Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **492,** 5811–5843 (2020) doi:10.1093/mnras/stz3610

Clusterix 2.0: a virtual observatory tool to estimate cluster membership probability

L. Balaguer-Núñez [1]★ M. López del Fresno,[2] E. Solano,[2] D. Galadí-Enríquez,[3] C. Jordi,[1] F. Jimenez-Esteban,[2] E. Masana,[1] J. Carbajo-Hijarrubia[1] and E. Paunzen [4]

[1]*Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, E-08028 Barcelona, Spain*
[2]*Departmento de Astrofísica, Centro de Astrobiología (CSIC-INTA), Spanish Virtual Observatory, ESAC Campus, Camino Bajo del Castillo s/n, E-28692 Villanueva de la Cañada, Madrid, Spain*
[3]*Observatorio de Calar Alto, Sierra de los Filabres, E-04550-Gérgal (Almería), Spain*
[4]*Departament of Theoretical Physics and Astrophysics, Marsaryk University, Brno, Czech Republic*

Accepted 2019 December 17. Received 2019 November 25; in original form 2019 October 10

ABSTRACT

Clusterix 2.0 is a web-based, Virtual Observatory compliant, interactive tool for the determination of membership probabilities in stellar clusters based on proper-motion data using a fully non-parametric method. In an area occupied by a cluster, the frequency function is made up of two contributions: cluster and field stars. The tool performs an empirical determination of the frequency functions from the vector point diagram without relying on any previous assumption about their profiles. Clusterix 2.0 allows us to search the appropriate spatial areas in an interactive way until an optimal separation of the two populations is obtained. Several parameters can be adjusted to make the calculation computationally feasible without interfering with the quality of the results. The system offers the possibility to query different catalogues, such as *Gaia*, or upload a user's own data. The results of the membership determination can be sent via Simple Application Messaging Protocol (SAMP) to Virtual Observatory (VO) tools such as Tool for OPerations on Catalogues And Tables (TOPCAT). We apply Clusterix 2.0 to several open clusters with different properties and environments to show the capabilities of the tool: an area of five degrees radius around NGC 2682 (M67), an old, well-known cluster; a young cluster NGC 2516 with a striking elongated structure extended up to four degrees; NGC 1750 and NGC 1758, a pair of partly overlapping clusters; the area of NGC 1817, where we confirm a little-known cluster, Juchert 23; and an area with many clusters, where we disentangle two overlapping clusters situated where only one was previously known: Ruprecht 26 and the new CLUSTERIX 1.

Key words: methods: statistical – virtual observatory tools – proper motions – open clusters and associations: general – open clusters and associations: individual: NGC 2682, CLUSTERIX 1.

1 INTRODUCTION

Open clusters (OCs) are coeval groups of stars formed from the same molecular cloud and thus having the same age and initial chemical composition. This makes them ideal targets to study the formation and evolution of stellar objects. Open clusters are also valuable tools for undertaking dynamic and kinematic studies of our Galaxy. For instance, clusters have been used to determine the spiral structure of the Galaxy and to investigate the dynamics and the chemical evolution of the Galactic disc (Frinchaboy & Majewski 2008). They are good tracers to follow the metallicity gradient of the Milky Way (Carrera & Pancino 2011; Netopil et al. 2016; Casamiquela et al. 2017) and its evolution through time (Magrini et al. 2009; Jacobson et al. 2016), providing insight on the formation of the Galactic disc. Studies of the kinematics of OCs and reconstruction of their individual orbits (Wu et al. 2009; Cantat-Gaudin et al. 2016) help us to understand the internal processes of heating (Quillen et al. 2018) and radial migration (Anders et al. 2017) and how they affect the chemodynamical evolution of the disc. Some very perturbed orbits might also provide evidence for recent merger events and traces of past accretion from outside the Galaxy (Cantat-Gaudin et al. 2016).

Open clusters are not only useful tracers of Milky Way structure but also interesting targets in their own right. Although most stars in the Milky Way are observed in isolation, it is believed that most (possibly all) stars form in clustered environments and spend at

★ E-mail: lbalaguer@fqa.ub.edu

© 2020 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

least a short amount of time gravitationally bound with their siblings (Clarke, Bonnell & Hillenbrand 2000; Portegies Zwart, McMillan & Gieles 2010), embedded in their progenitor molecular cloud. The majority of such systems will be disrupted in their first few million years of existence, due to mechanisms possibly involving gas loss driven by stellar feedback (Brinkmann et al. 2017) or encounters with giant molecular clouds (Gieles et al. 2006). Nonetheless, a fraction will survive the embedded phase and remain bound over longer time-scales.

Gaia Data Release 2 (DR2), published in 2018 April, provides a five-parameter astrometric solution (celestial position, proper motions in right ascension and declination, and parallax) and magnitudes in three photometric filters (G, G_{BP} and G_{RP}) for more than 1.3 billion sources with unprecedented precision and accuracy (Gaia Collaboration et al. 2018a). *Gaia* is a fundamental resource with which to study the known OCs and discover new ones, as well as to discard clusters previously identified (Cantat-Gaudin et al. 2018b, 2019; Castro-Ginard et al. 2018). With this aim, an accurate method to calculate cluster membership is essential in order to conduct further studies (Frinchaboy & Majewski 2008).

The determination of the mean properties of open clusters (like radial velocity or metallicity) requires prior knowledge of their member stars to optimize the costly process of obtaining and reducing high-resolution spectroscopic data on a large scale. Some information is already available, mainly for radial velocity (Soubiran et al. 2018; Carrera et al. 2019a). Moreover, knowing the membership probability of the stars in the cluster area helps to select member stars on photometric diagrams to determine the age of the cluster by isochrone fitting (e.g. Bossini et al. 2019). Hence, a precise identification of the stars that compose a cluster is critical or accurate determintion of the kinematic and fundamental parameters of the clusters (age, total mass, etc.), which are essential for studies on Galactic dynamics.

In order to exploit the wealth and quality of *Gaia* data efficiently, a variety of new approaches and new tools must be considered. Several automatic approaches have been developed to study the reality of known OCs (Cantat-Gaudin et al. 2018a) and to discover new ones (Castro-Ginard et al. 2018; Cantat-Gaudin et al. 2018b). However, to our knowledge, the only available code specific for performing automatic membership assignment in stellar clusters in the literature is the R version of UPMASK[1] by Krone-Martins & Moitinho (2015). In our case, we aim to provide a web tool to facilitate membership studies from proper-motion data to any user that requires a tailor-made study on any specific data set.

`Clusterix` 1.0 (Sezima et al. 2015) was a web tool based on the implementation of the non-parametric method for membership segregation (Galadi-Enriquez, Jordi & Trullols 1998c; Balaguer-Núñez, Galadí-Enríquez & Jordi 2007) developed at the Universitat de Barcelona (Spain) in collaboration with the Masaryk University (Czech Republic), designed as a complement to the WEBDA[2] database (Netopil, Paunzen & Stütz 2012) of observational data on stars in open clusters. In this article we present `Clusterix` 2.0,[3] an upgraded, much more powerful, VO-compliant version of the web tool jointly developed by the Universitat de Barcelona and the Spanish Virtual Observatory[4] (Balaguer-Núñez et al. 2017). Compliance with the Virtual Observatory is key for optimum gathering of additional information (such as parallaxes or radial velocities) from VO services and catalogues that can be used to refine the membership determination. Moreover, stellar physical parameters such as effective temperatures, radii, luminosities or metallicities can be estimated from VO services like VO Sed Analyzer (VOSA)[5] (Bayo et al. 2008). VO tools are deeply ingrained in `Clusterix` 2.0: in addition to using the Simple Application Messaging Protocol (SAMP) to send the results to any VO tool (e.g. TOPCAT[6]), the discovery and gathering of information in Step 1 is performed smoothly using VO protocols, in particular Table Access Protocol (TAP) and the Sesame Name Resolver.[7] Moreover, the photometric gathering in Step 3 used to build the VOSA file makes use of the Centre de Données astronomiques de Strasbourg (CDS) Xmatch service.[8]

The article is organized as follows: in Section 2 we present the formalism, Section 3 describes the implementation, while Section 4 presents different scientific cases to show the capabilities of the tool; finally, Section 5 contains our conclusions.

2 THE FORMALISM: MEMBERSHIP PROBABILITIES FROM PROPER MOTIONS

Kinematic segregation of stellar cluster members relies on the fact that the cluster displays a common spatial motion that distinguishes it from the field population. Kinematic segregation may relay on radial velocities, proper motions, or both.

The classical approach to cluster/field segregation from proper-motion data has traditionally been the parametric method. This method assumes the existence of two overlapping populations in the 2-D proper motion vector point diagram (VPD): *cluster* and *field*. The corresponding frequency functions are modelled as parametric Gaussian functions: a circular Gaussian model is adopted for the cluster distribution, while a bivariate (elliptical) Gaussian describes the field. Membership probabilities are later derived from the fits (see Sanders (1971) for a full description of the methodology). However, the parametric approach shows several drawbacks, mainly related to the fact that this Gaussian modelling of the frequency functions is not always realistic, and these disadvantages are not overcome fully by the several improvements introduced into this approach by different authors during the last decades (see the description of a quite sophisticated implementation of parametric methods, for instance, in Zhao & He 1990). More recently, this approach has been greatly improved to treat multidimensional data (photometric, astrometric) (Sarro et al. 2014) and even treat missing values (Olivares et al. 2018).

The cluster/field segregation from proper motions can also be approached by means of non-parametric methods, aimed at modelling the *cluster* and *field* proper-motion distributions that build up the particular VPD of the stars in the sky area under study empirically. Clustering algorithms have been applied to the assignment of membership probabilities in OCs from multidimensional data. UPMASK (Krone-Martins & Moitinho 2014) was developed to use only photometry and positions and was afterwards adapted to *Gaia* data (Cantat-Gaudin et al. 2018a, b) based only on proper motions and parallaxes. A similar approach using DBSCAN and machine-learning algorithms (Castro-Ginard et al. 2018, 2019) has

[1] https://cran.r-project.org/package = UPMASK
[2] http://webda.physics.muni.cz
[3] http://CLUSTERIX.cab.inta-csic.es
[4] http://svo.cab.inta-csic.es
[5] http://svo2.cab.inta-csic.es/theory/vosa/
[6] http://www.star.bris.ac.uk/mbt/topcat/
[7] http://cds.u-strasbg.fr/cgi-bin/Sesame
[8] http://cdsxmatch.u-strasbg.fr/xmatch/

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

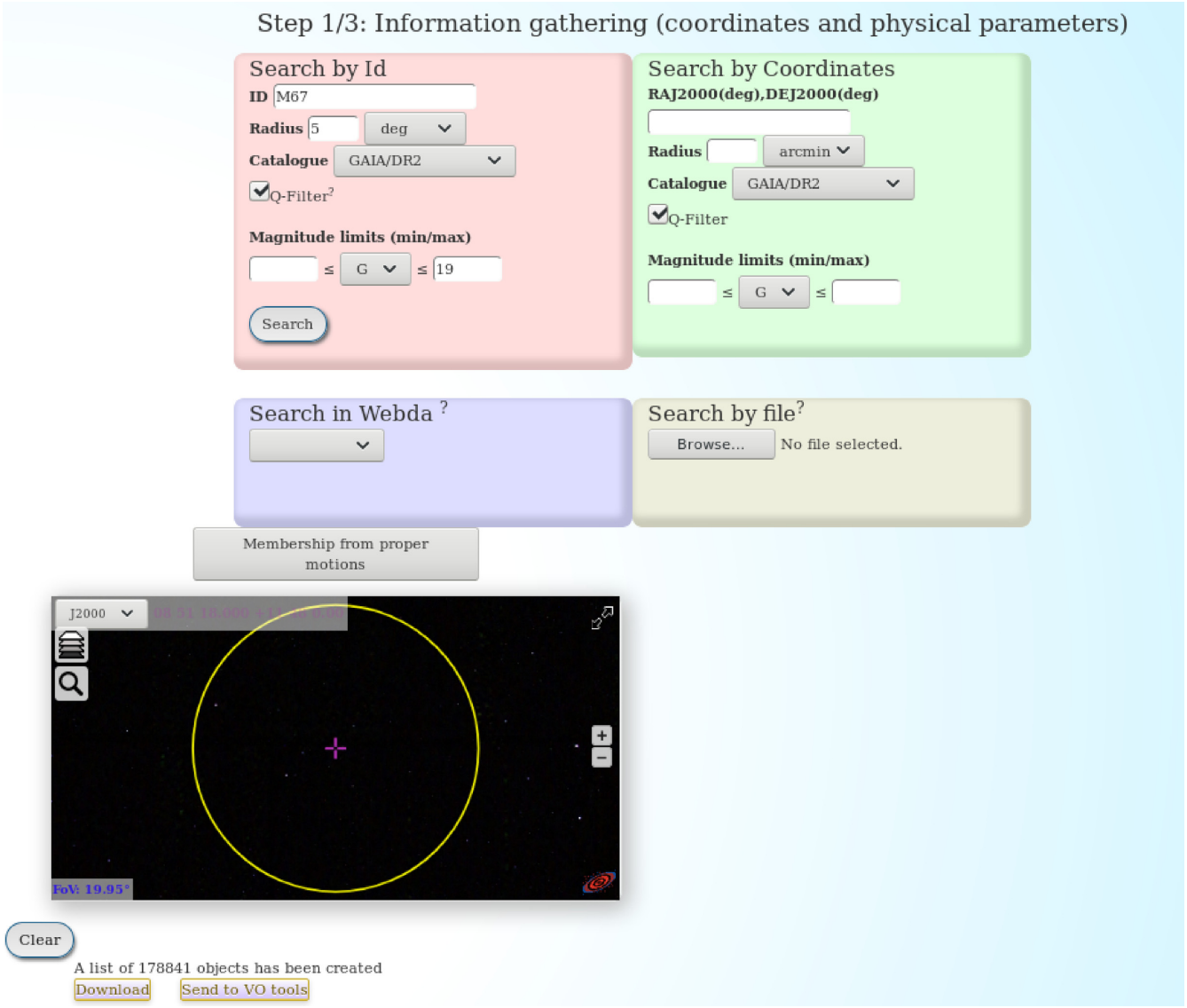


Figure 1. Step 1 of Clusterix 2.0. for the case of M67. The catalogue, the radius and the magnitude limit are chosen in the upper panels. Results can be downloaded or sent to Virtual Observatory tools. When the number of stars is bigger than 40 000, only a yellow circle is drawn to enclose the searched region. See text for more details. See Section 4.1 for details on the parameter selection for this cluster.

been conducted to discover new clusters from an automatic search, demonstrating that many nearby clusters still remain to be found (Cantat-Gaudin et al. 2019). A complex scenario with a huge amount of new moving groups (Faherty et al. 2018) and star-forming regions (Prisinzano, Magrini & et al. 2018) is starting to appear.

There are also other authors that have mixed both approaches in different ways (see e.g. Sampedro et al. 2017). There are specific differences between all these methods, depending on the type of data to be used or on some particular clusters.

2.1 The non-parametric approach as implemented in Clusterix 2.0

The non-parametric method implemented in Clusterix 2.0 follows a philosophy explained at length in Balaguer-Núñez et al. (2004a, 2007), Balaguer-Núñez, Jordi & Galadí-Enríquez (2005) and Galadí-Enríquez, Jordi & Trullols (1998a). However, several details are specific to Clusterix 2.0 and, for the sake of clarity, we will formalize the algorithm in the following paragraphs.

Some improvements introduced in Clusterix 2.0 are related to the normalization and scaling of frequency functions. Although these changes do not affect the final results, they lead to a formally clearer algorithm and to a simpler code as well. Observables and scaling factors derived from them are clearly identified and specified and the three kinds of resulting frequency function (raw, normalized and scaled) are named explicitly and distinguished formally. This way, the simplicity of the underlying concepts is much more evident. Also, the formal reshaping of the concepts and formulae may be useful for potential future generalizations of the formalism to spaces including other observables.

2.1.1 *Assumptions*

Clusterix 2.0 performs an empirical determination of the frequency functions without relying on any previous assumption about

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

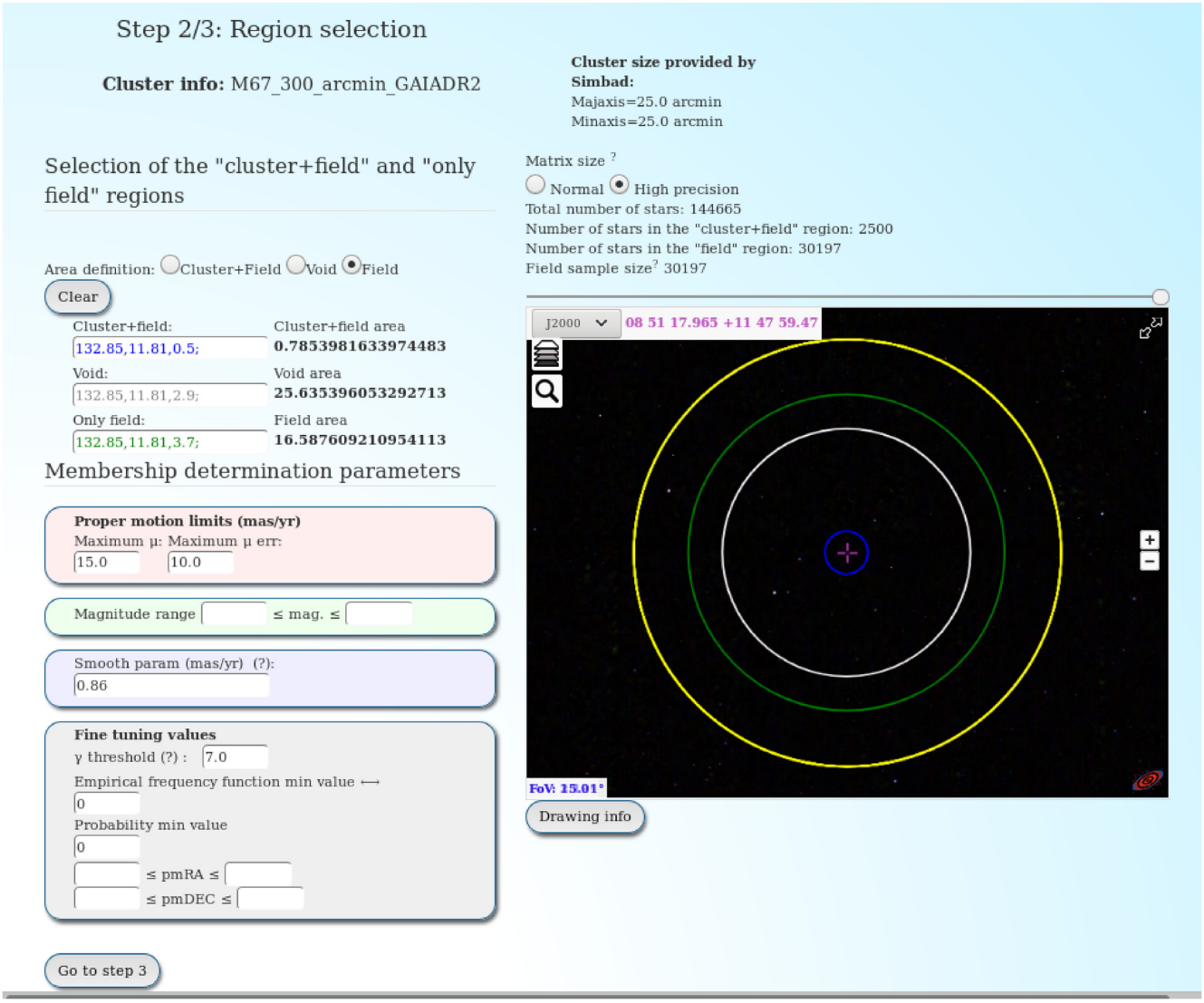


Figure 2. Step 2, where all the fine-tuning parameters of the `Clusterix` procedure are set for M67. In this step, c + f (blue), f (green), void (white) and total (yellow) areas are defined and parameters are chosen interactively. Tests can be performed at normal precision, while high precision is used here for final results. See the text for more details.

their profiles. First of all, some sky area has to be selected as the *work space* (we will use label 't' to refer to this area) with celestial coordinates and proper motions available for the stars inside that region. The procedure relies on several further assumptions.

(i) The population of field stars is spatially and kinematically uniform over the work space.

(ii) There is a non-field population that stands out, from the spatial density point of view, as an excess in the surface density of stars.

(iii) The non-field population does not occupy all of the work space, but is concentrated spatially, allowing us to distinguish two regions in the work space: the *only field* (label 'f') region, dominated by field stars, and the *cluster + field* (label 'c + f') region, which includes both field and non-field stars.

(iv) Field stars and non-field stars show proper-motion distributions on the VPD that are distinctive.

