Non-Uniform User Distribution in Non-Terrestrial Networks with Application to User Scheduling

Non-Uniform User Distribution in Non-Terrestrial Networks with Application to User Scheduling

Bruno De Filippo*, Bilal Ahmad*, Daniel Gaetano Riviello†, Alessandro Guidotti‡, Alessandro Vanelli-Coralli*
*Department of Electrical, Electronic, and Information Engineering (DEI), Univ. of Bologna, Bologna, Italy
†Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy
‡National Inter-University Consortium for Telecommunications (CNIT), Parma, Italy
{bruno.defilippo, bilal.ahmad6, a.guidotti, alessandro.vanelli}@unibo.it, daniel.riviello@polito.it

***Abstract*—Non-Terrestrial Networks (NTNs) have recently seen a surge in interest from both the industry and consumers, attracted by the promise of a global coverage with drastically increased performance. Such coverage is achieved through the use of satellite mega-constellations which field of view can span from a few tens to hundreds of kilometers within each beam. Usually, the performance of NTN systems is assessed assuming uniformly distributed users. This approach, which has been extremely effective in the past when the satellite user densities were limited, neglects the human tendency to settle in clusters. Thus, a strong approximation on the real distribution of users is introduced when the subscriber density starts approaching the terrestrial one. This leads to results that do not reflect the actual system performance and, hence, might lead to suboptimal designs. In this paper, we therefore propose a Cluster-Process-based non-uniform distribution that better reflects the actual user density on ground, thus providing an improved tool for the design of NTN systems. To this aim, we propose a general procedure to extract statistics from population count datasets and tailor the user distribution model to the coverage area. To show the effectiveness and importance of considering non-uniform distributions, we empirically analyze the user scheduling performance in a B5G Low Earth Orbit High Throughput Satellite system. The per-cluster sum-rate capacity and user throughput achieved under non-uniform and uniform distribution are compared, highlighting the inaccuracies that the uniform assumption can introduce.**

***Index Terms*—Non-uniform user distribution, Cluster process, Non-Terrestrial Networks, User scheduling**

I. Introduction

Non-Terrestrial Networks (NTNs) have rapidly grown in popularity within the telecommunications sector, with more and more satellite-based services being launched every year. From text messages on handheld devices to broadband internet access on small User Terminals (UTs), revenues from consumer satellite services amounted to $93B in 2022, approximately one-fourth of the global space economy revenues in the same year [1]. It is worth mentioning that NTN is also a key technology for IoT applications in remote or wide areas, such as smart agriculture and freight localization. Indeed, moving the Base Stations (BSs) away from the coverage area, with satellites acting as relay nodes or even BSs themselves, removes the need to deploy a dense net of gateways where it would not be economically or physically possible. Reflecting the needs of both consumers and industry, the 3rd Generation Partnership Project (3GPP) has recently integrated satellite-based communications within their studies, with NTN-related technical specifications being included since Release 17 [2].

By and large, the success of NTN services can be attributed to their capability of providing coverage extension and service continuity through satellite constellations consisting of a few to thousands of spacecrafts [3], each covering a portion of the Earth's surface. However, the design of such systems requires in-depth analyses based on realistic parameters. Indeed, the satellite communications (SatCom)-related literature has flourished in recent years, with several novel algorithms tailored to NTN being proposed for telecommunication systems procedures (*e.g.*, handover management, and user scheduling). However, the evaluation of such algorithms has often been carried out assuming uniformly distributed users, potentially overlooking effects caused by the real distribution of users. While the assumption has been fairly accurate for the past decades, the rise in number of subscribers has been pushing the distribution of users farther from uniformity and closer to the real inhabitants' distribution. As a consequence, the clustered nature of the population can hardly be neglected anymore when assessing the performance of telecommunications procedures in NTN. Nonetheless, the NTN literature that models the user distribution as non-uniform is limited. In this framework, the authors in [4] proposed a non-uniform traffic model based on the inverse sampling transform technique applied to traffic density databases. In [5], the authors developed a traffic simulator extracting data from Fixed Satellite Services, aeronautical and maritime traffic datasets. In both works, the algorithms tend to geographically replicate with high fidelity the input database, leading to the evaluation of NTN procedures on the specific coverage area rather than on a model that statistically represents the characteristics of the UT distribution. In [6], the well-known mega-constellations from Starlink and OneWeb are compared assuming bivariate-Gaussian-distributed users. Although non-uniform, such distribution tends to concentrate most of the users towards a single center, making the model far from realistic when large coverage areas are considered. On the terrestrial side, the authors in [7] proposed to model the number of users in a terrestrial cellular network through a homogeneous Poisson Point Process (PPP); then, the geographical position of the users with respect to the serving BS is either uniform, circular symmetric Gaussian with zero mean (*i.e.*, high user density

