Radar meteoroid orbit stream searches using cluster analysis

D. P. Galligan*†

Department of Physics and Astronomy, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

Received 2002 December 9; in original form 2002 June 21

ABSTRACT

The single-linkage cluster analysis technique is here applied to the data set provided by the AMOR meteoroid orbit radar to detect meteoroid streams present. This technique, which is much used particularly in photographic meteor work, is found to be inappropriate for use on radar data sets: while clusters corresponding to the major meteoroid streams are found in the AMOR data set, the algorithm is found to be very unstable and the statistical significance or otherwise of these clusters, and that of others which appear at similar levels in the agglomerative hierarchy, is unable to be validated.

Key words: meteors, meteoroids.

1 INTRODUCTION

Searching for streams within meteoroid orbit data sets is important. Techniques which assume no a priori knowledge are particularly useful as they allow us to objectively search the data set and discover new linkages with progenitor bodies (e.g. comets).

Single-linkage cluster analysis is a technique which has been used in many previous meteoroid stream searches (e.g. Southworth & Hawkins 1963; Lindblad 1971a). This method has been mainly used in the past to search photographic meteoroid orbit data sets, where the measurement uncertainty is an order of magnitude lower than for those derived from radar data. In the current study we wish to discuss the applicability of this method to searching radar detected data sets, in particular that provided by the Advanced Meteoroid Orbit Radar (AMOR), for the presence of streams.

We begin with a review of past stream search methods and then continue on to discuss the implementation and results obtained when single-linkage was applied to the AMOR data set.

2 A REVIEW OF PAST METEOROID CLUSTER ANALYSIS BASED STREAM SEARCHES

Non-instrument-based visual meteor shower observation techniques rely on the detection of an enhanced rate of meteoroids appearing to emerge from a small region of the sky called the ‘radiant’. Surveys using a large number of visual observers have yielded results of scientific value, for example Jenniskens (1994) presents a 10-yr study with an effective counting time of 4482 h of the activity (rate) curves of 50 major and minor showers. In these studies the meteor shower is defined in terms of radiant position (right ascension \(\alpha\) and declination \(\delta\)) and time of detection \((t_\odot)\).

Photographic and radar surveys offer the possibility of meteoroid speed determination and of more accurate radiant position determination. The meteor showers listed in works such as Lovell (1954) are defined on the basis of radiant position, geocentric velocity and time of detection \((\alpha, \delta, V_G, t_\odot)\). This system is sufficient to completely describe the shower: the heliocentric orbit of the meteoroid may be derived from these parameters when gravitational focusing by the Earth is included.\(^1\)

Southworth & Hawkins (1963) describe a search within 359 randomly selected meteoroid orbits photographed by Baker Schmidt meteor cameras between the period 1954 to 1957. They find that previous searches which relied on geocentric quantities for classification were insufficient. They instead use these quantities to derive heliocentric orbits which are described by the five orbital elements \((q, e, i, \omega, \Omega)\). A method of inter-comparing orbits is required: to this end Southworth and Hawkins define a dissimilarity function called a D-criterion \((D_{\text{SH}})\). \(D_{\text{SH}}\) allows the calculation of the difference between two orbits using all five elements of the orbit. The assumption inherent in this function is that differences between orbits are caused by physical orbital differences with a negligible component deriving from measurement uncertainties; Southworth & Hawkins show that \(D_{\text{SH}}\) is related to the amount of energy required to perturb one orbit into the other. This assumption is reasonable in the case of photographic orbits which have very low uncertainties \((\approx 0.1\) in angular elements).

Southworth and Hawkins define a meteoroid stream by the following two methods.

\(^1\) The geocentric system \((\alpha, \delta, V_G, t_\odot)\) is four-dimensional while the heliocentric orbit is described by five parameters: perihelion distance \((q)\), eccentricity \((e)\), inclination \((i)\), argument of perihelion \((\omega)\), and longitude of the ascending node \((\Omega)\). However \(q, e\) and \(\omega\) are linked by a necessary condition for Earth collision (see equation 4), leaving the orbit as a quasi-four-dimensional construct, which is completely defined by \((\alpha, \delta, V_G, t_\odot)\) with the addition of allowance for gravitational focusing by the Earth.

*E-mail: d.galligan@paradise.net.nz
†Present address: Defence Technology Agency, Devonport Naval Base, Private Bag 32901, Auckland, New Zealand.

© 2003 RAS
(i) Given a known mean stream orbit \( M \), any meteoroid orbit \( N \) is a member of the stream if \( D(M, N) \leq D_m \) where \( D_m \) is an appropriate cut-off value.

(ii) Two meteoroid orbits \( A \) and \( B \) are associated if \( D(A, B) \leq D_c \), where \( D_c \) is an appropriate association limit. A stream is defined as a group of meteoroid orbits where each orbit is associated with at least one other.

The direct-search approach (i) presupposes the knowledge of the correct mean orbit for the stream under consideration. It amounts to a search within a hyper-sphere of radius \( D_m \) in the \( D \)-criterion space centred on the mean. In this case the formulation of the \( D \)-criterion has a strong impact on the shape of streams that result from the search. A variant of this search approach is applied by Galligan (2003) to the same AMOR data set under study in the current paper.

