JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2020MS002229

# A Binomial Stochastic Framework for Efficiently Modeling Discrete Statistics of Convective Populations

**Roel A. J. Neggers**[1] and **Philipp J. Griewank**[2]

[1]Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany, [2]Institut für Meteorologie und Geophysik, Universität Wien, Vienna, Austria

**Key Points:**

- A 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 object life cycle effects
- Population subsampling effects in the convective gray zone are reproduced, while simple applications capture well-known convective behavior

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

R. A. J. Neggers,
neggers@meteo.uni-koeln.de

**Citation:**

Neggers, R. A. J. & Griewank, P. J. (2021). A binomial stochastic framework for efficiently modeling discrete statistics of convective populations. *Journal of Advances in Modeling Earth Systems*, *13*, e2020MS002229. https://doi.org/10.1029/2020MS002229

Received 7 JUL 2020
Accepted 10 FEB 2021



**Abstract** Understanding the coupling between convective clouds and the general circulation, as well as addressing the gray 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 discrete behavior in convective properties at small population sample sizes, ii) object age-dependence 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. The applicability to atmospheric convection as well as behavior implied by the formulation is assessed. Various simple applications of the BiOMi model (Binomial Objects on Microgrids) are explored, suggesting that important convective behavior can be captured at low computational cost. This includes i) subsampling effects and associated powerlaw scaling in the convective gray zone, ii) stochastic predator-prey behavior, iii) the downscale 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 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 well-known 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 globe. 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; Zuidema et al., 2012).

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 "gray zone problem" (Honnert et al., 2020; Wyngaard, 2004) 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 (Brast et al., 2018; Craig & Cohen, 2006; Kwon & Hong, 2017; Plant & Craig, 2008; Rochetin et al., 2014; Sakradzija et al., 2016). 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 (Nair et al., 1998; Sengupta et al., 1990; Weger et al., 1992) and more recently also simulations (Feingold et al., 2017; Neggers et al., 2019; Tompkins & Semie, 2017). 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 toward more discrete behavior for a significantly subsampled population, or even binary behavior in the limit that only one object is intermittently present in the gridbox. The way this happens is strongly affected by clustering (Neggers et al., 2019). Understanding and capturing this transition toward *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 (Dorrestijn et al., 2013; Khouider et al., 2010; 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 Plant and Craig (2008), Sakradzija et al. (2016), and Hagos et al. (2018). One step further downscale is the Lagrangian particle approach of (Böing, 2016), which tracks a multitude of interacting *subcloud* 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 gray 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 multiobject 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 discrete 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. An object age dimension is introduced, allowing discrete and explicit representation of life-cycle effects. The formulation of the framework allows for multiple coexisting cloud 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.

While some of these features have already been part of previously proposed convective population models, one of the main novelties of this framework lies in their combination, and in the way this is achieved. This introduces new behavior that is not just conceptual but also practical, adding new functionality in an efficient and general way which could potentially open new pathways to break the still existing convective parameterization deadlock (Randall et al., 2003; Sherwood et al., 2014; Vial et al., 2016). The net behavior resulting from the combination of these individual population model components is at this point still unknown, and remains unexplored. The main goal of this study is to gain more insight into this new behavior, among others through example applications representing atmospheric convection. The thorough

evaluation and performance testing of this population framework as part of a full convection scheme, as required before implementation in any operational weather or climate model, is for now considered as a future research topic.

Section 2 presents the basic formulation of the framework. In Section 3, behavior as implied by the formulation and its applicability to atmospheric convection is 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 binomial framework on microgrids, defined here as two-dimensional grids situated inside a gridbox of a larger-scale circulation model. This application is named BiOMi (Binomial Objects on Microgrids). The experiments are designed to represent atmospheric turbulence or convection, and include both single-species and multispecies setups. 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 subsampling of populations in a too small gridbox. In this "gray-zone" range of resolutions, only a few objects are present at varying stages of their life-cycle, which may lead to discrete behavior in their averaged properties. Adopting a discrete approach has direct implications for the formulation of all terms in the number budget.

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 $\Delta t$, as depicted in Figure 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 age-dependence, with $k$ being an integer number indicating discretized age, here referred to as an "age-stratum". For simplicity all objects of a species $i$ are assumed to have the same life-span $\tau_i$, by which the number of age strata $K_i$ is obtained through

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

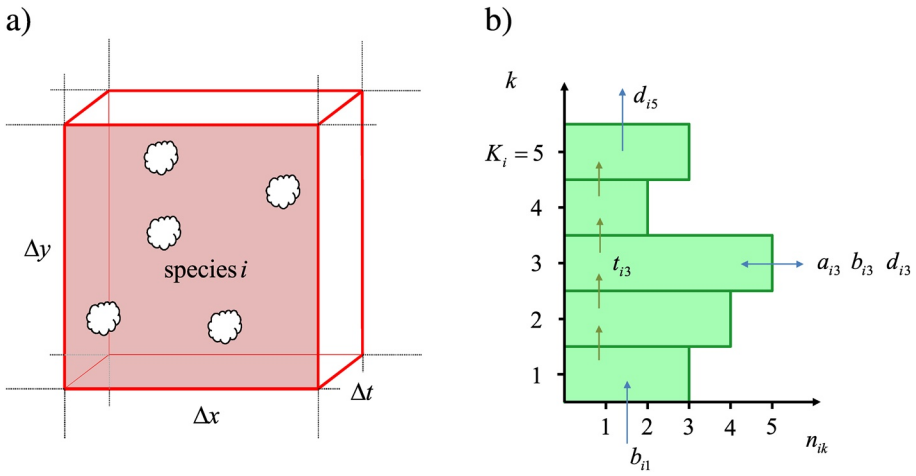


**Figure 1.** (a) Schematic illustration of a population of objects of species $i$ inside a three-dimensional space-time gridbox (red) with square horizontal area $\Delta x\Delta y$ and time-step $\Delta t$. (b) Schematic illustration of object age-dependence for a species $i$ with five age strata (defined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

In practice, the chosen time discretization determines how many age levels are maintained. The life times of objects are chosen to be a multiple of $\Delta t$, 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 $\Delta n_{ik}$ represents the change of $n_i$ at age level $k$ per time step $\Delta t$. 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. 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 age 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 theoretical 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 theoretical reference domain would be the whole globe, as this represents the theoretical upper limit of grid spacing 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 where shallow convection dominates. 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, $\mu_i$,

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

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 $B_i$ 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 grid spacing, in the birth probability of objects. 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 $\mu_i$ 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,

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

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 age array at the first (youngest) level. The birth rates $b_{ik}$ for $k > 1$ are set to zero for the moment.

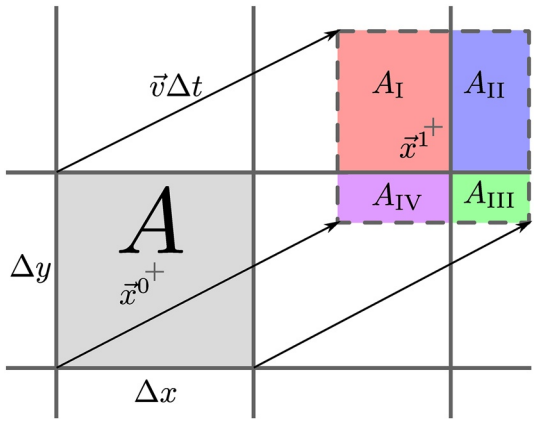


**Figure 2.** Schematic illustration of overlap between a displaced gridbox and the underlying grid. The arrows represent the displacement over one-time step, which is simply the horizontal wind multiplied by the time step duration. Gray crosses mark the mid of the gridbox before and after displacement.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

### 2.3. Object Life Cycle

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

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

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

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

Note that this death rate is automatic and discrete, in that it cannot create fractional object numbers. In this aspect it is different from Newtonian relaxation, which would be an alternative (but nondiscrete) formulation. Furthermore, 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 aging, 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 Figure 2. From this principle a conservative advection scheme can be derived that requires three 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 four 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

$$A = \sum_{j=1}^{\mathrm{IV}} A_j \tag{14}$$

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,

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

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

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

The number of objects advected into $A_{\mathrm{IV}}$ is then simply obtained as the residual,

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

While the clockwise ordering we used is arbitrary, it is critical that Equations 15–17 are calculated sequentially to guarantee that the remaining number of objects advected into the final subgrid box *IV* is not negative. Calculating advection separately for each age level $k$ ensures that age is conserved as objects are advected across the grid.

