

University of Western Ontario

Master's Project

Determination of the Observation Parameters for *Colibri*, a KBO Occultation Survey

Student: RICHARD BLOCH

Supervisor: STANIMIR METCHEV

November 19, 2018

Contents

1	Introduction	1
1.1	The Kuiper Belt	2
1.2	The method of Serendipitous Stellar Occultation (SSO)	4
1.3	The <i>Colibri</i> array	6
2	Finding the fields of observation	8
2.1	Defining the parameters of a field search sample	8
2.1.1	KBO sky surface density	8
2.1.2	Stellar sky surface density	9
2.1.3	Ranking of candidate fields	10
2.2	<i>Gaia</i> : producing a field search sample	10
2.2.1	The <i>Gaia</i> mission	11
2.2.2	<i>Gaia</i> Data Release 1 (DR1) and Data Release 2 (DR2)	11
2.2.3	Reliability of the <i>Gaia</i> DR2 radius estimates	12
2.3	An algorithm for the selection of optimal fields	15
2.3.1	Nearest-neighbour searching	15
2.3.2	Binary search trees and the optimized k-d tree	17
2.3.3	The field search algorithm	18
2.4	Performing the search	20
2.4.1	Errors and reliability of results	20
3	Simulating occultation events	26
3.1	Modelling low-diameter KBOs	26
3.1.1	The power law q and its break diameter	26

3.1.2	KBO sky surface density	27
3.1.3	KBO size distribution	29
3.2	Event detection algorithm	31
3.2.1	Generating a KBO path	32
3.2.2	The coarse search and spatial indexing	33
3.2.3	The high resolution followup and the Shapely module	35
3.2.4	Likelihood of detection	37
3.3	Simulating the observations	39
3.3.1	Choosing parameters: KBO diameter limits D_k and D_0	39
3.3.2	Choosing parameters: power law indices	41
3.3.3	Running the initial simulation	43
4	Results and Discussion	44
4.1	<i>Colibri</i> fields of observation	44
4.1.1	Results of <i>Colibri</i> field search	44
4.1.2	Discussion of the importance of <i>Colibri</i> 's limiting magnitude	44
4.1.3	Evaluating the <i>Colibri</i> fields	46
4.2	Simulation of occultation events	49
4.2.1	Core results	49
4.2.2	Discussion of the stellar angular size cutoff and detectable KBOs	50
4.3	Synthesized results	53
4.3.1	Extinction and cloud cover effects	53
4.3.2	Errors in detection rate	55
4.3.3	Estimated event detection rates	57
4.3.4	Future work	57

5 Conclusion	59
A Continuity and Transition	60
A.1 Field search	61
A.1.1 Storage	61
A.1.2 Overview	61
A.1.3 Dependencies	62
A.1.4 File descriptions	63
A.1.4.1 <code>findfields.py</code>	63
A.1.4.2 <code>utils.py</code>	64
A.1.4.3 <code>datasort.py</code>	64
A.1.4.4 <code>filehandling.py</code>	64
A.2 Occultation simulation	64
A.2.1 Storage	64
A.2.2 Overview	64
A.2.3 Dependencies	65
A.2.4 File descriptions	66
A.2.4.1 <code>runmodel.py</code>	66
A.2.4.2 <code>collisions.py</code>	67
A.2.4.3 <code>fieldgen.py</code>	67
A.2.4.4 <code>popgen.py</code>	68
A.2.4.5 <code>kbogen.py</code>	68
A.2.4.6 <code>colibri_sim_matching.py</code>	68
A.2.4.7 <code>fresnelModeler.py</code>	68

List of Figures

1	Light curve of a detected KBO occultation from an SSO survey	5
2	Simulated diffracted occultation shadow produced by a KBO	6
3	<i>Gaia</i> stellar population by G magnitude	12
4	Comparison of <i>Gaia</i> DR2 T_{eff} values against RAVE DR5 values	14
5	Sky density map of candidate program stars in the <i>Gaia</i> DR2	14
6	Illustrative example of a binary search tree	17
7	Locations of the 24 candidate fields on the sky	21
8	Stellar radii of stars in the <i>Gaia</i> field sample	21
9	HR diagram of stars in the <i>Gaia</i> field sample	22
10	Stellar angular size distribution in an optimal field	24
11	Candidate <i>Colibri</i> field population dependence on stellar angular diameter	25
12	KBO diameter assignment probability as a function of diameter	30
13	KBO diameters of simulated population	31
14	Illustrative example of the R-tree data structure	34
15	Effects of discretization on the detection of occultation events	35
16	Example of occultation geometry	36
17	Model light curve of occulting KBO for different star sizes	37
18	Model sampled light curve of occulting KBO for differing sampling rates	38
19	Kernel fitting of a noisy simulated occultation light curve	39
20	Small KBO population dependence on the value of adopted break diameter D_k . . .	40
21	Small KBO population dependence on the value of size power law index q	41
22	Small KBO population dependence on the value of radial power law index c	42
23	Stellar field populations ($\leq 100 \mu\text{as}$) as a function of limiting G magnitude.	45

24	Locations of the <i>Colibri</i> campaign fields on the sky	48
25	KBO diameter versus angular size of occulted star with detection likelihoods	51
26	Stellar angular sizes for detectable events	52
27	KBO diameter and stellar sizes for annual detected events	56

List of Tables

1	Top five <i>Colibri</i> field candidates from the search algorithm	20
2	Example of field coordinate-switching to preserve rank	24
3	Final 24 <i>Colibri</i> field candidates from the search algorithm	44
4	Final field selections for the <i>Colibri</i> observing campaign	47
5	Model and simulation parameters for the initial simulation.	49
6	Event detection results for final observing fields	49
7	Altitude distribution per field	54
8	Estimated extinction contribution per field	54
9	Estimated total clear nights per field	54
10	Final detection rate estimates	57
11	Default parameters for the field search algorithm	62
12	Default parameters for the occultation simulation	66

1 1 Introduction

2 The Kuiper Belt is a disk of small icy bodies beyond the orbit of Neptune, left over from the solar
3 nebula from which the planets were formed. Until the 1990's there was no direct observational
4 evidence for the existence of this disk. While Pluto (the largest member of the Kuiper Belt) was
5 discovered in 1930, it was considered its own planet in an empty region of space for decades. In
6 1992 Jewitt and Luu (1993) detected a new trans-Neptunian object which they suggested could be
7 the first of many such objects in the hitherto only hypothesized Kuiper Belt. In the ensuing years,
8 many more objects have been detected and confirmed to lie within the Kuiper Belt.

9 While the Kuiper Belt has a number of large (diameter of \sim 1000 km) members, the majority
10 of its population lies below \sim 25 km diameters (Kenyon, Bromley, O'brien, & Davis, 2007). Kuiper
11 Belt Objects (KBOs) of such a small size are too distant and dim to be directly observed from
12 Earth, either by ground or space-based telescopes (Schlichting et al., 2009). However, it has been
13 surmised (Bailey, 1976; Brown & Webster, 1997) that KBOs can occasionally pass the line-of-sight
14 to distant stars and produce an observable effect on the star's detectable light as seen from Earth.
15 The advantage to this technique is the KBO does not need to be bright to occult a background star.
16 Thus this technique, known as the method of Serendipitous Stellar Occultation (SSO) can be a tool
17 to probe the small KBO population.

18 The SSO method is already being used in both archival searches (e.g. Schlichting et al., 2009)
19 and in wide-field continuous monitoring surveys (e.g. Lehner et al., 2012) and is a burgeoning field
20 in the study of the Kuiper Belt. This report contains details of work done in preparation for the
21 upcoming wide-field continuous monitoring survey *Colibri* run by Dr. Stanimir Metchev.

22 This section contains an outline of the Kuiper Belt, the SSO method, and the *Colibri* project as
23 well as the purpose of this work within that project. Sections 2 and 3 outline the two major con-
24 tributions to *Colibri*: finding the fields of observation, and simulating those observations. Section 4
25 presents the results of this work, with complementary analysis and discussion of key results. Sec-
26 tion 5 presents the conclusion, and the appendix provides a transition document for the continuity
27 of the project.

28 **1.1 The Kuiper Belt**

29 The Kuiper Belt is a circumsolar disk of small icy bodies beyond the orbit of Neptune, between
30 roughly 30 and 50 AU. These objects are leftovers from the formation of the solar system, having
31 never accreted to produce a planet. This is the region in which the dwarf planet Pluto lies; it was
32 the first observed member of the Kuiper Belt, and remains among its largest and most massive.
33 Although it was at first thought to be a planet of its own, the increasing number of detections of
34 other KBOs in its orbital vicinity indicated Pluto was one of many members of a much larger feature
35 of the solar system.

36 The presence of an old circumstellar debris disk around a star of the age of the Sun is not
37 unexpected. Nearly all newly-formed stars are surrounded by larger, denser, and optically thick
38 disks which decrease in mass as the host star ages (Kenyon et al., 2007). The Kuiper Belt has many
39 similarities with other old debris disks around other stars (e.g. Greaves, Wyatt, Holland, & Dent,
40 2004); a close inspection of the Kuiper Belt can thus lead to a greater understanding of debris disks
41 in general.

42 The Kuiper Belt is not of uniform distribution. KBOs have orbits that are clustered around
43 different heliocentric distances due to the influence of the nearby giant planets (Luu & Jewitt,
44 2002). Each clustered group has its own unique dynamical properties, serving to provide three main
45 KBO groups: classical KBOs, resonant KBOs, and scattered KBOs (Luu & Jewitt, 2002). Classical
46 KBOs are thought to comprise the majority of the KBO population, lying between 42 and 48 AU
47 and having low eccentricities ($e \sim 0.1$). They lie outside mean motion resonances with Neptune,
48 and have orbits relatively unperturbed by the motion of Neptune on timescales of Gyr (Luu &
49 Jewitt, 2002). Resonant KBOs lie at the 4:3, 3:2, and 2:1 mean motion resonances with Neptune.
50 Due to gravitational perturbation, these KBOs tend to have on average higher eccentricities and
51 inclinations than classical KBOs (Luu & Jewitt, 2002). Scattered KBOs have been scattered through
52 gravitational interactions, and have high eccentricities and inclinations.

53 Classical KBOs represent the majority of KBOs, and due to their low inclination represent the
54 majority of KBOs in the ecliptic plane (Luu & Jewitt, 2002). With stable orbits over timescales on
55 the order of the age of the solar system, information about the formation and evolution of the solar
56 system is embedded in the classical KBO population.

57 Of particular interest is the process of planet formation, which was completed ~ 4.5 Gyr ago in

58 the inner solar system and its traces buried under billions of years of that dynamical environment.
59 Planet formation is thought to occur early in the life of a star during the transitory phase when
60 its optically thick circumstellar disk becomes optically thin (Kenyon et al., 2007). Since planet
61 formation was never completed in the Kuiper Belt, KBOs in the ecliptic contain information about
62 planet formation that was lost elsewhere in the solar system.

63 This process was likely halted because the velocity dispersion of planetesimals in this region grew
64 beyond the point where they could rapidly accrete material (Schlichting et al., 2012). While large
65 objects could still accrete smaller ones, collisional dynamics began to grind large objects into smaller
66 ones at a faster rate (Kenyon et al., 2007), preventing planetary accretion in the region and 'freezing'
67 the state of planetesimal growth in the large KBO population (Schlichting et al., 2012).

68 The size distribution of large KBOs follows a power law of index $q \approx 4.5$ which is well described
69 by coagulation models of planet formation and evolution (Schlichting et al., 2012). However, there
70 is ample evidence to suggest that q changes for KBOs smaller than some 'break' diameter to a
71 shallower slope, attributable to collisional dynamics affecting only smaller KBOs (Bernstein et al.,
72 2004; Schlichting et al., 2012; Schlichting et al., 2009). The value of the break diameter moves toward
73 larger diameters with time, and so carries with it the ability to constrain the collisional history of
74 the Kuiper Belt (Schlichting et al., 2012).

75 Coagulation models which successfully describe the size distribution of large KBOs also predict
76 wildly different slopes for small KBOs at various break diameters. Observational data of the small
77 KBO regime is required to further constrain these models. Indeed, information about the slope of q
78 below the break diameter can do more than inform models of planet formation: its particular value
79 is influenced by the material composition of the small KBO population. If small KBOs are held
80 together by their material strength, their size distribution should follow a power law index of $q = 3.5$
81 (Schlichting et al., 2012). If they are instead mainly held together by gravity, one would expect
82 $q \approx 3$ (Pan & Sari, 2005). Knowledge of the material strength and composition of small KBOs is
83 fundamental to understanding the collisional processes of the Kuiper Belt, and by extension similar
84 debris disks in other star systems. Therefore, both the diameter at which q breaks and the value of
85 q itself can constrain the collisional history and reveal the material properties of small KBOs.

86 Unfortunately due to their size and distance, observational data for KBOs with diameters of $\lesssim 20$
87 km cannot be acquired through conventional imaging (Schlichting et al., 2009). The deepest direct
88 detection of a KBO was performed with the Hubble Space Telescope at $m_R = 28.3$ (corresponding

89 to a diameter of ≈ 25 km) (Bernstein et al., 2004).

90 **1.2 The method of Serendipitous Stellar Occultation (SSO)**

91 KBOs too small for conventional imaging are ideal candidates for observation by serendipitous stellar
92 occultation, or the SSO method. As the name suggests, by random chance a KBO may cross the
93 line-of-sight to a distant star and produce an observable dip in the star's light output. Continuous
94 monitoring of the star at sufficiently high signal-to-noise (SNR) would thus facilitate the potential
95 discovery of KBOs. Since the star's light and not the reflected KBO light is being monitored, the
96 SSO method allows for the detection of KBOs too dim for direct observation. Thus, the SSO method
97 provides a useful tool for investigating the population of small diameter KBOs well below the current
98 ~ 20 km-diameter limit.

99 Attempts have been made to implement this method to detect occultations. Schlichting et al.
100 (2009) analyzed archival HST data and found a single likely occultation event corresponding to a
101 KBO of ~ 1 km diameter, and later expanded that analysis using the HST guide data to include
102 another KBO detection at about the same diameter (Schlichting et al., 2012). Other detections of
103 KBOs have been claimed by Chang et al. (2006) and Roques et al. (2006). Subsequent reanalysis
104 of the Chang et al. (2006) data and criticism of its statistical interpretation have put these detec-
105 tions in doubt (Bickerton, Kavelaars, & Welch, 2008; T. A. Jones, Levine, Morgan, & Rappaport,
106 2006). Additionally, Bickerton et al. (2008) argues that while the Roques and Moncuquet (2000)
107 events match expected event behaviour they are not necessarily statistically significant. Nonetheless
108 Schlichting et al. (2009, 2012) demonstrates the principle and outcomes are possible.

109 Aside from archival detections, dedicated wide-field surveys are becoming a new front in observa-
110 tions of the Kuiper Belt, the Taiwanese-American Occultation Survey (TAOS) being the prominent
111 example. The TAOS project was specifically designed to identify ~ 1 km-diameter objects beyond the
112 orbit of Neptune, and to measure the size distribution of KBOs with diameters $0.5 \text{ km} < D < 30 \text{ km}$
113 (Zhang et al., 2013). TAOS began operation in February 2005 with three 50 cm telescopes located
114 at the Lulin Observatory in Taiwan, and has since expanded to include another telescope (Zhang
115 et al., 2013). After seven years of observation it found no occultation events (Zhang et al., 2013),
116 although this was attributed to the lower-than-expected event rate (Lehner et al., 2012) and the low
117 (5 Hz) cadence of the camera systems. The TAOS survey is being followed by a second iteration,

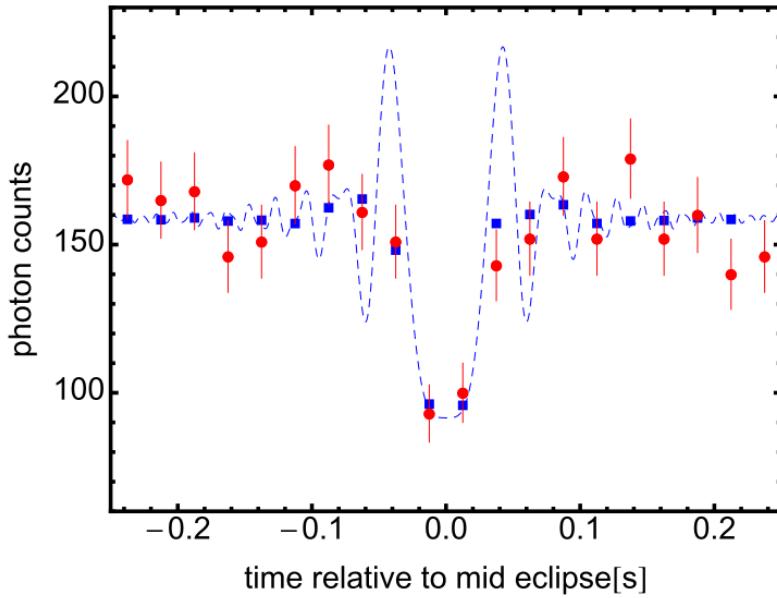


Figure 1: Occultation profile fit to light curve of a reported occultation event consistent with a ~ 1 km diameter KBO. From Schlichting et al. (2012, Fig. 11).

118 TAOS II, which will operate from 1.3 m telescopes at the Observatorio Astronómico Nacional at
 119 San Pedro M  rtir in M  xico (Lehner et al., 2012).

120 A major challenge for any continuous monitoring SSO survey is the short duration of an occul-
 121 tation event. With a relative observing velocity of ~ 30 km/s at solar opposition, typical events will
 122 have a duration of 100-200 ms (Schlichting et al., 2009; Zhang et al., 2013). In order to sufficiently
 123 sample the stellar light curve throughout the event duration, a survey will ideally operate its cameras
 124 at ≥ 40 Hz (Bickerton et al., 2008). The TAOS cameras were operating at 5 Hz, and TAOS II will
 125 operate at 20 Hz (Lehner et al., 2012); the tradeoff to operate at a higher cadence is a decrease in
 126 SNR, and poor SNR will decrease the likelihood an occultation event can be sufficiently separated
 127 from background noise.

128 Although these events are of a short duration, it is possible for an event to produce a diffraction
 129 effect which extends the observed duration of the event at the expense of the depth of the stellar
 130 light curve (Roques et al., 2006). This effect occurs strongest when the diameter of the occulting
 131 KBO is smaller than the so-called Fresnel scale:

$$F = \sqrt{\lambda d / 2} \quad (1)$$

132 Where d is the distance to the KBO and λ the wavelength of light of observation. This diffraction
 133 effect is known as Fresnel diffraction, a type of near-field diffraction that occurs when the objects
 134 are much larger than λ but much smaller than d (Roques & Moncuquet, 2000). Thus the effect is
 135 most pronounced both when the KBO diameter is $\lesssim F$ and when the angular size of the occulted
 136 star is smaller than the angular size of the occulting KBO.

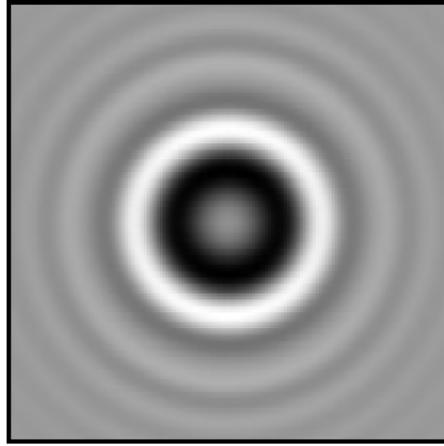


Figure 2: Simulated diffraction occultation shadow for a \sim 650 m-diameter KBO at 40 AU. From Fig. 11, Bickerton, Kavelaars, and Welch (2008).

137 For a survey operating in the R-band and an occulter at 40 AU, F corresponds to a KBO diameter
 138 of 1.3 km and an angular size of \sim 46 μ as. This means stars much larger than this angular diameter
 139 are expected to be poor sources for detecting \sim 1 km-wide KBOs.

140 1.3 The *Colibri* array

141 The *Colibri* array is a dedicated fast-photometry wide-field survey that will continuously moni-
 142 tor pre-selected stellar fields for KBO occultations, with particular emphasis on the detection of
 143 diffracted events and thus on KBOs of diameters $\lesssim 1$ km. It is run by Dr. Stanimir Metchev and
 144 is under construction at the Elginfield Observatory site 25 km north of London, ON. The array
 145 will consist of three 50-cm f/3 telescopes with a field-of-view of $1.48^\circ \times 1.48^\circ$ each equipped with
 146 high-sensitivity CMOS cameras operating in the R-band at a cadence of 40 Hz.

147 These specifications are primarily designed to overcome the issues related to short event dura-
 148 tions: the fast optics will facilitate short exposure times, since Bickerton, Welch, and Kavelaars
 149 (2009) have established that 40 Hz is the critical (Nyquist) sampling rate to detect KBOs, given the

150 relative speeds involved and the effects of sampling on the stellar light curve. The telescope place-
151 ment also provides advantages; for example, physical separation of the telescopes will provide a time
152 series of a given occultation event. This time series can be used in real-time analysis to reduce the
153 likelihood of false detections by avoiding correlated atmospheric effects across all telescopes (Pass,
154 Metchev, Brown, & Beauchemin, 2018).

155 This report details two main tasks completed for the *Colibri* project: the selection of stellar fields
156 for continuous monitoring, and a simulation of the observation of those fields.

157 Section 2 outlines the field selection process. Because the event rate is expected to be low and
158 random, it is important that *Colibri* monitor stellar fields that have an above average likelihood to
159 produce events. As that section will show, the choice of field can dramatically affect the detection
160 likelihood, demonstrating the crucial nature of an optimal choice.

161 Section 3 outlines the simulation of observations by *Colibri*. With the survey design work of
162 Bickerton et al. (2009) and previous work for a *Colibri* detection algorithm by Pass et al. (2018),
163 the simulation can reconstruct occultation events as seen by a virtual *Colibri*. As that section will
164 show, the information retrieved can be used to probe the parameter space of the array to elucidate
165 the capabilities and limitations of the project.

166 **2 Finding the fields of observation**

167 This section deals with the selection of observing fields for the *Colibri* array. Section 2.1 characterizes
168 an ideal observing field; section 2.2 outlines the data examined; section 2.3 details the algorithm
169 built to select the fields; and section 2.4 summarizes the results of the target field search.

170 **2.1 Defining the parameters of a field search sample**

171 The SSO method relies on random chance to detect an occultation event. The choice of observing
172 field can nonetheless have a great impact on the likelihood of detection, due to the non-uniform
173 distribution of stars and KBOs across the sky. Statistically, the greatest likelihood of detection will
174 occur in regions where both the KBO and stellar sky surface densities are large.

175 To estimate where a large KBO sky surface density might be found, it is useful to consider
176 the KBO populations outlined in section 1.1. Luu and Jewitt (2002) estimate the classical KBO
177 population to represent roughly 44% of the total KBO population. They also estimate the scattered
178 KBO population to represent roughly 35% of the total KBO population. These are the largest KBO
179 populations, and provide plenty of material for serendipitous stellar occultations.

180 However, despite the similarity in population size the classical KBO population will prove to be
181 a better population to consider for targeting. This is because scattered KBOs span a much wider
182 range of inclinations than classical KBOs, which are restricted to within 20° of the ecliptic plane
183 (Luu & Jewitt, 2002). As discussed in section 1.1, classical KBOs are also found mainly between 42
184 and 48 AU, while scattered KBOs have higher eccentricities and distances that are more difficult to
185 constrain. This suggests the classical KBO population is a well-constrained KBO population already
186 localized by ecliptic latitude, ideal attributes for a reliable KBO sample to have.

187 **2.1.1 KBO sky surface density**

188 The region of sky can be constrained even further on the basis of KBO ecliptic latitudes. The ecliptic
189 distribution of classical KBOs is not well constrained. However, sufficient observation and analysis
190 of KBO distributions (Bernstein et al., 2004; Elliot et al., 2005) have been made to justify an ecliptic
191 latitude boundary of $-5^\circ \lesssim b \lesssim 5^\circ$ between the classical KBO and scattered KBO populations. This
192 is in part because most KBOs in this latitude range are part of a dense, cold population composed

193 primarily of classical members (Elliot et al., 2005).

