

Artificial intelligence for the EChO mission planning tool

Alvaro Garcia-Piquer · Ignasi Ribas · Josep Colomé

Received: 16 January 2014 / Accepted: 11 July 2014 / Published online: 17 December 2014
© Springer Science+Business Media Dordrecht 2014

Abstract The Exoplanet Characterisation Observatory (EChO) has as its main goal the measurement of atmospheres of transiting planets. This requires the observation of two types of events: primary and secondary eclipses. In order to yield measurements of sufficient Signal-to-Noise Ratio to fulfil the mission objectives, the events of each exoplanet have to be observed several times. In addition, several criteria have to be considered to carry out each observation, such as the exoplanet visibility, its event duration, and no overlapping with other tasks. It is expected that a suitable mission plan increases the efficiency of telescope operation, which will represent an important benefit in terms of scientific return and operational costs. Nevertheless, to obtain a long term mission plan becomes unaffordable for human planners due to the complexity of computing the huge number of possible combinations for finding an optimum solution. In this contribution we present a long term mission planning tool based on Genetic Algorithms, which are focused on solving optimization problems such as the planning of several tasks. Specifically, the proposed tool finds a solution that highly optimizes the defined objectives, which are based on the maximization of the time spent on scientific observations and the scientific return (e.g., the coverage of the mission survey). The results obtained on the large experimental set up support that the proposed scheduler technology is robust and can function in a variety of scenarios, offering a competitive performance which does not depend on the collection of exoplanets to be observed. Specifically, the results show that, with the

A. Garcia-Piquer (✉) · I. Ribas · J. Colomé
Institute of Space Sciences (IEEC – CSIC), Campus UAB, Faculty of Science, C5 Tower - 2nd Floor,
08193 Bellaterra, Spain
e-mail: agarcia@ice.csic.es

I. Ribas
e-mail: iribas.@ice.csic.es

J. Colomé
e-mail: colome.@ice.csic.es

proposed tool, EChO uses 94 % of the available time of the mission, so the amount of downtime is small, and it completes 98 % of the targets.

Keywords Space applications · Observatory operations · Scheduling · Planning · Genetic algorithms · Constraint-based reasoning

1 Introduction

The primary objective of The Exoplanet Characterisation Observatory (EChO) [9, 27] is to study a representative sample of exoplanets around nearby stars, with masses ranging in sizes from Jupiter to a few Earths by using the differential technique of transit spectroscopy. Temporal variations in the observed signal from spatially unresolved observations of an exoplanet in orbit around its parent star, at different points in its orbit, will be used to determine the spectrum of the planetary atmosphere. This can be achieved using high-precision spectrophotometric observations of two types of events: (1) primary eclipse and (2) secondary eclipse, which can be also referred as transit and occultation, respectively. In order to yield measurements of sufficient Signal-to-Noise Ratio to fulfil the mission objectives, the events of each exoplanet may need to be observed several times. EChO can only examine one exoplanet event at a time, so the observations cannot be done simultaneously. In addition, several criteria have to be considered to carry out each observation: (1) exoplanet visibility, (2) time and duration of events, (3) number of events to be observed, and (4) exoplanet priority. The observations of the selected targets are only valid when the events required to reach the target Signal-to-Noise Ratio are observed. Finally, other tasks have to be considered along the mission for calibrating the telescope or transferring data from the spacecraft to stations on Earth.

A suitable long-term mission plan for all the mission time is expected to increase the efficiency of telescope operation, which will represent an important benefit in terms of scientific return and operational costs. However, this process becomes unaffordable for human planners due to the complexity in computing the huge number of possible combinations in search for an optimum solution. This class of optimization problems is considered NP-hard, and there are many mathematical tools to solve the planning/scheduling issue: from simple heuristics to more complex Artificial Intelligence (AI) approaches [8].

Although several tools for automatic computation of long-term plans have been previously developed [3, 12], they have been designed in the context of specific missions (e.g., constraints, scientific goals, proposal definition) and thus are unsuitable to the needs of the EChO mission. In this contribution we present a standalone Long Term Mission Planning Tool (LT-MPT) for EChO based on Genetic Algorithms (GAs) [15], which are an AI approach focused on emulating natural evolution by means of combining potential solutions using selection, combination and mutation operators [10]. This kind of algorithms attempt to generate solutions to optimization problems by exploring a large amount of potential solutions, including the most efficient ones [13]. A solution is considered efficient when it highly optimizes the objectives defined in the problem that, in our case, correspond to maximize the time

spent on scientific observations and the scientific return, measured in terms of the coverage of the mission sample. Taking into account that new exoplanets that may be of interest to EChO will be discovered until 2022, it is necessary to have a LT-MPT able to obtain optimal planning solutions in any exoplanet sample. For this reason, the capabilities of the LT-MPT have been tested with the sample of known exoplanets and with different hypothetical scenarios of 2022.

The paper is organized as follows. Section 2 briefly summarizes the related work on automatic mission planning. Section 3 describes the planning tool in the operational design of the mission. Section 4 details the EChO mission optimization problem. Section 5 presents the LT-MPT and how GAs are applied in the optimization process. Section 6 describes the experimentation done and discuss the results. Finally, Section Section 7 ends with conclusions and further work.

2 Related work

In the past few years, the ESA has increasingly focused its attention on the use of Automated Planning and Scheduling solutions to support space operations [20]. Several ESA tools have been developed for computing a long-term mission plan for each mission, such as AIMS [24] and Local Search Solution [19] for INTEGRAL mission or the scheduler of XMM-Newton mission [2]. Outside ESA we can find SPIKE, a well-known toolkit used to promote scientific research through the effective and efficient use of ground- and space-based astronomical observatories, which is used in the long-term planning of the Hubble Space Telescope [17]. All these tools use local search algorithms and specific heuristics for finding a long-term schedule of the mission, so they may find a solution in a local optimum that is far of the global optimum [26].

Global search algorithms, such as GAs, are powerful techniques for obtaining optimal solutions to a specific problem. The successful application of these optimization techniques has been properly demonstrated in scheduling problems [4, 28] and in other domains [7, 18, 22, 23, 25]. Other tools based on applying GAs in long-term scheduling of space missions are MrSPOCK for MARS EXPRESS mission [3] of ESA, the evolution of SPIKE that will be used in James Webb Space Telescope [12] and the scheduler of SOFIA mission [5]. To complement these examples, a summary of other planning and scheduling tools used in astronomical observatories can be found in [6].

3 Long term mission planning tool in the operational design

Regardless of the algorithm used to plan the observations, the scheduling time-cycle is a critical ingredient in space missions to determine the best design approach. Figure 1 illustrates the interaction of the LT-MPT with the other control modules in the ground segment data flow of EChO. It indicates a high level of interaction between the LT-MPT, the Science Operation Centre (SOC), the Mission Operation Centre (MOC), the Instrument Operations and Science Data Centre (IOSDC), and

