A novel generative method for star clusters from hydro-dynamical simulations

Stefano Torniamenti[1,2,3]

[1]Physics and Astronomy Department Galileo Galilei, University of Padova, Vicolo dell'Osservatorio 3, I–35122, Padova, Italy

[2]INFN- Sezione di Padova, Via Marzolo 8, I–35131 Padova, Italy

[3]INAF, Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 5, 35122 Padova, Italy
email: stefano.torniamenti@studenti.unipd.it

Abstract. Most stars form in clumpy and sub-structured clusters. These properties also emerge in hydro-dynamical simulations of star-forming clouds, which provide a way to generate realistic initial conditions for $N-$body runs of young stellar clusters. However, producing large sets of initial conditions by hydro-dynamical simulations is prohibitively expensive in terms of computational time. We introduce a novel technique for generating new initial conditions from a given sample of hydro-dynamical simulations, at a tiny computational cost. In particular, we apply a hierarchical clustering algorithm to learn a tree representation of the spatial and kinematic relations between stars, where the leaves represent the single stars and the nodes describe the structure of the cluster at larger and larger scales. This procedure can be used as a basis for the random generation of new sets of stars, by simply modifying the global structure of the stellar cluster, while leaving the small-scale properties unaltered.

Keywords. galaxies: star clusters, stellar dynamics, methods: numerical, methods: statistical

1. Introduction

A large fraction of star formation happens in clusters or associations Lada & Lada (2003). These star-forming systems are characterized by complex phase-space distributions, where sub-structures Larson (1995), fractality (Cartwright 2009; Kuhn *et al.* 2019), relative sub-clump motions Cantat-Gaudin *et al.* (2019), expansion (due to gas expulsion, Hills 1980), and possibly rotation Hénault-Brunet *et al.* (2012) are observed. The early evolution of these stellar systems is of fundamental importance for the comprehension of the present-day structure of older open and globular clusters, and can be modeled in a realistic way by means of direct $N-$body simulations. However, adequate initial conditions, able to take into account the observed phase-space complexity, are necessary for a correct comprehension of this evolutionary phase. In this sense, spherical equilibrium models like Plummer spheres Plummer (1911) or King models King (1966) cannot represent the initial spatial and kinematic distributions of young star clusters. A natural way to reproduce the observed properties of these stellar systems is by means of hydro-dynamical simulations of collapsing molecular clouds, where they naturally emerge (Ballone *et al.* 2020, 2021; Bate *et al.* 2009; Klessen & Burkert 2000; Wall *et al.* 2019). However, running hydro-dynamical simulations including all the relevant physics is computationally very expensive, and producing large sets of initial conditions would turn to be prohibitive in terms of computational time.

Table 1. Properties of the stellar clusters and their parent molecular cloud.

	m1e4	m2e4	m3e4	m4e4	m5e4	m6e4	m7e4	m8e4	m9e4	m1e5
$N_{\rm s}$	2523	2571	2825	2868	2231	3054	4214	2945	3161	3944
$M_{\rm s}\ [10^3\,{\rm M_\odot}]$	4.2	6.7	10.3	14.4	14.1	20.4	31.5	28.3	30.5	38.0
$M_{\rm mc}\ [10^3\,{\rm M_\odot}]$	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	100.0
$\epsilon_{\rm sf}$	0.42	0.33	0.34	0.36	0.28	0.34	0.45	0.35	0.34	0.38

Columns: [1] number and [2] total mass of the sink particles, [3] mass of the parent molecular cloud, and [4] the resulting star formation efficiency

Here, we introduce a novel method for producing new initial conditions for $N-$body runs without running additional independent hydro-dynamical simulations. Our approach relies on applying a clustering algorithm to an existing set of initial conditions. In particular, we adopt a hierarchical clustering algorithm, which learns a tree-like representation of the stellar cluster, describing its structure at different scales. In our generative model, new stellar clusters are then obtained by modifying selected nodes of the hierarchical tree.

