Exoplanet Characterisation Observatory (EChO)

Assessment Phase Payload Study

EChO Long Term Mission Planning Tool

ECHO-TN-0001-ICE

Issue 03

Prepared by: Alvaro Garcia-Piquer
Ignasi Ribas
Josep Colomé

Date: 13/09/2013

Checked / Approved by:

Date:

DOCUMENT CHANGE DETAILS

Issue	Date	Page	Description Of Change	Comment

DISTRIBUTION LIST

EChO Payload Consortium				External	
Co-PIs / Science Team Coordinators		Study Engineering Team Working Group Leads		European Space Agency	
✓	Giovanna Tinetti	✓	Paul Eccleston		Philippe Escoubet
	Hans Ulrik Nørgaard-Nielsen		Ranah Irshad		Ana Heras
	Jean-Philippe Beaulieu		Emanule Pace		Didier Martin
	Paul Hartogh		Gianluca Morgante		Kate Isaak
	Giusi Micela		Berend Winter		Ludovic Puig
	Ignasi Ribas		Marc Ferlet		Martin Linder
	Bruce Swinyard		Mercedes Lopez-Morales		Nicola Rando
	Mark Swain		Vince Coudé du Foresto		
	Tanya Lim		Alberto Adriani		
	Neil Bowles		Jean-Michel Reess		
	Enzo Pascale		Marc Ollivier		
	Gillian Wright		Gonzalo Ramos Zapata		
	Graziella Branduardi-Raymont		Neil Bowles		
	Marc Ollivier	**Other Engineering Team**			
	Pierre-Olivier Lagage		*As necessary for doc*		
	Vince Coudé du Foresto				
	Athena Coustenis				
	Emanuele Pace				
	Giuseppe Piccioini				
	Giuseppe Malaguti				
	Alessando Sozzetti				
	Maria Rosa Zapataro Osorio				
	Mercedes Lopez-Morales				
	Enric Palle				
	Christopher Jarchow				
	Denis Grodent				
	Allan Hornstrup				
	Geza Kovacs				
	Pierre Drossart				
	T Encrenaz				
	L Fletcher				
	D Pinfield				
	J Cho				
	F Forget				
	I Waldmann				
	P Deroo				
	I Mueller-Wodarg				
	F Selsis				
	O Grasset				
	L Stixrude				
	T Guillot				
	Others…				

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

1 PREAMBLE

1.1 SCOPE

This document is focused on the Long Term Mission Planning Tool (LT-MPT) for EChO based on Artificial Intelligence in the form of Genetic Algorithms.

1.2 PURPOSE

The purpose of this document is to present the design and implementation of the proposed LT-MPT and to explain how the Artificial Intelligence is introduced in it. Moreover, the document analyses the results obtained in different artificial and real scenarios in order to conclude if an automatic planning tool based on Genetic Algorithms can be useful in the EChO mission

1.3 INTELLECTUAL PROPERTY RIGHTS

The content of this document is property of the Institut de Ciències de l'Espai (CSIC-IEEC), and it is confidential. This document cannot be neither copied, totally or partially, nor shown to other people without the previous written authorization of Institut de Ciències de l'Espai (CSIC-IEEC).

1.4 ACRONYMS AND DEFINITIONS

AI	Artificial Intelligence
DSKO	Downlink and Station Keeping Optimization
EChO	Exoplanet Characterisation Observatory
GA	Genetic Algorithm
ICE	Institut de Ciències de l'Espai (CSIC-IEEC)
IOSDC	Instrument Operations and Science Data Centre
LT-MPT	Long Term Mission Planning Tool
LTMP	Long Term Mission Plan
MOC	Mission Operation Centre
MOEA	Multiobjective Evolutionary Algorithm
MRS	Mission Reference Sample
OPO	Observation Planning Optimization
SOC	Science Operation Centre
T14	Duration of the event of an exoplanet
TBC	To be confirmed
TBD	To be defined

1.5 APPLICABLE DOCUMENTS

AD #	APPLICABLE DOCUMENT TITLE	DOCUMENT ID	ISSUE / DATE
1			
2			
3			

1.6 REFERENCE DOCUMENTS

RD #	REFERENCE DOCUMENT TITLE	DOCUMENT ID	ISSUE / DATE
1	EChO – Science Requirements Document	SRE-PA/2011.037	3 (14/09/12)
2	EChO – Mission Requirements Document	SRE-PA/2011.038	3 (14/09/12)
3	EChO – Science Operations Assumptions Document	ECHO-SA-DC-0001	5 (06/05/13)

2 INTRODUCTION

The Exoplanet Characterisation Observatory (EChO) is one of the ESA M3 mission candidates currently in assessment for potential launch in 2022. EChO will be the first dedicated mission to investigate the physics and chemistry of exoplanetary atmospheres. The primary objective 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) secondary eclipse and (2) transit. 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 observations cannot be done simultaneously. In addition, several criteria have to be considered to carry out each observation: (1) target visibility, (2) time and duration of events, (3) number of events to be observed, and (4) target priority.

A suitable mission plan is expected to increase the efficiency of telescope operation, which will represent an important benefit in terms of scientific return and operational costs. Nevertheless, the planning for this mission has several constraints that must be respected for fulfilling the mission objectives. Thus, this process becomes unaffordable for human planers due to the complexity in computing the huge amount 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. In this contribution we present a Long Term Mission Planning Tool (LT-MPT) for EChO based on Evolutionary Algorithms (EAs), which are an AI approach focused on emulating natural evolution by means of combining potential solutions using selection, combination and mutation operators. This kind of algorithms attempts to generate solutions to optimization problems by exploring a large amount of potential solutions, including the most efficient ones. A solution is considered efficient when it highly optimizes the objectives defined in the problem that, in our case, correspond to maximizing the planning efficiency and the scientific return, measured in terms of the coverage of the mission sample.

The remainder of the document is organized as follows. In Section 3 we describe the planning tool in the operational design of the mission. In Section 4 we present the EChO mission planning optimization problem. In Section 5 we discuss the proposed approaches for the EChO planning tool. In Section 6 we describe the experimentation done and discuss the results. In Section 7 we present the Rosetta Stone targets and how to deal with them. Finally, in Section 8 we enumerate the main conclusions. In addition, in Annex A we include the main concepts of the AI techniques used, and in Annex B and Annex C we describe other approaches analysed for the EChO planning tool that where ultimately discarded.

3 LONG TERM MISSION PLANNING TOOL IN THE OPERATIONAL DESIGN

The main purpose of the mission planning tool is the allocation of tasks, while optimizing different objectives like minimizing operation time overheads or maximizing the mission scientific return. This planning tool is a key element in the control layer for the observatory time optimization. The large complexity of the process to handle the observing constraints (e.g., windows) for every target in the mission survey raises a big challenge concerning the scheduling of observations. This process must be carried out taking into account, for instance, the spacecraft configuration (e.g., attitude), the operation tasks and the state of the housekeeping variables. Suitable planning algorithms are needed to achieve the project goals and make optimum use of the instrumentation. Artificial Intelligence techniques based on optimization, such as Genetic Algorithms, Ant Colony Optimization or Multi-Objective Evolutionary Algorithms, can be useful to solve this kind of problems of high mathematical and computational complexity. The algorithm selection process must consider the scheduling/planning problem that best fits the observatory characteristics (e.g., job-shop problem).

In addition to the algorithm used to plan the observations, the scheduling time-cycle is a critical ingredient to determine the best design approach. Figure 3.1 illustrates the interaction of the LT-MPT with the other control modules in the ground segment data flow. 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 the EChO Archive. It can be observed that SOC updates the EChO Archive after the 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 targets (TBD) or to repeat some target observations.

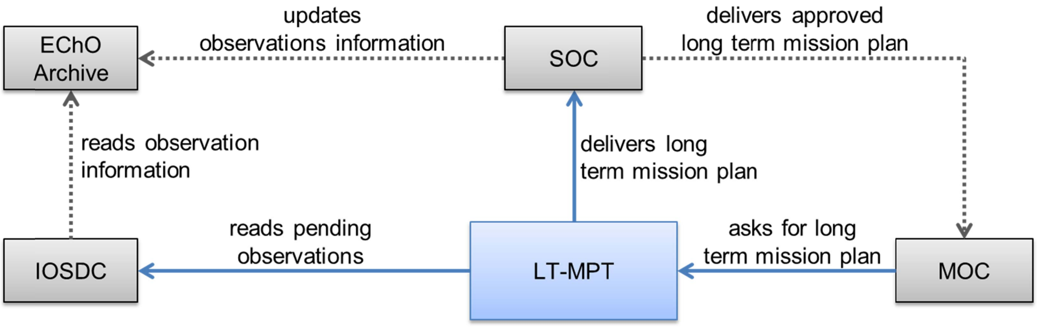

Figure 3.1. Basic interaction between LT-MPT and the other elements that communicate with it (SOC, MOC, IOSDC and EChO Archive). The origin of the arrow indicates the element that makes the action and the solid ones mean direct interaction between the LT-MPT and the other elements.

Two types of interactions can be identified in the operational design. The first one is focused on building a LTMP for being processed in the following six months (see the diagram sequence presented in Figure 3.2). The role of the functions presented in the diagram is described as follows:

- askForLTMP(): used by MOC for requiring a new mission plan to the LT-MPT.
- askForPendingObservations(): LT-MPT uses this action for getting the pending tasks from the IOSDC .
- calculateLTMP(observations): used by the LT-MPT for planning the required observations according to the remaining time of the mission. Downlinks and station keepings are not positioned.
- deliverPlan(LTMP): the LT-MPT uses this function for delivering the computed LTMP to SOC. The LTMP is optimized according the remaining time of the mission, but only the first six months are delivered to SOC.

- placeDownlinks(LTMP): SOC places downlinks and station keepings in the gaps between observations of the received LTMP. If any downlink or station keeping needs to remove an event observation, it is removed.
- deliverCompletePlan(LTMP): SOC uses this action for sending the LTMP with downlinks and station keepings to the LT-MPT.
- optimizeLTMP(LTMP,observations): the LT-MPT places in the LTMP the observations of the orbital phase curves of the rosetta stones, the calibrations and the observation of the priority events removed by SOC for positioning downlinks and station keepings.
- deliverOptimalPlan(LTMP): the LT-MPT uses this function for delivering the optimal LTMP to SOC.
- approvePlan(LTMP): used by the SOC for analysing the correctness of the LTMP and modifying some issues if necessary.
- deliverApprovedPlan(LTMP): SOC uses this action for sending the approved LTMP to MOC.
- processApprovedPlan(LTMP): MOC processes the LTMP.

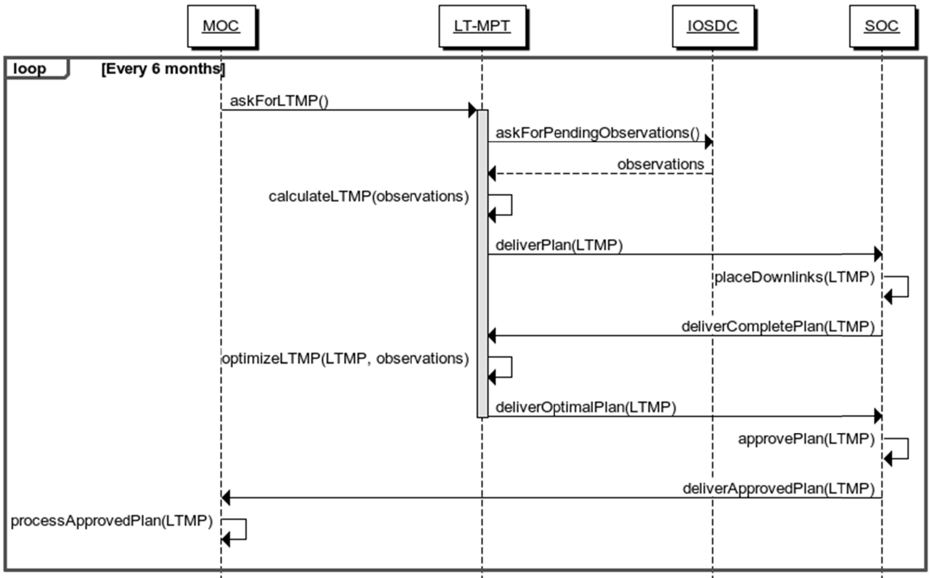

Figure 3.2. Sequence diagram of planning and sending an LTMP to MOC.

The second type of interaction is focused on responding to an unexpected problem in the observation of the planned events, as Figure 3.3 shows. The role of the functions presented in the diagram is described as follows:

- askForReplanningLTMP(LTMP,observations): used by MOC for requiring to the LT-MPT the replanning of the failed observations planned in the LTMP.
- replanLTMP(observations): used by the LT-MPT for replanning the required observations according to the remaining time of the LTMP. Downlinks and station keepings are still positioned in the LTMP and cannot be modified.
- deliverReplannedPlan(LTMP): the LT-MPT uses this function for delivering the replanned LTMP to SOC.
- approvePlan(LTMP): used by the SOC for analysing the correctness of the LTMP and modifying some issues if necessary.

- deliverApprovedPlan(LTMP): SOC uses this action for sending the approved LTMP to MOC.
- processApprovedPlan(LTMP): MOC processes the LTMP.

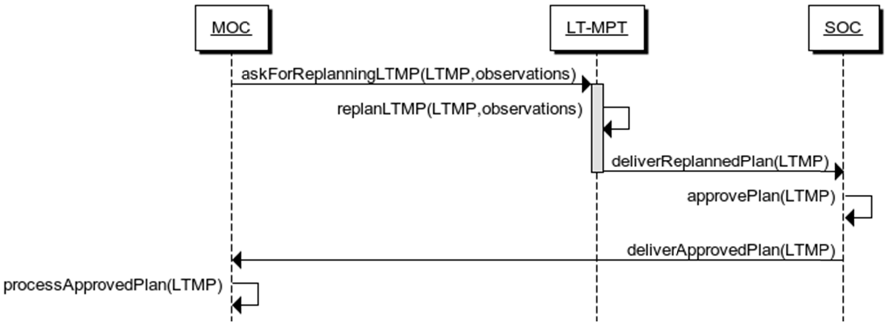

Figure 3.3. Sequence diagram of replanning and sending an LTMP to MOC

4 ECHO MISSION OPTIMIZATION PROBLEM

A complete and comprehensive list of all high-level mission requirements necessary to achieve the science goals detailed in [RD1] is provided in [RD2]. The aim of this section is to present the main aspects 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 (including science observations, downlinks, station keepings, and calibrations) that are described as follows (see [RD3] for more details):

- Science observations are the observations of target objects. A target event is defined as a time period when the exoplanet is transiting its host star. Each target event has a specific duration as defined by the Science Team.
- Downlink communication is established by ESA and it 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 directional antenna, requiring special communication attitudes during downlink. As a minimum requirement for spacecraft monitoring and control a biweekly link of 2 hour pass duration has been assumed, driven by MOC operational needs. Within the 2 hour pass, 1.75 hours will be available for science and housekeeping data downlink. Thus, preliminarily, downlink communication is initially planned every 3.5 days with some flexibility and has a duration of 2 hours (TBC).
- Station keeping operations are determined by ESA and they are defined to keep the spacecraft in the assigned orbit. They are initially defined to be carried out every 28 days with a duration of 8 hours (TBC).
- Calibration tasks are associated to science observations, they are defined in the same execution pattern and they are established by the EChO consortium. Specifically, several calibration items (e.g., instrument noise, instrument absolute wavelength or instrument pointing) will be monitored at intervals defined by the IOSDC (TBD). They are not considered in the current stage of the LT-MPT.

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 and the observation of any target 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. The next sections define the hard and soft constraints identified in the EChO mission.

4.2.1 Hard Constraints

Five hard constraints are identified in the EChO mission. Each constraint is explained in the next points.

4.2.1.1 Orbital Constraint

The satellite orbit constraints the visibility of targets (see Figure 4.1). Thus, it must be considered by the LT-MPT when computing the suitability for the observation of a target.

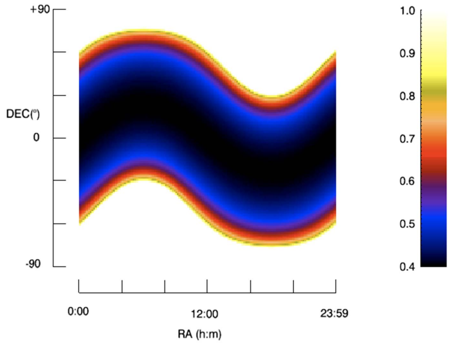

Figure 4.1. Overall sky visibility for a roll angle of 36°.

