Journal of Industrial and Management Optimization

https://www.aimspress.com/journal/jimo

JIMO, 22(7): 3362–3385.
DOI:10.3934/jimo.2026123
Received: 20 January 2026
Revised: 10 June 2026
Accepted: 11 June 2026
Published: 24 June 2026

Research article

Optimal spot-beam allocation by geostationary satellites

Gleb Beliakov[1,*], Juan Calle Salazar[1], Mark Jacobs[2], Jeewani Kodithuwakkuge[2] and Vicky Mak-Hau[1]

[1] School of Information Technology, Deakin University, Geelong, VIC 3125, Australia
[2] Defence Science and Technology Group, Adelaide, Australia

* **Correspondence:** Email: gleb.beliakov@deakin.edu.au; Tel: +61-39251-7475.

Abstract: We addressed the problem of optimally allocating of spot-beams by high-throughput satellites to the terminals on the ground (users), each with different transmission demands and user terminal characteristics. The geostationary satellite antenna can form spot-beams of different sizes, at different locations, and employ different bandwidths and polarisations (referred to as colours). The objective is to provide an optimal coverage of a geographical region so that the users' demands are satisfied, inter-beam interference is minimised, and beam transmission capacities are not exceeded. We formulated this challenging large-scale optimisation problem as a bilevel mixed integer programming problem and presented numerical experiments that illustrate the quality of the obtained solutions and scalability of the proposed approach for up to 200 spot-beams and 2000 terminals.

Keywords: high-throughput satellites; spot-beam placement; satellite coverage optimisation; integer programming; assignment problem
Mathematics Subject Classification: 65K05, 90B80, 90C11, 90C90, 90C30, 78A55

1. Introduction

High-Throughput Satellites (HTS) represent a paradigm shift in the satellite communication industry, offering significantly higher data transmission capacities than traditional satellite systems. By leveraging spot-beam technology, frequency reuse, and advanced modulation and coding techniques, HTS systems can achieve throughputs that are orders of magnitude greater than those of conventional satellites, enabling broadband services, high-speed internet, and seamless multimedia streaming even in remote and underserved regions of the world [1].

In commercial, government, emergency, and military applications, ground users may have diverse service demands and priorities, employ different terminals, and be distributed highly non-uniformly over a geographical region, although their locations are typically known. An HTS system can generate spot-beams at different locations, with varying footprints and bandwidths, and allocate the users to the most appropriate spot-beams. The aggregate traffic demand of the users allocated to a spot-beam should not exceed its overall transmission capacity, each beam footprint must be within the technical constraints of the satellite payload, and inter-beam interference should be minimised.

The optimal User-to-Beam Allocation and Colouring (UBAC) problem consists of determining an optimal allocation of users to spot-beams and an optimal beam colouring that satisfies the above-mentioned constraints. Beam colouring refers to optimisation of frequency reuse: each spot-beam is allocated a range of frequencies (and polarisations), collectively referred to as "colour"; spot-beams assigned the same colour but focused on different geographical regions can reuse the same frequencies without causing co-interference, whereas overlapping or nearby spot-beams must be assigned different colours. The term beam colouring derives from its similarity with the classical mathematical graph colouring problem, in which colours are assigned to the vertices of a graph so that adjacent vertices receive different colours while minimising the total number of colours used.

The UBAC problem constitutes a component of a broader system, which also involves HTS power distribution, carrier assignment, traffic scheduling, and prioritisation. Taken as a whole, such a system involves thousands of parameters and constraints, exhibits significant nonlinearity, and may not be practically solvable to optimality. In this contribution, we focus on the UBAC problem as one important component, which takes as its input users' positions and demands, as well as technical characteristics of the geostationary satellite and user terminals, and provides locally optimal assignment of users to spot-beams of different colours, as well as spot-beam locations and footprints. These outputs can then be used as inputs to subsequent resource allocation stages.

In this paper, we consider one-way communication from a satellite ground terminal (gateway) to user terminals, although practical satellite systems typically require bidirectional communication. The UBAC problem can be formulated and solved for user-to-gateway communication links in an analogous manner to optimise uplink communications. In this contribution, we focus on downlink spot-beam allocation and colouring.

It is assumed the geostationary satellite can generate N_B spot-beams steered in any direction in the coverage area, and shape them in a circular form with beamwidths in a given range (from 0.5 to 3 degrees, for instance). While in practice, phased array antennas are used for beam forming and steering, we assume parabolic antenna approximation, thus ignoring asymmetric distribution of the side lobes and focusing on the circular 3-dB beam footprints, which is consistent with the recent literature [2–4]. The interference arising from beam side lobes is addressed in a subsequent and separate power-allocation and link-budget optimisation stage.

We present formulations of the UBAC problem as a bilevel mixed integer linear program (MIP) in order to employ standard and well-developed commercial and open source MIP solvers such as Cplex, Gurobi, or GLPK, which often improve efficiency compared to alternatives (e.g., heuristics) [5–7].

The paper is structured as follows. After a literature review in Section 2, we present our system model and notation in Section 3. Problem formulation is presented in Section 4. In Section 5, we discuss variations of the beam allocation and colouring problem that incorporate some additional customer requirements. The numerical experiments are presented in Section 6, which are followed by conclusions.

2. Literature review

Pachler de la Osa [8] defined a framework to decompose the resource allocation problem in satellite communications into the following subproblems: power allocation, beam placement, beam shaping, user grouping, frequency allocation, satellite routing, and gateways routing. The two latter subproblems arise in non-stationary orbit satellites and are not of interest in this contribution.

In the framework defined in references [2, 8], the UBAC problem solves the beam placement, beam shaping, user grouping, and frequency allocation subproblems. The literature on the power and frequency allocation is rich. However, the literature on solving the beam placement, beam shaping, and user grouping is rather limited [9–11]. Recent works include [12–14].

In a review paper [4], the authors presented a comprehensive overview of multi-beam HTS from the perspective of their hardware, coding schemes, transmission requirements, beam-forming, and resource allocation. A rapidly growing body of publications is devoted to resource allocation, coverage, beam placement, scheduling, and beam hopping in non-geostationary orbit satellite systems (NGSO) [14–16]. The employed optimisation methods include evolutionary computing [17], AI/machine learning [18], simulated annealing [11], reinforcement learning [17, 19], and also mixed integer programming [20, 21] and heuristics [22, 23]. For recent works on geostationary satellites, the reader is referred to [3, 20, 24].

The beam placement and beam shaping subproblem defines the location of the spot-beams and their shape and size to satisfy the traffic demand of ground users. Traditionally, this problem has been solved by defining homogeneous shapes [8, 25, 26]. The authors of [3, 21, 24] also employed equal-size beam footprints steered to cover the desired region. However, recently heterogeneous shapes have been proposed to satisfy user demands better [9–11]. K-means clustering was used in [9] to define the corresponding Voronoi tessellation and to set up an ellipsoidal approximation to the Voronoi cells; we also refer to [21, 25] for a discussion of a similar approach. This is not applicable to scenarios where a user cluster cannot be served by a single spot-beam (it requires multiple overlapping beams of different colours). Capacitated beam placement [11] and a demand-driven beam densification approach [27] were adopted as models of uneven user demands. The possibility of multiple overlapping spot-beams makes the UBAC problem different from the traditional clustering problems, which aim to discover non-overlapping geometrical groups of points. Beam-forming is another technique applied recently for the same purpose [12, 28].

Camino et al. assumed a discrete set of circular shapes for the beams and proposed a mixed integer non-linear programming model that maximises the number of users covered while defining the placement, shape, and the user assignments [10]. It was shown that by irregularly placing beams across the same axis, the system throughput, measured in bits per second, can be increased in comparison to placing homogeneous beams evenly spread through the x-axis [29] . That work was formulated in the one-dimensional setting. The authors proposed an optimisation model that maximises throughput and solves it using numerical algorithms. Their control variables are beam centres, beam sizes, and the power allocated to each beam.

Various heuristic methods to optimise beam placements, width, colouring and transmission power were proposed by Mak-Hau et al. [22,23,30]. These authors proposed three different solution approaches. The first one integrates heuristics, integer programming, and constraint programming, and explicitly assigns users to beams, allocates colours to beams, and ensures that the beam transmission capacities are not exceeded. The second solution approach was to generate a number of capacity-feasible candidate beams and solve an integer programming model to select the best N beams from these candidate beams, whilst ensuring that no two neighbouring beams with the same colour are selected simultaneously. The third approach is again a combination of heuristics and integer programming. The authors further investigated the second method and developed a number of column generation-based approaches, and thoroughly compared all methods developed in both papers [30]. They also tested comprehensively the first method using different implementation scenarios to illustrate the benefit of using FlexBeamOpt as a simulation tool for analysing different satellite design and usage characteristics, e.g., the number of satellites, modulation and coding, the number of beams on each satellite, beam widths, clustered versus dispersed user distributions, and the number of users [22, 30].

