A binomial stochastic framework for efficiently modeling discrete statistics of convective populations

Roel Neggers[1] and Philipp Johannes Griewank[1]

[1]University of Cologne

November 22, 2022

Abstract

Understanding cloud-circulation coupling in the Trade wind regions, as well as addressing the grey zone problem in convective parameterization, requires insight into the genesis and maintenance of spatial patterns in cumulus cloud populations. In this study a simple toy model for recreating populations of interacting convective objects as distributed over a two-dimensional Eulerian grid is formulated to this purpose. Key elements at the foundation of the model include i) a fully discrete formulation for capturing binary behavior at small population sample sizes, ii) object demographics for representing life-cycle effects, and iii) a prognostic number budget allowing for object interactions and co-existence of multiple species. A primary goal is to optimize the computational efficiency of this system. To this purpose the object birth rate is represented stochastically through a spatially-aware Bernoulli process. The same binomial stochastic operator is applied to horizontal advection of objects, conserving discreteness in object number. Implied behavior of the formulation is assessed, illustrating that typical powerlaw scaling in the internal variability of subsampled convective populations as found in previous LES studies is reproduced. Various simple applications of the BiOMi model (Binomial Objects on Microgrids) are explored, suggesting that well-known phenomena from nature can be captured at low computational cost. These include i) subsampling effects in the convective grey zone, ii) stochastic predator-prey behavior, iii) the down-scale turbulent energy cascade, and iv) simple forms of spatial organization and convective memory. Consequences and opportunities for convective parameterization in next-generation weather and climate models are discussed.

A binomial stochastic framework for efficiently modeling discrete statistics of convective populations

R. A. J. Neggers[1] and P. J. Griewank[1],

[1]Institute for Geophysics and Meteorology, University of Cologne, Germany.

Key Points:

- An efficient scale-aware stochastic number generator based on a Bernoulli process is applied to model object births and advection on Eulerian grids.
- Discreteness in object number is conserved, while an age dimension is included to represent evolution of object demographic strata.
- Population subsampling effects in the convective grey zone are reproduced, while simple applications capture behavior as observed in nature.

Corresponding author: R. A. J. Neggers, `neggers@meteo.uni-koeln.de`

Abstract

Understanding cloud-circulation coupling in the Trade wind regions, as well as addressing the grey zone problem in convective parameterization, requires insight into the genesis and maintenance of spatial patterns in cumulus cloud populations. In this study a simple toy model for recreating populations of interacting convective objects as distributed over a two-dimensional Eulerian grid is formulated to this purpose. Key elements at the foundation of the model include i) a fully discrete formulation for capturing binary behavior at small population sample sizes, ii) object demographics for representing life-cycle effects, and iii) a prognostic number budget allowing for object interactions and co-existence of multiple species. A primary goal is to optimize the computational efficiency of this system. To this purpose the object birth rate is represented stochastically through a spatially-aware Bernoulli process. The same binomial stochastic operator is applied to horizontal advection of objects, conserving discreteness in object number. Implied behavior of the formulation is assessed, illustrating that typical powerlaw scaling in the internal variability of subsampled convective populations as found in previous LES studies is reproduced. Various simple applications of the BiOMi model (Binomial Objects on Microgrids) are explored, suggesting that well-known phenomena from nature can be captured at low computational cost. These include i) subsampling effects in the convective grey zone, ii) stochastic predator-prey behavior, iii) the down-scale turbulent energy cascade, and iv) simple forms of spatial organization and convective memory. Consequences and opportunities for convective parameterization in next-generation weather and climate models are discussed.

Plain Language Summary

Convective clouds in the Trade wind regions play a crucial role in Earth's climate. The way they interact with the atmospheric circulation is not well understood, and is associated with long-standing problems in weather forecasting and climate prediction. Recent research has suggested that the spatial structure of these cloud fields is a key factor in this problem, and that improving our understanding of such convective cloud patterns is crucial for making progress. This study explores a new model framework for generating such cloud patterns, consisting of populations of convective objects on small grids. The objects are born in a random way, complete a life cycle, and can freely move around on the grid. They can also interact and form larger clusters, obeying certain rules of interaction. The way the objects behave and move around features some key innovations compared to previous ecosystem models of this kind. These are introduced to optimize the performance and reduce run time on a computer. Various experiments are conducted to explore the new model, illustrating that observed behavior of convective populations is reproduced. These tests also highlight opportunities created for improving convection in weather and climate models.

1 Introduction

Convective cloud populations in Earth's atmosphere cover a broad range of spatial scales. Their occurrence acts on planetary scales, by persistently covering substantial areas of the marine subtropical Trade wind regions. On the other end, individual clouds have dimensions from a few meters up to tens of kilometers. The spatial structure of cumulus populations acts on the intermediate (meso)scales and can take many forms, including random-like distributions (Nair et al., 1998) but also more organized patterns including cold pool structures and convergence lines (Bony et al., 2020).

Understanding the spatial structure of cumulus populations is important for various reasons. Global weather and climate models require parameterizations to represent the impact of subgrid-scale processes on the resolved-scale flow. Until recently this still fully included cumulus convection, but ongoing advances in supercomputing have gradually created a "grey zone problem" (Wyngaard, 2004; Honnert et al., 2020) in which feasible gridspacings approaches typical neighbor spacings of cumulus clouds (Joseph & Cahalan, 1990). This means convective populations are no longer fully sampled in individual gridboxes, a situation for which existing convective parameterizations need to be adapted (Kwon & Hong, 2017; Brast et al., 2018). A second motivation for studying the spatial structure of cumulus populations is the role it plays in the cloud-climate feedbacks (Vogel et al., 2016; Wing et al., 2018).

The investigation of spatial patterns in convective cloud fields goes back decades, using large-domain covering observations (Sengupta et al., 1990; Weger et al., 1992; Nair et al., 1998) and more recently also simulations (Tompkins & Semie, 2017; Feingold et al., 2017; Neggers et al., 2019). What is clear is that spatial patterns consist of many individual convective objects. Zooming in on any pattern then leads to ever fewer elements being contained in the shrinking domain of interest. As a result, bulk population averages go from smoothly behaving for a fully sampled population towards binary behavior for a severely sub-sampled population. The way this happens is strongly affected by clustering (Neggers et al., 2019). Understanding and capturing this transition towards *discrete* behavior, including the role played by spatial organization, is key for developing scale-aware and stochastic convective parameterizations for next-generation weather and climate models.

Population models including many small convective elements can give useful new insights into this problem, and potentially provide new pathways for convective parameterization. For example, rules of interaction can be introduced that reflect known or observed physics, by which spatial patterns can emerge freely. Such rules are known from game theory (von Neumann, 1928; von Neumann & Morgenstern, 1944) and cellular automata (von Neumann, 1966; Gardner, 1970). A promising recent example is the lattice or microgrid approach (Khouider et al., 2010; Dorrestijn et al., 2013; Peters et al., 2017), which allows multiple *cloud-scale* structures to evolve naturally and gradually on a 2D grid. Other cloud-scale stochastic frameworks were recently proposed by Hagos et al. (2018) and Sakradzija et al. (2016). One step further down-scale is the Lagrangian particle approach of (Böing, 2016), which tracks a multitude of interacting *sub-cloud* scale elements as they form larger clusters on the grid. Although yielding powerful results, what remains relatively unexplored is how such systems behave in the grey zone, in particular their stochastic and discrete behavior resulting from population subsampling in a too small gridbox. One also wonders if the often considerable computational burden of such multi-object approaches might limit their use as part of a convective parameterization.

To gain further insight, in this study a simple toy model is formulated for recreating populations of interacting convective objects as distributed over a two-dimensional grid. A defining principle is its fully discrete formulation, aimed at capturing binary behavior at small population sample sizes. Another primary goal is to achieve a formulation that is generally applicable to many types of convection and convective object definitions, with a computational efficiency that is as high as possible. Object births are represented stochastically as a spatially-aware Bernoulli process, taking the form of a binomial number generator. The same operator is applied to horizontal advection of objects between gridboxes, making this process similarly stochastic and discrete. Object demographics are included, creating age strata and allowing discrete and explicit representation of life-cycle effects. The formulation of the framework allows for multiple co-existing species, as well as interactions to take place between individual convective objects. The formulation in terms of a Bernoulli process at multiple points in the model considerably enhances the computational efficiency.

Section 2 presents the basic formulation of the framework. In Section 3 behavior as implied by the formulation is briefly discussed, including an interpretation of implied scaling behavior, the advection operator, and the computational efficiency of the framework. Section 4 demonstrates simple applications of the framework on microgrids, including both single-species and multi-species setups. This application on microgrids is named BiOMi (Binomial Objects on Microgrids). Opportunities created by introducing simple physics-based rules of object interaction are explored, including predator-prey behavior, spatial organization and convective memory. Section 5 interprets these results in the context of limitations in the formulation, and compares to other recently proposed stochastic frameworks for atmospheric convection. Section 6 then summarizes the main conclusions and provides an outlook on future steps inspired by this study.

2 Formulation

In this section the framework for describing an evolving population of objects on a discretized grid is defined. At its foundation is a prognostic budget for object number that is discrete and includes various sources and sinks. We adopt the following guiding principles in its formulation:

1. The objects should have a stochastic birth rate and a finite lifespan;
2. The number of objects present in a gridbox should be both discrete and positive-definite, at any time;
3. The formulation should be general enough to be applicable to any type of convection.

Adopting the first and second principles is motivated by our primary goal of capturing the type of stochasticity that is introduced by the sub-sampling of populations in a too small gridbox. In this “grey-zone” range of resolutions, only a few objects are present at varying stages of their life-cycle, which may lead to binary (i.e. on-off) behavior in their averaged properties. Adopting a discrete approach has direct implications for the formulation of all terms in the number budget.

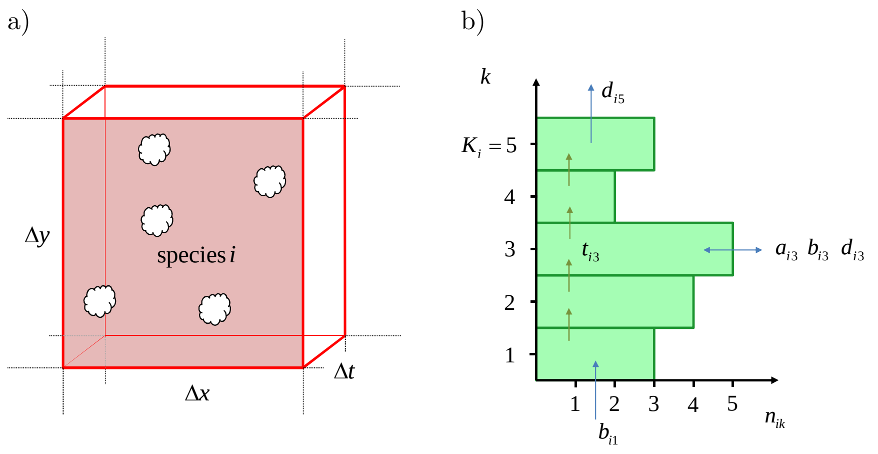

Figure 1. a) Schematic illustration of a population of objects of species i inside a three-dimensional space-time gridbox (red) with square horizontal area $\Delta x \Delta y$ and time-step Δt. b) Schematic illustration of object demographics for a species i with 5 age strata. The blue arrows indicate external sources and sinks of the demographics budget (2), while the green arrows indicate the internal aging process. Variables are explained in Section 2.1.