4.2.1.2 Transit Constraint

The strategy to execute the science observations depends on exoplanet transit or occultation events and their duration. This is included in the planning tool as the transit constraint. The exact occurrence of an ephemeris (T_c) can be calculated in advance. The duration of an event (T_{obs}) results from T14, which is the time between first and fourth eclipse contacts [RD3], and an additional time (T_b) devoted to determine the flux variation baseline. T_b is split in two time intervals: T_{before}, time allocated before first eclipse contact, and T_{after}, time interval after the fourth eclipse contact (see Equation 4.1).

We consider a time window of an exoplanet the duration of an event that is visible from the telescope. Figure 4.2 shows in blue colour a typical target event.

$$T_{obs} = T14 + T_b$$

$$T_b = T_{before} + T_{after}$$

$$T_{before} = \frac{T14}{2}$$

$$T_{after} = \frac{T14}{2}$$

Equation 4.1

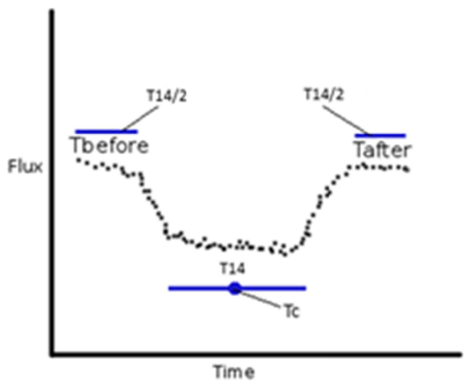

Figure 4.2. Transit light curve of an exoplanet with the total observation time of its event. Blue colour corresponds to T_{obs} and red colour to T_b. Note that T_c indicates the central time of the event.

4.2.1.3 Target Completeness Constraint

This constraint is related to the science observations. In terms of scientific interest, only the observations of complete targets are useful. A target is complete when it is observed between a minimum and maximum number of times. In the present, these two values are, respectively, 80% and 100% (both TBC) of the required number of event observations of each target. Note that these values are under consideration by the scientific team, so they can be modified in the future. Moreover, different limits could be defined according to each target class.

4.2.1.4 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.

4.2.1.5 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 Figure 4.3), including the time to reconfigure the system (i.e., slew time).

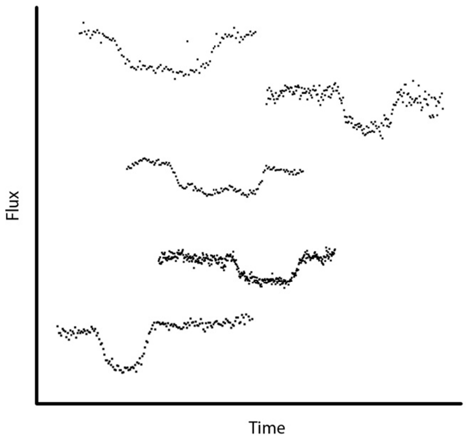

Figure 4.3. Transit light curves of different exoplanets. The LT-MPT selects the best target at any time to avoid overlapping.

4.2.1.6 Geometric Constraint

Eventually the mission could define avoidance zones for bright sources and potential sources of (out of field) straylight, for example planets. There may also be additional constraints placed on the need to avoid placing nearby stars in the slit. Such constraints will be defined when the final target list and calibration strategy are established.

4.2.2 Soft Constraints

In addition to the hard constraints identified, two soft constraints related to promote the scientific return can be defined in the EChO mission. Each constraint is explained in the next points.

4.2.2.1 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) from the targets of the same class, the less demanding targets in terms of number of events and duration. Equation 4.2 defines the criticality of a class, where *C* is the target class and *t* is a target of class *C*. The higher the value of *ClassCriticality*, the more critical the class is. Equation 4.3 defines the criticality of a target inside a class, where $Events_t$ is the number of events of target *t*, $T14_t$ is the duration of the target eclipse event in time units, and $Visibility_t$ is the number of time units where the target can be potentially observed. The higher the value of *TargetCriticality*, the more critical the target is.

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 a raising 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. Now, this scientific priority is not considered but, if necessary, the priority level can be modified (e.g., adding new priority stages) without affecting to the process of the LT-MPT.

$$ClassCriticality(C) = \sum_{t \in C} TargetCriticality(t)$$

Equation 4.2

$$TargetCriticality(t) = \frac{Events_t \cdot T14_t}{Visibility_t}$$

Equation 4.3

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.

4.2.2.2 Maximization of the Number of Targets Completed

Due to the target completeness constraint, the resulting LTMP only considers complete targets. However, the final LTMP should promote the planning of complete targets in order to obtain mission plans with higher scientific return. Thus, it will be preferred to obtain a LTMP with a high number of complete targets.

4.3 Objectives to Optimize

The objectives to be optimized by the LT-MPT are key factors for obtaining a suitable LTMP because they are focused on satisfying the soft constraints. In the EChO mission two kinds of objectives can be identified:

- Objectives focused on optimizing resources: Measures related to the time spent by an LTMP on doing some actions, such as observing events (planning efficiency) or slewing to target positions.
- Objectives focused on optimizing the scientific return: Measures to evaluate an observation scheduling from a scientific point of view, such as the priority of the planned targets, the number of events observed for each target, or the successful targets observed for each class.

The objectives used can be modified according to the characteristics of the optimization method used, in order to provide more flexibility to the LT-MPT. The definition of the objectives is included in Section 5, where each optimization approach is explained.

4.4 General Requirements

At this point we have introduced all the requirements that are the basis of the LT-MPT design in order to obtain a feasible and efficient mission plan. Table 4.1 provides a global view of these requirements.

LT-MPT Requirements
The LT-MPT has to optimize several objectives according to the optimization of resources and the scientific return
An input sample of exoplanet is used together with their observation constraints, to compute the LTMP
Each exoplanet has a specific priority according to several criteria

Each exoplanet must be observed a different number of event observations
Exoplanet visibility is known
The ephemeris and the T14 transiting time of an event are known a priory. The observation of an event has a duration of 2·T14 that is centred at the ephemeris time
Default strategy considered for downlinks: duration of 2 hours every 3.5 days
Default strategy considered for station keepings: duration of 8 hours every 28 days
Calibrations associated to science observations are not considered at the moment
Slew time has to be considered for changing the spacecraft pointing direction after each operation task
Operation tasks cannot overlap
Slewing cannot overlap operation tasks

Table 4.1. Summary of the requirements that the LT-MPT has to take into account for obtaining the final LTMP.

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, Artificial Intelligence techniques can allow us to find a solution near to the optimal one in a reasonable time. In this section we present an LT-MPT approach based on this kind of techniques for finding solutions to optimization problems that highly optimize the objectives defined. We propose a standard process to be followed for building an automatic LT-MPT for the EChO mission independently of the AI technique used.

5.1 LONG TERM MISSION PLANNING TOOL PROCESS

We propose the process showed in Algorithm 5.1 to implement the LT-MPT. This is designed to minimize conflicts between high priority targets, downlinks and station keepings, 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 downlinks and station keepings focused on restricting less priority targets, and (2) the optimization of the observation scheduling of each target event avoiding overlapping and optimizing specific objectives. Taking into account these considerations, we propose the LT-MPT process shown in Algorithm 5.1.

Step 0. Calculate the time windows of each target event.
Step 1. Clean-up impossible targets.
Step 2. Insert downlinks and station keepings minimizing potential conflicts with priority targets.
Step 3. Obtain observation planning avoiding overlaps and optimizing some specific objectives.
Step 4. Drop observations of incomplete targets
Step 5. Fill gaps with new observations or other operation tasks

Algorithm 5.1. Process followed in obtaining the LTMP.

The process specified in Algorithm 5.1 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 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 $2 \cdot T14_t$.
- **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 mission, cannot be observed at least *m*% of the requested number of events because of visibility limitations.
- **Step 2, downlink and station keeping optimization:** This part of the process is based on planning downlinks and station keepings. It is an optimization problem focused on minimizing potential overlapping with high priority targets. Other strategies to place downlinks and station keepings might be explored, like placing them after computing the LTMP.
- **Step 3, observation planning optimization:** This step places the observations of target events avoiding overlapping. This is another optimization problem that aims at planning the observations without conflicts and optimizing some specific objectives such as the planning efficiency or the scientific return.

- **Step 4, drop observations of incomplete targets:** This step responds to the target completeness constraint. Each one 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 to be completed, 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 5, fill gaps with new observations or other operation tasks:** This phase is devoted to fill the gaps of time that remain between the observations in the LTMP obtained at Step 4. It tries to plan as many new target events as possible not exceeding an upper limit of *M*% of the requested number of events. Moreover, operation tasks or other external observation can be planned in the gaps. It must be emphasized that if Step 2 is not applied, downlinks and station keeping could be planned in this step. The resulting observation plan resulting from this step is the final LTMP.

The optimization phases (Step 2 and Step 3) can be done by using several mathematical tools such as simple heuristics or AI techniques. We propose to apply Genetic Algorithms, a well-known AI technique of the Evolutionary Algorithms family, which are able to solve optimization problems and have the ability to be adapted to new environments and constraints. Other approaches based on the Evolutionary Algorithms family, presented in Annex C, have been analysed but proved to be less competitive.

5.2 LONG TERM MISSION PLANNING TOOL WITH GENETIC ALGORITHMS

This section presents the application of Genetic Algorithms in the Step 2 and Step 3 of the LT-MPT process described in Algorithm 5.1. The specific GA process used in our approach is shown in Figure 5.1. The optimization criterion corresponds to a specific number of generations. The description of each phase and the main concepts of the Genetic Algorithm theory are given in Annex A.

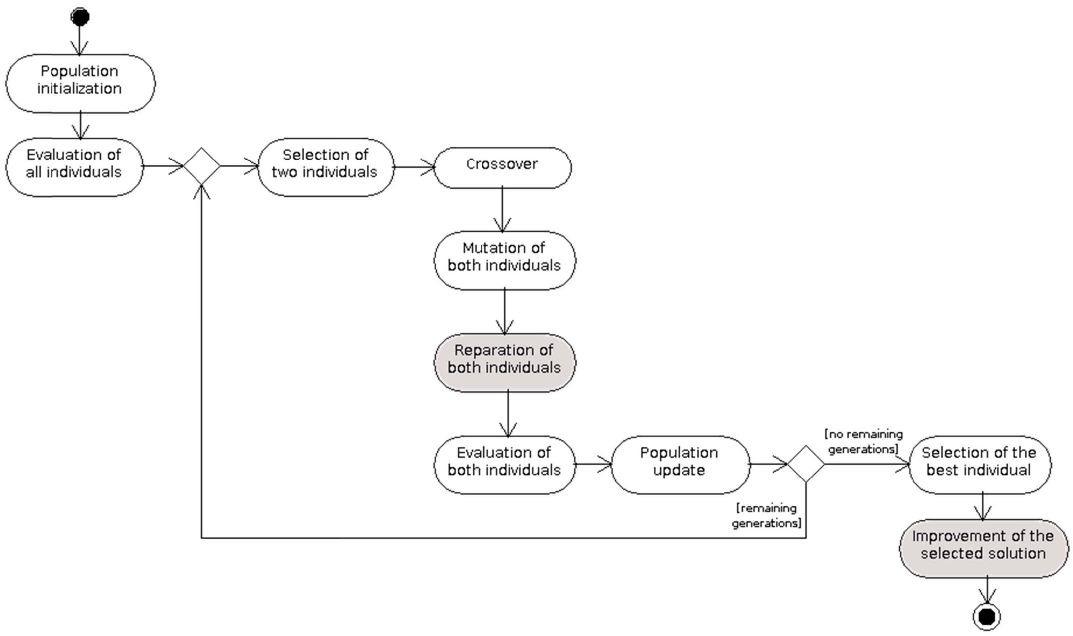

Figure 5.1. GA process proposed. Activities in grey colour correspond to steps that are only used in the Observation Planning Optimization.

5.2.1 Downlink and Station Keeping Optimization

The aim of the Downlink and Station Keeping Optimization (DSKO), which is the Step 2 in Algorithm 5.1, is to place these operations when they collide with less priority targets. We propose to optimize the planning of the downlinks, and maintaining a default positioning defined by the MOC for the station keepings (see Section 4.1). Thus, this process is only aimed to identify the optimal time slots for the downlinks. In particular, each downlink can have a specific date to be placed (default position defined by the MOC, see Section 4.1), but some flexibility is accepted. The procedure to add some flexibility in the allocation of time slots to downlinks is devoted to find an optimal shifting to the default position for each downlink. This will avoid any overlapping with high priority targets.

5.2.1.1 Individual Representation

Each potential solution in the genetic process is called individual and its representation is based on the definition of genotype presented in Annex A.1. The individual genotype is made up of real numbers which represent the shifting of the downlink. More specifically, each individual consists of $\overline{\overline{D}}$ genes { $s_1, \ldots, s_{\overline{\overline{D}}}$ }, where $\overline{\overline{D}}$ is the cardinality of the set of downlinks to be planned, and s_i the shifting assigned to downlink i. Moreover, d_i value has to be between the range [$-\delta, \delta$], where δ is the flexibility limit of a downlink (i.e., the maximum time that a downlink can be moved from its original position). The internal codification of the individual (genotypic representation) is transformed into the final positions of the downlinks (phenotypic representation) by adding the value s_i to each predefined downlink date d_i (i.e., the final date of a downlink i is $d_i + s_i$).

The initial population, which is the first step of a GA cycle, consists of building N_I new individuals assigning to each allele a random value between the range [$-\delta, \delta$].

5.2.1.2 Genetic Operators

The application of selection, crossover, mutation and replacement operators in the DSKO is described as follows:

- Selection: 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 (Freitas, 2002). The goal of this operator is to select two parents to be subsequently crossed. The idea is based on a randomly chosen set of S individuals from the population and selecting the best individual from this group as first parent. This process is repeated in order to obtain a second parent. The parameter S has a value between the range [1, $\overline{\overline{P}}$], where $\overline{\overline{P}}$ is the number of individuals in the population (P).

- Crossover: The aim of this operator is to mix the genetic information of the individuals. In this case, two new individuals are obtained from two previously selected parents. In particular, the new individuals are built by combining the genotypes of both parents. Our approach contemplates two crossover operators but only one of them is applied during all the GA cycle: (1) one-point crossover, and (2) uniform crossover. The first one is based on randomly selecting a cut-point and recombining the first part of first parent with the second part of the second parent to create one offspring and the second part of first parent with the first part of the second parent to obtain the other offspring. On the other hand, uniform crossover 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. Uniform crossover is at least as good as one-point crossover when elitism is considered in our population (Chen and Smith, 1999), which is the case of our algorithm. It is worth noting that 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 s_i' is obtained by adding a random value μ between the range [$-\delta, \delta$] to the allele of the gene s_i (i.e., $s_i' = s_i + \mu$). When s_i' is out of the range [$-\delta, \delta$] has to be modified to be into this range in order to obtain a correct mutated gene s_i''. So, if $s_i' < -\delta$, then $s_i'' = s_i' + \delta$, and if $s_i' > \delta$, then $s_i'' = s_i' - \delta$.

- Replacement: Our approach has been designed to use elitism in the population. This means that previous individuals of the population are not removed, so the offspring and their parents can coexist in the population. Specifically when the size of the population is smaller than N_P, which is the maximum number of individuals that the population can store, the new individual is added to the population. In the case that the population P is full (there are N_P individuals in P, $\bar{\bar{P}} = N_P$), the worst individual of the population is replaced by the new individual.

5.2.1.3 Objective Functions

The objective for the optimization is to avoid conflicts between downlinks and the high priority targets. The target priority is calculated according to the target criticality in Equation 4.3. Thus, in this case the priority of the targets is calculated in terms of duration, number of events and visibility, so the criticality of the classes is not considered. The fitness function F of an individual I is defined in Equation 5.1, where d is a downlink that belongs to the planned downlinks D_I for the individual I, t is a target from the initial T existing targets, w is a window from all the possible windows of target t, *TargetCriticality* is defined in Equation 4.3, and *Overlapping* indicates if two ranges of time are overlapped between the ranges of time defined by window w of a target and a window d of a downlink. *Overlapping* is defined in Equation 5.2, where s_{r_1} is a time unit of the range of time r_1 and s_{r_2} is a time unit of the range of time r_2. The more critical the target, the higher is the *TargetCriticality* function value and, thus, a small value of F is desired for avoiding a high number of critical conflicts (i.e., F has to be minimized in order to be optimized). Note that this objective function can be modified to obtain a more accurate downlink positioning (e.g., taking into account the class criticality).

$$F(I) = \sum_{d \in D_I} \sum_{t \in T} \sum_{w \in t} if\ Overlapping(d, w), then\ TargetCriticality(t)$$

Equation 5.1

$$Overlapping(r_1, r_2) = \exists s_{r_1} \in r_1\ \exists s_{r_2} \in r_2 : s_{r_1} = s_{r_2}$$

Equation 5.2

5.2.2 Observation Planning Optimization

The aim of the Observation Planning Optimization (OPO), which is the Step 3 in Algorithm 5.1, 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 maximize the planning efficiency and the scientific return that consists in maximizing the number of targets completed and planning targets with higher priority (see Section 4.2.2.1). 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.

