 3, 5, 6) \quad (14)$$

$$\Phi_h = \left\{ 1 - \sum_{i=1}^m a_i \right\}^2 \quad (h = 4) \quad (15)$$

$\lambda_h$  is a penalty coefficient for all  $h$ .

Eq. (14) is applied only to the irrigated areas with multiple croppings a year.

If a chromosome (denoted by  $\mathbf{p}$ ) violates any of the constraints (sowing date overlap and area fractions), respective penalties are imposed using the following rules (Eqs. (16) and (17)):

$$\text{if } (\text{sd}_j - \text{sd}_{j-1} - 100 \geq 0); \lambda_h = 0 \text{ otherwise } 10 \quad (h = 1, 2, 3; j = 3, 5, 6) \quad (16)$$

and

$$\text{if } \left( 1 - \sum_{i=1}^m a_i \neq 0 \right); \lambda_h = 20 \text{ otherwise } 0 \quad (h = 4) \quad (17)$$

Finally, the measure of a chromosome is quantified by the fitness function defined in Eq. (18):

$$\text{fitness}(\mathbf{p})_{\text{Max}} = \frac{1}{Z(\mathbf{k})} \quad (18)$$

fitness ( $\mathbf{p}$ ) is maximized if  $Z(\mathbf{k})$  is minimized (see Eqs. (11) and (12)), which is basically our aim; chromosome  $\mathbf{p}$  is defined below.

For sowing dates and area fractions determination, the chromosomes consist only of eight binary sub-strings because  $a_3$  can be expressed in  $a_1$  and  $a_2$ , i.e.  $a_3 = 1 - (a_1 + a_2)$ , reducing the length of the string (see [26]). The chromosome then is expressed as:

$$\mathbf{p} = \{\mathbf{sd}_{j=1,\eta}, \mathbf{a}_{i=1,2}\} \quad (19)$$

hence  $\mathbf{k} = \{\mathbf{p}, a_{i=3}\}$ , which means that  $\mathbf{p}$  is the only component of  $\mathbf{k}$  that is explicitly propagated in GA. We coded Eq. (19) as a binary string structure wherein all included mixing parameters in the chromosome are coded as binary sub-strings, as shown in Fig. 1. Eq. (20) concatenates the binary representation (base 2) of the parameter to its decimal form (base 10), where  $C_{\text{Max}}$  and  $C_{\text{Min}}$  correspond to the maximum range of the mixing parameter  $\rho$  (an element of  $\mathbf{p}$ );  $C$  is the decimal value of a binary sub-string  $\rho$ ;  $L_\rho$  is the length of the sub-string  $\rho$ ;  $\beta$  is the bit value—either 0 or 1 at position  $\ell$  in the sub-string.

$$C(\rho) = C_{\text{Min}}(\rho) + [C_{\text{Max}}(\rho) - C_{\text{Min}}(\rho)] \frac{\sum_{\ell=1}^{L_\rho} \beta_\ell 2^{\ell-1}}{2^{L_\rho} - 1} \quad \forall \rho \quad (20)$$

When water management practices are included as mixing parameters, the chromosomes are extended to this structure,  $\mathbf{p} = \{\mathbf{sd}_{j=1,\eta}, \mathbf{a}_{i=1,2}, \mathbf{w}_{i=2,3}\}$ .

### 2.3. Numerical experiments

#### 2.3.1. Numerical data

This study is the first stage of our effort to develop a methodology that can be practically applied to crop monitoring using low spatial resolution data. For this reason, we chose to use numerical experiments in testing the proposed approach to understand better its strengths and limitations. First, we simulated ET and

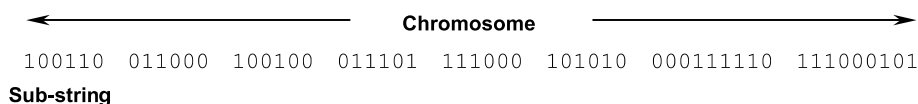


Fig. 1. A binary representation of a chromosome in GA for the mixed-pixel problem.

LAI for each agricultural land use with SWAP by assuming realistic values of the mixing parameters in the simulations. Then, we mixed the generated data from each land use to come up with the “observed” satellite data. The mixing parameters were estimated back using the proposed approach above. Several studies have used a similar procedure because they can easily control the externalities affecting the data and the uncertainties in the analysis, allowing them to understand better the processes involved (e.g. [24], [1,22]).

2.3.1.1. *SWAP input data.* Fig. 2 shows the daily data of solar radiation (SRAD), maximum and minimum temperature (TMAX, TMIN), wind speed and humidity used in running SWAP. The weather data were taken from the Asian Institute of Technology (AIT) weather database similar to that being developed for a proposed crop-monitoring project in Thailand. Table 1 shows the soil data used in the simulations. The depth of soil column considered is 200 cm partitioned into two layers, and discretized into a total of 33 compartments with a finer division at the top 10 cm from the soil surface. The crop we considered here is hypothetical, parameterized in WOFOST.

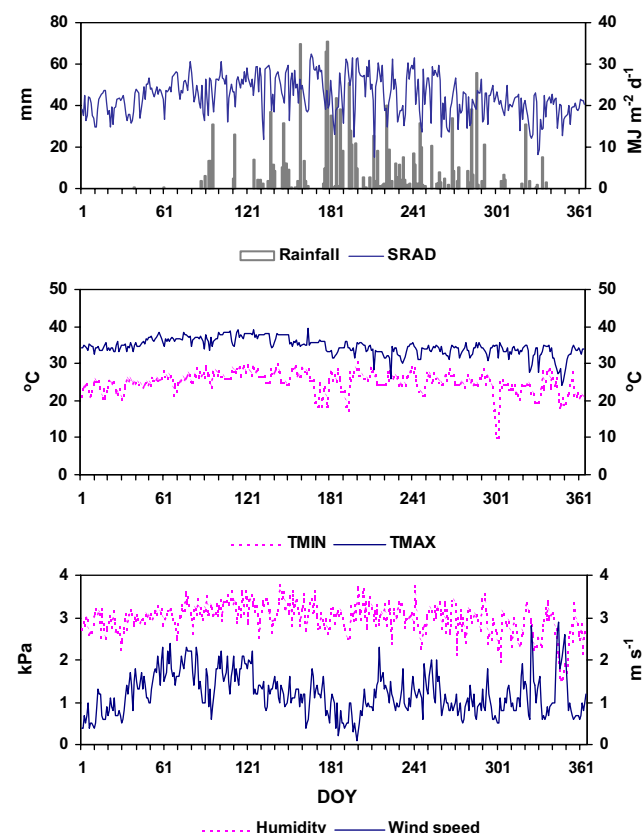


Fig. 2. Daily weather data used in the study.

Table 1  
Soil input parameters used in the simulations [21]

Soil hydraulic parameters <sup>a</sup>	Layer 1 (0–60 cm) 30.5%–50.1%–19.4% (S–Si–Cl)	Layer 2 (60–200 cm) 29.0%–48.8%–22.2%
$\alpha$	0.02100	0.01800
$\lambda$	–2.5230	–3.9970
$\theta_{res}$	0.000	0.000
$\theta_{sat}$	0.384	0.365
$K_{sat}$	15.282	10.188
$n$	1.2080	1.1370

Note:  $\theta_{res}$ —residual water content ( $\text{cm}^3 \text{cm}^{-3}$ );  $\theta_{sat}$ —saturated water contents ( $\text{cm}^3 \text{cm}^{-3}$ );  $\alpha$  ( $\text{cm}^{-1}$ ),  $n$  (–)—shape parameters of the soil water retention curve;  $\lambda$  (–) is a shape parameter of the hydraulic conductivity function;  $K_{sat}$ —saturated hydraulic conductivity ( $\text{cm d}^{-1}$ ) [for the Mualem–Van Genuchten equations]; S–Si–Cl—% of sand, silt and clay.