2. Sink particle distributions

We consider the sink particle distributions (also named as stars in the following) from 10 smoothed-particle hydro-dynamical simulations of molecular clouds, performed by Ballone *et al.* (2020). These simulations are initialized as spherical molecular clouds with total gaseous mass ranging between $10^4\,{\rm M_\odot}$ and $10^5\,{\rm M_\odot}$, uniform temperature $T = 10\,{\rm K}$ and uniform density $\rho = 2.5 \times 10^2\,{\rm cm^{-3}}$. Star formation is implemented during the simulation by means of a sink particle algorithm Bate *et al.* (1995). The star clusters are the result of the instantaneous gas removal at 3 Myr, mimicking the impact of the first supernova explosions (no stellar feedback was included in the simulations). We refer to Ballone et al. (2020,2021) (Ballone *et al.* 2020, 2021) for more details about the hydro-dynamical simulations. Table 1 resumes the main properties of the sink particle distributions under consideration.

3. Hierarchical clustering

Clustering algorithms are a class of unsupervised machine learning methods. In general, clustering identifies similar instances in a given sample and assigns them to groups (or clusters). In the specific case of hierarchical clustering, the algorithm proceeds in a hierarchical way, by connecting the most similar pair of clusters, starting from the individual instances (in this case the single stars), until a certain number of groups is reached† Kaufman *et al.* (1990). To identify similar elements in the sample, hierarchical clustering is provided with a similarity prescription, called linkage. For this case, we make use of Ward's linkage, which merges two clusters such that the variance within all clusters increases the least. This often leads to clusters that are relatively equally sized. For more details on such choice, we refer the reader to Torniamenti *et al.* (2022). Also, we use the implementation offered by the SCIKIT-LEARN library Pedregosa (2011).

3.1. *Application to stellar clusters*

We applied hierarchical clustering to the stellar clusters introduced in Section 2. Before applying the algorithm, we scaled the positions and the velocities by their standard deviations. Figure 1 shows different levels of the cluster hierarchy for the m1e4 cluster. The clumps are organized into a hierarchical tree-like structure $\mathcal{T}$, where the trunk

† When the hierarchy is built from the bottom up, the algorithm is also referred to as agglomerative clustering algorithm.
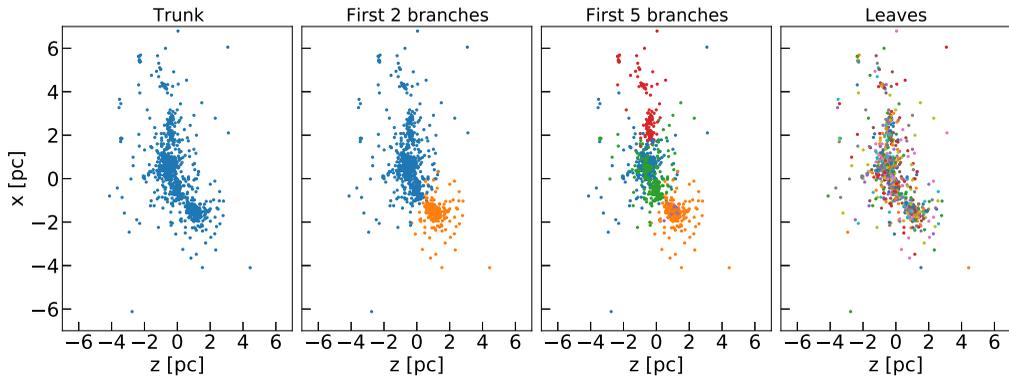
Figure 1. Different levels in the hierarchical tree of the `m1e4` simulation. The panels show different levels of the the tree: [1] the trunk, [2] the first two branches, [3] the first five branches, and [4] the leaves.

contains the whole set of stars and each subsequent node is a two-way split (with each branch being a sub-clump), down to the leaves, representing individual stars.