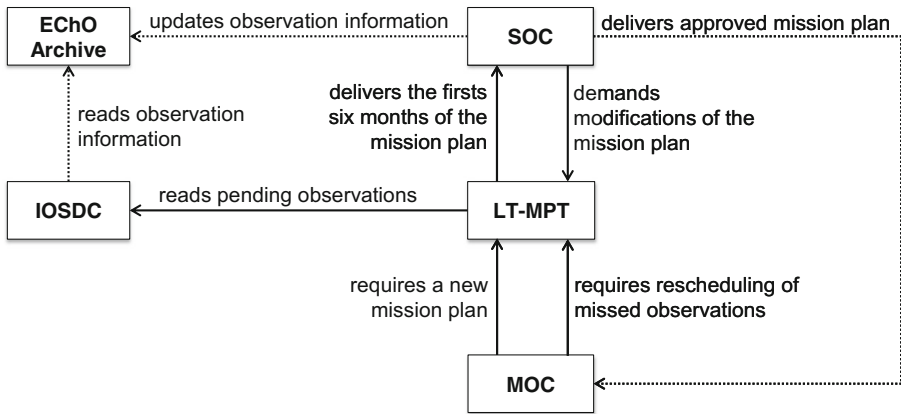


Fig. 1 LT-MPT interaction with the other elements that communicate with it, which are Science Operation Centre (SOC), Mission Operation Centre (MOC), Instrument Operations and Science Data Centre (IOSDC) and EChO Archive. The origin of the arrow indicates the element that prompts the action

the EChO Archive. It can be observed that SOC updates the EChO Archive after transferring data from the spacecraft to the station on Earth (downlink process), and manages the information making it available for the IOSDC. The EChO Archive sets as resolved all correct observations, so they will no longer be considered by the LT-MPT. Therefore, the LT-MPT is able to plan the remaining part of the mission when necessary, in order to add new objects to the set of exoplanets or to repeat some exoplanet observations.

Two types of interactions can be identified in the operational design between MOC and LT-MPT. The first one happens every six months and aimed at on building a new optimal mission plan according to the remaining time of the mission, and delivering the first six months to SOC. The second type of interaction is designed to respond to an unexpected problem in the observation of the planned events. In this case, the LT-MPT tries to reschedule the wrong observations in the current mission plan by placing them in the gaps between observations or replacing lower priority targets. Also, when LT-MPT delivers to SOC the first six months of the mission plan, SOC can propose some modifications (usually related to the scheduling of the downlink communications).

4 EChO mission optimization problem

The aim of this section is to present the main aspects of the mission that have to be considered for the LT-MPT design.

4.1 Operation tasks

The EChO mission will have to deal with a variety of observation patterns that are described as follows:

- Science observations are the observations of target events. A target event is defined as a time interval when the exoplanet is transiting its host star or is hidden by it.
- Downlink communication is used for transferring data from the spacecraft to stations on Earth. The design solution for nominal communications during science operations phase is a fixed antenna, requiring special communication attitudes during downlink. A downlink communication has to be planned every 3.5 days, with some flexibility, and has a duration of 2 hours.
- Station keeping operations are defined to keep the spacecraft in the assigned orbit. They are initially defined to be carried out every 28 days, with some flexibility, and with a duration of 8 hours.
- Calibration tasks based on several calibration items (e.g., instrument noise, instrument absolute wavelength or instrument pointing) will be done. One hour of calibration tasks should be planned each day if possible.

Several operation tasks have to be done in fixed slots of time and they involve a temporary stop of the scientific operations. Thus, any possible collision between them must be solved.

4.2 Constraints

A scheduling process can be considered a constraint satisfaction problem, which is a mathematical problem defined as a set of objects whose state must satisfy a number of constraints or limitations. Two kinds of constraints are identified: hard constraints and soft constraints. The first ones have to be necessarily satisfied, and the other ones express a preference of some solutions over other ones. Thus, the final scheduling solution must fulfil all the hard constraints and it should optimize the soft ones.

4.2.1 Hard constraints

The hard constraints identified in the EChO mission are:

- **Orbital Constraint:** The satellite orbit constraints the visibility of exoplanets (see Fig. 2). Thus, it must be considered by the LT-MPT when computing the suitability for the observation of a target event.
- **Transit Constraint:** The strategy to execute the science observations depends on exoplanet transit or occultation events and their duration. The exact occurrence of an event (T_c) can be calculated in advance. The duration of an event T_{obs} , see (1) results from T_{14} , which is the time between first and fourth eclipse contacts (i.e., time between the beginning and the end of an eclipse), and an additional time (T_b) devoted to determine the flux baseline. T_b is split in two time intervals: T_{before} , time allocated before the beginning of the eclipse, and T_{after} , time interval after the end of an eclipse. We consider the time window of an exoplanet as the duration of an event that is visible from the telescope (i.e., when the orbital constraint of the exoplanet is achieved). Figure 3 shows a typical target event.
- **Target Completeness Constraint:** This constraint is related to the science observations. In terms of scientific interest, only the observations of complete targets

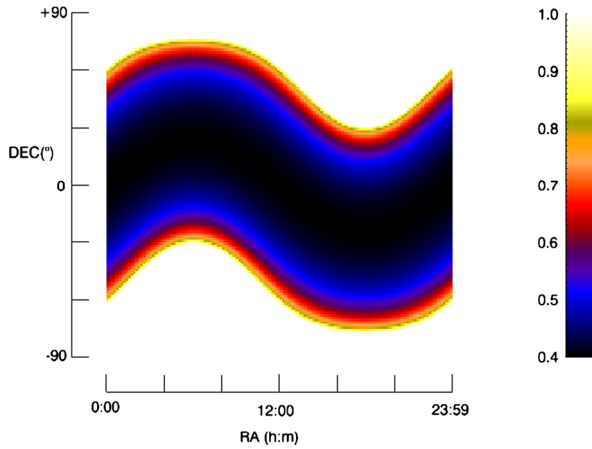


Fig. 2 Overall sky visibility for a roll angle of 36°

are useful. A target is complete when it is observed a minimum number of times. Presently, this value is adopted as 80 % of the required number of event observations of each target.

- **Slewing Constraint:** Pointing to a particular target and acquiring data requires a specific configuration. Thus, time to transfer to a new configuration must be taken into account when computing the mission planning. This reconfiguration time mainly depends on the slewing speed of the satellite.
- **Overlapping Constraint:** A key constraint in the EChO mission is that the telescope can only do one task at a time. Thus, the LT-MPT must plan the operation tasks avoiding overlapping between them (see Fig. 4), including the time to reconfigure the system (i.e., slew time).

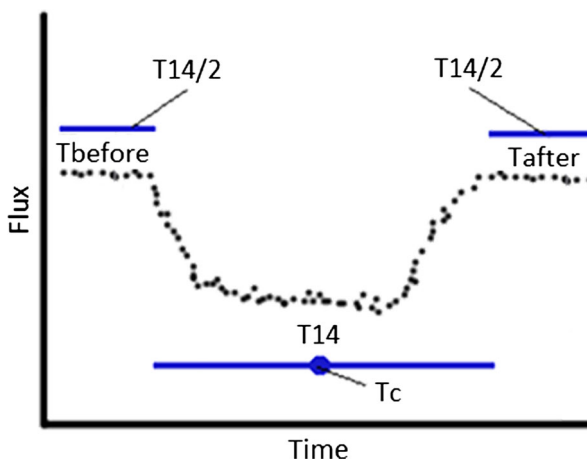


Fig. 3 Transit light curve of an exoplanet with the total observation time of its event marked with a solid line