<sup>a</sup> Derived using pedo-transfer functions [13].

2.3.1.2. *Initial and boundary conditions.* In the simulations, we assumed the initial wetness of the soil column to be at equilibrium condition. The top boundary condition (i.e. either infiltration or evaporation) is decided by SWAP during run-time based on a built-in decision tree [37]. While the groundwater level is a major determinant in the water balance, we did not consider in this example to simplify and speed up the calculations in SWAP. We considered the soil column to be well drained in the simulations.

2.3.1.3. *Simulated agricultural land uses.* As we have discussed earlier, we assumed three specific land uses to dominate the landscape in an agricultural region for illustration purposes: (i) rainfed agriculture (rainfed), (ii) irrigated with two croppings a year (irrigated\_2) and (iii) irrigated with three croppings a year (irrigated\_3).

Table 2a shows the true values of agricultural and water management practices we used in simulating the three agricultural land uses in the pixel. The mixing parameters considered are the sowing dates  $sd_1 \dots sd_6$  in day of year (DOY), area fractions  $a_1 \dots a_3$  and water management practices  $w_2$  and  $w_3$  (see also Section 2.1 for definition of indices). The mixing variable  $a_3$  is calculated from  $a_1$  and  $a_2$  knowing that the sum of the three area fractions is 1.0.

The water management practices in each irrigated land use could vary for each cropping season but for simplicity sake we assumed a standard irrigation cycle for the rest of the year, meaning that the criterion used for applying water in the first cropping season is practiced for the rest of the year. In this study, we used a minimum allowable ratio of  $T_{act}/T_{pot} = 0.75$  to indicate the timing of irrigation. We assumed a depth of irrigation of 100 mm per application, which is a conservative estimate for flooded irrigation. This means that when

Table 2

(a) Agricultural and water management variables and their representations in the genetic algorithm. (b) Summary of genetic algorithm parameters used in this study

Variables	True values	Search space		No. of bits
		Min. values	Max. values	
sd <sub>1</sub>	141	120	183	6
sd <sub>2</sub>	32	1	64	6
sd <sub>3</sub>	186	140	203	6
sd <sub>4</sub>	1	1	64	6
sd <sub>5</sub>	121	90	253	6
sd <sub>6</sub>	248	230	293	6
a <sub>1</sub>	0.15	0.0	1.0	9
a <sub>2</sub>	0.50	0.0	1.0	9
a <sub>3</sub> <sup>a</sup>	0.35	–	–	–
w <sub>2</sub> <sup>b</sup>	0.75	0.55	0.97	5
w <sub>3</sub>	0.75	0.55	0.97	5

GA parameters	Population size (n <sub>pop</sub> )		
	30	10	5
p <sub>creep</sub>	0.1 (I), 0.5 (II)	0.1, 0.5	0.1, 0.5
p <sub>cross</sub>	0.5	0.5	0.5
Random number seed	–1000	–1000	–1000
Maximum number of generations	150	150	150

Note: When sowing dates and area fractions (Cases 1 and 3) are to be determined, only sd<sub>1</sub>...a<sub>2</sub> are propagated in GA; when water management is added (Case 2), the whole set (sd<sub>1</sub>...a<sub>2</sub>, w<sub>2</sub>, w<sub>3</sub>) are propagated.

p<sub>creep</sub> and p<sub>cross</sub> are probabilities of creep mutation and crossover.

I—p<sub>creep</sub> = 0.1; II—p<sub>creep</sub> = 0.5.

<sup>a</sup> a<sub>3</sub> = 1 - ∑<sub>i=1</sub><sup>2</sup> a<sub>i</sub> = 0.35, thus excluded in chromosome p.

<sup>b</sup> w<sub>i</sub> = (T<sub>act</sub>/T<sub>pot</sub>)<sub>i</sub> where T<sub>act</sub> and T<sub>pot</sub> are the actual and potential transpiration (mm d<sup>-1</sup>), respectively; w<sub>2</sub> for irrigated<sub>2</sub>; w<sub>3</sub> for irrigated<sub>3</sub>.

the threshold of relative transpiration is exceeded (i.e. T<sub>act</sub>/T<sub>pot</sub> < 0.75) SWAP will irrigate with a depth of 100 mm.

Fig. 3a shows the daily actual ET for each of the components (top) and the mixed daily ET (weighted mean) in the pixel assuming that rainfed occupies 15% of the pixel area, and 50% and 35%, respectively, for irrigated<sub>2</sub> and irrigated<sub>3</sub> (bottom). Fig. 3b shows the signatures for LAI; the LAI curves are less variable since the daily variability of vegetation activity is relatively small.

Table 2a also shows the search space for each mixing parameters included in the chromosome used in GA. The number of bits indicates the accuracy of the discrete divisions between the minimum and maximum values of the parameters. These discrete values are the ones only represented by the binary representation of a mixing parameter.

2.3.1.4. Simulated “observed” RS data. Processed high temporal resolution data (e.g. NOAA, SPOT-VI, MODIS) are usually available in 10- or 30-day compos-

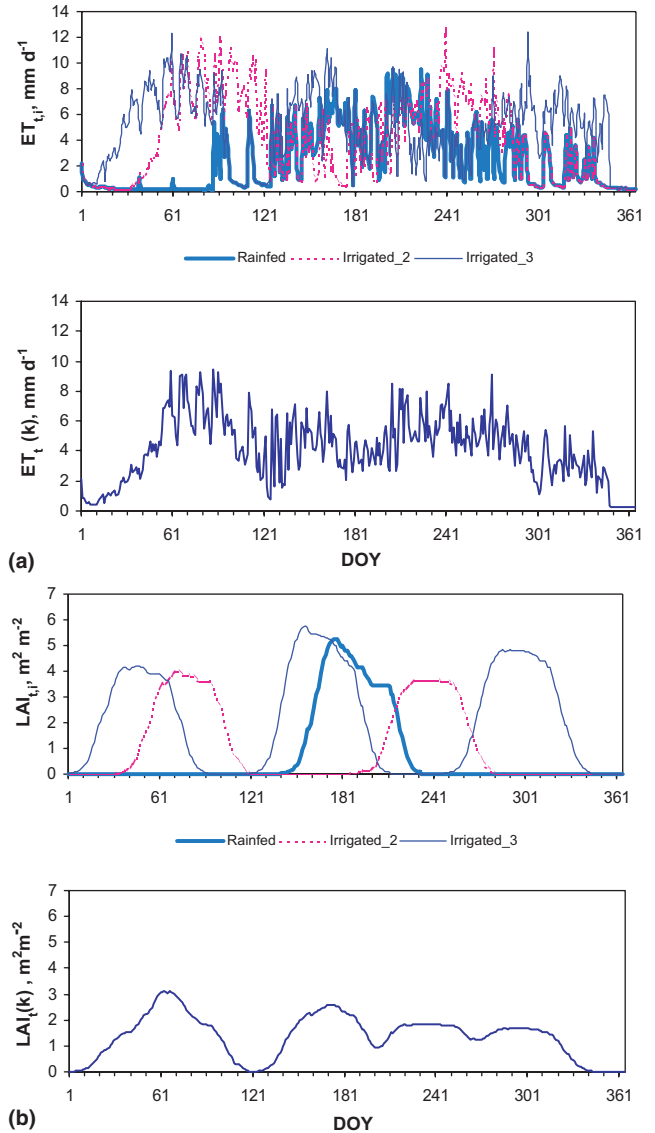


Fig. 3. (a) Generated daily hypothetical ET data to test the mixed-pixel model, (top) the individual ET signature under rainfed, irrigated<sub>2</sub> and irrigated<sub>3</sub> condition, and (bottom) under mixed environment. (b) Generated daily hypothetical LAI data to test the mixed-pixel model, (top) the individual LAI signature under rainfed, irrigated<sub>2</sub> and irrigated<sub>3</sub> condition, and (bottom) under mixed environment.