194 Since there is no reason to expect classical KBOs to have a preferential ecliptic longitude clus-
195 tering, the optimal field should lie somewhere on the $-5^\circ \lesssim b \lesssim 5^\circ$ strip of ecliptic latitude. The
196 latitude distribution of classical KBOs within this strip roughly follows the sum of two Gaussian
197 distributions with full width at half maximums of approximately 4.6° and $\sim 30^\circ$ (Elliot et al., 2005).
198 Then the number of KBOs expected at the $\pm 5^\circ$ boundaries is still more than half those expected
199 along the ecliptic plane, and KBO sky surface density does not fall sharply with deviation in ecliptic
200 latitude over the strip of interest.

201 **2.1.2 Stellar sky surface density**

202 Not all stars are equally likely to be associated with detectable occultation events. The optimal field
203 should have the maximal population of stars of angular sizes producing detectable events.

204 As discussed in section 1.2, when the angular size of the star is smaller than the occulting KBO
205 and the KBO has a diameter $\lesssim F$, the strongest Fresnel diffraction effects are produced. Since
206 eq. (1) gives $F \approx 1.3$ km for an occulting KBO diameter in the R-band, the associated angular
207 diameter at a distance of 40 AU is approximately 46 microarcseconds (μas). Stars smaller than this
208 angular diameter are expected to be responsible for Fresnel-diffracted events.

209 However stars of angular sizes $> 46 \mu\text{as}$ will still produce occultation events that are not in the
210 Fresnel-diffracted regime. An examination of simulated events for the *Colibri* array in section 4.2.2
211 indicates that stars with angular diameters as large as $100 \mu\text{as}$ are associated with detectable events.

212 The optimal field must then balance the falloff of KBO sky surface density away from the ecliptic
213 with the potential sky surface density of angularly small stars. Because the event rate is produced
214 through random processes, the number of events is proportional to the number of monitored stars.
215 The number of monitored stars will be roughly the number of angularly small stars in the field. The
216 likelihood that such stars vary only by a factor of ~ 2 among all candidate fields is low, since stellar
217 sky surface density is not uniform across ecliptic longitudes. Therefore while any given field should
218 be fully evaluated for the tradeoff between KBO and stellar sky surface densities, it is expected that
219 all or nearly all fields will have to prioritize stellar sky surface density over KBO sky surface density.

220 Nonetheless, the ecliptic latitude distribution of KBOs is not well understood beyond the $\pm 5^\circ$
221 boundaries (Elliot et al., 2005), and so the field selection process will not consider sky regions beyond

222 these bounds.

223 **2.1.3 Ranking of candidate fields**

224 As *Colibri* is a continuous monitoring survey, it will be operating year-round and thus requires
225 multiple fields to populate its observing campaign. Candidate fields can be ranked by order of
226 preference by examining the field's central ecliptic latitude (for KBO sky surface density) and number
227 of angularly small stars (for stellar sky surface density). The optimal field is the most important field
228 of the year: because the event rate is random, the best chance of detection will be when observing
229 the optimal field.

230 A corollary of this is that candidate fields *nearby* the optimal field should be excluded from the
231 observing program. Since the optimal field is most likely to have the largest event rate, it would be
232 counterproductive to switch observation to a lower-ranked field while the optimal field can still be
233 observed.

234 Of course, depending on the specifics of the field candidates, such a transition may be appropriate
235 when a lower-ranked field candidate can be observed on a given night longer than the optimal field,
236 at least proportionally to the likelihood of detection. This and similar other considerations are
237 important in designing the timing and targeting of the observing campaign.

238 **2.2 *Gaia*: producing a field search sample**

239 In order to determine the stellar sky surface density of angularly small stars, it is necessary to have
240 estimated distances and radii to all stars in the field search sample. Since the KBO ecliptic latitude
241 limits the field search sample to $-5^\circ \leq b \leq 5^\circ$, the data set on which to draw must have a high density
242 of stars in the ecliptic. While there are large spectroscopic surveys which produce batch quantity
243 estimates, these generally do not sample enough stars. The European Space Agency's (ESA) *Gaia*
244 satellite, launched in 2013 with data available since 2016, samples the parallax distances of over 1
245 billion bright ($G \lesssim 21$ mag) stars and thus provides a unique solution to this problem. The current
246 *Gaia* data release has three broad passbands: G , G_{RP} , and G_{BP} roughly corresponding to V , R ,
247 and B coverage. It also contains stellar radius estimates for over 160 million stars, with future data
248 releases planned to complete these estimates for all observed stars.

249 **2.2.1 The *Gaia* mission**

250 The *Gaia* mission is the ESA's latest astrometric mission. *Gaia*'s primary goal is to measure the
251 position and spectra of over one billion stars to provide high resolution three-dimensional spatial
252 and velocity maps of the wider stellar neighbourhood (Gaia Collaboration, Prusti, et al., 2016). One
253 of the key astrometric properties *Gaia* measures is trigonometric stellar parallax, which can provide
254 accurate distance estimates to each star. One of the key astrophysical properties *Gaia* estimates
255 is stellar radius, through a combination of apparent brightness and modelled effective temperature.
256 *Gaia* also estimates the extinction of each star, which it uses to correct the apparent brightness and
257 refine the inferred stellar radius. The radius can then be used with the parallax to estimate the
258 stellar angular diameter.

259 *Gaia* was launched in 2013 with a 5 year timeline and the promise of high precision measure-
260 ments. The accuracy targets for *Gaia* require multiple observations of the same sources, and the
261 promised accuracy and full data was recognized to be unlikely to arrive before 2022-2023 (Gaia Col-
262 laboration, Brown, et al., 2016). Thus *Gaia* was designed to stage intermediate data releases, with
263 the understanding that each release be treated as an independent catalogue and that measurements
264 may change in subsequent releases (Gaia Collaboration, Brown, et al., 2016). The first data release
265 (DR1) was made available in late 2016.

266 **2.2.2 *Gaia* Data Release 1 (DR1) and Data Release 2 (DR2)**

267 The *Gaia* DR1 contains approximately 1.1 billion sources, nearly all of which lie in the magnitude
268 range $11.2 \leq G \leq 21$ mag, observed over \sim 14 months. Astrometric positions are accurate to
269 within $300 \mu\text{as}$, and provide a reliable stellar sky surface density map for stars of all angular sizes.
270 However, positional errors are larger for stars in the ecliptic, which were observed less frequently
271 than other sources as a result of the observing pattern over the first 14 months of operation (Gaia
272 Collaboration, Brown, et al., 2016). The DR1 also does not contain stellar radius estimates, and
273 parallax measurements in the ecliptic are particularly lacking due to the observing pattern of the
274 spacecraft. Thus the DR1 provided insufficient information to well constrain optimal stellar field
275 candidates.

276 Nonetheless the DR1 provides a first look at the statistical properties of stars across the ecliptic
277 latitude strip of interest. The magnitude distribution of the DR1 was examined to estimate how steep

278 the sky surface density dependence was on stellar magnitude. Although there was no information
 279 on the stellar angular sizes, one would generally expect to find a greater number of angularly small
 280 stars where there is a greater number of all-sized stars. Thus the magnitude distribution could be
 281 used as a rough estimate to approximate the effect of *Colibri*'s sensitivity on the event rate.

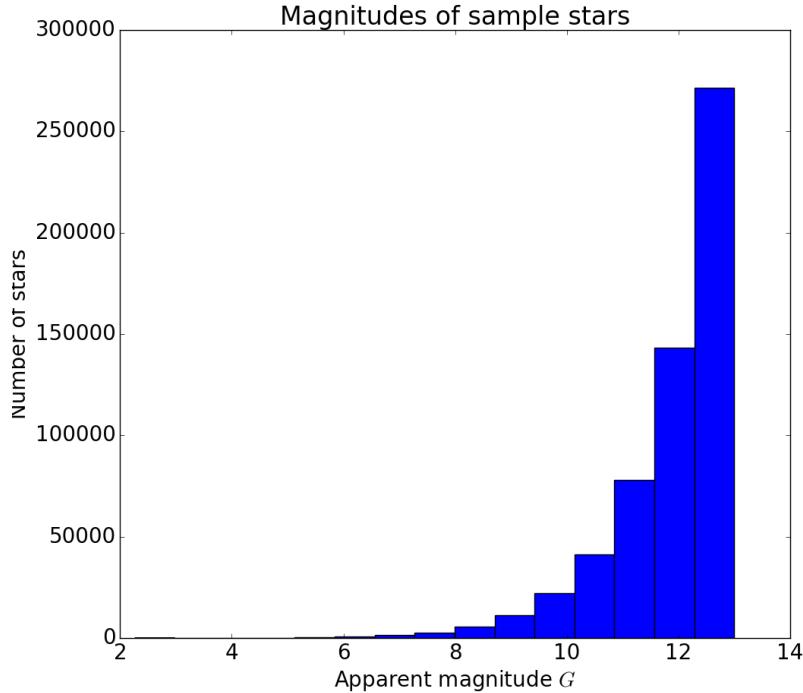


Figure 3: Number of stars within $-5^\circ \leq b \leq 5^\circ$ as a function of apparent G magnitude, extracted from the *Gaia* DR1.

282 The *Gaia* DR2 was made available in spring 2018. Based on 22 months of data (the 14 of
 283 DR1 plus an additional 8), parallax and position information for stars in the ecliptic were added
 284 and improved. Of particular interest was the introduction of radius estimates for 162 million stars
 285 brighter than $G \lesssim 17$ mag.

286 2.2.3 Reliability of the *Gaia* DR2 radius estimates

287 The *Gaia* DR2 radius estimates are calculated through the Stefan-Boltzmann law:

$$L = 4\pi R^2 \sigma T_{\text{eff}}^4 \quad (2)$$

288 The bolometric luminosity L is inferred from the three photometric bands (G , R_P , B_P); reddening
289 is estimated by comparing synthetic stellar spectra to the observed flux at the parallax distance
290 to improve the precision of L (Andrae et al., 2018). T_{eff} is not as easily approximated with data only
291 in these passbands, and Andrae et al. (2018) details the computational effort involved in arriving at
292 these estimates. The key takeaway is that T_{eff} is less reliable where reddening is substantial. This
293 has the potential to impact the quality of angular size calculations for candidate fields in the galactic
294 plane, and is discussed further in section 2.4.1.

295 The quality of the T_{eff} values can nonetheless be tested against other surveys with more precise
296 measurements of this quantity. One such survey is the Radial Velocity Experiment (RAVE), which at
297 the time of this analysis has made available its fifth data release (DR5). The RAVE DR5 is a survey
298 of nearly half a million stars randomly selected across the southern hemisphere with magnitudes
299 $9 < I < 12$ mag (Kunder et al., 2016). Crossmatching sources between the *Gaia* DR2 and RAVE
300 DR5 is not a trivial task, and some sources cannot be confidently matched across these two catalogues
301 (Marrese, Marinoni, Fabrizio, & Altavilla, 2018). Nonetheless the work done by Marrese et al. (2018)
302 suggests that at worst $\sim 1\%$ of RAVE sources must be excluded on this basis, and so the majority
303 of stars present in both catalogues can be compared across catalogues.

304 The RAVE DR5 contains temperatures obtained from spectroscopic data. A full analysis in
305 Kunder et al. (2016) concludes that the reported values of T_{eff} in the DR5 have uncertainties of ~ 150
306 K. This makes them a valuable benchmark against which to check the *Gaia* DR2 temperatures. A
307 good agreement increases confidence in the DR2 stellar radii, as they depend on the DR2 T_{eff} and
308 the T_{eff}^4 dependence in eq. (2) will make any estimation errors costly.

309 The current estimate of *Colibri*'s limiting is $G \approx 11.5$ mag when imaging at the optimal cadence
310 of 40 Hz. Restricting the stellar sample to *Gaia* DR2 stars with $G \leq 11.5$ mag that fall within
311 $-5^\circ \leq b \leq 5^\circ$ that also have radius and parallax estimates yields 80 129 stars to serve as the field
312 search sample. There are 16 492 stars within the ecliptic latitude and magnitude limits which do
313 not have temperature estimates in the DR2. As future data releases are expected to provide such
314 estimates for these stars, they represents a future $\sim 20\%$ expansion of the sample.

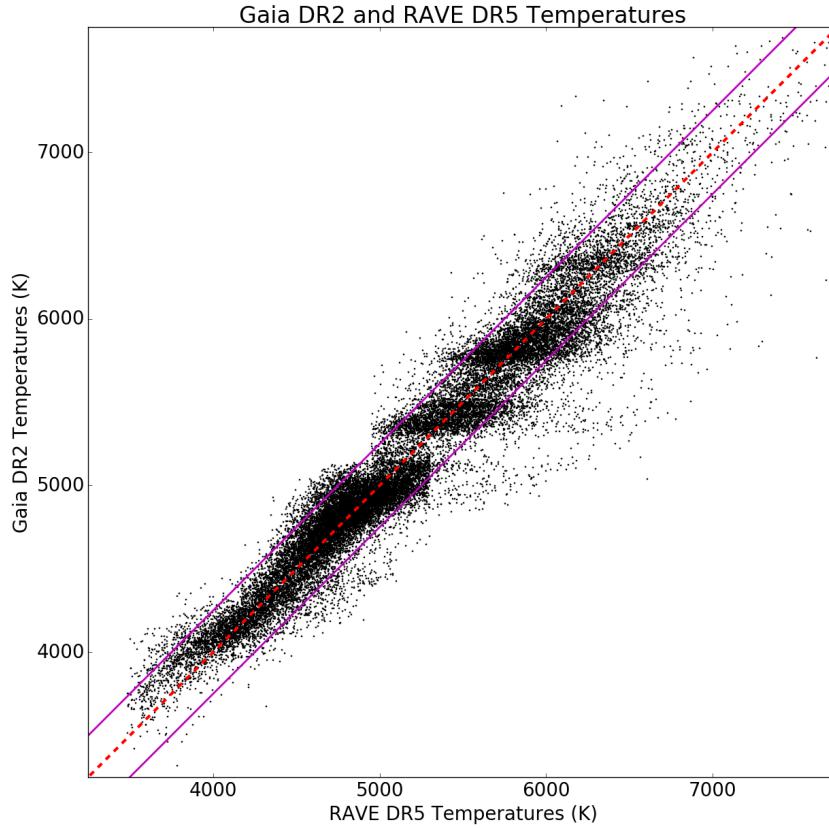


Figure 4: *Gaia* DR2 T_{eff} estimates compared against RAVE DR5 T_{eff} estimates. The dashed line indicates the 1:1 T_{eff} correspondence, and the solid lines indicate a ± 250 K window in which the majority of data lie.

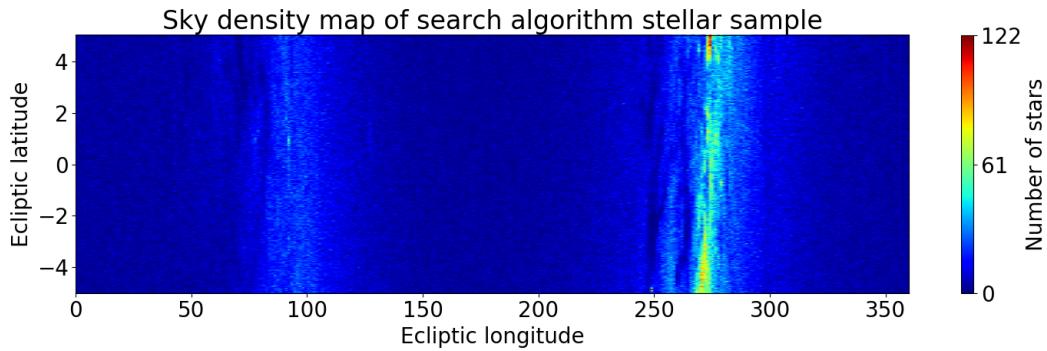


Figure 5: Density map of the 80 129 stars with estimatable angular sizes from the *Gaia* DR2. The sample includes only stars brighter than $G \leq 11.5$ mag between ecliptic latitudes $-5^\circ \leq b \leq 5^\circ$.

315 **2.3 An algorithm for the selection of optimal fields**

316 As discussed in section 2.1.2 the stellar sky surface density must be prioritized when searching the
317 sample for candidate fields. There are a number of ways to search for overdensities in the distribution
318 of stars. A common approach is to set the problem up spatially, by organizing the stars on a 2D
319 plane by their coordinates and subdividing that space into a grid of arbitrarily sized patches. A
320 number of techniques exist to sort these patches by their location and contents to investigate local
321 features, including an overdensity of points (Kogan, Nicholas, & Teboulle, 2006).

322 This approach is a type of cluster analysis, where objects are grouped according to similar
323 properties to better understand underlying group memberships and properties. Cluster analysis
324 has numerous applications in computer science such as data mining, machine learning, pattern
325 recognition, and is now even used to facilitate object recognition for intelligent computer systems
326 (Bewley & Upcroft, 2013; Kogan et al., 2006).

327 This approach has the advantage of being quick to program and will produce short and simple
328 output. It is also very adaptable; density plots like those of fig. 5 can be analyzed with imaging
329 software to identify overdensity hotspots and compare them to other local features (Bewley & Up-
330 croft, 2013). However, in general the arbitrary subdivision of the field will degrade the accuracy of
331 any patch-counting method and unnecessarily discritizes the possible results (Kogan et al., 2006).
332 Additionally, the computing time gains by using a small set of patches are lost when the desired
333 cluster resolution is high (corresponding to a greater number of patches) (Kogan et al., 2006). A
334 high resolution is needed to optimize the field candidates, and this type of cluster analysis quickly
335 becomes computationally heavy for $\sim 10^5$ star coordinates.

336 **2.3.1 Nearest-neighbour searching**

337 Another approach to locating overdensities is to utilize nearest neighbour searching (NNS). An NNS
338 algorithm operates over a metric space X (in this case the ecliptic coordinate system) populated
339 by data points in a sample S . The goal of the algorithm is to rapidly report for any queried point
340 $q \in X$ the closest data point in S to q (Maneewongvatana & Mount, 1999).

341 A brute force algorithm best illustrates the application of this concept; such an NNS algorithm
342 would calculate the distance from each star in the sample to every other star in the same sample.
343 Thus S represents both the set of coordinates in the sample and the list of field coordinates to

344 query. Then a field of view filter can be applied to each star, and whichever star retains the largest
345 number of neighbours represents the centre of the optimal field of view. The other centroids would
346 automatically be ranked in descending order of angularly small star population within the limiting
347 field of view.

348 The advantage to this method is that it outputs a full ranked list of field candidates as simultane-
349 ous output. As with the subdivision/patch approach there is still a discretizing bias since the density
350 resolution would be higher where there are more stars to consider as field candidates. However since
351 the goal of the field search is to find dense stellar fields this bias can be safely ignored.

352 The disadvantage to this method is the computation time, as there will be at best $N^2 - \sum_{n=0}^{N-1} n$
353 unique distances to compute for N stars: for the 80 129 stars in the field search sample, this requires
354 $\sim 3.2 \times 10^9$ calculations. Thus a brute force NNS algorithm has the promise of a robust and accurate
355 field search, but at the heavy expense of computing time. Attempts to implement this algorithm
356 directly have been unsuccessful due to the heavy memory cost of storing 10^9 ordered values.

357 A workaround of this problem was found by examining an analogous and well-known problem
358 in computer science called the 'travelling salesman problem' (TSP). The TSP considers N nodes
359 where each distinct pair of nodes has a well-defined distance between them. Laporte (2010) gives
360 an excellent introduction to TSP literature and development.

361 The problem itself is how to traverse all N nodes such that each node is visited once and the total
362 path taken across all nodes sums to the smallest possible value (Arora, 1998). This is analogous to
363 using stellar coordinates as nodes, where the 'path' travelled among them is the order of calculations
364 performed and its length is the sum of their distance values. By minimizing the number of distance
365 calculations that must be uniquely performed this necessarily reduces the total angular distance
366 traversed.

367 Attempts to locate or approximate solutions to the TSP have led to a rich development of
368 algorithm and optimization design (Arora, 1998; Laporte, 2010). One such area of development
369 has been the use of binary search tree data structures to optimize the paths of the TSP. Such data
370 structures can be used to significantly improve the performance of an NNS.

371 2.3.2 Binary search trees and the optimized k-d tree

372 A binary search tree is a data structure that stores elements of a sorted sequence in a tree-like
 373 structure that retains the sequence's order (see fig. 6) (Mehlhorn & Sanders, 2008). A query for
 374 an element of the sequence begins at the top of the tree, or the *root node*. At this and subsequent
 375 nodes there are exactly two possible paths the query can take depending on the query's relationship
 376 to the node's value. For example, in fig. 6 if the query is smaller than 17 it branches left at the root
 377 node, and otherwise branches right. Queries of binary trees find matches much faster than queries
 378 that iterate through a sorted sequence (Mehlhorn & Sanders, 2008).

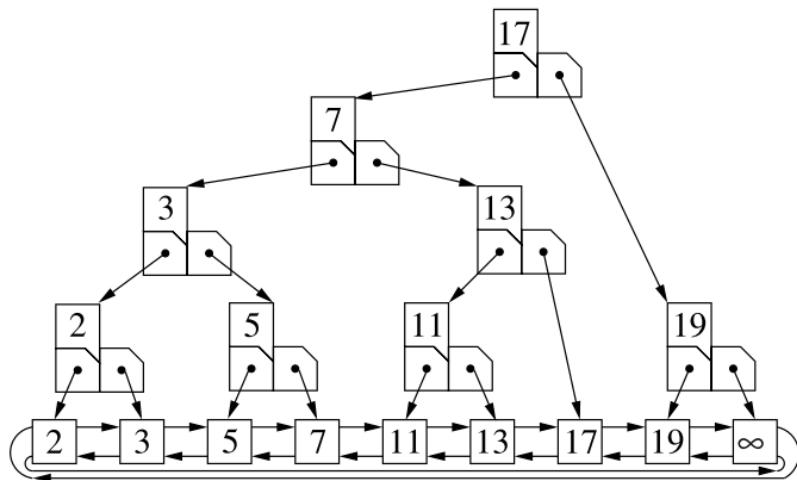


Figure 6: Example of the binary search tree representation of the ordered sequence of integers (2, 3, 5, 7, 11, 13, 17, 19) from Mehlhorn and Sanders (2008, Fig 7.2a). A query for any integer in the set begins at the top of the tree and branches either left or right, depending on its relation to the value of the node, terminating in the bottom 'leaves' of the tree. These leaves represent the complete and ordered set, verifying that the binary search tree structure retains the sequence order.

379 Typically for 2-dimensional data points (such as stellar coordinates) the binary search tree is
 380 replaced by a quadtree: a generalized binary tree optimized for 2-dimensional data (see Finkel &
 381 Bentley, 1974). An analysis of the TSP in Euclidean \mathbb{R}^2 by Arora (1998) shows that the use of
 382 quadtrees can be used to traverse paths in logarithmic time, as opposed to the polynomial time
 383 complexity of brute force methods. Further, Arora (1998) shows that the memory required to
 384 traverse a path using a quadtree is minimal when compared to brute force methods. A quadtree is
 385 thus a promising data structure in which to store the stellar coordinates.

386 Research has been done on the optimization of querying quadtrees; a useful outcome has been the

387 derivation of the *optimized k-d* tree, a data structure well-suited to minimizing NNS computation
 388 time (Freidman, Bentley, & Finkel, 1977). The optimized k-d tree was designed to generalize the
 389 quadtree structure with an NNS application in mind and offers benefits to memory storage and
 390 computation time over the classical quadtree.

391 Observation and calculation of how a query traverses nodes during an NNS have led to a gen-
 392 eralized algorithm that instructs the optimized k-d tree how best to construct itself to query the
 393 specific data set being used (Freidman et al., 1977; Maneewongvatana & Mount, 1999). The result
 394 is a major optimization of query times without laborious consideration of the optimal data structure
 395 to construct; this is particularly useful since the field search algorithm has $\sim 10^5$ coordinates to
 396 store in the tree.

