Journal of Quality Engineering and ProductionOptimization

Vol. 2, No. 1, PP.47-64, 2017

Scheduling a constellation of agile earth observation satellites with preemption

Saeed Hosseinabadi, Mohammad Ranjbar*, Sepehr Ramyar, Masoud Amel Monirian
Department of Industrial Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

***Corresponding Author:** Mohammad Ranjbar (E-mail: m_ranjbar@um.ac.ir)

Abstract- *In this paper, we consider a scheduling problem for a set of agile Earth observation satellites for scanning different parts of the Earth's surface. We assume that preemption is allowed to prevent repetitive images and develop four different preemption policies. Scheduling is done for the imaging time window and transmission time domain to the Earth stations as well. The value of each picture from different target regions and the limitations of the satellite constellation in terms of memory and energy cause high computational complexity for this problem and thus obtaining an optimum solution with a deterministic method is very time-consuming. Consequently, a genetic-based metaheuristic algorithm with a specific solution representation is developed in order to maximize the total value of the observation process by establishing heuristic rules in the initial population of this algorithm. Comparison of the results from the proposed model with the results of cases where repetition of observed areas is not ignored indicates that the proposed model can bring about a significant increase in profits in the planning horizon.*

Keywords*: Scheduling, agile Earth observation satellite, preemption, genetic algorithm.*

I. INTRODUCTION

The operation of Earth observation satellites is short-lived and thus the development and launching of these satellites is very costly. This necessitates an efficient utilization of this technology to guarantee the maximum return on the invested capital. Consequently, the scheduling of a satellite's mission is of considerable significance.

The scheduling of observation satellites can be divided into two major categories. The first is the former problem of scheduling a single observation satellite. But due to the increasing demand for satellite pictures and technological advances, the newer problem of scheduling a multitude of satellites in the form of a constellation has recently been posed and is the subject of this paper.

Another criterion for categorization of the observation satellites scheduling problem can be the type of the observation satellites. Before the advent of multi-directional satellites, single-directional satellites with only a single degree of freedom were used to image the Earth's surface. These satellites could only image the areas along their course and thus the start time of the process for observation of an area was predetermined. This transformed the satellite scheduling problem into a problem of selecting the areas for imagery. Multi-dimensional satellites, however, are capable of moving in three different directions. This enables the observation of various areas within a pre-specified region. The three movements of a satellite consist of rolling, pitching, and rotation. But because of the energy-inefficient and time-consuming nature of the rotational movement, this direction is virtually not used. There is a maximum value for each pitching or rolling movement of the satellite. Such satellites with three degrees of freedom are known as agile or multi-dimensional satellites. Figure 1 illustrates the rolling and pitching angle of an agile satellite.

This paper presents a scheduling scheme for the observation and transmission activities of a constellation of agile satellites in the presence of a preemption policy that is unprecedented in the literature. Given the limitations of this type of satellite, such as memory level, energy, and the narrow observation angle, this paper explores the scheduling of the imaging and transmission process in order to maximize the benefits, taking into account the value of each image. The proposed model takes account of spot and polygon targets and develops a metaheuristic method based on a genetic algorithm (GA) for the solution.

Manuscript Received: 16-June-2016 and revised: 02-Sep-2016
Accepted: 30-Oct-2016
ISSN: 2423-3781

The contribution of this paper is twofold: (a) considering the preemption of the observation process for a specific area that can be imaged by another satellite in order to avoid repetitive images while saving energy and memory; (b) presenting a GA using heuristic rules for the creation of an initial population for estimating the optimal solution.

The structure of this paper is as follows. Section 2 reviews the literature on the work. Section 3 describes the problem statement. Section 4 explores the solution method and the results are discussed in section 5. Ultimately, concluding remarks are drawn in section 6.

II. LITERATURE REVIEW

In one of the earliest studies for scheduling a single satellite, Gabrel et al. (1997) investigated the scheduling of an agile observation satellite. They used a heuristic graph theory method to maximize the number of images taken. Classification of targets based on the rolling angular limitation of the satellite camera for scanning into two categories of polygon and spot areas was introduced by Lemaître et al. (2002).

Using three heuristic methods (greedy search, constraint programming, and local search) and a deterministic dynamic programming approach, they maximized the profits gained from the imagery. Wang and Reinelt (2011) used a metaheuristic Tabu search algorithm combined with a partial enumeration method to maximize the profit function of an agile satellite. In a more recent study, Tangpattanakul et al. (2015) studied the scheduling of an agile satellite by maximizing the total profit of the selected acquisitions and simultaneously ensuring the fairness of resource sharing by minimizing the maximum profit difference between users. They used a multi-objective local search algorithm to solve the problem. Wang et al. (2016) considered satellite scheduling where observations may be significantly affected and blocked by clouds. They used the notion of forbidden sequences and developed a novel assignment formulation for satellite scheduling. To solve the problem, they used the idea of chance constraint programming and a branch-and-cup algorithm. Finally, Xu et al. (2016) studied an Earth observation scheduling problem from China's satellite platform. They developed priority-based indicators based on a cost–benefit analysis method. They employed a sequential construction procedure to generate feasible solutions and evaluated the performance of the proposed algorithms in various scenarios.

Some researchers have focused on the scheduling of a constellation of satellites instead of a single satellite. In an early study, Frank et al. (2001) investigated the scheduling of a constellation of single-directional satellites using a probabilistic greedy search, and defining heuristic criteria to maximize the number of high-priority requests for imagery. On the subject of scheduling a constellation of multi-directional observation satellites, Florio (2006) developed a heuristic algorithm to maximize the number of imaged areas and minimize the response time and transmission time for requests.

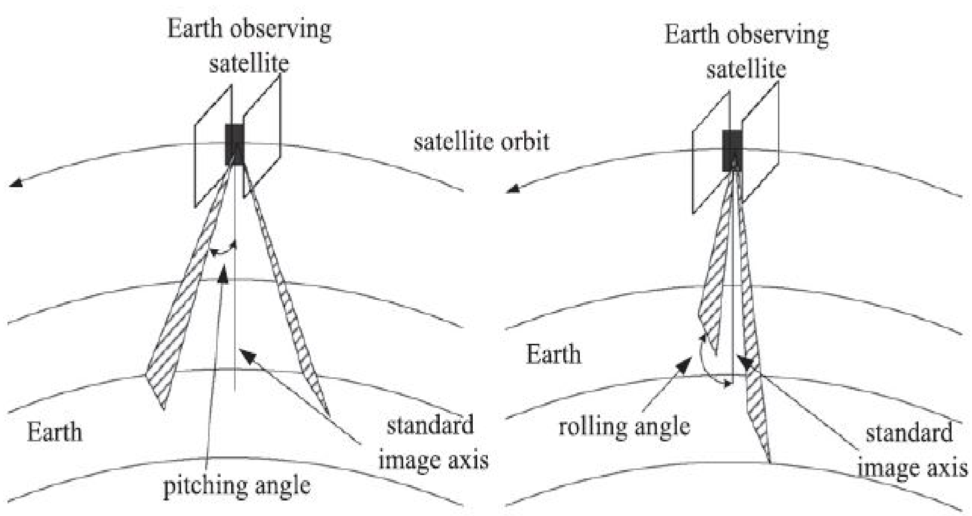

Fig. 1. Rolling and pitching angle for a satellite (Gabrel et al. 1997)

In another study, Bianchessi and Righini (2008) developed a constructive heuristic algorithm with a prediction capability for selecting and synchronizing the operations needed to acquire the requested images of the Earth's surface with the operations needed to transmit the image files to a set of ground stations. The objective of this algorithm was to maximize the number of images taken with the specified priority and transmission of the urgent images to ground stations at the earliest time using a FIFO policy. Wang et al. (2011) investigated the scheduling of the scanning and transmission of images to ground stations to maximize the profit in a constellation of single-directional satellites.