ite (i.e. the maximum value within the set time interval) to minimize the effect of clouds in the data. To simulate this, we generated two data types from Fig. 3a and b (bottom-plots): (i) every 10-day (ET<sub>10d</sub>, LAI<sub>10d</sub>), and (ii) averaged every 10-day data (ET<sub>10d ave</sub>, LAI<sub>10d ave</sub>). We have used them as the “observed” RS data for the investigated pixel. The subscripts 10d and 10dave indicate the non-averaged 10th day value and the aggregated 10-day values of ET or LAI, which are considered estimates of the 10-day composite data usually available from RS observations.

### 2.3.2. Case studies

We investigated three major case studies in this paper: (i) Case 1, using “observed” RS data (ET and LAI) to determine the agricultural practices and area fractions in the pixel; (ii) Case 2, same as Case 1 but adding water management practices determination; and (iii) Case 3, same as Case 1 but we added a random noise to the “observed” RS data (see Eq. (21)). For Case 3, we generated five data patterns by adding a relative error  $\xi$  (assumed to be 10%) to the “error-free” decadal and 10-day averaged “observed” RS data. The random error term could compose of some system induced errors to the actual RS data such as the effects of cloud, atmospheric condition or estimation errors of LAI and ET.

$$ET' = ET\{1 + x\xi\} \quad (21)$$

where

$$x \sim N(0, 1); -1 \leq x \leq 1$$

For Case 1, we applied both CPM and MPM to the mixed-pixel problem using  $ET_{10d}$  and  $ET_{10d\text{ave}}$  as search criteria, respectively (see Table 2b for GA parameters and treatment combinations). For the case of  $LAI_{10d}$  and  $LAI_{10d\text{ave}}$ , we applied only MPM in the solution—which was further followed for the rest of the case studies. In Case 3, we used all the data types ( $ET_{10d}$ ,  $ET_{10d\text{ave}}$ ,  $LAI_{10d}$  and  $LAI_{10d\text{ave}}$ ) in determining the mixing parameters. In Case 2, we used only  $ET_{10d\text{ave}}$  and  $LAI_{10d}$  as search criteria.

The case studies are partially virtual because the management practices used in the simulations are somewhat related to the actual situations in the field. The relevance of the work stems from the utilization of low spatial resolution data in crop area determination for water and food security applications in the region. Given the ability to determine the spatial extents of agricultural and water management practices would aid considerably agricultural and water managers to plan appropriately the delivery of services to the farmers to maximize the use of available resources and optimize the output from the agricultural system.

## 3. Results and discussion

### 3.1. Solution of the mixed-pixel model

Fig. 4 shows an example of a GA solution to the mixed-pixel problem. The best fitness depicts the measure of the best chromosome in each generation; the average fitness is a trace of the mean value of all the measures of the chromosomes; and the average error shows how the mean difference between the “observed” RS and simulated data, produced by the best chromosome, progresses in time. Based on the condition of optimality, GA stops for a solution. In this case, we have

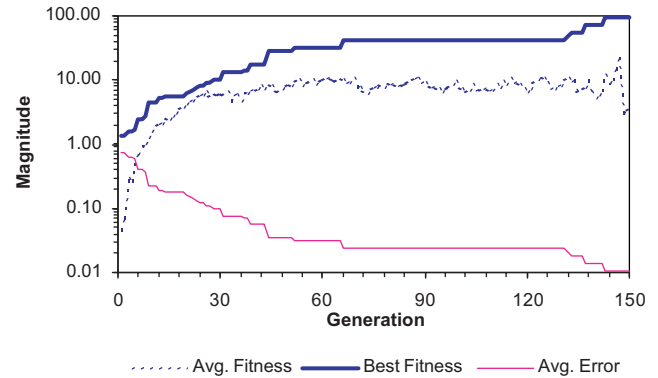


Fig. 4. An example of a GA solution: MPM using  $ET_{10d\text{ave}}$ ,  $n_{\text{pop}} = 30$ ,  $p_{\text{cross}} = 0.5$ ;  $p_{\text{creep}} = 0.1$ .

used the maximum number of generations, i.e. 150 generations.

Notice the prolonged plateau-like response of the best fitness, this is a normal response for GA especially when using a micro-population [21]. Other than the effects of crossover, the elusive solution may not be reached without random perturbations. GA is a greedy technique and tends to eliminate the weaker chromosomes in a short period of time. Note that crossover was strongly supplemented by creep mutation (see the perturbed average fitness) and has paid off after the 131st generation. The micro-population restarted before the 150th generation—the instance when a sudden peak in the average fitness is observed. At this point, most of the chromosomes are nearly similar and have converged accordingly.

#### 3.1.1. Case 1: Agricultural practices and area fractions

Table 3 shows the summary of results for Case 1 when using ET as search criterion. Fig. 5a and b also show the fitness magnitudes and average errors corresponding to each treatment used (see Table 2b). The reason why we applied here both CPM and MPM is to evaluate which method can aid considerably in solving the mixed-pixel problem, in terms of accuracy and speed in finding the best solution. It is evident in the table that MPM shows more promising results than CPM. It appears that the forcing in MPM is stronger, as it amplifies the penalties causing weaker chromosomes to vanish forcefully, hence favoring the propagation of the fitter chromosomes. For all the combinations we tested, a majority of the MPM results outperformed the CPM (Fig. 5a and b). When using  $ET_{10d}$  as search criterion, we found that MPM outperformed its classic counterpart four times while the latter succeeded only once. A best average error of  $0.01 \text{ mm d}^{-1}$  was recorded for MPM in one of the combinations we tested, while only  $0.07 \text{ mm d}^{-1}$  was achieved for CPM. Again, we observed the same performance for MPM when using  $ET_{10d\text{ave}}$  as search criterion. CPM, however, produced



Table 3  
Summary of results to the mixed-pixel problem using evapotranspiration as search criterion

Mixing parameters	Classical penalty method (CPM)				Modified penalty method (MPM)			
	ET <sub>10d</sub>		ET <sub>10dave</sub>		ET <sub>10d</sub>		ET <sub>10dave</sub>	
	Mean <sup>a</sup>	SD	Mean	SD	Mean	SD	Mean	SD
sd <sub>1</sub>	134.33	14.50	137.83	18.85	136.50	9.57	144.33	19.19
sd <sub>2</sub>	23.67	13.00	26.17	10.44	27.50	5.82	33.17	2.04
sd <sub>3</sub>	190.50	5.01	185.33	3.08	186.83	2.23	187.00	2.19
sd <sub>4</sub>	9.67	14.35	8.33	12.52	4.00	7.35	3.33	4.76
sd <sub>5</sub>	128.33	8.45	124.83	10.67	121.83	2.04	123.67	5.20
sd <sub>6</sub>	253.50	10.07	243.50	6.16	250.50	6.16	246.00	4.00
a <sub>1</sub>	0.20	0.05	0.15	0.09	0.17	0.06	0.18	0.05
a <sub>2</sub>	0.48	0.05	0.48	0.16	0.52	0.02	0.45	0.10
a <sub>3</sub>	0.32	0.07	0.36	0.08	0.31	0.07	0.38	0.06

Note: The sums of the means of  $a_i$  at two decimal places are not exactly equal to 1.0 due to rounding errors.