The second assumption implies that, at least in its current shape, `Clusterix` 2.0 yields accurate estimations about the expected number of cluster members for systems that show some spatial contrast against the background of field stars in the work space (see equations 1 and 2). This contrast helps, too, in the selection of field and cluster regions in the work space, at least in the very first steps of the process.

It is worth noting that these assumptions do not include any consideration about the shape of the proper-motion distributions, which may be highly asymmetric. In particular, we underline that the non-field population may be, in principle, composed from more than one cluster, for instance. In fact, two of the examples of scientific exploitation included in this article (see Sections 4.4 and 4.5) clearly illustrate this situation.

2.1.2 *Observables and scaling factors*

Once the three regions, t, f and c + f, have been defined, the first step leads to two observables for each of them: the spatial areas (A_t, A_f, A_{c+f}) and the number of stars they contain (n_t, n_f, n_{c+f}).

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

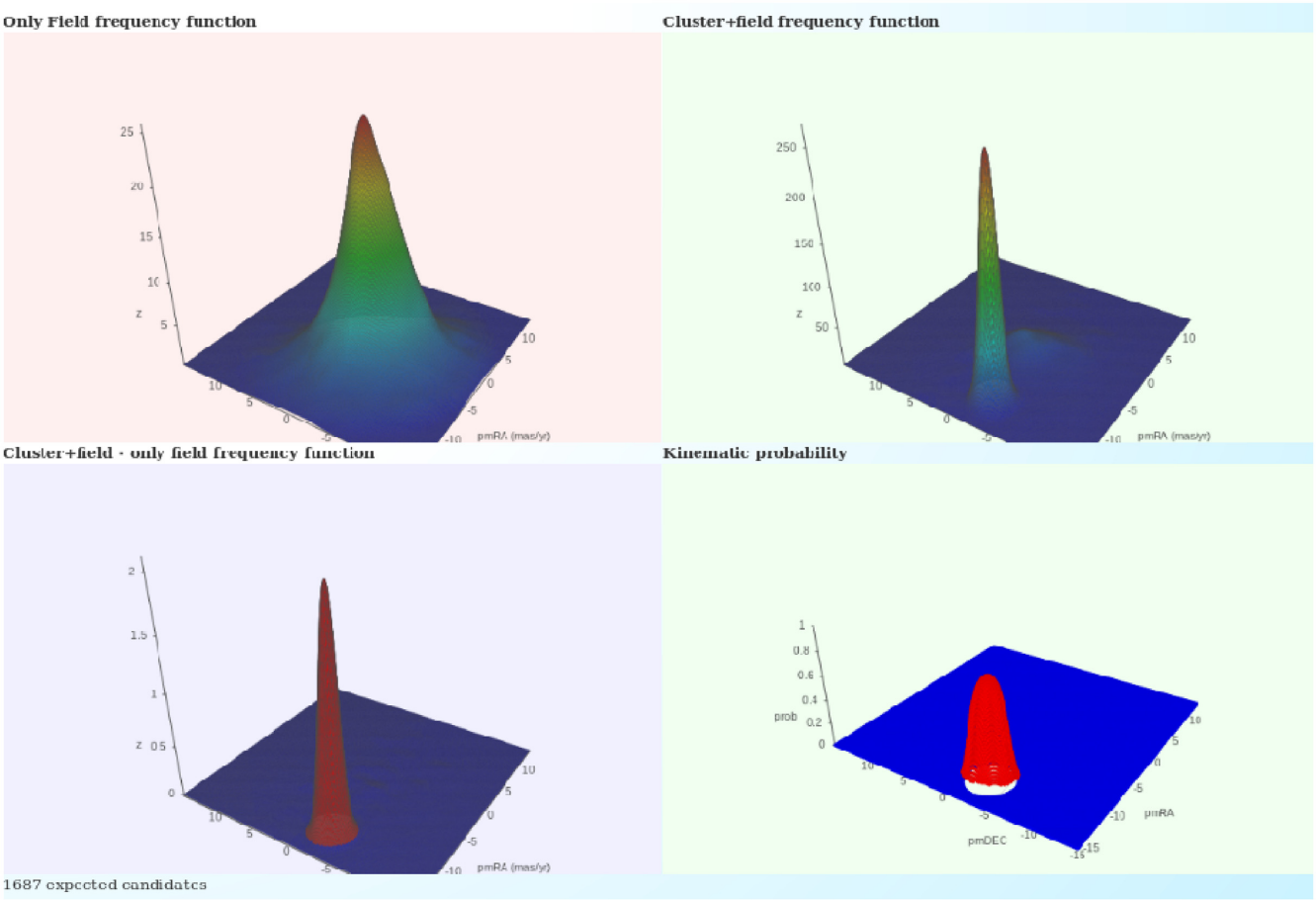


Figure 3. Interactive results of Step 2 of the Clusterix procedure for M67. Visualization of the frequency functions helps us to decide on the best set of areas and parameters: f and c + f (top) and c and probability (bottom). The user can rotate and change the angle of the plots as well as click in any cell to know its values. See text for more details.

The first quantity to be derived from these observables is the expected number of non-field stars in the work space A_t, N_c:

$$N_c = n_t - \frac{A_t}{A_f} n_f. \quad (1)$$

In equation (1), the computation of the expected total number of field stars in the work space, N_f, is implicit, computed from n_f through a straightforward scaling factor and relying on assumptions (i) and (ii):

$$N_f = \frac{A_t}{A_f} n_f. \quad (2)$$

The areas A_t, A_f and A_{c+f} will act several times as scaling factors in later steps of the procedure.

2.1.3 *Raw, normalized and scaled frequency functions*

Let us consider the VPD for an arbitrary set of N stars. The *frequency function* would provide the density of stars in this diagram, in units of stars per unit area. If, for instance, our proper motions are in units of milliarcseconds per year (mas yr^{-1}), the frequency function would be measured in stars per (mas yr^{-1})^2. Parametric methods try to fit functions to reproduce the observed density distribution in the VPD, but our non-parametric approach measures the true local density in the real diagram and provides the empirical distribution in the form of a table.

To this end, the VPD is divided into a (large) number of square cells. This number of cells is equal to an empirical value, which provides a good balance between the resolution needed for the VPD studied on the one hand and computing time on the other hand. The first consideration requires an adequate sampling of the VPD, taking into account the precision of the proper-motion data and the internal dispersion of the distributions (mainly that of the cluster). Limiting computer time is required, because Clusterix 2.0 benefits greatly from the interactive modification of parameters and, to this end, each step has to be kept within reasonable limits of CPU time. If the indications given in Section 3.2 to limit the overall size of the VPD are followed, then we have checked empirically that the number of cells used in the present implementation (100 × 100 for quick tests and 300 × 300 for higher precision) provides the intended balance. Each cell is described by its position in the array according to its row and column numbers, (i, j), and by the proper-motion values corresponding to its centre, (a_i, b_j). An empirical value of the frequency function is assigned to each cell, computed as follows:

$$\psi(a_i, b_j) = \sum_{k=1}^{N} K(a_k, b_k). \quad (3)$$

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026




```
#   Results were retrieved using Clusterix software
#   http://clusterix.cab.inta-csic.es/
#   In case of problems, please, report to: clusterix_archive_support@cab.inta-csic.es
#
#   Labels:
#
#       STAR_NO identificator of star retrieved from the input data
#       RA      right ascension of a star
#       DEC     declination of a star
#       PM_RA   proper motion in alpha
#       PM_DEC  proper motion in delta
#       PROB    probability that star belongs to evaluated open cluster
#       FLAG    M=Membership   NM=No membership
#
#   Parameters:
#
#       CLUSTER INFO: M67_300_arcmin_GAIADR2
#       PROPER MOTION CUTOFF: 15.00 mas/yr
#       PROPER MOTION ERR CUTOFF: 10.00 mas/yr
#       SMOOTH PARAMETER: 0.86
#       GAMMA FACTOR: 7.00
#       QUERY MIN MAG (STEP 1):
#       QUERY MAX MAG (STEP 1): 19
#       FILTERED MIN MAG (STEP 2):
#       FILTERED MAX MAG (STEP 2):
#
#       Boundaries can be circles (Ra(J2000)-center,Dec(J2000)-center, radius degrees) or
#       polygons (Ra,Dec Ra,Dec Ra,Dec ...). Each boundary is separated by ;
#
#       CLUSTER BOUNDARIES:
#       132.85,11.81,0.5;
#
#       FIELD BOUNDARIES:
#       132.85,11.81,3.7;
#
#       VOID BOUNDARIES:
#       132.85,11.81,2.9;
#
#       NUM STARS: 178748
#
#       EXPECTED NUMBER OF MEMBERS: 1687
#
#       For Webda clusters, STAR_NO is the merge of the Data source reference and the star number: Ref_Star
#
#       Region values: F = star was inside a "only field" region
#                      C = star was inside a "cluster+field" region
#                      V = star was outside the selected regions
```

Highest probability results sample (only 50 first shown).

##STAR_NO	RAJ2000	DECJ2000	pmRA	epmRA	pmDEC	epmDEC	Pll	ePll	BP	eBP	RP	eRP	G	eG	RV	eRV	PROB	Region
117145	132.809105	11.787142	-11.199	0.29	-2.878	0.213	1.2468	0.1719	18.7464	0.0199	16.8367	0.0071	17.8448	0.0019	NaN	NaN	0.84702224	V
117550	132.859049	11.897761	-11.002	0.069	-2.881	0.053	1.1579	0.0398	10.7649	7.0E-4	9.5266	6.0E-4	10.2024	2.0E-4	34.29	0.3	0.84702224	V

Figure 4. Step 3 of the `Clusterix` procedure for M67. All the parameters are saved together with the results in the file downloads as text, while the star list can be sent directly to TOPCAT/ALADIN via SAMP or a search of values from VOSA can be performed. Only the first rows of the results are shown.

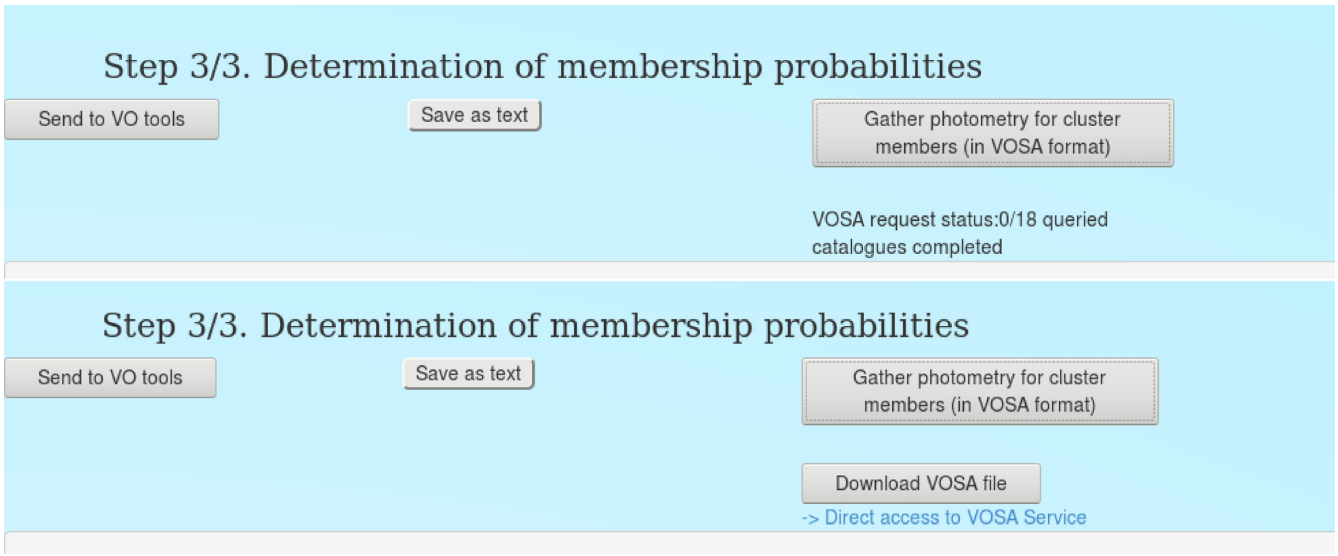


Figure 5. `Clusterix` 2.0. Step 3 for the gathering of photometry with VOSA. The top panel shows the system beginning the search. In the bottom panel, once the search is finished the user can download the photometry in the correct format for VOSA uploading.

Here, (a_k, b_k) are the coordinates (proper motions in our case) of the kth point of the observed sample with a total of N elements, and K is the so-called *kernel function* (see an explanation of this concept applied to this technique in Galadi-Enriquez et al. 1998c), a function that assigns to each element a weight that decreases with distance from the centre of cell (i, j). We use normal Gaussian kernel functions, defined as

$$K(a,b) = \frac{1}{2\pi h^2} \exp\left[-\frac{1}{2}\frac{(a-a_i)^2 + (b-b_i)^2}{h^2}\right]. \quad (4)$$

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

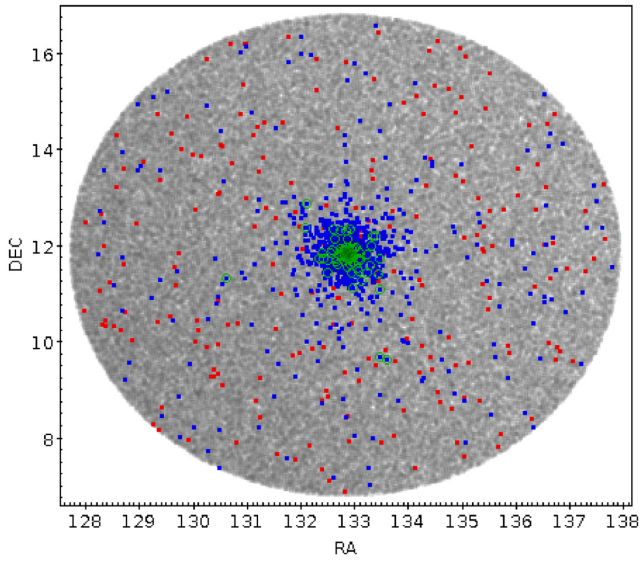


Figure 6. An area of 5 deg radius around M67 is studied. IAll stars up to magnitude $G = 19$ with proper motions from *Gaia* (178 748) are shown in grey. The selection made by Clusterix: 1699 stars (in red) is refined with a cut in parallax: 1440 members (in blue). Superimposed in green are the 84 stars with compatible *Gaia* radial velocity.

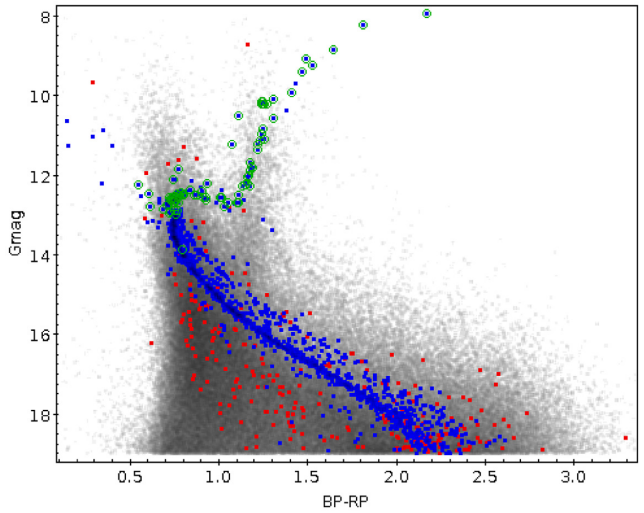


Figure 7. Colour–magnitude diagram of M67. All stars in a 5 deg radius around the centre up to magnitude $G = 19$ with proper motions (178 748) are shown in grey. The selection made by Clusterix: 1699 stars (in red and blue) is refined with a cut in parallax: 1440 members (in blue). Superimposed in green circles are the 84 stars with compatible *Gaia* radial velocity.

This way, a point located exactly at the centre of cell (i, j) provides the maximum contribution to compute the local density and all other points in the VPD contribute progressively less to the density evaluated at that cell. Number h is usually named *the smoothing parameter* in this context and it is measured in the same units as the proper motions. Clusterix 2.0 allows the user to modify h at will, although the system proposes by default a value computed by means of Silverman's rule (Silverman 1986):

$$h = \left(\frac{4}{d+2}\right)^{1/(d+4)} \sigma N^{-1/(d+4)}. \quad (5)$$

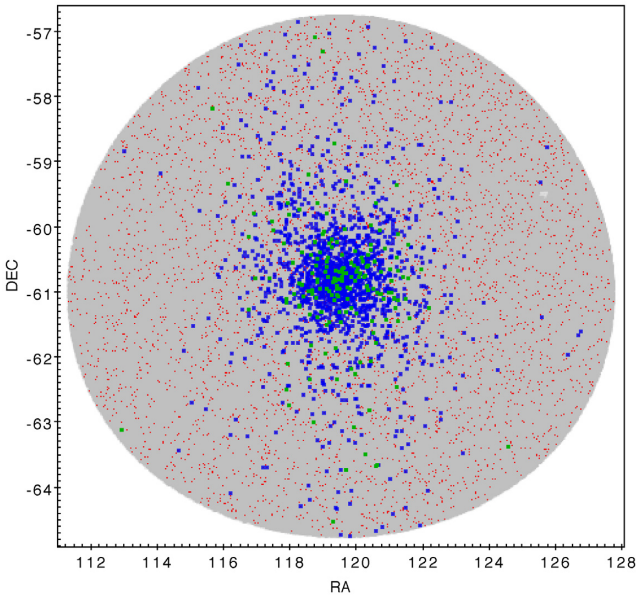


Figure 8. Area of NGC 2516: translucent grey shows all stars in a 4 degree radius around the centre (552 583 stars). The selection made by Clusterix, 6210 stars (in red and blue), is refined with parallax: 1819 members (in blue), where 175 in green show compatible *Gaia* radial velocities. A clear elongated structure is visible.

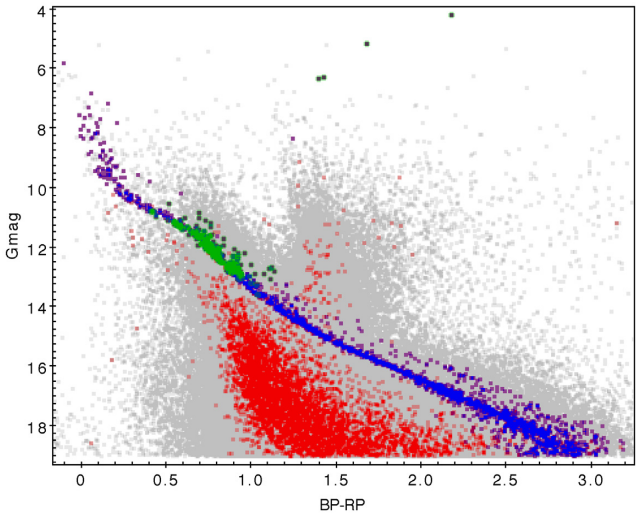


Figure 9. Colour–magnitude diagram of NGC 2516. Translucent grey shows all stars in a 4 degree radius around the centre (552 583 stars). The selection made by Clusterix, 6210 stars (in red and blue), is refined with parallax: 1819 members (in blue), where 175 in green show compatible *Gaia* radial velocities.

In equation (5), N is, again, the number of elements in the VPD that is being analysed, σ is the average marginal variance of the sample and d is the dimension of the space (in this case, $d = 2$).

The previous formalism is applied in Clusterix 2.0 to the VPDs observed in the regions c + f and f. The same value of h deduced from the sample in the c + f region according to equation (5) is used in the f region. This way, the empirical, raw frequency functions are computed for the mixed (c + f) and only field (f) samples: $\psi_{\rm c+f}$ and $\psi_{\rm f}$.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

Table 1. Comparison of our results for M67 with other authors.

	$\mu_{\alpha*}$ (mas yr^{-1})	μ_δ (mas yr^{-1})	ϖ (mas)	$N_{\rm mem}$	$G_{\rm lim}$ (mag)
This study	-10.99 ± 0.28	-2.94 ± 0.26	1.13 ± 0.11	1440	19
Cantat-Gaudin et al. (2018b)	-10.986 ± 0.193	-2.964 ± 0.201	1.135 ± 0.051	691	18
Gaia Collaboration et al. (2018b)	-10.974 ± 0.006*	-2.940 ± 0.006*	1.133 ± 0.001*	1520	20

*Note.** In the case of Gaia Collaboration et al. (2018b), the errors quoted are uncertainties and not standard deviations.

Table 2. Comparison of our results for NGC 2516 with other authors.

	$\mu_{\alpha*}$ (mas yr^{-1})	μ_δ (mas yr^{-1})	ϖ (mas)	$N_{\rm mem}$	$G_{\rm lim}$ (mag)
This study	-4.63 ± 0.43	11.15 ± 0.35	2.4 ± 0.1	1819	19
Cantat-Gaudin et al. (2018b)	-4.748 ± 0.441	11.221 ± 0.345	2.417 ± 0.045	798	18
Gaia Collaboration et al. (2018b)	-4.6579 ± 0.0075*	11.1517 ± 0.0075*	2.4118 ± 0.0006*	2518	20

*Note.** In the case of Gaia Collaboration et al. (2018b), the errors quoted are uncertainties and not standard deviations.