close to the BS) or its inverse. This model assumes a perfect hexagonal tessellation, requiring the user aggregation points to be evenly spaced. Thus, the proposed distribution is not suitable for NTN, where the distribution of clusters of users (*i.e.*, towns and cities) may be affected by geographic features such as valleys and rivers. Improving on this, [8] assesses the cell load considering two Poisson Cluster Processes (PCPs), namely the Thomas and Matérn Cluster Processes (MCP and TCP). It must be noted that Poisson distributions may not be suitable for NTNs, where the number of clusters per unit area corresponds to the density of inhabited centers.

Indeed, the literature analysis shows the need for a non-uniform distribution tailored to NTN that can represent the geographical distribution of the population. In this paper, we propose a Cluster Process (CP)-based non-uniform user distribution model, named Non-Terrestrial CP (NTCP), suitable for the assessment of NTN procedures. To obtain the statistics used in the model, we present a procedure to process a global population count dataset. While real satellite users' data would lead to more accurate statistics, such information is typically confidential and of scarce availability. On the other hand, data on population distribution is freely accessible and, if appropriately processed, can be assumed to be representative of satellite users' data. To exemplify the shortcomings of the uniformity assumption on NTN procedures, we apply the NTCP to the use case of user scheduling in multi-beam B5G LEO High Throughput Satellite (HTS) systems. Overall, three main sources of novelty can be identified in our work:

- we consider a Bivariate Normal Distribution (BND) to generate users' positions within each cluster, allowing for more realistic non-circular clusters;
- we present a procedure to extract statistics from population distribution datasets;
- we represent the number of clusters with an empirical distribution, leading to a CP tailored to the service area.

II. Non-Uniform User Distribution

As mentioned in Section I, there is a strong connection between the user distribution in NTN and the real population distribution, making the uniform users assumption unrealistic. When a clustering behavior can be identified in the coverage area, the correlation between the coordinates of different users inside each cluster cannot be neglected, making CPs a more accurate model than Point Processes to distribute users in heterogeneous networks [9]. A CP is a two-step process in which *parent points* independently generate *offspring points* around them. In PCPs, first introduced in [10], both the number of parent points and offspring points follow a Poisson distribution with intensity parameter λ_p and λ_o, respectively. Then, the parent points are distributed in the considered area and the respective offspring points are scattered around them, typically following either a uniform or normal distribution. From the SatCom point of view, parent points represent centers of high inhabitants density, *i.e.*, towns and cities, while their offspring points are active users inside them. Naming λ_{o_c} the average user activity in the inhabited center c, it is fair to assume the number of offspring points of parent c to be generated through a PPP of average λ_{o_c}. Furthermore, to reflect the decrease in user density as the distance from the inhabited center increases, the offspring points can be assumed to be distributed around the parent point as a BND. Hence, denoting with Ω_c the set of users belonging to cluster c, we can express the Probability Mass Function (PMF) of $|\Omega_c|$ (*i.e.*, the cardinality of Ω_c) as:

$$f_{|\Omega_c|}(k;\lambda_{o_c}) = Pr(|\Omega_c| = k) = \frac{\lambda_{o_c}^k e^{-\lambda_{o_c}}}{k!}, \tag{1}$$

while the joint Probability Density Function (PDF) of the coordinates of user $u \in \Omega_c$ is represented by the BND PDF:

$$\begin{aligned} p_c(x_u, y_u) &= \frac{1}{2\pi\sigma_{x_c}\sigma_{y_c}\sqrt{1-\rho_c^2}} \exp\left[-\frac{z}{2(1-\rho_c^2)}\right], \\ z &= \frac{(x_u - x_c)^2}{\sigma_{x_c}^2} - \frac{2\rho(x_u - x_c)(y_u - y_c)}{\sigma_{x_c}\sigma_{y_c}} + \frac{(y_u - y_c)^2}{\sigma_{y_c}^2}, \end{aligned} \tag{2}$$