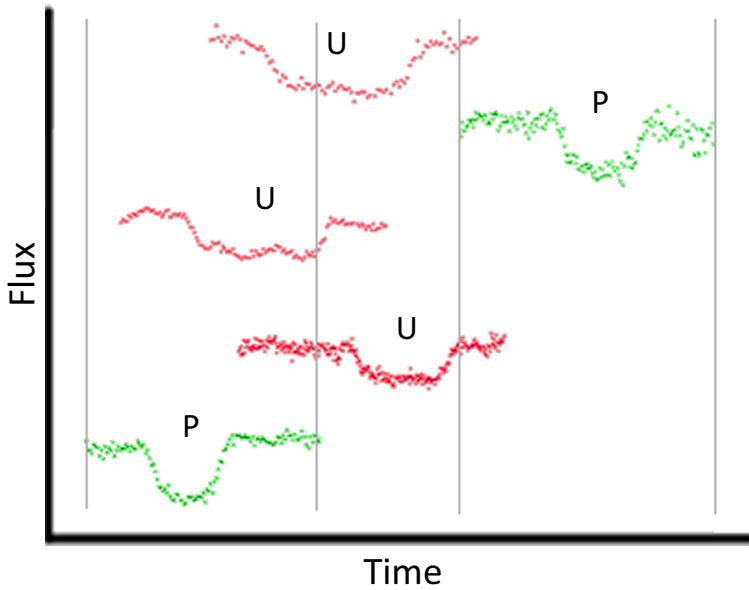


Fig. 4 Transit light curves of different exoplanets. The best target at a given time has to be selected to avoid overlapping. In this case, the planned targets are marked with *P* and the unplanned ones with *U*

- **Contamination Constraint:** To avoid contamination of the targets, these should be observed when the brightness of the Moon or light from other objects does not affect significantly the measurements.

$$\begin{aligned}
 T_b &= T_{before} + T_{after} \\
 T_{before} &= \frac{T_{14}}{2} & T_{obs} &= T_{14} + T_b \\
 T_{after} &= \frac{T_{14}}{2} & &= 2 \cdot T_{14}
 \end{aligned}
 \tag{1}$$

4.2.2 Soft constraints

In addition to the hard constraints identified, some soft constraints related to promote the exploitation of the resources and the scientific return can be defined. Each soft constraint is explained in the next points:

- Promotion of the time that the telescope is observing target events throughout the mission.
- Promotion of the reduction of the time spent in slewing for pointing to the targets in the entire mission.
- Promotion of the number of priority targets observed during the mission. The computation of the priority of each exoplanet is detailed in the next section.
- Promotion of the number of targets completed (observed, at least, 80 % of its required number of events) at the end of the mission, because the observations

of events related to incomplete targets will not be considered in the execution of the mission plan.

The optimization of these soft constraints is a key factor for obtaining a suitable mission plan because they are focused on obtaining a mission plan with an adequate exploitation of the resources and with a high scientific return. Thus, the LT-MPT proposed in the next section has the aim of obtaining a mission plan that optimizes these soft constraints and respects all the hard constraints.

4.2.3 Target priority

In EChO, the targets are classified in classes in order to guaranty that each class has some targets observed by promoting the most difficult ones to be planned. In particular, the priority of the targets is defined in two stages: (1) targets of classes with more criticality, which are the classes that are more difficult to plan in terms of number and duration of the events of their targets; and (2) within targets of the same class, the less demanding targets in terms of number of events and duration. Targets that belong to the most critical (i.e., most difficult) classes are the higher priority ones and, if several targets belong to the same class, the less critical targets (i.e., easier to observe) are the higher priority ones inside that class. A priority level is assigned to each target, this level is an increasing value starting at 1, where the lower values indicate the higher priority targets (i.e., a target with level 1 has higher priority than a target with level 2).

Finally, in terms of scientific criteria, some targets can be considered more important than other ones and should be promoted in the mission plan. This scientific priority is not considered in the current model but, if necessary, the priority level can be modified (e.g., adding new priority stages) without affecting to the process of the LT-MPT.

Note that other ad-hoc methodologies to determine target priorities can be defined. If, e.g., a specific object or a target class is deemed of great importance, individual priorities or class priority multipliers could be easily defined and accounted for. These ad-hoc target priorities are not considered in the present exercise.

5 Long term mission planning tool based on artificial intelligence

Planning of astronomical observations is an example of the classical task allocation problem known job-shop problem, where several tasks are assigned to identical resources while minimizing the total execution time. In most cases there is no single best approach to solve the planning system and, therefore, AI techniques can allow us to find a solution near to the optimal one in a reasonable time. In this section we present the LT-MPT approach for the EChO mission based on GAs [10], in spite of the standard process to be followed for building an automatic LT-MPT is independent of the AI technique used. This means that the global optimization process used can be based on other paradigms such as simulated annealing or swarm intelligence, which may perform similarly.

5.1 Long term mission planning tool process

The LT-MPT process is designed to minimize conflicts between high priority targets, downlink communications and station keeping operations, and to optimize the planning of science observations according to the constraints and objectives mentioned in Section 4. Note that the proposed process is independent from the operational design and it can be adapted to follow various approaches (i.e., order of observation and downlink placements, etc.). We can identify two main aspects based on the problem conditions described in Section 4: (1) the optimization of the positioning of downlink communications and station keeping operations focused on restricting lower priority targets, and (2) the optimization of the observation scheduling of each target event respecting the hard constraints and optimizing specific objectives. At the end of this process, a feasible and optimal Long Term Mission Plan (LTMP) is obtained. The LTMP is represented by a timeline for all the planned days (i.e., the remaining time of the mission) that indicates when each operation task will be done and its duration (see Fig. 5). Note that the LT-MPT sends to SOC only the first six months of the LTMP. Taking into account these considerations, we propose the LT-MPT process shown in Algorithm 1.

The process specified in Algorithm 1 has several steps that are explained below:

- Step 0, calculate the time windows of target events: This is an initial phase based on obtaining the windows of time where the target events can be scheduled. This

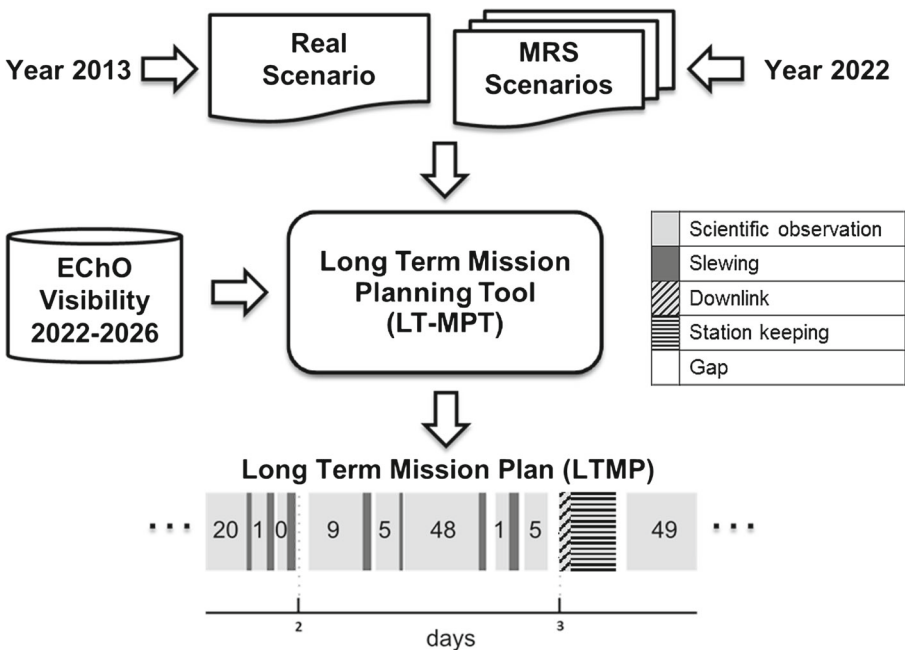


Fig. 5 Inputs needed by the LT-MPT to obtain the LTMP that indicates in a timeline the tasks to be done at each specific time

Algorithm 1 Process followed in obtaining the LTMP

Step 0. Calculate the time windows of each target event.

Step 1. Clean-up impossible targets.

Step 2. Obtain the observation planning respecting the hard constraints and promoting the soft ones.

