Parameter-less GA based Crop Parameter Assimilation with Satellite Image on High Performance Computing

KEIGO SAKAMOTO,[†1] AKHTER SHAMIM,[†1] KENTO AIDA[†1,†2] and CHEMIN YANN[†3]

1. Introduction

Monitoring of country level agricultural activities is now necessary to ensure the food security problem. The monitoring data, such as satellite images, can be used to estimate growth of agricultural products by using sophisticated estimation models proposed in the agricultural community, e.g. the SWAP (Soil, Water, Atmosphere, Plant) model[1]. However, satellite images do not provide complete information, or crop parameters, required by the sophisticated estimation model.

The SWAP Double Layer GA (SWAP-DLGA)[2] is one of methods to solve this problem. SWAP-DLGA assimilates missing crop parameters collected from satellite images by running the genetic algorithm (GA). A problem in SWAP-DLGA is that the user needs to define suitable parameters for running GA, or GA parameters such as population size and maximum generation, in advance. However, the suitable GA parameter needs to be found empirical way, and it is not realistic assumption that the user in the agricultural community can find the suitable GA parameter.

This paper proposes a method to run SWAP-DLGA with automatic generation of suitable GA parameters. The proposed method computes the suitable GA parameters using the Parameter-less GA (PLGA)[3] and runs SWAP-DLGA with the computed GA parameters. Both SWAP-DLGA and PLGA require huge computational time. This paper also presents implementation of the proposed method on a distributed computing system (DCS), or a cluster of clusters, to reduce the computation time.

2. Background

This section outlines our target application.

†1 Tokyo Institute of Technology
†2 National Institute of Informatics
†3 International Rice Research Institute

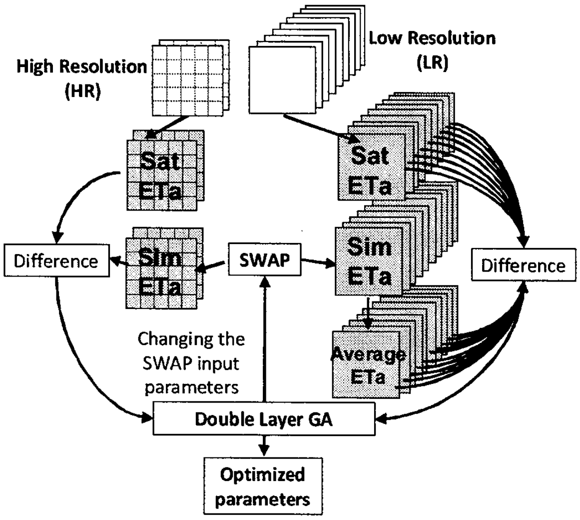

Fig. 1 DLGAmodel

2.1 SWAP-DLGA

The idea of SWAP-DLGA is assimilating missing information in satellite images by running GA. The evapotranspiration (ETa), which is obtained both from satellite images (SatETa) and results from running GA (SimETa), is used to evaluate the fitness. SWAP-DLGA uses two kinds of satellite images, daily obtainable Low Resolution (LR) and not daily obtainable High Resolution (HR). SWAP-DLGA runs two GAs in hierarchical manner. We call them the outside GA and the inside GA. The role of the outside GA is to recreate the new individual and the role of the inside GA is to evaluate all individuals and get their fitness values. Good individual, that has smaller difference between SatETa and SimETa, is selected for next generation.

2.2 PLGA (Parameter-less GA)

The idea of PLGA is searching suitable GA parameters using some rules. PLGA works well without concerning the setting of suitable GA parameters. Selection pressure s and crossover probability P_c are preset to fixed effective values ($s = 4, P_c = 0.5$) and mutation operator