397 2.3.3 The field search algorithm

398 The field search algorithm uses an optimized k-d tree to store the ecliptic coordinates of the 80 129
 399 stars in the search sample (section 2.2.3) for fast searching. Nonetheless, any filtering of data (such
 400 as imposing a limiting magnitude) is best performed prior to running the algorithm. The NNS is
 401 still computationally difficult even for optimized k-d trees, and the fewer coordinates in the tree the
 402 faster the algorithm can solve for candidate fields.

403 One simple filter to apply is in the *Gaia* search itself. For DR2 column headers `source_id`,
 404 `phot_g_mean_mag`, `parallax`, and `radius_val` a star must have non-null values to be considered for
 405 the algorithm. All stars have non-null `source_id` as their unique ID, but the other three columns
 406 can have null entries and can be used to reduce the input sample size.

407 Before constructing the optimized k-d tree it will be necessary to calculate the angular sizes of
 408 the stars. The angular size of a star is calculated as:

$$\theta \text{ (rad)} = \arctan \left(\frac{D}{d} \right) \quad (3)$$

409 Where D is the physical solar diameter of the star and d is its distance in the same units. For a
 410 *Gaia* star, `radius_val` is given in stellar radii, so $D = (D_{\odot})(\text{radius_val})$ where $D_{\odot} = 1.391 \times 10^9$ m.
 411 The parallax P is measured in milliarcseconds (mas) and so the distance is the inverse parallax
 412 converted to metres $d = \frac{\phi}{P}$ where $\phi = 3.085674 \times 10^{16} \frac{\text{m}}{\text{mas}}$. This gives the expression for the stellar

413 angular size (in μas):

$$\theta_{\text{GAIA}} (\mu\text{as}) = \arctan \left(\frac{(\text{parallax})(D_{\odot})(\text{radius_val})}{\phi} \right) \left(2.06265 \times 10^{11} \frac{\mu\text{as}}{\text{rad}} \right) \quad (4)$$

414 One last good filter is to omit all stars greater than some angular size for which *Colibri* is unlikely
 415 to find occultation events. This is fully addressed in section 4.2.2 and this limit is set to 100 μas ,
 416 with larger stars excluded from the sample. While the actual stellar fields will have stars of any
 417 angular size, this cutoff allows the search algorithm to base its decisions on stars of useful angular
 418 sizes only.

419 At this point the optimized k-d tree is constructed using the spatial.cKDTree class of SciPy (E.
 420 Jones, Oliphant, Peterson, et al., 2001). The coordinates of each star in the sample can then be
 421 queried in an NNS search with a boundary of the field of view of *Colibri*. The coordinates of the
 422 stars in each field candidate are retained and the field candidates are ranked by the number of stars
 423 they contain.

424 As discussed in section 2.1.3 it is important that the set of *Colibri* fields span all ecliptic longitudes
 425 to facilitate year-round observing. Therefore a field candidate overlap check is initiated at this stage
 426 of the search. The goal of this check is to preserve the ranking of field candidates while excluding
 427 candidates that lie too near to better-populated ones. This ensures the *current* field observed by
 428 *Colibri* year-round is the *best possible* field observable at that time.

429 This requires a user-specified minimum field separation by ecliptic longitude. It is difficult to
 430 specify an optimal separation for all purposes, as the final field choice will depend on many factors
 431 including the cloudiness of the season, the sky density of KBOs at the field's ecliptic latitude,
 432 other nearby field candidates, and even local obstruction at the observing site. Therefore it is
 433 recommended to set a minimum separation of approximately 10° to generate a number of options
 434 for final consideration.

435 The number of final fields returned by the program will be the number that distributes fields
 436 evenly by this separation across the whole longitude range. Thus a smaller separation will also result
 437 in a larger number of fields returned. Following the overlap check, the algorithm returns the top N
 438 fields ranked from most populous to least.

439 Note these field are based on the data of *angularly small* stars; the actual number of stars in the
 440 field bright enough to be seen by *Colibri* will differ, though they are not largely expected to produce

441 occultation events.

442 2.4 Performing the search

443 The search algorithm was performed on 80 129 valid *Gaia* DR2 stars with a magnitude limit of
 444 $G \leq 11.5$ mag, an angular size limit of $\theta \leq 100 \mu\text{as}$, and an ecliptic longitude spacing of $\Delta\ell = 12^\circ$.
 445 24 fields were returned, the top 5 of which are summarized in table 1. Figure 7 shows the sky density
 446 plot of stars in the sample overlaid with the locations of the 24 fields.

447 The search algorithm ran with a total duration of 104 seconds of which the optimized k-d tree
 448 search took 18 seconds. While previous attempts at a brute force NNS had overflowed the memory
 449 of a computer with 64 GB of RAM before completion the optimized k-d tree search never exceeded
 450 2 GB of memory use.

Field No.	Field Centre (ℓ, b)	$N_{\text{stars}} < 100 \mu\text{as} (G \leq 11.5 \text{ mag})$
1	(273.55, 4.75)	240^{+32}_{-29}
2	(287.48, 4.34)	137^{+13}_{-10}
3	(87.67, -2.60)	134^{+7}_{-21}
4	(102.07, 1.44)	130^{+4}_{-15}
5	(127.19, 1.03)	128^{+6}_{-10}

Table 1: Top five field candidates returned by the search algorithm.

451 2.4.1 Errors and reliability of results

452 The error in the number of angularly small stars in a field is directly tied to the error in the calculation
 453 of angular sizes and in the astrometric precision of the coordinates. Since the *Gaia* DR2 position
 454 errors ($\sim 40 \mu\text{as}$, Lindegren et al., 2018) are small compared to the field of view their contributions
 455 can be ignored and the field position contributes no error to the field population.

456 The radius error comes from the errors in inferred bolometric luminosity and modelled T_{eff} from
 457 Andrae et al. (2018); for this data the average radius error is $\delta R \approx^{+0.15}_{-0.19} R_{\odot}$. With an average radius
 458 of $1.87 R_{\odot}$ this represents an error of $^{+8.0\%}_{-10\%}$. Although it is expected that most stars have $R < 1R_{\odot}$,
 459 the strict magnitude limit of $G = 11.5$ mag is likely responsible for eliminating most of the smallest
 460 stars (figs. 8 and 9).

461 The average parallax error in the data is $\delta P \approx 74 \mu\text{as}$. Since the average parallax in the data

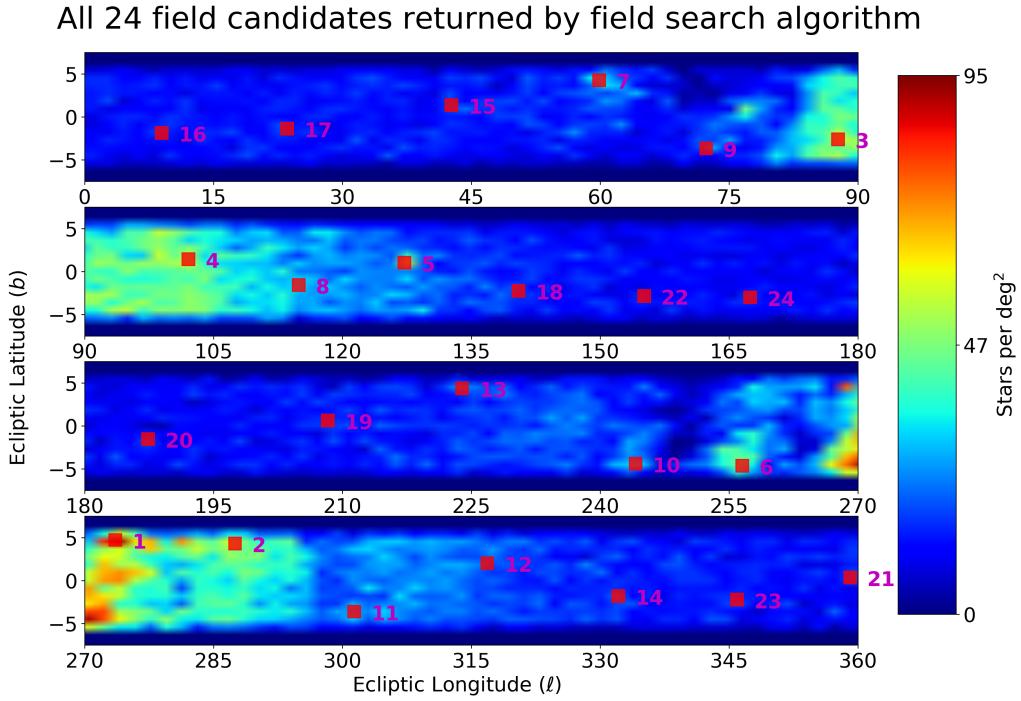


Figure 7: The 24 fields selected by the algorithm for the *Colibri* observing campaign. The red box indicates the field and field of view.

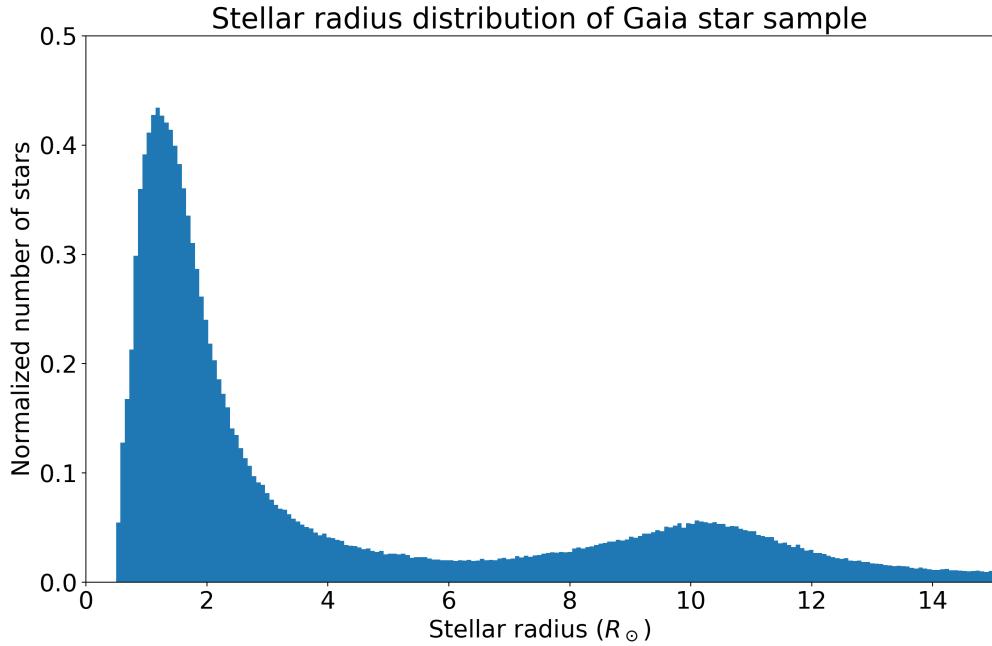


Figure 8: Stellar radii of stars in the *Gaia* field sample. The bump at $R \approx 10R_\odot$ is caused by the presence of giants in the star sample.

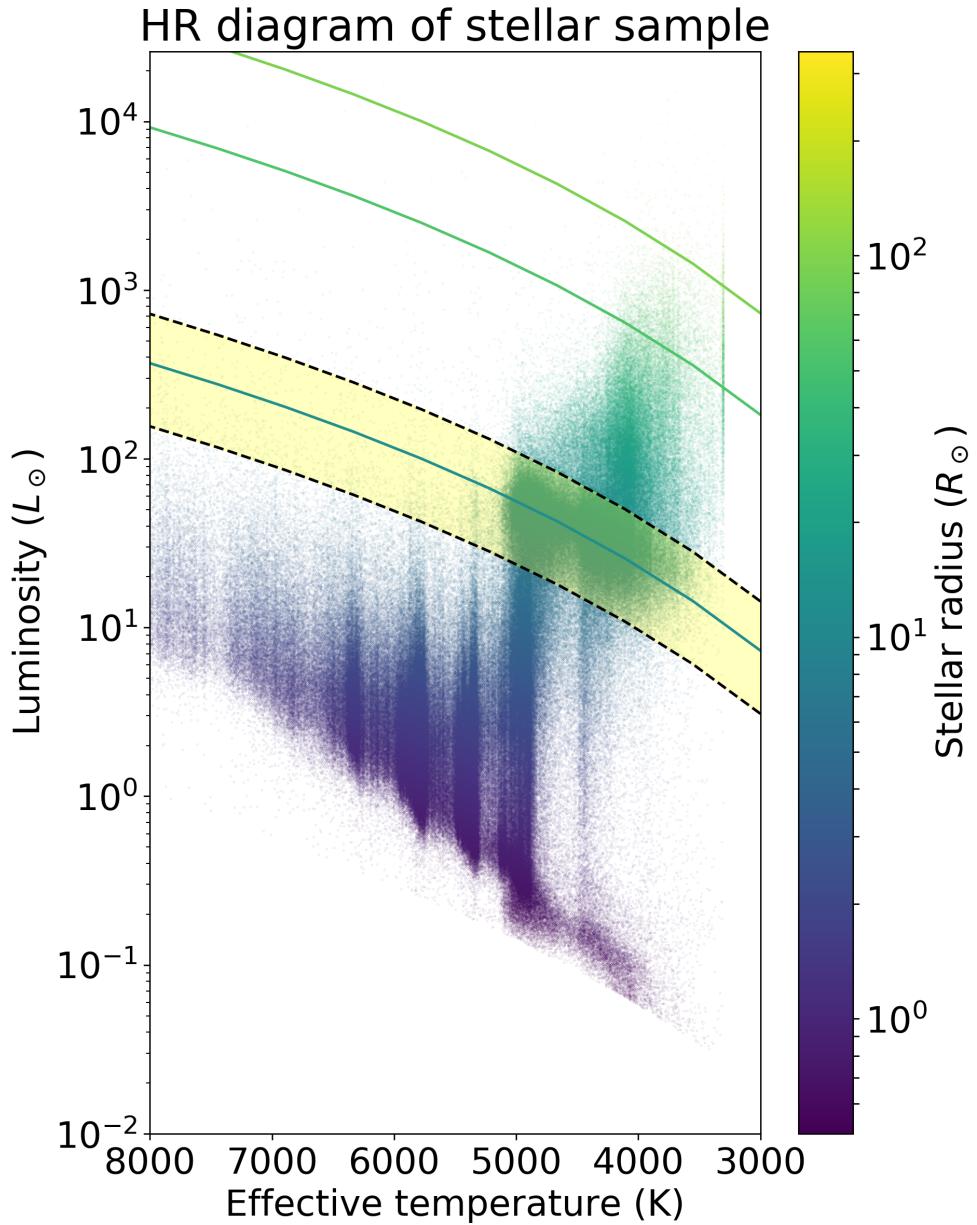


Figure 9: HR diagram of stars in the *Gaia* field sample. The yellow region indicates the $6.5 \lesssim R \lesssim 14$ radius range over which the bump in fig. 8 emerges. The line of constant radius $R = 10R_{\odot}$ is marked in the centre of the yellow region and passes through the giants population, producing the bump in fig. 8. Lines of constant radius at $R = 50R_{\odot}$ and $R = 100R_{\odot}$ are also included.

462 is 1225 μas this represents a fractional error of 6.0%. Thus while the radius error is the primary
 463 contributor to the error in angular size, the parallax error must also be taken into account.

464 The field search algorithm will return the locations of overdensities, but for the assessment of error
 465 the search algorithm estimates the errors in each field by fixing that particular field's coordinates
 466 and then recalculating the stellar angular sizes. This is a necessary step since the actual choice of
 467 field coordinates depends on the angular sizes of the stellar sample. Thus the reported *Gaia* values
 468 are used to find the fields, and the value errors are used to assess the uncertainty in the angularly
 469 small stellar population of candidate fields. Combining the effects of parallax and radius errors in
 470 one direction give the positive errors listed in table 1 and the negative errors come from the opposing
 471 direction.

472 A useful measure of the impact of the data quality on population error is the average fractional
 473 error over all N stars in the sample ($\frac{\bar{\delta}\theta}{\theta} = \sum_{n=1}^N \frac{\delta\theta_n}{\theta_n}$) where for each star the calculation is made:

$$\frac{\delta\theta}{\theta} = \sqrt{\left(\frac{(\delta R^+)^2 + (\delta R_-)^2}{4R^2}\right) + \left(\frac{\delta P}{P}\right)^2} \quad (5)$$

474 Where δR^+ and δR_- are the respective positive and negative radius errors and the substitution
 475 has been made that $D = 2R$. Since there is no reason to prefer one radius error over the other
 476 (Andrae et al., 2018) the error in D is the quadrature combination of the two radius error estimates.
 477 The computed value of $\frac{\bar{\delta}\theta}{\theta}$ is 0.0727; for a cutoff of 100 μas this corresponds to an error of $\approx 7 \mu\text{as}$.

478 Figure 10 shows the computed angular size distribution for the top field in the sample with this
 479 error marked, demonstrating an error of $\sim 7 \mu\text{as}$. While the region of error does not encompass a
 480 particularly notable number of stars, it makes a difference of ~ 30 stars in this field - not enough to
 481 change its ranking, though table 1 suggests that there may be some rank-switching for lower-ordered
 482 fields due to this effect. However, such switching would be over a few stars only and thus would be
 483 unlikely to impact *Colibri*'s expected event rate.

484 This can be generalized to examine how field population changes with the angular diameter
 485 cutoff. Figure 11 demonstrates this for cutoffs of $50 \leq \theta \leq 160 \mu\text{as}$. The ranking of fields remains
 486 unchanged, but the central position of the field does not as indicated by table 2. In some cases
 487 coordinates assigned to a field of some rank are reassigned to a field of another rank; this rank-
 488 switching is common for sparsely populated fields. In other cases entirely new positions on the sky
 489 are assigned as new field candidate locations. Only the location of the top field remained unchanged.

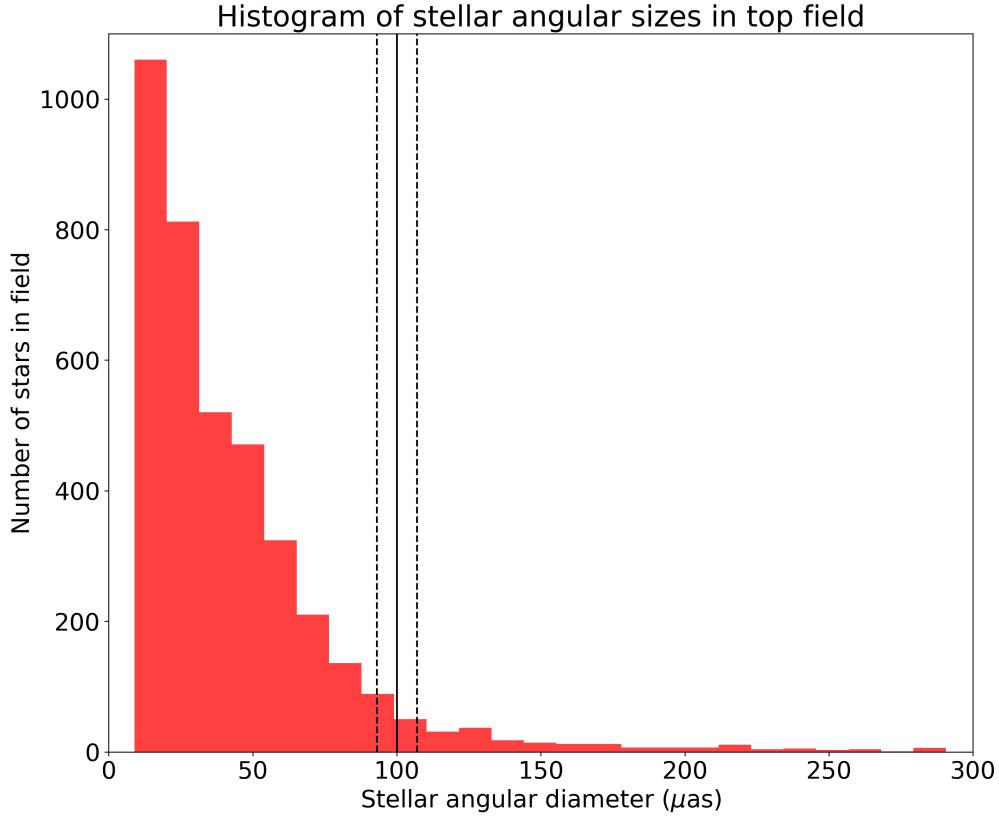


Figure 10: Histogram of the angular sizes of stars in the top field. The solid vertical line indicates the value of the size limit of the search, here $\theta = 100 \mu\text{as}$. The two dashed lines indicate the $\pm 7 \mu\text{as}$ error associated with θ at this value.

Field No.	Number of unique centroids
1	1
2	7
3	6
4	10
5	9

Table 2: Number of unique centroids for the top 5 field candidates as the θ -cutoff varies from $50 \mu\text{as}$ to $160 \mu\text{as}$ in fig. 11.

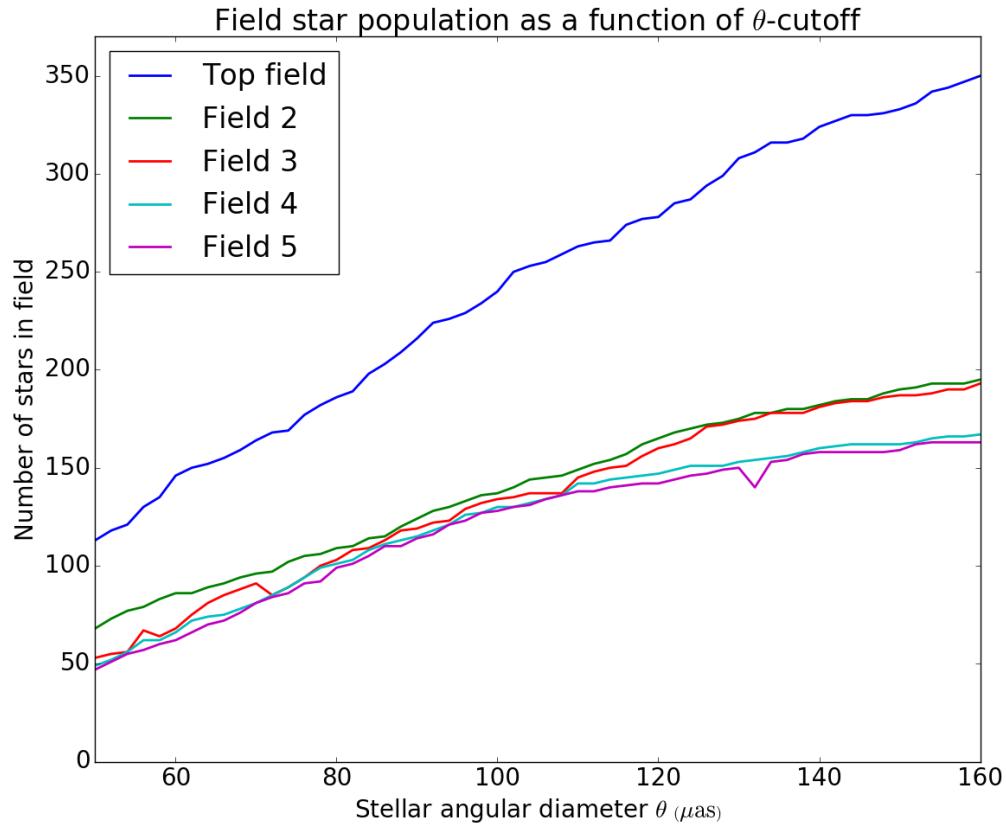


Figure 11: Stellar populations of the top five fields as a function of search cutoff size. The horizontal resolution of the graph is $2 \mu\text{as}$. For all examined cutoffs none of the top five fields changed ranking, but coordinate-switching occurred to preserve rank increasingly for low-ranking fields.