Step 3. Drop observations of incomplete targets.

Step 4. Fill gaps with downlink and station keeping operations.

Step 5. Fill gaps with calibration tasks.

is obtained by using the target ephemeris together with the restrictions coming from the spacecraft sky visibility data. The resulting information is a list of time windows for each target. Each time window will correspond to the duration of an event of its corresponding target t , so it will have a duration of $2 \cdot T_{14_t}$.

- Step 1, clean-up impossible targets: The aim of this step is to remove targets that are unschedulable. These are targets that, according to the time available for the entire mission, cannot be observed at least $m\%$ of the requested number of events because of visibility limitations.
- Step 2, observation planning optimization: This step places the observations of target events obtaining a LTMP. This is an optimization problem that aims at planning the observations respecting the hard constraints and optimizing some specific objectives related to promote the soft constraints. The only hard constraint that can be violated in this step is the target completeness constraint.
- Step 3, drop observations of incomplete targets: This step responds to the target completeness constraint. Each of the incomplete targets (i.e., targets observed less than an $m\%$ of the requested number of events) of the LTMP obtained at Step 3 is attempted for completion, and if the target cannot be completed it is removed from the LTMP. At the end of this step, a feasible LTMP for the EChO mission is obtained.
- Step 4, downlink and station keeping optimization: This part of the process is based on planning downlink communications and station keeping operations in the gaps between the observations by respecting their periodicity (with some flexibility). SOC can request modifications related to this operations according to the ground station conditions.
- Step 5, fill gaps with calibrations: This phase is aimed at placing calibrations in the gaps of time that remain between the scheduled tasks in the LTMP obtained at Step 4. Moreover, other operation tasks or external observations can be planned in the gaps. The first six months from the resulting observation plan produced from this step are sent to SOC.

The optimization phase of Step 2 can be done by using several mathematical tools such as simple heuristics or AI techniques. We propose to apply GAs, which are able to solve optimization problems and have the ability to be adapted to new environments and constraints.

5.2 Observation planning optimization

The aim of the Observation Planning Optimization (OPO), which is the Step 2 in Algorithm 1, is to plan the science observations finding the optimal time windows for their execution, and avoiding any overlap with other operation tasks. Moreover, the LTMP has to promote the soft constraints defined in the mission. Figure 6 describes the cycle followed by the OPO algorithm. In the next points, we present the individual representation, the genetic operators, how to repair an unfeasible individual, the objective functions more suitable to this optimization problem, and how to improve a final LTMP.

5.2.1 Individual representation

In a Pittsburgh-style GA, each potential solution in the genetic process is referred as individual and its representation is based on the definition of genotype, which is a set of genes that can have different values (alleles) [1, 15]. The proposed individual genotype is made up of integer numbers which represent the time windows where the targets are planned. Each individual consists of $\sum_{t \in T} \bar{E}_t$ genes $\{o_{1,1}, \dots, o_{1,\bar{E}_1}, \dots, o_{T,1}, \dots, o_{T,\bar{E}_T}\}$, where \bar{E}_t is the number of events of target t , \bar{T} is the cardinality of the set of existing targets T , and $o_{t,i}$ corresponds to observation i of target t . Moreover, the $o_{t,i}$ value has to be between the range $[-1, \bar{W}_t - 1]$, where -1 indicates that the corresponding event does not have a window assigned (i.e., it is not planned), and \bar{W}_t is the number of potential time windows of target t . The order of the targets in the genotype does not indicate a temporal sequence, it is only the order of the targets in the input file. The temporal sequence of targets is defined by the alleles, because they indicate the time window assigned to each target observation. Moreover, the individual represents all the observations required for each target but also there is not a temporal sequence in them. For instance, an

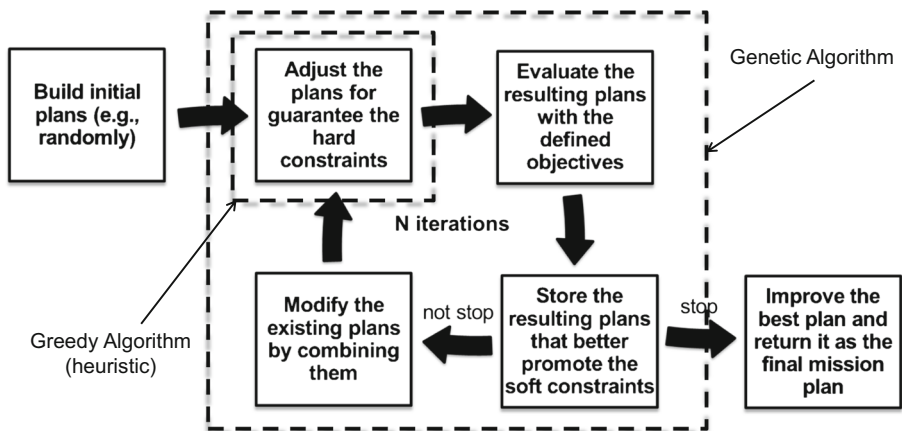


Fig. 6 Cycle followed in the OPO algorithm based on Genetic Algorithms

observation in the position i of the observations of a specific target can be planned in a time window previous to the time window of observation $i - 1$, and also observation $i - 2$ can be unassigned. An example is shown in Fig. 7, where the first part of the image describes three targets to be planned (the number of events to be observed, and the time windows where they can be observed) and the second part of the image presents a possible fragment of an LTMP (note that it is not the optimal one). In particular, the LTMP is depicted with the genotype of the individual and its interpretation, where the first two genes are referred to the first target (it has two events), and the third and fourth genes to the second and third target, respectively. The allele of each gene indicates the window assigned to each observation from the potential time windows of the corresponding target (the matching between alleles and windows of each target is shown in grey colour). Note that the third target is not planned.

Individuals consist in planning the observations of the events by assigning to each potential observation a time window where an event can be observed. However, the slew time between two consecutive observations is not considered. Therefore, the slew time must be added to the time window assigned to each observation, as described in Section 5.2.3.

The initial population is built by creating N_I new individuals assigning to each allele a value between the range $[-1, \bar{W}_t - 1]$. The process to build each individual is based on placing the observations of the targets, selected in a random order, and avoiding overlap. In case of overlapping, the event is dropped.

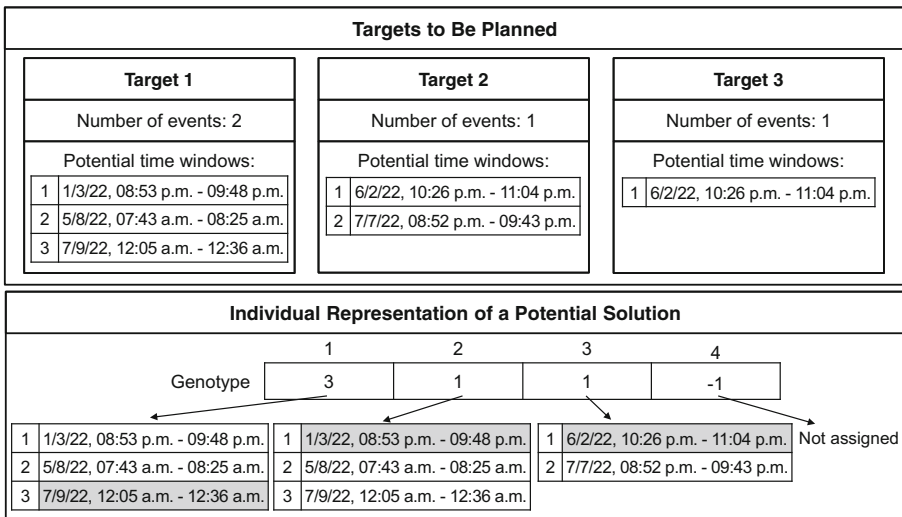


Fig. 7 The example provides some targets to be planned and a potential LTMP using the proposed individual representation. The targets to be planned are described with the number of events to be observed and the time windows where they can be observed. Each gene of the individual genotype refers to a target event, and its value (allele) indicates the window assigned to each observation from the potential time windows of the corresponding target (the grey colour indicates the matching between alleles and target windows)

5.2.2 Genetic operators

The GA process is roughly based on applying selection, reproduction (crossover) and mutation [10, 13] operators for several iterations.