The serial-association method (ii) is, in fact, an implementation of the classical single-linkage hierarchical cluster analysis algorithm – providing a snapshot of the hierarchy at a single cut-off level, \( D_c \). In contrast to the direct search approach this method allows a stream to define its own shape and does not require a priori knowledge of the stream structure present within the data set. A disadvantage is the possibility of the formation of long thin chains of orbits in the orbital element space if too large a cut-off value is used.

Southworth and Hawkins apply both techniques to their photographic data set. They find that cut-off levels set at 0.20, in both the direct- and serial-association approaches, retrieve the streams they expect; this expectation being based on previous more rudimentary searches in the data. One should note that \( D_m \) and \( D_c \) are not necessarily linked together and applications to larger data sets would require \( D_c \) to decrease. Assuming a four-dimensional point distribution Southworth and Hawkins suggest that \( D_c \) should vary inversely as the fourth root of the sample size.

Lindblad (1971a,b) continues the work of Southworth and Hawkins. He extends their data set from 360 to 865 orbits; these extra observations being obtained from various precise photographic studies. He determines an appropriate cut-off level for single-linkage analysis of this data set as

\[
D_{3H} = 0.20 \left( \frac{360}{865} \right)^{0.25} = 0.161,
\]

where the basis is on the assumption that Southworth and Hawkins’ suggested relationship between cut-off level and data set size, and their original cut-off level value, itself are appropriate.

Lindblad runs single-linkage searches at the 0.20, 0.15 and 0.10 cut-off levels on his data set. He finds that at the 0.20 level low-inclination streams tend to form very long duration pseudo-streams consisting of several well known streams: a stream is found extending from July to December including the \( \alpha \) Capricornids, the \( \chi \) Orionids, the Andromedids and the North and South Tau-rids; a similar stream is found featuring a combination of the Virginids stream with a number of sporadic meteors stretching from February to June. This effect illustrates a well known aspect of the single-linkage algorithm called ‘chaining’ where very dissimilar data set members are linked together by intermediates forming in effect a long chain with no associated physical meaning.

In order to remove these low-inclination pseudo-streams Lindblad abandons the \( D_{3H} = 0.20 \) cut-off level, even though it contains most of the known medium- and high-inclination streams, and instead uses the more conservative 0.15 and 0.10 levels. He finds that the optimum of these levels is \( D_{3H} = 0.15 \) and this is used throughout the remainder of his study.

The approach of Southworth, Hawkins and Lindblad to cluster analysis, using the \( D_{3H} \) dissimilarity function, has been repeated in many studies since that time. Such stream searches have, almost exclusively, been performed on photographically derived orbit data sets. One of the few departures from this is that of Jopek (1993a) who uses his \( D_H \) adaptation of \( D_{3H} \) to search for TV derived meteoroid streams at a cut-off level, \( D_H = 0.20 \). He defines 23 streams comprising 30 per cent of the 531 orbits. Most streams consist of ~5 orbits with a few larger streams of membership ~20. Due to the northerly latitude of the TV detectors (Ontario, Canada) the streams are almost exclusively of a northerly declination. The TV orbits used by Jopek are published in Sarma & Jones (1985) and Hawkes, Jones & Ceplecha (1984). The uncertainties on these orbits vary greatly.

Representative uncertainties are \( 2^\circ \) in angular position and 0.8 km s\(^{-1} \) in velocity (Sarma & Jones 1985); the former is similar to that achieved by the radar detection method while the latter is about half that typically expected from a radar. The number of TV meteors in recognizable showers is expected to be about an order of magnitude lower than that for photographic meteors. This behaviour is in common with that found for radar meteors (Jopek 1993a). It happens because both TV and radar measurements probe dust populations much smaller in particle size than do photographic techniques – the limiting optical magnitude for TV meteors is approximately 5 times lower than for Super Schmidt photographic meteors and the equivalent limiting (radio) magnitude for radar detected meteors is typically 5 to 10 times lower than for TV detections. Smaller dust particles are expected to have mostly lost their memory of the parent body and to have substantially merged into the sporadic background. This leads to an expectation of an increasingly smaller proportion of meteors in distinguishable streams as the detection equipment is changed from eye/camera to TV image intensifier and then to radar.

Weiss (1960), using a radar system at Adelaide, Australia, detects shower activity by changes in the meteor rate. By assuming that rate fluctuations in the sporadic background are essentially random events, he applies Poisson statistics to determine any significant fluctuations: such fluctuations are provisionally attributed to shower activity, with radar range characteristics being used to confirm whether a shower is present. Nilsson (1964) surveys southern hemisphere radar meteor orbits obtained from the Adelaide radar in 1961. He notes that this survey is the first in the southern hemisphere in which individual orbits have been calculated. The data set consists of 2200 orbits determined from meteors of limiting radio magnitude +6. It is estimated that 25 per cent of the meteors in this survey were associated with showers. The search method adopted defines association between two orbits when:

\[
|1/a_1 - 1/a_2| \leq 0.15 \text{ au}^{-1},
\]

\[
|e_1 - e_2| \leq 0.07,
\]

\[
|i_1 - i_2| \leq 7^\circ,
\]

and

\[
|\nu_1 - \nu_2| \leq 7^\circ,
\]