The frequency allocation problem (FAP) is concerned with defining the bandwidth used by each beam, so that the neighbouring beams use different bandwidths to avoid interference. Within the satcom resource allocation literature, the beam-colouring problem has been solved as a subsequent problem after defining the beam placement and shapes [2, 4, 21]. However, as mentioned by Kyrgiazos et al. [31], this approach could lead to infeasible solutions, and the authors proposed an alternative way to solve the problem integrally. In this work, we assume significant beam overlaps due to highly heterogeneous user demands [11, 27]. This makes clustering-based techniques [9, 21, 25] not applicable.

In our work, the inter-beam interference is not a hard constraint [31], and there are technological alternatives to reduce it in the subsequent steps, for example, by properly allocating the carriers to the users, as well as allocating transmission power to the carriers. For details, we refer the reader to [2, 17, 18].

The main contributions of this paper are threefold. First, we formulate the UBAC problem as a bilevel optimisation problem, where we iterate between two steps: user-to-beam allocation and beam colouring (for fixed spot-beam locations) as the Leader problem, and calculation the optimal beams centres and footprints (for a fixed user allocation) as the Follower problem. A key feature of the proposed approach is the simultaneous beam-shaping, positioning, and colouring of highly overlapping beams. An important advantage of the bilevel program formulation is that, at each step, we have either an explicit solution (the second step), or a classical MIP problem formulation (to which a range of existing methods and software can be applied, and which significantly outperform most heuristics). Second, we conduct numerical studies of scalability of our formulation on a variety of realistic scenarios, including much larger instances than those considered in most existing studies. Third, we offer present several extensions of the basic model that incorporate more sophisticated beam placement requirements and operational constraints. Another contribution is the selection of initial spot-beam positions based on the users' locations, which affects the required number of iterations and the quality of the resulting allocation.

We stress that the UBAC problem and our model are very different compared to models used in many commercial settings [14, 18, 29], where the beam footprints are typically the same, and their positions follow a regular pattern, because of the approximate uniformity of the geographical distribution of user terminals. The heterogeneity of user demands, service priorities, and terminal capabilities, combined with the highly non-uniform spatial distribution of users, introduces additional challenges that have received limited attention in the existing literature.

3. System model and notation

We assume a geostationary satellite capable of forming N_B circular spot-beams in Ka diapason [4, 32–34]. In our setting N_B ranges from 50 to 200. The uplink signals are sent by a gateway. There are N_U users (terminals) on the ground, $N_U > N_B$, and more than one user can be illuminated by one beam, in which case they share the bandwidth, see Figure 1. We shall use $\mathcal{U}, \mathcal{B}$ to denote the sets of users and beams, respectively.

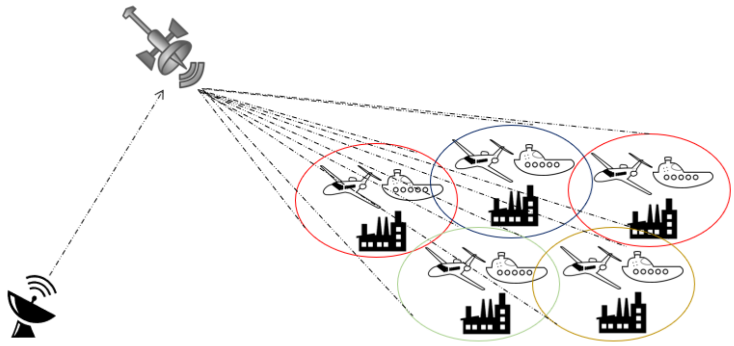

Figure 1. Multi-beam satellite illuminating different users by overlapping beams. The colouring problem intends to reduce interference by limiting overlap of same-coloured beams.

The total available bandwidth is divided into a number of colours—bandwidth ranges and polarisations, typically 4 or 8 colours N_Q. The objective of beam colouring is to reduce inter-beam interference when a user is illuminated by more than one spot-beam. It is assumed here that beams of different colours do not interfere.

The spot-beam position is characterised by its centre (x^c, y^c) and its size, the 3-dB diameter angle θ_{3dB}, which indicates where the transmitted power drops by 3 decibels relative to the beam centre. The users $u = 1, \ldots, N_u$ have coordinates (x_u^{usr}, y_u^{usr}) and transmission demands $\hat{R}_u$, x refers to the latitude and y to the difference between the user's and the satellite's longitudes (the geostationary satellite latitude is assumed to be 0).

We are interested in the downlink transmission rate whose bound can be calculated using Shannon's formula

$$R_u = A_u \log(1 + \Gamma_u), \tag{3.1}$$

where the signal to interference plus noise ratio (SINR) is

$$\Gamma_u(\mathbf{p}) = \frac{H_{ub} p_b}{\sum\limits_{b' \in \mathcal{B}, b' \neq b} H_{ub'} p_{b'} + C_u}, \tag{3.2}$$

the matrix H is computed from antenna gains and transmission losses as indicated below, p_b is the signal power applied to beam b, and C_u is the noise term, which depends on the user's terminal (established from the system temperature). By signal power p_b, we denote the total effect of various amplifiers, and in particular, the gains in the channeliser, which is the decision variable in power allocation (but not in the UBAC) problem.

Notice that Shannon's bound is a theoretical upper bound. In practice, the equipment implementations use a stepwise (piecewise-constant) function that relates the bandwidth efficiency to the SINR. Here the value A_u accounts for transmission efficiency and sharing bandwidth between the users illuminated by b. At this stage, we use the Shannon's formula following the literature. Simple modifications to Shannon's formula (via scaling) approximate the actual equipment transmission rates well.

The interference term in the denominator aggregates the signal powers received by u from the neighbouring beams b' of the same colour, whereas the numerator is the useful signal (carrier). The elements of the co-interference matrix H are calculated as $H_{ub} = g(\theta)g^R_{max}/L_{FS}$. All other losses (e.g., atmospheric, ionospheric) are treated as constants absorbed into L_{FS}. The receiving antenna gain is [1]

$$g^R_{max} = \left(\frac{\pi diam}{\lambda}\right)^2 eff^R, \tag{3.3}$$

where $diam$ is the diameter of the receiving (parabolic) antenna, eff^R is its efficiency, and λ is the wavelength $\lambda = c/f$. The maximum boresight transmitting gain is affected by the choice of spot-beam size θ_{3dB} [1] as

$$g^{\theta_{3dB}}_{max} = eff\pi^2 70^2/\theta^2_{3dB}. \tag{3.4}$$

When the user is not located at the beam centre, their respective gain is adjusted as [35, 36]

$$g(\theta) = g^{\theta_{3dB}}_{max} \frac{1}{C + \frac{1-C}{n+1}} \left(2C\frac{J_1(u)}{u} + \frac{1-C}{n+1} 2^{n+1}(n+1)! \frac{J_{n+1}(u)}{u^{n+1}}\right)^2, \tag{3.5}$$

where J_k is the Bessel function of the first kind of order k, n is the order of taper roll-off, C is the edge illumination (for tapered antennas), $u = \frac{M \sin\theta}{\sin\theta_{3dB}}$, and M is found numerically by solving $10 \log g^{\theta_{3dB}}_{max}/g(\theta) = 3$ dB. Some authors [36] use $n = 1$ and others use $n = 2$ and $C = -20dB$ (see [37] and subsequent works, e.g., [2, 12, 26]). The transmission losses $L_{FS} = L(4\pi R/\lambda)^2$, where R is the distance from the user to the satellite, and L are other constant losses, see [1].

The SINR is important when allocating power to each spot-beam in order to satisfy the transmission requests. We discuss this problem in a subsequent work, and we also refer to other recent works [2,14,17]. In this paper, we focus on minimising inter-beam interference governed by the entries of matrix H, which are inversely proportional to the distances between user u and the centres of the neigbouring beams of the same colour. Therefore our objective is to form the beams at such positions (and sizes) that the 3-dB footprints of spot-beams of the same colour overlap as little as possible. Thus our decision variables in this work will be the beam centres, their sizes θ^{3dB}_b, their colours col_b, and user-to-beam allocation binary variables z_{ub}.

Table 1. Nomenclature.