A tournament selection strategy has been chosen for this problem because it is one of the most widely used selection strategies in GAs and it works efficiently for a wide range of problems [10], and our approach has been designed to use elitism in the population. The crossover and mutation operators are described as follows:

- **Crossover:** In our case, two new individuals are obtained from two previously selected parents by using uniform crossover, which is based on assigning for each gene of the first child the allele of the same gene of the first parent or the second parent with a probability of 0.5. The alleles of the genes of each parent not assigned to the first child are copied in the corresponding genes of the second child. Parents are crossed with a specific p_c probability, which means that there are some situations where parents are not crossed and the two offspring are the parents themselves.
- **Mutation:** The mutation is applied to each gene of every new individual with a probability of p_μ , which means that some genes are not mutated. Usually, p_μ is a low value because only few genes have to be mutated in order to make minor changes to the individual, which is the key of diversity. In our approach, a mutated gene g' is obtained by changing its allele with one of the potential time windows of its corresponding target. Thus, a mutated gene g' changes its allele with a random value μ between the range $[0, \bar{W}_t - 1]$ (i.e., $g' = \mu$).

Note that the crossover of two feasible individuals can generate unfeasible offspring due to overlapping, and the mutation of a feasible individual can also generate an unfeasible solution. This is solved by a repairing procedure devoted to obtain new individuals with no overlapping, as the next point explains.

5.2.3 Repair of the individual

An individual represents the time windows assigned to target observations, but it does not consider the slew time between two observations. Thus, this aspect has to be considered for obtaining the final planning codified by each individual. This modification can produce an unfeasible individual because it can have conflicting observations (i.e., presence of overlaps in the observations). There are two ways for obtaining an unfeasible individual that requires repair during the GA process: (1) the individual has overlapping time windows between two or more observations, and (2) there is overlapping between two or more observations when slew time is added to each observation. We may find that it is necessary to repair the individuals after the mutation process in order to obtain feasible ones. So, the main idea of the repair operator is to solve all overlaps in the individual by changing the window assigned to the conflicting observations. The process is based on a hill climbing strategy that follows the steps of Algorithm 2. The hill climbing strategy is applied to each conflictive event, and it is considered that an individual is improved when: (1) a new

Algorithm 2 Process followed in repairing an unfeasible individual

Step 0. Identify all conflictive events in the individual (all the events planned in overlapping windows are considered conflictive).

Step 1. For each conflictive event:

Step 1.1. Find one of the event windows that does not overlap any other planned event in the individual.

Step 1.2. If no overlapping window has been found, unassign the event (the event will not be planned).

Step 1.3. Update the conflictive events (the modification of the window assigned to the event may solve the conflicts of other events).

window with no overlapping can be assigned to the conflictive event, the plan is improved because the conflict is solved and the event is still planned, and (2) the conflictive event is unassigned, the plan is improved because the conflict is solved. The first improvement is preferred and, for this reason, all possible windows are analysed before unassigning the event.

5.2.4 Objective functions

Two objectives to be optimized are proposed in order to promote the soft constraints of the mission. These two objectives are summarized in a single fitness function F (see 2) that evaluates each individual I . The first objective (F_G) computes the time that the telescope is not observing target events and the second one (F_T) computes the number of incomplete targets weighted with their priority (having more weight the incomplete targets with higher priority). F_G is defined in (3), where o is an observation from O_I , which contains all the target observations planned by individual I . $Target(o)$ returns the target associated to the observation o , $T14_i$ is the duration of the target event in time units of target i , and T is the collection of existing targets. The denominator of F_G indicates the total input time (i.e., the required time for observing all the required events). F_T is described in (4), where $Priority(t)$ indicates the priority of target t , and \bar{T} is the cardinality of the set of existing targets. It can be noticed that F , F_G and F_T have values between 0 and 1, and they are optimized when they are minimized. So, the individual of the population that minimizes F will be the best one. When an individual is evaluated with the fitness function, it represents a feasible plan that respects all the hard constraints except the target completeness constraint. The next two sections present how to obtain the final LTMP by removing the incomplete targets and filling the gaps between observations with downlink communications, station keeping operations, calibration tasks, and additional scientific observations.

$$F(I) = \frac{F_G(I) + F_T(I)}{2} \quad (2)$$

$$F_G(I) = 1 - \frac{\sum_{o \in O_I} 2 \cdot T14_{Target(o)}}{\sum_{t \in T} 2 \cdot T14_t} \quad (3)$$

$$F_T(I) = 1 - \frac{\sum_{t \in T} \text{if } t \text{ is incomplete in } O_I, \text{ then } \bar{T} - Priority(t)}{\sum_{t \in T} Priority(t)} \quad (4)$$

5.3 Improvement of the long term mission plan

The final individual obtained with the GA is the individual that best optimizes the defined fitness function in the final population. However, although the number of complete targets is promoted, some targets can be incomplete (not observed at least $m\%$ of their required number of events). For this reason, in order to obtain a feasible LTMP, the incomplete targets have to be entirely removed from the plan, increasing the number of gaps. Thus, the feasible LTMP can still be further improved by filling these gaps with additional tasks. The filling process attempts to add observations of targets with high priority. This can result in an overallocation of events of a particular target. Therefore, we have limited the times that a target can be planned and it is calculated as a percentage ($M\%$) according to their required number of events. This process is divided in three parts:

- The first one tries to place in the gaps between observations the downlink communications and the station keeping operations required with the specified periodicity and some flexibility (defined with the parameter f , which indicates a percentage of their periodicity). If these tasks cannot be placed in gaps according to their specific periodicity, they replace the observations planned in the conflictive time ranges.
- The second one tries to place in the remaining gaps the calibrations (one hour each day).
- The third one tries to complete each incomplete target and removes it if the target is still incomplete. After this, the gaps are filled with the observations of targets already included in the LTMP, starting with events of those targets with the highest priority.

After this procedure, a feasible and optimal LTMP with the planning of all the required operation tasks is obtained.

6 Experiments, results and discussion

The main goal of this section is to empirically analyse the performance of the proposed LT-MPT. The results obtained with different experiments are used to prove the suitability of the LT-MPT in EChO and, at the same time, to demonstrate the feasibility of the mission. The experiments have been carried out in several mock scenarios and in a real sample of known exoplanets.

6.1 Experimental methodology

This section describes the test samples used in the experimentation, the configuration of the LT-MPT and the comparison metrics applied in the evaluation.

6.1.1 Test samples and constraints

One of the key aspects of the EChO mission is the need to cover a broad area in the parameter space in terms of exoplanet and host star configurations. A statistical analysis has been carried out to estimate the future available parameter space for EChO together with the number of transiting planets expected in the year 2022. The resulting hypothetical so-called Mission Reference Sample (MRS) [9] covers the full range of exoplanetary host systems that EChO can potentially observe according to current Signal-to-Noise Ratio requirements and conservative assumptions on instrument performance.