490 These results indicate that the top fields found by the search algorithm are likely the best possible
 491 fields for the given search parameters and data quality. Although candidate field rank-switching may
 492 occur (especially for low-ranked fields), that the most populous fields are buffered against this effect
 493 and tend to retain not just their ranking but their location on the sky indicates the top fields will
 494 correspond to a significant overdensity for stars capable of producing detectable events.

495 **3 Simulating occultation events**

496 This section deals with the simulation of occultation events and their likelihood of detection by
497 *Colibri*. Section 3.1 outlines the KBO model used to overlay the fields of observation; section 3.2
498 contains an overview of the algorithm for detecting events; and section 3.3 provides justification for
499 the adopted model parameters.

500 **3.1 Modelling low-diameter KBOs**

501 In order to simulate occultation events it will be necessary to adopt a model on which to base
502 the KBO field population. Although little is known about the KBO population at the sizes being
503 simulated, there has been work done to estimate their size distribution based on the cumulative
504 luminosity function of observable KBOs (Bickerton et al., 2008; Gladman et al., 2001). As discussed
505 in section 1.1 the presence of a break in the power law index q that governs the size distribution of
506 KBOs complicates this extrapolation.

507 **3.1.1 The power law q and its break diameter**

508 The power law index for the size distribution of KBOs larger than some diameter D_k is known to be
509 a constant $q \approx 4.5$ (Schlichting et al., 2012). Previous SSO surveys have attempted to constrain the
510 value of q and the 'break' diameter D_k below which KBOs obey the shallower slope. Zhang et al.
511 (2013) presents a 95% c.l. $q = 3.82$ upper limit estimate in review of seven years of TAOS I data.
512 This upper limit is of particular interest to a *Colibri* simulation as 90% of the photometric data
513 analyzed to produce this value was acquired within 6° of the ecliptic (Zhang et al., 2013), which is
514 the same latitude strip *Colibri* will observe.

515 There is evidence that q remains constant only for KBOs above the break diameter, and that due
516 to collisional processes q may adopt a number of different values at different (small) KBO diameters
517 (Bickerton et al., 2008; Pan & Sari, 2005; Schlichting et al., 2012). However there are insufficient
518 observations of KBOs below the break diameter to constrain existing models (see section 1.1) and
519 so a constant q below the break diameter will be modelled.

520 The diameter D_k at which the break in q occurs impacts the number of KBOs below it. As D_k
521 moves to larger values fewer objects are produced below the break diameter. Unfortunately this

522 diameter is not well constrained. While it is widely accepted to lie somewhere below 100 km (Pan
523 & Sari, 2005) observational data is not currently strong enough to strongly support the adoption
524 of a particular value. Justifiable values range from ~ 25 km (Bickerton et al., 2009) to ~ 45 km
525 (Schlichting et al., 2009) to as high as ~ 90 km (Fuentes, George, & Holman, 2009; Zhang et al.,
526 2013); in many cases the uncertainty in these values is a substantial fraction of the value itself (e.g.
527 Fraser & Kavelaars, 2008). This variety in adopted break diameters reflects how little is known
528 about the initial conditions of the Kuiper Belt (Fuentes et al., 2009) and demonstrates the potential
529 knowledge SSO surveys can bring.

530 **3.1.2 KBO sky surface density**

531 The sky surface density of KBOs (Σ) is directly proportional to the expected occultation rate
532 (Bickerton et al., 2009, eq. 13). Since smaller KBOs are more abundant than larger ones the
533 number density of KBOs in general will depend on the size distribution adopted. In the ~ 1 km
534 regime where *Colibri* is best tuned for detection the number density of detectable KBOs will almost
535 entirely depend on the adopted values of q and D_k . Due to the uncertainty regarding the these values
536 care must be taken when modelling the number of KBOs beneath as-of-yet observed diameters.

537 A particularly useful model with respect to such concerns is the use of a cumulative luminosity
538 function to extend the observed size distribution of small KBOs into the no-observation regime
539 (Gladman et al., 2001). The slope of the size distribution below the break diameter mostly follows
540 the slope of the luminosity function of small KBOs (Gladman et al., 2001), and is expected to entirely
541 follow it for KBOs at the same heliocentric distance (Chiang & Brown, 1999). This dependence
542 allows the model to be scaled to fit with observational data by a single factor, which is convenient
543 for updating the model with the results of ongoing observing campaigns.

544 Like all models the drawback is that the accuracy of the predicted KBO population size re-
545 quires observational data of KBOs with diameters below D_k . However, the dependence of the size
546 distribution on the luminosity function at similar heliocentric distances is particularly useful for
547 extrapolating sizes in the classical belt with its relatively small range of 42-48 AU. This is the main
548 advantage of this method and is directly relevant to the *Colibri* campaign, as it is likeliest to detect
549 classical KBOs.

550 The KBO population and sizes can be derived from the model given by Gladman et al. (2001,

551 eq. 4) of the differential number of KBOs with diameter D and heliocentric distance r :

$$n(r, D) dr dD = Ar^{-c}D^{-q} dr dD \quad (6)$$

552 Where c and q are the respective radial and size power law indices. While c is not well constrained,
 553 a good approximation for the classical KBO population is $c \approx q - 1$ (Trujillo & Brown, 2001).

554 The constant A scales this distribution such that the integrated number of KBOs in eq. (6)
 555 matches the observed number of KBOs at those distances and sizes. To set this constraint the
 556 diameter is related to the luminosity by the R -band brightness scaling relation (Gladman et al.,
 557 2001, eq. 5):

$$R = C + 2.5 \log r^4 D^{-2} \quad (7)$$

558 C is a brightness scaling constant for KBOs at opposition: its value is ≈ 18.8 for an accepted
 559 KBO albedo of 4% (Bickerton et al., 2009; Gladman et al., 2001; Trujillo & Brown, 2001). The sky
 560 surface density of KBOs in the R band appears to follow (Gladman et al., 2001):

$$\Sigma = 10^{\frac{1}{5}(q-1)(R-R_0)} \quad (8)$$

561 R_0 is the R -band magnitude where the KBO density reaches 1 deg^{-2} in the ecliptic, and is the
 562 observational parameter that constrains the model. This value is well constrained by Gladman et al.
 563 (2001) as ≈ 23.5 mag through analysis of observational surveys and previous work. From eq. (7) this
 564 corresponds to a KBO diameter of 180 km at 40 AU.

565 Combining these equations and integrating over a range of heliocentric distances $r_{\min} \leq r \leq r_{\max}$
 566 yields the number N of small KBOs in some patch of sky S near the ecliptic plane. If this number
 567 is constrained such that $N = S\Sigma$ then the scaling constant A is fully determined by observational
 568 and model parameters (Gladman et al., 2001, eq. 9):

$$A = S(q-1)10^{\frac{1}{5}(q-1)(C-R_0)} \frac{3-2q-c}{r_{\max}^{3-2q-c} - r_{\min}^{3-2q-c}} \quad (9)$$

569 Thus A scales the model to the observed sky surface density of KBOs over the area of sky being
 570 observed. Integrating eq. (6) over the distances and sizes of interest one obtains (Gladman et al.,

571 2001, eq. 12):

$$N(D > D_0; D_0 < D_k) = \frac{A(r_{\max}^{1-c} - r_{\min}^{1-c})}{(1-c)(q-1)} D_k^{3.5-q} D_0^{-2.5} \quad (10)$$

572 Here N is the number of KBOs smaller than the break diameter D_k but larger than some
 573 minimum diameter D_0 that lie within the probed heliocentric distance range within a patch of sky
 574 S .

575 Thus eq. (10) provides a means to estimate (with a few parameter assumptions) the number of
 576 KBOs with diameters $\lesssim D_k$ populating a *Colibri* field. The choice of D_0 is arbitrary, though it
 577 should not be allowed to become arbitrarily small: eq. (10) diverges as $D_0 \rightarrow 0$, and since q is not
 578 expected to be constant for smaller and smaller diameters there is little point having D_0 very small.

579 Finally, the number of KBOs expected at the $b = \pm 5^\circ$ ecliptic latitude boundaries is well over
 580 half the number at a latitude of $b = 0$ (section 2.1.1). Along with the gradual nature of this latitude
 581 dependence, the tall span of ecliptic latitude of the *Colibri* field of view (1.48°) and the scarcity of
 582 events all combine to reduce the impact of an ecliptic latitude dependency on the simulated event
 583 rate. Thus the ecliptic latitude distribution for the model KBO population is assumed to be uniform.

584 3.1.3 KBO size distribution

585 Equation (10) returns the number of KBOs smaller than the break diameter D_k but larger than
 586 some arbitrary D_0 . When generating the KBO population, the simulation uses this equation as
 587 an analogous probability distribution as shown in fig. 12 and randomly draws all diameters from
 588 this distribution. In reality random algorithms deal with discrete quantities and not continuous
 589 functions, and so eq. (10) is binned.

590 The simulation automatically bins the diameters into 500 logarithmically evenly-spaced steps
 591 between D_0 and D_k . For $D_0 = 0.5$ km and $D_k = 90$ km (motivated in section 3.3.1) the difference
 592 in adjacent size bins is $\sim 1\%$ of the value of the size bins: for example, objects of ~ 500 m diameters
 593 have bin spacings of ~ 5 m. The logarithmic dependence ensures this is a fine diameter grid for
 594 the small objects of interest to *Colibri*. If $D_k = 90$ km and $D_0 = 0.5$ km then a simulated KBO
 595 has an uneven chance of having one of 500 possible diameter values, but $\sim 14\%$ of them lie in the
 596 $0.5 \leq D \leq 1.0$ km range.

597 Once the simulation has the diameter bins computed, it then calculates the number of KBOs in

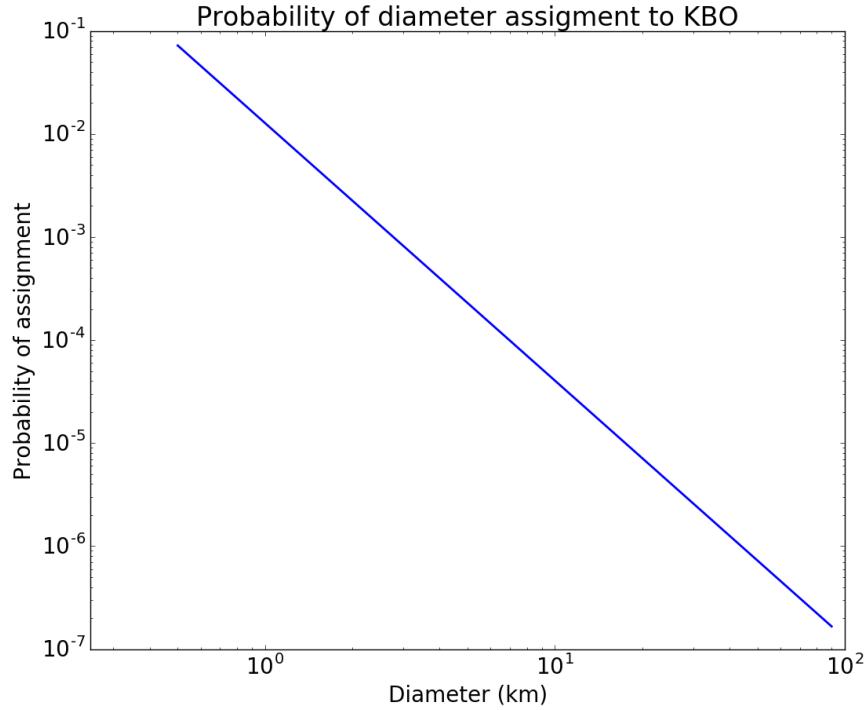


Figure 12: KBO diameter assignment probability as a function of diameter. Computed using eq. (10) with $c = 2.6$, $q = 3.6$, $R_0 = 23.5$ mag, $D_0 = 0.5$ km, $D_k = 90$ km, $42 \text{ AU} \leq r \leq 48 \text{ AU}$.

598 each bin. The fraction of the bin population to the total population then represents the likelihood
 599 of finding a KBO of that size in the field. Figure 13 shows the outcome of sampling fig. 12 as a
 600 probability distribution. At large diameters the histogram appears to become jagged. These large
 601 KBOs are so few that even with a population of $\sim 10^7$ KBOs their probability space has not been
 602 adequately sampled.

603 The total number of KBOs per field is the sum of all KBOs of all sizes $D_0 \leq D \leq D_k$ given
 604 by eq. (10). Objects above D_k are not simulated since the size distribution of large KBOs are
 605 well constrained and are capable of detection by direct observation. Since the ecliptic latitude
 606 distribution of KBOs is assumed to be uniform and there is no reason to expect a variation with
 607 ecliptic longitude, the number of KBOs is the same for all fields. The specific number will depend on
 608 the adopted model parameters: for $q = 3.6$, $c = 2.6$, $D_0 = 0.5$ km, $D_k = 90$ km, $42 \text{ AU} \leq r \leq 48 \text{ AU}$
 609 (the adoption of which are justified in section 3.3) this number is $N = 1.07 \times 10^7$.

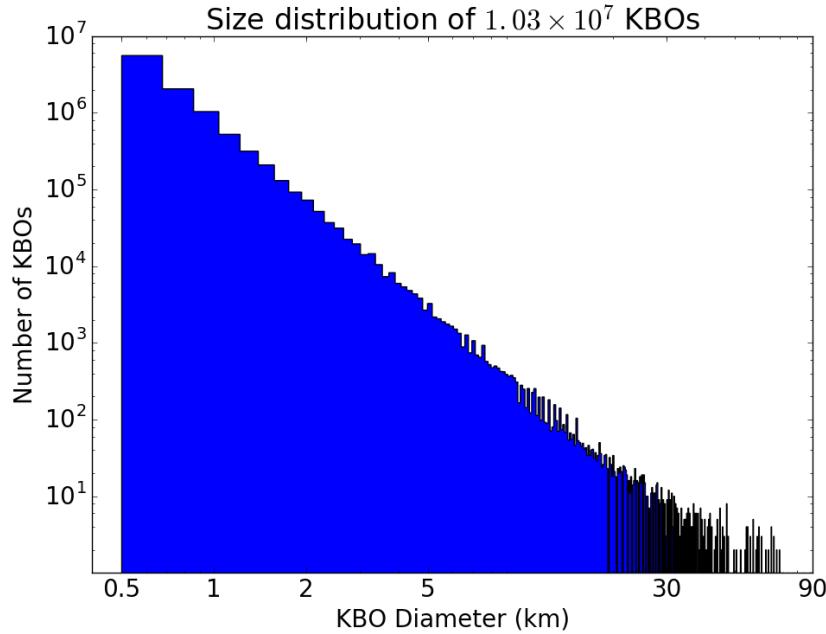


Figure 13: KBO diameters of simulated population. Computed using eq. (10) with $c = 2.6$, $q = 3.6$, $R_0 = 23.5$ mag, $D_0 = 0.5$ km, $D_k = 90$ km, $42 \text{ AU} \leq r \leq 48 \text{ AU}$.

610 3.2 Event detection algorithm

611 At the core of the simulation is an algorithm for rapidly detecting any occultation event among
 612 all the KBOs in the field by searching for 'collisions' in which a KBO or its path overlap a stellar
 613 disk. Before the algorithm can search for occultations it must first define the KBO's path on the
 614 sky (section 3.2.1).

615 Once the KBO paths are generated, the algorithm searches for collisions by using spatial indexing
 616 to locate candidate events (section 3.2.2). This approach provides a rapid search response at the
 617 cost of geometric precision and cannot be used to analyze any flagged collisions. It is thus referred
 618 to as the coarse search.

619 However, the geometry of a flagged collision must be analyzed to determine its impact parameter
 620 b : the shortest distance from the centre of the star to the centre of the occulter perpendicular to the

621 occulter's path. This parameter is required to predict the likelihood of an event detection. Therefore
 622 flagged collisions are examined by a geometric analysis package (section 3.2.3) before being assigned
 623 a likelihood of detection (section 3.2.4).

624 **3.2.1 Generating a KBO path**

625 Once the collision algorithm is called, its first step is to assign a relative perpendicular (projected)
 626 velocity to each KBO. The majority of this velocity is due to the orbital motion of the Earth about
 627 the Sun, and for observations near opposition will not significantly vary. However, since the expected
 628 event rate is proportional to this velocity (Bickerton et al., 2009, eq. 8) it is not a negligible variance.

629 Bickerton et al. (2009) notes that observing at opposition (within $\sim \pm 30^\circ$) maximizes the chance
 630 of detection when critically sampling. It will be part of the *Colibri* campaign design to ensure that
 631 the current field is always near opposition as the array will be critically sampling at 40 Hz.

632 The projected velocity of a KBO is given by Bickerton et al. (2009, eq. 10) in eq. (11); it is an
 633 approximation that is only strictly true for KBOs in the ecliptic plane with circular orbits, but holds
 634 well for the classical population considered.

$$v = \left(\frac{GM_{\odot}}{r_{\oplus}} \right)^{\frac{1}{2}} \left[\left[\frac{r_{\oplus}}{r_o} \left(1 - \left(\frac{r_{\oplus}}{r_o} \right)^2 \sin^2(\epsilon) \right) \right]^{\frac{1}{2}} + \cos(\epsilon) \right] \quad (11)$$

635 Where r_{\oplus} and r_o are the respective orbital radii of the Earth and KBO and ϵ is the solar
 636 elongation. Here G is the gravitational constant and M_{\odot} the mass of the Sun.

637 Once the KBOs are assigned their velocity they can be assigned a vector on the sky. This vector
 638 is scaled to the length of the observation period (i.e. one night's duration) and is assumed to move
 639 entirely along lines of constant ecliptic latitude. This vector is used as the geometric path of the
 640 KBO over the simulated night, with a width of the KBO angular diameter.

641 Out to a deviation in solar elongation of $\pm 30^\circ$ eq. (11) can be shown to yield a variation in v of
 642 $\sim 13\%$. Thus the solar elongation of an event is expected to play a minor but noteworthy role in its
 643 likelihood of detection.

644 3.2.2 The coarse search and spatial indexing

645 Since the runtime of the simulation primarily depends on the time taken by the algorithm to search
646 the field for collisions, minimizing this time cost is a priority. The time cost of conventionally
647 iterating through all KBOs and stars in the field to query their overlapping boundaries is too high
648 to be worth approximation. Code testing on even 10^6 KBOs lasted hours per single field. As
649 discussed in section 2.3 alternate data structures can prove effective in minimizing the time cost for
650 large data searches. In this case the data structure of interest is the R-tree.

651 The R-tree was introduced by Guttman (1984) as a solution to spatial-based queries for over-
652 lapping geometrical objects (fig. 14). The R-tree has a similar tree-like configuration as the binary
653 search tree (fig. 6), but like the optimized k-d tree each node can handle multi-dimensional data
654 (Guttman, 1984). As the query moves from the root node down through the tree it excludes non-
655 overlapping geometries at each subsequent node. When the query terminates we are left with only
656 the overlapping geometries.

657 The R-tree defines overlapping geometries as the overlapping of the bounding boxes of different
658 objects (fig. 14). The R-tree contains only the boundaries of these bounding boxes; this is why it
659 can rapidly identify overlapping geometries but cannot be used to analyze them.

660 For example, in fig. 14 the R8 box bounds a banana-like shape, which is contained in the R3 box,
661 which is contained in the R1 box. During construction the R-tree generates the R8 bounding box
662 and 'forgets' the object inside it, remembering only the four box vertices. With this information a
663 query can verify that the R8 box overlaps with the R3 and R1 boxes, but can provide nothing else.

664 The structure of the R-tree is determined by how each bounding box relates to the others. The
665 boxes are assigned unique indices (called spatial indices) and the arrangement of these indices across
666 the tree nodes encodes information about the potential of each bounding box to overlap another
667 (fig. 14).

668 Since a query to a well constructed tree runs in logarithmic time compared to the polynomial
669 time of an iterative approach (Arora, 1998) the R-tree provides a good solution to the time cost of
670 the coarse search.

671 For arbitrary shapes of arbitrary sizes spatial indexing can falsely detect overlapping geometries,
672 since a bounding box can make its containing object 'appear' to be much larger than it is (e.g. the
673 box R8 in fig. 14 is nearly half empty). However, this is not a problem for the simulation: since the

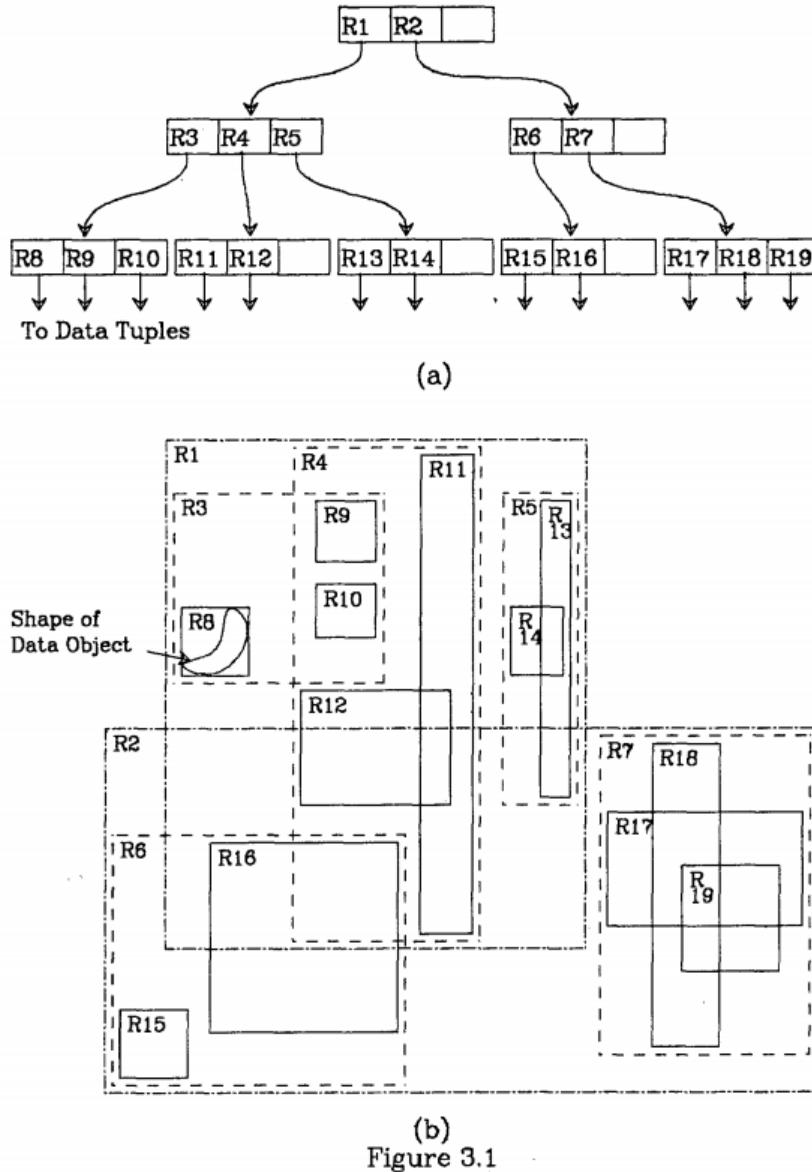


Figure 14: An example of the R-tree data structure from fig. 3, Guttman (1984). Above: The R-tree. Leaf nodes pointing to "data tuples" refer to the numerical data of the objects they represent. Below: The bounding boxes of the objects in their original spatial distribution.

674 (rectangular) KBO motion vector will be responsible for essentially all the area covered on the sky
 675 by the KBO, the bounding box of the KBO is an excellent representation of its containing object.
 676 False positives can nonetheless be detected by the high resolution followup for flagged collisions.
 677 Code testing of 1.07×10^7 field KBOs demonstrated the spatial indexing collision search provided
 678 results in ~ 660 seconds, a factor of 13 faster than the average time for a KBO-disk-only iterative

679 search. The Python module `rtree` was used to construct the R-tree and populate its spatial indices
 680 (Gillies & Butler, 2006). In all test cases there were no false positives: all events detected by the
 681 spatial index search were verified by the high resolution followup.