III. Problem statement

Once placed in its orbit, a satellite starts circling around the globe. The Earth orbit of a satellite is the area on the Earth's surface that lies along the course and observation angle of the satellite. Because of the Earth's rotation, this orbit is not fixed and each time the satellite circles the Earth, its orbit changes according to the angle it makes with the equator.

A constellation of multi-directional observation satellites consists of a number of satellites each having its own Earth orbit according to the horizon of the planning. Each of the Earth orbits covers parts or all of particular areas for observation. These areas are divided into two groups: polygon targets that cannot be imaged with a single passing of the satellite and spot targets that can be imaged in their entirety in a single passing. This condition is illustrated in Figure 2.

There are several assumptions for each multi-directional observation satellite, as described in the following sections.

A. Time windows for observation

In order to scan on a strip within an area, a satellite must have the target in its observation angle. Thus there is an earliest and latest start and finish time for observation of an area that constitutes the time window for observation.

The exact start time for the scanning process is determined by the pitching angle of the satellite, as shown in Figure 3**.**

B. Setup time

Since strips from sequential areas do not have identical geographical positions, it takes some time for the satellite to shift from one strip to another and this is indicated by the setup time, which depends on the pitching and rolling angle of entering the two sequential strips, which in this paper is accounted for by using a linear function of these parameters.

Therefore, the scanning of two such strips is only possible when the finish time for observation of the previous strip plus the setup time is less than the start time for the scanning process of the next strip.

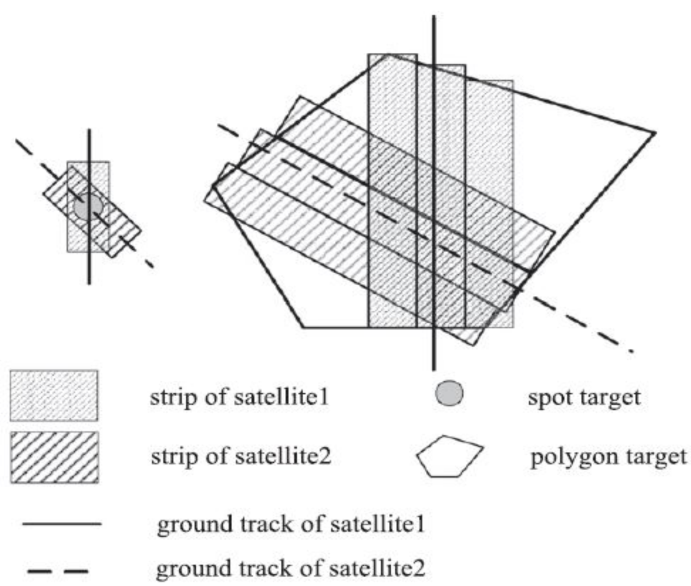

Fig. 2. Observation of spot and polygon targets (Wang et al. 2011)

C. Number of transitions

Changing the position of a satellite in order to move from one strip to another with a different pitching and rolling angle requires rotation of the satellite along its course (either in a backward-forward or a left-right direction) and this rotation consumes considerable energy. Thus the number of rotations in an orbit, while the satellite cannot absorb energy from its solar semi-orbit, is limited to a maximum value.

D. Memory capacity and duration of the observation process

The scanning of each strip requires consumption of energy that depends on the duration of the scanning process. Since the energy saved on a satellite's board is limited, the duration of the observation process is limited. Moreover, the memory board of a satellite can only contain a limited number of pictures until returning to the ground station.

E. Interference of strips and preemption of scanning

An area may be covered by the orbits of several satellites in the planning horizon. Therefore, if two distinct orbits share an area for scanning, the interfering parts of the strips may be imaged repetitively, which is a waste of energy and memory when further areas can be scanned.

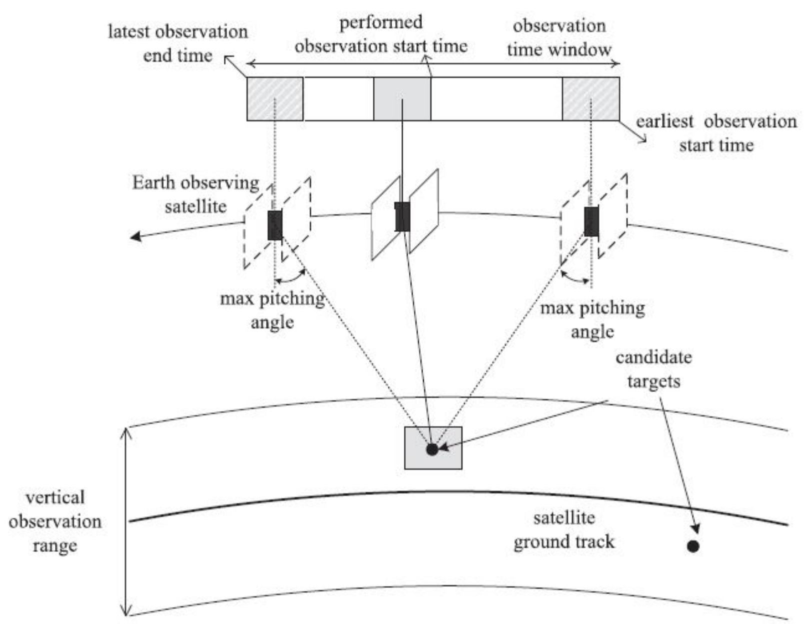

Fig. 3. Time window for observation of a target (Wang et al. 2011)

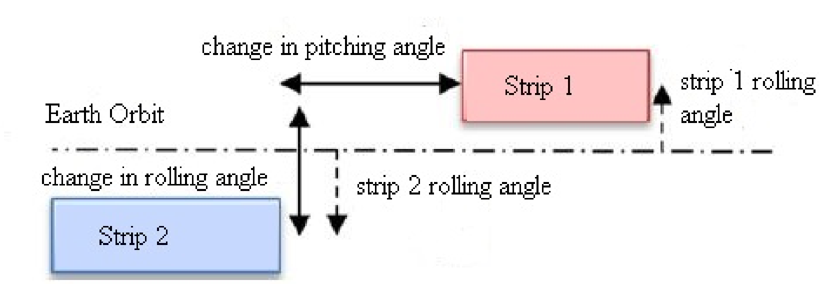

Fig. 4. Setup time for observation of two sequential strips

To avoid such circumstances, it only suffices to scan one of the two interfering strips and abort scanning the mutual area. However, due to the angle of the interfering strips, some areas may not be imaged. This is why several policies are defined to account for all possible scenarios. In this paper it is assumed that any two interfering strips completely cross each other.