where \( \nu \) is the true anomaly, \( a \) is the semi-major axis length of the orbit, and \( e \) and \( i \) are the eccentricity and inclination of the orbit. Additionally the total range of the grouping must not exceed twice that

\[2\] The longitude of the ascending node, \( \Omega \), is not used by Nilsson (1964) nor by Kashcheev & Lebedinets (1967) to help define the stream. The Adelaide and Ukraine surveys respectively consisted of only 5–10 d of coverage per month. Data from these months were analysed separately which effectively constrained the range of \( \Omega \) values (because \( \Omega \) is directly related to the instantaneous longitude of the Earth at the time the meteor was detected).
of these limits. These conditions are assumed to cover the real spatial spread of stream orbits and the central part of the measurement uncertainty induced spread. Nilsson (1964) was published soon after Southworth & Hawkins (1963) and notes the similarity of $D_{350}$ to the conditions imposed by equation (2). The essential difference between these methods is the dependence on an idealized dispersive mechanism assumed by Southworth & Hawkins, as compared to Nilsson’s assumption that most of the spread in radar detected streams is due to their uncertainties: there is an order of magnitude difference in the uncertainties between the photographic and radar methods (Nilsson 1964). In Nilsson’s study streams are accepted where they contain at least three orbits. There is a problem in the application of this method to a data set, such as that provided by AMOR, in that it is not clear how one prevents a grouping from growing to include all orbits through a chaining process similar to that discussed under the serial-association method. For a small and sparse data set, such as Nilsson’s, the problem does not arise as there are natural temporal boundaries which prevent chaining from proceeding too far: in the case of the near continuous AMOR data set this is not so and the method is quite unsatisfactory. Kashcheev & Lebedinets (1967) also find the use of the $D$-criterion serial-association method to be inappropriate for radar meteor studies due to the large uncertainty induced scatter in their stream orbits. Their adopted method involves partitioning the meteors for each month into overlapping groups according to their observed speeds. The radiants for these groups are ‘roughly’ reduced to the middle of the observational period in order to remove the expected variation over time of the apparent radiant coordinates.\footnote{The radiant coordinates of the shower do not necessarily move at exactly the same rate as that of the Sun in ecliptic longitude hence such a reduction to the centre of the month must be in error in some cases although it is the best possible result which can be achieved when applying such a method.} By studying relative enhancements in this ($\alpha, \delta, V_\alpha, t_d$) data space, regions of shower activity are noted. Meteors within these regions have their originating heliocentric orbits derived from such geocentric coordinates. All orbits are then compared with the mean stream orbit, i.e. that at the centre of the region, on the basis of similarities in the orbital elements: $q$, $e$, $i$ and $\omega$. Using known uncertainties in the orbital elements, meteors from adjacent months are declared to be associated with the stream if the following conditions are all met: \begin{align*}
[q - \bar{q}] &< 2\Delta q, \\
e - \bar{e} &< 2\Delta e, \\
i - \bar{i} &< 2\Delta i, \\
\omega - \bar{\omega} &< 2\Delta \omega.
\end{align*}
\label{eq:stream}
Gartrell & Elford (1975) apply two techniques to search for streams within radar data from their multi-station continuous wave and pulsed radar system in Adelaide, Australia. This system measured radio meteors down to a limiting magnitude, $M_R = +8$. They use a single-linkage search with cut-off levels at $D_{350} = 0.10$ and 0.20 to search their 1667 meteoroid orbit data set. They also apply a search in the form used by Nilsson by sorting on each orbital element in turn with the allowable differences for association set to $\approx$2 times the assumed measurement error. Using the single-linkage search they find 40 per cent of orbits being associated with at least one other orbit and 30 per cent of orbits being associated with two or more. Slightly fewer orbits are found to be grouped using Nilsson’s approach but essentially the same groupings are retained.

Gartrell & Elford note in this study that while the data being used are of similar quality to that used by Nilsson (1964), benefit is found from using the $D$-criterion based single-linkage search. They find ‘… numerous streams of undoubted reality in which the dispersion greatly exceeds that due to measurement error alone…’.

2.1 Statistical significance

It is important to test the results of a stream search in order to discover the probability of chance association. In their seminal work on the $D$-criteria Southworth & Hawkins (1963) perform such a check by randomizing the values of $\Omega$ and $i$ for meteors not originally classified as stream members. From a serial-association test on this ‘randomized’ data set they determine that $\approx$50 per cent of the groupings formed are chance associations. Furthermore these chance associations were found in the most densely populated regions of phase-space; this is to be expected as there are very strong biases on radar data sets, many of which are also applicable to those determined photographically (e.g. Galligan & Baggaley 2001). Baggaley & Galligan (1997) introduce the combination of the single-linkage method with a randomization technique in order to determine a reasonable cut-off level. This basic Monte Carlo technique involves determining the distribution in each of the orbital elements used by the $D$-criterion and creating pseudo-random data sets of orbits bearing the same large-scale distribution, but with all traces of shower activity removed. Several hundred such data sets are passed through the normal single-linkage analysis process and the point at which clusters larger than some minimum size (in that study $N \geq 5$) begin to appear is called the quasi-random level. This level is taken to be that at which groupings from the random background begin to masquerade as physically ‘real’ groupings. Jopek & Froeschlé (1997) present a similar method which is more generalized to provide the probabilities of the intrusion of random background orbits on groupings of a particular size at a particular cut-off level. Such randomization methods have also been previously used in related fields such as those of asteroid orbit cluster analysis using single-linkage (Zappalà et al. 1990, 1994, 1995). There are two reasons why the present paper disagrees with this method. Most low-inclination streams are found under a single-linkage search at relatively low cut-off levels while one must venture to higher levels to retrieve showers such as the retrograde $\eta$ Aquarids. In general higher inclination radar determined orbits have higher uncertainties, hence determining the probability of the existence of a stream at a particular level without reference to the area of the orbital element phase-space to which one is referring is erroneous. In contrast, Jopek & Froeschlé (1997) work with photographic orbits which have much lower uncertainties than radar orbits, to which the present study is addressed. It is accepted that the proportion of photographic orbits which are well defined within streams is an order of magnitude higher than that for radar orbits while the uncertainty regimes also increase by an order of magnitude between these methods. Hence the randomization method does remove much of the stream component in the case of photographic orbits but can do little to change the radar orbit local phase-space density as there are so few stream meteors to begin with.