682 An alternate collision strategy was tested by Mills (2018) in which the field was subdivided into a
 683 fine grid and only grid patches containing both a star and KBO were analyzed for collisions. However
 684 the spatial indexing approach proved faster, performing in half the time on tests of identical fields
 685 and in one quarter time on the full random simulation. The results between these methods were
 686 consistent, with the improved runtime motivating the use of spatial indexing.

687 3.2.3 The high resolution followup and the Shapely module

688 To investigate the geometry of the collision, the high resolution analysis requires only the identity
 689 of the KBO and star with intersecting bounding boxes. For the sake of precision, all geometry will
 690 be recalculated between these two objects using the Python module Shapely (Gillies et al., 2007),
 691 a geometric and topological analysis package developed as an assistive tool for use with geographic
 692 information systems (GIS).

693 Shapely represents circles (i.e. stellar disks) as point sources with a buffer radius. The radius
 694 chosen for a star is its angular radius. Shapely uses a discretized sample of points to represent the
 695 perimeter of this radius which form line segments approximating a circle. The resolution of the
 696 buffer can vary the effectiveness of this approximation (fig. 15): Gillies et al. (2007) estimates the
 697 default sampling will reproduce a polygon with 99.8% the area of the intended circle. Effects of
 698 discretization in the fine analysis are thus disregarded.

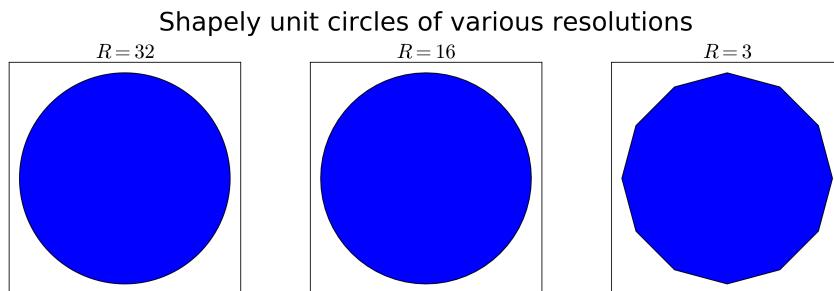


Figure 15: Shapely unit circles represented by the point-buffer method at different resolutions of a) oversampling ($R = 32$); b) default sampling ($R = 16$); and c) undersampling ($R = 3$). There is no discernable difference between the results of the default and oversampled resolutions. The undersampled circle demonstrates that discretizing happens only near minimal resolution.

699 Once the geometry has been verified as an occultation event, the algorithm computes the impact
 700 parameter b . The impact parameter will be used to generate the artificial sampled light curve that
 701 is used to estimate the likelihood of a detection by *Colibri*. The geometry of an occultation event
 702 and the length of this parameter are shown in fig. 16.

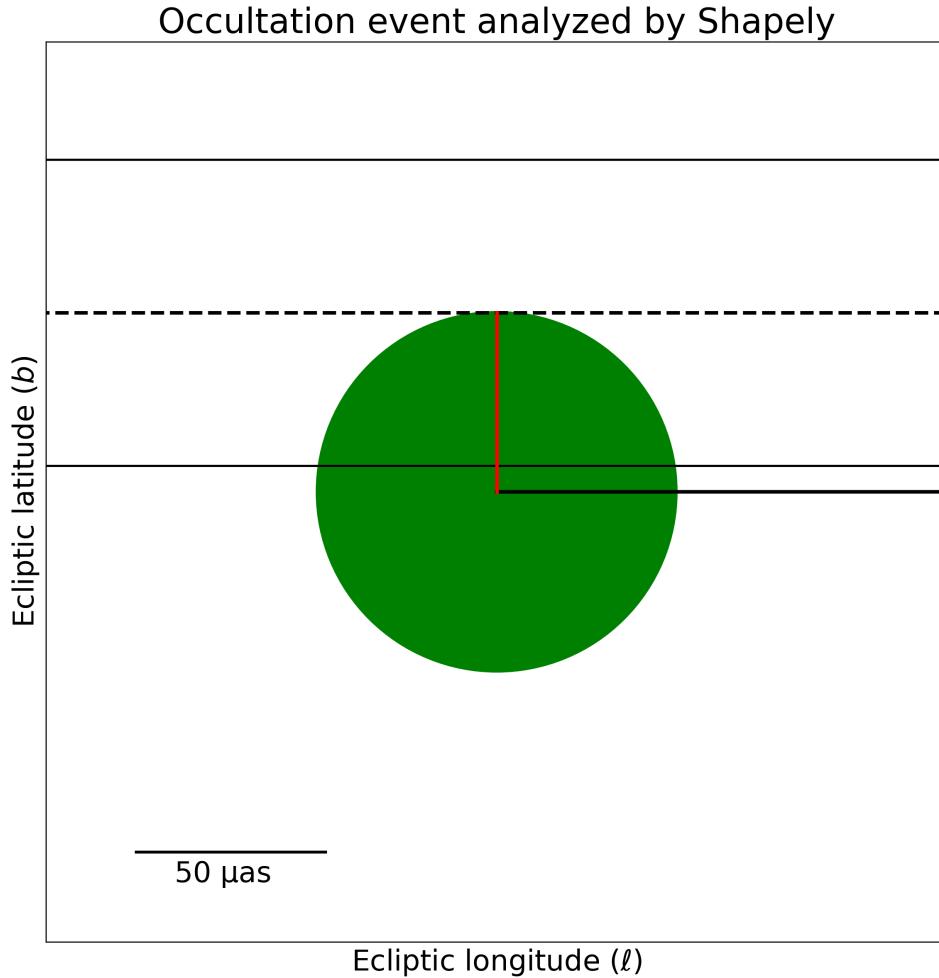


Figure 16: A 1.25 km-diameter KBO occults a $G \approx 10.1$ mag $47 \mu\text{as}$ -diameter star at 42.5 AU as constructed by Shapely. The green circle is the star. The dashed line indicates the motion of the centroid of the KBO over the observing period, and the thin solid lines indicate its radial extent. The red line indicates the perpendicular path from the centre of the KBO to the centre of the star: this is the impact parameter (here $b \approx 1.5$ km). The thick black line indicates the path the KBO centroid would follow for an impact parameter of zero.

703 3.2.4 Likelihood of detection

704 Once the impact parameter is known it is possible to estimate the likelihood of a detection by the
 705 *Colibri* system.

706 Work done by Pass et al. (2018) toward an automated detection algorithm for *Colibri* has pro-
 707 duced an accurate model of the expected light curves for arbitrary occultation events. This simula-
 708 tion produces real-valued parameter estimates for events expected to be detectable by *Colibri*. The
 709 ideal connection would be to synthesize stellar light curves of a candidate *Colibri* occultation event
 710 and estimate its likelihood of detection. Work done by Mills (2018) provides this crucial link.

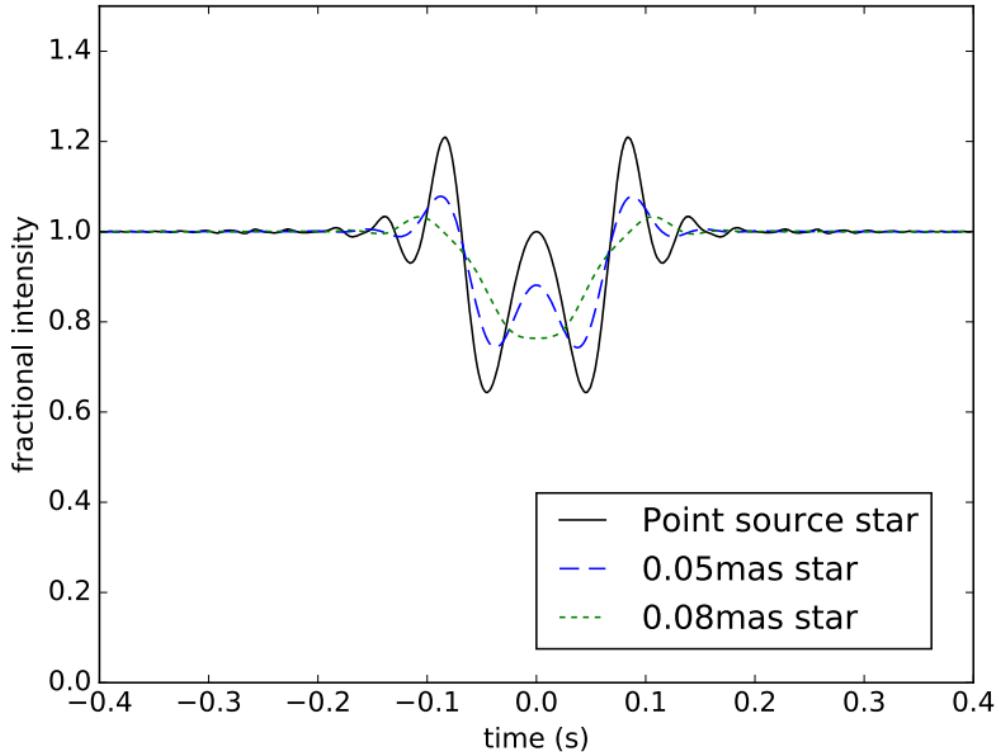


Figure 17: Model light curve of a 1 km diameter KBO at 40 AU occulting stars of 0, 50, and 80 μas . From Pass, Metchev, Brown, and Beauchemin (2018, fig. 3).

711 A fully parameterized occultation event is passed to an algorithm constructed by Mills (2018).
 712 This algorithm generates (fig. 17) and samples (fig. 18) the associated light curve using the work of
 713 Pass et al. (2018). The light curve is then populated with noise (predominantly scintillation noise
 714 which is significant for high cadence photometry. See: Osborn, Föhring, Dhillon, & Wilson, 2015).

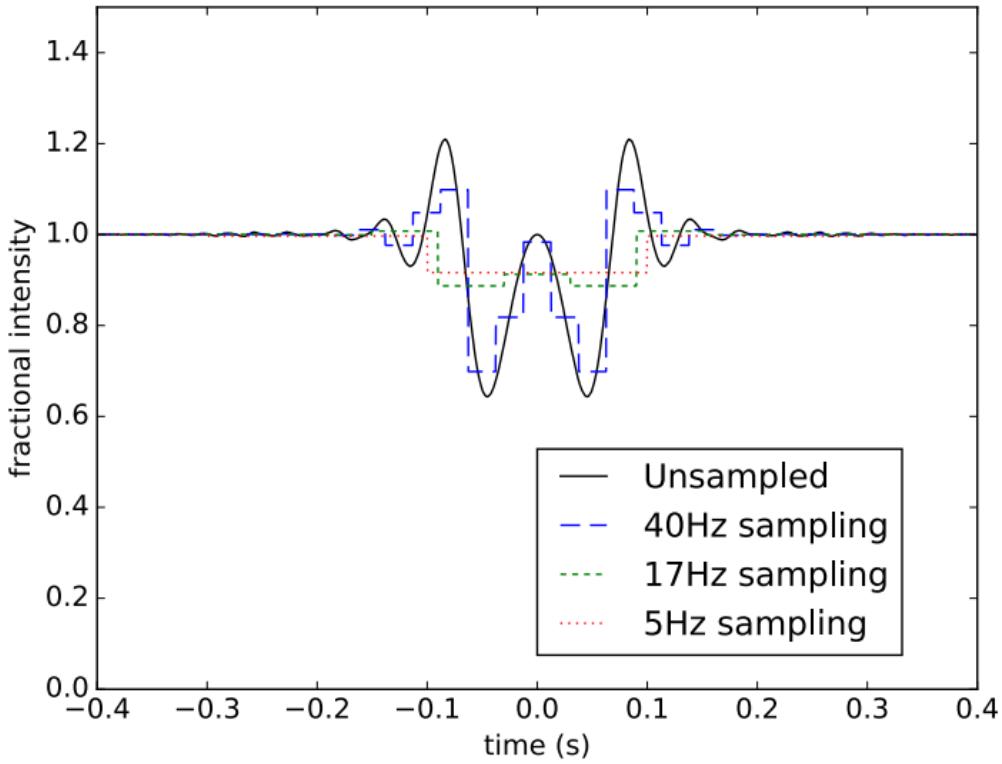


Figure 18: Model sampled light curve of occulting KBO for sampling rates of 40 Hz, 17 Hz, and 5 Hz. From Pass, Metchev, Brown, and Beauchemin (2018, fig. 4).

715 A pre-constructed kernel set spanning the detectable range of parameters for occultation events is
 716 used as a comparative library against which the current sampled, noisy light curve must be matched
 717 (fig. 19).

718 There is no detection for light curves which fail to match to kernels in the set. In order to assess
 719 the effect of statistical noise fluctuation on detection probability, the matching is performed many
 720 times. If a light curve is matched to a kernel only once the event is immediately assigned a 100%
 721 miss rate.

722 For all other cases, the miss rate is the fraction $1 - \frac{N_k}{N_n}$ where N_k is the number of iterations kernel
 723 matches were found and N_n is the total number of iterations performed. This value is stored along
 724 with the event and event parameters in the simulation output. Mills (2018) found that $N_k = 100$
 725 iterations are sufficient to consistently estimate the likelihood of an event detection.

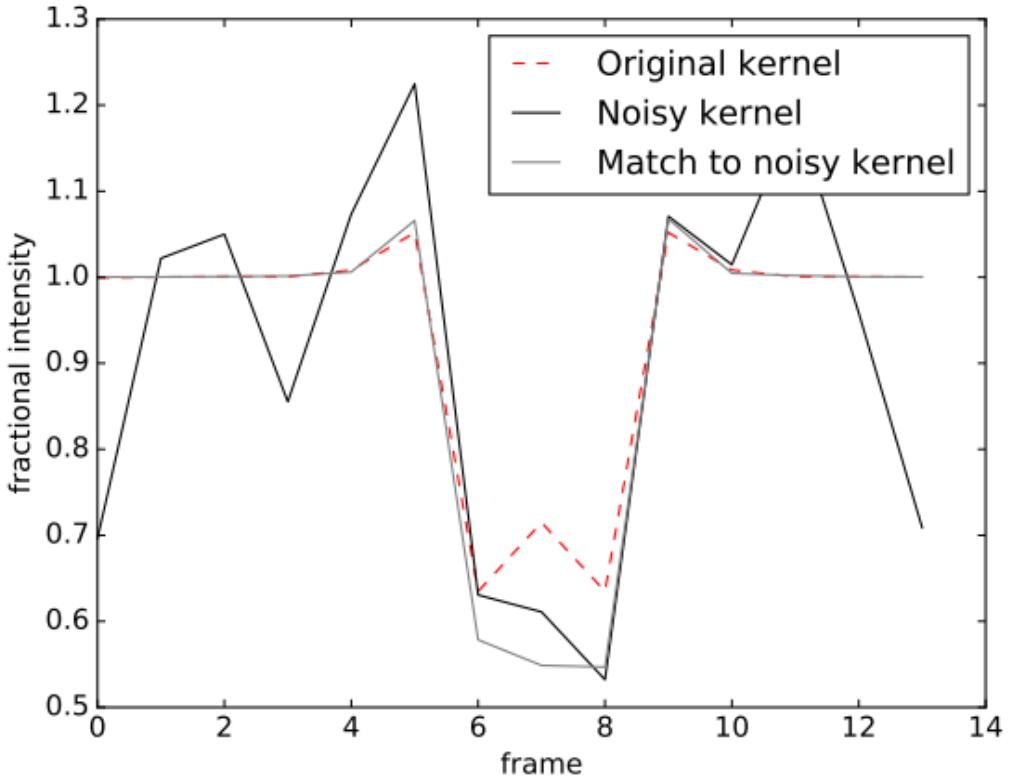


Figure 19: Kernel fitting of a noisy simulated occultation light curve. The light curve is consistent with a 2.3 km-wide KBO occulting a $10 \mu\text{as}$ diameter star at 40 AU with an impact parameter of $b = 200$ m. From Pass, Metchev, Brown, and Beauchemin (2018, fig. 9).

726 3.3 Simulating the observations

727 Each of the top fields were simulated for the detection of occultation events. Here some justification
 728 is presented for the choice of the initial simulation parameters. The heliocentric distance region
 729 considered is the classical population region of $42 \text{ AU} \leq r \leq 48 \text{ AU}$.

730 3.3.1 Choosing parameters: KBO diameter limits D_k and D_0

731 The KBO diameter range examined by the simulation runs from a choice of D_0 up to the break
 732 diameter D_k . Lower values of D_k increase the sky surface density of KBOs below this break diameter.
 733 The uncertainty in the literature associated with this value (section 3.1.1) indicates that a good
 734 approach is to make a conservative estimate (i.e. a larger value of D_k) and to adjust it downward
 735 as necessary. For the initial simulation, $D_k = 90$ km is set.

736 As *Colibri* begins its campaign, the simulation input variables can be tuned to better match
 737 any detections or lack thereof. The value of D_k represents one such tuning as a possible method
 738 to account for a higher-than-simulated rate of detection. Since the value of D_k has a less dramatic
 739 effect on KBO sky surface density than even small adjustments to q (fig. 20; fig. 21), it provides a
 740 valuable tool for fine-tuning the model concurrent with the main observing campaign.

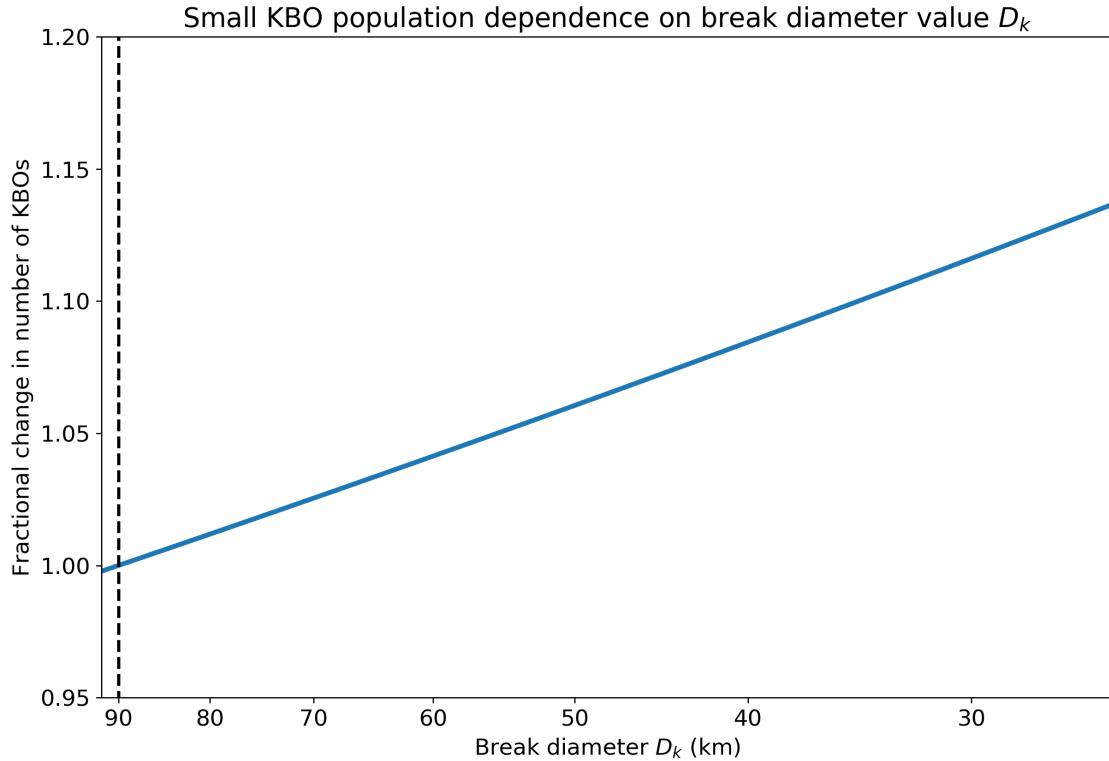


Figure 20: Small KBO population dependence on the value of adopted break diameter D_k for constant $q = 3.6$, $c = 2.6$, $D_0 = 0.5$ km. The number of KBOs is normalized to $D_k = 90$ km, the value adopted for the simulation, and is indicated by the dashed line.

741 The choice must also be made for the lower limit of KBO diameter to model, D_0 . Here a practical
 742 approach is to consider fig. 10, which indicates that there are no stars in the top field with angular
 743 diameters of $\lesssim 20 \mu\text{as}$ (brighter than $G = 11.5$ mag). For a KBO at ~ 40 AU this corresponds to
 744 a physical diameter of ~ 700 metres. Since the KBO disk should be larger than the stellar disk to
 745 produce a strong diffraction effect (section 1.2) some buffer room should be added to investigate the
 746 lower diameter limit of detection.

747 While there is no particular guidance on what this should be, the simulation output itself will

748 provide useful feedback. If the adopted value of D_0 is too large, a number of simulated detectable
 749 occultation events will occur with the minimum possible KBO diameter. If D_0 is small enough
 750 that the smallest simulated KBOs do not produce any detectable occultation events, this is an ideal
 751 situation. There is no disadvantage to the accuracy of the simulation if D_0 is much smaller than
 752 need be.

753 For the initial simulation, a value of $D_0 = 500$ m is adopted. As section 4 will show, there are
 754 no expected detections at diameters of this value and so smaller diameters need not be considered.

755 **3.3.2 Choosing parameters: power law indices**

756 There have been a number of attempts to constrain the value of q based on SSO surveys and optical
 757 detection of small $\lesssim 50$ km-diameter KBOs (e.g. Schlichting et al., 2012). A detection and analysis of
 758 reported detections by Schlichting et al. (2012) are used to justify a slope of $q = 3.6 \pm 0.2$ assuming
 759 small KBOs follow a uniform ecliptic latitude distribution (as is the case here).

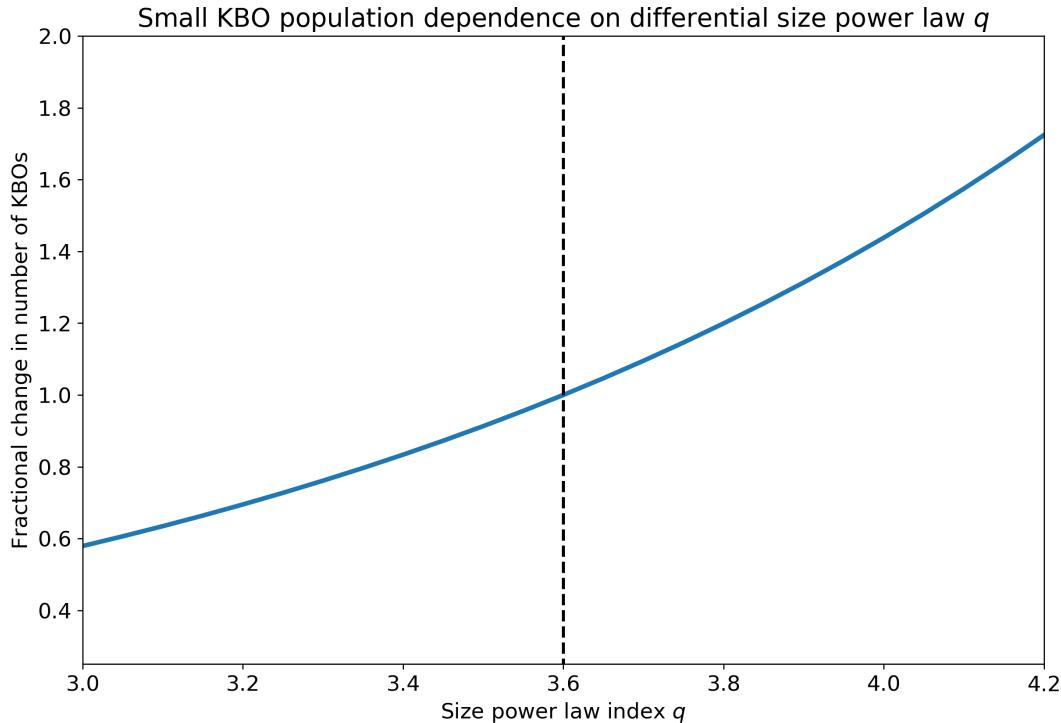


Figure 21: Small KBO population dependence on the value of differential size power law index q for constant $c = 2.6$, $D_k = 90$ km, $D_0 = 0.5$ km. The number of KBOs is normalized to $q = 3.6$, the value adopted for the simulation, and is indicated by the dashed line.