The experiments have been analysed in ten distinct artificial scenarios in order to test if the algorithm is able to obtain similar performance with different target lists. Each scenario corresponds to a realization of the MRS and each one characterizes 238 targets, which results in around 6200 events and requires about 27000 hours of observation. MRS realizations are calculated by randomizing several parameters according to astrophysically-sound assumptions.

Although the generated scenarios are created artificially, they are realistic in astrophysical terms and more complete than a list of currently known exoplanets. However, an additional scenario based on the known exoplanets (192 targets that result in 4065 target events and 22241 hours of observation) has been included in the experimentation with the aim of testing the proposed tool in the current time.

The EChO mission has a nominal duration of four years and a goal of six years. In the presented experiments, the scenarios are planned in five years (2022 – 2026) in order to allow extrapolating the results to four or six years. From these five years, the first six months are for commissioning, so the LT-MPT has to plan the scientific observations in the remaining four years and a half.

6.1.2 General test bench configuration

Some global assumptions for the problem constraints have to be specified when defining the test bench configuration:

- Tests cover 5 years (2022 – 2026), with the first 6 months for commissioning.
- 520 downlink communications with a duration of 2 hours and a periodicity of 3.5 days are considered.
- 65 station keeping operations with a duration of 8 hours and a periodicity of 28 days are assumed.
- 1825 calibrations of one hour every day.
- Slew time between observations of different targets is taken into account using a slew speed of 45 degrees per 10 minutes plus a flat 5-minute overhead.

6.1.3 Algorithm configuration

In terms of parameterization, the LT-MPT has several parameters related to the OPO algorithm and the improvement processes. Table 1 summarizes the parameter configuration used in the experiments done. Note that downlink communications and station keeping operations have a flexibility to their default position of 10 % (8.4 hours and 2.8 days respectively), which is a reasonable flexibility according to the ground station conditions.

6.1.4 Comparison metrics

Four metrics have been defined in order to evaluate and compare the performance of the obtained LTMPs in the aforementioned experiments. The first two metrics are related to the use of resources:

- **Event Observation Time:** It is computed as a percentage of the total input time (i.e., the required time for observing all the required events), which is the time covered by the whole sample.
- **Slew Time:** It is calculated as a percentage of the whole mission time (5 years).

The other two metrics are related to the scientific return:

- **Events Planned:** It is the percentage of events observed from the overall number of required events.
- **Targets Completed:** It is calculated as the percentage of complete targets from the overall sample of targets. Note that the LTMP only considers observations of complete targets, which are the targets observed at least $m\%$ (in this case, m is 80) of the required number of events.

6.2 Analysis of the experimentation results

This section presents the results obtained with the aforementioned configurations of the LT-MPT, which has been run several times using 25 different random seeds for each scenario. Each of these executions is referred hereafter as trial.

Table 1 Parameter configuration of the LT-MPT

Process	Parameter	Value
OPO	Generations	10000
	N_I	250
	N_P	1500
	p_c	0.9
	p_μ	0.005
Improvement	m	80
	M	100
	f	10

In Table 2, the results obtained for the MRS sample realizations (*MRS_rand_#*) and the real case (*Real_sample*) are described in terms of the four defined metrics (event observation time, slew time, events planned and targets completed). The mean and deviation values of each metric for the 25 trials are given. Moreover, the average of each metric for the sample realizations is shown (*MRS_rand Avg.*).

Table 2 shows that the results obtained in each scenario have a small deviation, so the GA is able to explore similar regions of the search space in each trial. Moreover, results show that there are slight differences between the results obtained in the sample realization scenarios because *MRS_rand Average* presents a small deviation in the four analysed metrics. Therefore, the LT-MPT is able to obtain solutions of similar quality in different scenarios.

It is important to compare the performance of the LT-MPT based on GAs with a LT-MPT based on heuristics. The proposed heuristic is similar to the way in which an operator plans the observations. The main idea is to place the observations of the targets in priority order avoiding overlaps. A target observation that cannot be planned without overlap is dropped. Finally, the obtained LTMP is improved by filling the gaps and removing the targets that do not have at least 80 % of the required events observed. The obtained results show that the LT-MPT based on GAs improves the Event Observation Time around 15 %. This implies that the spacecraft is collecting photons around 3500 hours more in the GA approach than in the heuristic approach. Moreover, the GA approach observes 30 % more events, so the observations are more efficiently planned. Therefore, the GA approach completes nearly 40 % more of targets than the heuristic approach and, consequently, the result based on GA has significantly more interest from a scientific point of view. The next two sections analyse in more detail the results of the LT-MPT based on GAs.

Table 2 Results obtained (mean and deviation) with the LT-MPT with DSKO and OPO in the defined scenarios after 25 trials

Name	Event			
	Observation Time	Slew Time	Events Planned	Targets Completed
MRS_rand_0	88.08% ±0.98	5.44% ±0.04	93.28% ±0.54	96.69% ±0.34
MRS_rand_1	94.55% ±0.74	5.26% ±0.04	96.62% ±0.60	98.75% ±0.37
MRS_rand_2	87.49% ±0.90	5.17% ±0.06	90.75% ±0.78	97.99% ±0.25
MRS_rand_3	80.21% ±1.47	5.22% ±0.11	85.03% ±1.36	95.50% ±0.64
MRS_rand_4	94.22% ±0.94	5.44% ±0.04	95.97% ±0.64	98.93% ±0.40
MRS_rand_5	87.91% ±1.32	5.27% ±0.07	91.72% ±1.03	97.29% ±0.32
MRS_rand_6	88.48% ±0.92	5.14% ±0.02	93.50% ±0.35	97.46% ±0.35
MRS_rand_7	89.38% ±0.83	5.06% ±0.03	93.64% ±0.36	97.44% ±0.36
MRS_rand_8	94.35% ±0.98	5.29% ±0.05	96.34% ±0.59	99.15% ±0.21
MRS_rand_9	91.89% ±1.40	5.49% ±0.07	94.22% ±1.12	98.81% ±0.43
MRS_rand Avg.	89.66% ±1.05	5.28% ±0.05	93.11% ±0.74	97.80% ±0.37
Real_sample	98.70% ±0.02	2.43% ±0.01	98.90% ±0.01	100.00% ±0.00

6.3 Use of resources analysis

In terms of the event observation time metric, the telescope is observing between 80 % and 95 % of the total input time (i.e., the required time for observing all the required events) for each sample realization scenario and around a 99 % of the time in the real sample. The duration of the available 4.5 years of the mission is 39420 hours, but one of the requirements of the mission is to allow around 20 % of the available time for doing tasks of missions unrelated with the EChO goal. Thus, 31536 hours are free for doing all the operation tasks required by EChO. The overall time that EChO will be working (observations, slews, downlink communications, station keeping operations and calibrations) is, in average, 29662 hours in the artificial scenarios (24300 hours of target event observations, 1643 hours of calibrations, 1408 hours of downlink communications and station keeping operations, and 2312 hours of slewing). These working time is equal to the 94 % of the free time of the mission (29662 hours worked of the 31536 hours available). Thus, due to the fact that these scenarios describe the hypothetical list of exoplanets to be observed in 2022, we can conclude that the optimization of the use of resources will be achieved.

Moreover, an interesting aspect to analyse is the number of gaps between operation tasks and their duration in the final LTMP (i.e., when all the operation tasks are planned). Figure 8 shows the number of gaps for several gap slot sizes in minutes for the artificial scenarios. It can be observed that there are a small number of gaps shorter than 10 minutes and longer than 300 minutes (5 hours). Thus, it seems that the gaps between observations are somehow commensurate to the visibility of each target event. This is supported by the fact that the algorithm obtains similar gap distribution for each scenario. Figure 9 shows the total duration of the gaps of each gap slot size in the artificial scenarios. It can be observed that the gap duration can slightly change in each scenario but they follow a similar pattern. As result of this analysis, it can be concluded that the resulting LTMP has around 9000 hours of gaps of considerable size that can be used for doing tasks unrelated with the EChO goal.