The introduced advection operator is a discrete and stochastic sibling to simple conservative semi-Lagranigan schemes that use a constant subgrid reconstruction. Accordingly, for large number of objects per gridbox, this discrete advection operator results in high gradient smoothing and fast dispersion. Less diffusive versions of the discrete advection operator can be derived using higher-order subgrid reconstructions (Staniforth & Côté, 1990), which abandon the assumption that objects are randomly distributed within a grid cell. For low object numbers, the stochastic nature of the operator 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 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 interspecies 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. Applicability to Atmospheric Convection

With the basic formulation of the framework concluded, some behavior can already be understood a priori its application in practice. These implied characteristics are what makes the frame work particularly suited to atmospheric convection. Stochasticity due to subsampling is the key to representing atmospheric convection consistently throughout its gray zone of resolutions, while advection is a fundamental aspect of atmospheric modeling which previous population frameworks have completely neglected. Furthermore, due to considerable computational burden of numerical weather prediction and climate simulation, any framework seeking to represent subgrid scale convection must be able to do so with as high efficiency as possible.

### 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

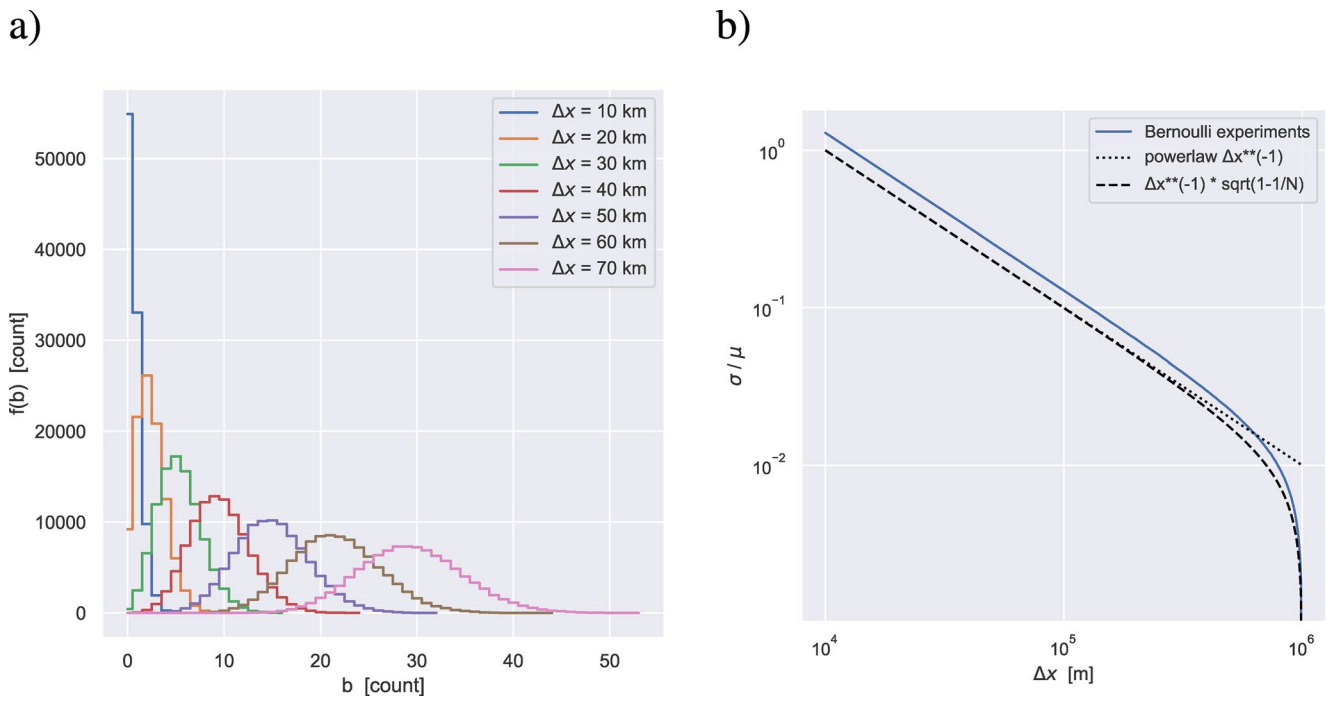


**Figure 3.** (a) Examples of binomial probability density $f$(b) as defined by 7 for various grid-spacings $\Delta x = \Delta y$, using a birth rate $\dot{B}_i = 10^{-10}$ m$^{-2}$ s$^{-1}$, a reference domain of size $L = 1{,}000$ km and an integration timestep $\Delta t = 60$ s. Results represent $10^6$ independent draws. (b) Associated functional form of the normalized standard deviation of the binomial distribution $\sigma/\mu$, as defined by 22. A pure powerlaw (black dotted) and modified powerlaw (black dashed) functional form are also shown, for reference.

in Figure 3, showing the binomial probability density function $f$(b) as defined by 7 for various gridspacings. Note that the shape of the distribution shifts from heavily skewed for small $\Delta x$ to more symmetric for large $\Delta x$. Such shapes are similar to previously proposed stochastic convection schemes (Cohen & Craig, 2006; Plant & Craig, 2008).

Both the mean $\mu_i$ and the width $2\sigma_i$ increase with gridspacing $\Delta x$, 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, $\sigma_i/\mu_i$. This can be understood by considering the definition of $\sigma_i$ for the binomial,

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

The standard deviation $\sigma_i$ normalized by the mean $\mu_i$ 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 $\mu_i$ 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 assumed to be proportional to the square root of the area surrounding each object that is free of other objects (on average). Substituting the first part of 21 for $\mu_i$ 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 $\Delta$ to the nearest neighbor spacing $l_i$, with larger values of $(\Delta/l_i)$ 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 Figure 3b, showing the functional dependence of the normalized standard deviation on gridbox size $\Delta$. In the range of gridspacings typical 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 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 (LES), 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 the stochasticity of such unorganized convective populations. Note that the framework also accounts for life cycle effects, which research has suggested is an essential and defining aspect of atmospheric convection as many of its properties depend on the stage of development (Chen & Houze, 1997; Genio et al., 2012). Also, the net episodic behavior in convective object number in an arbitrary domain that results from combining the binomial birthing formulation with life cycle effects is at the heart of gray zone problem (Honnert et al., 2020; Wyngaard, 2004).

### 3.2. Discrete Advection

The internal dynamics of a convective cloud field can be quite complex. While the population as a whole is embedded in the larger-scale flow, local smaller-scale perturbations in the flow exist that directly result from convective activity (Blyth et al., 1988; Heus & Jonker, 2008; Wing et al., 2018; Zuidema et al., 2012). When considering the life cycle of an individual convective object, one can state that in general these objects move horizontally because they are advected by the local flow directly surrounding them. Note that this principle does not contradict the propagation of multicell convective systems, because this reflects an entirely different process, namely the birth of new convective clouds in upstream areas that are lifted (Browning et al., 1976; Houze, 2014). The same applies to convective systems which are generated and maintained by permanent nonmoving orographic features, such as the infamous *Hector the Convector* phenomenon over the Tiwi Islands (Dauhut et al., 2016).

Taking this perspective, if the resolution of the model and the horizontal winds are high enough that the objects are expected to move between neighboring gridcells, advection should be taken into account. A defining novelty of the population framework presented here is that it includes an advection operator, which is also stochastic and discrete.