As discussed in Galligan & Baggaley (1998) it is likely that a randomization technique, applied to radar data sets, will simply indicate that the level at which clustering occurs is the level at which one should stop if one requires a low probability of random background orbit inclusion. This was the result found in the Galligan and Baggaley study and it is attributable to the predominant sporadic content of the radar data sets.
Figure 1. Normalized number density distribution of orbit size and shape showing the clear limits imposed by impact conditions. Note the marked inhomogeneity of the density distribution. The AMOR data set 1995–1999 has been used to provide the orbits shown in this graphic.

There are further problems in the production of pseudo-random data sets in that Baggaley & Galligan (1997), Jopek & Froeschlé (1997) and Valsecchi, Jopek & Froeschle (1999) all assume that the orbital elements are generally independent. The only exception made is for the very strong relationship between $q$, $e$ and $\omega$ necessary for Earth-intersection of the orbit. This is defined by

$$q = \frac{1 \pm e \cos \omega}{1 + e},$$

where $\pm$ refers to whether the intersection occurred at the ascending (+) or descending (−) node.

The histograms of the orbital elements in these studies are obtained and then the same number of orbits as in the original data set are recreated by randomly selecting parameter values within the large-scale distribution summarized by these histograms. In the case of $D_{\text{SH}}$ usage this implies independence for $q$, $e$, $i$ and $\Omega$, with equation (4) being used to determine a $q$ value from an $(e$, $\omega$) pair. Such a simplistic approach forms an unrealistic data set in several ways. First, as shown in Fig. 1, there is a strong bias towards certain regions of $q$–$e$ space (because of Earth collision probability), however there is no strict condition similar to equation (4) governing this which could be applied. Secondly, the solar longitude at detection directly defines the $\Omega$ value. There are frequent fluctuations in the rate of meteor detections by a radar system such as AMOR with sometimes interruptions of several days in temporal coverage, hence randomizing and reassigning $\Omega$ values to pseudo-orbits does not represent even the background orbit population expected. Finally owing to the influence of the different sporadic source regions (e.g. Jones & Brown 1993), there are distinctive differences between the types of orbits that are detected at different times of day. The orbits obtained in pseudo-random data sets should contain recognition of all of these conditions. In practice this ideal is unobtainable.

3 THE AMOR DATA SET

The AMOR meteoroid orbit radar facility (Baggaley et al. 1994; Baggaley & Bennett 1996; Baggaley et al. 2001) has been operational since 1990. Since that time it has been increasingly active, securing greater than $10^5$ meteoroid event records in each year since 1995. AMOR is very sensitive with a limiting radio magnitude of +14 corresponding to meteoroid diameter $40 \, \mu$m being achieved.

The data used for the current study were obtained between 1995 May and 1999 November. The system at that time consisted of three $\sim 8$ km separated sites – comprising a central site at which the transmitting and main receiving arrays and control and cataloguing computers were situated and two remote sites which each had receiving arrays and communicated with the central site by VHF telemetry.

Heliocentric orbits are recorded in this catalogue which have been fully corrected for all known effects related to observation on an Earth-based platform – this includes gravitational focusing. Typical orbital measurement uncertainties are $\sim 2^\circ$ in most angular elements and $\sim 5$ per cent on speed and size parameters (Baggaley et al. 1994). While these uncertainties are significantly larger than those on photographically determined orbits: the much greater number of meteoroids detected and the 24 h coverage more than compensate for this deficiency. The $\sim 5 \times 10^5$ orbit sample used here provides a very high degree of statistical reliability to the results obtained.

4 SINGLE-LINKAGE CLUSTER ANALYSIS

4.1 Theory

The single-linkage algorithm is a member of the hierarchical agglomerative class of cluster analysis algorithms. Such algorithms begin with $N$ items (meteoroid orbits in the present study) contained in $N$ groupings. They progressively agglomerate these groupings until, ultimately, only one remains containing all $N$ items. The rule governing this agglomeration is the defining characteristic of the particular algorithm. Single-linkage performs its agglomeration by joining the two entities which are closest to each other at each step, these entities may be single or groups of items. Where these entities are groupings of more than one item, the inter-group dissimilarity is based on the smallest dissimilarity between individual members of that entity and surrounding entities. Fig. 2 shows a 2D model example of the single-linkage grouping structure at a particular position in the hierarchy.