Adhering to the third principle makes it necessary to refrain from defining any closures that reflect specific physics behavior, as this by definition would make the framework no longer generally applicable. Accordingly, in this section the formulation of such physical parameterizations is for now left open. However, in Section 4 a few simple examples will be explored.

2.1 A discrete budget for object number

Consider a three-dimensional space-time gridbox covering a square horizontal area $\Delta x \Delta y$ and time-step Δt, as depicted in Fig. 1a. This grid box can contain a population of objects, potentially consisting of multiple species. The discrete number of objects of species i is indicated as n_i, with I being the total number of different species. How exactly species are defined is left open at this point, to maintain general applicability of the framework. Note that the vertical dimension is omitted because the altitude of objects is not considered in this framework.

We now introduce a fourth dimension, which is object age k. The number of objects of species i in a gridbox can then be written as $n_i(x, y, k, t)$. All four dimensions are discretized. As a result, the k-dimension introduces a discrete form of object demographics, with k being an integer number indicating an age-stratum. For simplicity all objects of a species i are assumed to have the same life-span τ_i, by which the number of age strata K_i is obtained through

$$K_i = \frac{\tau_i}{\Delta t} \tag{1}$$

In practice, the chosen time discretization determines how many demographics levels are maintained. The life times of objects are chosen to be a multiple of Δt, so that K_i is always an integer number.

The final step is to formulate a prognostic budget for each species i at each age level k. This gives

$$\Delta n_{ik} = b_{ik} - d_{ik} + a_{ik} + t_{ik}. \tag{2}$$

The left hand side Δn_{ik} represents the change of n_i at demographics level k per time step Δt. On the right hand side, b_{ik} and d_{ik} represent changes in n_{ik} due to births and deaths respectively, a_{ik} represents net advection of objects from neighboring gridboxes, and t_{ik} represents the process of object aging (demographics). Hereafter, lower-case notation indicates the property of a gridbox, while upper-case notation reflects the integral or average properties of a much larger domain. To shorten the notation only the species and age indices i and k are carried as subscripts. Each demographics level k thus has its own number budget. Note that all terms in (2) are still integer numbers.

2.2 Object births as Bernoulli trials

The first step in the closure of b_{ik} is to assume that objects of species i have a unique reference birth rate per unit area and unit time when diagnosed over an infinitely large area. Let us write this birth rate as $\dot{B}_i$. Because this rate depends strongly on the definition of the species, for now we assume this birth rate as a given, known property. By adopting this assumption we follow the recent study of Böing (2016).

Given $\dot{B}_i$, the next step is to consider a finite but still very large reference domain of horizontal size L in which the population of convective objects is still fully sampled. The average total number of births of species i within this reference domain during one time-step, B_i, can then be written as

$$B_i = \dot{B}_i \, L^2 \, \Delta t \tag{3}$$

A convenient choice of a reference domain would be the whole globe, as this represents the theoretical upper limit of gridspacing in any General Circulation Model (GCM) used for global weather and climate prediction. For smaller scale shallow convection one could also choose a smaller domain, for example the subtropical marine Trade wind region. When B_i is large and the reference domain is much larger than the individual gridbox, the binomial sampling approaches the Poisson distribution used by Sakradzija et al. (2015) to determine stochastic cloud births per gridbox.

Discretizing this reference domain at resolution $(\Delta x, \Delta y, \Delta t)$ results in a number of gridboxes N,

$$N = \frac{L^2}{\Delta x \Delta y}. \tag{4}$$

The total number of birth events in the reference domain, B_i, is spatially distributed over the grid, yielding an average number of birth events in a single gridbox, μ_i,

$$\mu_i = \frac{B_i}{N} \tag{5}$$

Let us assume for the moment that the spatial distribution is purely random (we will deviate from this condition later). Then for each of these N birth events the probability p that it takes place inside a specific gridbox is

$$p = 1/N. \tag{6}$$

Note that probability p is the same for each species, and is purely a property of the discretized grid. In that sense it introduces scale-awareness, or awareness of the gridspacing. Dependence on species is introduced by B_i.

The key step in defining the stochastic birth generator is to assume that the number of births in an arbitrary gridbox is independent of other gridboxes and timesteps. This means that object birth events can be considered as single, independent *Bernoulli trials*, associated with a specific success/failure probability p. With that assumption the full set of B_i birth events that takes place within the reference domain then becomes a Bernoulli process. Adopting the configuration as defined above this can be written as the following probability mass function,

$$f_i(b) = \binom{B_i}{b} p^b \, (1-p)^{(B_i - b)} \tag{7}$$

where the binomial coefficient is defined as

$$\binom{B_i}{b} = \frac{B_i!}{b! \, (B_i - b)!} \tag{8}$$

where we assumed for convenience that B_i can be rounded to the nearest integer. Function $f_i(b)$ can be interpreted as the probability of b births of objects of species i in an arbitrary gridbox, given a reference domain with properties B_i and p. The mean μ_i of this binomial distribution, or its expected value, is defined as

$$\mu_i \; = \; B_i \; p, \tag{9}$$

which, according to (5) and (6), corresponds exactly to the average number of object births per gridbox. Note that the actual average number of births on the grid might deviate from this expected value because each gridbox is sampled independently.

In practice, in each space-time gridbox the integer number of births of objects of species i is determined by randomly sampling the binomial distribution (7). This can be written as a binomial number generator,

$$b_{i1} = \mathcal{B}\left(B_i, p\right), \tag{10}$$

where $\mathcal{B}$ represents a single random sample of binomial function f_i. The number of births b_{i1} thus established for each gridbox can directly be used in budget equation (2), with subscript $k = 1$ reflecting that all newly born objects enter the demographics array at the first (youngest) level. The birth rates b_{ik} for $k > 1$ are set to zero for the moment.

2.3 Object demographics

The introduction of the age dimension k allows representing object life-cycle effects. At the start of every timestep, objects in one demographics level are *time-shifted* into the next (older) level. This process is illustrated in Fig. 1b (green arrows). This process of object aging is included in budget (2) through the operator t_{ik}, defined as

$$t_{ik} \; = \; \begin{cases} \quad\quad\quad\; - \, n_{ik} & \text{for} \quad k = 1 \\ n_{i,k-1} \; - \, n_{ik} & \text{for} \quad 2 \leq k < K_i \\ n_{i,k-1} & \text{for} \quad k = K_i \end{cases} \tag{11}$$

The time-shift out of the top (oldest) level represents object death due to old age ,

$$d_{ik} \; = \; n_{ik} \;\; \text{for} \;\; k = K_i \tag{12}$$

Note that this death rate is automatic and discrete, in that it can not create fractional object numbers. In this aspect it is different from Newtonian relaxation, which would

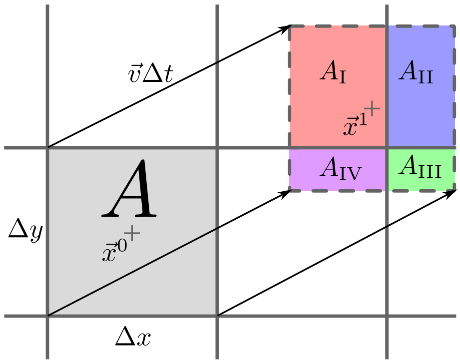

Figure 2. Schematic illustration of overlap between a displaced gridbox and the underlying grid. The arrows represent the displacement over one time step, which is simply the horizontal wind multiplied by the time step duration. Grey crosses mark the mid of the gridbox before and after displacement. See section 2.4 for full description.

be an alternative (but non-discrete) formulation. Futhermore, the amount of deaths per turn is not determined by the amount of objects currently alive, but is directly determined by the amount of births K_i time steps earlier. The death rates d_{ik} for $k < K_i$, which represent deaths caused by processes other than ageing, are set to zero for the moment.

2.4 A discrete advection operator

If horizontal advection is to be taken into account an advection approach must be chosen which preserves the total number of objects and their discrete nature. No fractions of objects are permitted.

The same Bernoulli process we use to distribute the number of births over a two-dimensional domain can be used to create a stochastic upwind advection scheme for discrete objects. At the core of this scheme is the assumption that the objects are randomly spatially distributed within each gridbox. From this assumption the probability of an object to be advected from one gridbox to another can be determined from the overlap area as shown in Fig 2. From this principle a conservative advection scheme can be derived that requires 3 sequenced Bernoulli trials per advected gridbox, age strata, and species.

The first step is to determine the arrival point $\vec{x}^1$ of the gridbox mid point after translation from its original location $\vec{x}^0$ due to advection by the horizontal wind $\vec{v}$,

$$\vec{x}^1 = \vec{x}^0 + \vec{v}\Delta t \tag{13}$$

The new gridbox is centered around the arrival point $\vec{x}^1$, making it overlap with 4 gridboxes. When the displacement is smaller than the grid box there is chance objects will remain in the original gridbox, if the displacement is larger all objects will move outside. The overlap areas A_j are labeled in clockwise direction from the topleft one, and obey

$$A = \sum_{j=\mathrm{I}}^{\mathrm{IV}} A_j \tag{14}$$

where $A = \Delta x \Delta y$. For each age level k, we now randomly select objects from the total number of objects in the original gridbox, n_{ik}, to arrive in each of these four areas A_j. To this purpose the binomial operator $\mathcal{B}$ as defined before is used,

$$a_{ik,\mathrm{I}} = \mathcal{B}(n_{ik}, \frac{A_\mathrm{I}}{A}) \tag{15}$$

$$a_{ik,\mathrm{II}} = \mathcal{B}(n_{ik} - a_{ik,\mathrm{I}}, \frac{A_\mathrm{II}}{A - A_\mathrm{I}}) \tag{16}$$

$$a_{ik,\mathrm{III}} = \mathcal{B}(n_{ik} - a_{ik,\mathrm{I}} - a_{ik,\mathrm{II}}, \frac{A_\mathrm{III}}{A - A_\mathrm{I} - A_\mathrm{II}}) \tag{17}$$

The number of objects advected into A_IV is then simply obtained as the residual,

$$a_{ik,\mathrm{IV}} = n_{ik} - \sum_{j=\mathrm{I}}^{\mathrm{III}} a_{ik,j} \tag{18}$$

Doing this separately for each age level k means that age is conserved as objects are advected across the grid

For large number of objects per gridbox this discrete advection operator behaves as a continuous first-order upstream approach with high gradient smoothing and fast dispersion. For low object numbers the stochastic nature becomes more visible, with the mean over all objects no longer smoothly tracking the wind. These aspects will be further illustrated in Section 3.2.