<sup>a</sup> Across treatments.

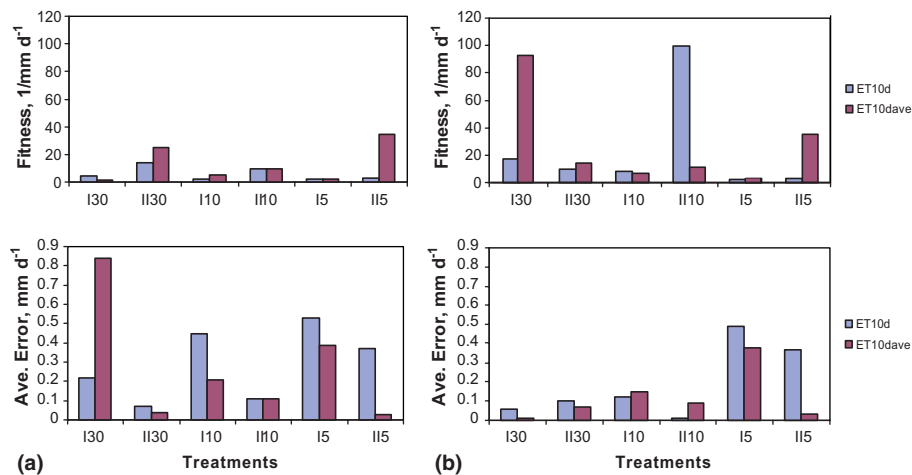


Fig. 5. (a) Fitness (top-left) and average errors (bottom-left) with classical penalty method and (b) fitness (top-right) and average errors (bottom-right) with modified penalty method using ET as search criterion. Note: I30 means  $p_{creep} = 0.1$  and  $n_{pop} = 30$ ; I130 means  $p_{creep} = 0.5$  and  $n_{pop} = 30$ .

a best average error of  $0.03 \text{ mm d}^{-1}$ . Interestingly, we observed the same outcomes regardless of using any fitness form and data type in the un-mixing process for the case of  $n_{pop} = 5$  with  $p_{creep} = 0.5$  (i.e. I15 in Fig. 5a and b).

The above results have practical implications. It is shown that both ET<sub>10d</sub> and ET<sub>10dave</sub> can be used to discriminate agricultural activities in a mixed-pixel environment, hence well-validated ET-RS methodologies e.g. SEBAL [3], ALEXI [28] can be employed to generate spatial ET data for the un-mixing application. Since computational time is precious for this kind of application, it is worth knowing that a small population sized GA is capable of arriving at a reasonable solution for the mixed-pixel problem. This is a step forward for the application of GA in solving actual linear mixture model for operational use.

We have tested the speed of the solutions qualitatively by observing the state of the elite chromosomes at the 100th generation. The question we seek to answer

in this part of the analysis is to whether the forcing in MPM really matters in speeding up the solution. Fig. 6 shows the state of the best chromosomes after the 100th generation for all the combinations we tested, when using ET<sub>10d</sub> and ET<sub>10dave</sub> as search criteria. It is shown in the figures that when  $p_{creep}$  is relatively low, the effect of the forcing in MPM is apparent and appears to be directly proportional to population size. This effect, however, seems to dissipate when a high perturbation rate is introduced to the population. The CPM fares well with MPM, although the latter has a more dominant performance than the former. The order of the results for ET<sub>10d</sub> and ET<sub>10dave</sub> are the same and given by population size: **5**: CPM<sub>01</sub> > MPM<sub>01</sub> > CPM<sub>05</sub>&MPM<sub>05</sub>; **10**: CPM<sub>01</sub> > MPM<sub>01</sub> > CPM<sub>05</sub> > MPM<sub>05</sub>; and **30**: CPM<sub>01</sub> > MPM<sub>05</sub> > CPM<sub>05</sub> > MPM<sub>01</sub>. ET<sub>10dave</sub>, in general, appears to be a more stable search criterion than ET<sub>10d</sub>.

It is clear at this point that a high creep mutation rate adds more speed in searching the solution than restart-

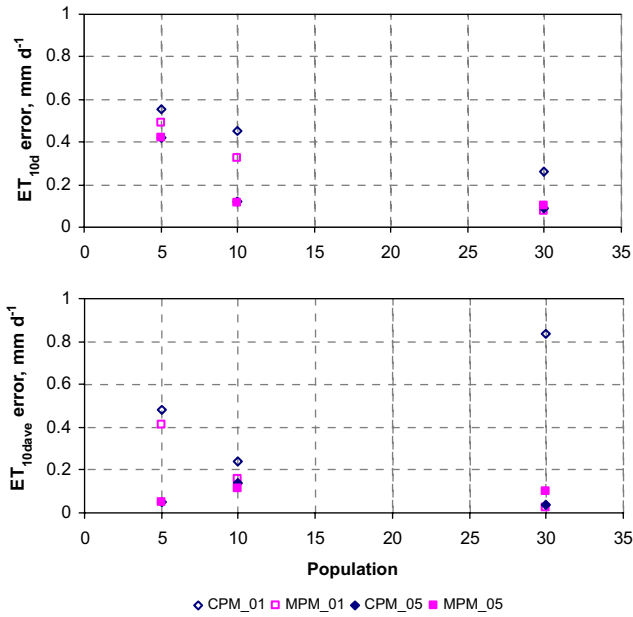


Fig. 6. State of the elite chromosome at the 100th generation (top) for ET<sub>10d</sub> and (bottom) ET<sub>10d\_ave</sub>. Note: CPM\_01 means CPM with  $p_{creep} = 0.1$ .

ing the micro-population. This is true because when  $p_{creep}$  is high, the micro-population seldom restarted within 150 generations. The forcing in MPM also works well in searching for the best results, as evident in Table 3 and Fig. 5. As an overall test, the spread of the solution with MPM is consistently narrower compared to CPM, except for  $sd_1$  when using ET<sub>10d\_ave</sub> as search criterion.

In addition to ET, we have also tested LAI as search criterion in solving the mixed-pixel problem. Nowadays, standard vegetation products, e.g. LAI, are available from satellite data (e.g. MODIS-LAI). Based on the above experiments, a micro-population of five still gives acceptable results while saving calculation time, thus we proceeded to the next step with  $n_{pop} = 5$ . We used only

Table 4  
Solutions of genetic algorithm to the mixed-pixel problem using leaf area index as search criterion

Mixing parameters	LAI <sub>10d</sub>	LAI <sub>10d_ave</sub>
$sd_1$	140	120
$sd_2$	32	31
$sd_3$	186	186
$sd_4$	2	1
$sd_5$	121	129
$sd_6$	246	249
$a_1$	0.16	0.12
$a_2$	0.50	0.52
$a_3$	0.34	0.36
Fitness	33.93	12.68
Error, $m^2 m^{-2}$	0.03	0.08

Note: Using modified penalty method;  $n_{pop} = 5$ ;  $p_{creep} = 0.5$ .

