

# Structured-Chromosome GA Optimisation for Satellite Tracking

Lorenzo Gentile  
Institute for Data Science,  
Engineering and Analytics, TH Köln  
Köln, Germany  
lorenzo.gentile@th-koeln.de

Cristian Greco  
Aerospace Centre of Excellence,  
University of Strathclyde  
Glasgow, United Kingdom  
c.greco@strath.ac.uk

Edmondo Minisci  
Aerospace Centre of Excellence,  
University of Strathclyde  
Glasgow, United Kingdom  
edmondo.minisci@strath.ac.uk

Thomas Bartz-Beielstein  
Institute for Data Science,  
Engineering and Analytics, TH Köln  
Köln, Germany  
thomas.bartz-beielstein@th-koeln.de

Massimiliano Vasile  
Aerospace Centre of Excellence,  
University of Strathclyde  
Glasgow, United Kingdom  
massimiliano.vasile@strath.ac.uk

## ABSTRACT

This paper presents a novel optimisation approach, called Structured-Chromosome Genetic Algorithm (SCGA), that addresses the issue of handling variable-size design space optimisation problems. This is based on variants of standard genetic operators able to handle structured search spaces. The potential of the presented methodology is shown by solving the problem of defining observation campaigns for tracking space objects from a network of tracking stations. The presented approach aims at supporting the space sector in response to the constantly increasing population size in the around-Earth environment. The test case consists in finding the observation scheduling that minimises the uncertainty in the final state estimation of a very low Earth satellite operating in a highly perturbed dynamical environment. This is evaluated by coupling the optimiser with an estimation routine based on a sequential filtering approach that estimates the satellite state distribution conditional on received indirect measurements. The solutions found by employing SCGA are finally compared to the ones achieved using more traditional approaches. Namely, the problem has been reformulated to be faced using standard Genetic Algorithm and another variable-size optimiser, the "Hidden-genes" Genetic Algorithm variant.

## CCS CONCEPTS

• **Mathematics of computing** → **Mixed-discrete optimisation**;  
• **Physical sciences and engineering** → *Aerospace*; • **Information systems** → Uncertainty; • **Computing methodology** → Modelling and simulation;

## KEYWORDS

Optimisation, Genetic Algorithms, Structured-chromosome, Satellite tracking

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic*

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6748-6/19/07...\$15.00

<https://doi.org/10.1145/3319619.3326841>

## ACM Reference Format:

Lorenzo Gentile, Cristian Greco, Edmondo Minisci, Thomas Bartz-Beielstein, and Massimiliano Vasile. 2019. Structured-Chromosome GA Optimisation for Satellite Tracking. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion)*, July 13–17, 2019, Prague, Czech Republic. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3319619.3326841>

## 1 INTRODUCTION

Tracking objects in space has become a major issue for the aerospace community and the general public. Operational satellites require accurate knowledge of their position and velocity to deliver increasingly precise services and solid scientific outputs. Non-collaborative objects tracking is needed for collision avoidance and re-entry prediction, events which pose a threat respectively to the around-Earth and the terrestrial environment. The continuous growth of the population of artificial objects and small debris orbiting the Earth requires an adequate response in tracking capabilities, scheduling, and processing routines. In particular, this necessity applies to the tracking of non-collaborative objects or small missions which cannot rely on dedicated extensive tracking networks.

Therefore, this research develops a scheduling approach to compute optimal observation campaigns for a generic object in space. The goal is to improve the knowledge of the object's state at the end of the tracking window while respecting an allocated budget. As an alternative, the developed approach can also reduce the resources to achieve an accuracy requirement.

In this problem the scheduling can be interpreted as the management of the following free variables: 1) the number of times a specific ground station is used, 2) in which observation's window this has to measure and 3) the number of measurements to perform. Consequently, without simplifying assumptions, the architecture and the number of design variables can vary among the solutions, and the observation scheduling optimisation falls under the area of the variable-size mixed-discrete global optimisation. To deal with dynamically varying search spaces, a number of additional challenges harden dramatically the complexity of the search algorithm. Genetic Algorithms (GAs) can overcome most of these issues by means of an appropriated encoding. As stated in [9], GAs is one of the most suited classes of optimisers for variable-size optimisation.

A variety of GAs variants for facing variable-sized global optimisation can be found in the literature, mostly employed for space trajectory design [2, 7, 8]. Of particular interest is the "Hidden-genes" GA (HGGA) adaptation introduced in [2]. In this algorithm, each candidate is represented using all the possible genes and additional "activation genes" indicating which genes have to be considered or when computing the objective and constraint functions. The main drawbacks are that the maximum dimensionality of the problem has to be known a priori and that this is further enlarged by the "activation genes". A more complex, but effective adaptation of GA is proposed in [7, 8]. In these cases, a hierarchical multi-level chromosome structure is adopted in place of the standard "string" one. Defining the concepts of vicinity and hierarchy of genes, the authors introduced a problem formulation that enhanced the exchange of information between chromosomes as well as the computational efficiency in respect to the HGGA. Moreover, this hierarchical formulation can produce a more meaningful exchange of information between chromosomes.

In light of these considerations, we decided to make use of a Genetic Algorithm variant based on hierarchical search spaces and apply it to the problem of generating optimal observation scheduling.

The structure of the paper is the following. Sec. 2 first introduces the general tracking problem and the scheduling formulation, then it presents the specific problem scenario and discusses the cost function. Sec. 3, describes the employed algorithms focusing in particular on the SCGA. The analyses of results are presented in Sec. 4. Finally, final considerations and outlooks of this research are given in Sec. 5.

## 2 MODEL

This section introduces the model for the scheduling of observations campaigns of space objects from terrestrial ground stations.