2.5 Object interactions

The framework allows introducing interactions between objects in two different ways. The first option is to make birth probability p appearing in (7) dependent on the presence of other objects in the vicinity of the gridbox. These could be locally present, inside the gridbox, but also in a wider area, covering multiple adjacent gridboxes. The spatial extent of such impacts depends on the physical/dynamical nature of the interaction

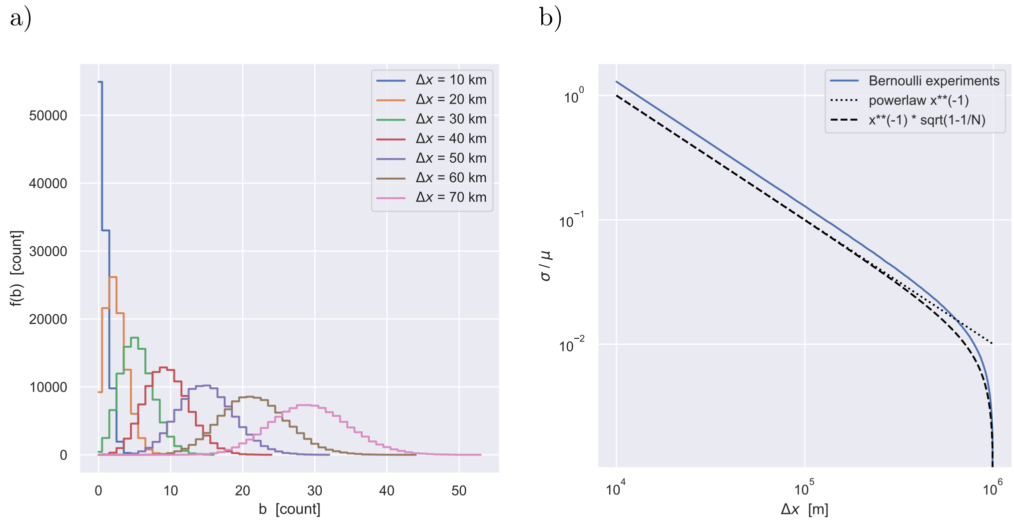

Figure 3. a) Examples of binomial probability density $f(b)$ as defined by (7) for various grid-spacings $\Delta x = \Delta y$, using a birth rate $\dot{B}_i = 10^{-10}$ m^{-2} s^{-1}, a reference domain of size $L = 1000$ km and an integration timestep $\Delta t = 60$ s. Results represent 10^6 independent draws. b) Associated functional form of the normalized standard deviation of the binomial distribution σ/μ, as defined by (22). A pure powerlaw (black dotted) and modified powerlaw (black dashed) functional form are also shown, for reference.

process of interest. The second option is to make the birth and death rates b_{ik} and d_{ik} dependent on the presence of other objects. This method is particularly suited to introduce inter-species interactions. For example, predator-prey dynamics can be introduced by making the death rate of one (prey) species dependent on the presence of another (predator) species. In Section 4 simple applications of the framework will be demonstrated that include both forms of interaction between objects.

3 Implied behavior

With the basic formulation of the framework concluded, some behavior can already be understood a priori its application in practice. The most relevant of these implied characteristics are discussed in this section.

3.1 Stochasticity due to subsampling

Describing object births on the grid as independent Bernoulli trials directly controls the behavior of stochasticity in object number at gridspacings at which the population is becoming subsampled. This is illustrated in Fig. 3, showing the binomial probability density function $f(b)$ as defined by (7) for various gridspacings. Both the mean μ_i and the width $2\sigma_i$ increase with gridspacing Δx, which is expected because p increases with gridspacing through (6). This results in more births per timestep in larger gridboxes. A more useful expression of stochasticity is provided by the relative width of the pdf, σ_i/μ_i. This can be understood by considering the definition of σ_i for the binomial,

$$\sigma_i^2 \;=\; B_i \; p \;\; (1-p) = \mu_i \;\left(1 - \frac{1}{N}\right) \tag{19}$$

The standard deviation σ_i normalized by the mean μ_i can then be written as

$$\frac{\sigma_i}{\mu_i} \;=\; \mu_i^{-\frac{1}{2}} \left(1 - \frac{1}{N}\right)^{\frac{1}{2}}. \tag{20}$$

Note that μ_i carries dependence on both spatial (grid) information and species properties, because it reflects that B_i births are randomly distributed over a discretized spatial domain. Through (5) this implies a relation for the average neighbor spacing l_i between objects born in the gridbox within the time-step,

$$l_i \;=\; \left(\frac{\Delta x \Delta y}{\mu_i}\right)^{\frac{1}{2}} \;=\; \left(\frac{1}{\dot{B}_i \;\Delta t}\right)^{\frac{1}{2}}. \tag{21}$$

Here the neighbor spacing is simply calculated as the square root of the area surrounding each object that is free of other objects (on average). Substituting the first part of (21) for μ_i in (20) then yields the following scaling relation,

$$\frac{\sigma_i}{\mu_i} \;=\; \left(\frac{\Delta}{l_i}\right)^{-1} \left(1 - \frac{1}{N}\right)^{\frac{1}{2}} \tag{22}$$

where we introduced $\Delta \;=\; \sqrt{\Delta x \Delta y}$ to shorten notation. On the right hand side only the variable l_i depends on the species, through the reference birth rate $\dot{B}_i$.

Each term between brackets in the product on the right hand side of (22) has its own specific meaning. The first term introduces a powerlaw dependency (with exponent -1) on the ratio of grid-spacing Δ to the nearest neighbor spacing l_i, with larger values of (Δ/l_i) suppressing the normalized standard deviation. This reflects that the population of object births of species i is better sampled at larger gridspacings, reducing stochasticity in object number. The second term depends purely on the grid, and acts to bring the standard deviation to zero in the limit of the grid spacing approaching the reference domain size.

This behavior is illustrated in Fig. 3b, showing the functional dependence of the normalized standard deviation on gridbox size Δ. In the range of gridspacings typical

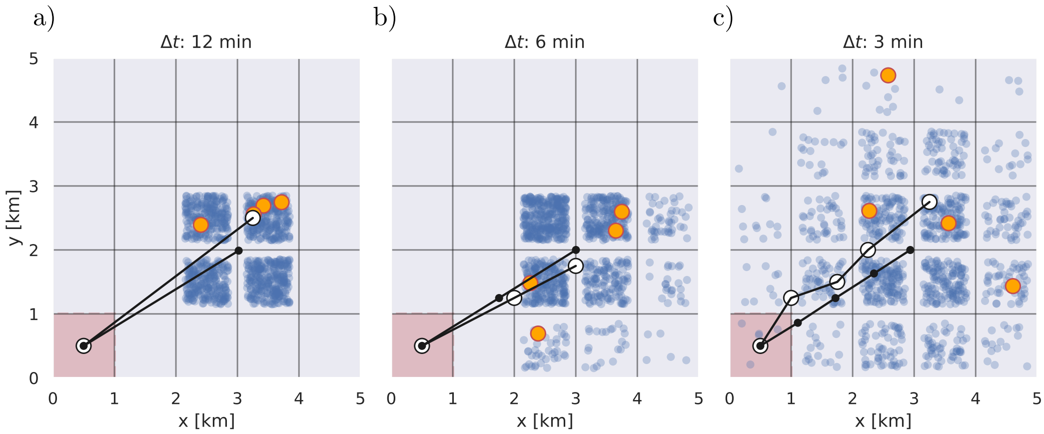

Figure 4. Example of discrete advection of objects on a 5x5 rectangular 1 km grid using the same initial conditions and grid but differing time step. The blue and orange objects behave identically, and differ only in the amount (1000 blue, 4 orange). Note that the individual objects have no specific x and y location within each gridbox, and are only plotted as such for visualisation purposes. The red square marks the gridbox in which all objects were initialized at t=0, and shown are the locations after 12 minutes of diagonal advection. The black line with small black circles marks the mean location at each time step of the blue objects, the large white circles the mean of the large orange objects.

of operational GCMs the second term is almost a constant, because $N \gg 1$. As a result, the dependence of the normalized standard deviation on grid-spacing approximately behaves as a powerlaw with exponent -1. When N approaches 1, the grid in effect becomes a slab model, and the variability is squeezed to zero.

The powerlaw scaling in the normalized standard deviation as implied by this formulation has recently been encountered in studies of the internal variability of shallow cumulus cloud size distributions. Neggers et al. (2019) performed subdomain analyses of unorganized shallow cumulus cloud populations in large-eddy simulations, and found that the variation across subdomains in the number of convective clouds of a given size follows scaling relation (22). This agreement provides support for the applicability of the Bernoulli process for reconstructing such unorganized convective populations.

3.2 Discrete advection

To illustrate the numerics of the discrete advection operator we run a highly idealized experiment in which all objects are initialized in the same gridbox before being advected diagonally (Fig 4). Objects do not interact with each other or have a life cycle, and all differences between the subplots of Fig 4 are due to the differing number and duration of the timesteps. This testcase was designed to maximize advective diffusion in order to highlight the randomness and discreteness of the stochastic advection operator.

For a large number of objects per gridbox the discrete advection operator behaves as a continuous first-order upstream approach with high gradient smoothing and fast dispersion (small blue dots). But in contrast to a continuous upstream approach, the discrete operator is positive definite and not limited by the Courant–Friedrichs–Lewy condition. How strong and in which direction the dispersion acts depends on the angle of the grid to wind direction, gridbox size, and the timestep. The impact of changing the timestep is shown in Fig. 4, illustrating that changing the timestep can not only affect the strength of the dispersion, but also the direction. As in the continuous analog, increasing resolution reduces diffusion (not shown). Despite this numeric diffusion, the mean over a sufficient number of objects will follow the wind direction closely. For low object numbers the stochastic nature becomes more visible, with the mean over all objects no longer smoothly tracking the wind (large white dots). A side effect of the stochastic nature is that an initially smooth field will become heterogeneous when advected. Similar to the stochastic subsampling this effect is more pronounced for low object numbers (not shown).

3.3 Computational viability

Given that efficiency is one of the core concepts of the introduced framework, this subsection briefly discusses the required processing cost and memory requirements of the framework and how they compare to Lagrangian approaches.