Notation	Definition
N_B	number of spot-beams
$\mathcal{B}$	set of spot-beam labels $b \in \{1, \ldots, N_B\} = \mathcal{B}$
$\mathcal{B}_{b1}$	conditioned set of beams $\{b2 \in \mathcal{B} : b2 > b1\}$
N_Q	number of colours
Q	set of colour labels $q \in \{1, \ldots, N_Q\} = Q$
N_U	number of users
$\mathcal{U}$	set of user labels $u \in \{1, \ldots, N_U\} = \mathcal{U}$
$\hat{R}_u$	requested data rate for user u
R_b	an estimate on the offered transmission capacity in beam b
	(given the number and bandwidth of subchannels and assuming spectral efficiency 1)
W^{max}	maximum beam width (measured in degrees, $\theta^{3dB} \leqslant W^{max}$)
W^{min}	minimum beam width (measured in degrees, $\theta^{3dB} \geqslant W^{min}$)
Δ^{sep}	separation factor
x_u^{usr}	location of user on x-axis (longitude)
y_u^{usr}	location of user on y-axis (latitude)
Variables	
$W_b = \theta_b^{3dB}$	size of beam b (diameter) in angular units
W_{b1b2}^{min}	$min(W_{b1}, W_{b2})$
θ_{ub}	the off-boresight angle of user u in beam b
θ_{b1b2}^c	the separation angle between beams $b1$ and $b2$
$z_{ub} \in \{0, 1\}$	1 if user u is allocated to beam b; otherwise, 0
$col_{bq} \in \{0, 1\}$	1 if colour q is used by beam b; otherwise, 0
$a_{b1b2q} \in \{0, 1\}$	1 if beam $b1$ and beam $b2$ are allocated colour q
$a_{b1b2}^{max} = \max_q a_{b1b2q}$	1 if beams b_1, b_2 have the same colour
$r_{ub} \in \mathbb{R}^+$	unserved data for user u in beam b when user u is allocated to beam b
$v_{b1b2} \in \mathbb{R}^+$	parameter indicating longitudinal intersection of beam $b1$ with beam $b2$
x_b^c	location of centre of beam b on x-axis (longitude)
y_b^c	location of centre of beam b on y-axis (latitude)

4. Problem formulation

Our objective now is to assign users to spot-beams in order to minimise unserved demand and inter-beam interference. This is a nonlinear and multiextremal problem in all control variables. However, with some variables kept fixed, it can be split into two simpler problems, a mixed integer linear problem and a quadratic convex problem that yields an explicit solution.

We formulate the overall problem as a bilevel optimisation program [5, 6, 38], in which the Leader problem uses the fixed optimal solution to the Follower problem. We solve

$$\text{minimise } w_1 UnmetDemand(z, r) + w_2 Compactness(z, x^c, y^c, \theta) + w_3 Overlap(\theta^c, col) \tag{4.1}$$

$$\text{s.t. } x^c, y^c, \theta \in \arg\min Steering(z), \tag{4.2}$$

$$OtherConstraints, \tag{4.3}$$

where the variables (in vector form) are listed in Table 1, and the objectives in the outer (Leader) problem refer to the total unmet demand, the sum of angular distances between beam centres and the positions of users allocated to that beam (Compactness), and the total degree of beam overlap (counting beams of the same colour). The objective in the inner (Follower) problem is optimal positioning and sizing the beams for a fixed user-to-beam allocation (vector z, provided by the Leader program).

Thus the outer problem has three competing objectives that are scalarised in a weighted sum [39], as it is frequently done in the satcom setting [16,20,26,40], and bilevel optimisation [41,42]. We will share more on this below. We note that the inner beam steering problem offers an explicit optimal solution we detail in the following. The constraints refer to various problem-specific semantic constraints (e.g., total beam capacity, each user allocated to a single beam) and technical constraints (e.g., each beam is allocated one colour only, overlap counts for two beams sharing the same colour, etc.). These constraints are made explicit in the allocation and colouring problem below.

The outer problem drives the overall process (see Algorithm 1) and aims at finding an optimal user-to-beam allocation and beam colouring. It supplies its current optimal allocation vector to the inner Follower problem, which provides the corresponding optimal beam centres and widths. Since the solution is explicit, one can recast the bilevel MIP as a single-level nonlinear integer problem (by introducing explicit functions $x^c = x^c(z), y^c = y^c(z)$ in $Compactness$), whose convergence in finite time is established via branch-and-bound techniques, see [5,38]. Next we look at the details of the inner and outer problems.

First, consider the inner problem in which user-to-beam allocation and colours are fixed. Then the optimal centre of a beam is the (weighted) centroid of the users allocated to that beam (as a solution of minimising the sum of squared distances problem). The size of the beam is then found as the smallest θ_b^{3dB} such that every allocated user is within the beam's footprint:

Beam Steering Problem

$$\theta_b^{3dB} = \max\{2\max_{u\in\mathcal{U}}\{\theta_{ub}z_{ub}\}, W^{min}\} \quad \forall b \in \mathcal{B}, \tag{4.4}$$

$$x_b^c = \frac{1}{\sum_{u\in\mathcal{U}} \hat{R}_u z_{ub}} \sum_{u\in\mathcal{U}} \hat{R}_u z_{ub} x_u^{usr} \quad \forall b \in \mathcal{B}, \tag{4.5}$$

$$y_b^c = \frac{1}{\sum_{u\in\mathcal{U}} \hat{R}_u z_{ub}} \sum_{u\in\mathcal{U}} \hat{R}_u z_{ub} y_u^{usr} \quad \forall b \in \mathcal{B}. \tag{4.6}$$

We use the standard 3-dB drop in antenna gain to demarcate the beam footprint boundary, but other thresholds, as well as a discrete set of possible beam widths, are possible. The meanings of all the sets, parameters, and variables are given in Table 1. The off-boresight angle θ_{ub} is calculated from simple geometric considerations. The factor 2 in (4.4) is to account for the meaning of 3-dB angle as the diameter of the spot-beam, which cannot be smaller than W^{min}.

Next, consider the outer user to beam allocation and colouring problem, in which the beam centres are fixed but their widths θ_b^{3dB} are constrained by the maximum and minimum beam widths W^{max} and W^{min}. Here, there are three objectives mentioned earlier. First, we want to minimise the total unserved demand. Second, we aim at locating the users close to their beam centres. Third, we want the smallest overlap of the beams of the same colour. We formulate this problem as the following linear mixed integer programming (MIP) problem.

Allocation and Colouring Problem

$$\min w_1 \sum_{b\in\mathcal{B}} \sum_{u\in\mathcal{U}} r_{ub} + w_2 \sum_{u\in\mathcal{U}} \sum_{b\in\mathcal{B}} \theta_{ub} z_{ub} + w_3 \sum_{b1\in\mathcal{B}} \sum_{b2\in\mathcal{B}_{b1}} v_{b1b2} \tag{4.7}$$

$$\text{s.t.:} \sum_{u\in\mathcal{U}} (\hat{R}_u z_{ub} - r_{ub}) \leqslant R_b \quad \forall b \in \mathcal{B}, \tag{4.8}$$

$$r_{ub} \leqslant \hat{R}_u z_{ub} \quad \forall u \in \mathcal{U}, b \in \mathcal{B}, \tag{4.9}$$

$$\sum_{b\in\mathcal{B}} z_{ub} = 1 \quad \forall u \in \mathcal{U}, \tag{4.10}$$

$$\sum_{q\in\mathcal{Q}} col_{bq} = 1 \quad \forall b \in \mathcal{B}, \tag{4.11}$$

$$\theta_{ub} z_{ub} \leqslant \frac{1}{2} W^{max} \quad \forall u \in \mathcal{U}, b \in \mathcal{B}, \tag{4.12}$$

$$a_{b1b2q} \leqslant col_{b1q} \quad \forall b1 \in \mathcal{B}, b2 \in \mathcal{B}_{b1}, q \in \mathcal{Q}, \tag{4.13}$$

$$a_{b1b2q} \leqslant col_{b2q} \quad \forall b1 \in \mathcal{B}, b2 \in \mathcal{B}_{b1}, q \in \mathcal{Q}, \tag{4.14}$$

$$a_{b1b2q} - col_{b1q} - col_{b2q} \geqslant -1 \quad \forall b1 \in \mathcal{B}, b2 \in \mathcal{B}_{b1}, q \in \mathcal{Q}, \tag{4.15}$$

$$a_{b1b2q} - v_{b1b2} \leqslant \frac{2\theta^c_{b1b2}}{(W_{b1} + W_{b2})} - \Delta^{sep} \quad \forall b1 \in \mathcal{B}, b2 \in \mathcal{B}_{b1}, q \in \mathcal{Q}, \tag{4.16}$$

$$r_{ub}, v_{b1b2} \geqslant 0, z_{ub}, col_{bq}, a_{b1b2q} \in \{0, 1\}.$$

Let us explain the objectives and constraints. The three mentioned objectives are combined with their respective relative importance weights $w_1, w_2, w_3 \geqslant 0$ using a traditional scalarisation method [39]. The weights also account for objectives expressed on different scales (unmet demand vs. angular distance vs. degree of overlap). Note that the problem variables are in lower case and the fixed parameters in upper case (except w_i). The first term of the objective is the unmet demand found from (4.8)–(4.9) (each term in (4.8) is simply $\max(\hat{R}_u z_{ub} - r_{ub}, 0)$, the satisfied demand, hence the total beam capacity R_b is not exceeded). The constraints (4.10) and (4.11) ensure that every user is allocated to exactly one beam and every beam has one specific colour. Constraint (4.12) ensures the maximum beam width is not exceeded.

The second objective term aims at allocating the users to one of their closest beams, hence it is simply the sum of the off-boresight angles. Note that we use off-boresight angles as coordinates in the satellite system, which facilitates accounting for the projections of the circular beams onto the Earth's surface.