First, the generic tracking problem is introduced in Sec. 2.1 in a state estimation probabilistic formulation. Then, the tracking scenario of space objects from ground stations is modelled in Sec. 2.2. The selection of a performance index for quantifying the estimation uncertainty follows in Sec. 2.3. Finally, the specific problem scenario to be used as reference test case is presented in Sec. 2.4.

### 2.1 Tracking

The system under consideration evolves according to deterministic ordinary differential equations  $f$  (ODE) describing its state evolution  $\mathbf{x}(t)$  in time

$$\begin{cases} \dot{\mathbf{x}} = f(t, \mathbf{x}) & (1) \\ \mathbf{x}(t_0) = \mathbf{x}_0 & (2) \end{cases},$$

where  $\mathbf{x}_0$  is the system initial state at the initial time  $t_0$ . The state at later times can be computed by classical numerical integration schemes for ODE.

In real-life scenarios, uncertainty always affects such systems and, therefore, the knowledge of the state at a later time  $\mathbf{x}_t = \mathbf{x}(t)$ . This uncertainty can result from a partial knowledge of the initial state, ambiguously specified model parameters, or unmodelled terms in the equations of motion. Therefore, measurements  $\mathbf{y}_t$  are employed to enhance the knowledge of the system state. Generally, it is not possible to observe the system state directly, and the

measurements are affected by noise. Hence, the state-observation relationship is modelled as

$$\mathbf{y}_t = h(t, \mathbf{x}_t) + \varepsilon, \quad (3)$$

where  $\varepsilon \sim \mathcal{N}(0; \sigma_y^2)$  is a Gaussian noise with standard deviation  $\sigma_y$ .

The addressed hidden dynamical process is affected by two sources of epistemic uncertainty: the initial state is only partially known, and its uncertainty is modelled with a normal density function  $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \sigma_{\mathbf{x}_0}^2)$ ; the received observation  $\mathbf{y}_t$  is noisy, and its uncertainty is described by a probability distribution  $p(\mathbf{y}_t | \mathbf{x}_t) = \mathcal{N}(h(t, \mathbf{x}_t); \sigma_y^2)$  conditional on the state value.

Given a sequence of observations  $\mathbf{y}_{1:l}$ , the tracking problem aims at combining the a priori information about the initial state with the available observations to compute an updated state distribution conditional on the observations  $p(\mathbf{x}_l | \mathbf{y}_{1:l})$ . In Bayesian filtering, this inference step is formulated by Bayes' rule [10]

$$p(\mathbf{x}_l | \mathbf{y}_{1:l}) = \frac{p(\mathbf{y}_l | \mathbf{x}_l) p(\mathbf{x}_l)}{p(\mathbf{y}_{1:l})}, \quad (4)$$

where the conditional independence of earlier measurements has been used to simplify the density  $p(\mathbf{y}_{1:l} | \mathbf{x}_l) = p(\mathbf{y}_l | \mathbf{x}_l)$ . The conditional distribution in Equation (4) is the complete solution of the tracking problem.

However, this update equation has no closed-form solution for general probability distributions. In this work, the inference step is solved using a Square-Root Unscented Kalman Filter (SR-UKF) [14], a sequential filtering approach suitable for normal distributions and non-linear dynamical and observational models. For this sequential scheme, the observations are processed as soon as they are available, and the updated posterior distribution propagated in future until the time of the next measurement. The unscented transform improves the estimation accuracy of standard Kalman Filters for an equal computational burden, whereas the square-root version has more robust numerical stability. The output conditional distribution from the SR-UKF is a multivariate normal distribution described by two parameters, i.e. mean and covariance matrix, as

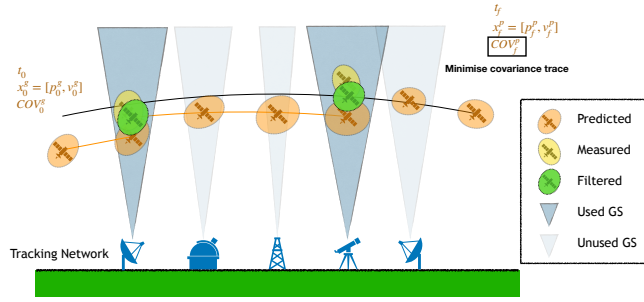
$$p(\mathbf{x}_l | \mathbf{y}_{1:l}) = \mathcal{N}(\hat{\mathbf{x}}_l; \sigma_{\mathbf{x}_l}^2) \quad (5)$$

The smaller are the elements of the covariance matrix  $\sigma_{\mathbf{x}_l}^2$  the more certain is the estimate of the system state.

The described tracking problem is depicted in Fig. 1 for a two-dimensional scenario, where only the first and fourth stations are used to take measurements of the satellite. The ellipses represent the satellite state and its uncertainty in time corresponding to the Gaussian distributions. The colours orange, yellow and green are used respectively for the predicted density, the measurement likelihood, and the filtered distribution.

### 2.2 Tracking Station Scheduling

It is intuitive to realise that two observation campaigns  $\mathbf{y}_{1:l}$ , differing for the employed sensors, times and number of observations, would generally result in different posterior distributions. As an example, two extreme possibilities are either to use no observation, or to measure the highest number of times possible, resulting respectively in an extremely uncertain scenario, or a very certain estimate.



**Figure 1: Graphical 2D representation of satellite single passage over network of ground stations. The dark faded blue field-of-view indicates that a station is used to take measurement, whereas the light faded blue indicates that the station is not operated. The ellipses symbolise a confidence region of the uncertain state of the satellite. The different colours orange, yellow and green are used to indicate respectively the predicted state, the measurement and the posterior distribution.**

**Table 1: General free variable  $\mathbf{u}_i$  scheme for observation optimisation.**

Description	Variable Type	Values
Use ground station $j$	Discrete	ON/OFF
Number of passages to use	Integer	$[0, p_{max}] \in \mathbb{N}$
Indexes of passages to use	Discrete	$[1, p_{max}] \in \mathbb{N}$
Number of observations per passage	Integer	$[0, Nobs_{max}] \in \mathbb{N}$
Times of observations per passage	Continuous	$t \in [T_{in}, T_{out}]_j^p$

In this paper, the challenge addressed is to find the observation campaign which produces the most certain estimate of the satellite final state  $\mathbf{x}_f$ , while respecting a budget constraint, given a network of variegated ground stations (GS). The budget constraint is introduced because, in real satellite operations, taking measurements from a tracking station has an associated expense, both in terms of human and monetary costs. This constraint prevents the aforementioned extreme case to be a feasible optimum. To further increase the model reality, each tracking station may be characterised by specific features depending on its geographical location, and characteristics of the instruments available in loco.