5.2.2.1 Individual Representation

The individual genotype for the OPO procedure is made up of integer numbers which represent the time windows where the targets are planned. Each individual consists of $\sum_{t \in T} \bar{\bar{E}}_t$ genes $\{o_{1,1}, \ldots, o_{\bar{\bar{E}}_1,1}, \ldots, o_{1,\bar{\bar{T}}}, \ldots, o_{\bar{\bar{E}}_{\bar{\bar{T}}},\bar{\bar{T}}}\}$, where $\bar{\bar{E}}_t$ is the number of events of target t, $\bar{\bar{T}}$ is the cardinality of the set of existing targets T, and $o_{i,t}$ corresponds to observation i of target t. Moreover, the $o_{i,t}$ value has to be between the range [-1, $\bar{\bar{W}}_t - 1$], where -1 indicates that the corresponding event does not have a window assigned (i.e., it is not planned), and $\bar{\bar{W}}_t$ is the number of potential time windows of target t (the limit is $\bar{\bar{W}}_t - 1$ because the value 0 corresponds to the first window). Note that 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 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 Figure 5.2, 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 green colour). Note that the third target is not planned.

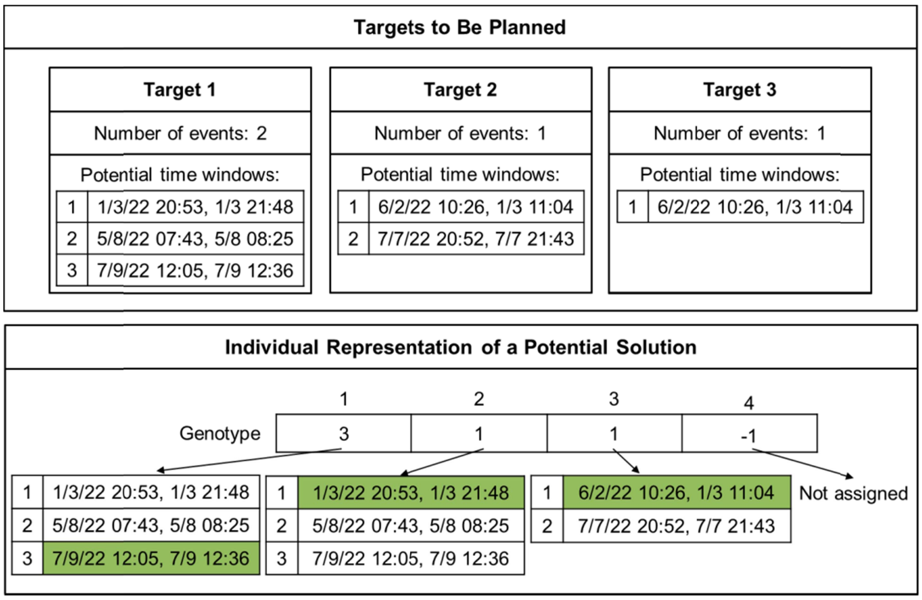

Figure 5.2. The example provides some targets to be planned and a potential LTMP using the proposed individual representation.

It must be emphasized that 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 5.2.2.1.

The initial population is built by creating N_I new individuals assigning to each allele a value between the range [-1, $\overline{\overline{W}}_t$ - 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. For more details, see Algorithm 5.2.

5.2.2.2 Genetic Operators

The genetic operators of selection, crossover and replacement are the same ones defined in Section 5.2.1.2 for the DSKO. However, we do not recommend the use of one-point crossover due to the fact that in our approach the exploration of the search space will be less exhaustive. This is because the phenotype of the offspring will not be considerably changed, with respect to the parents, with a single cut-off point. For this reason, we consider that uniform crossover will mix properly the genetic information of the parents for building the offspring. On the other hand, the mutation operator is modified and adapted to the individual representation, as explained below:

- Mutation: It is applied to each gene of every new individual with a probability of $P\mu$, which means that some genes are not mutated. In our approach, a gene is mutated by changing its allele with one of the potential time windows of its corresponding target. Thus, a mutated gene $o_{i,t}'$ changes its allele with a random value μ between the range [0, $\overline{\overline{W}}_t$ - *1*] (i.e., $o_{i,t}' = \mu$), where $\overline{\overline{W}}_t$ is the number of potential time windows of target *t*.

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.

```
Let Targs be the collection of targets
Let Ind be the vector that represents the individual to build
Let Plan be an empty collection of observations, where each observation will be a range of time

Add to Plan the ranges of time of each downlink and station keeping
foreach target t in Targs selected in random order do
  foreach gene g from Ind related to t do
    w_r ← random window of the potential time windows of target t
    added, Plan ← AddObservation(Plan, w_r, t)   //See Algorithm 5.5
    if added is true then
        Ind[g] ← the identifier of w_r in the list of potential windows of target t
    else
        Ind[g] ← -1    //Observation related to gene g is not planned
    endif
  endfor
endfor
return Ind
```

Algorithm 5.2. Process followed to build a new individual.

5.2.2.1 Reparation 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, downlinks or station keepings). There are two ways for obtaining an unfeasible individual that requires reparation 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 (see Figure 5.2) in order to obtain feasible ones. So, the main idea of the repair operator is to change the window assigned to the conflicting observations in order to avoid overlapping by using a Hill Climbing strategy (Brownlee, 2011). Algorithm 5.3 and Algorithm 5.4 show how conflicting observations are identified and repaired, obtaining a feasible individual. Moreover, Algorithm 5.5 shows how the plan represented by the individual is build considering the slew time between consecutive observations.

```
Let Targs be the collection of targets
Let Ind be the vector that represents the individual to repair
Let Plan be an empty collection of observations, where each observation will be a range of time

Add to Plan the ranges of time of each downlink and station keeping
foreach target t in Targs selected in random order do
  foreach gene g from Ind related to t do
    w ← the window from the list of potential windows of target t with the identifier Ind[g]
                                   //if Ind[g] is -1, the returned window is an empty time range
    conflict, Plan ← AddObservation(Plan, w, t)    //See Algorithm 5.5
    if conflict is true then
        Ind[g] ← Repair(g, t, ind, Plan) //See Algorithm 5.4
    endif
  endfor
endfor
return Ind
```

Algorithm 5.3. Process followed for identifying conflicting observations and repairing the individual.

```
Let c be the conflicting gene and w_c its assigned window
Let t be the target that corresponds to c
Let Plan be the collection of assigned observations, where each observation is a range of time

foreach window w_p in the list of potential windows of target t and different to w_c do
   added, Plan ← AddObservation(Plan, w_p, t) //See Algorithm 5.5
   if added is true then
      return the identifier of w_p in the list of potential windows of target t   //and exit from function
   endif
endfor
return -1   //the observation will not planned
```

Algorithm 5.4. Process followed to identify a non-conflicting time window.

```
Let Plan be the collection of assigned observations, where each observation is a range of time
Let w be the range time of the observation to be added to Plan
Let t be the target related to the observation to be added

o_p ← previous observation in Plan according to the starting date of w
o_n ← next observation in Plan according to the starting date of w
o_w ← build the time range of the observation with the initial time and final time of w
Add to the time range of o_p the slew time between the targets observed by o_p and w
Add to the time range of o_w the slew time between the targets observed by w and o_n
if ¬Overlapping(o_p, o_w) ∧ ¬Overlapping(o_w, o_n) then   //See Equation 5.2
   Add o_w to Plan
   Return true, Plan   //and exit from function
endif
return false, Plan
```

Algorithm 5.5. Process followed to add an observation without overlapping when adding the slew time.

5.2.2.2 Objective Functions

The objectives to be optimized in this approach are the planning efficiency and the scientific return. The planning efficiency corresponds to the observation time of the plan without considering the slew time and the rest of the operation tasks. The scientific return is based on computing the number of targets completed (observed at least m% of its required number of events) weighted according to their priority. The fitness function F of an individual I is defined in Equation 5.3 and it considers the measures F_G and F_T. The first one computes the time that the telescope is not observing and the second one computes the number of incomplete targets weighted with their priority (having more weight in the computation the incomplete targets that are more priority). F_G is defined in Equation 5.4, 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 Equation 5.5, where $PriorityLevel_t$ indicates the priority level of target t as it is defined in Section 4.2.2.1, and $\overline{\overline{T}}$ 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. It is important to highlight that when an individual is evaluated with the fitness function, it represents a plan without overlapping. However, this plan considers incomplete targets, so the target completeness constraint is not achieved. For this reason, the next section presents how to improve the final LTMP obtained with the GA by removing the incomplete targets and filling the gaps between observations.

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

Equation 5.3

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

Equation 5.4

$$F_T(I) = 1 - \frac{\sum_{t \in T} if\ t\ is\ incomplete\ in\ O_I,\ then\ \overline{\overline{T}} - PriorityLevel_t}{\sum_{t \in T} PriorityLevel_t}$$

Equation 5.5

5.2.2.3 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. These processes correspond to Step 4 and Step 5 of Algorithm 5.1.

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 this with a parameter called M, which limits the times that a target can be planned and it is calculated as a percentage according to their required number of events.

Algorithm 5.6 specifies the process followed to improve the LTMP. This process is divided in two parts. The first one tries to complete each incomplete target and removes it if the target is still incomplete. The second part fills the gaps by adding observations of targets already included in the LTMP, starting with events of those targets with the highest priority.

```
Let Targs be the collection of targets sorted from high to low priority (see Section 4.2.2.1)
Let Plan be the collection of the planned observations, where each observation will be a range of time

foreach target t in Targs selected in priority order do   //Target completeness constraint is achieved
  foreach window w in the list of potential windows of target t do   //Incomplete targets are tried to be completed
    if target t is not observed at least an m% of their required observations then
      Plan ← AddObservation(Plan, w, t)   //See Algorithm 5.5
    endif
  endfor
  if target t is not observed at least an m% of their required observations then //Incomplete targets are removed
      remove from Plan all the observations related to t
  endif
endfor

foreach target t in Targs selected in priority order and that are still planned in Plan do   //Gaps are filled
  foreach window w in the list of potential windows of target t do
    if target t is not observed more than M% of their required observations then
      Plan ← AddObservation(Plan, w, t)   //See Algorithm 5.5
    endif
  endfor
endfor
```

Algorithm 5.6. Process followed to obtain a feasible LTMP and to improve it.

5.2.3 Relevant Parameters

Several parameters have been defined in order to properly configure the GA used in the DSKO and OPO optimization processes. Moreover, both optimization algorithms need some additional parameters for their proper configuration. In order to have an overview of all the necessary parameters, they have been summarized in Table 5.1.

Algorithm	Parameter	Description	Range
GA	Generations	Number of times that the genetic cycle is repeated	[1, ∞)
	N_I	Number of individuals in the initial population	[2, ∞)
	N_P	Number of individuals in the final population	[2, ∞)
	S	Number of the individuals to be chosen from the population for selecting one parent	[1, $\bar{\bar{P}}$], where $\bar{\bar{P}}$ is the number of individuals in the population
	p_c	Crossover probability	[0, 1]
	p_μ	Mutation probability	[0, 1]
DSKO	δ	Maximum variation of the downlink position from the default one	[0, ∞)
OPO	m	Minimum number of times that a target can be observed. It is calculated as a percentage according to its required number of events	[0, 100]
	M	Maximum number of times that a target can be observed. It is calculated as a percentage according to its required number of events	[m, ∞)

Table 5.1. Summary of the parameters defined in DSKO and in OPO, and the ones defined in the GA used by both approaches.

6 EXPERIMENTS, RESULTS AND DISCUSSION OF THE LONG TERM MISSION PLANNING TOOL

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 selected approach to be considered in the final design of the LT-MPT and, at the same time, to demonstrate the feasibility of the EChO mission. The experiments have been carried out in several mock scenarios and in a real sample of known exoplanets. Moreover, a comparison of other algorithm approaches and configurations is provided in Annex B and Annex C.

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 Bench Configuration

The next points describe the test samples used and their constraints, as well as the general configuration of the test bench.

6.1.1.1 Test Samples and Constraints

Over 300 transiting planets have been discovered to date and, among those, a significant fraction of hot Jupiters with bright host stars. It is expected that future experiments will discover new targets of EChO interest, improving on those presently available. In the context of EChO, close to 75% of the exoplanets to be considered in the mission have been discovered over the last five years. For this reason, taking into account that EChO mission will be launched in the year 2022, we can predict that the available list of exoplanets to be observed will be modified. 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) 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. For more details about the definition of the MRS, the reader is referred to (TBD).

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 described with the parameters specified in Table 6.1, note that some parameters are not taken into account by the LT-MPT. MRS realizations are calculated by randomizing several parameters according to astrophysically-sound assumptions. These realizations are generically called MRS_rand and they are sized considering a usable science time of 31671 hours. At least, one target is generated for each of the generic classes indicated in the MRS. The value of the parameters for each target are obtained as follows:

- Stellar mass (*Ms*), stellar radius (*Rs*), stellar effective temperature (*Teff*), planet mass (*Mpl*) and planet radius (*Rpl*) are adopted from the MRS fiducial values.
- Number of targets following a homogeneous distribution in *K* magnitude from *K*=3 to *Klim*+0.5 (number of stars scaling as $10^{3K/5}$) (*Ks*).
- Scaled event duration and planet equilibrium temperature from period (T14, *Tpl*).
- Random origin of ephemeris (*T0*).
- Defined range of periods for each planet class (with overlap) and generated random values with a distribution flat in *log P* (*P*).
- Scaled distance from *K* magnitude (*Dist*).
- Uniform sky distribution in right ascension (*RA*) and declination (*Dec*).

- Scaled number of events from scaled period and *K* magnitude (*#ev*).
- *Typ* corresponds to the planet class to which the target belongs.

The target lists produced define hypothetical exoplanet scenarios to be observed by the EChO mission in the year 2022. These scenarios will be used in the experimentation in order to test the robustness of the proposed LT-MPT. It must be emphasized that 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 has been included in the experimentation with the aim of testing the proposed tool in the current time. Table 6.2 shows a summary of the main characteristics of the scenarios. For each scenario we indicate the number of targets of the sample, the number of targets that can be achieved (i.e., the *m%* of their required number of events can be potentially observed during the mission), the total number of required events considering all the achievable targets, and the total input time in hours (i.e., the required time for observing all the required events).

Related to	Parameter	Description	Unit
Star	Ms	Star mass	Solar masses
	Ks	Star brightness in the K band	Magnitudes
	Rs	Radius of the star	Solar radius
	Teff	Effective temperature of the star	Kelvins
Planet	Mpl	Mass of the exoplanet	Earth masses
	Rpl	Radius of the exoplanet	Earth radius
	Tpl	Equilibrium temperature of the exoplanet	Kelvins
	Typ	Class type of the exoplanet	String (e.g., HSE M4, HN E7...)
Orbit	T0	Exoplanet ephemeris	Julian days
	P	Period of planet event	Days
	T14	Duration of exoplanet event	Seconds
Location	Dist	Distance to the system	Parsecs
	RA	Right ascension of the exoplanet	Hours
	Dec	Declination of the exoplanet	Degrees
Observation	#ev	Number of events to be observed	Number according to 5-year mission

Table 6.1. Parameters that describe each one of the targets defined in an MRS_rand. The parameters that the LT-MPT takes into account are shown in grey colour, the other ones will be used for different scientific purposes.

Name	Number of Targets	Number of Achievable Targets	Total Number of Required Events	Total Input Time
MRS_rand_0	238	231	6338	27925.61 hours
MRS_rand_1	238	233	5854	25778.35 hours
MRS_rand_2	238	230	6161	26734.89 hours
MRS_rand_3	238	233	6669	29378.63 hours
MRS_rand_4	238	233	6095	26321.95 hours
MRS_rand_5	238	235	6271	27557.05 hours
MRS_rand_6	238	234	6013	26902.01 hours
MRS_rand_7	238	232	5866	25954.43 hours
MRS_rand_8	238	232	5966	24839.95 hours
MRS_rand_9	238	232	6276	26337.40 hours
Real_sample	122	121	2820	17392.05 hours

Table 6.2. Characteristics of each scenario used in the experiments.