The constraints (4.13)–(4.15) linearise the formula $a_{b1b2q} = \min\{col_{b1q}, col_{b2q}\}$ for binary variables [43], so that $a_{b1b2q} = 1$ indicates that both beams share colour q, and finally (4.16) expresses the variable v_{b1b2}, which is 0 for non-overlapping beams or beams of different colours, and otherwise gives the degree of overlap relative to the combined size of both beams. The separation factor $\Delta^{sep} \leqslant 0.2$ is an extra safety margin to account for a gradual decrease of power outside 3-dB radius: interference is still felt by users located within θ^{3dB} plus that safety margin.

The third term of the objective aims to minimise the total degree of beam overlaps. Hence it forces the solver to choose distinct colours for beams located close to each other. The objective weights are selected by first normalising the associated objectives and converting them into dimensionless units: w_1 has a factor $\frac{1}{|\mathcal{U}|Median(\hat{R}_u)}$, w_2 has a factor $\frac{1}{|\mathcal{U}|W^{max}}$, and the third objective is already dimensionless.

Then we have user-specified priorities, which are adjusted empirically. In our setting, we found that the weights $w_1 = \frac{1}{|\mathcal{U}|Median(\hat{R}_u)}$, $w_2 = \frac{400}{|\mathcal{U}|W^{max}}$, $w_3 = 1$ were satisfactory, and the resulting allocations were not very sensitive to the factor 400 in w_2 (one can halve or double it with little change in the result, see Section 6). Too-small values lead to all spot-beams being consistently larger (and with more overlap).

The proposed bilevel program [5, 38] iterates two steps: (1) beam steering, and (2) allocation and colouring. Starting with some random selection of initial spot-beams centres, the allocation and colouring problem (4.7)–(4.16) is solved providing the optimal allocation and colouring variables. Then the spot-beam centres are updated using beam steering (4.5)–(4.6) (centroids of the allocated users) and the allocation and colouring problem (4.7)–(4.16) is solved again with the new values θ_{ub} computed from the new beam centres. The process repeats for a specified number of iterations.

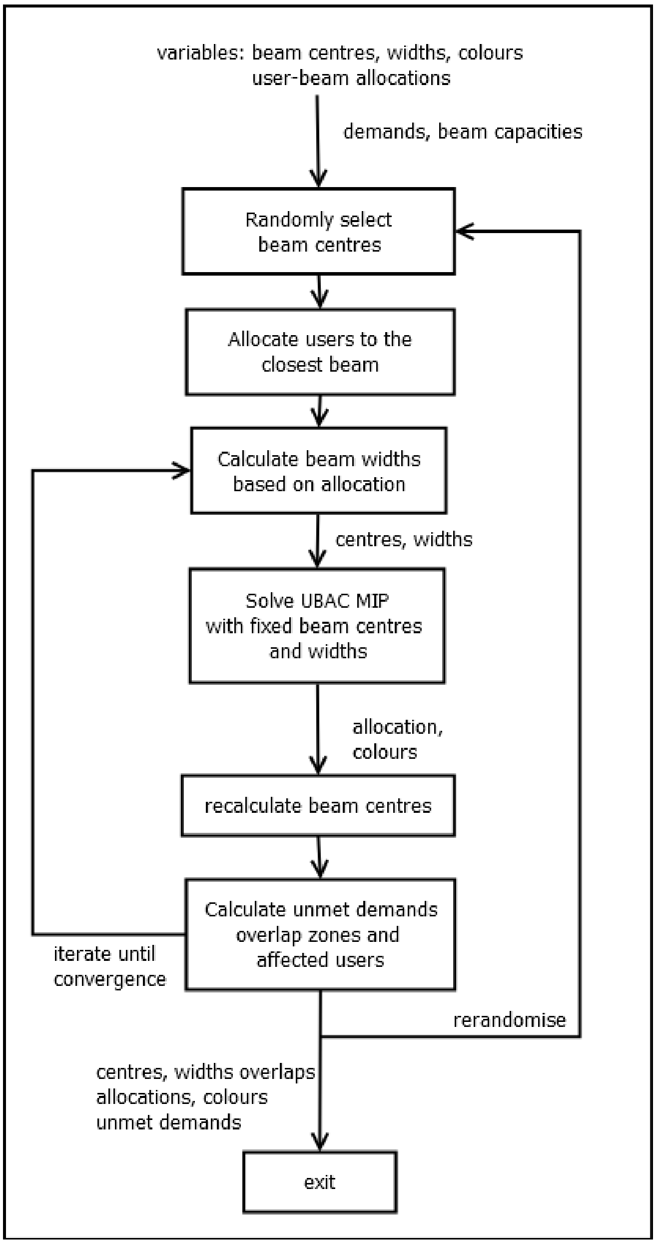

Figure 2. Data flow diagram for the UBAC problem.

The overall data flow diagram for the UBAC problem is presented in Figure. 2. There are internal iterations between the two steps of the bilevel program (backward arrow on the left) and also external iterations (backward arrow on the right), which re-randomise the initial beams centres. The latter are needed for convergence to a different solution from the initial random point, which can be potentially better. Thus the overall algorithm is similar to a random start.

Algorithm 1: Bilevel optimisation algorithm for the UBAC problem

Input: $\hat{R}, R, (x^{usr}, y^{usr})$ and other parameters from the first half of Table 1

Output: $z, col, r, (x^c, y^c), W^b$ (from the second half of Table 1)

1 **for** $k = 1, 2, \ldots, K^{extern}$ **do:** (external iterations)

1.1 $RandomSelect(x^c, y^c)$ and calculate angular distances θ, θ^c

1.2 **While** $\Delta(x^c_b, y^c_b) < Threshold$ **do**: (internal iterations)

1.2.1 Solve MIP problem (4.7)–(4.16)

1.2.2 Minimise $Steering(z)$ function explicitly by evaluating (4.5), (4.6)

1.2.3 Enforce the limit on beam size (4.4)

1.2.4 Calculate inter-beam distances θ^c and θ to update the MIP

1.3 Evaluate beam overlap (using θ^{3dB}_b), demand satisfaction criteria, and save current solution

2 Select the best solution $z, col, r, (x^c, y^c), W_b$

3 Return $z, col, r, (x^c, y^c), W_b$.

The internal iterations converge to one of the local optima and are executed a predefined number of times. We cannot provide theoretical insights as to the speed of convergence, but numerical evidence presented later suggests only a handful of iterations are necessary. Our empirical criterion of convergence is stabilisation of the solution to the inner problem, i.e., we stop iterations of the bilevel program once the beam centres stabilise within 10% of the minimal beam footprint width ($Threshold$ in Algorithm 1). The 10% mark is driven by pragmatic considerations: the balance between the computing budget and the stability of the solution. The external iterations provide new starting points for beam centres.

From the implementation point of view, all problems (4.7)–(4.16) are standard sparse MIPs of the same type (they differ between the iterations by parts of the right-hand sides and coefficients only), therefore the overhead of setting up the problem for the MIP solver is amortised.

There are various methods for choosing the initial random beam centres. Selecting random geometrical points in the region of interest is not suitable when the distribution of users on the ground is highly non-uniform, as many users may not be allocated to any spot-beam, whereas some beams may have no users due to the constraint on the largest spot-beam diameter. One can also select beam centres following some regular pattern (a tessellation, or a space-filling curve), but in our case, the region of interest may not be possible to cover fully with a given number of spot-beams of limited size.

Another approach is to select spot-beam centres as coordinates of randomly selected users. This method is also problematic as spot-beam centres tend to be selected within user clusters, leaving some remote users unserved by any beam.

The approach we chose is to follow random user coordinates, but dynamically reallocate probabilities of selecting a particular user coordinates as a beam centre based on its distances from the already selected beams. The probability of a user being selected is reduced when that user is within 3-dB radius of some beam. This way, remote users have a greater chance of being selected as a beam centre. Specifically, we use a discrete random generator which initially assigns equal weights to each user (another option is to use weights proportional to traffic demands). When each beam centre is allocated, the angular distances to each user are evaluated, and whenever a user is within W^{max} of the beam centre, its weight is reduced by half (a more aggressive option is to reduce it by $\frac{4}{5}$). This way, the random generator is less likely to select another beam centre in the vicinity of users that could be already covered by an existing beam. This approach does not unnecessarily prevent highly overlapping beams (they will be coloured differently to avoid interference, see the resulting figures in the following sections), as one or more of a large group of clustered users can still be selected as another beam centre despite the reduced weights.

Regardless of how the initial beam centres are selected, there is always a chance that some users are outside any spot-beam coverage. The second objective in problem (4.7) aims to cluster the allocated users around their beam centre, but the constraint (4.12) imposes a hard bound. That constraint can be relaxed by changing it to a penalty term in the objective, like $\sum_u \sum_b \max\{0, -\frac{1}{2}W^{max} + \theta_{ub} z_{ub}\}$.

On the other hand, repeating the whole process by means of external iterations will also present chances to select a feasible initial allocation. Dynamically reducing the parameter W^{max} from a larger initial value (e.g., 5 degrees) is another option.