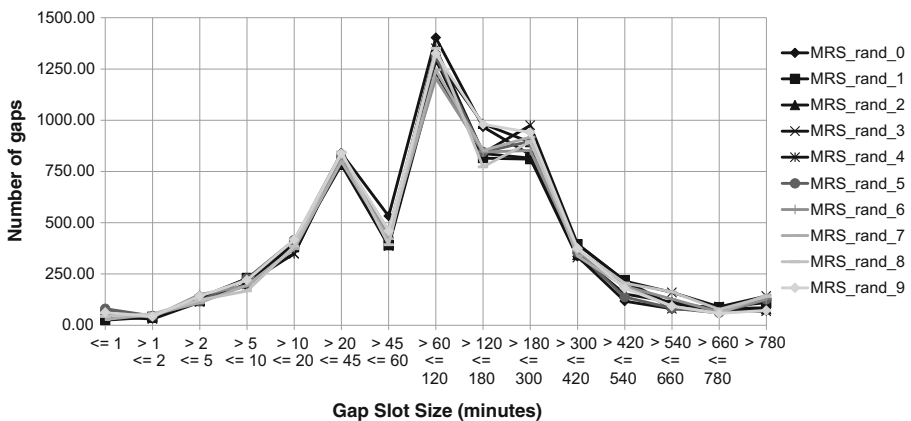


Fig. 8 Average number of gaps for each gap slot in the artificial scenarios

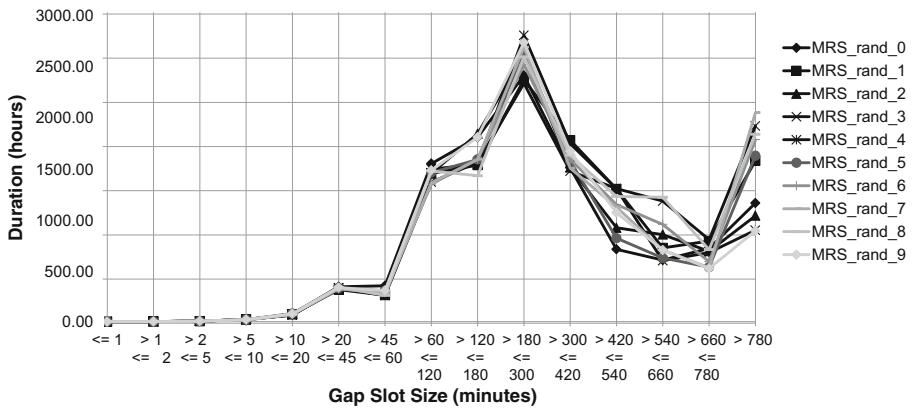


Fig. 9 Average duration in hours of the gaps of each gap slot in the artificial scenarios

6.4 Scientific return analysis

Table 6.4 shows that in the sample realization scenarios around a 93 % of the required observations are planned and about a 98 % of the targets are completed. In the case of the real sample, almost 99 % of the events are planned and 100 % of the targets are completed. These results show that the majority of the observations are fulfilled and almost all the required targets are completed (all these targets are observed for at least 80 % of their required events). Therefore, it is clearly shown that the optimization of the scientific return is also achieved due to the high number of target events observed and of targets completed.

Furthermore, an important issue to be analysed from a scientific point of view is the completeness of each target class, which is the number of targets completed for each class, and it is related with the priority of the targets. Figure 10 shows the average number of targets that belong to each class (first bar, in black colour), and the average number of targets planned for each one (second bar, in grey colour) in the artificial scenarios with the 25 trials executed. It can be observed that all classes have a high class completeness rate, even those that contain more targets.

6.5 Computational cost

The LT-MPT has been executed in a CPU Intel Core 2 Duo Processor E6600 2.40 GHz with 6GB of RAM, and the planning results of one trial are obtained in approximately 45 minutes for the sample realization scenarios and in 15 minutes for the real sample scenario.

The computational cost of the LT-MPT depends on the number of generations, the number of individuals evaluated in each generation (in our case, two individuals) and the cost of the evaluation of the objective functions and the repair of the individual. According to the individual representation described in Section 5.2.1, the total number of events required (i.e., sum of events to be observed for each target) is the aspect that affects the computational cost of the algorithm because the size of the individual

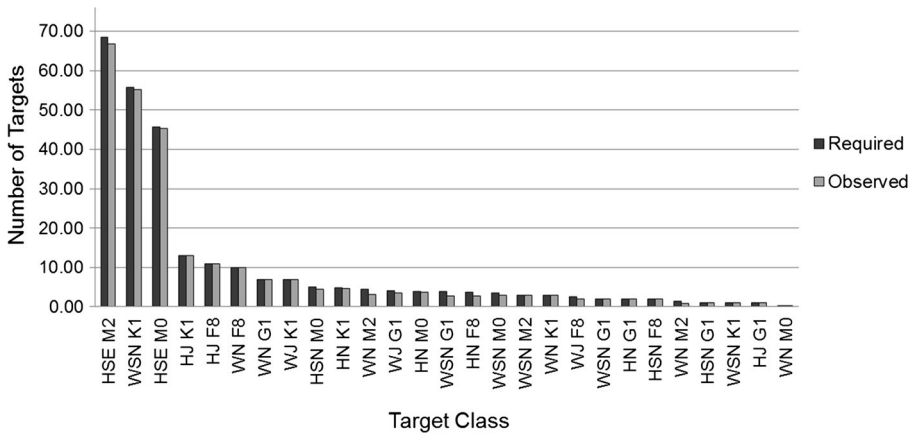


Fig. 10 Average of the target class completeness in the artificial scenarios. For each class, the first bar (in black colour) indicates the number of targets that belong to it, and the second bar (in grey colour) indicates the average number of targets finally planned

depends on it. The complexity of the evaluation of each individual described in terms of Big-O notation [21] is $O(2 \cdot e)$, where e is the total number of events required for all the targets; and the complexity of the repair algorithm of each individual is $O(e^2)$. In consequence, the cost expression of the LT-MPT is $g \cdot 2 \cdot (e^2 + 2 \cdot e)$, where g is the number of generations, and 2 refers to the number of individuals evaluated in each generation. Considering a high number of events, this results in a complexity of $O(e^2)$, which means that the LT-MPT has a quadratic cost.

On the other hand, to guarantee that the optimal plan is found, it is necessary to use a brute-force search algorithm. This algorithm will repair and evaluate each possible combination of the observation of events, and the cost expression will be $w^e \cdot (e^2 + 2 \cdot e)$, where w is the average number of windows where each event can be observed. Considering a high number of events, this results in a complexity of $O(w^e)$, which is an exponential cost and the computation of the algorithm will be unachievable.

In brief, the LT-MPT based on GAs finds a good solution in terms of use of resources and scientific return in a reasonable computing time.