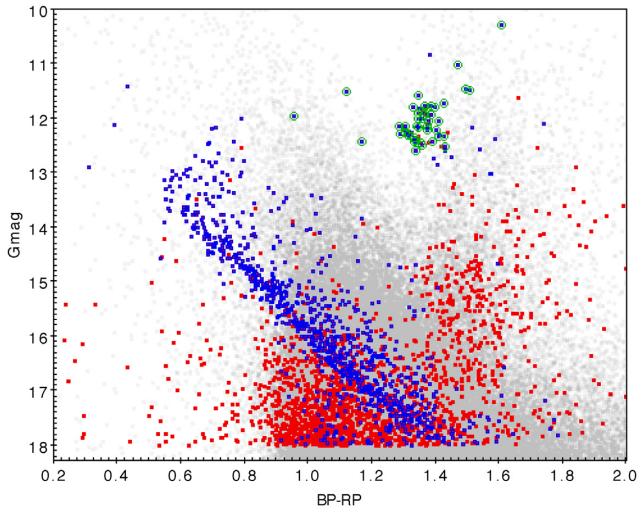


Figure 10. Colour–magnitude– diagram of NGC 1817. All stars in a 4 degree radius around the centre (102 313 stars) are shown in translucent grey. The selection made by `Clusterix`: 2898 stars (in red and blue), is refined with a cut in parallax: 1000 members (in blue), where 43 in green show compatible *Gaia* radial velocities.

Now the scaling factors are used to transform $\psi_{\rm c+f}$ and $\psi_{\rm f}$ into their area-normalized versions, $\Psi_{\rm c+f}$ and $\Psi_{\rm f}$, which would represent the frequency functions expected for a unit area on the sky (let us say stars per (mas yr^{-1})^2 and per solid angle unit):

$$\Psi_{\rm c+f} = \frac{1}{A_{\rm c+f}}\psi_{\rm c+f}\Psi_{\rm f} = \frac{1}{A_{\rm f}}\psi_{\rm f}. \quad (6)$$

These new functions are independent of the (arbitrary) areas selected for the regions and can be compared directly.

The next step will be to deduce the area-normalized frequency function expected for the non-field population. This population, which we will label as 'c' (because most often it will correspond to a cluster), is computed as the simple difference of the two previous ones:

$$\Psi_{\rm c} = \Psi_{\rm c+f} - \Psi_{\rm f} = \frac{1}{A_{\rm c+f}}\psi_{\rm c+f} - \frac{1}{A_{\rm f}}\psi_{\rm f}. \quad (7)$$

In a final manipulation of the frequency functions, the scaled versions are deduced. To do this, we first have to define the volumes of the frequency functions as the direct sum of the tabulated values. For the raw functions, we have

$$v_{\rm c+f} = \sum \psi_{\rm c+f} v_{\rm f} = \sum \psi_{\rm f}. \quad (8)$$

In the same way, for the area-normalized functions f and c we have

$$V_{\rm f} = \sum \Psi_{\rm f} V_{\rm c} = \sum \Psi_{\rm c}. \quad (9)$$

The scaled frequency functions are defined in such a way that their volumes are equal to the number of stars of their respective populations:

$$\overline{\Psi}_{\rm f} = \frac{N_{\rm f}}{V_{\rm f}}\Psi_{\rm f}\overline{\Psi}_{\rm c} = \frac{N_{\rm c}}{V_{\rm c}}\Psi_{\rm c}. \quad (10)$$

It can be easily shown that the first of these scaled frequency functions (the function for the field population, f) is equivalent to scaling the raw function to sample size $N_{\rm f}$:

$$\overline{\Psi}_{\rm f} = \frac{N_{\rm f}}{v_{\rm f}}\psi_{\rm f} = \frac{A_{\rm t} n_{\rm f}}{A_{\rm f} v_{\rm f}}\psi_{\rm f}. \quad (11)$$

However, the scaled function for the non-field population, c, would be much more complicated to reduce to directly observed quantities.

Scaled frequency functions can be regarded as expressed in units of stars per (mas yr^{-1})^2, i.e. for each population they measure the number of stars found at each of the cells of the VPD.

2.1.4 *Noise and membership probabilities*

The assumptions enumerated in Section 2.1.1 cannot be fulfilled perfectly. Spatial and kinematical uniformity of the field is only approximate in the real sky, a situation that worsens for larger fields of view. Cluster populations do not always stand out clearly as an overdensity in the plane of the sky, as in the case of faint clusters and also at the outer areas of even dense clusters. Irregularities in the field spatial distribution may reduce the contrast of the cluster against the field. Finally, open clusters are much more extended than previously thought, with wide coronas that may make it difficult to find areas completely devoid of cluster members. As a result, the frequency functions predicted by equations (7) and (10) for the non-field population, c, are affected by some noise. This noise level can be estimated from the negative values of the function $\overline{\Psi}_{\rm c}$, which obviously have no physical meaning. Assuming that these negative values are caused by noise and that this noise displays the same

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

distribution towards positive values, the noise level can be identified with

$$\sigma_- = \sqrt{\frac{\sum_- \overline{\Psi}_c}{M}}, \quad (12)$$

where the summation is performed over those cells with negative value and M is the number of negative-valued cells. This estimation allows us to evaluate the signal-to-noise ratio at each cell of this function. The third assumption from Section 2.1.1 implies that the contamination by cluster stars in the field area should be negligible. Under these conditions, equation (12) would provide a realistic measure of the perturbations due to the random nature of the spatial distribution of stars. However, in some cases, the corona of the cluster may introduce some overestimation of the field population, which would translate into a systematic negative bias that may increase the value of σ_-. This overestimation of the noise level, if present, may imply an underestimation of the number of stars belonging to the cluster, given the later use of σ_- in the procedure (see Section 3.4).

Finally, membership probabilities can be assigned to each cell of the VPD over the working space, as follows:

$$P = \frac{\overline{\Psi}_c}{\overline{\Psi}_c + \overline{\Psi}_f}. \quad (13)$$

Normally, the probability function P will be set to zero at those cells of the VPD where the signal-to-noise ratio of function $\overline{\Psi}_c$ is below some threshold $\gamma\sigma_-$, where γ is a parameter that has to be adjusted in each case, depending on the characteristics of the sample.

The simplicity of equation (13) is more apparent than real, in the sense that it is not straightforward to express that definition in terms of the observables and the raw frequency functions.

Given any star from the original sample, the probability of it belonging to the non-field population is identified with the value of function P at the cell of the VPD occupied by the star, according to its proper motions. The expected number of non-field stars, N_c (equation 1), may serve as a guide to set a probability cut-off.

3 THE CLUSTERIX WORKFLOW

The general, theoretical frame formulated in Section 2.1 is implemented into Clusterix 2.0 in the following way.

3.1 Data query

The first step of the procedure is the definition of the data source and the working space.

Users can gather information on a particular region in the sky from the following catalogues: *Gaia* DR2 (Gaia Collaboration et al. 2018a), *Tycho–Gaia* Astrometric Solution (TGAS: Gaia Collaboration et al. 2016a), PPMXL (Roeser, Demleitner & Schilbach 2010), United States Naval Observatory CCD Astrograph Catalog editions 4 (UCAC4: Zacharias et al. 2012) and 5 (UCAC5: Zacharias, Finch & Frouard 2017) and Hot Stuff for One Year (HSOY: Altmann et al. 2017). Queries can be made using equatorial coordinates and a search radius. Any object identifier that can be resolved by the SESAME[9] service can also be used. In both cases, the query can be filtered by specifying a range of magnitudes (G, G_{BP} or G_{RP} in the case of *Gaia*, J, H or K_S for 2MASS and so on). In the case of *Gaia* DR2, users can also select the Q filter, which follows equations (C.1) and (C.2) from Lindegren et al. (2018). These filters help us to work with an astrometrically clean subset (see also Arenou et al. 2018).

Alternatively, users can upload ASCII files with their own data. This comes very handy for custom datasets or synthetic data. Files must have comma-separated values (CSV files) and the fields can be in any order. Mandatory fields and labels are ID, RAJ2000, DEJ2000, pmRAcosdelta and pmDEC. If information on magnitude (MAG) and proper-motion uncertainties (epmRAcosdelta, epmDEC) is available in the input file, it will be possible to use this in subsequent steps. The magnitude is used only to apply limits to the workplace if desired. Proper-motion errors are also only used to apply limits that allow data of bad quality to be discarded, if present, but do not play any additional role in the definition of the frequency functions themselves. Depending on the precision of

[9] http://cds.u-strasbg.fr/cgi-bin/Sesame

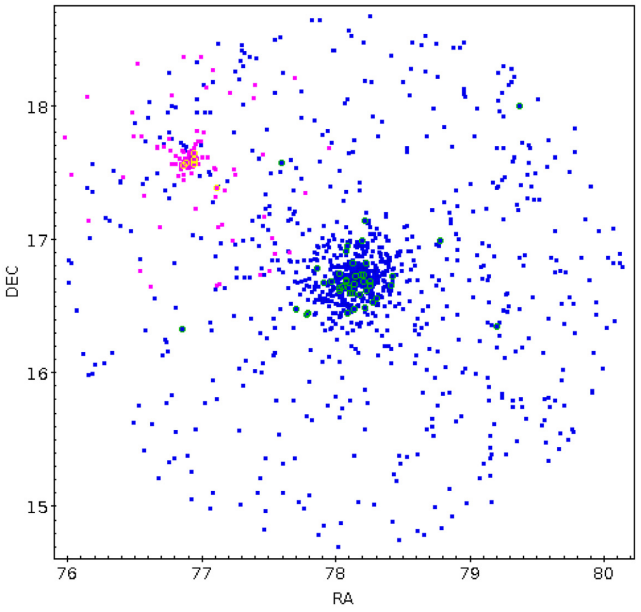


Figure 11. Area of NGC 1817 (members in blue) and Juchert 23 (members in pink). Radial velocity members are marked as open circles (green for NGC 1817, yellow for Juchert 23.).

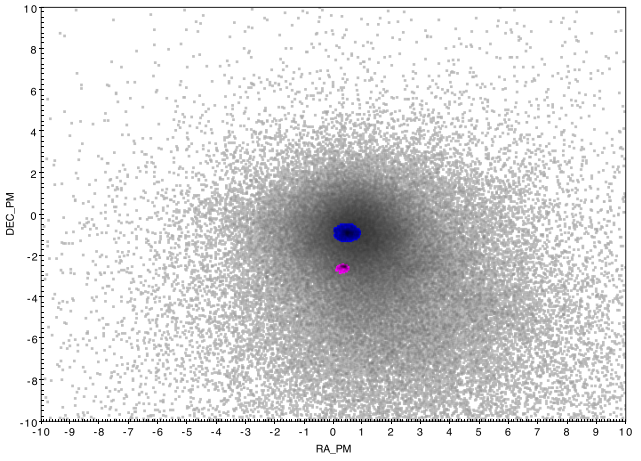


Figure 12. Vector point diagram of NGC 1817 (members in blue) and Juchert 23 (members in pink).

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

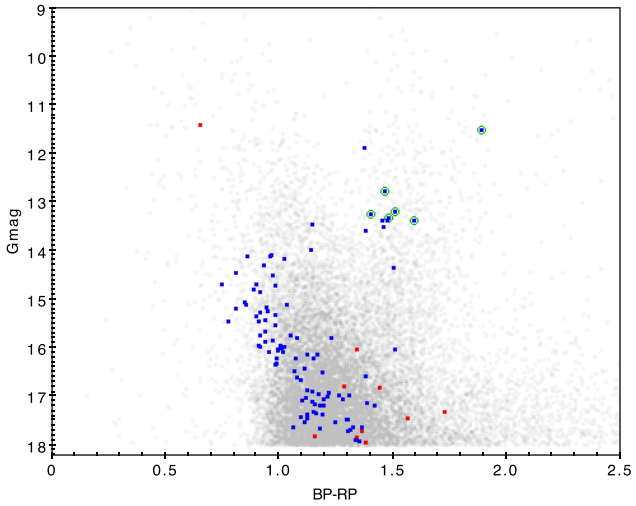


Figure 13. Colour–magnitude diagram of Juchert 23, in red and blue; 115 members by `Clusterix` and in blue, the 105 members stars with an additional cut in parallax. In green open circles, six stars with coherent *Gaia* radial velocities are shown.

the chosen catalogue, the user can define the value of the error in proper motion at which the measurements are considered of bad quality. In this article, using *Gaia* DR2 proper motions, we keep the limit on errors below 5 mas yr^{-1} to avoid outliers and limit the computation time, although the errors are currently much lower in all cases.

Extra columns present in the input file will be accepted by CLUSTERIX 2.0 but are ignored for calculations. By hovering the mouse over the question mark next to this section in the online article (Fig. 1), the user can download some examples of input files that can be used as templates. The privacy of user data is guaranteed, since the files uploaded by users are not kept in the system.

Finally, the main source of data in the previous version, `Clusterix` 1.0, the WEBDA database,[10] can still be used as a source of information on positions and proper motions.

The result of the query is visualized using ALADIN LITE,[11] where the objects obtained from the query are plotted as red diamonds. The user can zoom and pan, as well as select different background images. As the representation of tens of thousands of objects is quite demanding from a computational point of view, we limit the individual representation of objects (the red diamonds) to 40 000. If the result of a query exceeds this amount, then just a yellow circle enclosing the search region is drawn.

Then, the result of the query can be sent to the next step ('Membership from proper motions') or be downloaded as a CSV file or sent to VO applications using SAMP[12] for further analysis.

3.2 Selection of the *cluster* + *field* and *field* regions

The second step is to select the *cluster* + *field* (c + f) and *only field* (f) regions mentioned in Section 2.1.1. The f and c + f regions may be, in principle, arbitrary in shape. In fact, they do not even need to be connected spaces (each of them may be made up from separate pieces), nor simply connected spaces (they may even have holes).

[10] http://webda.physics.muni.cz/
[11] http://aladin.u-strasbg.fr/AladinLite/doc/
[12] http://www.ivoa.net/documents/REC/App/SAMP-20090421.html

The definition of these areas is one of the most critical decisions to take by the user and `Clusterix` 2.0 offers several ways to shape and reshape these areas interactively. In the current implementation, only circular areas are allowed.

The simplest option relies on mouse clicks to draw circles that define the c + f (inside the blue circle in Fig. 2) and f (inside the green circle in the same figure down to the following circle) regions. The system also includes an easy way to set up a 'buffer' void area around the c + f region that will be excluded from both areas (the area marked between the white and the blue circles in the figure). This is to define a clean field area, avoiding an intermediate region that could still have a significant number of cluster members, so no stars inside that buffer are taken into account for the frequency function calculations. See an example of the three areas in Fig. 2.

Alternatively, the user can specify the circular areas directly, writing their parameters (in decimal degrees) in the corresponding boxes (format: 'ra,dec,radius;').

These circles do not need to be concentric and other distributions are possible (see e.g. Section 4.5 and Fig. A18). If there is an overlap area of the two circles (f and c+f), that area is assigned only to the c + f and excluded from the f sample. The buffer (void area) is also excluded from the f sample.

Once the areas are chosen, the star sample can be filtered further in this second step according to the 'Membership determination parameters'.

(i) Proper motion limits. Maximum value of the total proper motion (to reduce the size of the VPD for the working space, discarding from the beginning objects that clearly cannot belong to the expected cluster population), and maximum value of the total proper motion error (to remove data of dubious quality).

(ii) Magnitude range to limit the selection further (with respect the magnitude chosen in Step 1), if desired.

(iii) Smoothing parameter. `Clusterix` 2.0 proposes a default value for the smoothing parameter h on the basis of equation (5) applied to the c + f sample and this can be modified by the user. The same h value is used to compute both frequency functions. As h represents the radius of the kernel window (i.e. the sigma of the Gaussian kernel K), a large value would blur out the details of the frequency functions, while a small value would yield noisy results, because a very low number of individuals would receive significant weights.

(iv) Fine-tuning values. Here we mainly fix the value of γ explained in the previous section (Section 2.1.4). To avoid meaningless probability values, `Clusterix` 2.0 restricts the probability calculations to stars with densities γ times higher than the noise.

All these parameters, shown in Fig. 2, can be chosen interactively, depending on the data used and/or the region studied, until satisfactory results are obtained.

3.3 Determination of the empirical frequency functions

The empirical determination of the frequency functions is performed according to the formalism of Section 2.1. The assumptions stated in Section 2.1.1 are critical for membership determination. Assumption (i) is usually fulfilled, but may be problematic for very large work spaces. Finding regions that fulfil assumption (iii) may be more difficult because, a priori, there is no information about the true extension of the cluster, and even its location on the sky may be uncertain. It is thus important to perform tests with areas of very different sizes, searching for a reasonable trade-off between cleanness (absence of a significant amount of cluster members) and

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

Table 3. Comparison of our results for NGC 1817 with other authors.

	$\mu_{\alpha*}$ (mas yr^{-1})	μ_δ (mas yr^{-1})	ϖ (mas)	$N_{\rm mem}$	$G_{\rm lim}$ (mag)
This study	0.48 ± 0.19	−0.89 ± 0.17	0.54 ± 0.07	1000	18
Cantat-Gaudin et al. (2018b)	0.485 ± 0.118	−0.89 ± 0.1	0.551 ± 0.056	460	18
Balaguer-Núñez et al. (2004a)	0.29 ± 0.10	−0.96 ± 0.07	–	169	14.5

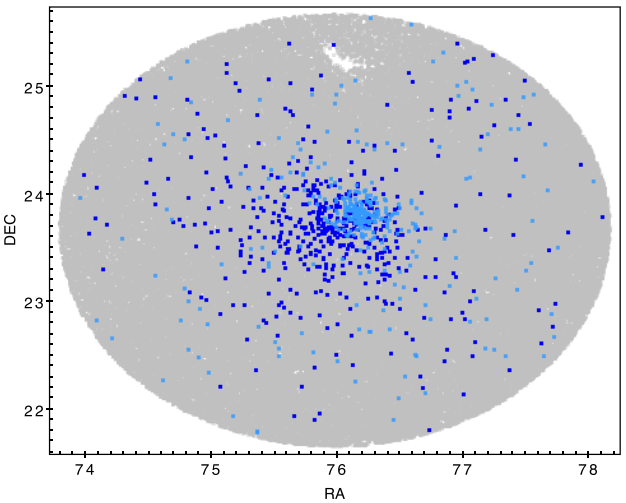


Figure 14. The two known overlapping clusters: NGC 1750 in blue with 504 member stars and NGC 1758 in light blue with 363 member stars. In grey, dots show all stars in a 2 degree radius around the centre.

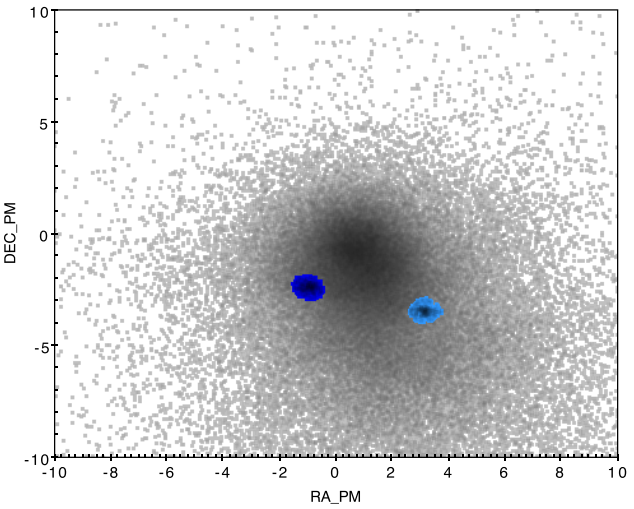


Figure 15. Vector point diagram of the two overlapping clusters: NGC 1750 in blue (504 stars) and NGC 1758 in light blue (363 stars). In grey, all stars in a 2 degree radius around the centre are shown.

signal-to-noise ratio (working area not too small, to avoid small number statistics).