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 (Figure 4). Objects do not interact with each other or have a life cycle, and all differences between the subplots of Figure 4 are due to the

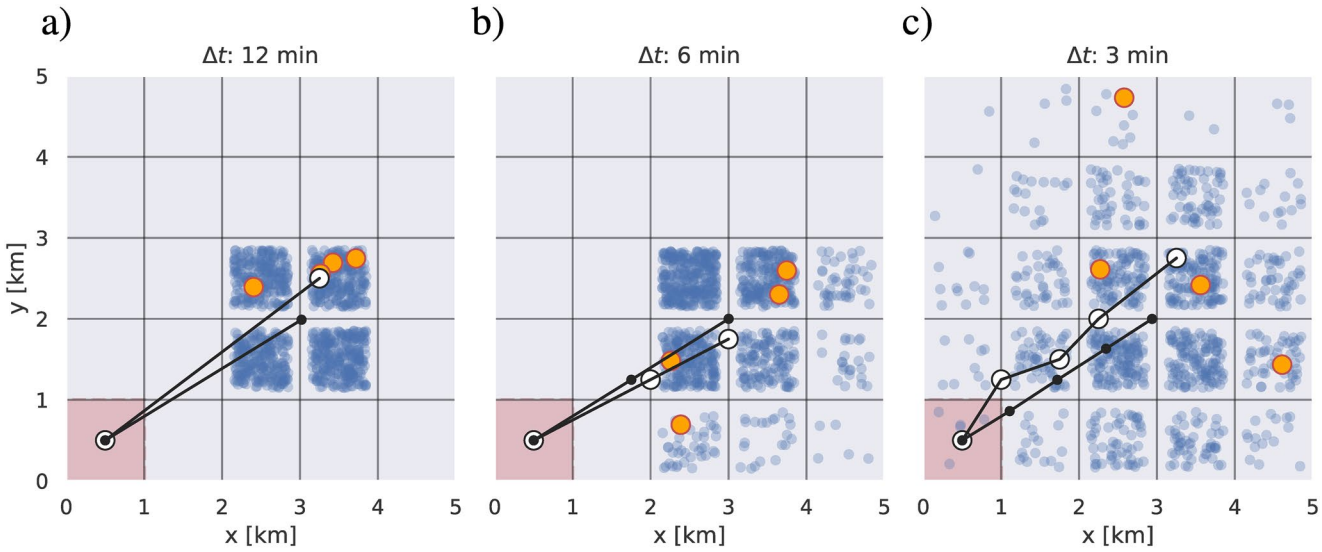


**Figure 4.** Example of discrete advection of objects on a 5 × 5 rectangular 1 km grid using the same initial conditions and grid but differing time step. The blue and orange objects behave identically, and differ only in the amount (1,000 blue, four orange). *Note*. that the individual objects have no specific x and y location within each gridbox, and are only plotted as such for visualization purposes. The red square marks the gridbox in which all objects were initialized at $t = 0$, and shown are the locations after 12 min 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.

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.

As expected, 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. 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 Figure 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. Keeping the computational burden of an object population model as low as possible is prerequisite for any implementation as part of a convection scheme in an operational weather or climate model. Generally speaking this numerical requirement has often acted as a hurdle for adopting more complex approaches in subgrid paramaterization in general circulation models, and was argued by Böing (2016) to apply to Lagrangian particle models in particular.

#### 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 operator 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

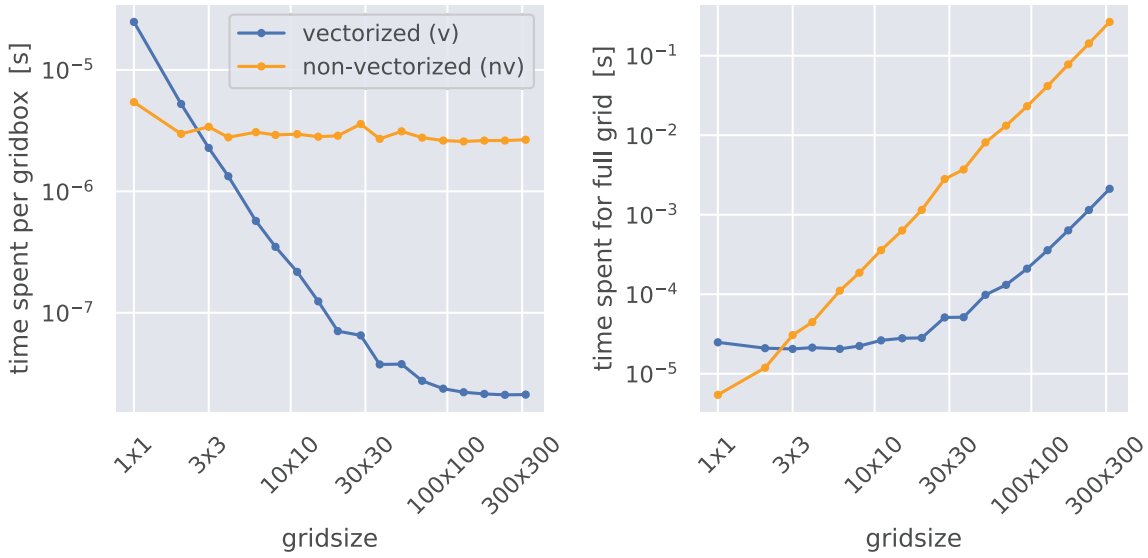


**Figure 5.** Results of a speed test of the binomial operator (10) as executed in Python on a single Intel i5-6400 2.7 Ghz CPU. (a) Time spent per gridbox as a function of gridsize, for a vectorized (*v*) and nonvectorized (*nv*) application. (b) Time spent for the full grid.

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 Figure 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 nonvectorized 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 $1 \times 1$ 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 $30 \times 30$, after which the vectorized version is 100× faster than the nonvectorized 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 microgrids 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 let's use the advection example shown in Figure 4. A Lagrangian approach would require the age, $x$, and $y$ location of each of the 1,004 objects to be tracked individually, resulting in the storage of 3,012 float values. Assuming an object lifetime of 24 min and a timestep of 12 min, as shown in the left subplot of Figure 4, the binomial memory footprint would be $25 \cdot 2 \cdot 2 = 100$ integer values (25 gridboxes, two species, two age strata). Reducing the timestep to 1 min while retaining a 24 min lifetime would increase the memory usage to 1,200 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 Convective Experiments

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"). All examples represent atmospheric turbulence or convection, which is also reflected in the definition of the species. Where possible a comparison is made to meaningful reference results, including LES results but also some well-known theoretical models. 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 is for now considered a future research topic.

The framework as applied on microgrids is hereby named *BiOMi* (Binomial Objects on Microgrids). Microgrids are here defined as two-dimensional grids that can be included inside gridboxes of larger-scale circulation models. Using small grids 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 gray zone stochastic noise can be superimposed onto this spectrum to represent the impact of subsampling of the population (Neggers, 2015; Plant & Craig, 2008), 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 object size with a slope of −2,

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

The proportionality constant $\lambda$ is scaled such that the birth rate is on average 256 per gridbox for the 50 m objects. A $1 \times 1$ grid is adopted with a grid spacing of 5 km, which is in the middle of the deep convective gray zone (Arakawa et al., 2011). The reference domain is 1,000 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 (Figure 6), with the ratio of subsampling variance to mean number becoming larger for the rarer object species. This behavior is a typical feature of cloud size distributions as sampled in LES (Neggers et al., 2019). Due to the implied scaling as discussed in Section 3.1, decreasing the domain size would increase the variability for a given object size, in a way that also resembles LES. This dependence of the stochasticity in object number on domain size is an expression of scale-awareness and scale-adaptivity.

Coupling the binomial framework to an operational spectral convection scheme would introduce these features in a weather or climate model. The simple "offline" experiment discussed here also highlights how stochasticity due to population subsampling would be captured in the convective gray 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); Lotka (1920) and 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 used to model various features in meteorology. Predator-prey approaches have been