3.3.1 Processing

The binomial operator (10) is a cornerstone of the framework, being applied to represent both object births and object advection. A computational benefit of this oper-

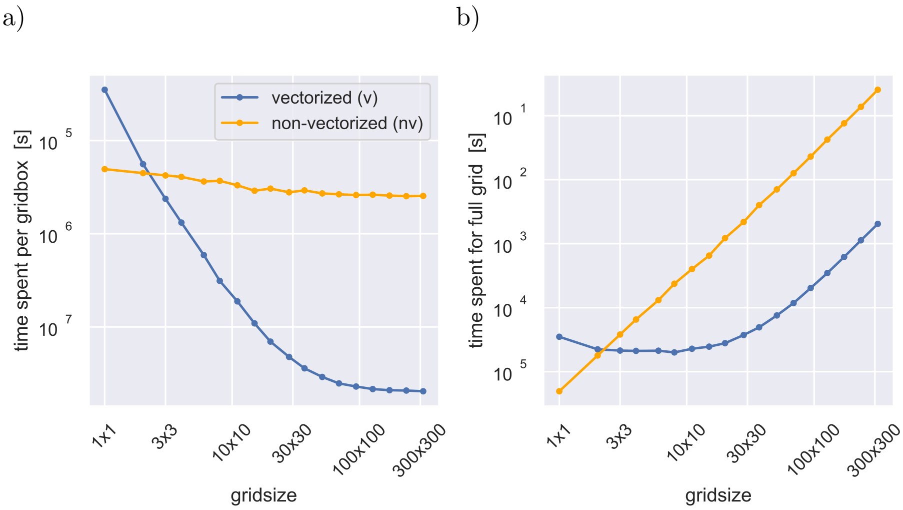

Figure 5. Results of a speed test of the binomial operator (10) as executed in Python on a single Intel i5-6400 2.7 Ghz CPU. a) Time spent per gridbox as a function of gridsize, for a vectorized (v) and non-vectorized (nv) application. b) Time spent for the full grid.

ator is that the operational cost becomes independent of the number of samples drawn from the distribution. This is a clear distinction from the Lagrangian particle approach in population dynamical modeling (Böing, 2016), which computes the evolution and movement of each particle individually. As a consequence, the cost of Lagrangian approaches scales with population size, while that of binomial approaches in principle scales with gridsize, species number, and age strata.

However, thanks to vectorization, the amount of CPU time needed to compute the binomial sampling need not scale linearly with gridsize, species number, and age strata. The results of the efficiency test shown in Fig. 5 shed some more light on this possibility. In the first panel the time spent by the binomial operator for each gridbox is shown as a function of gridsize. As can be expected, applying the operator in a non-vectorized way (i.e. a sample at each gridpoint) keeps this cost per gridbox more or less independent of gridsize (panel a). As a result, the total cost for the whole grid increases linearly with the gridsize (panel b). However, while a vectorized application of the binomial operator is slower for a 1x1 grid, it strongly reduces the computational cost in regards to the gridsize for larger grids. The vectorized version is almost independent of gridsize up until 30x30, after which the vectorized version is 100× faster than the non-vectorized version (panel b). We suspect that the precise gridsize when the cost of the vectorized version begins to increase with gridsize is related to the CPU memory. The boost in efficiency due to the vectorized application, combined with its independence on population size, is what allows the binomial approach to remain computationally viable as part of a convective parameterization, even for micrograds of substantial size. How these benefits hold up in practice will vary with hardware and implementation.

3.3.2 Memory

The memory usage of the binomial framework is not determined by the number of objects as would be the case for a Lagrangian approach (Böing, 2016). Instead memory depends linearly on the amount of species, the number of age strata, and the gridsize used. To illustrate memory consumption lets use the advection example shown in Fig. 4. A Lagrangian approach would require the age, x, and y location of each of the the 1004 objects to be tracked individually, resulting in the storage of 3012 float values. Assuming an object lifetime of 24 minutes and a timestep of 12 minutes, as shown in the left subplot of Fig 4, the binomial memory footprint would be $25 \cdot 2 \cdot 2 = 50$ integer values (25 gridboxes, 2 species, 2 age strata). Reducing the timestep to 1 minute while retaining a 24 minute lifetime would increase the memory usage to 1200 integers. An advantage of the discrete framework is that the memory required is static and evenly spread over the grid, which means it can be easily spatially decomposed into individual blocks with the rest of the atmosphere model to be run in parallel. In contrast, the memory usage of Lagrangian approaches grows and shrinks with the number of particles tracked, and particles moving from one memory domain to the other can complicate the parallelization process.

4 Simple applications

In this section the framework is further explored by means of simple experiments with four possible configurations, as applied to grids of small size ("microgrids"). The purpose is not to define ultra-realistic systems; instead, the goal is to explore basic behavior and highlight opportunities. Achieving a realistic configuration and calibration, including the use of observational datasets, is for now considered a future research topic. Most examples are loosely inspired by atmospheric convection, which is reflected in the definition of the species.

Table 1. Configuration of the four BiOMi experiments discussed in Section 4. Note that Exp 2 is an exception in that it is non-dimensional, age is neglected, and birthrates are derived from differential equations as explained in Subsection 4.2

Setting	Unit	Exp 1	Exp 2	Exp 3	Exp 4
Gridsize		1×1	1×1	15×15	100×100
					1000×1000
$\Delta x, \Delta y$	[m]	5000	1	100	100
L	[m]	1000000	5	1000000	1000000
Δt	[s]	60	1/10	60	60
I		10	2	5	1
τ_i	[s]	60	-	600	600
K_i		1	-	10	5
$\dot{B}_i$	[m^{-2} s^{-1}]	$\propto (100 \cdot i - 50)^{-2}$	$\dot{B}_1 = g(n_1, n_2)$	$\dot{B}_5 = 5 \cdot 10^{-6}$	$\dot{B}_1 = 2 \cdot 10^{-7}$
			$\dot{B}_2 = f(n_1, n_2)$		
Interactions		None	Inter-species	Inter-species	Spatial
$(u, v)_{\mathrm{adv}}$	[m s^{-1}]	$(0, 0)$	$(0, 0)$	$(0.3, 0.2)$	$(0, 0)$
r_f	[m]	-	-	-	300
C_f		-	-	-	2000

The framework as applied on microgrids is hereby named *BiOMi* (Binomial Objects on Microgrids). Using microgrids keeps the examples discussed in this section as simple and easy to understand as possible. But another important motivation for using microgrids is the associated high computational efficiency, which could allow its application as part of a convection scheme in operational general circulation models used for weather forecasting and climate prediction.

4.1 Exp 1: Single-column random sampler

The first experiment demonstrates how the BiOMi framework can be used to introduce stochastic noise in existing convection schemes in operational weather and climate models. Spectral convection schemes are perhaps best suited to this purpose. This class of convective parameterizations has been around since the early days of numerical weather forecasting (Arakawa & Schubert, 1974). A key assumption at the foundation of spectral schemes is the shape of the size distribution of convective elements that do the vertical transport. In the convective grey zone stochastic noise can be superimposed onto this spectrum to represent the impact of subsampling of the population (Neggers, 2015), for which the binomial number generator as proposed in this study can well be used.

As a demonstration a discretized spectrum of convective objects is considered, consisting of a histogram with 10 bins ranging linearly in size from 50 to 950 m. The reference birth rate of the objects is a power law of of object size with a slope of -2,

$$\dot{B}_i = \lambda \left(100 \cdot i - 50\right)^{-2} . \tag{23}$$

The proportionality constant λ is scaled such that the birth rate is on average 256 per gridbox for the 50 m objects. A 1×1 grid is adopted with a grid spacing of 5 km, which is in the middle of the deep convective grey zone (Arakawa et al., 2011). The reference domain is 1000 km, and the object distribution is sampled 50 times independently of each other to evaluate the stochasticity. In these 50 random samplings only the three smallest and most numerous object species are always present (Fig. 6), with the ratio of subsampling variance to mean number becoming larger for the rarer object species. This dependence of the stochasticity on size follows the implied behaviour as discussed in Section 3.1.

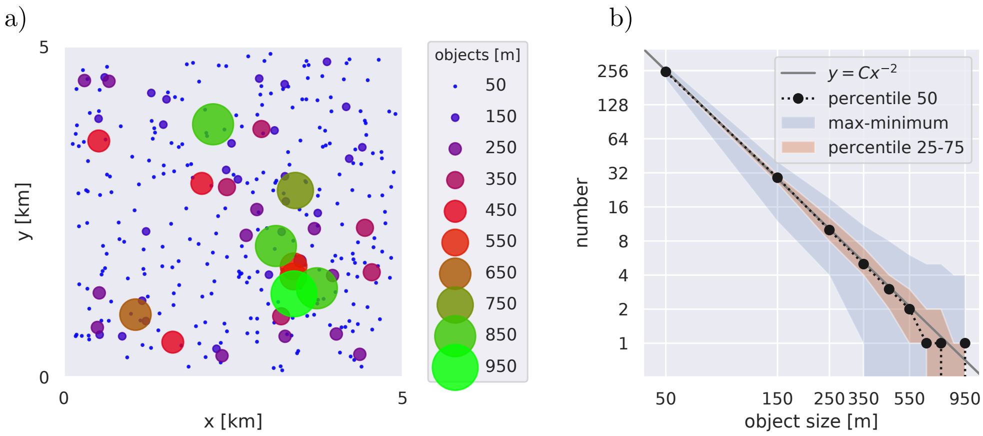

Figure 6. a) Scatter plot illustrating all objects of one of the 50 samples included in subplot b). The x and y position of each object is randomized for visualization. b) Object size distribution statistics of 50 random samplings of objects with decreasing birth rates as detailed in 4.1 with the parameters listed in 1.

This simple "offline" experiment thus shows how the binomial framework introduced in this paper can introduce not only scale-awareness and scale-adaptivity in a spectral convection scheme (through dependence on the grid spacing), but also stochasticity due to population subsampling in the grey zone. At the same time, the average number of objects over the grid is preserved.

4.2 Exp 2: Stochastic predator-prey system

This experiment is a translation of the continuous predator-prey system of Lotka (1910, 1920); Volterra (1926) to a discrete analog in which births and deaths are determined from Bernoulli trials. The intent of this experiment is to highlight the stochastic nature and to illustrate how the individual species can interact while conserving their discreteness. The predator-prey system was chosen as it a widely known problem that has been intensively studied in regards to stochasticity (Aguirre et al., 2013) and previously translated to a system of stochastic cellular automata by Guinot (2002) who studied under which conditions the behaviour of the cellular automata matches that of the continuous equations. Predator-prey approaches have also been used in Meteorology to describe cloud microphysics (Wacker, 1995) and cloud precipitation interactions (Koren & Feingold, 2011; Pujol & Jensen, 2019).

According to the classic formulation of the predator-prey equations, the prey x grows exponentially with a rate of α but is reduced by the hunting of the predator y which kills according to the product of prey and predator and β. The predator's growth is linked to the amount of hunting through δ, and the predator dies off with an exponential decay of strength γ. The equations have a periodic solution around a stable point when the populations of prey and predator, as well as the four parameters, are all positive.

$$\frac{dx}{dt} = +\alpha x \quad -\beta xy \tag{24}$$

$$\frac{dy}{dt} = -\gamma y \quad +\delta\beta xy \tag{25}$$

To switch to our discrete framework we neglect the age dimension and only look at the total number of prey n_1 and predators n_2, which simplifies equation (2) to:

$$\Delta n_1 = b_1 - d_1, \qquad \Delta n_2 = b_2 - d_2. \tag{26}$$

Bernoulli trials are used to determine specific numbers of births and deaths over Δt by sampling from a N times larger reference domain with the probability $p = 1/N$ that each birth or death of the reference domain occurs in a specific gridbox:

$$b_1 = \mathcal{B}\left(\alpha n_1 \cdot N\Delta t, p\right) \qquad d_1 = \mathcal{B}\left(\beta n_1 n_2 \cdot N\Delta t, p\right), \tag{27}$$

$$b_2 = \mathcal{B}\left(\beta\delta n_1 n_2 \cdot N\Delta t, p\right) \qquad d_2 = \mathcal{B}\left(\gamma n_2 \cdot N\Delta t, p\right). \tag{28}$$

Due to the number of deaths being stochastic the populations can become negative, which we avoid by introducing a limiter. The introduced stochasticity breaks the even cycle of the continuous solution, visible in the peaks and dips of the discrete prey in the ensemble quickly dispersing in the example shown in Fig. 7. The discrete nature is most visible in the less populous predator population. Once the predator population reaches zero the predator is extinct and can no longer recover. Once extinction occurs the prey can grow exponentially, as visible in the straight lines leaving the plot domain in Fig. 7. Note that extinction can occur in the continuous formulation as well when stochastic perturbations are added (Aguirre et al., 2013). The prey can also go extinct, though it is rarer for the parameters and initial conditions we choose to show.