To simplify the model, preemption policies for the observation of interfering strips is limited to the four policies. Figure 5 illustrates the policies for strip *l* and the interfering strip (i.e. strip *k*) in the area *j*. Strip *l* has crossed strip *k* at four points.

Point (1) is the first crossing point upon entry, point (2) is the last crossing point upon entry, point (3) is the first crossing point upon exit and point (4) is the last crossing point upon exit from strip *k*.

To demonstrate how the overlapping area and the corresponding memory savings are calculated, it is assumed that strip *k* in area *j* is wholly scanned. According to Figure 5, if W_{ojl} is the width of strip *l* from area *j* covered by a satellite in orbit *o*, *Q* is the direct distance from point (1) to point (2) and *P* the direct distance from point (2) to point (4), we have:

$$Q = W_{ojl} \times \tan\left(90 - \varphi_{oo'}\right) \quad \text{and} \quad P = \frac{W_{o'jk}}{\cos(90 - \varphi_{oo'})}.$$

assuming that the two strips *k* and *l* cross each other completely and in a way that results in four crossing points (see Figure 5). If V_{so} indicates the velocity of satellite *s* in orbit *o*, Table 1 indicates the overlapping area and the memory savings (on a time basis).

In this research, it is assumed that only two strips in an area may interfere with each other and in the case that both strips are to be scanned, only one will be scanned entirely. In some cases, however, a strip may cross several strips, which means that all interferences should be accounted for in order to calculate the overlapping areas and memory savings.

F. Objective of scheduling multi-directional observation satellites

Numerous requests from various agents are received in ground stations on a daily basis. Given the massive volume of requests, it is not possible to take all the required images in a short-term planning horizon of a single day and only a part of the polygon targets can be scanned in such a short time. Each request has a specific value or weight based on its importance and profitability. The value of each area has a significant impact on the selection of areas for scanning, as areas with higher value are prioritized.

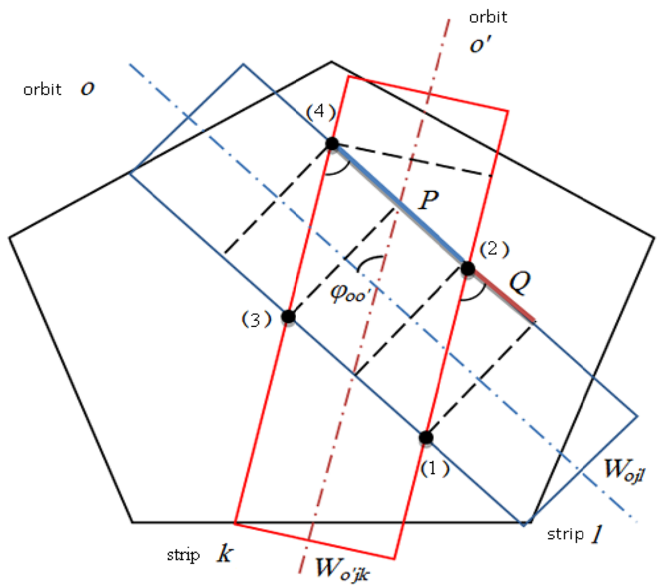

Fig. 5. Preemption policies for interfering strips

Therefore, the objective of this paper is to maximize the profits gained from observation and transmitting pictures to ground stations by scheduling a constellation of multi-directional observation satellites by taking into account the value of areas in the planning horizon, the net scanned area and the value of an area are the two most important items involved in determination of the profits.

Given the above-mentioned objective, the following should be determined upon completion of the scheduling: (1) what strips from which areas should be scanned and by which satellites? (2) what is the exact start time for scanning the selected strips? (3) what is the policy chosen for each strip in the event of an interference?

This problem can be modelled as a non-linear model, which cannot be solved by commercial software. The developed conceptual model is presented in Appendix 1

IV. GENETIC ALGORITHM

Due to the high computational complexity and NP-hard nature of the observation satellite scheduling problem, exact solution methods cannot be used to achieve the optimal solution (Chen et al., 2012). Therefore, in this paper a metaheuristic based on GA is used to solve the scheduling problem for agile observation satellites. Table 2 presents an overall description of the algorithm. Further specification is discussed in the following sections.

TABLE I. Preemption policies in case of crossing strips

Policy of a satellite	**Overlapping area**	**Memory saving**
Resuming observation of strip *l* without preemption	$W_{ojl}P$	0
Stop observation at the first crossing point at entry (point 1) until the first crossing point at exit (point 3)	$0.5W_{ojl}Q$	$\frac{P}{V_{so}}$
Stop observation at the last crossing point at entry (point 2) until the first crossing point at exit (point 3)	$W_{ojl}Q$	$\frac{P-Q}{V_{so}}$
Stop observation at the first crossing point at entry (point 1) until the last crossing point at exit (point 4)	0	$\frac{P+Q}{V_{so}}$
Stop observation at the last crossing point at entry (point 2) until the last crossing point at exit (point 4)	$0.5W_{ojl}Q$	$\frac{P}{V_{so}}$

TABLE II. General structure of the proposed GA

Steps	**Procedure**
Step 1	Construct the initial population: 1.1. Calculate the relative priority for strips from each area covered and create a chromosome based on the first rule. 1.2. Create the specified number of chromosomes based on the second rule. 1.3. Create the remaining chromosomes based on the third rule. 1.4. Repair all generated chromosomes of the initial population.
Step 2	Calculate the fitness function for the chromosomes of current generation, transfer the best chromosome to next generation and select the better chromosomes as parents.
Step 3	Apply crossover operator to parent chromosomes in order to create offspring.
Step 4	Apply mutation operator to generated offspring.
Step 5	Apply repairing operator to infeasible offspring.
Step 6	Repeat steps 3 to 5 for generation of offspring as many as the determined population size of the next generation.
Step 7	Repeat steps 2 to 6 until the end of time limit for execution of the algorithm

A. Solution representation

In order to construct an initial population, feasible solutions are required in which the number of satellites, Earth orbits, requested targets, and the scanning strips covered are known. Furthermore, the satellite's angle when entering a strip and the preemption policy are determined. The solution representation (chromosome) is in the form of a two-dimensional array with six fixed columns and several varying rows depending on the constraints, where each row indicates the scanning of a specific area and is considered as a gene from the chromosome. The first column includes the number of the satellite and the second column the corresponding orbit. The area chosen to be imaged by that satellite is specified in the third column. The fourth column indicates which strip from the set of strips partitions the selected area. The horizontal entry angle, i.e. the entry time to the selected strip, is indicated in column five. Finally, the sixth column represents the preemption policy. The following example demonstrates the representation of a solution and formation of a chromosome.

Example 1: Suppose five areas and two satellites are considered in the planning horizon, where orbits 1 and 2 are assigned to satellite 1 and orbits 3, 4, and 5 are allocated to satellite 2. Orbit 1 covers areas 1 and 2, orbit 2 area 3, orbit 3 areas 4 and 5, orbit 4 areas 3 and 4 and finally orbit 5 covers areas 1 and 2 for observation.