with (x_c, y_c) being the coordinates of cluster c's center, σ_{x_c} and σ_{y_c} representing the spatial size of cluster c over the x and y coordinates, respectively, ρ_c accounting for the rotation of the cluster and z being a support variable to improve readability. Considering all the clusters, the overall user distribution is a Gaussian Mixture Model (GMM) with $|\Pi|$ components, where Π is the set of parent points. The joint PDF of a user's coordinates can then be written as:

$$q(x_u, y_u) = \sum_{c=1}^{|\Pi|} w_c p_c(x_u, y_u), \tag{3}$$

where w_c is a weight that represents the number of offspring points generated by parent point c, *i.e.*:

$$w_c = \frac{|\Omega_c|}{\sum_{j=1}^{|\Pi|} |\Omega_k|}. \tag{4}$$

If $|\Pi|$ was also generated from a PPP, the overall CP would be a TCP as in [8]. However, there is no indication that this model would truthfully represent the distribution of inhabited centers. For this reason, a more suitable approach is to determine the empirical distribution of $|\Pi|$ from a population density dataset. Furthermore, the mentioned data can also be used to determine the statistics of the model parameters (e.g., σ_{x_c} and σ_{y_c}) over a specific coverage area.

III. Derivation of Population Statistics

We here present a procedure to extract statistical distributions for the proposed model from a population density dataset. It must be mentioned that using such dataset may lead to traffic underestimation when serving UTs with mobility in underpopulated areas, *e.g.*, for aeronautical or maritime use cases. Hence, further development may be needed to target such services, too. Furthermore, the proposed procedure can be adapted to process different data, e.g., spatial traffic distributions. Nonetheless, the accuracy of the extracted distribution with respect to the real user/traffic distribution depends on the

Algorithm 1 Dataset processing

Input: Geographical raster tile $\mathbf{T}$, number of sub-tiles N_{ST}
Output: Set of number of components per ST $\mathcal{N}$, set of covariance matrices per component $\mathcal{S}$, set of weights per component $\mathcal{V}$

Use case pre-processing
Generate $\mathbf{ST}_s$ for $s = 1, \ldots, N_{ST}$ s.t. $\bigcup_{s=1}^{N_{ST}} \mathbf{ST}_s = \mathbf{T}$ and $|\mathbf{ST}_s| = |\mathbf{T}|/N_{ST}$ for $s = 1, \ldots, N_{ST}$
Initialize $\mathcal{N} = \varnothing$, $\mathcal{S} = \varnothing$, $\mathcal{V} = \varnothing$
for $s := 1$ **to** N_{ST} **do**
  if Count('$x_{pop} < 0$', $x_{pop} \in \mathbf{ST}_s$) $\leq 0.1 \cdot |\mathbf{ST}_s|$ **then**
    Duplicate $\mathbf{ST}_s$ entries by their population count
    $L_s =$ HDBSCAN($\mathbf{ST}_s$)
    $[\mathbf{w}_s, \mathbf{\Sigma}_s] =$ GMM($\mathbf{ST}_s$, L_s)
    $\mathbf{v}_s \leftarrow \mathbf{w}_s \cdot |\mathbf{ST}_s|$
    $\mathcal{N} \leftarrow \mathcal{N} \bigcup$ Unique(L_s), $\mathcal{S} \leftarrow \mathcal{S} \bigcup \mathbf{\Sigma}_s$, $\mathcal{V} \leftarrow \mathcal{V} \bigcup \mathbf{v}_s$
  end if
end for

quality of the dataset. As a data source, we opted for the open database Global Human Settlement Layer (GHSL) by the Joint Research Centre of the European Commission (EC), mainly focusing on the GHS population grid multi-temporal dataset (GHS-POP R2023A) [11]. This dataset contains a geographical raster of the estimated population count in different periods and spatial resolutions. For this work, the year 2020 dataset with 100m resolution was used. Algorithm 1 describes the extraction of population distribution features. Focusing the study on central Europe, we applied the procedure on tile $\mathbf{T}$ = "R4_C19", a sub-raster of 10000x10000 elements. A pre-processing step adapts the data to the considered use case; without loss of generality, we first assume that the NTN users' distribution is similar to that of the population and skip this step. The algorithm splits $\mathbf{T}$ into $N_{ST} = 2500$ sub-tiles (STs), dropping those that contain water bodies for more than 10% of their surface (represented by a negative population count in each element). Each ST is then clustered using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), exploiting its ability to obtain an unspecified number of clusters with varying density [15]. From the generated labels, the algorithm produces the set of the number of clusters $\mathcal{N}$. The clustered data is then fed to a GMM clustering algorithm, obtaining the covariance matrix $\mathbf{\Sigma}$ (related to the cluster's geographical size) and weight w (related to the cluster's population count) of each bivariate Gaussian component. Merging data from different STs together, the algorithm outputs the set of covariance matrices $\mathcal{S}$ and the set of weights $\mathcal{V}$. The statistics of the number of parent points $|\Pi|$ (*i.e.*, the number of clusters) and of the weights are obtained through Kernel Density Estimation (KDE) using the Epanechnikov Kernel Function. The dip in the second PDF, reported in Figure 1, suggests that two classes of clusters may be present. Hence, defining a proper weights threshold v_{th}, we partition $\mathcal{S}$ and $\mathcal{V}$ into two subsets: clusters with $\mathbf{v}_s < v_{th}$ are said to be "small" ($\mathcal{S}_S$, $\mathcal{V}_S$), while the rest are called "large" ($\mathcal{S}_L$, $\mathcal{V}_L$). Hence, the distribution of the elements in $\mathcal{V}_S$