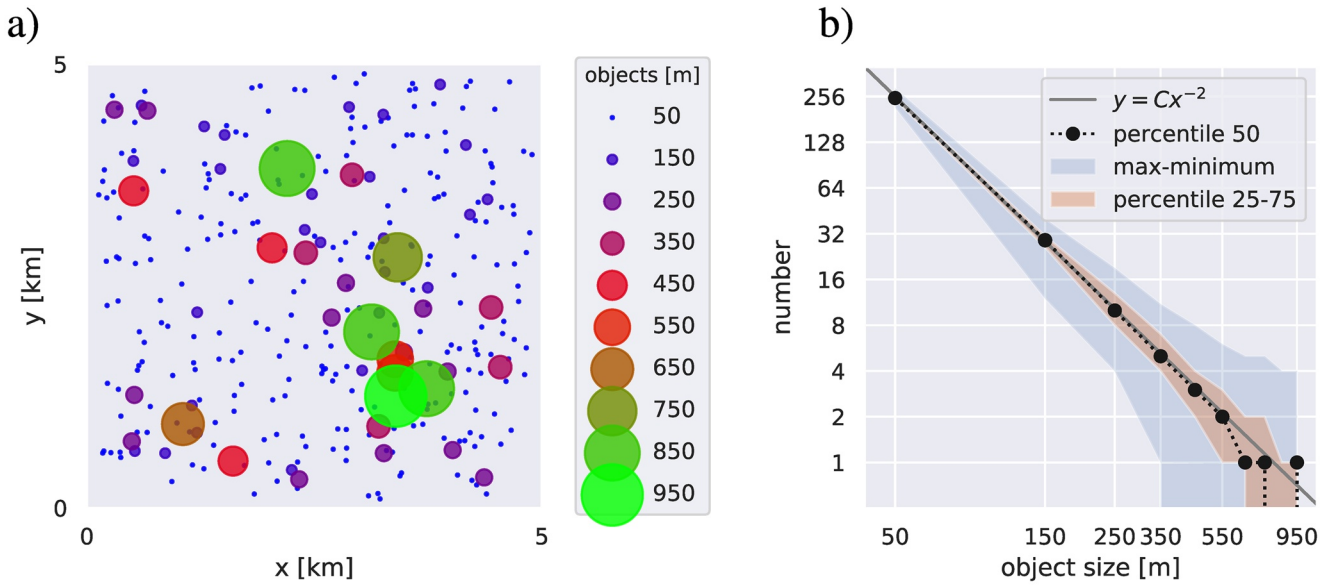


**Figure 6.** (a) Scatter plot illustrating all objects of one of the 50 samples included in subplot (b). The x and y position of each object is randomized for visualization. (b) Object size distribution statistics of 50 random samplings of objects as detailed in Section 4.1 with the parameters listed in Table 1. The larger the objects are, the lower their birth rate.

used in Meteorology to describe cloud microphysics (Wacker, 1995) and cloud precipitation interactions (Koren & Feingold, 2011; Pujol & Jensen, 2019). The Lotka-Volterra equations were used by Nober and Graf (2005) and Wagner and Graf (2010) to describe individual clouds competing for convective instability, in an analog to various animal species competing for a limited food source.

We will split this experiment into two halves. First we will look at the classic predator-prey equations in Section 4.2.1 which has been intensively studied. Both in regards to stochasticity (Aguirre et al., 2013), and how it can be translated to a system of stochastic cellular automata by Guinot (2002) who studied under which conditions the behavior of the cellular automata matches that of the continuous equations.

#### 4.2.1. Classic Predator-Prey

According to the classic formulation of the predator-prey equations, the prey $x$ grows exponentially with a rate of $\alpha$ but is reduced by the hunting of the predator $y$ which kills according to the product of prey and predator and $\beta$. The predator's growth is linked to the amount of hunting through $\delta$, and the predator dies off with an exponential decay of strength $\gamma$. 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 $\Delta t$ 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:

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

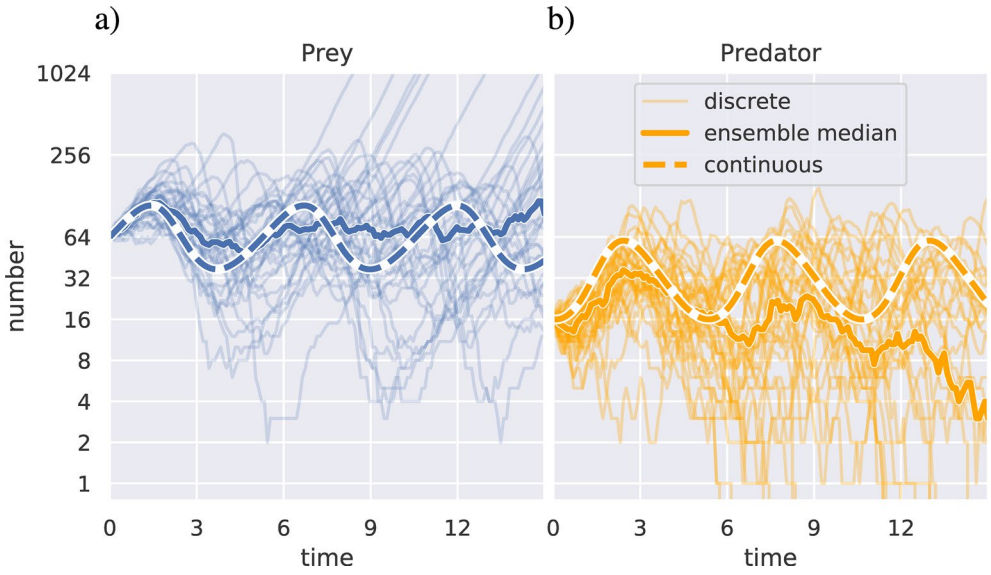


**Figure 7.** A 36 member ensemble of the predator prey system discussed in Subsection 4.2.1 using the parameters $\alpha = 1, \beta = 0.03, \gamma = 1.5, \delta = 0.75$ for Equation 24. Initial conditions are 64 (prey) and 16 (predator). Continuous solution is integrated numerically, discrete ensemble is generated using the values listed in Table 1.

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

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 Figure 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 Figure 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.

#### 4.2.2. Cloud Predator-Prey

This experiment is a modified version of the classic system to show in a very simple manner how the equations could be applied to a convection scheme. The prey now represents small clouds which feed on thermal instability, and predators are large clouds, representing the minority of small clouds which turn into larger and longer lasting clouds. This experiment is inspired by the idea of cloud types competing for convective instability used by Nober and Graf (2005) and Neggers (2015). We modify the classic equations by removing exponential growth, dropping $\delta$, and make $\alpha$, which represents the thermal instability the small clouds feed on, a function of time. This results in:

$$\frac{dx}{dt} = +\alpha(t) \quad -\beta xy \tag{29}$$

$$\frac{dy}{dt} = -\gamma y \quad +\beta x \tag{30}$$

We follow the same steps used in Section 4.2.1 to transform the continuous equations into a discrete and stochastic model, and use a sine function for $\alpha(t)$ in a crude imitation of the diurnal cycle of incoming solar radiation. An example of what this model can produce is shown in Figure 8, again using a 36 member ensemble with the same domain sizes as in Section 4.2.1. The small clouds form quickly after sunrise, leading to a peak in their number early in the day. The larger clouds are smaller in number and grow slower, but

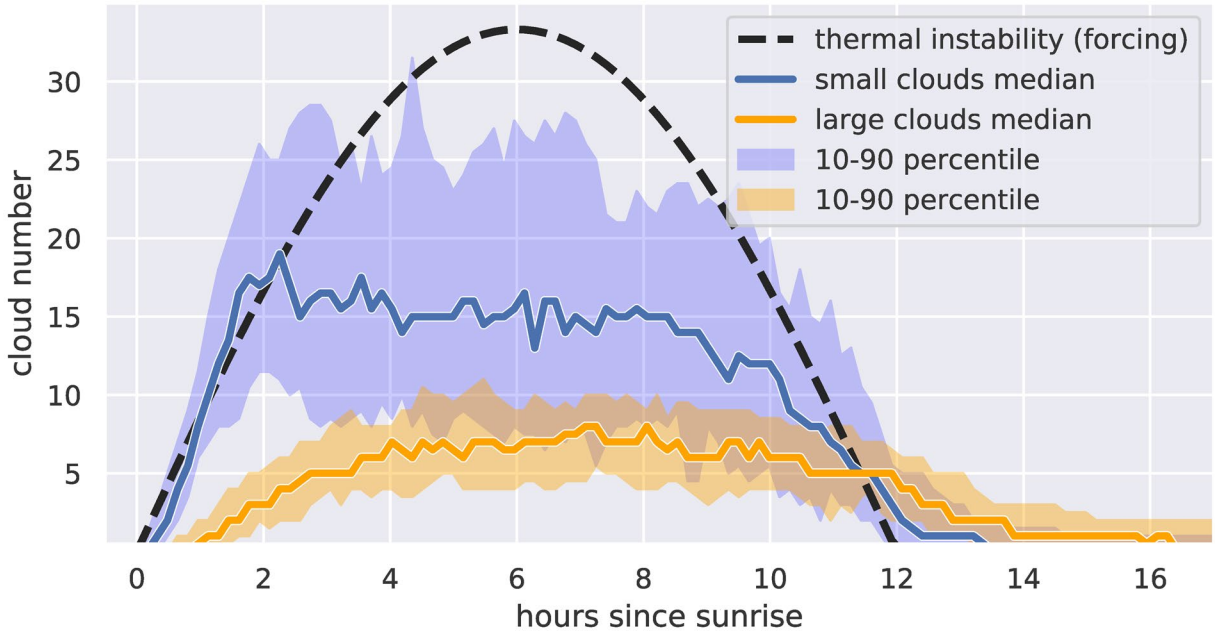


**Figure 8.** Median and 10–90 percentile spread of a 36 member ensemble of the discrete predator prey system described in Section 4.2.2. The dashed black line shows the prescribed cycle of thermal instability which determines the growth rate of the small clouds ($\alpha(t)$ in Equation 29). The large clouds in turn feed on the small clouds.