The hierarchical construction allows to identify groups of similar instances in the distribution of stars as well as drawing information about the structure of the star system at different scales. To describe the relevant physical properties of the cluster by means of the tree formalism, at each node we evaluate the distance vector between the centres of mass of the clumps $\mathbf{l}_i$, their relative velocity vector $\mathbf{u}_i$, and the mass ratio between the two clumps q_i, defined as the ratio between the lightest of the two resulting groups and the total mass of the node.

4. Generative method

Each node of the tree, $\mathcal{T}_i$, describes the relations between two sub-clumps departing from a parent branch. As a consequence, the relevant quantities $\mathbf{l}_i$, $\mathbf{u}_i$, and q_i can be used as instructions to progressively split clumps of stars in the phase space, starting from one reference mass. During this procedure, we can modify some selected nodes of the tree to obtain a new and different realization of the stellar cluster. The way in which the nodes are modified depends on what scales of the systems we want to preserve or alter. In our case, we aim to obtain new macroscopic configurations by preserving the small scale properties (such as their complex fractal structure), which make our clusters so realistic. For this reason, to obtain a new realization, we replace the first nodes of a reference tree, which describe the large scale distribution of sub-clumps, with the same quantities drawn from other trees. The method consists in the following steps:

• We consider a reference tree $\mathcal{T}$, and we replace the quantities $\mathbf{l}_i$, $\mathbf{u}_i$ and q_i, associated to the first $i < k$ nodes, with the same-level quantities from the tree $\mathcal{T}'$, learned from another set of sink particles. Here, we consider $k = 3$. For more details on such choice, we refer the reader to Torniamenti *et al.* (2022) Torniamenti *et al.* (2022).

• We consider one particle containing the total mass of the cluster we want to generate, placed at rest in the origin of the coordinate system. This particle is split into two particles, such that the resulting mass ratio is q_1 (the mass ratio relative to the first node of the new tree). The positions and velocities of the new particles are assigned such that their centre of mass is at rest in the origin of the system, and their distance and relative velocity vectors are $\mathbf{l}_1$ and $\mathbf{u}_1$, respectively.

• At each step i, we split a chosen particle into two new particles with mass ratio q_i and place them at a distance $\mathbf{l}_i$ from each other, moving with relative velocity $\mathbf{u}_i$. The particle-to-split is chosen by considering the same order of splitting as the original
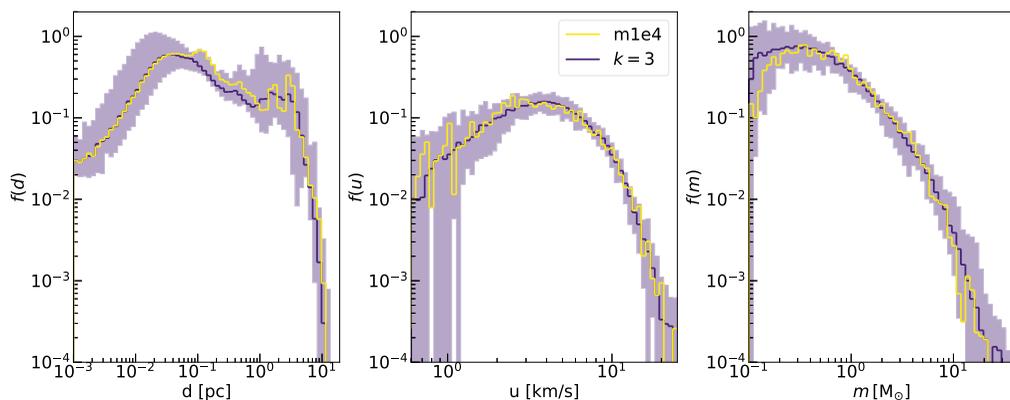
Figure 2. Distributions of inter-particle distances $f(d)$ (left), velocities $f(v)$ (center), and masses $f(m)$ (right) for the sink particles taken from the `m1e4` simulation (thick yellow line) and the distributions of new generations obtained by replacing the first 2 nodes (corresponding to $k = 3$, purple). The shaded area encloses the distribution of the new generations, and the solid line is the median of the distribution.

reference tree. This splitting procedure is then repeated until a cluster with the same number of particles as the reference one is obtained.