6.1.1.2 General Test Bench Configuration

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

- Tests will cover the total mission lifetime (five years, 2022 - 2026).
- 520 downlinks with a duration of 2 hours and a periodicity of 3.5 days are considered.
- 65 station keepings with a duration of 8 hours and a periodicity of 28 days are assumed.
- No calibrations are considered.
- Slew time between observations of different targets was taken into account using a slew speed of 45 degrees per 10 minutes plus a flat 5-minute overhead.

6.1.2 Algorithm Configuration

Several configurations of the LT-MPT, according to the process defined in Algorithm 5.1, have been proposed and analysed during the development of this work. The experimentation results presented in this section are referred to the configuration that we consider the best one according to its performance and to the final operational design of the EChO mission described in Section 3. This configuration consists of using the OPO algorithm defined in Section 5.2.2 for obtaining the observation planning (Step 3 of Algorithm 5.1) without considering downlinks and station keepings (i.e., the DSKO algorithm is not applied and the Step 2 of Algorithm 5.1 is omitted). Note that the rest of the steps of Algorithm 5.1 are applied in the same way defined in Section 5.2.2.

The description of the alternative configurations and the analysis of their performance are presented in Annex B and Annex C, and they are referred to downlink positioning strategies and to other OPO approaches, respectively. This analysis is summarized in Figure 6.1, where the selected configuration of the LT-MPT for the EChO mission is showed in green colour. Note that the OPO approach considered the best one is indicated as OPO-GAE and, from now, we will refer to it with this name.

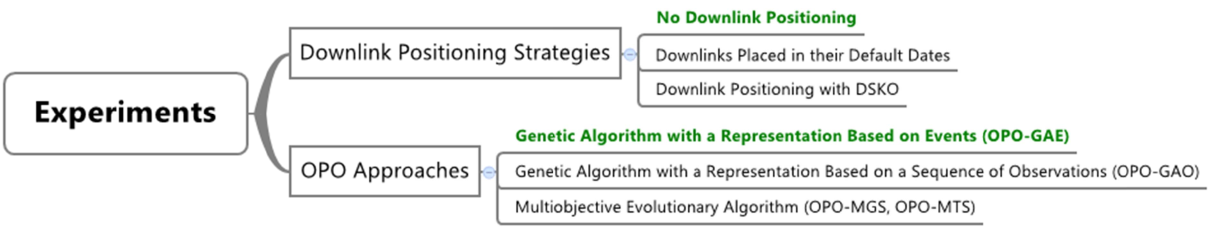

Figure 6.1. Experiments performed in the analysis of the LT-MPT. The methods selected are highlighted in green colour.

In terms of parameterization, the LT-MPT has several common parameters related to the GA used in the DSKO and the OPO algorithms, as well as some specific parameters used in each one of the processes. Table 6.3 summarizes the parameter configuration (see Table 5.1 for the parameter description) used in the experimentation done in Section 6.2, Annex B and Annex C.

Algorithm	Parameter	Value
GA	Generations	10000
	N_I	250
	N_P	1500
	S	$0.4\cdot\overline{\overline{P}}$
	p_c	0.9
	p_μ	0.005
DSKO	δ	10% of 3.5 days (8.4 hours)
OPO	m	80
	M	100

Table 6.3. Parameter configuration of the LT-MPT.

6.1.3 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 time optimization:

- **Planning Efficiency:** 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 overall time of the mission (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

As mentioned above, the selected LT-MPT configuration for the EChO mission is to use the OPO-GAE approach without placing downlinks and station keepings, because the process of placing them will be done by SOC. This process will consist of placing downlinks and station keepings in the gaps between the observations planned. It must be worth noting that the optimization algorithm OPO-GAE was run in a Monte Carlo fashion to better test the performance of the LT-MPT, using 25 different random seeds for each scenario. Each of these executions is referred hereafter as trial.

In Table 6.4, the results obtained for the sample realizations (*MRS_rand_#*) and the real case (*Real_sample*) are described in terms of the four defined metrics (planning efficiency, 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 Average*). In the next sections, these results are analysed according to each one of the defined metrics.

Table 6.4 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.

Name	Planning Efficiency		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 Average	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

Table 6.4. Results obtained (mean and deviation) with the LT-MPT based on OPO-GAE, without placing downlinks and station keepings, in the defined scenarios after 25 trials.

6.2.1 Planning Efficiency

In terms of the Planning Efficiency 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 total input time is smaller than the overall time of the mission, so several gaps between observations can be found along the mission. Thus, an interesting aspect to analyse is the number of gaps between observations and their duration. Figure 6.2 shows the number of gaps for several gap slots sizes in minutes for the 5 years of the mission. 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 related to the visibility of each target event. This is supported by the fact that the algorithm obtains similar gap distribution for each scenario. Figure 6.3 shows the overall duration of the gaps of each gap slot size for the 5 years of the mission. It can be observed that the gap duration can slightly change in each scenario but they follow a similar pattern. In the accumulated gap distribution shown in Figure 6.4 and Figure 6.5, it can be observed that there are around 6000 gaps longer than 10 minutes, and consequently there are around 18000 hours of gaps with considerable size. Therefore, it would be interesting to analyse how to fill these gaps with some additional observations to make the most of the mission.

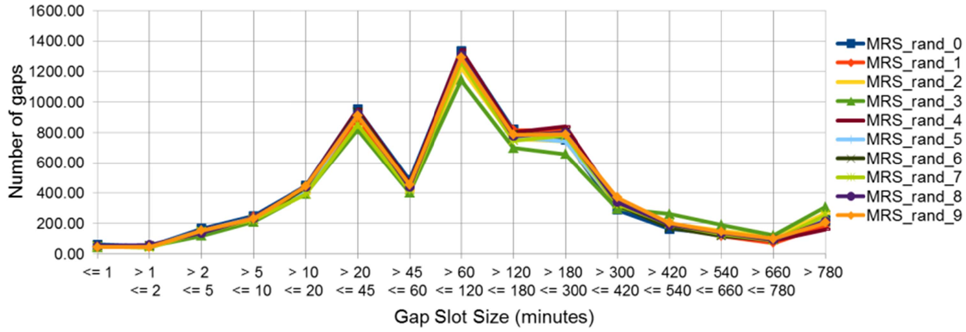

Figure 6.2 Number of gaps for each gap slot.

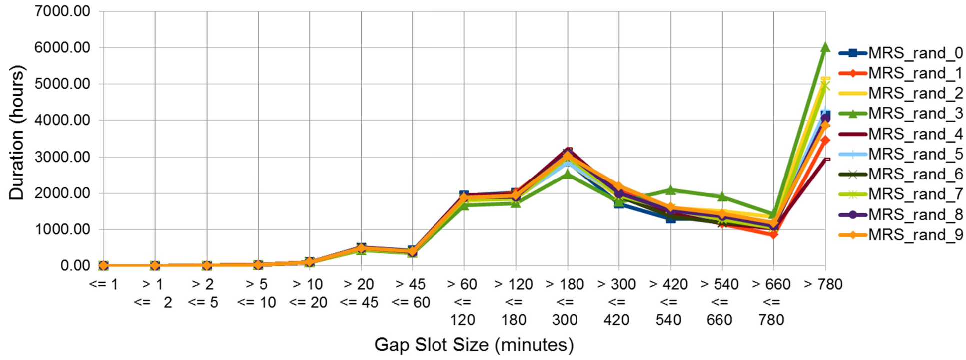

Figure 6.3. Overall duration in hours of the gaps of each gap slot.

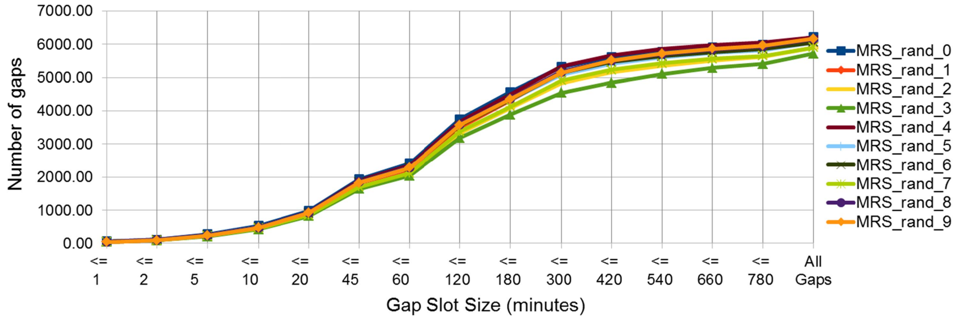

Figure 6.4. Accumulated number of gaps for each gap slot.

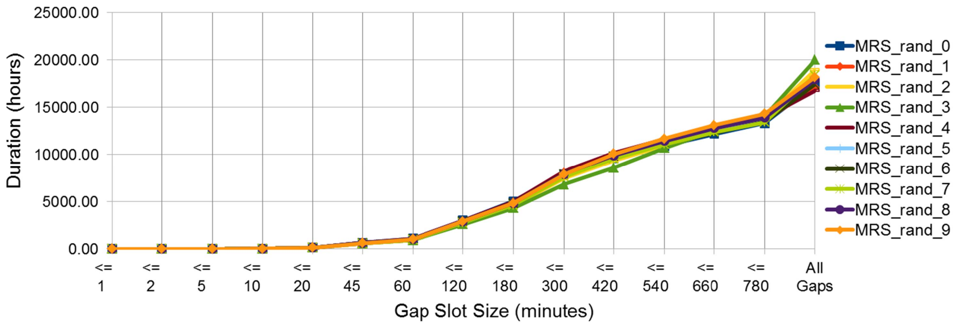

Figure 6.5. Accumulated overall duration in hours of the gaps of each gap slot.

6.2.2 Slew Time

Table 6.4 shows that the slew time duration is about 5.28% of the overall time of the mission (5 years) in the sample realization scenarios and 2.43% in the real sample.

Figure 6.6 shows, for each scenario, the number of trials with an overall slew time inside a specific time slot. It can be noticed that the trials of each scenario obtain similar results in the overall slew time duration and there is not an important deviation, and this can be confirmed in Table 6.4. Moreover, it can be observed that each scenario has a different peak slew time duration. This is because target observations are planned in different order in each scenario. After this analysis, it can be concluded that the results are similar for each scenario in terms of gap distribution due to target visibility, but each scenario needs a different optimal planning of the observations.

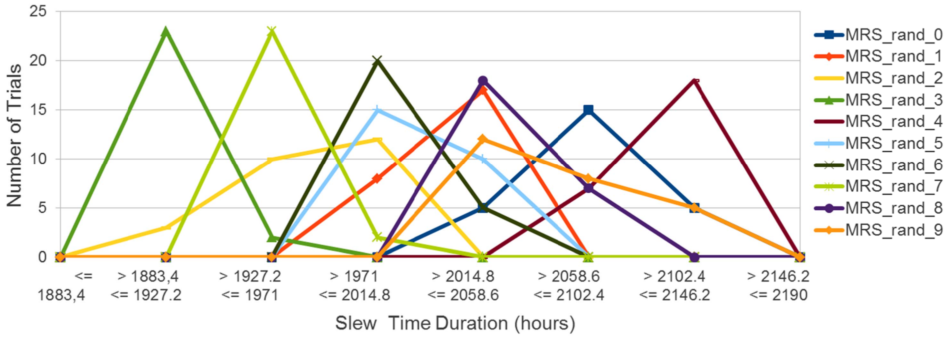

Figure 6.6. Number of trials with an overall slew time in each time slot. The number of trials considered in each scenario is 25 and the duration of the total slew time is presented in hours.

6.2.3 Events Planned and Targets Completed

Table 6.4 shows that in the sample realization scenarios around a 93% of the overall 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 (note that all these targets are observed for at least 80% of their required events).

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 observed for each class. Figure 6.7 shows the average number of targets that belong to each class (first bar, in red colour), and the average number of targets planned for each one (second bar, in green column) in the artificial scenarios with the 25 trials executed. It can be observed that all classes have a high completeness rate, even those that contain more targets. Another aspect to be considered is whether the final LTMP takes into account the priority of the targets.

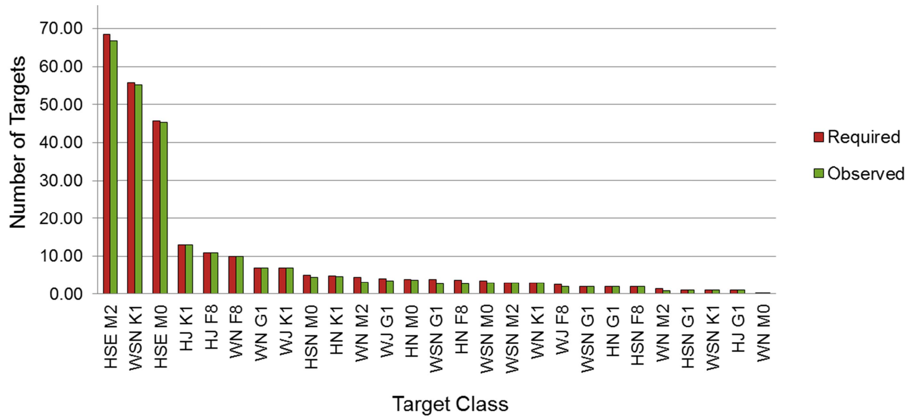

Figure 6.7. Average of the target class completeness. For each class, the first bar (in red colour) indicates the number of targets that belong to it, and the second bar (in green colour) indicates the average number of targets finally planned.

Table 6.5 details the results obtained in one scenario sorted by the target priority. For brevity, only 31 targets are shown (the higher priority and the lower priority). In particular, each target is sorted according to their priority level, which is obtained with the priority of each class and the priority of itself, and also for each target the number of events that are observed is indicated. It can be observed that the targets with high priority of each class tend to be observed more times than the ones with less priority. Consequently, it seems that the final LTMP takes properly into account the priority of the targets into each class.

Priority Level	**Class Priority**	**Target Priority**	**Target Identifier**	**Events observed**	**Priority Level**	**Class Priority**	**Target Priority**	**Target Identifier**	**Events observed**
Class HSE M2					Class HJ K1				
1	1	1	Target 0	482 from 482 (100%)	211	20	1	Target 88	11 from 11 (100%)
Class WJ F8					212	20	2	Target 101	12 from 12 (110%)
2	2	1	Target 195	17 from 18 (94%)	213	20	3	Target 211	37 from 37 (100%)
Class WJ G1					214	20	4	Target 197	55 from 55 (100%)
3	3	1	Target 191	5 from 5 (100%)	215	20	5	Target 206	14 from 14 (100%)
4	3	2	Target 192	15 from 15 (100%)	216	20	6	Target 213	54 from 54 (100%)
5	3	3	Target 194	12 from 12 (100%)	217	20	7	Target 215	70 from 70 (100%)
6	3	4	Target 190	15 from 17 (88%)	218	20	8	Target 102	11 from 11 (100%)
7	3	5	Target 193	0 from 23 (0%)	219	20	9	Target 106	14 from 14 (100%)
Class WSN G1					220	20	10	Target 201	40 from 40 (100%)
8	4	1	Target 28	21 from 21 (100%)	221	20	11	Target 208	23 from 23 (100%)
Class WN F8					222	20	12	Target 202	22 from 26 (100%)
9	5	1	Target 87	6 from 6 (100%)	223	20	13	Target 207	56 from 56 (100%)
10	5	2	Target 86	8 from 8 (100%)	224	20	14	Target 96	20 from 20 (100%)
11	5	3	Target 85	13 from 14 (92%)	225	20	15	Target 204	84 from 34 (100%)
Class WSN K1					226	20	16	Target 209	89 from 89 (100%)
12	6	1	Target 26	40 from 40 (100%)	227	20	17	Target 203	102 from 102 (100%)
13	6	2	Target 27	55 from 55 (100%)	228	20	18	Target 205	0 from 130 (0%)

Table 6.5. Extract of a target planning in one scenario (MRS_rand_8) and with one trial. The targets classes and targets within are sorted from high to low priority. The block of the left is referred to the targets with higher priority, and the block of the right is referred to the targets with lower priority. For each block of targets, the name of their class, their priority level, their class priority and the target priority inside the class, the target identifier and the number of events observed are indicated. The targets that are not planned are shown in red.

6.2.4 Downlinks as Gap Fillers

The obtained LTMP does not consider downlinks and station keepings, and it must be completed by placing the 520 downlinks and the 65 station keepings in the gaps between observations. In this experiment we want to test the feasibility of placing these operations with this strategy. For brevity, the experiment only considers the placement of downlinks, but the process is the same when station keepings are considered.