This value falls within the upper limit of $q \leq 3.82$ reported by Zhang et al. (2013) and lies near the expected $q = 3.5$ value indicating small KBOs are undergoing material strength-dominated collisional processes (Pan & Sari, 2005; Schlichting et al., 2012).

In particular, the analysis of Schlichting et al. (2012) takes into account two confirmed SSO events consistent with ~ 500 m diameters in the $40 \leq r \leq 45$ AU range; very similar to the parameter space *Colibri* will probe. Nonetheless, *Colibri* detection rates will inform the suspected value of q as small changes to this parameter have a substantial impact on the population size of KBOs (see fig. 21).

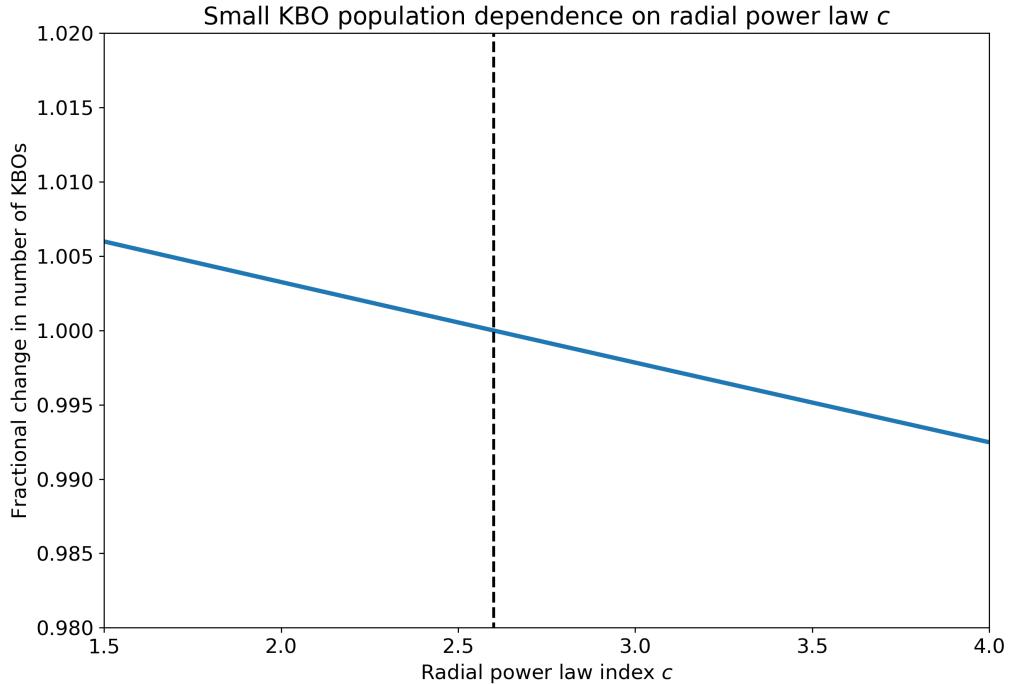


Figure 22: Small KBO population dependence on the value of radial power law index c for constant $q = 3.6$, $D_k = 90$ km, $D_0 = 0.5$ km. The number of KBOs is normalized to $c = 2.6$, the value adopted for the simulation, and is indicated by the dashed line.

The choice of radial power law index c is not very consequential. As fig. 22 shows, for large variation of the index there is little response in the KBO population size. Thus the approximation of $c \approx q - 1$ (section 3.1.2) will be used and a value of $c = 2.6$ adopted.

With the adoption of these values, the normalization constant from eq. (9) is computed to be $A = 2.55 \times 10^{10}$ AU^{4.2} and the number of KBOs per field is $N = 1.07 \times 10^7$.

772 **3.3.3 Running the initial simulation**

773 The output of the field search algorithm is a stored copy of the top fields found and the stars they
774 contain. This serves as input data for the simulation. The simulation will monitor all stars in this
775 field down to some limiting magnitude, not just angularly small ones. This avoids imposing an
776 artificial size cutoff that may bias the results and does not impact the simulation runtime. The
777 simulation and its output is detailed in section 4.

778 4 Results and Discussion

779 The results of the field search (section 2) are discussed in section 4.1. The results of the simulation
 780 of occultation events (section 3) are discussed in section 4.2. Together these results are synthesized
 781 into the *Colibri* observing campaign, presented in section 4.3 along with estimates of expected event
 782 detection rates.

783 4.1 *Colibri* fields of observation

784 4.1.1 Results of *Colibri* field search

785 The initial field search algorithm was performed on the 80 129 valid *Gaia* DR2 stars with a magnitude
 786 limit of $G \leq 11.5$ mag, an angular size limit of $\theta \leq 100 \mu\text{as}$, and an ecliptic longitude spacing of
 787 $\Delta\ell = 12^\circ$. 24 field candidates were returned and are listed in table 3.

Field No.	Field Centre (ℓ, b)	$N_{\text{stars}}^{\leq 100 \mu\text{as}}$	N_{stars}	Field No.	Field Centre (ℓ, b)	$N_{\text{stars}}^{\leq 100 \mu\text{as}}$	N_{stars}
1	(273.55, 4.75)	240^{+32}_{-29}	3952	13	(223.91, 4.40)	58^{+3}_{-5}	243
2	(287.48, 4.34)	137^{+13}_{-10}	1268	14	(332.10, -1.77)	53^{+2}_{-6}	274
3	(87.67, -2.60)	134^{+7}_{-21}	918	15	(42.68, 1.38)	50^{+4}_{-2}	260
4	(102.07, 1.44)	130^{+4}_{-15}	819	16	(8.97, -1.87)	49^{+2}_{-4}	162
5	(127.19, 1.03)	128^{+6}_{-10}	448	17	(23.56, -1.36)	49^{+2}_{-7}	178
6	(256.53, -4.62)	124^{+27}_{-11}	1455	18	(140.47, -2.25)	48^{+1}_{-3}	224
7	(59.87, 4.29)	94^{+16}_{-9}	423	19	(208.29, 0.65)	47^{+4}_{-3}	213
8	(114.89, -1.55)	80^{+8}_{-7}	465	20	(187.36, -1.50)	46^{+3}_{-3}	196
9	(72.31, -3.66)	78^{+7}_{-2}	440	21	(359.07, 0.37)	45^{+3}_{-2}	185
10	(244.10, -4.38)	75^{+14}_{-8}	587	22	(155.09, -2.85)	43^{+1}_{-1}	191
11	(301.36, -3.59)	74^{+5}_{-6}	461	23	(345.91, -2.20)	42^{+3}_{-5}	222
12	(316.84, 2.05)	66^{+1}_{-4}	342	24	(167.43, -3.00)	40^{+2}_{-1}	158

Table 3: The final 24 *Colibri* field candidates from the search algorithm.

788 4.1.2 Discussion of the importance of *Colibri*'s limiting magnitude

789 Of interest and potential impact is the final limiting magnitude of the *Colibri* array. Current
 790 estimates place this around $G \sim 11.5$ mag due to the large SNR required to confirm occultation

events. However, as the prototype array element becomes operational this value will likely change in accordance with test imaging data. Therefore it is a matter of prudence to consider how different limiting magnitudes may impact the initial campaign fields.

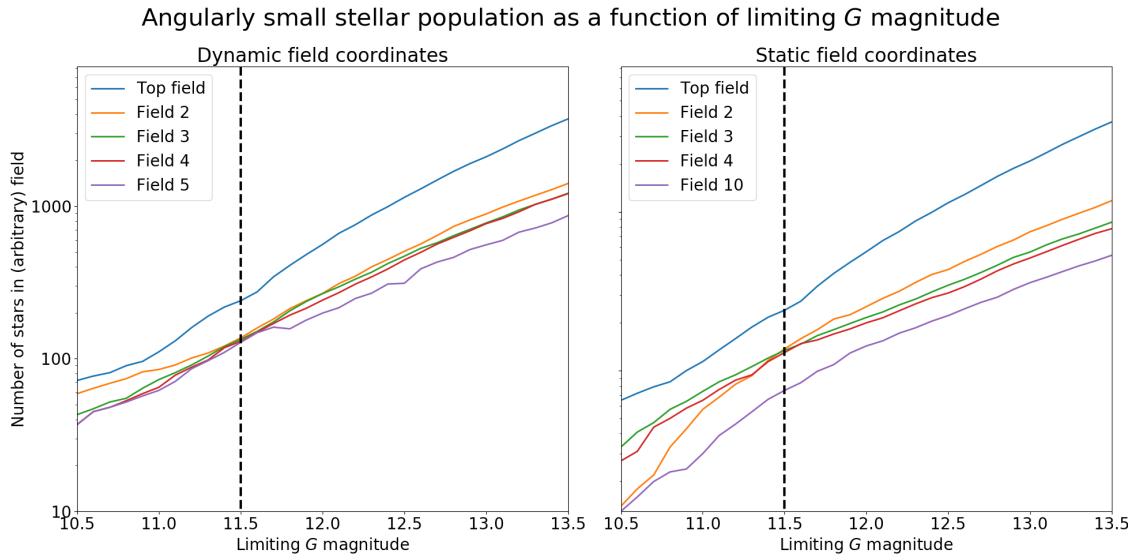


Figure 23: Stellar field populations ($\leq 100 \mu\text{as}$) as a function of limiting G magnitude. *Left:* Field candidates can have any coordinates so long as they maximize the angularly small population. *Right:* Fixed-coordinate field candidates for $G \leq 11.5$ mag and their dependence on *Colibri*'s operating limiting magnitude.

Figure 23(a) details the dependence of the top 5 field populations on the limiting magnitude of the system, where the top field coordinates represent the largest concentration of stars smaller than $100 \mu\text{as}$. There is a steep number dependence due to the logarithmic nature of the magnitude system. The dependence of angularly small stars on the limiting G magnitude of *Colibri* generally follows a logarithmic fit of power-index 0.61.

This is rather steep: for any field a $G \approx 13$ mag limit will allow *Colibri* to monitor stars ~ 4 times dimmer than at a $G \approx 11.5$ mag limit, but would increase the number of angularly small stars in the top field by a factor of ~ 8 . Thus any measure taken to improve the depth of observation with *Colibri* would greatly benefit the expected event detection rate as each incremental magnitude gained provides a wealth of new small stars.

Figure 23(b) details the limiting G magnitude dependence for the top fields, where their coordinates are fixed to the locations optimal for $G \leq 11.5$ mag. The dependence is nearly the same for the case where the field coordinates were not fixed, following a logarithmic fit of power-law index

807 0.62.

808 This indicates the dominance of limiting magnitude over the stellar population: if the top field
809 truly represented an unusual overdensity it would resist this trend for some range of magnitudes.
810 The opposite effect can be seen for field 2 in fig. 23(b), where for bright magnitude limits there are
811 too few stars to maintain the field population rank.

812 At each magnitude step in computing fig. 23(a) the chosen field coordinates were also monitored
813 to observe their behaviour (see section 2.4.1; table 2). Unlike the case of minor variation in the
814 angular size limit, each 0.1 magnitude produced a different enough density such that no field ever
815 retained unique coordinates.

816 Thus field optimization should be considered an ongoing process until a definitive operational
817 magnitude limit for *Colibri* is adopted, as even the optimal field's overdensity in one magnitude is
818 non-optimal in nearby magnitudes. Whether this monopoly of magnitude over population continues
819 unabated to stars dimmer than $G \lesssim 13.5$ is questionable, but would be beyond the domain of *Colibri*.

820 **4.1.3 Evaluating the *Colibri* fields**

821 Each of the fields must be considered in turn for its stellar sky surface density and its location on
822 the sky to determine its viability in the *Colibri* campaign. By examining the candidate fields of
823 table 3 the final fields of observation are thus determined and listed in table 4. They are shown in
824 a sky stellar density plot in fig. 24.

825 The ranking process consists of three guiding assumptions. First, because *Colibri* is not expected
826 to detect many events consistent with occultations of $\theta \gtrsim 100 \mu\text{as}$ stars, the stellar sky surface density
827 considered should be only that of angularly small stars. Second, as both this surface density and the
828 projected KBO velocity are proportional to the number of detectable events, higher-ranked fields
829 should only be switched out for lower-ranked fields when the fractional change in projected velocity
830 (eq. (11)) is larger than the fractional change in field population. Finally, since the event rate is also
831 proportional to the length of continuous observation, the duration of a night's observation must also
832 be weighed as a potential gain or loss when switching fields.

833 Table 3 also suggests that the top field and its immediate successors will provide the majority of
834 potential for occultation detection due to their population sizes. Therefore it will also be necessary to
835 prioritize the next-optimal fields to be observed at opposition, where the maximum detectable event

rate lies. The small populations of further candidates and their poor expected contributions to the annual event rate justify placing these (where possible) during months of above-average cloudiness and off-opposition if necessary.

It is clear from table 4 that the top field should be considered the optimal field due to its large population. Although its ecliptic latitude lies at $b = 4.7^\circ$ which may decrease the KBO sky surface density due to a gradual ecliptic falloff (Elliot et al., 2005), this is likely to be by much less than a factor of 2 (section 2.1.1). Even when considering observations off-opposition, eq. (11) implies that there is at most a 13% impact on the event rate for observations at a solar elongation of $\epsilon \pm 30^\circ$ from opposition.

Field No.	$N_{\text{stars}}^{\leq 100 \mu\text{as}}$	N_{stars}	Observing Period	Field Centre (ℓ, b)	Field Centre (α, δ)
1 (1)	240^{+32}_{-29}	3952	June 1 - July 31	(273.55, 4.75)	(273.735, -18.638)
2 (4)	130^{+4}_{-15}	819	January	(102.07, 1.44)	(103.263, 24.329)
3 (5)	128^{+6}_{-10}	448	February	(127.19, 1.03)	(129.869, 19.474)
4 (7)	94^{+16}_{-9}	423	August	(59.87, 4.29)	(56.684, 24.313)
5 (12)	66^{+1}_{-4}	342	December	(316.84, 2.05)	(318.657, -13.830)
6 (13)	58^{+3}_{-5}	243	May	(223.91, 4.40)	(222.785, -11.810)
7 (14)	53^{+2}_{-6}	274	September	(332.10, -1.77)	(334.741, -12.383)
8 (15)	50^{+4}_{-2}	260	November	(42.68, 1.38)	(39.791, 16.953)
9 (16)	49^{+2}_{-4}	162	October	(8.97, -1.87)	(8.974, 1.834)
10 (18)	48^{+1}_{-3}	224	March	(140.47, -2.25)	(142.138, 12.533)
11 (19)	47^{+4}_{-3}	213	April	(208.29, 0.65)	(206.512, -10.259)

Table 4: Optimal fields of observation for the *Colibri* campaign. The field number column indicates the final ranking of the field and the associated field number from table 3 in parentheses.

Simulation work (section 4.2) demonstrates that the top field is likely to be responsible for the major contribution to the annual event rate.

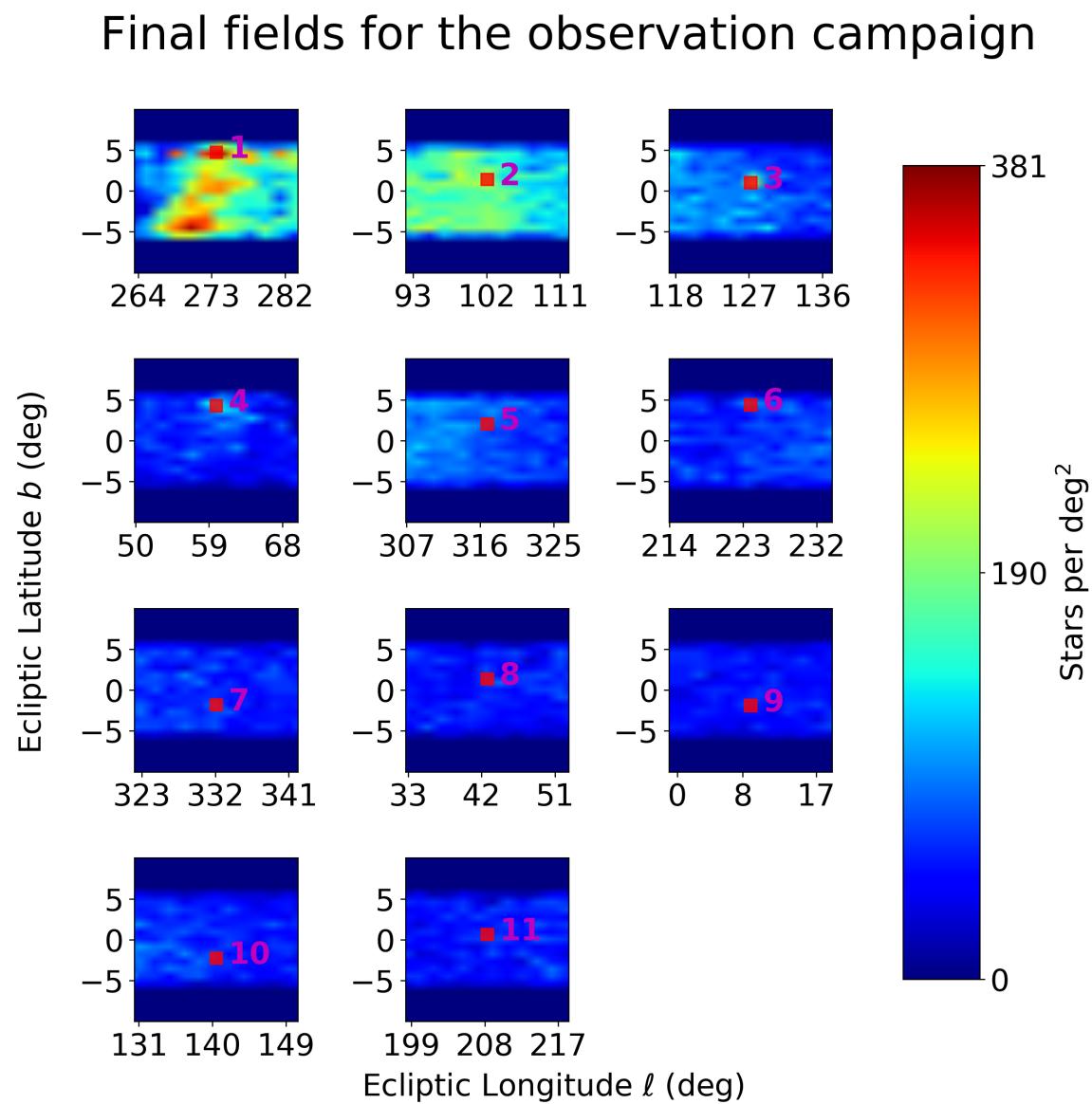


Figure 24: Locations of the 11 *Colibri* campaign fields on the sky. The red box indicates the field and field of view.

847 **4.2 Simulation of occultation events**848 **4.2.1 Core results**

849 The simulation was performed on the 11 *Colibri* fields (table 4) with the simulation parameters
 850 listed in table 5. The results are listed in table 6. Each field was simulated for 120 6-hour intervals,
 851 totalling $\sim 2.6 \times 10^6$ seconds (30 days) of continuous clear-sky observation. Each event is assigned
 852 a likelihood of detection. The majority of events were determined to be either certainly detectable
 853 or certainly undetectable, but for events with an intermediate probability a confidence level (c.l.) of
 854 95% was required to classify an event as a detection.

Parameter	Value
FOV (ℓ)	1.48°
FOV (b)	1.48°
D_0	0.5 km
D_k	90 km
r_{\min}	42 AU
r_{\max}	48 AU
c	2.6
q	3.6

Table 5: Model and simulation parameters for the initial simulation.

Field No.	$N_{\text{stars}}^{\leq 11.5 \text{ mag}}$	$N_{\text{stars}}^{\leq 100 \mu\text{as}}$	$N_{\text{evt}}^{95\%}$	$N_{\text{evt}}^{50\%}$	$N_{\text{evt}}^{\text{total}}$	$\mu \text{ (day}^{-1}\text{)}$
1	3952	240	131	412	3596	4.4
2	819	130	61	234	1847	2.0
3	448	128	72	212	2156	2.4
4	423	94	55	175	1296	1.8
5	342	66	48	142	1158	1.6
6	243	58	34	121	1091	1.1
7	274	53	31	139	982	1.0
8	260	50	27	107	725	0.90
9	162	49	19	97	696	0.63
10	224	48	21	86	412	0.70
11	213	47	18	82	457	0.60

Table 6: Event detection results for the final *Colibri* fields of observation. N_{stars} is the number of stars of any angular size in the field. $N_{\text{stars}}^{\leq 100 \mu\text{as}}$ is the number of stars less than the field-search cutoff (in this case 100 μas). $N_{\text{evt}}^{95\%}$ is the number of events *Colibri* is expected to detect. $N_{\text{evt}}^{50\%}$ is the number of events *Colibri* has a better than 50% chance of detecting in simulations. $N_{\text{evt}}^{\text{total}}$ is the number of events that occurred in the field. μ is the estimated event rate per 24 hours of continuous observation.

855 The final column μ of table 6 represents the daily detectable event rate. For the optimal field
 856 this is 4.4 day^{-1} for a full 24-hour day. Thus for each 6-hour cloudless observing period the optimal

857 field contributes on average 1 event. This is the highest event rate among all fields by a factor of 2.
858 Therefore it should be expected that the optimal field contributes the largest single share of events
859 per year.

860 **4.2.2 Discussion of the stellar angular size cutoff and detectable KBOs**

861 The relationship between the detection likelihood of an event, the KBO diameter, and the angular
862 size of the occulted star is shown in fig. 25. It shows that events which are nearly certain to be
863 detected occur most commonly for stars of angular diameters $\lesssim 100 \mu\text{as}$ with detections becoming
864 more sparse at larger angular sizes. Figure 26 shows a histogram of the number of detectable events
865 for different stellar angular diameters. Here it can be seen that stars above $\sim 100 \mu\text{as}$ are not
866 responsible for many events. Figure 27 shows that the vast majority of annual detections come from
867 stars below this size. Since the field search algorithm is designed to locate fields that produce many
868 events, an upper angular diameter limit of $100 \mu\text{as}$ is justified to produce the most lucrative fields.

869 Figure 25 shows that KBOs of all diameters produce detectable events across the range of stellar
870 angular diameters, but no KBOs of diameters $D \lesssim 1.2 \text{ km}$ are detectable. The smallest detectable
871 KBOs have $D \approx 1.2 \text{ km}$ and occur for stars within a few dozen μas of $60 \mu\text{as}$ (fig. 27).

872 Figure 25 also indicates that many KBOs are detectable in the small-KBO regime with diameters
873 $1.2 \lesssim D \lesssim 15 \text{ km}$. Detections at such sizes will be useful in constraining the values of q and D_k by
874 comparing the event frequency to the simulated event frequency for those diameters.

875 Figure 25 indicates that some KBOs with diameters $D \gtrsim 5 \text{ km}$ are not detectable. These events
876 have longer durations due to the large occulter sizes and the occultation kernel widths used in the
877 simulation are not sufficiently wide to accommodate such large objects. The result is a poor sampling
878 of the star's light curve before and after the event, producing difficulty for the algorithm to find a
879 matching kernel. These long duration events are likely to make an event more detectable, not less,
880 due to the length and depth of the occultation light curve. Therefore these events are expected to
881 be detected by *Colibri*.