The area and width of the scanning strips in the example are assumed so that each area is covered by 4 strips. Table 3 illustrates a solution to this example. In this table, parameter θ denotes the maximum change in the observation pitching angle and, for instance, the first row shows that strip 3 from area 1 is imaged by satellite 1 and with an angle of $-\theta$, and that interference policy 1 has been selected for the crossing of another strip imaged by satellite 2.

B. Construction of the initial population

Three rules are elaborated for the construction of the initial population and are used for filling the corresponding chromosomes. A single chromosome is filled based on the first rule, a specific number of chromosomes based on the second rule, and the remaining chromosomes from the initial population will be filled based on the third rule. As explained in the following sections, the first and second rules allow for the identification of good genes, in terms of profitability, in a chromosome and this enhances the probability of the algorithm reaching a favorable solution in a limited time. These three rules along with the strategy for the transmission of pictures to ground stations are described as follows.

First rule: the value of each area in proportion to the area of the corresponding strip is calculated using $PI_{ojl} = \frac{TW_j}{W_{ojl}-L_{ojl}}$.

The larger the proportion, the greater the profitability, because observation of smaller areas with higher values saves a considerable amount of energy and memory. This proportion determines the priority for observation of various strips from different areas covered by an orbit, as a larger proportion denotes a higher priority. Operationalization of this rule for other chromosomes results in the repetition of a chromosome in the initial population.

Evidently, the solution from this rule may be desirable, but is not necessarily the best solution at the end of the algorithm. Because of given limitations such as the time it takes a satellite to change strips and the maximum number of satellite deviations for transition, a combination with medium priority (strips 1 and 3), as illustrated in Figure 6, may lead to better solutions compared to combinations with higher or lower priority (strips 2 and 4). In this figure, the combination of strip 1 and strip 2 has the relative priority of 2.7 while the combination of strip 2 and strip 4 has the relative priority of 2.3.

TABLE III. Illustration of a solution as a chromosome for example 1.

Preemption Policy	Entry Angle	Strip No.	Area	Orbit	Satellite
1	-θ	3	1	1	1
0	θ	2	3	2	1
0	0	1	4	3	2
0	θ	1	1	4	2
3	θ	2	3	4	2
0	0	2	2	5	2

Second rule: areas are randomly selected and then the area of each strip is calculated as the multiplication of the length of the strip by its width. The strip for a polygon target is selected for scanning. The difference between the first and the second rule is that in the first rule, areas are selected based on their value and are prioritized accordingly, while the second rule chooses areas randomly. The second rule is used to fill a specific number of chromosomes because the operationalization of this rule for a large number of chromosomes may result in only local optimality and limit the search space of the algorithm. Given the possibility of repetition of a chromosome in the initial population, the second rule also does not guarantee that the best solution is achieved.

Third rule: in the third rule, the selection of areas, strips, and other characteristics of a chromosome is entirely random and the remaining chromosomes from the initial population are selected according to this third rule.

Transmission strategy: transmission of the pictures to the ground stations will be based on a FIFO policy as the first strips imaged will have higher priority to be transmitted to ground stations and, upon arrival at a station, the images will be transmitted.

C. Fitness function

Since the goal of the problem is to maximize the profits gained from the imagery, the objective function as specified in Appendix 1 is chosen as the fitness function, because there is a direct correlation between the profitability and the net area imaged.

D. Crossover operator

In this paper, one-point crossover is used, in which a single point is chosen as the crossover point in the chromosome (Goldberg, 1989). This point can be at the end of a row where the set of pictures taken by the satellite ends. Therefore, if *m* satellites are scheduled, (*m*-1) potential places exist for this point. It should be noted that the location of this point within the two crossover chromosomes may vary according to the variable number of rows in each chromosome. To further elaborate this process, Tables 4 and 5 (that relate to example 1) are presented. In these tables, a crossover point for two parent chromosomes is indicated by an asterisk "*", where the dark areas indicated in the two chromosomes are replaced and two children are made, as shown in Tables 6 and 7.

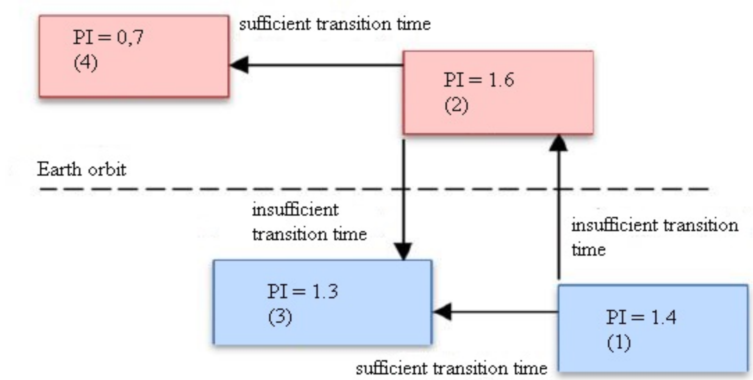

Fig. 6. Two possible combinations for observation of strips

TABLE IV. Parent chromosome 1

Preemption Policy	Entry Angle	Strip No.	Area	Orbit	Satellite	
2	-θ	3	2	1	1	
0	θ	2	3	2	1	*
0	0	1	5	3	2	
1	-θ	2	3	4	2	
0	0	1	4	4	2	
0	θ	2	2	5	2	

E. Mutation operator

In the mutation procedure, a specific number of rows are selected based on the mutation rate, and then the characteristics of that gene or row are modified from row 3 to row 6 of the chromosome. The characteristics include the scanning area, the strips for that area, the entry angle, and the type of interference that determines the preemption policy. For illustration purposes, assume that the mutation takes place for the chromosome of offspring 1 that is created after crossover. In Table 8, the mutation takes place for changing the characteristics specified in rows that are indicated as dark and randomly. Table 9 shows this chromosome after the mutation and illustrates the random change in the characteristics of that chromosome. These changes may lead to infeasible solutions.

TABLE V. Parent chromosome 2

Preemption Policy	Entry Angle	Strip No.	Area	Orbit	Satellite	
0	-θ	2	1	1	1	
1	θ	3	2	1	1	
0	0	1	3	2	1	*
0	-θ	2	5	4	2	
2	0	1	3	4	2	
3	θ	2	1	5	2	

TABLE VI. Offspring chromosome 1 after crossover

Preemption Policy	Entry Angle	Strip No.	Area	Orbit	Satellite	
0	-θ	2	1	1	1	
1	θ	3	2	1	1	
0	0	1	3	2	1	
0	0	1	5	3	2	
1	-θ	2	3	4	2	
0	0	1	4	4	2	

TABLE VII. Offspring chromosome 2 after crossover

Preemption Policy	Entry Angle	Strip No.	Area	Orbit	Satellite	
2	-θ	3	2	1	1	
0	θ	2	3	2	1	
0	-θ	2	5	4	2	
2	0	1	3	4	2	
3	θ	2	1	5	2	
0	-θ	3	2	5	2	

TABLE VIII. Offspring chromosome 1 before mutation

Preemption Policy	Entry Angle	Strip No.	Area	Orbit	Satellite
0	-θ	2	1	1	1
1	θ	3	2	1	1
0	0	1	3	2	1
0	0	1	5	3	2
1	-θ	2	3	4	2
0	0	1	4	4	2
0	θ	2	2	5	2

F. Repairing operator

During the crossover and mutation procedures, the solution may become infeasible. This may be due to:

a) Insufficient setup time for transition of the satellite’s position

b) Exceeding the overall allowed scanning time

c) Exceeding the maximum number of allowed deviations in an orbit

d) Repetition of scanned areas in a strip

e) Exceeding the maximum volume of the satellite’s memory

The repairing process in order to make solutions feasible consists of the following:

1- Change a satellite’s entry angle in the allowed spectrum: in this way, shortages for setup time and transition of the satellite can be eliminated. The smaller the entry angle of a satellite to the next strip, the later the scanning process of that strip starts and consequently more setup time becomes available.