5. More sophistication in the UBAC problem

In this section, we present a few variations and extensions which aim at incorporating some important nuances in the practical setting. We do not aim here for a comprehensive analysis but rather wish to outline possible variations which do not change the type of the problem nor the overall algorithm for its solution. Therefore, such variations can be incorporated into software with relative ease according to customer requirements. Hence, this section outlines future work on the subject. We leave the overall scheme in Figure 2 the same, and also keep the two steps of the UBAC problem in their respective classes, but modify them through additional terms and constraints.

First, we consider the UBAC problem for multiple satellites, for instance, two identical satellites for simplicity. In this case, a user can be served by one of the two satellites within their range of visibility. If one of the satellites is below, say, 20 degrees over the horizon from a particular user's u perspective, a set of hard constraints $z_{ub} = 0$ is added for all b served by that satellite. The spot-beams of the same colour but produced by different satellites are considered non-interfering. The allocation and colouring problem stays essentially the same, but some values a_{b1b2q} are now set to 0 explicitly (or implicitly, by eliminating such variables altogether) when beams $b1$ and $b2$ are produced by different satellites. Thus the UBAC problem becomes easier to solve because of fewer variables and constraints, provided the total number of beams stays the same. On the other hand, multiple satellites typically offer more spot-beams, which makes such problem instances larger.

We should mention that the UBAC problem cannot be split into two smaller uncoupled problems, for each individual satellite, because they share common variables r_{ub}, z_{ub} and the constraint (4.10).

Second, we incorporate beam transmission capacity as a function of beam width W_b. Clearly, wider spot-beams provide less boresight antenna gain, and hence less signal power. In problem (4.7)–(4.16), the values R_b were constants, but in fact, the resulting lower SINR affects R_b through Shannon's formula or its modifications.

We approximate the reduction of R_b for wider beams through a linear function, replacing (4.8) with

$$\sum_{u\in\mathcal{U}}(\hat{R}_u z_{ub} - r_{ub}) \leqslant \frac{R_b(W^{max} - W_b) + R_b^{min}(W_b - W^{min})}{W^{max} - W^{min}} \tag{5.1}$$

$$= \frac{R_b^{min} - R_b}{W^{max} - W^{min}} W_b + \frac{R_b W^{max} - R_b^{min} W^{min}}{W^{max} - W^{min}},$$

where R_b^{min} is a given smallest beam capacity (of the widest possible beam $W_b = W^{max}$). While here, more accurate nonlinear expressions for $R_b(W_b)$ are possible (the MIP problem remains in the same class), we used a linear function as a relatively accurate approximation, taking into account that the UBAC problem is followed by the power allocation problem (to be presented elsewhere) which properly accounts for SINR and link budgets as functions of beam widths and radiation patterns. Moreover, the subsequent power allocation problem can also incorporate radiation patterns produced by phased array antennas [28], which will provide an even more accurate treatment of inter-beam interference. Because of that subsequent power allocation stage, it suffices to have just a reasonable approximation at the UBAC stage in the overall process.

Third, we also incorporate the users' demand priorities at two stages: at the random beam initialisation stage and at the UBAC stage. In the random beam initialisation strategy that we chose, the beams' centres are selected as some user's coordinates. The probability of a particular user to be chosen is affected by their demands, and also by their transmission priority (relative to the rest). These probabilities are recalculated each time a new beam center is chosen in order to reduce the probabilities of the users already covered by one or more beams (but not completely, as we allow for overlapping beams of different colours). At the UBAC stage we incorporate priorities in two ways: in formulas (4.5)–(4.6) by an extra priority weighting factor under the sum, and by an extra weight in the first objective (unmet demand) for each user u.

Another variation is related to softening the hard constraint that each user must be allocated to one of the spot-beams. While that was our modelling choice in the current work, we also considered an alternative formulation in which not all users were required to be in any spot-beam footprint, and hence not served. While the first term in our objective (4.7) specifically penalises unserved demands, it is effective only under constraint (4.10). Hence we also added an extra penalty $\sum_u \sum_b (1 - z_{ub})\hat{R}_u$ and changed (4.10) to a "less than or equal to" constraint. Under these conditions remote users would be unallocated if random initialisation failed to choose those users as spot-beam centres. The addition of the fourth term into the objective (with its own weight) changes considerably the interplay between different objectives. A detailed study of this scenario and its implications are left to future work.

We should mention another variation, which is not directly applicable to downlink spot-beam allocation, but is useful for uplink allocation (i.e., when communication links are from users to the gateway), which is essentially the same UBAC problem. Here, the gateways are illuminated by narrow beams of full bandwidth (i.e., all colours) serving the gateways exclusively due to their high transmission demand. The additional challenge here is the interference with these fixed spot-beams by beams of any colour. Thus the UBAC problem should avoid placing other beams in the vicinity of the fixed gateway beams. This requirement is incorporated into the UBAC by setting the relevant $a_{gbq} = 1$, where g is the label of the fixed gateway beam.

Potential infeasibility of the UBAC MIP problem subject to the additional constraints can be dealt with by converting the above constraints into penalty terms in the objective.

In all these variations, we kept the UBAC problem in the same class of problems (MIP) by means of introducing additional linear constraints and extra weights. It means that the solution strategies remain the same, and the same MIP solvers can be used in the software. It also means that the computational costs of the numerical solution could be approximately the same. In saying so, we know that the introduction of even a single constraint may increase the difficulty of the MIP, so that locating its global optimum by the MIP solvers becomes much more costly. In our application, we take a pragmatic approach of not necessarily solving the MIP to optimality but allocating to it a certain budget, and selecting the best feasible solution.

6. Numerical experiments

6.1. Instance generation

To test the algorithms presented in the previous section, we generate users within a box that contains Australia and its close neighbours. The more users generated, the larger the box. To generate users in the box, we generate N_C clusters, where $N_C = \frac{N_U}{15}$ (but the number of users in each cluster is a random variable, with the total number of users N_U). There are two types of clusters:

- uniform: user coordinates are generated using a uniform distribution within the original box. There is one uniform "cluster".
- normal: user coordinates are generated using a normal distribution. Each normal cluster has a random centre (mean) within the box, r_C, and a standard deviation, σ_C, calculated as: $\frac{min_{corner}\{\|r_C - r_{corner}\|\}}{max\{N_U/10, 1\}}$, where the index *corner* runs through the corners of the enclosing box. There are $N_C - 1$ normal clusters.

The transmission rate requested by each user $[Kbps]$ is generated using a uniform distribution, $\hat{R}_u \sim U(3750, 31250)$. Each instance of the problem is defined by the following factors: number of users, number of beams, transmission level, and W^{max}. Table 2 presents the range of each of these factors used to create the set of scenarios.

Factor	Levels
N_U	$\{150, 200, 300, 500, 1000, 1500, 2000\}$
N_B	$N_U/10$
Transmission level	$\{Medium, High\}$
$W^{max}[degrees]$	$\{2, 4\}$

Table 2. Instance settings.

The Medium level of transmission rate means that $R_b \sim U(\frac{\sum_u \hat{R}_u}{N_B}, 1.1\frac{\sum_u \hat{R}_u}{N_B})$ and the High level means that $R_b \sim U(1.1\frac{\sum_u \hat{R}_u}{N_B}, 2\frac{\sum_u \hat{R}_u}{N_B})$. The former creates a tight problem, an the beam transmission capacities are modest. The latter creates a more relaxed problem with high overall beams capacities.

Figures 3–5 illustrate data instances with few compact and less compact clusters, and also show the resulting solutions to the UBAC problem (circular beam footprints projected onto the Earth's surface) from a geostationary satellite located at longitude 110 degrees East. The intersection of beams of the same colour is minimised. Some clusters of users were required to be allocated to the same beam. We used four colours in the UBAC, but other values are also possible.

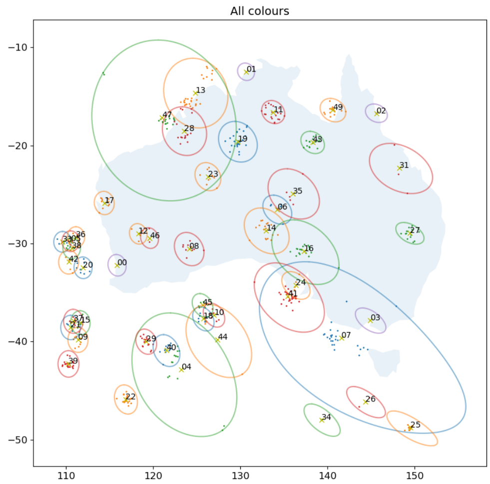

Figure 3. The data distribution (compact clusters of dots, few uniformly distributed users) and the resulting user-to-beam allocation and beam colouring. 500 users, 50 beams, and 4 gateways.