last somewhat longer than the smaller clouds which disappear almost immediately after sunset. A peak of small clouds early in the morning and a delayed shift to larger clouds in the afternoon is realistic behavior known from LES studies of shallow convective days (e.g., van Laar et al., 2019). The stochastic nature of the model is evident in the spread, with the discrete nature most obvious during the slow decay of the large clouds once the small clouds have disappeared.

While this simplified experiment yields interesting results it remains a proof of principle with many questions still unanswered. For example, which physical/dynamical processes do the prescribed interactions represent, would the method remain computationally efficient if many interacting species are employed, and how far horizontally should interactions extend? Some of these questions we will expand upon in Experiment 4.

### 4.3. Exp 3: A DownScale 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 downscale energy cascade typical of atmospheric turbulence (Frisch, 1995; Kolmogorov, 1941a, 1941b). 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,

$$b_{i1}^{\text{casc}} = 2\, d_{i+1,10}, \tag{31}$$

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 downscale 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 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 31. In addition, a weak mean wind is applied, so that the objects are slowly advected in the direction marked by the arrow in Figure 9. As a result of the advective diffusion illustrated in Section 3.2, the population starts to resemble a widening plume initiated at a fixed location and being advected downwind. 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 9a 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 9b 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 (Hernandez-Deckers & Sherwood, 2016; Morrison & Peters, 2018; Scorer & Ludlam, 1953). 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" (Bengtsson et al., 2011; Gardner, 1970).

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 and in LES, 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 preexisting instability, favoring subsequent thermals to thrive and thus form "thermal-chains" (Blyth & Latham, 1993; Damiani et al., 2006; 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:

$$p_i^{'} = C_f \; f_p \sum_k n_{ik} \tag{32}$$

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

$$f_p = \begin{cases} 1 - r \,/\, r_f & \text{for} \quad r < r_f \\ 0 & \text{for} \quad r \geq r_f \end{cases} \tag{33}$$

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.

**Table 1**
*Configuration of the Four BiOMi Experiments Discussed in Section 4*

| Setting | Unit | Exp 1 | Exp 2 | Exp 3 | Exp 4 |
|---|---|---|---|---|---|
| Gridsize | | $1 \times 1$ | $1 \times 1$ | $15 \times 15$ | $100 \times 100$ |
| | | | | | $1{,}000 \times 1{,}000$ |
| $\Delta x$, $\Delta y$ | [m] | 5,000 | 1 | 100 | 100 |
| $L$ | [m] | 1,000,000 | 5 | 1,000,000 | 1,000,000 |
| $\Delta t$ | [s] | 60 | 1/10 | 60 | 60 |
| $I$ | | 10 | 2 | 5 | 1 |
| $\tau_i$ | [s] | 60 | – | 600 | 600 |
| $K_i$ | | 1 | – | 10 | 5 |
| $\dot{B}_i$ | $[\mathrm{m^{-2}\,s^{-1}}]$ | $\propto (100 \cdot i - 50)^{-2}$ | $\dot{B}_1 = g(n_1, n_2)$ | $\dot{B}_5 = 5 \cdot 10^{-6}$ | $\dot{B}_1 = 2 \cdot 10^{-7}$ |
| | | | $\dot{B}_2 = f(n_1, n_2)$ | | |
| Interactions | | None | Interspecies | Interspecies | Spatial |
| $(u,v)_{\mathrm{adv}}$ | $[\mathrm{m\,s^{-1}}]$ | (0, 0) | (0, 0) | (0.3, 0.2) | (0, 0) |
| $r_f$ | [m] | – | – | – | 300 |
| $C_f$ | | – | – | – | 2,000 |

*Note.* Exp 2 is an exception in that it is nondimensional, age is neglected, and birthrates are derived from differential equations as explained in Subsection 4.2.

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,

$$p = \frac{1}{N} \frac{p_c}{\langle p_c \rangle}, \tag{34}$$

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 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 ($\sim$ 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× 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 10a shows a snapshot of the $100 \times 100$ gridsize experiment at 13 h 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 11 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

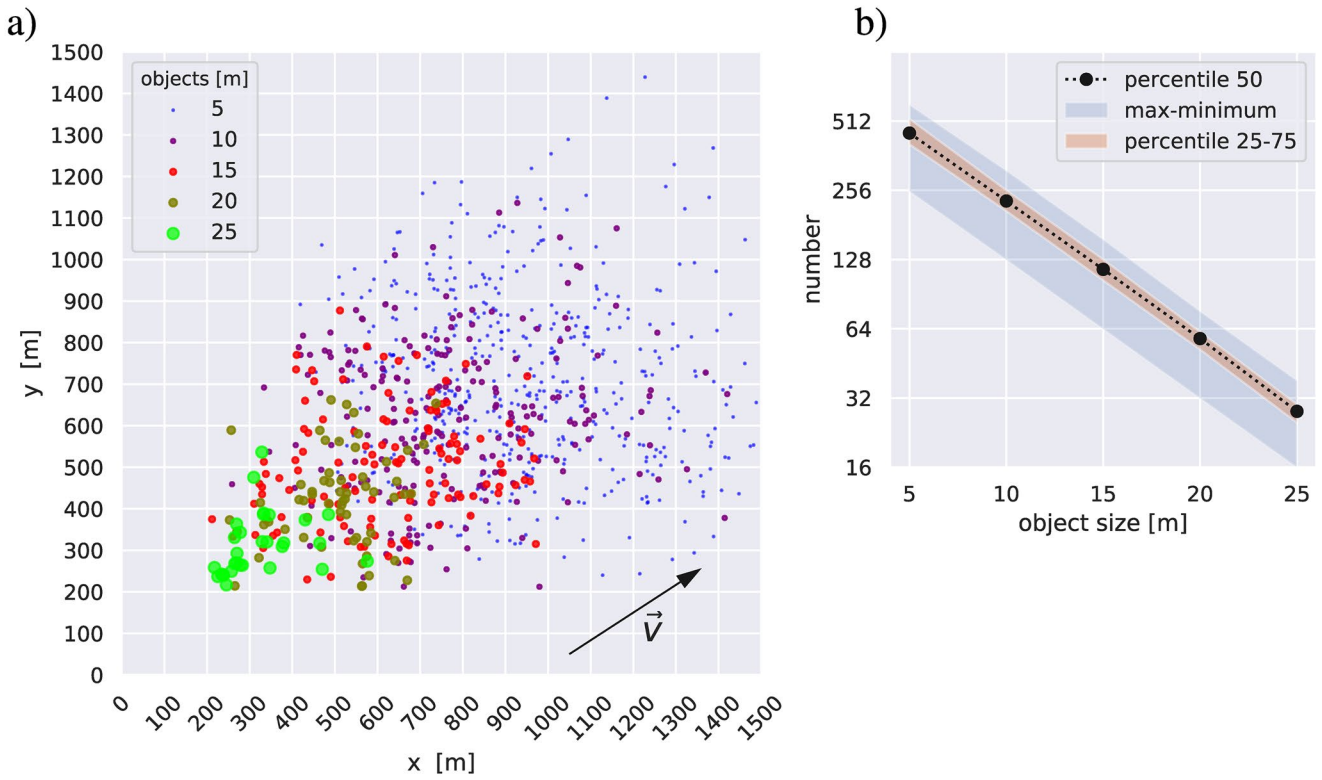


**Figure 9.** (a) Snapshot during an experiment with Binomial Objects on Microgrids (BiOMi) in the five-species energy-cascade configuration as described in Section 4.3 with an arrow showing the wind speed direction advecting the objects. The number of each species per gridbox is shown, with each species 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.

have two or more objects in them. Figure 11a shows the resulting clusters on the grid, while Figure 10b 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 gray zone (Neggers et al., 2019).