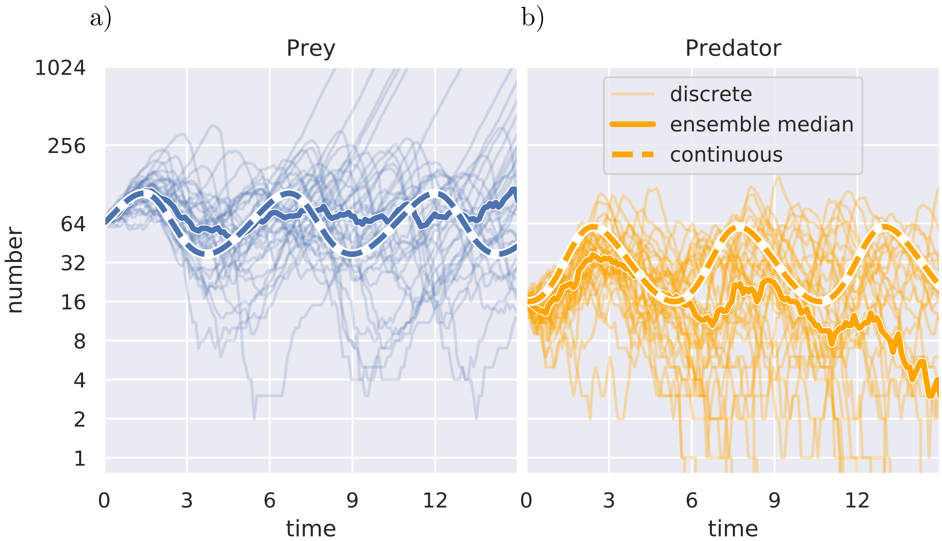

Figure 7. A 36 member of ensemble of the the predator prey system discussed in subsection 4.2 using the parameters $\alpha = 1$, $\beta = 0.03$, $\gamma = 1.5$, $\delta = 0.75$ for equation 24. Initial conditions are 64 (prey) and 16 (predator). Continuous solution is integrated numerically, discrete ensemble is generated using the values listed in table 1.

4.3 Exp 3: A down-scale energy cascade

In the third experiment the model is configured as an ecosystem consisting of five species, without spatial interaction. The goal of this simple experiment is to mimic the down-scale energy cascade typical of atmospheric turbulence (Kolmogorov, 1941a, 1941b; Frisch, 1995). To this purpose each species represents an individual size-class of turbulent structures. Only the largest species experiences births, which is conform the idea that the turbulent energy in an unstable turbulent layer is injected at the largest possible scale. At the end of its life-cycle the object then breaks up into two objects of half its size, which are injected as births in the species-category one size-class smaller,

$$b_{i1}^{\text{casc}} = 2\ d_{i+1,10}, \tag{29}$$

where we used that $K_i = 10$ for all species. This additional birth process is added to the default birth term b_{i1} in budget (2). This process is applied at all scales (species), which in effect establishes a simple form of species interaction in down-scale direction across the spectrum. This process is analogous to the flow of energy across the inertial subrange in turbulence. When an object of the smallest species dies it is simply removed

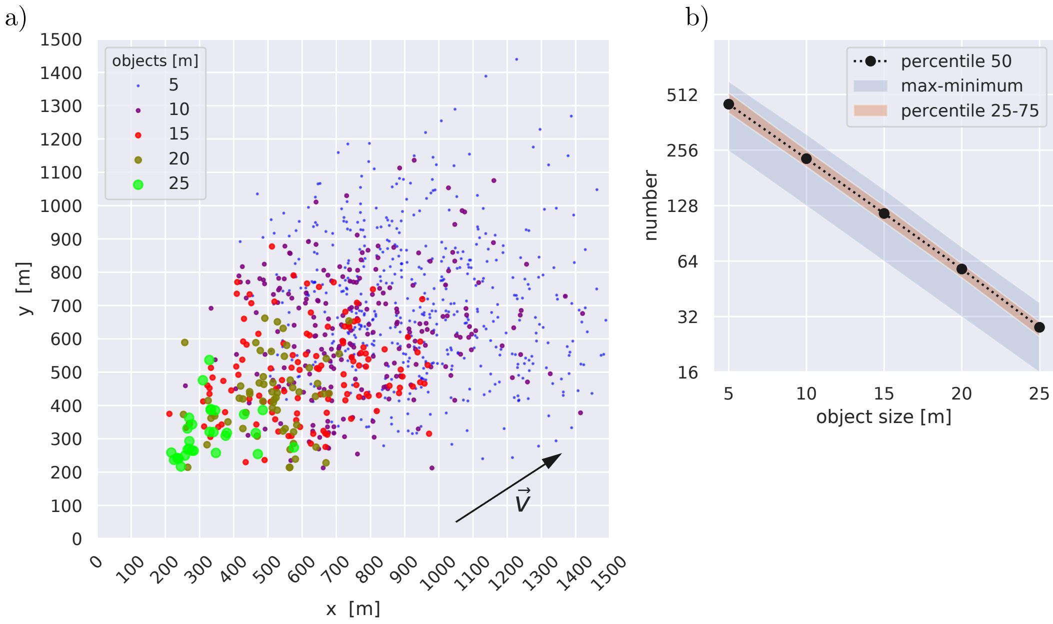

Figure 8. a) Snapshot during an experiment with BiOMi in the five-species energy-cascade configuration as described in Section 4.3 with an arrow showing the wind speed direction advecting the objects. The number of each species per gridbox is shown, with each species having a different size and color. The position of each object within the gridbox is randomized, for visualization. b) Associated size density of object number. The y-axis is plotted in log scale to highlight exponential dependency. The 25-75% range is shaded in red, the maximum and minimum range in blue, and the median is shown as a dotted black line.

from the grid, a process analogous to viscous dissipation of turbulent kinetic energy at molecular scales.

To give the experiment another twist, the births of the largest size-class ($i = 5$) are only allowed to occur in a single specific gridbox $(3, 3)$. For all other species, $\dot{B}_i = 0$ everywhere on the grid. This means the other (smaller) species can only form through the cascade process described by (29). In addition, a weak mean wind is applied, so that the objects are slowly advected in the direction marked by the arrow in Fig. 8. As a result of the advective diffusion illustrated in Subsection 3.2, the population starts to resemble a widening plume initiated at a fixed location and being advected down-wind. This could be a chimney, a forest-fire, or a convective cell creating a slowly dissipating outflow or anvil cloud. All other settings of the BiOMi model as used for this five-species experiment are summarized in Table 1.

Figure 8a shows a snapshot of the population of objects during this experiment, an animation of which is also provided as a digital supplement to this paper (Supporting Information). Similar to Exp 1 multiple species are present, but they now cover multiple gridboxes. The results highlight the stochastic nature of both object birth and advection. The largest objects (green) are born in a single gridbox. As they age, they are advected by the mean wind, but also break up into two objects half their size (red) when they complete their life-cycle. This process continues across multiple life-cycles. As a result, the distance from the birthing-gridbox becomes proportional to age, on average. However, because advective movement contains a random element, this creates a spreading plume of particles that "dissipates" when the life cycle of the smallest objects has been completed. Figure 8b shows the associated size density of object number, which carries a clear exponential dependence. Such exponential functionality in the spectrum is typical of a turbulent energy cascade. The spread in object number is caused by the stochastic birth rate and also decreases exponentially with size (i.e. it is constant on the logarithmic y-axis). This reflects that all objects have the same life span.

4.4 Exp 4: Spatial organization in a single-species population

The fourth experiment considers only a single species, here assumed to represent the smallest building block of convection: the short-lived bubble or thermal (Scorer & Ludlam, 1953; Hernandez-Deckers & Sherwood, 2016; Morrison & Peters, 2018). Simple rules of spatial interaction are introduced to let thermals respond to each other's presence, by which they can collaborate or compete to let larger-scale coherent convective structures self-organize and emerge on the grid. This behavior introduces convective memory that acts on time-scales much longer than the life-time of individual objects. The use of such rules is known from cellular automata, there often referred to as "transition rules" (Gardner, 1970; Bengtsson et al., 2011).

Two rules of interaction are adopted, both working through the probability field p. These rules reflect atmospheric physics and dynamics, and are inspired by the recent study by (Böing, 2016). The first rule reflects the "pulsating growth" behavior as observed in individual shallow cumulus clouds in nature, consisting of a series of subsequent individual pulses (Anderson, 1960; French et al., 1999; Heus et al., 2009). The idea is that the first pulse breaks down pre-existing instability, favoring subsequent thermals to thrive and thus form "thermal-chains" (Blyth & Latham, 1993; Damiani et al., 2006;

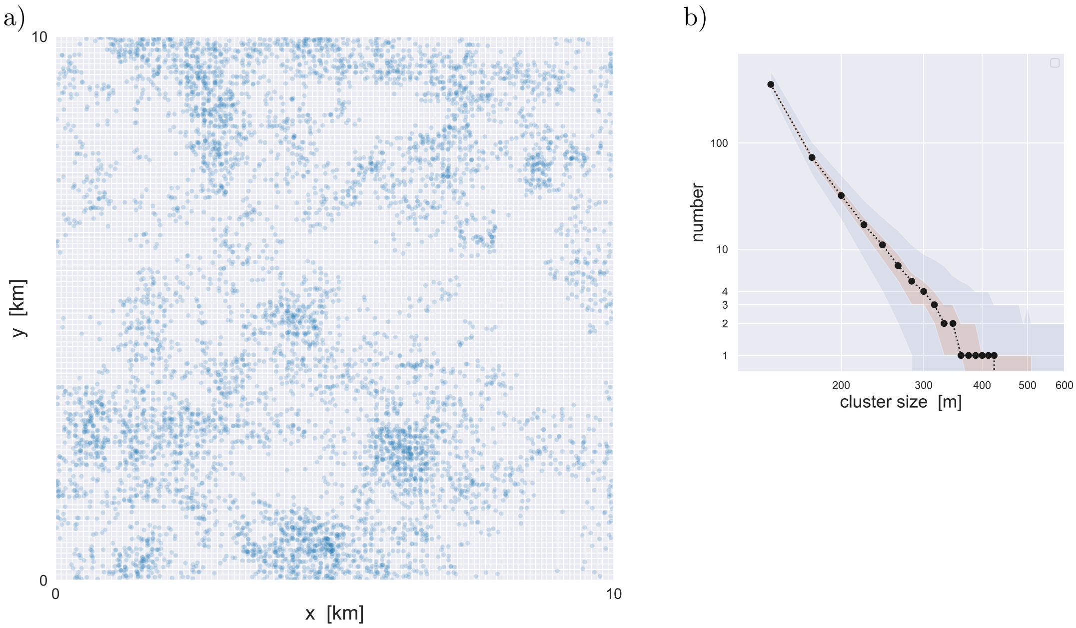

Figure 9. a) Snapshot during an experiment with BiOMi in the single-species configuration with two rules of interaction between objects, as described in Section 4.4. The position of each object within the gridbox is randomnized, for visualization. The opacity of each object is 0.2, to highlight clusters. b) Associated size density of cluster number. Log scale is used on both axes for highlighting powerlaw dependency. The 1-99% and 25-75% ranges are shaded blue and red, respectively, while the median is shown as dotted black.