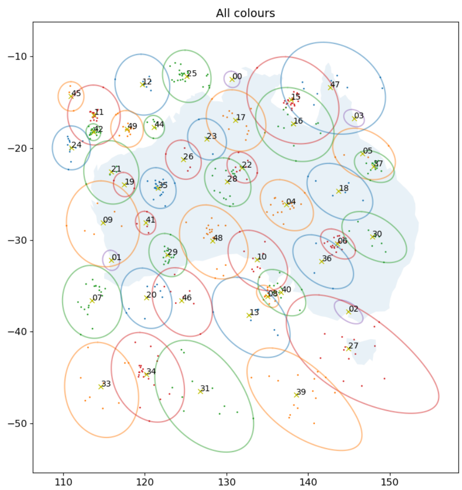

Figure 4. The data distribution (less compact clusters of dots, more uniformly distributed users) and the resulting user-to-beam allocation and beam colouring. 500 users, 50 beams, and 4 gateways.

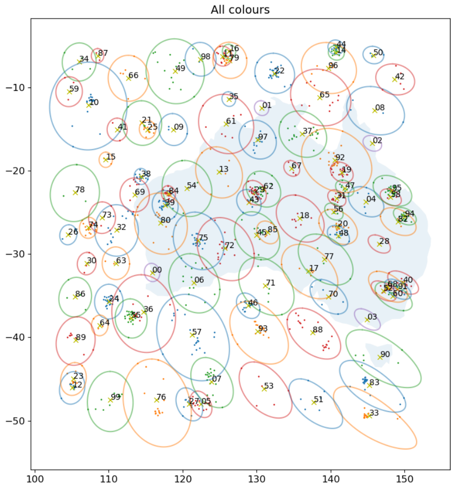

Figure 5. The data distribution (many compact clusters of dots, many users uniformly distributed across the box) and the resulting user-to-beam allocation and beam colouring. 1000 users, 100 beams, and 4 gateways.

6.2. Implementation and platform

The UBAC problem formulated in Section 4, as well as its variations reported in Section 5, were implemented in C++ language. As the MIP solver, we opted for the open source GNU Linear Programming Kit (GLPK) [44] (implemented in C). We also developed a version that uses commercial Cplex software, but did not observe significant differences in solution quality and computational costs.

For testing, we used a modest Windows 10-based workstation, Intel i7 CPU at 1.90GHz with 32GB of RAM. The algorithms ran using a single processor core for fair comparison. Microsoft Visual Studio 2019 compiler was used.

6.3. Numerical results

In this section, we present two types of numerical results. First, we performed scalability studies, interested in the CPU time from relatively small to larger problems. We measured average CPU wall time for various problem instances, specifically focusing on increasing the number of users N_U, and separating harder and easier problems (based on the chosen transmission level). The number of beams was fixed at $N_U/10$, so that 10 users on average were allocated to one beam, excluding the gateway beams which served a single gateway each.

In the second type of experiment, we evaluated the solution quality using different metrics [4, 12, 20]. Typically, the following metrics are used to evaluate the quality of transmission [45–51]:

Unmet demand: $\sum_{u \in \mathcal{U}} \max(\hat{R}_u - R_u, 0)$

Useful capacity: $\sum_{u \in \mathcal{U}} \min\{R_u, \hat{R}_u\}$

$$\text{Satisfaction SF: } \frac{\sum_{u\in\mathcal{U}} \min\{R_u, \hat{R}_u\}}{\sum_{u\in\mathcal{U}} \hat{R}_u}$$

$$\text{Offered ratio: } \sum_{u\in\mathcal{U}} \frac{R_u}{\hat{R}_u}$$

$$\text{Allocation fairness: } \min_{u\in\mathcal{U}} \frac{R_u}{\hat{R}_u}$$

The unmet demand is meant to be minimised whereas the remaining metrics are meant to be maximised.

However, those metrics also require the power allocation step in order to calculate transmission rates, also accounting for non-circular beam footprints produced by the phased array antenna, which we solve as a separate problem. That step will affect inter-beam interference (which extends beyond the 3-dB footprint, especially for phased array antennas [28]) and the spectral efficiency, and hence the transmission rates. In this instance, we only measure the quality of user-to-beam allocation using the following:

- The percentage of overlapping beam footprints
- The percentage of users in the overlap zones

The scalability results are presented in Table 3. We fixed the number of external iterations (random initialisation) at 5. Clearly, a larger number of external iterations would improve the quality of the solution, but we were limited by our overall computing budget. We did not observe very significant improvements when increasing that number in small problem instances, see Figure. 6. We see that the CPU time increases significantly with the increase of the number of user terminals and beams, which is not surprising as the UBAC problem is based on the MIP, which is an NP-hard problem. The practical operating range of $N_B \in [50, 150]$, so it seems that optimal beam allocation can be found with acceptable effort. In addition, problems with higher beam capacity seem to be easier to handle by the algorithm.

The quality metrics were averaged across various random scenarios and are presented in Table 4. For small numbers of beams and users, the percentage of beam overlap and of the affected users is low (1 beam out of 20), and it grows to approximately the 10 % mark. This number is acceptable, as being located in the overlap zone does not immediately imply poor transmission. These users' demands can be satisfied by allocating sufficient transmission power to their respective channels, and also allocating different channels to the users in the overlapping beams.

Table 3. Average running time (sec) and the number of internal iterations for instances of the UBAC problem with high and medium transmission levels of each beam. The number of iterations is an indicator of convergence.

N_U	N_B	High Tx rate	Medium Tx rate	Iterations High Tx	Iterations Medium Tx
150	15	7	9	3	3
200	20	45	84	5	4
300	30	321	503	5	7
500	50	2391	5324	7	7
1000	100	7312	10069	7	9
1500	150	12162	19321	7	9
2000	200	> 24h	> 24h	7	7

Table 4. Average percentage of overlapping beams (of the same colour) and the percentage of users affected.

N_U	N_B	overlapping beams %	users in overlap zones %
150	15	0	0
200	20	5	6
300	30	5	4
500	50	8	8
1000	100	9	11
1500	150	11	11
2000	200	14	12

We should note that there are more challenging distributions of users on the ground that can make the UBAC problem harder to solve and also affect the quality of the solution, for example, by locating many users in the same geographical cluster, so that their demands cannot be served by only four overlapping beams of different colours, or locating only a few users at the outskirts of the coverage area. Such problem instances require special attention and potentially human intervention, for example, by fixing some spot-beams at certain points, which can be added to the UBAC problems in the form of fixed values or constraints.

Table 5. Average percentage of overlapping beams (of the same colour) , the percentage of users affected,and unmet demand by clustering and greedy colouring vs. the UBAC method.

N_U	N_B	Overlapping beams %	Users in overlap zones %	Unmet demand %
Clustering				
150	15	2	6	15
200	20	10	8	19
300	30	12	12	22
500	50	18	24	18
UBAC				
150	15	0	0	0
200	20	5	6	3
300	30	5	4	3
500	50	8	8	8

While for assessing UBAC problem solutions, we have not used the unserved demand criterion (which is the first term in our objective), because it will be significantly affected in the power allocation step due to co-interference, beamforming and link budget calculations, for benchmarking purposes, we used the value of our objective as an estimate of unserved demand in user to beam allocation step. As a benchmarking baseline, we chose some of the standard clustering algorithms (k-means, fuzzy c-means, hierarchical clustering) and a subsequent greedy beam colouring algorithm. We fixed the number of clusters at N_B in each instance and four colours, and computed the resulting percentages of overlapping beams (of the same colour), users in the overlap zones, and an estimate of unmet demand based on calculating the excesses of beam capacities by the allocated users. We used k-means, Gaussian mixture models, and agglomerative clustering, and selected the best clustering result out of three based on the unmet demand, and then performed greedy colouring to find the overlaps. The CPU time for baseline clustering was under 20 seconds (we used 100 independent runs), but the user to beam allocation did not account for fixed beams capacities, and as a consequence, various beams capacities were exceeded by a large margin. The colouring step was effective in reducing the overlaps, but often more than four clusters were in close proximity, which made overlaps unavoidable. Also, in our experiments, the maximum beam widths were not exceeded, even though these were not part of clustering criteria. We report the comparisons in Table 5, which indicates the improvements made through our solution approach.

In addition, we compared our results to those in [22, 30], although indirectly, because these methods included beam transmission powers among the optimisation variables (and hence also aimed at minimising co-interference), which we perform in a separate step. Also the methods in [22, 30] allowed some users not to be allocated to any spot-beams, which was different from our setting. We noticed that the mentioned methods from [22, 30] offer 3% to 10% of users not covered by any spot-beams (for 500 users and 72 spot-beams), which translates into 5% to 10% unmet demand, comparable to our results in Table 4. However, in larger instances (5000 users, 72 spot-beams), the unserved demand increases up to 35%, which presents a challenge and a scope for future work involving large instances.

To illustrate sensitivity of the resulting user allocation and colouring to the weights of the objectives in (4.7), we performed experiments with different w_2 values on a number of smaller problem instances (due to a limited computing budget), and report the results in Table 6. We see here that smaller values of w_2 favour less compact beams with more overlap, but because of beam-colouring step, even larger spot-beams did not result in very significant degradation of the solution quality.