Another important aspect of the clustering behavior is highlighted by Figures 11b, showing convective memory on the grid as expressed by the time evolution of the size of the largest cluster, $l_{\max}$. Two grid-sizes 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 that 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).

At this point, it makes sense to qualitatively compare this behavior to LES results. Figure A1 shows the time evolution of the size of the maximum convective cluster in two experiments of the RICO shallow cumulus case. Clusters are defined as enclosed areas in two-dimensional fields of the vertically projected cloud core, the latter here defined as gridboxes with both cloud condensate and positive buoyancy (Siebesma & Cuijpers, 1996). This definition is applied to exclude stratiform cloudiness. The same behavior as produced by BiOMi is apparent, with the largest convective cluster growing gradually in size over a period of 10 h or so. This expresses the gradual emergence of larger clusters in the LES, and is a form of convective memory. Another feature that is similar is the largest cluster size being inhibited in the experiment with small domain

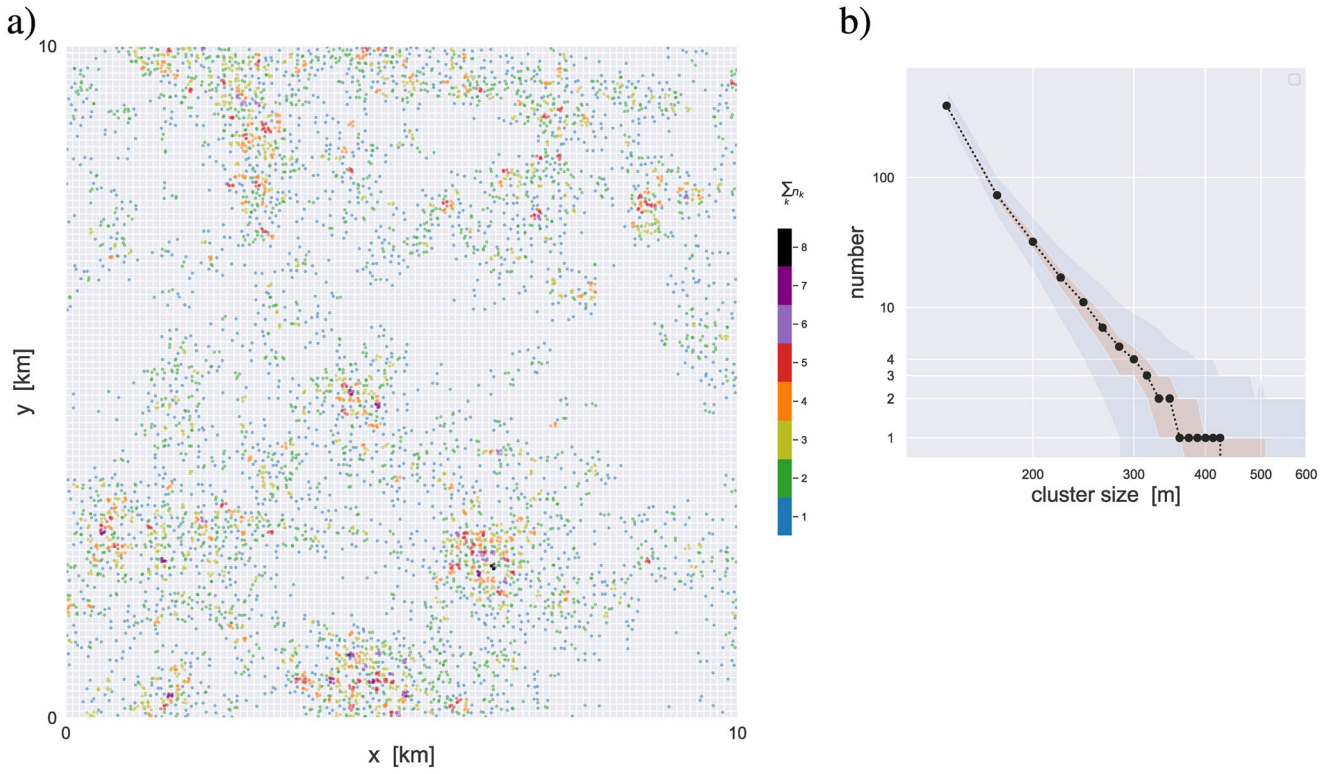


**Figure 10.** (a) Snapshot during an experiment with Binomial Objects on Microgrids (BiOMi) in the single-species configuration with two rules of interaction between objects, as described in Section 4.4. The position of each object within the gridbox is randomized, for visualization. The coloring indicates the number of objects in a grid cell, to highlight clustering. (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.

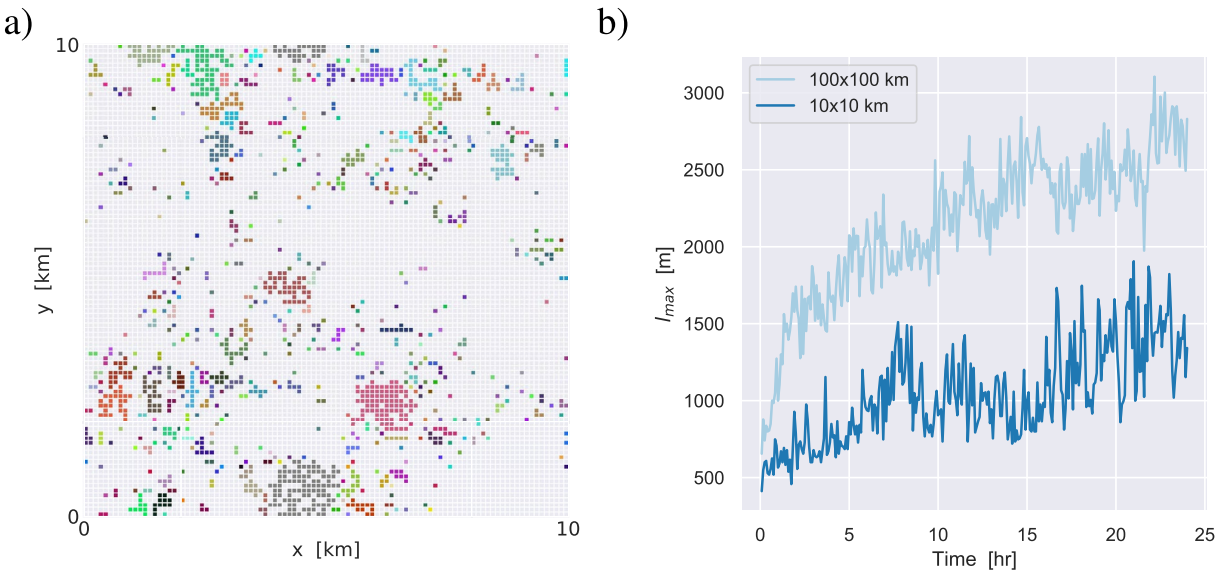


**Figure 11.** Results of cluster analysis using threshold $n_1 > 1$. (a) Spatial distribution of the clusters at the last timestep for the experiment with the 10 × 10 km domain (100 × 100 gridsize). Each cluster is assigned a unique color. (b) Time evolution of the size of the largest cluster on the grid. Results with two domain sizes are shown, 10 × 10 km (dark blue) and 100 × 100 km (light blue).

size. While these similarities are encouraging, differences also exist, such as the magnitude of $l_{max}$ being larger in the LES. This at least partially reflects that various processes are not captured by the two simple rules of interaction. These might include precipitation effects, cold pool dynamics and stratiform cloudiness that become apparent in the second half of the simulation.