Varble et al., 2014). On a microgrid this behavior can simply be introduced by perturbing the p field at locations where objects already exist. The perturbation-field p'_i surrounding a single gridpoint containing n_{ik} objects could be modeled as follows,

$$p'_i \;=\; C_f \; f_p \; \sum_k n_{ik} \tag{30}$$

where f_p is a two-dimensional spatial impact field of radius r_f. In this experiment f is assumed to be cone-shaped,

$$f_p \;=\; \begin{cases} 1 - r/r_f & \text{for} \quad r < r_f \\ 0 & \text{for} \quad r \geq r_f \end{cases} \tag{31}$$

where r is the distance to the gridpoint of interest, and C_f is a constant of proportionality carrying the efficiency of objects in modifying their environment. The perturbation field p'_i is calculated at every gridpoint and added to the spatially uniform reference probability $p\,=\,1/N$, yielding a new cumulative field p_c that can be spatially heterogeneous.

The second rule is a constraint on the perturbed p field which ensures that averaged over the whole grid the mean birth rate always equals $\dot{B}_i$. To this purpose the new cumulative probability field including all perturbations, p_c, is suitably normalized,

$$p \;=\; \frac{1}{N}\,\frac{p_c}{\langle p_c\rangle}, \tag{32}$$

where the brackets indicate the average over the grid. Comparison to (6) shows that the grid-dependent probability $1/N$ is multiplied by a spatially varying factor. This means that while on average the birth rate of the number of objects on the grid B_i remains controlled by external forcings, locally strong deviations can develop in the p field. In effect, this reduces the probability p in areas where few objects are present. This behavior can loosely be interpreted as environmental deformation caused by convective objects through for example gravity waves and compensating subsidence (Bretherton & Smolarkiewicz, 1989).

The model settings for this single-species experiment are also summarized in Table 1. An important difference with the third experiment is that the mean wind is zero, so that objects stay quarantined in their gridbox. In addition, object births are not limited to a specific single gridbox but can freely occur everywhere on the grid. Thermal size is implicitly assumed to be on the order of the grid-spacing (~ 100 m). As a result, any coherent spatial structures resulting from object interactions can be resolved. The thermals are short-lived while their spatial impact does not exceed beyond $3\times$ their size. As a consequence, thermals have to cooperate to let larger-scale structures emerge on the grid.

Animations of Exp 4 for two gridsizes are provided as digital supplements to this paper (Supporting Information). Figure 9a shows a snapshot of the 100×100 gridsize experiment at 13 hours after initialization. At this time spatial organization is apparent in the population, featuring dense clusters but also areas that are almost free of objects. In those areas the probability of birth is very low. By eye this spatial distribution including both dense and sparsely populated areas is not unlike the organization visible in high-resolution satellite images of Trade wind cumulus cloud populations (Bony et al., 2020).

Figure 10 shows results from a cluster analysis of this population, using the density-based GRIDCLUS algorithm (Schikuta, 1996). The clustering threshold is $n > 1$, meaning that only gridboxes are included that have two or more objects in them. Figure 10a

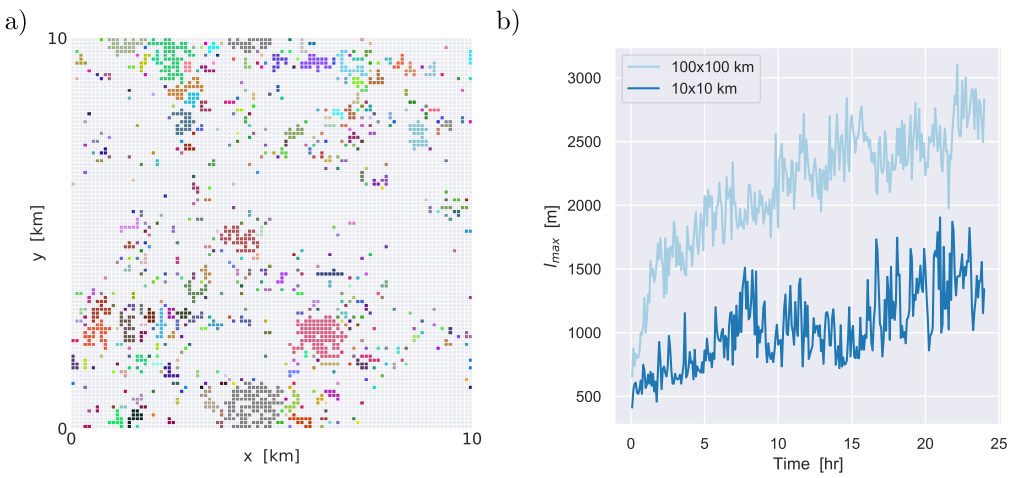

Figure 10. Results of cluster analysis using threshold $n_1 \geq 2$. a) Spatial distribution of the clusters at the last timestep for the experiment with the 10x10 km domain (100x100 gridsize). Each cluster is assigned a unique color. b) Time evolution of the size of the largest cluster on the grid. Results with two domain sizes are shown, 10x10 km (dark blue) and 100x100 km (light blue).

shows the resulting clusters on the grid, while Fig. 9b shows the associated size density of cluster number, with size calculated as the square root of the cluster area. In contrast to Exp 3 a clear powerlaw dependency is apparent, featuring a negative exponent. This means that small clusters are very frequent and big clusters are rare. Such powerlaw scaling is frequently observed for shallow cumulus cloud fields in nature (Benner & Curry, 1998; Neggers et al., 2003; Wood & Field, 2011). The widening spread at large cluster sizes shows that the clusters at those sizes become subsampled, which is a defining feature of the convective grey zone (Neggers et al., 2019).

Another important aspect of the clustering behavior is highlighted by Fig. 10b, showing convective memory on the grid as expressed by the time evolution of the size of the largest cluster, l_{max}. Two gridsizes are compared, one with a mesoscale domain size ($D =$ 10 km) and one with a macroscale domain ($D = 100$ km). Both domains feature a gradual increase in l_{max}. However, on the mesoscale domain the growth of l_{max} is markedly slower, featuring temporary peaks and failing to grow beyond 1.5 km. This suggests the cluster growth becomes limited by the domain size. This is not the case for the macroscale domain, where growth is unimpeded and follows a parabolic evolution (see also the provided animation). What these results suggest is that under simple rules of interaction, convective memory can be created and carried on the grid. Introducing this behavior in convective parameterizations is a long standing ambition that has not yet been achieved (Khairoutdinov & Randall, 2006; Grabowski et al., 2006). If population models on two-dimensional microgrids can solve this problem is a future research topic.

5 Discussion

5.1 Limitations

The formulation of the framework contains a few important limitations. These were consciously introduced, in order to explore a system that is as low-complexity and transparent as possible. However, it is important to consider these limitations and their impact on the results. In addition, possible future modifications can be considered that might make the system better reflect realistic conditions.

The first limitation is the assumption of a constant object birth rate $\dot{B}_i$ which is sufficient for the purposes of this study. However, what external factors control this birth rate remains a fundamental question and depends strongly on the definition of the species to be represented by the model. In the case of surface-driven convection in a viscous fluid, the number of plumes has been observed to depend on the heating rate at the surface, as expressed by the surface Rayleigh number (Zhong, 2005). Dependence of object birth rates on thermodynamic conditions can be investigated using large-eddy simulations, for example for convective cloud populations (Garrett et al., 2018). Such dependencies can easily be implemented in this framework.

The choice to adopt a discrete formulation introduces opportunities but also makes the framework less flexible in some regards. For example, the object lifespan must be a multiple of the timestep, which suggests that adaptive time-stepping would no longer be possible. However, this could be remedied by applying separate timestepping for the microgrid.

The use of the binomial advection operator introduces some numerical diffusion which is an unavoidable side effect of any Eulerian advection scheme. The strength and direction of the diffusion is dependent on the horizontal gradients, grid spacing, timestep, and the angle between grid orientation and wind. To achieve a controlled and consistent diffusion one could easily combine the advection operator with aspects of the classic Gaussian plume model (Sutton, 1932) that is often used to model dispersion in the atmosphere.

The rules of interaction between convective objects as adopted in Exp 4 are still very simple. While being successful in demonstrating opportunities, important interactions acting in atmospheric moist convection in nature are still missing. These include i) latent heat effects due to cloud formation, ii) impacts of wind shear on spatial organization, iii) formation of cold pools due to evaporation of precipitation. Additional rules can well be added in the system. But before introducing such rules they should be carefully calibrated and trained against relevant datasets, for example using machine learning techniques.

5.2 Comparisons to other stochastic frameworks

The BiOMi framework as applied in the previous section shares some features with other recently proposed population models, but also differs in some key aspects. These similarities, differences and novelties are briefly highlighted here, for reference.

The STOMP framework (STOchastic Model for Population dynamics of convective clouds, Hagos et al. (2018)) is at its core also discrete and stochastic, consisting of size distributions of convective cells that interact by exchanging "convective pixels". In contrast to BiOMi's predetermined number *species* that can represent differing convective objects, STOMP is explicitly defined in terms of cloud size distributions. BiOMi also differs fundamentally by the inclusion of an explicit age dimension, the use of binomial sampling to determine births and advection, and the possibility to use a microgrid spatially. As a result, objects in BiOMi can overlap, allowing in principle the representation of thermal chains that are oriented vertically, as illustrated in Exp4.

Recent studies by Stechmann and Hottovy (2016) and Khouider and Bihlo (2019) proposed stochastic models based on principles from statistical mechanics that represent convective regimes as phase transitions. BiOMi adheres to this principle, in that spatial patterns associated with convective regimes can freely emerge on the grid under certain rules of transition. A key conceptual difference concerns the main stochastic budget equation; while these models use integrated humidity, BiOMi considers the evolution of object number. These interacting objects can also freely move around on the grid, taking object demographics into account as an additional dimension. This in effect combines an object-based approach with a microgrid approach, which is a novelty. The representation of horizontal movement is another difference, which in BiOMi takes place through stochastic advection instead of stochastic diffusion. Finally, the rules of transition reflect different processes. While in the above studies the rules reflect behavior of cloudy areas as embedded in open- or closed cell stratocumulus, in BiOMi Exp4 the rules reflect the physics and dynamics of individual sub cloud-scale convective thermals in fair-weather cumulus cloud fields.