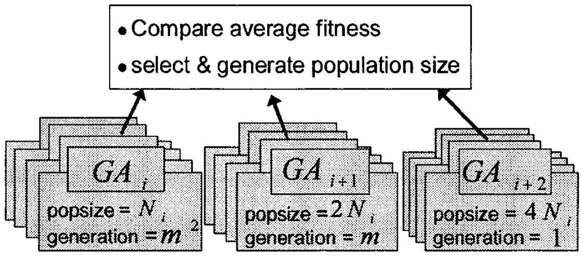

Fig. 2 The model of PLGA

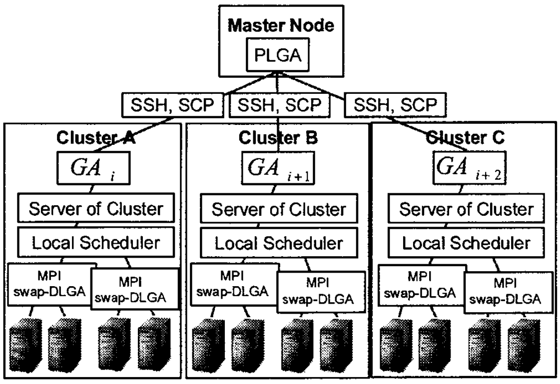

Fig. 3 Implementation scheme on DCS

is ignored to ensure the convergence of populations. PLGA runs some GAs with the above values simultaneously and these GAs maintain the following rules. When the GA_i (i is the GA indicator) sets the population size (popsize) as N_i, the GA_{i+1} sets the population size as $2N_i$, and the GA_i is allowed to run for m (m is called the counter and $m \geq 2$) times more generations than the GA_{i+1}. When the average fitness of GA_i is overtaken the average fitness of GA_{i+1}, the GA_i is deleted and the $GA_{i+n} (n \geq 2)$ is generated. In this way, population size is updated.

3. Experiments

We implemented the proposed method on the distributed computing system illustrated in **Fig. 3**. We run PLGA to find suitable GA parameters for the outside GA in SWAP-DLGA. Our preliminary experiments show that the performance of the inside GA is not significantly affected by selection of GA parameters. The proposed algorithm is parallelized using the master-worker model[4] and implemented using ssh and MPI. A master node gathers results of all GAs and compares their average fitness, then decides whether to continuously run GA_i or to eliminate GA_i and to generated GA_{i+3}. **Table 1** shows computing resources used in the experiments, where 50 cores in each cluster were used to run the proposed method. In the initial setting, the population size N_0 and the counter m are $N_0 = 25$ and $m = 2$, respectively. **Fig. 4** presents comparison between ETa and SimETa. From the result, the accuracy of assimilation with PLGA is better than the one with Simple SWAP-DLGA. Distributed SWAP-DLGA with PLGA was executed in 9,842 seconds and got 75% performance efficiency in computation time.

Table 1 The specifications of clusters

Machine	Specification
suzuk	Core2Duo 2.13GHz, 36nodes
tsukuba	Xeon E5410 2.33GHz ×2, 16nodes
kyoto	Core2Duo 2.13GHz, 26nodes

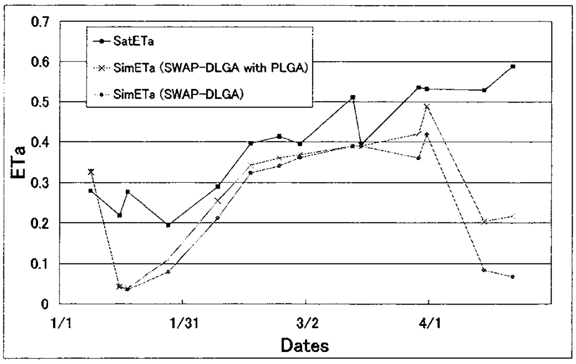

Fig. 4 Simulated ETa against LR pixel ETa

4. Conclusions

This paper proposed SWAP-DLGA with PLGA and its implementation scheme on DCS. This methodology enables to get better crop activity parameters without concerning the setting of the suitable parameters within acceptable time.

Acknowledgment

The work presented in this paper is partially supported by the MEXT Grant-in-Aid for Scientific Research on Priority Area, "Design and Development of Advanced IT Research Platform for Information Explosion Era".

References

1) http://www.swap.alterra.nl/
2) Y. Chemin, and K. Honda, "Spatio-temporal Fusion of Rice Actual Evapotranspiration with Genetic Algorithms and an Agro-hydrological Model," *IEEE Trans on Geogscience and RS*,Vol.44, No.11, 2006
3) G. Harik, and F. Lobo, "A parameter-less genetic algorithm," *GEC conf-99* , July 13-22, Florida(1999)
4) S. Akhter, K. Osawa, M. Nishimura, K. Aida, "Experimental Study of Distributed SWAP-GA Models on the Grid," *IPSJ Trans*, Vol.1 No.2 193-206 (2008)