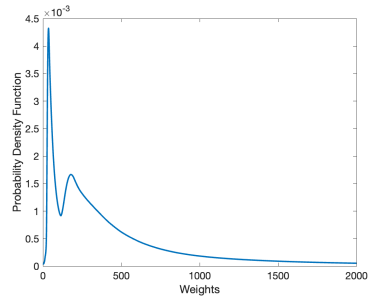

Fig. 1. PDF of the GMM components' weights.

and $\mathcal{V}_L$ is separately assessed, obtaining $g_{v_S}(v)$ and $g_{v_L}(v)$, respectively. Each element of $\mathcal{S}$ can be expressed as:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}, \tag{5}$$

where σ_x and σ_y represent the coordinates' standard deviation in the cluster over the x axis and y axis and ρ is the correlation between the two coordinates. A statistic of the clusters' geographical size can be obtained using KDE jointly on σ_x^2 and σ_y^2 (*i.e.*, irrespectively of the coordinate) for small and large clusters, resulting in $g_{\sigma_S^2}(\sigma^2)$ and $g_{\sigma_L^2}(\sigma^2)$. Then, the eccentricity of each cluster can be taken into account by computing the parameter e as follows:

$$e = \frac{\min(\sigma_x^2, \sigma_y^2)}{\max(\sigma_x^2, \sigma_y^2)}, \tag{6}$$

such that $0 \leq e \leq 1$. As for the weights and the variances, the distributions $g_{e_S}(e)$ and $g_{e_L}(e)$ can be obtained. Finally, as the correlation term ρ represents the cluster's rotation, its PDF is assumed to be uniform in [-1, 1] without loss of generality.

In this paper, we focus on users who are underserved or completely unserved by terrestrial networks, *i.e.*, on suburban and rural areas. Thus, the *use case pre-processing* step aims at setting zero population in GHS-POP cells corresponding to high-density clusters. These centers were identified processing the GHS-POP, GHS-BUILT-S [12] and GHS-LAND [13] datasets using the Degree of Urbanization definition: *"contiguous grid cells of 1 km^2 with a density of at least 1500 inhabitants per km^2 and a minimum population of 50000"* [14]. Figure 2 shows the empirical PMF of $|\Pi|$, $g_{|\Pi|}(|\Pi|)$, in the considered area, compared with a Poisson distribution with same average. As hypothesized in Section II, the goodness of fit is visibly poor, justifying the rejection of TCP as a model for the NTN users' distribution. As in the complete dataset, a dip in the GMM components' weights PDF was observed, leading to setting $v_{th} = 370$. Due to space constraints, we here report only the clusters' variance PDF $g_{\sigma^2}(\sigma^2)$ for small and large clusters (Figure 3). It must be mentioned that every plot shows different statistics for small and large clusters. In Figure 3, one can spot a higher probability density for small variances in small clusters than in large clusters. This suggests that some level of correlation between the clusters' variance and the clusters' weight is taken into account through the definition of small and large classes.

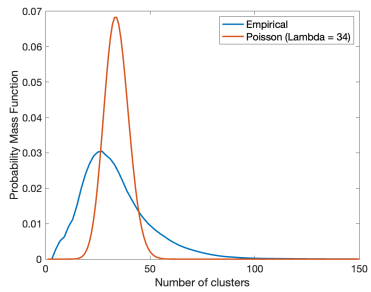

Fig. 2. Comparison between the empirical distribution of the number of clusters (blue line) and a Poisson distribution with equal mean (orange line).