Detectability of KBO diameters at varying stellar angular diameters

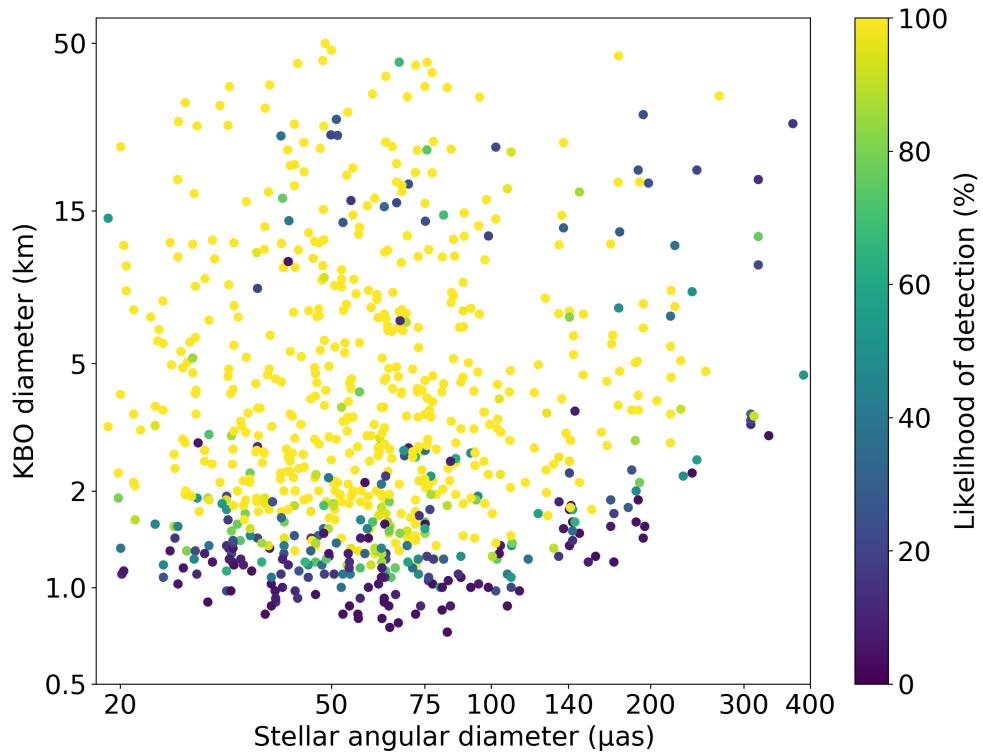


Figure 25: KBO diameter versus angular size of occulted star with detection likelihoods on a colour scale. Some KBOs with diameters $D \gtrsim 5$ km go undetected in simulations but are expected to be detectable. This is an expected artifact of the simulation.

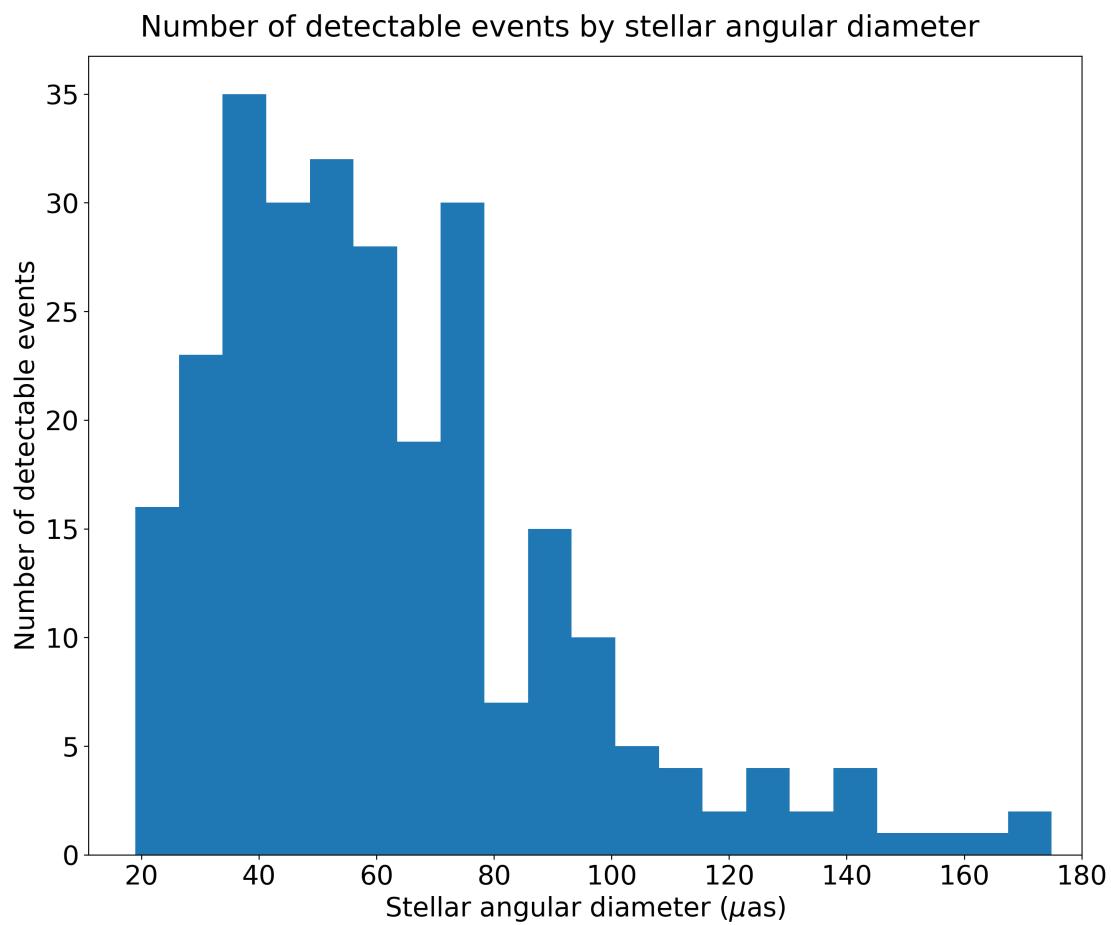


Figure 26: Stellar angular sizes for detectable events.

882 **4.3 Synthesized results**

883 In order to put the simulated detection rate and field selection into a more practical context, it
884 is necessary to examine the effects of seasonal cloud cover and extinction on *Colibri*'s observing
885 program. This is discussed in section 4.3.1. The error in detection rates is discussed in section 4.3.2,
886 and final detection rate estimates are presented in section 4.3.3.

887 **4.3.1 Extinction and cloud cover effects**

888 Operating in the *R*-band one can expect a ~ 0.1 magnitude / airmass contribution from atmospheric
889 extinction. An airmass of ~ 3 occurs around 20° altitude, corresponding to a 0.3 magnitude loss of
890 dim stars. For the chosen fields the relevant behaviour is shown in fig. 23(b).

891 With a slope of 0.62, a loss of 0.3 magnitudes down from $G = 11.5$ mag corresponds to a factor
892 of ~ 1.5 decrease in the sky surface density of $\lesssim 100 \mu\text{as}$ stars. Since this density is proportional to
893 the event rate, observing at 20° altitude can cut the event rate by as much as 33% for fields that
894 can also be observed near the zenith. For fields that remain low in the sky through the night the
895 potential loss is lower, as the limiting magnitude at the meridian is already extincted.

896 Because the fields are all chosen to be within $-5^\circ \leq b \leq 5^\circ$ and the Earth has a significant tilt with
897 respect to the ecliptic plane, some fields are more extincted than others even at their best position
898 due to the revolving nature of the ecliptic. Therefore it is appropriate to examine the average time
899 spent by a field centre at each 5° altitude increment to determine the cumulative average extinction
900 over the observing period of the field.

901 These extinction estimates are presented in table 7. The summer fields are clearly impacted by
902 the low altitude of the ecliptic, with most of the observing time in the winter months being spent
903 at high altitudes for most of the night. Nonetheless the field population difference between winter
904 and summer fields is such that the majority of detections are expected in the summer.

905 A good approximation for the airmass X is $X = \sec(z)$, where z is the zenith distance, as it is
906 roughly valid down to $\sim 20^\circ$. Combined with the data of table 7 a total fractional event rate loss can
907 be found by summing over the 5° increments. The equivalent fractional sky surface density retained
908 through atmospheric extinction is given in table 8.

909 In addition to nightly atmospheric extinction it is practical to consider the effects of cloud cover.
910 Cloud cover data is available from NASA's Modern-Era Retrospective analysis for Research and

Field No.	20°-25°	25°-30°	30°-35°	35°-40°	40°-45°	45°-50°	50°-55°	55°-60°	60°-65°	65°-70°
1	0.5	0.5								
2	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.22
3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.24	
4	0.5	0.5								
5	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.18
6	0.33	0.33	0.33							
7	0.21	0.21	0.57							
8	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.22		
9	0.13	0.13	0.13	0.13	0.25	0.25				
10	0.11	0.11	0.11	0.11	0.11	0.11	0.16	0.21		
11	0.17	0.17	0.17	0.33	0.17					

Table 7: Fractional time per observing season spent in the given altitude ranges.

Field No.	Fraction retained
1	0.761
2	0.844
3	0.837
4	0.761
5	0.848
6	0.781
7	0.794
8	0.838
9	0.821
10	0.833
11	0.808

Table 8: Assuming the limiting magnitude of 11.5 is for a star at the zenith, this table shows the fractional population of stars retained through extinction for each field per year.

Field No.	Clear nights
1	0.500
2	0.160
3	0.200
4	0.550
5	0.195
6	0.385
7	0.530
8	0.300
9	0.430
10	0.300
11	0.350

Table 9: Average fractional clear nights cover over the observing period per field. Data acquired from the Global Modeling and Assimilation Office (2015).

911 Applications, v.2 (MERRA-2) (Global Modeling and Assimilation Office, 2015). The data obtained
912 was for London, ON from 1980 through the end of 2015, and generally indicated the most clear
913 nights occur between July and September.

914 This is fortunate for the *Colibri* program as the top field is at opposition on July 15 and is
915 observable within the $\pm 30^\circ$ solar elongation limit from June 1 - July 31. Within this time period
916 the field can only be observed for ≈ 5 hours per night above the 20° altitude limit ($\sim 3x$ airmass).
917 Nonetheless the prevalence of clear nights and field population will make this the most lucrative
918 field of the year for *Colibri*.

919 Data from MERRA-2 in table 9 indicates the usable time per field season. While extinction can
920 affect the fractional expected event rate by $\sim 24\%$ in the worst case (table 8) a cloudy night will
921 result in no detections at all. Thus it is expected that the cloud behaviour patterns play a larger
922 role in the eventual detected event rate than atmospheric extinction.

923 **4.3.2 Errors in detection rate**

924 The primary sources of error for the detection rate will be the number of stars per field producing
925 detectable events and the variability of cloud cover from table 9. Although stars with diameters
926 $\gtrsim 100 \mu\text{as}$ can produce detectable events, the number of events produced by these stars is small
927 compared to those produced by stars with diameters $\leq 100 \mu\text{as}$ (fig. 27). It is also expected that the
928 event rate is proportional to the number of stars producing detectable events. Therefore the error
929 in the number of stars with diameters $\leq 100 \mu\text{as}$ in a field provides a useful first approximation to
930 the uncertainty of the number of detections in that field. (See table 4 for the errors in the number
931 of stars per field.)

932 Additionally, the cloud cover data from MERRA-2 has a reasonably consistent error of 10% for
933 the fraction of all clear nights per field. This gives an estimated error in the detection rate of:

$$\frac{\delta\mu^*}{\mu^*} = \sqrt{\left(\frac{\delta N_*^{\leq 100 \mu\text{as}}}{N_*^{\leq 100 \mu\text{as}}}\right)^2 + 0.01} \quad (12)$$

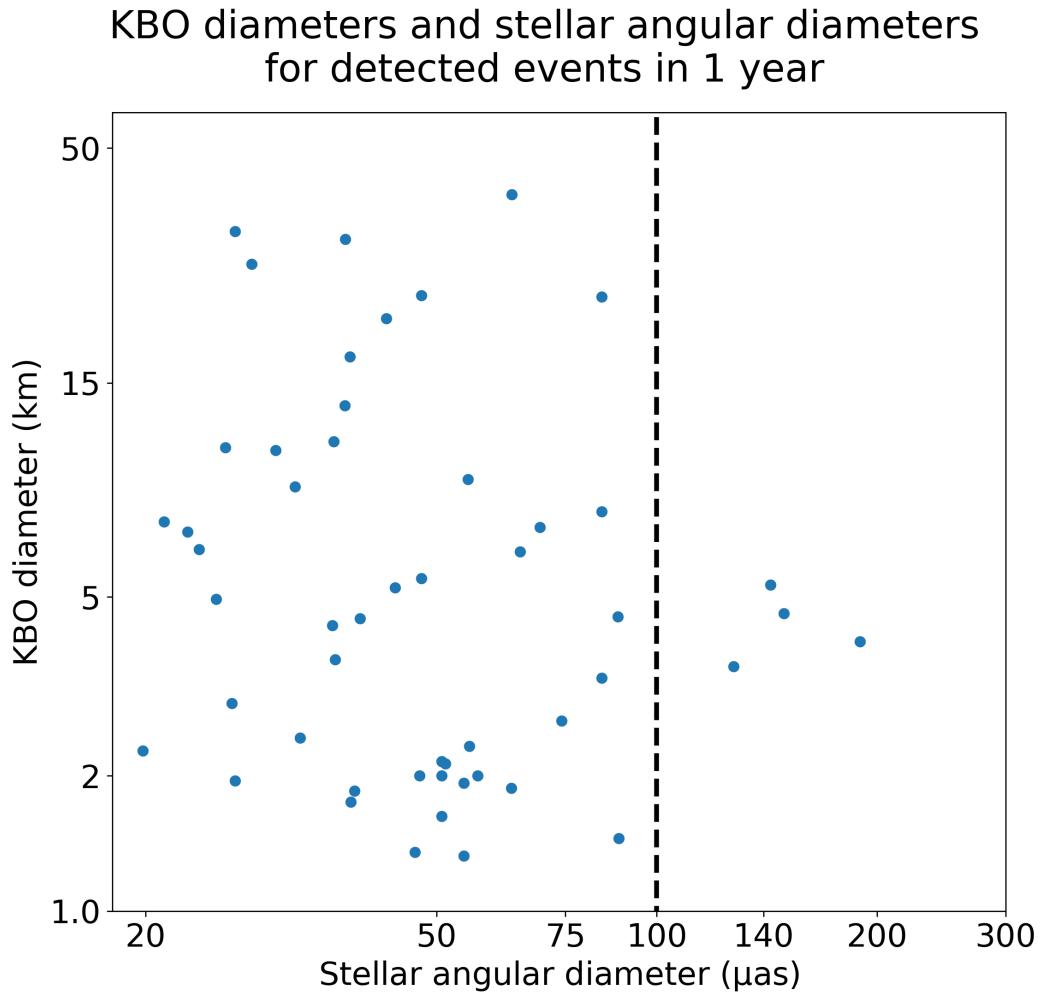


Figure 27: KBO diameter versus angular size of occulted star for annual detected events. The dashed line represents the $100 \mu\text{as}$ cutoff applied to the field search sample. Only 4 events are produced by stars larger than this size. To maximize the chance of detecting events *Colibri* must observe stars likely to produce events. This motivates the need and this value for an angular size cutoff in the field search algorithm.

934 **4.3.3 Estimated event detection rates**

935 With estimates of atmospheric extinction and cloud cover the expected event rate can be combined
 936 from table 6 to yield estimates of the *Colibri* detection rate. These are presented in table 10. As
 937 expected, the optimal field is responsible for the majority of detected occultation events, though it
 938 accounts for about 40% of the expected events per year.

Field No.	Days of observation	Fractional clear nights	Fractional stars after extinction	μ^* (yr^{-1})
1	12.71	0.500	0.761	21^{+4}_{-3}
2	14.85	0.160	0.844	$4.1^{+0.4}_{-0.6}$
3	11.67	0.200	0.837	$4.7^{+0.5}_{-0.6}$
4	6.46	0.550	0.761	$5^{+1}_{-0.7}$
5	14.85	0.195	0.848	$3.9^{+0.4}_{-0.5}$
6	7.75	0.385	0.781	2.6 ± 0.3
7	7.50	0.530	0.794	$3.3^{+0.3}_{-0.5}$
8	11.25	0.300	0.838	2.5 ± 0.3
9	9.04	0.430	0.821	$2.0^{+0.2}_{-0.3}$
10	10.33	0.300	0.833	1.8 ± 0.2
11	8.75	0.350	0.808	1.5 ± 0.2

Table 10: Final estimates for the expected detected number of events per observing field per year. The expected events are rounded. μ^* is the expected detected event rate per field per year. The total annual event rate across all fields is estimated to be $\mu^* = 52 \pm 4$ events per year for *Colibri*.

939 Table 10 demonstrates the expense of a cloudy location on the expected detected event rate.
 940 Because the field locations are already optimized for the limiting magnitude of *Colibri*, the only
 941 major solution to counter effects deleterious to the event rate is to increase the photometric depth
 942 of *Colibri*. However, as the array is still under construction, further testing to constrain the limiting
 943 magnitude is required.

944 **4.3.4 Future work**

945 Due to the magnitude dependence of the fields listed in table 4 it will be necessary to periodically
 946 re-evaluate these selections, with increasing priority toward the start of the *Colibri* campaign. In
 947 particular, work done on constraining the operational limiting observing magnitude of *Colibri* will
 948 require fields to be re-assessed.

949 To that end, it would be valuable to incorporate the weather data and a sky model into the field
950 search algorithm, such that the final fields can be selected automatically by the computer. If such a
951 model were constructed and validated, it would greatly reduce the time to optimize field candidates
952 based on their likelihood to produce detections.

953 It is highly encouraged to keep these fields up to date with future *Gaia* releases as that dataset
954 has provided a robust resource for this selection process. At the time of this report the latest
955 available release is the DR2, but there are expected in total to be 3 further data releases by 2022-23.

956 As *Colibri* enters its initial observing phase, the simulation parameters should be modified to
957 best reproduce the observed detection rate. The simulation can be used as a tool to help constrain
958 these parameters and compare newly estimated values with other ongoing SSO survey results.

959 **5 Conclusion**

960 The optimal fields for the *Colibri* observing campaign have been found (table 4). These fields are
961 likely to provide the best event detection rate for the array at Elginfield. As discussed in section 4.1
962 the dominant factor affecting the choice of field is the limiting magnitude of the array. *Colibri* would
963 stand to greatly benefit if deeper observations can be made.

964 These fields were simulated against a KBO population as described in section 3. The results
965 of these simulations, taking into account the time of year, altitude of the field, cloud cover, and
966 atmospheric extinction indicate that most fields will not be lucrative, though all will contribute
967 events to the annual total (table 10). Cloud cover is minimized in the summer, but atmospheric
968 extinction is maximized due to the low altitude of the ecliptic. Once again the dominant factor
969 determining expected event rate is the number of angularly small stars available for monitoring, and
970 thus the limiting magnitude of *Colibri*.

971 Nonetheless the optimal summer field provides the best chance of detecting occultations. It is
972 therefore imperative to the operation of *Colibri* that the array remain operational throughout the
973 months of June and July. For a full year of operation *Colibri* can expect to detect 52 ± 4 occultations
974 of KBOs with diameters $1.2 \leq D \leq 50$ km.

975 The field search algorithm and simulation together provide a package for future re-optimization
976 of the observing parameters appendices A.1 and A.2. Further automation of the few manual tasks
977 that remain in these algorithms would make future simulation results more efficient to compile.

978 The simulation in particular offers a tool for concurrent analysis of *Colibri* results once the main
979 campaign begins. Differences between the observed event rate and the simulated event rate can
980 be examined in the context of observables (actual cloud cover, limiting magnitude from data as in
981 section 4.3.1) to investigate the accuracy of the KBO model population.

982 Particularly poorly constrained values like D_k and sharp-dependency variables like q can be
983 estimated with increasing precision as *Colibri* operates. Depending on the sizes of detected KBO
984 occulters it is also possible to reconstruct a KBO sky surface density as a function of KBO diameter
985 to compare with upper limits and other estimates from the literature. Thus the *Colibri* array is well
986 positioned to make detections in the SSO regime and to help constrain model parameters key to the
987 burgeoning SSO field.

988 A Continuity and Transition

989 The field search algorithm and occultation simulation are separate programs that can be run inde-
990 pendent of one another. The output of the field search is intended to be a quick-access small-size file
991 containing all the stars and field information, from which the simulation can read. The simulation
992 works on any stellar field, but is optimized to read from this file.

993 To ensure the field algorithm output can be fed directly into the simulation both programs
994 (detailed separately in appendix A.1 and appendix A.2) should be stored in the same directory with
995 the following directory structure:

```
- ./
  |- common/
  |- DATA/
  |- Fields/
  |- Simulation/
  |- __init__.py
```

996 This directory structure can be recreated anywhere but a copy of all files and data contained in
997 this structure is available on Mango at /mnt/data/rbloch/Colibri.

998 The **common/** folder is intended to hold the outputs of the two programs. For the field search
999 algorithm this is the default location to save the field data. For the occultation simulation this is
1000 the default location to save the simulated event data.

1001 The **DATA/** folder is intended to hold input data for the programs. At the time of this writing
1002 only the field search algorithm intakes data from external sources, in this case the *Gaia* DR2. This
1003 is the default location the program will search before asking for user guidance.

1004 The **Fields/** folder contains the field search algorithm and is detailed in appendix A.1.

1005 The **Simulation/** folder contains the occultation simulation and is detailed in appendix A.2.

1006 The **__init__.py** file is required in all top-level directories for cross-module imports. It is
1007 unnecessary in this folder at the moment except by convention, though its clones in the **Fields** and
1008 **Simulation** folders are strictly necessary for those programs to call on their custom modules. The
1009 presence of **__init__.py** will have no further discussion.

1010 A.1 Field search

1011 A.1.1 Storage

1012 The field search algorithm can be found at https://github.com/rpbloch/colibri_fields.

1013 The search algorithm can be downloaded as a self-contained archive, but will require a separate
1014 data file on which to draw. The data file used to find the initial fields is located on Mango
1015 at /mnt/data/rbloch/Colibri/DATA/Fields/Colibri_DR2.vot. The data structure of the field
1016 search algorithm follows, with code descriptions in appendix A.1.4.

```
- Fields/
  |- findfields.py
  |- utils.py
  |- __init__.py
  |- tools/
    |- datasort.py
    |- filehandling.py
    |- __init__.py
```

1017 A.1.2 Overview

1018 The field search algorithm operates out of the `Fields/` folder with the main code `findfields.py`.
1019 The field search algorithm searches the `DATA/` folder for the VOTable file `Colibri_DR2.vot`. If this
1020 file is not present, it will prompt the user to enter an alternate name: this name can be a relative
1021 path from the `Fields/` directory or can be an absolute path, terminating with the filename and
1022 extension in both cases. The default filename and storage location for input data can be changed
1023 within the `filehandling` module of the program.

1024 The field search algorithm then performs a nearest neighbour search of all the stars in the input
1025 data based on their ecliptic coordinates (see section 2.3). It returns a ranked list of candidate
1026 fields and iterates through them to exclude fields that are too close together to be worth separate
1027 observing time. The iteration occurs from the top field down so that the spaced-out fields are the
1028 most populous possible.

1029 The control variables of the search can be found at the top of the main program file. They are
1030 given with default values and explanations in table 11.

1031 **Note:** the variable REQUIRED_PARAMS is 'hidden' beneath all the data string declarations
1032 in `findfields.py`. It cannot be moved as it requires these definitions to be populated with data

Variable Name	Default	Description
NREGIONS	1000	Number of regions to find. The search terminates when it has enough top fields to span the sky. For high values this variable ensures the sky is spanned, but can be set low for more specific queries.
FOV	1.48	The field of view (in degrees) of the field. The program assumes a square field of view.
MAG_LIM	11.5	The limiting G magnitude of <i>Colibri</i> , given by the <i>Gaia</i> data column <code>phot_g_mean_mag</code> .
MAS_LIMIT	0.05	The upper angular stellar diameter size cutoff for the nearest neighbour search. Units are milliarcseconds (mas).
SPACING	(10., 0.)	The ecliptic longitude and latitude minimum spacings, respectively, within which field candidates are rejected in favour of more populous neighbours.
REQUIRED_PARAMS	[...]	The data column headers required to have valid data for a star to pass the initial quality check. The default names of these values are defined at the top of <code>findfields.py</code> and refer to the <i>Gaia</i> data columns <code>source_id</code> , <code>phot_g_mean_mag</code> , <code>parallax</code> , and <code>radius_val</code> .