2- Replace the selected strip by a shorter strip in the same area: in this way, less memory will be used and the total duration of the satellite’s trip reduces while allowing more setup time for scanning the next strip.

3- Change the preemption policy: operationalization of this alteration (for example, changing the preemption policy from 0 to 1 for a strip when both interfering strips have chosen policy 0) enables the prevention of repetition in scanned areas and consequently less consumption of memory as well as a shortened duration of the whole operation.

4- Change the scanned area: when it is not possible to provide the necessary setup time for scanning of an area or if the memory or remaining time for observation of the strips of that area are not sufficient, or when no further deviation in that orbit is possible, the area can be changed in order to scan another area with more favorable conditions to achieve feasible solutions.

5- Remove a strip: when the previous measures are not effective in obtaining a feasible solution, the strip in question will be removed.

Using one or a set of the above-mentioned measures, an infeasible solution can become feasible. As in example 1, assume that the chromosome illustrated in Table 3 is an infeasible solution after crossover and mutation processes in which the transition time from strip 1 of area 1 to strip 2 of area 3 under the entry angle θ for scanning in orbit 4 is not enough and the total scanning time in orbit 5 has exceeded the specified limit.

Now, in order to make the solution feasible, the angle for entering area 3 in orbit 4 can be changed from θ to $-\theta$, which allows for an increased setup time for transition. As for orbit 5, strip 2 from area 2 can be replaced by strip 3 from the same area (assuming that strip 3 is shorter) to decrease the scanning time. These alterations are illustrated in Table 10 in dark.

TABLE IX. Offspring chromosome 1 after mutation

Preemption Policy	Entry Angle	Strip No.	Area	Orbit	Satellite
0	$-\theta$	2	1	1	1
3	$-\theta$	4	2	1	1
0	0	1	3	2	1
0	0	1	5	3	2
2	θ	3	3	4	2
0	0	1	4	4	2
0	θ	2	2	5	2

TABLE X. Making a solution feasible by repairing operator

Preemption Policy	Entry Angle	Strip No.	Area	Orbit	Satellite
1	$-\theta$	3	1	1	1
0	θ	2	3	2	1
0	0	1	4	3	2
0	θ	1	1	4	2
3	$-\theta$	2	3	4	2
0	0	3	2	5	2

V. COMPUTATIONAL RESULTS

The proposed GA is implemented in a Visual C++ 2010 environment and all the test problems were run on a computer with CPU 3.5 GHz core i7, 5 GB RAM, and the Windows XP operating system. The planning horizon for execution of the proposed problem was determined as 24 hours. A total of four satellites were considered for the observation process. The parameters and technical characterization of the hypothetical satellites in the problem are selected as similar to those of a RADARSAT-2 (Morena et al., 2014). It should be noted, however, that the considered planning horizon in the problem is shorter than that of RADARSAT-2. Moreover, five ground stations for receiving the images have been assumed in the problem.

The requested spot targets are selected as latitudes and longitudes uniformly selected from the Earth's surface with an area of 10000 km^2 and a value of 1 to 10 assigned randomly. Polygon targets are assumed to be 30000 to 300000 km^2 and valued randomly as an integer in the interval [5, 40]. Table 11 demonstrates six classes of targets requested to be imaged and each class is studied under three different scenarios for the planning horizon. Therefore, a total of 18 test problems are given. In the first scenario a single satellite, in the second scenario three satellites and in the third scenario four satellites are scheduled. JSatTrak software version 4.1 developed by NASA is used to determine the specification and number of areas to be scanned and the corresponding time windows. This software uses a Simplified General Perturbation (SGP) algorithm to estimate the past, present and future positions of a satellite. For implementation of the GA, a crossover rate of 0.8, mutation rate of 0.05, and an initial population of 100 chromosomes are used.

One chromosome is made using the first rule, 30 chromosomes by the second rule and the remaining chromosomes are made by the third rule. These parameters are determined by the design of experiments, the results of which are shown in Table 12**.** This table illustrates the number of spot and polygon targets imaged (either partially or wholly) and the corresponding obtained profits. These values are presented for cases with and without preemption policy. The results from all the test problems in the various scenarios are also illustrated. The results are the average of the best solutions under a time constraint of 100 seconds. Also, based on the DOE techniques, we found that the best values for the population size, crossover operator rate and mutation operator rate are 100, 0.8 and 0.05, respectively.

A. Analysis of results

Given the significant difference between the results in the presence and absence of a preemption policy for the various scenarios, it can be seen that the profits increase when a preemption scheme is in place. This can be attributed to energy and memory savings from avoiding the scanning of repetitive areas, which ultimately resulted in a 14.2 % increase in profits and 12.58 % increase in the number of areas scanned as compared with the case where no preemption policy existed. Furthermore, as indicated in Table 14, it is shown that in the presence of a preemption policy, the profits of the third scenario are 18.96 % higher than those of the second scenario and the profits of the second scenario are 22.09 % more than the profits of the first scenario. This is because scenario 3, as compared to scenario 2, and scenario 2 as compared to scenario 1, consist of more satellites and consequently larger overall scanned areas, which evidently impacts the profits obtained. Where in each scenario there are more requested targets with particularly large areas, the profits accordingly increase.

TABLE XI. Number of targets in test problems

Target Classes	Number of spot targets	Number of polygon targets
1	100	100
2	150	100
3	100	150
4	200	100
5	100	200
6	150	150

B. Impact of the population creation method on the algorithm performance

To create the initial population, a method based on three rules was specified. If the method used for creation of the initial population was based on random selection of areas and strips, and the entry angle and all the chromosomes of the initial population were made accordingly, the performance of the proposed methodology for creation of the initial population can be assessed under the given time constraint.

Figure 7 shows the average percent deviation (APD) for the best solutions found in the implementation of the algorithm under a 100-second time limit for the various scenarios in the presence of a preemption policy for the two types of initial population.