A cloud population model with a stochastic scale-aware birthrate very similar to that of BiOMi was developed by Sakradzija et al. (2015) for use in a shallow convection scheme (Sakradzija et al., 2016; Sakradzija & Klocke, 2018). In their approach the cloud birth rates are sampled from a Poisson distribution instead of a binomial, and further differs from BiOMi in that each cloud has an individual continuous duration and there are no fixed species. For a high number of clouds their approach requires a large amount of memory as the birth time and duration of each cloud is saved individually.

6 Conclusions and outlook

In this study a computationally efficient stochastic binomial framework is formulated for representing discrete populations of objects on a two-dimensional grid. A defining feature of the BiOMi framework (Binomial Objects on Microgrids) is its binomial number generator based on a Bernoulli process. This stochastic and scale-aware operator is applied to both object birth and object advection, by which discreteness in object number is preserved in both processes. A discrete prognostic budget for object number is combined with an age dimension, allowing representation of life-cycle effects and object demographics. In addition, multiple co-existing species can be represented, making the framework suitable for multiple modes of application. Interactions between objects can be introduced in various ways, by adopting concepts from game theory and cellular automata. Finally, due to its reliance on binomial sampling the BiOMi system is also computationally cheap to operate.

The BiOMi framework is tested and explored in various simple configurations, designed to reflect key aspects of atmospheric turbulence and convection. This yielded the following main conclusions:

- The binomial number generator is effective in introducing stochasticity in object number due to population subsampling in the convective grey zone;
- The binomial operator also introduces stochasticity in advection of objects between gridboxes;
- The framework can successfully reproduce key characteristics of the classic predator-prey problem while preserving discreetness and introducing stochastic variations;
- Behavior as observed in nature can be reproduced by the system, including i) the down-scale energy cascade in atmospheric turbulence, and ii) spatial organization in convective cloud populations resulting from interactions between objects;
- The arrangement of binomially generated populations on a microgrid is a form of convective memory, evolving on timescales much longer than the lifespan of individual objects;
- The computational efficiency is high enough to allow application as part of convection schemes in operational weather and climate models.

While the framework has many possible applications, its potential use as part of a convective parameterization for weather and climate models has always been a primary motivation behind this study. These opportunities are further explored in an ongoing related study, in which the BiOMi system as applied to a population of single-sized, short-lived but interacting convective thermals as explored in Exp 4 is implemented in a discretized spectral convection scheme (ED(MF)^n, Neggers (2015)). BiOMi then acts to provide cluster size densities that emerge on its microgrid, replacing one of the existing closures at the foundation of the scheme. In effect, this equips ED(MF)^n with subgrid convective memory and introduces awareness of spatial organization - both longstanding bottlenecks in convective parameterization. For testing the ED(MF)^n-BiOMi system is implemented as a subgrid transport scheme in a simplified circulation model and explored for prototype cumulus cases. Impacts on the onset of precipitation in diurnal cycles of continental convection are investigated, as well as behavior in the range of resolutions spanning the convective grey zone.

BiOMi offers further opportunities when applied within GCM gridboxes. Firstly, existing convection schemes can be equipped with the 1D random sampler as explored in Exp 1 to introduce stochastic noise in the grey zone. Secondly, the microgrid can be used to make surface-atmosphere interactions more sophisticated. For example, awareness of small-scale surface heterogeneity can be introduced by coupling the BiOMi microgrid to similarly high-resolution maps of surface properties. Convective triggering can then respond in areas which are known to affect this process, such as mountains or areas of different vegetation.

Acknowledgments

This research was supported by the U.S. Department of Energy's Atmospheric System Research, an Office of Science Biological and Environmental Research program, under grant DE-SC0017999. We thank Brian Mapes, Timothy Garrett and Steve Sherwood for discussions on early versions of the binomial framework, and Andreas Griewank for advice related to the framework definitons. The BiOMi code is publicly accessible through GitHub at *https://github.com/pgriewank/BioMi*. Animations of Exp 3 and Exp 4 with the BiOMi framework as discussed in Section 3 are provided as Supporting Information to this publication.

References

Aguirre, P., González-Olivares, E., & Torres, S. (2013). Stochastic predator–prey model with allee effect on prey. *Nonlinear Analysis: Real World Applications*, *14*(1), 768 - 779. Retrieved from `http://www.sciencedirect.com/science/article/pii/S1468121812001575` doi: https://doi.org/10.1016/j.nonrwa.2012.07.032

Anderson, C. E. (1960). A Study of the Pulsating Growth of Cumulus Clouds. *Dissertation. Massachusetts Institute of Technology*.

Arakawa, A., Jung, J.-H., & Wu, C.-M. (2011). Toward unification of the multiscale modeling of the atmosphere. *Atmospheric Chemistry and Physics*, *11*(8), 3731–3742. Retrieved from `https://www.atmos-chem-phys.net/11/3731/2011/` doi: 10.5194/acp-11-3731-2011

Arakawa, A., & Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.*, *31*, 674-701.

Bengtsson, L., Körnich, H., Källén, E., & Svensson, G. (2011). Large-scale dynamical response to subgrid-scale organization provided by cellular automata. *Journal of the Atmospheric Sciences*, *68*(12), 3132-3144. doi: 10.1175/JAS-D-10-05028.1

Benner, T. C., & Curry, J. A. (1998). Characteristics of small tropical cumulus clouds and their impact on the environment. *Journal of Geophysical Research: Atmospheres*, *103*(D22), 28753-28767. doi: 10.1029/98JD02579

Blyth, A. M., & Latham, J. (1993). Development of ice and precipitation in New Mexican summertime cumulus clouds. *Quarterly Journal of the Royal Meteorological Society*, *119*(509), 91-120. doi: 10.1002/qj.49711950905

Böing, S. J. (2016). An object-based model for convective cold pool dynamics. *Mathematics of Climate and Weather Forecasting*(1). doi: 10.1515/mcwf-2016-0003

Bony, S., Schulz, H., Vial, J., & Stevens, B. (2020). Sugar, gravel, fish, and flowers: Dependence of mesoscale patterns of trade-wind clouds on environmental conditions. *Geophysical Research Letters*, *47*(7), e2019GL085988. doi: 10.1029/2019GL085988

Brast, M., Schemann, V., & Neggers, R. A. J. (2018). Investigating the Scale Adaptivity of a Size-Filtered Mass Flux Parameterization in the Gray Zone of Shallow Cumulus Convection. *Journal of the Atmospheric Sciences*, *75*(4), 1195-1214. doi: 10.1175/JAS-D-17-0231.1

Bretherton, C. S., & Smolarkiewicz, P. K. (1989). Gravity Waves, Compensating Subsidence and Detrainment around Cumulus Clouds. *Journal of the Atmospheric Sciences*, *46*(6), 740-759. doi: 10.1175/1520-0469(1989)046⟨0740:GWCSAD⟩2.0.CO;2

Damiani, R., Vali, G., & Haimov, S. (2006). The Structure of Thermals in Cumulus from Airborne Dual-Doppler Radar Observations. *Journal of the Atmospheric Sciences*, *63*(5), 1432-1450. doi: 10.1175/JAS3701.1

Dorrestijn, J., Crommelin, D. T., Biello, J. A., & Böing, S. J. (2013). A data-driven multi-cloud model for stochastic parametrization of deep convection. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1991), 20120374. doi: 10.1098/rsta.2012.0374

Feingold, G., Balsells, J., Glassmeier, F., Yamaguchi, T., Kazil, J., & McComiskey, A. (2017). Analysis of albedo versus cloud fraction relationships in liquid water clouds using heuristic models and large eddy simulation. *Journal of Geophysical Research: Atmospheres*, *122*(13), 7086-7102. doi: 10.1002/2017JD026467

French, J. R., Vali, G., & Kelly, R. D. (1999). Evolution of small cumulus clouds in

Florida: observations of pulsating growth. *Atmospheric Research*, *52*, 143-165. doi: doi:10.1016/S0169-8095(99)00024-1

Frisch, U. (1995). *Turbulence: The legacy of a. n. kolmogorov.* Cambridge University Press. doi: 10.1017/CBO9781139170666

Gardner, M. (1970). Mathematical Games - The Fantastic Combinations of John Conway's New Solitaire Game 'Life'. *Scientific American*, *223*, 120–123. doi: 10.1038/scientificamerican1070-120

Garrett, T. J., Glenn, I. B., & Krueger, S. K. (2018, Aug). Thermodynamic constraints on the size distributions of tropical clouds. *Journal of Geophysical Research: Atmospheres*, *123*(16), 8832–8849. Retrieved from `http://dx.doi.org/10.1029/2018JD028803` doi: 10.1029/2018jd028803

Grabowski, W. W., Bechtold, P., Cheng, A., Forbes, R., Halliwell, C., Khairoutdinov, M., ... Xu, K.-M. (2006). Daytime convective development over land: A model intercomparison based on LBA observations. *Quarterly Journal of the Royal Meteorological Society*, *132*(615), 317-344. doi: 10.1256/qj.04.147

Guinot, V. (2002, 06). Modelling using stochastic, finite state cellular automata: Rule inference from continuum models. *Applied Mathematical Modelling*, *26*, 701-714. doi: 10.1016/S0307-904X(01)00078-6

Hagos, S., Feng, Z., Plant, R. S., Houze Jr., R. A., & Xiao, H. (2018). A stochastic framework for modeling the population dynamics of convective clouds. *Journal of Advances in Modeling Earth Systems*, *10*(2), 448-465. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001214` doi: 10.1002/2017MS001214

Hernandez-Deckers, D., & Sherwood, S. C. (2016). A Numerical Investigation of Cumulus Thermals. *Journal of the Atmospheric Sciences*, *73*(10), 4117-4136. doi: 10.1175/JAS-D-15-0385.1

Heus, T., Jonker, H. J. J., Van den Akker, H. E. A., Griffith, E. J., Koutek, M., & Post, F. H. (2009). A statistical approach to the life cycle analysis of cumulus clouds selected in a virtual reality environment. *Journal of Geophysical Research: Atmospheres*, *114*(D6). doi: 10.1029/2008JD010917

Honnert, R., Efstathiou, G. A., Beare, R. J., Ito, J., Lock, A., Neggers, R., ... Zhou, B. (2020). The atmospheric boundary layer and the "gray zone" of turbulence: A critical review. *Journal of Geophysical Research: Atmospheres*, *n/a*(n/a), e2019JD030317. doi: 10.1029/2019JD030317

Joseph, J. H., & Cahalan, R. F. (1990). Nearest neighbor spacing of fair weather cumulus clouds. *J. App. Meteor.*, *29*, 793-805.

Khairoutdinov, M., & Randall, D. (2006). High-resolution simulation of shallow-to-deep convection transition over land. *Journal of the Atmospheric Sciences*, *63*(12), 3421-3436. doi: 10.1175/JAS3810.1

Khouider, B., Biello, J., & Majda, A. J. (2010, 03). A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, *8*(1), 187-216.