Despite these differences, the agreement on basic behavior with LES does suggest that under simple rules of interaction, convective memory can be created and carried on the microgrid. Introducing this behavior in convective parameterizations is a long standing ambition that has not yet been achieved (Grabowski et al., 2006; Khairoutdinov & Randall, 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, possibly inviting or requiring additional layers to represent more complex three-dimensional flow structures. But before introducing such rules they should be carefully calibrated and trained against relevant datasets, ideally carrying information on spatial organization. For example, machine learning could be applied to train the framework against high-resolution satellite datasets of cloud populations. This is a future research topic.

### 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 as a prognostic variable, BiOMi considers the evolution of object number. These interacting objects can also freely move around on the grid, taking object life cycle 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, although reduced complexity can help in making such schemes more efficient (Machulskaya & Seifert, 2019).

## 6. Conclusions and Outlook

In this study, a computationally efficient stochastic binomial framework is formulated for representing discrete populations of convective 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. 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 population dynamics and cellular automata. Finally, due to its reliance on binomial sampling the BiOMi system is also computationally cheap to operate.

The combination of these features makes the BiOMi framework particularly applicable to atmospheric turbulence and convection. This is illustrated by testing the framework in various simple configurations. Characteristic convective behavior that is reproduced includes i) classic predator-prey behavior, preserving discreetness and introducing stochastic variations; ii) gray zone behavior as expressed by stochasticity in convective object number, iii) the downscale energy cascade and advective dispersion typical of atmospheric turbulence; and iv) spatial organization in convective clusters resulting from interactions between individual thermals. We find that the spatial arrangement of binomially generated population of clusters on a microgrid is a form of convective memory, evolving over timescales much longer than the lifespan of individual objects.

While the BiOMi 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. The high computational efficiency should facilitate this effort. 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 is coupled to a size-resolved 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 the spectral scheme with subgrid convective memory and introduces awareness of spatial organization.

BiOMi offers further opportunities when applied within GCM gridboxes. First, existing convection schemes can be equipped with the framework in one dimensional (1D) random sampler mode, as explored in Exp 1, to introduce stochastic noise in the gray zone. Second, 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.

## Appendix A: Clustering analysis in LES

Figure A1 shows results from a cluster analysis of simulated cloud populations for a prototype marine shallow cumulus case based on the RICO field campaign (Rauber et al., 2007). The simulations are generated with the Dutch Atmospheric Large-Eddy Simulation code (DALES) (Heus et al., 2010). The exact case configuration as described by vanZanten et al. (2011) is applied, while the double moment warm microphysics scheme of Seifert and Beheng (2001) is used. Two square domain sizes are simulated, with horizontal sizes $L = 12.8$ km and $L = 102.4$ km and a ceiling of 5 km. These domain sizes are chosen to approximately match the BiOMi Experiment 4 as described in Section 4.4. Clusters are defined as enclosed areas in the two-dimensional field of the vertically projected cloud core, defined as LES grid points which have cloud condensate and are also positively buoyant. Using this definition excludes stratiform cloudiness, and yields convective clusters that are conceptually comparable to the object-clusters emerging on the microgrid in Exp 4.

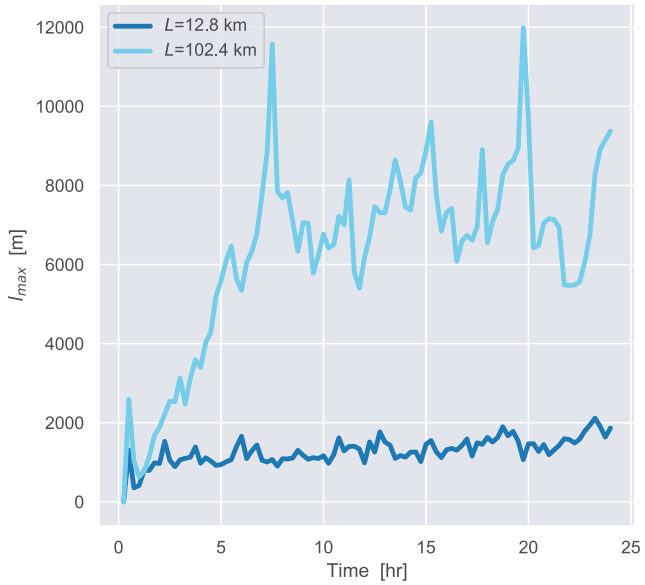


**Figure A1.** Large-Eddy Simulations (LES) results for the RICO shallow cumulus case. Shown is the time evolution of the size of the largest convective cluster in the domain. Clusters are calculated from the projected cloud core area, as explained in the text. Results for two domain sizes are shown, 12.8 × 12.8 km (dark blue) and 102.4 × 102.4 km (light blue).

## Data Availability Statement

The Gauss Center for Supercomputing e.V (www.gauss-centre.eu) is acknowledged for providing computing time on the GCS Supercomputer JUWELS at the Jülich Supercomputing Center (JSC) under project CHKU28 for the LES experiments used in this study. The BiOMi code used to prepare this manuscript is archived and freely accessible at https://doi.org/10.5281/zenodo.3932807. The BiOMi code repository is managed on Github at https://github.com/pgriewank/BioMi. Animations of Exp 3 and Exp 4 with the BiOMi framework as discussed in Section 4 are provided as supporting information to this publication.

**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. The authors thank two reviewers and the assistant editor for their constructive and knowledgeable comments that significantly improved this manuscript. The authors further 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 definitions.

## 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. 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. https://www.blackhistory.mit.edu/publications/study-pulsating-growth-cumulus-clouds-1960

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. https://doi.org/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. *Journal of the Atmospheric Sciences*, *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. https://doi.org/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: Atmosphere*, *103*(D22), 28753–28767. https://doi.org/10.1029/98JD02579

Blyth, A. M., Cooper, W. A., & Jensen, J. B. (1988). A study of the source of entrained air in Montana cumuli. *Journal of the Atmospheric Sciences*, *45*(24), 3944–3964. https://doi.org/10.1175/1520-0469(1988)045⟨3944:ASOTSO⟩2.0.CO;2

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. https://doi.org/10.1002/qj.49711950905

Böing, S. J. (2016). An object-based model for convective cold pool dynamics. *Mathematics of Climate and Weather Forecasting*. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1175/1520-0469(1989)046⟨0740:GWCSAD⟩2.0.CO;2

Browning, K., Frankhauser, J., Chalon, J.-P., Eccles, P., Strauch, R., Merrem, F., et al. (1976). Structure of an evolving hailstorm part V: Synthesis and implications for hail growth and hail suppression. *Monthly Weather Review*, *104*(5), 603–610. https://doi.org/10.1175/1520-0493(1976)104⟨0603:SOAEHP⟩2.0.CO;2

Chen, S. S., & Houze, R. A., Jr. (1997). Diurnal variation and life-cycle of deep convective systems over the tropical pacific warm pool. *Quarterly Journal of the Royal Meteorological Society*, *123*(538), 357–388. https://doi.org/10.1002/qj.49712353806

Cohen, B. G., & Craig, G. C. (2006). Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *Journal of the Atmospheric Sciences*, *63*(8), 2005–2015. https://doi.org/10.1175/JAS3710.1

Craig, G. C., & Cohen, B. G. (2006). Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *Journal of the Atmospheric Sciences*, *63*(8), 1996–2004. https://doi.org/10.1175/JAS3709.1

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. https://doi.org/10.1175/JAS3701.1

Dauhut, T., Chaboureau, J.-P., Escobar, J., & Mascart, P. (2016). Giga-les of hector the convector and its two tallest updrafts up to the stratosphere. *Journal of the Atmospheric Sciences*, *73*(12), 5041–5060. https://doi.org/10.1175/JAS-D-16-0083.1

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 & Engineering Sciences*, *371*(1991), 20120374. https://doi.org/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: Atmosphere*, *122*(13), 7086–7102. https://doi.org/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. https://doi.org/10.1016/S0169-8095(99)00024-1

Frisch, U. (1995). *Turbulence: The legacy of A. N. Kolmogorov*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139170666

Gardner, M. (1970). Mathematical games - The fantastic combinations of John Conway's new solitaire game 'life'. *Scientific American*, *223*, 120–123. https://doi.org/10.1038/scientificamerican1070-120

Garrett, T. J., Glenn, I. B., & Krueger, S. K. (2018). Thermodynamic constraints on the size distributions of tropical clouds. *Journal of Geophysical Research: Atmosphere*, *123*(16), 8832–8849. https://doi.org/10.1029/2018jd028803