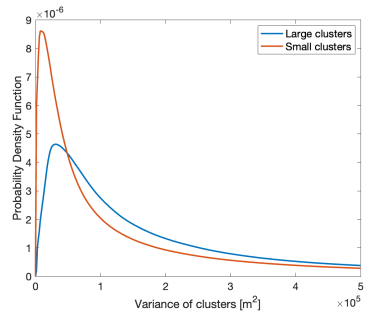

Fig. 3. Estimated PDFs of the clusters' variance.

TABLE I
NTCP MODEL PARAMETERS

Parameter	Source	Distribution or formula
$\lvert\Pi\rvert$	Empirical distribution	$\lvert\Pi\rvert \sim g_{\lvert\Pi\rvert}(\lvert\Pi\rvert)$
x_c	Theoretical distribution	$x_c \sim$ Uniform (x_{min}, x_{max})
y_c	Theoretical distribution	$y_c \sim$ Uniform (y_{min}, y_{max})
v_c	Empirical distribution	$v_c \sim \{g_{v_S}(v), g_{v_L}(v)\}$
$\lvert\Omega_c\rvert$	Theoretical distribution	$\lvert\Omega_c\rvert \sim$ Poisson $(\lambda_{UT} \cdot v_c)$
$\sigma^2_{x_c}$	Empirical distribution	$\sigma^2_{x_c} \sim \{g_{\sigma^2_S}(\sigma^2), g_{\sigma^2_L}(\sigma^2)\}$
e_c	Empirical distribution	$e_c \sim \{g_{e_S}(e), g_{e_L}(e)\}$
$\sigma^2_{y_c}$	Analytical formula	$\sigma^2_{y_c} = \sigma^2_{x_c} \cdot e_c$
ρ_c	Theoretical distribution	$\rho_c \sim$ Uniform $(-1, 1)$

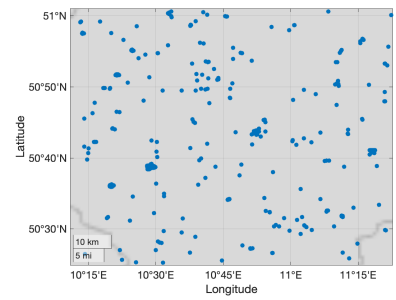

Fig. 4. Detail of a generated user distribution (each dot represents one UT).

With this data, the non-uniform distribution is fully described. To simulate users, a specific area delimited by $[x_{min}, x_{max}]$ and $[y_{min}, y_{max}]$ over the x and y coordinates is first fixed, naming its size A_{cov} and its average UT activity per inhabitant λ_{UT}. Then, the NTCP model parameters can be generated as in Table I. It must be noted that, in order to take into account the different area size with respect to the area of the STs A_{ST}, $|\Pi|$ has to be multiplied by the scaling factor $\frac{A_{cov}}{A_{ST}}$. Finally, the NTCP can be obtained using Equations 3 and 2. A detail of a generated user distribution is reported in Figure 4.

IV. Application of NTCP to User Scheduling

To exemplify the shortcomings of the uniformity assumption, we present an empirical analysis of user scheduling algorithms in NTN HTS systems. The results of the evaluation are reported in Section V.

User scheduling is the technique through which a BS can allocate resources to UTs. Two main categories of scheduling algorithms can be identified: user selection and user grouping. Schedulers of the first type select a single user group that, if scheduled, would optimize a chosen metric, *e.g.*, the SR. However, the propagation channel in NTN is mostly characterized by slow fading effects due to atmospheric events, *e.g.*, rain and clouds; hence, a user selection algorithm would tend to always schedule the same set of UTs. While this may indeed maximize the SR of the system, it would leave the majority of the users unserved. Instead, user grouping algorithms generate multiple user groups such that every UT is part of at least one group. Then, each user group is scheduled for a specific time frame over a scheduling window, ensuring fairness.

