Proceedings of the *FOSS/GRASS Users Conference* - Bangkok, Thailand, 12-14 September 2004

Genetic algorithm for assimilating remotely sensed evapotranspiration data using a soil-water-atmosphere-plant model. Implementation issues.

Yann Chemin*, Kiyoshi Honda*, Amor V. Ines**

* STAR program, Asian Institute of Technology, 58 Moo 9, Km. 42, Paholyothin Highway, Klong Luang, Pathum Thani 12120, Thailand, tel. +6625160110, fax +6625162126, e-mail yann.chemin@ait.ac.th, honda@ait.ac.th

** International Research Institute for Climate Prediction, Columbia University, 61 Rt. 9W, Palisades, NY 10964-8000, USA, tel. +18456804468 fax +18456804688 e-mail ines@iri.columbia.edu

1 Introduction

Agricultural monitoring is necessary for efficient food security management at country level. Typically, monitoring requirement from the point of view of an agricultural/irrigation manager would be to “see” each field at a regular interval to which 15 days is reasonable. Evapotranspiration (ETa) is converting the water into crop, and is therefore a crucial indicator of crop productivity. ETa can be estimated from satellite remote sensing [1] [2] [3]. However, on the side of satellite platforms specifications, high spatial resolution is at about size of the largest fields (~1 ha), but is available only few times a year practically, while low spatial resolution is available daily (even 8days composites are ready from Internet).

A potential solution would be to match the two type of satellite images evapotranspiration by running instances of crop models at both resolutions with proper parameters. Those crop model input parameters are changing on pixel-to-pixel basis, therefore using a data assimilation method is most interesting to try and solve this problem. Because the search domain of this assimilation problem is multi-dimensional and highly non-linear, using an evolutionary search algorithm like the genetic algorithm (GA; [4]) is preferred. Similar work by [5] and [6] used some remotely sensed information combined with GAs and the Soil Water Air Plant model (SWAP) in the objective of optimization of soil hydraulic parameters.

This paper describes the implementations issues of the program [GA+SWAP] whereby the ETa data from remote sensing images could be used to generate the ETa data during the periods when satellites are not available. SWAP is used by a GA to estimate pixel-based plant/water parameters controlling the pixel ETa observed by the satellite images. Two open source options have been investigated for geographical linkage, the first one being GRASS GIS [7] and the second one being a remote sensing image handling library [8].

2 Methodology

As found in Figure 1, the ETa dataset is created from the processing of the Surface Energy Balance Algorithm for Land (SEBAL, [9]) and actual Solar radiation correction. The ETa data to be used as target for the optimization process is made of 12-24 images of Terra-Modis and 2-3 images of Landsat/Aster.

This target ETa dataset has to be matched by the ETa output of successive optimization runs of SWAP model. The SWAP model is a one-dimensional transient model to simulate water flow in a heterogeneous soil-root system, which can be under the influence of groundwater [11][12]. It has been recently modified to include solute transport, heat flow and crop growth in the air-plant-soil environment [13]. The SWAP input parameters under optimization are the starting date of cropping, the time extent of cropping and the groundwater depth in 1st January and in 31st December. It is expected to use rice pixels with double cropping as a case study.

The Genetic Algorithm will feed the newly proposed parameters to SWAP according to the evaluation of the difference between SWAP output ETa values and the target ETa values. The genetic algorithm [10] used in this study is a simple implementation of the processus of evolution. Genetic algorithms deals with the evolution of genes from individuals belonging to a given population. Changes in the genes of the individuals from a given population permits the selection of certain group of genes that are most important in fitting the environment pressure on the population. In this study, the environment pressure is the SWAP model ETa output that has to match the remote sensing ETa target. When a minimum-difference defined threshold will be reached, SWAP parameters will be stored for reconstruction of high spatial ETa for any required day in the cropping season. The search domains for the dates of starting of cropping will require a non-overlapping restriction of about 90-100 days for soil preparation essentially. The time extent of the cropping season will be between 3 to 5 months. The groundwater level maybe ranging from 0 to 500cm depth but for the purpose of the case study it may be narrowed according to some general information about the area in order to improve the time efficiency for convergence.

Consider F the fitness function (inverse of the cost), having (*x*,*y*,*d*,*p*) parameters, *x* the *longitude [0-180/E-W]*, *y* the *latitude [0-90/N-S]*, *d* the *date [yyyymmdd]* and *p* the *pixel size [90/1000]*. The fitness of an individual having *xydp* characteristics will be the inverse of the cost function aiming at minimizing the differences between SWAP simulation and target ETa, i.e. $F_{xydp} = 1 / Eval_{xydp}$:

$$F_{xyd[90;1000]} = \frac{1}{\frac{1}{n}\Sigma_e^f \left(ETa_{XYD1000} - \frac{1}{m}\Sigma_i^j ETa_{SWAP\ xyd90} \right)^2} + \frac{1}{\frac{1}{mq}\Sigma_i^j \left(ETa_{xyd90} - ETa_{SWAP\ xyd90} \right)^2}$$

With the value of parameter A at p=1000 being $A_{XYD1000}$ and its value at p=90 being A_{xyd90}, $(x,y,d) \in \{[(x_{1000}-500);(x_{1000}+500)],[(y_{1000}-500);(y_{1000}+500)],[i,...,j]\}$ and *m* being the count of small pixels in one large pixel, *n* being the number of dates for the large pixels $\in$ [e,...,f] and *q* the number of dates for the small pixels $\in$ [i,...,j].