The predefined position of the downlinks is known at the beginning of the mission (see Section 4.1), but the operation constraints may allow some flexibility. Therefore, the process consists in seeking a gap in the range [$d - \beta$, $d + \beta$] to place each downlink; where, d is the default position of a downlink and β is the flexibility that is given as a rate over the cadence of the downlinks. Moreover, each gap must be of at least the duration of a downlink (see Section 4.1). Table 6.6 summarizes the downlinks rate that can be placed according to several flexibilities (from 0% to 50%). It can be observed that 30% of the overall downlinks can be placed without considering flexibility in the sample realizations and around a 50% in the real sample. However, the number of downlinks placed is considerably increased with 10% of flexibility. The number of downlinks allocated raises when the flexibility increases, being able to place almost all downlinks with a flexibility of 30% in the sample realizations and of 20% in the real sample. The downlinks not allocated in this procedure will be inserted by replacing the observations in the same time interval.

Figure 6.8 shows an example of how downlinks can fill the gaps in a single sample realization scenario with a flexibility of 10% and 50%. The horizontal axis indicates each day of the first 100 days of the mission and the blue lines indicate the gaps in the LTMP with the duration needed to place a downlink. The first row of points indicates the theoretical position of the downlinks. The second row of points indicates the final position of the downlinks, where the green squares represent the downlinks placed in the position of gaps and the red crosses represent the downlinks that have not been successfully allocated. Thus, downlinks indicated with red crosses will replace the observations planned in the same range of time.

Name	3.5days +/- 0%	3.5days +/- 10%	3.5days +/- 20%	3.5days +/- 30%	3.5days +/- 40%	3.5days +/- 50%
MRS_rand_0	26.70	80.11	94.15	97.85	99.15	99.69
MRS_rand_1	28.40	80.30	93.99	97.62	98.92	99.42
MRS_rand_2	32.56	84.54	96.24	98.46	99.30	99.73
MRS_rand_3	32.10	84.25	96.27	98.81	99.71	99.90
MRS_rand_4	26.56	80.97	94.39	98.06	99.30	99.67
MRS_rand_5	28.10	80.82	95.03	98.21	99.44	99.71
MRS_rand_6	29.76	82.25	94.11	97.79	99.02	99.53
MRS_rand_7	29.92	82.02	95.14	98.21	99.42	99.73
MRS_rand_8	28.10	81.07	93.74	97.69	98.84	99.44
MRS_rand_9	29.04	81.58	94.99	98.23	99.21	99.67
MRS_rand Average	29.12	81.79	94.80	98.09	99.23	99.65
Real_sample	46.93	94.80	99.30	99.86	99.90	99.92

Table 6.6. Percentage of downlinks placed in the LTMP gaps with respect to the total downlinks. Each column shows a different flexibility in the downlink position for the defined scenarios after 25 trials.

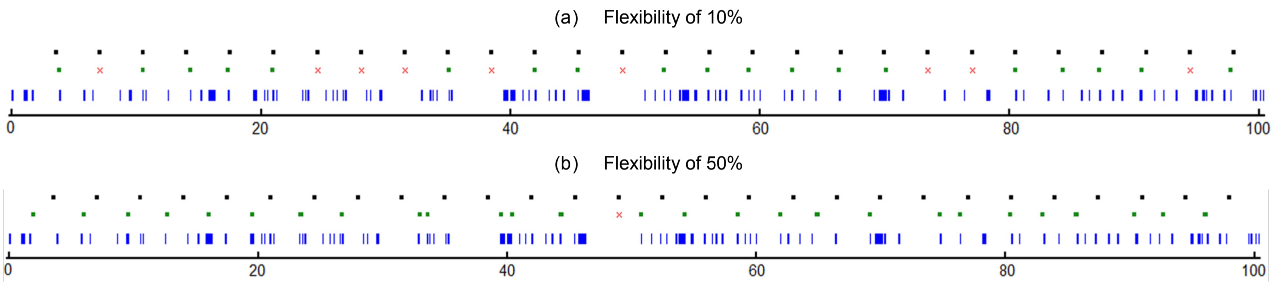

Figure 6.8. Downlink positioning in the first scenario (MRS_rand_0) with flexibilities of (a) 10% and (b) 50% in the first 100 days of the mission, which are represented in the horizontal axis. The blue lines indicate the gaps of the LTMP and the first row of points indicates the predefined position of the downlinks. The second row of points shows the final position of the downlinks, indicating with red crosses the downlinks that have not been successfully allocated.

6.2.5 Computational Cost

The algorithm 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.

7 ROSETTA STONE TARGETS

The Rosetta Stone targets are exoplanets that require observation of both transit and occultation events and whose orbital phase curve has to be fully observed. A list of present Roseta Stone targets (to be used as an example) is showed in Table 7.1. These exoplanets are added to the sample realization scenarios and the process to plan the observation of their occultation and transit events is the same than the used with the other targets. However, at the end of the process of the LT-MPT, the gaps between observations in the LTMP are used for placing the rest of the orbital phase curve of these exoplanets. Due to the fact that it is difficult to find gaps with the size of the full curves, they are divided into smaller parts and each of those parts is placed within a gap.

Name	T14	Period	Transit Events	Occultation Events
55Cnce	1.66	17.68	20	40
Gliese 1214 b	0.91	37.93	20	40
Gliese 436 b	0.70	63.45	20	40
HAT-P-11 b	2.43	117.31	20	20
HAT-P-32 b	3.56	51.60	20	20
HAT-P-41 b	4.13	64.66	20	20
HD 189733 b	1.86	53.25	20	20
HD 209458 b	3.02	84.59	20	20
HD 80606 b	4.23	104.47	20	20
WASP-13 b	3.79	89.65	20	20
WASP-17 b	3.34	88.65	20	20
WASP-54 b	6.21	87.90	20	20
WASP-79 b	2.11	73.63	20	20
WASP-80 b	1.66	17.68	20	20

Table 7.1. Sample list of exoplanets that can be considered Rosetta Stones. T14 is in hours, period is in hours, and the number of transit and occultation events is indicated.

The main considerations for dividing the orbital phase curve of each exoplanet are listed as follows:

- The observation of the segments of the orbital phase curve of each exoplanet can correspond to different periods (i.e., the observation of the curve is done by repeating visits to the system, but never to the same point in the orbit).
- In order to avoid the division of the orbital phase curve in a large number of small segments, the minimum size of a segment has been set to *k*·T14. In the experiments, several *k* values have been used.
- Neighbouring segments of the phase curve have an overlapping baseline of 0.5·T14.
- The orbital phase curve of each exoplanet is considered complete when it is observed, at least, for 90% of its total duration.
- The overall time interval over which a planet's phase curve is obtained is minimized (to avoid intrinsic changes in the star and/or planet).

First, it is necessary to define *k*, in the expression *k*·T14, in order to set the minimum size of each one of the segments in which the orbital phase curve is divided. We have experimented with values of *k* between 1 and 5, and the results obtained in *MRS_rand_0* scenario are showed in Table 7.2. It can be observed that, when the minimum size of the curve is increased, it is more difficult to complete the observation of the orbital phase curves because it is more difficult to find large gaps. We have selected a *k* value of 2 because it is not very restrictive and it obtains a reasonable number of parts.

Table 7.3 shows the results obtained observing the transit and occultation of Rosetta Stone targets but without considering the observation of the orbital phase curves. On the other hand, Table 7.4 shows the results obtained when their orbital phase curves are observed, considering a minimum size of each part of 2·T14.

Name	k=1	k=2	k=3	k=4	k=5
55Cnce	F	F	F	F	F
Gliese 1214 b	F	F	F	F	F
Gliese 436 b	F	F	F	F	F
HAT-P-11 b	F	F	F	F	N (46.67%)
HAT-P-32 b	F	F	F	N (56.90%)	N (56.90%)
HAT-P-41 b	F	F	F	N (56.39%)	N (56.39%)
HD 189733 b	F	F	F	F	F
HD 209458 b	F	F	F	F	F
HD 80606 b	U	U	U	U	U
WASP-13 b	F	F	N (54.05%)	U	U
WASP-17 b	F	F	N (54.23%)	N (54.23%)	U
WASP-54 b	F	F	F	U	U
WASP-79 b	F	F	U	U	U
WASP-80 b	F	F	F	F	F

Table 7.2. Results obtained (for one run) for each Rosetta Stone target in the *MRS_rand_0* scenario. Each column indicates a value of *k* used for setting the minimum size of each part of the orbital phase curve. *F* indicates that the orbital phase curve of the exoplanet is fully observed (100%), *N* indicates that only some parts of the curve are observed, and *U* indicates that no part of the curve is planned.

Name	Planning Efficiency	Slew Time	Events Planned	Targets Completed
MRS_rand_0	81.13	5.19	89.30	93.90
MRS_rand_1	88.39	5.05	93.04	96.30
MRS_rand_2	79.85	4.90	86.78	94.22
MRS_rand_3	72.06	4.80	79.65	91.92
MRS_rand_4	85.38	5.16	91.46	94.61
MRS_rand_5	79.50	5.02	87.13	93.65
MRS_rand_6	80.25	4.81	88.76	94.30
MRS_rand_7	82.49	4.76	89.53	94.59
MRS_rand_8	87.34	5.08	92.73	96.96
MRS_rand_9	83.21	5.13	89.18	95.61
MRS_rand Average	81.96	4.99	88.75	94.60

Table 7.3. Results obtained (mean) with the LT-MPT without observing the orbital phase curves of the Rosetta Stones.

Name	Planning Efficiency	Slew Time	Events Planned	Targets Completed
MRS_rand_0	84.22	5.27	89.30	93.90
MRS_rand_1	91.71	5.13	93.04	96.30
MRS_rand_2	83.16	4.98	86.78	94.22
MRS_rand_3	75.24	4.90	79.65	91.92
MRS_rand_4	88.79	5.26	91.46	94.61
MRS_rand_5	82.72	5.12	87.13	93.65
MRS_rand_6	83.65	4.90	88.76	94.30
MRS_rand_7	85.83	4.85	89.53	94.59
MRS_rand_8	90.92	5.16	92.73	96.96
MRS_rand_9	86.69	5.24	89.18	95.61
MRS_rand Average	85.29	5.08	88.75	94.60

Table 7.4. Results obtained (mean) with the LT-MPT considering the observation of the orbital phase curves of the Rosetta Stones.

Table 7.5 shows the complete curves (fully placed or almost fully placed) and the incomplete ones in each scenario. It can be observed that virtually all Rosetta Stone targets can be fully observed in each scenario. Note that the exoplanet *HD 80606 b* is not considered by the LT-MPT because it can be calculated in advance that the required number of events (in this simulation) cannot be placed in a 5-year mission according to the period of the exoplanet and its visibility from the spacecraft. Table 7.6 indicates the maximum distance between the periods where each part of the orbital phase curve of each Rosetta Stone target is placed. It can be observed that the results depend on the scenario but the example illustrates that it is possible to obtain phase curves that are covered over a relatively short period of time. Note that some optimization can still be done if internal target priorities are set.

Name	0	1	2	3	4	5	6	7	8	9
55Cnce	F	F	F	F	F	F	F	F	F	F
Gliese 1214 b	F	F	F	F	F	F	F	F	F	F
Gliese 436 b	F	F	F	F	F	F	F	F	F	F
HAT-P-11 b	F	F	F	F	F	F	F	F	F	F
HAT-P-32 b	F	F	F	F	F	F	F	F	F	F
HAT-P-41 b	F	F	A (99.23%)	F	F	F	F	F	F	F
HD 189733 b	F	F	F	F	F	F	F	F	F	F
HD 209458 b	F	F	F	F	F	F	F	F	F	F
HD 80606 b	U	U	U	U	U	U	U	U	U	U
WASP-13 b	F	F	F	F	F	F	A (98.12%)	F	F	F
WASP-17 b	F	F	F	F	F	F	F	F	F	F
WASP-54 b	F	F	F	F	F	F	F	F	F	F
WASP-79 b	F	N (73.02%)	F	F	N (74.14%)	F	F	F	F	F
WASP-80 b	F	F	F	F	F	F	F	F	F	F

Table 7.5. Results obtained (for one run) for each Rosetta Stone target. Each column indicates the *MRS_rand_#*, and *F* indicates that the orbital phase curve of the exoplanet is fully observed (100%), *A* indicates that it is almost fully observed (at least, 90%), *N* indicates that only some parts of the curve are observed, and *U* indicates that no part of the curve is planned.

Name	0	1	2	3	4	5	6	7	8	9	Orbits in a 5-year mission
55Cnce	0	0	0	0	0	0	12	10	17	1	2478
Gliese 1214 b	7	4	2	4	4	4	52	15	44	4	1155
Gliese 436 b	8	4	5	4	9	6	6	49	53	5	690
HAT-P-11 b	13	16	10	16	36	10	9	18	46	22	373
HAT-P-32 b	6	4	0	5	30	1	71	67	49	7	849
HAT-P-41 b	3	12	3	7	8	3	16	9	9	24	677
HD 189733 b	4	4	3	2	3	6	11	11	8	5	823
HD 209458 b	12	13	4	6	51	14	7	6	2	45	518
WASP-13 b	49	82	119	78	74	118	6	7	8	174	419
WASP-17 b	51	13	42	44	106	10	4	2	4	47	489
WASP-54 b	57	53	50	53	138	44	5	7	4	152	494
WASP-79 b	68	104	40	50	123	17	5	3	3	99	498
WASP-80 b	18	12	7	9	54	11	0	0	0	55	595

Table 7.6. Period distance between the parts of the orbital phase curve of each Rosetta Stone target (for one run). Each column indicates the *MRS_rand_#* and the last one indicates the number of periods that are visible in a 5-year mission.

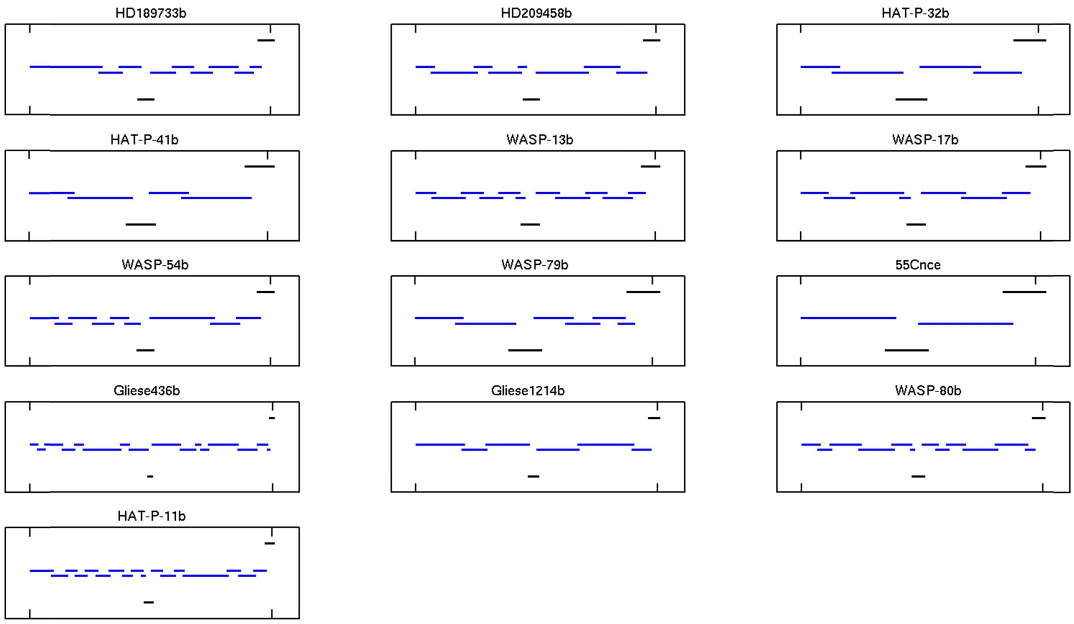

Figure 7.1. Segments of each orbital phase curve of the exoplanets in *MRS_rand_0*. The segments are showed in the exoplanet phase. Lines in black indicate the occultation or transit observations, and blue lines indicate the observations of the other parts of the curve.