The frequency functions are determined empirically as tables that contain cells defined in the VPD. To facilitate the first, tentative tests, Clusterix 2.0 works with two different matrix sizes to discretize the VPD: 100 × 100 ('Normal') or 300 × 300 cells ('High precision'). The first size is useful for fast queries, while the

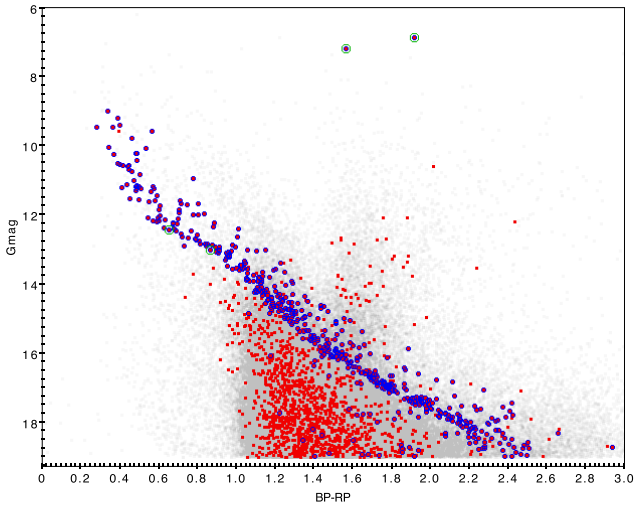


Figure 16. Colour–magnitude diagram of NGC 1750: in grey are the 111 799 stars in the field, in red and blue the 1827 candidates and in blue the final selection of 504 stars. Green open circles denote the four stars with coherent *Gaia* radial velocities. See text for more details.

second one allows more precise data representation and results. See Figs 2 and 3 for the example at High precision.

One of the interesting capabilities of Clusterix 2.0 is that, after selecting the c + f and f regions, the tool provides a visualization of the resulting empirical, raw frequency functions $\psi_{\rm c+f}$ and $\psi_{\rm f}$. This allows the user to evaluate the suitability of the selected regions. Usually, if a cluster is present, $\psi_{\rm f}$ should display a broad shape with only one maximum and $\psi_{\rm c+f}$ would include that same profile with an additional, normally sharper, peak due to the cluster (Fig. 3, top panel). Figures can be interactively rotated and zoomed in and out. Also, by placing the cursor on any point of the graphics, its coordinates (pmRA, pmDEC, frequency value) are displayed.

There is another parameter that can be fixed for the case of large sample size: 'Field sample size'. Frequency function calculations can be optimized by reducing the number of stars used. Statistically we only need a subset of stars to obtain the relevant information that is needed for determining the functions. Lower values of this parameter can also be used for quick testing. The suggested upper limit of 50 000 stars is set, as no significant differences are obtained with greater numbers and the calculations can then take too long to complete. See example in Section 4.2 and Fig. A2.

3.4 Visualizing the results

After each redefinition of the c + f and f regions, the system recomputes and displays the empirical frequency functions and also calculates and shows the scaled cluster frequency function $\overline{\Psi}_{\rm c}$ (equation 7) and probability distribution P (equation 13). In these two graphs, high signal-to-noise ratio areas are marked in red (see

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

Table 4. Comparison of our results for NGC 1750 and NGC 1758 with other authors.

	$\mu_{\alpha*}$ (mas yr^{-1})	μ_{δ} (mas yr^{-1})	ϖ (mas)	$N_{\rm mem}$	$G_{\rm lim}$ (mag)
NGC 1750					
This study	-0.97 ± 0.23	-2.40 ± 0.21	1.34 ± 0.16	504	19
Cantat-Gaudin et al. (2018b)	-0.96 ± 0.246	-2.366 ± 0.201	1.361 ± 0.09	439	18
NGC 1758					
This study	3.13 ± 0.23	-3.45 ± 0.20	1.08 ± 0.15	363	19
Cantat-Gaudin et al. (2018b)	3.156 ± 0.146	-3.465 ± 0.129	1.103 ± 0.059	146	18

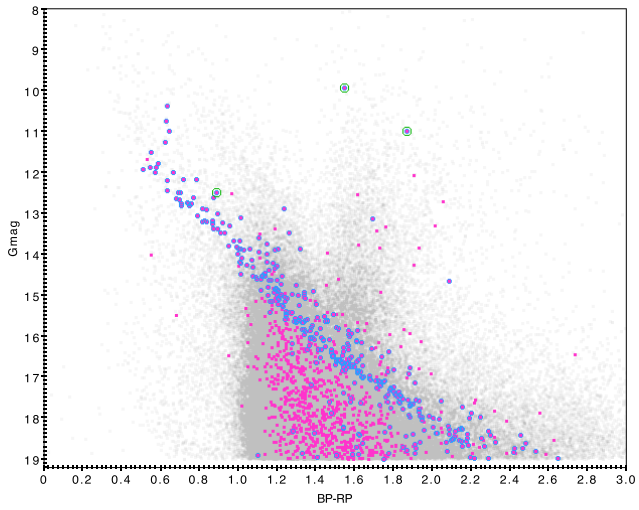


Figure 17. Colour–magnitude diagram of NGC 1758: in grey are the 111 799 stars in the field, in pink and light blue the 1162 candidates and in light blue the final selection of 363 stars. Green open circles denote the three stars with coherent *Gaia* radial velocities. See text for more details.

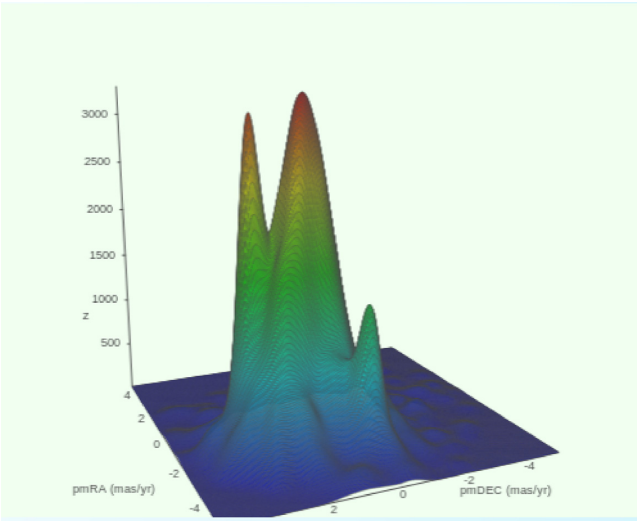


Figure 18. Frequency function of Ruprecht 26 area with the three kinematically differentiated populations: two clusters and the field in the middle.

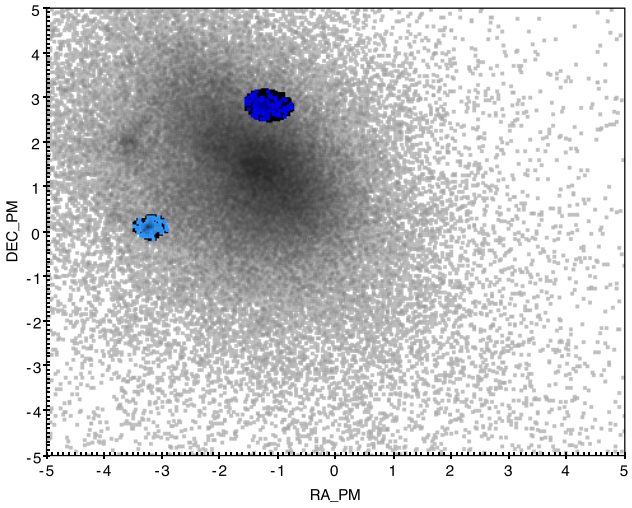


Figure 19. Vector point diagram of Ruprecht 26 (in light blue) and CLUSTERIX 1 (in blue). In black all stars with probability greater than 0.5 are shown and grey shows all stars in a 1 degree radius around the centre.

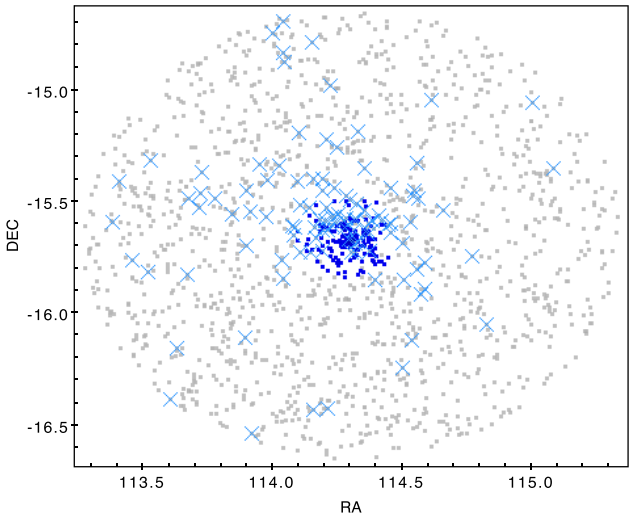


Figure 20. Area of Ruprecht 26 in light blue crosses and CLUSTERIX 1 in blue dots. Grey dots show all stars with probability greater than 0.5.

Fig. 3, bottom panel). According to what was stated in Section 2.1.4, P is null for those cells of the VPD with $\overline{\Psi}_{\rm c} < \gamma\sigma_{-}$, where the γ factor is set by the user.

Once the configuration of c + f and f regions and the other parameters are considered satisfactory enough, the user executes the command 'Go to step 3' to trigger computation of the results for the whole work space and the catalogue is created. The results are displayed in a new page and they can be downloaded as a CSV file, or sent to any VO tool (e.g. TOPCAT) using the SAMP protocol. A summary of the results is shown directly in the browser, along with

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

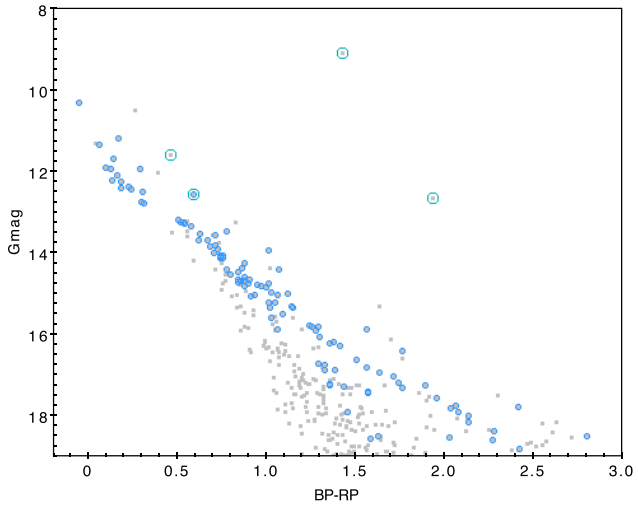


Figure 21. Colour–magnitude diagram of Ruprecht 26. The 109 members are in light blue, with four stars with coherent *Gaia* radial velocities shown as open green circles. In grey, the 331 candidate stars are shown. See more details in the text.

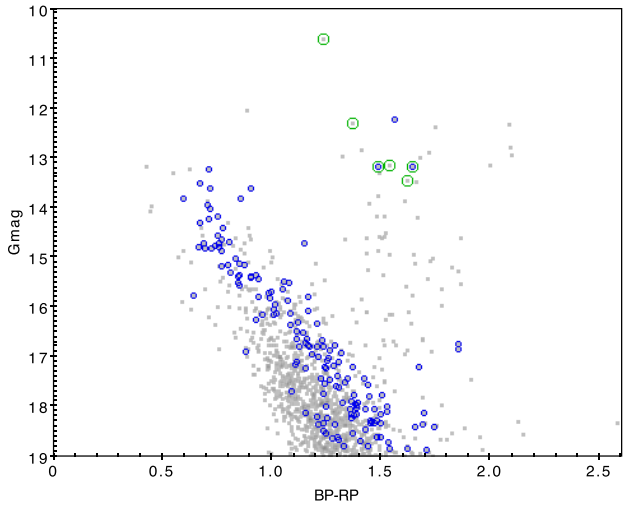


Figure 22. Colour–magnitude diagram of CLUSTERIX 1. The 157 members are shown in blue, with six stars with coherent *Gaia* radial velocities shown as open green circles. In grey, the 1238 candidate stars are shown. See more details in the text.

the parameters used for the calculations, so the experiment can be replicated again in the future if needed (see Fig. 4).

An estimation of the number of cluster members is provided, based on the expected overpopulation of the cluster region provided in step 2, but this number may not be exact, due to an inexact region determination or a non-homogeneous field distribution. This expected number of members, N_c, is given as a result and would correspond to the first N_c stars with higher probability of being members. However, the full sample is included in the given file with the individual probabilities, allowing users to apply their own decision using further criteria if wished.

For those N_c stars with higher probability of being cluster members, it is possible to gather additional photometric data using the VOSA[13] service. The user can download this photometry, make their own selection, and upload it again to work with VOSA. Clusterix 2.0 implements a simple login system in order to access past VOSA queries. Also, as they can take a long time, the system checks periodically whether a photometry query is finished and sends an email to the user. This user authentication system is not needed for any other Clusterix 2.0 functionality. See Fig. 5.

4 THE SCIENCE CASES

The tool discussed in the previous sections is applied to a few specific cases using *Gaia* DR2 to demonstrate its performance in both easy and complex scenarios. We have chosen five areas with clusters of different sizes, ages and distances, as well as areas where more than one cluster is present, from known overlapping clusters to unexpected new ones.

We follow the recommendations by Arenou et al. (2018) and Lindegren et al. (2018, equations C.1 and C.2) for an astrometrically clean subset. Although the *Gaia* limiting magnitude is $G = 21.0$, the median uncertainty for bright sources ($G < 15$ mag) is 0.06 mas yr^{-1} in proper motion and goes up to 0.3 mas yr^{-1} for $G = 18$.

Depending on the cases, we only consider sources brighter than $G = 18$ or 19. Fainter objects have larger uncertainties that may blur the separation between the cluster and the field in proper motion space. Also, this cut in magnitude significantly reduces the amount of sources to be managed, as 80 per cent of *Gaia* DR2 sources are fainter than $G \sim 18$.

The positions and main parameters of the stars in the areas of the clusters discussed in this section are given in Table 6 in electronic format only, at the CDS. [14]

4.1 NGC 2682 (M67)

M67 is an old cluster (∼3.6 Gyr: Bossini et al. 2019) at about 900 pc with a near-solar metallicity and low reddening. It is one of the best studied open clusters, considered a cornerstone of stellar astrophysics and used as a calibrator in many surveys. Still, there was no study covering a large area, in spite of some studies on its corona showing that it is an extended cluster. Gaia Collaboration et al. (2018b) studied an area of 1 degree with $G < 20$ and found 1520 members.

We study the data from *Gaia* DR2 within a radius of 5 degrees from the cluster centre and magnitude limit $G < 19$ (178 748 stars). The extended halo of this cluster has been studied by Carrera et al. (2019b), based on the data explained here. In the search for the furthest objects, their study extended up to 10 degrees (but $G < 18$) and shows that the cluster radial distribution is $r_{85\ \mathrm{per\ cent}} = 38$ arcmin, well under our chosen 5 degrees. The bigger the field, the worse the representation of the field results (see Appendix A in their article) and the noisier the solution. See the area covered in this study in Fig. 6.

In step 2 of Clusterix, we apply a constraint in proper motions ($\mu \leq 15$ mas yr^{-1}) and in proper-motion errors (below 10 mas yr^{-1}) to avoid outliers and limit the computation time (see Fig. 2).

Clusterix 2.0 gives us an estimated number of members of 1687, which corresponds to a probability cut at 0.794 79. With this probability cut, we get 1699 candidate members using only proper motions. If we apply a 1σ cut in parallax to that selection, we obtain

[13] http://svo2.cab.inta-csic.es/theory/vosa/
[14] Table 6 is only available online at the CDS.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

Table 5. Comparison of our results for Rup 26 with those of other authors.

	$\mu_{\alpha*}$ (mas yr^{-1})	μ_δ (mas yr^{-1})	ϖ (mas)	$N_{\rm mem}$	$G_{\rm lim}$ (mag)
This study	-3.20 ± 0.12	0.11 ± 0.10	0.92 ± 0.06	109	19
Cantat-Gaudin et al. (2018b)	-3.2 ± 0.094	0.071 ± 0.092	0.923 ± 0.049	81	18

a final list of 1440 member stars (see the colour–magnitude diagram in Fig. 7). From this sample, we check those with radial velocities from *Gaia* DR2 (88 stars) and use it only to calculate a median radial velocity with a 2σ clipping, resulting in $v_{\rm rad} = 34.5 \pm 2.0$ km s^{-1} from 84 stars. Soubiran et al. (2018) find $v_{\rm rad} = 33.80 \pm 1.06$ km s^{-1} from 64 member stars.

The average proper motion and parallax of the cluster are $\mu_{\alpha*} = -10.99 \pm 0.28$ mas yr^{-1}; $\mu_\delta = -2.94 \pm 0.26$ mas yr^{-1}; $\varpi = 1.13 \pm 0.11$ mas from 1 440 member stars. The average parallax from the 84 radial velocity members is $\varpi = 1.16 \pm 0.06$ mas.

All the errors given represent the standard deviation of the distributions. Comparing this result with the 691 member stars found in the study in the 30 arcmin central area with a magnitude limit of $G = 18$ from Cantat-Gaudin et al. (2018b), we find 677 members in common (98 per cent agreement). From the 1251 members of Gaia Collaboration et al. (2018b) with $G < 19$, we have 1116 members in common (89 per cent agreement). The comparison of the final values is shown in Table 1.

4.2 NGC 2516

NGC 2516 is a young cluster ($\sim$ 150 Myr), at a distance of about 400 pc and with $E(B-V) \sim 0.12$ mag. It has a known radial structure already visible in fig. D.24 of Gaia Collaboration et al. (2017) from *Gaia* DR1 (Gaia Collaboration et al. 2016b).

We have studied an area of 4 deg radius around the centre and magnitude limit ($G < 19$) (552 583 stars). It is a very populated region and we tested different areas to maximize the contrast between the field and the cluster. The final result (see Appendix for more information on the areas and fine-tuning parameters chosen) gives a number of expected members of 2537. However, in this case, a very close cluster, the parallax is a greatly discriminant parameter. We therefore accept all stars with probability greater than 0 (which in this case, due to the γ noise level cut, corresponds to $P > 0.6$) and then we apply a cut in the parallax distribution that, given the obvious overdensity, is quite straightforward ($2 < \varpi < 3$ mas), resulting in 1819 member stars (see Fig. 8). The resulting colour–magnitude diagram shows the clean selection (see Fig. 9).

From the 175 stars with compatible *Gaia* radial velocities, the average radial velocity of the cluster is $v_{\rm rad} = 24.4 \pm 2.6$ km s^{-1}.

The proper motion and parallax of the cluster are $\mu_{\alpha*} = -4.63 \pm 0.43$ mas yr^{-1}; $\mu_\delta = 11.14 \pm 0.35$ mas yr^{-1}; $\varpi = 2.4 \pm 0.1$ mas for the 1819 member stars. The average parallax from the 175 radial velocity members is $\varpi = 2.42 \pm 0.08$ mas.

The member stars are distributed as seen in Fig. 8 where the elongated distribution is clear even among the 175 radial velocity members.

Comparing this result with the 798 member stars found in the study in the 1 deg central area with a magnitude limit of $G = 18$ by Cantat-Gaudin et al. (2018b), we find 696 in common with our member list (87 per cent agreement). From the 2125 members by Gaia Collaboration et al. (2018b) with $G < 19$ in a 2 deg area, we have 1620 members in common with our final selection (76 per cent). The comparison of the final values is shown in Table 2.

Comparing our membership with the radial velocity study by Bailey et al. (2018), from high-resolution spectroscopy of 126 main sequence stars in the core 30 arcmin of NGC 2516, we have 43 stars in common and a 93 per cent agreement in membership classification. Bailey et al. give $v_{\rm rad} = 24.50 \pm 0.12$ km s^{-1} from 81 member stars.

4.3 NGC 1817 and Juchert 23

In spite of the work by Balaguer-Núñez et al. (2004a, 2004b) concerning the open cluster region NGC 1817, where only one very extended cluster was found and the asterism NGC 1807 was confirmed as a non-physical cluster, subsequent general catalogues have included NGC 1807 (Kharchenko et al. 2016; Krone-Martins et al. 2010) and even given physical parameters (age, extinction, etc.). The precision and completeness of the *Gaia* data drive us to study this area again.

We have studied an area of 2 deg radius around NGC 1817 with $G < 18$ (102 313 stars) to check if any sign of NGC 1807 is present. NGC 1817 is far, located at around 1700 pc, and because of its small proper motion it is more difficult to disentangle from the field than the previous examples of NGC 2682 and NGC 2516. `Clusterix` finds 2898 members in the area. For a cleaner list of members, we cut around the overdensity in parallax ($0.4 < \varpi < 0.7$ mas), which makes the cluster very conspicuous in the colour–magnitude diagram. From the 1000 resulting members in that selection (see Fig. 10), there are 43 stars with coherent *Gaia* radial velocities, giving an average of $v_{\rm rad} = 66.3 \pm 1.9$ km s^{-1}.

The proper motion and parallax of the cluster are $\mu_{\alpha*} = 0.48 \pm 0.19$ mas yr^{-1}; $\mu_\delta = -0.89 \pm 0.17$ mas yr^{-1}; $\varpi = 0.54 \pm 0.07$ mas for 1000 member stars. The average parallax from the 43 radial velocity members is $\varpi = 0.54 \pm 0.04$ mas.