MPM ( $p_{creep} = 0.5$ ) in this case since we have shown above that MPM is more robust than CPM. Table 4 shows the results of the GA solution for both cases of LAI<sub>10d</sub> and LAI<sub>10d\_ave</sub>. It is shown in the table that by using LAI, it is possible still to derive the mixing parameters with good accuracy. Interestingly, LAI<sub>10d</sub> appears to be more stable than LAI<sub>10d\_ave</sub>, compared to ET<sub>10d\_ave</sub> as a better criterion when using ET in the search. Note that DOY 120 is the minimum possible value of  $sd_1$  in the search space (Table 2a), which implies that the fitness function is non-sensitive in this region as the solution yielded still to a competitive average error of  $0.08 m^2 m^{-2}$ . At this point, however, it is difficult to compare which search criterion is better between LAI and ET because the average errors are not in the same units. Nevertheless, the individually derived mixing parameter values (not shown) clearly show that LAI and ET are both desirable. The decision then of using either ET or LAI should be based upon the available RS data and level of expertise in the institution. What is interesting to note is that GA performs better with 10<sub>d</sub> data when using LAI, while with 10<sub>d\_ave</sub> data when using ET. The reason for this is still trivial because average data types are expected to have more recovery capabilities, as they contain a higher memory of the system, thus more likely to perform better in a sparse data series, in this case, 36 data points for one year. This can be true for ET because of its high variability with time, hence the averaging of data points could improve the search, but not necessarily for the smoother LAI curve (see Fig. 3a and b). To support this hypothesis, it can be useful to mention that the plot of LAI<sub>10d\_ave</sub> is like a silhouette of LAI<sub>10d</sub> (not shown) slightly shifted to the right ( $R^2 = 0.93$ ), whereas ET<sub>10d</sub> and ET<sub>10d\_ave</sub> curves appear to be entangled, with  $R^2 = 0.70$ .

### 3.1.2. Case 2: Agricultural and water management practices and area fractions

We extended the model in Section 2.2.1 to include water management practices as mixing parameters, i.e.  $\mathbf{k} = \{sd_{j=1..n}, a_{i=1..m}, w_{i=2..m}\}$  (see Table 2a). Our initial experiments showed that the combination of  $w_2$  and  $w_3$  in the chromosome resulted in a severe sensitivity problem with the fitness function. It seems that they cannot be determined simultaneously in their present forms. We speculated that  $w_2$  and  $w_3$  are highly correlated in real space such that any change in one of them could lead to similar or nearly similar effect on the fitness function causing their identification problem (see [36,22]). To overcome this, we transformed  $w_2$  and  $w_3$  in such a way that they are forced to behave opposite to the real space. The transformations,  $w'_2 = \ln(w_2)$  and  $w'_3 = -\ln(w_3)$ , tested best for the extended mixed-pixel problem.

We have written a separate module for this case study to accommodate the 10 mixing parameters in the chromosomes [as  $a_3 = f(a_1, a_2)$  where  $f$  denotes a function

Table 5  
Solutions of genetic algorithm to the mixed-pixel problem with water management variables

Mixing parameters	ET <sub>10d ave</sub>	LAI <sub>10d</sub>
sd <sub>1</sub>	146	120
sd <sub>2</sub>	32	32
sd <sub>3</sub>	184	186
sd <sub>4</sub>	2	1
sd <sub>5</sub>	121	131
sd <sub>6</sub>	248	249
a <sub>1</sub>	0.16	0.14
a <sub>2</sub>	0.51	0.51
a <sub>3</sub>	0.33	0.35
w <sub>2</sub>	0.78	0.78
w <sub>3</sub>	0.76	0.76
Fitness	12.12	12.70
Error	0.08	0.08

Note: Using modified penalty method; n<sub>pop</sub> = 5; p<sub>creep</sub> = 0.5.

of] and the evaluation of the derived water management practices in the simulations. We used the best search criteria, excerpted from the previous experiments, to evaluate the extended mixed-pixel problem. Table 5 shows the results of the two experiments for LAI<sub>10d</sub> and ET<sub>10d ave</sub>. Apparently, the uncertainty in the solutions has increased with the extended number of unknown parameters. This is manifest particularly with LAI<sub>10d</sub>, which could be attributed in part to the low memory of the 10<sub>d</sub> or non-averaged data, although the sensitivity problem of the mixing parameters could have played a major role in the process. Nevertheless, aside from sd<sub>1</sub> and sd<sub>5</sub>, the other parameters were estimated fairly well. For curiosity sake, we tried to solve the extended linear mixture model with LAI<sub>10d ave</sub> as search criterion and the resulting values for sd<sub>1</sub> and sd<sub>5</sub> have improved to DOY 146 and 121, respectively. The result for ET<sub>10d ave</sub> further demonstrates the strength of averaged data in solving extended mixing problems. Majority of mixing parameters were estimated with good accuracy regardless of the increased number of parameters.

3.1.3. Case 3: With random error

Generally, actual RS data contains errors due to atmospheric conditions, cloud cover, and errors in the RS/relational models used to estimate ET or LAI. For this reason, we tested the procedure assuming that some degree of error is present in the dataset. As mentioned in Section 2.3.2, we generated five datasets for this case study, which we used to solve the linear mixture model formulated in Case 1. The best GA parameters (based on accuracy and efficiency) derived from Case 1 were selected to implement these experiments. They are n<sub>pop</sub> = 10 and p<sub>creep</sub> = 0.5 for ET<sub>10d</sub>, n<sub>pop</sub> = 5 and p<sub>creep</sub> = 0.5 for ET<sub>10d ave</sub>. For LAI, we assumed that it was practical to use n<sub>pop</sub> = 5 and p<sub>creep</sub> = 0.5 for both cases.

Table 6 shows the summary of results of the experiments. We found that even with added random noise in the “observed” RS data it is still possible to derive a conservative solution for the mixed-pixel problem. The quality of the solutions depends greatly on the randomness and severity of error in the dataset. It is interesting to note that LAI performs better than ET in this case study. The reason why LAI is more stable than ET with noisy data is not yet clear but could be attributed to the characteristics of the LAI curve. Fig. 7a and b show the fitness magnitudes and average errors for each pattern when using ET and LAI in the search.