Hence, a station  $GS_j$  is uniquely specified by defining: its geodetic coordinates, namely latitude  $lat_j$  and longitude  $lon_j$ ; the measurements model  $h_j(t, \mathbf{x}_t)$ ; observation covariance  $\sigma_y^2$  at 90 deg, whereas at lower elevations the covariance elements are assumed to degrade as  $\sim 1/\sin(EL)$  [11], where  $EL$  is the object elevation angle over the local horizon at the observation time. Given such specifications, the object to be tracked may orbit in each station field-of-view (FOV) once, multiple times, or never, depending on the tracking window time-span, station-object relative geometry, and the object initial orbital parameters.

Clearly, it is not possible to pick an observation realisations  $\mathbf{y}_i$  directly, but there is an optimisable action  $\mathbf{u}_i$  which influences the measurement  $\mathbf{y}_i(\mathbf{u}_i)$ . In the most general case, the action  $\mathbf{u}_i$ , associated to the ground station  $GS_j$ , instructs the model whether to take a measurement from  $GS_j$  or not, how many, at which passage, and when in the passage the measurements are to be taken. The variables enclosed in an action  $\mathbf{u}_i$  are summarised in Tab. 1.

Hence, the free variables of the scheduling problem are the sequence of actions  $\mathbf{u}_{1:l}$  that reduces a performance index on the final state uncertainty, while respecting imposed budget levels.

### 2.3 Cost Function

Selecting a performance index which quantifies the accuracy of the filtered state estimate is a non-trivial task [11]. First, since the true state is unknown, a direct measure of the estimation error cannot be computed. Second, each scalar measure would only provide limited information out of the multi-dimensional posterior distribution characterising the state uncertainty.

The Kalman Filter solution returns the Gaussian posterior distribution in terms of mean and covariance matrix, which intuitively expresses how spread the density function is. In this work, the trace of the covariance is chosen as performance index to express how confident we are on each element of the state vector. Indeed, although the covariance matrix is usually an optimistic measure, for scheduling purposes only a relative indication of the accuracy between two observation campaigns is needed.

Hence, the scheduling problem can be formally stated as finding the optimal sequence of actions  $\mathbf{u}_{1:l}$  that reduces the uncertainty on the state at the end of the tracking window, conditional on the observations, as

$$\min_{\mathbf{u}_{1:l}} J = \text{Tr} \{ \sigma_{\mathbf{x}_f}^2 (\mathbf{x}_f | \mathbf{y}_{1:l}(\mathbf{u}_{1:l})) \}. \quad (6)$$

Graphically, this performance index corresponds to shrinking the final orange ellipse in Fig. 1 as much as possible. Being the covariance diagonal composed only of positive terms, the theoretical minimum is zero, i.e. the case of perfect knowledge of the satellite state.

### 2.4 Reference Scenario

The scenario investigated in this research is characterised by a network of nine ground stations characterised by different geodetic coordinates. Each ground station can take measurements composed of three scalar quantities, namely the relative satellite range, azimuth, and elevation. The observation covariance at  $EL = 90\text{deg}$  is set as three-dimensional diagonal with elements equal to  $1.0e^{-5}$ .

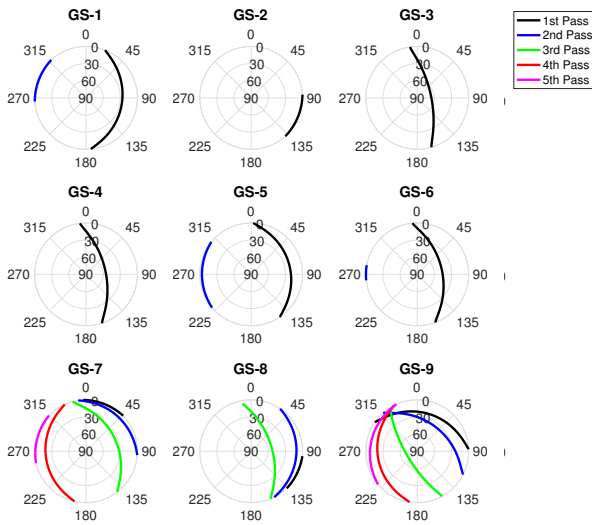
The satellite to track is in very low-Earth orbit, and the observation window of interest spans from the 2018 October 29 12:00 UTC to the 2018 October 29 20:00 UTC. The initial state estimate in Keplerian elements is

$$\begin{aligned} \mathbf{x}_0 &= (a[\text{km}], e[-], i[\text{rad}], \Omega[\text{rad}], \omega[\text{rad}], \theta[\text{rad}]) \\ &= (6608.17, 1.61e^{-3}, 1.685, 5.662, 1.199, 1.589). \end{aligned}$$

After conversion to Cartesian coordinates [13], this estimate is the mean of the Gaussian initial state distribution with covariance set as

$$\begin{aligned} \sigma_{\mathbf{x}_0}^2 &= \text{diag}(1.0e^{-2} \text{ km}, 1.0e^{-2} \text{ km}, 1.0e^{-2} \text{ km}, \\ &1.0e^{-4} \text{ km/s}, 1.0e^{-4} \text{ km/s}, 1.0e^{-4} \text{ km/s}) \end{aligned}$$

The dynamical model determining the satellite evolution in a low Earth orbit has several force contributions [6]: the gravitational force modelled employs the EGM96 geopotential model up to degree and order 10; atmospheric drag according to Jacchia-Gill model; third-body disturbances due to the Moon and Sun gravitational pull; solar radiation pressure (SRP) with a conical shadow model. The satellite drag cross-section is set to  $15.0 \text{ m}^2$ . The mean solar flux



**Figure 2: Sky plots of satellite passes over ground stations; the circles correspond to different elevation levels, while the angular quantity indicates the azimuth measured eastwards from the local north. Different colours indicate different satellite passes over the same station.**

considered is 106.4 in solar flux units, with a mean SRP cross-section of  $1.625 \text{ m}^2$  and a SRP coefficient of 1.3.