The user must make two important decisions when using any of these agglomerative algorithms. First, a suitable measure of dissimilarity between items must be defined, and secondly, a method of determining a cut-off level in the hierarchy at which meaningful groupings may be found must be specified.

For the cluster analysis of orbits the standard measures of dissimilarity are the so-called $D$-criteria, a summary of which is provided in Galligan (2001). The $D$-criteria are generally defined in terms of the orbital elements $(q$, $e$, $i$, $\omega$, $\Omega)$ with various weighting schemes for element comparisons being introduced by Southworth & Hawkins (1963) and Drummond (1981): these are hereafter referred to as $D_{\text{SH}}$ and $D_{\text{DP}}$ respectively. Jopek (1993b) introduces a $D$-criterion which combines parts of $D_{\text{SH}}$ and $D_{\text{DP}}$: this function will not be explored further here due to its similarity to these functions. Valsecchi et al. (1999) introduce another scheme based upon geocentric parameters of the orbits under comparison: this function is referred to as $D_{\text{NS}}$. Using these functions, at a particular level corresponding to the cut-off $D_{\text{C}}$, membership in a particular grouping is defined such that

$$\min[D(A, B)] < D_{\text{C}},$$

where $A$ and $B$ are any individual orbits.

Certain characteristics are clear in such a process. As the cut-off level is increased the clusters become increasingly inhomogeneous. The highest quality clusters are those found in their final forms at low
levels of the hierarchy. A cluster may be defined as highly compact and distinct against the background if it is found at such a level and it survives with little agglomeration until a much higher cut-off level. A lower quality cluster is one which is only found at higher levels of the hierarchy and quickly merges into the background as one proceeds to still higher levels, such clusters may be labelled as disperse and indistinct.

4.2 Implementation

Theoretically, for \(N\) orbits, the single-linkage algorithm (implemented here by high-level computer language) must make \((N - 1)\) complete passes through the inter-orbit dissimilarity matrix stored in memory. This is obviously time and memory consuming. The memory usage is proportional to \(N(N - 1)/2\) and the time usage is proportional to \(N^2(N - 1)/2\), i.e. the process is computational order \(O(N^2)\) in memory and \(O(N^3)\) in time. It is possible to extract the cluster analysis for a number of levels within the clustering hierarchy without performing a complete formal analysis: a useful feature of this algorithm is the independence of cut-off levels within the hierarchy. Therefore when searches such as those of Lindblad (1971a) were performed, only the clusters present at three cut-off levels at \(D_{\text{cut}} = 0.10, 0.15\) and 0.20 were calculated for study.

The programs developed for the current study perform the discrete level cluster analysis in two steps. The data is first processed by a C program which accepts input of a file of orbits, indication of the particular dissimilarity function to use on these orbits and a maximum \(D\) value at which pairings should be accepted for further analysis. The program determines all orbital pairings satisfying this criteria and stores them in a specially compressed output file. This file effectively contains the dissimilarity matrix with only entries listed for pairings with suitably small dissimilarities. Another C program is then applied to the file to determine the clustering hierarchy at each of a specified set of regularly spaced cut-off levels. This hierarchy is output to a second file which contains a single large matrix representing it, where each column represents a cut-off level, each row a particular orbit and the particular numbers stored in each matrix location refer to the group number to which the orbit belongs [the group number is always the \((0 \ldots N - 1)\) row number of the lowest number orbit in the grouping]. A description file contains a list of the cut-off levels to which each column of the clustering matrix refers. Data handling programs are readily able to read these file pairs in order to further analyse the grouping structures contained therein.

4.3 Data analysis

In the current study the dissimilarity functions \(D_{\text{SH}}\) and \(D_{\text{SH}}\), in particular, are tested in the application of single-linkage analysis to various data sets provided by AMOR. Tests on the various dissimilarity functions have already been performed in Galligan (2001) for direct-search techniques. This section aims to study their performance in the context of single-linkage cluster analysis using some practical examples.

The use of randomization tests, in an attempt to provide meaningful cut-off levels in the resulting agglomeration hierarchies, is found to be unsuccessful in practice. As discussed in Section 2.1 these tests consisted of creating a large number of pseudo-random valid Earth-intersecting orbit sets with large-scale distributions similar to the directly observed data set. The assumption is that each of these data sets presents one possible instance of the background orbit population, with all meteor shower traces removed.

By running \(\sim 10^2\) such randomized data sets through the same cluster analysis process used for each corresponding original data set, a picture may be obtained of the level at which clusters begin to appear out of the sporadic (random) background.\(^4\) Galligan & Baggaley (1998) discusses the results of the application of this method to data sets of \(\sim 6 \times 10^4\) orbits. From data obtained from the vernal equinox in 1995 to the same time in 1998, eight data sets are formed each based on 180° of inter-equinoctial years. Randomization studies on these data (as described above) yield the result that the 95/99 per cent probability levels of background intrusion correspond well with the first appearance of structure in the original data set. Well-defined showers, such as the \(\eta\) Aquarids (ETA), are therefore deemed to be structures appearing out of the random background with no statistical reality. The Southern \(\delta\) Aquarids (SDA), the largest shower detected in the AMOR data set and also having much lower individual orbit measurement uncertainties than the ETA, only begins its formation at the random-simulation determined cut-off level.