In order to better understand what is happening in the process, we documented the individual daily ET and LAI outputs for all the agricultural land uses produced by the mean values of the solutions of the linear mixture model derived using ET<sub>10d ave</sub> and LAI<sub>10d ave</sub>, as search criteria. Fig. 8a and b show the simulated ET and LAI curves using the mean values in Table 6 versus the ones using the actual agricultural practices in the pixel (see Table 2a). It is clear in Fig. 8a that using ET as search criterion conditions the ET outputs for both the individual ET curves for each agricultural land use and the mixed ET signature. The area fractions for

Table 6  
Summary of results to the mixed-pixel problem using evapotranspiration and leaf area index as search criteria with added random noise

Mixing parameters	ET				LAI			
	ET <sub>10d</sub>		ET <sub>10d ave</sub>		LAI <sub>10d</sub>		LAI <sub>10d ave</sub>	
	Mean <sup>a</sup>	SD	Mean	SD	Mean	SD	Mean	SD
sd <sub>1</sub>	151.60	22.83	139.00	9.57	138.60	5.13	136.60	8.76
sd <sub>2</sub>	32.40	1.14	30.80	4.55	31.40	0.55	31.80	0.45
sd <sub>3</sub>	182.40	4.93	185.20	4.32	185.40	0.89	186.60	2.70
sd <sub>4</sub>	1.20	0.45	5.00	6.20	1.40	0.55	1.20	0.45
sd <sub>5</sub>	121.00	0.00	120.80	7.95	121.05	0.00	121.57	2.83
sd <sub>6</sub>	247.20	1.79	248.20	1.48	248.00	0.71	248.20	0.84
a <sub>1</sub>	0.14	0.02	0.17	0.07	0.16	0.01	0.18	0.05
a <sub>2</sub>	0.51	0.04	0.48	0.04	0.51	0.01	0.49	0.04
a <sub>3</sub>	0.35	0.03	0.35	0.05	0.34	0.01	0.33	0.01

Note: The sums of the means of a<sub>i</sub> at two decimal places are not exactly equal to 1.0 due to rounding errors.

<sup>a</sup> Across patterns.

each land use served as weights and play a significant role in the conditioning, which is apparent for irri-

gated\_2 ( $a_2 = 0.50$ ). The above statements are also true with LAI (Fig. 8b) as search criterion, but with one

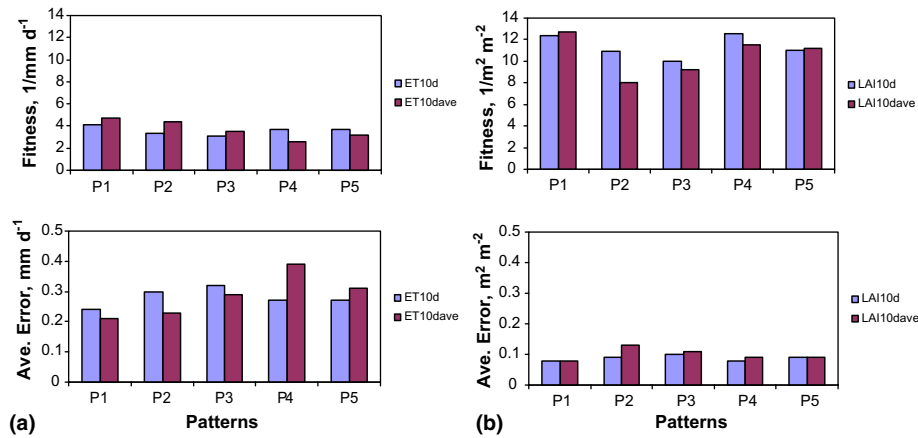


Fig. 7. (a) Fitness (top-left) and average errors (bottom-left) using ET with different noise patterns added to the error-free “observed” RS data. (b) Fitness (top-right) and average errors (bottom-right) using LAI with different noise patterns added to the error-free “observed” RS data.

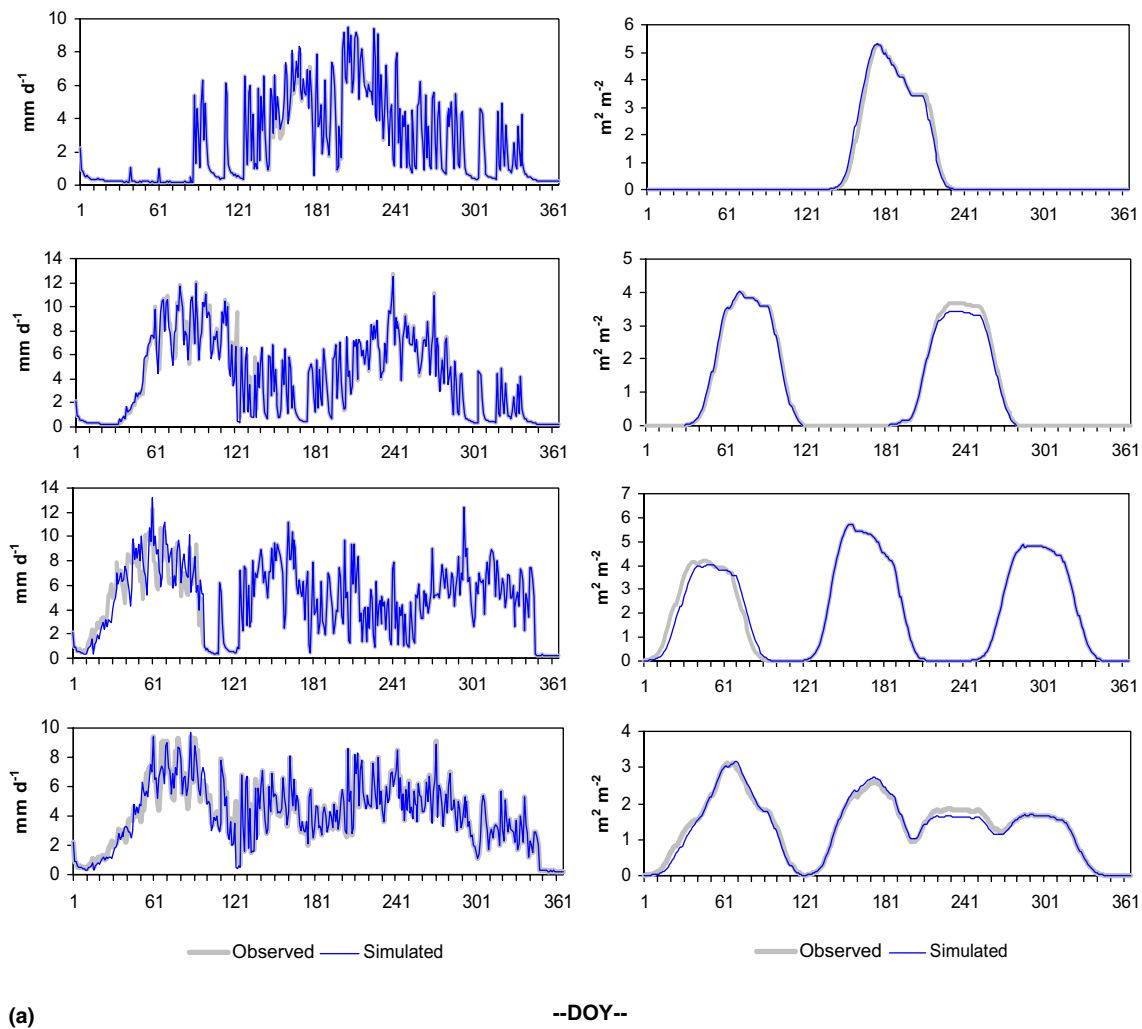


Fig. 8. (a) Simulated versus observed ET (left) and LAI (right) using ET<sub>10dave</sub> with random errors. (b) Simulated versus observed ET (left) and LAI (right) using LAI<sub>10dave</sub> with random errors. From top to bottom: rainfed, irrigated\_2, irrigated\_3, mixed signature.