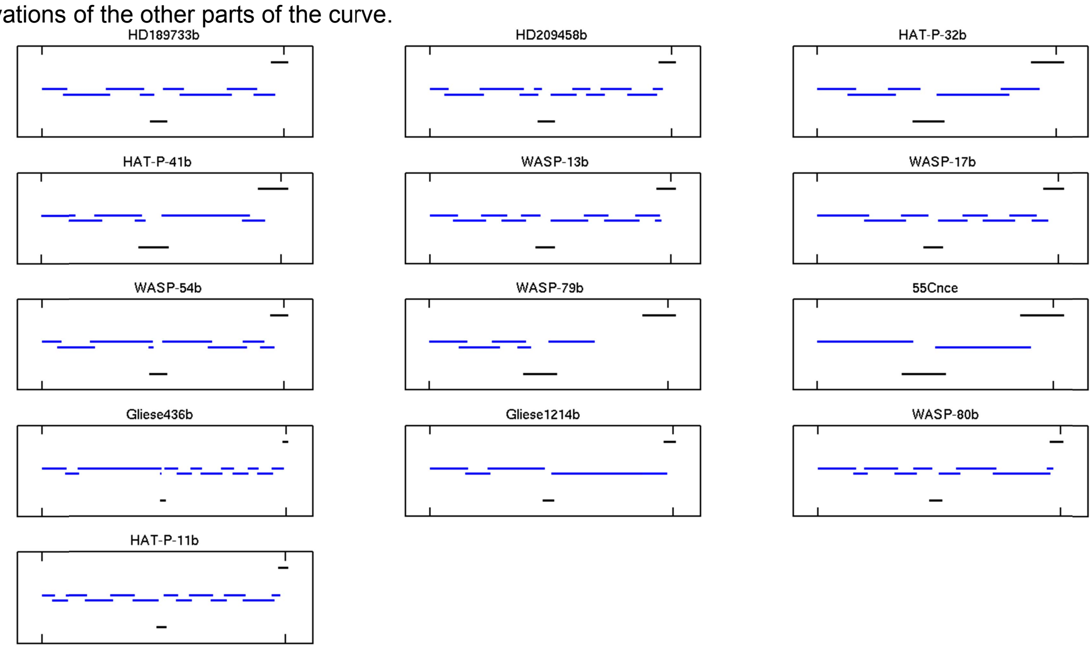

Figure 7.2. Sections of each orbital phase curve of the exoplanets in the *MRS_rand_1*. The sections are showed in the exoplanet orbital phase. Llines in black indicate the occultation or transit observations, and blue lines indicate the observations of the other parts of the curve.

Finally, Figure 7.1 and Figure 7.2 show all segments that the orbital phase curve of each exoplanet has been divided into, according to the exoplanet phase, for *MRS_rand_0* and *MRS_rand_1* scenarios, respectively. We can clearly observe that, for each Rosetta Stone target, the number of curve segments and their length depends on each scenario. Note that, in Figure 7.2, there is a gap in the orbital phase curve of the exoplanet *WASP-79b*. The orbital phase curve of this exoplanet is not fully observed in the *MRS_rand_1* scenario, so the gap corresponds to the small interval of the curve that is finally not observed.

8 CONCLUSIONS

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 scientific success. However, an enormous amount of possible combinations exists and finding an optimal solution becomes unaffordable for human planners. 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 based on six steps for an automatic planning of the mission. These steps 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) inserting downlinks by minimizing potential conflicts with critical targets, (4) obtaining LTMP by avoiding overlap and maximizing the planning efficiency and the scientific return, (5) removing observations of the targets that are not planned at least *m*% of their required events, and (6) filling gaps with new observations or other operation tasks. The key steps of this process are the downlink and station keeping positioning and the observation planning, which are considered two optimization problems. Moreover, we have proposed a definition of the target priority according to class priority and to the priority of the targets inside each class. Note that this priority calculation can be changed without affecting the presented process (e.g., assigning a specific priority to targets in the exoplanet sample).

Two algorithms have been proposed for addressing both optimization problems in the LT-MPT. The first one is called DSKO and is focused on placing the downlinks when they restrict a smaller number of targets with high criticality. The second one is called OPO and is focused on planning target observations by avoiding any overlap, and maximizing the planning efficiency and the scientific return. The OPO approach presented in the document has been called OPO-GAE. Both optimization algorithms (DSKO and OPO-GAE) are resolved with Artificial Intelligence techniques based on Evolutionary Algorithms (EAs), which are methods based on the way nature solves the problem of living entities by means of natural selection and evolution. In particular, we have used Genetic Algorithms (GAs), which are a search and optimization paradigm that makes possible to explore the regions of the search space where the best solutions are located. We have described in Annex A the main characteristics of these algorithms in order to help the reader understand the explanation of the proposed approaches.

A detailed experimentation of the LT-MPT using several configurations of the positioning of downlinks and of the OPO approach is discussed in Section 6, Annex B and Annex C. Specifically, the LTMPs obtained for five year mission in ten different artificial samples (sample realizations) and in a real 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 with a high coverage of the defined classes. 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 go on discovering new targets of EChO interest, modifying the available list of exoplanets. It is noteworthy that the conducted experiments allow us to determine the sizing of the mission with the aim of guarantee the feasibility of the mission, as well as to establish a proper design of the LTMP prototype.

Finally, this work concludes that Genetic Algorithms are useful in obtaining an optimal LTMP for the EChO mission. Three methodologies for placing the downlink operations have been identified in Section 5.2.1 and Annex B, and the experimentation has showed that downlinks can affect the planning efficiency and the scientific return of the resulting LTMP. According to the operational design of the EChO mission described in Section 3, the final LT-MPT will likely be based on combining the OPO-GAE approach without placing downlinks and station keepings, and subsequently placing them using the gaps between the planned observations.

ANNEX A. INTRODUCTION TO GENETIC ALGORITHMS

Evolutionary Algorithms (EAs) are an Artificial Intelligence paradigm that includes the learning algorithms which is based on the way nature solves the problem of living entities (Cordón et al., 2001; Freitas, 2002) by means of natural selection (Darwin, 1859) and evolution (Mendel, 1865). Genetic Algorithms (GAs) are one of the techniques included in the EAs, and they are focused on finding candidate solutions to an optimization problem (Holland, 1992). They are theoretically and empirically proven to provide a robust search in complex spaces, thereby offering a valid approach to problems requiring efficient and effective searches (Cordón et al., 2001; Goldberg, 1989). GA process is roughly based on selection, reproduction and mutation. This paradigm makes it possible to explore all regions of the search space, which is a vast area with a large amount of potential solutions, in search of the best solutions. This kind of algorithms begins with a set of initial solutions that are improved through an iterative cycle based on evaluating, selecting, recombining and mutating them. The key aspect for finding high quality solutions lies in individual representation and genetic operators. The main shortcoming of these techniques is that they are expensive in terms of computational time and memory usage, and this is clearly noticeable when they are applied to a very large amount of data. However, these limitations are not significant when real-time response is not required and there are no memory restrictions. The next two sections introduce the main concepts of a GA, its process, and some strategies for improving its scalability and how to optimize several objectives.

A.1 NATURAL PRINCIPLES

These concepts of natural selection and evolution are introduced in the GAs to do a guided search of the solution in wide solution spaces where an exhaustive or random search cannot be sufficiently accurate. It is considered that this search is directed because the population is guided towards the desired solution through the fitness landscape. One of the key elements of the GA process is how it represents a suitable individual (or potential solution) for the problem to solve. Figure A.1 shows the fundamentals of an individual representation (genotype or chromosome, phenotype, gene and allele) with the aim of introducing the reader to the main concepts of a GA. In terms of natural selection emulation, it is done by means of selection, crossover, mutation and replacement of the individuals of the population. On the other hand, the emulation of evolution is done with a fitness function that evaluates the best individuals of a population, being the fittest ones the individuals with more probabilities to survive.

A.2 PROCESS

As mentioned before, the process of a GA emulates the natural selection and evolution principles in which it is based. The main steps of this emulation are detailed as follows (Bacardit, 2004):

1) Generation of the initial population: The individuals used in the first generation of the GA are created according to the individual representation chosen. Each individual is a candidate solution to the problem to solve.
2) Evaluation of the fitness function: Each individual of the population is evaluated according to the defined fitness function. That is, how well the individual performs in solving the problem.
3) Selection of the parents: Some individuals of the population are selected as parents to produce offspring.
4) Crossover and mutation: The individuals selected as parents are combined with the crossover operator, and the new individuals (or offspring) generated are slightly modified with the mutation operator (Freitas, 2002). The aim of the mutation operator is to be a mechanism of introducing diversity in the population. Figure A.1 shows an example of crossover and mutation operators. In this example, a new individual is generated from two previous individuals using the crossover operator, which mixes the first half of the genotype of the first parent with the second half of the genotype of the second parent. Finally, the generated individual is slightly modified with the mutation operator, which changes the allele of one gene, obtaining the final genetic information of the new individual. This final individual is evaluated with the same fitness function used in the second point.

5) Replacement of the population: When the final new individuals are obtained, it is necessary to add them to the population. Given the original population and some new individuals (i.e., the offspring), the replacement operator is responsible for merging them, thus obtaining a new set of solutions for subsequent iterations.
6) Achievement of the optimization criteria: It defines when the algorithm ends (e.g., likelihood reached, or number of generations done). If the ending criterion is reached, the algorithm returns the best individual of the population according to the fitness function. If it is not reached, it returns to the third point.

It must be emphasized that, to successfully apply a GA to real-world problems, it is important to choose a suitable individual representation according to the domain of the problem, due to the fact that the representation defines the search space. Thus, individual representation and how it is initialized are two important issues in evolutionary algorithms, and the selection of them is directly related to the domain characteristics. Moreover, the genetic operators have the goal of exploring new areas in the search space, but an uncontrolled application could lose the focus on the right search way in some individual representation. Thus, it is necessary to do a previous analysis of the suitability of applying each genetic operator in the individual representation used.

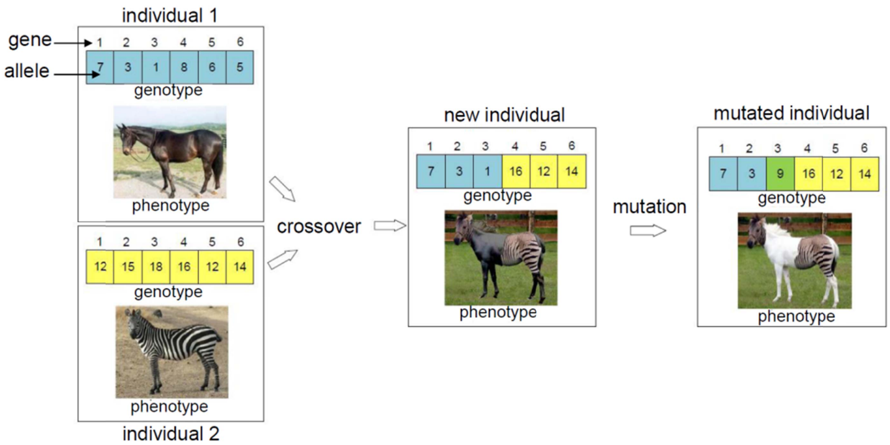

Figure A.1. The main concepts of an individual representation (genotype or chromosome, phenotype, gene and allele) and the genetic operators (crossover and mutation) are shown. Adapted from Garcia-Piquer (2012).

Finally, it is important to highlight that there are two main approaches related to the interpretation of the individuals in the GAs (Freitas, 2002): Pittsburgh and Michigan. Simplifying the explanation of the differences between them, in the Pittsburgh approach each individual is a candidate solution to the problem, thus the individuals are independent. In the Michigan approach each individual is part of the solution to the problem, so all individuals make up the complete solution. This approach is useful due to the fact that the individuals can be simpler in comparison with the Pittsburgh approach. Nevertheless, in the Pittsburgh approach the fitness function will measure the performance of an individual without taking into account the other individuals. In a scheduling problem, if each individual represents a part of the plan, the fitness function can only evaluate the quality of this part, but it cannot evaluate the quality of the overall scheduling solution. Because of this, the Pittsburgh approach is more suitable to solve our scheduling problem, due to the fact that when the fitness function evaluates an individual it is evaluating the quality of the overall scheduling solution.

A.3 Scalability

The main lack of GAs is their high computational time and memory usage, being this effect worse when they are applied to large data sets (Freitas, 2002). In this section, we briefly comment on how to speed up a GA by reducing its computational cost, and how memory usage can be optimized with a suitable individual representation according to the problem.

Two main approaches for speeding up GAs are commonly used. The first one is based on parallelizing some steps of the GA (Cantu-Paz, 2000), and the second one consists in using only a subset of the available data to evaluate the individuals (Bacardit, 2004). The approach based on parallelizing the algorithm does not affect to the accuracy of the method, due to the fact that the idea is to distribute the computational load of some expensive steps of the GA among several processors without changing the way that the individuals are evaluated. However, this approach requires adapting the algorithm and introduces communication costs related to the coordination of the task executions. Commonly, the most expensive step is the individual evaluation, so the fitness evaluation is usually parallelized. For example, the set of individuals of the GA are distributed across all the processors, evaluating each one a different subset of individuals. The second approach is focused on evaluating the individuals using subsets of data instances from the complete data set to evaluate the individuals instead of using the complete data set. This approach does not require the adaptation of the algorithm, but the accuracy of the method can be reduced because the individuals are evaluated with less information that can be biased. For example, in a scheduling problem, if the fitness function consists in evaluating how many tasks have been completed, this approach will evaluate the individual using only the count of some tasks.

The memory usage of GAs is related to the amount of data to manage but it mainly depends on the individual representation used. For example, two ways of representing an individual of a scheduling problem, focused on planning several tasks in a year, can be: (1) with a genotype where each gene represents a time slot of the full year and the allele the task assigned to the time slot, and (2) with a genotype where each gene is the task to be planned and the allele the time slot of the year when it starts. Assuming that there are less tasks to be planned than time slots in the year, representing the individuals employing the second approach is substantially less expensive in terms of memory usage. Therefore, to optimize memory usage it is necessary to analyse several representations in order to use the one that simplifies the problem. Moreover, in some situations, a suitable individual representation can improve the computational cost of the individual evaluation.

Finally, it is worth noting that the application of these strategies is not essential when computation time and memory usage are not a limitation.

A.4 MULTIOBJECTIVE OPTIMIZATION IN GENETIC ALGORITHMS

The Multiobjective Optimization Problem (MOP) can be defined as the problem of finding a vector of decision variables which satisfies constraints and optimizes a vector function whose elements represent the objective functions (Osyczka, 1985). These functions form a mathematical description of performance criteria which are usually in conflict with each other. Hence, the term “optimize” means finding such a solution which would give the values of all the objective functions acceptable to the decision maker (Coello, 1999). It is rarely the case that there is a single point that simultaneously optimizes all the objective functions of a MOP. Therefore, in these problems it is necessary to look for trade-offs, rather than single solutions. The concept of Pareto Optimality (Pareto, 1896) defines that we can consider a Pareto optimal when it exists no feasible vector of decision variables that would decrease some criterion without causing a simultaneous increase in at least one other criterion. Thus, this concept almost always gives not a single solution, but rather a set of solutions called the Pareto optimal set. All the solutions included in the Pareto optimal set are non-dominated. The plot of the objective functions whose non-dominated vectors are in the Pareto optimal set is called the Pareto front (Coello, 2001). The reader is referred to (Coello, 1999) for the details.

Multiobjective Evolutionary Algorithms (MOEAs) are GAs focused on optimizing several objectives simultaneously obtaining a trade-off among them (Garcia-Piquer, 2012). They were proposed by Rosenberg in 1967 (Rosenberg, 1967), but the first algorithm (VEGA) was created by Schaffer in 1984 (Schaffer, 1985). There are two different kinds of MOEA: Non-Pareto based and Pareto based. The first one does not incorporate directly the concept of Pareto optimum, so it is incapable of producing certain portions of the Pareto front (Coello, 2001). The second one consists of a selection scheme based on the concept of Pareto optimality. The most representative algorithms are SPEA, SPEA2, NSGA-II, MOMGA, MOMGA-II, PAES, PESA and PESA-II (Coello et al., 2007; Zitzler et al., 2000).

ANNEX B. DOWNLINK POSITIONING STRATEGIES

In this annex, several alternative strategies for downlink positioning are compared with the strategy exposed in Section 5.2.1 (see Figure B.1) for analysing their performance combining them with the OPO approach based on GAs presented in Section 5.2.2 (OPO-GAE). The two alternative strategies for placing the downlinks are: (1) the LTMP is obtained with OPO-GAE considering that downlinks are placed in their default dates (i.e., one downlink each 3.5 days), and (2) downlinks and station keepings are placed with the DSKO algorithm and next the LTMP is obtained with OPO-GAE. The configuration of the algorithms is the same indicated in Section 6.1.2. The next sections analyse, from the point of view of the metrics defined in Section 6.1.2, the results obtained for the three strategies executed in combination with the OPO-GAE and the configuration described in Section 6.1.3.