Comparing this result with the member stars found in the study in the 1 deg central area with a magnitude limit of $G = 18$ by Cantat-Gaudin et al. (2018b), we find 447 member stars in common (97 per cent agreement). From the 169 astrometric members found by Balaguer-Núñez et al. (2004a) with $V < 14.5$ in a 1.5 deg $\times$1.5 deg area, we have 92 members in common (54 per cent agreement), which shows the superior quality of the *Gaia* data. A comparison of the final values is shown in Table 3.

After many tests in the area of NGC 1817, we were not able to find any hint of NGC 1807 as a physical cluster. However, we can confirm Juchert 23 (DSH J0507.6 + 1734), a very small, little known cluster (see full area in Fig. 11 and the VPD of both in Fig. 12). It is not in the lists of Dias et al. (2002), Kharchenko et al. (2016) or Cantat-Gaudin et al. (2018b) and it only appears in the list of suspected open cluster candidates from 2MASS and DSS in an additional table 2e only in electronic form from Kronberger et al. (2006). They only give coordinates and a visual diameter of 7 arcmin. We study an area of one degree radius around its centre with $G < 18$ (15 145 stars). `Clusterix` 2.0. gives us an estimated number of members of 115. With 105 members after a sigma cut in parallax, we get a clean colour–magnitude diagram (Fig. 13).

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

The proper motion and parallax of the cluster are $\mu_{\alpha*} = 0.34 \pm 0.09$ mas yr^{-1}; $\mu_\delta = -2.56 \pm 0.09$ mas yr^{-1}; $\varpi = 0.35 \pm 0.09$ mas for 105 member stars. There are six stars with coherent *Gaia* radial velocity ($v_{\rm rad} = 0.5 \pm 2.1$ km s^{-1}) that produce an average parallax of $\varpi = 0.34 \pm 0.04$ mas.

4.4 Overlapping clusters: NGC 1750 and NGC 1758

NGC 1750 and NGC 1758 are a pair of partly overlapping open clusters at around 700 pc with $E(B-V) = 0.34$ (Straižys et al. 2003; Galadi-Enriquez et al. 1998b,c; Galadí-Enríquez et al. 1998a; Tian et al. 1998).

In an area of 2 deg around NGC 1750 (111 799 stars with $G < 19$, see Fig. 14), Clusterix finds two clusters with 2989 expected members in total. With these results, we plot the proper-motion diagram and we can clearly separate the two populations: NGC 1750 with 1827 candidate members and NGC 1758 with 1162 (see Fig. 15). By simply introducing a further cut in parallaxes, one can easily clean out the colour–magnitude diagrams (see Figs 16 and 17). The average proper motion of the clusters is $\mu_{\alpha*} = -0.97 \pm 0.23$ mas yr^{-1}; $\mu_\delta = -2.40 \pm 0.21$ mas yr^{-1}; $\varpi = 1.34 \pm 0.16$ mas for 504 member stars in NGC 1750. There are only four stars with coherent *Gaia* radial velocities ($v_{\rm rad} = -11.5 \pm 0.7$ km s^{-1}). Regarding NGC 1758, we obtain $\mu_{\alpha*} = 3.13 \pm 0.23$ mas yr^{-1}; $\mu_\delta = -3.45 \pm 0.20$ mas yr^{-1}; $\varpi = 1.08 \pm 0.15$ mas for 363 member stars. There are only three stars with coherent *Gaia* radial velocities ($v_{\rm rad} = 7.2 \pm 3.4$ km s^{-1}).

Comparing this result with the member stars found in the study in the 1.5 deg area with a magnitude limit of $G = 18$ from Cantat-Gaudin et al. (2018b), we find 378 member stars in common for NGC 1750 (86 per cent agreement) and 146 member stars in common for NGC 1758 (99 per cent agreement). A comparison of the final values is given in Table 4.

4.5 Ruprecht 26 and Clusterix 1

In order to check areas with multiple clusters, we have also evaluated an area of one degree around Ruprecht 26, where many other known clusters exist (NGC 2428, NGC 2425, NGC 2414, Ruprecht 151 and Alessi 17). Unexpectedly, we have found in the central area of Ruprecht 26 two kinematically distinct populations, besides the field one, as can be seen in Figs 18 and 19. One of these populations corresponds to Ruprecht 26 and the other does not correspond to any of the previously mentioned known clusters.

In cases like this, where the field is full of other clusters, it is difficult to find a patch of sky mostly free of clusters. Nevertheless, after different trials we have managed to find a reasonable trade-off. You can see the parameters used in the Appendix. In this case, due to the low constrast of the populations in the spatial plane, the expected number of cluster members is not reliable and we need to apply other criteria (photometry, parallax, spatial distribution, etc.) to clean the membership from proper motions. We accept as tentative members for both clusters all stars with membership probabilities higher than 0.5 (1569 candidate stars).

The two populations found are superposed in space; while one population with 1238 stars (see Fig. 20) is very well concentrated, the other with 331 stars (Ruprecht 26) is more disperse and poorer. The differentiated population (which we called Clusterix 1 from now on) has only been discovered thanks to the exquisite quality of the *Gaia* DR2 proper motions. To confirm their separate existence, we study their colour–magnitude diagrams and parallax distribution.

Ruprecht 26 candidate members show a distinct parallax, so, applying a cut selection ($0.8 < \varpi < 1.1$ mas), we can see a clear main sequence (see Fig. 21). From this final selection of 109 members, we find an average proper motion of $\mu_{\alpha*} = -3.20 \pm 0.12$ mas yr^{-1}; $\mu_\delta = 0.11 \pm 0.10$ mas yr^{-1}; $\varpi = 0.92 \pm 0.06$ mas. There are only four stars with coherent *Gaia* radial velocities, $v_{\rm rad} = 44.8 \pm 4.8$ km s^{-1}.

In the case of Clusterix 1, its parallax is small and it is difficult to separate from the field, so we do not use parallaxes and just select a core area ($r < 0.2$ deg) that shows a cleaner distribution in the colour–magnitude diagram (see Fig. 22). With that selection, we found an average proper motion of $\mu_{\alpha*} = -1.20 \pm 0.16$ mas yr^{-1}; $\mu_\delta = 2.80 \pm 0.13$ mas yr^{-1} from 157 members and a parallax of 0.41 ± 0.15 mas. With these parameters, we find six stars with coherent *Gaia* radial velocities in the whole area (three of them inside the core area), giving $v_{\rm rad} = 37.3 \pm 5.2$ km s^{-1}.

Comparing the Ruprecht 26 result with the member stars found in the study in the 1 deg area with a magnitude limit of $G = 18$ by Cantat-Gaudin et al. (2018b), we find 69 member stars in common (85 per cent agreement). The comparison of the final values is given in Table 5.

5 CONCLUSIONS

Many automatic tools have been developed in recent years to separate cluster members from the field. We present here an open-access web tool, VO compliant, to facilitate membership studies from proper-motion data for any user that requires a tailor-made study on any specific data set. Different tools are able to apply to different cases and all of them have been able to discover new clusters. The census of open clusters seems far from complete.

We present the first results of Clusterix 2.0 for different cases to show the capabilities of the tool, its flexibility and its adaptability to different environments: an area of five degrees around NGC 2682 (M67), an old, well-known cluster; a young close-by cluster NGC 2516 with a striking elongated structure extended up to four degrees that deserves further study; NGC 1750 and NGC 1758, a pair of partly overlapping clusters found without applying any a priori knowledge; the area of NGC 1817, where we confirm the non-existence of NGC 1807 and the existence of a little known cluster, Juchert 23; and an area with many neighbouring clusters, where we are able to disentangle the existence of two clusters where only one was previously known: Ruprecht 26 and the new Clusterix 1.

ACKNOWLEDGEMENTS

The authors acknowledge the anonymous referee for very useful comments. LBN acknowledges P.L. White for continuous support. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This work was partly supported by the Spanish State Research Agency (AEI) projects ESP2017-87676-C5-1-R, ESP2016-80079-C2-1-R and RTI2018-095076-B-C21 (MINECO/FEDER, UE), and MDM-2014-0369 of ICCUB and MDM-2017-0737 Centro de Astrobiología (CSIC-INTA) (Unidades de Excelencia 'María de Maeztu'), as well as the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement GENIUS FP7–606740 and Horizon H2020 Framework Programme under

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

grant agreement ASTERICS–653477. This publication makes use of VOSA, developed under the Spanish Virtual Observatory project supported by the Spanish MINECO through grant AyA2017-84089. This work has made extensive use of TOPCAT (Taylor 2005) and of NASA's Astrophysics Data System. This research has made use of the WEBDA database, operated at the Department of Theoretical Physics and Astrophysics of the Masaryk University.

REFERENCES

Altmann M., Roeser S., Demleitner M., Bastian U., Schilbach E., 2017, A&A, 600, L4
Anders F. et al., 2017, A&A, 600, A70
Arenou F. et al., 2018, A&A, 616, A17
Bailey J. I., Mateo M., White R. J., Shectman S. A., Crane J. D., 2018, MNRAS, 475, 1609
Balaguer-Núñez L., Jordi C., Galadí-Enríquez D., Zhao J. L., 2004a, A&A, 426, 819
Balaguer-Núñez L., Jordi C., Galadí-Enríquez D., Masana E., 2004b, A&A, 426, 827
Balaguer-Núñez L., Jordi C., Galadí-Enríquez D., 2005, A&A, 437, 457
Balaguer-Núñez L., Galadí-Enríquez D., Jordi C., 2007, A&A, 470, 585
Balaguer-Núñez L., Galadí-Enríquez D., López del Fresno M., Solano E., Jordi C., Sézima T., Paunzen E., 2017, in Highlights on Spanish Astrophysics IX. SEA, Bilbao, p. 328
Bayo A., Rodrigo C., Barrado Y Navascués D., Solano E., Gutiérrez R., Morales-Calderón M., Allard F., 2008, A&A, 492, 277
Bossini D. et al., 2019, A&A, 623, A108
Brinkmann N., Banerjee S., Motwani B., Kroupa P., 2017, A&A, 600, A49
Cantat-Gaudin T. et al., 2018a, A&A, 615, A49
Cantat-Gaudin T. et al., 2018b, A&A, 618, A93
Cantat-Gaudin T. et al., 2019, A&A, 624, A126
Cantat-Gaudin T., Donati P., Vallenari A., Sordo R., Bragaglia A., Magrini L., 2016, A&A, 588, A120
Carrera R. et al., 2019a, A&A, 623, A80
Carrera R. et al., 2019b, A&A, 627, A119
Carrera R., Pancino E., 2011, A&A, 535, A30
Casamiquela L. et al., 2017, MNRAS, 470, 4363
Castro-Ginard A., Jordi C., Luri X., Julbe F., Morvan M., Balaguer-Núñez L., Cantat-Gaudin T., 2018, A&A, 618, A59
Castro-Ginard A., Jordi C., Luri X., Cantat-Gaudin T., Balaguer-Núñez L., 2019, A&A, 627, A35
Clarke C. J., Bonnell I. A., Hillenbrand L. A., 2000, in Mannings V., Boss A. P., Russell S. S., eds, Protostars and Planets IV. University of Arizona Press, Tucson, p. 151
Dias W. S., Alessi B. S., Moitinho A., Lépine J. R. D., 2002, A&A, 389, 871
Faherty J. K., Bochanski J. J., Gagné J., Nelson O., Coker K., Smithka I., Desir D., Vasquez C., 2018, ApJ, 863, 91
Frinchaboy P. M., Majewski S. R., 2008, AJ, 136, 118
Gaia Collaboration et al., 2016a, A&A, 595, A1
Gaia Collaboration et al., 2016b, A&A, 595, A2
Gaia Collaboration et al., 2017, A&A, 601, A19
Gaia Collaboration et al., 2018a, A&A, 616, A1
Gaia Collaboration et al., 2018b, A&A, 616, A10
Galadí-Enríquez D., Jordi C., Trullols E., 1998a, Ap&SS, 263, 307
Galadi-Enriquez D., Jordi C., Trullols E., Ribas I., 1998b, A&A, 333, 471
Galadi-Enriquez D., Jordi C., Trullols E., 1998c, A&A, 337, 125
Gieles M., Portegies Zwart S. F., Baumgardt H., Athanassoula E., Lamers H. J. G. L. M., Sipior M., Leenaarts J., 2006, MNRAS, 371, 793
Jacobson H. R. et al., 2016, A&A, 591, A37
Kharchenko N. V., Piskunov A. E., Schilbach E., Röser S., Scholz R. D., 2016, A&A, 585, A101
Kronberger M. et al., 2006, A&A, 447, 921
Krone-Martins A., Moitinho A., 2014, A&A, 561, A57
Krone-Martins A., Moitinho A., 2015, Astrophysics Source Code Library, record ascl: 1504.001
Krone-Martins A., Soubiran C., Ducourant C., Teixeira R., Le Campion J. F., 2010, A&A, 516, A3
Lindegren L. et al., 2018, A&A, 616, A2
Magrini L., Sestito P., Randich S., Galli D., 2009, A&A, 494, 95
Netopil M., Paunzen E., Stütz C., 2012, Astrophysics and Space Science Proceedings, 29, 53
Netopil M., Paunzen E., Heiter U., Soubiran C., 2016, A&A, 585, A150
Olivares J. et al., 2018, A&A, 617, A15
Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, ARA&A, 48, 431
Prisinzano L., Magrini L. et al., 2018, White Paper for the cadence optimization call of LSST, preprint (arXiv:1812.03025)
Quillen A. C., Nolting E., Minchev I., De Silva G., Chiappini C., 2018, MNRAS, 475, 4450
Roeser S., Demleitner M., Schilbach E., 2010, AJ, 139, 2440
Sampedro L., Dias W. S., Alfaro E. J., Monteiro H., Molino A., 2017, MNRAS, 470, 3937
Sanders W. L., 1971, A&A, 14, 226
Sarro L. M. et al., 2014, A&A, 563, A45
Sezima T., Galadí-Enríquez D., Paunzen E., Jordi C., Balaguer-Núñez L., Jilkova L., 2015, in Highlights of Spanish Astrophysics VIII. SAE, Teruel, p. 594
Silverman B. W., 1986, Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability. Chapman and Hall, London
Soubiran C. et al., 2018, A&A, 619, A155
Straižys V. et al., 2003, Baltic Astron., 12, 323
Taylor M. B., 2005, ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac., San Francisco, p. 29
Tian K. P., Zhao J. L., Shao Z. Y., Stetson P. B., 1998, A&AS, 131, 89
Wu Z.-Y., Zhou X., Ma J., Du C.-H., 2009, MNRAS, 399, 2146
Zacharias N., Finch C. T., Girard T. M., Henden A., Bartlett J. L., Monet D. G., Zacharias M. I., 2013, AJ, 145, 44
Zacharias N., Finch C., Frouard J., 2017, AJ, 153, 166
Zhao J. L., He Y. P., 1990, A&A, 237, 54

APPENDIX A: `CLUSTERIX` PROCEDURE FOR OUR SCIENCE CASES

We present here as case examples some of the choices made in `Clusterix` 2.0 to produce the results explained before.

For NGC 2516, see Figs A1, A2, A3 and A4.
For NGC 1817, see Figs A5, A6, A7 and A8.
For Juchert 23, see Figs A9, A10, A11 and A12.
For NGC 1750 and NGC 1758, see Figs A13, A14, A15 and A16.
For Ruprect 26 and CLUSTERIX 1, see Figs A17, A18, A19 and A20.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

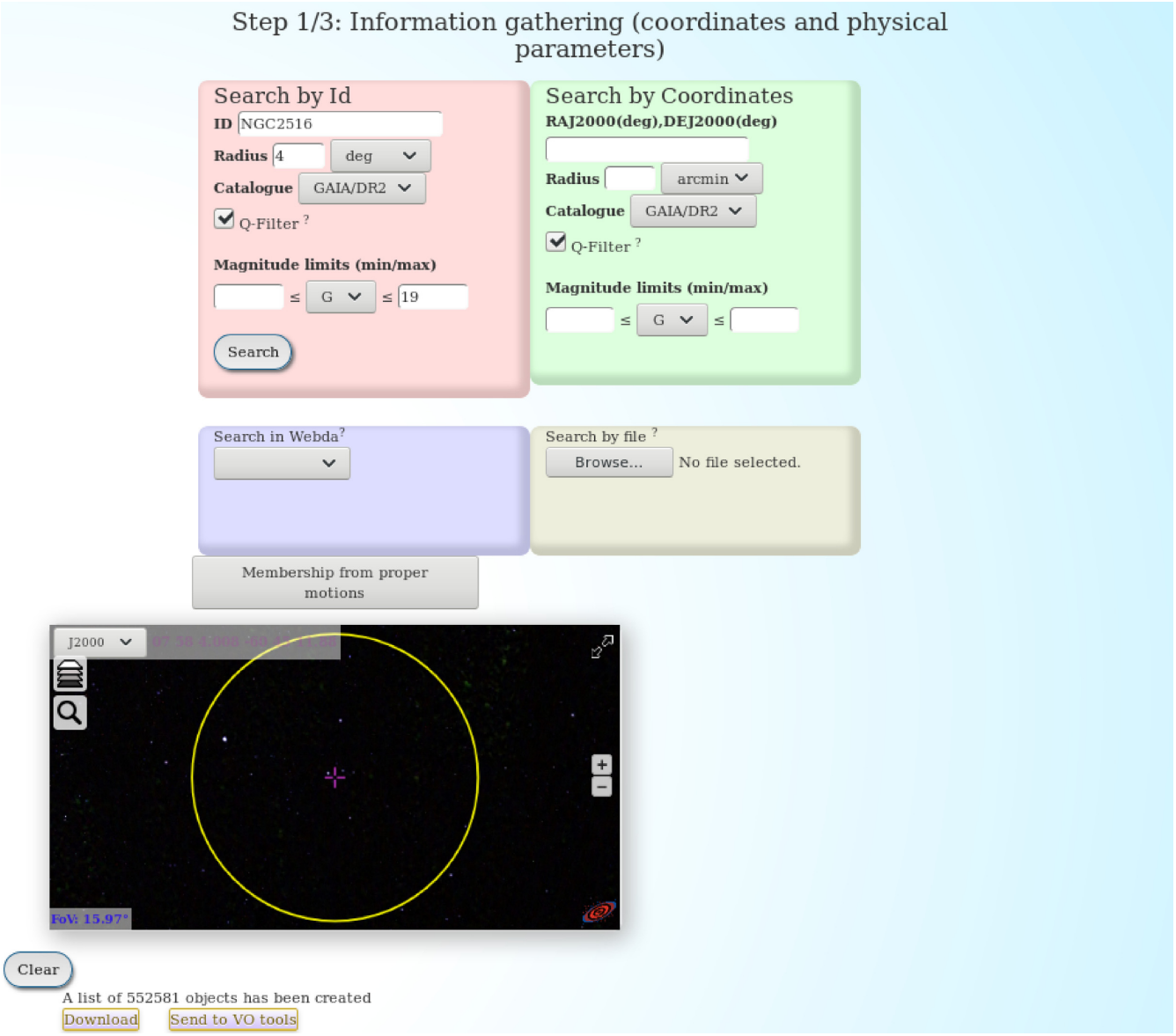


Figure A1. Step 1 of the Clusterix procedure for NGC 2516.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