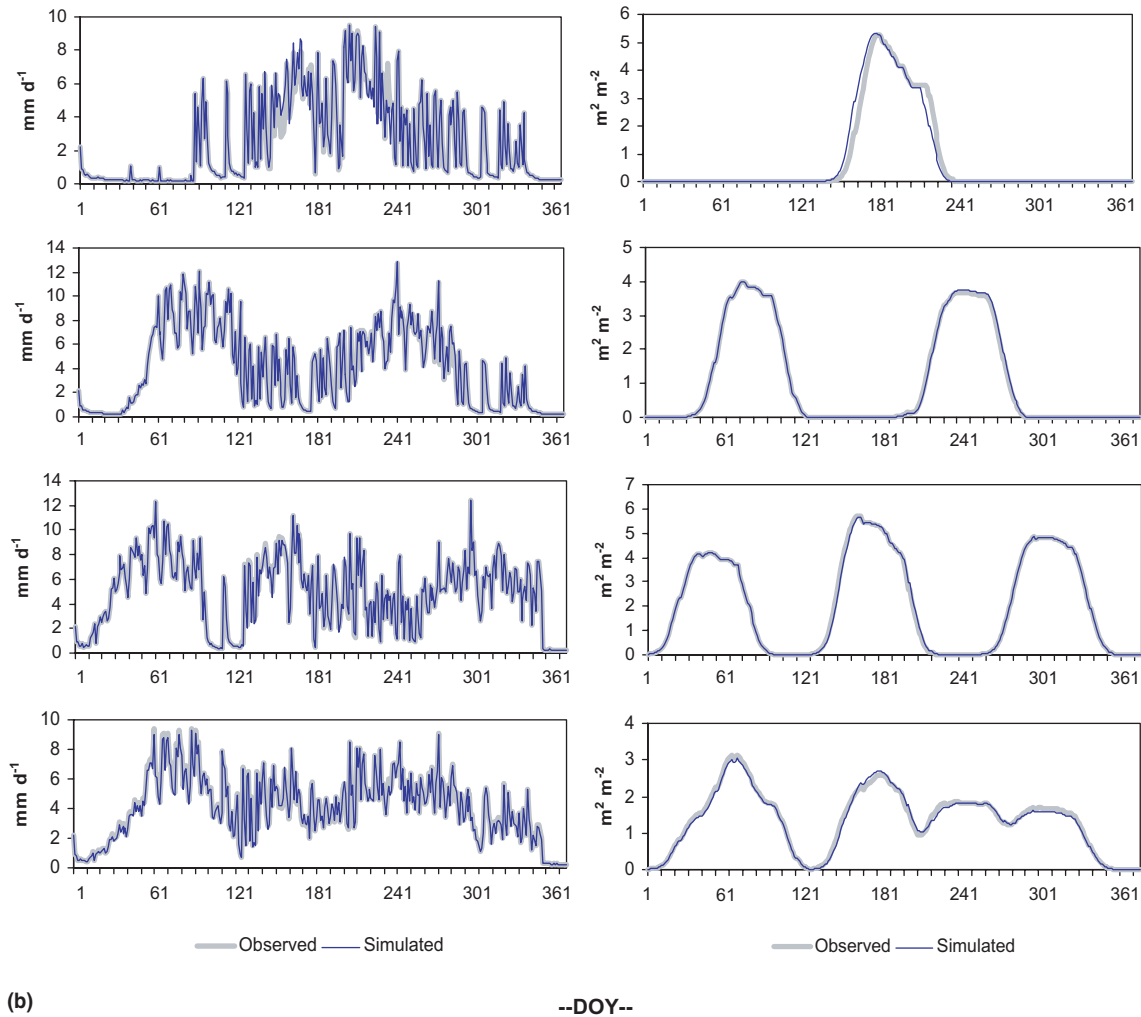


Fig. 8 (continued)

exception—LAI conditions also the ET outputs aside from conditioning itself. This is true because  $ET_{act} = f(LAI)$ . In SWAP, LAI is used to partition  $ET_{pot}$  into two components,  $E_{pot}$  and  $T_{pot}$  (see Section 2.2.2), and then these potential values are reduced according to actual environmental conditions to arrive at a value for  $ET_{act}$ , which is basically the sum of the actual soil evaporation ( $E_{act}$ ) and plant transpiration ( $T_{act}$ ).

#### 4. Conclusions and recommendations

In this paper, we presented a methodology to solve the mixed-pixel problem associated with the use of low spatial resolution RS data for agricultural and water management applications. Our numerical experiments showed the substantial potential of determining both the agricultural and water management practices from low spatial resolution data using the appropriate search criterion and data type to solve the linear mixture model. To minimize the effect of non-uniqueness in the

solution, we recommend conducting exploratory works before the actual applications. A sensitivity analysis of the mixing parameters could reveal parameters that can be removed from the formulation. Priori-information on the existing agricultural and water management practices in the area can be very useful in the process.

We tested two biophysical variables, ET and LAI, as search criteria to solve linear mixture models at varying degrees of complexity and found that both yielded satisfactory results. We also explored the effect of data types in the solution and found some interesting results to improve the effectiveness of the solutions. Shorter memory data type such as non-averaged LAI are more applicable for lesser complex linear mixture models, while higher memory data type such as 10-day average ET are more applicable for more complex linear mixture models e.g. the extended problem in Case 2. Based on the results, however, LAI is preferred for mixed-pixel applications when a high degree of noise is present in RS data.

In this study, the average computational time used to evaluate a chromosome is summarized as follows: one,

two and three seconds for rainfed simulation, irrigated\_2, and irrigated\_3 simulations, respectively. For a micro-population of five, about 1.5 h is needed to complete 150 generations in a Pentium 4 processor, 1.8 GHz speed with 256 MB RAM. It is clear that for larger domain problems, improving the computational time is necessary.

Thus, in line with this effort, we are trying several approaches to circumvent the constraints of computational time. One approach we are exploring is by using cluster computing to implement the procedure presented above where the calculations are done in parallel fashion, either at the pixel-level or within the GA-chromosome evaluations. The AIT cluster computer OPTIMA is being used for this purpose. We are also exploring the possibility of extending the application to the AP-GRID (Asia Pacific GRID) clusters. Another option is by using a heuristic approach where we use a look up table to evaluate the individual chromosomes instead of a dynamic link with the simulation model. These procedures will be reported in future publications.

Finally, we recommend a sensitivity analysis of the penalty coefficient  $\lambda_h$  in the un-mixing algorithm because its optimal value could aid speeding up the search for the solution of the mixed-pixel problem. An optimal  $\lambda_h$  may control the rate of demise of the weaker but above averaged chromosomes that may contain desirable or essential genetic imprints that could hasten the search for a solution. This is particularly important when GA is implemented in a way that it can automatically stop as the best solution is achieved, saving time in the calculations.

## Acknowledgements

This research is partly supported by the Royal Thai Government AIT Joint Research Project on Near Real Time Agriculture Activity Monitoring using Multi-Temporal MODIS Earth Observation Satellite Data, and the International Water Management Institute (IWMI) of Colombo, Sri Lanka as part of their project on Global Irrigated Area Mapping (GIAM) under their Comprehensive Assessment of Water Food and Agriculture initiative.