The geographical software should give access to specific pixels at different resolutions and provide them to the [GA+SWAP] program.

3 Implementation in GRASS GIS

Initial implementation was done only for optimization of a large pixel (Figure 1), without considering the dimension of small pixels [14]. It was taking 30-45 minutes to optimize one pixel. At that point a GRASS GIS model (r.gaswap [15]) was developed to provide pixels input to the model as a proper GIS modelling functionality, in the hope that further GIS functionalities from within GRASS GIS could be used to develop the final model with combined resolutions. One of the functions that was of interest at that point in time, was the g.region. The need is to get one large pixel (i.e. Modis) at a certain location for many dates, for that same large pixel location, a set of small pixels from other satellites (i.e. Landsat, Aster) should be selected within the area encompassed into the area of the large pixel. The idea is to provide a location of a center of a large pixel, collect their values, then do a g.region resolution transform to collect all small pixels corresponding to the large one. Additional GIS/RS functionalities from the environment of GRASS GIS would be convenient in the process of database preparation and handling for the pixels to be accessed and new images written. However, the development went as far as porting the [GA+SWAP] program for only the large pixels optimization.

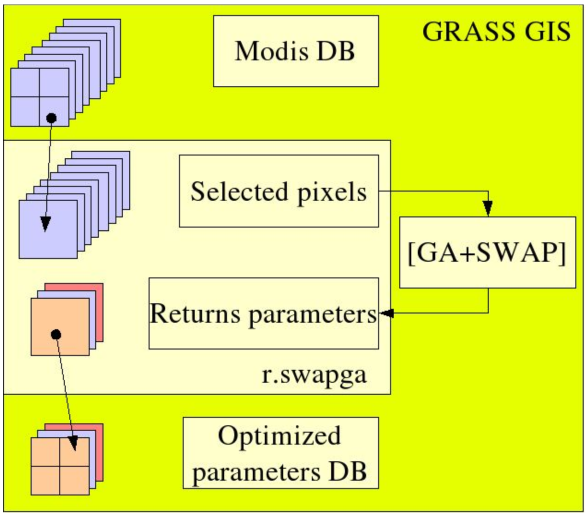

Figure 1. GRASS GIS implementation

The main loop of the GRASS GIS bridge (r.gaswap) is following here. GIS/RS input parameters are the soil textures (line 1-10) and the ETa from n satellites (Line 12-19). The [GA+SWAP] program (Line 21) is run externally by the system() command, assimilates the parameters and returns a text file (grass_cell.txt, line 23). Main ouputs are the assimilated variables (NVAR, Line 24), each variable being directed to its own GRASS GIS layer (Line 28):

```
1       soilFile = fopen("soil.txt","w")
2       f = ((CELL *) inrast_psand)[col];
3       fprintf(soilFile,"%lf \n",f);
4       f = ((CELL *) inrast_psilt)[col];
5       fprintf(soilFile,"%lf \n",f);
6       f = ((CELL *) inrast_pclay)[col];
7       fprintf(soilFile,"%lf \n",f);
8       f = ((CELL *) inrast_pomat)[col];
9       fprintf(soilFile,"%lf \n",f);
10      fclose(soilFile);
11      /***********************************/
12      satFile = fopen("satellite.eta","w")
13      for(j=1;j<nfiles+1;j++)
14      {
15              f = ((FCELL *) inrast[j])[col];
16              f = f_calc(f); /* convert to cm/day */
17              fprintf(satFile,"%5.3f \n",f);
18      }
19      fclose(satFile);
20      /***********************************/
21      system("./gaswap");
22      /***********************************/
23      if((swapFile=fopen("grass_cell.txt","r"))==NULL)
        24      for (j=1;j<NVAR+2;j++)
25      {
26              fscanf(swapFile,"%f",&fe[j]);
27              printf("%f\n",fe[j]);
28              ((FCELL *) outrast[j])[col] = fe[j];
29      }
30      fclose(swapFile);
31      break;
```

It became immediately evident that it would become problematic to run more than a dozen of pixels at a time on a single computer because of the inherent processing time needed for each pixel. Therefore another processing direction was open, besides the development of the model itself.

4 Implementation in image handling library

[16] have investigated the distributed processing of the model with large pixels only (Figure 2) in a Beowulf-style cluster (www.optima.ait.ac.th). Early reports are encouraging.