TABLE XII. Comparison between results in presence and absence of a preemption policy

Scenario	Area Classes	With Preemption		Without Preemption		Average Percent Deviation in Profits
		Total No. of observed targets	Gained Profits	Total No. of observed targets	Gained Profits	
1	1	96	792.53	83	667.28	18.77
	2	123	985.14	105	841.36	17.09
	3	130	1027.85	113	891.72	15.27
	4	145	1058.61	129	911.15	16.18
	5	153	1329.42	136	1134.67	17.16
	6	159	1268.15	142	1085.23	16.86
2	1	121	982.64	107	841.72	16.74
	2	142	1068.53	126	961.34	11.15
	3	153	1241.64	139	1121.45	10.72
	4	188	1362.18	165	1189.35	14.53
	5	197	1686.92	178	1473.92	14.45
	6	194	1547.28	183	1381.46	12.00
3	1	143	1138.24	126	998.61	13.98
	2	161	1291.76	144	1138.24	13.49
	3	182	1442.65	167	1297.70	11.17
	4	218	1723.81	191	1560.98	10.43
	5	244	1912.55	203	1661.81	15.09
	6	236	1876.34	214	1695.92	10.64

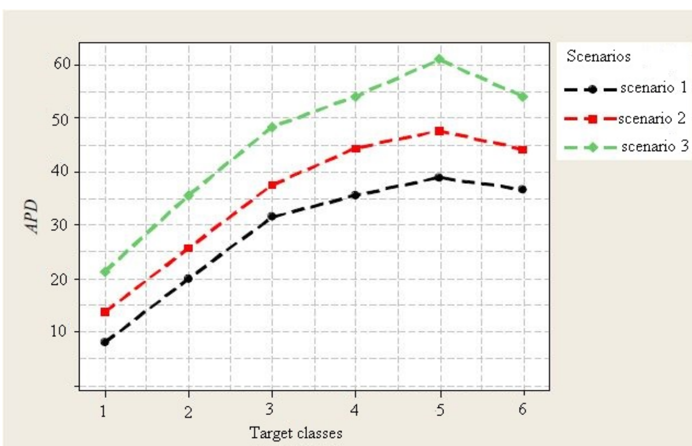

Fig. 7. Average Percent Deviation for the obtained results of creation of the initial population randomly and by the specified rules in three different scenarios

Appendix 1: Conceptual modelling of the problem

TABLE XIII. Symbols and Parameters

Symbol	Description
R_1	Set of spot targets
R_2	Set of polygon areas
$R=R_1 \cup R_2$	Set of requested targets
O	Set of satellite orbits
O_j	Set of orbits covering all or parts of area j and $j \epsilon R$
$O_{ij}=O_i \cap O_j$	Set of orbits covering both areas i and j where $i,j \epsilon R$ and $i \neq j$
$\varphi_{oo'}$	The angle between crossing orbits o and o' where $o,o' \epsilon O$ and $o \neq o'$
$R^O=R_1^O \cup R_2^O$	Set of spot and polygon targets that can be covered and imaged from orbit o
TI_o	Maximum allowed time for observation of given the constraints of the satellite in orbit o
NS_o	Maximum number of allowed angular transition given the energy constraints in orbit $o \epsilon O$
S	Set of satellites
$s\ (o)$	The satellite to which orbit o belongs and $o \epsilon O$
V_s	Velocity of satellite s and $s \epsilon S$
MC_s	Maximum memory capacity for satellite s and $s \epsilon S$ (on a time basis)
h_s	Average distance of satellite s from earth's surface ($s \epsilon S$)
$_s\omega$	Width of scanning area for satellite s while in an orthogonal position against its earth orbit ($s \epsilon S$)
β_s	Half of satellite s observation angle while in an orthogonal position against its earth orbit ($s \epsilon S$)
G	Set of ground stations
G_o	Set of ground stations covered by satellites in orbit $o \epsilon O$
NG_o	Number of ground stations covered by satellites in orbit $o \epsilon O$
T_g	Average time for setting a ground station's antenna from one orbit to another in order to receive transmissions from two sequential satellites and $g \epsilon G$
$_s\theta$	Maximum change in pitching angle of satellite $s \epsilon S$
n_{oj}	Number of strips in area j covered by satellite in orbit o where $j \epsilon R,\ o \epsilon O_j$.
ES_{ojl}	Earliest start time for observation of strip l from area j by a satellite in orbit o under the pitching angle of $\theta_{s(O)}$ while $1 \leq l \leq n_{oj}$ and $j \epsilon R,\ o \epsilon O_j$
LS_{ojl}	Latest start time for observation of strip l from area j by a satellite in orbit o under the pitching angle of $\theta_{s(O)}$ while $1 \leq l \leq n_{oj}$ and $j \epsilon R,\ o \epsilon O_j$
L_{ojl}	Length of strip l from area j covered by the satellite in orbit o when $j \epsilon R,\ o \epsilon O_j,\ 1 \leq l \leq n_{oj}$
ρ_{ojl}	Rolling angle of strip l from area j against the orthogonal axis of satellite in orbit o where $j \epsilon R,\ o \epsilon O_j$, $1 \leq l \leq n_o$
W_{ojl}	Width of strip l from area j covered by satellite in orbit o where $j \epsilon R,\ o \epsilon O_j,\ 1 \leq l \leq n_{oj}$
SU_{ojl}	Area of strip l from area j covered by satellite in orbit o where $j \epsilon R,\ o \epsilon O_j,\ 1 \leq l \leq n_{oj}$
$offset_{ojl}$	Ineffective area of strip l from area j covered by satellite in orbit o where $j \epsilon R,\ o \epsilon O_j,\ 1 \leq l \leq n_{oj}$
I_{ojl}	Set of strips interfering with strip l from area j covered by satellite in orbit o when $j \epsilon R,\ o \epsilon O_j,\ 1 \leq l \leq n_{oj}$
PI_{ojl}	Relative priority of being imaged for strip l from area j covered by the satellite in orbit o and $j \epsilon R,\ o \epsilon O_j,\ 1 \leq l \leq n_{oj}$
m_{ojlk}	Number of policies for observation of strip l from area j covered by the satellite in orbit o and $j \epsilon R,\ o \epsilon O_j,\ 1 \leq l \leq n_{oj},\ k \epsilon I_{ojl}$
ESD_{og}	Earliest start time for transmission of images to ground station g where $o \epsilon O,\ g \epsilon G$.
LFD_{og}	Latest start time for transmission of images to ground station g where $o \epsilon O,\ g \epsilon G$.
$\mu(ojl,oif)$	Duration of rolling angular transition from strip l of area j to strip f from area i covered by orbit $o \epsilon O_{ij}$
η	Energy consumption coefficient between 0 and 1 for each transmission of images to ground stations
B	Start of the planning horizon
H	End of the planning horizon
TW_j	Total value of imagery from area $j \epsilon R$
TA_j	Total area of area $j \epsilon R$