Given this scenario, the geometry of this satellite passes above the considered network of tracking stations is visualised in the sky plots shown in Fig. 2, which shows the azimuth and elevation of the satellite when it is over the stations' local horizon. Hence, the station-satellite relative geometry and the extension of the tracking window determine a number of possible observations windows equal to 21. The best passage over one station's local horizon is registered for the tracking station GS-9 with an elevation of  $\approx 71 \text{ deg}$ , which furthermore sees the satellite in five different passes. Stations GS-3, GS-4 and GS-8 see the satellite with decent elevation angles  $\geq 50 \text{ deg}$ , while the worst passages are realised over stations GS-2 and GS-5. Therefore, it can be predicted that an optimal schedule would take measurements from stations which see the satellite at higher elevation angles, as in the model the observation covariance degrades when the elevation decreases.

Generally, the more observations are used the more reliable the computed estimate is, hence the lower the value of the cost function will be. As discussed, this new information comes at a cost in a real-life scenario. For this reason, the maximum number of measurements will be imposed as a constraint and varied over the different runs. In order to quantitatively assess the impact of the maximum available number of the observations over the uncertainty in satellite tracking, the analysed optimisation strategies have been repeated varying this in a range from 15 to 65 with the span of 5. These values are representative of an operative mission scenario that would benefit more from the employment of the proposed algorithm.

### 3 OPTIMISATION PROBLEM

This section introduces the methodologies adopted for minimising the uncertainty associated with the state of an object in space varying the scheduling of the observation campaign. The optimisation strategies, as well as the problem formulations, will be described and compared in detail. In this research, the proposed Structured-Chromosome Genetic Algorithm (SCGA) is compared to the standard Genetic Algorithm and an implementation of the "Hidden-genes" Genetic Algorithm (HGGA). All of them have been used for minimising the uncertainty associated with the state of the satellite at the end of the tracking window.

The optimal resources allocation for the object's tracking is the common aim of all the optimisers employed but, as will be shown in the following sections, the problem formulations are different. In particular, a throughout explanation of the proposed optimiser and its problem formulation is given in section 3.2 and the others are summarised in section 3.1.

#### 3.1 Standard GA and "Hidden-genes" GA

The performances of the proposed optimiser have been compared to the ones of a standard Genetic Algorithm [5] and the HGGA variant proposed in [2]. As well known, the standard GA is not suitable for handling variable-sized optimisation problems. Therefore, applying this strategy for facing the presented problem requires making assumptions on the maximum dimensionality. In this particular problem, this entails assuming that the GSs are always used. If that, the quantity to be managed by the optimiser is limited to the budget for measurements at each satellite pass over the GSs. This assumption, in theory, should not tragically compromise the strategy's capability of optimally allocating the budget but, as the results will show, extremely harden the task.

The HGGA makes use of activation genes for activating or deactivating genes. As in the previous case, it is assumed that the GSs are used at every pass so the dimensionality of the problem is the greatest possible. Then, at every gene an "activation" one is associated. Its value indicates whether to consider or not the associated gene in the decision space. By means this trick, it is possible to encode and manage solutions with different architectures using standard genetic operators.

In the scenario considered, the maximum number of decision variables is 21, one for each time the satellites falls in the FOV of one of the GS composing the tracking network. In the standard GA problem formulation, these constitute the complete decision space whereas, in the HGGA one, the 21 additional activation genes are introduced.

Two different implementations of the HGGA have been tested. In one the crossover operator performs a one-point crossover acting at one on the overall chromosome (original+activation genes) in the second, this is applied separately on the decision and activation genes as proposed in [2].

Both standard GA and the HGGA variants have been implemented using the GA R package [12] and using the default options but for the selection operator. For consistency, a selection operator identical to the one used in the SCGA has been used.

### 3.2 Structured-Chromosome Genetic Algorithm

**3.2.1 Problem Formulation.** The adopted problem formulation aims at condensing the information in the free variables (see Tab. 1) making use of the concept of hierarchy for enhancing the operators capability of reaching good performing solutions. One of the main advantages of this problem formulation is that it is not required assuming a priori the number of ground stations to use and the number of passes in which these are activated. This implies that the characteristics of the candidate observation campaign are contained in a flexible number of decision variables.

In this problem formulation, the search space has been structured in a hierarchical way imposing dependencies between genes. Consequently, the target of the operators is not a single gene but rather an entire gene structure. This structure is actively helping the operators to perform "meaningful" transformations over the candidates. In fact, it helps to preserve the overall information limiting the extrapolation of one gene from its context. Due to the elevated interactions, blindly exchanging genes might lead to a high information destruction rate.

Differently from classical genetic algorithms, in SCGA a chromosome (solution) is represented by a matrix either indicating the value of the gene either its position in the hierarchy of the chromosomes. The genes are grouped into *gene classes* that characterise them by data type, dependent genes, and bounds. For the observation scheduling optimisation, the hierarchical structure is stratified in three levels.