A. System Model

The system model we here consider is adopted from the work we already presented in [16]. For the sake of brevity, we will only report the most relevant information. For this analysis, we consider a multi-beam LEO satellite equipped with a Uniform Planar Array (UPA) made of N_F feeds, capable of forming beams through feed space digital beamforming (BF) techniques. The satellite provides downlink connectivity to a total of K_T users, each communicating through a Very-Small-Aperture Terminal. Given $K_T \gg N_F$, the scheduler identifies P disjoint groups of UTs, with K_p denoting group p's size. Then, the scheduling window is split into P frames, and the symbols belonging to users in the same group are transmitted during the corresponding time frame using Spatial Division Multiplexing (SDM). Hence, each UT in the scheduled group can receive its symbols during the same time frame, over the same subcarrier, and using the same polarization. To achieve this, the gNB uses the most recent channel coefficients to compute the group's minimum mean square error BF matrix $\mathbf{W}_p$ with Sum Power Constraint (SPC) normalization. The vector of channel coefficients between UT i and the UPA can be expressed as:

$$\mathbf{h}_i = G_i^{(rx)} \frac{\lambda}{4\pi d_i} \sqrt{\frac{L_i}{\kappa B T_i}} e^{-j\frac{2\pi}{\lambda} d_i} \mathbf{a}(\vartheta_i, \varphi_i), \tag{7}$$

where $G_i^{(rx)}$ is the i-th UT's reception gain, λ is the carrier wavelength, B is the channel bandwidth, d_i represents the slant range, L_i includes additional losses as in [16], T_i is the receiver's noise temperature, and $\mathbf{a}(\vartheta_i, \varphi_i)$ is the array factor

of the UPA in the direction of the i-th user. It is worth stressing that the channel coefficients are normalized with respect to the noise power. The Signal to Interference plus Noise Ratio (SINR) experienced by the i-th UT in group p can then be written as:

$$\text{SINR}_i^{(p)} = \frac{\left\|\mathbf{h}_i \mathbf{w}_i^{(p)}\right\|^2}{1 + \sum_{k=1,k\neq i}^{K_p} \left\|\mathbf{h}_k \mathbf{w}_i^{(p)}\right\|^2}. \tag{8}$$

where $\mathbf{w}_i^{(p)}$ is user i's vector of BF coefficients. Finally, we can define the SR in group p as:

$$\Gamma_p = B \sum_{i=1}^{K_p} \log_2(1 + \text{SINR}_i^{(p)}), \tag{9}$$

and the throughput experienced by the i-th user in group p as:

$$R_i^{(p)} = \frac{B}{P} \log_2(1 + \text{SINR}_i^{(p)}). \tag{10}$$

B. Scheduling Algorithms

In this work, two state-of-the-art schedulers are considered. For both of the algorithms, a Coefficient of Correlation (CoC) matrix $\mathbf{\Psi}$ is computed as:

$$[\mathbf{\Psi}]_{i,j} = \frac{\left|\mathbf{h}_i \mathbf{h}_j^H\right|}{\|\mathbf{h}_i\| \, \|\mathbf{h}_j\|}. \tag{11}$$

Then, each scheduler involves different processing:

1) Graph-based Max Clique Scheduler [16]: A graph is constructed from $\mathbf{\Psi}$, where vertices represent users and edges are based on the dissimilarity between the corresponding users' channels. The user clustering strategy is based on a greedy procedure in which the graph is iteratively searched for its maximum clique and pruned with a constant threshold. The version assessed in this work is an improved algorithm proposed in [18], which automatically optimizes the pruning threshold at each iteration to reach a target graph density ϵ_G. In order to maximize the system performance, a heuristic optimization of ϵ_G is performed.
2) MADOC Scheduler [17]: The Multiple Antennas Downlink Orthogonal Clustering (MADOC) algorithm uses the CoC matrix as a metric to find ϵ-orthogonal user groups, *i.e.*, groups of users in which the CoC does not exceed a predefined threshold ϵ_M. As for the first scheduler, the value of this parameter is optimized to maximize the overall system capacity.

V. Results

In this section, we analyze the scheduling performance of an NTN LEO satellite serving NTCP-distributed users and compare the results with the system performance under the uniform distribution assumption. The assessment was carried out by means of Montecarlo simulation on MATLAB with 10^5 iterations. The set of parameters was chosen as in [16], except for the carrier frequency and channel bandwidth, which are here set to 20 GHz ($\lambda \simeq 1.5cm$) and 400 MHz, respectively.