There are therefore two possibilities: either the randomization method is too harsh in determining a significant cut-off level or the major showers known to be present within the AMOR data set are so weak that, while they may be visually detectable from the dendrograms, they are marginal when it comes to statistically sound detection. If the latter is true then the Monte Carlo simulation work has performed its job well and nothing more can be gained by continuing. Galligan & Baggaley (1998) assume the former case, however,

\(^4\) It is desirable in a Monte Carlo simulation process to produce many more than \(\sim 10^2\) randomized sample sets, especially given the four-dimensional nature of the orbital data sets under study. Unfortunately each sample run takes 1–2 h for the \(\sim 6 \times 10^4\) orbits studied. This time requirement increases steeply as the data set size is increased. Studies such as Jopek, Valsecchi & Froeschle (1999) have no such problems as they are only dealing with a set of \(\sim 10^3\) orbits and therefore can perform \(10^3\) randomized runs or more with no difficulty.
and continue to develop new techniques based upon the simple single-linkage algorithm. The first method uses the percentage of orbits within any groupings at each cut-off level in order to line up the agglomerative hierarchies from several separate years of data. Relative cut-off levels, corresponding (for example) to the levels at which 5 per cent and 10 per cent of the data set are clustering into streams, are then used to compare grouping structure in the different searches. The assumption is that the relative cut-off level in the hierarchy is the same for similar data sets apart from a multiplier due to differences in their relative sizes; this is in line with the assumption of Southworth & Hawkins (1963), as discussed in Section 2. This method is found to be inadequate due to the lack of statistical basis to the cut-off level and also due to the totally different numbers and types of orbits in streams found in different years at the same relative level. In fact a general finding emerging from work using the single-linkage algorithm on radar orbits is that the agglomerative hierarchy is often highly unstable: very minor changes in cut-off level may lead to disproportionately large changes in the stream structure obtained. Simulations have been performed where an original data set has its orbits changed slightly within the bounds of the measurement uncertainty. Retrieval of groupings, similar to those originally obtained, is by no means guaranteed and often the stronger showers are then found to have dwindled to the point of non-detectability.

Another single-linkage based method, developed by the author, involves taking several years of data and separately forming their agglomerative hierarchies. For each year’s hierarchy, the computer program proceeds downwards from the highest agglomeration level. All groupings, of a suitable size, at each level, having a suitably small intra-grouping mean dissimilarity and range in mean solar longitude are accepted. A further run is then made on this set of mean orbits to form clusters of similar mean values. All of the major streams were found using this method which, while more robust than a simple single-linkage algorithm, is also now rejected as the new showers found by it are not verifiable by other means. The algorithm is thought to most likely be producing large clusters made up of smaller individually insignificant clusters which in some cases occur at very different cut-off levels.

4.4 Sample results

Some results from single-linkage runs on three data sets are now presented in order to give examples of the types of output one obtains from this method. These data sets cover the activity times of the four major showers previously identified in the AMOR data set. The first and second sets are composed of prograde orbits recorded over solar longitudes $\lambda = [110^\circ, 160^\circ]$ and $\lambda = [160^\circ, 210^\circ]$ in 1995 May to 1999 November respectively while the last one is composed of retrograde orbits recorded over $\lambda = [20^\circ, 70^\circ]$ in the same years; these data sets will be referred to as Set I, II and III respectively here on. These time-frames/orbital orientation selections correspond to regions where the Southern $\delta$ Aquarids (SDA) and $\alpha$ Capricornids (CAP), Daytime Sextants (DSX); and finally $\eta$ Aquarids (ETA) meteor showers are known to be active. Results, based on the analysis of these files, are presented for $D_{\text{SH}}$ and in some cases for $D_{\text{N}}$. Both functions produced similar results apart from the case of the ETA where higher uncertainties produce a scenario worthy of further discussion.

Fig. 3 shows dendrograms corresponding to parts of the single-linkage hierarchy obtained for data set I. Cut-Off levels have been determined separately for the SDA and CAP searches in this data set. The method used to determine these levels is very simple: it is based on the assumption that meteor showers derive homogeneous

\[ \begin{align*}
\text{Cutoff Level} & = 1.80 \\
\text{Number of Orbits Scale} & = 0 – 2000
\end{align*} \]

Typically, the mean solar longitude range has a maximum set at $35^\circ$ in order to prevent chaining over large time intervals of orbits which are obviously unrelated.

In the past single-linkage searches on meteoroid orbits have generally been applied to data sets containing the complete range of orbital inclinations, i.e. both prograde and retrograde orbits. The examples presented here are for data sets which contain prograde or retrograde orbits exclusively. It is thought to be very improbable that streams will appear which cross over this boundary. Generally in the current study complete data sets were used but, for the sake of execution speed in the current examples, this orbital partitioning has been applied.