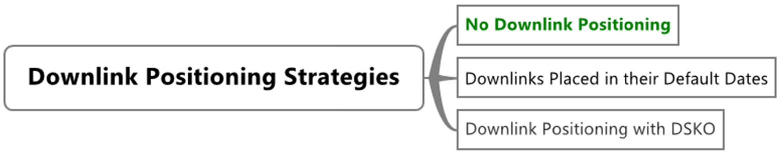

Figure B.1. Downlink positioning strategies proposed and compared, in combination with the OPO-GAE approach, in the LT-MPT. The strategy identified as the best one is highlighted in green colour.

B.1 PLANNING EFFICIENCY

The results of the Planning Efficiency metric obtained in the three downlink positioning strategies are shown in Table B.1. The three methods show a high planning efficiency, but it can be observed that to place the downlinks in their default positions has the lower planning efficiency. This is because the observation of some exoplanet events is restricted by the rigid positioning of the downlinks. On the other hand, the DSKO strategy has a higher planning efficiency due to the fact that it can avoid some restrictions of the downlinks by placing them with some flexibility. The best planning efficiency is obtained when downlinks are no placed, because they do not restrict any observation of exoplanet events. It must be emphasized that in the real sample scenario, the DSKO results have an improvement of more than a 10% of planning efficiency compared with the default downlinks strategy, and it obtains similar results to the strategy that does not place downlinks.

Name	Default Downlinks	DSKO	No Downlinks
MRS_rand_0	82.43	85.21	88.08
MRS_rand_1	89.82	92.02	94.55
MRS_rand_2	81.82	83.73	87.49
MRS_rand_3	74.39	76.77	80.21
MRS_rand_4	88.69	90.93	94.22
MRS_rand_5	81.72	84.98	87.91
MRS_rand_6	83.87	87.41	88.48
MRS_rand_7	83.22	87.56	89.38
MRS_rand_8	89.90	91.83	94.35
MRS_rand_9	85.10	87.72	91.89
MRS_rand Average	84.10	86.82	89.66
Real_sample	87.58	98.57	98.70

Table B.1. Results obtained for the Planning Efficiency metric (mean) for each one of the downlink positioning strategies in combination with the OPO-GAE in the defined scenarios.

B.2 SLEW TIME

Table B.2 shows the results of the Slew Time metric. It can be observed that the slew time is related with the planning efficiency, so the strategies with higher planning efficiency have a larger value.

Name	Default Downlinks	DSKO	No Downlinks
MRS_rand_0	4.78	4.86	5.44
MRS_rand_1	4.58	4.66	5.26
MRS_rand_2	4.50	4.58	5.17
MRS_rand_3	4.48	4.58	5.22
MRS_rand_4	4.79	4.86	5.44
MRS_rand_5	4.56	4.66	5.27
MRS_rand_6	4.50	4.62	5.14
MRS_rand_7	4.36	4.53	5.06
MRS_rand_8	4.67	4.71	5.29
MRS_rand_9	4.76	4.83	5.49
MRS_rand Average	4.60	4.69	5.28
Real_sample	1.81	2.00	2.43

Table B.2. Results obtained for the Slew Time metric (mean) with the LT-MPT for each one of the downlink positioning strategies in combination with the OPO-GAE in the defined scenarios.

B.3 EVENTS PLANNED AND TARGETS COMPLETED

Table B.3 and Table B.4 show the results obtained for the Events Planned and Targets Completed metrics, respectively. The three downlink positioning strategies have similar values for both metrics. However, it can be observed that placing the downlinks in their default position has the lower results, and the other two strategies can place almost a 10% more of events.

Name	Default Downlinks	DSKO	No Downlinks
MRS_rand_0	90.34	91.69	93.28
MRS_rand_1	93.57	95.09	96.62
MRS_rand_2	87.46	88.98	90.75
MRS_rand_3	81.05	82.76	85.03
MRS_rand_4	92.82	94.01	95.97
MRS_rand_5	88.09	89.65	91.72
MRS_rand_6	91.03	93.04	93.50
MRS_rand_7	90.00	92.81	93.64
MRS_rand_8	94.10	94.91	96.34
MRS_rand_9	90.15	91.81	94.22
MRS_rand Average	89.86	91.48	93.11
Real_sample	89.50	98.78	98.90

Table B.3. Results obtained for the Events Planned metric (mean) with the LT-MPT for each one of the downlink positioning strategies in combination with the OPO-GAE in the defined scenarios.

Name	Default Downlinks	DSKO	No Downlinks
MRS_rand_0	94.83	95.84	96.69
MRS_rand_1	96.72	97.75	98.75
MRS_rand_2	96.01	96.75	97.99
MRS_rand_3	93.25	94.06	95.50
MRS_rand_4	96.94	97.93	98.93
MRS_rand_5	95.49	96.67	97.29
MRS_rand_6	95.48	96.95	97.46
MRS_rand_7	95.30	96.85	97.44
MRS_rand_8	97.81	98.56	99.15
MRS_rand_9	97.00	97.63	98.81
MRS_rand Average	95.88	96.90	97.80
Real_sample	99.17	100.00	100.00

Table B.4. Results obtained for the Targets Completed metric (mean) with the LT-MPT for each one of the downlink positioning strategies in combination with the OPO-GAE in the defined scenarios.

B.4 DISCUSSION OF THE COMPARISON

Finally, we can conclude that placing the downlinks with the optimization algorithm DSKO or by filling the gaps are two appropriate procedures, as both maximize the science output of EChO. Nevertheless, it should be emphasized that in the DSKO the final position of each downlink is not moved more than a 10% (8.4 hours) from its default position. On the other hand, placing the downlinks by filling the gaps shows that, with the same 10% of flexibility, there are around 20% of downlinks that cannot be placed without being in conflict with some observations. Thus, the conflictive event observations must be removed and the performance of the LTMP will be reduced. For this reason we can conclude that the DSKO strategy is the most useful approach for placing the downlinks since they restrict a lower number of event observations. However, according to the operational design of the EChO mission, the DSKO strategy may be difficult to be implemented due to the fact that it does not consider conflicts with the booking of Ground Station by other missions. Therefore, it is likely that the final method will be close to that of placing downlink operations within gaps by taking all constraints that the MOC considers appropriate.

ANNEX C. ALTERNATIVE APPROACHES FOR THE OBSERVATION PLANNING OPTIMIZATION

The OPO approach used in the experimentation presented in Section 6.2 (OPO-GAE) is based on a GA with an individual representation based on the events to be planned, and it seems that the results are promising for the EChO mission. However, other approaches have been tested in the current version of the LT-MPT with the aim of analyse if the obtained results can be improved (see Figure C.1). One of the approaches uses a GA with an individual representation based on a sequence of observations (OPO-GAO). The other one uses a Multiobjective Evolutionary Algorithm as optimization tool, and we have tested it with two different pairs of objectives: one based on optimizing the planning efficiency and the slew time (OPO-MGS), and another one based on optimizing the number of complete targets and the slew time (OPO-MTS).

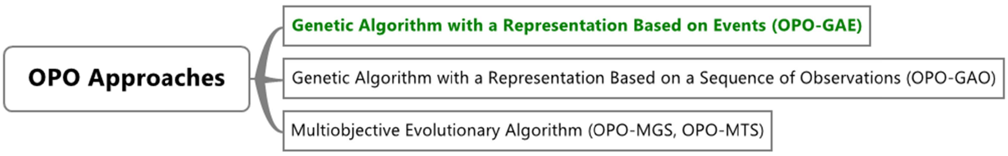

Figure C.1. OPO approaches proposed and compared, in combination of the DSKO strategy, in the LT-MPT. The approach identified as the best one is highlighted in green colour.

C.1 GENETIC ALGORITHM REPRESENTATION BASED ON AN OBSERVATION SEQUENCE

This section presents an alternative approach for OPO based on GAs but using a different individual representation to that used in Section 5.2.2. The main idea of this approach is to build only individuals that represent feasible LTMP, avoiding the repair phase. The differences with the representation presented in Section 5.2.2 are the definition of the genetic operators, due to the fact that the individual representation is changed, and the reparation of the individual is removed.

C.1.1 Individual Representation

The individual genotype is made up of pairs of integer numbers which represent the operation tasks (i.e., targets, downlinks and station keepings) and the time windows where they are planned. Each individual consists of $\overline{\overline{O}}$ genes $\{o_1, o_2, \ldots, o_{\overline{\overline{O}}}\}$, where $\overline{\overline{O}}$ is the number of operation tasks planned sorted by their initial date, and o_i correspond to operation task *i*. Note that the order of the targets in the genotype indicates a temporal sequence, so an operation task in the position *i* is previous in time than an operation task in position *i + 1*. An example is shown in Figure C.2, where the same situation presented in Figure 5.2 is used. The image describes three targets to be planned presents a possible fragment of an LTMP (note that it is not the optimal one). Specifically, the LTMP is depicted with the genotype of the individual and its interpretation, where each gene is referred to an observation of a target event sorted by initial date of the observation. The allele of each gene indicates the target and the window assigned to each observation (the matching between alleles and windows of each target is shown in green colour). Note that, unlike the approach presented in Section 5.2.2, only the planned operation tasks are considered in the genotype. Therefore, the individuals of the population can have different size.

It must be emphasized that individuals represent feasible LTMPs without any additional process, because this representation does not allow overlapped observations and the slew time between each pair of consecutive observations can be directly calculated.

The initial population is built by creating N_I new individuals placing the observations of the targets, selected in a random order, and the rest of operation tasks by avoiding overlap. In the situation where a target observation cannot be planned without overlap, this is dropped. For more details, see Algorithm C.1.

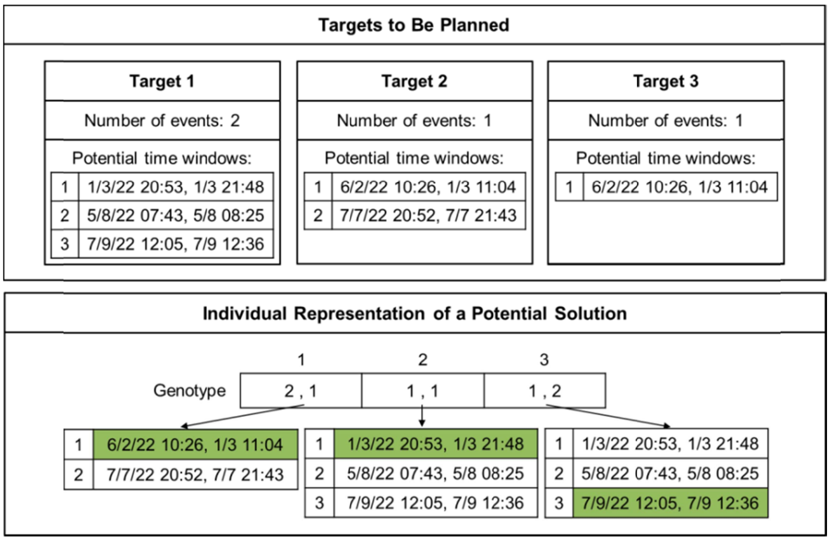

Figure C.2. The example provides some targets to be planned and a potential LTMP using the individual representation proposed in the alternative individual representation of the GA approach.

Let *Targs* be the collection of targets
Let *Ind* be an empty collection of genes

Add to *Ind* each downlink and station keeping sorted by their initial date
foreach target *t* in *Targs* selected in random order **do**
  foreach event e from *t* **do**
    $w_r \leftarrow$ random window of the potential time windows of target *t*
    Ind $\leftarrow$ *AddObservation*(*Ind*, w_r, *t*) //It is equivalent to the process presented in Algorithm 5.5
  endfor
endfor
return *Ind*

Algorithm C.1. Process followed for building a new individual in the alternative GA approach.

C.1.2 Genetic Operators

The genetic operators of selection and replacement are the same ones defined in Section 5.2.1.2 for the DSKO. The crossover and the mutation operators are modified and adapted to the individual representation, as explained below:

- Crossover: the process of the uniform and one-point crossover operators is the same presented in Section 5.2.1.2 but with some modification because of the individuals can only represent feasible plans. Thus, each gene that represents a target observation is not added to the corresponding offspring, if it causes overlapping with any operation task previously planned in the offspring. Moreover, we do not recommend the use of uniform crossover due to the fact that in this approach the exploration of the search space will be less exhaustive. This is because the phenotype of the offspring will be considerably changed, with respect to the parents, since each conflictive gene will not be considered in them. For this reason, and taking into account that individuals have different size, we consider that one-point crossover will mix properly the genetic information of the parents for building the offspring.

- Mutation: in this approach, an individual is mutated by randomly choosing to add or drop target observations. Specifically, if the mutation selects to add targets, a random number of target observations (with random windows) are added to the individual if it is not overlapping and if the particular target is not observed more than *M%* of its required number of events. On the other hand, if the mutation chooses to drop targets, a random number of target observations are selected from the individual and removed from it.

Note that in this approach, crossover and mutation operators always obtain feasible individuals.

C.2 MULTIOBJECTIVE APPROACH

One interesting aspect in OPO from the point of view of reduction of the operational costs is to optimize the slew time, trying to obtain a solution that minimizes it. However, in EChO, the optimization of the slew time is in conflict, mainly, with the planning efficiency and the target completeness. This is because the reduction of the slew time means to do fewer observations (i.e., the only way to have a slew time of 0 is to do not observe any target or to only observe all the required events of a single target). In this situation, it can be interesting to use the MOP concept (see Annex A.4) to obtain a trade-off between the optimization of the slew time and the optimization of the planning efficiency or the target completeness. Thus, Pareto based MOEAs can be a useful technique to carry out this evaluation.

In this section we present how to apply this kind of MOEAs to the OPO approach presented in Section 5.2.2. The main differences with the approach of Section 5.2.2 is the definition of some genetic operators, the objective functions and the need for a new step that is focused on retrieving the most suitable plan from the set of solutions included in the Pareto optimal set.

C.2.1 Genetic Operators

The GA process used in Section 5.2.2 has been adapted to solve multiobjective problems using the NSGA-II (Deb et a., 2002) procedure, which is one of the most well-know MOEA. This algorithm is a non-domination based genetic algorithm for multiobjective optimization. In the NSGA-II, the individual population is initialized as usual, and the individuals in the population are sorted in several fronts based on non-domination. The first front includes a set of non-dominated individuals according to the current population, the second front includes the individuals that are only dominated by the individuals in the first front, and so it continues until the last of the fronts. The individuals have a rank assigned according to the front in which they belong to. Thus, individuals in the first front have a rank value of 1, individuals in the second one have a rank value of 2, and so on. For computing if an individual is dominated or non-dominated it is necessary to assign to them a value for each one of the objectives evaluated in the algorithm. In terms of the genetic operators, crossover and mutation are the same ones used in Section 5.2.2. However, the selection and replacement operators are different:

- Selection: The process is similar to the one presented in Section 5.2.2, but in this case $\bar{\bar{P}}$ parents are selected from the current population P by using a tournament selection based on the rank and crowding distance. The crowding distance is a measure of how close an individual is to its neighbors according to the value of the evaluated objectives. Therefore, an individual is selected if the rank is lower than the other ones or if crowding distance is greater than the other ones. After selecting the parents, the crossover and mutation operators generate $\bar{\bar{P}}$ offspring.
- Replacement: The offspring obtained after the application of the crossover and mutation operators is merged with the individuals of the current population in a temporal one. The temporal population is sorted in non-dominated fronts. Next, only the best N_p individuals are added to the population used in the next generation. Note, that individuals are selected based on their rank and on their crowding distance if they belong to the same front.

Finally, when the last generation of the NSGA has been achieved, the algorithm return the optimal front of the population (i.e., all the individuals that belong to the first front, which are non-dominated by any individual of the rest of the population). Thus, all the individuals have a different trade-off between objectives but there is no individual better than the other ones. For this reason, it is necessary to add a new step at the end of the algorithm for identifying the most suitable solution according to some specific criteria.

C.2.2 Objective Functions

In EChO, we can identify two pair of objectives that are in conflict between them:

- F_G and F_S: the minimization of the time that the telescope is not observing (F_G, defined in Equation 5.4) and the minimization of the time that the telescope is doing slewing (F_S, defined in Equation C.1).

- F_T and F_S: the minimization of the number of incomplete targets weighted with the priority of the targets (F_T, defined in Equation 5.5) and F_S.