In another set of experiments, we measured the improvements to the best objective function values after a specified number of external iterations K^{extern}. Here, we kept all algorithm parameters fixed and recorded the best values of the objective (4.7), which were later normalised to the best objective after 50 external iterations (taken as an approximation to the global optimum). Figure 6 shows that in relatively small problem instances, the biggest improvements happen in the range $K^{extern} \in [1, 10]$, whereas larger numbers of external iterations provide only marginal benefit.

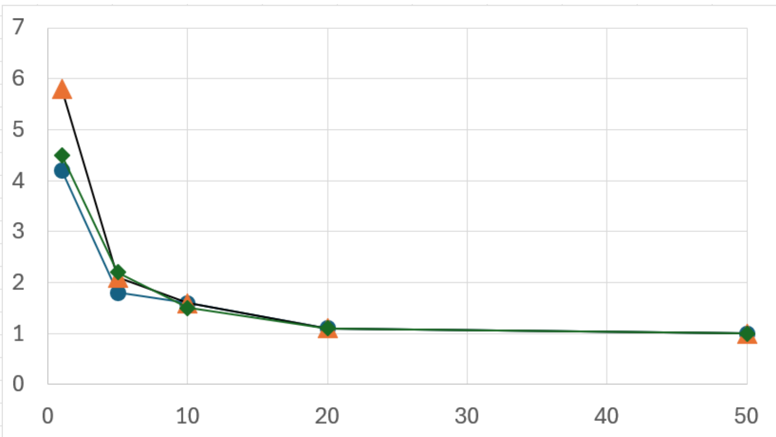

Figure 6. The dependence of the normalised value of the objective on the number of external iterations. The objective values are taken relative to the best objective after 50 external iterations $Obj_{norm} = \frac{Obj_{K^{extern}}}{Obj_{50}}$. The values are averaged over 10 instances. △ corresponds to $N_U = 200, N_B = 20$, ○ corresponds to $N_U = 150, N_B = 15$, and ◇ corresponds to $N_U = 300, N_B = 30$.

Table 6. Dependence of the measured solution characteristic as a function of objective weight coefficient w_2, averaged over 10 instances.

N_U	N_B	w_2	Overlapping beams %	Users in overlap zones %
200	20	100	8	8
		200	7	8
		400	5	6
		800	5	6
300	30	100	7	7
		200	5	7
		400	5	4
		800	5	5

7. Conclusion

We present a mathematical programming formulation of the problem of user-to-spot-beam allocation and beam colouring, which arises in satellite communications, and examine the computational performance of the proposed methods. The overall problem is formulated as a bilevel optimisation problem. At one level, the user-to-beam allocation and colouring problem is formulated as a mixed-integer linear programming problem; at the other level, the calculation of spot-beam centres and widths is solved explicitly. The algorithm iterates between the two problems, starting from an initial random choice of beam centres.

We performed a computational study of the scalability of our approach, up to 2000 users and 200 spot-beams, and also evaluated the quality of the numerical solutions using different metrics.

The problem formulation and computational algorithms presented here constitute an important step in satellite resource allocation, followed by carrier and power allocation, which will be addressed elsewhere. The proposed methods are applicable to multi-satellite systems, two-way communication (via uplink and downlink spot-beam allocation), and a variety of communication topologies. Future research will address more challenging problem instances, as well as subsequent stages, such as power allocation and frequency allocation, in order to further minimise inter-beam interference.

Author contributions

All authors: Conceptualization, G. Beliakov: Methodology, Formal analysis, Investigation, Software, Supervision, Writing–original draft, Writing–review and editing; V. Mak-Hau: Methodology, Investigation, Supervision, Resources, Writing–review and editing, Funding acquisition, Project administration; J. Calle Salazar: Investigation, Software, Data curation, Methodology, Validation, Writing–original draft, Writing–review and editing; M. Jacobs: Investigation, Methodology, Validation, Funding acquisition, Project administration, Writing–review and editing; J. Kodithuwakkuge: Investigation, Methodology, Validation, Resources, Data curation, Project administration, Writing–review and editing.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to acknowledge support from the Defense Science and Technology Group under grant DSP10369.

Conflict of interest

The authors declare no conflict of interest.

References

1. G. Maral, M. Bousquet, Z. L. Sun, *Satellite communications systems: Systems, techniques and technology*, John Wiley and Sons, 2020. https://doi.org/10.1002/9781119673811
2. N. Pachler, J. J. Garau-Luis, E. F. Crawley, B. G. Cameron, A unified resource allocation framework and impact evaluation for NGSO satellite constellations, *Int. J. Satell. Commun. Netw.*, **43** (2025). 77–96. https://doi.org/10.1002/sat.1547
3. X. Tong, N. X. Li, J. X. Liu, S. Peng, An efficient downlink resource allocation scheme in power-limited GEO satellite systems, *Digit. Commun. Netw.*, **11** (2025), 1751–1761. https://doi.org/10.1016/j.dcan.2025.05.007
4. R. Chen, W. X. Long, B. Wang, Y. He, R. Sun, N. Cheng, et al., Multibeam high throughput satellite: Hardware foundation, resource allocation, and precoding, *IEEE Commun. Surv. Tutor.*, **28** (2026), 5379–5415. https://doi.org/10.1109/COMST.2026.3673850
5. J. F. Bard, *Practical bilevel optimization*, Springer, New York, 2010. https://doi.org/10.1007/978-1-4757-2836-1
6. S. Dempe, A. Zemkoho, *Bilevel optimization*, Springer Nature, 2020. https://doi.org/10.1007/978-3-030-52119-6
7. X. W. Chen, Q. L. Liu, D. A. Ralescu, A bi-level optimization model for the asset-liability management of insurance companies, *J. Ind. Manag. Optim.*, **19** (2023), 3003–3019. http://dx.doi.org/10.3934/jimo.2022074
8. N. D. I. O. Pachler, *A complete resource allocation framework for flexible high throughput satellite constellations*, Massachusetts Institute of Technology, 2022.
9. P. J. Honnaiah, N. Maturo, S. Chatzinotas, S. Kisseleff, J. Krause, Demand-based adaptive multi-beam pattern and footprint planning for high throughput GEO satellite systems, *IEEE Open J. Commun. Soc.*, **2** (2021), 1526–1540. https://doi.org/10.1109/OJCOMS.2021.3093106
10. J. T. Camino, C. Artigues, L. Houssin, S. Mourgues, MILP formulation improvement with k-means clustering for the beam layout optimization in multibeam satellite systems, *Comput. Ind. Eng.*, **158** (2021), 107228. https://doi.org/10.1016/j.cie.2021.107228
11. N. Torkzaban, A. Zoulkarni, A. Gholami, J. S. Baras, Capacitated beam placement for multi-beam non-geostationary satellite systems, *IEEE Wirel. Commun. Netw. Conf.*, 2023, 1–6. https://doi.org/10.1109/WCNC55385.2023.10118814

12. J. F. Zhu, Y. H. Sun, X. A. Feng, S. G. Wang, L. Liu, M. Peng, Beam footprint design, scheduling, and spectrum assignment in low earth orbit mega-constellations, *IEEE Trans. Vehic. Tech.*, **74** (2025), 13968–13984. https://doi.org/10.1109/TVT.2025.3560320
13. O. B. Yahia, Z. Garroussi, O. Bélanger, B. Sansò, J. F. Frigon, S. Martel, et al., Evolution of high-throughput satellite systems: A vision of programmable regenerative payload, *IEEE Commun. Surv. Tutor.*, **27** (2025), 1565–1597. https://doi.org/10.1109/COMST.2024.3450292
14. J. Y. Song, S. Chan, H. W. Kim, B. J. Ku, D. Oh, S. Kim, Dynamic resource allocation for LEO satellite systems under heterogeneous traffic distributions, *Int. J. Satell. Commun. Netw.*, **44** (2026), 199–209. https://doi.org/10.1002/sat.70015
15. Y. He, K. Q. Qi, X. Feng, R. Chen, Load-balanced and full-coverage beam layout for multibeam satellite systems, *Int. Conf. Commun. Sig. Proc. Appl.*, 2024, 1–5. https://doi.org/10.1109/ICCSPA61559.2024.10794213
16. J. F. Zhang, W. Li, Y. Li, H. M. Wang, S. L. Li, A framework for joint beam scheduling and resource allocation in beam-hopping-based satellite systems, *Electronics*, **14** (2025), 2887. https://doi.org/10.3390/electronics14142887
17. Q. Wei, Z. Li, J. Shi, Z. Bai, X. Liao, P. Xiao, et al., Evolutionary reinforcement learning for resource efficiency optimization in downlink multi-leo communication system with cross-layer interference, *IEEE Trans. Cogn. Commun. Netw.*, **12** (2026), 4223–4236. https://doi.org/10.1109/TCCN.2025.3632396
18. S. Kashyap, N. Gupta, Resource allocation techniques in multibeam satellites: Conventional methods vs. AI/ML approaches, *Int. J. Satell. Commun. Netw.*, **43** (2025), 97–121. https://doi.org/10.1002/sat.1548
19. P. S. Eftetahi, L. Cai, A. Sepahi, Fair beam scheduling in leo satellite networks with reinforcement learning, *IEEE Trans. on Cogn. Commun. Netw.*, **12** (2026), 5438–5450. https://doi.org/10.1109/TCCN.2026.3657104
20. T. S. Abdu, S. Kisseleff, E. Lagunas, S. Chatzinotas, Flexible resource optimization for GEO multibeam satellite communication system, *IEEE Trans. Wireless Commun.*, **20** (2021), 7888–7902. https://doi.org/10.1109/TWC.2021.3088609
21. T. Van Chien, N. M. Quan, T. N. Do, C. Le, T. N. Nguyen, S. Chatzinotas, Fast beam placement for ultra-dense leo networks, *IEEE Wireless Commun. Lett.*, **14** (2024), 621–625. https://doi.org/10.1109/LWC.2024.3516818
22. A. Gaudry, R. Li, V. M. Hau, M. Jacobs, J. Kodithuwakkuge, Resource optimisation for geostationary communication satellites, *Proc. Int. Congr. Model. Simul.*, 2024. https://doi.org/10.36334/modsim.2023.gaudry
23. R. Li, A. Gaudry, V. M. Hau, CG-FlexBeamOpt: Advanced solution methodology for high throughput GEO satellite beam laydown and resource allocation, *Int. J. Satell. Commun. Netw.*, **42** (2024), 286–312. https://doi.org/10.1002/sat.1513
24. H. Shehata, H. Inaltekin, I. B. Collings, S. V. Hanly, P. Whiting, Beam direction optimization for next-generation geo satellite networks, *Asia Pacific Conf. Commun.*, 2023, 267–272. https://doi.org/10.1109/APCC60132.2023.10460601