7 Dendrograms are convenient devices for the graphical display of the hierarchical clustering process. They require input of a minimum size for grouping display and a cut-off level at which significant groupings are present. Parent groupings of those present at the cut-off level are shown at higher levels of agglomeration (and therefore $D_c$ value) while, to continue the analogy, child groupings are found at lower levels of agglomeration. Groupings cannot be shown on the dendrogram which are not related to those at the chosen statistically significant cut-off level.
groupings of similar orbits and that the mean intra-group dissimilarity is a good measure of this homogeneity. The agglomeration of the shower groupings are monitored until their mean $D_{3SH}$ exceeds 0.20, the level below this point is declared the cut-off level. The maximum $D_{3SH}$ reached under this regime is 0.199 for the SDA and 0.153 for the CAP. The reason for the lower CAP $D_{3SH}$ is clear from the corresponding dendrogram where the shower stalactite\(^8\) merges into a $\sim 1.3 \times 10^4$ orbit grouping at the level above its cut-off.

The two dendrogram figures of Fig. 3 demonstrate many of the problems which hinder use of the single-linkage method on the AMOR data set. The SDA at its $D_{3SH} = 0.0244$ acceptance level merges gradually into the background with no common-sense cut-off point being visibly present. The orbit set obtained for the SDA at this level, as shown in Table 1, has a spread (standard deviation) similar to that expected due to the individual stream orbit measurement uncertainties. This is in agreement with the findings of Galligan & Baggaley (2002) and leads to the conclusion that any grouping partitions appearing at lower levels of this shower, as shown on the dendrogram, have no physical significance, and that perhaps even a higher cut-off level might be desirable. Also shown in Table 1 is an $84$ orbit group which appears at $D_{3SH} = 0.0272$ and then merges much higher in the hierarchy into a general SDA/background grouping. Such groupings show a particular problem with the use of $D_{3SH}$ in orbital analysis. The mean inclination on this group is double that of the SDA and therefore according to $D_{3SH}$ very different. Members of this group are believed to be outliers on the inclination distribution of the SDA occurring due to the large inclination angle uncertainties in this shower’s member orbits. The mean radiant, speed and detection solar longitudes of the main SDA grouping are very similar to this grouping, further backing up the opinion that these are ‘real’ SDA. However, the height in the hierarchy at which this merges with the main SDA grouping is such that if only the single-linkage hierarchy were considered then the association between these groupings could not be sensibly entertained.

The CAP shower stalactite is quite different from the SDA. Its cut-off level, $D_{3SH} = 0.0384$, is very much higher than that of the SDA, its data set companion. Indeed at the level of the CAP cut-off, the SDA has already been assimilated into a grouping of $1.2 \times 10^4$ orbits, clearly a grouping which is much too large and inhomogeneous to be called a shower. The reason for this difference in cut-off levels is due to the relative strength of the showers: the SDA has approximately $8$ times as many orbits as the CAP. The CAP is the weakest shower detected in the wavelet probe searches of Galligan & Baggaley (2002); it also appears to set a limit on the size of shower which might be detected using the single-linkage method. It is clear that a weaker shower would simply not be able to be discerned when one notices the high cut-off level at which it is found and the small number of levels over which this shower’s stalactite remains separated from the background. It is generally expected that the probability of obtaining significant groupings at lower levels in a hierarchy is greater when one is searching for a smaller grouping rather than a larger grouping. The SDA is therefore more likely to be ‘real’ astronomically than the CAP, as any cut-off level generated by a randomized simulation which did accept the SDA in data set I would most likely ignore the CAP.

The DSX is found in the $D_{3SH}$ based search of data set II. Again a mean $D_{3SH}$ maximum of $0.20$ is set with the closest achieved being $D_{3SH} = 0.19$. The corresponding dendrogram is shown in Fig. 4 and

\(^8\) As the groupings appear at different hierarchical levels on dendrograms in a manner resembling a stalactite, they are often named as such.
The grouping in question is composed of low-inclination (\(i < 10^\circ\)) associated with stream searches in a large meteoroid orbit data set. An unrelated stalactite appears on this figure which appears to be larger at each level than the DSX: if one considers the DSX to be astronomically ‘real’ then surely one should consider this to be ‘real’ too. This provides an example of one of the major problems associated with stream searches in a large meteoroid orbit data set. The grouping in question is composed of low-inclination (\(i < 10^\circ\)) orbiting meteoroids. The detection of these ‘stream’ meteoroids is spread over 50° of solar longitude in this 2,130 orbit grouping at the DSX cut-off level. One can pose the thesis that as this grouping is similar in size to the SDA and it appears at the same cut-off level as the DSX, therefore it must be ‘real,’ assuming these streams are ‘real.’ This grouping is an example of ‘chaining’ as discussed in Section 2. It contains orbits of meteors detected over the complete time-frame of the data set, a very long time for an astronomically ‘real’ shower to last. The longitude of the ascending node, for all orbits, has negligible uncertainty and indeed for low-inclination orbits combines with the argument of perihelion to produce a single parameter: the longitude of perihelion. It is therefore particularly easy to form a chain of orbits based on the daily changes in \(\lambda\) (and therefore \(\Omega\)) and this is what is believed to have happened here. Such chains act to confuse randomization techniques (like that considered earlier). The study of the original data sets leads to the combination of highly homogeneous groupings (showers) producing, in some cases, grossly inhomogeneous groupings for very small increases in cut-off level. Only very strong showers, outside the low-inclination regime, can resist such effects. The example discussed above points out the futility of single-linkage algorithm use on the large bias and orbital uncertainty laden radar meteoroid data sets: it is only possible to see the largest showers using this technique and it is not possible to prove whether these showers are statistically significant or not.