The approach based on F_G and F_S is called OPO-MGS, and the approach based on F_T and F_S is called OPO-MTS. It is important to highlight that in a multiobjective strategy, the objectives are individually evaluated. Thus, the value of each one of the objective functions defined is stored for each individual.

$$F_S(I) = \frac{\sum_{o \in O_I} slew\ time\ between\ o\ and\ the\ following\ observation}{overall\ mission\ time}$$

Equation C.1

C.2.3 Selection of the Most Suitable Solution

The results of the MOEA algorithm used are a set of solutions that correspond to the optimal front of the final population. Each one of these individuals has a different trade-off between the defined objectives. Thus, there is not an individual better than the other ones. For this reason, it is necessary to define the criteria to select the most suitable solution for EChO. These criteria depend on the pair of objectives used for the optimization step, and they are listed as follows:

- OPO-MGS: the criterion used in this case is to select the individual that minimizes the F_T measure, which will be the individual from the optimal front with more targets completed.
- OPO-MTS: in this situation, the criterion is based on selecting the individual that minimizes the F_G measure, which will be the individual from the optimal front with higher planning efficiency.

Finally, the selected solution is improved by removing the incomplete targets and filling gaps with new observations. It must be emphasized that only two objectives are used in the optimization process although three objective functions are used in the overall process. The third objective does not change the obtained LTMPs but it is used for selecting one of them.

C.3 OBSERVATION PLANNING OPTIMIZATION APPROACHES COMPARISON

In this section we want to analyse the performance of the presented OPO approaches (OPO-GAE, OPO-GAO, OPO-MGS and OPO-MTS) executed in combination with the DSKO algorithm in order to compare the results when downlinks and station keepings are placed. The configuration of OPO-GAO is the same specified in Section 6.1.2 but with a number of generations of 25000. The configuration of OPO-MGS and OPO-MTS is also the same specified in Section 6.1.2 (the GA parameters correspond to the MOEA parameters).

C.3.1 Planning Efficiency

Table C.1 summarizes the results obtained with the analysed OPO approaches for the Planning Efficiency metric. It can be observed that the OPO-GAE approach is considerably better for this metric. On the other hand, the OPO-GAO is significantly worse than the other ones. Moreover, the multiobjective approaches (OPO-MGS and OPO-MTS) have similar values, but the OPO-MGS is slightly better due to it considers the planning efficiency as an objective to be optimized (F_G).

Name	OPO-GAE	OPO-GAO	OPO-MGS	OPO-MTS
MRS_rand_0	85.21	69.05	79.02	80.01
MRS_rand_1	92.02	71.06	86.54	84.78
MRS_rand_2	83.73	68.33	79.07	76.03
MRS_rand_3	76.77	62.74	71.85	71.36
MRS_rand_4	90.93	69.98	84.80	78.58
MRS_rand_5	84.98	69.04	76.75	76.70
MRS_rand_6	87.41	70.52	78.63	74.02
MRS_rand_7	87.56	66.49	80.41	78.82
MRS_rand_8	91.83	73.68	82.60	82.00
MRS_rand_9	87.72	68.88	80.92	80.25
MRS_rand Average	86.82	68.97	80.05	78.25
Real_sample	98.57	76.86	97.31	87.09

Table C.1. Results obtained for the Planning Efficiency metric (mean) with the LT-MPT based on DSKO and the OPO approaches in the defined scenarios.

C.3.2 Slew Time

Table C.2 shows the Slew Time metric values of the three approaches. We can see that the slew time is related to the planning efficiency, due to the fact that the approaches with lower planning efficiency have lower slew time.

Name	OPO-GAE	OPO-GAO	OPO-MGS	OPO-MTS
MRS_rand_0	4.86	4.21	4.64	4.66
MRS_rand_1	4.66	3.88	4.40	4.43
MRS_rand_2	4.58	4.09	4.36	4.42
MRS_rand_3	4.58	3.87	4.33	4.29
MRS_rand_4	4.86	4.04	4.60	4.39
MRS_rand_5	4.66	4.17	4.35	4.39
MRS_rand_6	4.62	4.02	4.22	4.23
MRS_rand_7	4.53	3.72	4.23	4.18
MRS_rand_8	4.71	3.94	4.45	4.51
MRS_rand_9	4.83	4.07	4.58	4.63
MRS_rand Average	4.69	4.00	4.41	4.41
Real_sample	2.00	1.65	1.75	1.64

Table C.2. Results obtained for the Slew Time metric (mean) with the LT-MPT based on DSKO and the OPO approaches in the defined scenarios.

C.3.3 Events Planned and Targets Completed

Table C.3 and Table C.4 summarize the results for the Events Planned and Targets Completed metrics, respectively. It can be observed that OPO-GAE is the approach with the highest percentage of events planned and of targets completed. In the other hand, the OPO-GAO approach obtains the lowest percentage of events planned but has a considerable percentage of targets completed.

The multiobjective approaches can place a similar number of events in the sample realizations. However, the OPO-MTS approach is able to complete more targets due to the fact that it consider the number of targets completed (F_T) in the objectives to be optimized.

Name	OPO-GAE	OPO-GAO	OPO-MGS	OPO-MTS
MRS_rand_0	91.69	80.85	88.06	88.92
MRS_rand_1	95.09	80.03	92.28	91.03
MRS_rand_2	88.98	79.66	86.53	85.00
MRS_rand_3	82.76	70.51	79.40	79.11
MRS_rand_4	94.01	78.56	91.04	86.61
MRS_rand_5	89.65	80.34	85.89	85.95
MRS_rand_6	93.04	81.36	88.19	84.80
MRS_rand_7	92.81	77.67	88.31	88.00
MRS_rand_8	94.91	80.1	90.78	90.30
MRS_rand_9	91.81	76.98	88.30	87.87
MRS_rand Average	91.48	78.60	87.87	86.75
Real_sample	98.78	81.67	97.73	89.29

Table C.3. Results obtained for the Events Planned metric (mean) with the LT-MPT based on DSKO and the OPO approaches in the defined scenarios.

Name	OPO-GAE	OPO-GAO	OPO-MGS	OPO-MTS
MRS_rand_0	95.84	89.59	80.01	93.68
MRS_rand_1	97.75	91.14	84.78	95.20
MRS_rand_2	96.75	89.55	76.03	92.91
MRS_rand_3	94.06	88.56	71.36	91.14
MRS_rand_4	97.93	89.3	78.58	92.62
MRS_rand_5	96.67	89.74	76.70	92.31
MRS_rand_6	96.95	89.34	74.02	92.28
MRS_rand_7	96.85	89.26	78.82	92.59
MRS_rand_8	98.56	91.11	82.00	94.44
MRS_rand_9	97.63	91.11	80.25	94.44
MRS_rand Average	96.90	89.87	78.25	93.16
Real_sample	100.00	97.52	100.00	99.17

Table C.4. Results obtained for the Targets Completed metric (mean) with the LT-MPT based on DSKO and the OPO approaches in the defined scenarios.

C.3.4 Computational Cost

This point presents the computational cost of the OPO approaches executed in a CPU Intel® Core™2 Duo Processor E6600 2.40 GHz with 6GB of RAM. Table C.5 summarizes the duration of each one of the approaches in the sample realization and real sample scenarios.

As mentioned in Section 6.2.5, the OPO-GAE approach provides the planning results of one trial in approximately 45 minutes for the sample realization scenarios and in 15 minutes for the real sample scenario. The OPO-GAO is less time consuming because it is not necessary to translate the new individuals to a plan and to repair them. Specifically, the planning results of one trial are provided in approximately 15 minutes for the sample realization scenarios and 5 minutes for the real sample scenario. In short, OPO-GAO is three times faster than OPO-GAE. However, this is not an important issue because there are no time limitations for obtaining the mission plan in the EChO mission.

The OPO-MGS and the OPO-MTS have a high computational cost due to the fact that the NSGA-II algorithm requires a high number of computations. The time needed to obtain the results in the sample realization scenarios is around 4 hours and approximately 50 minutes in the real sample scenario.

	Sample Realization	Real Sample
OPO-GAE	45 minutes	15 minutes
OPO-GAO	15 minutes	5 minutes
OPO-MGS / OPO-MTS	240 minutes	50 minutes

Table C.5. Execution time in minutes of the LT-MPT according to each one of the analysed OPO approaches.

C.3.5 Pareto Optimal Front of the Multiobjective Approach

Figure C.3 and Figure C.4 depict the Pareto optimal front obtained for OPO-MGS and OPO-MTS respectively. Note that the optimal front only stores non-dominated solutions, so there is not a solution better than another one for all the objectives. These figures symbolize each solution with a bullet, and each one of the objectives is represented in each axis. Thus, the value for each one of the objectives can be identified for each solution. Both figures show the solutions obtained at the end of the MOEA cycle, so incomplete targets are not removed. If we analyse the shape of the Pareto optimal fronts, it can be observed that the minimization of the time that the telescope is not observing (F_G, defined in Equation 5.4) or the minimization of the number of incomplete targets weighted with the priority of the targets (F_T, defined in Equation 5.5) increase the time that the telescope is doing slewing (F_S, defined in Equation C.1), and vice versa. Therefore, returning a solution that minimizes the slew time (i.e., minimum value of F_S) means that the obtained solution is the one with worse scientific return. Finally, the strategy for retrieving the most suitable solution from the optimal front according to the maximization of the planning efficiency or the scientific return tends to obtain a LTMP with a high slew time.

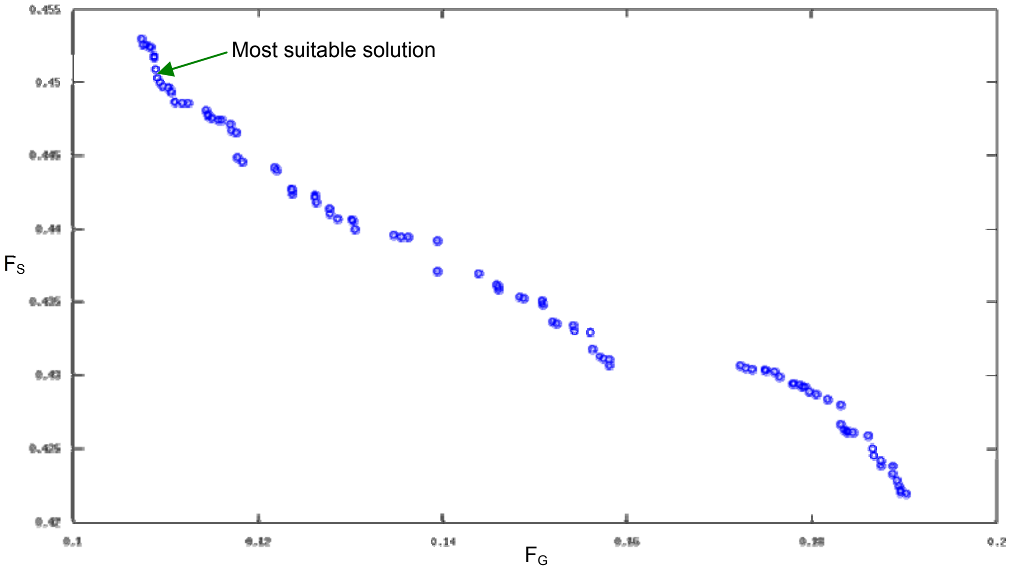

Figure C.3. Pareto optimal front of one sample realization (MRS_rand_8) obtained by OPO-MGS.

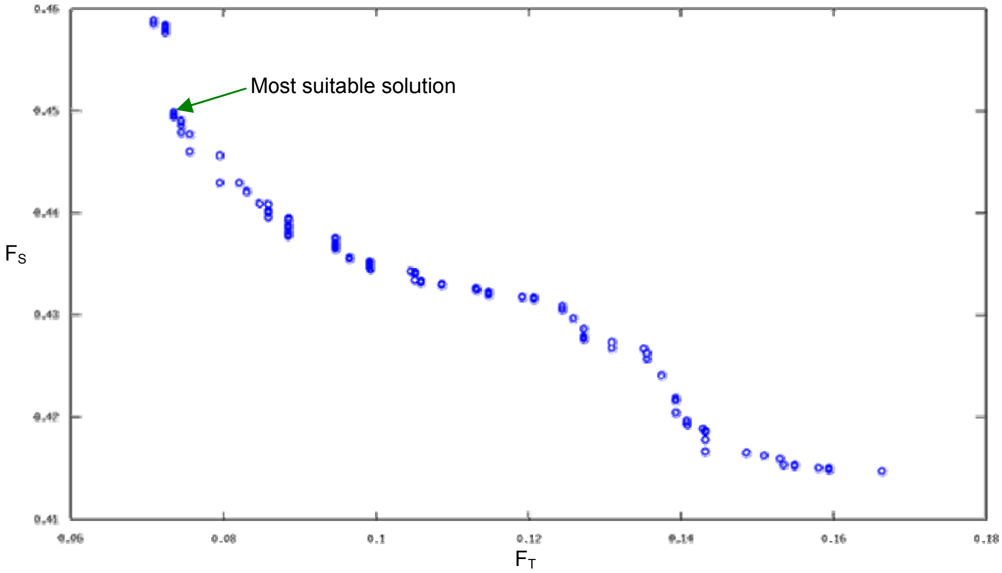

Figure C.4. Pareto optimal front of one sample realization (MRS_rand_8) obtained by OPO-MTS.

C.3.6 Comparison and Discussion

If we globally compare the results of OPO-GAE and OPO-GAO, we can observe that the main difference in terms of the performance of the obtained plans is that OPO-GAO plans a small number of events. This means that OPO-GAO completes the targets by carrying out a smaller number of observations. This aspect has a direct impact on reducing the planning efficiency metric and, obviously, the slew time because a smaller number of configuration transfers is needed. Thus, in terms of planning accuracy (i.e., planning efficiency and targets completed), OPO-GAO is significantly worse than OPO-GAE. For this reason, due to the fact that the aim of the LT-MPT in the EChO mission is to maximize the scientific return of the obtained plan, it seems that OPO-GAO is not useful at this time. However, this OPO-GAO characteristic can be relevant in other problems.

On the other hand, if we compare OPO-GAE with OPO-MGS and OPO-MTS we can see that to return a solution with a trade-off between the slew time and the planning efficiency or the number of targets completed is not a useful method in the EChO mission, where the goal is to maximize the scientific return and the planning efficiency of the LTMP. Nevertheless, this multiobjective approach can be interesting in other kind of missions where a trade-off between conflictive measures is a challenge.

Finally, according to all these aspects, we can conclude that the OPO-GAE approach is the best method for planning the observations in the EChO mission.

REFERENCES

Bacardit, J. (2004). *Pittsburgh Genetic-Based Machine Learning in the Data Mining era: Representations, generalization, and run-time*. PhD Thesis, La Salle, Universitat Ramon Llull.

Brownlee, J. (2011). *Clever Algorithms: Nature-Inspired Programming Recipes*. LuLu Press.

Cantu-Paz, E. (2000). *Efficient and Accurate Parallel Genetic Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA.

Chen, S., and Smith, S. F. (1999). Putting the Genetics back into Genetic Algorithms: Reconsidering the Role of Crossover in Hybrid Operators. In *Proceedings of the Fifth Workshop on Foundations of Genetic Algorithms*, pages102–115.

Coello, C. A. (1999). A comprehensive survey of evolutionary-based multiobjective optimization techniques. *Knowledge and Information Systems*, 1:269–308.

Coello, C. A. (2001). *A Short Tutorial on Evolutionary Multiobjective Optimization*.

Coello, C. A., Lamont, G. B., and Veldhuizen, D. A. V. (2007). *Evolutionary algorithms for solving multi-objective problems*. Springer-Verlag New York, Inc.

Cordón, O., Herrera, F., Hoffmann, F., and Magdalena, L. (2001). *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*, volume 19 of *Advances in Fuzzy Systems–Applications and Theory*. World Scientific Publishing Co. Pte. Ltd., Singapore.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182 –197.

Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Garcia-Piquer, A. (2012). *Facing-up Challenges of Multiobjective Clustering Based on Evolutionary Algorithms: Representations, Scalability and Retrieval Solutions*, PhD Thesis, La Salle, Universitat Ramon Llull.

Goldberg, D. E. (1989). *Genetic Algorithm in Search, Optimization, and Machine Learning*. Addison-Wesley, Inc., Boston, MA, USA.

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, Cambridge, MA, USA.

Mendel, G. (1865). Experiments in Plant Hybridization. In *Natural History Society of Brunn in Bohemia*.

Osyczka, A. (1985). Multicriteria Optimization for Engineering Design. In Gero, J. S., editor, *Design Optimization*, pages 193–227. Academic Press.

Pareto, V. (1896). *Course d'Économie Politique*. F. Pichou, Lausanne and Paris.

Rosenberg, R. (1967). Simulation of genetic populations with biochemical properties. *Dissertation Abstracts International*, 28(7):67–17.

Schaffer, J. D. (1985). Multiple objective optimization with vector evaluated genetic algorithms. In *Proceedings of the 1st International Conference on Genetic Algorithms*, pages 93–100, Hillsdale, NJ, USA. L. Erlbaum Associates Inc.

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary computation*, 8(2):173–195.