• Finally, we remove the very low-mass stars (which may result in planet-sized objects) by setting a cutoff mass to the minimum mass of the original stars on which $\mathcal{T}$ was learned.

In Fig. 2 we compare the distance, velocity and mass distributions of `m1e4` to those of a set of new generations. All the distributions of the new realizations are consistent with those of the original simulation. In particular, the new distance distributions are altered at large scales but, moving toward small distances, they recover the same trend as the original simulation, as meant for this method. Figure 3 and 4 show the spatial distributions of the original cluster and of three new generations per each, for all the sink particle distributions of our sample. The new generations are qualitatively indistinguishable from the original clusters.

5. Summary

We introduced a new method for generating a number of new realizations from a given set of initial conditions from hydro-dynamical simulations. This method is based on a hierarchical clustering algorithm, which learns a tree-like representation of the stellar system. This tree encodes the structural properties of the sink particle distributions and can be turned into new macroscopic realizations by modifying its first branches. This procedure results in different large scale realizations (e.g., the number of main clumps and their distances), while approximately preserving the characteristics of the small scale structure responsible for most of dynamical evolution. The new realizations are qualitatively similar to the original simulations when visualized in the three-dimensional space, and present consistent velocity, mass and pairwise distance distributions.

This generative method leads to a speedup in computation of several orders of magnitude: generating initial conditions from hydro-dynamical simulations, in fact, requires hundreds of thousands core hours per simulation, while our procedure takes about some core seconds to generate a new realization. Also, our scheme is very flexible, allowing to set how deep we alter the tree structure by choosing the number of initial branches we modify.
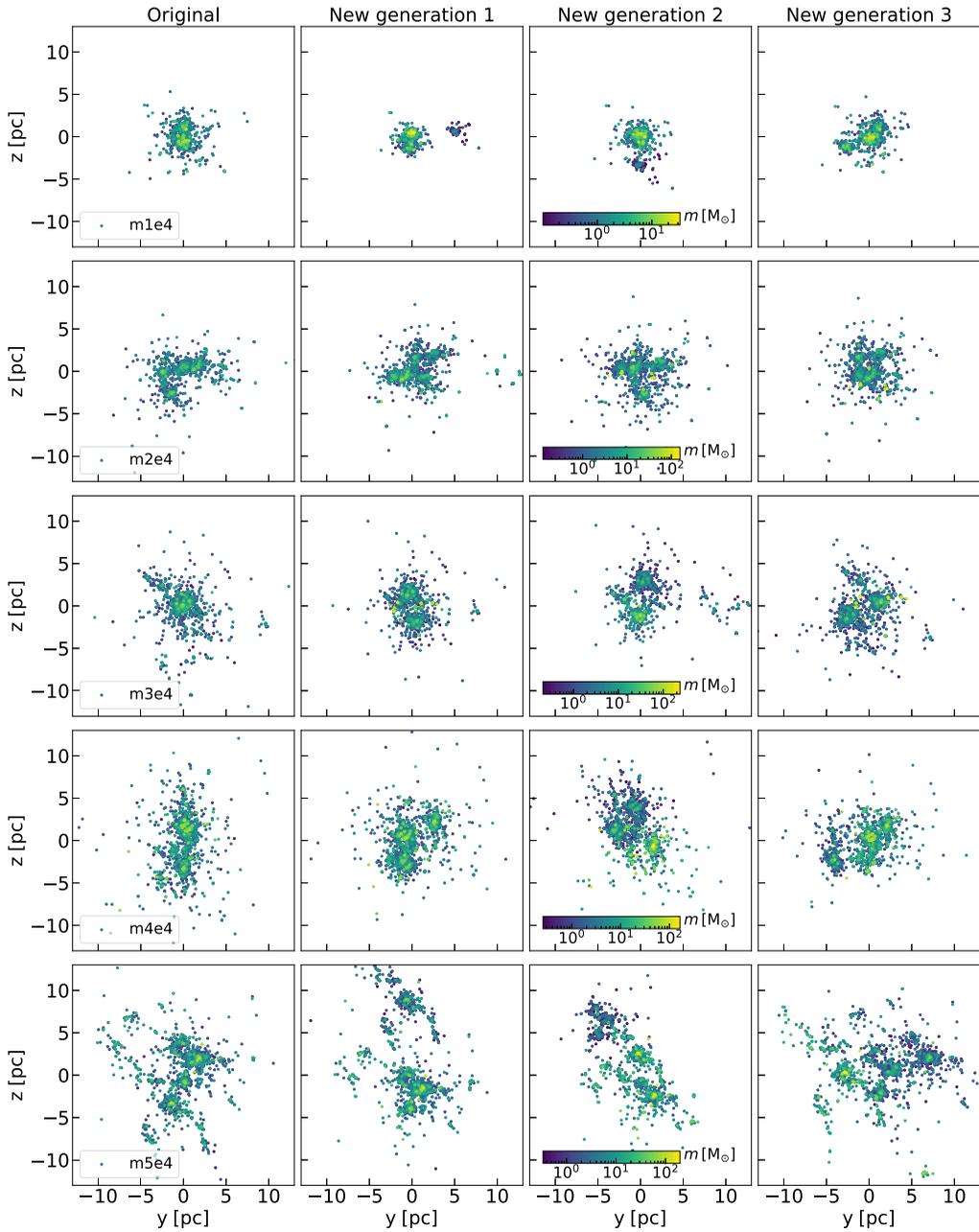
Figure 3. Projections in $y - z$ of the 5 least massive star clusters (left), and of three different generated clusters per each. The colour map marks the mass of the individual stars.
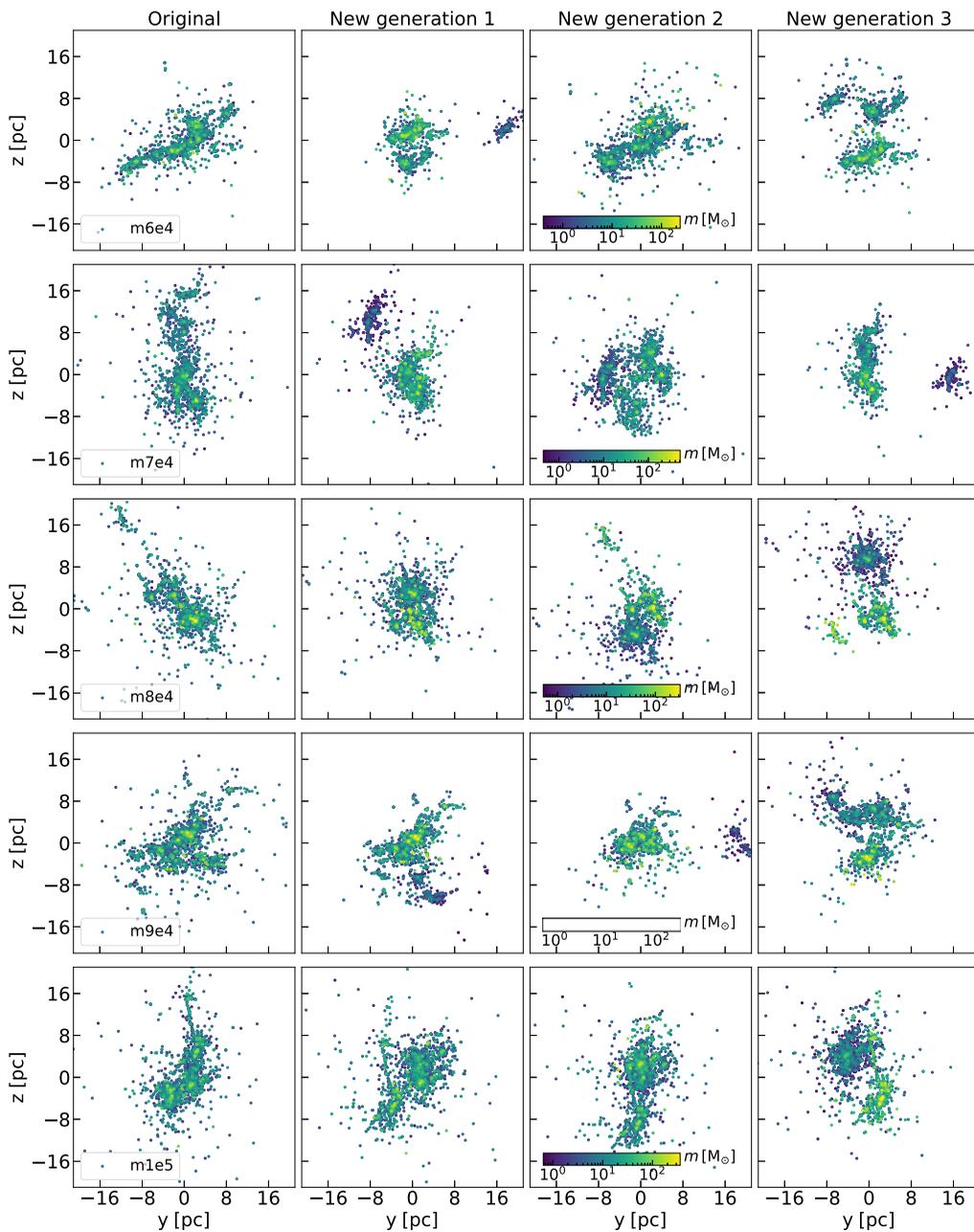
6. Acknowledgements