The gene class *Ground Station* forms the top of the hierarchical structure. The value of this gene indicates the number of different satellite passes in which the specific ground station will be employed to measure the satellite state. Thus, given that the modelled GS network is constituted by nine ground stations, nine genes belonging to this class are present in all the solutions, one for each GS. However, the limiting upper bound of these genes is determined by the referring GS. In fact, shown in Fig. 2, every GS has the satellite in its FOV a different number of times. So the genes belonging to the gene class *Ground Station* don't have all the same bounds.

The gene class *Orbit index* (*OI*) constitutes the second level of hierarchy. This variable specifies the index of the pass of the satellite in the FOV of the associated GS to allocate the budget for measurements specified by the subordinated gene class *Budget for Measurements* (*BfM*). Obviously, a one-to-one correspondence between the bounds assigned to *Ground Station* and *Orbit index* genes has to exist.

As mentioned, as a rule of thumb, the more the observations the more reliable the estimate should be. To simulate a real-life scenario, a cost associated to each observation has been introduced to simulate the restrictions of a real observation campaign. Consequently, a

budget limitation constraining the number of observations has been imposed and it has been varied into the independent runs in order to have a quantitative indication of its effect on the uncertainty in the final state estimation.

The *Budget for Measurements* gene class indicates the percentage of the total available budget allocated for a set of measurements performed by a given GS at a given pass of the satellite in its FOV. The relation between the decision variables in the optimisation *Budget for Measurements* and the value requested by the model *Number of Measurements* is expressed by the equation:

$$NoM = \left\lfloor \frac{BfM}{Cost} \right\rfloor \quad (7)$$

where *NoM* is the *Number of Measurements* and *BfM* the allocated *Budget for Measurements*. This indirect measure of the number of measurements has been preferred in the perspective of testing the proposed approach assigning different levels of reliability and sensibility to the measurements of every GS. Accordingly, the unitary cost of the measurements by different GSs will also differ. In this experiment, for the sake of simplicity, the same unitary cost has been assigned to each GS.

**3.2.2 The algorithm.** The adopted algorithm is a population-based heuristic optimiser that relies on two operators to pursue the search of the global optimum: the crossover and the mutation. These operators, nowadays established in stochastic fixed-length mixed-discrete optimisation, have been redefined in order to manipulate candidates characterised by different length and structure. Then, these strategies are integrated into the classical structure of genetic algorithms [5]. In the following, a throughout explanation of the most relevant aspects of the proposed operators will be given.

**3.2.3 Initial population.** The optimiser has to deal with bounds and incompatibility constraints that considerably reduce the search space. In the test case, the number of times the object falls in the FOVs of the different GSs is different. This means that the upper bound imposed to the gene classes *Ground Station* and *Orbit index* cannot be one value but rather a set of values. Furthermore, another problem is that the values of the genes belonging to the gene class *Orbit index* and dependent on the same gene cannot assume the same value. An additional limitation is imposed on the gene class *Budget for measurements* for not exceeding the overall budget. Hence, creating feasible solutions is not a trivial problem. Therefore, an ad-hoc strategy for defining an initial feasible population has been created. The first step in the presented algorithm consists in the definition of the first set of candidates (population). In classical global optimisation, this often relies on techniques that aim at maximising the information-gain distributing samples in the search space following some strategy. However, none of them can be carelessly adopted in variable-size optimisation problems [9]. In the SCGA an iterative algorithm that creates feasible candidates has been developed. The creation starts from the gene at the top of the hierarchy and defines the value using random uniform sampling on the range of feasible values. Once a value has been assigned to a gene, the ranges of the possible values for all the dependent genes are recomputed in order to guarantee their feasibility. Nevertheless, a population size equals to 30 has been chosen for this experiment.

**Table 2: Decision variable of the Observation Scheduling optimisation Problem.**

Gene	Type	Lower Bound	Upper Bound
<i>Ground Station</i>	Integer	$0 \vee GS_1^9$	[2, 1, 1, 1, 2, 1, 5, 3, 5]
<i>Orbit index</i>	Discrete	$1 \vee GS_1^9$	[2, 1, 1, 1, 2, 1, 5, 3, 5]
<i>Budget</i>	Real	$0 \vee GS_1^9$	$1 \vee GS_1^9$

**3.2.4 Selection.** In stochastic optimisation the way candidates are selected is crucially important. This, greatly affects the velocity of convergence towards a region, determining the balance between exploration and exploitation of the search space. This step is strongly dependent on the method used for fitness assignment. The fitness shall reflect the "goodness" of a candidate orienting the selection to the most promising chromosomes. In the observation scheduling optimisation problem the objective function can assume a wide range of values that can differ in 7 order of magnitude or even be impossible to compute because of model divergence. In light of this, it has been decided to use ranking fitness assignment filtering the objective function differing less than 3 order of magnitude from the best of the current population. For selecting the chromosomes for crossover, the Tournament selection with tournament size equal to 6 has been used. In addition, the best 10% members of the population are preserved immutably. This can sometimes have a dramatic impact on performance by ensuring that the algorithm does not waste time re-discovering previously found solutions.

**3.2.5 Crossover.** The crossover is an operator that exchanges genes between two different chromosomes (parents) to produce two new candidates (children). This aims at combining and transferring the information contained in the parents to the children. In such a way, hopefully, the children will contain the relevant characteristics that originated the performance of their parents.

In classical fixed-size algorithms, all the genes lie on the same level and have a well-defined position and meaning. Genes in the same positions in the strings of two different chromosomes represent the same variable. This is not the case for structured chromosomes. Here, swapping genes among parent chromosomes on the basis of their position may result in selecting genes that represent different variables and creating unfeasible and meaningless solutions. Hence, a different strategy for selecting genes to swap based on their class has been proposed by [7].