TABLE II
Schedulers' optimal parameters and results

	Scheduler	Uniform	NTCP
Optimal ϵ	Max Clique	0.945	0.950
	MADOC	0.480	0.855
TP (10%-ile)	Max Clique	42.42 Mbps	16.72 Mbps
	MADOC	43.61 Mbps	25.53 Mbps
SR (10%-ile)	Max Clique	112.22 Gbps	11.67 Gbps
	MADOC	131.38 Gbps	51.01 Gbps

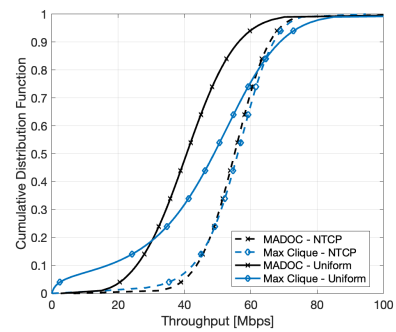

Fig. 5. CDF of the user throughput, comparing the proposed model (solid line) with the uniform distribution (dashed line).

We considered a coverage area with an average UT density of 0.05 UTs/km^2. The propagation model was based on the Line of Sight model reported in 3GPP TR 38.811 [19].

First, we determined the values of the schedulers' ϵ parameters that maximize the SR (Table II). One can notice that MADOC's optimal parameter for the uniform distribution would not be suitable in a real deployment. Specifically, ϵ_M in NTCP is significantly higher than its uniform counterpart, suggesting that more users need to be scheduled in the same group to maximize the SR in NTCP. However, this also suggests that more interference is to be expected, leading to users with low SINR and TP.

Figure 5 reports the TP CDFs. The graph shows that 10% of the non-uniformly distributed UTs can expect a maximum perceived throughput of 25.53 Mbps with MADOC. However, when users are assumed to be deployed uniformly, the same fraction of UTs experience a TP of at most 43.61 Mbps. Similarly, the low 10%-ile of the NTCP Max Clique CDF is 16.72 Mbps, compared to 42.42 Mbps under uniform distribution. It must be noted that, despite the large overestimation of the TP 10%-iles that comes with the uniformity assumption (70% with MADOC and 154% with Max Clique), the NTCP reveals that the number of UTs with high throughput (TP > 70 Mbps) scheduled using Max Clique is underestimated. In general, it is clear that tests on uniformly distributed UTs lead to an overestimation of the system fairness. Figure 6 shows an even worse trend for the SR. Specifically, while the uniform distribution promises a sum-rate capacity with MADOC of at most 131.38 Gbps in 10% of the clusters, the figure drops to 51.01 Gbps in NTCP. The worst performance difference is with Max Clique, an overestimation of the 10%-iles of SR of 862%, from 112.22 Gbps to just 11.67 Gbps. Table II

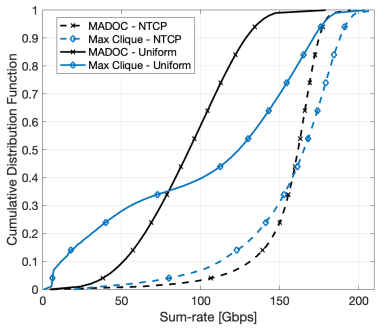

Fig. 6. CDF of the per-cluster sum-rate capacity, comparing the proposed model (solid line) with the uniform distribution (dashed line).

reports the mentioned 10%-iles for a direct comparison. The results prove that the uniform assumption is far from being reliable for the evaluation of scheduling in NTN and, more in general, of SatCom procedures, when the user density is large enough to reveal the UTs' clustered nature. This is due to the large correlation that non-uniform distributions introduce between users' channel vectors. With high correlations and large optimal ϵ parameters, the amount of interference strongly limits the system performance. Furthermore, while one may consider the uniform distribution as a benchmark to evaluate algorithms for NTN procedures, it must be noted that their relative performance may not necessarily be representative of the real-world difference. Indeed, Figure 5 shows that the two schedulers could be considered equivalent in terms of TP if assessed under uniform distribution, but the NTCP reveals a clear advantage in using MADOC to improve the TP 10%-iles and the fairness. Finally, this study motivates the need for algorithms to tackle challenging UT distributions.

VI. Conclusions

In this work, we presented a procedure to simulate user distributions in NTN systems based on CPs. The model represents more accurately the real population distribution for a considered use case, overcoming the approximations introduced with the more often used uniform distribution. To tailor the model to the statistics of the coverage area, we extracted population statistics by means of clustering over population count datasets. Empirical results show that the uniform assumption may severely overestimate critical NTN KPIs. In the case of user scheduling, the SR 10%-iles may be more than an order of magnitude smaller than what is estimated assuming uniformly distributed users. Hence, we conclude that it is of primary importance to assess the performance of NTN algorithms under NTCPs. Future works may further extend the analysis to several SatCom procedures, such as Random Access and interference management in NTN, and propose novel algorithms tailored for NTCPs.