The above examples have been for prograde meteoroid streams. An example is now given of a retrograde stream: the ETA. This shower is clearly detected using both \(D_{SH}\) and \(D_N\). The resulting hierarchies are shown in Fig. 5 with the statistics of the groupings obtained listed in Table 1. Due to the large measurement uncertainty driven spread on this shower higher mean dissimilarities are used in order to include a reasonable proportion of the grouping: \(D_{SH} = 0.30\) and \(D_N = 0.20\) are used. The stalactite for the \(D_{SH}\) case appears to have the clearest delineation with only very minor showers present at the cut-off level apart from it. The \(D_N\) dendrogram is interesting for two reasons. First the \(D_N\) required to retrieve a reasonable proportion of the grouping is closer to that which is typically used to retrieve prograde showers; this differs from the \(D_{SH}\) situation where a much higher value than usual is required due to the increased spread from uncertainties. \(D_N\) effectively measures the compactness of radiant positions in addition to geocentric speed. Galligan & Baggaley (2002) show that the radiant distribution of the ETA is much more compact than that for such showers as the SDA. Geocentric speed is the major uncertainty laden parameter of the ETA due to its large size. The derived orbital elements almost all suffer large uncertainties due to this speed uncertainty. Relatively large \(D_{SH}\) dissimilarities are found when inter-comparing members of the shower owing to these large orbital element spreads. The inter-orbit dissimilarities are at a lower relative level when \(D_N\) is used as the spread caused by the shower’s speed uncertainty is partially offset by the greater uncertainty in the radiant positions of the shower.

![Figure 4](https://academic.oup.com/mnras/article-abstract/340/3/899/1747157/DownloadedFromhttps://academic.oup.com/mnras/article-abstract/340/3/899/1747157)

Figure 4. Dendrograms for the DSX shower obtained using \(D_{SH}\) in single-linkage cluster analysis on data set II. The cut-off level on this shower is set at \(D_{SH} = 0.042\). Only groupings of 40 orbits, or more, are shown.

![Figure 5](https://academic.oup.com/mnras/article-abstract/340/3/899/1747157/DownloadedFromhttps://academic.oup.com/mnras/article-abstract/340/3/899/1747157)

Figure 5. Dendrograms for the ETA shower obtained using \(D_{SH}\) (upper) and \(D_N\) (lower) for single-linkage cluster analysis on data set III. The cut-off levels on these figures are set at \(D_{SH} = 0.0448\) and \(D_N = 0.0368\) respectively. Only groupings of 40 orbits, or more, are shown.

other showers: in fact the number of stream orbits included in the grouping using $D_N$ at its cut-off level (1119) is much greater than that obtained at the very high cut-off required for $D_{38}$ where 885 orbits are retrieved. The second reason for interest in the $D_N$ dendrogram is the stalactite appearing to the right of the ETA stalactite representing a grouping of 259 meteors. By analogy with the SDA program is the stalactite appearing to the right of the ETA stalactite orbits because most of the major showers have members which, due to measurement uncertainty, often fall within the low-inclination zone. Apart from this, the removal of the most populous portion of the orbital population in such an arbitrary way is difficult to justify if one assumes no a priori knowledge of the situation. Searching for showers smaller than the major showers will be unsuccessful under the single-linkage method.

5 CONCLUSIONS

None of the single-linkage related search variants are found to be suitable for radar meteoroid stream determination (where angular uncertainties of $\sim 2^\circ$ are achievable by that technique). The early result which showed by randomized simulation that in a set of half-year data sets no significant structure is present, appears in general to be vindicated. Exceptions to this are in the case of the major showers, which do appear visually in single-linkage searches and are supported by other methods and surveys. Due to the high ‘noise’ level in the radar data sets it is assumed that these major showers are simply not strong enough to compete in a statistically sound test of the single-linkage algorithm. A substantial proportion of this noise is caused by the concentration of highly randomized orbits at low inclinations. Unfortunately one cannot simply remove such orbits because most of the major showers have members which, due to measurement uncertainty, often fall within the low-inclination zone. Apart from this, the removal of the most populous portion of the orbital population in such an arbitrary way is difficult to justify if one assumes no a priori knowledge of the situation. Searching for showers smaller than the major showers will be unsuccessful under the single-linkage method.

ACKNOWLEDGMENTS

The author wishes to thank the University of Canterbury for the provision of support via a Doctoral Scholarship during the time this work was undertaken. Professor Jack Baggaley is acknowledged for his support and guidance. Helpful discussions with Niamh Gallage are acknowledged. The anonymous referee is thanked for the insightful comments and for meticulously checking the detail of the manuscript.

REFERENCES


Baggaley W., Galligan D., 1997, Planet. Space Sci., 45, 865

Baggaley W., Bennett R., Steel D., Taylor A., 1994, QJRAS, 35, 293


Drummond J., 1981, Icarus, 45, 545


Jopek T., 1993b, Icarus, 106, 603


Lindblad B., 1971a, Smithsonian Contrib. Astrophys., 12, 1

Lindblad B., 1971b, Smithsonian Contrib. Astrophys., 12, 14


This paper has been typeset from a TeX/LaTeX file prepared by the author.