Khouider, B., & Bihlo, A. (2019). A new stochastic model for the boundary layer clouds and stratocumulus phase transition regimes: Open cells, closed cells, and convective rolls. *Journal of Geophysical Research: Atmospheres*, *124*(1), 367-386. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029518` doi: 10.1029/2018JD029518

Kolmogorov, A. (1941a). Dissipation of energy in locally isotropic turbulence. *Dokl. Akad. Nauk SSSR*, *32*, 16–18.

Kolmogorov, A. (1941b). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk SSSR*, *30*, 301–305.

Koren, I., & Feingold, G. (2011). Aerosol–cloud–precipitation system as a predator-prey problem. *Proceedings of the National Academy of Sciences*, *108*(30), 12227–12232. Retrieved from `https://www.pnas.org/content/108/30/12227` doi: 10.1073/pnas.1101777108

Kwon, Y. C., & Hong, S.-Y. (2017). A Mass-Flux Cumulus Parameterization Scheme across Gray-Zone Resolutions. *Mon. Wea. Rev.*, *145*(2), 583–598. doi: 10.1175/MWR-D-16-0034.1

Lotka, A. (1910). Contribution to the Theory of Periodic Reaction. *J. Phys. Chem.*, *14*, 271-274.

Lotka, A. (1920). Analytical Note on Certain Rhythmic Relations in Organic Systems. *Proc. Natl. Acad. Sci. U.S.*, *6*, 410-415.

Morrison, H., & Peters, J. M. (2018). Theoretical Expressions for the Ascent Rate of Moist Deep Convective Thermals. *Journal of the Atmospheric Sciences*, *75*(5), 1699-1719. doi: 10.1175/JAS-D-17-0295.1

Nair, U. S., Weger, R. C., Kuo, K. S., & Welch, R. M. (1998). Clustering, randomness, and regularity in cloud fields: 5. the nature of regular cumulus cloud fields. *J. Geophys. Res.: Atmospheres*, *103*(D10), 11363–11380. doi: 10.1029/98jd00088

Neggers, R. A. J. (2015). Exploring bin-macrophysics models for moist convective transport and clouds. *Journal of Advances in Modeling Earth Systems*, *7*(4), 2079-2104. doi: 10.1002/2015MS000502

Neggers, R. A. J., Griewank, P. J., & Heus, T. (2019). Power-Law Scaling in the Internal Variability of Cumulus Cloud Size Distributions due to Subsampling and Spatial Organization. *Journal of the Atmospheric Sciences*, *76*(6), 1489-1503. doi: 10.1175/JAS-D-18-0194.1

Neggers, R. A. J., Jonker, H. J. J., & Siebesma, A. P. (2003). Size Statistics of Cumulus Cloud Populations in Large-Eddy Simulations. *Journal of the Atmospheric Sciences*, *60*(8), 1060-1074. doi: 10.1175/1520-0469(2003)60⟨1060:SSOCCP⟩2.0.CO;2

Peters, K., Crueger, T., Jakob, C., & Möbis, B. (2017). Improved MJO-simulation in ECHAM6.3 by coupling a Stochastic Multicloud Model to the convection scheme. *Journal of Advances in Modeling Earth Systems*, *9*(1), 193-219. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016MS000809 doi: 10.1002/2016MS000809

Pujol, O., & Jensen, A. (2019). Cloud–rain predator–prey interactions: Analyzing some propertiesof the koren–feingold model and introduction of a new species-competition bulk system with a hopf bifurcation. *Physica D: Nonlinear Phenomena*, *399*, 86 - 94. Retrieved from http://www.sciencedirect.com/science/article/pii/S0167278918301520 doi: https://doi.org/10.1016/j.physd.2019.04.007

Sakradzija, M., & Klocke, D. (2018). Physically constrained stochastic shallow convection in realistic kilometer-scale simulations. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2755-2776. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001358 doi: 10.1029/2018MS001358

Sakradzija, M., Seifert, A., & Dipankar, A. (2016). A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, *8*(2), 786-812. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/

2016MS000634 doi: 10.1002/2016MS000634

Sakradzija, M., Seifert, A., & Heus, T. (2015). Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Processes in Geophysics*, *22*(1), 65–85. Retrieved from https://www.nonlin-processes-geophys.net/22/65/2015/ doi: 10.5194/npg-22-65-2015

Schikuta, E. (1996). Grid-clustering: an efficient hierarchical clustering method for very large data sets. *Proceedings of the 13th International Conference on Pattern Recognition*, *2*, 101-105.

Scorer, R. S., & Ludlam, F. H. (1953). Bubble theory of penetrative convection. *Quarterly Journal of the Royal Meteorological Society*, *79*(339), 94-103. doi: 10.1002/qj.49707933908

Sengupta, S. K., Welch, R. M., Navar, M. S., Berendes, T. A., & Chen, D. W. (1990). Cumulus Cloud Field Morphology and Spatial Patterns Derived from High Spatial Resolution Landsat Imagery. *Journal of Applied Meteorology*, *29*(12), 1245-1267. doi: 10.1175/1520-0450(1990)029⟨1245:CCFMAS⟩2.0.CO;2

Stechmann, S. N., & Hottovy, S. (2016). Cloud regimes as phase transitions. *Geophysical Research Letters*, *43*(12), 6579-6587. doi: 10.1002/2016GL069396

Sutton, O. G. (1932). A theory of eddy diffusion in the atmosphere. *Proc. Roy. Soc. London, A*, *135*, 143-165.

Tompkins, A. M., & Semie, A. G. (2017). Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *Journal of Advances in Modeling Earth Systems*, *9*(2), 1046-1068. doi: 10.1002/2016MS000802

Varble, A., Zipser, E. J., Fridlind, A. M., Zhu, P., Ackerman, A. S., Chaboureau, J.-P., ... Shipway, B. (2014). Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations: 1. Deep convective updraft properties. *Journal of Geophysical Research: Atmospheres*, *119*(24), 13,891-13,918. doi: 10.1002/2013JD021371

Vogel, R., Nuijens, L., & Stevens, B. (2016). The role of precipitation and spatial organization in the response of trade-wind clouds to warming. *Journal of Advances in Modeling Earth Systems*, *8*(2), 843-862. doi: 10.1002/2015MS000568

Volterra, V. (1926). Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei Roma*, *2*, 31-113.

von Neumann, J. (1928). Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, *100*, 295-320. doi: 10.1007/BF01448847

von Neumann, J. (1966). Theory of Self-Reproducing Automata, A.W. Burks (Ed.). *University of Illinois Press, Urbana, IL.*

von Neumann, J., & Morgenstern, O. (1944). Theory of games and economic behavior. *Princeton University Press*.

Wacker, U. (1995). Competition of precipitation particles in a model with parameterized cloud microphysics. *Journal of the Atmospheric Sciences*, *52*(14), 2577-2589. Retrieved from `https://doi.org/10.1175/1520-0469(1995)052<2577:COPPIA>2.0.CO;2` doi: 10.1175/1520-0469(1995)052⟨2577:COPPIA⟩2.0.CO;2

Weger, R. C., Lee, J., Zhu, T., & Welch, R. M. (1992). Clustering, randomness and regularity in cloud fields: 1. theoretical considerations. *Journal of Geophysical Research: Atmospheres*, *97*(D18), 20519-20536. doi: 10.1029/92JD02038

Wing, A. A., Reed, K. A., Satoh, M., Stevens, B., Bony, S., & Ohno, T. (2018). Radiative–convective equilibrium model intercomparison project. *Geoscientific Model Development*, *11*(2), 793–813. Retrieved from `https://www.geosci-model-dev.net/11/793/2018/` doi: 10.5194/gmd-11-793-2018

Wood, R., & Field, P. R. (2011). The Distribution of Cloud Horizontal Sizes. *Journal of Climate*, *24*(18), 4800-4816. doi: 10.1175/2011JCLI4056.1

Wyngaard, J. C. (2004). Toward Numerical Modeling in the "Terra Incognita". *Journal of the Atmospheric Sciences*, *61*(14), 1816-1826. doi: 10.1175/1520-0469(2004)061⟨1816:TNMITT⟩2.0.CO;2

Zhong, S. (2005, 07). Dynamics of thermal plumes in three-dimensional isoviscous thermal convection. *Geophysical Journal International*, *162*(1), 289-300. doi: 10.1111/j.1365-246X.2005.02633.x

Supporting Information for "A binomial stochastic framework for efficiently modeling discrete statistics of convective populations"

R. A. J. Neggers[1] and P. J. Griewank[1],

[1]Institute for Geophysics and Meteorology, University of Cologne, Germany

Contents of this file

Additional Supporting Information (Files uploaded separately)

1. Captions for Movies S1 to S3

Introduction The Supporting Information (SI) provided with this manuscript consists of three animations of simulations with the BiOMi framework. More precisely, one animation (S1) shows results for the Exp3 experiment, while two animations (S2 and S3) correspond to the Exp4 experiment. The main manuscript includes various two-dimensional snapshots of the population of objects during these experiments (see Figs. 8, 9 and 10). These movies have the purpose of providing additional information about their time-evolution, which should help putting these snapshots into better perspective. The movies are designed to highlight object interactions as well as clustering behavior on the microgrid. The experiment settings are fully defined in Table 1 and the text in Section 4.

Corresponding author: R. A. J. Neggers, Institute for Geophysics and Meteorology, University of Cologne, Pohligstr. 3, 50969 Cologne, Germany. (neggers@meteo.uni-koeln.de)

Movie S1. The first animation (ms01.wmv) shows the evolution of objects on the microgrid during Exp 3 with the BiOMi framework, as described in Section 4.3 and defined in Table 1. A 15×15 grid is used, with a grid spacing of 100 m and a time step of 60 seconds. Five species are included, each representing a different size-class of convective objects. The largest objects are shown as light green dots, while the smaller objects are shown in dark green, red, purple and then blue for the smallest. The radius is proportional to the object size. At the end of a life cycle the objects break up into two smaller objects, thus in effect creating a down-scale energy cascade. Only the largest objects are born randomly, exclusively taking place in gridbox (3,3). A weak horizontal advection is applied, so that objects age as they move away from the point source.

Movie S2. The second movie (ms02.wmv) shows the evolution of objects on the microgrid during Exp 4 with the BiOMi framework, as described in Section 4.4 and defined in Table 1. A 100×100 grid is used, with a grid-spacing of 100 m and a time step of 60 seconds. This grid is referred to in the text as the mesoscale domain with horizontal size $D = 10$ km. Only a single species is included, representing a small convective bubble or thermal. The objects are born randomly at a fixed rate and spatially interact under two rules of transition, as described in detail in the text. In contrast to S1 no advection is applied, so objects stay in their gridbox. All objects on the grid are shown as filled circles with slightly reduced opacity. Their color indicates the number of objects in the gridbox, to highlight clustering. The duration of the experiment is 24 hours.

Movie S3. The third movie (ms03.wmv) is similar to the S2 movie, but shows the population of objects during the simulation of Exp4 using a 1000×1000 grid. This grid is referred to in the text as the macroscale domain with horizontal size $D = 100$ km. Note that this time the objects are not shown individually; instead, a two-dimensional mesh plot is used, with the color indicating the number of objects per gridbox. The duration of the experiment is also 24 hours.