Genio, A. D. D., Wu, J., & Chen, Y. (2012). Characteristics of mesoscale organization in WRF simulations of convection during TWP-ICE. *Journal of Climate*, *25*(17), 5666–5688. https://doi.org/10.1175/JCLI-D-11-00422.1

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

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

Hagos, S., Feng, Z., Plant, R. S., Houze, R. A., Jr., & 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. https://doi.org/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. https://doi.org/10.1175/JAS-D-15-0385.1

Heus, T., Heerwaarden, C. C., Jonker, H. J. J., Siebesma, A. P., Axelsen, S., van den Dries, K., et al. (2010). Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications. *Geoscientific Model Development*, *3*, 415–444. https://doi.org/10.5194/gmd-3-415-2010

Heus, T., & Jonker, H. J. J. (2008). Subsiding shells around shallow cumulus clouds. *Journal of the Atmospheric Sciences*, *65*(3), 1003–1018. https://doi.org/10.1175/2007JAS2322.1

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: Atmosphere*, *114*(D6). https://doi.org/10.1029/2008JD010917

Honnert, R., Efstathiou, G. A., Beare, R. J., Ito, J., Lock, A., Neggers, R., et al. (2020). The atmospheric boundary layer and the "gray zone" of turbulence: A critical review. *Journal of Geophysical Research: Atmosphere*. https://doi.org/10.1029/2019JD030317

Houze, R. (2014). *Cloud dynamics*. Amsterdam: Elsevier Science. https://books.google.de/books?id=GXEpAgAAQBAJ

Joseph, J. H., & Cahalan, R. F. (1990). Nearest neighbor spacing of fair weather cumulus clouds. *Journal of Applied Meteorology*, *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. https://doi.org/10.1175/JAS3810.1

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: Atmosphere*, *124*(1), 367–386. https://doi.org/10.1029/2018JD029518

Khouider, B., Biello, J., & Majda, A. J. (2010). A stochastic multicloud model for tropical convection. *Communications in Mathematical Sciences*, *8*(1), 187–216.

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

Kolmogorov, A. (1941b). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Doklady Akademii 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. https://doi.org/10.1073/pnas.1101777108

Kwon, Y. C., & Hong, S.-Y. (2017). A mass-flux cumulus parameterization scheme across gray-zone resolutions. *Monthly Weather Review*, *145*(2), 583–598. https://doi.org/10.1175/MWR-D-16-0034.1

Lotka, A. (1910). Contribution to the theory of periodic reaction. *Journal of Physical Chemistry*, *14*, 271–274.

Lotka, A. (1920). Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences of the United States of America*, *6*, 410–415.

Machulskaya, E., & Seifert, A. (2019). Stochastic differential equations for the variability of atmospheric convection fluctuating around the equilibrium. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2708–2727. https://doi.org/10.1029/2019MS001638

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. https://doi.org/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. *Journal of Geophysical Research: Atmosphere*, *103*(D10), 11363–11380. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1175/1520-0469(2003)60⟨1060:SSOCCP⟩2.0.CO;2

Nober, F. J., & Graf, H. F. (2005). A new convective cloud field model based on principles of self-organisation. *Atmospheric Chemistry and Physics*, *5*(10), 2749–2759. https://doi.org/10.5194/acp-5-2749-2005

Peters, K., Crueger, T., Jakob, C., & Möbis, B. (2017). Improved MJO-simulation in ECHAM 6.3 by coupling a stochastic multicloud model to the convection scheme. *Journal of Advances in Modeling Earth Systems*, *9*(1), 193–219. https://doi.org/10.1002/2016MS000809

Plant, R. S., & Craig, G. C. (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, *65*(1), 87–105. https://doi.org/10.1175/2007JAS2263.1

Pujol, O., & Jensen, A. (2019). Cloud-rain predator–prey interactions: Analyzing some properties of the Koren—Feingold model and introduction of a new species-competition bulk system with a Hopf bifurcation. *Physica D: Nonlinear Phenomena*, *399*, 86–94. https://doi.org/10.1016/j.physd.2019.04.007

Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547–1564. https://doi.org/10.1175/BAMS-84-11-1547

Rauber, R. M., Stevens, B., Ochs, H. T., III, Knight, C., Albrecht, B. A., Blyth, A. M., et al. (2007). Rain in shallow cumulus over the ocean: The Rico campaign. *Bulletin of the American Meteorological Society*, *88*(12), 1912–1928. https://doi.org/10.1175/BAMS-88-12-1912

Rochetin, N., Couvreux, F., Grandpeix, J.-Y., & Rio, C. (2014). Deep convection triggering by boundary layer thermals. Part I: LES analysis and stochastic triggering formulation. *Journal of the Atmospheric Sciences*, *71*(2), 496–514. https://doi.org/10.1175/JAS-D-12-0336.1

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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1002/qj.49707933908

Seifert, A., & Beheng, K. D. (2001). A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmospheric Research*, *59–60*, 265–281.(13th International Conference on Clouds and Precipitation). https://doi.org/10.1016/S0169-8095(01)00126-0

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. https://doi.org/10.1175/1520-0450(1990)029⟨1245:CCFMAS⟩2.0.CO;2

Sherwood, S., Bony, S., & Dufresne, J. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, *505*, 37–42. https://doi.org/10.1038/nature12829

Siebesma, A. P., & Cuijpers, J. W. M. (1996). Evaluation of parametric assumptions for shallow cumulus convection. *Journal of the Atmospheric Sciences*, *52*(6), 650–666. https://doi.org/10.1175/1520-0469(1995)052<0650:EOPAFS>2.0.CO;2

Staniforth, A., & Côté, J. (1990). Semi-Lagrangian integration schemes for atmospheric models—A review. *Monthly Weather Review*, *119*(9), 2206–2223. https://doi.org/10.1175/1520-0493(1991)119<2206:SLISFA>2.0.CO;2

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

Sutton, O. G. (1932). A theory of eddy diffusion in the atmosphere. *Proceedings of the Royal Society of 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. https://doi.org/10.1002/2016MS000802

van Laar, T. W., Schemann, V., & Neggers, R. A. J. (2019). Investigating the diurnal evolution of the cloud size distribution of continental cumulus convection using multiday LES. *Journal of the Atmospheric Sciences*, *76*(3), 729–747. https://doi.org/10.1175/jas-d-18-0084.1

vanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Burnet, F., et al. (2011). Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *Journal of Advances in Modeling Earth Systems*, *3*(2). https://doi.org/10.1029/2011MS000056

Varble, A., Zipser, E. J., Fridlind, A. M., Zhu, P., Ackerman, A. S., Chaboureau, J.-P., et al. (2014). Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations: 1. Deep convective updraft properties. *Journal of Geophysical Research: Atmosphere*, *119*(24), 13891–13918. https://doi.org/10.1002/2013JD021371

Vial, J., Bony, S., Dufresne, J.-L., & Roehrig, R. (2016). Coupling between lower-tropospheric convective mixing and low-level clouds: Physical mechanisms and dependence on convection scheme. *Journal of Advances in Modeling Earth Systems*, *8*(4), 1892–1911. https://doi.org/10.1002/2016MS000740

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. https://doi.org/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. *Mathematische Annalen*, *100*, 295–320. https://doi.org/10.1007/BF01448847

von Neumann, J. (1966). In A. W. Burks (Ed.), *Theory of self-reproducing automata*. Chicago, IL: University of Illinois Press.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: 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. https://doi.org/10.1175/1520-0469(1995)052⟨2577:COPPIA⟩2.0.CO;2

Wagner, T. M., & Graf, H.-F. (2010). An ensemble cumulus convection parameterization with explicit cloud treatment. *Journal of the Atmospheric Sciences*, *67*(12), 3854–3869. https://doi.org/10.1175/2010JAS3485.1

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*, *97*(D18), 20519–20536. https://doi.org/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. https://doi.org/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. https://doi.org/10.1175/2011JCLI4056.1

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

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

Zuidema, P., Li, Z., Hill, R. J., Bariteau, L., Rilling, B., Fairall, C., et al. (2012). On trade wind cumulus cold pools. *Journal of the Atmospheric Sciences*, *69*(1), 258–280. https://doi.org/10.1175/JAS-D-11-0143.1