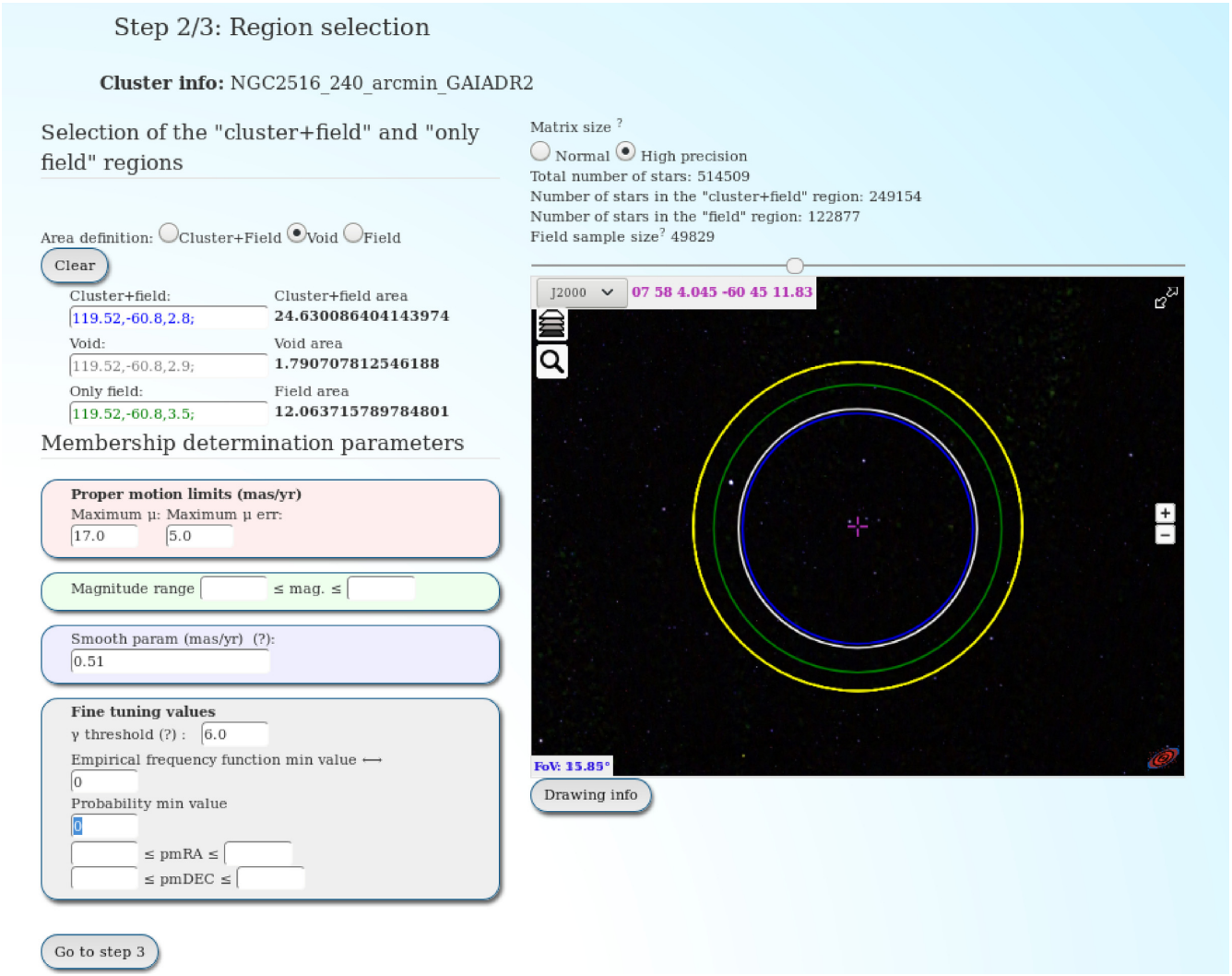


Figure A2. Step 2, in the case of NGC 2516 we have also limited the number of stars in the field to around 50 thousand. High precision is used here for the final results.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

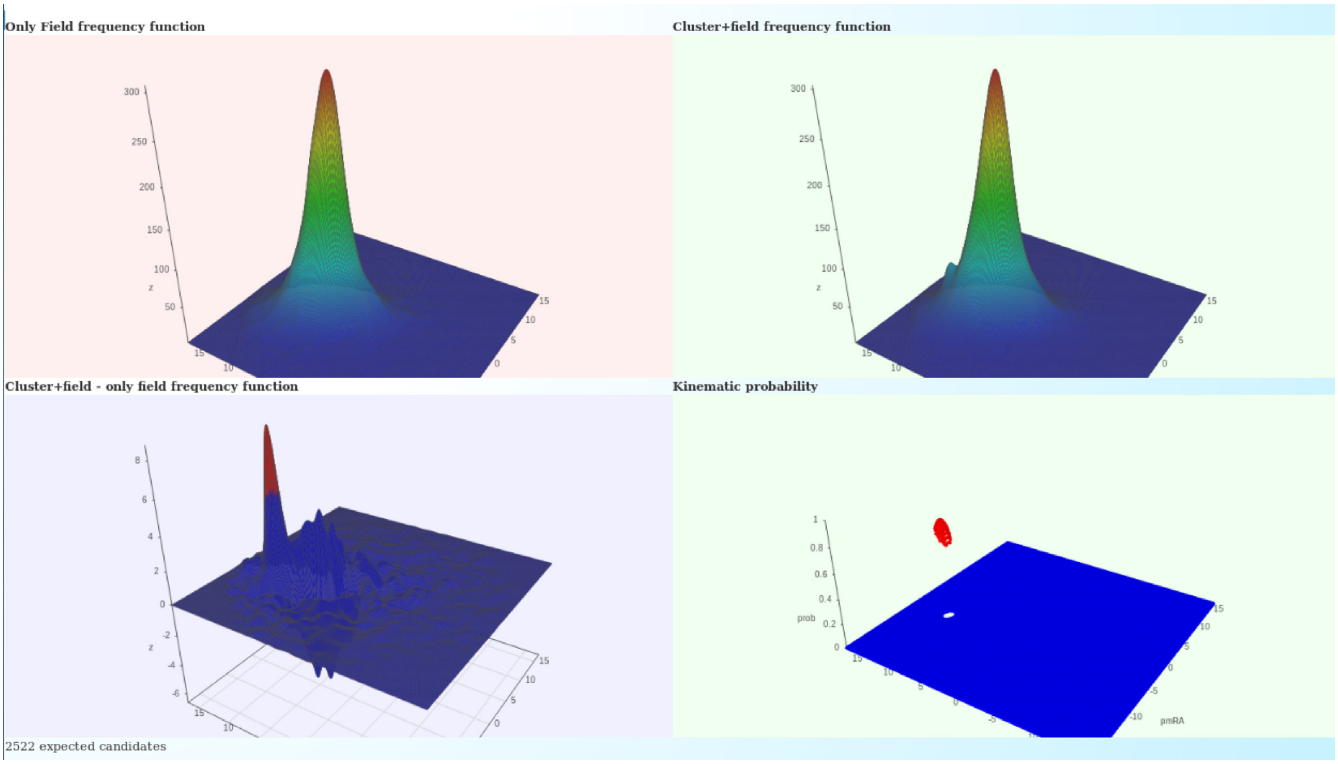


Figure A3. Interactive results of Step 2 of the Clusterix procedure for NGC 2516.

Step 3/3. Determination of membership probabilities

Send to VO tools | Save as text | Gather photometry for cluster members (in VOSA format)

```
#   Results were retrieved using Clusterix software
#   http://clusterix.cab.inta-csic.es/
#   In case of problems, please, report to: clusterix_archive_support@cab.inta-csic.es
#
#   Labels:
#
#       STAR_NO identificator of star retrieved from the input data
#       RA      right ascension of a star
#       DEC     declination of a star
#       PM_RA   proper motion in alpha
#       PM_DEC  proper motion in delta
#       PROB    probability that star belongs to evaluated open cluster
#       FLAG    M=Membership   NM=No membership
#
#   Parameters:
#
#       CLUSTER INFO: NGC2516_240_arcmin_GAIADR2
#       PROPER MOTION CUTOFF: 17.00 mas/yr
#       PROPER MOTION ERR CUTOFF: 5.00 mas/yr
#       SMOOTH PARAMETER: 0.51
#       GAMMA FACTOR: 6.00
#       QUERY MIN MAG (STEP 1):
#       QUERY MAX MAG (STEP 1): 19
#       FILTERED MIN MAG (STEP 2):
#       FILTERED MAX MAG (STEP 2):
#
#       Boundaries can be circles (Ra(J2000)-center,Dec(J2000)-center, radius degrees) or
#       polygons (Ra,Dec Ra,Dec Ra,Dec ...). Each boundary is separated by ;
#
#       CLUSTER BOUNDARIES:
#       119.52,-60.8,2.8;
#
#       FIELD BOUNDARIES:
#       119.52,-60.8,3.5;
#
#       VOID BOUNDARIES:
#       119.52,-60.8,2.9;
#
#       NUM STARS: 552560
#
#       EXPECTED NUMBER OF MEMBERS: 2522
#
#       For Webda clusters, STAR_NO is the merge of the Data source reference and the star number: Ref_Star
#
#       Region values: F = star was inside a "only field" region
#                      C = star was inside a "cluster+field" region
#                      V = star was outside the selected regions
```

Highest probability results sample (only 50 first shown).

##STAR_NO	RAJ2000	DECJ2000	pmRA	epmRA	pmDEC	epmDEC	Pll	ePll	BP	eBP	RP	eRP	G	eG	RV	eRV	PROB	Region
311985	123.137545	-58.364635	-4.579	0.169	11.46	0.164	0.6116	0.0861	18.1376	0.0111	16.7461	0.0046	17.5075	0.001	NaN	NaN	0.8024464	C
343362	111.334290	-61.146801	-4.604	0.271	11.392	0.306	0.3724	0.1262	18.7873	0.0229	17.3952	0.0124	18.1415	0.0016	NaN	NaN	0.8024464	V

Figure A4. Step 3 of the Clusterix procedure for NGC 2516. In the file we can see that the value of expected members is 252; based on that value, we looked for the probability value corresponding to the 2522nd star and found a value of 0.76047766, so we accept as candidate members all stars with equal or greater probability, which gives us a value of 2537 candidate members.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

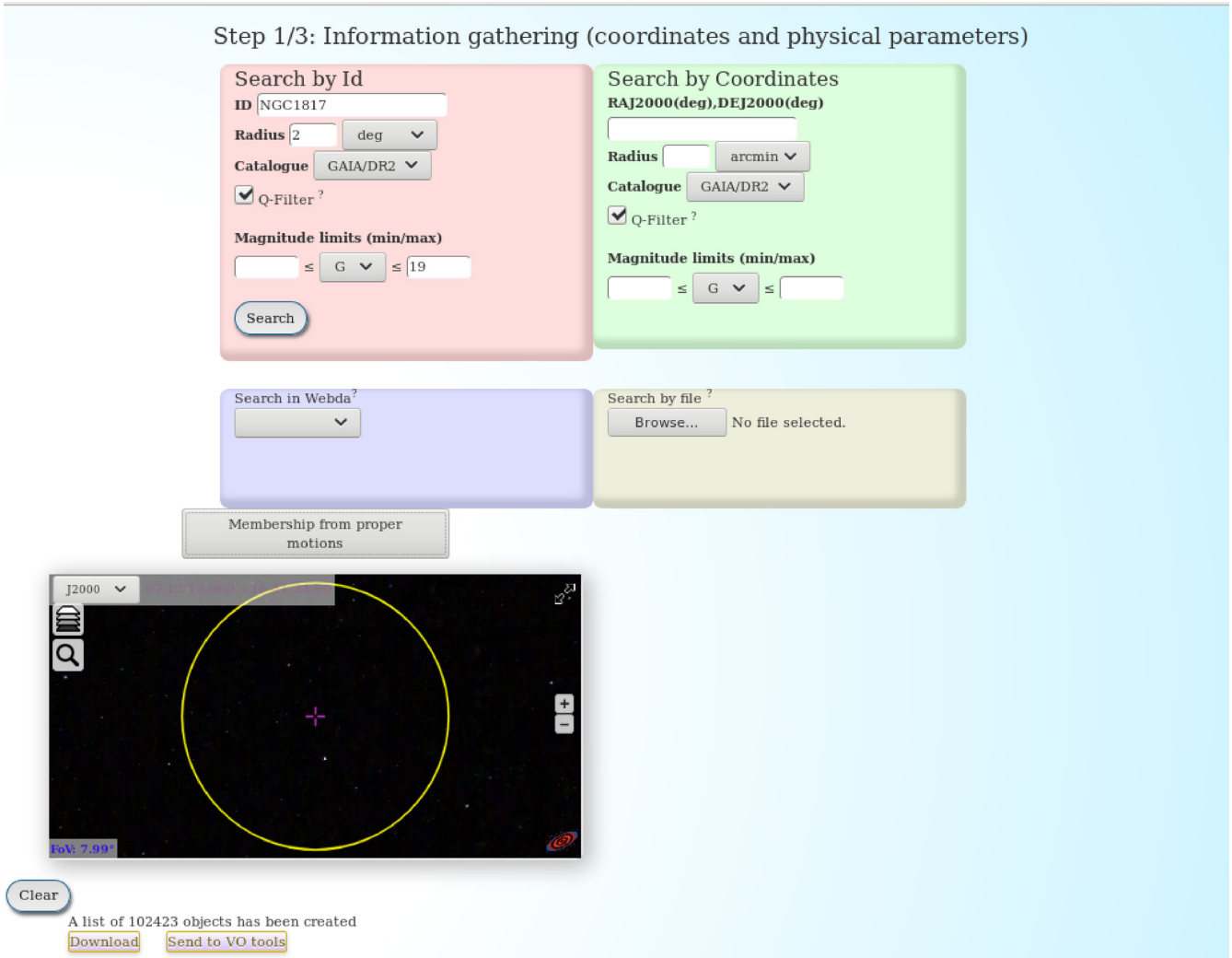


Figure A5. Step 1 of the `Clusterix` procedure for NGC 1817.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

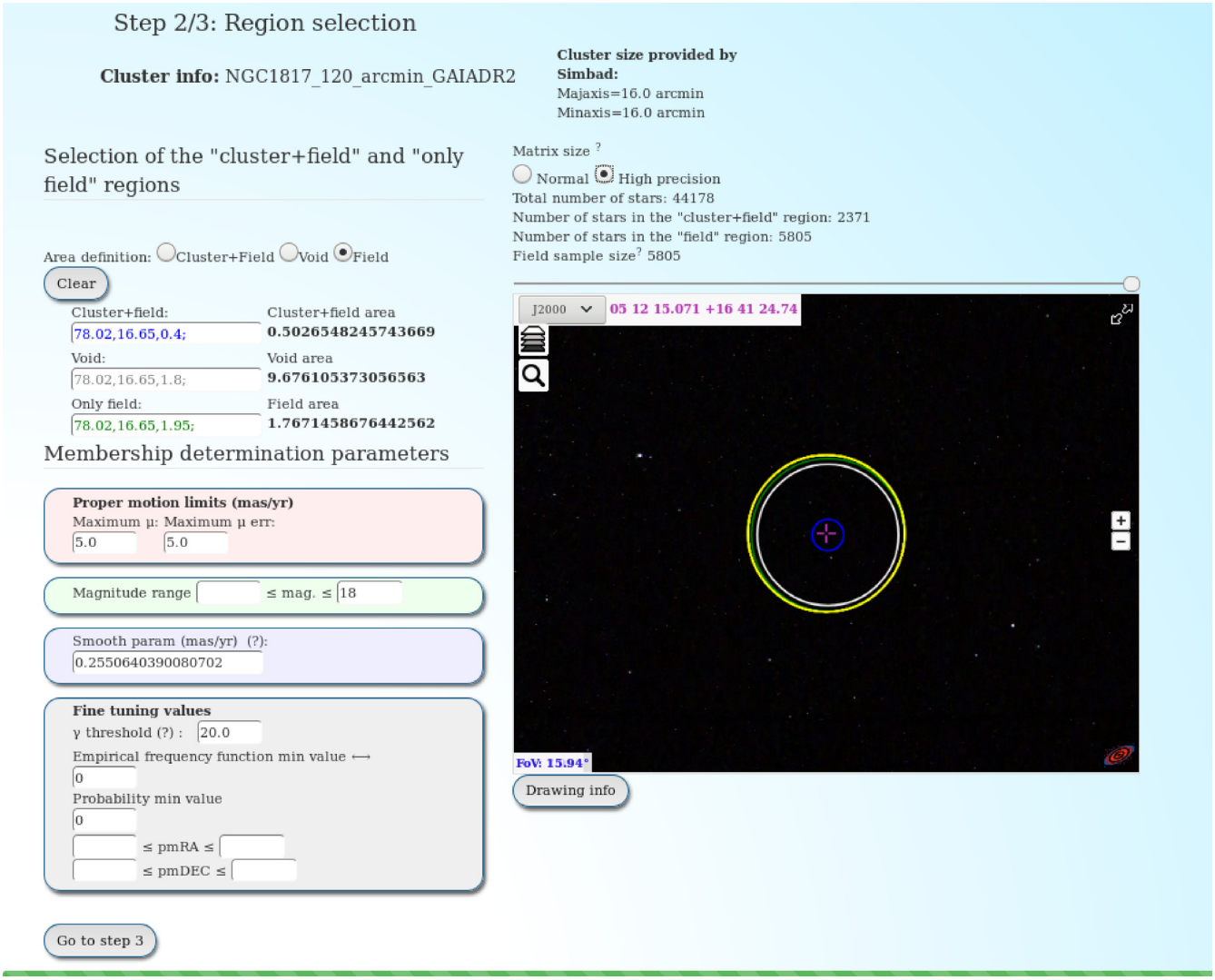


Figure A6. Step 2, where all the fine-tuning parameters of the `Clusterix` procedure are set for NGC 1817.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

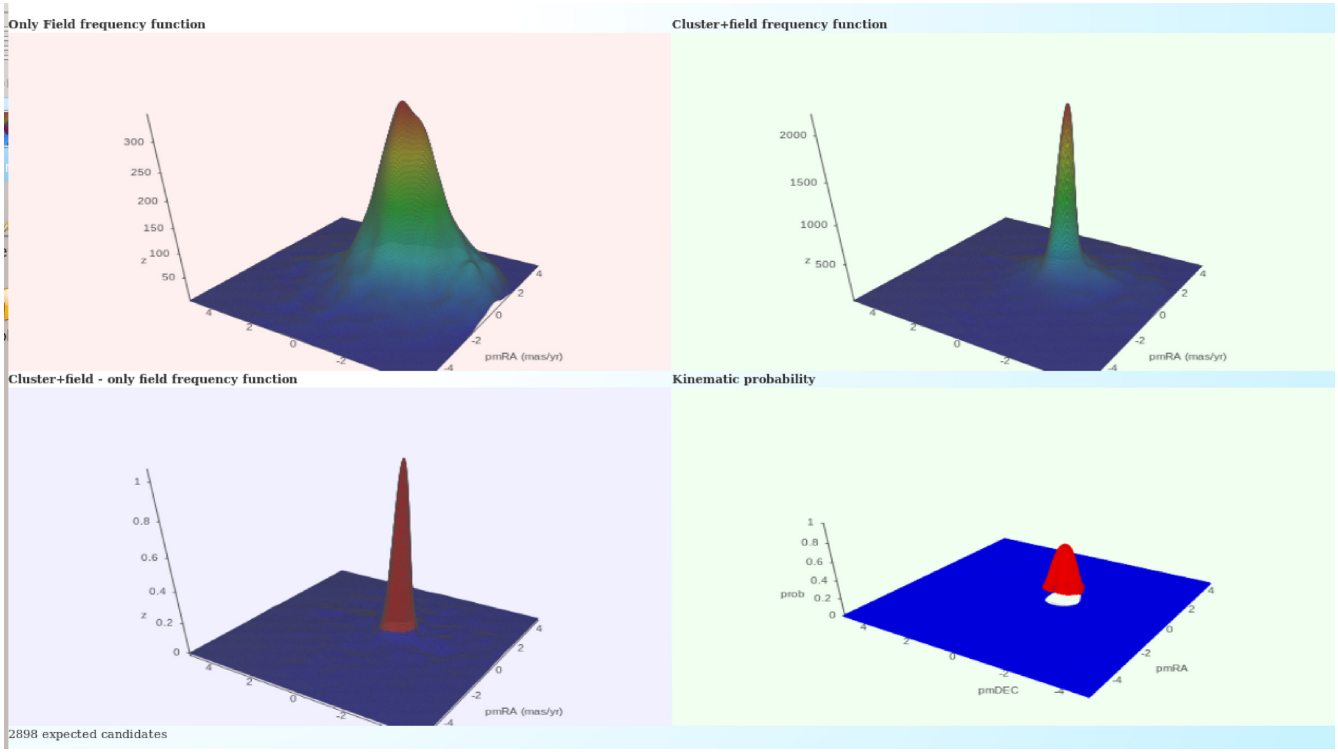


Figure A7. Interactive results of Step 2 of the `Clusterix` procedure for NGC 1817.