Table 11: Default parameters for the field search algorithm. These are defined and set at the top of the `findfields.py` file.

1033 column names.

1034 The field search algorithm is currently tuned to input from the *Gaia* DR2 (see section 2.2). To
 1035 accommodate future data releases, the dictionary keywords which are associated with the *Gaia* data
 1036 column headers must be located and changed (if applicable).

1037 A.1.3 Dependencies

1038 The following external dependencies exist for this program:

1039 • **NumPy**: NumPy handles most math operations for the efficiency of their array format.

1040 • **SciPy**: The `scipy.spatial` package provides the `cKDTree` class for the search.

1041 • **Pickle**: Pickle is used to store the output of the field search.

1042 • **Astropy**: The `astropy.io` module provides the `votable` package used for reading the *Gaia* data.

1043 A.1.4 File descriptions

1044 A.1.4.1 findfields.py

1045 This is the main program file which controls the search variables (table 11) and executes the search.
1046 The data that is input is initially sorted: if a star does not contain valid entries for the specified
1047 headers, it is discarded immediately. Only stars within the magnitude limit, and with quantities
1048 producing an angular size estimate, are useful for the search (see section 2.2). The star data is then
1049 appended with the appropriate angular sizes and their errors. Stars above the angular size limit are
1050 removed from the data sample.

1051 The data is then run through the nearest neighbour search to provide the top field candi-
1052 dates' coordinates and populations. The search is executed in a single function in which the
1053 `scipy.spatial.cKDTree` class constructs an optimized k-d tree from the ecliptic coordinates of
1054 the star (see section 2.3.2). The direct results of the search are the coordinates of the contained
1055 stars of each field. These are sorted immediately by population and the ranked list is produced.

1056 This list is fed through another function that prevents fields from being chosen within the spacing
1057 limit set at the top of the program. The output of that function is the list of top candidate fields
1058 spaced by the minimum distance. With default values the fields will span the sky and thus self-
1059 enforce a limit to the number of fields returned. If the number of fields queried is low enough that
1060 there are a larger number of candidate fields, this is when the ranked list is truncated to produce
1061 the desired number of fields.

1062 Thus the optimized k-d tree produces only the field coordinates and the coordinates of its popu-
1063 lation. **No other associative information can be passed through the tree, by the nature**
1064 **of its data structure.** For this reason, when the program moves on to calculate the error, it
1065 cannot evaluate the magnitudes of the stars in the field because this information was not passed
1066 through the tree. Any pairing algorithm to re-associate the population's coordinates with the *Gaia*
1067 magnitudes takes longer than searching the tree again. Thus it is fastest to let the program redo
1068 the search down the tree.

1069 Since the stellar information in the fields does not pass through the tree, and this information
1070 is crucial to have, the program saves a modified subset of the original *Gaia* input instead. The
1071 modifications are the angular size estimates appended to the stellar data. These stars are a subset
1072 of the original data as they are restricted by data validity, angular size, and magnitude limit, and thus

1073 represent a much smaller file to store and load repeatedly. This file, along with the field candidate
1074 coordinates, are stored as output.

1075 **A.1.4.2 utils.py**

1076 This is a utility function module. It was more populated in previous versions of the search, but now
1077 contains the angular size calculation called on by `findfields.py`.

1078 **A.1.4.3 datasort.py**

1079 This module contains the sorting functions that stars undergo before being constructed into the
1080 optimized k-d tree. Each filter is segmented into its own function so that `findfields.py` can call
1081 on the singular umbrella function at the end.

1082 **A.1.4.4 filehandling.py**

1083 This module contains all the file handling protocols for the program. It contains the default file
1084 location assumptions and has functions to read the *Gaia* VOTables and to write the search output.

1085 **A.2 Occultation simulation**

1086 **A.2.1 Storage**

1087 The occultation simulation can be found at <https://github.com/rpbloch/colsim>. The search
1088 algorithm can be downloaded as a self-contained archive, but will require a separate data file on
1089 which to draw. The data file containing the field search output for the initial fields is located on
1090 Mango at `/mnt/data/rbloch/Colibri/common/regions.p`. The data structure of the occultation
1091 simulation follows, with code descriptions in appendix A.2.4.

1092 **A.2.2 Overview**

1093 The occultation simulation operates out of the `Simulation/` folder with the main code `runmodel.py`.
1094 The simulation is meant to take input from the output of the field search algorithm. Any data can
1095 be fed into it that mimics the *Gaia* data structure, but the advantage of using the search output
1096 is the simulation will prompt the user to choose the field to simulate. Each simulation takes one

```
- Simulation/
|- runmodel.py
|- collisions.py
|- __init__.py
|- models/
    |- fieldgen.py
    |- popgen.py
    |- kbogen.py
    |- __init__.py
|- support/
    |- colibri_sim_matching.py
    |- fresnelModeler.py
    |- __init__.py
```

1097 thread, so many simulations of different fields can run simultaneously. This will degrade the runtime
1098 of each simulation, but is still far faster than running them sequentially.

1099 The simulation then populates the field with the input stars. Using its support modules it
1100 constructs a KBO population across the field (section 3.1), and then examines the resulting geometry
1101 to identify event candidates (section 3.2.2). Any detected candidates are examined again in a high
1102 resolution manner to compute the impact parameter and verify the event (section 3.2.3). Events
1103 are stored in the top-level common/ folder as pickled dictionaries. All event data is stored: stellar
1104 angular size and G magnitude; KBO physical diameter, heliocentric distance, and projected velocity;
1105 the impact parameter and the coordinates and geometry of the collision.

1106 The control variables of the simulation can be found at the top of the main program file. They
1107 are given with default values and explanations in table 12 and most are model parameters from
1108 section 3.3.

1109 **A.2.3 Dependencies**

1110 The following external dependencies exist for this program:

1111 • **NumPy**: NumPy handles most math operations for the efficiency of their array format. It
1112 also provides valuable quick random selection algorithms.

1113 • **R-tree**: The rtree.index package is used to construct the r-tree and perform the spatial in-
1114 dexing search to detect KBO/star overlap.

1115 • **Shapely**: Shapely is used to perform the high resolution geometric analysis of candidate
1116 events.

Variable Name	Default	Description
FOVX, FOVY	1.48	Field of view in X and Y, respectively. For <i>Colibri</i> these values will be the same.
EXPOSURE_TIME	21 600.	The number of seconds of the observation period. This default value should be changed to reflect the actual expected observation period duration depending on the field being simulated.
MIN_DIAM	0.5	The minimum KBO diameter to probe.
MAX_DIAM	90.	The maximum KBO diameter to probe.
MIN_RAD	42.	The minimum heliocentric distance to model.
MAX_RAD	48	The maximum heliocentric distance to model.
RAD_LAW	2.6	The radial distribution power-law index.
SIZE_LAW	3.6	The differential size distribution power-law index.
C	18.8	An albedo-based brightness scaling constant for TNOs in this model. Leave as default.
R0	23.5	The R magnitude at which KBOs are expected to reach 1 per square degree in the ecliptic. This is also model-dependent and should be left as default.
NBINS	500	The number of bins in logarithmic space to sample the diameter range $D_0 \leq D \leq D_k$. At default values the difference between adjacent diameter bins will be $\sim 1\%$ of the bin value.
N_NIGHTS	20	The number of nights to simulate observation of the field. The simulation terminates on completion; there is no upper limit.

Table 12: Default parameters for the occultation simulation. These are defined and set at the top of the `runmodel.py` file.

- 1117 • **Pickle:** Pickle is used to store the output of the field search.
- 1118 • **Astropy:** The astropy.convolution package is used in the estimation of the detection likelihood
- 1119 to match events to the detection model.

1120 **A.2.4 File descriptions**

1121 **A.2.4.1 runmodel.py**

1122 This is the main program file from which the simulation will run. Once stellar data and a field are
 1123 specified (by default this is the output from the search algorithm). The bulk of the simulation is
 1124 performed by the various support modules; this file acts as a top-level control module in which they

1125 are individually called.

1126 Its functions iterate through the process of populating a field with star data, simulating a KBO
1127 field on top, and evaluating the output to decide on the next steps. It is this file which decides
1128 whether events have been located or qualify for the high resolution analysis, and which ultimately
1129 saves the event data to file.

1130 **A.2.4.2 collisions.py**

1131 This is the geometrical heart of the simulation. It contains two occultation event detection methods:
1132 a coarse quick search using the `rtree` package and a fine search using the `Shapely` package. This
1133 algorithm is motivated and explained in section 3.2.

1134 The functions contained in this module are ideally called in a particular order as specified in
1135 `runmodel.py` but can be called in any order for custom queries. Preferably, the coarse search is run
1136 first as it is the fastest. Function labels indicate whether they belong to the coarse or fine approach
1137 by including the geometry package name.

1138 The 'vectorize' functions turn the KBO sky motion into data types that are handled by the
1139 geometry packages and must be called before the event search begins. In the case of the coarse
1140 search this simply identifies a rectangular area over which the KBO will travel during the observation
1141 period. In the case of the Shapely search this will include the Shapely LineString and Point objects
1142 to reconstruct the KBO path and disk, respectively.

1143 The LineString object is also used to represent the vectors from which the impact parameter is
1144 calculated. These vectors are stored in the simulation event output and can be used to reconstruct
1145 the geometry of the event and to plot the event. This allows the user to visually check the decision-
1146 making of the search.

1147 **A.2.4.3 fieldgen.py**

1148 This module is called on by `runmodel.py` to populate the field with stars. It has two functions.
1149 One calculates the angular size of the star, and can be used to repopulate the input data should it
1150 not already contain that information as the field output does. The second function sorts the stellar
1151 data to ensure the validity (to avoid errors from bad values) and to throw out stars beyond the field
1152 limits. Each star is represented by a coordinate pair, a G magnitude, and an angular size.

1153 **A.2.4.4** `popgen.py`

1154 This module generates a KBO population across the sky. It is designed to be a parent module to
1155 `kbogen.py` in that the higher level `runmodel.py` only needs to call on this module to generate the
1156 entire population. It has functions that contain the model information from Gladman et al. (2001)
1157 to construct the number of KBOs and their individual sizes. Most of the functions in this module
1158 call on `kbogen.py` to produce KBO data; with the exception of the model information this module
1159 only acts as a control module much like `runmodel.py`.

1160 The model information must be invoked at this level because the size distribution is a statistical
1161 distribution. To ensure the random choice algorithm of NumPy strictly follows this distribution, all
1162 sizes must be drawn and assigned at once. Therefore while `kbogen.py` handles some KBO physical
1163 parameters, the diameters must be assigned here.

1164 **A.2.4.5** `kbogen.py`

1165 This module generates all the remaining KBO properties for population members: the heliocentric
1166 distance, coordinates, and projected velocity on the sky.

1167 The coordinates generated are based on the field of view provided at the start of the simulation.
1168 They are randomly generated as deviations from the centre of the field of view, such that the total
1169 population is uniformly scattered across the field. When these deviations are (later) added to a
1170 central field coordinate, they assume coordinates in that field in the given coordinate system. Thus
1171 the KBO coordinates can be generated at this level.

1172 **A.2.4.6** `colibri_sim_matching.py`

1173 This module was contributed by Tristan Mills and determines the likelihood of detection of a simu-
1174 lated occultation event. Events which are verified by the high resolution model pass their parameters
1175 a key function of this module `get_occultation_miss_rate` which returns a float between 0 and 1.
1176 A value of 1 represents a complete miss while a value of 0 represents a certain detection.

1177 **A.2.4.7** `fresnelModeler.py`

1178 This module is a support module to `colibri_sim_matching.py` which generates the sampled oc-
1179 cultation light curve used for detection analysis. It was originally made by Emily Pass as part of

1180 the detection algorithm outlined in Pass et al. (2018) but is a part of the software package provided
1181 by Tristan.

1182 References

- 1183 Andrae, R., Fouesneau, M., Creevey, O., Ordenovic, C., Mary, N., Burlacu, A., ... Bailer-Jones,
1184 C. A. L. (2018). *Gaia Data Release 2: first stellar parameters from Apsis*. Retrieved from
1185 <https://arxiv.org/pdf/1804.09374.pdf>
- 1186 Arora, S. (1998). Polynomial time approximation schemes for Euclidean traveling salesman and other
1187 geometric problems. *Journal of the ACM*, 45(5), 753–782. doi:10.1145/290179.290180
- 1188 Bailey, M. (1976). Can ‘invisible’ bodies be observed in the Solar System? *Nature*, 259, 290–291.
1189 doi:10.1038/260170a0
- 1190 Bernstein, G. M., Trilling, D. E., Allen, R. L., Brown, M. E., Holman, M. J., & Malhotra, R. (2004).
1191 *The size distribution of trans-neptunian bodies*. Retrieved from [https://arxiv.org/pdf/astro-](https://arxiv.org/pdf/astro-ph/0308467.pdf)
1192 [ph/0308467.pdf](https://arxiv.org/pdf/astro-ph/0308467.pdf)
- 1193 Bewley, A. & Upcroft, B. (2013). Advantages of exploiting projection structure for segmenting dense
1194 3D point clouds. *Australasian Conference on Robotics and Automation, ACRA*, 2–4.
- 1195 Bickerton, S. J., Kavelaars, J. J., & Welch, D. L. (2008). A search for sub-km kuiper belt objects
1196 with the method of serendipitous stellar occultations. *Astronomical Journal*, 135(3), 1039–
1197 1049. doi:10.1088/0004-6256/135/3/1039
- 1198 Bickerton, S. J., Welch, D. L., & Kavelaars, J. J. (2009). Kuiper belt object occultations: Expected
1199 rates, false positives, and survey design. In *Astronomical journal*. doi:10.1088/0004-6256/137/
1200 5/4270
- 1201 Brown, M. & Webster, R. (1997). *Occultations by Kuiper belt objects*. Retrieved from [https://](https://academic.oup.com/mnras/article-abstract/289/4/783/1062123)
1202 academic.oup.com/mnras/article-abstract/289/4/783/1062123
- 1203 Chang, H. K., King, S. K., Liang, J. S., Wu, P. S., Lin, L. C. C., & Chiu, J. L. (2006). Occultation
1204 of X-rays from Scorpius X-1 by small trans-neptunian objects. *Nature*, 442(7103), 660–663.
1205 doi:10.1038/nature04941
- 1206 Chiang, E. I. & Brown, M. E. (1999). *Keck pencil-beam survey for faint Kuiper Belt Objects*. Retrieved
1207 from <http://iopscience.iop.org/article/10.1086/301005/pdf>
- 1208 Elliot, J. L., Kern, S. D., Clancy, K. B., Gulbis, A. A. S., Millis, R. L., Buie, M. W., ... Meech,
1209 K. J. (2005). The Deep Ecliptic Survey: A Search for Kuiper Belt Objects and Centaurs. II.
1210 Dynamical Classification, the Kuiper Belt Plane, and the Core Population. *The Astronomical
1211 Journal*, 129(2), 1117–1162. doi:10.1086/427395

- 1212 Finkel, R. A. & Bentley, J. L. (1974). Quad Trees: A Data Structure for Retrieval on Composite
1213 Keys. *Acta Informatica*, 4(1), 1–9.
- 1214 Fraser, W. C. & Kavelaars, J. J. (2008). A derivation of the luminosity function of the Kuiper belt
1215 from a broken power-law size distribution. *Icarus*, 198, 452–458. doi:10.1016/j.icarus.2008.08.
1216 009
- 1217 Freidman, J. H., Bentley, J. L., & Finkel, R. A. (1977). An Algorithm for Finding Best Matches
1218 in Logarithmic Expected Time. *ACM Transactions on Mathematical Software*. doi:10.1145 /
1219 355744.355745
- 1220 Fuentes, C. I., George, M. R., & Holman, M. J. (2009). A Subary pencil-beam search for $m_R \sim 27$
1221 trans-neptunian bodies. *The Astrophysical Journal*, 696, 91–95. doi:10.1088/0004-637X/696/
1222 1/91
- 1223 Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J., Mignard, F., ... Co-
1224 authors, 5. (2016). Gaia Data Release 1. Summary of the astrometric, photometric, and survey
1225 properties. *I. Serraller Vizcaino*, 595, 68. doi:10.1051/0004-6361/201629512
- 1226 Gaia Collaboration, Prusti, T., J de Bruijne, J. H., A Brown, A. G., Vallenari, A., Babusiaux,
1227 C., ... Pigozzi, E. (2016). The Gaia mission. *E. Anglada Varela*, 595. doi:10.1051/0004-
1228 6361/201629272
- 1229 Gillies, S. & Butler, H. (2006). R-Tree spatial index for Python GIS. Retrieved from <http://toblerity.org/rtree/>
- 1230 Gillies, S., Tonnhofer, O., Arnott, J., Toews, M., Wasserman, J., & Others, A. (2007). Shapely.
1231 Retrieved from <https://github.com/Toblerity/Shapely>
- 1232 Gladman, B., Kavelaars, J. J., Petit, J.-M., Morbidelli, A., Holman, M. J., & Loredo, T. (2001). The
1233 Structure of the Kuiper Belt: Size Distribution and Radial Extent. *The Astronomical Journal*,
1234 122(2), 1051–1066. doi:10.1086/322080
- 1235 Global Modeling and Assimilation Office. (2015). MERRA-2 tavgM_3d_cld_Np: 3d,Monthly mean,Time-
1236 Averaged,Pressure-Level,Assimilation,Cloud Diagnostics V5.12.4. Greenbelt, MD, USA: God-
1237 dard Earth Sciences Data and Information Services Center (GES DISC).
- 1238 Greaves, J. S., Wyatt, M. C., Holland, W. S., & Dent, W. R. F. (2004). The debris disc around
1239 Tau Ceti: A massive analogue to the Kuiper Belt. *Monthly Notices of the Royal Astronomical
1240 Society*, 351(3), 54–58. doi:10.1111/j.1365-2966.2004.07957.x

- 1242 Guttman, A. (1984). *R-trees: a dynamic index structure for spatial searching*. University of California,
1243 Berkeley.
- 1244 Jewitt, D. & Luu, J. (1993). Discovery of the candidate Kuiper belt object 1992 QB1. *Nature*,
1245 362(6422), 730–732. doi:10.1038/362730a0
- 1246 Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open Source Scientific Tools for Python.
1247 Retrieved from <http://www.scipy.org/>
- 1248 Jones, T. A., Levine, A. M., Morgan, E. H., & Rappaport, S. (2006). Millisecond Dips in Sco X-1
1249 are Likely the Result of High-Energy Particle Events, 1–5. arXiv: 0612129v2 [astro-ph]
- 1250 Kenyon, S. J., Bromley, B. C., O'brien, D. P., & Davis, D. R. (2007). *Formation and collisional*
1251 *evolution of Kuiper Belt Objects*. Retrieved from <https://arxiv.org/pdf/0704.0259.pdf>
- 1252 Kogan, J., Nicholas, C., & Teboulle, M. (2006). *Grouping multidimensional data: Recent advances*
1253 *in clustering*. doi:10.1007/3-540-28349-8
- 1254 Kunder, A., Kordopatis, G., Steinmetz, M., Zwitter, T., Mcmillan, P. J., Casagrande, L., ... Navarro,
1255 J. F. (2016). The RAdial Velocity Experiment (RAVE): Fifth data release. *Joss Bland-Hawthorn*,
1256 20, 30. doi:10.17876/rave/dr.5/005
- 1257 Laporte, G. (2010). A concise guide to the Traveling Salesman Problem. *Journal of the Operational*
1258 *Research Society*, 61(1), 35–40. doi:10.1057/jors.2009.76
- 1259 Lehner, M. J., Wang, S.-Y., Alcock, C. A., Cook, K. H., Furesz, G., Matthew Lehner, a. J., ...
1260 Lee, W. H. (2012). The Transneptunian Automated Occultation Survey (TAOS II). 8444(17).
1261 doi:10.1117/12.925541
- 1262 Lindegren, L., Hernández, J., Bombrun, A., Klioner, S., Bastian, U., Ramos-Lerate, M., ... Vec-
1263 chiato, A. (2018). *Gaia Data Release 2 - The astrometric solution*. Retrieved from <https:////archives.esac.esa.int/gaia>.
- 1264
- 1265 Luu, J. X. & Jewitt, D. C. (2002). Kuiper Belt Objects: Relics from the Accretion Disk of the Sun.
1266 *Annual Review of Astronomy and Astrophysics*, 40(1), 63–101. doi:10.1146/annurev.astro.40.
1267 060401.093818
- 1268 Maneewongvatana, S. & Mount, D. M. (1999). Analysis of approximate nearest neighbor searching
1269 with clustered point sets. Retrieved from <http://arxiv.org/abs/cs/9901013>
- 1270 Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. (2018). *Gaia Data Release 2. Cross-match*
1271 *with external catalogues Algorithms and results*. Retrieved from <https://www.cosmos.esa.int/web/gaia/dr2-papers>
- 1272

- 1273 Mehlhorn, K. & Sanders, P. (2008). *Algorithms and data structures - the basic toolbox*. doi:10.1007/
1274 978-3-540-77978-0
- 1275 Mills, T. (2018). Private Communication.
- 1276 Osborn, J., Föhring, D., Dhillon, V. S., & Wilson, R. W. (2015). *Atmospheric Scintillation in Astro-*
1277 *nomical Photometry* (tech. rep. No. 2). Retrieved from <https://arxiv.org/pdf/1506.06921.pdf>
- 1278 Pan, M. & Sari, R. (2005). Shaping the Kuiper belt size distribution by shattering large but strength-
1279 less bodies. *Icarus*, 173(2), 342–348. doi:10.1016/j.icarus.2004.09.004
- 1280 Pass, E., Metchev, S., Brown, P., & Beauchemin, S. (2018). *Pipeline for the Detection of Serendip-**
1281 *itous Stellar Occultations by Kuiper Belt Objects with the Colibri Fast-Photometry Array* *.
1282 Retrieved from <https://arxiv.org/pdf/1711.00358.pdf>
- 1283 Roques, F., Doressoundiram, A., Dhillon, V., Marsh, T., Bickerton, S., Kavelaars, J. J., ... Tozzi,
1284 G. P. (2006). Exploration of the Kuiper Belt by High-Precision Photometric Stellar Occulta-
1285 tions: First Results. *The Astronomical Journal*, 132(2), 819–822. doi:10.1086/505623
- 1286 Roques, F. & Moncuquet, M. (2000). A Detection Method for Small Kuiper Belt Objects: The Search
1287 for Stellar Occultations. *Icarus*, 147(2), 530–544. doi:10.1006/icar.2000.6452
- 1288 Schlichting, H. E., Ofek, E. O., Sari, e., Nelan, E. P., Gal-Yam, A., Wenz, M., ... Livio, M. (2012).
1289 *Measuring the Abundance of sub-kilometer sized Kuiper Belt Objects using Stellar Occultations*.
1290 Retrieved from <https://arxiv.org/pdf/1210.8155.pdf>
- 1291 Schlichting, H. E., Ofek, E. O., Wenz, M., Sari, R., Gal-Yam, A., Livio, M., ... Zucker, S. (2009).
1292 *A single sub-km Kuiper Belt object from a stellar Occultation in archival data*. Retrieved from
1293 <https://arxiv.org/pdf/0912.2996.pdf>
- 1294 Trujillo, C. A. & Brown, M. E. (2001). *The radial distribution of the Kuiper Belt*. Retrieved from
1295 <http://iopscience.iop.org/article/10.1086/320917/pdf>
- 1296 Zhang, Z. W., Lehner, M. J., Wang, J. H., Wen, C. Y., Wang, S. Y., King, S. K., ... Schwamb, M. E.
1297 (2013). The TAOS project: Results from seven years of survey data. *Astronomical Journal*.
1298 doi:10.1088/0004-6256/146/1/14