The crossover operation is permitted only on genes belonging to the same class. This approach guarantees a semantically correct crossover where the picked gene, as well as the dependent genes, are swapped. Furthermore, the adopted crossover operator aims at maximising the information exchange per crossover operation. However, in the preliminary stage of this research, the strategy proposed in [7] has been used but appeared to be too destructive, making the information contained in the parents disappear over the generations. For this reason, a revised one has been developed. In this, the number of exchanging genes belonging to each class is computed in regards to the structure of the two parents chromosomes. This helps to homogenise the crossover operation all over the hierarchy of the chromosome. Moreover, the already swapped genes are removed from the list of eligible genes for crossover. This helps to prevent the repetition of the crossover operation on the same genes that would reduce the exchange of information. It is worth to mention that, although the crossover on genes related to the same GS and *OI* is encouraged, the probability of exchanging of information between uncorrelated genes is preserved. The feasibility then is not guaranteed and the candidates are always checked and, if needed, repaired. The procedure adopted is then able to create meaningful children that respect the hierarchical structure of the parents. However, the respect of bounds or constraints is

not guaranteed and a step of repair is necessary for evaluating the response of the candidates.

**3.2.6 Mutation.** The mutation operator is characteristic of most of the population-based optimiser. Many different variants of this operator can be found in literature for standard fixed dimension optimisation [3]. As in the case of the crossover, a new mutation operator has been introduced to deal with structured chromosome. In this case, two main steps can be recognised in the mutation operation. First, the mutation of the chosen gene to be mutated and in particular its value, then, treating the context of the mutated gene. The former finds many similarity points with the standard mutation operators. Indeed, it aims essentially at perturbing the current value of the gene in order to introduce randomness in the chromosomes' evolution. In the problem of tracking observation campaigns determination, variables of different types have to be managed. Hence, the basics of the mutation operations described in [4] have been borrowed to be used as gene variable perturbation operators. The mutation of the offspring relies on operations acting differently on real, integer and discrete variables, all respecting the requirements for a mutation strategy in the search spaces: Accessibility, Feasibility, Symmetry, Similarity, Scalability, and Maximal Entropy [4]. The operator achieves this by adding normally distributed noise to real-valued variables. For integer variables, the distribution is based on the difference of two geometrical distributions. Categorical variables are simply re-sampled (uniform randomly) with some probability. The probability and the magnitude of mutation are controlled by hyperparameter, the *strategy parameters*, that evolve during the optimisation in a self-adaptive fashion. To each candidate, a set of hyperparameter is associated. This is initialised as described in Alg. 1 and evolves undergoing the same crossover operations of the associated chromosome. However, the crossover operation for the mutation hyperparameter consists in simply averaging the hyperparameter sets associated to the chromosomes selected by the operator. Hence, the mutation strength itself is also governed by an evolutionary process. The philosophy behind self-adaptation is that the evolutionary process can solve two problems simultaneously: the determination of the best strategy variables, and the determination of the best object variables.

The second step of the mutation consists in treating the dependent genes. In particular, their value has to undergo the mutation operation and their number has to be consistent with the new value assumed by the higher-level gene. For better clarify the process, the pseudocode of the mutation is reported in Alg. 2.

In this case as well, the respect of all the constraints is not guaranteed. Then, the repair function is employed to guarantee feasible solutions.

**3.2.7 Repairing.** For ensuring candidates feasibility after the evolutionary operations, a repairing strategy has been introduced. This guarantees that all the candidates are semantically correct, that the genes' values lie in their feasibility range, and the budget constraint is still respected. Firstly all the dependent genes are checked. In the specific, it is secured that the number of genes belonging to the class *Orbit index* corresponds to the one indicated by the *Ground Station* gene value associated. If not, if the referring GS can potentially be used in all the passes indicated by the *Orbit index* genes, its value is modified in order to have a consistent chromosome. Else,



---



---

**input:** Chromosomes, LB, UB;

$\text{NGenesPerType} \leftarrow$  average number of type of the genes in the chromosomes;

$$\tau = \frac{1}{\sqrt{2 \cdot \text{NGenesPerType}}};$$

$$\rho = \frac{1}{\sqrt{2 \cdot \sqrt{\text{NGenesPerType}}}};$$

$$\sigma_{\text{continuous}} = (UB - LB) \times 0.1;$$

$$\sigma_{\text{integer}} = (UB - LB) \times 0.33;$$

$$\sigma_{\text{categorical}} = 0.1;$$


---



---

**Algorithm 1:** Initialisation of the mutation hyperparameters for Observation scheduling optimisation. The global learning rates are denoted by  $\tau$ , the local learning rate by  $\rho$  and step sizes by  $\sigma$ .

the *Orbit index* genes are removed to be maximum the number of times the satellites falls in the ground station's FOV and their values are modified in order to minimise the difference between the unfeasible and the feasible solution. The second operation aims at identifying and correcting the *Orbit index* genes that don't respect the semantic constraints. In example, whether more *Orbit index* genes refer to the same satellite's pass or whether an *Orbit index* gene refers to an unfeasible satellite's pass. Finally, the algorithm checks whether the budget limit is exceeded. If this happens, all the *BfM* are scaled to be have an overall used budget equal to the maximum allowed.

## 4 RESULTS

The developed tool is tested in a quasi-realistic scenario for the tracking of a satellite in a very low Earth orbit (LEO), where the dynamics is highly nonlinear, as presented in Sec. 2.4.

Three optimisation strategies have been tested on the satellite tracking observation campaigns problem: the standard GA, the HGGA with the one-point and two-points crossover, the SCGA with adaptive mutation, and the SCGA with mutation hyperparameters fixed to the initial ones. Specifically, these algorithms have been tested varying the maximum number of possible measurements in a range from 15 to 65 with a span of 5. For each configuration 5 independent runs have been run in order to have more statistically significant results.

All the optimisation runs have been limited to 500 generations, his results in 13503 function evaluations for the SCGA, and 15000 for the others algorithms.