7 Conclusions and further work

The allocation of tasks, by optimizing different objectives and respecting some constraints, is a key point for the success of astronomical space missions. A proper planning of the observations for EChO can represent a big difference in its efficiency and ultimate scientific performance. However, an enormous amount of possible combinations exists and a manual search for an optimal solution becomes unrealistic. Hence, it will be necessary to increase the efficiency of telescope operation by means of automatic processes that can find solutions close to optimal. Artificial Intelligence techniques can be applied in such optimization problems with high mathematical and

computational complexity. In the first part of the document we have identified the role of the LT-MPT into the EChO operations design and we have discussed the main considerations of the EChO mission to be taken into account by the planning tool. In the second part of the document we have proposed a process to be followed for an automatic planning of the mission. The steps of this process are mainly based on (1) calculating the windows of time where each target event is visible, (2) cleaning up targets that cannot be observed the required number of times, (3) planning the target observations by respecting the hard constraints and maximizing the soft constraints, (4) removing observations of the targets that are not planned at least $m\%$ of their required events, and (5) filling gaps with downlink communications, station keeping operations and calibration tasks. The key step of this process is the observation planning, which is considered an optimization problem. The optimization algorithm proposed is called OPO and it defines a process with a hybrid optimization step [14] based on (1) a global search carried out with Pittsburgh-style GAs focused on planning target observations, and (2) a local search for guarantee the feasibility of the potential solutions. Moreover, in GAs the definition of the individual (i.e., knowledge representation) and the genetic operators are a key factor for obtaining suitable results [11, 16]. For this reason, the proposed algorithm makes special emphasis in these two aspects in order to properly guide the exploration of the search space to the region where the optimal solutions are.

A detailed experimentation of the LT-MPT has been presented and discussed. Specifically, the LTMPs obtained for five year mission in ten different artificial samples (sample realizations) and in a real exoplanet sample have been analysed. The results are very promising and show that the proposed LT-MPT is useful to solve this problem at a high level of optimization. In particular, the experimentation has shown that the proposed LT-MPT is able to plan observations of almost all the targets. Moreover, it has been shown that the process is able to explore similar regions of the search space in each trial and it obtains the results in a reasonable computing time. It obtains solutions of similar quality in different scenarios, so our proposal is a robust LT-MPT whose performance does not depend on the exoplanet sample. This is an important property because it is expected that the ongoing and future transit search experiments will continue discovering new targets of EChO interest, adding to the available list of exoplanets. Besides establishing a proper design of the LTMP, the experiments that we have conducted, and described here, have been used to size the EChO mission, with the aim of guaranteeing its feasibility (duration of survey, planning efficiency, slew statistics).

Finally, the proposed LT-MPT methodology is applicable to other ground and space missions because the optimization process does not use any heuristic that can limit the search of the plan, so it is only necessary to adapt the representation of the individual and the genetic operators to the problem to be solved.

Acknowledgments This work has been supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and the "Fondo Europeo de Desarrollo Regional" (FEDER) through grant AYA2012-39612-C03-01.

References

1. Bacardit, J.: Pittsburgh genetic-based machine learning in the data mining era: representations, generalization, and run-time. PhD thesis, *ingeniería i arquitectura la salle* (2004)
2. Castellini, F., Lavagna, M.: Advanced planning and scheduling initiative XMAS tool: aI for automatic scheduling of XMM-newton long term plans. in the 6th international workshop on planning and scheduling for space, IWPSS-09 (2009)
3. Cesta, A., Cortellessa, G., Fratini, S., Mrsposck, A.O.: A long-term planning tool for Mars express in the 6th international workshop on planning and scheduling for space, IWPSS-09 (2009)
4. Chung, S.H., Chan, H.K.: A two-level genetic algorithm to determine production frequencies for economic lot scheduling problem. *IEEE Trans. Ind. Electron.* **59**(1), 611–619 (2012)
5. Civeit, T.: Automated long-term scheduling for the sofia airborne observatory in aerospace conference, *IEEE* (2013)
6. Colome, J., Colomer, P., Guàrdia, J., Ribas, I., Camprecis, J., Coiffard, T., Gesa, L., Martínez, F., Rodler, F.: Research on schedulers for astronomical observatories. *Proc. SPIE* **8448**, 84481L–84481L–18 (2012)
7. Corral, G., Garcia-Piquer, A., Orriols-Puig, A., Fornells, A., Golobardes, E.: Analysis of vulnerability assessment results based on CAOS. *Appl. Soft Comput. J.* **11**, 4321–4331 (2011)
8. Donati, A., Reinhold, B., Martinez-Heras, J.A., Policella, N.: Why introducing innovative technology in operations? In The 12th international conference on space operations (2012)
9. European Space Agency: EChO assessment study report (Yellow Book) (2013)
10. Freitas, A.A.: *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2002)
11. Garcia-Piquer, A., Fornells, A., Bacardit, J., Orriols-Puig, A., Golobardes, E.: Large-scale experimental evaluation of cluster representations for multiobjective evolutionary clustering. *IEEE transactions on evolutionary computation*, in press
12. Giuliano, M.E., Hawkins, R., Rager, R.: A status report on the development of the JWST long range planning system. In the international workshop on planning and scheduling for space, IWPSS (2011)
13. Goldberg, D.E.: *Genetic algorithm in search, optimization, and machine learning* addison-wesley (1989)
14. Harman, M., McMinn, P.: A theoretical and empirical study of search-based testing: Local, global, and hybrid search. *IEEE Trans. Softw. Eng.* **36**(2), 226–247 (2010)
15. Holland, J.H.: *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor (1975)
16. Hruschka, E.R., Campello, R.J.G.B., Freitas, A.A., de Carvalho, A.C.P.L.F.: A survey of evolutionary algorithms for clustering. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **39**(2), 133–155 (2009)
17. Johnston, M., Miller, G.: *Intelligent scheduling chapter spike: intelligent scheduling of hubble space telescope observations*. Morgan Kaufmann Publishers (1994)
18. Khodadad, F.S., Jahed, M.: Optimization of a cascading tmr system configuration using genetic algorithm. In The 10th IEEE international conference on Industrial Informatics (INDIN) (2012)
19. Kitching, M., Policella, N.: A local search solution for the INTEGRAL long term planning. In In proceedings of the 7th international workshop on planning and scheduling for space, IWPSS11 (2011)
20. Kitching, M., Policella, N.: An automated approach to support alphasat TDP operations. In The 12th symposium on advanced space technologies in robotics and automation (2013)
21. Knuth, D.E.: *The Art of Computer Programming, Volume 1: Fundamental Algorithms*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA (1997)
22. Liao, L.: Discovering prognostic features using genetic programming in remaining useful life prediction. *IEEE Trans. Ind. Electron.* **61**(5), 2464–2472 (2014)
23. Nguyen, V.D., Nguyen, T.T., Nguyen, D.D., Lee, S.J., Jeon, J.W.: A fast evolutionary algorithm for real-time vehicle detection. *IEEE Trans. Veh. Technol.* **62**(6), 2453–2468 (2013)
24. Pralet, C., Verfaillie, G.: AIMS: a tool for long-term planning of the ESA INTEGRAL Mission. In The 6th international workshop on planning and scheduling for space, IWPSS-09 (2009)
25. Qiao, F., Ma, Y., L., Yu, H.: A petri net and extended genetic algorithm combined scheduling method for wafer fabrication. *IEEE Trans. Autom. Sci. Eng.* **10**(1), 197–204 (2013)
26. Tawarmalani, M., Sahinidis, N.V.: *Handbook of global optimization, chapter Exact algorithms for global optimization of mixed-integer nonlinear programs, pp. 65–86*. Kluwer Academic Publishers (2002)

27. Tinetti, G., et al.: EChO - Exoplanet Characterisation Observatory. *Experimental Astronomy*, this volume (2014)
28. Yahyaoui, A., Fnaiech, N., Fnaiech, F.: A suitable initialization procedure for speeding a neural network job-shop scheduling. *IEEE Trans. Ind. Electron.* **58**(3), 1052–1060 (2011)