25. V. N. Ha, E. Lagunas, T. S. Abdu, H. Chaker, S. Chatzinotas, J. Grotz, Large-scale beam placement and resource allocation design for meo-constellation satcom, *IEEE Int. Conf. Commun. workshops*, 2023, 1240–1245. https://doi.org/10.1109/ICCWorkshops57953.2023.10283806

26. V. P. Bui, T. Van Chien, E. Lagunas, J. Grotz, S. Chatzinotas, B. Ottersten, Joint beam placement and load balancing optimization for non-geostationary satellite systems, *IEEE Int. Mediter. Conf. Commun. Netw.*, 2022, 316–321. https://doi.org/10.1109/MeditCom55741.2022.9928761

27. P. J. Honnaiah, E. Lagunas, S. Chatzinotas, J. Krause, Demand-driven beam densification in multibeam satellite communication systems, *IEEE Trans. Aerospace Elec. Sys.*, **59** (2023), 6534–6554. https://doi.org/10.1109/TAES.2023.3278253

28. L. Yu, J. X. Wan, Y. K. Liu, K. Zhang, The latest developments of multi-beam phased array antennas for satellite communications, *Int. Conf. Microwave and Millimeter Wave Tech.*, 2022. https://doi.org/10.1109/ICMMT55580.2022.10023090

29. M. Takahashi, Y. Kawamoto, N. Kato, A. Miura, M. Toyoshima, Adaptive multi-beam arrangement for improving throughput in an hts communication system, *IEEE Int. Conf. Commun.*, 2020, 1–6. https://doi.org/10.1109/ICC40277.2020.9148622

30. A. Gaudry, R. Li, V. Mak-Hau, FlexBeamOpt: hybrid solution methodologies for high-throughput GEO satellite beam laydown and resource allocation, *Int. J. Satell. Commun. Netw.*, **41** (2023), 515–538. https://doi.org/10.1002/sat.1481

31. A. Kyrgiazos, B. Evans, P. Thompson, Irregular beam sizes and non-uniform bandwidth allocation in HTS multibeam satellite systems, *AIAA Int. Commun. Sat. Syst. Conf.*, 2013. https://doi.org/10.2514/6.2013-5617

32. M. Cooley, Phased array fed reflector (PAFR) antenna architectures for space-based sensors, *IEEE Aerospace Conf. Proc.*, 2015. https://doi.org/10.1109/AERO.2015.7118963

33. G. Toso, P. Angeletti, C. Mangenot, Multibeam antennas based on phased arrays: An overview on recent ESA developments, *The 8th Europ. Conf. Antennas and Propagation*, 2014, 178–181. https://doi.org/10.1109/EuCAP.2014.6901721

34. D. Sikri, R. M. Jayasuriya, Multi-beam phased array with full digital beamforming for SATCOM and 5G, *5G Phased Array Techn.*, **30** (2019), 1–20.

35. R. E. Collin, *Antennas and radiowave propagation*, McGraw-Hill, New York, 1985.

36. A. Kyrgiazos, B. Evans, P. Thompson, P. T. Mathiopoulos, S. Papaharalabos, A terabit/second satellite system for european broadband access: A feasibility study, *Int. J. Satell. Commun. Netw.*, **32** (2014), 63–92. https://doi.org/10.1002/sat.1067

37. C. Caini, G. E. Corazza, G. Falciasecca, M. Ruggieri, F. Vatalaro, A spectrum- and power-efficient ehf mobile satellite system to be integrated with terrestrial cellular systems, *IEEE J. Sel. Areas Commun.*, **10** (1992), 1315–1325. https://doi.org/10.1109/49.166759

38. S. Dempe, V. Kalashnikov, G. A. Perez-Valdes, N. Kalashnikova, *Bilevel programming problems; Theory, algorithms and applications to energy networks*, Energy Systems, Springer, New York, 2015. https://doi.org/10.1007/978-3-662-45827-3

39. Y. Collette, P. Siarry, *Multiobjective optimization*, Springer, 2011. https://doi.org/10.1007/978-3-662-08883-8

40. Y. C. Zhang, X. Hu, R. Chen, Z. L. Zhang, L. Q. Wang, W. D. Wang, Dynamic beam hopping for DVB-S2X satellite: A multi-objective deep reinforcement learning approach, *IEEE Int. Conf. Ubiquit. Comput. Commun. and Data Sci. Comput. Intel. Smart Comput., Netw. Serv.*, 2019. https://doi.org/10.1109/IUCC/DSCI/SmartCNS.2019.00056
41. Y. B. Lv, Z. P. Wan, Linear bilevel multiobjective optimization problem: Penalty approach, *J. Ind. Manag. Optim.*, **15** (2019), 1213–1223. http://dx.doi.org/10.3934/jimo.2018092
42. F. Fali, M. Moulaï, Solving discrete linear fractional bilevel programs with multiple objectives at the upper level, *J. Ind. Manag. Optim.*, **19** (2023), 2639–2654. https://doi.org/10.3934/jimo.2022059
43. G. G. Brown, R. F. Dell, Formulating integer linear programs: A rogues' gallery, *INFORMS Trans. Educ.*, **7** (2007), 153–159. https://doi.org/10.1287/ited.7.2.153
44. A. Makhorin, GNU Linear programming kit, 2012. Available from: https://www.gnu.org/software/glpk/.
45. Y. X. Wang, D. M. Bian, J. Hu, J. Y. Tang, C. Wang, A flexible resource allocation algorithm in full bandwidth beam hopping satellite systems, *IEEE Adv. Inf. Manag. Commun. Electron. Autom. Control Conf.*, 2019. https://doi.org/10.1109/IMCEC46724.2019.8984132
46. X. Alberti, J. M. Cebrian, A. Del Bianco, Z. Katona, J. Lei, M. A. Vazquez-Castro, et al., System capacity optimization in time and frequency for multibeam multi-media satellite systems, *Adv. Satell. Multimed. Syst. Conf. Signal Process. Space Commun. Workshop*, 2010, 226–233. https://doi.org/10.1109/ASMS-SPSC.2010.5586902
47. R. Alegre, N. Alagha, M. A. V. Castro, Heuristic algorithms for flexible resource allocation in beam hopping multi-beam satellite systems, *AIAA Int. Commun. Satell. Syst. Conf.*, 2011. https://doi.org/10.2514/6.2011-8001
48. M. W. Baidas, E. Alsusa, K. A. Hamdi, Performance analysis and SINR-based power allocation strategies for downlink NOMA networks, *IET Commun.*, **14** (2020), 723–735. https://doi.org/10.1049/iet-com.2018.6112
49. J. Lei, M. A. Vázquez-Castro, Joint power and carrier allocation for the multibeam satellite downlink with individual sinr constraints, *IEEE Int. Conf. Commun.*, 2010, 1–5. https://doi.org/10.1109/ICC.2010.5502506
50. M. Schubert, H. Boche, QoS-based resource allocation and transceiver optimization, *Found. Trends Commun. Inform. Theory*, **2** (2005), 383–529. https://doi.org/10.1561/0100000010
51. G. Zheng, S. Chatzinotas, B. Ottersten, Generic optimization of linear precoding in multibeam satellite systems, *IEEE Trans. Wireless Commun.*, **11** (2012), 2308–2320. https://doi.org/10.1109/TWC.2012.040412.111629