TABLE XIV. Model Variables

Variables	**Description**
X_{ojl}	Binary variable denoting selection of strip l from area j covered by orbit o; $j\epsilon R$, $o\epsilon O_j$, $1\leq l\leq n_{oj}$.
Y_{ojlk}	Non-negative integer variable representing the scanning policy for strip l from area j covered by orbit o interfering with strip k where $0\leq Y_{ojlk}\leq 4$, $k\epsilon I_{ojl}$
ST_{ojl}	Start time for observation of strip l from area j covered by orbit o
FT_{ojl}	Finish time for observation of strip l from area j covered by orbit $o\epsilon O_{ij}$
θ_{ojl}	Pitching angle for entry to and exit from strip l of area j covered by orbit $o\epsilon O_{ij}$
$\gamma(\theta_{ojl},\theta_{oif})$	Duration for changing pitching angle from strip l of area j to strip f from area i covered by orbit o under entry and exit angle of θ_{ojl} and $o\epsilon O_{ij}$
RS_{ojlif}	Binary variable for selection of two sequential strips l from area j and strip f from area i with different rolling angles that are covered by satellite in orbit o. When l and f are sequential and $\rho_{ojl}\neq\rho_{oif}$ this variable takes a value of 1 and otherwise 0.
PS_{ojlif}	Binary variable for selection of two sequential strips l from area j and strip f from area i with different pitching angles that are covered by satellite in orbit o. When l and f are sequential and $\rho_{ojl}\neq\rho_{oif}$ this variable takes a value of 1 and otherwise 0 when $\theta_{ojl}\leq\theta_{s(o)}$, $\theta_{oif}\leq\theta_{s(o)}$.
DW_{og}	Binary variable for selection of time window for transmission of images to ground station g by satellite in orbit o
STD_{og}	Non-negative integer variable indicating the start time for transmission of images to ground station g by satellite in orbit o in the specified time window
FTD_{og}	Non-negative integer variable indicating the finish time for transmission of images to ground station g by satellite in orbit o in the specified time window
DTD_{og}	Non-negative integer variable indicating the duration of transmission of images to ground station g by satellite in orbit o in the specified time window
II_{ojlif}	Binary variable for selection of two sequential strips, l from area j and f from area i, by a satellite in orbit o for scanning
DI_{ojlg}	Binary variable for transmission of images to ground station g and selection of strip l from area j for observation of as two sequential activities by satellite in orbit o
ID_{ojlg}	Binary variable for selection of strip l from area j for observation of and transmission of images to ground station g as two sequential activities by satellite in orbit o.
$DD_{og'g}$	Binary variable for transmission of images to ground station g' and transmission of images to ground station g as two sequential activities by satellite in orbit o
LM_{ts}	Non-negative integer variable indicating memory level at time t for satellite s when $s\epsilon S$,$B\leq t\leq H$
OV_{ojl}	Overlapping areas of strip l from area j covered by orbit o interfering with other strips and $j\epsilon R$, $o\epsilon O_j$ $1\leq l\leq n_{oj}$
f	A function of $X_{o'jk}$,$Y_{o'jkl}$,Y_{ojlk} denoting the overlapping areas of strip X_{ojk} interfering with strip $X_{o'jk}$
SD_{ojl}	Time savings in case of interference for strip l from area j covered by orbit o; $j\epsilon R$, $o\epsilon O_j$ $1\leq l\leq n_{oj}$
h	A function of $X_{o'jk}$,$Y_{o'jkl}$,Y_{ojlk} denoting the savings when strip X_{ojk} interferes with strip $X_{o'jk}$
OA_j	Continuous non-negative variable denoting the areas imaged ($j\epsilon R$) by all satellites regardless of the overlapping

The conceptual model for scheduling the satellites is as follows:

$$Max\ Z=\sum_{j\in R_1}TW_j\sum_{o\in O_j}X_{ojl}+\sum_{j\in R_2}TW_j\frac{OA_j}{TA_j} \tag{15}$$

$S.t.$

$$\sum_{o\in O_j}X_{ojl}\leq 1,\qquad \forall j\in R_1,\ l=1 \tag{16}$$

$$\sum_{l=1}^{n_{oj}} X_{ojl} \le 1, \quad \forall j \in R_2, o \in O_j \tag{17}$$

$$X_{ojl} \le Y_{ojlk} \le m_{ojlk} X_{ojl}, \quad \forall j \in R_2, o \in O_j, 1 \le l \le n_{oj}, k \in I_{ojl} \tag{18}$$

$$OV_{ojl} = 0, \quad \forall j \in R_1, o \in O_j, l = n_{oj} = 1 \tag{19}$$

$$OV_{ojl} = \bigcup_{o' \ne o \in O_j, k \in I_{ojl}} f(X_{o'jk}, Y_{o'jkl}, Y_{ojlk}), \quad \forall j \in R_2, o \in O_j, 1 \le l \le n_{oj} \tag{20}$$

$$SD_{ojl} = 0, \quad \forall j \in R_1, o \in O_j, l = n_{oj} = 1 \tag{21}$$

$$SD_{ojl} = \bigcup_{o' \ne o \in O_j, k \in I_{ojl}} h(X_{o'jk}, Y_{o'jkl}, Y_{ojlk}), \quad \forall j \in R_2, o \in O_j, 1 \le l \le n_{oj} \tag{22}$$

$$OA_j = \sum_{o \in O_j} \sum_{l=1}^{n_{oj}} (SU_{ojl} X_{ojl} - V_{s(o)} SD_{ojl} W_{ojl} - OV_{ojl}), \quad \forall j \in R_2 \tag{23}$$

$$ES_{ojl} \le ST_{ojl} \le LS_{ojl}, \quad \forall j \in R, o \in O_j, 1 \le l \le n_{oj} \tag{24}$$

$$FT_{ojl} = ST_{ojl} + LT_{ojl}, \quad \forall j \in R, o \in O_j, 1 \le l \le n_{oj} \tag{25}$$

$$\begin{aligned} & X_{ojl} + X_{oif} \le 1, \quad \forall i \ne j \in R, o \in O_{ij}, 1 \le l \le n_{oj}, 1 \le f \le n_{oi} \\ & \quad If: \max\{FT_{ojl}, FT_{oif}\} - \min\{ST_{ojl}, ST_{oif}\} \le FT_{ojl} - ST_{ojl} + FT_{oif} - ST_{oif} \\ & \qquad + \gamma(\theta_{ojl}, \theta_{oif}) + \mu(ojl, oif) \end{aligned} \tag{26}$$

$$\sum_{g \in G_o} DW_{og} \le NG_o, \quad \forall o \in O \tag{27}$$

$$ESD_{og} \le STD_{og} \le LFD_{og}, \quad \forall o \in O, g \in G_o \tag{28}$$

$$FTD_{og} \le LFD_{og}, \; FTD_{og} = STD_{og} + DTD_{og}, \quad \forall o \in O, g \in G_o \tag{29}$$

$$\begin{aligned} & DW_{og} + DW_{o'g} \le 1, \quad \forall g \in G, o \ne o' \in O \\ & \quad If: \max\{FTD_{og}, FTD_{o'g}\} - \min\{STD_{og}, STD_{o'g}\} \le DTD_{og} + DTD_{o'g} + T_g \end{aligned} \tag{30}$$

$$\sum_{j \in R} \sum_{l=1}^{n_{oj}} X_{ojl} (FT_{ojl} - ST_{ojl} - SD_{ojl}) + \sum_{g \in G_o} \eta \, DW_{og} \, DT_{og} \le TI_o, \quad \forall o \in O_j \tag{31}$$

$$\sum_{l=1}^{n_{oj}}\sum_{f=1}^{n_{oi}} RS_{ojlif} + \sum_{l=1}^{n_{oj}}\sum_{f=1}^{n_{oi}} PS_{ojlif} \le NS_o, \qquad \forall i \ne j \in R^o, o \in O \tag{32}$$

$$LM_{Bs} = LM_{Hs} = 0, \quad \forall s \in S \tag{33}$$

$$0 \le LM_{ts} \le MC_s, \qquad \forall t = FTD_{og},\ o \in O, g \in G_o, s = s(o) \in S \tag{34}$$