Figure 4. Projections in $y - z$ of the 5 most massive star clusters (left), and of three different generated clusters per each. The colour map marks the mass of the individual stars. The colour map marks the mass of the individual stars.

References

Ballone A., Mapelli M., Di Carlo U. N., Torniamenti S., Spera M., Rastello S., 2020, *MNRAS*, 496, 49

Ballone A., Torniamenti S., Mapelli M., Di Carlo U. N., Spera M., Rastello S., Gaspari N., Iorio G., 2021, *MNRAS*, 501, 2920

Bate M. R., Bonnell I. A., Price N. M., 1995, MNRAS, 277, 362B

Bate M. R., 2009, *MNRAS*, 392, 1363

Cantat-Gaudin T., et al., 2019, *A&A*, 626, A17

Cartwright A., 2009, *MNRAS*, 400, 1427

Hénault-Brunet V., et al., 2012, *A&A*, 545, L1

Hills J. G., 1980, *ApJ*, 235, 986

Kaufman L., Rousseeuw P. J., 1990, Finding groups in data. an introductionto cluster analysis, Wiley Series in Probability and Statistics.

King I. R., 1966, *AJ*, 71, 64

Klessen R. S., Burkert A., 2000, *ApJS*, 128, 287

Kuhn M. A., Hillenbrand L. A., Sills A., Feigelson E. D., Getman K. V., 2019, *ApJ*, 870, 32

Lada, C. J., Lada, E. A., 2003, *ARAA* 41, 57

Larson R. B., 1995, 272, 213

Pedregosa F., et al., 2011, *Journal of Machine Learning Research*, 12, 2825

Plummer H. C., 1911, *MNRAS*, 71, 460

Torniamenti S., Pasquato M., Di Cintio P., Ballone A., Iorio G., Mapelli M., 2022, *MNRAS*, 510, 2097

Wall J. E., McMillan S. L. W., Mac Low M.-M., Klessen R. S., PortegiesZwart S., 2019, ApJ, 887, 62