Figure A8. Step 3 of the `Clusterix` procedure for NGC 1817.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

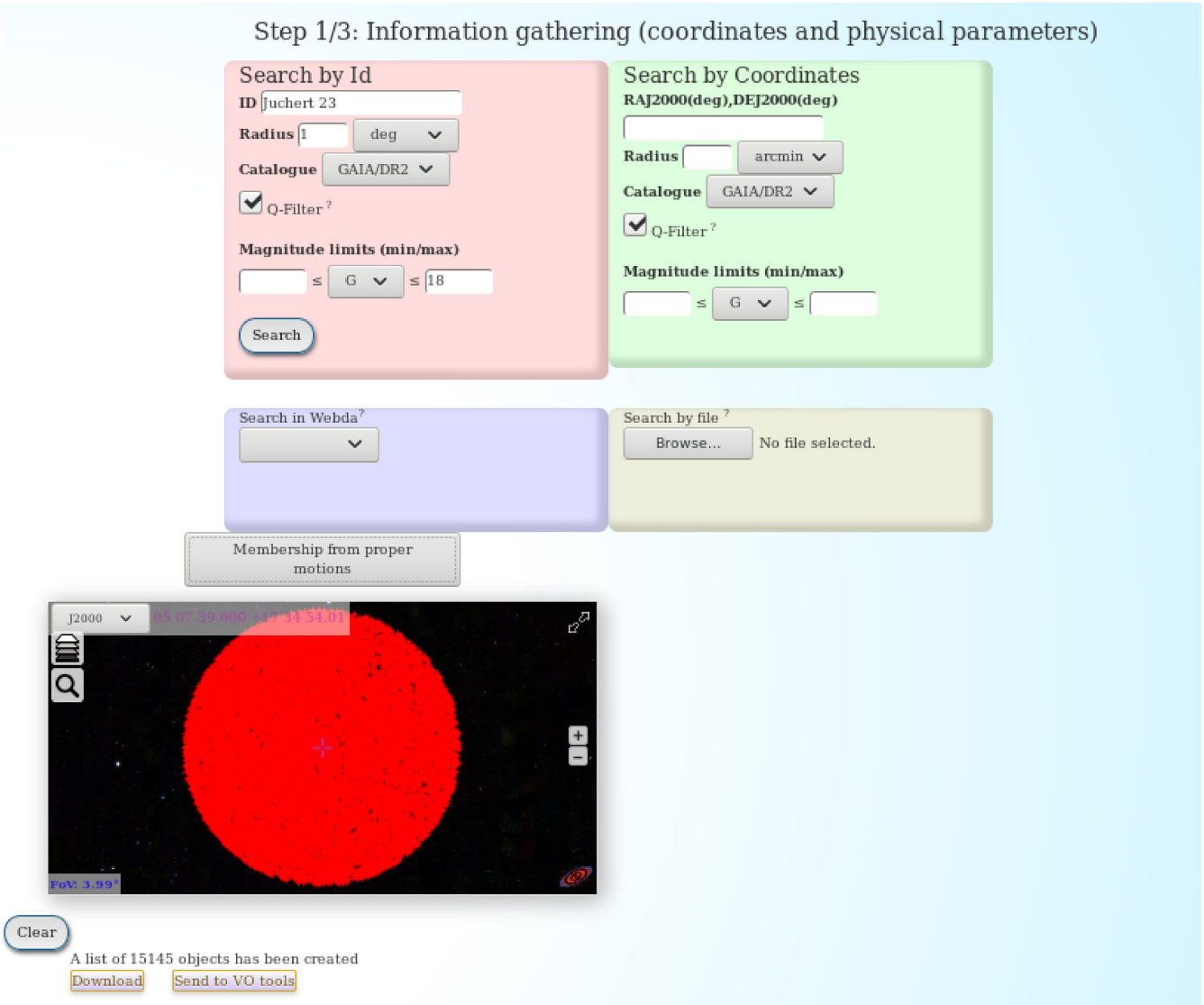


Figure A9. Step 1 of the Clusterix procedure for Juchert 23. We chose the database, the radius and the magnitude limit and can download the results of this search.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

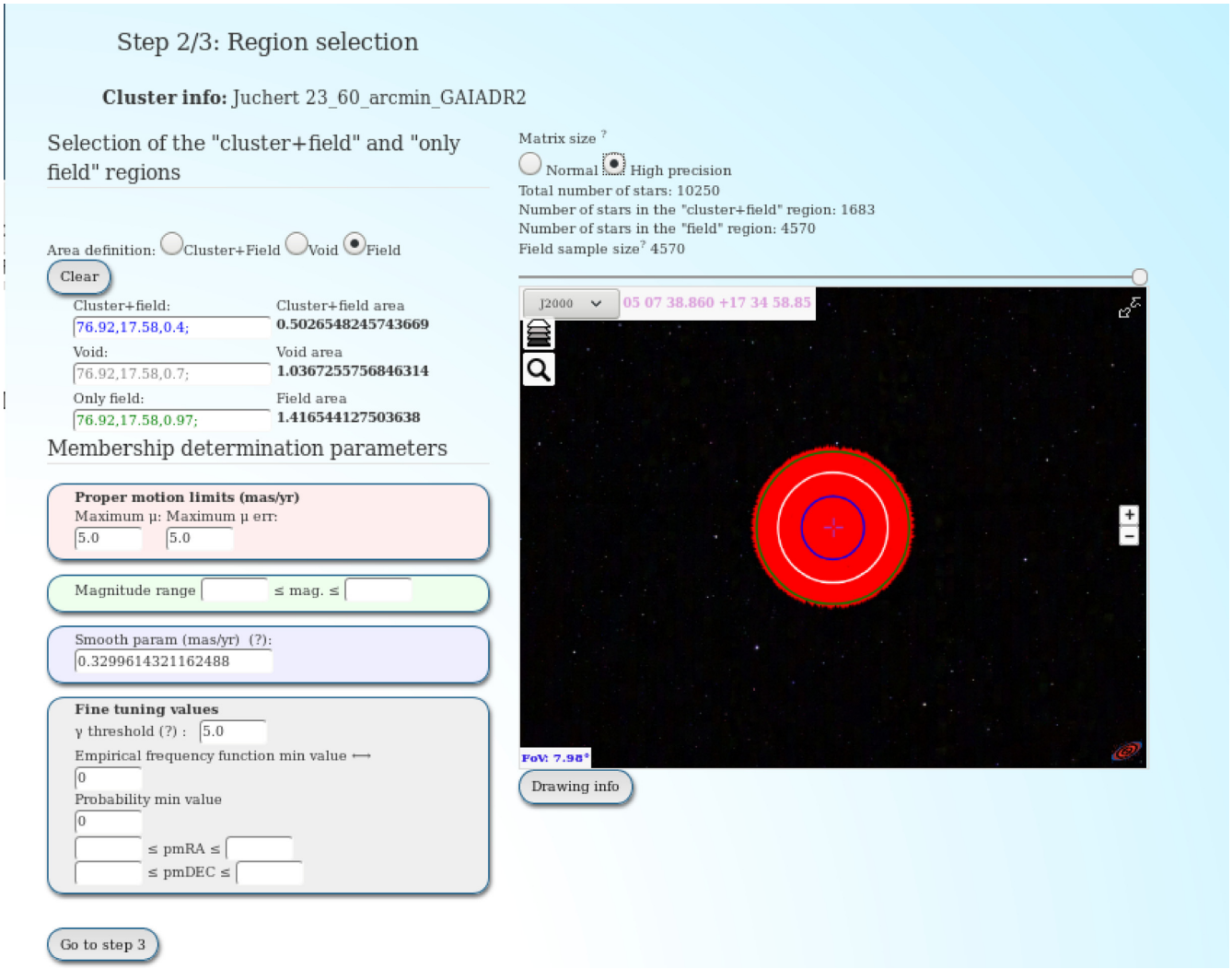


Figure A10. Step 2, where all the fine-tuning parameters of the `Clusterix` procedure are set for Juchert 23.

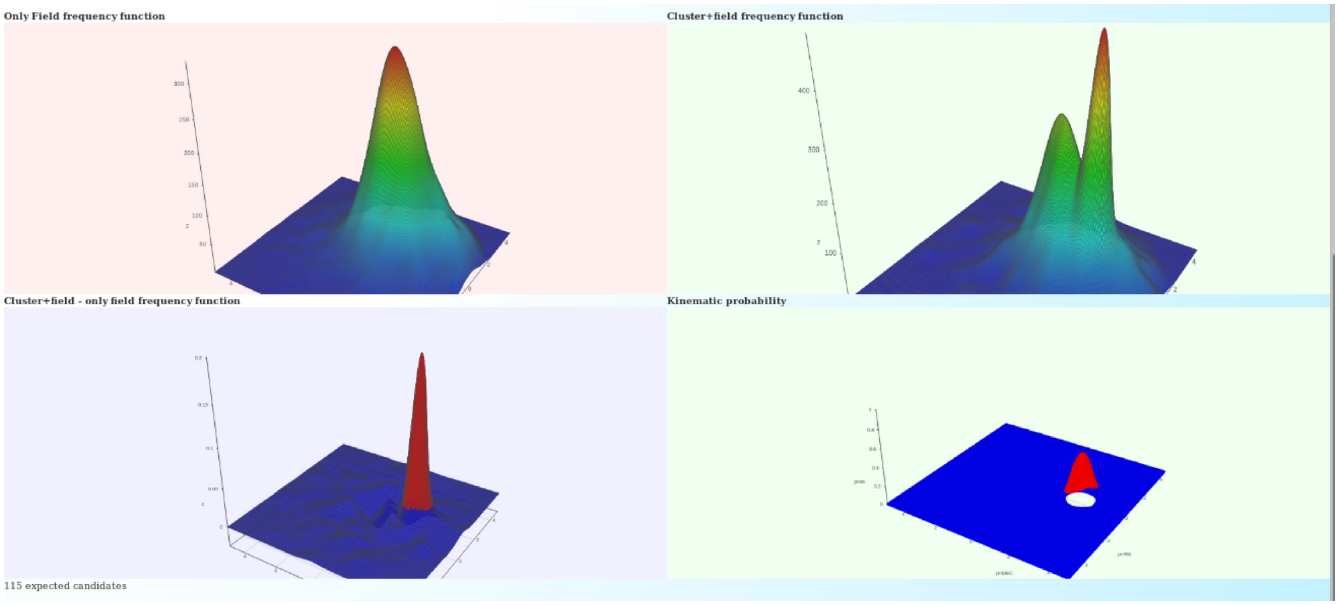


Figure A11. Interactive results of Step 2 of the `Clusterix` procedure for Juchert 23.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026



Step 3/3. Determination of membership probabilities

Send to VO tools | Save as text | Gather photometry for cluster members (in VOSA format)

```
#   Results were retrieved using Clusterix software
#   http://clusterix.cab.inta-csic.es/
#   In case of problems, please, report to: clusterix_archive_support@cab.inta-csic.es
#
#   Labels:
#
#       STAR_NO identificator of star retrieved from the input data
#       RA      right ascension of a star
#       DEC     declination of a star
#       PM_RA   proper motion in alpha
#       PM_DEC  proper motion in delta
#       PROB    probability that star belongs to evaluated open cluster
#       FLAG    M=Membership   NM=No membership
#
#   Parameters:
#
#       CLUSTER INFO: Juchert 23_60_arcmin_GAIADR2
#       PROPER MOTION CUTOFF: 5.00 mas/yr
#       PROPER MOTION ERR CUTOFF: 5.00 mas/yr
#       SMOOTH PARAMETER: 0.33
#       GAMMA FACTOR: 5.00
#       QUERY MIN MAG (STEP 1):
#       QUERY MAX MAG (STEP 1): 18
#       FILTERED MIN MAG (STEP 2):
#       FILTERED MAX MAG (STEP 2):
#
#       Boundaries can be circles (Ra(J2000)-center,Dec(J2000)-center, radius degrees) or
#       polygons (Ra,Dec Ra,Dec Ra,Dec ...). Each boundary is separated by ;
#
#       CLUSTER BOUNDARIES:
#       76.92,17.58,0.4;
#
#       FIELD BOUNDARIES:
#       76.92,17.58,0.97;
#
#       VOID BOUNDARIES:
#       76.92,17.58,0.7;
#
#       NUM STARS: 15145
#
#       EXPECTED NUMBER OF MEMBERS: 115
#
#       For Webda clusters, STAR_NO is the merge of the Data source reference and the star number: Ref_Star
#
#       Region values: F = star was inside a "only field" region
#                      C = star was inside a "cluster+field" region
#                      V = star was outside the selected regions
```

Highest probability results sample (only 50 first shown).

##STAR_NO	RAJ2000	DECJ2000	pmRA	epmRA	pmDEC	epmDEC	Pll	ePll	BP	eBP	RP	eRP	G	eG	RV	eRV	PROB	Region
1701	76.539691	16.757669	0.308	0.055	-2.613	0.04	0.2286	0.0352	15.0508	0.002	13.5464	0.0013	14.3474	4.0E-4	NaN	NaN	0.51018196	V

Figure A12. Step 3 of the Clusterix procedure for Juchert 23.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

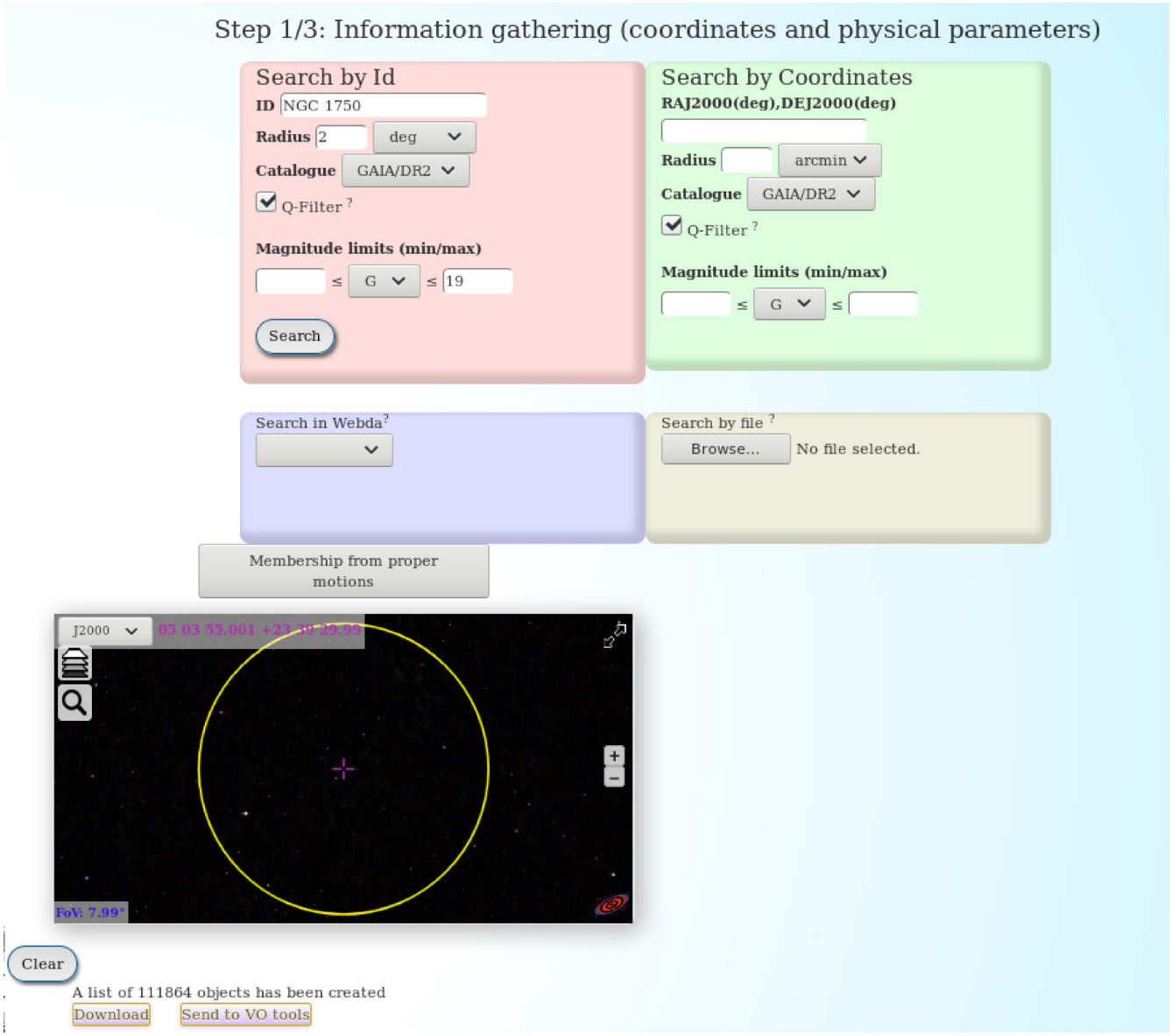


Figure A13. Step 1 of the `Clusterix` procedure for NGC 1750 and NGC 1758.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

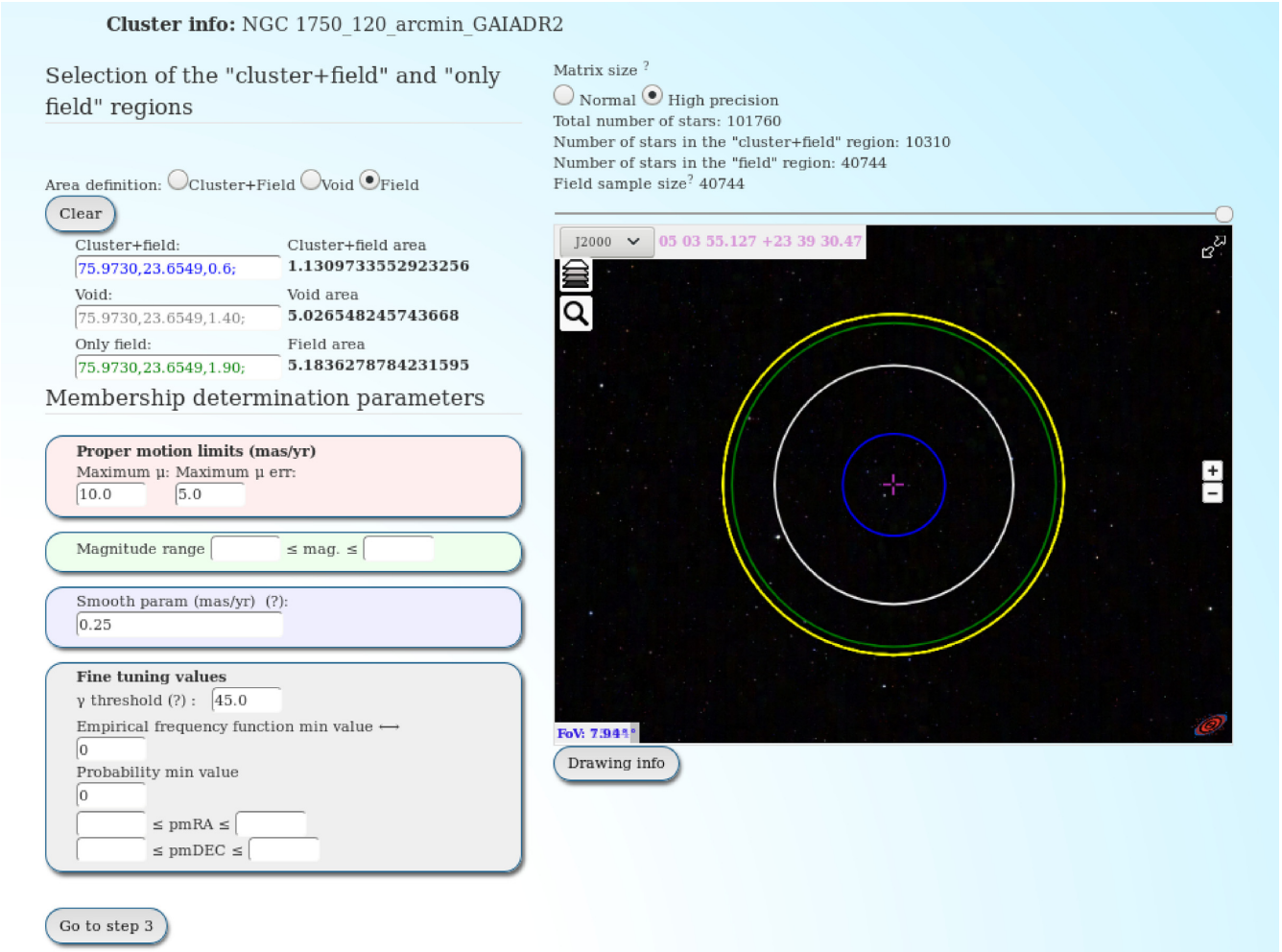


Figure A14. Step 2 where all the fine-tuning parameters of the Clusterix procedure are set for NGC 1750 and NGC 1758.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