$$0 \le LM_{ts} \le MC_s, \qquad \forall t = FT_{ojl},\ j \in R, o \in O_j, s = s(o) \in S \tag{35}$$

$$If: II_{ojlif} = 1 \Rightarrow LM_{(FT_{ojl})s} + FT_{oif} - ST_{oif} - SD_{oif} = LM_{(FT_{oif})s}, \quad \forall i \ne j \in R, o \in O_{ij}, s = s(o) \in S \tag{36}$$

$$If: DI_{ogjl} = 1 \Rightarrow LM_{(FTD_{og})s} + FT_{oif} - ST_{oif} - SD_{oif} = LM_{(FT_{oif})s}, \quad \forall j \in R, o \in O_j, g \in G_o, s = s(o) \in S \tag{37}$$

$$If: ID_{ogjl} = 1 \Rightarrow LM_{(FT_{ojl})s} - DTD_{og} = LM_{(FTD_{og})s}, \quad \forall j \in R, o \in O_j, g \in G_o, s = s(o) \in S \tag{38}$$

$$If: DD_{ogg'} = 1 \Rightarrow LM_{(FTD_{og})s} - DTD_{og'} = LM_{(FTD_{og'})s}, \quad \forall g \ne g' \in G_o, o \in O, s = s(o) \in S \tag{39}$$

The objective function in Equation (15) is made up of two parts: the first constitutes the sum of values for spot targets, and the second part shows the sum of values for all or parts of polygon targets. Constraint (16) indicates that each area may only be imaged once. Constraint (17) refers to the fact that for polygon targets, each satellite circling in the orbit covering that area may, at most, scan one of the strips in that area. Constraint (18) guarantees that preemption policies may only be selected once a strip is chosen for scanning. Constraints (19) and (20) indicate the overlapping area in the case of interfering strips for spot and polygon targets respectively. Constraints (21) and (22) also represent the memory savings based on time in the case of interfering strips for spot and polygon targets respectively. Equation (23) shows the net area scanned for each polygon target. Constraints (24) and (25) demonstrate the intervals allowed for the start and finish times for scanning of an area. Constraint (26) guarantees that a satellite will only scan one of the strips when its orbit contains two strips from two distinct areas and the setup time for transition from one strip to another is not sufficient. The maximum number of ground stations for receiving images from a satellite in a specific orbit is shown in Equation (27). Constraints (28) and (29) denote the interval for the start time and finish time for the transmission of images respectively. Constraint (30) assures that a ground station will only allow reception of images from a single satellite and simultaneous transmission of images from two satellites is not allowed. Constraint (31) refers to the fact that the sum of the scanning time and transmission time for a satellite in a specific orbit must not exceed the maximum time allowed. Constraint (32) shows that the number of deviations for the transition of a satellite in its orbit must not exceed a maximum value. Constraint (33) guarantees that the memory level will be set to zero at the start and end of the planning horizon to assure transmission of all images to ground stations. Constraint (34) and (35) indicate that the memory level will not exceed the maximum capacity. Constraints (36) and (37) show the change in memory level after each observation process. Finally, constraints (38) and (39) denote the change in memory level after each transmission process.

The results indicate that using the proposed rules to create the initial population expedites the algorithm in finding a desirable solution. In the same time period for execution of the algorithm, the overall profits gained using the proposed rules were on average 36.56 % higher than in the case where the initial population is only randomly created. This superiority is further enhanced when a larger number of satellites or areas is considered. This can be attributed to the existence of good genes in the initial population, which are passed on to next generations and eventually result in faster solutions.

VI. Conclusions

This paper investigates the scheduling of a constellation of agile satellites, which is a problem with a high degree of complexity. This study introduces a preemption policy for energy and memory savings and to avoid the repetitive scanning of areas; this is unprecedented in the literature and at the same time adds to the complexity of the problem. A conceptual programming model for a better statement of the problem and a metaheuristic based on a genetic algorithm for solving the problem are developed. The computational results indicate that the total value of the imagery in the presence of the specified preemption policies, to avoid the repetition of imaged areas and save the consumption of energy and memory in a 24-hour planning horizon, has increased by an average of 14.2 %. This increase in the profits, taking into account the massive expenses of development, launching, and control of satellites, is of considerable significance.

Future studies may investigate the scheduling of this problem under several other objectives, including the minimization of the response time to customers. Furthermore, application of a hybrid algorithm for solving this problem may result in shortened computation times and future studies may explore this possibility.

REFERENCES

Bianchessi, N., & Righini, G. (2008). Planning and scheduling algorithms for the COSMO-SkyMed constellation, Aerospace Science and Technology, v12(7), 535–544.

Chen, Y., Zhang, D., Zhou, M., & Zou, H. (2012). Multi-satellite Observation Scheduling Algorithm Based on Hybrid Genetic Particle Swarm Optimization. In D. Zeng (Ed.), Advances in Information Technology and Industry Applications (Vol. 136). Berlin, Heidelberg: Springer Berlin Heidelberg.

Florio, S. De. (2006). Performances Optimization of Remote Sensing Satellite Constellations: a Heuristic Method, In Proceedings of the 5th international workshop on planning and scheduling for spacev, http://www.stsci.edu/largefiles/iwpss/20069151043Paper.pdf.

Frank, J., Jónsson, A., & Morris, R. (2001). Planning and scheduling for fleets of earth observing satellites, In Proceedings of Sixth Int. Symp. on Artificial Intelligence, Robotics, Automation & Space, http://citeseerx.ist.psu.edu/viewdoc/summary?.

Gabrel, V., Moulet, A., Murat, C., & Paschos, V. T. (1997). A new single model and derived algorithms for the satellite shot planning problem using graph theory concepts, Annals of Operations Research, 69, 115–134.

Lemaître, M., Verfaillie, G., Jouhaud, F., Lachiver, J.-M., & Bataille, N. (2002). Selecting and scheduling observations of agile satellites, Aerospace Science and Technology, 6(5), 367–381.

Morena, L. C., James, K. V, & Beck, J. (2014). An introduction to the RADARSAT-2 mission. Canadian Journal of Remote Sensing, 30(3), 221–234.Wang, P., & Reinelt, G. (2010). A Heuristic for an Earth Observing Satellite Constellation Scheduling Problem with Download Considerations, Electronic Notes in Discrete Mathematics, 36, 711–718.

Tangpattanakul, P., Jozefowiez, N., & Lopez, P. (2015). A multi-objective local search heuristic for scheduling Earth observations taken by an agile satellite, *European Journal of Operational Research*, 245(2), 542–554.

Wang, P., Reinelt, G., Gao, P., & Tan, Y. (2011). A model, a heuristic and a decision support system to solve the scheduling problem of an earth observing satellite constellation, *Computers & Industrial Engineering*, 61(2), 322–335.

Wang, J., Demeulemeester, E., & Qiu, D. (2016). A pure proactive scheduling algorithm for multiple earth observation satellites under uncertainties of clouds, Computers and Operations Research, 74, 1-13.

Xu, R., Chen, H., Liang, X., & Wang, H. (2016). Priority-based constructive algorithms for scheduling agile earth observation satellites with total priority maximization, Expert Systems with Applications, 51, 195-206.