## References

- [1] Aires F, Rossow WB, Chédin A. Rotation of EOFs by the independent component analysis: toward a solution of the mixing problem in the decomposition of geophysical time series. *J Atmos Sci* 2002;59:111–23.
- [2] Anser GP, Wessman CA, Privette JL. Unmixing the directional reflectances of AVHRR sub-pixel landcovers. *IEEE Trans Geosci Remote Sens* 1997;35:868–78.
- [3] Bastiaanssen WGM, Menenti M, Feddes RA, Holtslag AAM. A remote sensing surface energy balance algorithm for land (SEBAL) 1. Formulation. *J Hydrol* 1998;212–213:198–212.
- [4] Bateson C, Asner GP, Wessman CA. Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis. *IEEE Trans Geosci Remote Sens* 2000;38:1083–94.
- [5] Black TA, Gardner WR, Thurtell GW. The prediction of evaporation, drainage and soil water storage for a bare soil. *Soil Sci Soc Am J* 1969;33:655–60.
- [6] Boesten JJTI, Stroosnijder L. Simple model for daily evaporation from follow tilled soil under spring conditions in a temperate climate. *Neth J Agric Sci* 1986;34:70–5.
- [7] Bouman BAM. Linking physical remote sensing models with crop growth simulation models, applied for sugar beet. *Int J Remote Sens* 1992;13:2565–81.
- [8] Bouman BAM, Van Diepen CA, Vossen P, Van Der Wal T. Simulation and systems analysis tools for crop yield forecasting. In: Teng PS et al, editors. *Applications of systems approaches at the farm and regional levels*. Kluwer Academic Publishers; 1997. p. 325–40.
- [9] Carroll DL. GA Fortran Driver version 1.7. Available from: <<http://www.cuaerospace.com/carroll/ga.html>> 1998.
- [10] Chan-Hilton AB, Culver TB. Constraint handling for genetic algorithms in optimal remediation design. *J Water Resour Plan Manage* 2000;125:128–37.
- [11] D'Urso G, Menenti M, Santini A. Regional application of one-dimensional water flow models for irrigation management. *Agric Water Manage* 1999;40:291–302.
- [12] Defries RS, Hansen MC, Townshend JRG. Global continuous fields of vegetation characteristics: a linear mixture model applied to multi-year AVHRR data. *Int J Remote Sens* 2000;21:1389–414.
- [13] Droogers P. PTF: pedo-transfer function version 0.1. Colombo, Sri Lanka: International Water Management Institute; 1999.
- [14] Droogers P, Bastiaanssen W. Irrigation performance using hydrological and remote sensing modeling. *J Irrig Drain Eng* 2002;128:11–8.
- [15] Droogers P, Bastiaanssen WGM, Beyazgül M, Kayam Y, Kite GW, Murray-Rust H. Distributed agro-hydrological modeling of an irrigation system in Western Turkey. *Agric Water Manage* 2000;43:183–202.
- [16] Goldberg DE. *Genetic algorithms in search and optimization and machine learning*. USA: Addison-Wesley; 1989.
- [17] Holben BN, Shimabukuro YY. Linear mixing applied to coarse spatial resolution data from multispectral satellite sensors. *Int J Remote Sens* 1993;14:2231–40.
- [18] Holland JH. *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press; 1975.
- [19] Ines AVM. Improved crop production integrating GIS and genetic algorithms. Doctoral Thesis, Asian Institute of Technology (AIT), Bangkok, Thailand, 2002.
- [20] Ines AVM, Droogers P. Inverse modeling to quantify irrigation system characteristics and operational management. *Irrig Drain Syst* 2002;16:233–52.
- [21] Ines AVM, Droogers P. Inverse modeling in estimating soil hydraulic functions: a genetic algorithm approach. *Hydrol Earth Syst Sci* 2002;6:49–65.
- [22] Jhorar RK, Bastiaanssen WGM, Feddes RA, Van Dam JC. Inversely estimating soil hydraulic functions using evapotranspiration fluxes. *J Hydrol* 2002;258:198–213.
- [23] Kerdiles H, Grondona MO. NOAA-AVHRR NDVI decomposition and subpixel classification using linear mixing in the Argentinean Pampa. *Int J Remote Sens* 1995;16:1303–25.
- [24] Kool JB, Parker JC. Analysis of the inverse problem for transient unsaturated flow. *Water Resour Res* 1988;24:817–30.

- [25] Krishnakumar K. Micro-genetic algorithms for stationary and non-stationary function optimization. SPIE: Intelligent Control and Adaptive Systems, vol. 1196. Philadelphia, PA, 1989.
- [26] Michalewicz Z. Genetic algorithms + data structures = evolution programs, programs. third ed. USA: Springer; 1996.
- [27] Mualem Y. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour Res* 1976;12:513–22.
- [28] Norman JM, Anderson MC, Kustas WP, French AN, Mecikalski J, Torn R, et al. Remote sensing of surface. *Water Resour Res* 2003;39:1221.
- [29] Oleson KW, Sarlin S, Garrison J, Smith S, Provete JL, Emery WJ. Unmixing multiple land-cover type reflectances from coarse spatial resolution satellite data. *Remote Sens Environ* 1995;54:98–112.
- [30] Sakthivadivel R, Thiruvengadachari S, Amerasinghe U, Bastiaanssen WGM, Molden D. Performance evaluation of the Bhakra irrigation system, India, using remote sensing and GIS techniques. Research Report 28. Colombo, Sri Lanka: International Water Management Institute; 1999. 22p.
- [31] Sarwar A, Bastiaanssen WGM, Boers ThM, Van Dam JC. Evaluating drainage design parameters for the fourth drainage project, Pakistan by using SWAP model: Part I—calibration. *Irrig Drain Syst* 2000;14:257–80.
- [32] Shimabukuro YE, Smith JA. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. *IEEE Trans Geosci Remote Sens* 1991;29:16–20.
- [33] Supit I, Hooijer AA, VanDiepen CA, editors. System description of the WOFOST 60 crop growth simulation model. Brussels, Luxembourg: Joint Research Centre, Commission of the European Communities; 1994.
- [34] Takeuchi W, Tamura M, Yasuoka Y. Estimation of methane emission from West Siberian wetland by scaling technique between NOAA AVHRR and SPOT HRV. *Remote Sens Environ* 2003;85:21–9.
- [35] Takeuchi W, Yasouka Y, Tamura M. Estimation of methane emission from paddy fields in Central Plain of Thailand by scaling technique. In: Proceedings of the 22nd Asian Conference on Remote Sensing, Singapore, 2001. p. 1–6.
- [36] Van Dam JC. Field-scale water flow and solute transport. SWAP model concepts, parameter estimation and case studies. Doctoral Thesis, Wageningen University, The Netherlands, 2000.
- [37] Van Dam JC, Huygen J, Wesseling JG, Feddes RA, Kabat P, Van Walsum PEV, et al. Theory of SWAP version 2.0: Simulation of water flow and plant growth in the soil–water–atmosphere–plant environment. Technical Document 45, Wageningen Agricultural University and DLO Winand Staring Centre, The Netherlands, 1997.
- [38] Van Genuchten MTh. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci Soc Am J* 1980;44:892–8.
- [39] Wesseling JG, Kroes JG. A global sensitivity analysis of the model SWAP. Report 160, DLO Winand Staring Centre, Wageningen, The Netherlands, 1998.
- [40] Yang CM, Su MR. Seasonal variations of reflectance spectrum and vegetation index in rice vegetation cover. In: Proceeding of the 3rd Asian Crop Science Conference. Taichung, Taiwan: Chinese Society of Agronomy; 1998. p. 574–93.
- [41] Yang CM, Su MR. Correlation of spectral reflectance to growth of rice vegetation. In: Proceedings of the 19th Asian conference on Remote Sensing, Manila, Philippines, 1998. p. 1–3.