The results in all the problem configurations are summarised in Fig. 3 where the covariance trace of the final optimal solutions found in the different runs is depicted by means box plots. For all the budgets, the GA is significantly outperformed by the others. This straightens the hypothesis that traditional fixed-size optimisation strategies actually struggle to face variable-size problems, even when they have been reformulated. To compare the other algorithms, the Conover's rank-based pairwise multiple-comparison [1] tests have been used. These tests indicate that there is no significant difference between all the other algorithms' performance. For further considerations, a visualisation of the best-found solutions in the configuration with the maximum number of measurements = 25 has been reported in Fig. 4. From this Figure, it can be seen that

---



---

**input:** geneToMutate,  $\sigma$ ,  $\tau$ ,  $\rho$ , LB, UB;

$\text{genesPresence} \leftarrow \text{lengths}(\text{genesPerClass});$

**for**  $i \in \text{geneToMutate}$  **do**

$\text{geneClass} \leftarrow \text{geneClass of } i;$

**if** *gene type of i is real* **then**

$\text{geneValue} \leftarrow$

$\text{geneValue} + \sigma_{\text{geneClass}} \cdot N(0, 1) \cdot (UB - LB);$

$\text{geneValue} \leftarrow \max(\text{geneValue}, LB_{\text{geneClass}});$

$\text{geneValue} \leftarrow \min(\text{geneValue}, UB_{\text{geneClass}});$

**end**

**if** *gene type of i is integer* **then**

$\text{NGene} \leftarrow \text{genePresence}_{\text{geneClass}};$

$p \leftarrow \frac{\sigma}{\text{NGene} \cdot \sigma_{\text{geneClass}}};$

$p \leftarrow \frac{1-p}{1+\sqrt{1+p^2}};$

$p \leftarrow \max(p, 1e-3);$

$G1 \leftarrow \text{rgeom}(\text{NGene}, \text{prob} = p);$

$G2 \leftarrow \text{rgeom}(\text{NGene}, \text{prob} = p);$

$\text{geneValue} \leftarrow \text{geneValue} + (G1 - G2) \cdot (UB_i - LB_i);$

$\text{geneValue} \leftarrow \max(\text{geneValue}, LB_i);$

$\text{geneValue} \leftarrow \min(\text{geneValue}, UB_i);$

**end**

**if** *gene type of i is categorical* **then**

$\text{Activation} = N(0, 1) > \sigma_{\text{geneClass}};$

**if** *Activation* **then**

$\text{geneValue} \leftarrow$  pick one random possible feasible value

**end**

**end**

**if** *i has dependent genes* **then**

**if** *number of dependent genes has to be changed* **then**

**if** *genes have to be added* **then**

                initialise new dependent genes

**end**

**if** *genes have to be removed* **then**

                randomly pick dependent genes to be removed

**end**

**end**

        Mutate all the dependent genes

**end**

**end**

**for**  $i \in \text{geneClass}$  **do**

$\text{NGene} \leftarrow \text{genePresence}_i;$

**if** *gene type of i is real or integer* **then**

$\sigma_i \leftarrow \max(1, \sigma_i \cdot e^{\tau_i * N(0,1) + \rho_i * N(0,1)});$

**end**

**if** *gene type of i is categorical* **then**

$\sigma_i \leftarrow \frac{1}{1 + \left(\frac{1-\sigma_i}{\sigma_i}\right) \cdot e^{\tau_i * N(0,1) + \rho_i * N(0,1)}};$

**end**

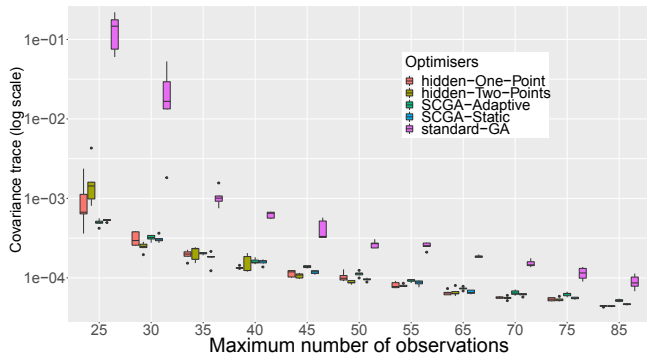
**end**

---

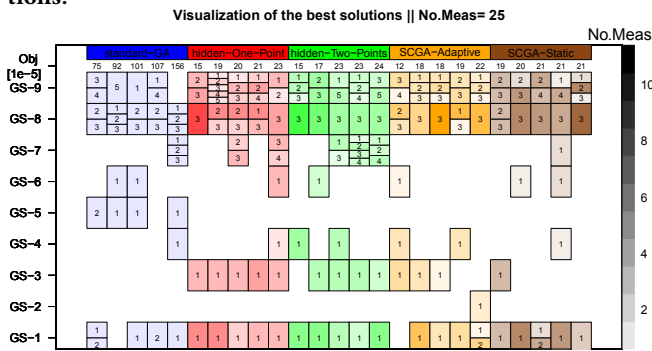


---

**Algorithm 2:** Algorithm summarising the main steps of the mutation operator. The notation  $N(0, 1)$  denotes a random number generated from a normal distribution centred in 0 having a standard deviation = 1.



**Figure 3: Box-plots of representing the best results found by all the algorithms in all the analysed problem's configurations.**



**Figure 4: Visualisation of the best solutions found with the maximum number of measurements=25. On the vertical axis, the  $J$  value and the Orbit indexes at which every GSs is used are reported. The colour shades represent the budget allocated to the specific observation: the darker the more measurements employed. The colour indicates the optimiser employed.**

the solutions found by the standard GA differ significantly from the ones found by the competitors: the standard GA does not make use of all the available budget, and only the observation campaigns found by the standard GA employ the GS-5 that is one of the two (with GS-2) with the lowest elevation angle.