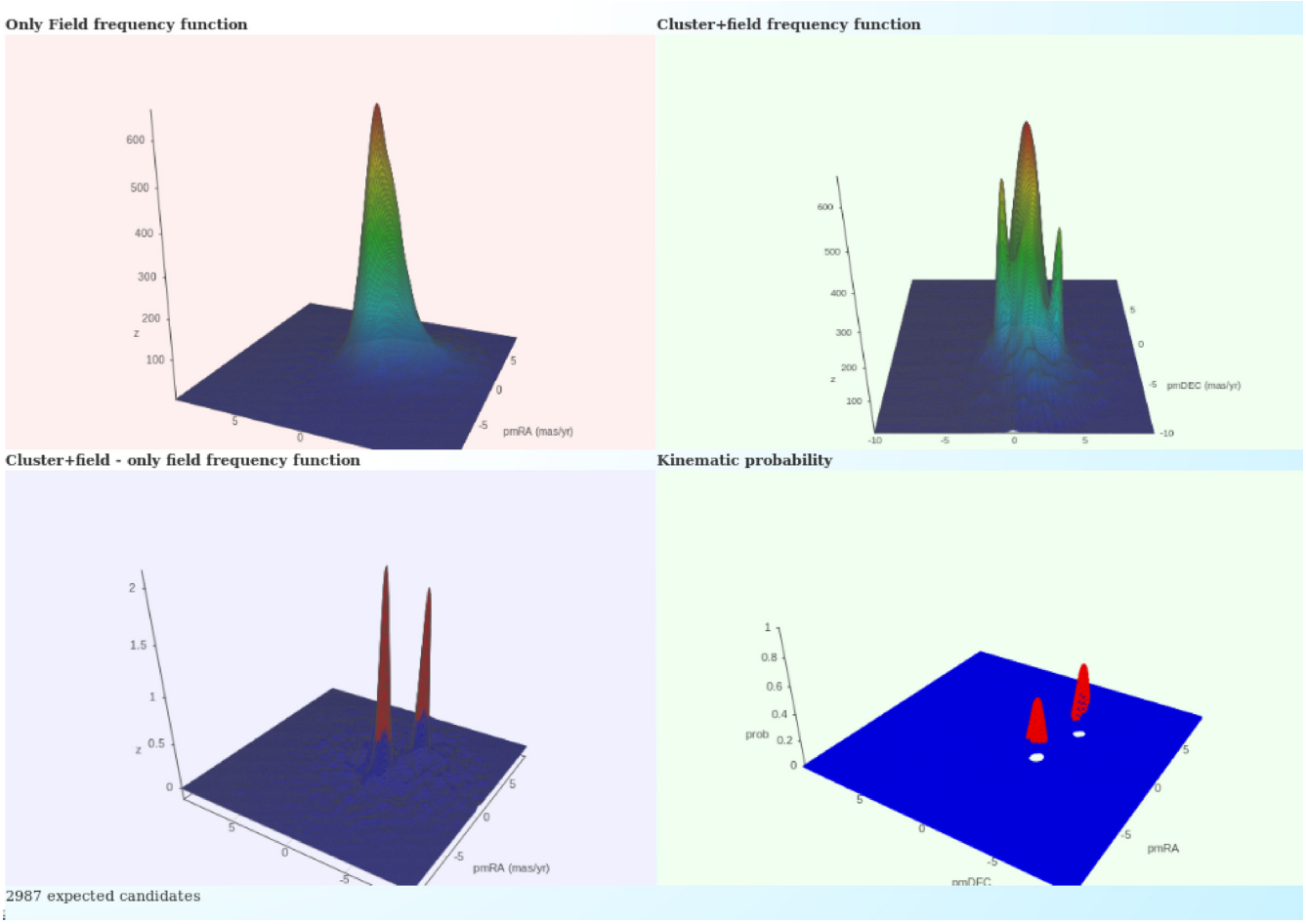
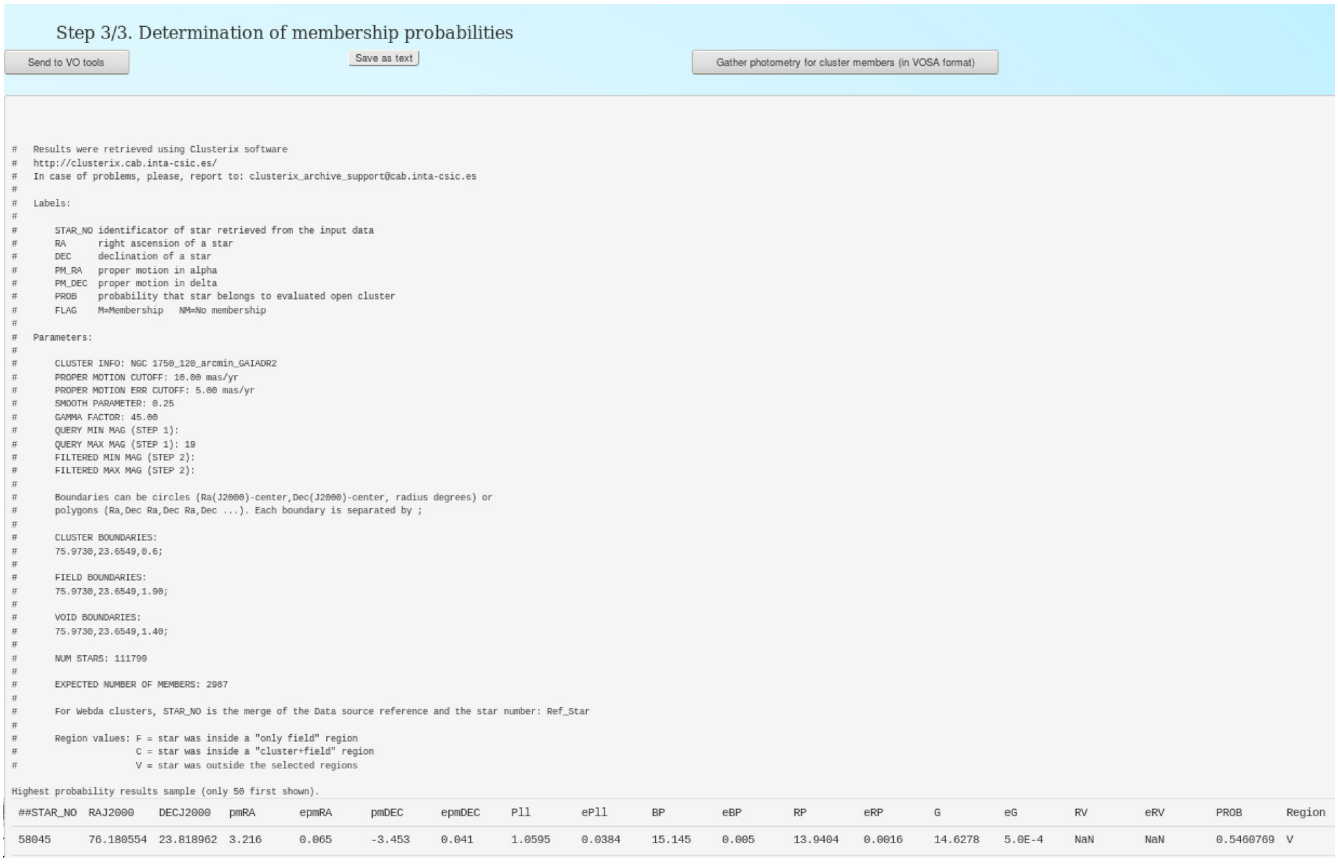


Figure A15. Interactive results of Step 2 of the `Clusterix` procedure for NGC 1750 and NGC 1758.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026



Step 3/3. Determination of membership probabilities

Send to VO tools | Save as text | Gather photometry for cluster members (in VOSA format)

```
#   Results were retrieved using Clusterix software
#   http://clusterix.cab.inta-csic.es/
#   In case of problems, please, report to: clusterix_archive_support@cab.inta-csic.es
#
#   Labels:
#
#       STAR_NO identificator of star retrieved from the input data
#       RA      right ascension of a star
#       DEC     declination of a star
#       PM_RA   proper motion in alpha
#       PM_DEC  proper motion in delta
#       PROB    probability that star belongs to evaluated open cluster
#       FLAG    M=Membership   NM=No membership
#
#   Parameters:
#
#       CLUSTER INFO: NGC 1750_120_arcmin_GAIADR2
#       PROPER MOTION CUTOFF: 10.00 mas/yr
#       PROPER MOTION ERR CUTOFF: 5.00 mas/yr
#       SMOOTH PARAMETER: 0.25
#       GAMMA FACTOR: 45.00
#       QUERY MIN MAG (STEP 1):
#       QUERY MAX MAG (STEP 1): 19
#       FILTERED MIN MAG (STEP 2):
#       FILTERED MAX MAG (STEP 2):
#
#       Boundaries can be circles (Ra(J2000)-center,Dec(J2000)-center, radius degrees) or
#       polygons (Ra,Dec Ra,Dec Ra,Dec ...). Each boundary is separated by ;
#
#       CLUSTER BOUNDARIES:
#       75.9730,23.6549,0.6;
#
#       FIELD BOUNDARIES:
#       75.9730,23.6549,1.90;
#
#       VOID BOUNDARIES:
#       75.9730,23.6549,1.40;
#
#       NUM STARS: 111790
#
#       EXPECTED NUMBER OF MEMBERS: 2987
#
#       For Webda clusters, STAR_NO is the merge of the Data source reference and the star number: Ref_Star
#
#       Region values: F = star was inside a "only field" region
#                      C = star was inside a "cluster+field" region
#                      V = star was outside the selected regions
```

Highest probability results sample (only 50 first shown).

##STAR_NO	RAJ2000	DECJ2000	pmRA	epmRA	pmDEC	epmDEC	Pll	ePll	BP	eBP	RP	eRP	G	eG	RV	eRV	PROB	Region
58045	76.180554	23.818962	3.216	0.065	-3.453	0.041	1.0595	0.0384	15.145	0.005	13.9404	0.0016	14.6278	5.0E-4	NaN	NaN	0.5460769	V

Figure A16. Step 3 of the Clusterix procedure for NGC 1750 and NGC 1758.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

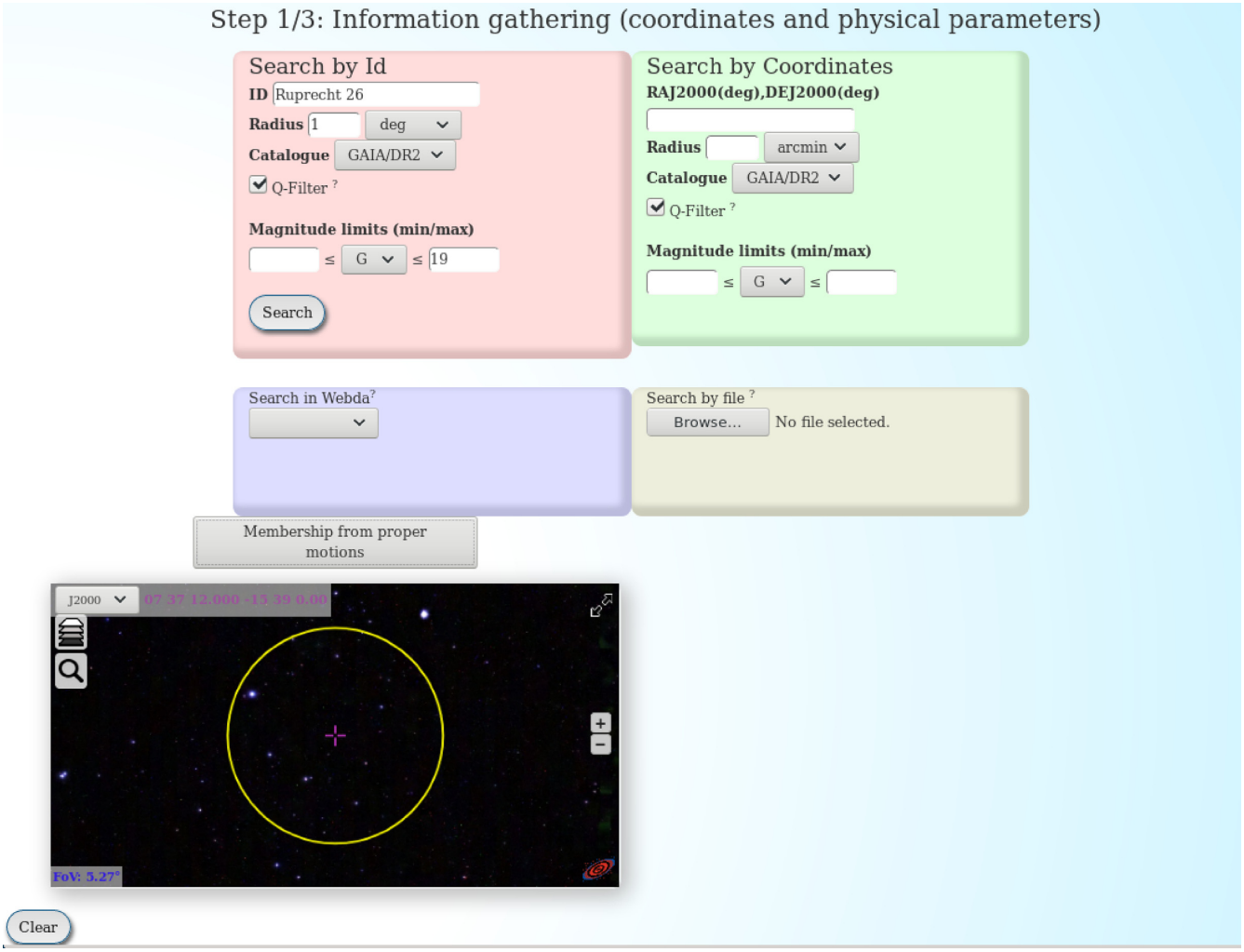


Figure A17. Step 1 of the `Clusterix` procedure for Ruprecht 26 and CLUSTERIX 1.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

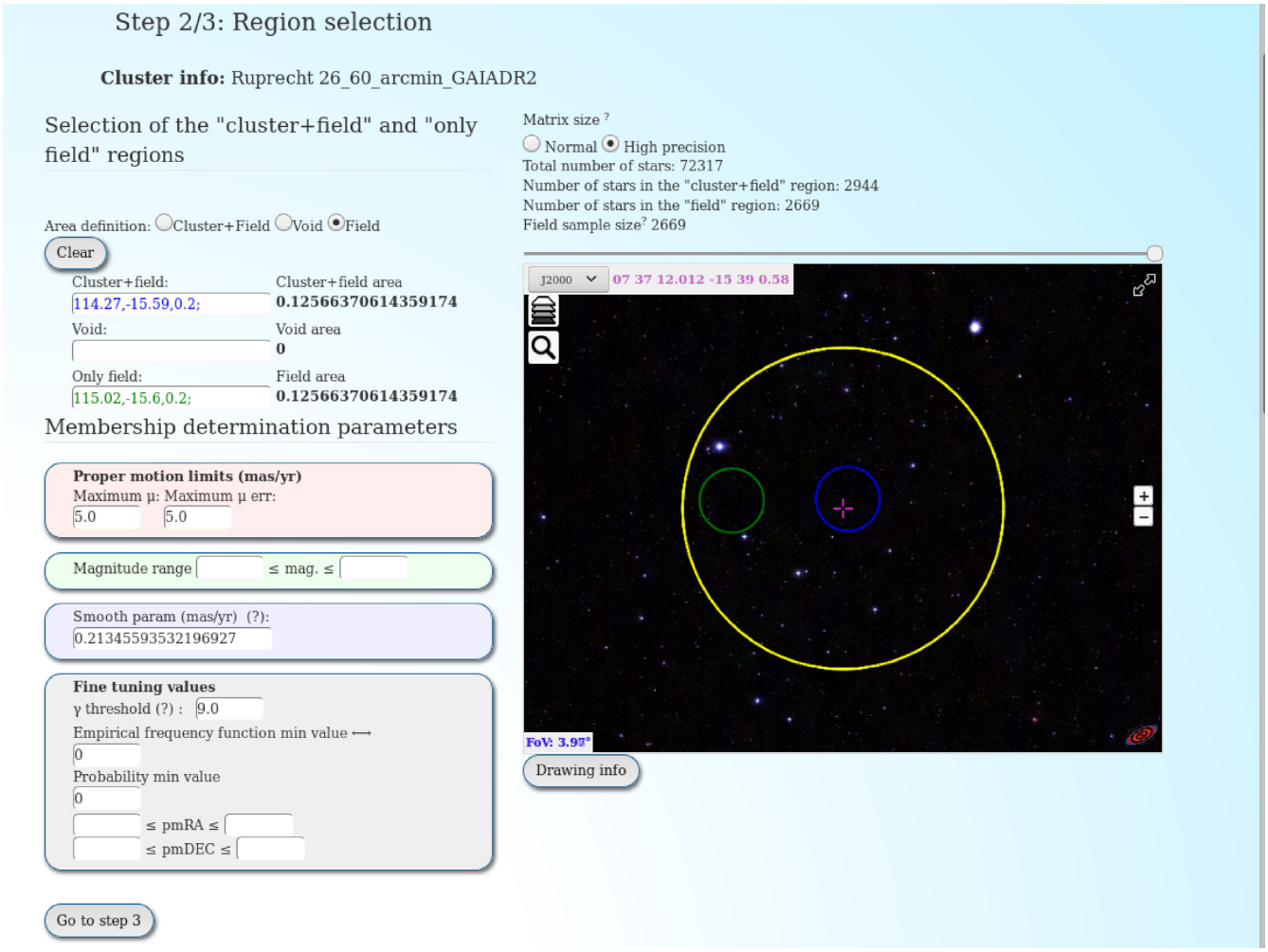


Figure A18. Step 2, where all the fine-tuning parameters of the Clusterix procedure are set for Ruprecht 26 and CLUSTERIX 1.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026

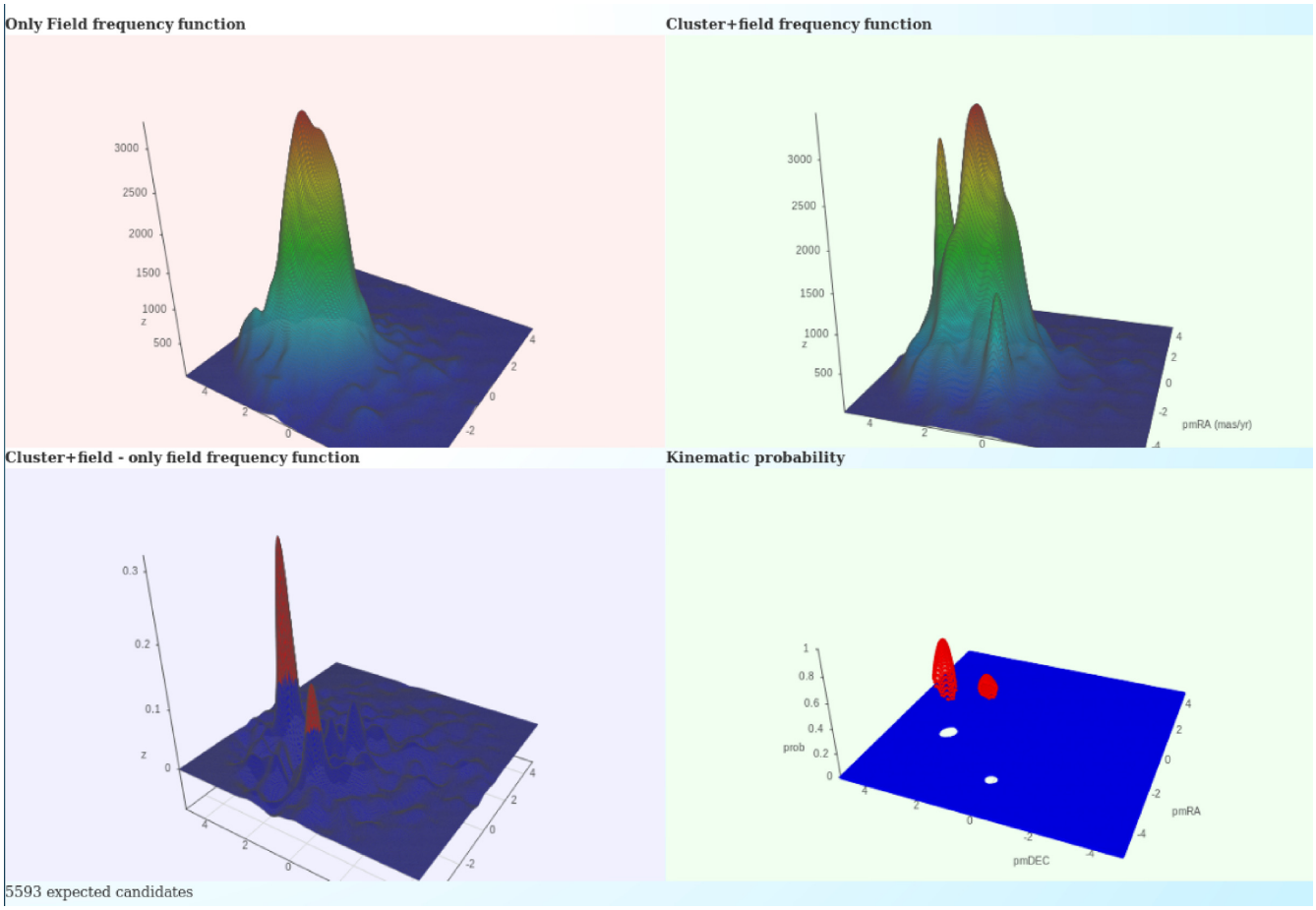


Figure A19. Results of Step 2 `Clusterix` procedure for Ruprecht 26 and CLUSTERIX 1.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026



Figure A20. Step 3 of the Clusterix procedure for Ruprecht 26 and CLUSTERIX 1.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/492/4/5811/5733679 by guest on 12 February 2026