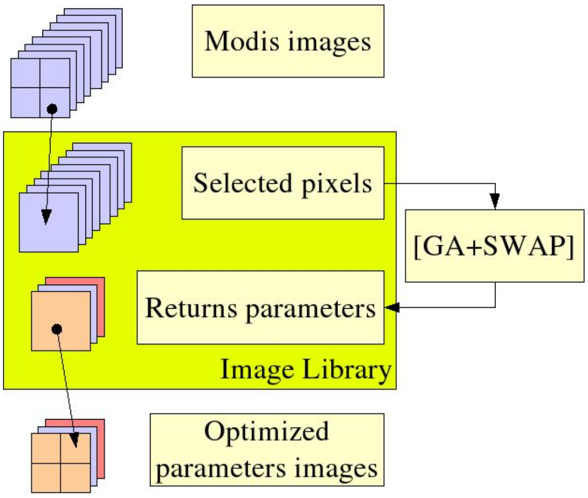

Figure 2: Image handling library implementation

However, the complete program (with two spatial resolutions) would likely require a large size cluster to process a medium size area within short time spans. Since using GRASS GIS environment in Cluster computers is not documented and maybe cumbersome, a small-size, portable image library [8] will be preferred once the main model will be completed and reimplemented over a distributed programming language. Initial directions [17] are 1) process one pixel at a time by distributing each genetic chromosome to each node at the same time, and 2) run one pixel per node.

5 Conclusion

While this methodology is still not applied, the scope of research has been well defined, and the preliminary analysis for one spatial resolution is encouraging. A bridge to GRASS GIS was made, and worked successfully but the processing time involved was too long on a stand-alone computer when considering more than a dozen of pixels. A parallel computing development has started with a small GPL image handling library.

Implementation of the methodology with two spatial resolutions is still in early stage (programming and processing data) at the time of the writing of this paper.

References

[1] Kustas, W.P., and Norman, J.M. 1996. Use of remote sensing for evapotranspiration monitoring over land surfaces. Hydrological Sciences – Journal – des Sciences

Hydrologiques, 41(1):495-516.

[2] Bastiaanssen, W.G.M., 1998. Remote sensing in water resources management: the state of the art. International Water Management Institute (IWMI), Colombo, Sri Lanka.

[3] Menenti, M., 2000. Evaporation. In Remote Sensing in Hydrology and Water Management, Gert A. Schultz and Edwin T. Engman (Eds.), Springer Verlag, Berlin, Heidelberg, New York: 157-188.

[4] Goldberg, D.E., 1989. Genetics Algorithms in Search, Optimization, and Machine Learning. Reading, MA: Addison-Wesley.

[5] Ines, A.V.M., 2002. Improved crop production integrating GIS and Genetic Algorithms. Doctoral Thesis. AIT, Bangkok, Thailand. 236pp.

[6] Ines, A.V.M., and Droogers, P., 2002. Inverse modeling in estimating soil hydraulic functions: a Genetic Algorithm approach. Hydrology and Earth System Sciences. 6:49-65.

[7] http://grass.itc.it/

[8] Honda, 1995. Image handling library. GPL>=2 license. http://www.star.ait.ac.th/~honda/textbooks/advdip/imagelib.zip

[9] Bastiaanssen, W.G.M., 1995. Regionalization of surface flux densities and moisture indicators in composite terrain. A remote sensing approach under clear skies in Mediterranean climates. Agricultural Research Department, Report 109, Wageningen, The Netherlands.

[10] Michalewicz, Z., 1996. Genetic algorithms + data structures = evolution programs. 3rd reviewed and extended edition. Springer-Verlag, Berlin Heidelberg.

[11] Feddes, R.A., Kowalik, P.J., and Zaradny, H., 1978. Simulation of water use and crop yield. Simulation Monography, Pudc, Wageningen, 188pp.

[12] Belmans, C., Wesseling, J.G., and Feddes, R.A., 1983. Simulation of the water balance of a cropped soil: SWATRE. Journal of Hydrology, 63:217-286.

[13] Van Dam, J.C., Huygen, J., Wesseling, J.G., Feddes, R.A., Kabat, P., and Van Walsum, P.E.V., 1997. Simulation of transport processes in the Soil-Water-Air-Plant environment. SWAP user's manual. DLO-Winand Staring Centre, Wageningen, The Netherlands.

[14] Chemin, Y., Honda, K., Ines, A.V.M., 2004. Real coded genetic Algorithm for assimilating remotely sensed evapotransiration data using a soil-water-atmosphere-plant model. A methodology. In proceedings of the AFITA/WCCA Conference, Bangkok, Thailand, 2004.

[15] http://203.159.10.13/~yann/coding/r.gaswap.tar.gz GPL license.

[16] Akhter,S., Honda, K., Chemin,Y., Uthayopas, P., 2004. Input assimilation of Soil

Water Atmosphere and Plant (SWAP) Model with GA using Cluster Computer processing. In proceedings of ACRS 2004, Chiang Mai, Thailand. Submitted.

[17] Akhter, S. 2004. Implementation of SWAP-GA model in cluster computing. M.Sc. Proposal, Asian Institute of Technology, Khlong Luang, Thailand.