The solutions found by the HGGGA and the SCGA are also interesting. Notably, one of the convenient stations for the early part of the tracking window, GS-1 and GS-3, are employed in almost all solutions whereas, for the late part, both the high elevation ones GS-8 and GS-9 are always employed. However, Fig. 4 shows great variability in the selected orbit indexes and the budget allocation, that reveals a high multimodality and complexity of the search space.

Finally, the impact of the maximum number of measurements can be addressed by comparing the uncertainty in the final state for the best solutions when varying this parameter. This has a strong influence on the configurations with small budgets, whereas it appears to vanish with the growth of the available budget.

The results indicate that only the variable-size optimisers are able to find and select the most important features while discarding the bad characteristics.

## 5 CONCLUSIONS

This paper presents a novel strategy, Structured-Chromosome Genetic Algorithm, that aims at tackling variable-size design space optimisation problems advantaging of the concept of genes hierarchy.

This algorithm has been applied for solving the problem of autonomously generating optimal satellite tracking observation campaigns under a limited budget. This concerns the aerospace community as nowadays the presence of objects orbiting the Earth is being pushed to the limit. Enhancing the efficiency of the observation schedules aims at making the tracking of space objects sustainable for the future. The presented scheduling formulation requires to optimise the sequence of observations in order to improve the knowledge of the satellite position and velocity at the end of the tracking window. Each observation results from a free action associated with each tracking station in the considered network. Such observations are processed using the SR-UKF to compute the posterior distribution of the state uncertainty, by efficiently merging different sources of information, namely the prior knowledge, the dynamics, and the measurements. The trace of the associated covariance matrix has been selected as performance index to synthetically measure the uncertainty associated with each component of the satellite state.

The proposed method has been tested on a quasi-realistic scenario in which nine ground stations were available to track a satellite in very low Earth orbit. For comparison, the standard GA and the "Hidden-gene" GA have been run on the same test case. The influence of the budget on the uncertainty in final state estimation as well on the searching algorithms' performance has been addressed by varying the number of available measurements.

The results indicate that the presented methodology can successfully enhance resources allocation strategies in space object tracking problems if compared to traditional approaches. However, the limited number of repetitions made, the complexity and the apparent deceptiveness of the objective function make further analyses absolutely needed for a definitive claim.

The SCGA has been implemented to work with any dynamic, measurement model and station network such that different test cases can be tested in future. The apparent parity of the SCGA and "Hidden-genes" GA performances can be partially due to the simplicity of the problem formulation. In the authors' opinion, introducing additional features such like a cost associated to the measurements dependent by the GS employed or differentiating the GSs by their accurateness could accent the SCGA's capacity of handling complex problem formulations. A further possible outlook of this research would be the definition of a similarity measure in order to embed this optimiser in a Surrogate model-based optimisation framework. Finally, a real-time approach that employs this observation schedule as first guess could be developed to improve the solution optimality in online applications.

## ACKNOWLEDGMENTS

This work was funded by the European Commission's H2020 programme, through the H2020-MSCA-ITN-2016 UTOPIAE Marie Curie Innovative Training Network grant agreement No.: 722734.



## REFERENCES

- [1] William Jay Conover and Ronald L Iman. 1979. On multiple-comparisons procedures. *Los Alamos Sci. Lab. Tech. Rep. LA-7677-MS* (1979), 1–14.
- [2] Shadi A Darani and Ossama Abdelkhalik. 2017. Space Trajectory Optimization Using Hidden Genes Genetic Algorithms. *Journal of Spacecraft and Rockets* 55, 3 (2017), 764–774.
- [3] Intiaz Korejo and Shengxiang Yang. 2012. Comparative Study of Adaptive Mutation Operators for Genetic Algorithms. In *in MIC 2009: The VIII Metaheuristics International Conference, 2009. 196 Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 2. Citeseer*.
- [4] Rui Li, Michael TM Emmerich, Jeroen Eggermont, Thomas Bäck, Martin Schütz, Jouke Dijkstra, and Johan HC Reiber. 2013. Mixed integer evolution strategies for parameter optimization. *Evolutionary computation* 21, 1 (2013), 29–64.
- [5] Melanie Mitchell. 1998. *An introduction to genetic algorithms*. MIT press.
- [6] Oliver Montenbruck and Eberhard Gill. 2012. *Satellite orbits: models, methods and applications*. Springer Science & Business Media.
- [7] Hui Meen Nyew, Ossama Abdelkhalik, and Nilufer Onder. 2012. Autonomous Interplanetary Trajectory Planning Using Structured-Chromosome Evolutionary Algorithms. In *AIAA/AAS Astrodynamics Specialist Conference*. 4522.
- [8] Hui Meen Nyew, Ossama Abdelkhalik, and Nilufer Onder. 2015. Structured-chromosome evolutionary algorithms for variable-size autonomous interplanetary trajectory planning optimization. *Journal of Aerospace Information Systems* 12, 3 (2015), 314–328.
- [9] Julien Pelamatti, Loïc Brevault, Mathieu Balesdent, El-Ghazali Talbi, and Yannick Guerin. 2017. How to deal with mixed-variable optimization problems: An overview of algorithms and formulations. In *World Congress of Structural and Multidisciplinary Optimisation*. Springer, 64–82.
- [10] S. Sarkka. 2013. *Bayesian Filtering and Smoothing* (1st ed.). Cambridge University Press, New York.
- [11] Bob Schutz, Byron Tapley, and George H Born. 2004. *Statistical orbit determination*. Elsevier.
- [12] Luca Scrucca et al. 2013. GA: a package for genetic algorithms in R. *Journal of Statistical Software* 53, 4 (2013), 1–37.
- [13] D. A. Vallado and W. D. McClain. 2007. *Fundamentals of Astrodynamics and Applications* (3rd ed.). Springer, New York.
- [14] Rudolph Van Der Merwe and Eric A Wan. 2001. The square-root unscented Kalman filter for state and parameter-estimation. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, Vol. 6. IEEE, 3461–3464.