VII. Acknowledgments

This work has been funded by the 6G-NTN project, which received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101096479. The views expressed are those of the authors and do not necessarily represent the project. The Commission is not liable for any use that may be made of any of the information contained therein.

References

[1] Satellite Industry Association (SIA) prepared by Bryce Tech, "2023 State of the Satellite Industry Report", 2023. https://www.sia.org/
[2] "38.108 - NR; Satellite Access Node radio transmission and reception", June 2022.
[3] SpaceX, "SECOND GENERATION STARLINK SATELLITES", Feb. 2023. https://www.starlink.com/resources
[4] P. Angeletti and R. de Gaudenzi, "Optimizing Massive MIMO Design for Non Uniform Traffic in Broadband Telecommunication Satellites Networks," in IEEE Access, vol. 11, pp. 113493-113513, 2023.
[5] H. Al-Hraishawi, E. Lagunas and S. Chatzinotas, "Traffic Simulator for Multibeam Satellite Communication Systems," 2020 10th Advanced Satellite Multimedia Systems Conference and the 16th Signal Processing for Space Communications Workshop (ASMS/SPSC), Graz, Austria, 2020, pp. 1-8.
[6] S. Xia, Q. Jiang, C. Zou and G. Li, "Beam Coverage Comparison of LEO Satellite Systems Based on User Diversification," in IEEE Access, vol. 7, pp. 181656-181667, 2019.
[7] X. Wu, W. T. Toor, H. Jin and B. C. Jung, "On CDF-based scheduling with non-uniform user distribution in multi-cell networks," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2017, pp. 388-392.
[8] M. Shi, J. Li, X. Gao, K. Yang, D. Niyato and Z. Han, "Load Distribution in Poisson Networks With Spatially Coupled Users," in IEEE Transactions on Vehicular Technology, vol. 72, no. 1, pp. 1331-1336, Jan. 2023.
[9] H. ElSawy, A. Sultan-Salem, M. -S. Alouini and M. Z. Win, "Modeling and Analysis of Cellular Networks Using Stochastic Geometry: A Tutorial," in IEEE Communications Surveys & Tutorials, vol. 19, no. 1, pp. 167-203, Firstquarter 2017.
[10] P. J. Diggle, "The Statistical Analysis of Spatial Point Patterns.", London: Edward Arnold, 2003.
[11] M. Schiavina, S. Freire, A. Carioli, K. MacManus, "GHS-POP R2022A - GHS population grid multitemporal (1975-2030)", European Commission, Joint Research Centre (JRC), 2023.
[12] M. Pesaresi, P. Politis, "GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030)", European Commission, Joint Research Centre (JRC), 2023.
[13] M. Pesaresi, P. Politis, "GHS-LAND R2022A - Land fraction as derived from Sentinel2 image composite (2018) and OSM data", European Commission, Joint Research Centre (JRC), 2022.
[14] Dijkstra, Lewis & Poelman, Hugo. (2014). A harmonised definition of cities and rural areas: the new degree of urbanisation.
[15] R. Campello, D. Moulavi, J. Sander, "Density-based clustering based on hierarchical density estimates", in: J. Pei, V.S. Tseng, L. Cao, H. Motoda, G. Xu, "Advances in Knowledge Discovery and Data Mining", PAKDD 2013, "Lecture Notes in Computer Science()", vol 7819. Springer, Berlin, Heidelberg.
[16] D. G. Riviello, B. Ahmad, A. Guidotti and A. Vanelli-Coralli, "Joint Graph-based User Scheduling and Beamforming in LEO-MIMO Satellite Communication Systems," 2022 11th Advanced Satellite Multimedia Systems Conference and the 17th Signal Processing for Space Communications Workshop (ASMS/SPSC), Graz, Austria, 2022, pp. 1-8.
[17] K. -U. Storek and A. Knopp, "Fair User Grouping for Multibeam Satellites with MU-MIMO Precoding," GLOBECOM 2017 - 2017 IEEE Global Communications Conference, Singapore, 2017, pp. 1-7.
[18] B. Ahmad, D. G. Riviello, A. Guidotti and A. Vanelli-Coralli, "Improved Graph-Based User Scheduling For Sum-Rate Maximization in LEO-NTN Systems," 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), Rhodes Island, Greece, 2023, pp. 1-5.
[19] "38.811 - Study on New Radio (NR) to support non-terrestrial networks", 2020.